CHALMERS
UNIVERSITY OF TECHNOLOGY

Creating a Forecasting Model for a Volatile Environment

Evaluating the Accuracy of Different Forecasting Methods at a Container Terminal

Master's thesis at the department of Technology Management and Economics

Axel Camitz

Mattias Johansen

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS

DIVISION OF TECHNOLOGY MANAGEMENT AND ECONOMICS
CHALMERS UNIVERSITY OF TECHNOLOGY
REPORT NO. E2021:074
Gothenburg, Sweden 2021
www.chalmers.se

Report No. E2021:074

Creating a Forecasting Model for a Volatile Environment

Evaluating the Accuracy of Different Forecasting Methods at a Container Terminal

Axel Camitz
Mattias Johansen

Department of Technology Management and Economics
Division of Technology Management and Economics
Chalmers University of Technology
Gothenburg, Sweden 2021

Creating a Forecasting Model for a Volatile Environment
Evaluating the Accuracy of Different Forecasting Methods at a Container Terminal
Axel Camitz
Mattias Johansen

Supervisor: Jesper Svensson, APM Terminals
Supervisor: Carl Sjöberger, Technology Management and Economics
Examiner: Carl Sjöberger, Technology Management and Economics

Report No. E2021:074
Department of Technology Management and Economics
Division of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Creating a Forecasting Model for a Volatile Environment.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Abstract

The container terminal in the Port of Gothenburg, Skandia Container Terminal, is operated APM Terminals that is part of the Danish consortium A.P. Møller - Mærsk A/S. In the terminals hinterland operations, containers arrives and departs from the terminal by either rail-road or truck. For the rail-road, there is a sufficient system for controlling the volumes containers arriving to and departing from the terminal. No such system is in place when i comes to the truck traffic. Many container terminals in Sweden provide hauliers with slot times for when they can visit the terminal. However, APM Terminals, do not use such slot times. Instead, the haulers issue a visit code that gives them access to enter the terminal anytime during the upcoming two week. Since the visit codes are valid for such a long time, there is no sufficient indication of future volumes. To provide planners and management with predictions of future volumes, a forecast system needs to be developed. To forecast future demand in predictable environments are relatively easy and can be done by using simple models. For more volatile environments where the fluctuation of demand can seem irregular, the development of forecast methods are not as simple.

The aim of this study was to develop a forecast model that could predict the daily volumes of trucks arriving at the Skandia Container Terminal. To produce the most accurate and usable forecast system the performance of different models were compared. The best performing time series model was determined by using the measures of RMSE, MAPE, MaxAE and MaxAPE. Parallel to the time series models, two causal models were developed, one simple linear regression model and one multiple linear regression model.

The findings suggest that the time series model based on ARIMA yields a forecast with the highest predictive ability. Furthermore, it was seen that the operations of container freight has potential for digitization and more advanced method of forecasting. This study shows how more data and more advanced model could be used to potentially make models with higher predictive ability.

Keywords: Forecasting, Time series forecasting, Machine Learning, LSTM, Container Operations, Hauliers, Freight.

Acknowledgements

We would like to thank Carl Sjöberger, our supervisor at Chalmers University of Technology, for his assistance to writing this paper. We are grateful for his support and how he has guided us through periods of uncertainty and his sense making in times of confusion.

We also want to give a thank to the people at APMT that have given us the opportunity to write this paper and all support they have gave us. A special thanks to Jesper Svensson, our supervisor at APMT, for his time and knowledge. Another thank you to Henrik Holmén, Executive Manager at APMT, because he initiated the idea to project.

Lastly, we would like to give a special thanks to Ludvig Lifting for coaching and helping us through out the programming and building processes of the machine learning models

Axel Camitz & Mattias Johansen, Gothenburg, May 2021

Contents

List of Figures

List of Tables

1

Introduction

This chapter will present the essential background of the company and the problem statement. It will also include the following sections limitations, aim, delimitations and specification of the issue investigation.

1.1 Background

The port of Gothenburg is owned by the municipal company *Göteborgs Hamn AB* and is by far the largest container terminal in Sweden Göteborgshamn (2021). As owners, Göteborgs Hamn AB is responsible for the maintenance and the marketing, but also developing the port. Their mission is to strengthen businesses both locally and nationally. However, th the port is divided into three parts that are operated by three separate privately owned companies. *Skandia Container Terminal*, the container terminal in the Port of Gothenburg is operated by APM Terminals (APMT). APMT is part of the Danish consortium A.P. Møller - Mærsk A/S. They operate terminals all over the world and have since 2011 operated in the Port of Gothenburg.

APMT that operates the terminal is responsible for load and discharging the vessels that are arriving at the port. Export goods are being loaded and import goods are being discharged. They are therefore accountable for, and in charge of, the cranes, carriers and other machines in the terminal. To load and discharge a container ship, container cranes are used. The cranes are able to lift one container at a time. When discharging a container ship, the crane grabs or locks, the container with an equipped spreader in one locking point in each corner of the container called the corner castings. The container is then lifted onto the quay where a straddle carrier picks it up with its equipped spreaders just like the crane. The carrier then moves the container to a predetermined spot in the container yard where the container will stay until pickup by truck, train, or if the container for some reason needs to be moved.

1.1.1 The goods

A container used for shipping is called an intermodal container and can, as the name indicates, be used for multiple forms of transportation such as semi-trailer trucks, railroad transport, aircraft, or container ships. The intermodal container enables containerization, a system where containers can be stacked on top of each other and handled and transported efficiently due to the standardized dimension. There are different standard dimensions, but the two most common lengths of containers are 20 and 40 feet. A great advantage is that cargo can be loaded and discharged without the need for any additional handling of the cargo when changing the form of transportation. This increases the security for both humans and the cargo, and the reduced handling time allows the transportation process to be smoother and faster. In this report, intermodal containers will be referred to as containers only.

The content of the containers in the terminal is not specified. The only tag used for specifying the content is if it contains any dangerous goods or hazardous material. This type of goods can for example be chemical substances or compressed gases and liquids.

1.1.2 Planning and Forecasting at APM Terminals

At APMT, long term forecasts are produced by the finance department and the marketing department. These forecasts are considering the yearly volume and are mostly used for budgeting and the planning of purchases. Due to the low resolution of these forecasts, the daily operations can not make use of these forecasts. To plan the daily operations there are three different type of planners, *yard planner*, *rail* *planner* and *vessel planner*.

1.1.2.1 Yard Planner

The yard planner is responsible for planning the operations in the container yard. The operations involve moving, loading, and discharging containers with straddle carriers and cranes from both trucks and vessels. The yard planner is also responsible for allocating resources and making sure that the real capacity can meet the demand. The planning horizon for the yard planner is one week, and therefore a short term forecast is needed. A qualitative forecast of the capacity demand for the upcoming week is developed by the yard planner and is used as the basis for resource allocation.

1.1.2.2 Rail Planner

As the name indicates, the rail planner is responsible for planning the freight transported by railroad. Unlike trucks, the railroad traffic is having slot times, and the rail planner therefore needs to have frequent communication with different railroad companies to coordinate the operations.

Rail operations have to deal with serving customers within certain time frames. The rail operations have a demand forecast which is based upon customer orders and slot times for the arriving trains. This enables the planning process to be more optimized as sudden increases or decreases are limited. While trains can be late there is often room for re-planning which enables customers to still get a satisfactory service level.

1.1.2.3 Vessel Planner

The vessel planner is the one responsible for communicating with the vessels and creating a stowage plan over how each container is going to be positioned at the vessel. The stowage plan will be sent to the captain of the vessel who is either going to accept or reject the plan. If the stowage plan is rejected the vessel planners need to adjust it accordingly. When the stowage plan is accepted, the containers in the yard will need to be placed in a way at the yard that makes the loading of the ship to be done as quickly as possible.

1.1.3 Actors in the Shipping Process

The port is only one of the actors active in the process of shipping goods. This section will clarify the role of the other actors active when exporting or importing goods. Since this report only covers the transport of containers in and out of the port by truck, the process of rail transport will be disregarded.

1.1.3.1 Supplier

The supplier can be a manufacturer or other actor that is going to export goods. Before shipment, the supplier is packing and preparing the container for pickup by truck. The cargo in a container can be stuffed differently. An efficient stuffed container can therefore carry more cargo, which leads to lower shipping costs for the supplier.

1.1.3.2 End Customer

The end customer can be a manufacturer or wholesaler that has imported goods.

1.1.3.3 Haulage Contractor

In general terms, a haulage contractor is a business that transports goods either by railroad or by truck. However, in this report, the term haulage contractor will only refer to contractors transporting goods by truck. Trucks that are used for transporting containers can be of two different types of trucks, with or without side lifters. Trucks equipped with side lifter are able to load and unload containers from the ground without the need for any additional equipment. By using a pair of hydraulic-powered cranes mounted on the chassis, one crane on each end, the truck can load one 20 feet container, one 40 feet container, or even two 20 feet containers. The cranes can move along the chassis by hydraulic power to be of suitable length for the loaded container. Trucks without the side lifter need a reach stacker or a straddle carrier to load and unload containers.

1.1.3.4 The Port

In this report, the port refers to the Skandia Container Terminal in the Port of Gothenburg. The port is a facility that connects the sea with land by handling containers for import or export and is responsible for loading and discharging container ships.

1.1.3.5 Shipping Line

The shipping line is a company that is responsible for the transportation by sea. The transportation is made by container ships that are loaded with containers. Each ship has its own route, both regarding destinations and time of arrivals. The routes are generally one week long and the same ship is therefore striving to arrive at the port on the same day and at the same time every week. However, things can happen along the route and affect the arrival times.

1.1.3.6 Freight Forwarder

A freight forwarder is a company or person that is responsible for organizing the logistics of shipments of goods. Multiple actors are involved in the shipping process and the freight forwarder's purpose is to act as an expert in this network of actors. When shipping internationally the freight forwarder is often responsible for preparing and handling customs documents. In some cases, the freight forwarder represents the same company as the shipping line, but not necessarily.

1.1.4 The Flow of Freight

In this section, the role of the different actors in the shipping process will the outlined, both for export and import.

1.1.4.1 Export

The export process starts with the freight forwarder receiving an order from the supplier to ship a container abroad. The freight forwarder contacts the shipping line and orders transport by sea for the container. The shipping line will then inform the port about which container is going to be loaded on which ship. Simultaneously, an order for pickup of the freight at the supplier is sent to the haulage contractor that will deliver the freight to the port. For the haulier to enter the port, a visit code is needed. The visit code can be issued two weeks before the arrival of the truck and does not specify any day or time of the delivery. When the haulier has picked up the freight and delivered it to the port, straddle carriers place the container at an appropriate spot in the yard to later load it on a ship. How long a container stays in the port can vary greatly.

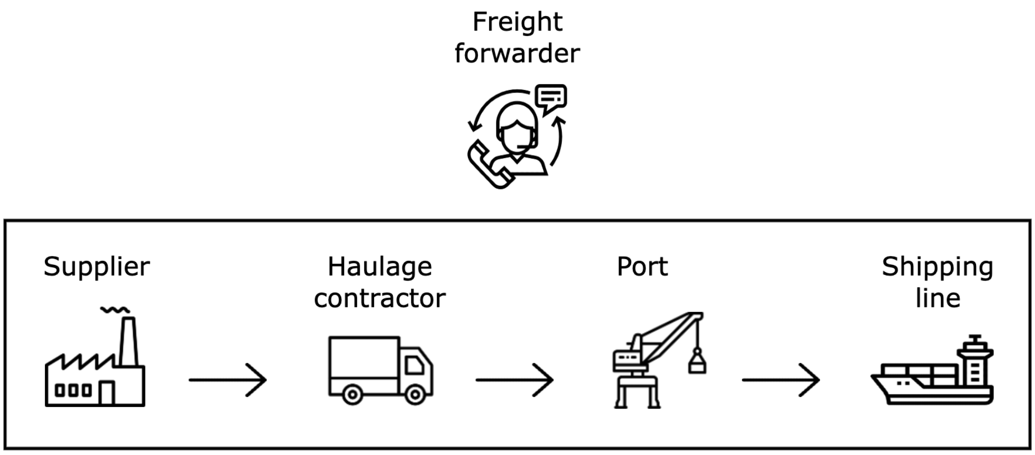

Figure 1.1: An illustration over the flow of freight for the export process

1.1.4.2 Import

The import process is similar to the export process but reverse. The freight forwarder has arranged so that the freight is shipped by the shipping line and arrives at the port where the freight is discharged and placed in the container yard. The freight forwarder then contacts the haulage contractor somewhere in between a week in advance or even when the container is already at the port and ready for pickup. As for export, the haulage contractor needs a visit code to enter the port. The haulier's time of arrival is first announced when the truck is at the gate of the port. The container is thereafter loaded onto the truck and transported to the end customer.

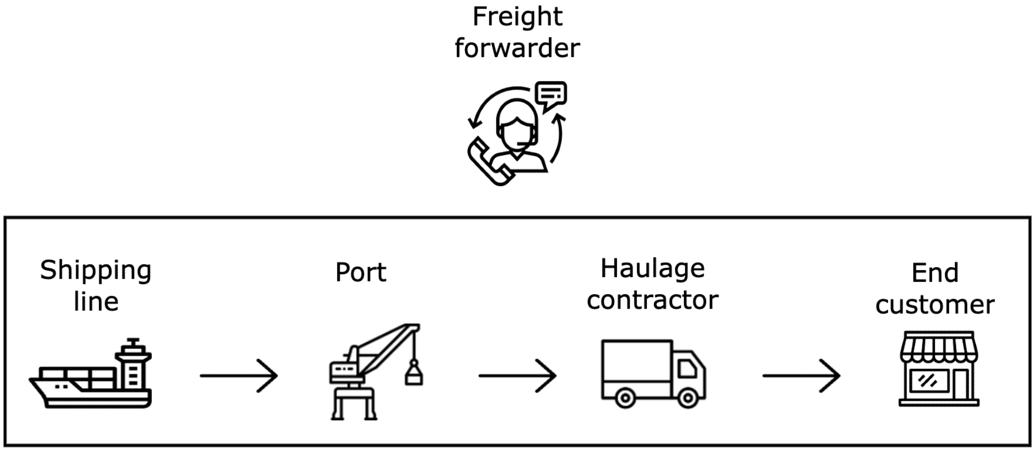

Figure 1.2: An illustration over the flow of freight for the import process

1.1.5 Hinterland Operations

Hinterland refers to "the land behind" and is a term for the area behind the coast such as the roads, railroads, or the yard of the port. Hinterland operations are associated with all handling, transportation and storage of goods in this area. APMTs hinterland operations include transporting containers within the container yard and loading and unloading them onto trucks and train carts.

A competitive advantage that APMT has in the Skandia Container Terminal is that they offer flexibility to haulage contractors in their hinterlands operations. A haulier can schedule an import or export visit in the terminal any time during the day and is free to enter during visit hours depending on when it suits the haulier. Other terminals most commonly offer time slots to hauliers in order to better control the workload of the hinterland operations (Hu et al., 2019).

While the truck drivers are offered great flexibility in the Skandia Container Terminal the truck turnaround time might vary greatly during the course of a day. In practice, this creates a volatile workload where significant ups and downs can occur. Long waiting time could suggest high utilization of resources, meaning that all available resources are in use to serve arriving hauliers. Long waiting times will also negatively impact the experience for the hauliers and are often associated with penalty fees.

The yard planner can increase the capacity for serving hauliers by allocating more resources and increasing the number of straddle carriers available during the work shift. However, it is costly to increase resources and overcapacity is an unwanted waste. Ideally, the amount of resources would be flexible to vary over time and be able to adjust when demand increases or decreases. Today, this is only possible to some extent by reallocating resources to other chores, but the set time for this is high and will therefore reduce the overall capacity in the terminal during the set time.

Rather than the forecast used today created by the yard planner, APMT wishes to have a quantitative forecasting system to estimate the volume of trucks arriving at the port in order to improve the resource planning. Since the yard planner is planning one week ahead of time, the quantitative forecast should ideally forecast one week as well. The output from the forecast should be the volume of trucks on a daily basis.

1.1.5.1 Volume of Trucks

In Figure 1.3, the fluctuation of volume can be seen for the year 2020. The y-axis represents the volume of trucks arriving at the terminal, the scale is linear but does not start in origin. The volume is of sensitive nature and will therefore not be shown in this report. However, the figure displays the variation of volume over the course of a year. It is important to consider that the graph only shows the volume during 2020, the year that the COVID-19 pandemic broke out in Europe. The pandemic could potentially have affected the pattern of both import and export, but this cannot be concluded with certainty. The graph only contains days with truck traffic in the terminal. On weekends and other red-letter days, the terminal is closed for hauliers even if the imports arrive by sea.

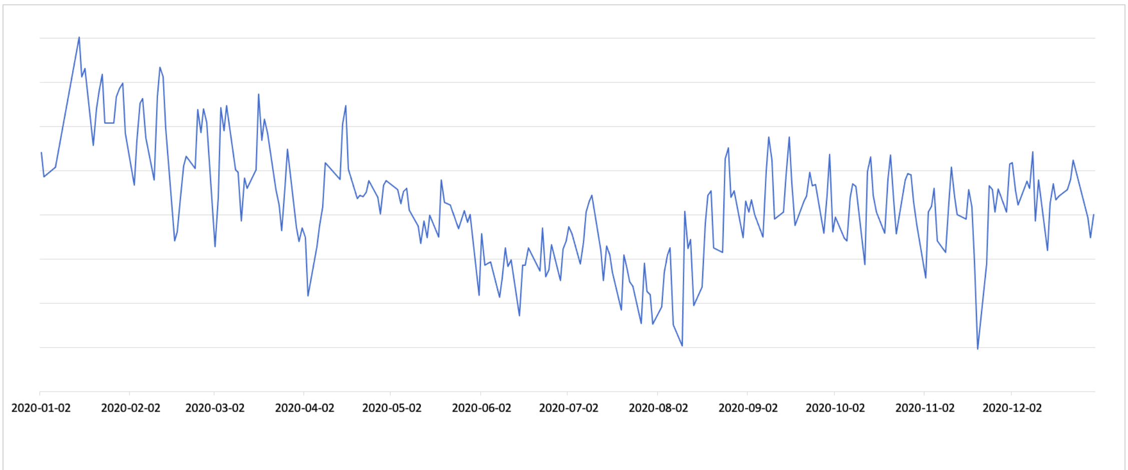

Figure 1.3: An overview of the volume during the year

The terminal is open for hauliers between 6:00 am until 8:00 pm. Figure 1.4 displays the distribution of trucks arriving on the average day of 2020. The y-axis represents the number of hauliers arriving at a certain hour presented in percentage of the total volume of trucks that day. The volume peaks between 1:00-2:00 pm and are steadily declining until closing hour.

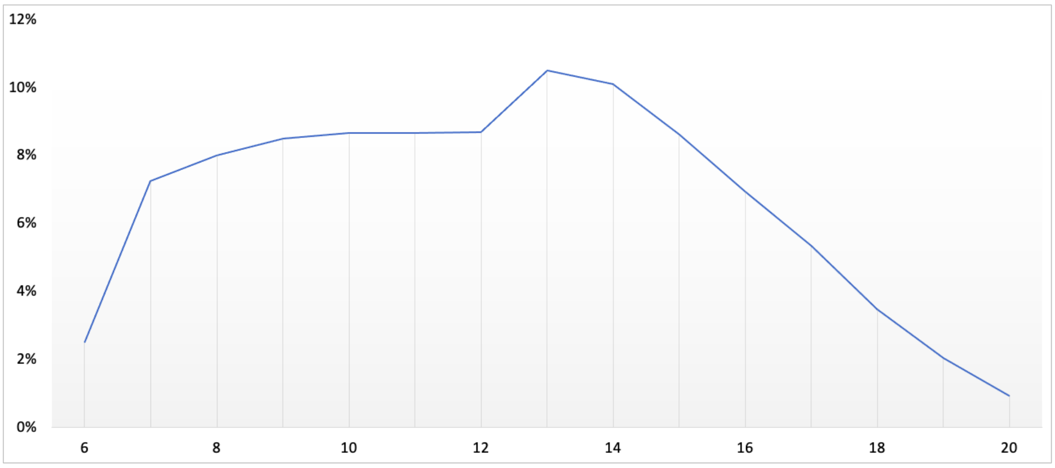

Figure 1.4: An overview of the volume during the average day

1.1.6 The haulier's arrival

In Skandia Container Terminal there are two gates to which the hauliers arrive, gate 3 and gate 4. The arrangement of the gates is similar with some small differences such as that gate 3 are handling empty containers. Both gates are holding 15 slots for trucks, five of these slots in gate 3 are allocated for trucks with empty containers.

In figure 1.5, the five slots in the *"truck grid empty"* are the ones used for trucks with empty containers.

When the hauliers arrive at the gate they enter *security in* where a security guard is checking so that the visit code and other security related codes match the container loaded on the chassis. The truck is then driven past the *OCR-portal in* where the container is automatically scanned and checked again. When the truck arrives at *gate in* it is allowed to pass if there are any free slots in the *grids*. If all slots are occupied the trucks need to wait, this can create a queue in front of *gate in*. When a slot is available in the grid the truck is driven to the slot where they unload a container for export, load a container for import, or both. Thereafter the truck is driven through the *OCR-portal out* where the containers are scanned to control that the container loaded on the chassis is correct. Then the haulier leaves the gate through *gate out*.

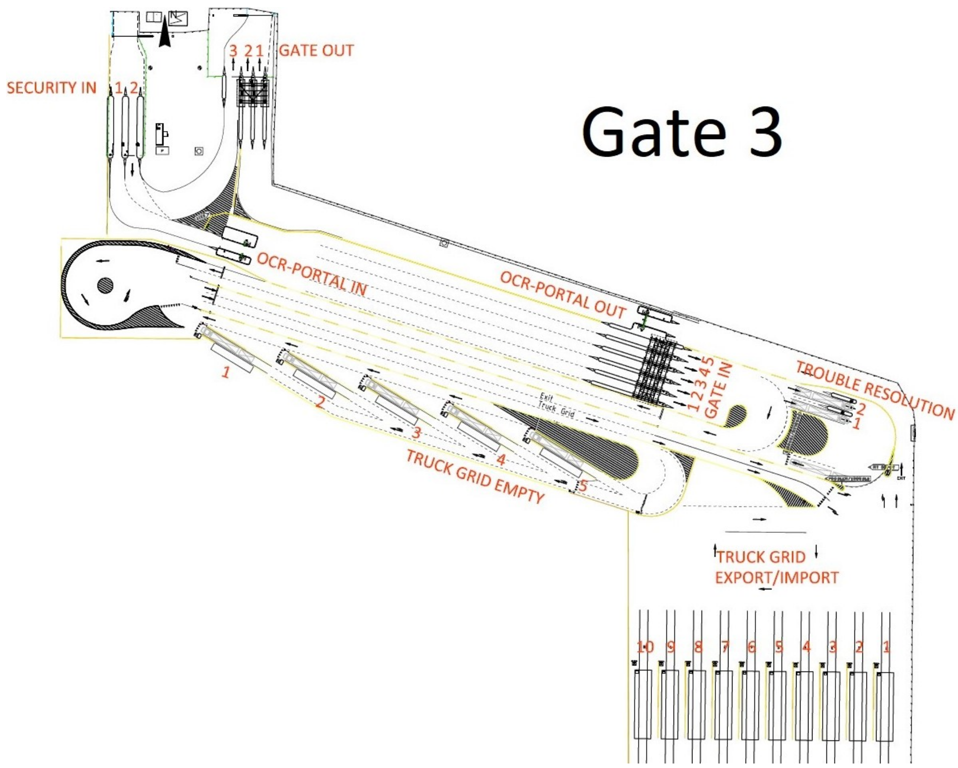

Figure 1.5: An overview of gate 3

1.2 Limitations

The greatest limitations of this study was the limited amount of available data. The limited data was impart due to the short time span in which APMT has gathered data, as well as what data is gathered.

Another limitation of this study is the author's lack of knowledge in programming and model creation. Furthermore, the lack of processing power is another limiting factors that limits the complexity of the models that can be created.

1.3 Aim

The aim of this report is to develop a forecast system for predicting demand that could be used in an environment where demand cannot be controlled or known in advance. The forecast system should provide APMT the daily volume of trucks and foresee one week ahead of time.

1.4 Delimitations

To formulate the issues of investigation below, some delimitations and boundaries for this project had to be defined.

Together with APMT the researchers decided not to implement the forecasting system but only develop it. The project was initiated without a clear plan on how the system would be integrated at APMT and how it would be used to improve operations. How the forecast would be presented was therefore not considered in this project. Furthermore, the forecast system would be developed without any changes in APMTs or any other actors operations or sharing of information.

The forecasts should only consider the upcoming week, thus a long term forecast to support the finance department and the marketing department in their predictions was not considered. The transportation by rail was completely ignored since the forecast would only predict the volume of trucks.

1.5 Specification of Issue of Investigation

- How can the volume of trucks arriving at Skandia Container Terminal be forecasted?
- How would such a forecasting system be designed?
- How accurate could such a forecast system be?

2

Theory

In this chapter, essential theories on forecasting and statistics will be provided. An overview of both qualitative and quantitative forecasts will be presented. Statistical formulas and concepts for different forecasting methods will be clarified, followed by measures for evaluating the accuracy of different forecasting methods.

2.1 Forecasting

In the planning process of resources and materials, forecasting is an integral part. By providing planners and management with a forecast of the future, uncertainty can be reduced and a more well-founded plan can be developed. In a study of the airline industry, Pölt (1998) states that a 20 percent reduction in forecast error will have a positive effect on the revenue with approximately one percent. However, it is important to have in mind that a forecast is only a number and only a basis for a decision, not a decision as such. Long term demand forecast is required for establishing long term plans while short term demand forecasts are used for monthly, weekly, or even daily operation planning to deal with more immediate concerns (Kim et al., 2016).

Mobarakeh et al. (2017) states that in environments where the nature of the demand is intermittent and sporadic, traditional forecast methods such as moving average and exponential smoothing are not always as appropriate as more sophisticated forecasting methods. Another downside with traditional methods is that nonlinear data patterns are sometimes neglected Mobarakeh et al. (2017). Suh and Ryerson (2019) states in an article about aviation demand forecasting that techniques leveraging machine learning with the focus on short term demand are more suitable for the planning of the daily operations. By using big data analysis to identify short term fluctuations, large volumes of data can be used and a more accurate forecasting model can be developed (Kim et al., 2016).

2.2 Quantitative and Qualitative Forecasting Methods

There are two main approaches when conducting a forecast, either a quantitative or a qualitative method. The quantitative methods are objective and hold the ability to process a large amount of data, which gives it a good level of consistency. However, since the forecast is based on data, the quality of the data input must be high. The qualitative methods are in contrast based on subjective judgments and few formal calculations. There are risks related to subjectivity, such as overconfidence bias and wishful thinking (Jonsson and Mattsson, 2009). Instead of basing the judgment on pure facts, the judgments can be optimistically misleading. One of the advantages of qualitative methods is that they can contemplate changes in the environment and other factors neglected in a quantitative forecast.

2.3 Quantitative Forecasting Methods

In this section, the concepts of *times series forecasting* and *causal forecasting models* will be clarified. For each concept, there are some commonly used models for forecasting that will be presented.

2.3.1 Time Series Forecasting

Time series forecasting methods use raw historical data as the foundation of the model. A fundamental difference between time series forecasting and causal forecasting models is that the former assumes that past events will repeat themselves. These events are typically seasonal trends, cycles, volumes, etc. The use of time series forecasting models is usually based on areas such as Business and Sales operations. Where it is thought that the model can help predict future trends, patterns, growth rates and irregularities. Commonly used time series forecasting models are:

- Moving Average
- Auto Regressive
- Exponential Smoothing
- Seasonal Indexation
- ARIMA (Auto Regression Integrated Moving Average)
- Box Jenkins
- State Space
- Spectral Analysis

2.3.1.1 Moving Average

The *moving average method* $MA(q)$ is a method where past data from a time series is used to forecast a future value for a certain period. The model is based around the statement that the past will represent the future Chen et al. (2010). When calculating the moving average, the formulas below are used (Jonsson and Mattsson, 2009). A_t represents the actual volume for period t, F_{t+1} represents the forecasted volume for period $t+1$ and value of n determines the number of time periods considered in the forecast.

$$F_{t+1} = \frac{A_t + A_{t-1} + ... + A_{t-n+1}}{n} \tag{2.1}$$

If n is increased, the number of historical observations included in the forecast is increased. It will result in the forecast being more stable in respect to random variation, but also react slower to systematic fluctuations.

2.3.1.2 Auto Regressive

The *auto regressive model* $AR(p)$ model is a model that requires the data to be stationary, this means that we want the distribution of data to depend only on differences in time and not on the location in time. The AR model forecast a value based on only the previous value in time is called AR(1) of order one.

$$Y_t = \omega + \phi Y_{t-1} + e_t \tag{2.2}$$

Where Y_t is the target variable, ω the intersect, Y_{t-1} is the lagged variable, ϕ is the slope coefficient and e_t is the error. Furthermore, it is possible to look further back in time to predict the future value. This can be done by going p steps back in time - AR(p), see equation 2.3.

$$Y_t = \omega + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + e_t \tag{2.3}$$

Depending on the chosen value of p the model works by regressing p steps back in time until reaching the first target value Y_{t-p}. Thus, each present value will have a small effect on the target value Y_t, which will depend on the value of ϕ. See equation 2.4

$$Y_t = \omega + \phi Y_{t-2} + e_{t-1} \tag{2.4}$$

2.3.1.3 ARIMA

The autoregressive integrated moving average (ARIMA) model is a model that combines differencing with the $AR(p)$ and $MA(q)$ model. This makes the model suitable for non-stationary data. The ARIMA model has, compared to the $AR(p)$ and $MA(q)$, three parameters that can be used to fit the model. This is denoted as $ARIMA(p, d, q)$. Where p is the order related to the $AR(p)$ model. d is the order of difference used. q is the order related to the $MA(q)$ model. The difference in the time series is referred to as the change between consecutive data points in the data set. Thus, the first order change is can be observed in equation 2.5 and the second order can be observed in equation 2.6.

$$Y_t' = Y_t - Y_{t-1}. \tag{2.5}$$

$$Y"_t = Y_t' - Y_{t-1}'. \tag{2.6}$$

The $ARIMA$ model will depend on the p, d, and q being very different. $ARIMA(0, 0, q)$ is the same as referring to the $MA(q)$ model. Likewise, referring to $ARIMA(p, 0, 0)$ is the same as $AR(p)$ model (Ho and Xie, 1998).

$$\alpha_1 X_{t-1} - ... - \alpha_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \tag{2.7}$$

$$1 - \sum_{i=1}^{p'} \alpha_i L^i X_t = 1 + \sum_{i=1}^{q} \theta_i L^i \varepsilon_t \tag{2.8}$$

2.3.1.4 SARIMA

The seasonal autoregressive integrated moving average (SARIMA) is similar to the ARIMA model but with an additional step, seasonal trends (Graves, 2020). The model is denoted as SARIMA$(p, d, q)(P, D, Q, s)$ where (p, d, q) are the ARIMA model and $(P, D, Q, sm$ represents the parameters for the seasonal model. Meaning that P and p are both representing the aggressive terms with the difference that P are seasonal. The same goes for d and D as well as q and Q. The s in the model represents the seasonal length in the data.

2.3.1.5 Exponential Smoothing Methods

Simple exponential smoothing forecasting is a univariate forecasting method that uses past data observations of the target variable Y to predict future events. The method is similar to the MA method. However, the method uses past observations and gives more recent observations an increased or decreased exponential weight. The variable X_t represents the known data at the present time. The variable Y_{t-1} represents the predicted target value from the previous step in time. α represents the smoothing factor, where $0 < \alpha < 1$. A value closer to 1 will make the model more responsive to the values closer in time. See equation 2.9

$$Y_t = \alpha X_t + (1-\alpha)Y_{t-1} + (1-\alpha)^2 Y_{t-2} + ..., \tag{2.9}$$

2.3.1.6 Double Exponential Smoothing Forecasting

Double exponential smoothing forecasting is similar to the Simple method. However, it also adds a trend variable β in the forecast. The value of β can be chosen between $0 < \beta < 1$, depending on how strong the trend component is considered to be. The initial value of b is then selected depending on the desired forecasting horizon and the initial value of X_t. The predicted value is then calculated as the value of Y_t is added by the trend value b multiplied by the desired forecasting time h.

The smoothing equation 2.10 serves to adjust the level component based on the previous trend value by adding it to the previous smoothed trend value Y_{t-1}. The second equation 2.11 serves to update the trend component by adjusting the difference of the level component at t and $t-1$ with the trend parameter β.

$$Y_t = \alpha X_t + (1-\alpha)(Y_{t-1} + b_{t-1}) \tag{2.10}$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1} \tag{2.11}$$

$$Y_{t+h} = Y_t + hb_t \tag{2.12}$$

2.3.1.7 Seasonal Exponential Smoothing Forecasting

The Holt-Winters' additive seasonal method is a seasonal exponential smoothing forecasting method that combines a smoothing forecasting method to account for both trends and seasonality. See equation 2.13. The method works by combining the the forecasting equation with three smoothing equations: ℓ_t , b_t and s_t.

Each of the smoothing equations has smoothing parameters, namely: α, β and γ. These parameters then have to be fitted to the data set where the model can be used to make predictions. Furthermore, m has to be decided pending on the frequency of seasonality and h depending on far the predictions will account for. See equation 2.14.

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \tag{2.13}$$

$$\begin{aligned} \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned} \tag{2.14}$$

2.3.2 Causal Forecasting Models

Causal forecasting models assume that the predicted variable has a cause-effect relationship with independent variables. This means that independent variables, one or more, are used as input in the model. Whereas the dependent variable is the output of the model. The independent variables can vary of sort and could be any factor that in some way is correlated with the dependent variable. A drawback of this type of model is that it must be updated if the relationship between the dependent and the independent variable changes. Some common causal models are:

- Simple Linear regression
- Multiple Linear regression
- Artificial neural networks (ANN)

The forecast period in causal models is usually medium- to long-term, therefore the models offer great flexibility. The model can be used to forecast different segments of a large set of data where different factors affect the chosen dependent variable in a specific manner. Depending on the chosen model, the cost of forecasting can be moderately low (Simple or Multiple Linear Regression) or costly (Econometric Models and ANN).

For a causal model to be suitable there must be plentiful historical data available, preferably at least two years. Furthermore, the models are rather easy to test and the implementation time ranges from weeks to quarters depending on the models' level complexity. (Zhang et al., 2015)

2.3.2.1 Simple Linear Regression

Linear regression is a statistical model which consists of a linear relationship between a dependent variable also known as the forecast variable y_t and an independent variable also known as input variables x_t. The correlation between y_t and x_t is then calculated, often with methods such as the least square method. While the simple Linear regression has predictive value it has in practice never full predictive power as the dependent variable often is affected by other unknown factors. This is illustrated by the β_0 variable, which accounts for the effects of the unknown factors. Furthermore, is the error term ε_t used to account for the forecasting error. See equation 2.15 (Späth, 2014).

For regression models to work sufficiently, a few criteria should be covered.

- The value of Y should be normally distributed for any chosen value of the variable x_t.
- For any observations in the data set they should be independent of each other.
- The mean of the dependent variable should be linearly related to the independent variable.

- The variance of the error term ε_t should not depend on the independent variable and thus, be the same regardless of when it is observed.

The benefit of linear regression is that it does not require the data set to be linear, rather the input variable can be of a different kind of data. Though, the output of the model will show a linear connection between the two variables(Späth, 2014).

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{2.15}$$

2.3.2.2 Multiple Linear Regression

Multiple Linear regression models can offer better predictive value compared to simple Linear regression models when the output of the model is affected by more than one input variable. Thus, rather than a linear function, the output of the model will be a plane that best fits the relationship between the dependent and the independent variables. The effect of each variable x_t is modified by each correlation coefficient β_n, which makes the model measure the marginal effect of the forecasting variables (Späth, 2014). See equation 2.16.

$$Y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_n x_{n,t} + \epsilon_t \tag{2.16}$$

2.3.2.3 Artificial Neural Networks Model

An *artificial neural network (ANN)* is a type of network that can be thought of as "neurons" that are organized in different structures and layers, see figure 2.1. The simple form of an ANN consists of two layers. A base layer with input variables, also known as *predictors*. A second layer that displays the output of the network. The a ANN, also known as a Feed Forward Network. It is in a sense similar to the linear regression models. Each node in the base layer is given a weight of how much it should affect the output layer. This is in a sense how a multiple regression model as there is a linear combination of the inputs that causes a certain output.

For the simple ANN to work the model has to be "trained" on a representable data set that is desired to forecast. In this context, the training of the model is referred to as analyzing how the model operates on a training data set. This is typically done by using a cost function that serves to analyze how good the model is. The end result of the "training" is then suitable weights for each node in the network.

The ANN model can be further developed by adding "hidden" layers in the network. These nodes can provide further accuracy and increase the efficiency of the model.(Batista et al., 2004)

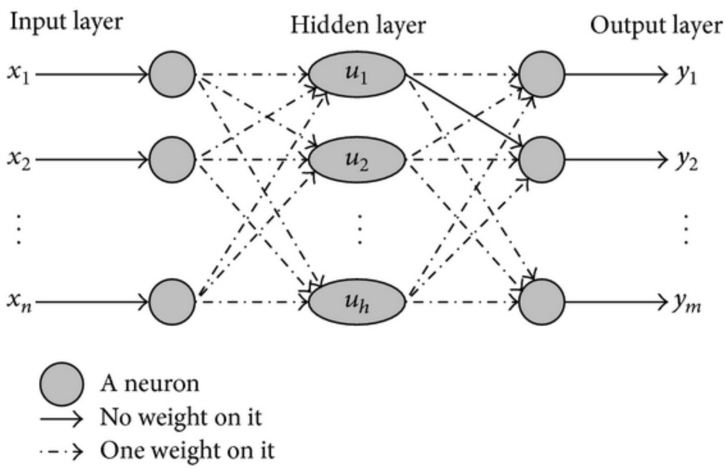

Figure 2.1: An illustration of Feed Forwards Artificial Neural networks by (Xie et al., 2014)

The research within the field of ANN is growing and further improvements in learning algorithms and more accurate methods could serve to make the single variable models yield forecasts with higher accuracy than today (Ray et al., 2020; Bui et al., 2020; Pang et al., 2020).

2.3.2.4 Long Short Term Memory

Long Short Term Memory (LSTM) is a type of Recurrent Neural Network architecture (RNN). Its main attributes is related to deep learning and is suitable for different kinds of forecasting environments. The theory behind LSTM is that in contrast to the ANN architecture it allows for backpropagation rather than just a feed forward structure like the classical ANN. This means that the network does not start from zero on each iteration the network is executed. The backpropagation allows the network to store data from the previous calculation to improve the performance of the next iteration. This type of backpropagation will however cause the model to not perform well on long- term dependencies (Sherstinsky, 2020). Thus, normal RNN models will have problems learning from specific data sets (Ouyang et al., 2017). The LSTM architecture does not have this problem. By utilizing a sigmoid function in each cell state, also known as the passing gate, the LSTM cell can keep information relevant for a longer set of iterations. Where the normal RNN architecture only contains a single layer in each cell has the LSTM architecture four different interactive layers. These layers make the model able to remove or add information based on set parameters in the passing gates. In conclusion, the LSTM is a model which can remember long term dependencies while also responding to short term fluctuations in the selection of weights to a forecasting model (Zhao et al., 2017).

2.3.2.5 Adam

Adaptive Moment Estimation (Adam) is a gradient descent optimization algorithm that enables the machine learning algorithm to be reliable, quick to learn and provide good accuracy in forecasts. The model operates by computing adaptive rates of learning for all parameters in the network. The theory behind the model is to store exponentially decaying averages of past gradients in square. The squared averages v_t can be seen in equation 2.17, where β is the decay rates which makes the model more likely to keep improving without either overflowing or converging to zero. (Bock and Weiß, 2019)

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{2.17}$$

The model also stores the non squared value of the exponentially decaying average of past gradients. See equation 2.18

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{2.18}$$

For Adam to operate as efficiently as possible Ruder (2016) suggests that β_1 and β_2 should be selected to be as close to 1 as possible. This will in turn make the model biased towards diverging to zero. Thus, this is prevented by calculating the biased estimates before the bias corrected estimates. See equation 2.20 and 2.19

$$\bar{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{2.19}$$

$$\bar{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{2.20}$$

The algorithm then works by updating the weights θ_t according to equation 2.21, where n is the learning rate. A higher learning rate will yield a method that takes larger "steps" in each epoch, thus, updating the weights of the model with a greater magnitude. (Ruder, 2016)

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{\bar{v}_t} + e}\bar{m}_t \tag{2.21}$$

2.4 Qualitative Forecasting

When conducting a qualitative forecast few formal calculations are used and subjective judgments are the basis for the forecast even though computerized support can be used (Jonsson and Mattsson, 2009). The computerized support can illustrate historical information such as the actual demand compared to the forecasted demand. These numbers can be obtained in an ERP system or other supportive systems. Using a qualitative assessment of future demand is according to Jonsson and Mattsson (2009) suitable in the situation of long term forecast and are therefore better when forecasting annual demand rather than weekly demand. Jonsson and Mattsson (2009) describes three approaches for qualitative forecasting; *the sales management*, a top-down approach; *the grassroots*, a bottom-up approach; and *pyramid forecast*, an approach that combines the other two approaches. The downside with judgemental approaches as these three is the tendency of thinking wishfully rather than realistically and therefore being over-optimistic in the forecast.

2.4.1 Sales Management Approach

The sales management approach is a top-down approach, based on the management's judgement(Jonsson and Mattsson, 2009). For management to achieve consensus about future demand, managers are gathered in one or more meetings. In these meetings, the composition of managers can vary depending on the size of the company and the relevance of insights from different functions at the meeting. The biggest advantage of the sales management approach is the speed at which a forecast can be produced.

2.4.2 Grass Root Approach

This approach intends to include all personnel in direct connection to the market by letting them separately making their own proposals for forecasts (Jonsson and Mattsson, 2009). These forecasts are then collected and processed centrally. How the processing looks differ from company to company depending on the structure of sales activities and other market connections. The result of the processing is a common forecast. One upside of this approach is that the ones responsible for realizing the forecast are also the ones responsible for producing it.

2.4.3 Pyramid Forecasting

Pyramid forecasting is an approach that combines the two above approaches by comparing both of the forecasts generated from the two approaches (Jonsson and Mattsson, 2009). The procedure for this is to see the deviation between the two and adjust them proportionally so that the total forecast is the same for both approaches. The benefit of using this combined approach is that the width of insights is greater. Generally, managers are better at taking an overall perspective while grass root personnel are usually closer to the market.

2.5 Trends and Patterns in the Quantitative Data

In this section, a review of the investigation of trends and patterns in quantitative data used in this study will be presented. It will cover *white noise* and *stationarity.*

2.5.1 White Noise

White noise refers to a statistical signal whose samples are independent with unrelated random variables. It is a time series with a mean of zero, a standard deviation that is constant with time and with zero correlation of lags (Shao, 2011). zero correlation of lags means that there is no correlation between the time series and a lagged version of the time series. If a time series has white noise, the values can not be predicted.

One can investigate if there is any white noise in a time series in several ways. One way is by making a visual test, look at the graphs and see if any obvious violation of the three criteria mentioned earlier exist. A more quantified way of testing the white noise is by performing a Ljung-Box test and checking the ACF (Boshnakov, 2021). By performing an ACF test the hypothesis of not finding any lags which is non-correlated is examined. If no lags are statistically different from zero, no correlation of lags can be assumed. A Ljung-Box test is not testing randomness at each lag, but instead tests the global randomness based on several lags.

2.5.2 Stationarity

In statistics, stationary means that all statistical properties are unchanged with time (Lindgren, 2012). A random function is called "stationary" if the statistical distribution is identical when looking at any sample of data of the same size. For a time series to be classified as stationary three conditions needs to be satisfied (Radecic, 2020). First, the mean needs to be constant; secondly, the variance needs to be constant; and thirdly, the covariance of different periods with identical distance needs to be constant. if these conditions are not satisfied the time series is classified as "non-stationary", meaning that the statistical distribution is varying with time.

A common way to test stationarity is to perform an Augmented Dickey-Fuller test, which examines the null hypothesis (Mushtaq, 2011). To test if the null hypothesis can be rejected a P-value is calculated. The P-value is based on the given information and can be produced by performing an *analysis of variance* (ANOVA) using a program for statistics such as *SPSS* or *Microsoft excel*. If the P-value is < 0.05 the null hypothesis can be rejected and the time series can be classified as stationary (Mushtaq, 2011).

2.6 Accuracy of Quantitative Forecasting Models

The most important criterion when choosing a forecasting method is the accuracy of the output, how well it represents the actual value. Therefore, estimating the error of a certain forecast method and comparing it to other methods is an essential step in the process of selecting an appropriate method for forecasting. When evaluating different forecast methods, it is of great importance to find a suitable measure to assess the forecast errors (Davydenko and Fildes, 2016). However, multiple different measures have been developed by Shcherbakov et al. (2013), but no common agreement of the most suitable accuracy metric exists (Hoover et al., 2006). In this section, methods for assessing the accuracy of forecasting methods will be presented and explained in further detail.

2.6.1 Mean Absolute Error

Calculating the *mean absolute error* (MAE) is relatively simple. The calculation involves determining the absolute difference between the actual demand (A_t) and the forecasted demand (F_t) for all data points (t) (Shcherbakov et al., 2013). The total error will be calculated by summing the magnitude of all errors. The total error will then be divided by the number of observations considered (n) to get the arithmetic value. Below, the equation for MAE can be seen.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \tag{2.22}$$

MAE is the most simple measure to use and Willmott and Matsuura (2006), among others, states that it is the most natural measure when assessing the magnitude of the average error and is unambiguous. Since MAE is scaled dependent, meaning that the resulting measurements will be presented as a relative error, all series needs to be on the same scale to be comparable. In environments where all series are on the same scale, the MAE may be preferred based on the simplicity of the method (Hyndman and Koehler, 2006).

2.6.2 Mean Absolute Percentage Error

Mean absolute percentage error (MAPE) is one of the most commonly used method for assessing the accuracy of a forecast and are used in multiple different fields and for a great range of forecasting methods (Gupta and Dhingra, 2012; He et al., 2017; Fadare, 2010). Two reasons for MAPE's popularity are that the method is scale-independent and the interpretation of the results is relatively intuitive and simple (Byrne, 2012). MAPE is defined as the equation below where n represents the number of observations considered, A_t and F_t, are the value of actual demand and the forecasted demand at observation t.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100 \tag{2.23}$$

MAPE may be one of the most popular assessment methods, however, there are some limitations and drawbacks of the method. MAPE is sensitive to outliers, meaning that irregularities or other deviations can lead to deceptive results (Tayman and Swanson, 1999; Davydenko and Fildes, 2016). In Tayman and Swanson

(1999), an article on a measure of population forecast accuracy, one can see that MAPE's tendency to overstate errors increases with a higher degree of asymmetry. Further, the results from this study also indicate that when the forecast exceeds an error of five percentage the MAPE tends to overstate the forecast errors. A common problem is that the measures can be misinterpreted when trying to average the APEs since percentage errors are affected by positive errors to a higher degree than negative errors (Tayman and Swanson, 1999; Davydenko and Fildes, 2016).

Another drawback with the use of MAPE is its inability to produce defined values when the actual demand, A_t, is zero, due to the nature of the formula. In environments where the demand constantly exceeds zero, this will not be a problem, but in environments where the demand is intermittent and occasionally drops down to zero, the value of the error will be infinite or undefined (Kim and Kim, 2016).

2.6.3 Mean Squared Error

The *mean squared error* (MSE) is a method for evaluating forecasting methods where each error $(A_t - F_t)$ is squared and then summed up. To get the arithmetic value the total square error is divided by the number of observations n Shcherbakov et al. (2013).

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (A_t - F_t)^2 (2.24)$$

Since each error is squared, the total error will be influenced in relation to its square instead of its magnitude (Willmott and Matsuura, 2006). This results in large errors having a greater impact on the total square error. Therefore, MSE are sensitive to outliers and might perform better in environments where errors are small (Hyndman and Koehler, 2006; Prayudani et al., 2019).

2.6.4 Root Mean Square Error

The *root mean squared error* (RMSE) is similar to the MSE with the only addition that the RMSE is the root square of MSE Shcherbakov et al. (2013). The formula below shows how RMSE is calculated

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (A_t - F_t)^2 (2.25)}$$

In relation to MAE, RMSE is more sensitive to outliers (Hyndman and Koehler, 2006). The difference is that RMSE does not weigh all errors equally but is instead weighing errors with large absolute values higher (Chai and Draxler, 2014). This gives the RMSE the ability to better uncover differences in performance for models (Chai and Draxler, 2014).

3

Methods

This study's research strategy was a mixed methods research that integrated quantitative and qualitative research characteristics. The research methodology used in this study was an inductive research strategy to present the relationship between theory and research by generating theory based on research (Bryman et al., 2019). The nature of the research question demanded mixed methods research. The understanding of how the volume of trucks arriving at the Skandia Container Terminal could be forecasted required an understanding of processes, information flow, and other relevant aspects and was examined with a qualitative approach. However, the creation of a forecasting system required a quantitative approach including an examination of historical data of the operations in the terminal.

A mixed methods design such as the one used in this study is called exploratory sequential design (Bryman et al., 2019). As seen in figure 3.1, in exploratory sequential design, the collection of qualitative data is foregoing the quantitative data collection. This sequencing is associated with the researchers wanting to develop hypotheses that could be explored by using quantitative research (Bryman et al., 2019).

Figure 3.1: Illustrating the exploratory sequential design by Bryman et al. (2019)

Even though the study was a mixed methods research, the exploratory sequential design was used in combination with the process of qualitative research outlined by Bryman et al. (2019), see figure 3.2. A qualitative approach was used for the first four steps, followed by a quantitative approach for the fifth step. The findings and conclusions were both qualitative and quantitative. As the exploratory sequential design suggests, the first four steps were used as preparation for the quantitative research. It would not have been possible to comprehend the relevance of the quantitative data without qualitative preparation and interpretation.

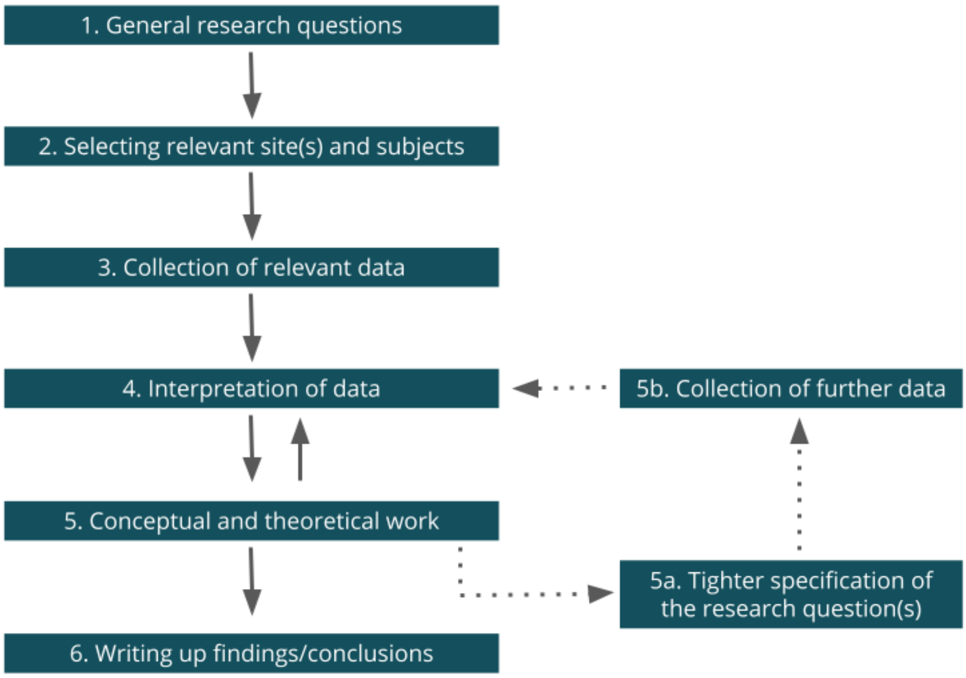

Figure 3.2: Illustrating the main steps of qualitative research by Bryman et al. (2019)

3.1 Literature Review

In the initial phase of the research, a review of existing literature on the field of forecasting and statistic was carried out. The literature review was critical for the researchers to gain an initial understanding of the field of research as well as understanding key areas that need to be considered (Hart, 2018).

The literature review followed the steps proposed by Rowley and Slack (2004), scanning existing literature, making notes, creating the structure for the literature review, and lastly writing the literature review and building the bibliography. To scan and find existing literature on the topic, platforms such as *Google scholar* and *Chalmers library* were used as well as various books by reputable authors. To find relevant literature keywords such as *demand forecasting*, *forecasting methods*, *short-term forecasting*, *time-series*, and *forecasting assessment* was used. Throughout the scanning process, notes were taken and added to a spreadsheet along with comments on the perception of the content, the relevance of the information, and its usefulness.

3.2 Collection of Data

Since a mixed methods strategy was used for this research, both qualitative and quantitative data was used. The qualitative data was mainly gathered through interviews and public material published by Göteborgs Hamn AB or APMT. The quantitative data used in this research came from the terminal operating system *Navis N4*.

3.2.1 Qualitative Data Collection

An important aspect when gathering the qualitative data was the prior knowledge of the operations in the container terminal by one of the researchers who had worked for APMT at dispatch in the Skandia Container Terminal. This prior knowledge was not considered as valid information for the research but was rather considered facilitating the process of collecting further data that was considered as valid. To map out the environment the forecasting system, interviews with people inside and outside APMT were conducted. The interviews were carried out in a semi-structured manner by using a set of general questions in a predetermined sequence (Bryman et al., 2019). By using semi-structured interviews the interviewee had the option to elaborate on subjects he or she considered important and the interviewer also had the possibility to go off the script and ask follow up questions or even change the sequence of questions to improve the flow of the interview and cover important subjects that would otherwise be missed. All semi-structured interviews were recorded and notes were taken throughout the interviews.

Since the research questions addressed aspects regarding the haulers, the perspective of the haulage contractors was chosen to investigate further. An interview with an employee at a Swedish haulage company was conducted to gain an understanding of their situation. The interview covered the haulage contractor's role in the process of shipment and the contact they had with freight forwarders, suppliers, end customers, and APMT. Interviews with employees at APMT provided valuable information about the container terminal's internal operations, the role of the terminal in the process of shipment, and the communication they had with haulage contractors, shipping lines, and freight forwarders.

Four interviews were conducted with staff from APMT, two blue collars, and two white collars. These were chosen by using theoretcial sampling approach that is a type of purposive sampling. The interviewees were considered to be the ones who could provide most insights with regards to the field of study. The interviews intended to answer questions regarding the forecast system today, how it could be improved, how accurate it needs to be for it to be useful, and the advantages of a well working forecast system.

3.2.2 Quantitative Data Collection

All quantitative data collected for this research was extracted from the terminal operating system, Navis N4. Navis N4 is a Terminal operating system used for enhancing operational efficiency in the terminal by facilitating yard management and spotting performance related issues proactively. The operating system stores a great deal of information about each container that arrives at or leaves the terminal. The historical data available for this research covered all containers arriving at the Skandia Container Terminal during the period of 30th of December 2019 until 30th of December 2020. Additional information was given about all containers shipped from the 30th of December 2019 until the middle of April. The content of this data will be clarified later in the chapter.

3.3 Data Analysis

In this section, the procedures of how the analysis of both the qualitative and the quantitative data was carried out will be described.

3.3.1 Analysis of Qualitative Data

The qualitative data was mainly used to map out the environment in which the forecasting system was supposed to be used. The coding process included transcribing interviews and structuring data collected during interviews. The information that was considered valuable for mapping out the environment of the forecasting system was organized into categories to enhance the comprehensibility of the data. The categories that was used were *internal operations* and *external operations*.

3.3.2 Analysis of Quantitative Data

To analyze and summarize the quantitative data extracted from Navis N4, a spreadsheet was used. With the data from the spreadsheet, graphs of the demand over the available time period were plotted to see if there were any apparent patterns or trends in the time series.

To see if and how the collected data was usable in a forecasting environment, a number of tests had to be performed. The tests were based on the different types of forecasting methods found in the literature review.

3.3.3 Screening and Testing of Quantitative Data

To see which forecasting method was most suitable for the collected data, four different tests were performed. The first test was to check the data for outliers and missing values. The collected data covered weekends even though no trucks are arriving at weekends. This created a lot of missing values that needed to be accounted for. To solve this, the data set were fitted to a time axis with the help of Pythons'*Date_time* library. The empty rows in the data with Nan-values, days the terminal was closed, were then filled with Pythons' *Fillna* library which propagates the last numeric value to the missing rows.

As different types of forecasting methods were to be evaluated, many different metrics were involved in the data set. Five variables were collected, with containers per day being the dependent variable. Furthermore, the four other measured independent variables were related to import, export and storage relationship, ship dwell time within the terminal per day as well total weight handled in gate 2 and gate 4. These five variables were chosen to consider since they had a relatively high correlation with the volume.

In order to transform the data into something useful, it first had to be aggregated to weekdays. Furthermore, non-numeric data such as ship variation and import and export were illustrated by assigning a numeric variable to the different states the variable could take. This means that import and export could take 3 different values, being 1 to 3 i.e. import, export, and storage. Each different vessel name within the data set was handled in the same way which led to a vast amount of different variables. Each of the different variables was then summarized in a new column, and each row could take the number of how many times the variable appears in the data set on that specific weekday. The summarized data were then converted into CSV-files for easier data integration.

3.3.3.1 Trends and Patterns in the Quantitative Data

To understand the statistical environment that is active when looking at how the volume will differ over time, some trends and patterns needed to be investigated. These trends and patterns would provide limitations and guidance as to which methods would be useful for forecasting the daily volume of trucks. Two things that were investigated were white noise and stationarity. White noise was investigated by carrying out a Ljung-Box test, and stationarity was tested with an augmented Dickey-Fuller test.

The Ljung-Box test was performed in IBM's platform for statistical analysis, SPSS. The Ljung-Box test tests the hypothesis of zero correlation of lags in the time series. In SPSS the auto-correlations were tested by inserting the data of volume of trucks each day and plotting an ACF-graph. In the generated graph, a lower and a higher confidence limit was plotted. For each lag, a staple representing the value of the coefficient for that specific lag. If the coefficients did not exceed the higher confidence limit or go below the lower confidence limit, they were considered not statistically significant.

The augmented Dickey-Fuller test was carried out by performing an ANOVA and deciding the p-value in Microsoft Excel, the choice of software is important due to the mechanism behind the different software. The input to the analysis was three columns of data, differenced volume, lagged volume, differed lagged volumes. In the list below, formulas for determining each input parameter are displayed. $V(n)$ represents the volume of the day n.

- Differenced volume: $DV(n) = V(n) - V(n-1)$
- Lagged volume: $LV(n) = V(n-2)$
- Differed lagged volumes: $DLV(n) = DV(n) - DV(n-1)$

In Microsoft Excels the data analysis tool was used, and *regression* was chosen to estimate the p-value. By using the differenced volume column as *input Y range* and the columns of both lagged volume and differed lagged volumes as *input X range* the ANOVA has generated.

3.3.3.2 Screening of Multivariate Data

To not solely rely on past volume, i.e. time series data. A number of variables were collected. These variables were analyzed in a persons correlation matrix. The test shows how the different variables correlate with each other. This test wast used to figure out which variables were of significance when building the multivariate forecasting models.

The data set was plotted to screen for outliers and values which seemed illogical or unlikely. As all multivariate data were aggregated from different files it was also compared to the original volume file to see if the daily volume variable were identical.

3.4 Forecasting models

In this section, the methods for creating the forecast models and determining the different parameters and variables used in these models will be outlined. First, the creation of time series models will be presented, then the causal models, single variable LSTM and multivariate LSTM, will be presented.

3.4.1 Creation of Time Series Models

In order to make predictions based on the collected data, it was concluded that a few different methods should be tested to see which one is most appropriate for the forecasting environment based on the established measurements.

3.4.1.1 Moving Average

The first forecasting method to test was the moving average. The moving average model is one of the simplest methods used for forecasting. Therefore, the method was chosen as a reference to how well a simple forecast model could perform in an environment such as the one studied for this research.

The input to the formula is the actual value during previous time periods and the number of time periods being considered. The actual value of historical events can not be altered, meaning that the number of periods being considered is the only parameter that can be altered. To determine the number of data points to consider for producing the highest performing model, a python code inspired by Newton's method for iteratively producing better and better approximations to the roots was used. Instead of using the root-finding algorithm, the code tried to find the lowest RMSE-value possible. The way this was done began with an initial guess and comparing it to a second guess. The n-value that generated the lowest RMSE-value becomes indicative of the next guess. This procedure was repeated until the lowest RMSE is found.

The forecast was generated by using SPSS and using all available data of volumes of containers. To carry out a moving average forecast with the optimal n-value an ARIMA(0,0,q) was used where q represents the n-value.

3.4.1.2 Exponential smoothing

There are different types of exponential smoothing methods for forecasting, and therefore both a simple exponential smoothing forecast and a double exponential smoothing method would be tested and compared. The method that performed best with regards to RMSE would later be compared to the other methods. The difference between the two methods is that the double exponential smoothing method consider trends in contrast to the simple version. By considering trends, the model will faster react to sudden changes in the volume.

To create the forecast models, SPSS was used. By using SPSS the optimal value of the smoothing factor α and the trend parameter β were automatically used by the software. If no trends can be found in the data, the trend parameter would be zero, and then produce the same forecast as the simple exponential smoothing method would.

3.4.1.3 SARIMA

As the moving average model and exponential smoothing model seemed promising it was considered to see if they would perform better combined into one model. SARIMA, also known as Seasonal Auto Regressive Integrating Moving Average, as a combination of the above mentioned methods. Similar to ARIMA it has parameters that make the different methods more or less impactful in the forecasting. The parameters which are as follows: SARIMA$(p, d, q)(P, D, Q, s)$, can make around 1300 different models depending on the different settings. To see which settings yielded the lowest error with regards to RMSE on the current data set, a Python function was made. The function was made by building a for loop with 9 layers of independent for loops. Every for loop did then run the SARIMA-model with a specific value. Then store the parameters in a data frame and the results in a vector. When the code was fully executed a new function was created to find the position of the lowest error. The SARIMA settings were then printed out by selecting the corresponding position in the data frame which stored the settings.

To verify the findings, the three best settings were tested in SPSS on the same data set containing the daily volume. The model building tool was used to build the model and the SPSS toolbox for statistics was used to evaluate the findings.

3.4.2 Single Variable LSTM

To make a model which does not only rely on fulfilling the prerequisites of the time series methods, such as stationarity etc. It was concluded that a machine learning model would be suitable.

To make a working model with the limited time and processing power available, it was concluded that the model will be based around the LSTM method. The coding of the model was performed in Python V.3,7 and the main libraries which had been used in similar projects was chosen to be Keras Tensorflow and SKlearning (Géron, 2019). LSTM was chosen to be suitable as the data set only covers one year. Thus, seasonal trends are hard to identify and it allows the model to hopefully make good predictions based on a rather short training window.

The data was imported from a CSV-file and stored into a data frame for easier manipulation. The data frame which now contained the daily volume and a corresponding date, was changed into a vector with the date being an index for each row. The high variation in the data set could pose a problem that could yield instability and training problems such as exploding gradient or decaying gradient (Batista et al., 2004). Thus, making the model unable to "learn". This was resolved by normalizing the data frame with with a $Min_Max_function$. While this causes the model to only predict values from 0 to 1, it can still provide valuable information regarding how well the predictions compare with the real normalized values. This trade-off was thought of as necessary due to a lack of time and processing power.

The normalized data frame was then reshaped to the correct tensor alignment, in order to fit the prerequisites of the Keras library (Géron, 2019). Three LSTM layers and one Dense layer were added to the model and training loss was initially selected to 0,01. These settings were changed throughout the testing phase and multiple changes were made to see which of the settings performed best. This process of selecting correct parameters were as seen in similar research projects an iterative process which required trial and error to find optimal settings (Bouktif et al., 2018; Sumi et al., 2012; Prudêncio and Ludermir, 2004). The metric to optimize for was selected to be accuracy and RMSE was chosen as the loss metric to optimize against. The output of the model was tested and by calculating MAPE between the predicted values and the real values.

3.4.3 Multivariate LSTM

To consider more parameters for the forecast than only historical volume of trucks, a multivariate model was tested. While many models fit this description, a machine learning model was especially suitable. The multivariate LSTM model is relatively easy to use and offers the option to be re-evaluated daily to constantly have an updated forecasting model. The model is also quite easy to program in such a way that the independent variables are chosen based on recent correlations. This means that the variables that the model used can be changed depending on influences such as external factors, that potentially can influence the daily volume.

The multivariate LSTM model is a type of RNN based machine learning model which works well with time series data such as the one available in the collected data. Since there is a high variation with the daily volume, it would be hard to use a simple neural network model such as one based on the feed forward method (Sun et al., 2020).

Due to the authors' limited knowledge in model creation, a simple program language was chosen. Python 3.7 was chosen due to prior knowledge and its suitable machine learning libraries. This means that the model was exclusively built in Python. The coding was based on trending machine learning libraries such as Keras, Tensorflow, and Scikit-learn.

Firstly, The desired data was imported from a CSV-file and then converted into a data frame that fitted the variables on a timeline by using the *date_time* function in python. Since the truck terminal only operates during weekdays, weekend resulted in rows with missing values. This led to a number of missing variables which were filled in by using the *Fillna*, with the forward fill setting. While this leads to a model which does not entirely predict the real volume. The model will still illustrate the concept of predicting the daily volume.

The data was then separated into a set of training data which the model will "learn" from, and a test data set on which the model will be evaluated. Furthermore, the data also normalized to range from $0 < x < 1$ to make the model easier to run with limited time and processing power.

The data sets were then imported into a multivariate LSTM model with the Keras library. To make the model both easy to run and quickly diverge to a low error. The *Adam* optimizer was chosen as suggested by many other studies with similar requirements (Chang et al., 2018; Fu et al., 2016; Sakinah et al., 2019).

To see how the model improved during training, a loss function was created to see how many training iterations were needed for the model to stop improving significantly. A suitable value based on the loss function was then selected as the number of epochs.

The model was then tested and debugged to find any obvious issues with the model.

The model then tested on the testing data. The two vectors which one were the predicted volumes, and the other one the real volume on the same date was then compared and plotted on an axis. The RMSE function from the $SK-learn$ library in python was then used to compare and calculate the RMSE on the predicted values and the real values of the daily volume.

3.5 Assessment of Forecast Models

The assessment of the forecasting models differed depending on if the model were a time series model or a causal model. The times series model was all assessed with the same measurements and comparisons between the models were made. The purpose of this was to identify how good a time series model could be. Single variable LSTM used the same measurements but with a different sample size then the time series models. For the multivariate LSTM, some of the measurements used for the other models was unsuitable, while others worked just as well.

3.5.1 Assessment of Time Series Models

The time series models generated and assessed in SPSS forecasted the volume for the next data point in the time series, meaning that the predicted value for a certain day represents what was forecasted based on the data known the day before.

To evaluate the accuracy of the different forecasts, measurements needed to be established. It was decided to use different methods for measuring the accuracy of the forecasts. By using more than one measurement, the assessment would be more robust and reliable since it would not only be primed to perform well with regards to one specific measurement. It was decided to not use any weighing or prioritization of the measurements. Instead, the performance of the forecast models was analyzed and assessed with the chosen measurements as a basis.

One of the chosen methods was the MAPE due to its simplicity and easiness to interpret (Tayman and Swanson, 1999; Davydenko and Fildes, 2016). An advantage with this measurement is that persons not familiar with methods of assessing the accuracy of forecasting methods can, to some extent, understand the values and evaluate the performance of the forecasts. The other method chosen for evaluating the accuracy was RMSE. The main reason for using this measurement was due to its sensitivity to large numbers, meaning that larger errors are more easily uncovered (Chai and Draxler, 2014).

To assess the applicability of the different forecast models, two additional measures were used, *MaxAE* and *MaxAPE*. MaxAE represents the largest error for a specific data point in the model in absolute value, and MaxAPE represents the largest error for a specific data point in the model in percentage. These measures are relevant because they will expose the weakest performance of the model and therefore enhance the understanding of how well it will perform during more volatile periods. A too high value of MaxAE or MaxAPE will reduce the applicability of the model since it will reveal the biggest shortcoming in the model. Great errors in the models would affect the operations in the terminal negatively and potentially lead to improper resource planning.

A less quantified way to assess the accuracy of the forecast models was by plotting the predicted values from the model and comparing it to the real value in a graph with the volume on the y-axis and time on the x-axis. In the next chapter, these graphs will be shown, the value of the volume will not be visible due the sensitive nature of that information. The y-axis starts at the origin, and the x-axis reaches over the time period of available data. Both axes are linear. The purpose of these graphs were that they provided indications of how well the compliance of the different models was. By looking at heavy increases and decreases in real volume and see how well the forecast model was able to predict these variations a compliancy assessment was performed. The MaxAE and MaxAPE only examine the point with the highest absolute error and the highest absolute percentage error. It can therefore be interesting to analyze the compliance of the forecast model and consider more data points than just the absolute largest error.

3.5.2 Assessment of Causal Models

To assess the causal models and compare predicted values to actual values, the predictions had to be made based on a data set not used for training the model. Meaning that the model was trained on the data from January to November and tested against the actual values in December for the single variable LSTM. For multivariate LSTM, the model was trained on the period January to March and the model was then tested for April. This allowed a much lower sample size for the assessment of the causal models

The single variable LSTM model used the same measurements, RMSE, MAPE, MaxAE, and MaxAPE, as the time series models while the multivariate LSTM model did not. The reason for this was that the multivariate LSTM model were based on gradient decent optimization to learn the model. This in turn made it necessary to normalize all values to make the model work. The normalization of data also makes the model only able to yield an output in the same range. While this in theory could possible have been resolved, the authors considered the lack of knowledge and time as valid reasons for accepting the normalized output.

With normalized values, some of the measurements would be misleading and irrelevant. The benefit with MAPE and MaxAPE is that the scale used is not relevant since they are percentage based. Therefore, MAPE and MaxAPE could be used to assess the causal models.

To measure RMSE would be misleading when comparing the multivariate LSTM model of the time series models due to the nature of the formula for calculating RMSE. MaxAE is another measurement that would be misleading to use and is therefore disregarded for the multivariate LSTM model.

The compliance will, for the causal models, be assessed in the same way as with the time series models. Since the multivariate LSTM model did not use as many measurements as the other models, more weight was put on the assessment of the visual compliance in the plotted graph of real and predicted values for the multivariate LSTM model.

3.6 Analysis of the Results

The analysis started with dividing the models into two groups, the time series models and the causal models. These two groups were analyzed separately due to the nature of the models.

The time series models, moving average, exponential smoothing and SARIMA (p,d,q)(P,D,Q,s) were analyzed by using the measurements of RMSE, MAPE, MaxAE, and MaxAPE. RMSE and MAPE measured how well the models performed overall while MaxAE and MaxAPE provided some indications of how wrong the predictions could be at their greatest. This was important to investigate since it is crucial for the prediction to not diverge from the real value too much. If the errors are too great, the forecasts will no be applicable. This is especially true if major errors recur frequently. To analyze how well the models are performing with regards to major errors the graphs of volume over time was studied. Models, where major errors occurred frequently, were considered unqualified as forecast models for APMT.

The machine learning model single variable LSTM and multivariate LSTM were also analyzed based on statistical metrics. The single variable LSTM model was analyzed using RMSE, MAPE, MaxAE, and MaxAPE. While the Multivariate model was due to its normalized values only analyzed with MAPE and MaxAPE.

The five forecast models were analyzed by comparing the performance of their predictions concerning the measurements and compliance. The measurements were compared to see how well the different models performed for the different measures. The time series model that was considered to have the best predictions over the time period available was the one considered to represent how good a time series could be in an environment such as the volume of trucks in a container terminal.

4

Results

In this chapter, the results from this research will be objectively presented. No analysis of or discussion about the data will be given. First, the results from the interviews will be presented. Then an examination of trends and patterns in the quantitative data and how these will affect the different forecasting methods will be performed. Lastly, the forecasting models will be presented and evaluated based on the decided methods for accuracy assessment.

4.1 Interviews

From the interviews with APMT, it was found that the greatest advantage of a properly functioning forecasting system would be to make resource planning more cost effective and reliable. Today, the forecasts are considered to be of qualitative nature with some support of simple computations. They are based on historical data and the pattern of seasons, the total volume of export and import are also considered. The forecasts are done one week ahead of time. These forecasts are performing with a MAPE of approximately $15-20$ %. Based on this margin of error, a model with a MAPE of 10 % would be usable and 5 % would be great.

When the interviewees were asked about potential parameters that could be used to improve the forecast, some proposals were made. The first proposal was the use of a "pre-advice" system where the hauliers send a notification earlier of which containers that are going to be delivered or picked up. Secondly, the implementation of some kind of *"First delivery date"* system was proposed, meaning that the first day a container can be delivered to the terminal before export shipment. Thirdly, it was proposed to compare the forecasts of volume for the railroad and the forecasts of total export and import volumes. Both railroad transportation's and sea traffic are having relatively well working forecast systems and therefore these can be used. By comparing the differences between these two, one can figure out the weekly volume of containers that are going to be either delivered or picked up by truck. The last proposal was to consider trends in the industry and follow an order index to see future deliveries.

All interviewees agreed that an alternative way to improve the resource utilization in the hinterland operations without using a forecasting system would be to implement, as many other container terminals have, slot times for the hauliers to arrive. If they do not arrive within the slot time a penalty fee will be placed.

4.2 Trends and Patterns in the Quantitative Data

In the section, the outcome of the Ljung-Box test and the augmented Dickey-Fuller test will be presented.

4.2.1 Ljung-Box-test

When the Ljung-Box-test was performed in SPSS, a figure illustrating the autocorrelation was generated, figure 4.1.

As can be seen in the figure below, some white noise, i.e. non predictable sections, occurs in the data set. However, if it is too much white noise to use the data to make a proper forecast can not be determined. The evidence for white noise is not strong enough to exclude the fact that values can be predicted.

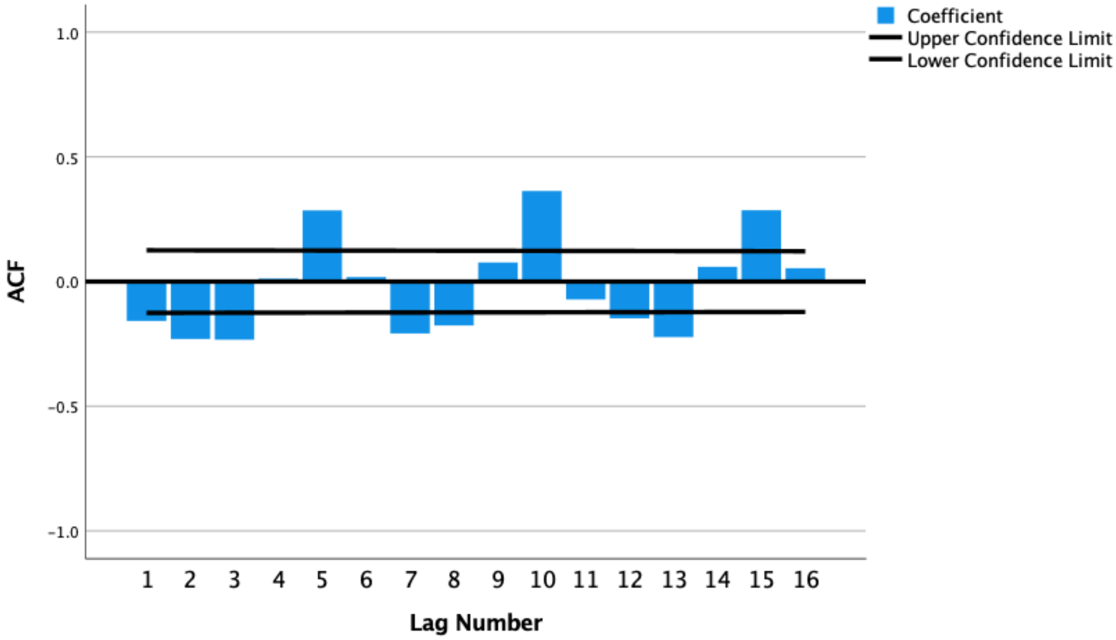

Figure 4.1: Illustrating the ACF over 16 lags

4.2.2 Augmented Dickey-fuller Test

The augmented Dickey-fuller test however clearly indicates that the time series is stationary. The p-value generated from ANOVA using Microsoft Excel is much lower than the critical value of 0.05. The p-value is estimated to be approximately $5 * 10^{-14}$. A clear indication of stationarity.

4.3 Assessment of Forecast Models

To assess the different forecast models, the established measurements will be used. Each forecast model will be evaluated as equally as possible. For moving average, exponential smoothing, and SARIMA the assessment will be identical. This is possible since the same software, SPSS, was used to create these three forecast models. To create the single variable LSTM and the multiple variable LSTM Python was used.

4.3.1 Moving Average

The first step in creating a moving average model was to determine n, see formula 4.1.

$$F_{t+1} = \frac{A_t + A_{t-1} + ... + A_{t-n+1}}{n} \tag{4.1}$$

When the Python code that optimized the n-value was executed, the best preforming model used a n-value of 47. An ARIMA(0,0,47) was executed in SPSS and the performance of the forecasting model was generated. The measurements and how well the model performed with regards to the measurements can be seen in Table 4.1.

Table 4.1: Performance of the moving average model with an n-value of 47.

Moving Average	
Measurements	Values
RMSE	73
MAPE	$5,51$ %
MaxAE	306
MaxAPE	$26,82$ %

To visually evaluate the compliance of the model a graph with the real volumes and the predicted volumes was plotted, see Figure 4.2. The graph is used to facilitate the understanding of the performance showed in table 4.1 rather than a measurement itself. The y-axis represents the volume of trucks arriving at the terminal and the x-axis represents the time where each data point is one day of trucks arriving. The axes are not numerical because the volumes are considered as sensitive information for APMT.

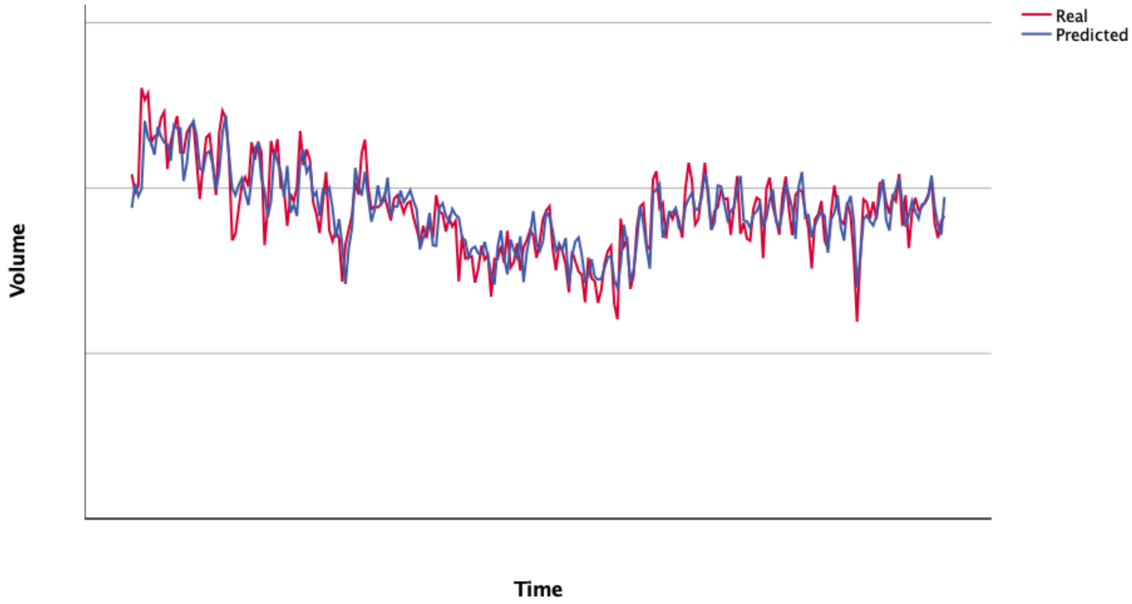

Figure 4.2: Illustrates the compliance of the moving average model.

4.3.2 Exponential Smoothing

Before assessing the exponential smoothing method, the type of exponential smoothing method to be used needed to be determined. When running both the simple and the double exponential smoothing method in SPSS only minor differences was found, the β-value was almost zero. One can therefore conclude that no trends can be considered to produce a more accurate forecast by looking at the available data. The performance of the exponential smoothing forecast can be seen in Table 4.2.

Table 4.2: Performance of the exponential smoothing model.

Exponential Smoothing	
Measurements	Values
RMSE	85
MAPE	$7,23$ %
MaxAE	298
MaxAPE	$48,05$ %

As with the moving average forecast, a graph of the real volumes and the predicted volumes was plotted, see Figure 4.3, to visually see how well the forecast model's compliancy was over time.

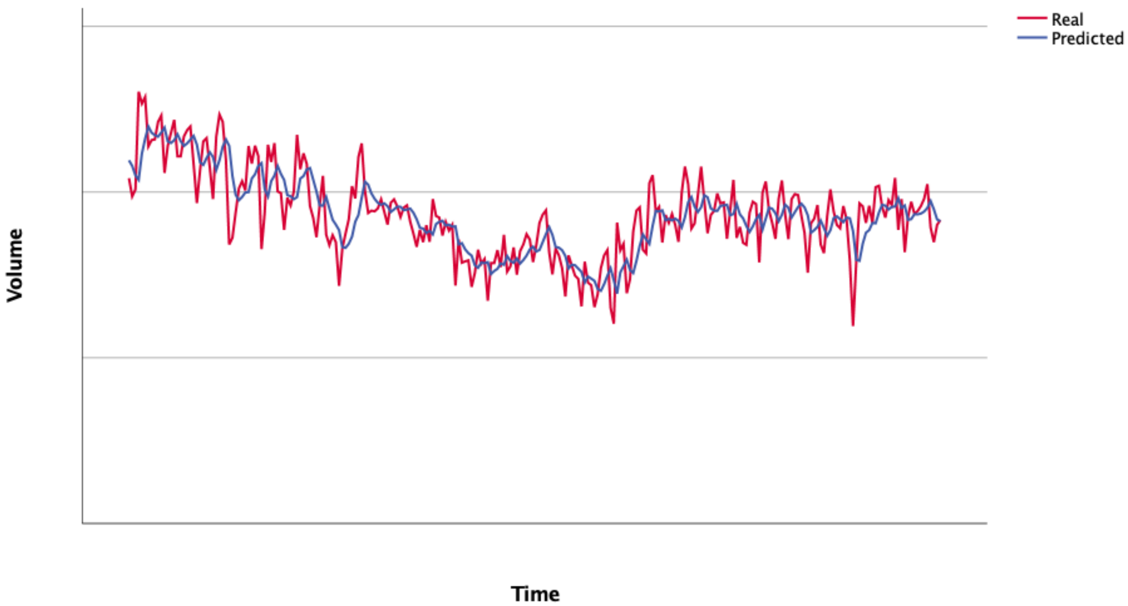

Figure 4.3: Illustrates the compliance of the exponential smoothing model.

4.3.3 SARIMA

When the Python code was run to see which settings yielded the lowest error in regards to RMSE it was found that no seasonal trends could be found to improve the forecast. The setting that yielded the lowest RMSE was SARIMA(2,1,47)(0,0,0,c), or ARIMA(2,1,47). The performance of the model can be seen in Table 4.3

Table 4.3: Performance of the ARIMA(2,1,47) model.

ARIMA(2,1,47)	
Measurements	Values
RMSE	73
MAPE	$5,50$ %
MaxAE	283
MaxAPE	$23,63$ %

In Figure 4.4 one can see the compliance of the predicted value in relation to the real value over time.

4.3.4 Single Variable LSTM

The single variables LSTM model was able to stabilize around 20 epochs and was improving until around 100 epochs, see figure 4.5. Going further than 100 epochs

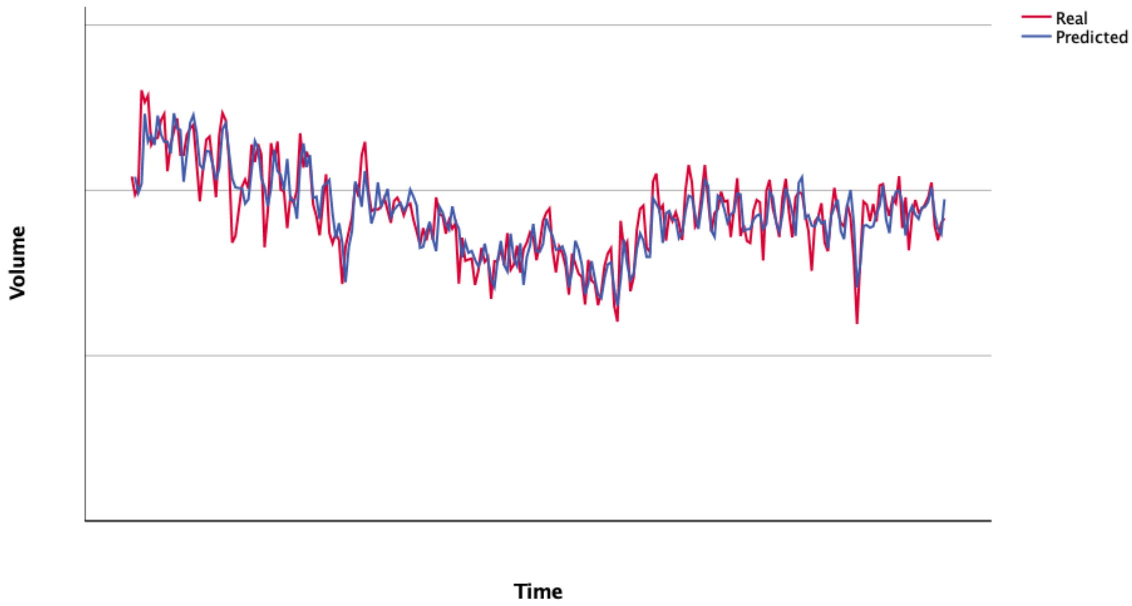

Figure 4.4: Illustrates the compliance of the ARIMA(2,1,47) model.

did not improve the overall performance of the model.

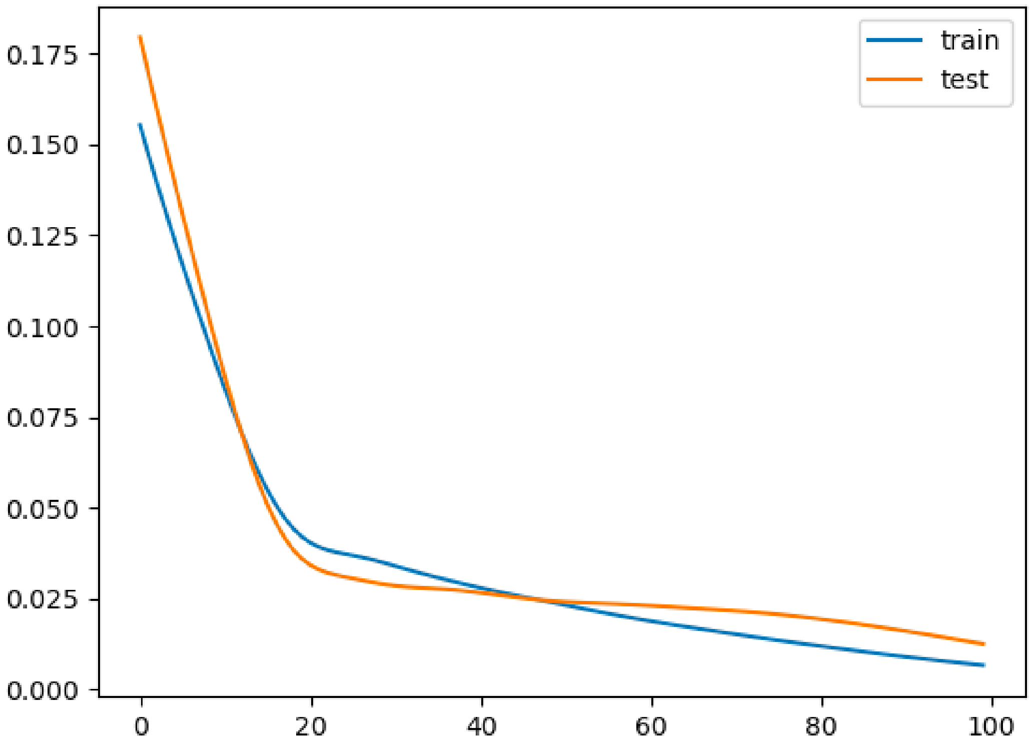

Figure 4.5: Illustrating the learning behaviour of the single variable LSTM model. It can be seen that the model keeps improving until around 100 epochs.

The compliance of the model when it had been improved with 100 epochs can be seen in figure 4.6. The y-axis represents the days that the model was tested on. The forecasting accuracy of the model for the model can be seen in table 4.4

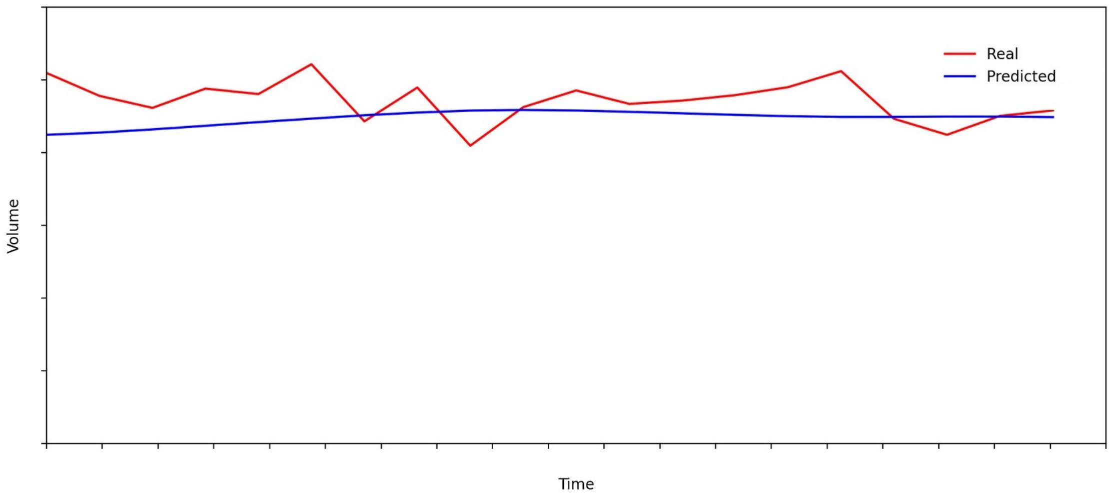

Figure 4.6: Illustrates the compliance of the single variable LSTM model.

Table 4.4: Performance of the single variable LSTM.

Single Variable LSTM	
Measurements	Values
RMSE	57
MAPE	22,41 %
MaxAE	188
MaxAPE	38,50 %

4.3.5 Multivariate LSTM

The LSTM-based model which used multiple variables to forecast the daily volume of trucks was supposed to act more like a casual model rather than the other time series based models. The theory behind the model was to find available metrics from the daily operations of the container terminal and then see if these could be correlated with the daily volumes of trucks.

The preliminary findings of understanding which variables that were relevant to base the model on indicated that most variables were uncorrelated. The variables which were found significant can be seen in table 4.5. The model which was constructed ended up being based around the LSTM framework with Adam as an optimizer.

See figure 4.7 for the logic behind the algorithm.

Table 4.5: Correlation Matrix

	Export	Import	Storage	Dwell time	Volume
Export	1				
Import	0.03	1			
Storage	0.350	0.80	1		
Dwell time	0.03	0.02	0.03	1	
Volume	−0.02	0.28	0.23	0.15	1

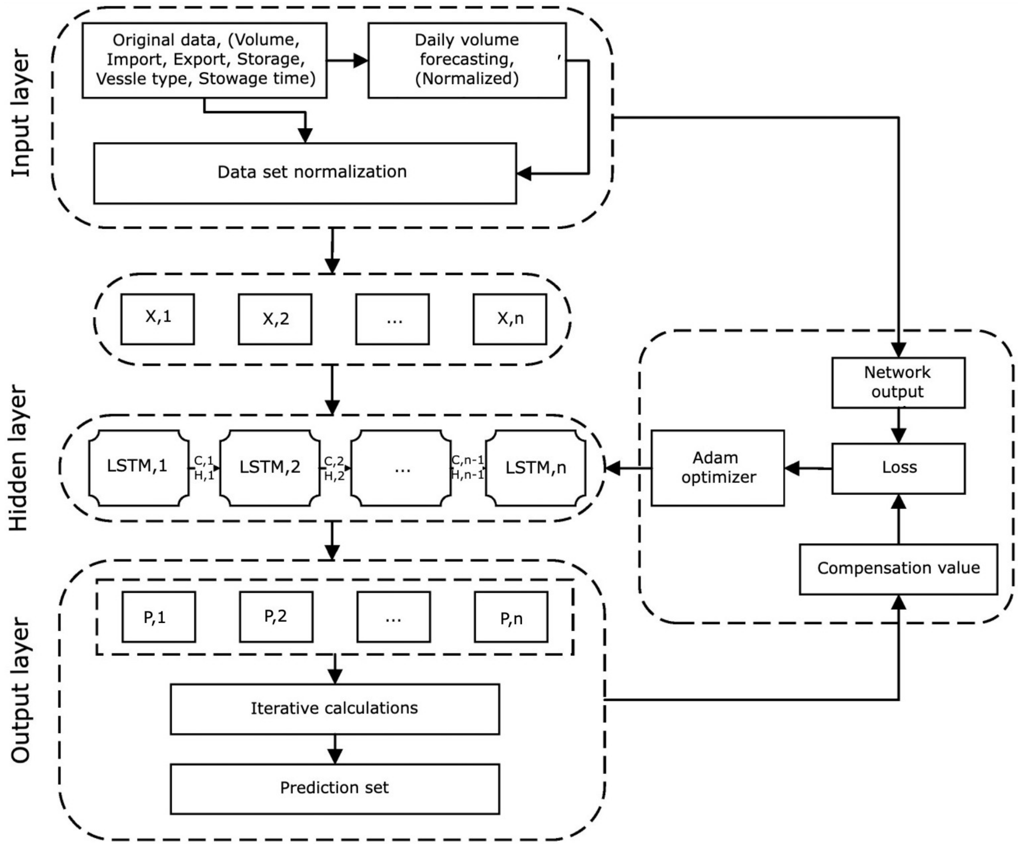

Figure 4.7: Logic behind the algorithm with Adam.

It was shown that the model gained increased accuracy as the number of training iterations increased. When training the model, different learning rates were tested to see if any improvement could be found. See figure 4.8 and 4.9.

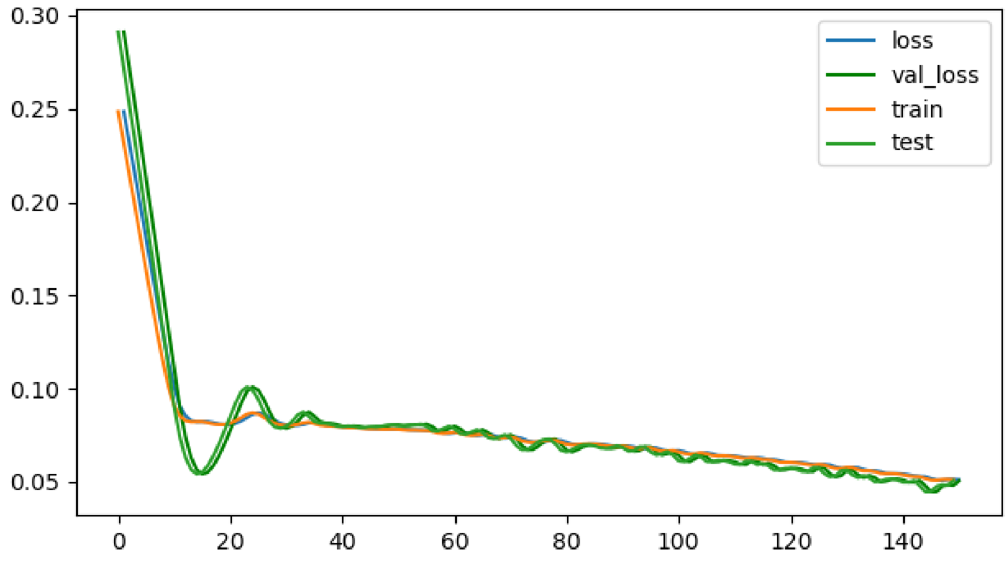

Figure 4.8: The multivariate model's training performance with high learning rate. The figure illustrate how the model stabilize around 150 epochs, without being under or over fitted.

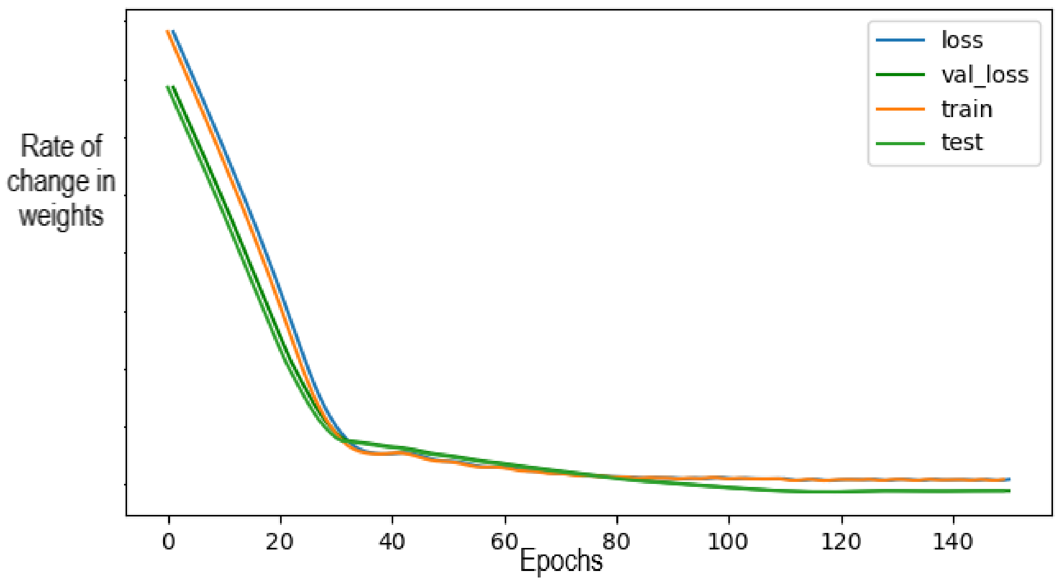

Figure 4.9: The multivariate model's training performance with low learning rate. It can be seen that a low learning rate reduce the problems with over and under fitting. It can also be seen that the model stops improving around 110 epochs.

The model in figure 4.8 with a high learning rate seems to yield a model which is a bit more unstable, it also does not reach a steady state as soon as the model seen in figure 4.9. The predictive capabilities of the model were not noticed to differ depending on the difference in the two learning rates.

To test the model on the normalized data set different lags and training horizons were selected. It was noticed that the model's predictive ability improved while increasing the lag of the model. It was also noticed that using more data to train the model yielded a forecast which seems to follow the test data set more closely.

The first findings presented in figure 4.10 indicated that the model is somewhat able to predict the daily volumes of trucks. While it does seem to follow changes it is not able to make correct predictions regarding the size of the volume.

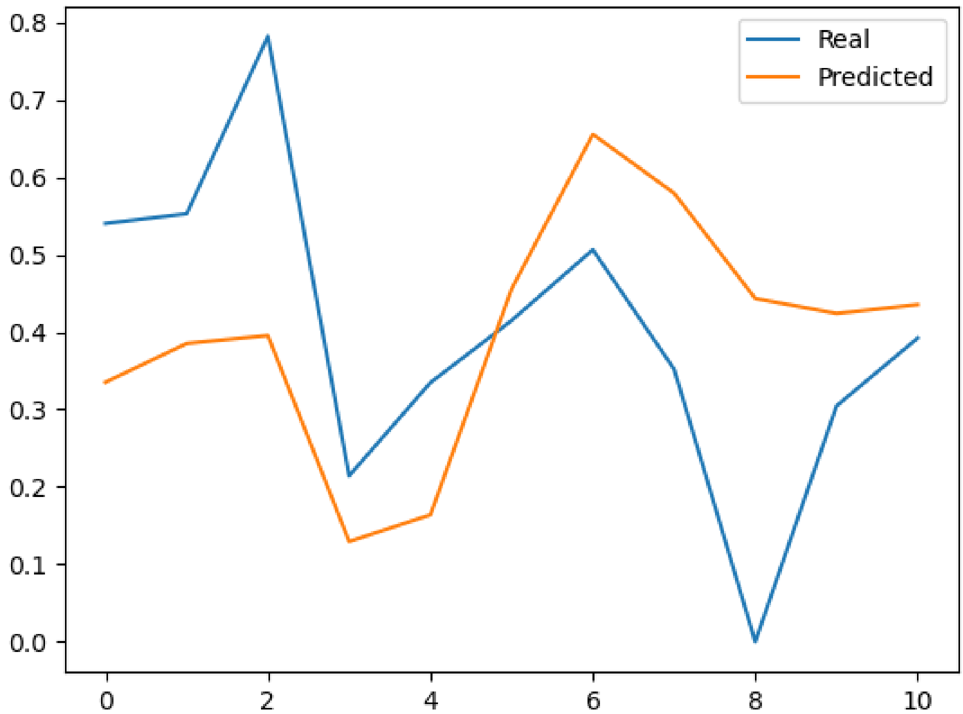

Figure 4.10: Plot illustrating performance with very low lag and low test size. The Y-axis illustrate the normalized volume, and the X-axis illustrate the timeline.

Further testing with higher lag, from 2 to 4 days and changes in test size did not improve the model as seen in figure 4.11. The responsiveness of the model decreased and it was not able to correctly make predictions.

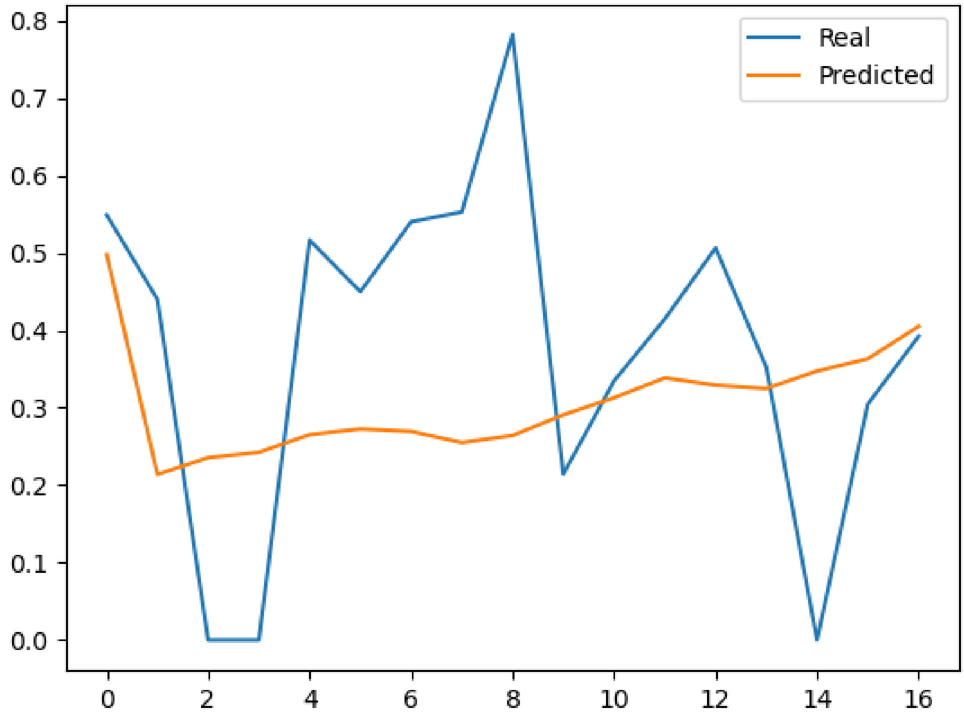

Figure 4.11: Plot illustrating performance with low lag and low test size. The Y-axis illustrate the normalized volume, and the X-axis illustrate the timeline.

By further increasing the lag of the model it was shown that the responsiveness of the model increased. Though, the model could not follow the test data set as the variety of the test data was high.

Lastly, the time lag was increased further and the test size was doubled. This, in turn, made the model able to follow changes in the test data set. As well as following the proportion of the changes in the test data. It was thus, concluded that a higher lag and higher test size were a better fit for the model.

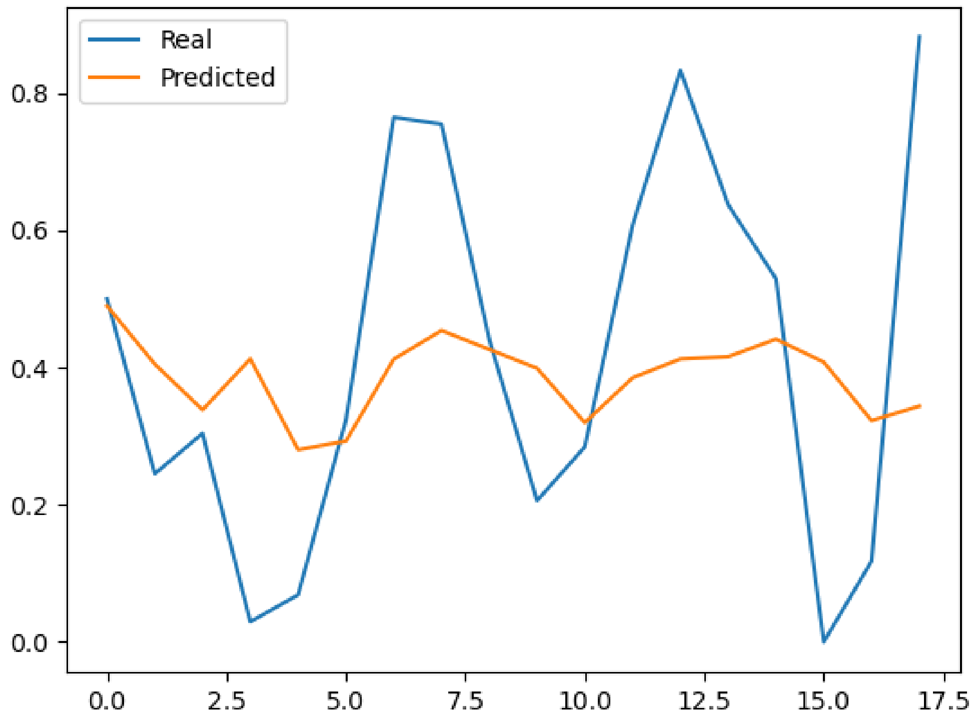

Figure 4.12: Plot illustrating performance with moderate lag and moderate test size. The Y-axis illustrate the normalized volume, and the X-axis illustrate the timeline.

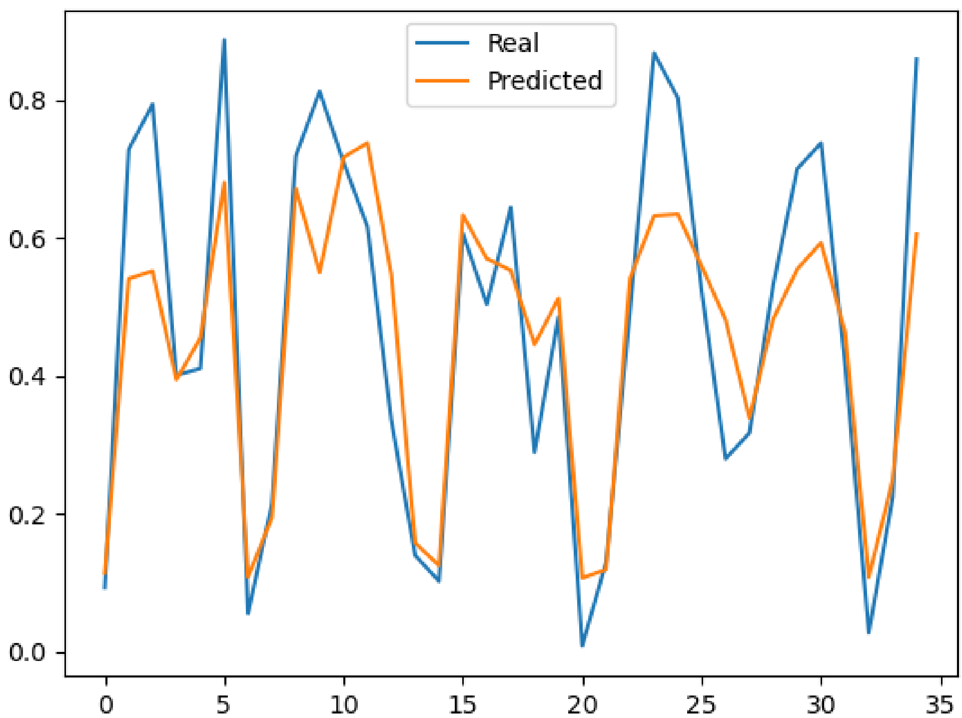

Figure 4.13: Plot illustrating performance with high lag and high test size. The Y-axis illustrate the normalized volume, and the X-axis illustrate the timeline.

The final model which were created was able to follow changes and the descriptive statistics related to the forecasting accuracy can be found in table 4.6.

Table 4.6: Performance of the multivariate LSTM.

Multivariate LSTM	
Measurements	Values
MAPE	$6,90$ %
MaxAPE	$38,50$ %

5

Analysis

In this chapter, an analysis of the results will be presented. The evaluation of the different models and their accuracy and applicability will be covered here. First, the three time series models, moving average, exponential smoothing and ARIMA(2,1,47) will be evaluated. Then, the two casual models, single variable LSTM and multivariate LSTM, will be evaluated

5.1 Evaluation of Time Series Models

A comparison of the time series models' performance with regards to the four measurements can be seen in Table 5.1. Since four different measurements were used to assess the accuracy of the forecasting models, the assessment of which model is most accurate does not necessarily have to be obvious since no weighing or prioritization of the measurements was used. A subjective judgment had to be applied to analyze the performance. However, as can be seen in Table 5.1, there is a clear indication that the exponential smoothing model performs worse. One can also see that ARIMA(2,1,47) is the best performing model. It performs better or as well as the other models for all measurements. The moving average model performs equally well for both RMSE and MAPE. The difference in MAPE for the two models is not significant. When comparing the performance for MaxAE and MaxAPE, ARIMA(2,1,47) performs better with a MaxAE of 23 units better and $3,19$ percentage points better when comparing MaxAPE.

Table 5.1: Comparison of the performance of moving average, exponential smoothing and ARIMA(2,1,47).

Comparison of performance				
Models	RMSE	MAPE	MaxAE	MaxAPE
Moving average	73	$5,51$ %	306	$26,82$ %
Exponential smoothing	85	$7,23$ %	298	$48,05$ %
ARIMA (2,1,47)	73	$5,50$ %	283	$23,63$ %

When evaluating the compliance in figure 4.2 and 4.4 one can see that the predicted volume for both the moving average model and the ARIMA(2,1,47) model is following the real value with noticeable compliance. Both models are having problems forecasting some drops and peaks, but the overall impression of the compliance is good.

The compliance of the exponential smoothing model in figure 4.3 is not as obvious as for the moving average model and the ARIMA(2,1,47) model. The mean is relatively good, which a MAPE of $7,23$ % indicates. However, when evaluating the graph it is obvious that the exponential smoothing model does not have the ability to foresee the increases and decreases in volume to the same extent as the moving average model and the ARIMA(2,1,47) model. This makes the model inapplicable to use as a basis for daily resource planning. The model would maybe be sufficient when forecasting volume over longer periods of time where the mean is more important than following sporadic increases and decreases in volume. The exponential smoothing method's inability to follow increases and decreases in real volume disqualifies the model for forecasting the volume of containers on a daily basis.

As mentioned, no significant difference can be found when assessing the performance of the moving average model and the ARIMA(2,1,47) model. That these two models perform as similar as they do is not surprising since the same lag is used in both models, 47. The reason for the ARIMA(2,1,47) to better predict the sudden increases and decreases is due to its additional parameters, p and d.

The ARIMA(2,1,47) performs slightly better for the measurements of MaxAE and

MaxAPE. However, these measurements are only assessing the one predicted value with the highest absolute error and the one with the highest absolute percentage error. Totally, 251 values are predicted, but since there exist no differences in RMSE and MAPE these two measurements will therefore be decisive. The ARIMA(2,1,47) is the most appropriate model for forecasting the volume when considering time series models.

It is an important requirement for the forecasting model to be used for planning to foresee sudden and heavy increases and decreases in volume. If not, the yard planner who is planing the resources available in the terminal would assume a forecast that differs significantly from reality and therefore plan a large abundance or shortage of resources. This would either result in unnecessary expenditure of resources or penalty fees due to the inability to meet the high demand and high lead times in hinterland operations.

5.2 Evaluation of Causal Models

The two different LSTM models which were created were with the correct fine tuning of parameters able to make predictions which could act as guidance for the yard planner. While the output of the models would yield predictions below MAPE of ten percent, it was also the case that the models sometimes crashed due to exploding vector gradient, see Chapter 2.3.2.5. Thus, the model failed to make any predictions at all. Over some periods the error would also rise above 20 percent, which according to interviews with staff at APMT is too high to be of use to a planner.

While these findings suggest that the models have the potential to work., the results also impose some of the drawbacks of using machine learning for prediction. It seems that the smoothing component of the time series model will yield a forecast which is more optimal for shorter forecast horizons. For example, a forecasting horizon below two days, while the LSTM models, in theory, would perform better on longer forecasting horizons. The findings of testing different parameters and horizons could not derive a correlation of lower forecast horizon and higher accuracy.

Since the time series models use moving average the weight of the model can be shifted to be shortsighted and thus, be more appropriate for shorter horizons. While the model can be trained on more short term data which is closer in time. The model was not able to "learn" and the model would just overflow. Which results in the model being unable to yield an output.

5.2.1 Single variable LSTM

The LSTM model based around only one variable seems to not offer any apparent advantages compared to the other models which have been construed and tested throughout the study. The method can make predictions, but the margin och error is higher compared to all other tested models. While it has the potential to be improved further by increasing the training data set and improving parameters in the model. It seems that the model, as it is constructed now, has limits. Increasing the training time yields marginal improvements but there is no guarantee that using more data will make the model better.

The time series models on the other hand could gain from more data and would thus, be a better candidate to improve upon. Controversially, a time series model can only be improved so much. The results from findings the best parameters of the time series models explored around 1300 different configurations. This way of using for loops and finding optimal settings is possible for a machine learning model but would require more data processing capabilities and time than available in this study.

5.2.2 Multivariate LSTM

Where as the single variable LSTM model has limitations the multivariate model seems to cover up for these in some areas. The multivariate model is able to perform better in regards to forecast accuracy and potential for improvements. The forecast accuracy, as of now, measured in MAPE is lower, thus being better than the single varaible LSTM model, see table 5.2. While, it is higher that the ARIMA(2,1,47). It has a average forecasting accuracy around 10 percent which, according to planners in the industry is good enough.

Table 5.2: Comparison of the performance of LSTM and multivariate LSTM.

Comparison of performance				
Models	RMSE	MAPE	MaxAE	MaxAPE
LSTM	57	$22,41$ %	188	$38,5$ %
Multivariate LSTM	$*$	$6,90$ %	$*$	$38,5$ %

The model is also able to be developed further by finding new variables which can be introduced to increase the accuracy of the model. Using more variables will make the model harder to learn, thus, potentially harder to reliably use. This trade off, however, could prove to be worth implementing.

Another advantage of using a multivariate model is that new variables can be introduced depending on how the forecasting environment. If a planner has a gut feeling that a certain variable in the operations is relevant. The hypothesis can be tested and a new variable can quite easily be introduced in the model. If the model improves with the addition of a new variable it can be kept and thus, the model has been upgraded. This will make the forecasting model more time consuming and resource intensive to use. Though, it can potentially become very good. By finding new correlations between measurable events and the volume of daily trucks. It could potentially be possible to forecast events that could be considered as "outliers. This would in turn make the model better at reducing large deviations, compared to a time series model.

6

Discussion

This chapter will begin with a discussion about the results and the analysis, first of the time series models, then of the causal models. Secondly, the usability of the forecast model will be discussed as well. Thirdly, a discussion of the method used for this study and the validity and reliability. Lastly, some recommendations for further research will be given.

6.1 Time Series Models

In the previous chapter, it was demonstrated that ARIMA(2,1,47) was the best time series model for predicting the future volume of trucks arriving at the terminal. The performance of the model could be considered to be very good, a MAPE of $5,5$ % is good. The model is having some problems with sudden and heavy variations in volume, to expect something else from a time series model would be unreasonable since the predictions are based on only historical data.

If the model will be usable when the yard planner is planning resources is hard to tell. The results presented are based on predictions made one day ahead and today, the yard planner is using a planning horizon of one week. Generally, ARIMA models are bad at predicting further than a few days. Depending on how the forecast model would be used will determine if the ARIMA(2,1,47) is usable. The reason for only predicting the volumes for one day ahead was that the software used for testing the models was designed to predict the next data point in the time series, the next day.

When determining which settings would yield the lowest error for the SARIMA$(p, d, q)(P, D, Q, s)$ it was found that the ARIMA(2,1,47) yielded the lowest error. As with the exponential smoothing model, no seasonal trends were found. One reason for this could be that the data available only covered one year. If data over a longer period of time, preferably several years, would be available, some seasonal trends would presumably be found. If so, it would have increased the accuracy of the model and most likely extended the number of days ahead of time that could be forecasted.

It is important to understand that the optimal settings found, (2,1,47), are only optimal based on the data available now. In the future, new settings can yield a lower error. Therefore, if a SARIMA-model were applied at APMT, the Python code described in section 3.4.1.3 must be executed regularly to ensure that the settings used at that time are the ones that also yield the lowest error.

If the predictions from the SARIMA-model are not accurate enough or do not predict the volumes far enough in the future, the use of the model could be altered. Instead of predicting the daily volume, the model could be used for predicting the weekly volume. Then the fluctuations in volume would be lower and easier to predict. This method would not be sufficient to use as the basis of resource planning but could provide the yard planner with a hint of the total volume for the upcoming week. In such a model the input would be the weekly volume instead of the daily volume. It is not necessary that the SARIMA-model would be most appropriate, there is a likelihood that an exponential smoothing model would be the best fit even if it performed worse among those compared.

6.2 Causal Models

Casual models have the advantage of relying on outside factors and events which enable the models to be made which accounts for changes in the forecasting environment. Rather than just using past data for predictions. While it is hard to quantify and make outside factors and events into data it has been seen that this is in some sense possible with some minor configurations to already existing data. As there seems to be a trend to gather more and more data throughout the industry, it can be assumed that data availability and quality will be higher in the near future. This will in turn enable the building of more complex and potentially more accurate models. The limitations of this research project were that only data available by courtesy of APMT was used in the modeling. If more elaborate methods to extract data are developed the casual models could be modeled very differently.

Different methods for finding data that has correlation or auto correlation with the daily arrivals of trucks could be constructed. Either manual or automatic. The parameters could also be modeled to be changed depending on the lowest error or similar to how they were chosen in section 3.4.1.1. Choosing to include variables based on correlation is not necessarily the most optimal solution for building a good model. However, it was thought of to be a good starting point.

6.2.1 Single Variable LSTM

The single variable LSTM model is as of now not a model which has any apparent advantages over the other models. It has low accuracy and offers no real advantages to the other time series models, other than that it could be used over longer forecasting horizons. Since the other time series model will trend down due to the moving average component. In relation to the multivariate LSTM model. It is superior in the sense that it is easier to train and program. However, in a performance comparison, it seems to not offer any advantages. Thus, APMT should probably be a lot better to look into more advanced models related to time series or machine learning.

Though, only LSTM has been used in the modeling process of the ANN models. Thus, the findings suggest that using a single variable LSTM is not suitable. There could be other ANN methods that potentially can perform well in a forecasting environment like the one with the daily volumes of trucks. The research within the field of ANN is growing and further improvements in learning algorithms and more accurate methods could serve to make the single variable models yield forecasts with high accuracy(Ray et al., 2020; Bui et al., 2020; Pang et al., 2020).

6.2.2 Multivariate LSTM

The multivariate LSTM model has been able to reliably make accurate predictions. While the model was a lot trickier and more complex to build than the others it can provide some advantages related to how it can be improved and optimized for different environments. The test which was used to evaluate the model will however have some degree of error and the validity of the tests should be considered as lower compared to the other methods. In part, due to the low training data and testing data set. Furthermore, the data set used for testing and training was also altered to fit the model prerequisites. These were that no zero numbers can be present and the numbers have to be fitted to a continuous date. While it is possible to resolve this, the lack of knowledge and time made the authors decide to keep the model as is. Thus, the model is based on data with missing values which has been filled with established methods. This could in theory make the model less valid, though, it still goes to show that the reliability of the model is high.
while it has some values which have been filled in with the missing values formulas in python. The model should still be reasonably good at predicting real volumes in the truck terminal.

As mentioned with the Single Variable model, it is likewise possible for the Multivariate LSTM model to be improved upon. Both in terms of code efficiency and optimization of parameters. Furthermore, a change in the algorithm architecture could also serve to improve the model, in terms of forecasting accuracy and reduced training time. A real advantage that has been mentioned previously with the Multivariate LSTM is the possibility to keep expanding the model (Sagheer and Kotb, 2019). As more data gets available the model can be grown larger with relative ease. External data such as economical, geographical and societal could be used to expand the variable list of the model (Xu et al., 2020; Alhirmizy and Qader, 2019). With the hope of gaining increased accuracy related to uncommon events which cause high variation in the daily volume of trucks.

Thus, the multivariate model could be considered to be the model with the most improvement potential. In contrast to the ARIMA or SARIMA models which in the best case would need a data set including different years. The multivariate LSTM could potentially perform well on less data. The results, however, only suggest that the multivariate model work on a shorter data set. Since the data set used to evaluate the multivariate were only 4 months. In contrast to the ARIMA model which used a full year in the model creation.

6.2.3 Data Quality

While the main content of the Thesis is related to predicting values based on gathered data, it is of great importance that the gathered data is based on reality. In other terms, the validity of the data. While it is possible to check the reliability of the models and the data with statistical tests. It is a lot harder to see if the validity of the data is high. While all the data that has been used in the Thesis has been provided by APM Temrinals Sea Traffic Management system, and thus, has been gathered automatically. The chance of human errors is low and the data could therefore be considered to be of high quality.

While it has been mentioned in the report that more data could be useful in creating a forecasting model. It could also be useful in the sense of being more picky with which data to train the models on. As more data is available the models can be trained on variables which has high correlation with the arrival of trucks.

6.3 Usability of the Forecast Model

How a forecast model would be used by APMT in the Port of Gothenburg is unknown at the time of writing this report. Different applications had been discussed but nothing has been decided. Therefore, the use of a forecast model will be discussed and elaborated on in this section. First, the use of a forecast model as a basis for resource planning will be examined. Secondly, an alternative way of using the forecasts will be elaborated on.

6.3.1 Planning in Accordance with Forecast

If the predictions generated from the forecast model are of sufficient accuracy to plan resources in accordance with, that would be a suitable use of the model. By using a quantitative forecast model instead of a qualitative would eliminate the downsides with subjective judgments. The quantitative forecast is less biased and does not have the same tendency of thinking wishfully rather than realistically. However, a qualitative forecast can consider many parameters and adjust which parameters that are relevant and consider from week to week without the need of reprogramming the forecast model. As have been concluded earlier, the time series models are having some problem with foreseeing sudden changes in volume and would perform worse when forecasting further forward in time, which can be costly if planned in accordance with.

To partly overcome the problems with sudden and heavy changes in volume, an alternative solution could be to sometimes abandon the quantitative forecast and use a qualitative forecast instead. There are scenarios when things happen that the quantitative forecast cannot consider. For example, on the 23 of March 2021 the massive container ship ran aground in the Suez Canal and blocked the passage for other vessels to pass (Samaan et al., 2021). This created turbulence and affected the import volumes for the near future. The effects of the blockade in the Suez Canal could not be considered by a quantitative model. A time series model would not have foreseen the sudden drop in volume but would after some days stabilize to the new low levels of volume and when the rise in volume comes after the ship has been towed away the same procedure would be repeated until it stabilizes on the ordinary level of volume.

It is important to remember that a forecast is only a forecast and not a final decision. The forecast should only be considered as a tool. By understanding what the forecast is basing its predictions on, one can make assumptions when parameters not embedded in the forecast model can change the actual volume considerably, as with the blockade of the Sues Canal. In a situation like that, it could be beneficial to disregard the quantitative forecast and instead of using a well-founded subjective judgment as the basis for the resource planning.

6.3.2 Change the Hauliers Behavior

An alternative way of using a forecast system would be by trying to even out the volume over the day by encouraging the hauliers to visit the terminal during hours with generally lower volume. By distributing the volume more evenly over the visit hours, both the resource efficiency and the flow efficiency could be improved.

An idea of how such as system could work is by providing the hauliers with an estimated waiting time at the gate if they arrive at a certain time. As has been seen in figure 1.4, the number of haulier arrivals peaks between 1:00-2:00 pm. If an assumption is made that the resources are constant over a day, which they are to some degree, the longest waiting time for the hauliers would be between 1:00 pm until 4:00 pm. If this waiting time was communicated to the hauliers and the haulage contractor they could see how much time they could save by arriving a couple of hours earlier or later than planned. Expectantly, this would lead to the hauliers arriving during hours with lower waiting time in order to increase their efficiency.

The forecast models presented in this report only predict the daily volume and could not without modifications or additions be used for this purpose. The forecast models are not good at creating so high-resolution predictions. There are several different ways of increasing the resolution, one way to do it would be to assume that the distribution of volume during a day is represented by the average day in figure 1.4. Meaning that, for example, approximately 10 % of the trucks during a day arrive between 2:00-3:00 pm and 7 % arrive between 7:00-8:00 pm. By assuming this and knowing the daily volume, one can come up with an estimate for the volume for each hour by multiplying the total volume with the portion of trucks arriving a certain hour. See formula 6.1 where F_h represents the forecasted volume for hour h, F_t represents the forecast volume for day t and P_h represents the percentage of daily volume during hour h

$$F_h = F_t * P_h \tag{6.1}$$

When the forecasted volume for a certain hour, F_h, is determined, the waiting time, W_h, if the haulier arrives that hour can be estimated by dividing the F_h with the number of available resources R_h with the unit of the number of trucks that can be processed per hour. See formula 6.2.

$$W_h = \frac{F_h}{R_h} \tag{6.2}$$

If an implementation of a forecast system such as this would be used as described and give the desired effect, the distribution of trucks arriving would change over time. For the waiting time to remain reliable, the P_h needs to be updated repeatedly.

6.4 Method Discussion

In this section, a reflection on the research methodology will be given.

6.4.1 Data Collection

Here, a reflection on the effects of the absence of a pilot study and the lack of data that was available will be provided.

6.4.1.1 Absence of a Pilot Study

Since the research method was mixed with a combination of quantitative and qualitative, data was collected in line with both methods. An initial plan was to conduct a pilot study at companies that are, as APMT, active in an industry with a volatile demand with no option to govern the demand with, for example, booking systems. interesting businesses were for example fast food restaurants, grocery stores, mail delivery or, emergency rooms. The idea of conducting the pilot study was to create references to benchmark against and to understand how other businesses solved the issues with forecasting. Unfortunately, no pilot study was carried out. When the intended participants were contacted, the level of interest was generally low, they considered this information as sensitive to their business and therefore did not want to participate in the pilot study. The number of intended participants that were interested to participate was considered too low and the idea to conduct a pilot study was abandoned.

The pilot study would have added another dimension to benchmarking. Instead of only comparing the different forecasting models with each other, it would have been interesting to compare different businesses and understand how one model performs in different environments. A pilot study would also facilitate the first sequence of the study, the qualitative phase, by enhancing the understanding of how forecasting systems are created and used in other industries.

6.4.1.2 Lack of Quantitative Data

The data of volumes of containers in the terminal only reached from 30th of December 2019 until 30th of December 2020. In early 2020, the COVID-19 pandemic broke out and can potentially have affected the volumes. Due to the fact that the data only covers a relatively short time period and a very special time period, this study only provides answers in the context of a pandemic.

If data over a longer time period would have been available, other findings could have been found. If looking at the figure 1.3, one can see that volume is high at the beginning of the year, slowly decreases until the summer months. After the summer the volume slowly increases and then stabilizes in September. With the data available, one can only assume that this pattern is similar every year.

When creating the multivariate LSTM model, more variables than historical volumes were consider. The additional information about the containers that were needed for this only reached from January 2020 until April 2020. With only four months of data, only a small sample size could be used which decreases the reliability of the model. However, with the sample size that was available, inclinations were identified.

6.5 Sustainability Discussion

With a proper forecast system at place a few advantages related to sustainability could be gained. As of know, how the forecasts will be used at APMT is not yet decided, and which advantages that could be gained are dependent on the use of the forecasts. However, if the forecasts would be used to facilitate the yard planner with allocating and planning resources, a more efficient planning could be made. With a higher efficiency, the same effect could be produced with less resources which could be considered a development towards sustainability. This is only one of a few advantages that could be gained, but we could expect more sustainability related advantages when a proper forecasting system is at place.

6.6 Further Research

This section will provide recommendations for further research, things that would have been investigated if appropriate time, resources and data would have been available.

6.6.1 Access to More Data

As mentioned above, the data only covered a relatively small time period. With more data, other findings could have been made. For further research, the recommendation is to investigate it seasonal trends can be found when looking at data over a period of several years.

It would also be interesting to have access to richer data about the container to consider even more variables in the multivariate LSTM model. With richer data about the specific containers, correlating variables not established in this study could be found. If so, the multivariate LSTM model could probably be enhanced.

6.6.2 Investigate Possibility to Track the Trucks

Investigate the possibility to track the trucks was not included in the scope of this study, even though it could be interesting. This could possibly enhance the forecast in short term, on an hourly basis. Similar systems exist for sea traffic.

6.6.3 Investigate How the Model Could be Used

This study did not cover how a forecasting system of trucks arriving at the container terminal could be used. It would be a natural next step to investigate this. For a thorough analysis of the usability of a forecasting system, how the system will be used are necessary. When the use of the system is determined, a decision regarding how the system will facilitate the operations, or the planning of the operations is needed.

7

Conclusion

This study has explored different methods for building and constructing forecasting models to be used for the forecast of trucks visiting a container terminal. Several methods have been tested and used to construct models which then have been tested on real truck data gathered from the previous year (2020). The models have been evaluated on accuracy as well as how they could fit into the operations of a container terminal.

Out of the five different models that were constructed, three were based on time series methods and two were based on artificial neural networks and machine learning. It was shown that the time series models were on average better at predicting a forecasting value more in line with the real world data. Furthermore, it was also shown that all of the constructed models had the potential to be improved upon with more data. The time series models could gain increased accuracy by adding the seasonality and trend component of testing the models on multiple years. Furthermore, could the LSTM models be improved with more training data and the use of more relevant independent variables.

As it seems the random variation of sudden spikes in the volume of truck visits was unable to be countered by a time series model. Perhaps since the outlier events might not be autocorrelated with past volumes, or that they might just be random appearances. While this might be the case, a strong argument for using LSTM models is that it might exist external or internal variables related to the business of container freight and terminal operations which are related to the outlier events.

If these variables are identified and incorporated into the model, the Multivariate LSTM model could be of great use to APMT as outlier events might be somewhat predictable. Moreover, it was shown in the model creation phase of the multivariate LSTM model that small changes in the models could yield decent improvements. Though, as it seems, on average the multivariate LSTM model performs worse than the tested ARIMA model. A conclusion can thus be made, that the ARIMA model is the superior choice if a model was to be implemented into the business in the near future. However, if more resources are invested in the multivariate LSTM model, it is highly possible that this model would outperform the ARIMA model.

A conclusion to be made from the reasoning above is therefore that it might be more profitable to invest the time and resources to construct a more advanced LSTM model. Especially so, if the model could gain an advantage related to predicting outlier events which in turn could cause big problems in the operations of the container terminal.

Bibliography

Alhirmizy, S. and Qader, B. (2019). Multivariate time series forecasting with lstm for madrid, spain pollution. In *2019 International Conference on Computing and Information Science and Technology and Their Applications (ICCISTA)*, pages 1–5. IEEE.

Batista, G. E., Prati, R. C., and Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1):20–29.

Bock, S. and Weiß, M. (2019). A proof of local convergence for the adam optimizer. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Boshnakov, G. N. (2021). Autocorrelations and white noise tests.

Bouktif, S., Fiaz, A., Ouni, A., and Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, 11(7):1636.

Bryman, A., Bell, E., and Harley, B. (2019). *Business research methods*. Oxford university press.

Bui, D.-K., Nguyen, T. N., Ngo, T. D., and Nguyen-Xuan, H. (2020). An artificial neural network (ann) expert system enhanced with the electromagnetism-based firefly algorithm (efa) for predicting the energy consumption in buildings. *Energy*, 190:116370.

Byrne, R. F. (2012). Beyond traditional time-series: Using demand sensing to improve forecasts in volatile times. *Journal of Business Forecasting*, 31(2).

Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250.

Chang, Z., Zhang, Y., and Chen, W. (2018). Effective adam-optimized lstm neural network for electricity price forecasting. In *2018 IEEE 9th international conference on software engineering and service science (ICSESS)*, pages 245–248. IEEE.

Chen, C.-Y., Lee, W.-I., Kuo, H.-M., Chen, C.-W., and Chen, K.-H. (2010). The study of a forecasting sales model for fresh food. *Expert Systems with Applications*, 37(12):7696–7702.

Davydenko, A. and Fildes, R. (2016). Forecast error measures: critical review and practical recommendations. *Business forecasting: Practical problems and solutions*, 34.

Fadare, D. (2010). The application of artificial neural networks to mapping of wind speed profile for energy application in nigeria. *Applied Energy*, 87(3):934–942.

Fu, R., Zhang, Z., and Li, L. (2016). Using lstm and gru neural network methods for traffic flow prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 324–328. IEEE.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Graves, A. (2020). Time series forecasting with a sarima model.

Gupta, A. and Dhingra, B. (2012). Stock market prediction using hidden markov models. In *2012 Students Conference on Engineering and Systems*, pages 1–4. IEEE.

Göteborgshamn (2021). Störst blir större – göteborgs hamn tar marknadsandelar.

Hart, C. (2018). Doing a literature review: Releasing the research imagination.

He, Y., Deng, J., and Li, H. (2017). Short-term power load forecasting with deep belief network and copula models. In *2017 9th International conference on intelligent human-machine systems and cybernetics (IHMSC)*, volume 1, pages 191–194. IEEE.

Ho, S. L. and Xie, M. (1998). The use of arima models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2):213–216.

Hoover, J. et al. (2006). Measuring forecast accuracy: Omissions in today's forecasting engines and demand-planning software. *Foresight: The International Journal of Applied Forecasting*, (4):32–35.

Hu, Q., Wiegmans, B., Corman, F., and Lodewijks, G. (2019). Integration of inter-terminal transport and hinterland rail transport. *Flexible Services and Manufacturing Journal*, 31(3):807–831.

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

Jonsson, P. and Mattsson, S.-A. (2009). *Manufacturing planning and control*. McGraw-Hill Education.

Kim, S. et al. (2016). Forecasting short-term air passenger demand using big data from search engine queries. *Automation in Construction*, 70:98–108.

Kim, S. and Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679.

Lindgren, G. (2012). *Stationary stochastic processes: theory and applications*. CRC Press.

Mobarakeh, N. A., Shahzad, M., Baboli, A., and Tonadre, R. (2017). Improved forecasts for uncertain and unpredictable spare parts demand in business aircraft's with bootstrap method. *IFAC-PapersOnLine*, 50(1):15241–15246.

Mushtaq, R. (2011). Augmented dickey fuller test.

Ouyang, P., Yin, S., and Wei, S. (2017). A fast and power efficient architecture to parallelize lstm based rnn for cognitive intelligence applications. In *Proceedings of the 54th Annual Design Automation Conference 2017*, pages 1–6.

Pang, Z., Niu, F., and O'Neill, Z. (2020). Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renewable Energy*, 156:279–289.

Pölt, S. (1998). Forecasting is difficult–especially if it refers to the future. In *AGIFORS-reservations and yield management study Group meeting proceedings*, pages 61–91.

Prayudani, S., Hizriadi, A., Lase, Y., Fatmi, Y., et al. (2019). Analysis accuracy of forecasting measurement technique on random k-nearest neighbor (rknn) using mape and mse. In *Journal of Physics: Conference Series*, volume 1361, page 012089. IOP Publishing.

Prudêncio, R. and Ludermir, T. (2004). Using machine learning techniques to combine forecasting methods. In *Australasian Joint Conference on Artificial Intelligence*, pages 1122–1127. Springer.

Radecic, D. (2020). What is stationarity in time series and why should you care.

Ray, A., Halder, T., Jena, S., Sahoo, A., Ghosh, B., Mohanty, S., Mahapatra, N., and Nayak, S. (2020). Application of artificial neural network (ann) model for prediction and optimization of coronarin d content in hedychium coronarium. *Industrial Crops and Products*, 146:112186.

Rowley, J. and Slack, F. (2004). Conducting a literature review. *Management research news*.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Sagheer, A. and Kotb, M. (2019). Unsupervised pre-training of a deep lstm-based stacked autoencoder for multivariate time series forecasting problems. *Scientific reports*, 9(1):1–16.

Sakinah, N., Tahir, M., Badriyah, T., and Syarif, I. (2019). Lstm with adam optimization-powered high accuracy preeclampsia classification. In *2019 International Electronics Symposium (IES)*, pages 314–319. IEEE.

Samaan, M., Deng, S., El Sirgany, S., Salem, M., and Said-Moorhouse, L. (2021). Suez canal blocked by traffic jam after massive container ship runs aground. *CNN*.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, pages 312–343.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., Kamaev, V. A., et al. (2013). A survey of forecast error measures. *World Applied Sciences Journal*, 24(24):171–176.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306.

Späth, H. (2014). *Mathematical algorithms for linear regression*. Academic Press.

Suh, D. Y. and Ryerson, M. S. (2019). Forecast to grow: Aviation demand forecasting in an era of demand uncertainty and optimism bias. *Transportation Research Part E: Logistics and Transportation Review*, 128:400–416.

Sumi, S. M., Zaman, M. F., and Hirose, H. (2012). A rainfall forecasting method using machine learning models and its application to the fukuoka city case. *International Journal of Applied Mathematics and Computer Science*, 22:841–854.

Sun, Y., Wang, Y., Fu, K., Wang, Z., Yan, Z., Zhang, C., and Ye, J. (2020). Fmaeta: Estimating travel time entirely based on ffn with attention. *arXiv preprint arXiv:2006.04077*.

Tayman, J. and Swanson, D. A. (1999). On the validity of mape as a measure of population forecast accuracy. *Population Research and Policy Review*, 18(4):299–322.

Willmott, C. J. and Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science*, 20(1):89–102.

Xie, M.-Q., Li, X., Zhou, W.-L., and Fu, Y.-B. (2014). Forecasting the short-term passenger flow on high-speed railway with neural networks. *Computational intelligence and neuroscience*, 2014:375487.

Xu, D., Cheng, W., Zong, B., Song, D., Ni, J., Yu, W., Liu, Y., Chen, H., and Zhang, X. (2020). Tensorized lstm with adaptive shared memory for learning trends in multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1395–1402.

Zhang, K., Wang, Z., Zhang, J., and Schölkopf, B. (2015). On estimation of functional causal models: general results and application to the post-nonlinear causal model. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(2):1–22.

Zhao, Z., Chen, W., Wu, X., Chen, P. C., and Liu, J. (2017). Lstm network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2):68–75.

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden

www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY