IMPROVING THE PLANNNG PROCESS BY DEVELOPING A FORECASTING MODEL

Predicting future demand based on historical information

by

Lotte Verkerk

A Bachelor's Thesis

Submitted for the degree of Industrial Engineering and Management

University of Twente

August 2022

Supervised by Dr. M.C. van der Heijden (UT), Dr. Ir. L.L.M. van der Wegen (UT), S.F. van den Bos (Mainfreight), and T.J. Bijl (Mainfreight).

UNIVERSITY OF TWENTE.

MAINFREIGHT

Improving the planning process by developing a forecasting model

Predicting future demand based on historical information

Date
31 August 2022

Author
L. Verkerk
2193205
Bachelor Industrial Engineering and Management
University of Twente

Supervisors

Dr. M.C. van der Heijden	(UT)
Dr. Ir. L.L.M. van der Wegen	(UT)
S.F. van den Bos	(Mainfreight)
T.J. Bijl	(Mainfreight)

University of Twente
Faculty of Behavioural, Management and Social sciences
PO Box 217, 7500 AE Enschede
The Netherlands
+31(0)534899111
www.utwente.nl

Mainfreight Forwarding Netherlands
Industriestraat 10, 7041 GD, 's-Heerenberg
The Netherlands
www.mainfreight.com/the-netherlands/en-nz

This report was written as the graduation assignment in module 12 of the bachelor Industrial Engineering and Management educational program.

Preface

Dear reader,

You are about to read my thesis 'Improving the planning process by developing a forecasting model'. This research is executed in collaboration with Mainfreight's 's-Heerenberg Forwarding as a graduation assignment of the bachelor Industrial Engineering and Management at the University of Twente. The thesis addresses the development of a forecasting model in order to improve the planning process and the distribution of the collective transport capacity of Mainfreight's 's-Heerenberg Forwarding.

There are several people I would like to thank for their involvement in the realization of my research. First, I would like to thank my supervisors Matthieu van der Heijden and Leo van der Wegen who guided me during the research. Their advice and feedback enabled me to take the research to a higher academic level. I am also grateful to my internship supervisors, Sjoerd van den Bos and Thomas Bijl, who made it possible to conduct my graduation assignment at Mainfreight. Thank you for giving me the opportunity to come up with a research topic together and for the flexibility and unwavering support during the process. I would also like to thank my colleagues and managers at Mainfreight for their help answering my questions and for the fun and positive environment at the workplace. Finally, I would like to thank my family and friends for their support during this journey.

I hope you enjoy reading my thesis.

Lotte Verkerk

Enschede, Aug 31, 2022

Management summary

Introduction

This research is conducted on behalf of the domestic planning department at the branch Transport s'-Heerenberg of Mainfreight. Mainfreight is a global logistic provider which offers various supply chain services such as air & ocean freight, warehousing, and transport services. This research addresses the lack of reliable predictive information of the domestic planning of the Branch Transport 's-Heerenberg. At the moment, hardly any use is made of predictive information within the company. As a result, predictive information could contribute to the solution of several problems, such as vehicle capacity allocation and the routing process. However, not all those problems benefit from the same type of predictive information. Based on the collective interest, this research focuses on predictive information that contributes to a better distribution of the shared capacity of vehicles between the domestic and international planning departments. This focus leads to the following research goal: *develop a forecast model that forecasts the demand, expressed in loading meters for the domestic planning team with a time interval of one day and lead times from one day to two weeks ahead,* meaning a forecast is produced for one day ahead up to and including two weeks ahead. The forecast is used to predict the required transport capacity of the domestic planning.

Methods

Historical demand data is obtained from the transport management system of Mainfreight which collects historical shipment information. The obtained dataset needed to be thoroughly cleaned and transformed before it could be used. After cleaning and transforming the data to the required format, a data analysis is conducted to obtain insight in the factors that have an influence on the demand forecast of the domestic planning. The data analysis showed that the within-week seasonality dominates the demand, but the within-year seasonality, based on week numbers, also provides additional information about the demand. Furthermore, national holidays also significantly affect the demand on the day itself and surrounding days. Considering these characteristics of the demand, two forecasting models are proposed, called exponential smoothing and seasonal ARIMA, to forecast the demand of the domestic planning. Demand on days affected by national holidays require an alternative approach, called rule-based forecasting. Rule-based forecasting allows us to forecast days affected by the national holidays separately, according to a 'special day' rule, from days that are not affected by national holidays. In order to arrive at the required transport capacity of the domestic planning, the forecast is converted into an optimal required transport capacity based on the trade-off between the cost of understocking and overstocking vehicles.

Results

Domestic planning lacks a forecast, but there is some documentation on which we can reconstruct the current forecasting approach. The reconstructed current forecasting model is used as a benchmark for the developed model. The results show that exponential smoothing performs best if the effect of special days is not treated separately. When comparing the performance of this model with the reconstructed current forecasting model, we find that this model has increased performance by 23% in terms of mean absolute deviation. An interesting result is the stable performance of the forecasting model across the lead time, meaning that the forecast for one day ahead is approximately just as accurate as two weeks ahead.

By zooming in on the forecasting performance of the exponential smoothing on the national holidays and surrounding days, we proved that the model is not able to produce a reasonable forecast for days that are affected by national holidays. Applying the 'special day' rule of the rule-based forecasting model on just days affected by national holidays approximately doubles the forecasting performance, in terms of mean absolute deviation, compared to the previously mentioned exponential smoothing model.

The results show that applying rule-based forecasting, which allows us to forecast days not affected by national holidays according to the Exponential smoothing model and to forecast affected days according

to the 'special day' rule, leads to the highest forecasting performance. This model is considered the final forecasting model. It still has the stable forecasting performance across the one to fourteen steps forecast. By comparing the final model with the reconstructed current forecasting model, we proved that a professional forecast model decreases the mean absolute deviation from 235.87 to 130.76 loading meter, which represents an 44.2% increase in forecasting performance. However, there are still opportunities for higher forecasting performance, especially in the area of the effect of national holidays.

Conclusions and recommendations
We provide the following conclusions and recommendations to Mainfreight and, in particular, to domestic planning:

1. We recommend domestic planning to use the developed forecasting model over the current forecasting model because we conclude that the developed forecasting model resolves the underestimation problem of the current model and increases the forecasting accuracy by 44.2%.
2. We conclude that the developed model has a constant forecasting accuracy, meaning that the forecast for one day ahead is approximately just as accurate as fourteen days ahead. It is more valuable to Mainfreight to know the demand two weeks ahead with a certain accuracy than one day ahead with the same accuracy. Therefore, the forecasts that predicts more than a week ahead have a relatively higher added value than those for less than a week ahead.
3. We recommend the domestic planning to use the forecast model with its corresponding capacity advice as a starting point for the planning if no actual demand information is known. However, if a part of the demand is known, we recommend that the domestic planning revisit the advised capacity based on the latest actual demand information. The added value of the model lies in a better and timely estimation of the required vehicle capacity so that domestic planning does not have to upscale and downscale much at the last minute.
4. The forecast enables domestic planning to give an indication of the expected demand which can be used to arrive at a better distribution of the collective vehicle capacity between domestic and international planning. Therefore, we also recommend the domestic planning to use the advised vehicle capacity obtained from the forecasting model to communicate with Trucks and Drivers and international planning.
5. To increase the forecasting performance, we advise further research on the influence of anomalous demand caused by national holidays because the research showed that the forecast performance around national holidays is lower than average.
6. Finally, the significantly contaminated data is a major obstacle to easily implement the forecast. We recommend domestic planning to keep track of the observed demand by constructing a database from which the demand, expressed in the load meter, can be derived directly. For validity, it is also essential to use consistent conversion rules to indicate shipment sizes, such as gross weight, cubage, and loading meter, in the various databases. Furthermore, we advise the branch to set up KPIs regarding transport capacity utilization rate to keep track of the efficiency of the collective vehicle capacity.

Table of Contents

List of figures

List of tables

Abbreviations

The following abbreviations are frequently used in this research:

LDM	Loading meters
MSE	Mean squared error
RMSE	Root mean squared error
MAD	Mean absolute value
(S)ARIMA	(Seasonal) Autoregressive Integrated Moving Average
CSL	Optimal cycle service level
O*	Optimal ordered transport capacity

1. Introduction

In this chapter, we provide an introduction to our research. Section 1.1 gives a concise description of Mainfreight and the importance of the planning process. Section 1.2 addresses the problem statement by introducing the stakeholders in the planning process of the Transport Branch in 's-Heerenberg and the main problems encountered. From these problems, we chose the problem that will be solved in this research. Section 1.3 gives a brief introduction to the current situation. Section 1.4 describes the research goal, and Section 1.5 provides the research question, which answers are needed to accomplish the research goal.

1.1 Company description

Mainfreight is a global logistic provider which offers various supply chain services such as air & ocean freight, warehousing, and transport services. With 297 branches in 26 countries, Mainfreight possesses a powerful global reach. Mainfreight's largest cross-docks and warehouses in Europe are located in 's-Heerenberg. The location is convenient because it lies near the Rhine and on the German border. Sea containers entering the port of Rotterdam can be transported by inland vessel to the warehouses in 's-Heerenberg. This easy access, in combination with the large warehouses and cross docks located there, has ensured that these branches supply a large part of the Dutch and European markets. Consequently, most road transport activities from the Netherlands to Europe and from Europe to the Netherlands are planned here. This research is conducted on behalf of the Branch Transport 's-Heerenberg and is aimed at the planning department.

1.2 Problem statement

This section addresses the problem statement of this research by introducing the stakeholders in the planning process of the Transport Branch in 's-Heerenberg and the main problems encountered in the planning process. From the problem context, we chose the problem that will be solved in this research.

The branch in 's-Heerenberg accounts for a large of the supply to the European network. The team responsible for this road supply is divided into domestic and international transport. International planning is responsible for all road transport from 's-Heerenberg to Europe and from Europe to 's-Heerenberg, whereas domestic planning is responsible for road transport from 's-Heerenberg to the Netherlands and from the Netherlands to 's-Heerenberg. For example, a small shipment to be transported from Enschede to Madrid is picked up by domestic planning in Enschede and taken to 's-Heerenberg, after which the international planning takes over and ensures that it is transported to Madrid.

The domestic and international share a collective capacity of vehicles managed by another team called Truck and Drivers. The international and domestic planning teams consult individually with Trucks and Drivers about the needed vehicles on a daily basis. This system of dividing vehicles presents problems due to the little communication about the distribution of vehicles between the two teams with a collective capacity. Furthermore, as a rule, international planning takes precedence over domestic planning. A problem caused by this system of communication and precedence is the varying capacity of vehicles available for domestic planning. The variety of available capacity can cause idle vehicles or less efficient use of the vehicles and, therefore, suboptimal use of the shared capacity of vehicles. Due to the varying capacity of domestic planning in contrast to international planning and on the advice of the branch director of Transport s'-Heerenberg, we decided to focus the research on the domestic planning team. All problems mentioned further in this report are therefore related to domestic planning.

At the moment, domestic planning cannot change this way of allocating capacity because they do not possess reliable predictions regarding the expected demand to make capacity allocation requirements. This can be explained by looking closer at the order placement process of the customers because if a customer places an order before 5 p.m., then the transport is arranged for the next day. This policy ensures that the domestic planning has not yet received a part of the orders one day in advance when the last changes to the capacity allocation are made. If there is a reliable prediction regarding the expected demand, then a pull relationship can arise between domestic planning and truck and drivers instead of the current push relationship. Moreover, if international planning also has reliable predictive information, the collective capacity can be distributed more efficiently among the two parties.

The late arrival of part of the orders also causes inconvenience for the planning process because the planning teams already start planning for the next day at noon even though not all orders have been received yet. This early start of planning is necessary because the driver's departure times depend on the planned routes and the drivers need to receive that information at the latest in the evening since they need to know how late they must be at Mainfreight the next day. Starting with planning without knowing all the orders that must be arranged results in poor routes and, therefore, suboptimal use of the own vehicle capacity. As a solution for the initial poor routes, the planning team updates the planning during the day accordingly to the newest information. However, this continuous process of updating the routes according to the newest information is time-consuming.

The domestic planning team does not have a professional and structural demand forecasting model, and the demand fluctuates. The lack of a professional forecast and the highly fluctuating demand cause ignorance of the expected demand. Planning without a reliable forecast of the missing information also contributes to the poor quality of the routes and, therefore, the time-consuming planning process. In addition, a part of the shipments is sold to third parties, such as couriers and charters. Couriers are partner transport companies with whom Mainfreight does business because of their strong network in regions where Mainfreight's network is somewhat smaller. Mainfreight has fixed agreements with couriers about the shipment area and the bandwidth for the number of shipments they can perform for Mainfreight. Charters are hired for specific individual shipments, mostly FLT or LTL. However, unlike couriers, there are no fixed agreements with charters, and charters must buy individual shipments. The ignorance of the expected demand ensures an unreliable number of shipments that must be sold to charters, and the later the shipments are offered for sale, the smaller the chance that they will be bought. The poor quality of the routes and the inaccurate volume sold to charters also cause suboptimal use of the Mainfreight capacity, in other words, suboptimal use of the own vehicle capacity.

The two core problems are the late availability of a part of the shipping information and the absence of a professional demand forecasting model. The late availability of a part of the shipping information is a problem that cannot be solved entirely because it is a strategic objective needed to compete in the competitive transport business. There are, however, opportunities for improvement, such as improving communication between the warehouses and the transport teams. The problem that will be solved in this research, which is also commissioned by Mainfreight, is the absence of a professional demand forecasting model. As discussed, a professional forecast model can contribute to better communication between the planning team and ensure that domestic planning can also express substantiated expectations for the number of necessary capacity vehicles. In addition, a better distribution of vehicles can increase the capacity utilization efficiency of domestic planning vehicles. Furthermore, busy periods can be considered early by arranging extra capacity with couriers or charters.

1.3 Current situation

The domestic planning team focuses primarily on the demand that needs to be fulfilled tomorrow, meaning the shipments that must be transported tomorrow. Due to this one-day forward-looking approach, the team primarily responds to the demand instead of anticipating it. They are overwhelmed when there is very little or very little demand. In a day's slack, there are only limited options to increase capacity in the event of too much demand and to decrease the capacity in the event of too little demand. For example, international vehicles are often away for a few days, so receiving more vehicles one day in advance from the international planning team is not always possible. However, suppose the domestic planning team has an indication of the expected demand a week in advance. In that case, the international planning team can take this into account by chartering fewer or more shipments. In addition, responding to demand creates a higher workload among planners because scaling down and scaling up capacity often has to be done at the last minute. To look further ahead, reliable predictive information is needed, but the problem is a lack of reliable predictive information. So, without reliable predictive information that can be obtained from a professional forecast, anticipating demand is impossible.

Until now, we have said that there is no professional forecasting model, but there is an excel file that can be reconstructed into a forecast model. The limitations of the current approach are explained in Section 2.2. However, the main problem is that the forecasting performance of the current method is not established, which makes domestic planning unknown to what extent their current approach is an accurate method of forecasting demand. The ignorance of the accuracy makes domestic planning cautious about using the forecast, which is logical since a forecast without the accuracy information is trivial and unreliable. Therefore, a desirable result of this research is to determine the accuracy of the current forecasting method.

1.4 Statement of research goal

As indicated in Section 1.3, this research is aimed to solve the problem of the lack of reliable predictive information due to the lack of a professional forecast with the associated accuracy information. We have identified several factors that benefit from reliable predictive information, but not all factors benefit from predictive information at the same aggregate level, time bucket, and forecast horizon. Therefore, in the remainder of the Section, we will first look at which forecast model has the most value for Mainfreight. Then we will explain the needed aggregation level, time bucket, and lead of that forecast model.

For operational assistance in the planning of domestic planning, a very disaggregated forecasting model is required with a time bucket of one day and lead times of one day to one week ahead. However, the domestic planning team is only one part of Mainfreight. Therefore, it is good to also look at the forecast model with a zoomed-out perspective, meaning that it would be better if a forecast model could also contribute to improvements in other departments instead of being only beneficial for the domestic planning team. After all, Mainfreight's capacity of vehicles is shared between domestic and international and managed by Trucks & Drivers. If we look at the collective interest, then a forecast model that contributes to a more optimal distribution of the shared capacity of vehicles has more value. Therefore, we decided to focus on a forecast that can estimate how much capacity domestic planning needs on a daily basis.

In order to forecast the daily needed capacity of domestic planning, a forecast bucket of one day is needed. Appropriate lead times are forecast for one day ahead up to 2 weeks. In other words, we forecast the demand for tomorrow up to two weeks ahead. As will be explained in Section 2.1, the demand needs to be disaggregated into three transportation flows because the relationships between these flows determine the necessary transport capacity of domestic planning. Finally, we need to determine the forecast unit. The demand of the domestic planning team can be expressed in different units, such as volume weight, cubage, and loading meters. The domestic planning team plans their routes in loading meters, the space that a shipment takes up in a vehicle. Therefore, the demand in the forecast model will also be expressed in loading meters.

Thus, the goal of this research is to develop a forecast model that forecasts the demand, expressed in loading meters for the domestic planning team at the Branch Transport s'-Heerenberg with a time bucket of a day and lead times from one day to twee weeks ahead for each transportation flow in order to predict the required transport capacity of the domestic planning.

1.5 Research questions

At this stage, Mainfreight is introduced, and the problem and research goal are identified. In order to achieve the research goal, research questions must be defined. Each research question provides an interim result needed input for a later question. This research considers the following research question in order to reach the research goal.

1. What is the forecasting performance of the reconstructed current forecasting method of the domestic planning team at the Branch Transport s'-Heerenberg?

The answer to this research question provides insight into the forecasting performance of the current forecasting method. One of the major limitations of the current method is that domestic planning does not know the accuracy of this current method. A desirable result of this research is, therefore, to determine the accuracy of the current forecasting method. In addition, the performance of the current forecasting model is used as a benchmark for the developed forecasting method. The answer to the research question is provided in Chapter 2.

2. Which factors significantly influence the demand forecast of the domestic planning team at the Branch Transport s'-Heerenberg?

The answer to this research question provides insight into the factors that influence the demand forecast of domestic planning. We have to analyse the demand characteristics because they are important for the choice of the forecasting model. In the analysis, we look for patterns and remarkable events in the demand data of the domestic planning team, which can help us to develop an accurate forecasting model. Before this analysis, we have to investigate what type of data is available and for which time there is reliable data. The answer to the research question is provided in Chapter 3.

3. Which forecasting models are most appropriate to forecast the demand, expressed in loading meters, of the domestic planning team at the Branch Transport s'-Heerenberg?

The answer to this research question is found by reviewing the literature for suitable forecasting methods. In this review, we use the factors found in the previous research question to help determine whether a forecasting method is appropriate. The answer to the research question is provided in Chapter 4.

4. How much does a professional forecast model increase the performance of demand forecast compared to the current situation of the domestic planning team at the Branch Transport s'-Heerenberg?

The answer to research question 1 provides the forecasting performance of the reconstructed current forecasting model, and the answer to research question 3 provides a forecasting model. With these answers, we compare the forecasting performance of the current forecasting method with the developed forecasting method to measure the performance improvement of a professional forecasting model. The answer to the research question is provided in Chapter 5.

5. How much transport capacity does the domestic planning team at the Branch Transport s'-Heerenberg require on a daily basis?

To answer this research question, we use the forecast to arrive at the required transport capacity on a daily level. The answer to the research question is provided in Chapter 6.

2. Current forecasting approach

This chapter addresses the current forecasting performance of domestic planning. Before we explain and evaluate the current forecasting model, Section 2.1 first explains the three types of shipments flows which are handled by the domestic planning. Then, Section 2.2 addresses the reconstructed current forecasting model. Section 2.3 provides an overview of forecasting measures and Section 2.3 evaluates the reconstructed current forecasting model using the selected forecasting measures. Section 2.5 summarized the findings of this chapter.

2.1 Shipment flows

Domestic planning mainly handles the following three types of shipment flows:

1. Pick-up flow; from a warehouse or customer in the Netherlands to the hub in 's-Heerenberg,
2. Drop-off flow; from the hub in 's-Heerenberg to a warehouse or customer in the Netherlands,
3. Direct flow; from a warehouse or customer directly to another warehouse or customer.

The pick-up flow consists of export and domestic shipments, whereas the drop-off flow consists of import and domestic shipments. However, the domestic shipment flow only consists of domestic shipments.

International shipments, i.e., import and export shipments, can be transported via groupage or direct transport. Groupage shipments are picked up at the customer warehouse and transported to a nearby cross-dock owned by Mainfreight or a partner. At the cross-dock, shipments with similar destinations are gathered and sent to a cross-dock near the final destination in a linehaul. There, shipments with the same or near destination are again gathered and transported to the final destination. Direct shipments are transported from origin to destination in one trip.

Domestic planning executes the drop-off of an import shipment if the shipment is transported via the groupage transport, whereas domestic planning executes the pick-up of an export shipment if the shipment is transported via the groupage flow. Suppose an import or export shipment is transported directly. In that case, domestic planning is not responsible for any transport movement, and therefore, this research will not consider these shipments. Note that in import and export shipments, the domestic planning only executes the drop-off respectively and pick-up. Other departments handle other transport activities; therefore, this research will only handle transport activities executed by domestic planning.

Domestic shipments are both picked up and dropped off within the Netherlands, plus a piece of the German Ruhr area, as depicted in Figure 1. This type of shipment can also be handled via groupage or direct transport. Domestic groupage shipments go over the cross-dock in 's-Heerenberg, which means the shipment is transported via two trips, usually in two days. In the direct transport, the shipment is picked up and dropped off on one trip and usually on the same day.

Whether a domestic shipment is transported via groupage or direct transport is a choice of domestic planning. In reality, groupage or direct transport choice depends on various factors, such as the pick-up and drop-off location, shipment size, and available capacity. We cannot identify whether domestic planning transported shipments via groupage or directly from the data. However, it is possible to define a rule of thumb based on the planners' experience. In consultation with the planners, we decided to only look at the shipment size as the rule of thumb because the shipment size is the most critical factor, and the shipment size can be easily extracted from the data, in contrast to the locations and available capacity. According to domestic planning, 7.5 loading meters can be considered the threshold to transport a shipment directly. So, as a rule of thumb, a domestic shipment is said to be transported directly if the number of loading meters is larger than 7.5. Otherwise, the shipment is transported by groupage transport.

Figure 1 Delivery area of the domestic planning

Table 1 provides an overview of the ratios between pick-up, drop-off, and direct flow per year. The table shows that 6.1 percent of the loading meter is transported directly. The other 93.9 percent is transported indirectly via the groupage flow. The ratio between pick-up and drop-off is essential for the efficiency of domestic planning because a truck can use its loading capacity twice per trip. An optimally used vehicle departs 's-Heerenberg full of drop-off shipments and returns full of shipments they must pick up. Table 1 shows that the ratio between the pick-up and drop-off volume is close to zero. Theoretically, the fleet can be used for 0.96 percent if this ratio continues in the daily data and other constraints for capacity are ignored.

Table 1 Ratios between pick-up, drop-off, and direct transport per year

Pick-up (%)	Drop-off (%)	Direct (%)
45.1%	48.9%	6.1%

2.2 Current forecast model

Recall that in Section 1.3, we said that domestic planning does not possess a professional forecasting model, but there is some documentation that can be reconstructed into a forecast model. The major limitation of the current approach is its unknown forecasting performance. The ignorance of the accuracy makes domestic planning cautious about using the forecast, making the evaluation of the forecasting performance one of the desirable results of the research. In this section, we start by describing this current forecasting approach.

The documentation that we can reconstruct into an official forecasting method consist of an overview of the total number of trucks deployed per day over the past four years. Based on this overview, the domestic department estimates the number of needed trucks per day. This approach to forecasting has several limitations. First, this approach makes no distinction between the several types of trucks and their corresponding loading capacities. Not distinguishing the load capacity of the vehicles ensures that the domestic department cannot make a valuable forecast of the number of loading meters because the available types of vehicles differ per day. Second, the estimation procedure is not defined, with the consequence that the forecast has to be calculated manually, and the estimation procedure is subjected to the interpretation of the forecaster. This own interpretation of the forecaster leads to an instable model. Finally, the approach ignores the type of transport flow, which is important because the ratio between pick-up and drop-off is an essential efficiency indicator, as a truck can use two times the loading capacity per trip.

We reconstructed this existing overview of the total number of trucks deployed per day into an actual forecasting model by defining the estimation procedure together with the domestic planning team. The defined estimation procedure has to be applied to a dataset containing loading meters to predict the number of loading meters instead of the number of trucks because, as mentioned before, the number of trucks is not a reliable unit for forecasting demand.

The forecast focuses on weeks and weekdays because it assumes that the weekday and week numbers indicate the expected demand. For example, the number of loading meters transported on Monday of week two last year indicates the expected number of loading meters that must be transported on Monday of week two this year. In the same way, the forecast considers the number of loading meters of the concerning day two and three years ago, but through weighting, more emphasis is placed on nearby years. The forecasting model also includes a trend based on the department's targeted growth rate (4%). The reconstructed current forecast is mathematically expressed in the following formula:

$$LDM_{y,w,d} = (0.2 \cdot LDM_{y-3,w,d} + 0.2 \cdot LDM_{y-2,w,d} + 0.6 \cdot LDM_{y-1,w,d}) \cdot 1{,}04 \quad (1)$$

where $LDM_{y,w,d}$ is the number of loading meters transported in year y, week w, and weekday d.

In order to evaluate the performance of the current model, we need to determine suitable forecasting measures. Therefore, the next section first gives an argumentation for suitable forecasting measures.

2.3 Forecasting measures

The main challenge of selecting a suitable forecasting performance measure is that different measures may lead to different conclusions. Even scientists have not agreed on which performance accuracy measure should be preferred to compare forecasting methods. This chapter discusses which performance measures are the most suitable to evaluate and compare the various forecasting methods. We identify the most popular forecasting performance measures utilized in the literature and analyse their differences. Then, an explanation will be given about the forecasting measure preference.

The forecast accuracy of a model is rarely 100%. The forecast can be slightly higher or lower than the actual values (Klimberg, Sillup, Boyle, & Tavva, 2010). The difference between the forecast value and the actual value is called the forecasting error:

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h} \quad (2)$$

where e_{t+h} is the forecasting error at time t + h, y_{t+h} is the actual value at time t + h, and $\hat{y}_{t+h}$ is the forecasted value at time t + h. Time t is the time at which the forecast is made, and h is the lead time of the forecast.

In general, accuracy measures can be split into three groups: directional measures, scale-dependent measures, and scale-independent measures. Directional performance measures are useful to measure the direction of the error. Scale-dependent performance measures are useful when comparing forecasting methods applied to similar scaled data, whereas scale-independent is more appropriate for comparing forecasting methods applied to different scaled data.

A forecasting measure that evaluates the direction of the forecasting error is the bias (or mean error). The bias can be calculated as the average of the forecasted errors:

$$Bias_{t+h} = \frac{\sum_n e_{t+h}}{n} \quad (3)$$

where e_t is the forecasting error at time t and n is the number of forecasting values. The bias measures the degree by which the forecasting model either overestimate or underestimate the actual values. The expected value of the bias should be close to zero. A forecast with a positive bias tends to underestimate the actual values, so the actual values are on average greater than the forecasted values. Conversely, a forecast with a negative bias tends to overestimate the actual values.

Scale-dependent performance measures are measures for which the error's size depends on the data's scale. Popular scale-dependent performance measures in the literature are the mean square error (MSE), root mean square error (RMSE), mean absolute error deviation (MAD), and mean absolute error (MAE), defined as follows:

$$RMSE_{t+h} = \sqrt{MSE_{t+h}} = \sqrt{\frac{1}{n}\sum_{n} e_{t+h}^{2}} \tag{4}$$

$$MAD_{t+h} = \frac{1}{n}\sum_{n} |e_{t+h}| \tag{5}$$

where e_{t+h} is the forecasting error at time t + h and n is the number of forecasting values.

The RMSE (and MSE) is related to the variance of the forecast error. In fact, the random component of the demand has a mean of 0 ($E[random] = 0$) and variance of RMSE ($Var[random] = MSE$. In other words, the RMSE measures the spread of the errors. RMSE penalizes large errors much more significantly than small errors because all errors are squared. According to Sunil Chopra (2019), MSE is an appropriate performance measure to compare forecasting methods if the cost of a large error is much higher than the gains from very accurate forecasts. In addition, Chopra argues that the MSE is only appropriate when the forecast error has a symmetric distribution around zero. If the forecast errors are normally distributed, then the MAD can also be used to estimate the standard deviation of the random component. The standard deviation of the random component is the MAD multiplied by 1.25.

The MAD is more appropriate than RMSE if the forecast error does not have a symmetric distribution. Even when the error distribution is symmetric, MAD is more appropriate if the cost of a forecast error is proportional to the size of the error because RMSE gives high penalizes to large errors relative to small errors. The smaller the MAD or RMSE, the more accurate the forecasting model. A major shortcoming of MAD and RMSE is that they are ignorant of the magnitude of the actual values. Consequently, there is no context that says something about the accuracy of the model.

A popular scale-independent performance measures is the mean absolute percentage error (MAPE). A popular scale-independent performance measures is the mean absolute percentage error (MAPE). The MAPE considers the effect on the magnitude of the actual values. The MAPE is defined as follows:

$$MAPE_{t+h} = \frac{1}{n}\sum_{n} \frac{|e_{t+h}|}{y_{t+h}} \tag{6}$$

where e_{t+h} is the forecasting error at time t + h, y_{t+h} is the actual value at time t + h, and n is the number of forecasting values. MAPE is an appropriate forecasting performance measure if the underlying forecast has significantly seasonality, which is the case, and if demand varies significantly from one period to the next (Sunil Chopra, 2019). MAPE has the major limitation that it gives infinite or undefined values for zero values or close to zero values. On weekends and public holidays, it is common that the actual value is zero or close to zero and this leads to infinite or undefined MAPE values, which makes the performance measure useless. In addition, the measure is based on percentage errors puts different penalty on positive errors than on negative errors, so those types of measures are asymmetric.

There have been attempts to resolve this problem. Makridakis (1993) proposes to exclude outliers from the averaging of the absolute error, where an outlier can be defined as an actual value that has a value of less than one or as an absolute percentage error value greater than MAPE plus three standard deviations. However, in this approach, removing the outliers can be a problem, and the exclusion of outliers can distort the results, mainly because the data involves multiple small actual values (Kim & Kim, 2016). In addition, Makridakis proposed the symmetric MAPE, which is more robust to values close to zero. However, the robustness of the symmetric MAPE is not strong enough to overcome the zero values during the weekends and national holidays because on those days, the symmetric APE value is regularly around 200%, with the result that those values influence the forecasting performance too much.

Considering the benefits and drawbacks of the forecasting measures, we conclude that the mentioned scale-dependent forecasting measures are more appropriate than the scale-independent ones. Because the MAPE and symmetric MAPE do not have the required robustness to zero values, and all forecasting methods will be evaluated on the same dataset, making the scale independence unnecessary. Regarding the scale-dependent RMSE, and MAD forecasting measures, we prefer the MAD because the forecast errors do not have symmetric distributions for the most forecasting models which are introduced later in this paper. In addition, domestic planning prefers a forecast for regular days, and they do not want to emphasize incidents, such as customer promotions too much. Despite the aforementioned advantages of MAD, we will also compare the RMSE values of the forecast models to check whether they lead to the same conclusions.

2.4 Evaluation current forecasting method

In order to evaluate the forecasting model, we need to collect data containing information about the number of loading meters per day because we said that the forecasting method defined in equation 1 has to be applied to a dataset containing loading meters. For the explanation of the data collection method and the cleaning and transforming process, we refer to Section 3.2. In order to fairly compare the current forecast performance with the forecast performance of other methods, the forecasting model is evaluated on the same data set, which is the last 306 days of 2021. For the explanation of the testing set, we refer again to Section 3.2.

Table 2 presents the forecasting performance of the reconstructed current forecast model. The model consists of a weighted average of actual demands from one, two, and three years ago on the corresponding days, with the result that a forecast with a lead time of one day gives the same accuracy as a forecast with lead times up to a year. The bias shows that the forecasting model tends to underestimate the actual demand. The mean absolute deviation is also relatively high because, on average, the forecast is 15.1 percent off by the actual demand.

Table 2 Forecasting performance of the reconstructed current forecast model.

	ME	RMSE	MAD
LDM	170.34	407.19	235.87
LDM/actual (%)	(10.90%)	(26.5%)	(15.09%)

Note. The forecasting performance is based on the test dataset.

In conclusion, even if forecasting equation 1 is applied to a data set containing information about the number of load meters, the forecast still needs to be improved. The positive bias is especially problematic because the forecast underestimates the number of loading meters on average by 170.34 meters. In addition, the forecast has a mean absolute deviation of 236.67, which is approximately 15.89% of the total demand.

2.5 Conclusion

In this chapter, we have addressed the first research question: What is the forecasting performance of the reconstructed current forecasting method? Domestic planning handles three types of shipment movements which can be referred to as pick-up, drop-off, and direct flow. A truck can a truck can use its loading capacity twice per trip. Therefore, it is essential to distinguish these flows because the ratio between pick-up and drop-off is crucial for the efficiency of domestic planning. Domestic planning lacks a professional forecast, but there is some documentation that is reconstructed into an actual forecasting model. This reconstructed forecasting model has some limitations, including its inability to distinguish these three types of transport flows. In addition, the overview uses the number of needed trucks as a forecast unit instead of loading meters. Finally, the evaluation of the model showed that even applying the reconstructed forecasting model to the obtained data set containing information about the number of load meters; the forecast still needs to be improved. In order to improve forecasting accuracy, we will propose and evaluate alternative forecasting methods in the remaining of this paper. Chapter 3 addresses the data analysis where the demand characteristics are identified. Chapter 4 describes some potential model to forecast the demand of domestic planning.

3. Data analysis

The chapter provides insight into the factors that influence the demand forecast of domestic planning. Section 3.1 explains which forecasting class is the most suitable to forecast the demand and the associated required data. Next, Section 3.2 describes the data collection and cleaning process. Section 3.3, identifies the factors that influence demand based on plot analysis. Then, we decompose the time series in Section 3.4. Next, in Section 3.5, we address the effect of national holidays on demand, and Section 3.6 summarizes the main finding of this chapter.

3.1 Selection of forecasting class and associated data

A forecasting class needs data, but the type of data required depends on the forecasting class. This section aims to identify the needed company-specific data by determining the most appropriate forecasting class. We will not evaluate the specific forecasting methods. So, in this section, we will evaluate the suitability of the forecasting classes in order to identify the required data. Later, in Section 4.1, we will provide an evaluation of suitable forecasting methods of the most suitable forecasting class determined in this section.

Forecasting methods can be classified as quantitative and qualitative methods. Quantitative methods need data that can be analysed in terms of numbers of equations, and qualitative methods do not have that need. The data requirement of quantitative methods is fulfilled because Mainfreight has historical records, including the demand of the domestic planning team. Due to the objective nature and some characteristics of quantitative methods, such as numerical values, quantitative methods have an advantage over qualitative methods in this situation. Therefore, we will focus on quantitative forecasting classes.

From the literature, we found the following three classes of quantitative forecasting methods:

1. Time-series forecasting methods
2. Causal forecasting methods
3. Artificial intelligence forecasting methods

Each quantitively forecasting class consists of various forecasting methods. Ghalehkhondabi, Ardjmand, Weckman, & Young (2017) concluded that, based on their studied review articles about demand forecasting methods in the energy sector, there is an agreement among all authors that none of the methods outperforms the others in all situations. So, the most suitable forecasting method depends on domestic planning-specific circumstances.

The main advantage of artificial intelligence over statistical methods is its ability to deal with complex and highly nonlinear problems and randomness (Velasquez, Zocatelli, Estanislau, & Castro, 2022). Particularly, artificial neural networks are commonly used in the literature as an artificial intelligence forecasting method. However, artificial neural networks do not provide a systematic means to improve the understanding of the system, which is desirable. Furthermore, the company's maturity regarding forecasting is still in its early stages, so no forecasting methods are currently used. Since simplicity is preferred over complexity, more straightforward methods, such as statistical methods, should be evaluated first to see if they can provide satisfactory results.

Causal methods are helpful when understanding the factors influencing the dependent variable is desired. However, identifying company-specific independent variables that explain the demand of the domestic planning team is difficult. It is also costly and time-consuming to obtain the entire data set for the independent variables. Therefore, we also omit regression.

Time series methods are the most used quantitative methods for estimating future transport demand when reliable historical data is available (Profillidis & Botzoris, 2018). Time series are series of data recorded and analysed in a time order, where the only independent variable is time. The fundamental assumption of time series is that all other factors affecting the transport demand follow the same path in the future as in the past, with the same characteristics and degree of influence. The shorter the forecasting horizon, the more likely the assumption is close to reality (Profillidis & Botzoris, 2018). The forecasting horizon of the forecast for the domestic planning team is a week; therefore, it is likely that the assumption holds. Besides that, the forecasting class seems suitable because the Mainfreight's data availability consists of time series. Therefore, we decide to forecast based on time series.

3.2 Data collection and cleaning process

Times series forecasting methods need time series data. This need for data is by Mainfreight's historical records of the executed shipments. The dataset with historical shipment records is obtained from the information stored in the transport management system (TMS) of Mainfreight. The ability of TMS to collect transportation data makes it possible to view it as a database in which information about shipment movements is stored. This information, stored by a TMS, can be used to perform data analysis in order to forecast future demands. The same dataset can also be used to develop the forecast model.

Before using the dataset, we have to clean and transform the dataset to the required format. Data cleaning is the process of preparing data for analysis by removing or modifying incorrect, corrupt, or inaccurate observations. Transformation is the process of converting data into a convenient format. For the cleaning process, we need multiple steps because the data has a high degree of contamination. The cleaning steps are performed carefully, but not all contaminating elements may have been filtered from the data due to the high degree of contamination of the dataset. We suspect that the residual contamination is negligible because we compared our dataset with a dataset based on trip lists. However, this suspicion cannot be established with certainty because the loading meters in the trip list database are not calculated in the same way as those for our database. The transforming process mainly consists of aggregating the data to both the total daily aggregated level (all flow) and the daily aggregated level separated by flows (pick-up, drop-off, direct flow). A detailed explanation of the cleaning and transforming process can be found in Appendix A.2 Data cleaning and transforming.

In general, empirical evidence based on the out-of-sample forecast is considered more reliable than empirical evidence based on in-sample forecast performance. Therefore, the given dataset is divided into a training set used for initialization and parameterization of the forecasting methods and a testing sample used to evaluate forecasting performance. The first 1155 observations of the data are used to estimate method parameters, and the remaining 306 observations to evaluate the post-sample forecasting performance. This gives a training/test ratio of approximately 4:1, which is a commonly used ratio.

3.3 Plot analysis

The cleaned and transformed dataset can be used to identify the characteristics of the demand. A good start to identifying trends and seasonality is to plot the data. Figure 2 depicts the time plot for the total daily demand, expressed in loading meters, of domestic planning. In this graph, no distinction is made between the three transport directions, i.e., pick-up, drop-off, and direct flow. The plot reveals a slightly positive trend that seems linear, i.e., an additive trend. However, additional research is needed to confirm this presumption. For an explanation of the composition of time series, we refer to Appendix A.3 Time series composition. As for seasonality, the data shows multiplicative seasonality because the magnitude of the seasonal fluctuations does vary over time. The figure reveals that seasonality consists of multiple seasonal cycles. It shows a short and long seasonal cycle where the long cycle is mainly visible during the summer and Christmas holidays. The shorter seasonal cycle dominates the data, and it is probably a within-week cycle with a period of seven days. The length of longer cycle, within-year cycle, is harder to identify. It requires further investigation by looking closer at some potential seasonal cycles.

Figure 2 Total daily loading meter demand of Domestic planning from 1-1-2018 to 31-12-2021.

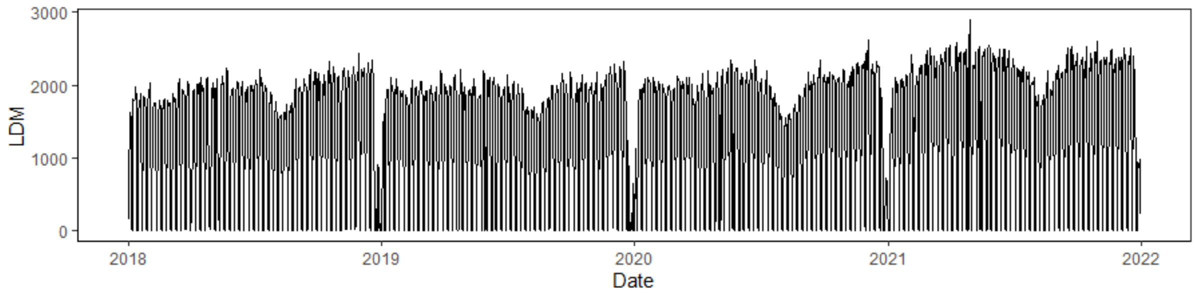
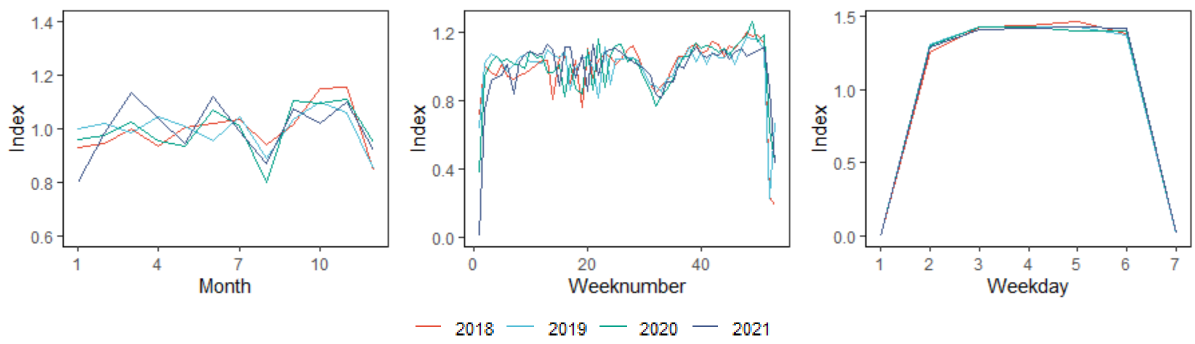


Figure 3 shows three seasonal cycles. The month and week number plots represents the two potential within-year seasonal cycles, and the weekday plot represents the within-week seasonal cycle. In the figure, the index represents the ratio between the average cycle period and the total yearly average. The plot with the monthly seasonal cycles does not show robust seasonality. The monthly plot does show a weak seasonality in the second part of the year, in contrast to the first part of the year. The plot with the week number seasonality reveals a more robust seasonality. As a result, we prefer the weekly seasonal cycle over the monthly one as the within-year seasonal cycle. From now on, the within-year seasonality indicates the week number seasonality. The weekly plot also exhibits some noise in the year's first half, especially between weeks 14 and 24. However, this noise is mainly caused by the effect of national holidays, such as Eastern, Ascension Day, and White Monday, which fall on different days and weeks each year and significantly influence the demand for domestic planning. Suppose we correct the effect of national holidays according to the smoothing-out process introduced in Section 5.3. In that case, we observe a reduction of the noise between weeks 14 and 24 and a more robust week seasonality. The weekday plot shows a strong within-week seasonality where Sunday (period 1) and Saturday (period 7) have much lower demand than the weekdays. To conclude, the within-week seasonality dominates the demand for domestic planning, but the week seasonality also provides additional information about the demand.

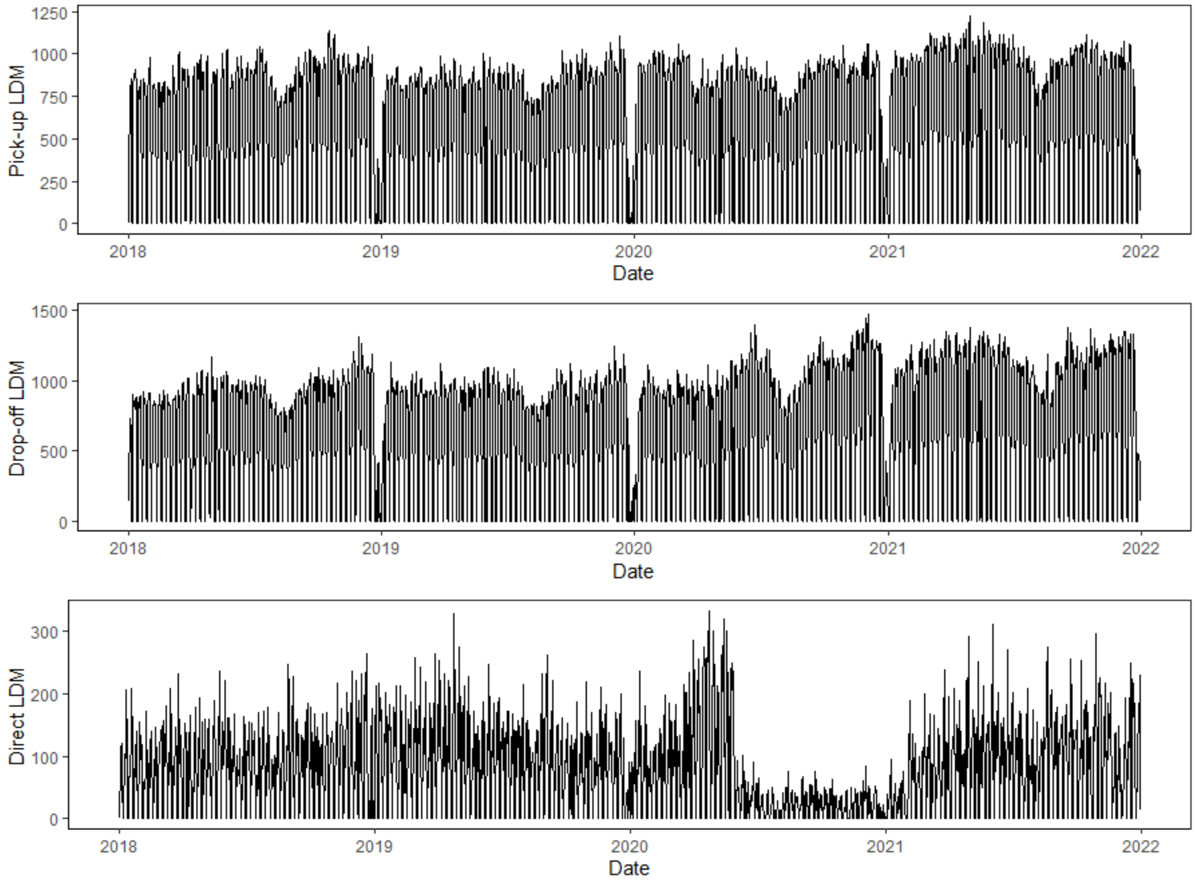
Figure 3 Season plots for three potential seasonal cycles.



The ratio between pick-up and drop-off is also essential for optimal vehicle capacity because a vehicle can use its loading capacity twice per trip. An optimally used truck departs 's-Heerenberg full of drop-off shipments and returns full of pick-up shipments. Figure 3 depicts the daily demand, expressed in loading meters, for the three transportation flows of domestic planning. The figure reveals that the pick-up and drop-off demand are similar and match's trend and seasonality. However, the direct flow exhibits a different demand pattern. The direct flow shows a within-week seasonality, but the week seasonality disappeared. A remarkable feature of the direct plot is the low demand in the second half of 2020. The available data cannot explain this event, and domestic planning cannot explain this either.

We also investigate the three potential seasonal cycles (monthly, weekly, and within-week) for the three transportation flows. These seasonal plots, depicted in Figure 20 in Appendix A1, confirm the transport activities' trend and seasonality assumptions made for the total daily demand.

Figure 4 Daily demand in loading meter for the three flows.



3.4 Time series decomposition

According to Brockwell & Davis (2016), there are multiple methods to estimate and eliminate the trend and seasonality. The classical decomposition model is the first method to estimate and eliminate trends and seasonality, as described by Brockwell & Davis (2016). The second method eliminates the trend and seasonality by differencing and finding an appropriate stationary model for the differenced series. In this section, we will use the classical decomposition method to eliminate the trend and the within-week and within-year seasonality. Later, in Section 4.4, we will use the second method, differencing, for ARIMA forecasting models.

In the decomposition method, we assume that the seasonal components are constant in time. It starts with deseasonalizing the demand by applying a moving average filter, defined by Sunil Chopra (2019) to eliminate the seasonal components. As the time series suggests a within-week (weekday) seasonality and the within-year seasonality (week number), we set the length of the within seasonality to 7 (m1=7), and the within-year seasonality is set accordingly to the ISO week numbers where a year has 52 or 53 weeks. The long years, with 53 weeks, occur on years that start on Thursday and on leap years that start on Wednesday, which is the case in 2020. This extra week is referred to as the leap week. For all other periods in the time series, the within-year cycle consists of 52 weeks. The moving average filter is expressed in the following equation:

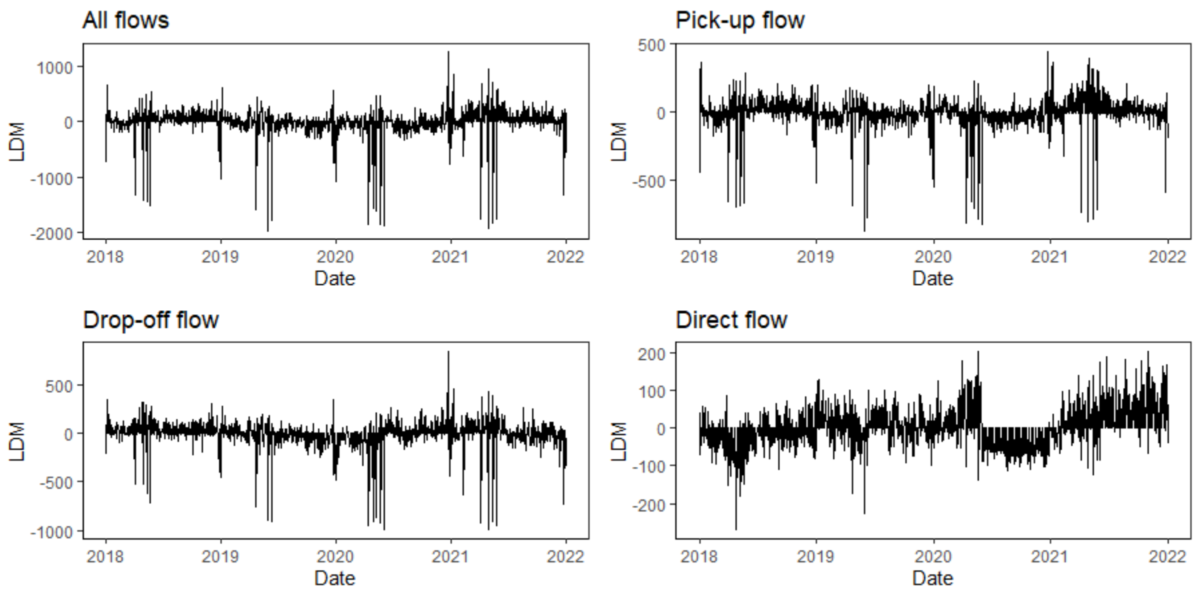
$$\bar{y}_t = \begin{cases} \dfrac{\left[y_{t-\left(\frac{m}{2}\right)} + y_{t+\left(\frac{m}{2}\right)} + \sum_{i=t+1-\left(\frac{m}{2}\right)}^{t-1+\left(\frac{m}{2}\right)} 2y_i\right]}{2\cdot m}, & if\ m = 364\ days \\ \dfrac{\left[\sum_{i=t-\left[\frac{m-1}{2}\right]}^{t+\left[\frac{m-1}{2}\right]} y_i\right]}{m}, & if\ m = 371\ days \end{cases} \tag{7}$$

where m is equal to 364 in a year with 52 weeks and 371 in a year with 53 weeks.

The level and trend components are estimated by fitting a linear trend using linear regression on $\bar{y}_t$. The initial level estimate is equal to the intercept coefficient, and the trend is equal to the slope coefficient of the linear trend. In the case of multiplicative decomposition, the seasonal factors are computed by averaging the ratio of the actual demand to deseasonalized demand obtained from the linear regression. For additive decomposition, the seasonal factors are computed by the average deviations between the actual demand and deseasonalized obtained from the linear regression. The seasonal factors are normalized so that they add to m for multiplicative seasonality and add to zero for additive seasonality.

Figure **5** shows that the random components of the time series, with the additive trend and multiplicative seasonality, are not white noise, meaning there is information left in the residuals. For an explanation of the composition of time series, we refer to Appendix A.3 Time series composition. The spikes in the pick-up and drop-off flow graphs confirm the suspicion that the national holidays significantly influence demand. The effect of national holidays on demand is further addressed in the Section 3.5. The direct flow graph shows that linear regression is an inadequate approach to capturing the trend component for the direct flow. However, the direct flow only represents 6.1% of the total demand, which raises the question of whether estimating a higher-order polynomial fit significantly affects the total demand. To model demand adequately, special day effects must be considered, along with the trend and seasonality.

Figure 5 Random component of the time series.

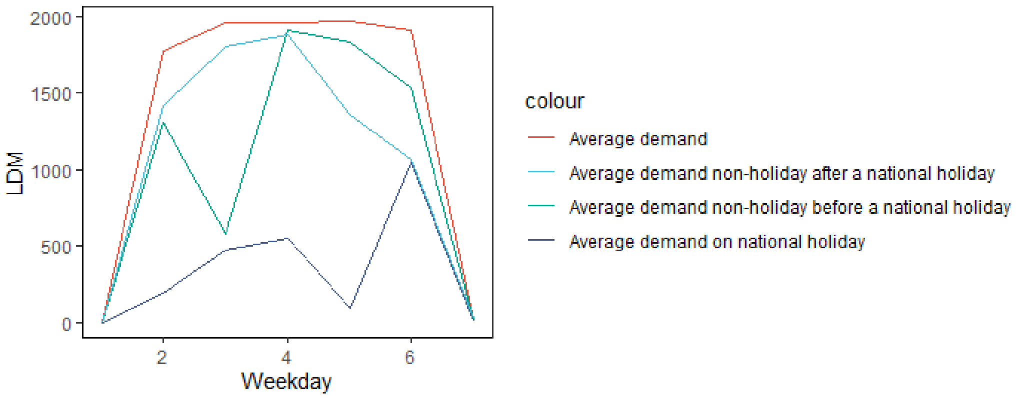


3.5 Anomalous demand

We concluded that the national holidays cause noise in the seasonality and spikes in the random component. Later in the study we will see even more effects of the national holidays. Therefore, national holidays significantly influence the data and should be incorporated into the forecasting model.

Figure 6 depicts some averages regarding special days calculated using the dataset. Special days consist of national holidays, listed in Table 7 in Appendix A.1 Figures and tables, and the day before and after the holiday because we know from experts in domestic planning that the days around national holidays also show a deviation from normal demand. This deviation can also be seen in the figure. It shows that the average demand is generally higher on normal days than on special days, with the highest difference on national holidays. The low average of a normal day before a national holiday on day 3 (Tuesday) is remarkable, but this can be explained by the fact that this event only occurred once in the dataset on December 24, 2019. In addition, the average demand after a national holiday is relatively low on Thursday and Friday, probably caused by long weekends.

Figure 6 Holiday plot from 2018 to 2021.



Note. The week starts on Sunday (weekday 1) and ends on Saturday (weekday 7).

3.6 Conclusion

In this chapter, we have addressed the second research question: Which factors significantly influence the demand forecast of the domestic planning team at the Branch Transport s'-Heerenberg? We conclude that the only forecasting methods based on time series are suitable for forecasting the demand of domestic planning. In addition, we determined that the data shows a strong seasonal cycle where the within-week seasonal cycle dominates the demand. However, even though the within-week seasonal cycle dominates the data, the week number seasonality also provides additional information about the demand. From the time series decomposition, we found that the direct flow cannot be captured adequately with linear regression, but the direct flow only represents 6.1% of the total demand. Therefore, we conclude that estimating a higher order polynomial fit probably does not significantly affect the total demand. Furthermore, we found that national holidays significantly influence the data and should be incorporated into the forecasting model.

4. Forecasting methods

In this chapter, we select appropriate forecasting models to predict demand for domestic planning. Section 4.1 discusses some well-known time series methods, and from these methods, we select Exponential smoothing and seasonal ARIMA models to forecast the demand. Then, in Section 4.2, we elaborate on Exponential smoothing, and in 4.3, we discuss the initialization and estimation procedure of Exponential smoothing. Next, Section 4.4 elaborates on the ARIMA models, and the associated estimation procedure is discussed in Section 4.5. In Section 4.6, we propose Rule-based forecasting for demand on special days because we argued that univariate methods could generally not produce reasonable forecasts. Finally, we summarize the finding of this chapter in Section 4.7.

4.1 Selection of time series forecasting methods

This section addresses the following four well-known time series forecasting methods: Moving average, Exponential smoothing, Autoregressive integrated moving average, and Trend projection. We start by explaining why the Moving average and Trend projection are not appropriate methods to forecast the demand for domestic planning. Then, we argue that Exponential smoothing and Autoregressive integrated moving averages are appropriate methods to forecast the demand for domestic planning. Finally, Exponential smoothing and Autoregressive integrated moving averages are weighed against each other.

Moving average is a simple forecasting method that averages the most recent observations and does not incorporate a trend and seasonality, which makes it inappropriate to forecast the demand of domestic planning. Trend projection projects historical statistical data to the future, requiring a large amount of reliable data, at least 7 to 10 years (Profillidis & Botzoris, 2018). Only four years of historical data are available, making Trend projection also inappropriate to forecast the demand of the domestic planning.

Exponential smoothing methods are recursive time series where new historical data values update the forecast. As the name exponential smoothing suggests, the weights of past observations decrease exponentially as the observations are more distant in time. Exponential smoothing was first introduced in the literature without citation to previous work by Brown in 1956. In 1957, it was expanded by Holt. Exponential smoothing is a well-known method for forecasting time series. Tayler (2003) states that the robustness and accuracy of exponential smoothing methods are the reason for their widespread use in applications where many series require an automated procedure. Exponential smoothing is a class of forecasting methods consisting of models that can capture various trends and seasonalities. This suggests that exponential smoothing with trend and seasonality might be a reasonable candidate for the demand forecasting of domestic planning.

Autoregressive integrated moving average (ARIMA) is a sophisticated econometric model that uses correlations between historical data at various times. The ARIMA model is a combination of three parts, the autoregressive part (AR), the moving average part (MA), and the integrated part (I). In addition, a seasonality index should be added to account for the seasonality in the demand data of the company. The ARIMA model can be generalized to a class of multiplicative ARIMA models to accommodate a seasonal cycle. These models are also known as seasonal ARIMA models or SARIMA models. This suggests that ARIMA models might be a reasonable candidate for the demand forecasting of domestic planning.

Both exponential smoothing and seasonal ARIMA might be appropriate methods to forecast the demand of the domestic planning team. However, Brockwell & Davis (2016) state that general theoretical statements about the appropriateness of the various forecasting methods for particular problems are challenging. However, they also argue that actual data is rarely generated by a simple mathematical model such as an ARIMA process. Therefore, heuristic models, such as exponential smoothing, should be seriously considered in practical forecasting problems. In addition, Beaumont, Makridakis, Wheelwright, & McGee (1984) say that the data characteristics are essential for choosing a forecasting model. When the trend-cycle is the dominant component in the data, sophisticated models such as ARIMA are appropriate. However, if the randomness is dominant, then simple methods are preferred, such as exponential smoothing (Li, Rose, & Hensher, 2010). Furthermore, from the point of view that simple methods which provide reasonable results with a satisfactory degree of accuracy should be preferred above complex methods and that real data is rarely generated by a simple mathematical model such as an ARIMA, we conclude that exponential smoothing methods. Particularly Holt Winter's model should be considered first. However, we will also investigate the possibilities of the ARIMA model as a second approach.

4.2 Description of Exponential smoothing

Exponential smoothing is a class of forecasting methods. The class consists of various forecasting methods with the property that forecasts are a weighted combination of past observations. If the error component is ignored, then there are fifteen exponential smoothing models, given in Table 3. This classification of exponential smoothing methods originated with Pegel's taxonomy in 1969, and it was later extended or modified respectively by Gardner (1985), Hyndman et al. (2002), and Taylor (2003). Some of these methods are better known under different names. For example, the “N,N” describes the simple exponential smoothing method and the “A,N” describes Holt's linear method, and the “A,M”. The “A,A” is also known as Holt-Winters with additive trend and “A,M” as Holt-Winters with multiplicative trend.

Table 3 Classification of exponential smoothing methods.

Trend component	Seasonal component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
A_d (Additive damped)	A_d,N	A_d,A	A_d,M
M (Multiplicative)	M,N	M,A	M,M
M_d (Multiplicative damped)	M_d,N	M_d,A	M_d,M

We will investigate the methods listed in Table 3 except for the damped trend methods. Damped trends dampen the trend as the length of the forecast horizon increases. Therefore, damped trends improve forecast accuracy over a long lead time (R. Hyndman, Koehler, Ord, & Snyder, 2008). However, the lead time of this research is up to 2 weeks, so the damped trends are unnecessary. The data analysis suggests that the time series contain an additive trend and multiplicative seasonality, but we also see value in testing the other models and benchmarking them.

In 2003, Taylor adapted the Holt-Winters (AM) exponential smoothing formulation to accommodate two seasonal cycles and showed that the forecasts produced by the new double seasonal Holt-Winters method outperform exponential smoothing with one seasonal cycle. In the plot analysis in 3.3, we saw that the within-week and within-year seasonal cycle shows some pattern where the within-week seasonal cycle is dominant over the within-year seasonal cycle. Therefore, the within-week seasonality cycle will be used for exponential smoothing methods that can accommodate one seasonal cycle. For Taylor's double seasonal Holt-Winters method, the within-week and within-year seasonalities will be used.

In this section, we will only show the equations Taylor's adapted Holt-Winters with double seasonality because that model has the highest expectation based on the demand characteristics found in Chapter 3. For the other exponential smoothing method, we refer to Appendix A.4 Formulas of the Exponential smoothing methods. Taylor's double seasonal Holt-Winters model (2003) with an additive trend and two multiplicative seasonalities gives the following component form:

$$Forecast\ equation: \quad \hat{y}_{t+h} = (l_t + hb_t) \cdot s_{t+h-m_1} \cdot w_{t+h-m_2} \tag{8}$$

$$\textit{Level equation:} \qquad l_t = \alpha \frac{y_t}{s_t \cdot w_t} + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{9}$$

$$\textit{Trend equation} \qquad b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{10}$$

$$\textit{Season 1 equation:} \qquad s_t = \gamma \frac{y_t}{l_{t-1} \cdot w_{t-m_2}} + (1-\gamma)s_{t-m_1} \tag{11}$$

$$\textit{Season 2 equation:} \qquad w_t = \delta \frac{y_t}{l_{t-1} \cdot s_{t-m_1}} + (1-\delta)w_{t-m_2} \tag{12}$$

where $0 \leq \alpha \leq 1$ is the smoothing parameter for the level, $0 \leq \beta \leq 1$ is the smoothing parameter for the trend, $0 \leq \gamma \leq 1$ is the smoothing parameter for the within-week seasonal cycle 1, $0 \leq \delta \leq 1$ is the smoothing parameter for the within-year (week number) seasonal cycle. Smoothing parameters control the rate at which the weights decrease. A forecast with high smoothing parameter values is responsive to recent observations, whereas a forecast with a low value of α is more stable and less responsive to recent observations.

4.3 Initialization and parameterization of Exponential smoothing

In this section, we give an argumentation for the choice of initialization and parameterization methods. We use the training data set for initialization and parameterization. For more information about the training set, we refer to Section 3.2.

Koutsandreas, Spiliotis, Petropoulos, & Assimakopoulos (2021) argue that scientists disagree about the appropriate measures and approaches that should be used for training a forecasting model, i.e., optimally selecting its parameters. Matching the utilization of the parameter with the forecasting performance measure is widespread practice in machine learning but not statistical forecasting. They state that forecasting models are usually parameterized by optimizing information criteria or by minimizing the in-sample one-step ahead sum of squared errors, regardless of the measure used for evaluating the forecasting accuracy in the test set. In 2021, they studied the effect of mismatching the training and testing measures on forecasting models' performance, and they concluded that mismatching has only a minor effect on forecasting accuracy. In addition, Sunil Chopra (2019) states that smoothing parameters should be parametrized by minimizing performance measures, such as MSE, MAD, and MAPE, which are also commonly used to evaluate the test set. Therefore, we will use both the MSE and MAD performance measures to estimate the parameters and evaluate the accuracy of the test set. The MAD measure derives the parameter values by minimizing the sum of absolute one-step ahead forecast errors using a nonlinear optimalisation method. In contrast, the MSE measure derives the parameter values by minimizing the sum of squared one-step ahead forecast errors.

In order to calculate the forecasts using exponential smoothing, we need to specify the initial values. Depending on the smoothing method, the set of initial values consists of a scalar value for the initial level and initial trend and a column vector of length m for the initial seasonality factor. For the initialization, we use both the method proposed by Sunil Chopra (2019) and R. J. Hyndman, Koehler, Snyder, and Grose (2002).

For the initialization process of the exponential smoothing models by Chopra, we refer to Section 3.4 except for simple exponential smoothing. The initial level value of simple exponential smoothing is estimated by taking the average of all historical data because simple exponential smoothing ignores trend and seasonality. Hyndman, Koehler, Snyder, and Grose determine the initial values slightly different. The main differences are that they also provide a method for calculating the initial multiplicative trend and additive seasonality. In addition, they refined the initial values by estimating them along with the parameters. The smoothing parameters and the initial states are estimated by minimizing performance metrics, which are the one-step ahead mean squared error or the one-step ahead mean absolute error in this research.

When estimating smoothing parameters, we set a minimum bound on the level smoothing parameter because the forecasting models tend to select zero as the optimum parameter value. The consequence of a zero-level parameter is that the model does not include a trend, while Section 3.3 shows that there is indeed a trend in the data. We suspect this tendency is due to the low trend, as we use a daily time unit. The effect of the national holidays also might be a factor that disrupts the level component because the low demand around national holidays unduly lowers the level component.

4.4 Description of ARIMA models

ARIMA models forecast the demand for domestic planning based on autocorrelations in the data. Autocorrelation is conceptually similar to correlation, but autocorrelation measures the relationship between the current demand of domestic planning and its past demands. We will explain ARIMA model by outlining each of its components.

The autoregressive component is the component that forecasts the variable of interest, and the demand of domestic planning, using a linear combination of past values of the variable. The term autoregression indicates regression of the variable against its historical values. Thus, the autoregressive part of order p can be written as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{13}$$

where c is the intercept, ϕ_1,…, ϕ_p are the parameters, and ε_t is white noise. i.e., it is independent and identically distributed with a mean of zero.

Rather than using past values of the demand of domestic planning in a regression, the moving average component uses a linear combination of past forecast errors. In other words, the forecast can be considered a weighted average of past forecast errors. Thus, the moving average component of an order op q can be written as:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{14}$$

where c is the intercept, θ_1,…, θ_q are the parameters, and $\{\varepsilon_t\}$ is white noise. Note that the word "moving average" in the context of ARIMA models differs from its use in the forecasting method Moving average.

The integrated component addresses the stationarity requirement of ARIMA models. A time series is stationary if its probability distribution does not change over time (Stock & Watson, 2019). Section 3.3 identified trend and seasonality components in the data, which means that the data is non-stationary because the trend and seasonal components will affect the data at various times. The advantage of stationary time series over non-stationary time series is its relatively simple predictability since the statistical properties can be assumed to be the same in the future as in the past. Non-stationary time series can be converted into stationary time series using mathematical transformations, such as differencing.

Differencing is a method of transforming non-stationary time series into stationary time series by taking differences between certain values. Differencing can contribute to the stabilization of the mean of time series by removing changes in the time series level and therefore eliminating or reducing trend and seasonality. For differencing, it is important to distinguish the lag operator and the difference operator. The lag operator, also called the backshift operator, operates on an observation of the time series to produce the previous observation. Formally defined, the lag operator (or backward shift) L is an operator that maps Y_t to Y_{t-1}. The difference operator gives the number that the time series is differenced.

As said in Section 4.1, the ARIMA model can be generalized to a class of multiplicative ARIMA models to accommodate a seasonal cycle. The multiplicative seasonal ARIMA model with one seasonal pattern can be expressed as ARIMA(p,d,q) × $(P,D,Q)_m$, where p and P are the orders of the autoregressive part, *d* and *D* are the degrees of differencing, *q* and *Q* are the orders of the moving average part, and m is the number of periods in a seasonal cycle. The model is multiplicative in the sense that the polynomial functions L and L^m are multiplied on both sides of equation 15. Note that the word "multiplicative" in the context of seasonal ARIMA models differs from its use in exponential smoothing. The backshift notation of the multiplicative seasonal ARIMA model with one seasonal pattern, for the time series y_t, can be written as:

$$\phi_p(L)\Phi_P(L^m)\nabla^d\nabla_m^D y_t = \theta_q(L)\Theta_Q(L^m)\varepsilon_t \quad (15)$$

where L is the lag operator, ∇ is the difference operator (1-L), ∇_m is the seasonal difference operator, (1-L^m), ε_t is the error term, and ϕ_p,Φ_p, θ_q and Θ_Q are polynomial functions of orders *p*, *P*, *q*, and *Q*, respectively.

4.5 Initialization and parameterization of ARIMA models

As with exponential smoothing, the training set is used for the initialization and parameterization of the ARIMA models. When fitting the ARIMA model to the forecasting flows (all flows, pick-up flow, drop-off, and direct flow), each forecasting flow has completed the following steps. In the remainder of this section, we will explain each of these steps in detail.

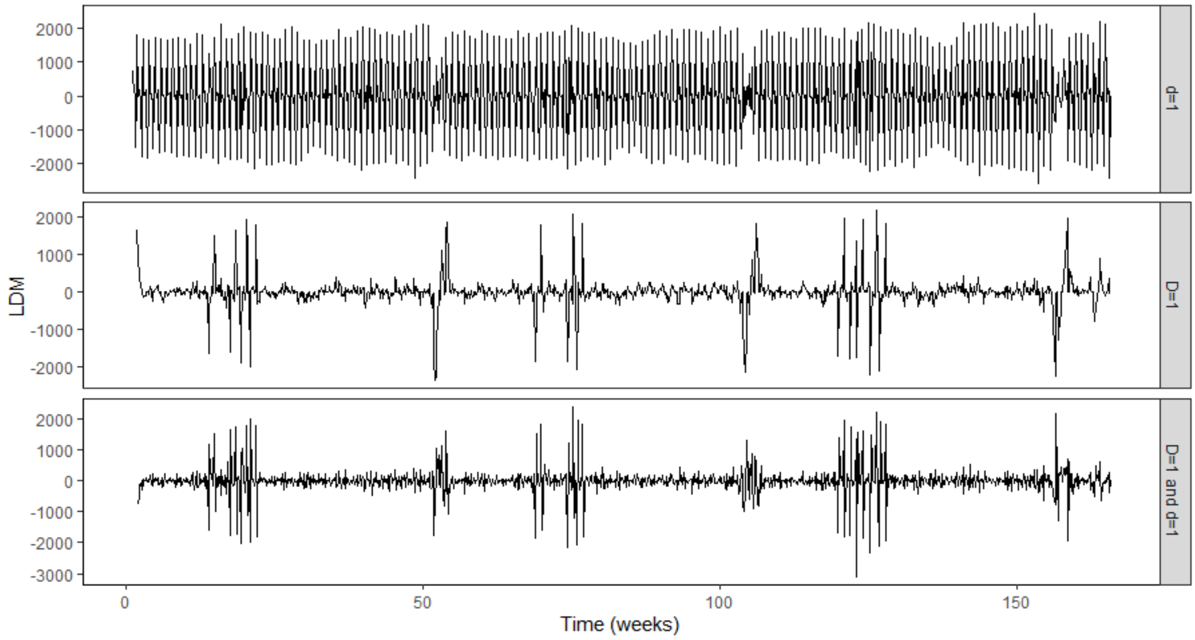
1. Transform the datasets to stationary time series.
   a. If necessary, the data is transformed using a Box-Cox transformation to stabilize the variance.
   b. If necessary, the data is stationarized through differencing to stabilize the mean.
2. Examine the autocorrelation function (ACF) and the partial autocorrelation function (PACF) to identify possible candidate models.
3. Compare all models with polynomials up to order two using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).
4. Estimate the values of the parameters using maximum likelihood estimation
5. Conduct goodness-of-fit checks on the residuals to investigate whether the model describes the data adequately by inspecting if information is left in the residuals.

We investigated logarithmic transformation for the demand in the training dataset to stabilize the variance because we determined that the magnitude of the seasonal fluctuations varies over time. Box-Cox transformation was used to resolve this instability of the variance in Section 3.3. The method of Guerrero (1993) was applied to the training set to estimate the lambda of the Box-Cox transformation. A problem with using a Box-Cox is that the back-transformed point forecast usually reports the median of the forecast distribution instead of the mean (source forecasting: principles and practice). To extract the mean point forecast, we used bias-adjusted back-transformation.

A method to remove the data trend is to take the first difference, which is the difference between consecutive observations (d=1). A method to remove the weekly seasonality of the data is to use seasonal differencing. A seasonal difference is the difference between an observation and past observations from the same season. Seasonal differencing uses the lag-d differencing operator, where the d represents the period of the season. Equation 15 accommodates only one seasonal cycle. Due to the dominancy of the within-week seasonal cycle over the within-year seasonal cycle, this will be the within-week seasonal cycle (D=1).

Figure 7 shows the time series of the total demand after taking the first trend difference (d=1), the first seasonal difference (D=1), and the first seasonal differencing and trend difference (D=1 and d=1). We also analysed this for the pick-up, drop-off, and direct flow, but we do not report these results because the plots are very similar, leading to the same conclusion as in Figure 7. In all three graphs, the mean seems to approach zero, and the trend seems to be removed. In the first difference, the weekly seasonality is strongly present, but the weekly seasonality is removed in the seasonal difference. There is, however, another type of pattern related to the national, summer, and Christmas holidays.

Figure 7 Various types of differencing on the total demand of the domestic planning.



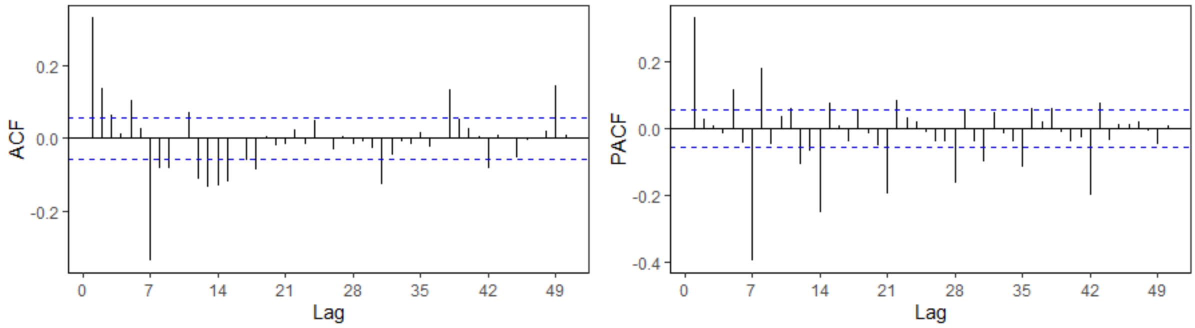
Note. The x-as represents the time in weeks, starting at week 1 of 2018.

A more objective method to determine whether differencing is required is to use a unit root test. Wang, Smith, & Hyndman's (2006) test determines whether seasonal differencing (seasonal period of one week) is required. This test indicates that the four-time series requires one seasonal difference. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and the Augmented Dickey-Fuller (ADF) test are used to determine whether trend differencing is required after seasonal differencing. The KPSS test has a null hypothesis of stationary data, but the ADF test has a null hypothesis of non-stationary. Both tests indicate that the time series does not require a trend differing at a 5% significance level. Therefore, only one seasonal difference will be applied to the four-time series to make it stationary.

Figure 8 plots the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the Box-Cox transformed and seasonal differenced training data of the total aggregated dataset. The ACF and PACF are plotted after seasonal differencing because we concluded that all four-time series requires one seasonal difference with the seasonal cycle of 7 periods. The ACF and PACF are used to see the correlation between the point, up to and including the lag unit. The ACF gives the autocorrelation with its lagged values, but the PACF gives the autocorrelation with its lagged values after removing the effects already explained by earlier lags.

The ACF and the PACF are used to identify possible candidate models. The ACF and PACF for the pick-up, drop-off, and direct flow are shown in Figure 21 in Appendix A.1 because they display a similar plot in terms of patterns, and the same conclusion is drawn for those time series. The PACF in Figure 8 shows an exponential decay in seasonal lags, and the ACF only has two significant seasonal lags, lags 7 and 14, which suggest a seasonal MA(2). Furthermore, the ACF shows an exponential decay in non-seasonal lags, whereas the PACF shows a significant lag 1, followed by insignificant lag two, which suggests a non-seasonal AR(1) lag. So, based on the ACF and PACF plots, we suspect that an ARIMA(1,0,0)(0,1,2) model can capture the characteristics of time series.

Figure 8 ACF and PACF plot for total daily aggregated trainings dataset.



Note. The training dataset used in this figure is already Box-Cox transformed, and the first seasonal difference with period seven is already taken.

We start with considering all models with polynomials up to order two and compare them using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). If necessary, the autocorrelation function of the residuals for higher-order autocorrelations is also checked. The AIC and BIC are measures of model performance that incorporate a penalty for the number of used parameters to avoid overestimation. According to the AIC, the best models are the ARIMA(2,0,2) × $(1,1,2)_7$, ARIMA(1,0,2) × $(2,1,2)_7$, ARIMA(2,0,1) × $(1,1,1)_7$, and ARIMA(2,0,1) × $(1,1,2)_7$ for the all, pick-up, drop-off, and direct flow, respectively. The best models according to BIC is the same as the best models for AIC for the drop-off and direct flow, but the BIC selects the ARIMA(1,0,0) × $(0,1,2)_7$ and the ARIMA(1,0,2) × $(0,1,2)_7$ as the best model for the total and pick-up flow, respectively. This difference is caused by the BIC penalizing additional parameters more than AIC.

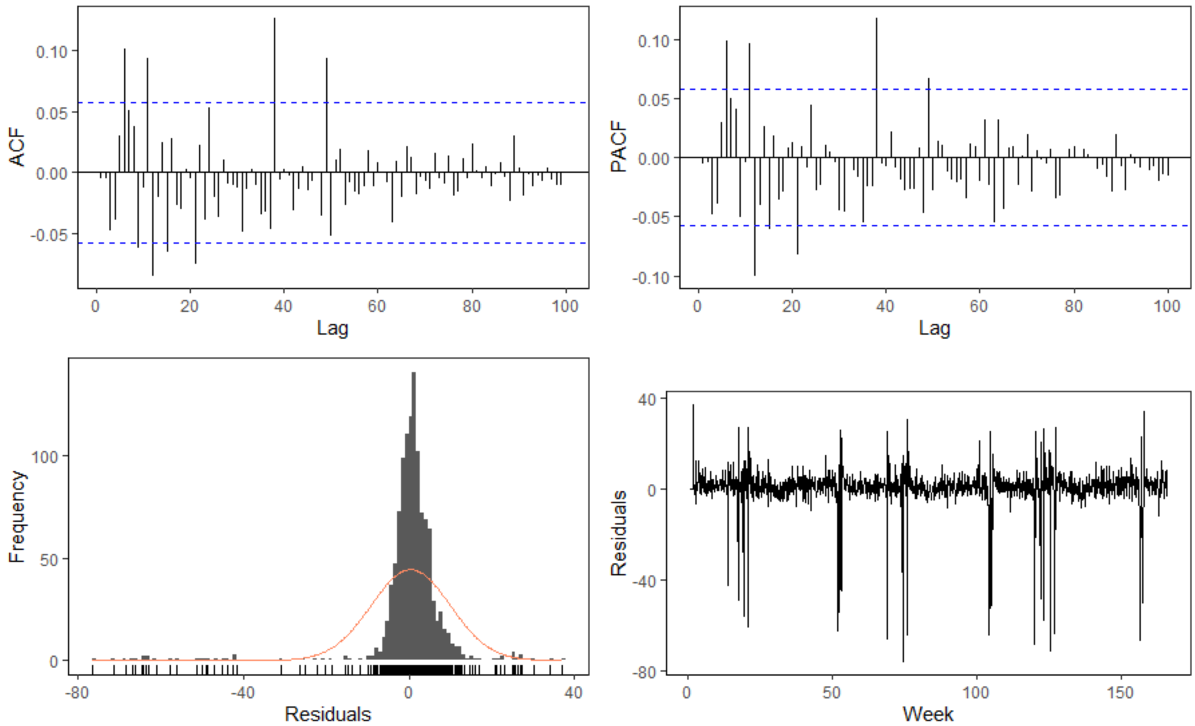
An investigation of the in-sample and out-of-sample performance for both models reveal that ARIMA(2,0,2) × $(1,1,2)_7$ outperforms the ARIMA(1,0,0) × $(0,1,2)_7$ both in-sample as out-of-sample based on both the RMSE and MAD. Therefore, we choose the ARIMA(2,0,2) × $(1,1,2)_7$ to forecast the total flow demand. Regarding the pick-up flow, there is hardly any difference between the RMSE and MAD of the AIC and BIC models. The MAD of the BIC model is slightly lower than the MAD of the AIC model for both in-sample and out-of-sample, but the RMSE of the AIC model is slightly lower than the BIC model. We choose to use the BIC model, the ARIMA(1,0,2) × $(0,1,2)_7$ model, because overfitting is undesired, and this model corresponds best with the conclusions of the manual ACF and PACF inspection. From now on, we shall refer to the chosen models as the seasonal ARIMA model.

Once the model's non-seasonal, seasonal, and differencing order has been identified, the parameter values are estimated using maximum likelihood estimation. Maximum likelihood estimation finds the parameter values such that they maximize the likelihood of obtaining the observed data. According to Brockwell & Davis (2016), maximum likelihood estimation can also be used for choosing the parameter coefficients and as a measure of goodness of fit of non-Gaussian distributions. For the ARIMA equations with parameter values of the four-time series, we refer to Appendix A.6 Formulas and parameter values of the ARIMA models.

Goodness-of-fit checks are conducted on the residuals to investigate whether the model adequately describes the data by inspecting if information is left in the residuals. The residuals should be white noise with zero mean, constant variance, and uncorrelated in time. In other words, y_t should be covariance stationery. In addition, the residuals should be strictly stationary to produce confidence intervals, which means that the residuals should also follow a gaussian distribution. The time series are stable because the inverse roots of the coefficients lie inside the unit circle.

Figure 9 shows four plots that check whether the residuals are normally independently distributed for the total flow. We do not report the results of the pick-up, drop-off, and direct flow because the plots are very similar, leading to the same conclusion as in Figure 9. The ACF and PACF plot shows that the residuals are autocorrelated because there are significant lags in the residuals. Many of these significant lags probably correspond with the national holidays. For example, lag 11 represents the number of days between Ascension Day and White Monday, lag 39 represents the number of days between Easter Monday and Ascension Day, and lag 50 represents the number of days between Easter Monday and White Monday. The national holidays also cause residual outliers in the residual time plot. The presumption of autocorrelation is supported by the Ljung-Box test, which rejects the null hypothesis that the residuals do not show autocorrelation. Finally, the histogram shows that the residuals are not gaussian distributed, and the Jarque-Bera test supports this conclusion. To conclude, the model's fit is not optimal but increasing the number of polynomials is not the solution because the suboptimal fit is caused by the national holidays and including 49 lags would cause major overfitting. We will use this model to produce a point forecast.

Figure 9 Goodness-of-fit checks for the ARIMA(2,0,2) × $(1,1,2)_7$ of the total flow.



4.6 Anomalous demand forecasting method

In Chapter 3, we concluded that the national holidays cause noise in the seasonality, and spikes in the random component. In addition, later in Section 4.5, we saw that ignoring national holidays leads to significant lags in the ACF and PACF plots of the residuals corresponding with the national holidays. Therefore, they contribute to the autocorrelation in the residuals of the ARIMA forecasting models. Thus, national holidays significantly influence the data and should be incorporated into the model.

In the literature, numerous forecasting methods have been proposed for forecasting the demand for normal days. However, modelling anomalous demand has often been ignored in the research literature by choosing periods with no special days, such as national holidays and long weekends. Anomalous demand poses modelling challenges because of its infrequent occurrence and the significant deviation from normal demand (Arora & Taylor, 2013). Taylor already argued in 2003 that demand on these days significantly differs from the normal demand and that univariate methods generally cannot produce reasonable forecasts. This statement is consistent with the findings found so far, such as the spikes in the random component and with finding in later sections.

Existing models have been modified to account for special days. For example, Smith (2000) accounts for special days by treating them as Sundays, and Taylor (2010 & 2012) smooths the special days out before modelling. However, Smith's method will lead to inaccurate estimations for special days because Smit assumes that all special days have similar demand profiles to Sundays, which is incorrect. Regarding Tayler's approach, smoothing out the special days will exclude the important potential contribution of the forecast to domestic planning because especially an accurate forecast for special days and the day around the special days can make a major contribution to domestic planning. In addition, smoothing out special days was not felt appropriate because it means that 28 days per year are excluded from the sample, which is almost one month per year.

An approach for short-term forecasting of normal and anomalous demand is based on regression, where dummy variables incorporate the effects of special days (Cancelo, Espasa, & Grafe, 2008). To avoid over-parameterization, some authors assume that the demand profile for different special days can be treated as similar or classified. The data shows that the assumption of treating special days as similar is too restrictive and classifying similar special days will still result in over-parameterization. Other approaches to incorporating special days are based on rules or Artificial neural networks. The latter method is disregarded because, as mentioned in Section 3.1, the maturity of the company regarding forecasting is still in its early stages. Therefore, more straightforward statistical methods should be considered first. We choose to use a rule-based approach to incorporate the effects of special days because the method gives the freedom to allow the incorporation of the special days' characteristics into a statistical modelling framework without over-parameterization.

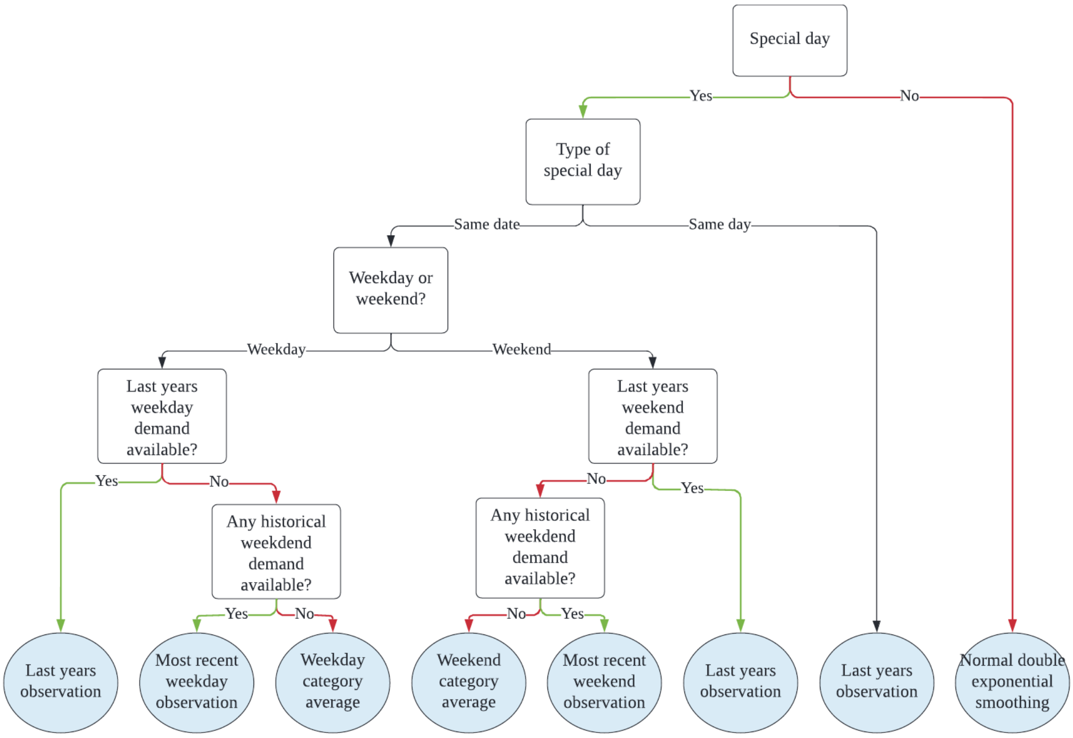
The prior knowledge and the data properties are quantified and expressed in the form of a rule. The rule aims to estimate a suitable point forecast for the special days in the formulations of double seasonal Holt-Winters because this method has the highest forecasting accuracy, which will be later seen in 5.2. The mathematical expression of the Rule-based double seasonal Holt-Winters with additive trend and multiplicative seasonality is given as follows:

$$\textit{Forecast equation:} \quad \hat{y}_{t+h} = I_{N_{t+h}}\left[(l_t + hb_t) \cdot s_{t+h-m_1} \cdot w_{t+h-m_2}\right] + I_{S_{t+h}}[a_{t+h}] \qquad (16)$$

where the $\left[(l_t + hb_t) \cdot s_{t+h-m_1} \cdot w_{t+h-m_2}\right]$ is the forecast equation for double seasonal Holt-Winters. The a_{t+h} value is the forecast for the special day on day t + h according to the special day rule, and the $I_{S_{t+h}}$ and $I_{N_{t+h}}$ are the binary indicator terms for the occurrence of a special day. The binary indicator $I_{S_{t+h}}$ equals one if t occurs on a special day and zero otherwise, whereas $I_{N_{t+h}}$ equals one if t occurs on a normal day and zero otherwise. So, at any given day t, the sum of $I_{S_{t+h}}$ and I_{N_t+h} is equal to one. The forecast of the special day a_t is determined according to the special day rule, explained in the following paragraph.

The special day rule treats weekdays as having a distinct impact on anomalous demand compared to weekends. There are two special days: special days that occur on the same day of the week each year (e.g., Ascensions Day and Pentecost) and special days that occur on the same date each year (e.g., New Year's Day and King's Birthday). We know that seasonality has a major impact on demand within the week. However, there is only reliable historical data available for four years, so we cannot extract the effect of national holidays on every day of the week from the data. We also know that demand on the weekend days is somewhat similar. The same applies to weekdays; therefore, this rule will only distinguish between weekends and weekdays. However, for several special days, there is no historical data available for the special days. In that case, the average demand of the historical days for that weekday or weekend day will be taken. For special days that occur on the same day of the week each year, this rule refers to the historical demand observed on the same special day from the previous year. The special day rule is depicted in Figure 10.

Figure 10 Graphical representation of special day rule.



4.7 Conclusion

In this chapter, we have answered the third research question: Which quantitative forecasting models are most appropriate to forecast the demand, expressed in loading meters, of the domestic planning team at the Branch Transport s'-Heerenberg? We concluded that exponential smoothing and seasonal ARIMA might be appropriate methods to forecast the normal demand of the domestic planning team because both models can capture various trends and seasonalities. In addition, we argued that univariate methods generally could not produce reasonable forecasts due to their infrequent occurrence and the significant deviation from normal demand. Therefore, we proposed a Rule-based forecasting method where the demand on special days is forecasted according to the special day rule and on normal days according to the double seasonal Holt-Winters because this method has the highest forecasting accuracy, which will be later seen in Section 5.2.

5. Forecasting performance

This chapter evaluates the performance of the proposed forecasting methods. In Section 5.1, we start with describing the model evaluation procedure. Then, Section 5.2 compares the forecasting performance of the proposed exponential smoothing and ARIMA models. Next, Section 5.3 evaluates the performance of Rule-based forecasting for special days. We investigate whether Rule-based forecasting, used in conjunction with the best model of Section 5.2, also leads to better forecast accuracy for normal days. Section 5.4 discuss whether the rule-based double exponential smoothing model with the smoothing and updating adjustment can be further improved by investigating the errors. Section 0 provides prediction intervals in order to evaluate the uncertainty associated with the forecast. Finally, Section 5.6 summarizes the main findings of this chapter.

5.1 Model Evaluation Procedure

This section describes the procedure of determining the forecasting performance in terms of root mean squared error (RMSE) and mean absolute error (MAD) for the h-step ahead forecasting error (h = 1, 2, …, 14).

The parameters for the methods are estimated using the training set according to the estimation methods described in Chapter 4. These parameters are used to produce an out-of-sample forecast for the testing sample. Starting at observation y_t where t = 1156, we forecast the demand, in terms of loading meters, for lead times from one- up to and including 14-step ahead. In other words, forecasts are made one day to two weeks ahead. After the forecasts at time 1156 are produced, the actual demand observation is added to the set of known observations. Then, forecasts are produced for $\hat{y}_{1157+h}$ after which the actual demand observation at time 1157 is added to the set of know observations. This procedure repeats until the last observation of the testing set, observation y_{1463} is added to the set of known values. The forecast for observations beyond the testing set, for example, $\hat{y}_{1550+14}$, are omitted from the set of forecasts because the actual observation of the demand for it is not in the testing set.

The performances of the one- up to and including 14-step ahead lead times are evaluated as the average of the error measure of the respective step-ahead forecasts, i.e., the performance of the 1-step ahead error is the average of all 1-step ahead errors in the testing sample and the performance of the 2-step ahead error is the average of all 2-step ahead errors in the testing sample, etc.

5.2 Comparison of forecasting methods

Exponential smoothing and seasonal ARIMA are appropriate methods to forecast the normal demand of the domestic planning team. The performance of those proposed models for normal days is estimated according to the model evaluation procedure and compared using the MAD and RMSE. We have only proposed and tested one model regarding ARIMA models, but for exponential smoothing, we tested several models. In order to keep the result figures clear, we first give three observations regarding some exponential smoothing models.

First, exponential smoothing models with multiplicative trend cannot capture the deterministic components correctly. Its inability to capture the deterministic components results in large values of RMSE and MAD. This result is consistent with earlier findings. For example, Section 3.3 already suggested that a multiplicative trend is inappropriate. The conclusion is also consistent with Hyndman and Athanasopoulos (n.d.) statement that multiplicative trend methods tend to produce poor forecasts. Second, the results of exponential smoothing with both additive trend and seasonality are very similar those of exponential smoothing with additive trend and multiplicative seasonality. Therefore, we decide only to show the results of Holt-Winters with multiplicative seasonality to avoid plotting models with similar results. Finally, the results of the forecasting models initialized and parameterized by Chopra are very similar to those of the models initialized by R. J. Hyndman, Koehler, Snyder, and Grose (2002).

Figure 11 compares the out-of-sample forecasting performance of the exponential smoothing and the ARIMA models for lead times up to 14 days ahead for the four-time series (all flows, pick-up flow, drop-off flow, and direct flow). The figure depicts the MAD, the preferred performance measure in this research, as explained in Section 2.3. Note that the figure has a logarithmic scale to fit the models' forecasting performance in the plot. In addition to the MAD lead times, we also calculated the RMSE lead times, but we do not report these results here because the relative performances of the methods for this measure were very similar to those of the MAD. The results for the RMSE lead times can be found in Figure 22 in Appendix A.1 Figures and tables.

The reconstructed current forecast can only be seen in the “all flows” plot because the current approach cannot forecast the demand of the three shipment flows. The reconstructed current forecasting model depends on observed historical demands of at least one year ago, resulting in a straight line at 235,87 for the "all flows" time series in Figure 11. The reconstructed forecasting model outperforms simple and trend exponential smoothing models and the ARIMA model at all lead times, except for the ARIMA model with a lead time of one. However, the reconstructed forecasting model is outperformed by Holt-Winters and Double seasonal Holt-Winters. In the remaining Section 5.2, we will take a closer look at the performances of the proposed forecasting models.

Figure 11 shows that the MAD of the simple and trend exponential smoothing methods reflects the seasonality of the demand of the domestic planning in the all, pick-up, and drop-off flows. The performance of those models is inadequate due to their inability to capture seasonality. This inability of simple and trend exponential smoothing to capture trend causes the smoothing parameters to go to extremes, with the MAD optimized smoothing parameters often approaching one, resulting in a model with high responsibility to recent observations. A notable result is that trend exponential smoothing performs worse than simple exponential smoothing, while we determined that the data contains a trend in Chapter 3. This event is due to the initial parameters being calculated based on the training set, where the trend is relatively flat compared to the trend in the test sample. In addition, the high responsiveness causes the forecast to be one day behind, which ensures that the forecast error is greatest when the demand differs from the previous day, i.e., on Monday and Saturday. The prediction error is the slightest if we predict seven days in advance because the prediction is made precisely for the same day of the week. The trend component does not capture the trend correctly due to the large differences between the levels due to the high responsiveness. It makes the forecast more extreme, resulting in trend exponential smoothing performing worse than simple exponential smoothing.

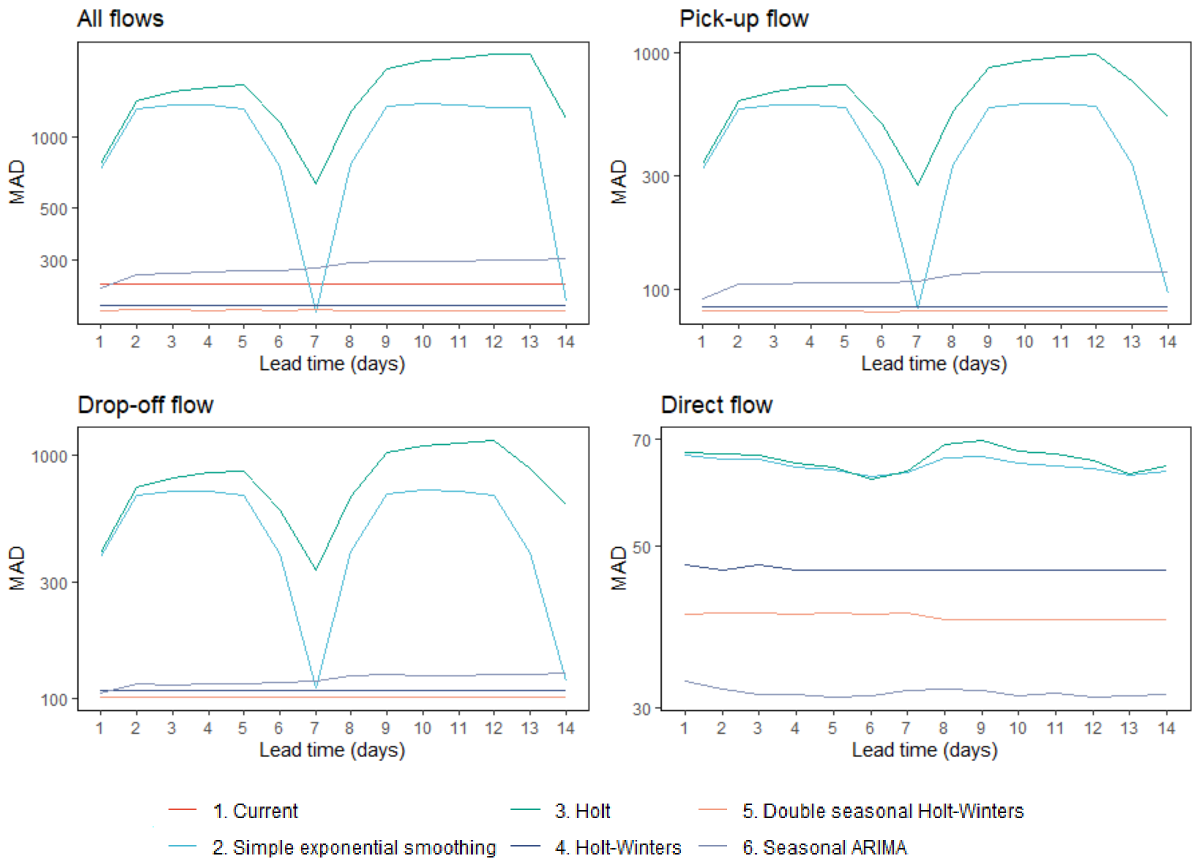
The ARIMA model is not competitive with the seasonal exponential smoothing methods for the time series, except for the direct flow. This result is consistent with the statement made by Li, Rose, and Hensher (2010) in 4.1 that ARIMA is preferred if the trend-cycle component is dominant in the data and exponential smoothing if the randomness is dominant. In the Section 3.3, it can be seen that the trend-cycle component is only dominant in the direct flow, which explains that ARIMA does outperform double exponential smoothing for the direct flow but not for the other flows.

Despite ARIMA's better performance of the direct flow forecast, we will forecast the direct flow using Double seasonal Holt-Winters for several reasons. Firstly, the difference in performance between double exponential smoothing and ARIMA is not large enough to significantly influence the required vehicle capacity because the share of direct flow is only 6.1% of the total demand. In addition, ARIMA models require more statistical knowledge than exponential smoothing models, and the difference in performance does not outweigh the more straightforward model of exponential smoothing for domestic planning. Finally, it is easier to integrate the forecasts if they are predicted using the same method.

The Holt-Winters outperform simple exponential smoothing, Holt's model, and ARIMA across all lead times and flows, except the direct flow. However, Holt-Winters is outperformed by the Double seasonal Holt-Winters model across all lead times and flows. Thus, incorporating within-year seasonality based on week numbers results in a better forecast performance in terms of mean squared error. In the Holt-Winters and double seasonal Holt-Winters, we observe two notable events:

1. Notable is the relatively constant performance of the Holt-Winters and double seasonal Holt-Winters across the lead times for all flows. The forecast for one day ahead is approximately just as accurate as 14 days ahead. We can explain this consistent performance by looking at the effect of the level, trend, and seasonality components on the forecast value. As said in Section 4.3, the level smoothing parameters tend to select the minimum bound (0,0001) as the optimal parameter value. The low smoothing parameter causes a stable level component. In addition, the short time bucket (one day) and the stable level component keep the trend component small and relatively constant. So, if we forecast a short period ahead, the seasonality component(s) mainly determine the forecasting values. In a forecast with lead times of one to fourteen days, the within-week seasonal cycle has the greatest influence on the forecasting values. Each within-week seasonal factors (Sunday, Monday, ..., Saturday) is updated once a week, which ensures that the forecasting performance is fairly constant for lead times of one to seven days. A shift between lead times seven and eight is therefore caused by a seasonal factor derived from data from two weeks ago instead of one.
2. Remarkably, despite the minimum bound on alpha, the model does not capture the trend in the beta parameters in the intended way. It seems that the model is trying to capture the trend in the seasonality factors. This event may be caused by the relatively high incorrect update of the smoothing parameters on special days due to the major deviations from the observed demand. This unfavourable update on national holidays, in combination with the flat trend from 2018 to 2019, could have resulted in the benefits of the level parameter update being secondary to the unfavourable update on national holidays.

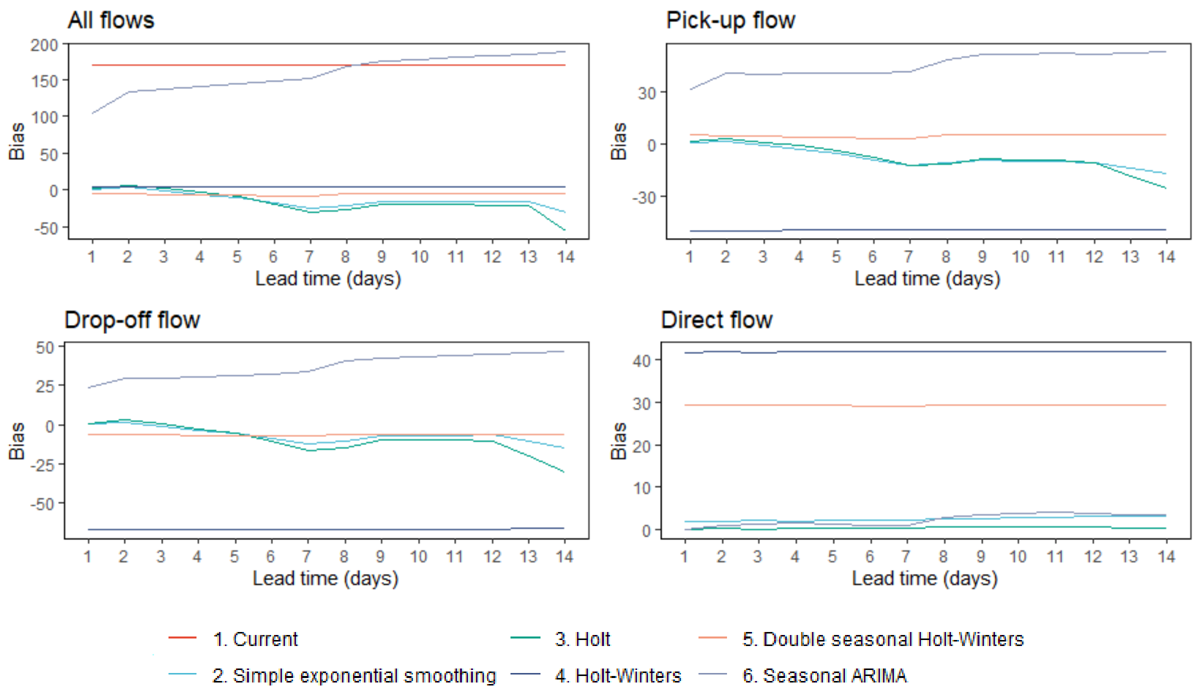
Figure 11 Comparison of methods for normal days using MAD.



In Figure 12, we also briefly look at the direction of the forecasting error using the bias. The current forecast has a bias of 170.34. The positive bias is especially problematic because the forecasts underestimate the number of loading meters on average by 170.34. This under-forecasting tendency of the current approach can be explained by the fact that the trend from 2018 to 2020 is much lower than in 2021, as a result of which the current approach gives too low forecast values. The ARIMA models also give a high bias for all lead time for the all, pick-up, and drop-off flow. The double seasonal Holt-Winters do not have a large directional deviation, except for the direct flow.

To summarize, Double seasonal Holt-Winters outperforms the reconstructed current forecasting models and all other proposed models across all lead times and flows, except the direct flow. The forecast based on ARIMA has the highest performance for the direct flow due to the dominance of the trend-cycle component in that flow. However, despite the higher performance of the ARIMA model, we will forecast the direct flow with Double seasonal Holt-Winters because the difference in required vehicle capacity is not big enough to justify the difference in necessary statistical knowledge and the more difficult forecast integration. In the next section, we evaluate the performance of double exponential smoothing, where a rule is included to forecast the demand on special days.

Figure 12 Comparison of methods for normal days using bias.



5.3 Performance of rule-based forecasting

This section aims to evaluate the results of rule-based forecasting. The special rule has the purpose of estimating a suitable point forecast for special days in the formulations of double exponential smoothing. For a description of the special rule and the equation of rule-based forecasting equation, we refer to Section 4.6. First, we compare the results of double exponential smoothing and the rule only for special days to investigate whether the rule leads to a better forecast on special days. Then, we compare the forecasting performance of the original double exponential smoothing and the rule-based version across normal days to determine whether the rule-based version also leads to better forecast accuracy for normal days.

Table 4 presents the results of the rule for just the special days. The figure shows that the original double exponential smoothing method is outperformed by the rule-based approach for the four flows. The root mean squared error (RMSE) and the mean absolute value (MAD) has been more than halved for all flows except the direct flow. The RMSE and MAD also decreased by approximately one-third for the direct flow. The table also shows that the rule-based forecast underestimates the value of the forecast, which is reflected in the positive bias. This underestimation is caused by the trend not being included in the rule and by not distinguishing whether an after-holiday day can constitute a long weekend with a special day. The trend is not included in the forecast because the special day rule looks to the most recent value of the special day that is representative of the special day this year, i.e., it looks at the most recent value of the special day falls on the weekend if it falls on the weekend this year and weekday if it falls this year falls on a weekday.

Table 4 Forecast accuracy on special days expressed in loading meters.

	Total aggregated flow		
	Bias	RMSE	MAD
Original double seasonal Holt-Winters	-328.17	801.96	496.23
Special day rule	97.30	351.75	187.56
	Pick-up aggregated flow		
	Bias	RMSE	MAD
Original double seasonal Holt-Winters	-152.99	348.20	206.57
Special day rule	42.03	147.15	76.36
	Drop-off aggregated flow		
	Bias	RMSE	MAD
Original double seasonal Holt-Winters	-162.43	402.68	243.02
Special day rule	49.71	198.82	110.72
	Direct aggregated flow		
	Bias	RMSE	MAD
Original double seasonal Holt-Winters	-19.33	68.96	44.19
Special day rule	5.56	47.87	28.50

Note. The original double seasonal Holt-Winters model is the model with the lowest out-of-sample MAD and MSE in Section 5.2 Data 2019 to 2021.

Now that we have concluded that the rule-based approach on special days performs better than the original double exponential smoothing method, we present the forecast accuracy of the double exponential smoothing when used in conjunction with the rule as specified in equation 16 in Section 4.6. In addition, we will present three rule-based double exponential models with the following adjustments.

Adjustment 1. No updating on special days.

In Section 5.2, we suspected that the special days caused major deviations resulting in a relatively high incorrect update of the smoothing parameters. Therefore, we need to test whether not updating the smoothing parameters on special days results in better forecasting performance. So, on normal days, the smoothing parameters are updated according to equations 9 to 12 of Section 4.2, whereas there will be no updating on special days.

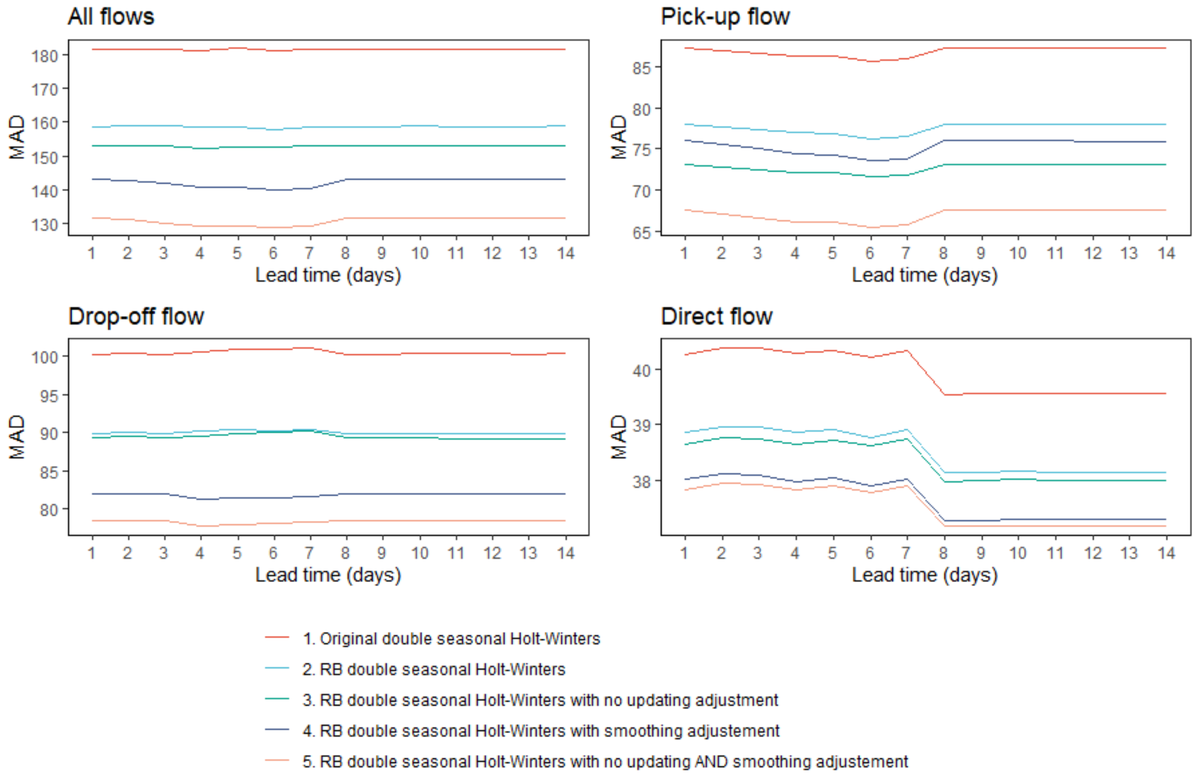
Adjustment 2. Specials days are smoothed out.

In weeks with special days, the forecast model generally underestimates the demand because special days lower the week number seasonal factors. Note that smoothing out in this context has a different meaning than in exponential smoothing. In this context, smoothing out means that we compensate for the low demand on special days in order to derive better initial seasonal factors for within-year seasonality (week number factors). The smoothing out process is described in Appendix A.7 Smoothing out process.

Adjustment 3. No updating on special days and special days are smoothed out.

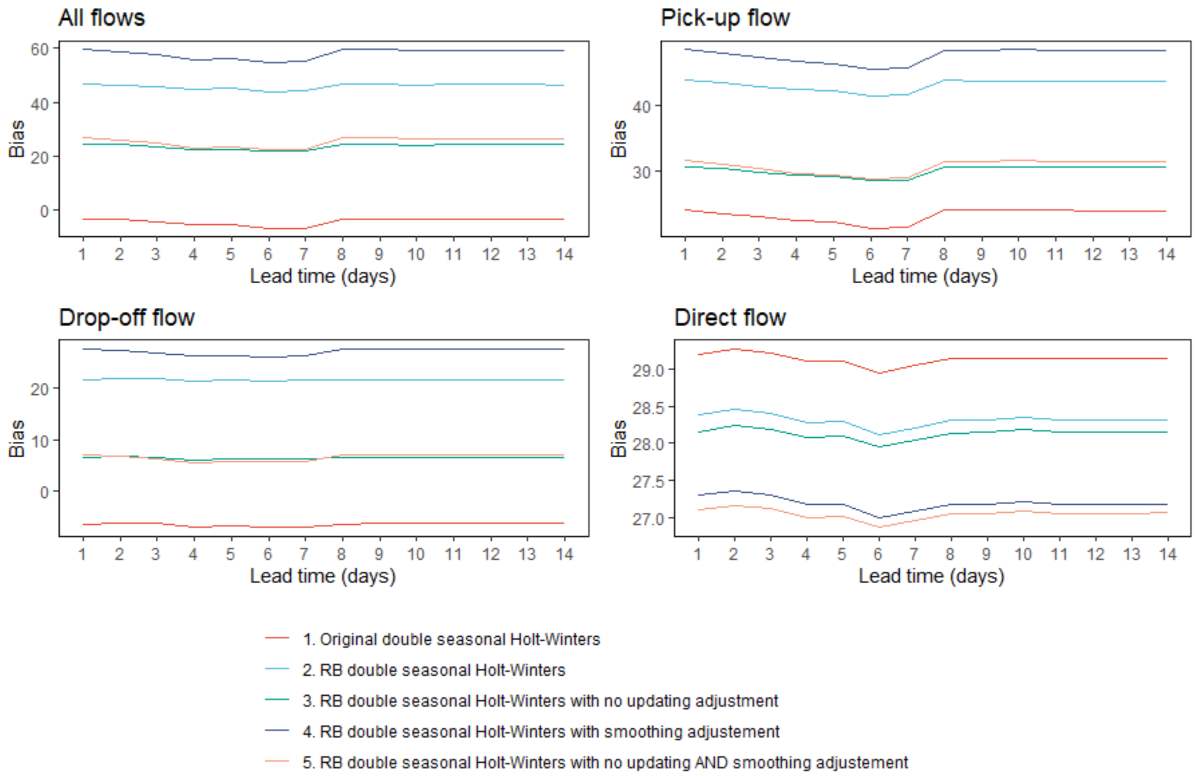
The last adjustment is a combination of adjustments 1 and 2. With this adjustment, we can test whether the combination increases the performance of original rule-based double exponential smoothing models and whether it increases the individual performance of adjustments 1 and 2. Figure 13 presents the forecasting performance using the MAD of the four rule-based double seasonal Holt-Winters methods. The RMSE shows very similar results and can be found in Figure 23 in Appendix A.1 Figures and tables. As benchmarks, we use the original double seasonal Holt-Winters methods. The figure shows that the original model is outperformed by all versions of the rule-based forecasting models. Thus, we can conclude that the rule-based double seasonal Holt-Winters also leads to better forecast performance for normal days. Furthermore, rule-based double seasonal Holt-Winters with adjustment 2, smoothing special days, outperforms the models that do not have the smoothing adjustment. Adjustment 1, no updating during special days also increases the performance of the model compared the rule-based model without adjustments. Combining the two adjustment gives the highest forecasting performance in terms of MAD.

Figure 13 Forecast performance of Rule-based double seasonal Holt-Winters using MAD.



In Figure 14, we also briefly examine the direction of the forecasting errors using the bias measure. The figure reveals that rule-based double seasonal Holt-Winters increases the bias across all lead times and flows compared to the original double seasonal Holt-Winters model, except for the direct flow. This increase in positive bias is mainly caused by special days. In Table 4, it can be seen that the bias on special is very negative for the original model and positive for the special day rule. Figure 14 also shows that adjustment 2, not updating on special days, increases the forecasting performance in terms of bias. Not updating the smoothing parameters results in a lower bias across all lead times and flows. The smoothing parameters of the double seasonal Holt-Winters model with adjustment can be found in Appendix A.5 Smoothing parameters of the final developed forecasting model.

Figure 14 Forecast performance of Rule-based double seasonal Holt-Winters using the bias.

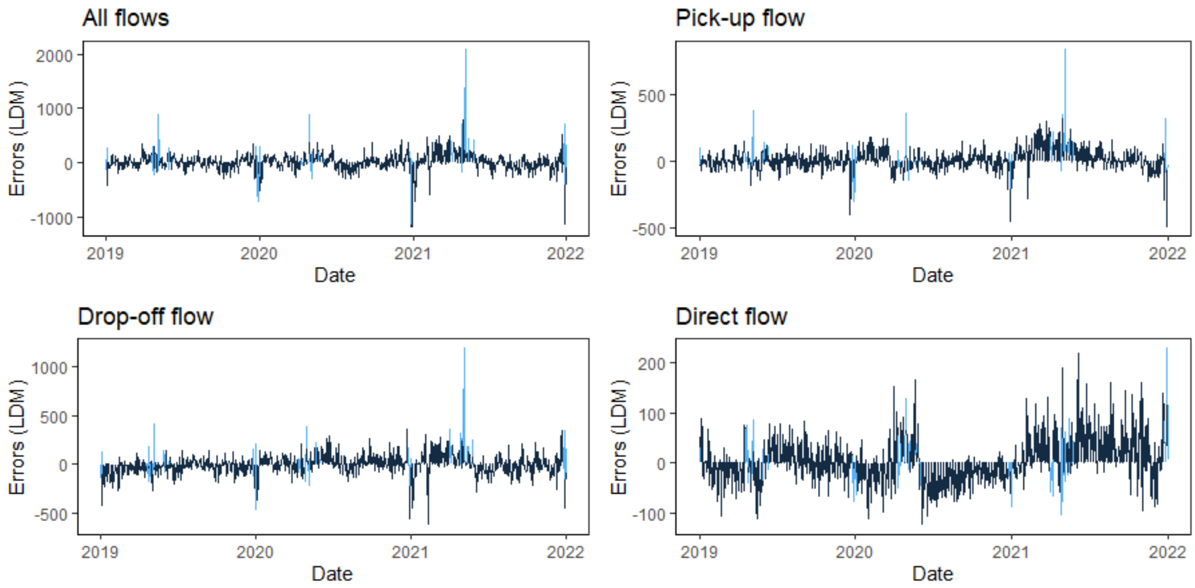


5.4 Error analysis

In this section, we discuss whether the rule-based double seasonal Holt-Winters model with the smoothing and updating adjustment can be further improved by investigating the forecasting errors. Recall that we said that the random component of the time series should be white noise, meaning no pattern can be found in the errors because that means that there is information unused.

Figure 15 shows the forecast errors of applying smoothed Rule-based double seasonal Holt-Winters without updating on special days on the four-time series. The figure shows the forecast errors of the training and testing set so that we can study the difference between the in-sample and out-of-sample. In the errors of the flows, an undulating pattern can be seen, indicating unused information left in the errors. In addition, the magnitude of the undulating pattern also seems to increase as we approach the test sample. Furthermore, the figure shows that the most prominent errors still occur on special days, and there seems to be a pattern on some special days. We also observe high errors on days around special days, especially on non-special days that are the missing link for long weekends and on non-special days in holiday periods, such as Christmas. Therefore, we conclude that the national holidays significantly affect more days than just the day before and after the national holiday. Thus, even though the special rule has already halved the MAD and MSE, improving the performance of the forecast model is still possible. Particularly in the area of the effect of national holidays on demand for domestic planning.

Figure 15 Forecasting errors for the time series.



The black errors are errors on normal days, and the blue ones are on special days.

5.5 Prediction intervals

The point forecasts express the uncertainty associated with the forecast. Forecast distributions can be used to express the forecasts' uncertainty because they describe the probability of observing future values. The point forecast is the mean of the forecast distribution. The prediction interval gives an interval within the expected demand of domestic planning with a specific probability, called the coverage distribution. The prediction interval for the h-step forecast is:

$$\hat{y}_t \pm z\hat{\sigma} \ , \qquad \hat{\sigma} = 1.25 \cdot MAD \tag{17}$$

where $\hat{\sigma}$ is an estimate of the standard deviation of the forecast distribution, and z is the multiplier of the normal distribution.

This prediction interval assumes that the forecast errors are normally distributed, but this assumption is incorrect, as seen in Figure 24 in Appendix A.1 Figures and tables. However, the Central Limit Theorem states that the errors approximately follow a normal distribution when the sample size is large enough. The test sample consists of 306 observations, allowing us to assume that the errors are approximately normally distributed.

Table 5 exhibits the standard deviation of the reconstructed current forecasting method and the Rule-based double seasonal Holt-Winters method in loading meters which are needed to calculate the prediction interval. The table shows that the prediction interval bounds for the rule-based double seasonal Holt-Winters have become smaller for the total demand (all flows). This shows less uncertainty in the point forecast because the width of the prediction intervals of all probability coverage is decreased by 44.2%. Domestic planning can determine the necessary daily fleet capacity based on the desired probability coverage. Note that the constant performance of the forecasting model ensures that these prediction intervals apply to all lead times from one day to two weeks.

Table 5 Standard deviation for the prediction interval.

Model	Flow	Probability coverage			
		75%	80%	90%	95%
Reconstructed current forecasting model	All	339,07	377,40	483,54	577,89
Rule-based double seasonal Holt-Winters	All	189,14	210,53	269,74	322,37
Rule-based double seasonal Holt-Winters	Pick-up	97,16	108,14	138,56	165,59
Rule-based double seasonal Holt-Winters	Drop-off	112,63	125,36	160,62	191,96
Rule-based double seasonal Holt-Winters	Direct	54,40	60,55	77,58	92,72

5.6 Conclusion

In this chapter, we have answered the fourth research question: How much does a professional forecast model increase the performance of the demand forecast compared to the current situation of the domestic planning team at the Branch Transport s'-Heerenberg?

We performed an out-of-sample evaluation to compare the reconstructed forecasting model with proposed models based on exponential smoothing, ARIMA, or Rule-based forecasting. We used the parameters, estimated using the training dataset, as input to compare the forecasting models on de testing sample. Forecasting performances of the one-up to and including 14-step ahead lead times are evaluated to compare forecasting methods. As a forecasting performance measure, we used MAD and RMSE.

The reconstructed current forecast can only be compared with the daily total aggregated demand (all flows) because the forecast cannot forecast the demand of the individual three shipment flows. We conclude that the reconstructed current forecasting model is outperformed by Holt-Winters and double seasonal Holt-Winters across all lead times, whereas Holt-Winters is outperformed by the double seasonal Holt-Winters model across all lead times and flows. Thus, double seasonal Holt-winter has the best forecast performance if special days are not treated separately.

A remarkable result is the relative constant performance of the Holt-Winters and the double seasonal Holt-Winters models. This means that the forecast for one day ahead is approximately just as accurate as 14 days ahead. The reconstructed current model also has consistent performance, but that is because the model relies on the observed historical demand from at least a year ago. The consistent performance allows us to extend the conclusions drawn for the one-step forecast to the other lead times used. Thus, the double seasonal Holt-Winters increase the forecast performance, in terms of MAD, by 23% compared to the reconstructed current forecasting method for all lead times.

Furthermore, we conclude that the double seasonal Holt-Winters cannot produce a reasonable forecast for special days. Rule-based forecasting for just the special days approximately halves the MAD and MSE for the time series, except for the direct flow, where the MAD and MSE decreased by approximately one-third. When the rule for special days is used in conjunction with double seasonal Holt-Winters for normal days, we see that the original model is outperformed by the rule-based forecasting across all lead times and flows. The two proposed adjustments for the rule-based forecasting model even increased this forecast performance. Like the original double seasonal Holt-Winters, the rule-based double seasonal Holt-Winter with the two proposed adjustments also has a constant forecasting performance across all lead times for all flows.

In order to answer the research question, we compare the forecasting performance of the best-proposed model, Rule-based double seasonal Holt-Winters, with the reconstructed current forecasting model. We conclude that a professional forecast model increases the performance of the demand forecast by 44.2%. In addition, the width of the prediction intervals with probability coverage of 0,75, 0,80, 0,90, and 0,95 are also decreased by 36.8%. Despite this improvement in forecasting performance, the forecasting errors show an undulating pattern which indicates unused information left in the errors which can be used to increase the performance of the forecasting model. Especially the effect of holidays can be investigated further to improve the forecast performance.

6. Required transport capacity

This chapter addresses the process of converting the forecast into the required transport capacity of domestic planning. Section 6.1 explains how the optimal order capacity can be determined using the newsboy problem. Then, Section 6.2 investigate whether the forecast can better predict the required capacity than what has actually been deployed. Finally, Section 6.3 summarizes the main findings of this chapter.

6.1 Optimal order capacity from Trucks & Drivers

This research aims to develop a forecasting model that contributes to the vehicle allocation problem. Domestic planning (and international planning as well) do not possess reliable predictions regarding the expected demand, which makes it difficult to distribute collective capacity optimally. The previous chapters focused on developing the forecast model, and this section concentrates on converting the predictive information of forecasting into an optimal ordered transport capacity at Truck & Drivers by using the newsboy problem.

The newsboy uses the trade-off between the cost of understocking and overstocking of vehicles to determine the required transport capacity. Understocking means there is not enough vehicle capacity to transport all shipments, and overstocking means there are idle or redundant vehicles. We assume the cost of overstocking (C_u) by one vehicle is €100,00, and the cost of understocking (C_o) is approximately three times as high. These costs are based on the expert opinion of Mainfreight and can be refined if desired.

The optimal cycle service level (CSL) can be calculated according to the following equation:

$$CSL = \frac{C_u}{C_u + C_o} = 0{,}75 \tag{18}$$

Then, the optimal ordered transport capacity at Truck & Drivers (O*) can be computed with the following equation:

$$O^* = F^{-1}(CLS, \mu, \sigma) \tag{19}$$

where μ is the point forecast and σ is the standard deviation of the point forecasts in the test set.

As mentioned before, the pick-up/drop-off ratio is essential for the efficiency of domestic planning because a vehicle can use its payload twice per trip. The larger of the two determines the necessary capacity for the groupage shipments. The required capacity is equal to the sum of the necessary capacity for the groupage and direct flow shipments. In addition to the flow, we also distinguish the days of the week as the standard deviations are significantly different. The optimal order capacity of loading meters for the first week of the test set is given in Table 6. Recall that we have concluded, in Chapter 5, that the forecast for one day ahead is approximately just as accurate as 14 days ahead. As a result, the optimal order capacity calculated two weeks ahead is just as accurate as one day ahead.

Table 6 Optimal ordered capacity of the first week of the test set.

Date	Optimal order capacity			
	O^* Pick (LDM)	O^* Drop (LDM)	O^* Direct (LDM)	O^* Total (LDM)
01/03/2021	865,57	1.092,71	102,09	2.287,50
02/03/2021	959,68	1.142,10	89,53	2.373,72
03/03/2021	971,94	1.169,72	98,08	2.437,52
04/03/2021	967,23	1.167,09	129,45	2.463,64
05/03/2021	957,18	1.170,96	151,36	2.493,28
06/03/2021	4,81	67,32	0,06	134,70
07/03/2021	0,65	0,48	0,07	1,36

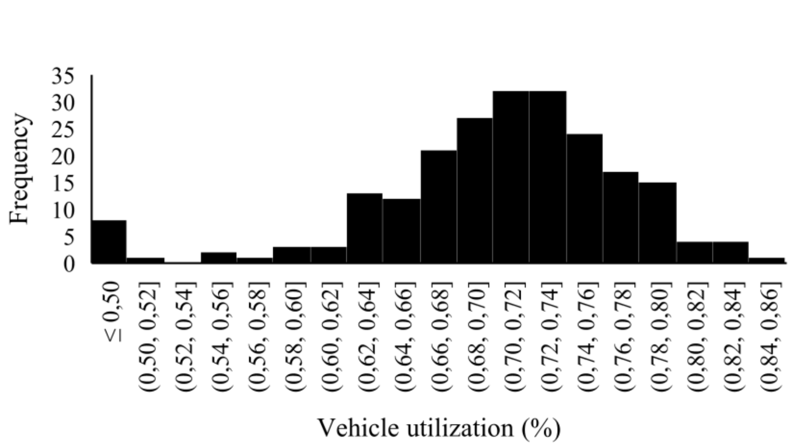
Until now, the demand is expressed in loading meters, but domestic planning orders vehicles at Trucks & Drivers who manage the collective transport capacity. So, the number of loading meters must be converted into vehicles. A potential problem for this conversion is the varied loading capacities of the various vehicles. To avoid this problem, this study will use a standard-size vehicle of 13.6 loading meters. The collective vehicle capacity has been converted into standard-sized vehicles, and from now on, the word 'vehicle' refers to the standard-sized vehicle. The previously mentioned under and overstocking costs per vehicle are also based on the standard-sized vehicle.

A standard-sized vehicle has a transport capacity of 13.6 loading meters. However, there are other limiting factors for the number of loading meters a vehicle can transport per day, such as the number of stops, distance, and driving time. To determine the vehicle capacity utilization, the total available capacity is divided by the used capacity for each day in the test sample. The exact deployed capacity per day is unknown, but we have estimated the deployed capacity based on the following assumptions:

1. Domestic planning has a standard fleet of vehicles consisting of several types of vehicles with different capacities. Table 8 of Appendix A.1 gives overview of the domestic standard fleet of vehicles.
2. If less than the standard capacity is deployed, the deployed capacity is the ratio between the deployed vehicles and standard fleet vehicles multiplied by the standard fleet capacity.
3. If more than the standard capacity has been deployed, we assume that the extra capacity consists of vehicles with a transport capacity of 13.6.

Figure 16 depicts a histogram of the daily capacity utilization in the test sample. The figure shows that the most commonly achieved capacity utilization is between 70 and 74%. In the optimal situation, we want a high-capacity utilization rate on the one hand, but on the other hand, the workload for the drivers should not be too high. This consideration leads to an optimal capacity utilization rate of around 78%. The actual capacity utilization is lower than depicted here because not all types of capacity deployment are included due to insufficient information. Nevertheless, because the performance of all forecasting models and the capacity advised by the forecast are both based on this number, it does not heavily influence on the outcome of this analysis.

Figure 16 Histogram capacity utilization.

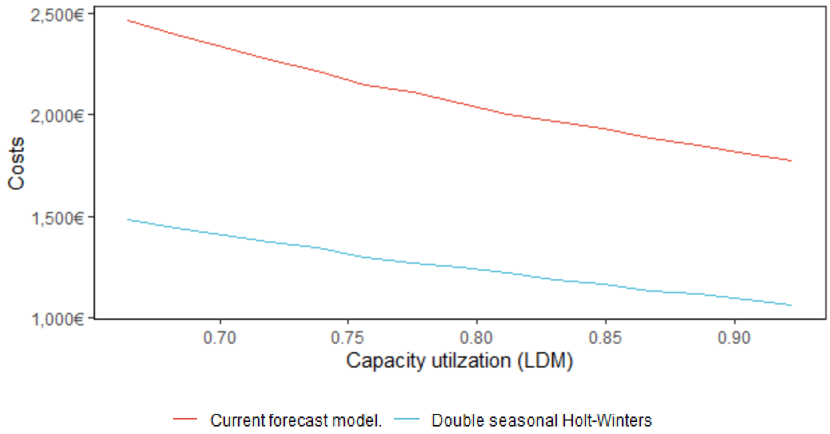


Note. The capacity utilization is based on the weekdays of the test set because the deployed vehicle capacity on weekend days is unknown.

Based on the optimal capacity utilization rate and the observed demand in the dataset, it is possible to calculate the required transport capacity for each day. Dividing this required transport capacity by the standard-sized vehicle capacity utilization gives the required capacity in standard-sized vehicles.

Figure 17 shows the theoretical average daily under and overstocking costs of domestic planning during the test period. This figure confirms that the double seasonal Holt-Winters outperforms the reconstructed current forecasting model. 'Theoretical' refers to the fact that in real life, domestic planning does not only rely on the recommended optimal order quantity from the forecast but also includes other factors that may not always have been included in the forecast, such as already received demand information. For example, suppose much more demand for tomorrow has already been confirmed at a certain time today than domestically expected based on the forecast. In that case, they will respond to this situation by hiring extra capacity. The costs that should actually be compared are the costs of additional renting and scaling down capacity in a certain period of time. However, those costs are not well established at the moment. Therefore, we decided to use €100,00 and €300 as over and understocking costs, respectively, because these could be determined with sufficient certainty. Suppose these scaling up and down costs can be determined with sufficient accuracy in the future. In that case, domestic planning can use the same approach to compare the actual costs of the reconstructed current forecasting model and the developed double seasonal Holt-Winters.

Figure 17 Theoretical average daily under and overstocking costs.

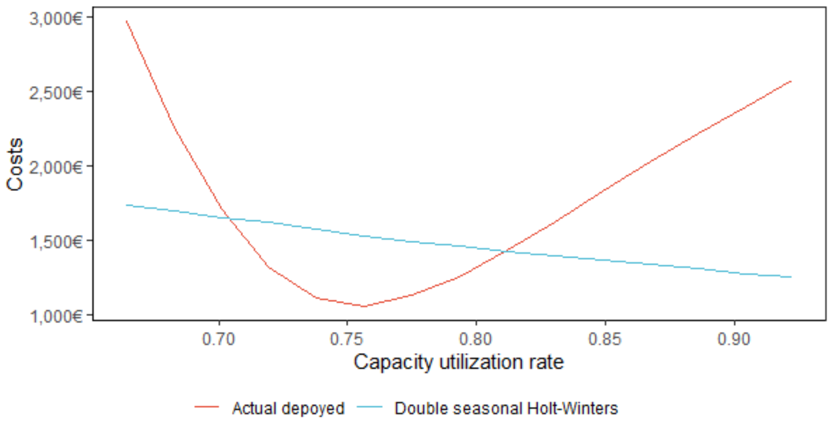


6.2 Forecast vs. reality

In this section, we examine whether the advised optimal order capacity from Section 6.1 is competitive with the actually deployed capacity. The outcome of the analysis actually determines whether the forecast can be trusted blindly until the last moment. In reality, the domestic can scale up and down the transport capacity to a point until the evening based on the actual number of incoming demand.

Figure 18 depicts the average daily under and overstocking costs of domestic planning if the advice according to the developed forecast model is 100% followed and the costs of what actually happened. The figure shows that the forecast is not competitive with the actual deployed capacity between capacity utilization rates of 0,71 and 0,80. Utilization rates above 0.80 almost do not occur in the test sample, as can be seen in Figure 16, and rates below 0,71 are not desired. Thus, the forecasting model is competitive with the actually deployed capacity. It is not advised to rely blindly on a forecasting model. The last-minute upscaling and downscaling of capacity based on the actual number of incoming demands is more competitive.

Figure 18 Costs of understocking or overstocking.



Note. The average daily costs are based on the weekdays of the test set because the deployed vehicle capacity on weekend days is unknown. The total daily aggregated flow is used.

It should be noted that this is a somewhat skewed comparison. The forecast model and the associated required transport capacity are based on the situation that no demand is known yet. In contrast, the actual deployed capacity is partly based on the observed demand. What should actually be compared with the starting situation at least one full day in advance, on which the largest part of the demand is not yet known because the minimum lead time of the forecast model is also one day. However, unfortunately, this information is not available. Despite the skewed comparison, we chose to perform the analysis because it was a request from the company.

6.3 Conclusion

In this chapter, we have answered the last research question: How much transport capacity does the domestic planning team at the Branch Transport s'-Heerenberg require on a daily basis? The required daily transport capacity of the domestic planning is calculated using the newsboy problem, which calculates the optimal ordered transport capacity at Truck & Drivers based on the trade-off between the cost of understocking and overstocking of vehicles. By comparing the average daily costs of under and overstocking, we conclude that the developed model (double seasonal Holt-Winters) outperforms the reconstructed current forecast model across all plausible vehicle capacity utilization rates. Furthermore, the advised transport capacity, based on the forecasting model, is not competitive with the actual deployed capacity. The added value of the model lies not in predicting the demand when part of the demand is already confirmed but in the better and timely estimation of the required capacity so that domestic planning does not have to upscale and downscale much at the last minute.

7. Conclusions and recommendations

This chapter provides the conclusion and recommendations of our research conducted to solve the main problem of the domestic planning of the Transport Branch in 's-Heerenberg. Section 7.1 states the main conclusions of the research, and Section 7.2 addresses some discussion points. Section 7.3 provides our recommendations and suggestions for future research.

7.1 Conclusion

This research aims to solve the lack of reliable predictive information on the domestic planning of the Branch Transport 's-Heerenberg. We argued that predictive information contributes to solving several problems, such as vehicle capacity allocation and the inefficient routing process. However, these do not all benefit from the same type of predictive information. Looking at the collective interest of the stakeholders, we decided to focus on a forecast that can estimate the required domestic transport capacity daily. In the remainder of this section, we first give our conclusion regarding the current forecasting method. Then, we propose some forecasting methods based on the demand characteristics. Next, we evaluate and compare the forecasting performance of the proposed forecasting models and we close with our conclusion on the capacity predictions.

The current forecasting approach cannot be evaluated because the used forecasting unit is inconsistent, and the estimation procedure is not defined. We needed to reconstruct the current forecasting approach into an actual forecasting model in order to determine the forecast performance. The evaluation of the reconstructed current forecasting model showed that the model underestimates the actual demand by 10.9%. In addition, the mean absolute error of the forecasting is 15.1%, suggesting that the forecast performance is inadequate. Note that an improvement was already made during the reconstruction since the current model could not be evaluated. Thus, the forecast performance showed that even the reconstructed model is inadequate to forecast the demand of domestic planning. Therefore, we proposed and evaluated some alternative forecasting methods.

Suitable forecasting methods are determined based on the characteristics of the demand. The data analysis showed that within-week seasonality (Sunday, Monday, …, Saturday) dominates the demand, but the within-year seasonality (week numbers) also provides additional information about the demand. In addition, national holidays also significantly affect the demand on the day itself and its surrounding days. Considering these demand characteristics, we concluded that exponential smoothing and seasonal ARIMA might be appropriate methods to forecast demand. Demand on days affected by national holidays, called special days, requires an alternative approach. We concluded that rule-based forecasting is appropriate to forecast the demand on special days.

The forecasting models are evaluated for one-up to and including 14-step ahead lead times using the mean absolute deviation (MAD) and the root mean squared error (RMSE). The out-of-sample evaluation reveals that the reconstructed current forecasting model is outperformed by Holt-Winters and double seasonal Holt-Winters across all lead times, while Holt-Winters is outperformed by the double seasonal Holt-Winters model across all lead times and flows. So, double seasonal Holt-winter has the best forecast performance if special days are not treated separately.

In line with our expectations of the data analysis, we conclude that the double seasonal Holt-Winters cannot produce reasonable forecast for special days. Using a separate rule that specifies the demand on special days approximately halves the MAD and MSE for the time series, except for the direct flow where the MAD and MSE decreased by approximately one-third. Rule-based forecasting combines the separate rule for special days with double seasonal Holt-Winters for normal days. Rule-based forecasting outperforms earlier double seasonal Holt-Winters, where no distinction is made between normal and special days. Like the original double seasonal Holt-Winters, the Rule-based double seasonal Holt-Winter (with two small adjustments) has a constant forecasting performance across all lead times for all flows. This constant performance suggests that this model may also be constant for longer lead times. Due to the constant performance, forecasts with longer lead times have a higher value than forecasts with shorter lead times. As a result, the forecast for two weeks ahead has the most value for domestic planning because it has the same accuracy as for one day ahead.

Comparing this final model with the reconstructed current forecasting model, we determine that a professional forecast model increases the performance of the demand forecast by 44.2% in terms of mean absolute error. In addition, the width of the prediction intervals with probability coverage of 0.75, 0.80, 0.90, and 0.95 are also decreased by 44.2%. Finally, we conclude that the final model can still be improved, especially the effect of the holidays, as there is still information left in the forecast errors.

The forecast is converted into an optimal ordered transport capacity at Truck & Drivers by using the newsboy problem. This method calculated the optimal ordered capacity based on trade-off between the vehicle under and overstocking costs. We conclude that the added value of the model lies not in the predicting the demand when part of demand is already be confirmed, but in the better and timely estimation of the required capacity, so that domestic planning does not have to upscale and downscale much at the last minute.

The forecast is converted into an optimal ordered transport capacity at Truck & Drivers using the newsboy problem. This method calculated the optimal ordered capacity based on a trade-off between the vehicle under and overstocking costs. We conclude that the added value of the model lies not in predicting the demand when part of the demand is already confirmed, but in the better and timely estimation of the required capacity, so that domestic planning does not have to upscale and downscale much at the last minute.

Regarding insight for science, the research confirmed the statement made by Li, Rose, & Hensher in 2010). They stated that simple methods, such as exponential smoothing, are appropriate when the trend-cycle is the dominant component in the data and sophisticated models, such as ARIMA, when the randomness is dominant. We proved that exponential smoothing outperforms ARIMA when the trend-cycle is not dominant in the data, such as in the pick-up and drop-off flow. In contrast, ARIMA outperforms exponential smoothing when the trend-cycle is dominant, such as in the direct flow.

7.2 Discussion points

This research is accompanied by several assumptions and decisions that has led to a set of limitations and points of discussion.

The research only considers statistical methods because Mainfreight's maturity regarding forecasting is still in its early stages. We argued that simplicity is preferred over complexity. Therefore, we decided to evaluate statistical methods over artificial methods in order to investigate whether they can provide satisfactory results.

Secondly, the dataset obtained from Mainfreight is significantly contaminated. In order to use this dataset, multiple steps needed to be taken to clean and transform the data to the appropriate format. The cleaning and transforming processes have been carefully performed, but not all contaminating elements may have been filtered from the data. We suspect that the residual contamination is negligible because we compared our dataset with a dataset based on trip lists. However, this suspicion cannot be established with certainty because the loading meters in the trip list database are not calculated in the same way as those for our database.

Next, in the initialization and estimation procedure of the smoothing parameters, we set a minimum bound on the level smoothing parameter to prevent the parameter of selection zero as the optimum value. Despite this minimum bound, the seasonal smoothing parameters tend to capture a part of the trend through the update. We suspect this also contributes to the undulating pattern in the forecasting errors in Section 5.4. The cause of this behaviour of the seasonal smoothing parameters is the effect of the national holidays. The results show that national holidays significantly affect more days than just the day before and after the national holiday.

Lastly, the analysis regarding the question the forecast can better predict the required capacity than what has actually been deployed is not fair. The forecast model and the associated required transport capacity are based on the situation that no demand is known yet. In contrast, the actual deployed capacity is partly based on the observed demand. In addition, the minimum lead time of the forecasting model is one full day, while the actually deployed data is based on information less than one day ahead.

7.3 Recommendations

Based on this research, we provide several recommendations for the domestic plannings department at the Branch Transport 's-Heerenberg of Mainfreight. The recommendations consist of advice for domestic planning based on this research and suggestions for further research.

We start by discussing the data availability and reliability. The significantly contaminated data is a major obstacle to easily implementing the forecast. At the shipment level, there is only a database that contains all shipments executed by the Transport Branch. This database does not indicate which department was involved in transporting the shipment. As a result, it is impossible to determine straightforwardly whether the domestic was involved in the shipment and was, therefore, part of the demand of the domestic planning. In addition, the branch uses multiple rules to convert the shipment size into units such as gross weight, cubage, and loading meter. We advise domestic planning and the Transport Branch in 's-Heerenberg to construct or modify an existing database so that the number of loading meters performed by the domestic planning per day can be directly derived. It is also convenient to use consistent conversion rules in the various databases. The available data regarding the collective vehicle capacity efficiency can also be improved. In order to keep track of the vehicle utilization, we recommend domestic planning to set up a KPI regarding transport capacity utilization rate.

Based on the conclusion of this research, we recommend domestic planning to employ rule-based forecasting with the two adjustments, as described in Section 5.3, as an indicator of the expected demand instead of the current forecasting method. The rule-based forecasting model reduces the under-forecasting tendency of the current model, and it increases the forecast performance by 44.2% in terms of mean absolute error compared to the current model.

To increase the forecasting performance and to better understand the effect of national holidays on their surrounding days, we advise further research on the influence of anomalous demand caused by national holidays. The special rule used to forecast the demand on special days has already significantly increased the forecasting performance, but the forecast errors show that it is possible to improve the performance. Furthermore, we suspect that the national holidays significantly affect more days than just the day before and after the national holiday. Thus, there are also opportunities for higher forecasting performance in that area.

We do not recommend domestic planning to follow the forecast blindly if a part of the demand is already known. However, we advise domestic planning to take the forecast model with the associated capacity as a starting point if the demand is unknown. The added value of the model lies in a better and timely estimation of the required capacity so that domestic planning does not have to upscale and downscale much at the last minute. The forecast enables domestic planning to give an indication of the expected demand which can be used to arrive at a better distribution of the collective vehicle capacity between the two planning parties.

The forecast for two weeks ahead has the most added value for domestic planning because it has the same accuracy as for one day ahead. This constant performance suggests that this model may also be relatively constant for longer lead times. Further research is needed to confirm this suggestion. Suppose domestic planning would like to use the forecast as an operationally assistant of the planning team during the routing. In that case, we suggest further research in forecasting models that can incorporate live data, such as already received demand information.

Lastly, we advise the Transport Branch to extend the research to international planning in 's-Heerenberg. Mainfreight and other Transport companies with little forecasting experience as well can use this paper as a guideline to introduce predictive information in the company, especially forecasting information. The forecasting methods proposed for domestic planning are also appropriate for international planning. Those methods can be initialized, parameterized, and compared according to the procedures described in this research. As the ultimate goal, we recommend that the site integrates the forecast from domestic and international planning to arrive at the optimal collective vehicle allocation.

References

[1]. Arora, S., & Taylor, J. W. (2013). Short-term forecasting of anomalous load using rule-based triple seasonal methods. *IEEE Transactions on Power Systems, 28*(3), 3235–3242. https://doi.org/10.1109/TPWRS.2013.2252929

[2]. Beaumont, C., Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1984). Forecasting: Methods and Applications. *The Journal of the Operational Research Society*, *35*(1), 79. https://doi.org/10.2307/2581936

[3]. Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (Third). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-29854-2

[4]. Cancelo, J. R., Espasa, A., & Grafe, R. (2008). Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *International Journal of Forecasting*, *24*(4), 588–602. https://doi.org/10.1016/j.ijforecast.2008.07.005

[5]. Ghalehkhondabi, I., Ardjmand, E., Weckman, G. R., & Young, W. A. (2017). An overview of energy demand forecasting methods published in 2005–2015. *Energy Systems*, *8*(2), 411–447. https://doi.org/10.1007/S12667-016-0203-Y

[6]. Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/J.IJFORECAST.2003.09.015

[7]. Hyndman, R. J., & Athanasopoulos, G. (n.d.). *Forecasting : principles and practice*.

[8]. Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). *A state space framework for automatic forecasting using exponential smoothing methods*. *International Journal of Forecasting* (Vol. 18). Retrieved from www.elsevier.com/locate/ijforecast

[9]. Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). *Forecasting with Exponential Smoothing*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71918-2

[10]. Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/J.IJFORECAST.2015.12.003

[11]. Klimberg, R. K., Sillup, G. P., Boyle, K. J., & Tavva, V. (2010). Forecasting performance measures - What are their practical meaning? *Advances in Business and Management Forecasting*, *7*, 137–147. https://doi.org/10.1108/S1477-4070(2010)0000007012

[12]. Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2021). On the selection of forecasting accuracy measures. *Article in Journal of the Operational Research Society*. https://doi.org/10.1080/01605682.2021.1892464

[13]. Li, Z., Rose, J. M., & Hensher, D. A. (2010). Forecasting automobile petrol demand in Australia: An evaluation of empirical models. *Transportation Research Part A: Policy and Practice*, *44*(1), 16–38. https://doi.org/10.1016/J.TRA.2009.09.003

[14]. Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, *9*(4), 527–529. https://doi.org/10.1016/0169-2070(93)90079-3

[15]. Profillidis, V. A., & Botzoris, G. N. (2018). *Modeling of transport demand: Analyzing, calculating, and forecasting transport demand*. *Modeling of Transport Demand: Analyzing,*

Calculating, and Forecasting Transport Demand. Elsevier. https://doi.org/10.1016/C2016-0-00793-3

[16]. Smith, M. (2000). *Modeling and Short-Term Forecasting of New South Wales Electricity System Load*. *Source: Journal of Business & Economic Statistics* (Vol. 18).

[17]. Stock, J. H., & Watson, M. W. (2019). *Introduction to econometrics* (Fourth). Pearson.

[18]. Sunil Chopra. (2019a). *Supply Chain Management: Strategy, Planning, and Operation*.

[19]. Sunil Chopra. (2019b). *Supply Chain Management: Strategy, Planning, and Operation* (Seventh). Pearson.

[20]. Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, *54*(8), 799–805. https://doi.org/10.1057/palgrave.jors.2601589

[21]. Taylor, James W. (2012). Short-term load forecasting with exponentially weighted methods. *IEEE Transactions on Power Systems*, *27*(1), 458–464. https://doi.org/10.1109/TPWRS.2011.2161780

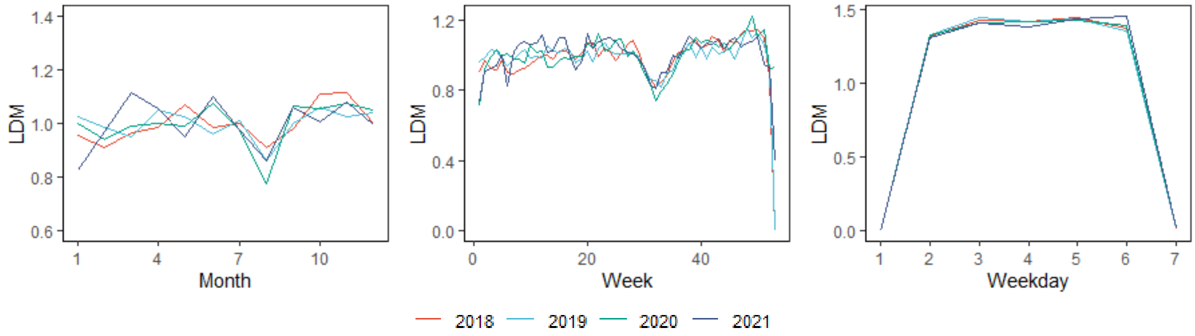
[22]. Verma, Y. (2021, August 24). Why Decompose a Time Series, and How? Retrieved May 20, 2022, from https://analyticsindiamag.com/why-decompose-a-time-series-and-how/

[23]. Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, *13*(3), 335–364. https://doi.org/10.1007/s10618-005-0039-x

Appendices

A.1 Figures and tables

Figure 19 Season plots for three potential seasonal cycles with modified for special days.



Note. The seasonal plots are modified for special days according to the adjustment 2 explained in Section 5.3.

Figure 20 Season plots of the pick-up, drop-off and direct flow.

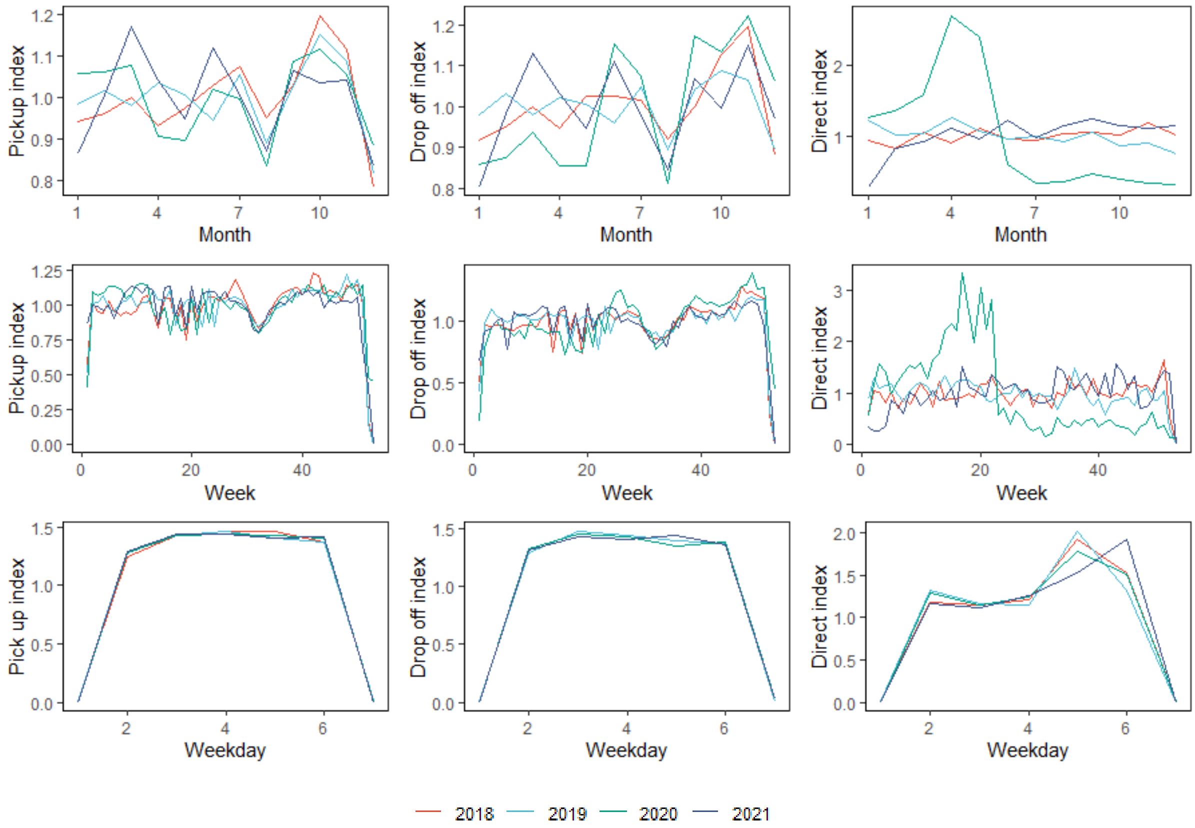
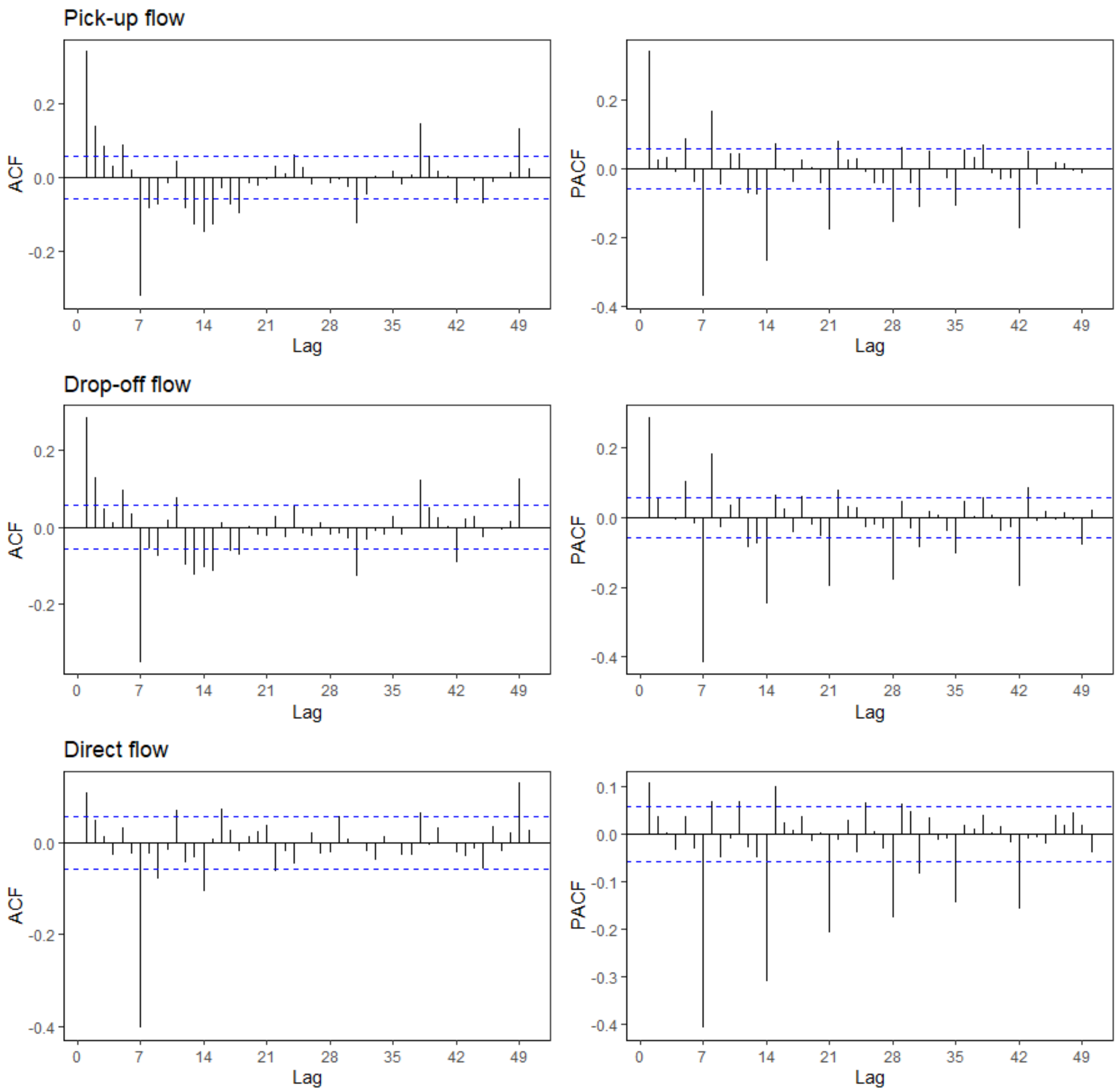


Figure 21 ACF and PACF plot for pick-up, drop-off, and direct flow.



Note. The trainings data sets used in this figure are already Box-Cox transformed and the first seasonal difference with a period 7 are already taken.

Table 7 National holidays

Name national holiday	Date	Weekday
New Year's Day	1 January	-
Good Friday	-	Friday
Easter Day	-	Sunday
Easter Monday	-	Monday
King's Birthday	27 April	-
National Remembrance Day	4 May	-
Liberation Day	5 May	-
Ascension Day	-	Thursday
Pentecost Sunday	-	Sunday
Whit Monday	-	Monday
Christmas Day	25 December	-
St. Stephen's Day	26 December	-
New Year's Eve	31 December	-

Note. National holidays with a data occur on the same date each year and national holidays with a weekday occur on the same day of the week.

Figure 22 Comparison of methods using RMSE

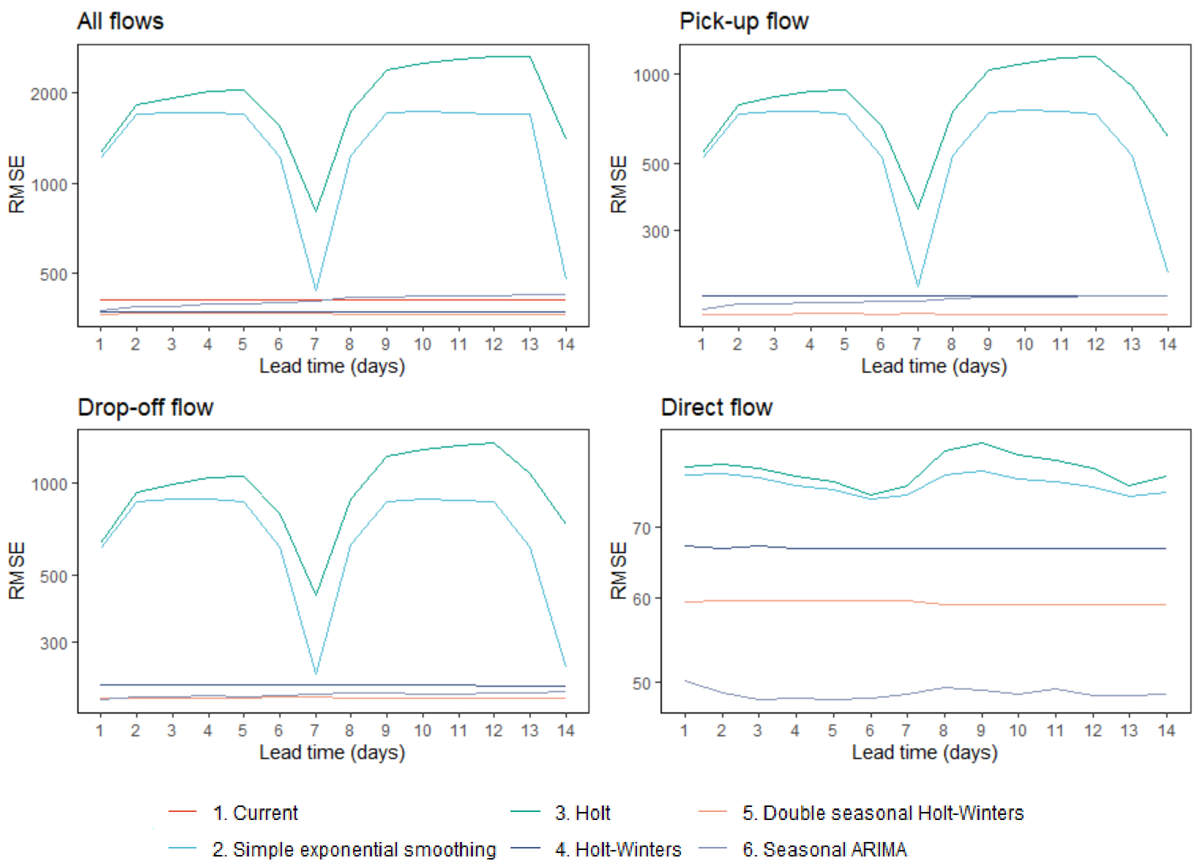


Figure 23 Comparison of Rule-based forecasting adjustments using RMSE

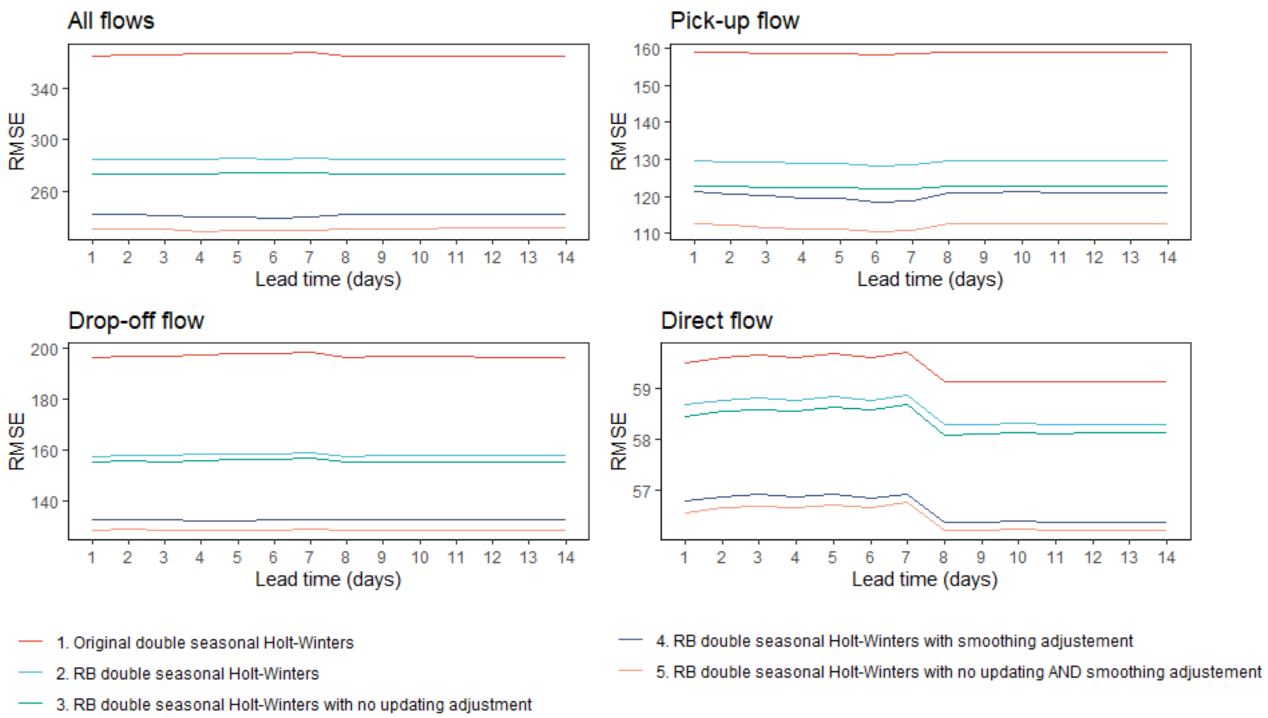


Figure 24 Histogram of forecast errors for the four-time series

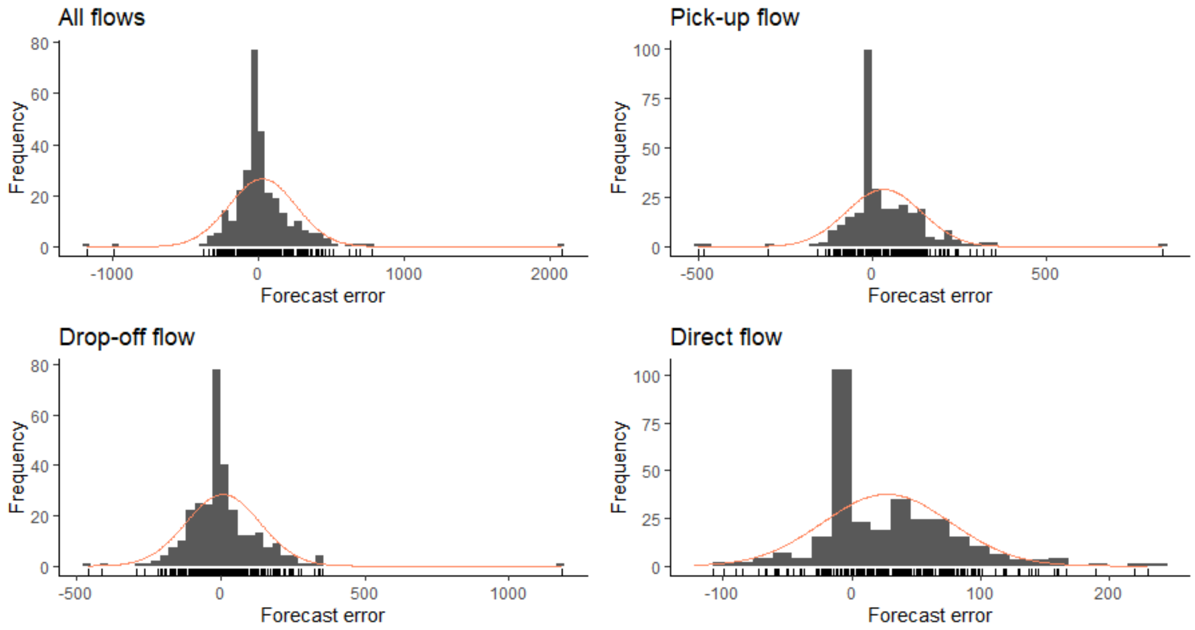


Table 8 Overview of domestic planning's standard fleet of vehicles.

Type of vehicle	Available vehicles	Loading meters capacity
Trailer	74	13.6
Combi	37	14.4
Lf	24	7.2
Sprinter	5	1.5
Total	140	3,439.00

A.2 Data cleaning and transforming

Data cleaning is the process of preparing data for analysis by removing or modifying incorrect, corrupt, or inaccurate observations. Transformation is the process of converting data into a convenient format. The data is cleaned with the following steps:

Step 1: Remove duplicates and irrelevant observations: 6 filters
Step 2: Fix structural errors
Step 3: Filter unwanted outliers
Step 4: Handle missing data

Step 1: Remove duplicates and irrelevant observations.

The TMS database consists of all shipments executed by the 's-Heerenberg Transport Branch. Chapter 1 states that the forecast is developed for the domestic planning department, which is mainly responsible for domestic transport and therefore we have to extract the domestic shipments from the database. Domestic shipments are extracted from the TMS database through the following six filters, depicted in figure 1.

1. Waybill or pick-up request

   A domestic waybill and pick-up request indicate that the shipment is linked to domestic planning. Shipments without waybill or pick-up requests are removed from the data.

2. DPD

   DPD shipments are small shipments that DPD transports. These shipments have no distribution costs, and they have no load on the distribution device of domestic planning.

3. Standard-pick up

   Domestic planning has agreements with several customers that they reserve a certain amount of space in the trucks on certain days as standard. These types of pick-ups are called standard pick-ups. Because the number of shipments to pick up is often not known in advance, a fictitious number of shipments is placed on the pick-up request. In theory, this fictitious pick-up request should be replaced with a pick-up request with the exact number of shipments after the pick-up. Due to a system error, this does not always happen, and with this filter, these fictitious shipments are removed.

4. Dummy

   Dummy shipments are used if the shipment has not yet been officially booked, but it is known that the shipment is coming. These dummy shipments are fictitious shipments that are later replaced by official bookings. As with standard pick-ups, these fictitious shipments are not always automatically deleted due to a system error.

5. Tilburg and Genk

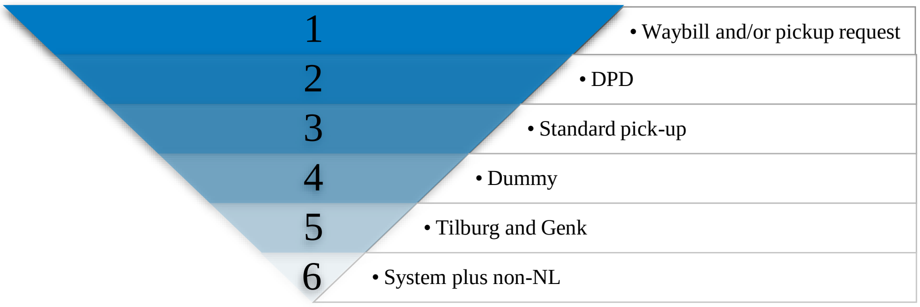
   Most international shipments from and to the delivery area of Tilburg are transported by line hauls to 's-Heerenberg because a lot of volume passes through 's-Heerenberg. The transport branch in Tilburg has recently been opened, and as a result, among other things, little volume passes through the branch for international line hauls. The shipments in the delivery area are transported via a linehaul from 's-Heerenberg to Tilburg. In addition, Tilburg has line hauls with several Belgian branches. The linehaul between 's-Heerenberg and Tilburg does not burden the domestic planning; therefore, these shipments will be removed from the data.

Due to the location, the transport branch Genk is responsible for transport in South Limburg; therefore, these shipments are also removed from the data.

6. System plus non-NL

   System plus shipments are special shipments transported only in the Netherlands by domestic planning.

Figure 25 Filtering process of the data



Step 2: Fix structural errors.

The given number of loading meters per shipment is sometimes incorrectly zero. Therefore, we check for every shipment whether the zero loading meters are correct by looking at the cubage and gross weight. A shipment with a gross weight of less than 30 kg rightly has a loading meter value of zero because these are boxes and not pallets and can therefore be placed on something which does not take up floor space. If a shipment has a gross weight of 30 kilograms or more, the load meter value of zero is considered incorrect. In the case of an incorrect zero value, the number of loading meters is derived from the cubage value. If the cubage value is unknown or equal to zero, the number of loading meters is derived from the gross weight. Table 9 shows the origin of the loading meters per year.

Table 9 Origin of loading meters

	Year				
Origin LDM	2017	2018	2019	2020	2021
Shipment (%)	92.19%	93.80%	94.94%	95.04%	95.54%
CMB (%)	6.81%	5.69%	4.81%	4.75%	4.29%
Gross weight (%)	0.99%	0.52%	0.25%	0.22%	0.17%

Step 3: Filter unwanted outliers.

Outliers are shipments with a negative loading meter number and shipments with more than 13,6 loading meters. Only the latter type of outliers is found in the data. Shipments with more than 13,6 loading meters are considered an outlier because a shipment can, in principle, not exceed 13.6 loading meters. However, we do not delete shipments with more than 13,6 loading meters; we adjust the number of loading meters of those shipments to 13.6 loading meters.

Step 4: Handle missing data.

A problem is the shipments of which the trip list is missing a date. All shipments contain the number of transported load meters, but some of the shipments do not have the trip list date. The missing shipments can be attributed to a year. Table 10 gives an overview of the number of loading meters that misses a trip list since the forecast focuses on loading meters for which the trip list date is missing.

Table 10 Number of loading meters without a trip list date

	LDM that miss a trip list date			
Year	Unload (%)	Load (%)	Direct (%)	All (%)
2017	8.49%	19.10%	0.69%	13.03%
2018	2.05%	8.11%	0.58%	4.71%
2019	0.45%	1.57%	2.32%	1.08%
2020	0.33%	1.39%	4.42%	0.96%
2021	0.30%	1.54%	0.59%	0.86%

The missing percentage decreases over the years, influencing the trend estimation of the time series. Removing the shipments with missing trip list dates will negatively bias the trend estimates, meaning that the parameters are underestimated. Suppose that the shipments with a missing trip list date are not correlated with the date, loading meters, transport type, region, or loading/unloading, which seems to be the case. In that case, it is possible to correct the data for this by distributing the loading meters with missing trip list dates accordingly to the distribution of the loading meters of the days of the year. This correction will only be applied to the data for 2018 to 2021 because 2017 has too many updated loading meters to correct for.

A.3 Time series composition

According to Profillidis & Botzoris (2018), a time series is composed of one or more of the following components:

- trend,
- seasonal,
- cyclical
- random.

The trend component reflects the long-term increase or decrease in the data. The trend does not have to be linear. Seasonal components exist when a series exhibits regular fluctuations based on a season, such as a week, month, quarter, or year. Seasonality is always a fixed and known period. The cyclic component exists when data shows the rise and fall that are not of a fixed period. These fluctuations are usually caused by economic conditions and are often connected to the "business cycle" (Klimberg et al., 2010). The length of cycles is longer than the length of seasonal patterns. The random component, sometimes called the remainder, noise, residual, or irregular component, includes everything that the deterministic components cannot capture.

The appropriateness of the forecasting model is related to the degree to which it can capture the deterministic components adequately. In order to remove the deterministic components, the time series can be decomposed into stationary and deterministic components. The decomposition can be mathematically expressed as follows:

$$Y_t = f(T_t, S_t, E_t) \tag{20}$$

where

Y_t is the number of loading meters at period t;
T_t is a deterministic trend-cycle or general movement component;
S_t is a deterministic seasonal component;
E_t is the irregular (stationary) component.

The functional form depends on the decomposition method. There are multiple decomposition methods in which the time series components can be combined. Sunil Chopra (2019) mentions three types of these functional forms: additive, multiplicative and mixed.

Additive means that the components are added together to give the number of loading meters at period t. A time series is additive if the time series' increasing or decreasing pattern is similar throughout the series (Verma, 2021). The mathematical function of an additive time series can be represented by:

$$Y_t = T_t + S_t + E_t \tag{21}$$

Multiplicative means that the components are multiplied together. A times series is multiplicative if the increasing or decreasing pattern is not similar throughout the series. The mathematical function of an additive time series can be represented by:

$$Y_t = T_t \cdot S_t \cdot E_t \tag{22}$$

Both trend and seasonality can be additive or multiplicative. An additive trend indicates a linear trend, and a multiplicative trend indicates a non-linear trend. An additive seasonality is appropriate if the magnitude of the seasonal fluctuations does not vary over time, and a multiplicative model is appropriate if the seasonal fluctuations increase or decrease proportionally with time. If an additive trend is combined with a multiplicative seasonality or if a multiplicative trend with an additive seasonality, then the functional form can be called mixed.

A.4 Formulas of the Exponential smoothing methods

Simple exponential smoothing uses weighted averages where the weights decrease exponentially as observations come from further in the past. The forecast at time t+1 is equal to a weighted average between the most recent observation:

$$\hat{y}_t = \alpha y_t + (1-\alpha)\hat{y}_t \tag{23}$$

where $0 \le \alpha \le 1$ is the smoothing parameter that controls the rate at which the weights decrease. A forecast with a high value of α is more responsive to recent observations, whereas a forecast with a low value of α is more stable and less responsive to recent observations.

Simple exponential smoothing can also be represented in the component form. The component representation consists of a forecasting equation $\hat{y}_{t+h}$ and a smoothing equation for each component included in the method. Simple exponential smoothing only includes the level (l_t) component, which gives the following component form:

$$\text{Forecast equation:} \quad \hat{y}_{t+h} = l_t \tag{24}$$

$$\text{Level equation:} \quad l_t = \alpha y_t + (1-\alpha) l_{t-1} \tag{25}$$

where $0 \le \alpha \le 1$ is the smoothing parameter for the level.

Holt (1957) extended the simple exponential smoothing method by incorporating a trend. Holt's model with an additive trend gives the following component form:

$$\text{Forecast equation:} \quad \hat{y}_{t+h} = l_t + hb_t \tag{26}$$

$$\text{Level equation:} \quad l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{27}$$

$$\text{Trend equation:} \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1} \tag{28}$$

where $0 \le \alpha \le 1$ is the smoothing parameter for the level and $0 \le \beta \le 1$ is the smoothing parameter for the trend.

Holt and Winters extended Holt's method by incorporating seasonality. Holt-Winters' model with an additive trend and multiplicative seasonality gives the following component form:

$$\text{Forecast equation:} \quad \hat{y}_{t+h} = (l_t + hb_t) \cdot s_{t-m} \tag{29}$$

$$\text{Level equation:} \quad l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{30}$$

$$\text{Trend equation:} \quad b_t = \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1} \tag{31}$$

$$\text{Season equation:} \quad s_t = \gamma \frac{y_t}{l_{t-1}} + (1-\gamma) s_{t-m} \tag{32}$$

where $0 \le \alpha \le 1$ is the smoothing parameter for the level, $0 \le \beta \le 1$ is the smoothing parameter for the trend, $0 \le \gamma \le 1$ is the smoothing parameter for the within-week seasonal cycle, and m is the length of the within week seasonal cycle.

A.5 Smoothing parameters of the final developed forecasting model

Table 11 Smoothing parameters of the Rule-based double seasonal Holt-Winter (with two small adjustments)

Smoothing parameters	All flows	Pick-up flow	Drop-off flow	Direct flow
α	0.0001	0.0001	0.0001	0.0001
β	0.0319	0.0135	0.0006	0.0001
γ	0.0961	0.0436	0.1162	0.0498
δ	0.1664	0.0427	0.1432	0.0000

A.6 Formulas and parameter values of the ARIMA models

$$Total\ flow: \quad (1 - 1{,}02L + -0{,}12L^2)(1 + 0.84L^7)y_t = (1 - 0{,}61L - 0{,}13L^2)(1 - 0{,}10\ L^7 - 0{,}90L^{14})\varepsilon_t \tag{33}$$

$$Pickup\ flow: \quad (1 - 0{,}85L)y_t = (1 - 0{,}44L - 0{,}20L^2)(1 - 0{,}85L^7 - 0{,}14L^{14})\varepsilon_t \tag{34}$$

$$Drop\ off\ flow: \quad (1 - 1.12L + 0{,}21L^2)(1 - 0{,}12L^7)y_t = (1 - 0{,}78L)(1 - 0{,}98L^7)\varepsilon_t \tag{35}$$

$$Direct\ flow: \quad (1 - 1.13L + 0{,}14L^2)(1 + 0{,}80L^7)y_t = (1 - 0{,}93L)(1 - 0{,}08L^7 - 0{,}82L^{14})\varepsilon_t \tag{36}$$

A.7 Smoothing out process

The specials day are smoothed out by replacing the original demand with the multiplication of the initial within-week seasonal factors and the average demand on that day of the week in the training set. The equation for the smoothing demand is mathematically expressed as follows:

$$Smoothed\ demand\colon\ d_t = I_{N_t} \cdot y_t + I_{S_t} \cdot \bar{y} * s_0 \quad (37)$$

where s_0 represents the initial within-week seasonal factors determined according to Chopra in Section 4.3 for the four flows and $\bar{y}$ is the average demand of the training dataset of the corresponding flow. The I_{N_t} and I_{S_t} are the binary indicator terms for the occurrence of a special day. Recall that the binary indicator I_{S_t} equals one if t occurs on a special day and zero otherwise, whereas I_{N_t} equals one if t occurs on a normal day and zero otherwise. Thus, on normal days the smoothed demand equals the original demand and on special days the multiplication of initial within-week seasonal factors and the average demand of the training dataset of the corresponding flow.