NTNU

Norwegian University of
Science and Technology

Predicting Shipping Freight Rate Movements Using Recurrent Neural Networks and AIS Data

On the tanker route between the Arabian Gulf and Singapore

Stian Røyset Salen

Gisle Hoel Århus

Industrial Economics and Technology Management
Submission date: June 2018
Supervisor: Sjur Westgaard, IØT

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Preface

This master's thesis represents the finalisation of our second Master of Science (MSc) degrees at the Norwegian University of Science and Technology (NTNU), after completing our MSc degrees in Marine Technology at NTNU in 2016. The current thesis is part of the specialisation program Financial Engineering in the study program Industrial Economics and Technology Management, and it was written in the spring of 2018.

There are several people to whom we would like to express our sincere gratitude, for making this thesis possible. Firstly, we would like to thank our supervisor, Professor Sjur Westgaard at the Department of Industrial Economics and Technology Management at NTNU. Further, we want to thank Professor Keith Downing, at the Department of Computer and Information Science at NTNU, for kindly helping us with machine learning related issues. Moreover, the Norwegian Coastal Administration, Professor Bjørn Egil Asbjørnslett at the Department of Marine Technology at NTNU, and particularly PhD in Marine Technology, Carl Fredrik Rehn, and MSc in Marine Technology, Bjørnar Brende Smestad, deserve a thanks for providing AIS data, inspiration, useful discussions and help along the way. Lastly, we give thanks to our friends who have supported us, proofread the thesis and provided useful advices.

Trondheim, June 11, 2018
Gisle Hoel Århus and Stian Røyset Salen

Abstract

The purpose of this thesis is twofold. *Firstly*, we want to predict shipping freight rates in the liquid crude oil tanker market, applying a machine learning technique suitable for processing time-series (sequential) data, such as *Long Short-Term Memory (LSTM) Neural Networks (NN)*. *Secondly*, an important objective of this thesis is to investigate the predictive ability of satellite *Automatic Identification System (AIS)* data. AIS potentially provides real-time data on for example the current position, past routing and headed destination of ships world-wide. We examine whether this type of data adds valuable information, on top of other fundamental and financial data, with respect to the prediction of freight rate movements in the tanker shipping market. A total of 17 explanatory variables are included in the study, selected based on a thorough literature review of tanker freight rate determinants; eleven of the variables are derived from AIS data, and the other six are fundamental and financial variables gathered from Thomson Reuters Eikon. The applied data set contains time series of these 17 variables in the sample period 4 January 2012 to 24 December 2015.

Specifically, a type of recurrent neural networks, an LSTM NN model developed based on TensorFlow, is utilised to predict future price fluctuations on the major tanker route between the Arabian Gulf (AG) and Singapore, over the following short-term horizons: one day-, five days- and ten days ahead. This route is defined as TD2 by the Baltic Exchange. Furthermore, this model is applied to determine whether it achieves better prediction (or classification) accuracy by adding AIS data as input variables. A three-dimensional solution space is examined: Numerous experiments are performed, investigating the use of various subsets of explanatory variables (combinations of the 17 selected variables), over three different forecasting horizons, using two degrees of model complexity. Moreover, a multivariate linear regression model is utilised to provide benchmark results.

The findings of this thesis indicate that the LSTM NN model provides promising results when forecasting ten days ahead; the LSTM NN model outperforms the benchmark model in all six experiments carried out. However, on the shorter forecasting horizons (one day- and five days ahead), the benchmark model competes well with the LSTM NN model. The results further indicate that the LSTM NN model does not achieve significantly better prediction (or classification) accuracy by adding AIS data as model input. However, the quality of the applied AIS data is questionable. Thus, by sampling new and more recent AIS data, the forecasting results could be improved. Moreover, certain measures and techniques can be applied to attempt to enhance the LSTM NN model performance.

Sammendrag

Formålet med denne oppgaven er todelt. *For det første* ønsker vi å predikere fraktrate-endringer i det likvide oljetanker-markedet ved anvendelse av en maskinlæringsteknikk egnet for prosessering av tidsseriedata, som for eksempel *Long Short-Term Memory (LSTM) Neural Networks (NN)*. *For det andre*, et viktig formål med denne oppgaven er å undersøke den prediktive evnen til satellitt-data for skipstrafikk (*automatisk identifikasjonssystem (AIS)-data*). AIS gir potensielt sanntidsdata på for eksempel den nåværende posisjonen, tidligere ruter og forventet destinasjon for alle sjøgående skip verden over. Vi undersøker om denne typen data gir verdifull informasjon, i tillegg til andre fundamentale og finansielle data, når det kommer til prediksjon av fraktrater i oljetanker-markedet. I alt 17 forklarende variabler er inkludert i dette studiet, valgt ut basert på en grundig litteraturgjennomgang av typiske fraktratedeterminanter i tankmarkedet; elleve av variablene er hentet fra AIS data, og de andre seks er fundamentale og finansielle variabler innhentet fra Thomson Reuters Eikons database. Det anvendte datasettet inneholder tidsserier av disse 17 variablene i perioden 4. januar 2012 til 24. desember 2015.

Mer spesifikt, så bruker vi en type rekurrent nevralt nettverk, en LSTM NN-modell utviklet basert på TensorFlow, til prediksjon av fremtidige prisendringer på den viktige oljetanker-ruten mellom Persiabukten og Singapore, over følgende tidshorisonter: en dag-, fem dager- og ti dager frem i tid. Denne ruten er definert som TD2 av The Baltic Exchange. Modellen blir deretter brukt til å avgjøre om det oppnås bedre prediksjonsnøyaktighet (eller klassifikasjonsnøyaktighet) ved å inkludere AIS-data som inputvariabler. Et tredimensjonalt løsningsrom blir dermed undersøkt: mange eksperimenter utføres, med flere sett inputvariabler (ulike kombinasjoner av de 17 utvalgte variablene), over tre forskjellige tidshorisonter, med to grader av modellkompleksitet. Til slutt anvendes en multivariabel lineær regresjonsmodell til å gi referanseverdier som LSTM NN-modellresultatene kan sammenlignes med.

Resultatene i denne masteroppgaven indikerer at LSTM NN-modellen gir lovende resultater ved fraktrateprediksjon ti dager frem i tid. LSTM NN-modellen overgår benchmarkmodellen i alle seks utførte eksperimenter på denne prediksjonshorisonten. På de kortere horisontene (en dag og fem dager frem i tid) konkurrerer referansemodellen godt med LSTM NN-modellen. Videre viser resultatene at LSTM NN-modellen ikke oppnår signifikant bedre prediksjonsnøyaktighet ved å inkludere AIS-data som inputvariabler til modellen. Kvaliteten på AIS-dataen anvendt i dette studiet er imidlertid tvilsom. Ved å samle inn og anvende oppdatert, samt nyere AIS-data, kan det hende at prediksjonsresultatene blir bedre. I tillegg kan visse tiltak og teknikker brukes for å forsøke å heve ytelsen til LSTM NN-modellen.

Contents

List of Figures

List of Tables

1

Introduction

Shipping market participants, like actors in any industry, make decisions based on their expectations about the future. For shipowners and charterers, two important commercial decisions are whether and when to operate in the spot market (spot charter), where the daily freight rate can vary significantly, or to fix a ship on a period contract (time charter) for a lump sum (Stopford, 2009). Reducing the uncertainty linked to such decision-problems makes it easier for shipowners and charterers to predict future cash flows and profitability. This can be done through forecasting and risk management. To achieve as good forecasts as possible, it is necessary to obtain and analyse the right information about the present. As for any free market, the prediction of shipping freight rates is all about predicting the supply-demand balance. However, considering the bad reputation of maritime forecasting, this does not seem to be an easy task (Stopford, 2009). Explanations for this could be that prediction models developed until now either are too simple or too sophisticated, or that the right information is hard to discover.

1.1

Background

Regulations adopted by the International Maritime Organization (IMO) in late 2004 require all international voyaging vessels with above 300 Gross Tonnage (GT) and all passenger ships to carry an Automatic Identification System (AIS) transmitter. Ship-to-ship and ship-to-shore positions are now frequently reported through AIS messages, within time intervals of only a few seconds. Moreover, AIS messages include static information, such as ship identity, ship type, and physical appearance, like draught, destination and estimated time of arrival. They also contain dynamic data, such as ship speed, course and rate of turn. The AIS signals are transmitted using very-high-frequency (VHF) radio waves, and can be captured by terrestrial land-based antennas (T-AIS) and Low Earth Orbit satellites (S-AIS). In theory, this provides real-time data on the past routings, current positions and expected destinations of all ocean-going vessels. Automated processing of AIS data combined with expert knowledge, and perhaps other data sources on the type and size of cargo onboard the vessels, could thus potentially enable the generation of accurate and up-to-date trade flows data (such as true sailing distance and better estimates of individual vessel cargo sizes) (Adland et al., 2017).

According to classical maritime economic literature, freight rate levels in shipping are determined by the interplay between supply and demand. When it comes to supply, the overall cargo-carrying capacity and operational efficiency (or productivity) of the fleet are important factors. Operational efficiency is often considered constant in shipping market analysis, because reliable data on such dynamics are usually not available as empirical time series (Olsen and da Fonseca, 2017). On the demand side, the volume of cargo transported by sea and the sailing distance over which the cargo is transported are decisive factors. The true demand is thus measured on a tonne-mile basis [1]. A problem in terms of shipping market analysis have been the lack of accessible disaggregate data for ship demand on a tonne-mile basis. Having to rely on infrequently updated customs trade data or ship movement data, gathered and structured by shipbroking firms, maritime researchers and analysts have until recently been forced to take a simplified view of demand dynamics (Adland et al., 2017, 2016).

The introduction of satellite AIS data (*S-AIS* data, hereafter referred to as *AIS* data) has changed this situation. With the ability to aggregate data by the ship type or ship size group, this enables in theory the estimation of shipping demand on a tonne-mile basis with much greater accuracy than before. According to Adland et al. (2017), any maritime research dealing with market fundamentals, could benefit from AIS-derived tonne-mile demand data, as long as the cargo is observable and homogenous. This is typically the case with cargoes that are loaded from single-use terminals, like in the case of commodities such as crude oil, coal and iron ore. Furthermore, information on ship specifications, ship capacity utilisation and dynamic data like ship speed, which can be derived from AIS data, makes it possible to measure and account for operational efficiency on a per-shipment basis. This may lead to more accurate predictions of supply (Olsen and da Fonseca, 2017). Therefore, AIS-derived data could be beneficial to use as input in a maritime forecasting model.

1.2

Motivation and Problem Description

Using unconventional machine learning techniques, such as *Long Short-Term Memory (LSTM) Neural Networks* to exploit the potential in AIS-derived data for maritime forecasting, is the motivation behind this thesis. Moreover, the highly liquid crude oil tanker market will be investigated because it provides a lot of data and, as mentioned above, is suitable when it comes to the estimation of AIS-based trade volumes. The purpose of the thesis is twofold:

- Use neural networks to predict future price fluctuations in shipping freight rates. Specifically, develop a model that decides the future price-change outcome for a freight rate F, on a specific route S_F, over a time horizon t.
- Investigate the predictive ability of AIS data: Examine whether the neural network model achieves better prediction accuracy by adding AIS data as input variables, in addition to other fundamental and financial data.

[1] Tonne mile = *Metric tonnes shipped* multiplied by the *shipped distance* (measured in *miles*).

1.3

Contribution to the Literature

This thesis contributes to the existing body of literature with research that combines the use of machine learning, specifically LSTM Neural Networks, and AIS-derived data to forecast freight rate movements in shipping. According to the literature review presented in Chapter 2, i.e. as far as the authors are concerned, this has not been attempted before.

1.4

Limitations

The freight rate in question will be the price for transporting crude oil on the major tanker route between Ras Tanura, the largest oil port in Saudi Arabia, and Singapore. Saudi Arabia is the third largest oil producing country in the world behind the U.S. and Russia. It is located in the Arabian Gulf (AG) in the Middle East, which is by far the largest oil producing and oil exporting region in the world (see Section 4.2). This specific route between Saudi Arabia and Singapore, named *TD2* by the Baltic Exchange [2], is thus one of the world's most traded tanker routes (see e.g. Γολας (2012)).

Concerning the forecasting time horizons, the current thesis is limited by the availability of historical AIS data (four years of data). Thus, forecasts will be done on an operational level, using three different horizons: one day ahead, five days ahead and ten days ahead. These time frames are relevant for a fundamental shipping decision: whether to spot-charter a ship, for example tomorrow or several days or weeks ahead. According to Stopford (2009), this decision problem requires a short-term view of the market and conventional forecasting techniques are not much helpful. Usually, little reliable data is available on such short time frames; traditionally, decision-makers have relied on the gut feeling of shipbrokers working the market and their own intuitive models. However, modelling is not entirely out of the question for companies or pools with a strong information base (Stopford, 2009).

When applying a suitable machine learning technique instead of conventional forecasting techniques, the availability of AIS data is expected to compose a strong enough information base to provide good forecasts on these short time frames. Additionally, if the developed approach turns out to be promising, the length of the forecast could perhaps be extended in the future, to for instance a few months ahead, because the quantity and quality of AIS-data is expected to increase with time (See Chapter 3). Longer time frames are relevant for another fundamental shipping decision: whether and when to time-charter a ship. This decision-problem focuses on the probable future level of spot earnings over the time-charter period, compared with available time charter rate and the residual value of the ship at the end of the charter (Stopford, 2009).

[2] The Baltic Exchange provides a dirty tanker base index on the TD2 route for the settlement of freight derivatives agreements.

1.5

Thesis Structure

Inspired by the widely used *IMRAD (Introduction, Methodology, Results and Discussion)* style, the rest of the thesis is structured as follows: Chapter 2 reviews relevant literature, Chapter 3 introduces the reader further to AIS data and Chapter 4 provides insight to the world of tanker shipping and maritime economics. Chapter 5 then presents a summary of academic discussions and conclusions regarding tanker freight rate determinants, before Chapter 6 describes the selection, preparation and handling of data, in addition to descriptive statistics. In respective manner, Chapter 7 and 8 presents machine learning theory and the applied methodology. In Chapter 9, the results are presented and discussed, before a conclusion is drawn and potential further work is discussed in Chapter 10.

2

Literature Review

This thesis encompasses a wide range of disciplines, including shipping, maritime economics, empirical and quantitative methods in finance, big data handling and machine learning. The following literature review starts with a review of previous research on tanker freight rate prediction, continuing with research on the application of machine learning as a forecasting tool in shipping markets and ends with relevant applications of AIS data [3] .

2.1

Freight Rate Modelling Techniques

The development of modelling techniques, or modelling approaches, researchers in the past have applied to forecast or indicate tanker freight rates will be discussed in this section [4] . The academic reviews by Ringheim and Stenslet (2017) and Molvik and Stafseng (2018) reveal that early research generally focussed on structural models with a manifold of variables (see e.g. Koopmans (1939); Zannetos (1966); Strandenes (1984)). According to Ringheim and Stenslet (2017), the research by Beenstock and Vergottis (1993) together with development in econometrics in the 1990s, made researchers shift focus towards advanced time series models.

Several studies have examined the time-series properties of freight rates, such as their dependence on past values (Alizadeh and Talley, 2011). They have utilised univariate or multivariate time-series models, built on aggregate data and macroeconomic variables, to attempt to capture the dynamics and fluctuations of shipping freight rates and accordingly use the models for forecasting purposes (see e.g. Veenstra and Franses (1997); Alizadeh and Kavussanos (2002); Kavussanos (2003); Adland and Cullinane (2005, 2006); Batchelor et al. (2007)).

As opposed to structural models, which were used in much of the early research, recent studies have devoted more attention to the co-integration relationships of variables. Some recent studies have also used time-series models to investigate the time-varying structure and non-linear dynamics of freight rates (see e.g. Adland and Cullinane (2005, 2006) and Alizadeh and Talley (2011)). Furthermore, recent econometrics time-series models include e.g. Vector Autoregressive models (VAR) (Kavussanos and Alizadeh-M, 2001; Veenstra and Franses, 1997), varieties of Autoregressive Conditional Heteroskedasticity models (ARCH) (Kavussanos, 1996; Kavussanos and Alizadeh-M, 2002), as well as other stochastic time-series models (see e.g. Benth et al. (2014) and Askari and Montazerin (2015)). Two even more recent and relevant studies, predicting freight rates in the tanker market, are the masters theses by Ringheim and Stenslet (2017) and Molvik and Stafseng (2018). In consensus with the conclusions of Kavussanos and Alizadeh-M (2001) and Kavussanos and Alizadeh-M (2002), Ringheim and Stenslet (2017) discovered that the best out-of-sample result in terms of predictive accuracy was achieved using a univariate seasonal model. Molvik and Stafseng (2018) found inspiration in research fields such as oil and electricity price modelling. This led them to explore a combination of Markov regime-switching and multiple regression applied to specific crude oil and oil products tanker routes, with some promising results.

According to Alizadeh and Talley (2011), the abovementioned freight rate prediction models (in studies published before 2011) have underperformed. They suggest that the underperformance may be attributed to the utilisation of aggregate and macroeconomic data. They also argue that such macro-forecasts could be useful for medium to long-term investment purposes, while shipowners and charterers need micro-forecasts, e.g. of freight rates for specific routes, for making operational decisions, cash flow analyses, and budgeting. Therefore, Alizadeh and Talley (2011) investigated the importance of vessel and contract specific factors in the determination of tanker freight rates and laycan periods in shipping contracts using a system of simultaneous equations.

Kavussanos and Visvikis (2006) reveal that much of the existing literature on freight rate forecasting is concerned with the relationship between forward rates and future spot rates. According to Olsen and da Fonseca (2017), such studies find theoretical support in the unbiasedness hypothesis, in which forward rates are expected to be an unbiased predictor of future spot rates in markets for non-storable commodities. For example, Kavussanos et al. (1999) used a Vector Error Correction Model (VECM) to examine the lead-lag relationship between spot rates and the former BIFFEX futures contracts. They found VECM to perform relatively better than univariate ARIMA and random walk in forecasting spot freight rates, which indicate that such contracts fulfil their unbiased role. In line with Kavussanos et al. (1999), Batchelor et al. (2007) found evidence in favour of the unbiasedness hypothesis when investigating whether Freight Forward Agreements (FFAs) contain information about future spot rates.

2.2

Machine Learning and Maritime Forecasting

Previous research papers providing insight in using artificial intelligence (AI) or, more specific, machine learning (ML) as a forecasting tool in shipping markets are for example Lyridis et al. (2004), Leonov and Nikolov (2012) and Fan et al. (2013).

Lyridis et al. (2004) used monthly data in the period October 1979 to December 2002 and artificial neural networks (ANNs) to attempt to forecast Very Large Crude Carrier (VLCC) spot freight rates. Specifically, they analysed this period to detect possible causes of fluctuations and determine the independent variables, and then use them to construct reliable ANNs. They aimed to reduce error while

[3] Chapter 3 introduces the reader further to AIS data.

[4] A thorough review of freight rate modelling factors or freight rate determinants can be found in Chapter 5.

allowing the model to maintain a stable error variance during high volatility periods, and found that "*ANNs can, with the appropriate architecture and training, constitute valuable decision-making tools especially when the tanker market is volatile; the use of variables in differential form enhances the ANN performance in high volatility periods while variables in normal form demonstrated better performance in median periods; ANN demonstrated mean errors comparable to the naive model for 1-month forecasts but significantly outperformed it in the 3-, 6-, 9- and 12-month cases; finally, the use of informative variables such as the arbitrage between types of crude oil as well as Capesize rates can improve ANN performance (Lyridis et al., 2004)*".

Leonov and Nikolov (2012) used a wavelet neural network (WWN) model to study fluctuations in the freight rates of the Baltic Panamax route 2A and 3A. They argue that using wavelet transformations, trends such as clustering and rupture effects can be captured in the data, which gives valuable insight in the freight rates volatility dynamics. They further argue that their model can be used as a spot price discovery tool, if applied to the implied volatility and derivative contracts.

Fan et al. (2013) also used WWN in their study, but here applied on the Baltic Dirty Tanker Index (BDTI). Fan et al. (2013) used six input time-series, including AMEX Oil index and Dow Jones Industry Average Index as inputs in a WNN model to predict the BDTI index. They compared the forecasts by the WNN with ARIMA forecasts and found that there is no significant difference in the performance between the ARIMA and the WNN model. Regarding longer periods, however, the WNN model showed some superiority.

When it comes to previous research on LSTM Neural Networks applied to maritime forecasting, none directly relevant papers have been found. However, books such as Deep Learning by Goodfellow et al. (2016) and Deep Learning with R by Chollet and Allaire (2018) are useful with respect to the objectives of the current thesis.

2.3

Relevant Applications of AIS Data

Since the use of satellites to receive AIS data is a relatively recent development, the amount of literature on relevant applications of AIS data is limited. However, with more AIS satellites being launched, both the quality and availability of AIS data will be improved. Thus, an increasing number of maritime researchers, shipping market analysts, port authorities, to mention a few, are now directing their attention towards the possibilities offered by AIS data. Most AIS based research thus far have focussed on safety and environmental issues. Examples of application areas are utilisation of AIS data to mitigate accidents at sea and operate ships more efficiently to save fuel costs and reduce emissions (Leonhardsen, 2017; Smestad, 2015). Two important studies on the latter issue are the Third IMO Greenhouse Gas Study 2014 by Smith et al. (2015) and the Assessment of Shippings Efficiency Using Satellite AIS data published by Smith et al. (2013).

With respect to the objectives of the current study, the masters thesis by Olsen and da Fonseca (2017) is highly relevant. In the case of crude oil tanker rates on the TD3 route from the Arabian Gulf (AG) to Chiba in Japan, they investigated the predictability of AIS data using VAR and ARIMA models. Their results suggest that multivariate models perform relatively better than univariate models in forecasting freight rates. Moreover, they found weak evidence in favour of including information about tonne-mile demand and operational efficiency, derived from AIS data, in the forecasting models.

Besides Olsen and da Fonseca (2017), the relevant literature on the application of AIS data is focused on ship movements based on arrival and departure records for each ship (Kaluza et al., 2010), aggregation and visualisation of real-time seaborne trade flows on the individual ship level (Jia et al., 2017), the reliability of AIS-based trade volume estimates (Adland et al., 2017), operational issues like ship speeds (Adland and Jia, 2016, 2018; Leonhardsen, 2017), ship capacity utilisation (Adland et al., 2016) and onboard cargo sizes (Jia et al., 2015).

On trade flows, the current literature is generally concerned with either ship routing and network analysis at the micro level (see e.g. Kaluza et al. (2010)) or theoretical models of international seaborne trade at the macro level (Adland et al., 2017). According to Adland et al. (2017), there is a gap in the literature between the macro-level and micro-level studies, and a clear need to employ bottom-up analysis in studies of international seaborne trade. In this respect, the study by Adland et al. (2017) contributes to the literature as the first comparison of aggregate AIS-derived trade volumes with official export statistics based on customs data. They also show how the use of AIS data allows for the construction of a richer data set on seaborne trade, and argue that this is beneficial for the estimation of shipping demand, as the distribution of vessel sizes by load country and across time can be properly accounted for. Researchers must often take a simplified view of demand dynamics due to the lack of accessible disaggregate data for tonne-mile ship demand, and any research that deals with market fundamentals could in general benefit from AIS-derived tonne-mile demand data Adland et al. (2017).

As mentioned in the introduction, operational efficiency (or fleet productivity) affects supply and depends on for example ship speed and utilised cargo-carrying capacity. In this thesis, the tanker fleet's utilised capacity may be measured by the aggregation of a capacity utilisation ratio (load factor) multiplied by the onboard cargo-carrying capacity (dead weight tonnage, dwt) for each ship. Respectively, the studies by Smith et al. (2013) and Jia et al. (2015) propose models for the estimation of capacity utilisation and the cargo size onboard a ship primarily based on the ship's draught [5]. The draught models by Jia et al. (2015) are based on a large sample of dry bulk cargoes from port agent reports, and predict cargo sizes with a standard deviation of 7-9%. Smith et al. (2013) used AIS-reported draughts, and found the capacity utilisation to be approximately 80% or less in the case of crude

[5] A ship's *draught* may be defined as the vertical distance between the ship's waterline and the bottom of the ship's outer hull (*keel*) (Magnussen et al., 2014)

oil tankers. Such models may be useful in the absence of data on cargo size and capacity utilisation, because a ships draught is directly observable in the AIS data. A drawback with this approach is that it relies heavily on the accuracy of the manual draught reports by the ships crew. The quality of the AIS data also depends largely on the geographical coverage of the AIS data (Adland et al., 2017; Jia et al., 2015). It may be less accurate in congested or highly trafficked areas, such as important ports, straits and canals, due to interference of AIS signals (Smestad, 2015).

3

AIS Data

As mentioned in the introduction, IMO adopted new regulations in 2004, forcing all international voyaging vessels with above 300 Gross Tonnage (GT) and all passenger ships to carry an *Automatic Identification System (AIS)* transmitter for the reporting of ship positions to other ships and coastal authorities (Smestad, 2015). In this chapter, we briefly present the AIS technology and the content of AIS messages.

AIS is a ship-to-ship and ship-to-shore reporting system, i.e. an automatic ship tracking system, intended to increase safety of life at sea by improving the monitoring and control of maritime traffic (Skauen et al., 2013). AIS signals are transmitted as very high frequency (VHF) radio waves and captured by terrestrial land-based antennas (T-AIS) and Low Earth Orbit satellites (S-AIS). The introduction of S-AIS around 2010 enabled the reception of global AIS messages. The messages include static information like ship type and the ships identity number (e.g. MMSI), voyage-related data such as draught, heading, cargo and estimated time of arrival, and dynamic data like speed, course and rate of turn (Adland et al., 2017). This information is broadcasted to ships and shore stations within range of the VHF transmission. AIS messages reach about 70 kilometres (km) horizontally at sea level. In vertical direction, the signals can reach AIS receivers on satellites up to 400 km (Skauen et al., 2013).

The International Telecommunication Union (ITU) has defined 27 different types of AIS messages (ITU, 2014). The five most common ones are listed in Table 1, where *ID* in the leftmost column refers to the message type number. Additionally, in respective manner, Table 2 and 3 present the contents of message type 1 and 5. According to Smestad (2015), 72.5% of all AIS messages are of message type 1. Detailed information on the content of the different AIS message types is provided by the International Maritime Organisation (IMO, 2004).

Furthermore, the ship type reported in message type 5 is a double-digit number between ten and 99. As shown in Table 4, the first digit represents the ship type, while the second digit may state whether a cargo is dangerous, hazardous or a marine pollutant (USCG, 2018). Moreover, the transmitting frequency of AIS messages vary. Regarding dynamic data, the frequency depends on the ship's operational status. The different time intervals are shown in Table 5. When it comes to static and voyage-related data, messages are sent every sixth minute or upon request.

Table 1. *The five most common AIS message types.*

ID	*Name*	*Description*
1	Position report	Scheduled position report
2	Position report	Assigned scheduled position report
3	Position report	Special position report
4	Base station report	Position, UTC, date and current slot number of base station
5	Static and voyage report	Scheduled static and voyage-related ship data report

Table 2. *The content of AIS message type 1.*

Information	*Description*
Unixtime	Number of seconds elapsed since 1 January 1970
Position	Coordinates, longitude and latitude
Speed	Speed over ground (SOG), measured in knots
Course	Course over ground (COG)
MMSI	Maritime Mobile Service Identity (Ship ID)

Table 3. *The content of AIS message type 5.*

Information	*Description*
Unixtime	Number of seconds elapsed since 1 January 1970
Ship specifications	Length and breadth, measured in meters
Draught	Current draught, measured in meters
IMO number	International Maritime Organisation number (Ship ID)
Origin	Origin of current voyage
Destination	Destination of current voyage
ETA	Estimated time of arrival, measured in Unixtime
MMSI	Maritime Mobile Service Identity (Ship ID)
Ship type	Ship type category

Table 4. *Ship types reported in AIS message type 5.*

First digit	*Ship type*
1	Reserved for future use
2	Wing in ground (WIG)
3	Other vessels
4	High-speed carrier or vessels < 100 Gross Tonnes
5	Special craft
6	Passenger ships > 100 GT
7	Cargo ships
8	Tankers
9	Other types of ships

Table 5. *AIS reporting interval.*

Operational status	*General reporting interval*
At anchor	3 min
0-14 knots sailing speed	12 sec
0-14 knots sailing speed and changing course	4 sec
14-23 knots sailing speed	6 sec
14-23 knots sailing speed and changing course	2 sec
> 23 knots sailing speed	3 sec
> 23 knots sailing speed and changing course	2 sec

4

The World of Tanker Shipping and Maritime Economics

This chapter aims to provide a fundamental understanding of the economics of the shipping industry, with a focus on the supply-demand dynamics of the tanker market. Gaining this insight is a natural step in the process of identifying relevant freight rate determinants.

4.1

The Shipping Market Model

Figure 1 illustrates the classic maritime supply-demand model on a macroeconomic level. It is known as the shipping market model (Stopford, 2009). A more detailed version of the model can be found in Appendix A.

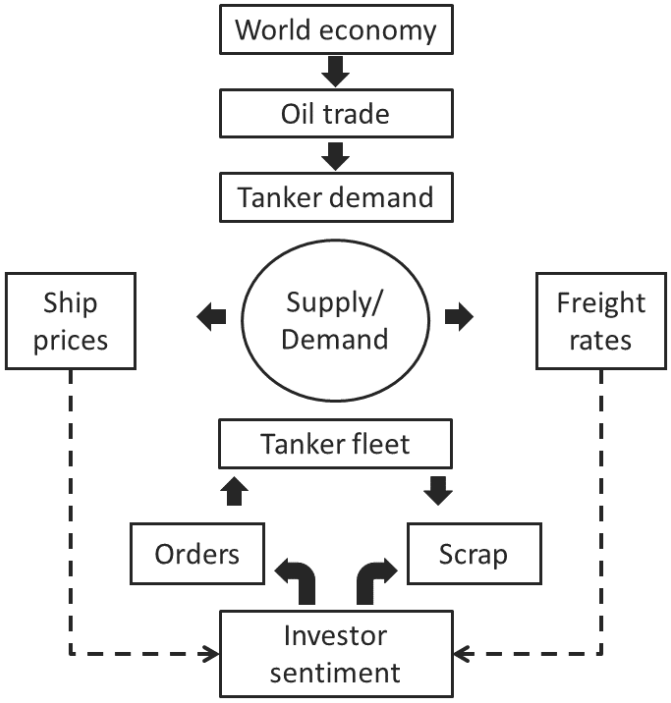


Fig. 1. Inspired by Stopford (2009) *The shipping market model adapted to the tanker market.*

As the above model illustrates: the tanker freight market is characterised by the interplay between supply and demand for tanker shipping services. Tanker shipping companies make a living by transporting liquid bulk from one terminal to another, whereas the income from seaborne transportation, determined by the freight rate, is the main driver in shipowners positive cash flow [6] . Freight rates thus constitute the link between supply and demand (Koopmans, 1939; Zannetos, 1966; Stopford, 2009).

4.2

Global Oil Production and -Tanker Trade

"*Maritime economics is a practical discipline, and there is not much point in being an expert on the economics if we cannot find the ports on a map! (Stopford, 2009)*". This statement implies that it is important to be aware of the worlds largest producers, exporters and importers of oil, and their geographical distribution, when investigating supply-demand dynamics of the tanker market; these actors have significant impact on the global seaborne trade volume and trade dynamics.

4.2.1

The World's Oil Production

Respectively, Figure 2 and 3 show the top five OPEC [7] and non-OPEC oil-producing countries during March 2018. The measurements are in million barrels per day (mbpd) and in percentage of total OPEC and non-OPEC production. The numbers in Figure 2 and 3 reveal that the U.S., Russia and Saudi Arabia are the three dominating countries in the world in terms of oil production. Saudi Arabia is a part of the AG and the Middle East, together with four other major crude oil-producing OPEC countries: Iraq, Iran, United Arab Emirates (UAE) [8] and Kuwait. Thus, according to the numbers in Figure 2, the AG is by far the largest oil-producing region in the world with a production of 24 mbpd.

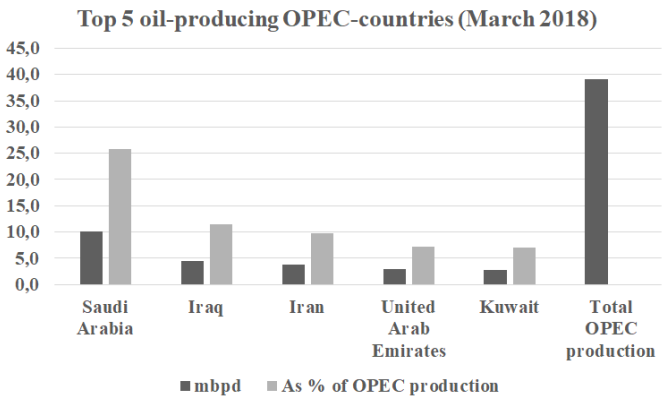


Fig. 2. Thomson Reuters Eikon (2018) *Top five oil-producing OPEC countries during March 2018.*

[6] This may vary based on the market situation. E.g. if a shipowner in a period sells a ship in the second-hand market or scraps a ship in the demolition market, this can dominate the positive cash flow.
[7] OPEC is the abbreviation of the Organization of the Petroleum Exporting Countries (http://www.opec.org/opecweb/en/).
[8] UAE is a federation of seven emirates, including Dubai and the capital of the federation, Abu Dhabi.

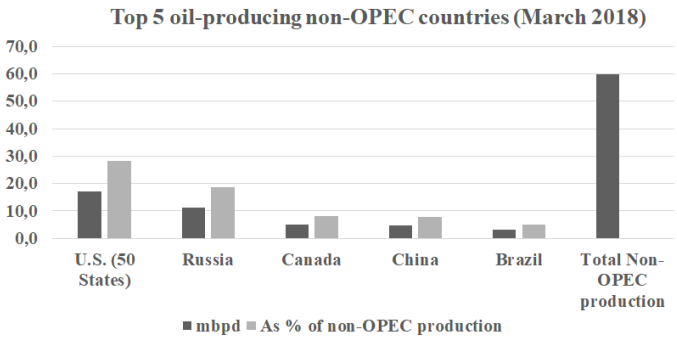


Fig. 3. Thomson Reuters Eikon (2018) *Top five oil-producing non-OPEC countries during March 2018.*

4.2.2

The World's Oil Export and -Import

Table 6 and 7 re-present data provided by Clarkson Research Services Limited, originally collected and presented by Molvik and Stafseng (2018). According to Table 6, the Arabian Gulf (AG)[9] is by far the largest crude oil *exporting* region in the world measured by tonne-miles. Regarding crude oil *imports*, also measured by tonne-miles, Table 7 shows that China is almost twice as important as North America and almost twice as important as Japan and India combined.

Table 6. Molvik and Stafseng (2018) *Seaborne crude-oil trade in 2017, measured in billion tonne-miles.*

Region	*Export*	*Region*	*Import*
AG	5,178	China	3,030
Caribs.	1,399	N. America	1,721
WAF	1,320	Japan	1,000
U.S.	242	India	732
UK/Cont.	217	UK/Cont.	577

Table 7. Molvik and Stafseng (2018) *Seaborne oil-products trade in 2017, measured in billion tonne-miles.*

Region	*Export*	*Region*	*Import*
Far East	203	China	216
USG	190	N. America	136
UK/Cont.	149	Japan	94
AG	137	India	90
Baltic	101	UK/Cont.	89

4.2.3

Global Tanker Trade

When it comes to seaborne crude oil trade, there literally has been an enormous development in terms of vessel sizes and technicalities, since the first seaborne transportation of oil found place in 1861 to the first Very Large Crude Carrier (VLCC) sat sail in 1966 (Molvik and Stafseng, 2018). The first tankers carrying oil in bulk, using the outer ship-hull as part of tank compartments, were not seaborne before 1886. After the Suez Canal opened for tanker trade in 1892, voyage distances were significantly shortened. Since then, the Suez Canal has contributed to a cyclical tanker market through several shut downs, re-openings and enlargements. In the mid-1900s, tonne-miles demand and average haul grew rapidly, especially due to an increased oil trade flow from the Middle East to Western Europe through the Suez Canal. At the time, oil majors were shipowners as well. They faced high costs by building supertankers such as VLCCs in the mid-1960s, exploiting the economies-of-scale principle.

Onwards from the mid-1960s, there was a rapid growth in the tanker fleet, until a decline in oil trade led the tanker market into recession and eventually a crash in the late 1970s. The average haul was reduced by the introduction of short-haul trade in e.g. the North Sea, openings of Middle East refineries and pipelines, increased domestic production in importing regions such as North-America, and the re-opening of the Suez Canal in 1975 following the closure in 1967. This was advantageous for the smaller ship classes. However, the freight market improved again in 1986, as low oil prices supported oil demand from the Middle East. From 1990 up until today, seaborne oil trade has grown significantly despite some cyclical downturns. Regarding oil products trade, which is related but different from crude oil trade, the interested reader is referred to the master's thesis by Molvik and Stafseng (2018).

Today, seaborne oil trade is more speculative than in the early days; it is subject to a highly volatile and market-regulated industry, influenced by a manifold of different stakeholders all over the world. Trading patterns are affected by the geographical distribution and availability of natural resources, which have resulted in an extensive network of both crude oil and oil products trade routes. The most traded routes in the world are commonly referred to by route codes, presented by the Baltic Exchange indices (see Ch. 4.6 and Appendix B for more information and an overview of the routes). Respectively, the TD and TC codes refer to the major tanker routes listed under the Baltic Exchange Dirty Tanker Index (BDTI) and Baltic Exchange Clean Tanker Index (BCTI).

4.3

Tanker Demand

4.3.1

Tanker Demand in Numbers

The tanker shipping market is by far the largest shipping sector of the world in terms of trading volume and weight. During 2008, the tanker fleet transported 2795 million metric tonnes (mmt) of liquid bulk commodities worldwide, out of which 2043 mmt was crude oil and 752 mmt was petroleum products (Alizadeh and Talley, 2011). In 2016, these figures were 1942 mmt and 1069 mmt,

[9] The Arabian Gulf is also commonly referred to as the Persian Gulf.

corresponding to 24.6% and 13.5% of the total bulk trade that year, respectively (Ringheim and Stenslet, 2017) [10] .

To properly quantify the demand for seaborne transportation, one must account for the distance over which the cargo has been transported (commonly referred to as the *haul*). This is done by multiplying the average haul of the trade by the amount of respective cargo traded, resulting in the transport demand being measured in *tonne-miles*. The so-called *distance effect* on demand is revealed by the Table 8, which presents an overview of annual tonne and tonne-mile demand in 2016 for the most important bulk cargoes. For example, crude oil's share of total volume is larger than its share of tonne-miles, implying that crude oil tankers' average haul is shorter than the average for total bulk cargo (Ringheim and Stenslet, 2017).

Table 8. Ringheim and Stenslet (2017) *2016 annual seaborne bulk trades in tonnes and tonne-miles.*

Cargo	*Million tonnes*	*Share*	*Billion tonne-miles*	*Share*
Crude oil	1,942	24.6%	9,399	23.7%
Iron ore	1,418	18.0%	8,035	20.3%
Coal	1,130	14.3%	4,903	12.4%
Grain	471	6.0%	3,376	8.5%
Minor bulk	1,860	23.6%	10,819	27.3%
Oil products	1,069	13.5%	3,104	7.8%
Total bulk trade	7,890	100%	39,636	100%
Total seaborne trade	11,101	13.5%	54,936	

4.3.2

Demand Dynamics

On a macro level, the demand for tanker shipping services is derived from the international oil trade, which in turn is dependent on world economic activity and imports and consumption of energy commodities (Alizadeh and Talley, 2011; Stopford, 2009). Stopford (2009) points out Gross Domestic Product (GDP) as the single most important indicator for future ship demand, but he also argues that development in seaborne commodity routes and trades are principal indicators. In addition, Stopford (2009) states that the following three factors have significant influence on ship demand: average haul, political events and transport costs (Jugović et al., 2015; Stopford, 2009). Regarding tanker demand, Stopford (2009) comments on the relationship between tanker freight rates and the oil price. He states that an increase in the oil price tends to alter the global energy mix, which reduces demand in the tanker market because coal to a certain extent substitutes oil. Moreover, the global economic growth usually fluctuates in periodic movements referred to as business cycles, and thus seaborne trade should approximately follow the same pattern. The interaction between consumption and investment, time-lags between economic decisions and implementation, and build-up of inventories are among the causes of business cycles (Jugović et al., 2015; Ringheim and Stenslet, 2017; Stopford, 2009).

[10] These statistics as well as the numbers in Table 8, are originally provided by Clarkson Research Services Limited and collected by Ringheim and Stenslet (2017), for tankers with more than 10k dwt capacity.

4.4

Tanker Supply

4.4.1

The Tanker Fleet

Early 2018, the cargo-carrying capacity of the world tanker fleet was 561 million tonnes (million dwt) and the number of tankers exceeded 6100. This accounts for approximately 40% of the total world shipping fleet. As Table 9 reveals, the tanker fleet consists of different size segments, some named after the canals the tankers are able to transit. For example, Panamax ships can transit the Panama Canal. VLCCs [11] are the largest tankers with a capacity of more than 200,000 dwt. Furthermore, crude oil is the most important liquid bulk cargo with more than 2200 ships and a cargo-carrying capacity of more than 400 million dwt. This equals approximately 30% of the total world shipping fleet. Crude oil is transported in tankers from production facilities to refineries, where it is refined into gasoline and other petroleum products. These products are then transported in smaller tankers, called product or clean tankers, to their destinations (Ringheim and Stenslet, 2017).

Table 9. Thomson Reuters Eikon (2018) *The world's tanker fleet by January 2018.*

Tanker type	*No. tankers*	*Tanker size (k.dwt)*	*Fleet size (mill.dwt)*	*Main cargo*
VLCC	720	200+	216	Crude oil
Suezmax	559	120−200	136	Crude oil
Aframax	955	80−120	94	Crude oil, oil products
Panamax	433	60−80	90	Oil products
Handysize	3439	10−60	89	Oil products
Tanker fleet	6106		561	

4.4.2

Supply Dynamics

As illustrated in Figure 4, the shipping industry deals with four closely related and dynamic markets in which shipping companies can participate: the freight market, the newbuilding market, the second-hand market and the demolition market. Together these markets form the supply characteristics of shipping (Stopford, 2009).

Shipowners are the primary decision-makers in the four shipping markets. They order new ships from the shipyard, buy and sell used ships with other shipowners in the second-hand market, trade ship services with charterers in the freight market, and sell old ships to scrapyards in the demolition market (Stopford, 2009). It is well recognised in the literature that shipping cycles, as illustrated in Figure 5, are driven by the way shipowners trade in these four

[11] Ultra Large Crude Carriers (ULCCs) are included in this segment.

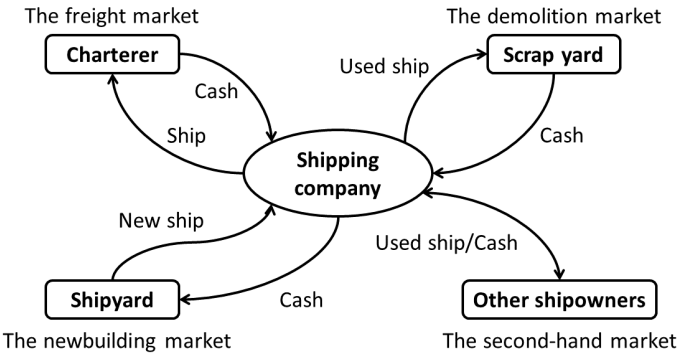


Fig. 4. Stopford (2009) *The four shipping markets.*

markets (Lun et al., 2010; Randers and Göluke, 2007; Stopford, 2009). The lag between investment decisions and market impact intensify the shipping cycles, and each cycle is different in terms of nature, circumstances, magnitude and length. There are no general rules about the timing and length of the cycle periods. Therefore, it is difficult for shipping market participants to determine cyclical turning points (Hawdon, 1978; Randers and Göluke, 2007; Stopford, 2009).

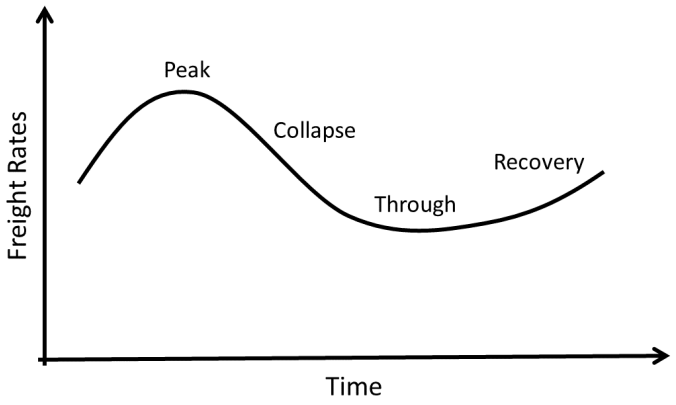


Fig. 5. *The four shipping cycle stages.*

According to Stopford (2009), *short-term* [12] shipping cycles range from four to 12 years, averaging eight years, and have the following four stages: *through* (when supply-demand ratio is at a maximum level and freight rates are at a minimum level), *recovery* (when the market adjusts towards supply-demand balance and freight rates increase), *peak* (when demand-supply ratio is at a maximum level and freight rates are at a maximum level) and *collapse* (excess supply and falling freight rates). At the recovery stage in the shipping cycle, market optimism starts to grow after a period of low freight rates, because the future appears brighter in terms of higher demand and higher freight rates. In this case, shipowners are stimulated to provide more transport. Consequently, second-hand prices and newbuilding orderbooks tend to increase. Eventually, when the newbuilds enter the market (in the order of 2-3 years), the peak of the shipping market is usually reached, and supply begins to offset demand. This puts downwards pressure on freight rates, making returns less profitable. Furthermore, the market moves into recession and eventually collapses, resulting in a through with longer periods of low freight rates. This results in a distressed environment, where some shipowners may be threatened by bankruptcy. Shipowners then sell ships cheaply in the second-hand market or scrap them, to become more profitable and to increase liquidity. The through continues until enough ships have exited the market and supply falls to the levels of demand, or until demand surges and reaches the levels of supply (Ringheim and Stenslet, 2017; Stopford, 2009).

Ringheim and Stenslet (2017) argue that the second-hand market plays a vital economic role in the shipping industry, because it allows direct entry and exit to the freight market. This means that ships quickly can shift owners in this market, as opposed to the newbuilding market where it can take up to three years to build a new ship. Thus, by inspecting second-hand prices relative to newbuild prices, inferences about future market expectations can be done. For example, if the freight rates are low, the market can still expect the freight rates to be profitable enough the next few years, to justify the buy of a second-hand ship even though its life expectancy is relatively shorter than a newbuild.

Shipowners are not the only decision-makers that influence the supply in shipping. Three other groups are also important: *charterers* can influence shipowners by becoming shipowners themselves, like oil companies that ship their own crude oil; *banks* and *investors* finance the shipping industry and to a large extent determine shipowners leeway by their willingness to lend or invest; and finally, *regulators* can affect the industry by introducing new legislation and regulatory frameworks. Additionally, *shipyards* and *shipbrokers* are influential stakeholders in this industry. In conclusion, supply in shipping is behavioural and dependent on a small group of players. This latter fact makes shipping supply very prone to changes (Jugović et al., 2015; Ringheim and Stenslet, 2017).

4.5

Freight Rates

This section delves into the tanker freight market, and aims to explain the freight rate process, which is the mechanism that links shipping cycle theory with supply-demand theory.

4.5.1

Freight Market Participants

The freight market is one of the four shipping markets and normally the main driver in shipowners positive cash flow. In this market, shipowners receive payments from charterers, and charterers receive payments from shippers, in return for shipping services. The shipowner obviously owns the ship, the charterer is an individual or a company hiring the ship to transport the cargo, and the shipper is an individual or a company that needs the cargo shipped.

[12] Stopford (2009) also defines *seasonal cycles* and *long-term cycles*. This is discussed in Section 4.5.4.

The contract agreement that sets out the terms on which the shippers get the cargo shipped, or the terms on which the charterer hires a ship, is called the *charter-party*. The way it works is perfectly simple: The two parties negotiate contractual terms (the charter-party), maybe through a shipbroker acting as a link between them, to determine the freight rate. When the charter-party is agreed upon, the ship is said to be "fixed" (Furset and Hordnes, 2013). In theory, the agreed freight rate should reflect the balance of ships (supply) and cargo (demand) currently available in the market. However, many types of freight agreements, different ship sizes, various cargo types and numerous trading routes, et cetera, complicate the picture.

4.5.2

Freight Contract Agreements

There are four main types of charter-parties between shipping companies and charterers, all different in the way they distribute responsibilities and costs (Stopford, 2009).

The *voyage charter (VC)* contract provides transport for a specific shipload of cargo from a load port to a discharge port. A VC contract thus covers a single, route-specific voyage, and is therefore often referred to as a "spot contract". Furthermore, the shipper pays a price for the shipment to the charterer, who in turn pays the shipowner a pre-agreed freight rate on a per-tonne or a lump-sum basis. The contractual terms of the transport include, for example, the freight rate, load port, discharge port, cargo type, cargo quantity, speed, laytime [13] and demurrage [14]. A deviation from the charter-party may result in a claim. Regarding trade-related costs, the shipowner is fully responsible. For tankers, including VLCCs, VC contracts are the most common arrangement (Furset and Hordnes, 2013; Gilleshammer and Hansen, 2010).

Contract of Affreightment (CoA) is another type of contract agreement. CoA is a bit more complex than voyage charter, but very similar. With CoA, a series of cargo, for example one shipload each month over a few months ahead in time, are transported for a fixed price per tonne. Again, the shipowner pays all costs.

The third charter-party type, *time charter (TC)* contracts, are most common in the dry-bulk segment (Gilleshammer and Hansen, 2010). With TCs, the charterer pays the shipowner an agreed day-rate over a certain period and gets full operational control of the ship. The shipowner pays for most of the operational costs (OPEX), while the charterer pays the voyage costs (VOYEX) including port fees and bunker costs. TC contracts can be divided into two different contract types, namely spot and term (or period) contracts. The difference lies in the duration of the contracts, where spot contracts usually have some duration less than three months. They are normally written only a few days before the start of operations, and re-negotiated frequently. Thus, spot freight rates (or spot rates) typically vary from one day to the next (Molvik and Stafseng, 2018).

Finally, the fourth and last type of charter-party is the *bare boat charter (BBC)*. Here, the charterer gets full operational control of the ship, and the contract generally stretches over several years (Stopford, 2009).

4.5.3

Freight Costs

The costs of shipping normally determine the so-called *refusal rate* [15] and, accordingly, the ships'assumed lay-up point. The total costs consist of capital costs (CAPEX), operation costs (OPEX), voyage costs (VOYEX) and cargo-handling costs. However, these costs vary by the type of charter-party, as illustrated in Figure 6.

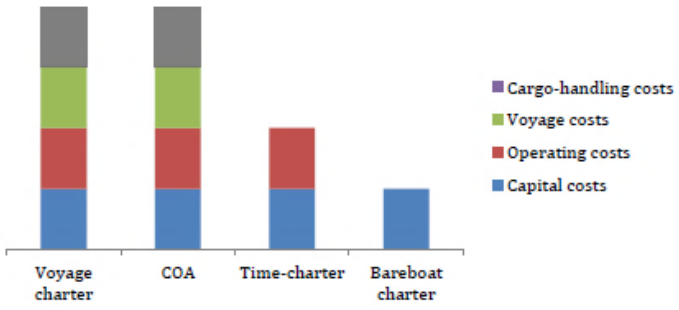


Fig. 6. Alizadeh and Nomikos (2009) *Shipping cost allocation from a shipowner perspective under different charter-parties.*

According to Alizadeh and Nomikos (2009), CAPEX includes interest and capital repayments on a ship, which are affected by the current market situation, the financial structure of the purchase, and future market expectations. OPEX is fixed, although the ship is inactive, and involves crew wages, maintenance, insurance, inspections and renewal of certificates. VOYEX incur for a specific voyage, and is mainly determined by fuel costs, canal dues, pilotage and port charges. Cargo-handling costs cover loading, stowage, lightering and discharging of the shipped cargo.

4.5.4

Freight Rate Dynamics

Stopford (2009) argues that freight rates are a mix of current and future expectations, thus it is important to be precise about which time-frame that is used when explaining freight rate dynamics. Three time-periods may be considered: *momentary*, *short-term* and *long-term*. Momentary (hours, days or weeks) is the time-scale of charterers, shipbrokers and traders, where a decision problem

[13] Laytime is the time window where the charterer load and discharge the cargo, without incurring extra costs.

[14] Demurrage is the daily amount paid to the shipowner by the charterer if the number of port days exceed the agreed laytime. Correspondingly, if the laytime exceeds the number of port days, the shipowner pays a despatch to the charterer.

[15] The refusal rate is the freight rate, subtracted the lay-up costs, where the shipowner is assumed to lay up the ship, because otherwise the ship operates with a loss. Thus, this is also known as the *lay-up point*.

can be whether to fix a ship on a period contract (in the order of months) or to operate in the spot market (in the order of days or weeks). In the short-term (up to a few months or few years), owners and charterers have time to respond to changes in the market by sending ships in or out of lay-up. In the long-term (several years), the fleet size can be adjusted by the ordering or scrapping of ships. According to Stopford (2009), these time frames constitute three markets with considerably different dynamics, and long-term is the only time frame that adequately can be explained by fundamental variables.

According to shipping cycle theory, the framework for the shipping cycle is set by the combination of a volatile demand and a significant time-lag for the supply to adjust accordingly. Stopford (2009) distinguishes between three shipping cycle periods: long-term, short-term and seasonal. Shipping cycles are periodic and not symmetric, implying that they last for various lengths of time and each cycle can be very different from the previous cycle. There will always be fluctuations in the balance between supply and demand, thus cycles are deemed to occur. Especially the short-term cycle, which can last three to 12 years, has the function of coordinating the supply-demand balance. Therefore, *short-term cycles* are noticeable and often subject to analysis. *Long-term cycles* on the other hand, can last from decades to entire centuries, driven by secular trends such as technical innovations (e.g. the steam engine invention, diesel engines replacing steam engines, containerization and the bulk shipping revolution). It is hard to identify exactly when freight rates are affected by long-term cycles. *Seasonal cycles* are seasonal changes in the freight rate within a year, mainly determined by the seasonal demand; supply does not change much within a year.

The *freight rate process* is the mechanism that links shipping cycle theory with supply-demand theory. From a microeconomic point of view, the freight rate process can be analysed using three key economic concepts: *supply function*, *demand function* and *equilibrium price* (Stopford, 2009). The supply function defines the shape of the supply curve and the demand function likewise shapes the demand curve. These two curves intersect at the equilibrium price, the theoretical freight rate equilibrium, where supply and demand is perfectly balanced. This is illustrated in Figure 7, where it also is shown how the demand curve shifts when supply changes.

In a perfectly competitive spot freight market, the theoretical freight rate equilibrium is normally determined by the marginal cost of the marginal ship required to satisfy the demand for seaborne transportation(Koekebakker et al., 2006). The J-shaped supply curve in Figure 7 is composed of each ship's individual supply curve and thus indicates the amount of transportation, measured in tonne-miles, the fleet is willing to supply at a given freight rate (Alizadeh and Nomikos, 2011). Classical maritime economic literature, as first introduced by Koopmans (1939) among others, characterizes the short-term supply curve by two distinct regimes: whether the fleet is fully employed or not.

When the available supply exceeds demand, i.e. when the fleet is not fully employed, the most cost-efficient ships will contribute to the lower left part of the supply curve. This is because freight rates are relatively low at this point. If

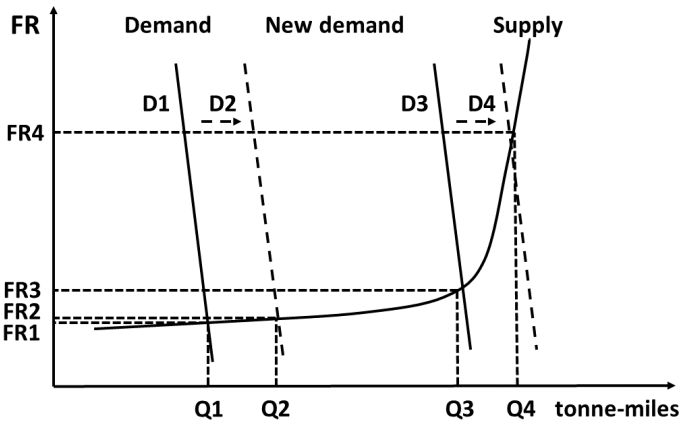


Fig. 7. Alizadeh and Nomikos (2011) *Supply-demand framework in shipping freight-rate determination.*

the freight rate drops below the acceptable level for a given ship, previously referred to as the ship's refusal rate and lay-up point, a decision must be made whether the ship in the short-term should be laid-up or operated with a loss. In this decision, switching costs related to laying up the ship should be considered too (Dixit, 1989; Koekebakker et al., 2006; Mossin, 1968). In a bit longer time-frame, say between one and three years, the shipowner may consider scrapping or selling the ship, if the freight rate stays below the lay-up point.

Conversely, if demand for transportation and hence freight rates increase, less cost-efficient ships enter the market. This leads to less ship unemployment and a series of perfectly elastic steps in the short-term supply function. The supply curve is price elastic up to the point where the fleet sails at close to maximum capacity. Here, the curve becomes inelastic, because the fleet cannot react to a short-term increase in demand. At maximum fleet capacity, where the fleet is fully employed, the only way to increase the supply of seaborne transportation is through higher utilisation of the existing ships (Koekebakker et al., 2006). This can be achieved by, for example, increasing ship speed, delaying regular maintenance, reducing port-time, shortening ballast legs and fully utilising the ship's cargo-carrying capacity; as shown in Figure 8, Adland et al. (2016) discovered that there is a positive correlation between a ship's capacity utilisation [16] and freight rates. Technical constraints and higher marginal operation cost, due to increased fuel consumption and more wear and tear, does however put a ceiling on this additional increase in supply (Adland et al., 2016; Koekebakker et al., 2006).

Furthermore, the demand curves in Figure 7 describe the required amount of supply from the operator at a given freight rate. As the figure illustrates, the demand curve is inelastic all the way, and demand can shift quite extensively compared to supply in the short-run. The reason for this is that operators are dependent on continuous delivery of supplies, because delay in operations (downtime) is very costly in terms of lost income. Moreover, it is not always in the interest of the shipping industry to operate at the

[16] Adland et al. (2016) defined a ship's capacity utilisation as the ratio of cargo size divided by dead weight tonnes (dwt).

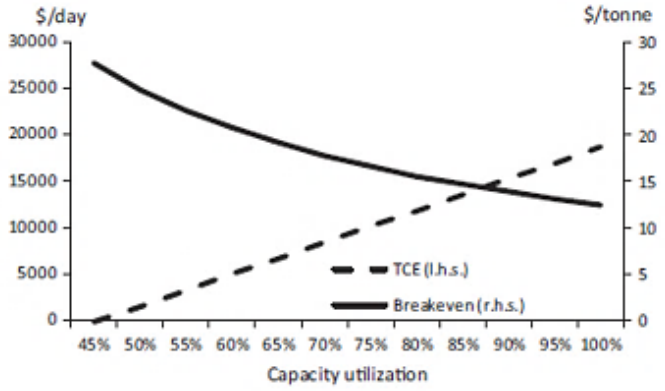


Fig. 8. Adland et al. (2016) *The relationship between shipping freight rates and capacity utilisation.*

theoretical freight rate equilibrium. Operating companies may have incentives to take on some of the risk in times of tough market conditions, and hence does not always benefit from negotiating the lowest possible freight rate. Over time it is beneficial for both shipowners and operators to keep supply at a certain level (Molvik and Stafseng, 2018).

Adland et al. (2016) discuss microeconomic analysis of freight rate formation for *individual* voyages, where the concern is about the immediate equilibrium between the number of cargoes and ships in a specific loading area. Cargoes need to be shipped within a certain time window (*laycan*), and hence supply is determined by whether ships are commercially available and physically able to meet the laycan. Subject to capacity constraints, cargoes are then matched with ships in either an auction-like process or a perfectly competitive micro-market, depending on whether there is excess demand or excess supply in the loading area, respectively. Furthermore, if we ignore speed changes and inter-temporal substitution, it follows that available supply (dwt capacity) is fixed in the short run (in the order of days). Thus, an increase in short-run demand and, consequently, a higher freight rate, will result in a higher contemporaneous capacity (dwt) utilisation. In the longer run (in the order of weeks) supply is not fixed; high freight rates in one region of the world would attract more ships; a scenario which could push freight rates and capacity utilisation down.

4.6

The Freight Derivatives Market

According to Alizadeh and Kavussanos (2002), spot and period rates (TC rates) are related through the term structure expectation hypothesis. Based on this argument, Alizadeh and Nomikos (2011) argue that period rates are a form of forward freight rates [17], and that forward freight curves thus can be constructed at any point in time by comparing spot and period rates with different durations.

Depending on the shape of the forward curve, which can change over time, the shipping market may be characterised as being normal, in *contango* or in *backwardation*. In contango, the spot rate is below long-term period earnings, while in backwardation the spot earnings are higher than period earnings. The forward curve is normally backwardated in commodity markets when there is a high temporal demand for shipping services (inelastic supply curve). Correspondingly, a contango forward curve can be associated with low temporal shipping demand (elastic supply curve).

The freight derivatives market, which is a part of the freight market, allows charterers and shipowners to hedge their freight risk or speculate by making Freight forward agreements (FFAs). FFAs are financial contracts settled against the value of a base index on the date specified in the agreement and traded in over-the-counter (OTC) markets. In the tanker market, a tanker FFA contract is an agreement between two parties to fix a freight rate in Worldscale units, over a period, on a predetermined tanker route. The fixed forward price is settled at the end of each month against the value of a tanker base index (e.g. the TD2 index for the route between the Arabian Gulf and Singapore). Tanker base indexes are published by the Baltic Exchange and classified under the Baltic Dirty Tanker Index (BDTI) routes (see Appendix B) or the Baltic Clean Tanker Index (BCTI) routes (see Appendix B) (Kavussanos and Visvikis, 2006).

Besides the OTC forward (FFA) contracts, shipping market agents can use freight futures, which are traded in organized exchanges such as the International Maritime Exchange (IMAREX) and New York Mercantile Exchange (NYMEX). According to Kavussanos and Visvikis (2006): "*IMAREX is a professional freight derivatives exchange for the maritime industry, founded in spring 2000. It provides a marketplace for freight derivatives (freight futures and FFAs) and in partnership with the Norwegian Options and Futures Clearing-House (NOS) offers clearing services for these derivatives*". The Oslo-based IMAREX utilises mostly the indices built by the Baltic Exchange, in addition to some indices from Platts, to write freight rate derivatives upon. IMAREX is accepted by the US Commodity Futures Trading Commission to operate an electronic facility as an Exempted Commercial Market (ECM) (Γολας, 2012). For contract details of IMAREX tanker derivatives, see Appendix B.

A third available derivatives tool for risk management and investment purposes, in addition to FFAs and freight futures, is the option contract, offered by the same brokers that trade FFA contracts. The first cleared tanker IMAREX Freight Option (IFO) contract, cleared through NOS, was launched in 2005 on route TD3 (VLCCs sailing on the trade lane between the Arabian Gulf and Japan). The IFOs are settled against the Baltic Exchange quotes, structured as monthly call and put Asian style options and available for trading and clearing for all IMAREX and NOS members. For the tanker routes, the settlement prices (measured in Worldscale points) are calculated as

[17] "*A forward contract is an agreement to buy or sell an asset at a certain future point in time at a certain price. Forward contracts are traded in the OTC-market, usually between two financial institutions, or a financial institution and a client. At maturity the buyer of the forward contract receives the underlying asset (Gilleshammer and Hansen, 2010)*".

the arithmetic average across all trading days in a calendar month (Kavussanos and Visvikis, 2006).

4.7

Market Efficiency in the Freight Market

Adland and Strandenes (2006) argue that the traditional form of the efficient market hypothesis (EMH) does not apply to the freight rate process, because the freight rate is the price of a transportation service that cannot be traded or stored. However, they argue that the notion of market efficiency still applies to the freight market: Under the hypothesis that the freight market is semi-strong-form efficient, i.e. the current price of an asset in this market incorporates all publicly available information, it should not be possible to earn excess profit by taking chartering positions based on public information. Such information could be past levels of the spot freight rate or the shape of the term structure of freight rates. Adland and Strandenes (2006) further discuss that the current spot tanker freight rate does not necessarily reflect all public information like, for instance, OPEC revealing today that the they will reduce the oil output in three months time. This is because spot freight rates are a result of the near-term (in the order of weeks) effective supply-demand balance in each loading area.

Based on the arguments above, Adland and Strandenes (2006) argue that the analysis of past price patterns, referred to as technical analysis, may contain useful information about future freight rate changes. To test market efficiency in the bulk freight market, Adland and Strandenes (2006) utilise technical analysis based on the history of spot freight rates and examine the profitability of chartering strategies for a tanker operator. Their empirical results suggest that a large tanker operator, with a pool of tankers, could have earned significant profits without investing in ships by trading on relevant public information (like past levels of the spot freight rate or the shape of the term structure of freight rates).

In line with the abovementioned findings, Stopford (2009) states that shipping markets operate under conditions of nearly perfect competition, meaning that shipping markets are close-to efficient. According to Investopedia (2018), five characteristics must be apparent for a market to have a perfect competition structure. These have been translated into the tanker market by Molvik and Stafseng (2018): *Firstly*, shipowners offer almost identical products, since tankers within the various classes are almost perfectly interchangeable. *Secondly*, all shipowners are price takers, since the freight mechanism determine the price. *Thirdly*, the tanker market is fragmented to a large extent, with no single shipowner owning more than 2.5% of the total tanker fleet capacity. *Fourthly*, freight rate manipulation is difficult, because the market is highly transparent. Charterers are fully informed about the tankers available for chartering and about the Worldscale rates charged on the various routes. *Lasty*, with the necessary financing in place, the tanker market is both easy to enter and exit. When it comes to freight rate prediction based on empirical data, these five market characteristics may serve as theoretical support and motivation, as they possibly indicate predictability of tanker market dynamics over time (Molvik and Stafseng, 2018).

5

Review of Tanker Freight Rate Determinants

As described in Chapter 4, the evolution of freight rates is determined by the interplay between supply and demand, both on a macroeconomic and microeconomic level. Additionally, financial and other non-fundamental determinants may impact the level of freight rates. A decisive step in the process of forecasting tanker freight rates is to identify key freight rate determinants. This section therefore provides a recap of academic discussions and conclusions regarding supply, demand, microeconomic, financial and other non-fundamental determinants.

5.1

Supply- and Demand-driven Determinants

Koopmans (1939), among others, lay the foundation for the dynamic supply theory of shipping. Several empirical models are built upon this theory, describing the dynamic relationship between freight rates and the supply of seaborne transportation (See e.g. Zannetos (1966), Hawdon (1978), Strandenes (1984), Hampton (1990), Beenstock and Vergottis (1989), Beenstock and Vergottis (1993), Engelen et al. (2006) and Randers and Göluke (2007)). These studies discovered, for example, that shipping freight rates are dependent on factors such as world economic activity, growth in industrial production, seaborne commodities trade, oil prices, availability of ship tonnage, newbuilds on order, newbuild deliveries and scrapping rates. In other words, these findings imply that freight rates are determined by the balance between demand and the active fleet size. Moreover, Beenstock and Vergottis (1989) published an econometric model for the tanker market, which was developed using theory from the model of expected second-hand ship prices presented by Beenstock (1985). Here, supply was modelled as a function of freight rates, fleet size, operational costs and costs of lay-up, while demand was modelled as exogenous.

Furthermore, the recent studies by Dikos et al. (2006) and Randers and Göluke (2007) apply macroeconomic variables in a system dynamic setting to model and forecast tanker freight rates. Like Beenstock and Vergottis (1993) and Engelen et al. (2006), Randers and Göluke (2007) treated demand as exogenous, while freight rates, newbuild orders, average building time, average lifetime of ships, fleet productivity, utilisation changes and scrapping rate were included on the supply side. Moreover, Randers and Göluke (2007) explain that supply (measured by tonne-miles) is flexible and affected by the way shipowners operate their ships; when there is excess demand, shipowners can improve profitability by, for example, speeding up their fleet, utilising more of their ships capacity, postpone regular maintenance and shorten port time. Thus, supply measured in tonne-miles is not constant, but rather dynamic and continuously adjusting to market conditions.

As shown in Table 10, Stopford (2009) explains the shipping market model by ten variables in total, five supply variables and five demand variables. Regarding fleet productivity on the supply side, Stopford (2009) argues in line with Randers and Göluke (2007) that it mainly depends on the following factors: speed, port time, fleet utilisation and loaded days at sea. Additionally, Stopford (2009) explains that supply in the long-run is determined by the fleet size, driven by scrapping and newbuild deliveries.

Table 10. Stopford (2009) *Ten variables in the Shipping Market Model.*

Demand	*Supply*
1. The world economy	1. The world fleet
2. Seaborne commodity trades	2. Fleet productivity
3. Average haul	3. Shipbuilding production
4. Political events	4. Scrapping and losses
5. Transport costs	5. Freight rates

When it comes to the tanker market, Lyridis et al. (2004) forecasted VLLC spot freight rates using Artificial Neural Networks (ANNs). In their study, they considered the following variables to be the most important factors: demand for oil transportation (measured in tonne-miles), active fleet, crude oil production, crude oil price, surplus as a percentage of active fleet, TC rates, newbuild prices, second-hand prices, bunker oil prices, scrap prices and oil stock building. Much in line with Lyridis et al. (2004), Alizadeh and Talley (2011) found that supply depends on, for example, the tanker fleet size, the tonnage available for trading, tanker shipbuilding activities, the scrapping rate of the fleet, bunker prices, and the tanker fleet productivity at any point in time.

Ringheim and Stenslet (2017) predict monthly dry bulk (BDI) and tanker (BDTI) freight rates, by applying a general-to-specific methodology as outlined by Campos et al. (2005), and using a total of 44 dry bulk and 37 tanker variables. Consistent with the findings of Poulakidas and Joutz (2009), Ringheim and Stenslet (2017) conclude that the single most significant tanker predictor is the oil price. The oil price being an important indicator for tanker demand is also pointed out by Stopford (2009).

Molvik and Stafseng (2018) forecast tanker freight rates on four major dirty tanker routes (TD1, TD3, TD7 and TD12) and on two major clean tanker routes (TC1 and TC2). These routes are explained in Appendix B. Furthermore, they included seven groups of supply-driven variables: five related to fleet size, one to fleet age and one to vessel prices. Six groups of demand-driven variables were included: oil demand, oil import and export, vessel fixtures, crude oil production, refinery output and refinery utilisation. Additionally, they included ten groups of economic and non-fundamental variables: GDP, TC rate, exchange rate, CPI and money supply, interest rate, industrial production, crude oil and oil products price, bunker price, shipping index and stock index. In total, this amounts to 169 variables.

Olsen and da Fonseca (2017) applied two sets of variables to their forecasting models (see Chapter 2): one set included what they refer to as "publicly available data", and

the other set included both "publicly available data" and AIS data. The "publicly available data" set consisted of seven variables:

- Average production volume of crude oil in the Middle-East (local demand)
- VLCC fleet development (global supply)
- VLCC tonnage available for chartering in the Arabian Gulf (local supply)
- Gasoline price in Singapore (local supply/demand)
- A variable called "diffussing" (probably accounting for changes in the US dollar−Singapore dollar exchange rate) [18]
- Refinery margin in Asia (local demand)
- Brent-Dubai crude oil price spread (local demand)

The AIS data set contained information supposed to capture the effects of operational efficiency and tonne-mile demand (both globally and locally). The following three determinants accounted for operational efficiency:

- Fleet speed
- Voyage speed
- Load factor

Furthermore, the following four AIS-derived determinants accounted for tonne-mile demand:

- The number of VLCCs in the Arabian Gulf (local tonne-mile demand)
- The number of tankers other than VLCCs in the Arabian Gulf (local tonne-mile demand)
- The number of VLCCs in other parts of the world (global tonne-mile demand)
- The number of tankers other than VLCCs in the rest of the world (global tonne-mile demand)

The latter four variables were included to attempt to accurately account for the distance-effect on demand, which in theory is possible using AIS data (Adland et al., 2017). As mentioned in Chapter 2, they found weak evidence in favour of including AIS-derived information about tonne-mile demand and operational efficiency in their forecasting models.

As opposed to Olsen and da Fonseca (2017), Tham (2008) applied Bayesian selection methods to identify leading predictors of TD3 front month swaps without using AIS data, resulting in the following significant price drivers: refining margin in Asia, crude oil production in the Middle East, capacity utilisation and Brent-Dubai crude oil price spread.

5.2

Microeconomic Determinants

The literature on *microeconomic* determinants of shipping freight rates is limited to the studies by Tamvakis (1995) Tamvakis and Thanopoulou (2000) and Alizadeh and Talley (2011). Tamvakis (1995) found no significant empirical results to support the hypothesis that the construction and employment of double-hull tankers, as mandated by the Oil Pollution Act 1990 (OPA), would create an additional freight rate premium for the hiring of double-hull tankers, relatively to the hiring of single-hull tankers. Tamvakis and Thanopoulou (2000) investigated the existence of a two-tier dry-bulk ship charter market, reflecting the ship age. This was based on the period 1989-1996, which covered different shipping cycle stages. However, their empirical results revealed no significant difference between freight rates paid for newer versus older ships.

Alizadeh and Talley (2011) investigated specific vessel and voyage determinants of shipping freight rates, as well as the timing of charter contracts (i.e. laycan periods) in the tanker freight market. Their estimation results indicate that the duration of the laycan period is a significant determinant of the shipping freight rate and vice versa. Other freight rate determinants include the ships hull type, fixture deadweight utilisation ratio, age of the ship, and voyage routes. For the laycan period, determinants include the former determinants in addition to the Baltic Dirty Tanker Index (BDTI) and its volatility.

5.3

Financial and Other Non-fundamental Determinants

The literature study by Ringheim and Stenslet (2017) reveals several relationships between shipping freight rates and factors that are not directly linked with the supply or demand of seaborne transportation. Ringheim and Stenslet (2017) define these factors as either *financial determinants* or *non-fundamental determinants*: Financial determinants are prices and measures that are determined by or traded in financial markets, and non-fundamental variables can for instance be political events, enforcement of new regulations, or wars taking place. Furthermore, Stopford (2009) highlights the importance of behavioural aspects and other shipping market participants like banks and regulators. Shipowners' and charterers' decisions are often influenced by the behaviour of these participants and, thus, both financial and non-fundamental determinants could impact the shipping market's expectations about future supply and demand.

The two studies by Alizadeh and Talley (2011) and Bakshi et al. (2011) both reveal a link between real and financial markets, by showing that the Baltic Dry Index (BDI) have predictive power in both markets. They both found that the BDI was an indicator for stock market returns, while Bakshi et al. (2011) also showed that the BDI could predict global economic growth and commodity indices. Moreover, Bakshi et al. (2011) discovered that their model was applicable across international stock indices.

Some recent studies have revealed that financial markets could lead shipping freight rates (Ringheim and Stenslet, 2017). As indicated in Chapter 2 and in Section 4.6, futures and forward contracts might contain information about future spot freight rates. Kavussanos and Alizadeh-M (2001) discovered for instance a bidirectional lead-lag relationship in daily returns and volatilities between

[18] This variable is not explained by Olsen and da Fonseca (2017). However, we assume that it accounts for the difference or ratio between US dollars and Singapore dollars and, thus, currency exchange rate effects.

spot freight rates and FFAs. Moreover, Kavussanos and Visvikis (2006) call for awareness regarding currency risk, as many shipowners have income in US dollars while their payments often are in a local currency.

Finally, the study by Fan et al. (2013) investigated the possibility of using the following variables to predict the BDTI: the oil price, the CBOE SPX Volatility Index, and the SP Global 1200 Index. Ringheim and Stenslet (2017) on the other hand, included a total of 19 financial and non-fundamental variables in their models to forecast the BDI and the BDTI.

6

Data

In this section, the input data and freight rate determinants are presented and reasoned. The process of gathering, filtering and transforming raw data from AIS data to proper time-series data is also briefly described. Finally, descriptive statistics of the explanatory variables are presented and discussed before their correlation with the dependent variable is studied.

6.1

Data Selection

The selected determinants, which may influence the tanker freight rate on the TD2 route, are summarised in Table 11. Eleven of these explanatory variables are derived from AIS data, while six of them are not. The time-series data for the latter variables are gathered from Thomson Reuters Eikon. Moreover, the freight rate in question, the dependent variable in this study, is a rate calculated by Thomson Reuters and reflects the price level on the TD2 route (for more information about this tanker route, see Chapter 1). Based on the review of freight rate determinants presented in Chapter 5, the arrows in the rightmost column of Table 11 indicate the impact a positive change in the respective explanatory variable is expected to have on the freight rate. Furthermore, several data sets of explanatory variables will be investigated in this study. This will be further explained in Chapter 8 and 9.

6.2

AIS-derived Data

6.2.1

Fleet Productivity (Supply)

Stopford (2009) lists fleet productivity as the second-most important supply determinant. Fleet productivity is also influential for tanker supply (Alizadeh and Talley, 2011). The fleet productivity, or operational efficiency, is assumed to be largely dependent on the sailing speed and the utilisation of cargo-carrying capacity (dwt). The average speed of the global VLCC fleet (`Speed_global`) and the local VLCC fleet (`Speed_local`) are therefore included as two separate supply-determinants to capture the effects of global and local speed changes. Additionally, two load factor determinants are included to account for capacity utilisation of the VLCC fleet globally (`LF_global`) and locally (`LF_local`). Here, local refers to the VLCCs currently located in the areas around and between the origin and destination ports and global refers to the VLCC world fleet, i.e. we account for all existing VLCC ships. The rationale for including global variables is simply that global market conditions are suspected to potentially have ripple effects on local market conditions.

Both Koekebakker et al. (2006) and Randers and Göluke (2007) explain that supply is affected by the way ships are operated; when there is excess demand, the fleet is fully employed and freight rates are high, profitability can be improved up to a certain point [19] by, for example, speeding up the fleet and utilising more of the cargo-carrying capacity. These arguments indicate that freight rates lead factors such as sailing speed and capacity utilisation, but not necessarily the other way around. However, maritime economic theory states that higher sailing speed increases productivity and, thus, supply. Eventually, more supply should in theory result in lower freight rates. This does not necessarily mean that there is a negative correlation between fleet speed and the freight rate, it might just as well be positive, because in shipping there is often a lag between decisions and market impact. Furthermore, regarding the load factor determinants, Adland et al. (2016) found a positive relationship between capacity utilisation and the freight rate, which confirmed their hypothesis that poor market conditions with low freight rates force ships to compete for lower-than-optimal stem sizes. Again, this indicates that freight rate movements lead changes in capacity utilisation, but not necessarily vice versa. From a microeconomic perspective, Olsen and da Fonseca (2017) argue that the relationship between the load factor and the freight rate is inverse, because the marginal transport cost and, thus, the freight rate decreases when the load factor increases. On the other hand, a higher average load factor for the overall tanker fleet at any point in time, could imply a lower immediate availability of supply, which in turn gives rise to a higher freight rate. Thus, there are some uncertainties regarding which way the load factor determinants impact the freight rate.

To calculate average speed for each time step using the AIS dataset, (1) is applied. For simplicity it is assumed that the ship has moved in a straight line from the previous position to the current position, although this not necessarily is the case. The AIS position is given in longitude and latitude, so basic mathematics (Pythagoras' equation) is applied to calculate the distance of the straight line (hypotenuse) between the positions. The sailing time between the positions are directly derived from the AIS data set. The unit of this factor is measured in kilometres (km) per day.

$$\text{Average speed} = \frac{\sum_{n \in N} \frac{\text{Distance sailed}}{\text{Sailing time}}}{N \text{ tankers}} \tag{1}$$

The load factor is simply calculated as the ratio of the ships current draught and the ships maximum draught [20] . The average load factor for the global and local VLCC fleet, is

[19] There is an upper boundary, because higher speed leads to higher fuel cost, which, depending on demand levels and ship specifications, eventually will cancel out the additional income due to increased utilisation.

[20] Maximum draught within the time-frame of the AIS data set (2012-2015).

Table 11. *Overview of selected freight rate determinants - input variables in the neural network model.*

Variable	Unit	Description	Expected impact
Dependent Variable:			
`Rate`	\$/tonne	Tanker spot freight rate on the BDTI TD2 route (the major tanker route between the Arabian Gulf and Singapore)	↕
AIS-derived data:			
Fleet productivity (Supply)			
`Speed_global`	km/day	Average speed of the global VLCC fleet	↕
`Speed_local`	km/day	Average speed of VLCCs in the area around and between origin (Arabian Gulf, AG) and destination (Singapore, SIN) ports	↓
`LF_global`	[−]	Average load factor (cargo-carrying capacity (dwt) utilisation) of the global VLCC fleet	↕
`LF_local`	[−]	Average load factor of the local VLCC fleet	↕
Tanker fleet activity (Supply)			
`N_tankers_op`	[Integer]	Number of (No.) VLCCs in origin port area (AG)	↓
`N_tankers_heading_op`	[Integer]	No. VLCCs heading to origin port area (AG)	↓
`N_tankers_dp`	[Integer]	No. VLCCs in destination port area (SIN)	↕
`N_tankers_heading_dp`	[Integer]	No. VLCCs heading to destination port area (SIN)	↓
`N_tankers_bp`	[Integer]	No. VLCCs in the area between ports (AG - SIN)	↕
Tonne-mile Demand			
`TMD_dp`	[tonne · km]	Aggregation of tonne-mile demand for all VLCCs heading to destination port area (SIN)	↑
`TMD_omp`	[tonne · km]	Aggregation of tonne-mile demand for all VLCCs heading to other major ports: AG, China, Japan, USA and West Africa	↕
Non-AIS-derived data:			
Supply determinants			
`Bunker_price`	\$/tonne	Bunker oil price in Singapore	↕
Demand determinants			
`Refinery_margin`	\$/bbl	Approximation of refinery profitability in Asia; the difference between petrol price in Singapore and crude oil price in Dubai; a proxy for crude oil demand in Asia	↑
Financial determinants			
`Oil_spread`	\$/bbl	The spread between Dubai crude oil front month (1-month) futures and 3-month futures	↕
`FX_USD_SR`	USD/SR	U.S. dollar-Saudi Riyal exchange rate	↕
`FX_USD_SD`	USD/SD	U.S. dollar-Singapore dollar exchange rate	↕
`BDTI`	[−]	Baltic Exchange Dirty Tanker Index (BDTI)	↑

thus calculated using (2). A weakness with this approach is that it relies heavily on the accuracy of the manual draught reports by the ships' crew. Another issue is that ships commonly carry ballast water when sailing *ballast legs*, i.e. when they transit without carrying payload [21], and the AIS data does not directly capture whether a tanker is filled with oil or ballast water. However, tankers are typically carrying less weight when sailing ballast legs, which constitutes a difference that may justify the inclusion of the load factor variables.

$$\text{Average load factor} = \frac{\sum_{n \in N} \frac{\text{Current draught}}{\text{Maximum draught}}}{N \text{ tankers}} \quad (2)$$

6.2.2

Tanker Fleet Activity (Supply)

Five different variables accounting for VLCC supply are included, in terms of the VLCCs' current position and headed destination. Our hypothesis is that a high number of VLCCs headed to or located in or around the origin port (Ras Tanura, Saudi Arabia) and destination port (Singapore), will lead to a high level of VLCC supply in these areas. We also believe that a higher number of VLCCs located between the ports can increase supply. In theory, more supply should put downwards pressure on the freight rate.

6.2.3

Tonne-mile Demand

As mentioned in Section 4.3, tanker demand is calculated on a tonne-mile or, in this case, tonne-km basis to account for the distance-effect on demand. According to Adland et al. (2017), the estimation of tonne-mile demand on a per-shipment basis is an application area where it can be beneficial to use AIS data. Thus, demand on a per-shipment basis is first calculated per time step for all VLCCs that, according to the AIS data, are heading to a specific destination. As (3) shows, tonne-mile demand is then calculated at each time step as the VLCC's load factor multiplied by the VLCC's cargo-carrying capacity (dwt) and the distance (km) from the VLCC's current position to the destination port. Both the load factor and the distance are derived from the AIS data, while the cargo-carrying capacity (dwt) is retrieved from the Sea-web online database [22]. A ship registry such as Sea-web was used here, because the AIS data does not contain ship size measured by dwt. From Sea-web a .csv file containing IMO numbers was extracted. This was obtained after specifying and matching the capacity interval for VLCCs (see Table 9) with the IMO numbers in the AIS data. Furthermore, when dealing with the formation of freight rates, the estimation of tonne-mile demand is only relevant on an aggregate level. Therefore, tonne-mile demand for all VLCCs heading to the specific destination for each time step has been aggregated.

$$\text{T.m. demand} = \sum_{n \in N} \text{load factor} \cdot \text{capacity} \cdot \text{distance} \quad (3)$$

Two variables are included when it comes to the above-mentioned tonne-mile demand: `TMD_dp` and `TMD_omp`. The former variable accounts for the tonne-mile demand, at each time step, attached to all VLCCs heading for Singapore. An increase in this variable is expected to result in a higher freight rate on the TD2 route. The latter variable accounts for the total tonne-mile demand, at each time step, linked to all VLCCs heading for five other important areas (the AG, the U.S., China, Japan, and West Africa), to account for ripple effects. The relationship between this variable and the freight rate is not obvious.

Again, a possible drawback with the applied approach is the way the load factor is calculated. As previously explained, it relies on the manual draught reports by the ships' crew, which can be subject to human error. However, these two variables are included because after removal of erroneous data from the AIS dataset, the remaining data is more available, up-to-date and, thus, potentially richer than weekly or monthly data reports from e.g. customs or shipbroking firms. Furthermore, since the load factor appears on both sides of the freight rate equation, this will affect both supply and demand. Thus, the inaccuracy issue could to some extent be balanced out.

6.3

Non-AIS-derived Data

Regarding input data not derived from the AIS data set, a total of six variables have been selected.

6.3.1

Supply Determinants

Only one supply-driven variable, not derived from AIS data, is included besides the AIS-derived supply-driven variables: `Bunker_price`. The bunker price, or fuel price, is assumed to make up most of the voyage costs (VOYEX) for the VLCC operator; VOYEX is low when the fuel price is low and vice versa. The fuel price indirectly affects supply through a shipowner's or a shipping operator's wish to maximise profit, because in shipping, profit maximisation is closely related to the optimisation of fuel consumption. For example, as discussed in Section 4.5.4, when demand exceeds supply and the tanker fleet capacity is maximised, profitability can be further improved by increasing supply through speeding up the fleet, although this leads to a higher fuel consumption. Unless demand is extremely high in this case, the fuel price would, given all else equal, determine the ceiling where it is no longer profitable to increase the speed. Thus, the fuel price definitively impacts

[21] Ships carry sea-water when sailing *ballast legs* (not sailing with cargo), to obtain adequate ship stability (mainly for safety reasons) and a sailing speed closer to the design speed (for cost-efficiency).
[22] http://maritime.ihs.com/. The dwt capacity could also have been roughly estimated using AIS data.

supply, but in which direction is not straightforward to tell: A higher fuel price makes supply costlier due to higher VOYEX, which could imply that supply decreases and, in theory, pushes the freight rate upwards. At the same time, the fuel price closely follows the fluctuations of the oil price, which is the commodity that tankers transport. Poulakidas and Joutz (2009) investigated the relationship between weekly spot tanker rates and the oil market from 1998 to 2006. They concluded that the spot tanker market is related to the crude oil prices in such a way that higher oil prices put upwards pressure on spot tanker rates. Contrarily, as mentioned in Section 4.3, Stopford (2009) comments as follows on the relationship between the oil price and tanker freight rates: An increase in the oil price tends to alter the global energy mix, which reduces demand in the tanker market because coal to a certain extent substitutes oil. A reduction in tanker demand would, in theory, put downwards pressure on the tanker freight rate. Therefore, we argue that the `Bunker_price` variable could both have a positive and negative impact on the tanker freight rate.

The active global VLCC fleet size at any point in time is also a relevant supply determinant. However, only yearly data from Thomson Reuters is available for this variable. Since the time horizon of the forecast is on an operational level and, thus, significantly shorter than one year, this variable was not included. Additionally, the relevant and active VLCC fleet size at any point in time is effectively accounted for, by using the AIS dataset to count the number of VLCCs located in or headed to the areas in, around and between the origin ports in Saudi Arabia and the destination port in Singapore.

6.3.2 Demand Determinants

Inspired by Tham (2008) and Olsen and da Fonseca (2017), an approximation of the refinery margin in West Asia (`Refinery_margin`) is included as one of the demand determinants. This variable is supposed to be a proxy for crude oil demand in Asia, or Southeast Asia, and its value is determined by taking the difference between the petrol price in Singapore and the crude oil price in Dubai. The rationale for including this variable is as follows: Refineries are the only conventional buyers of crude oil and, therefore, the volume of crude oil transported should be positively and highly correlated with refinery profitability. For example, when refinery margins are low, one would expect less crude oil- and tanker demand, which in turn would impact tanker freight rates negatively.

6.3.3 Financial Determinants

The rationale for including the (`Oil_spread`) variable is to capture potential effects from the futures curve; when e.g. the futures curve is steep upwards, traders would buy oil, sell forward or futures derivatives and use tankers to store the oil meanwhile. Consequently, one would expect tanker demand and, thus, freight rates to increase following an increase in the futures spread level. Furthermore, the BDTI is a tanker base index reflecting the price levels on all the major tanker routes in the world (see Appendix B for an overview of these routes). Hence, a positive relationship between the BDTI and the VLCC freight rate on the TD2 route from Saudi Arabia to Singapore is expected. Finally, two variables are included for the U.S. dollar-Singapore dollar and U.S. dollar-Saudi Riyal exchange rates, because shipowners have their revenues in U.S. dollars and could pay costs in local currencies, such as the Singapore dollar or the Saudi Riyal. A strengthening of the U.S. dollar relative to the Singapore dollar or the Saudi Riyal, in this case, could impact freight rates in the positive direction as it would improve shipowners' profitability. On the other hand, it could also lead to more newbuild contracting, higher supply expectations and, thus, a downwards pressure on freight rates (Ringheim and Stenslet, 2017).

6.4 AIS Data Handling

6.4.1 Collection, Preparation and Sampling of AIS Data

The AIS data used in this study is gathered from an AIS database created by Bjørnar Brende Smestad, who received raw data from the Norwegian Coastal Administration. The same database was used in the master's theses by Smestad (2015) and by Leonhardsen (2017), both former NTNU students at the Department of Marine Technology. Additionally, a list with IMO numbers of ships with certain specifications was downloaded from the Sea-web online database; ships classified as oil tankers with a cargo-carrying capacity larger than 200,000 dwt were searched for. Then all messages of type 1 and 5 from the AIS database were matched with the ships from the Sea-web list by their IMO numbers. SQLite, a relational database management system, was a vital tool here. Furthermore, this resulted in a new database containing the location, destination and draught for in total 446 VLCCs. Further, necessary calculations were done and time series for all the AIS-derived variables, described in Table 11, were created using a self-developed Python script.

AIS data is sampled at different intervals, with only a few seconds between each signal, as discussed in Chapter 3. Since only short forecasting horizons are considered, daily data is the most important. Arguably, a higher frequency of the AIS data could have been used, but that is left for another research problem. Thus, when it comes to the sampling frequency of the AIS-messages, the last received message of the day was included in our time series.

6.4.2 Uncertainty and Dealing with Missing Data

As mentioned in Section 6.1, the use of AIS data has a few drawbacks. First, the AIS satellites are not always able to cover all ocean areas simultaneously. Therefore, there are instances with gaps, where the satellites do not

receive any messages[23]. Second, in crowded areas like major ports and heavily trafficked straits or canals, such as the Strait of Malacca or the Suez Canal, the transmitted AIS messages from the many ships may interfere with each other. Interference results in more missing data, because the AIS signals do not reach the satellites. A third drawback, which was briefly mentioned in Section 6.2.1, is that the static data must be typed in manually by the ships' crew. Static data include for example the IMO number, ship type, and physical appearance like draught, destination and estimated time of arrival. Thus, the static data may be wrong, as the manual reporting can be subject to human errors.

Together, the abovementioned drawbacks give rise to many sources of uncertainty. Regarding current position of the ships, where data is missing, interpolation of the value of the previous day with the following day was done. When it comes to draught and destination, the missing data was filled using the previous day's data.

6.5

Descriptive Statistics

In this section, descriptive statistics, including correlation matrices, are presented for the AIS-derived and the non-AIS-derived input data summarised in Table 11 and described in 6.1-6.3. This data consist of time series of daily levels from the sample period 4 January 2012 to 24 December 2015 and originally contain 1450 data points. However, most of the non-AIS-derived input data are only quoted at trading days, not weekends et cetera. Consequently, there is a lot of missing values in the sample period. This is also the case with the AIS-derived data, for reasons explained in Section 6.4.2. This issue was solved by simply removing all observations on the days where one or more of the input time-series contain(s) a missing value, reducing the data sets to 913 observations.

6.5.1

AIS-derived Data

Descriptive statistics of the AIS-derived data and a correlation matrix for the levels of the AIS-derived data, are presented in Table 12 and Table 13, respectively. Additionally, time-series plots and normality plots can be found in Appendix C and D, respectively.

From Table 12 it can be observed that most of the explanatory variables derived from AIS data are quite volatile. For the higher moments, *excess kurtosis* and *skewness*, values different from zero indicate a non-normal distribution of the sample data. High *adjusted Jarque-Bera (AJB)* test results, which is observable in the second-rightmost column of Table 12, also indicate non-normality. In this case, most of the AJB values are way above critical values. Additionally, all except two p values are zero. The null hypothesis that the data is normally distributed can thus be rejected with 100% confidence, for the cases with zero p values. Moreover, a higher positive excess kurtosis implies that the distribution of a given time series has a fatter tail, which in turn indicates a higher probability of extreme events occurring more frequently. At the same time, positively skewed distributions imply more frequent small negative events and less frequent extreme positive events. Opposite is true for negatively skewed distributions, which indicate frequent small positive events and some few extreme negative events (Alexander, 2009).

From Table 13 it can be seen that all explanatory variables, derived from AIS data, are positively correlated with the freight rate. Not surprisingly, based on the discussion in Section 6.2, the local speed variable (`Speed_local`) has the strongest correlation with the freight rate among these variables. In comparison, the global speed variable (`Speed_global`) is significantly less positively correlated with the freight rate. This makes sense because `Speed_local`, as opposed to `Speed_global`, is directly linked to the trade lane between the AG and Singapore. However, the fact that these correlations are positive, instead of negative, could indicate that it is the freight rate that leads changes in sailing speed and not vice versa. When it comes to the load factor variables, `LF_global` and `LF_global`, both are close to uncorrelated with the freight rate. An explanation could be that the load factor calculations are inaccurate, perhaps because ships are sailing in ballast condition when not carrying cargo (oil), resulting in a too large load factor (when using the approach outlined in Section 6.2.1). Another reason could be that the draught reports by the ships' crew members sometimes are wrong (as discussed in Section 6.2.1 and 6.4.2).

Furthermore, the `N_tankers_heading_op` and `N_tankers_heading_dp` variables have relative strong correlations with the freight rate; and stronger than the `N_tankers_op` and `N_tankers_dp` variables. In other words, it seems that the number of ships located in the areas surrounding the ports in the AG and Singapore are less important than the number of tankers heading to these destinations. Finally, the tonne-mile demand variables are less correlated with the freight rate than expected, again based on our discussion in Section 6.2. Additionally, it is not clear why global tonne-mile demand seems to be more important (in terms of correlation numbers) than local tonne-mile demand.

6.5.2

Non-AIS-derived Data

Descriptive statistics and a correlation matrix for the data (levels) not derived from AIS, are presented in Table 14 and Table 15, respectively. Additionally, time-series plots and normality plots can be found in Appendix C and D, respectively.

From Table 14 and the time-series plots in Appendix C it can be observed that all non-AIS-derived variables are quite volatile, except the two exchange rate variables. At the same time, the U.S. dollar-Saudi Riyal exchange rate stands out when it comes to the higher moments, excess kurtosis and skewness. A higher positive excess kurtosis implies that the distribution of a given time-series has a fatter tail, which in turn indicates a higher probability of extreme events occurring more frequently. The AJB values

[23] See Smestad (2015) for further explanation on this phenomenon.

are way above critical values and the p values are all zero; the null hypothesis that the data is normally distributed can be rejected with 100% confidence.

Table 15 shows that all non-AIS-derived time-series are correlated with the VLCC freight rate to a significant extent, with correlations varying between -0.58 and 0.62. Only one of the variables, the bunker oil price in Singapore, is negatively correlated with the freight rate. In Section 6.3.1, it was argued that the `Bunker_price` variable both could have a positive and negative impact on the freight rate. However, in the sample period, a positive change in the bunker oil price seems to have a negative effect on the freight rate, or vice versa. As expected, the `BDTI` and `Refinery_margin` variables are both positively correlated with the freight rate. For the three remaining variables, the `Oil_spread` and the two currency exchange rate variables (`FX_USD_SR` and `FX_USD_SR`), we were not sure what to expect. However, all three variables are positively correlated with the freight rate.

Table 12. *Descriptive statistics of the AIS-derived data.*

	Mean	*Median*	*Std. Dev.*	*Min*	*Max*	*Excess kurtosis*	*Skewness*	*Count*	*AJB*	*AJB p value*
Rate	8,2	7,8	2,3	4,8	14,5	2,5	0,6	913	57,4	0,0
Speed_global	330,5	313,9	51,0	224,1	537,9	4,0	1,2	913	250	0,0
Speed_local	340,2	342,5	41,3	213,6	460,4	2,6	-0,1	913	6,5	0,0
LF_global	0,8	0,8	0,018	0,7	0,8	3,8	0,4	913	49,1	0,0
LF_local	0,7	0,7	0,042	0,6	0,9	3,5	-0,2	913	15,2	0,0
N_tankers_op	64,3	65,0	11,8	33,0	94,0	2,4	0,0	913	12,2	0,0
N_tankers_heading_op	21,6	22,0	10,0	2,0	47,0	2,0	0,0	913	39,4	0,0
N_tankers_dp	55,5	55,0	9,0	30,0	89,0	3,4	0,2	913	12,1	0,0
N_tankers_heading_dp	10,7	11,0	3,8	4,0	23,0	2,4	0,3	913	25,7	0,0
N_tankers_bp	128,3	128,0	12,0	96,0	164,0	2,8	0,1	913	2,1	0,3
TMD_dp	1,59E+10	1,50E+10	7,39E+09	1,74E+09	3,94E+10	2,5	0,4	913	34,1	0,0
TMD_omp	3,43E+10	3,42E+10	1,07E+10	4,34E+09	6,68E+10	2,8	0,1	913	3,3	0,2

Table 13. *Correlation matrix for the AIS-derived data.*

	Rate	Speed _global	Speed _local	LF _global	LF _local	N_t _op	N_t _h_op	N_t _dp	N_t _h_dp	N_t _h_bp	TMD_dp	TMD_omp
Rate	1,00	0,21	0,49	0,08	0,10	0,32	0,40	0,33	0,37	0,20	0,26	0,33
Speed_global	0,21	1,00	0,39	0,03	0,04	0,11	0,16	0,05	0,14	-0,03	0,11	0,17
Speed_local	0,49	0,39	1,00	-0,16	-0,09	0,39	0,56	0,15	0,43	0,03	0,35	0,48
LF_global	0,08	0,03	-0,16	1,00	0,64	-0,21	-0,31	-0,08	-0,09	0,06	-0,10	-0,23
LF_local	0,10	0,04	-0,09	0,64	1,00	-0,04	-0,21	0,02	0,01	0,17	-0,07	-0,11
N_tankers_op	0,32	0,11	0,39	-0,21	-0,04	1,00	0,71	0,49	0,60	0,43	0,51	0,54
N_tankers_heading_op	0,40	0,16	0,56	-0,31	-0,21	0,71	1,00	0,44	0,64	0,28	0,59	0,67
N_tankers_dp	0,33	0,05	0,15	-0,08	0,02	0,49	0,44	1,00	0,44	0,54	0,40	0,38
N_tankers_heading_dp	0,37	0,14	0,43	-0,09	0,01	0,60	0,64	0,44	1,00	0,31	0,83	0,58
NTBP	0,20	-0,03	0,03	0,06	0,17	0,43	0,28	0,54	0,31	1,00	0,33	0,24
TMD_dp	0,26	0,11	0,35	-0,10	-0,07	0,51	0,59	0,40	0,83	0,33	1,00	0,50
TMD_omp	0,33	0,17	0,48	-0,23	-0,11	0,54	0,67	0,38	0,58	0,24	0,50	1,00

Table 14. *Descriptive statistics of the non-AIS-derived data.*

	Mean	*Median*	*Std. Dev.*	*Min*	*Max*	*Excess kurtosis*	*Skewness*	*Count*	*AJB*	*AJB p value*
Rate	8,2	7,8	2,3	4,8	14,5	2,5	0,6	913	57,4	0
Bunker_price	540,8	609	157,3	163	761	2,4	-0,9	913	125,1	0
Refinery_margin	13,6	12,3	8,8	-11	42,3	2,9	0,4	913	20,1	0
Oil_spread	-0,3	-0,7	1,4	-3,5	3,4	2,5	0,5	913	46,9	0
FX_USD_SR	3,8	3,8	0	3,7	3,8	26,3	4	913	23554	0
FX_USD_SD	1,3	1,3	0,1	1,2	1,4	3	1,1	913	186,1	0
BDTI	740,2	711	120,4	577	1344	5,3	1,2	913	424,8	0

Table 15. *Correlation matrix for the non-AIS-derived data.*

	Rate	Bunker_price	Refinery_margin	Oil_spread	FX_USD_SR	FX_USD_SD	BDTI
Rate	1,00	-0,58	0,25	0,51	0,28	0,58	0,71
Bunker_price	-0,58	1,00	-0,46	-0,76	-0,29	-0,92	-0,34
Refinery_margin	0,25	-0,46	1,00	0,41	0,34	0,34	0,20
Oil_spread	0,51	-0,76	0,41	1,00	0,38	0,71	0,43
FX_USD_SR	0,28	-0,29	0,34	0,38	1,00	0,17	0,29
FX_USD_SD	0,58	-0,92	0,34	0,71	0,17	1,00	0,38
BDTI	0,71	-0,34	0,20	0,43	0,29	0,38	1,00

7

Machine Learning Theory

The intention of this chapter is to provide a basic understanding of machine learning, with a focus on the techniques applied in the current thesis: classification and artificial neural networks. Readers familiar with machine learning or artificial neural networks theory could skip this chapter and go straight to the methodologies outlined in Chapter 8. Others are advised to continue reading.

7.1

Machine Learning and Artificial Neural Networks

Machine learning has become a popular and powerful data-processing tool in recent years due to its wide range of usefulness: Machine learning algorithms can help companies across many industries by learning systems how to operate or predict future outcomes based on patterns in data. Mitchell et al. (1997) describes machine learning with the following definition: *"A computer program is said to learn from experience E with respect to some class of task T and performance measure P, if its performance at tasks in T as measured by P, improves with experience E"*.

Two tasks commonly solved by machine learning algorithms are *classification problems* and *regression problems*. Simply put, with classification problems a machine should separate some input into some category or classes. With regression problems, the machine should calculate a numeric value given some input. Regarding classification problems the performance can be measured in accuracy, i.e. the amount of correct classified input, if given some label or blueprint to follow. When it comes to regression problems, performance usually is measured using some residual.

Furthermore, *experience* can be thought of as the amount of data that is fed into the machine learning algorithm. For example, when algorithms are trained to recognise handwritten digits, they are typically fed some data set containing several thousand examples of handwritten digits ranging from zero to nine. The algorithms then *experience* this data and this process of learning through experience is called *training* in the domain of machine learning.

The development of *Artificial Neural Networks (ANNs)*, a type of machine learning or neural network algorithm, was originally inspired by how learning occurs in a biological system such as a human or animal brain. ANNs have become increasingly powerful, as computational capacity has dramatically improved since the 80s. ANNs, and neural networks in general, consist of *units* often called *neurons*, organised in *layers*, where the units in each layer are connected to units in the neighbouring layers. Data is received through the first layer (often placed on the left side in illustrations), which contains the *input units* and is called the *input layer*. Further, the data is sent and processed through the network's centre layers, the *hidden layers*, towards the final layer, the *output layer*. The units in the network are connected with *weights* deciding the amount of data that will be transferred from the previous unit. In a classification problem, the *output* represents the network's guess on which class the *input* (it was fed) belongs to.

7.2

Sequence Modelling and Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a type of ANN suitable for sequential-data processing such as time-series processing, because RNNs can store information; they have memory. This is possible because their units are connected in loops. As opposed to regular feed-forward networks, whereas the information flows one way through the network, the information cycles back into the network in RNNs. The basic idea of RNNs, how they can be designed to perform different tasks, and why they are favourable when doing sequence modelling, will be described in the following section.

7.2.1

Unfolding Computations

A traditional neural network carries out a set of computations, such as mapping some target y, given some input x, within some example or training set (x_i, y_i). Whereas traditional feed-forward networks cannot distinguish one example from another along some time-frame, because the sets are independent, RNNs have the ability to process information from previous test examples and use it to predict future outcomes along some time-frame. This can be mathematically described as a recursive dynamic system (see Goodfellow et al. (2016) for a more detailed description):

$$\mathbf{h}^{(t)} = f(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}, \theta) \tag{4}$$

This is called a recurrent system because the definition of the variable $\mathbf{h}$ is referred to the previous definition of itself, where $\mathbf{h}$ is called the state of the system. RNNs are built on these systems. Thus, (4) shows how RNNs not only depend on the input at time t, but also on the previous state at time $t-1$.

Figure 9 depicts a computational graph of a RNN that produces some output o_t, given some input x_t at time t. The illustration shows how the units at each time-step not only get input from the current x values, but also from the previous states. Mathematically this is described by Equation (4), which is applied *t - 1* times. Moreover, Figure 9 illustrates how one can represent the cyclical behaviour of recurrent systems by unfolding them. The unfolded network can be thought of as many copies of traditional feed-forward networks, but with inputs also from the past, which connects the chain. Thus, each state contains information from the whole past. This ability

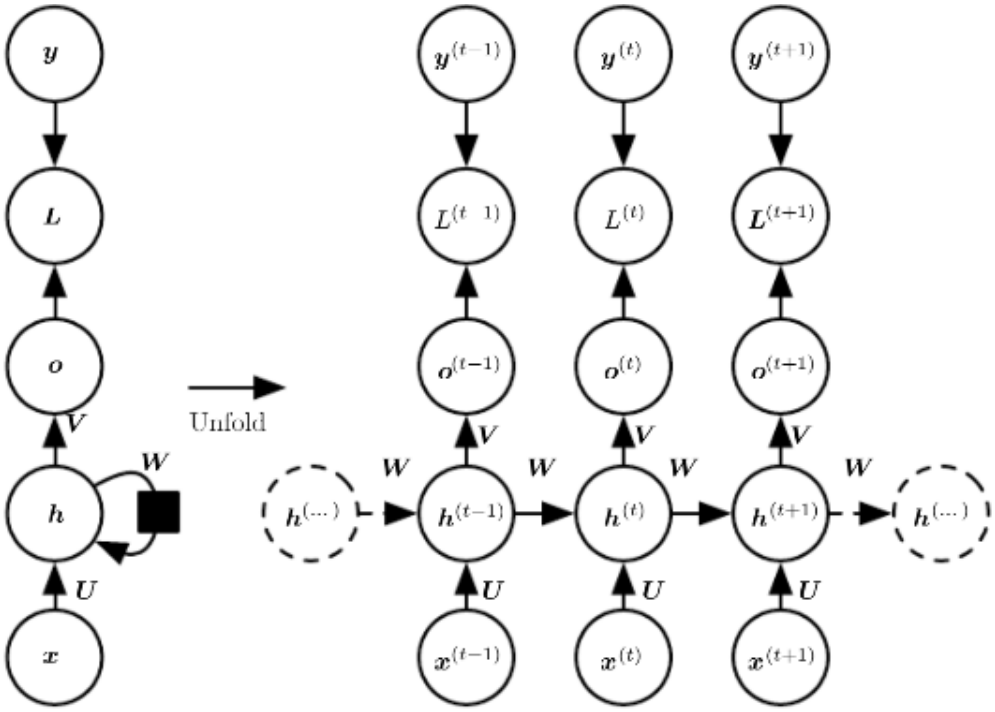


Fig. 9. Goodfellow et al. (2016) *Unfolding computational graph illustrating the cyclical behaviour of recurrent neural networks (RNNs). The output (o^t) is mapped by the state- (**W**), input- (**U**) and output (**V**) connections (weights), the input ($\boldsymbol{x}^t$) and the previous state ($\boldsymbol{h}^{t-1}$). The state output (**o**) is compared to the target (**y**). The performance of the network is measured by the loss (**L**).*

make RNNs suitable for processing sequence data like for instance financial time-series.

7.2.2

Training Neural Networks

Each neural network unit is connected with an *activation function* $\boldsymbol{a}$. The input to this function is a *weight* $\boldsymbol{w}$ multiplied by the input $\boldsymbol{x}$ it receives from its connected unit, added to a *bias number* $\boldsymbol{b}$, and σ is a transformation, usually nonlinear such as the sigmoid or tanh function(see Chapter 8 for definitions).

$$\mathbf{a} = \sigma(\mathbf{xw} + \mathbf{b}) \tag{5}$$

The weights make each neuron unique and they are tuned so the network can output as accurate results as possible. The weights are updated through a process called *training*. Initially, the weights are random. This makes the network's accuracy low, as a result of wrongly produced output. In the training phase, the correct outcome is known. It is important not to *overfit* the model on the training data, or else the network will not generalise well on data sets other than the training set. The training process is often to minimise an optimisation function, often called a *loss* or *cost function*. The loss function tells the network how well it is performing with the output it produces, given the input x and the weights w. The minimisation is commonly performed by the *back propagation algorithm*: When the data flows from the input units, through the hidden layers and finally to the output layers (*forward propagation*), one compares the output of the network with the correct output, using the loss function. Thereafter, information is sent back through the network, which allows the network to update the weights. This results in more correct output the next time the input is forward propagated, and so on. This cycle of sending information back and forth internally in the network is the core idea of the back propagation algorithm.

When it comes to RNN and LSTM neural networks (which will be described in the forthcoming sub-section), a modified version of the back propagation algorithm must be used: the *back propagation algorithm trough time*, which is thoroughly explained by Goodfellow et al. (2016).

Furthermore, a typical way to optimise the loss function is to apply *gradient descent*, a method that finds a way to identify the minimum of a function: Given an arbitrary loss function $L(x)$, the derivative dL/dx can be used to minimise the function by telling how x can be changed to minimise L. By moving in small steps in the negative direction of the sign of the gradient or derivative $L(x - \lambda \cdot sign(dL/dx))$, the minimum will be located. Here, the step size is called the *learning rate*. The lower the learning rate, the slower the network trains. Correspondingly, the higher the learning rate, the faster the network trains. However, at this end of the scale, the network will be at risk of not learning properly. Thus, there is a trade-off, as the following example shows:

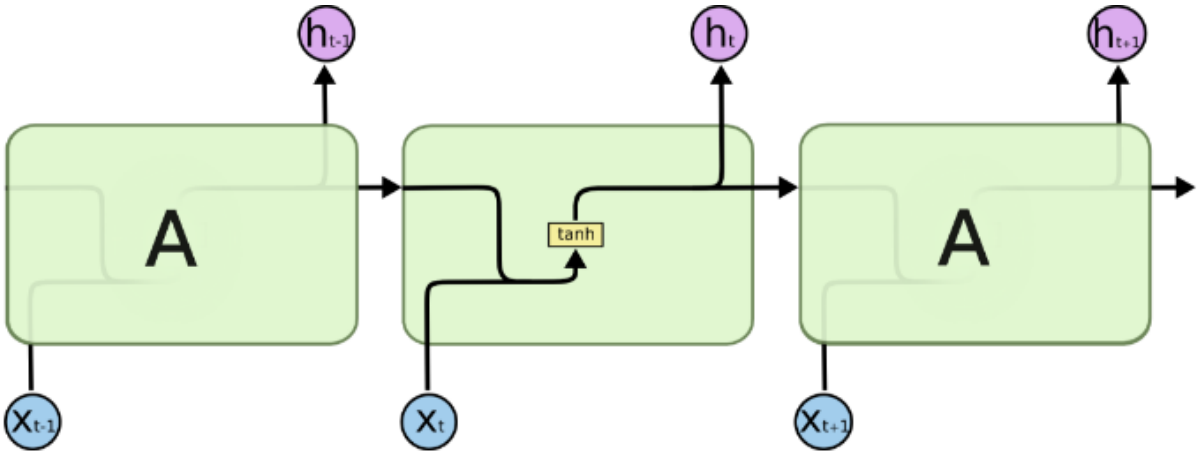
Lets denote the learning rate λ, the loss function x^2 and the derivative $2x$. Additionally, say $x = 0.03$ and $\lambda = 0.001$. Then the loss is

$$L = 0.03^2 = 9 \cdot 10^{-4}. \tag{6}$$

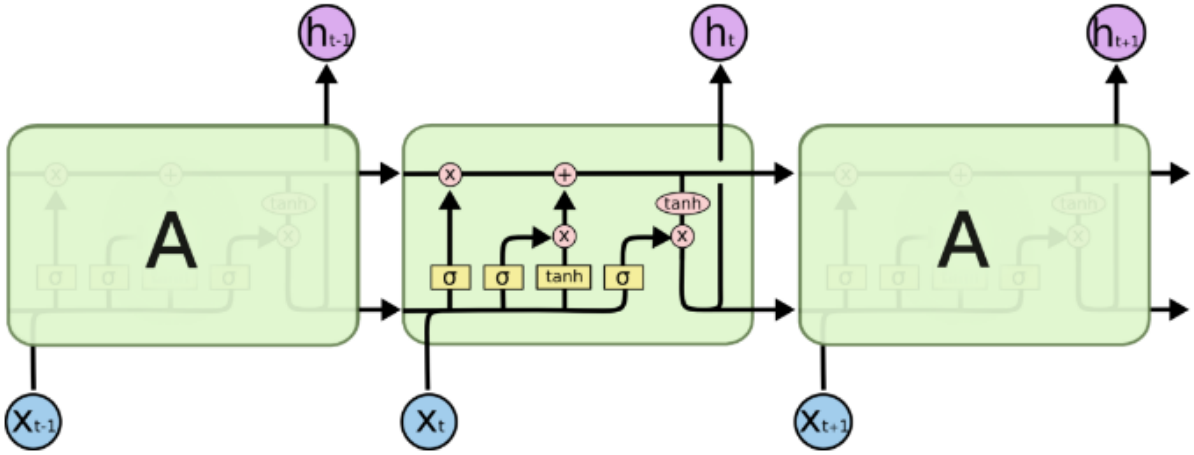
By applying gradient descent, the new loss after taking one step is

$$L_{new} = (0.03 - 0.001 \cdot sign(2 \cdot 0.03))^2 = 8.41 \cdot 10^{-4}. \tag{7}$$

The new loss is smaller than before. Moreover, using a higher learning rate, say 0.1, the answer would have been $4.9 \cdot 10^{-3}$, which illustrates the importance of applying a correct learning rate for the gradient descent optimiser. The gradient descent method is the standard (loss function minimisation) procedure to make the neural network train and tune its weights. However, other more sophisticated methods exist, which may be faster at learning, such as the *Adam optimiser* discussed in Chapter 8 (Kingma and Ba, 2014).



(a) The repeating module in a standard RNN contains a single layer.



(b) The repeating module in an LSTM contains four interacting layers.

Fig. 10. Olah (2015). Yellow rectangular operators represent layers containing activation functions (sigmoid or tanh functions). Pink circular operators represent pointwise operations, like vector addition (+ sign) or multiplication(× sign). Black lines denote content flow, separating lines denote copying content, and merging lines mean concentrating content.

7.2.3 Long Short-Term Memory

During network training, the error gradient tunes the weights in the right direction and with the right magnitude. When the gradient is updated in training, sometimes it can vanish or explode. This is a result of the recurrence relation working similarly to the power calculation, where weights are multiplied together for each time step (see Chapter 10 in Goodfellow et al. (2016) for further explanation on this topic). The *vanishing gradient problem* makes it difficult for RNNs to learn tasks where long-term memory is required (Bengio et al., 1993, 1994). To solve this problem, the *long short-term memory (LSTM) model* was developed (Hochreiter and Schmidhuber, 1997). This type of RNN allows the gradient to flow for a long duration. Moreover, the similarities between a standard RNN cell and LSTM cell is displayed in Figure 10.

Furthermore, whereas the regular RNNs only have one neuron (unit) with a corresponding activation function, LSTM cells (the green boxes in Figure 10b) have four activation functions, also called the *gates* of the cells. In addition to the input unit (the yellow *unit* second from the right in Figure 10b), there are three other gates, called the *input gate* (second from the left), *forget gate* (leftmost) and the *output gate* (rightmost). The gated units of the LSTM cells have *sigmoid* activation functions, while the input unit usually has the *tanh* activation function.

The most important component of an LSTM cell is the *state unit*, sometimes referred to as the *cell state*, or *self loop unit*. The cell state allows the LSTM cell to store information from the past states, with only minor interactions. In Figure 10 this is illustrated as the horizontal top line going straight through the cell to the next cells. The forget gate decides what information is being kept and what is disregarded by the non-linear activation expression in (8). Similar to a regular RNN unit, the LSTM forget gate is fed the input data x_t and the previous state data $h_{(t-1)}$.

$$f_t = \sigma(b^f + U^f x_t + W^f h_{(t-1)}) \tag{8}$$

Here, f_t is the output data of the forget gate at time t, b^f is a bias at the forget gate, U^f is the weights at the forget gate corresponding to the input data x_t at time t, and W^f is the weights at the forget gate corresponding to the previous state data at time $t-1$. The input unit is activated with a *sigmoid* activation function (9), while the input gate is usually activated with a *tanh* function (10).

$$g_t = \sigma(b^i + U^i x_t + W^i h_{(t-1)}) \tag{9}$$

$$i_t = tanh(b^g + U^g x_t + W^g h_{(t-1)}) \tag{10}$$

In the above expressions, g_t and i_t are the output data of the input gate and the input unit at time t, b^g and b^i are the biases, U^g and U^i are the weights corresponding to the input data x_t at time t, and W^g and W^i are the weights corresponding to the previous state data at time $t-1$. The input gate decides if the new input should be allowed to enter the cell. At last, the output gate is similar to the other gated units, as expressed by (11).

$$O_t = \sigma(b^o + U^o x_t + W^o h_{(t-1)}) \tag{11}$$

Here, the internal cell state (self loop state) is updated using (12).

$$s_t = f_t \cdot s_{(t-1)} + g_t \cdot i_t \tag{12}$$

The output of the entire cell, the hidden state, is calculated using (13).

$$h_t = tanh(s_t) \cdot O_t \tag{13}$$

The LSTM cell avoids the exploding and vanishing gradient problem because there is no activation function in the recurrent aspect of the cell, because the self loop is linear. Additionally, the self loop allows the LSTM cell to remember information in the long-term and it automatically decides what information to forget and what information to let into or out of the cell. Finally, there are also other variants of the LSTM model, in addition to the version explained in this section, but they will not be elaborated further as they are not relevant for this thesis. Good reading material on LSTM is provided by, for example, Olah (2015); Goodfellow et al. (2016).

8

Methodology

This chapter outlines the methodologies used to predict or classify freight rate movements on the TD2 tanker route between the Arabian Gulf and Singapore. Specifically, an attempt will be made to classify whether the freight rate goes up or down with certain amounts or within specified intervals (*classes*). Numerous experiments will be performed, investigating several sets of explanatory variables (as mentioned in Section 6.1), different forecasting time-horizons and various degrees of model complexity. In other words, a three-dimensional solution space is examined, as illustrated in Figure 11. Furthermore, an 80/20 ratio is applied when dividing the data into a training set and a test set; the less recent 80% of the data constitutes the in-sample period for the training set, while the most recent 20% of the data is earmarked for validation and makes up the out-of-sample period for the test set (see Chapter 6 for more information about the data).

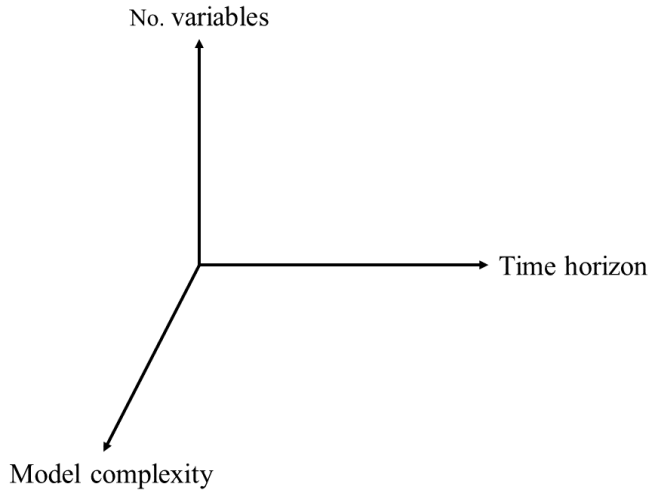


Fig. 11. *A three-dimensional solution space.*

8.1

Benchmark Model: Multivariate Linear Regression

According to Domingos (2012) it could pay to try the simplest classification techniques first. Therefore, the results from a classic and relatively simple method, multivariate linear regression, will be used as a baseline for the LSTM NN model results.

As (14) reveals, the method is named multivariate linear regression, because the general linear statistical model includes more than one explanatory variable Alexander (2009). As mentioned, this model will be run on two sets of explanatory variables, one with six variables (not including AIS-derived data) and the other with 17 variables (all variables, including AIS-derived data).

$$Y_t = \alpha + \beta_1 X_{t1} + ... + \beta_k X_{tk} + \epsilon_t,\ \epsilon_t \sim i.i.d.(0, \sigma^2) \quad (14)$$

for $t = 1, ..., T$. There are T equations in the k unknown parameters ($\beta_1, ..., \beta_k$), one equation for each data vector. Thus, the k index will either be equal to 6 or 17, depending on the set of variables we fit the model to. The error, ϵ_t, is assumed to be independent and identically distributed [24]. Furthermore, model fitting or model estimation involves solving these equations using a selected method, such as ordinary least squares (OLS). Formulas for OLS estimators are presented in, for example, Alexander (2009). The fitted value of Y_t can be expressed as

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta}_1 X_{t1} + \hat{\beta}_2 X_{t1} + ... + \hat{\beta}_k X_{tk}, \quad (15)$$

for $t = 1, ..., T$. The difference between the actual and predicted value of Y_t is the residual e_t. Thus, (15) can be re-expressed as

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta}_1 X_{t1} + \hat{\beta}_2 X_{t1} + ... + \hat{\beta}_k X_{tk} + e_t, \quad (16)$$

for $t = 1, ..., T$. When it comes to the input data, for which descriptive statistics were presented and discussed in Section 6.5, one important modification is done both in the training set and in the test set: all data vectors, including the freight rate time-series, are normalised using (17).

$$\text{Normalised value} = \frac{\text{value} - \text{mean value}_{training\ set}}{\sigma_{training\ set}} \quad (17)$$

The data vectors of the training set and the test set are both normalised using the mean and standard deviation of the training set's data vectors because, when running the model on the test set, the future values are "unknown". This will be further explained in the forthcoming section, as the same normalisation procedure is used for the LSTM NN model. Furthermore, since predictions are performed at different time horizons, time lags equal to the different time horizons between the dependent variable (the freight rate, Y_t) and the explanatory variables are used both in the training set and the test set. For example, when predicting one-day ahead and 20-days ahead changes, (15) transforms to (18) and (19), respectively:

$$\hat{Y}_{t+1} = \hat{\alpha} + \hat{\beta}_1 X_{t1} + \hat{\beta}_2 X_{t1} + ... + \hat{\beta}_k X_{tk}, \quad (18)$$

$$\hat{Y}_{t+20} = \hat{\alpha} + \hat{\beta}_1 X_{t1} + \hat{\beta}_2 X_{t1} + ... + \hat{\beta}_k X_{tk}, \quad (19)$$

[24] The independence assumption implies that there is no autocorrelation, and the identical distribution implies that homoscedasticity is apparent. For a proper explanation of these statistical properties, see Alexander (2009).

for t = 1, ..., *T-1* and t = 1, ..., *T*-20. When the values of the test set are predicted, using the multivariate linear regression model fitted to the training set's data, the rate of return [25] of the predicted freight rate is calculated for each time step by multiplying by the standard deviation and adding back the mean of the freight rate's rates of return in the training set.

Furthermore, to calculate the prediction or classification accuracy, the abovementioned rates of return must be compared with the actual rates of return from the test set for each time step. The classification accuracy will depend on the forecasting time-horizon and on the number of classes used. In this study, different time horizons (in the order of days or weeks) and two sets of classes are investigated.

Say R denotes the rate of return, then the applied sets of classes are as follows:

- 3 classes: (R = 0, R > 0, or R < 0)
- 5 classes: (R = 0, R = [0, 1σ], R = [0, -1σ], R = [1σ, ∞], or R = [-1σ, -∞])

Above, σ is the standard deviation of the freight rate returns in the training set. The reason for using this standard deviation as interval limits of the classes, instead of a constant value measured in percentage (e.g. 5%, 10% and 20% change in return), will be explained in the forthcoming section.

Finally, the classification accuracy is calculated for each combination of time horizon and class set (e.g. 3 x 2 = 6 combinations in total if three different time horizons are used). This is done by dividing the number of correct classifications (identified by comparing the prediction results with the values of the test set for each time step) with the total number of predictions (which is equal to the number of days in the test set).

8.2

LSTM Neural Network Model

8.2.1

Model Overview

As discussed in Chapter 7, LSTM neural networks are suitable for processing sequential data. Therefore, an LSTM NN model will be developed in this thesis, to read sequential data in the form of time-series data of the explanatory variables summarised in Table 11. The goal of the LSTM NN model, presented in this section, is to predict whether the rate of return of the freight rate F at time t, R_F^t, will be positive or negative at some future time $t + \tau$. In addition, the magnitude of the freight rate change will be predicted. Furthermore, using the definition of a learning algorithm provided by Mitchell et al. (1997), the model will "*[...] learn from experience E with respect to some class of task T and performance measure P, if its performance at tasks in T as measured by P, improves with experience E*". Moreover, a single output, an integer, will be produced by the model. Such an output is commonly referred to as a *label* in the domain of machine learning. In this study, different outcomes (freight rate movements, $R_F^{t_f}$) will be assigned different labels. Thus, these labels can for example represent a positive or negative freight rate movement within a pre-defined interval of magnitude. This interval is referred to as a *class*. The sets of classes used in the rest of this study were defined in Section 8.1.

Specifically, the *task* of the model is to learn the label $y_{t_0}^{t^f}|x_{t_p,t_0}$, given some input vector x_{t_p,t_0}. The label $y_{t_0}^{t^f}$ refers to which class the rate of return, $R_F^{t_f}$, belongs to at a future time t^f. Furthermore, x_{t_p,t_0} is a vector containing all the explanatory variables (freight rate determinants) in Table 11, within the time interval from a previous time t_p up to and including the present day t_0. The class s, an interval of possible rate of return values of the freight rate, can be any one within a set S of classes: $S \in \{s_1, s_2, ...s_n\}$. The arithmetic rate of return R, at some future day t_f, is defined by (20).

$$R(t_f) = \frac{R(t_f) - R(t_0)}{R(t_0)} \tag{20}$$

The model's *performance* is measured by *classification accuracy*. This is just the number of correct guesses divided by the total number of guesses. For example, if 180 guesses were correct in a test sample of 200 data points, the classification accuracy would be 90%. When it comes to model *experience*, this is simply the training set (80% of the original data set) with the given labels.

When looking at different time horizons to train over, the abovementioned labels may change with the horizon, as exemplified in Table 16. Dependent on the horizon, label "1" represents a positive change, label "2" represents a negative change and "0" represents no change.

Table 16. *An example showing how labels change with the time horizon of the forecast.*

Date	*Freight rate*	*1 day ahead*	*2 days ahead*	*3 days ahead*
04/1/12	6.46	1	1	1
05/1/12	6.47	1	1	1
06/1/12	7.11	1	1	2
09/1/12	7.36	0	2	0
10/1/12	7.36	2	0	0
11/1/12	6.80	1	1	1
12/1/12	7.36	0	1	1
13/1/12	7.36	1	1	-
16/1/12	7.50	1	-	-
17/1/12	7.91	-	-	-

[25] *Rate of return* is in this study defined as the current price, subtracted the original price, divided by the original price (this is the arithmetic rate of return as defined by (20). Multiplying this by 100 yields the rate of return expressed in percentage, which may be referred to as *the percentage return*.

8.2.2

Network Architecture

The architecture of the LSTM NN model, depicted in Figure 12, is inspired by the works of Chevalier (2017). The network has one input layer, two hidden LSTM layers and one output layer. The layers are also referred to as cells. Furthermore, at each time step the input layer of the network receives the time-series data of the explanatory variables (which were summarised in Table 11 in Chapter 6). The input layer is activated by the non-linear sigmoid activation function, expressed by (21), with values ranging between zero and one.

$$\sigma = \frac{1}{1 + e^{-z}} \tag{21}$$

Above, $z = wx + b$ is the input at a given node, multiplied by the weight w and added a bias b. From the input layer, the data is passed through the hidden LSTM layers, before the network produces the output: a guess on which label and, thus, class the input belongs to. The number of input *nodes* (units) in the input layer is the same as the number of explanatory variables investigated. Thus, the number of input nodes depends on which data set is applied. The number of output nodes will depend on how many classes that are used. The more classes, the higher model "complexity". For example, if three classes are used, the model can classify whether the price moves up, down or remains unchanged. To also predict how much the price moves up or down, more classes must be added. Moreover, the two hidden layers contain 15 hidden units each. The number of hidden units in the these layers determine the capacity of the network, i.e. how many previous states is used to calculate the output. In other words, the short-term memory in the LSTM is set to 15 days. Additionally, all layers are fully connected.

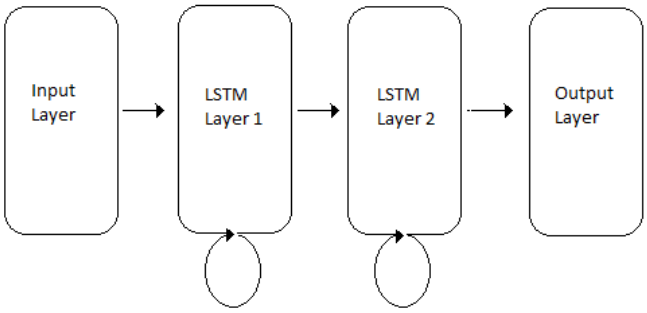


Fig. 12. *The architecture of the LSTM NN model.*

The LSTM layers are built using TensorFlow basic LSTM cells (for documentation, see TensorFlow (2015)). Moreover, the LSTM cells use the sigmoid activation function at the gates and the tanh activation function expressed by (22), ranging between -1 and 1, at the inner states.

$$tanh = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \tag{22}$$

Again, $z = wx + b$ takes in the input x at a given node in the LSTM layer, multiplied by the weight w and added a bias b. The final layer, the output layer, is a *softmax* layer, which calculates the probability of how much the freight rate will change, given some forecast horizon. The softmax function, expressed by (23), represents the probability of predicting one specific class in a set of different classes.

$$softmax(z)_i = \frac{exp(z_i)}{\Sigma_i^n exp(z_i)} \tag{23}$$

Again, $z = wx + b$.

8.3

Data Processing

Before the data is fed into the network, it must be handeled in certain ways. *Firstly*, the whole data set has to be divided into a training set and a test set. As mentioned in the introduction to Chapter 8, an 80/20 ratio is applied, which gives 731 training samples and 183 testing samples. *Secondly*, the data sets must be normalised, using (24), which is the same procedure as described in Section 8.1. Here, x_i is the value of the variable x at instance i, while μ_x and σ_x represent the variables' mean and standard deviation, respectively.

$$z_i = \frac{x_i - \mu_x}{\sigma_x} \tag{24}$$

As mentioned in Section 8.1, the test set is standardised using the training samples' mean and standard deviation, because if new sample points were to be tested on the model, one would not possess a "test set"; one would only know the data points up until today, future test points are unknown. This normalisation procedure is commonly applied on data fed into ANNs, because if not applied, some data might have a much larger scale than other data, which will negatively affect the tuning of the weights and give poor results.

Thirdly, a rolling-window technique is applied when turning the normalised data set into input data for the network. The input data consists of *data points*, where each data point is a *window* (time series) of past data up until the present day. This window, T_{window}, of data corresponds to a label, which denotes the known movement of the freight rate. For example, one can choose to make the network train on a data window of 30 days prior to today. Furthermore, when evaluating the model on the test set, the corresponding unknown labels are to be predicted.

Lastly, during training, the input data, often referred to as the input *examples*, are fed into the network in so-called batches. The reason being, to train on all the input

examples simultaneously, each training cycle would be very memory-costly and require a lot of computational power, especially if the data set is large. One *batch* contains a set of data points. As mentioned above, one data point contains past data of the explanatory variables within the chosen time window (T_{window}), in addition to the corresponding label. Moreover, the number of data points in the batch is usually referred to as the batch size. A typical number chosen is 64 data points per batch.

8.4

Network Training and Optimisation

During training, the weights in the LSTM network is tuned using the *Adaptive Moment Estimation (Adam) optimiser*, introduced by Kingma and Ba (2014). This optimiser is regarded as being more advanced, but robust, and it usually trains neural networks faster than the gradient descent optimiser (which was described in Chapter 7). Whereas the gradient descent optimiser only has one learning rate, the Adam optimiser has one learning rate for each weight in the network. Moreover, it is called adaptive, because the learning rates of all model parameters are adapted separately as the network-learning progresses. The method uses estimates of the first and second moment of the gradients of the loss function to compute the adaptive individual-learning-rates in an iterative process. Furthermore, the *update-equations* of the Adam optimiser can be written as

$$g_t = \nabla_\theta f_t(\theta_{t-1}) \tag{25}$$

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{26}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \tag{27}$$

$$\hat{m_t} = m_t / (1 - \beta_1^t) \tag{28}$$

$$\hat{v_t} = v_t / (1 - \beta_2^t) \tag{29}$$

$$\theta = \theta_{t-1} - \alpha \cdot \hat{m_t} / (\sqrt{\hat{v_t}} + \epsilon), \tag{30}$$

where g is the gradient at iteration time step t of the objective function f_θ with parameters θ. Moreover, m_t and v_t are the first and second raw moment biased estimations, where β_1 and β_2 are decay rates for the estimates and g_t^2 indicates the element-wise square of g_t. Further, $\hat{m_t}$ and $\hat{v_t}$ are bias-corrected estimates of the first and second moment of the gradients of the objective function. Finally, the parameters are updated using (30). α is the step size used in the iterations and ϵ is a small constant for numerical stability. Kingma and Ba (2014) has shown that the Adam optimiser is a robust stochastic optimisation algorithm, which is why it is applied in this study.

The cross-entropy loss function in (31) is used to calculate the prediction or classification error made by the network. The output of this function is used to adjust the weights of the network during the training phase. The cross-entropy function for classification problems with i (and more than two) classes is defined as

$$L = -\sum_{i=1}^{C} p_i ln(q_i), \tag{31}$$

where L is the cross-entropy loss function, i is the ith class in class set C, p_i is the probability that outcome i occurs, made by the neural network, and q_i is the actual probability that outcome i occurs in the data. For example, if the model predicts whether the freight rate moves up (label 1), down (label 2) or remains unchanged (label 0), there are thus three labels: $y \in \{1, 2, 0\}$. Furthermore, lets say that some input x_t is fed into the model and that the actual outcome for this input is the positive movement of the freight rate, so the true label is "1". Using the popular machine learning term and approach "one hot encoding", the actual probability distribution can be described as follows:

$$P(y = 1) = 1,\ P(y = 2) = 0,\ P(y = 0) = 0 \tag{32}$$

The network's job is to produce the probability distribution of the labels. If the results are

$$P(1) = 0.63,\ P(2) = 0.12,\ P(0) = 0.25, \tag{33}$$

it means the model guesses that the correct label is "1", but only with about 63% confidence. Further, the cross-entropy loss for this outcome is

$$L = -(1 \cdot ln(0.63) + 0 \cdot ln(0.12) + 0 \cdot ln(.25)) = 0.46, \tag{34}$$

which is what the network will try to minimise during training. The weights are then adjusted by sending the information by the back-propagation algorithm through time. The training occurs in cycles, where data is sent trough the network in batches. Each batch is sent trough the network and then the value of the loss function is propagated backwards in order to adjust the weights. When all the batches are sent trough the network, one training *epoch* is finished. One can choose how many times to repeat this in order to tune the network, referred to as the number of training epochs.

8.5

Regularisation

Overfitting occurs when a function or model has an overly good fit with one (training) data set, and therefore is not able to learn generally on other (test) data sets. This is a problem that may occur when using neural networks to perform some given task. Overfitting can be detected by plotting the model's accuracy and loss on both the training and test sets. Simply put, if the accuracy and loss are much better on the training set compared to the test set, the model is said to be overfitted on the training set. To avoid this, certain regularisation techniques can be applied. Goodfellow et al. (2016) lists several useful regularisation techniques (in Chapter 7), such as *L2 regularisation* and *dropout regularisation*.

L2 regularisation, also known as *Tikhonov regularisation*, is a method that adds a regularisation term to the loss function, L. The new loss function, L', is expressed by (35). Here, a penalty is given, which is proportional to the sum of the square of the weights in the network.

$$L' = L + 0.5 \cdot \lambda \cdot ||w||^2 \tag{35}$$

Here, λ is a coefficient that determines how much the weights should penalise the regularised loss function L'. If λ is zero, there is no weight regularisation. Moreover, if λ has a value, this regularisation technique prevents the network from tuning the weights during training too finely on the training set, due to the penalty on the weights.

Dropout is a regularisation technique that, simply put, randomly drops units in the network during training. This means that at random occurrences, information-flow from a unit is multiplied by zero. This will prevent the network from overfitting, because the units will be less prone to a phenomenon called *co-adapting* (for more details, see Srivastava et al. (2014)). In this thesis, dropout regularisation is the selected technique that will be used on the LSTM layers to prevent overfitting. In accordance with the paper by Baldi and Sadowski (2013), a 50% probability of dropping units is applied.

9

Results and Discussion

This chapter presents the research findings of the current thesis. Several experiments are carried out to answer the research problems presented in Section 1.2. The first section of this chapter describes the experimental setups. The second section follows up with a presentation and a brief discussion of the experimental results. Finally, the last section summarises and analyses the aforementioned results.

As thoroughly described in Chapter 8, the model's task is to predict the interval (class) in which the future rate of return of the freight rate will belong to, over a specified forecasting time-horizon. This rate of return, denoted R_f, over some time span $t = f - i$, is defined as

$$R_f = \frac{F_f - F_i}{F_i}, \tag{36}$$

where F_f is the final value of the freight rate over the time span and F_i is the initial value of the freight rate. The model's task can thus be defined as follows: to find the probability distribution of the discrete outcomes y_f^o, in outcome $O \in [o_1, o_2, ..., o_n]$ of R_f at time i, given some historical input data, x. As mentioned in Chapter 8, the model should for example be able to predict whether the future rate of return of the freight rate is positive, negative or zero, when using three classes.

9.1

Experimental Setups

Figure 7 in Chapter 8 displays the possible solution space, with model complexity constituting one of the three dimensions. Model complexity is simply the number of classes used in the experiments. Furthermore, the experimental setup is defined by the three axes in Figure 7: the forecast time-horizon, number of variables used, and the model complexity as defined above.

Forecasting Time-horizon The LSTM NN model will predict future freight rate movements (rate of return) over different time horizons, within the following discrete space:

$$T \in [1\,day,\, 5\,days\,,\, 10\,days]. \tag{37}$$

Model Complexity As mentioned, the model complexity will vary. This means that the number of labels the model must predict will vary correspondingly. Moreover, the model will produce a discrete probability distribution for n outcomes and the label with the highest probability attached, will be the model's prediction. Given the distribution of the rate of return of the freight rate, intervals (classes) that the model must predict the label to be within are defined. These intervals or classes are measured in standard deviations of the rates of return of the freight rate in the training set (as defined in Section 8.1), e.g. between zero and one standard deviation, between zero and minus one standard deviation, and so on. Furthermore, the two different difficulty levels or model complexities used in this study are listed below.

- ***Easy*** *model complexity*: The model will classify the rate of return of the freight rate (R_f^i) at the future day f, predicted at day i, thus over the time span $t = f - i$, as one of three possible outcomes (up, down or unchanged); mathematically described as the model's guess on whether $R_f^i > 0$, $R_f^i < 0$ or $R_f^i = 0$.
- ***Hard*** *model complexity*: The model will classify the predicted rate of return of the freight rate (R_f^i) as one of five possible outcomes, over the time span t as defined above; mathematically described as the model's guess on whether $R_f^i \in [0, \sigma_R]$, $R_f^i \in [0, -\sigma_R]$, $R_f^i \in [\sigma_R, \infty]$, $R_f^i \in [-\sigma_R, -\infty]$ or $R_f^i = 0$.

The standard deviation of the rate of return of the freight rate (σ_R), is calculated on the freight rate in the training set, using a standard method (for explanation and method, see Chapter 8).

Input Variables Finally, the number and type of input variables will be varied using different subsets of the selected explanatory variables (for an overview of the selected freight rate variables, or determinants, see Table 11). Thus, the performance of the network using the complete data set, will be compared to its performance using various subsets of the selected explanatory variables.

Experiments Table 17 displays all the performed experiments, and describes their setups. All experiments are carried out using the LSTM NN model and the benchmark model (multivariate linear regression), both presented in Chapter 8. The results obtained using the LSTM NN model are presented in the forthcoming section, while a summary of these results and the results obtained using the benchmark model are gathered and discussed in Section 9.3.

Table 17. *Experimental setups.*

Experiment	**Forecast horizon**	**Model complexity**	**Input variables**
			Complete data set:
A1	1 day	Easy	All 17 variables in Table 11
A2	5 days	Easy	All 17 variables in Table 11
A3	10 days	Easy	All 17 variables in Table 11
B1	1 day	Hard	All 17 variables in Table 11
B2	5 days	Hard	All 17 variables in Table 11
B3	10 days	Hard	All 17 variables in Table 11
			Subsets:
C1	1 day	Easy	All six non-AIS variables in Table 11
C2	5 days	Easy	All six non-AIS variables in Table 11
C3	10 days	Easy	All six non-AIS variables in Table 11
D1	1 day	Easy	`Speed_local`, `BDTI`
D2	5 days	Easy	`Speed_local`, `BDTI`
D3	10 days	Easy	`Speed_local`, `BDTI`
E1	1 day	Easy	`Speed_local`, `N_tankers_heading_op`, `N_tankers_heading_dp`
E2	5 days	Easy	`Speed_local`, `N_tankers_heading_op`, `N_tankers_heading_dp`
E3	10 days	Easy	`Speed_local`, `N_tankers_heading_op`, `N_tankers_heading_dp`
F1	1 day	Easy	`FX_USD_SD`, `Bunker_price`, `BDTI`
F2	5 days	Easy	`FX_USD_SD`, `Bunker_price`, `BDTI`
F3	10 days	Easy	`FX_USD_SD`, `Bunker_price`, `BDTI`

9.1.1

LSTM Model Configuration and -Parameters

The network parameters, also called *hyperparameters*, and other important features of the LSTM NN model are gathered in Table 18. The parameters are tuned by trial and error to obtain results that do not fluctuate too much between each run [26] and the Adam optimiser parameters are chosen as the default values recommended by Kingma and Ba (2014). The selected number of training epochs, 500, makes sure the model trains long enough to give stable predictions on the test set. When it comes to the capacity of the model, it increases with the number of hidden units. At the same time, the run time shortens, but the memory cost increases. The number of hidden units is therefore chosen to somewhat balance the trade-off between high memory cost and short run time. Moreover, the learning rate is not listed in Table 18, because it is tuned automatically for all parameters by the Adam optimiser. Finally, the time it takes to train the network is about two minutes per run (for each experiment). There are some variances in the results: the model provides slightly different results for each run, due to the random initialisation of the weights. Therefore, to get reliable results, each experiment is run ten times and these ten results are then averaged.

[26] A grid search could have been performed to optimise the parameters (or hyperparameters) of the LSTM NN model. This will be further discussed in Chapter 10.

Table 18. *Definition and description of LSTM NN model parameters.*

Parameter	**Description**	**Value**	**Role and importance**
α	Step size	0.01	The step size in the Adam optimiser.
β_1	1st moment decay rate	0.9	Controls the exponential decay rate of the exponential moving-average of the gradient in the Adam optimiser. The moving averages are estimates of the 1st moment (mean) of the gradient (Kingma and Ba, 2014).
β_2	2nd moment decay rate	0.999	Controls the exponential decay rate of the exponential moving-average of the gradient in the Adam optimiser. The moving averages are estimates of the 2nd moment (variance) of the gradient (Kingma and Ba, 2014).
σ	Nonlinear activation function	Sigmoid in LSTM layers, sigmoid in input layer, softmax in output layer.	Determines how the units in each layer is activated.
N_{hidden}	Hidden units	10	Number of sequential LSTM units in each LSTM layer.
$N_{Training}$	Training epochs	500	Number of times the entire input data-set is sent trough the network to adjust the weights.
b	Mini-batch size	64	The number of data points (training examples) sent trough the network at each training iteration. The network does not receive the entire data set each time, only subsets, called batches.
τ_{window}	Rolling-window size	21 days	The size of each data point or training example. 21 days means that each training example looks at 21 days of past data as input, for all explanatory variables in the input data set.
$N_{outcome}$	Outcome size	$N_{outcome} \in [3, 5]$	Number of outcomes (or classes) the model must predict: 3, using easy model complexity, and 5, using hard model complexity.
T	Forecasting time-horizon	$T \in [1\,day, 5\,days, 10\,days]$	One of the dimensions investigated in the 3-dimensional solution space (see Figure 11).
λ	L2 regularisation parameter	0.001	The penalty coefficient used in weight decay. 0 gives no regularisation, which is unfavourable, because it increases the chance of overfitting. At the same time, a large regularisation value may lead to underfitting, which also is undesirable. Thus, there is a trade-off.

9.2

Experimental Results

This section provides the results of the performed experiments, which were described in Table 17. As mentioned in Section 9.1, each experiment is performed ten times and the average is presented. Additionally, Figure 13 displays the freight rate plotted over the entire sample period, which includes both the training set and test set periods.

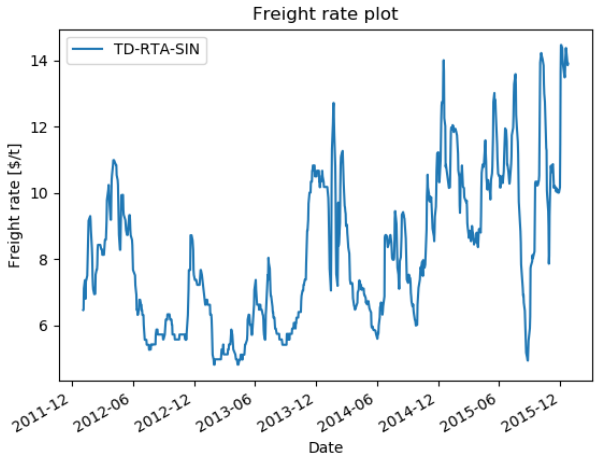


Fig. 13. *Time-series plot of the freight rate in the sample period.*

9.2.1

A Experiments

In the A experiments (A1, A2 and A3), the easy model complexity with three classes is used. For each experiment, the model thus predicts, for each time step (each day), whether the freight rate movement is positive, negative or zero, from the day of prediction to the end of the time horizon specified in Table 17. Additionally, as listed in Table 17, all explanatory variables in Table 11 are included in these experiments.

For illustration purposes, Figure 14 displays the freight rate in the top window, the forecasts made by the LSTM NN model in the middle window, and the test set's actual outcomes in the bottom window, over the test set period. The values on the vertical axis of the two latter plots represent the labels of the forecasts and actual outcomes, respectively. "2" means that the freight rate moves down the following day, "1" means it moves up and "0" means it remains unchanged. The plots indicate that the forecasts are not particularly good.

Moreover, Table 19 presents the LSTM NN model's ability to forecast over different time horizons, given the experimental setup defined in Table 17. Here, the performance is measured by the accuracy and loss on the training set and test set, i.e. by how many correct guesses the model makes and by the cross-entropy loss, respectively.

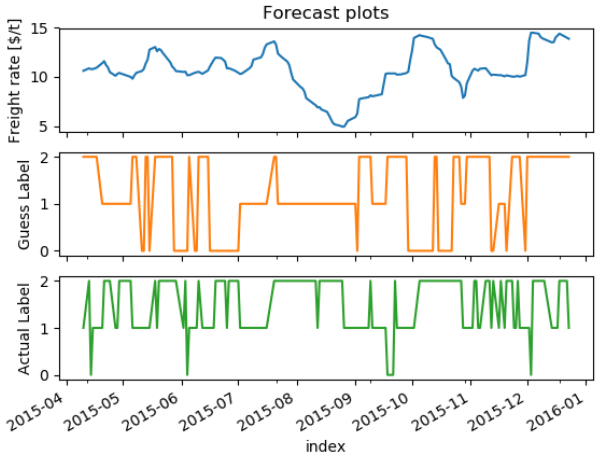


Fig. 14. *Time series of the freight rate in the test set (blue line), along with a comparison of the model's predicted outcome (orange line) and the actual outcome in the test set (green line) over the entire test set period.*

Again, for illustration purposes, the model's accuracy on the training set (blue line), versus the model's accuracy on the test set (green line), is presented in Figure 15. As explained in Section 8.2, accuracy is measured as the total number of correct predictions divided by the total number of predictions. At the end of each training epoch, the model makes one prediction for each day in the training and testing sample. As Figure 15 shows, the training accuracy is better than the test accuracy, which means that the model is able to tune its weights to give good results in-sample, but not as well out-of-sample. Furthermore, Figure 16 shows that the model is not able to bring the test loss down along with the training loss.

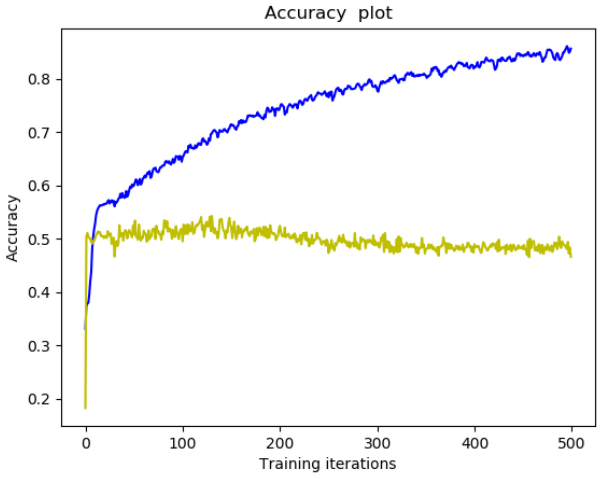


Fig. 15. *Training accuracy (blue line) versus test accuracy (green line).*

As revealed by Figure 14 and Table 19, the model does not produce good forecasts using this experimental setup. Table 19 shows that the model is not able to produce forecasts with accuracy better than 50% for one-day-ahead and five-days-ahead forecasts, and only slightly better

Fig. 16. *Training loss (red line) versus test loss (green line).*

than 50% for ten-days-ahead forecasts. The model thus produces slightly better results at the forecasting horizon of ten days, compared to horizons of one day and five days. This is a trend of the easy model complexity, as later results will show.

In conclusion, when using the complete data set (including all variables in Table 11), it seems difficult for the network to find a pattern.

9.2.2

B Experiments

In the B experiments (B1, B2 and B3), the difficulty level of the forecasting is raised. Now, the hard model complexity (five classes) is used, on the same input data applied in the A experiments. In other words, the model does not only predict whether the freight rate movement is positive, negative or zero, but also by how much it moves up or down, as explained in Section 9.1.

As Table 20 reveals, the forecasts are once again not very accurate. However, it was expected that the classification accuracy would drop when raising the difficulty level of the forecast (using more classes). Furthermore, Table 20 shows that the best forecast, using this experimental setup, is done on the one-day-ahead horizon. Compared to the accuracy of pure guessing, which would be 20% as there are five possible outcomes (after a large number of guesses), the results of the B experiments are not too bad.

Similar to the A experiments, the model is able to tune the weights to the training set, but not the test set. This can be seen in Appendix F, Figure F.2, by comparing the test accuracy and test loss with the training accuracy and training loss, respectively. One can see that the test loss goes upwards when the training loss goes downwards. This means that the model is able to learn a pattern in the training data, but not in the test data. One could thus argue that the model is overfitted on the training set.

Furthermore, in the A and B experiments, the complete data set (all 17 explanatory variables listed in Table 11, Chapter 6) was used. By reducing the number of input variables, the model may find it easier to discover a pattern in the data. This issue will be addressed in the next experiments, where the model will be fed subsets containing only a few of the explanatory variables listed in Table 11.

9.2.3

C Experiments

In the C experiments (C1, C2 and C3), all AIS variables are removed from the model's input data, but the six non-AIS variables listed in Table 11 remains. The intention of these experiments is twofold: *Firstly*, to investigate the effect of using (or not using) AIS data as input, with respect to test accuracy. *Secondly*, to see if the model better can find a pattern in the input data when reducing the number of variables.

Table 21 displays the results of the C experiments. Compared to the A experiments, one can see that the test accuracy of the C experiments has improved slightly for the two shortest forecasting horizons (one day and five days ahead), but unchanged for ten days ahead. Surprisingly, it appears that removing the AIS variables as input to the model slightly improves the forecasts. However, it still seems that the model struggles to find a pattern among this many variables (six variables), based on the plots in Appendix F, Figure F.3. Regarding the second intention of the C experiments, the reduction from 17 to six input variables did not improve the model's forecasting performance in terms of test accuracy.

9.2.4

D Experiments

Only two variables and the easy model complexity are used in the D experiments (D1, D2 and D3): `Speed_local`, the AIS variable with the highest correlation to the freight rate, and `BDTI`, the non-AIS variable with the highest correlation to the freight rate (for correlation matrices, see Table 13 and 15 in Chapter 6).

Table 22 displays the results of the D experiments, which reveal that the model performs best when forecasting ten days ahead, with about 63% test accuracy. Moreover, this result is the best one across all experiments (A - F). Further, Figure F.4 in Appendix F indicates that this result is stable after about 100 training epochs, because the test accuracy does not tend to move up or down with higher numbers of training epochs. However, the loss is increasing, which is an indication of model overfitting on the training data.

By comparing the results in Table 19 with the results in Table 21 and Table 22, it seems that the model performs better when using fewer explanatory variables. Especially the five-days-ahead and ten-days-ahead forecasts achieve higher test accuracy with fewer input variables. In addition, the loss results are better on all forecasting horizons.

9.2.5

E Experiments

Similar to the C experiments, the E experiments (E1, E2 and E3) try to answer the second research problem: whether the use of AIS-derived input data provides additional value in terms of classification test accuracy when forecasting freight rate movements. In these experiments, as opposed to the C experiments, solely AIS-derived input data is used. Based on the results from the D experiments, the amount of input variables should be limited to only a few. Thus, only three AIS variables, the ones with the strongest correlation to the freight rate, are selected: `Speed_local`, `N_tankers_heading_op` and `N_tankers_heading_dp`.

Table 23 shows that the test accuracy is about 50% for all forecasting horizons. Compared to the C experiments, the model performs marginally worse in terms of test accuracy on five-days-ahead and ten-days-ahead forecasts, and exactly the same when forecasting ten days ahead. With respect to test loss, the E experiments perform worse on all horizons. Moreover, in line with previous experiments, revealed by Figure F.5 in Appendix F, the loss is increasing with the number of training epochs. As mentioned, this could be a sign of overfitting. Regarding the second research problem, these results may indicate that AIS-derived input data, applied in this thesis, does not add extra value when forecasting freight rates in the tanker market. This hypothesis will be investigated further in the next and final experiments.

9.2.6

F Experiments

In the F experiments (F1, F2 and F3), the opposite approach to the approach used in the E experiments is attempted to try to answer the second research problem. Here, none AIS-variables are included, only financial variables; the ones with the highest correlation to the freight rate are selected: `FX_USD_SD`, `Bunker_price` and `BDTI`.

Table 24 shows that the test accuracy in the F experiments is about 50% for the one-day-ahead and five-days-ahead forecasts, while the ten-days-ahead forecasts achieve about 60% accuracy. Compared to the E experiments, the model performs better on test loss for all forecasting horizons and better on test accuracy when forecasting ten days ahead. Compared to the D experiments, the results are about the same. Thus, the results are slightly better in the F experiments, which indicates that the hypothesis presented in sub-section 9.2.5 may be correct. However, there are numerous potential sources of error, particularly related to the AIS data used in this thesis. This will be further discussed in Section 9.3.

Table 19. *Results of the A experiments: A1 (1 day), A2 (5 days), A3 (10 days).*

Forecasting horizon	Test accuracy	Test loss	Train accuracy	Train loss
1 day	0.47	1.39	0.83	0.42
5 days	0.46	3.23	0.96	0.22
10 days	0.54	2.47	0.97	0.18

Table 20. *Results of the B experiments: B1 (1 day), B2 (5 days), B3 (10 days).*

Forecasting horizon	Test accuracy	Test loss	Train accuracy	Train loss
1 day	0.34	2.27	0.80	0.60
5 days	0.29	5.30	0.94	0.26
10 days	0.26	4.71	0.96	0.20

Table 21. *Results of the C experiments: C1 (1 day), C2 (5 days), C3 (10 days).*

Forecasting horizon	Test accuracy	Test loss	Train accuracy	Train loss
1 day	0.52	1.58	0.80	0.48
5 days	0.52	3.31	0.94	0.16
10 days	0.54	3.39	0.97	0.10

Table 22. *Results of the D experiments: D1 (1 day), D2 (5 days), D3 (10 days).*

Forecasting horizon	Test accuracy	Test loss	Train accuracy	Train loss
1 day	0.46	1.41	0.76	0.57
5 days	0.52	2.35	0.92	0.21
10 days	0.63	1.77	0.93	0.18

Table 23. *Results of the E experiments: E1 (1 day), E2 (5 days), E3 (10 days).*

Forecasting horizon	Test accuracy	Test loss	Train accuracy	Train loss
1 day	0.51	1.46	0.78	0.55
5 days	0.50	3.12	0.93	0.21
10 day	0.54	2.34	0.95	0.16

Table 24. *Results of the F experiments: F1 (1 day), F2 (5 days), F3 (10 days).*

Forecasting horizon	Test accuracy	Test loss	Train accuracy	Train loss
1 day	0.51	1.24	0.78	0.53
5 days	0.51	2.89	0.91	0.26
10 days	0.61	1.99	0.96	0.14

9.3

Benchmarking and Discussion

9.3.1

Benchmarking

Table 25 presents a comparison between the results obtained using the LSTM NN model and the benchmark model (the multivariate linear regression model described in Section 8.1), in terms of classification test accuracy (for the definition of classification accuracy, see Section 8.2).

For the A experiments, the results reveal that the benchmark model outperforms the LSTM NN model on one-day-ahead and five-days-ahead forecasts. The opposite occurs when forecasting ten days ahead. Additionally, neither of the two models achieves better than 60% accuracy on any forecasting horizon in the A experiments. Regarding the B experiments, where the hard model complexity is applied, the LSTM NN model outperforms the benchmark model on one-day-ahead and ten-days-ahead forecasts. When forecasting five days ahead, both models perform at about 30% accuracy. This latter result is not necessarily poor, considering that five classes are used. Further, in the C experiments, the benchmark model again outperforms the LSTM NN model on the shortest horizon, but opposite is true for five days ahead and ten days ahead. When it comes to the D experiments, the results are similar to the C experiments, except for the magnitude of the results on the longest forecasting horizon; the LSTM NN model achieves accuracy above 60%. The benchmark model gets below 60% accuracy on all horizons. Further, both models perform quite similarly in the E experiments: around 50% accuracy on all horizons. Lastly, the LSTM model outperforms the benchmark model on all horizons in the F experiments, particularly on the ten-days-ahead forecasts. Here, again, the model achieves higher than 60% accuracy on the longest horizon, ten days ahead, which is promising.

In fact, the LSTM NN model outperforms the benchmark model on the longest forecasting horizon, ten days ahead, in all experiments. On the two shorter horizons, the benchmark model generally competes well with the LSTM NN model, and actually outperforms the LSTM NN model on all one-day-ahead forecasts except for the B and F experiments. With respect to the first research problem presented in Section 1.2, the ten-day-ahead forecasting results are promising. In further work, it would thus be interesting to investigate longer forecasting horizons.

9.3.2

Findings in the Three-dimensional Solution Space

As described in the beginning of Chapter 8 and illustrated in Figure 11, a three-dimensional solution space has been investigated in this thesis. With respect to the time-horizon dimension, the discussion above reveals that the LSTM NN model achieves the best results on the longest forecasting horizon investigated (ten days ahead), while opposite is true for the benchmark model.

Model complexity constitutes another dimension. The experimental results presented in Section 9.2 show that the LSTM NN model generally produces forecasts with better but not significantly better test accuracy than random guessing [27], neither with easy nor hard model complexity. Exceptions are the D and F experiments when forecasting ten days ahead, where the test accuracy for both D and F is above 60%. Furthermore, the test accuracy generally dropped, as expected, when moving from easy model complexity in the A experiments to hard model complexity in the B experiments, i.e. when using five instead of three classes. However, the test results are still better than the guessing approach would provide.

Further, the input variables applied in the experiments constitute the third and last dimension in the investigated solution space. The results show that the use of only AIS input variables produces forecasts with test accuracy comparable to what the guessing approach would provide. Moreover, not using AIS input variables actually gives better results when forecasting ten days ahead, using the LSTM NN model. For the benchmark model it is the other way around. Furthermore, reducing the number of input variables seems to improve the forecasting results, as long as the applied input variables are somewhat correlated with the freight rate. A hypothesis is that too many variables makes it difficult for the LSTM NN model to learn a general pattern for both the training and test data.

9.3.3

Limiting Factors

There are many factors that could have led the LSTM NN model to not produce forecasts with better accuracy than about 60% in this thesis. Firstly, neither the quantity nor the quality of the data applied in this study is optimal. Regarding data quantity, only four years of data have been used, resulting in a total of 913 data points after cleaning the data. Neural networks are known to perform well on large data samples. For example, the MNIST data sample popularly used for handwriting recognition has 60,000 images for training and 10,000 images for testing. Furthermore, another issue is the quality of the AIS data applied in this thesis, which contains several gaps where data is missing. As mentioned in Section 2.3 and 6.4, these gaps may be a result of interference of the AIS messages, or simply a consequence of the AIS satellites not being able to cover all ocean areas simultaneously in the sample period. In this study, this missing data problem has been handled by removal of some data and interpolation, an assumption which is questionable, because it is not certain that ships have moved in straight lines or with constant speed. An additional issue concerning the quality of the AIS data is the reported destination for each ship, in the

[27] With three classes, the probability of randomly guessing the correct outcome is 1/3 (three possible outcomes). Likewise, with five classes, this probability is 1/5 (five possible outcomes). However, it could be argued that, in reality, shipping domain experts would guess future freight rate movements to be positive or negative more often than unchanged (due to the high volatility in shipping markets). Thus, closer to 1/2 and 1/4 guess accuracy with three and five classes, respectively, are more realistic.

AIS messages. Such static data is reported by the ships' crew manually, thus it could be subject to human error such as wrong-typing or the crew forgetting to report the information. Therefore, the reliability of this data is also questionable. A final uncertainty related to AIS data in this study is the AIS-derived input variables that depend on the load factor. Tankers commonly carry ballast water when not carrying crude oil or oil products (when sailing ballast legs). The reported draughts used to calculate the load factor in this study could thus falsely imply too high dwt capacity utilisation for the VLCCs, if the VLCCs sail ballast legs, and thus overestimate tonne-mile demand. In conclusion, not regarding the latter issue, the LSTM NN model could perhaps obtain better results after sampling more recent AIS data with higher quality.

Furthermore, the LSTM NN model results could perhaps be improved by adjusting the model configuration and optimise the tuning of the hyperparameters through a grid search (see Goodfellow et al. (2016), page 420). It is possible that some of the parameter values make the LSTM NN network unable to learn from the input data. Further, the applied regularisation technique may not be sufficient. Other techniques that might be applied are the early stopping technique and the L1 regularisation, explained by Goodfellow et al. (2016). Adding other regularisation techniques could reduce the overfitting problem, which in turn could result in better forecasting accuracy.

Moreover, maybe other input variables should have been included in the study. Some of the variables listed in Table 11 in Chapter 6 do not correlate strongly with the freight rate. To investigate this potential issue, experiments with only a few input variables, the ones with the highest correlation to the freight rate, were performed. These experiments achieved in general slightly better results. Finally, it is not certain that ship traffic data is suitable for the prediction of future freight rate movements. It could for instance be the other way around, that freight rate movements are better suited to predict future ship movements and behaviour of shipping operators.

Table 25. *Comparison of classification accuracy in the test set between the LSTM NN model and the benchmark model (a multivariate linear regression model).*

Comparison of test set classification-accuracy between the LSTM NN model and the benchmark model:				
Forecasting time-horizon:	**1**: 1 day ahead	**2**: 5 days ahead	**3**: 10 days ahead	*Model complexity:*
Benchmark model:				
Experiment **A**	0.58	0.54	0.46	Easy
Experiment **B**	0.29	0.30	0.21	Hard
Experiment **C**	0.57	0.51	0.45	Easy
Experiment **D**	0.57	0.51	0.53	Easy
Experiment **E**	0.53	0.51	0.54	Easy
Experiment **F**	0.49	0.48	0.39	Easy
LSTM NN model:				
Experiment **A**	0.47	0.46	0.54	Easy
Experiment **B**	0.34	0.29	0.26	Hard
Experiment **C**	0.52	0.52	0.53	Easy
Experiment **D**	0.46	0.52	0.63	Easy
Experiment **E**	0.51	0.50	0.54	Easy
Experiment **F**	0.51	0.51	0.61	Easy

10

Concluding Remarks and Further Work

The final chapter of this thesis concludes our results and findings, as well as outlines a plan for further work, which encourages more research on the topics discussed.

10.1

Conclusions

The aim of this thesis was to examine the ability of machine learning algorithms, specifically neural networks, to predict future shipping freight rates. Additionally, whether the inclusion of AIS-derived explanatory variables enhances the forecasting performance of the applied network (the LSTM NN model), has been investigated. The use of the tanker market has been motivated by the fact that it is highly liquid, and therefore provides sufficient amounts of data. Furthermore, using a recurrent neural network with long short-term memory (LSTM) cells, and various subsets of explanatory variables, an attempt has been made to predict the tanker freight rate on the major tanker route between Ras Tanura (Saudi Arabia) and Singapore, over the following short-term horizons: one day ahead, five days ahead and ten days ahead. The performance of the LSTM NN model has been measured based on whether the model was able to classify (predict) the correct movement of the freight rate over the specified forecasting horizon.

In this thesis, the LSTM NN model provides promising results on the forecasting horizon of ten days ahead; the LSTM NN model outperforms the benchmark model on all experiments carried out on this forecasting horizon. However, on shorter horizons (one day ahead and five days ahead), the benchmark model competes well with the LSTM NN model. Further, it is expected that the performance of the LSTM NN model can be improved, which is discussed in Section 10.2. Thus, regarding the first research problem listed in Section 1.2, whether neural networks are suited to forecast freight rate movements in shipping markets, the findings of this thesis show that LSTM neural networks have potential.

When it comes to the second research problem listed in Section 1.2, the results presented and discussed in Chapter 9 indicate that the applied AIS-derived input data did not provide a significant additional value when forecasting freight rates in the tanker market. For reasons discussed in Section 9.3 it is expected that the LSTM NN model can obtain better results if new and more recent AIS data, with higher quality, is sampled and fed the model.

10.2

Further Work

Based on the discussions in Section 9.3 and the conclusions in Section 10.1, further studies should focus on the topics listed below.

- *AIS data quality:* New, more recent AIS data should be sampled. AIS data from 2012 to 2015 has been used as LSTM NN model input. This data contains gaps with missing data. As mentioned in Section 2.3, more satellites being launched will lead to better quality of AIS data. Therefore, by using new AIS data, the number of these gaps could be reduced, as the number of satellites receiving AIS messages has increased since 2012. With higher quality AIS data, it is expected that the LSTM NN model could perform better, using the same input variables; the LSTM NN model may find it easier to discover a general pattern for the training set and the test set, and avoid or reduce the overfitting issue discussed in Section 9.2.

- *AIS data quantity*: Sampling of more AIS data is an option. In this thesis, four years of sample data was used (from the beginning of 2012 to the end of 2015). This data amount is perhaps not sufficient for the model to learn general patterns. As mentioned in Section 9.3, it is not rare that neural networks have to train on tens of thousands of data points to obtain high accuracy on test sets. Thus, if the data is available and its quality is good enough, the sample data window should be increased.

- *Hyperparameter optimisation:* As discussed in subsection 9.3.3, optimisation of the model's hyperparameters, e.g. through a grid search, is a possibility. Tuning the hyperparameters optimally can give better forecasting accuracy. For example, as explained in Chapter 7, using the wrong learning rate can make training much slower. The same logic applies to all the hyperparameters of the model.

- *Apply other regularisation techniques:* A possible solution to the overfitting problem [28], discussed in Section 9.2, is to apply additional regularisation techniques, or change the one already used. For example, the λ_2 parameter in the L2 regularisation could be increased. Other regularisation techniques that could be attempted is dropout regularisation, as described in Chapter 8, or early stopping, described in Goodfellow et al. (2016).

- *Use a different neural network:* The LSTM NN model is based on the most basic form of LSTM neural networks. There are other methods that could have been attempted as well, such as bidirectional networks or a combination of recurrent and convolutional networks, which could give better results.

[28] All the experiments showed that the test loss increased with increasing training iterations, which is a sign of overfitting.

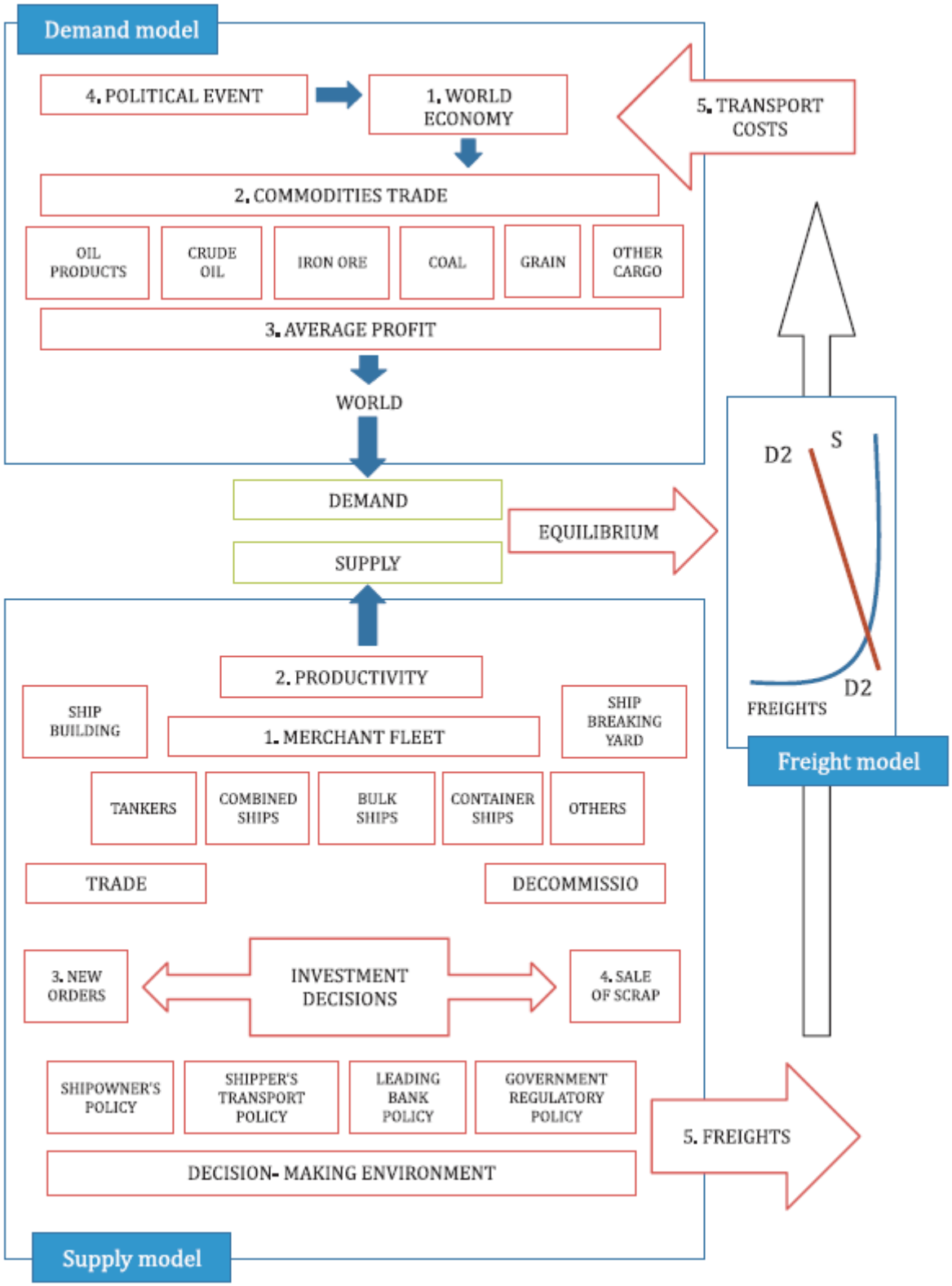
- *Different input variables/subsets of input variables:* The input variables (explanatory variables) and the subset selections used in this thesis are perhaps not the optimal ones with respect to the current research problems. When it comes to the selection of explanatory variables, perhaps further studies should focus on finding variables that are more strongly correlated with the freight rate in question. Further, regarding subset selections, mathematical and statistical methods could be applied to identify the best combinations of variables. Additionally, as mentioned in Section 2.1, Alizadeh and Talley (2011) criticised in their paper the utilisation of aggregate and macroeconomic data in maritime forecasting. They further suggested that shipowners and charterers need micro-forecasts, e.g. of freight rates for specific routes, for making operational decisions, cash flow analyses and budgeting. Therefore, Alizadeh and Talley (2011) investigated the use of micro economic determinants such as vessel and contract specific factors, and e.g. found that the duration of the laycan period is an important freight rate determinant and vice versa. Thus, further studies that perform maritime forecasts on very short horizons, such as the ones attempted in this thesis, should consider to include micro economic determinants, as the ones mentioned above, if such data is accessible.

- *Extend the forecasting horizon:* The LSTM NN model has shown potential when forecasting ten days ahead, i.e. on the longest forecasting horizon attempted in this study. As mentioned in Section 9.2, it would be interesting to see if the LSTM NN model can perform well on even longer forecasting horizons, e.g. several weeks or months ahead.

- *Compare the results with other benchmark models:* In this thesis, the LSTM NN model's performance is only compared with the results of one benchmark model, a multivariate linear regression model. In further studies, other standard forecasting methods should be attempted for bench marking purposes, such as an Auto-regressive Integrated Moving Average (ARIMA) model or a Vector Autoregressive (VaR) model.

Appendix A The Shipping Market Model

The below figure is adapted from Jugović et al. (2015), and is a more detailed version of the classic maritime supply-demand model, or shipping market model, presented in Chapter 4.



Appendix B Overview of Tanker Shipping Routes

The tables below are adapted from Γολας (2012), but the Baltic Exchange is the original source of information. The tables present an overview of the Baltic Exchange tanker route definitions for the Baltic Clean Tanker Index (BCTI) and the Baltic Dirty Tanker Index (BDTI), respectively:

Table B.1. *Baltic Clean Tanker Index (BCTI) - Route Definitions.*

Routes	**Vessel size (k.dwt)**	**Route description**	**Vessel type**
TC1	75	Arabian (Middle East) Gulf to Japan: Ras Tanura to Yokohama	Aframax
TC2	37	Continent to US coast: Rotterdam to New York	Handysize
TC3	38	Carribean to US coast: Aruba to New York	Handysize
TC5	55	Arabian Gulf to Japan: Ras Tanura to Yokohama	Panamax
TC6	30	Algerie to Euromed: Skikda to Lavera	Handysize
TC8	65	Arabian Gulf to UK-Cont.: Jubail to Rotterdam	Panamax
TC9	22	Baltic to UK-Cont.: Ventspils to Le Havre	Handysize

Table B.2. *Baltic Dirty Tanker Index (BDTI) - Route Definitions.*

Routes	**Vessel size (k.dwt)**	**Route description**	**Vessel type**
TD1	280	Arabian (Middle East) Gulf to US Gulf: Ras Tanura to Loop	VLCC
TD2	260	Arabian Gulf to Singapore: Ras Tanura to Singapore	VLCC
TD3	260	Arabian Gulf to Japan: Ras Tanura to Chiba	VLCC
TD4	260	West Africa to US Gulf: Port of Bonny Offshore Terminal to Loop	VLCC
TD5	130	West Africa to US coast: Port of Bonny Offshore Terminal to Philadelphia	Suezmax
TD6	135	Black Sea to Mediterranean: Novorossiyk to Augusta	Suezmax
TD7	80	North Sea to Continent: Sullom Voe to Wilhelmshaven	Aframax
TD8	80	Kuwait to Singapore: Mena al Ahmadi to Singapore	Aframax
TD9	70	Carribean to US Gulf: Puerta La Cruz to Corpus Christi	Panamax
TD10	50	Carribean to US coast: Aruba to New York	Panamax
TD11	80	Cross Mediterranean: Banias to Lavera	Aframax
TD12	55	Amsterdam-Rotterdam-Antwerp range to US Gulf	Panamax
TD14	80	SE Asia to EC Australia	Aframax
TD15	260	West Africa to China	VLCC
TD16	30	Black Sea to Mediterranean: Odessa to Augusta	Handysize
TD17	100	Baltic to UK-Cont.: Primorsk to Wilhelmshaven	Aframax
TD18	30	Baltic to UK-Cont.: Tallinn to Rotterdam	Handysize

Appendix C Time-series Plots

Fig. C.1 and C.2 show time-series plots of the AIS-derived data and Fig. C.3 presents time-series plots of the non-AIS-derived data:

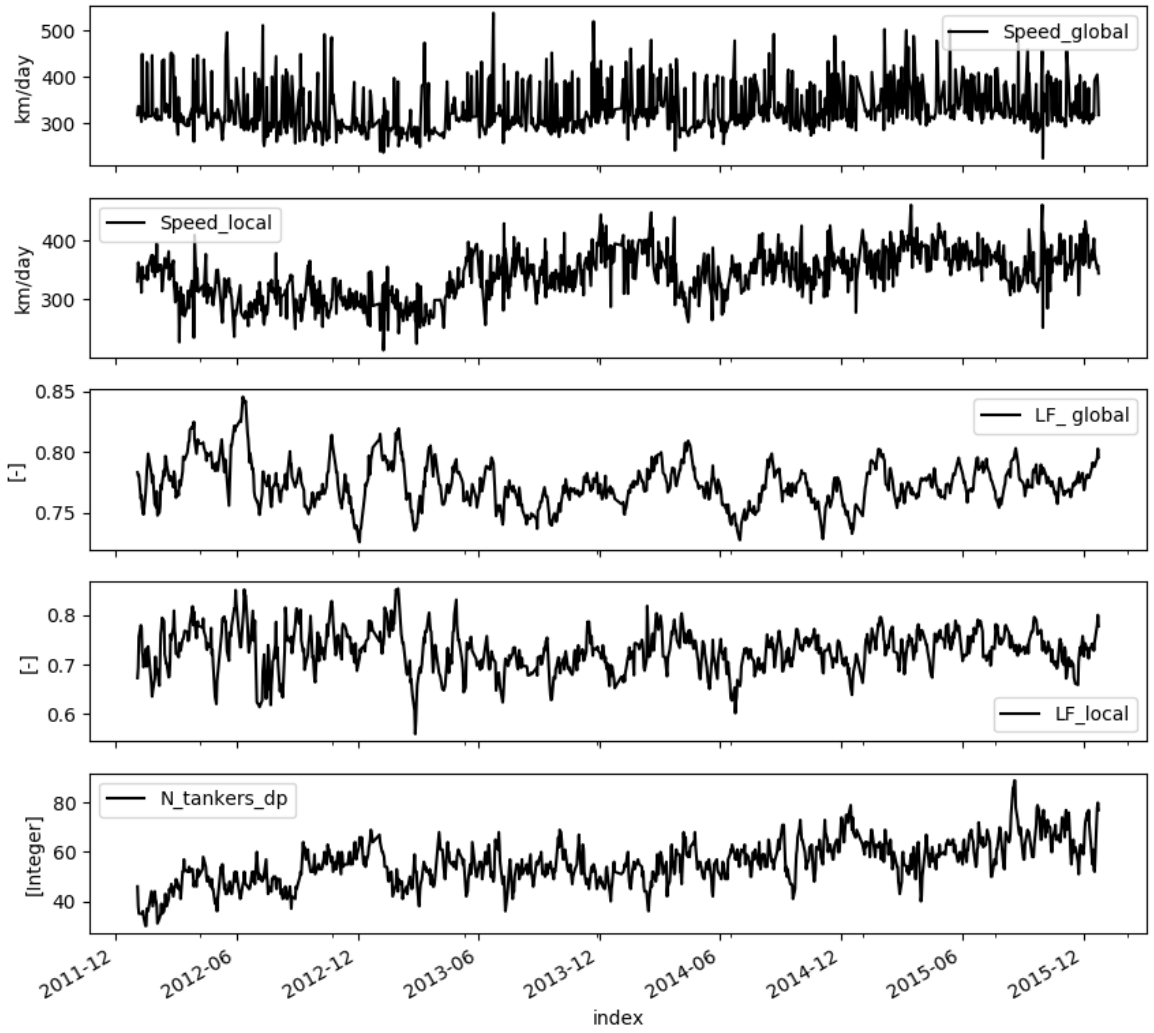


Fig. C.1. *Time-series plots of the AIS-derived data.*

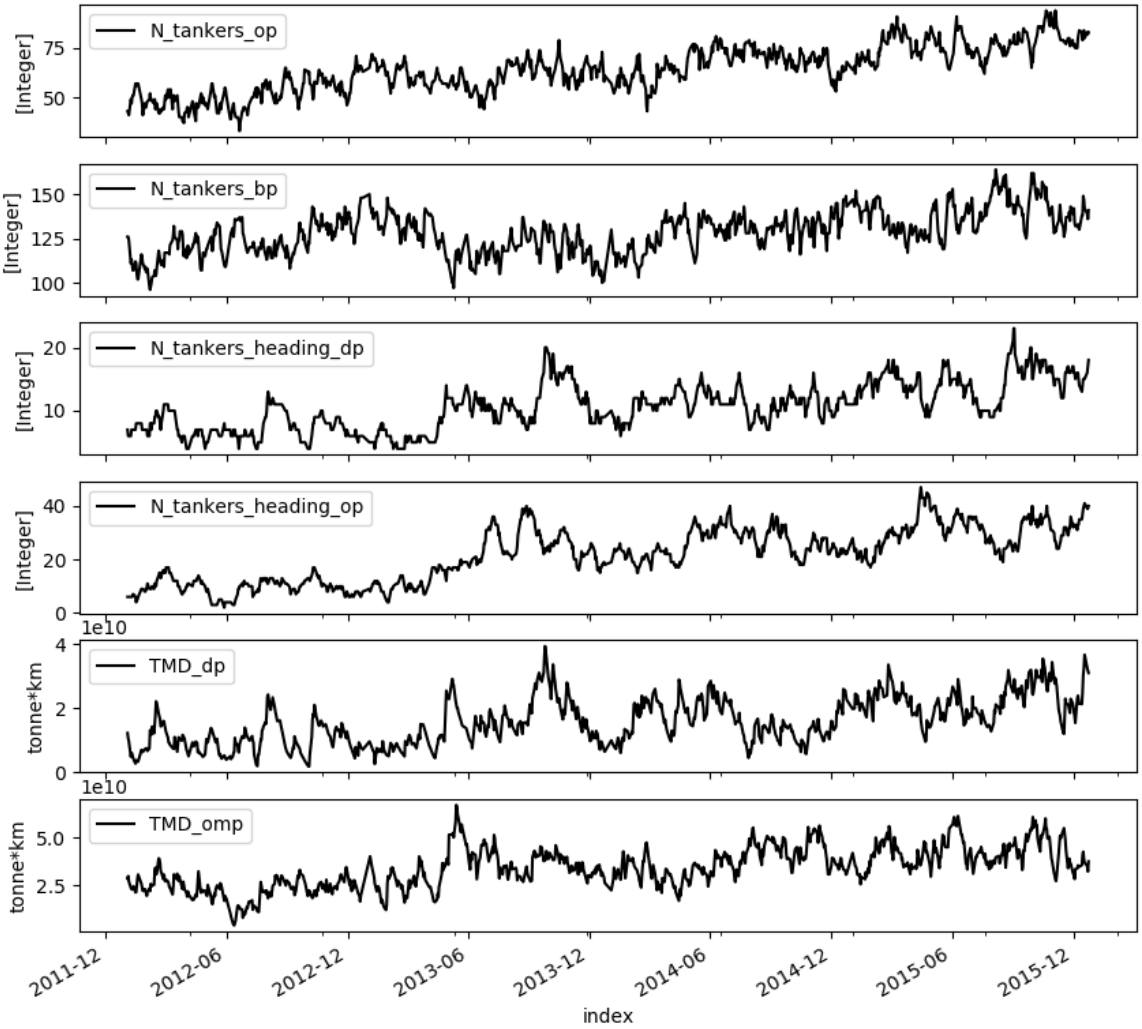


Fig. C.2. *Time-series plots of the AIS-derived data.*

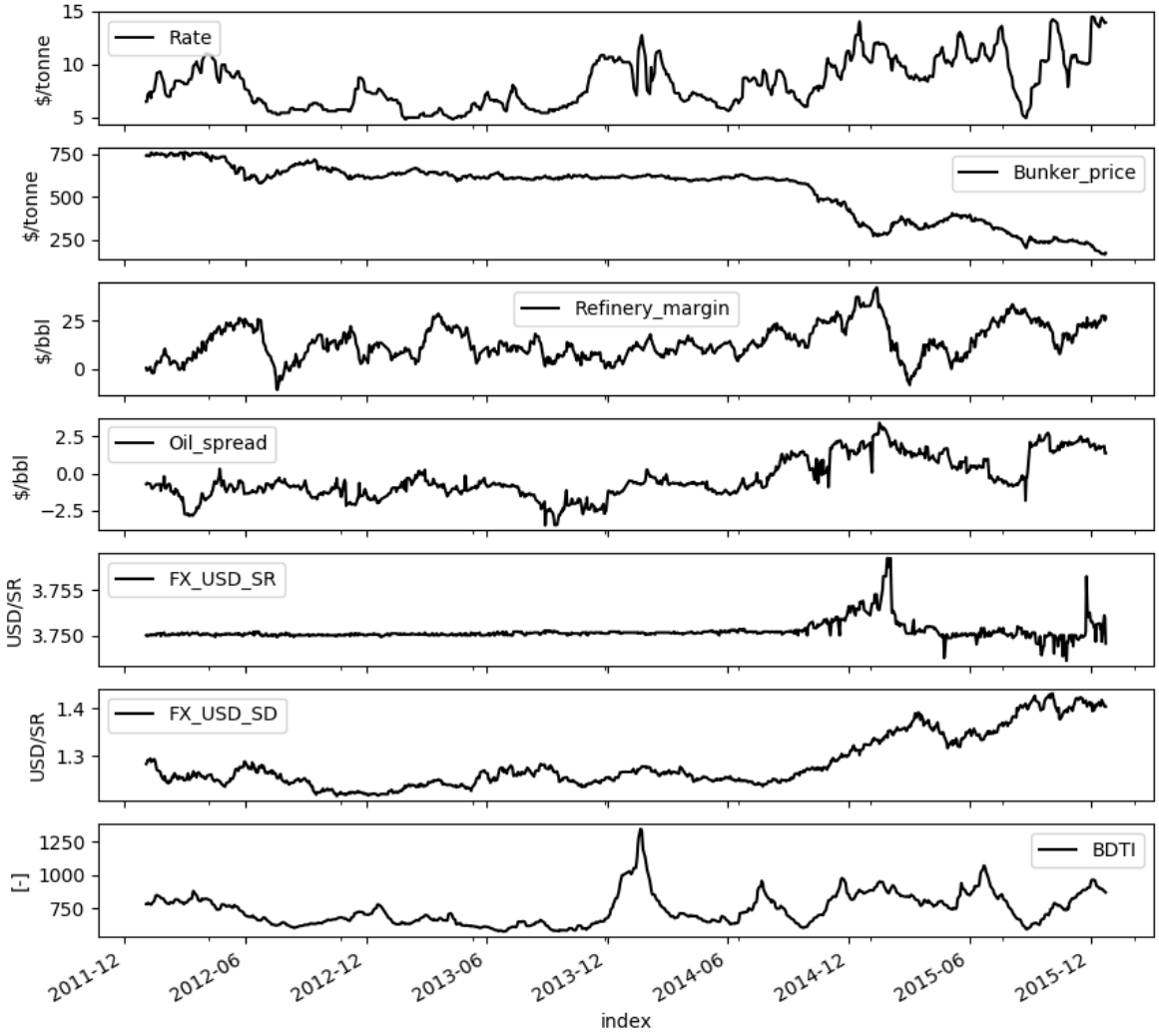


Fig. C.3. *Time-series plots of the non-AIS-derived data.*

Appendix D Normality Plots

Fig. D.1 and D.2 show normality plots of the AIS-derived data and Fig. D.3 presents normality plots of the non-AIS-derived data:

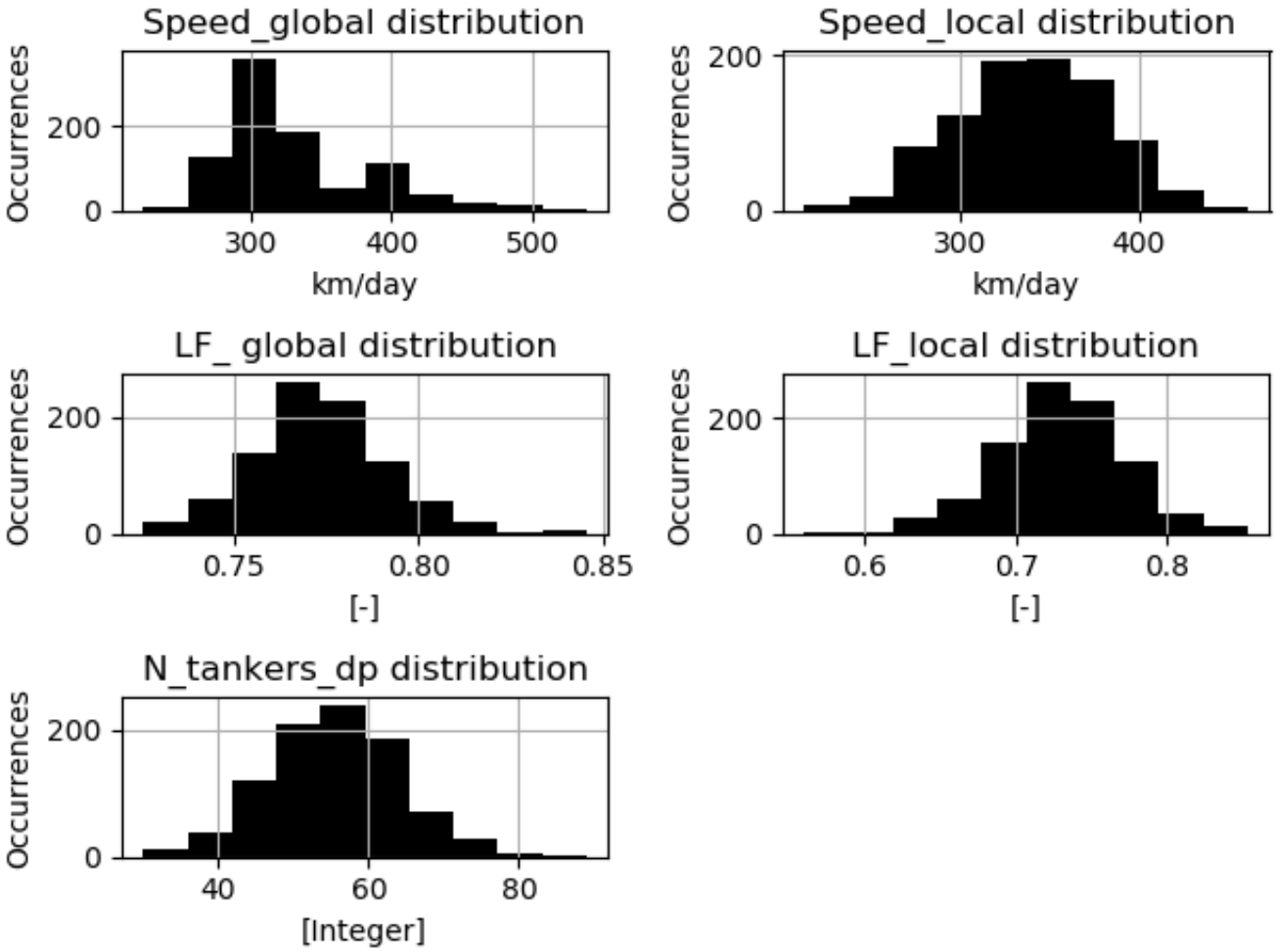


Fig. D.1. *Normality plots of the AIS-derived data.*

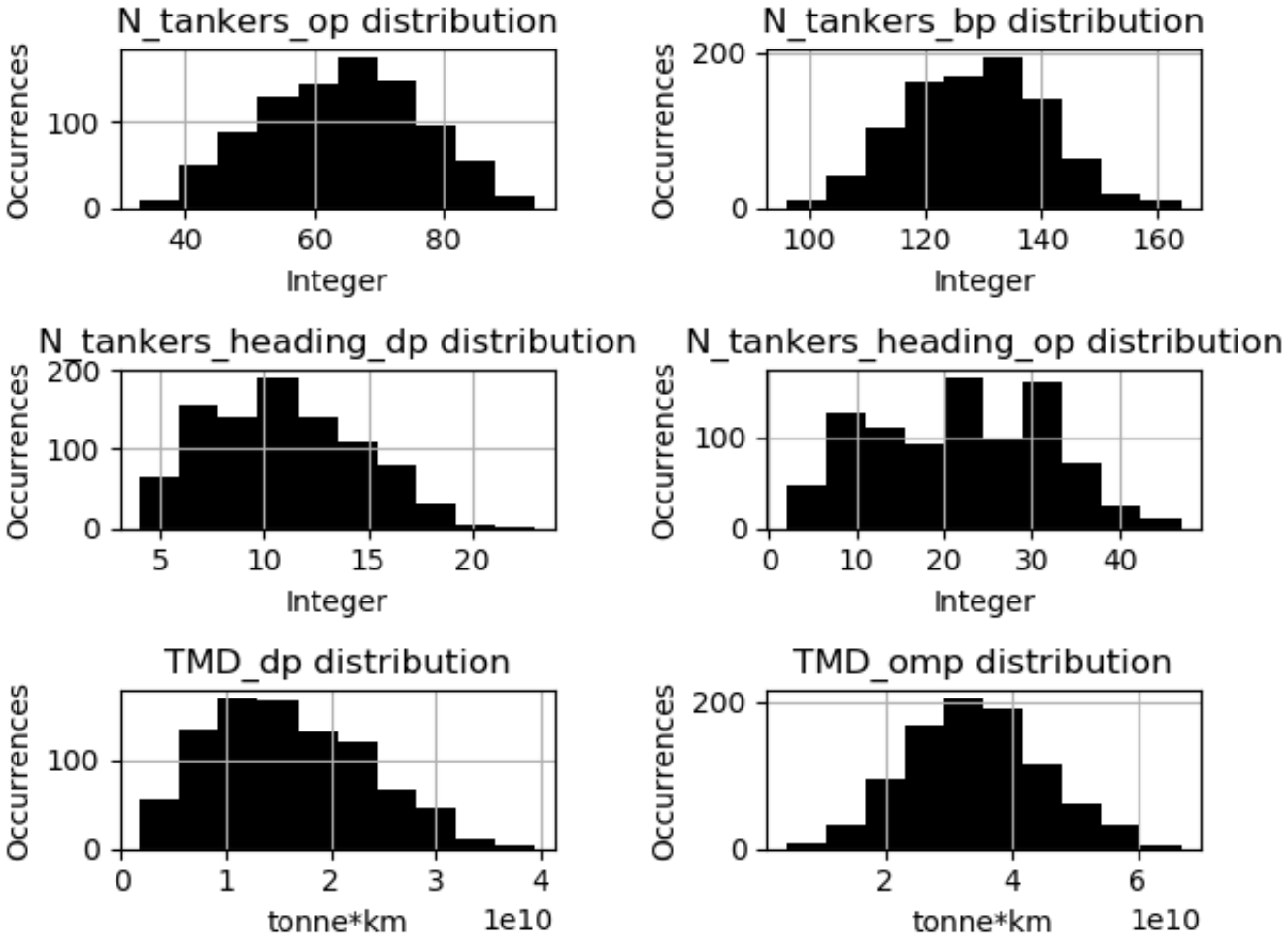


Fig. D.2. *Normality plots of the AIS-derived data.*

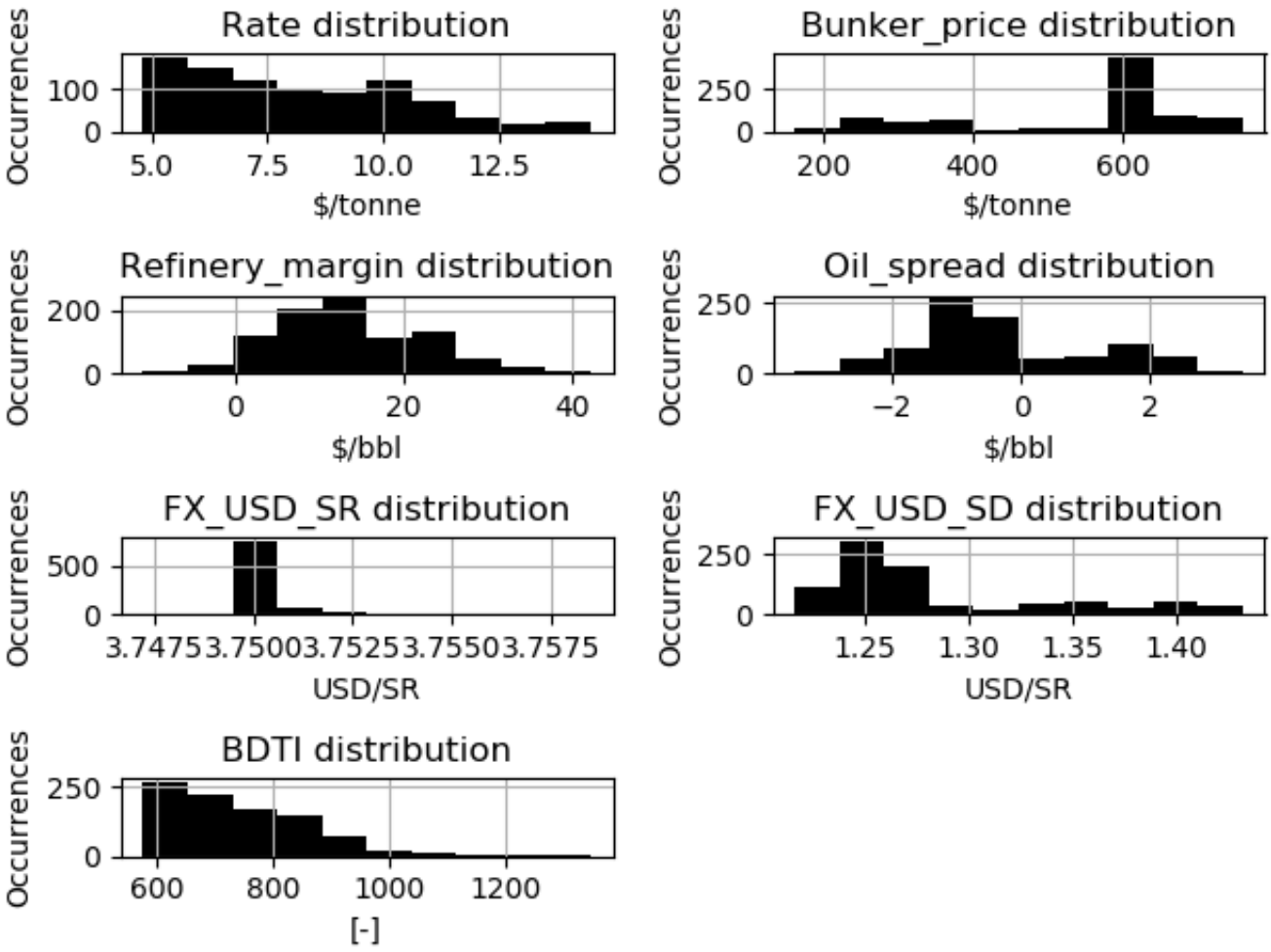


Fig. D.3. *Normality plots of the non-AIS-derived data.*

Appendix F Accuracy Plots

A experiments: Fig. **F.1** illustrates accuracy (the plots to the left) and loss (the plots to the right) for the training data and the test data. Blue lines represent training accuracy, light green lines represent test accuracy, red lines represent training loss and dark green lines represent test loss. Moreover, the top plots represent one-day-ahead forecasts, the middle plots represent five-days-ahead forecasts and the bottom plots represent ten-days-ahead forecasts.

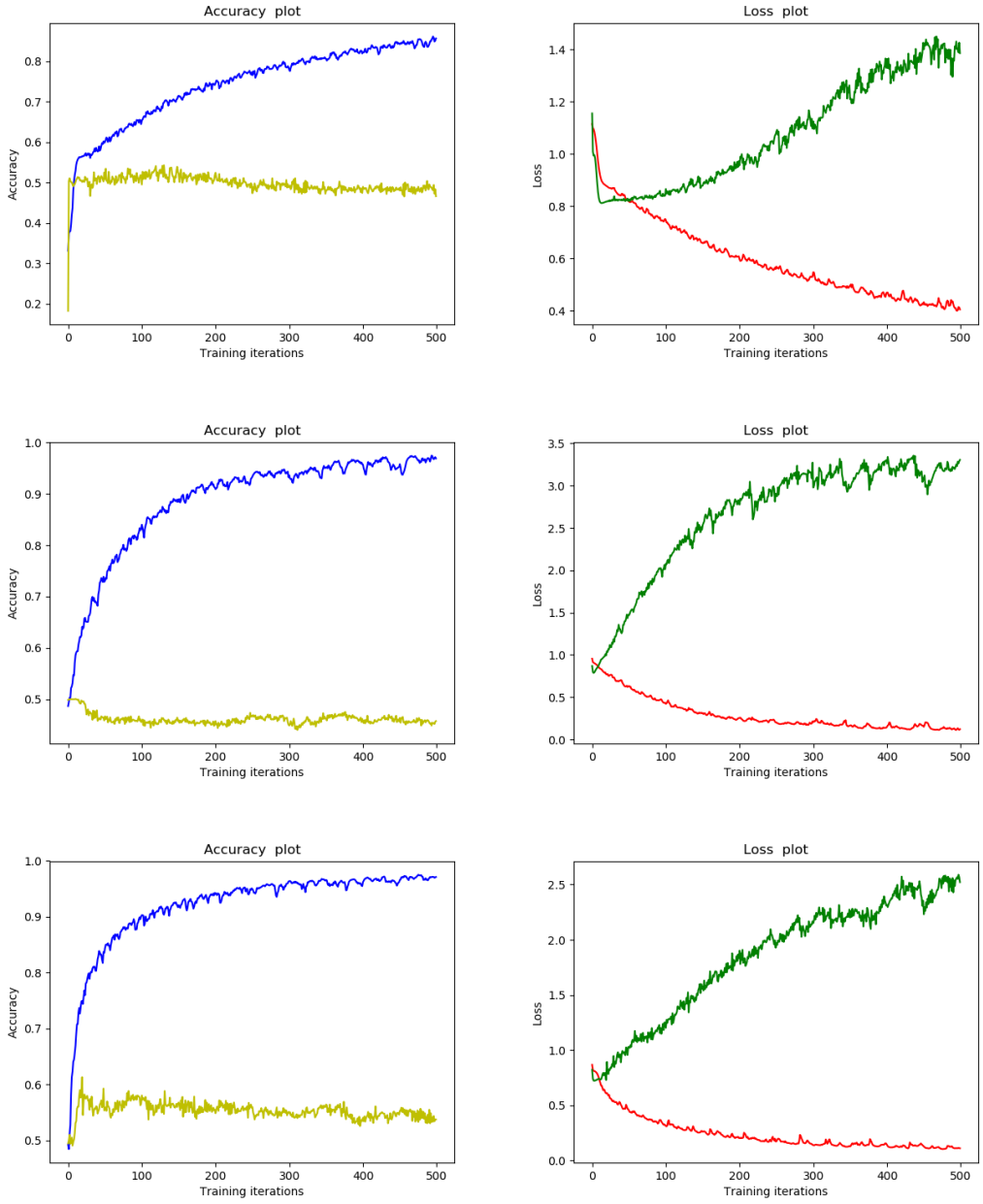


Fig. F.1. Plots displaying training accuracy, test accuracy, training loss and test loss for the A experiments.

B experiments: Fig. **F.2** illustrates accuracy (the plots to the left) and loss (the plots to the right) for the training data and the test data. Blue lines represent training accuracy, light green lines represent test accuracy, red lines represent training loss and dark green lines represent test loss. Moreover, the top plots represent one-day-ahead forecasts, the middle plots represent five-days-ahead forecasts and the bottom plots represent ten-days-ahead forecasts.

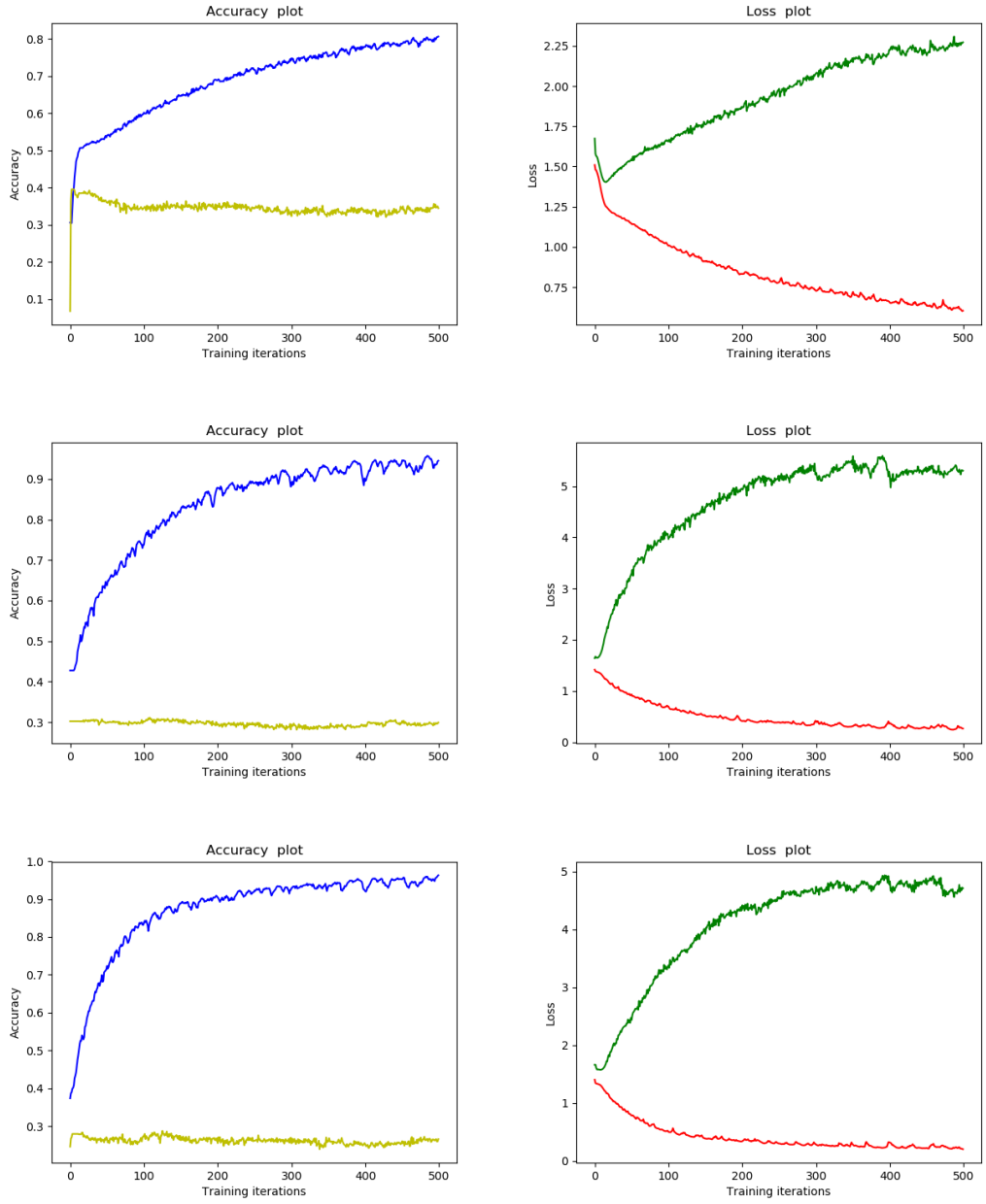


Fig. F.2. Plots displaying training accuracy, test accuracy, training loss and test loss for the B experiments.

C experiments: Fig. **F.3** illustrates accuracy (the plots to the left) and loss (the plots to the right) for the training data and the test data. Blue lines represent training accuracy, light green lines represent test accuracy, red lines represent training loss and dark green lines represent test loss. Moreover, the top plots represent one-day-ahead forecasts, the middle plots represent five-days-ahead forecasts and the bottom plots represent ten-days-ahead forecasts.

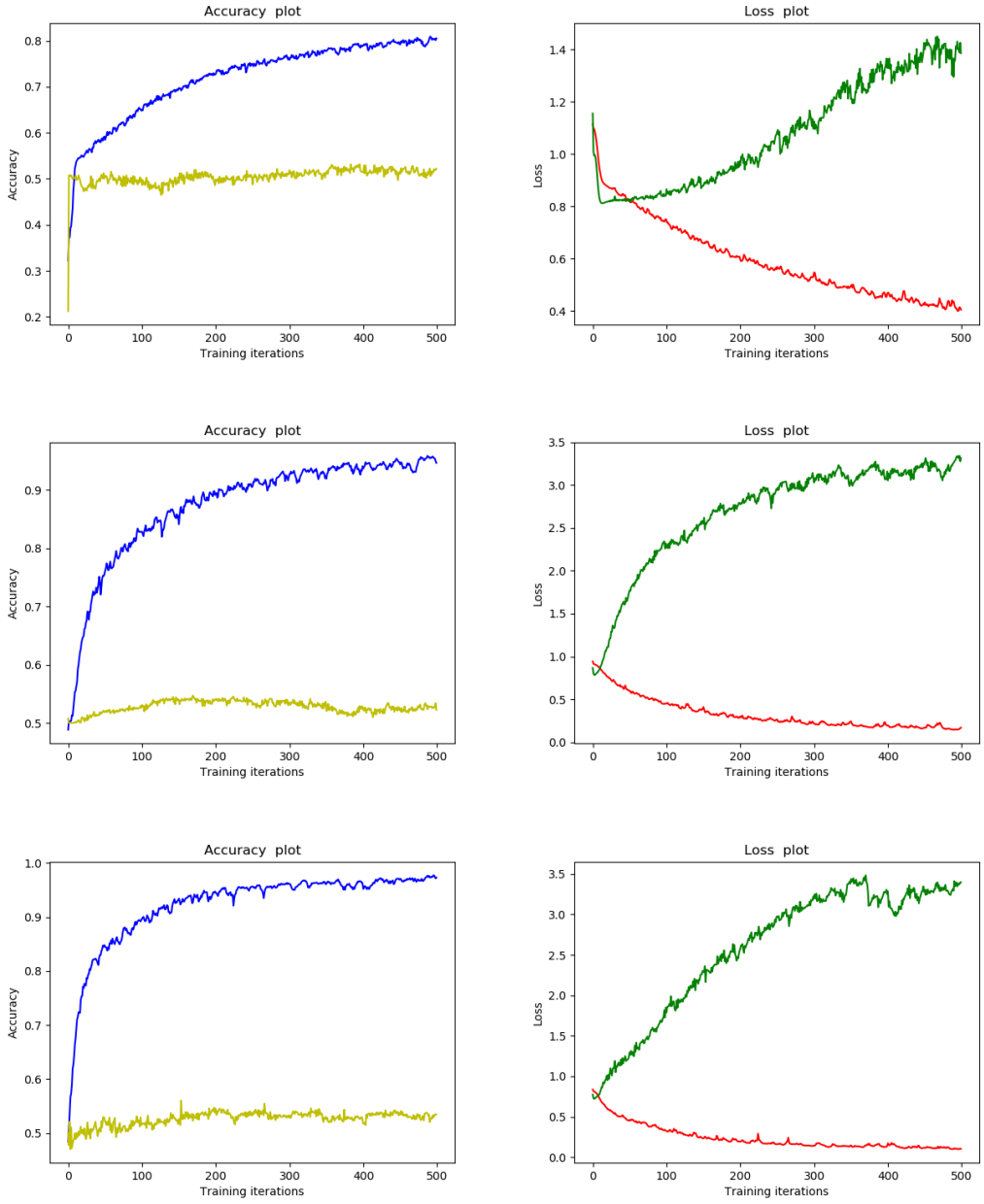


Fig. F.3. Plots displaying training accuracy, test accuracy, training loss and test loss for the C experiments.

D experiments: Fig. **F.4** illustrates accuracy (the plots to the left) and loss (the plots to the right) for the training data and the test data. Blue lines represent training accuracy, light green lines represent test accuracy, red lines represent training loss and dark green lines represent test loss. Moreover, the top plots represent one-day-ahead forecasts, the middle plots represent five-days-ahead forecasts and the bottom plots represent ten-days-ahead forecasts.

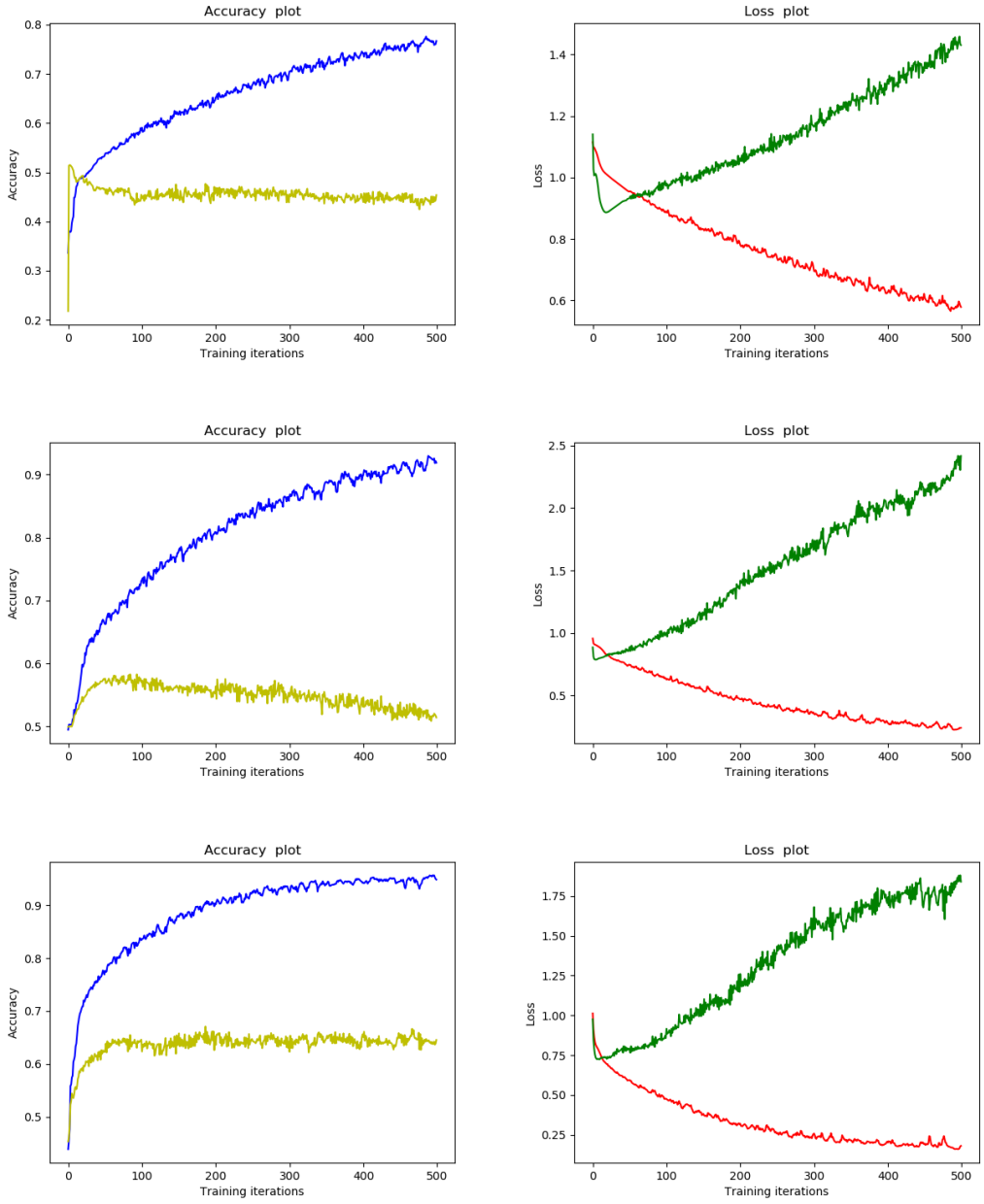


Fig. F.4. Plots displaying training accuracy, test accuracy, training loss and test loss for the D experiments.

E experiments: Fig. **F.5** illustrates accuracy (the plots to the left) and loss (the plots to the right) for the training data and the test data. Blue lines represent training accuracy, light green lines represent test accuracy, red lines represent training loss and dark green lines represent test loss. Moreover, the top plots represent one-day-ahead forecasts, the middle plots represent five-days-ahead forecasts and the bottom plots represent ten-days-ahead forecasts.

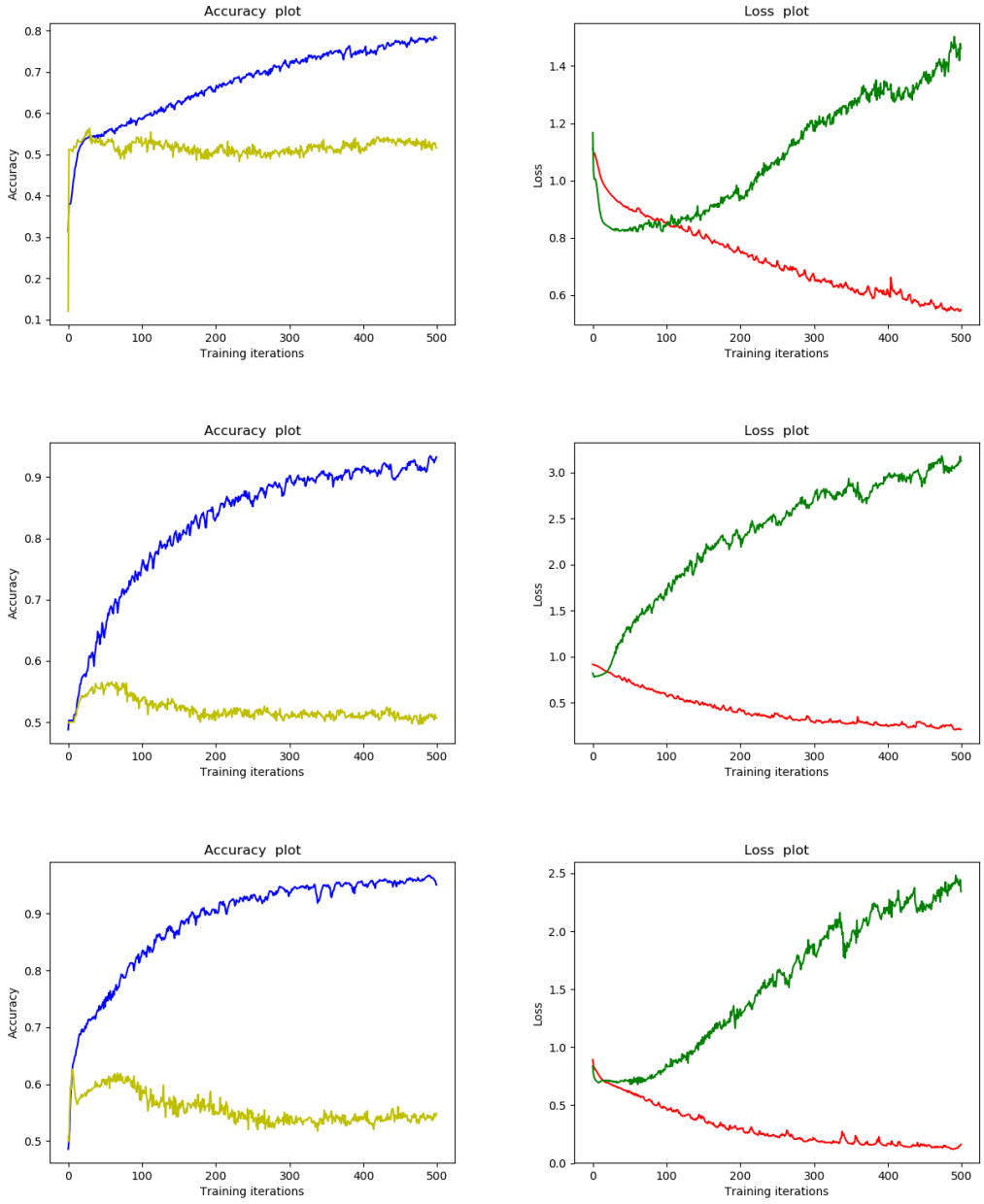


Fig. F.5. Plots displaying training accuracy, test accuracy, training loss and test loss for the E experiments.

F experiments: Fig. **F.6** illustrates accuracy (the plots to the left) and loss (the plots to the right) for the training data and the test data. Blue lines represent training accuracy, light green lines represent test accuracy, red lines represent training loss and dark green lines represent test loss. Moreover, the top plots represent one-day-ahead forecasts, the middle plots represent five-days-ahead forecasts and the bottom plots represent ten-days-ahead forecasts.

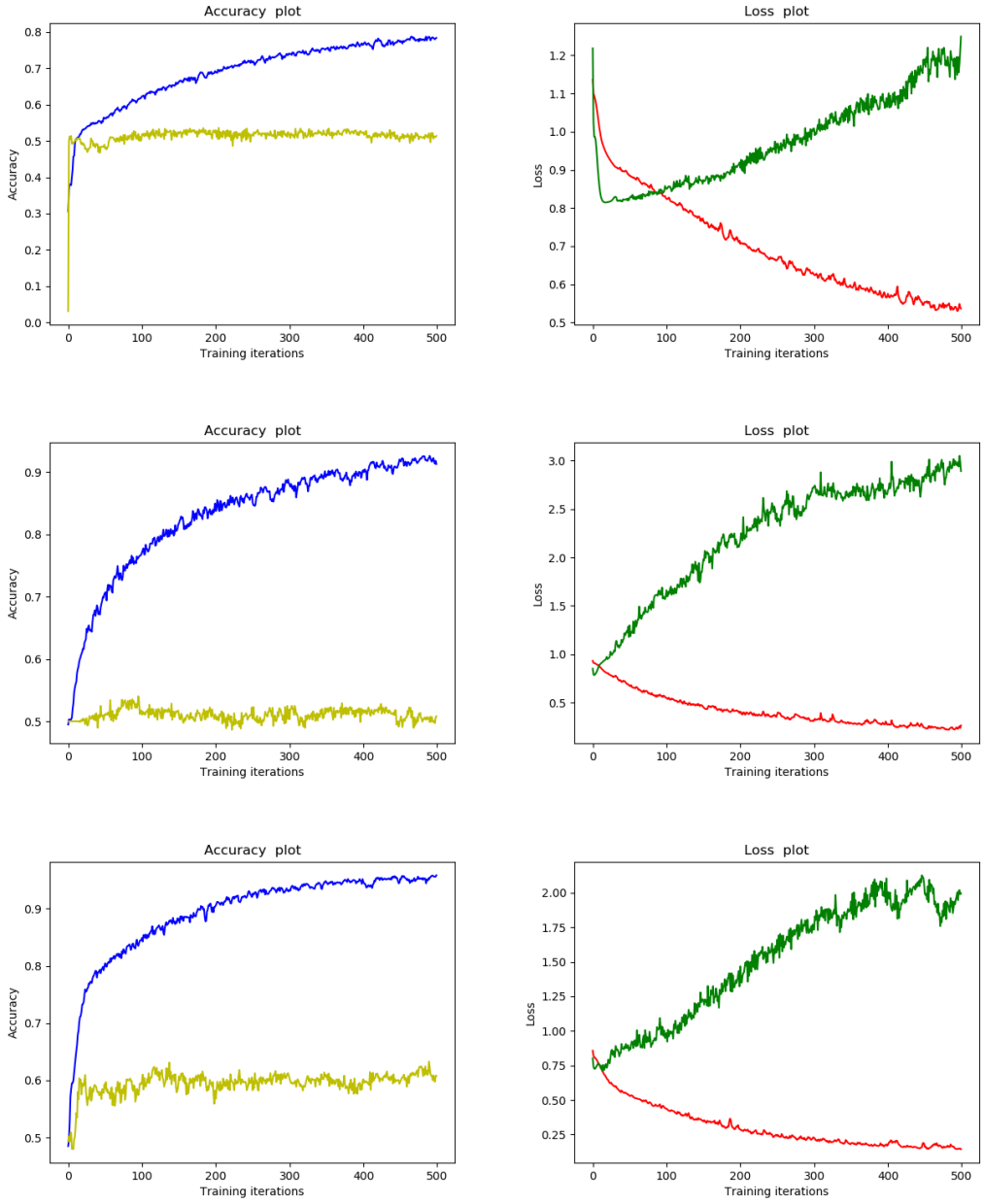


Fig. F.6. Plots displaying training accuracy, test accuracy, training loss and test loss for the F experiments.

References

Adland, R. and Jia, H. (2016). Vessel speed analytics using satellite-based ship position data. In *Industrial Engineering and Engineering Management (IEEM), 2016 IEEE International Conference on*, 1299–1303. IEEE.

Adland, R. and Cullinane, K. (2005). A time-varying risk premium in the term structure of bulk shipping freight rates. *Journal of Transport Economics and Policy (JTEP)*, 39(2), 191–208.

Adland, R. and Cullinane, K. (2006). The non-linear dynamics of spot freight rates in tanker markets. *Transportation Research Part E: Logistics and Transportation Review*, 42(3), 211–224.

Adland, R. and Jia, H. (2018). Dynamic speed choice in bulk shipping. *Maritime Economics & Logistics*, 20(2), 253–266.

Adland, R., Jia, H., and Strandenes, S.P. (2016). The determinants of vessel capacity utilization: The case of brazilian iron ore exports. *Transportation Research Part A: Policy and Practice*.

Adland, R., Jia, H., and Strandenes, S.P. (2017). Are ais-based trade volume estimates reliable? the case of crude oil exports. *Maritime Policy & Management*, 44(5), 657–665.

Adland, R. and Strandenes, S. (2006). Market efficiency in the bulk freight market revisited. *Maritime Policy & Management*, 33(2), 107–117.

Alexander, C. (2009). *Market Risk Analysis, Quantitative Methods in Finance*, volume 1. John Wiley & Sons.

Alizadeh, A. and Kavussanos, M. (2002). The expectations hypothesis of the term structure and risk premia in dry bulk shipping freight markets; an egarch-m approach.

Alizadeh, A. and Nomikos, N. (2009). *Shipping derivatives and risk management*. Springer.

Alizadeh, A.H. and Nomikos, N.K. (2011). Dynamics of the term structure and volatility of shipping freight rates. *Journal of Transport Economics and Policy (JTEP)*, 45(1), 105–128.

Alizadeh, A.H. and Talley, W.K. (2011). Vessel and voyage determinants of tanker freight rates and contract times. *Transport Policy*, 18(5), 665–675.

Askari, S. and Montazerin, N. (2015). A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering. *Expert Systems with Applications*, 42(4), 2121–2135.

Bakshi, G., Panayotov, G., and Skoulakis, G. (2011). The baltic dry index as a predictor of global stock returns, commodity returns, and global economic activity.

Baldi, P. and Sadowski, P.J. (2013). Understanding dropout. In *Advances in neural information processing systems*, 2814–2822.

Batchelor, R., Alizadeh, A., and Visvikis, I. (2007). Forecasting spot and forward prices in the international freight market. *International Journal of Forecasting*, 23(1), 101–114.

Beenstock, M. (1985). A theory of ship prices. *Maritime Policy and Management*, 12(3), 215–225.

Beenstock, M. and Vergottis, A. (1989). An econometric model of the world tanker market. *Journal of Transport Economics and Policy*, 263–280.

Beenstock, M. and Vergottis, A. (1993). The interdependence between the dry cargo and tanker markets. *Logistics and Transportation Review*, 29(1), 3.

Bengio, Y., Frasconi, P., and Simard, P. (1993). The problem of learning long-term dependencies in recurrent networks. In *Neural Networks, 1993., IEEE International Conference on*, 1183–1188. IEEE.

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2), 157–166.

Benth, F.E., Koekebakker, S., and Taib, C.M.I.C. (2014). Stochastic dynamical modelling of spot freight rates. *IMA Journal of Management Mathematics*, 26(3), 273–297.

Campos, J., Ericsson, N.R., and Hendry, D.F. (2005). General-to-specific modeling: an overview and selected bibliography.

Chevalier, G. (2017). Lstms for human activity recognition. URL `https://github.com/guillaume-chevalier/LSTM-Human-Activity-Recognition`.

Chollet, F. and Allaire, J. (2018). Deep learning with r.

Dikos, G., Marcus, H.S., Papadatos, M.P., and Papakonstantinou, V. (2006). Niver lines: a system-dynamics approach to tanker freight modeling. *Interfaces*, 36(4), 326–341.

Dixit, A. (1989). Entry and exit decisions under uncertainty. *Journal of political Economy*, 97(3), 620–638.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78–87.

Engelen, S., Meersman, H., and Voorde, E.V.D. (2006). Using system dynamics in maritime economics: an endogenous decision model for shipowners in the dry bulk sector. *Maritime Policy & Management*, 33(2), 141–158.

Fan, S., Ji, T., Gordon, W., and Rickard, B. (2013). Forecasting baltic dirty tanker index by applying wavelet neural networks. *Journal of Transportation Technologies*, 3(01), 68.

Furset, O. and Hordnes, E. (2013). *The VLCC tanker market: the present, past and future: a historical fleet analysis followed by a stochastic partial equilibrium model of the spot freight market*. Master's thesis.

Γούλας, L. (2012). *Forecasting freight rates: evidence from the baltic exchange indices and the IMAREX freight futures*. Master's thesis.

Gilleshammer, P.K. and Hansen, J.Ø. (2010). *Hedging risks in shipping using futures contracts traded on Imarex*. Master's thesis.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Hampton, M.J. (1990). *Long and short shipping cycles: the rhythms and psychology of shipping markets*. Cambridge Academy of Transport.

Hawdon, D. (1978). Tanker freight rates in the short and long run. *Applied Economics*, 10(3), 203–218.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.

International Maritime Organisation (IMO, 2004), title=Resolution A.956(23): AMENDMENTS TO THE GUIDELINES FOR THE ONBOARD OPERATIONAL USE OF SHIPBORNE AUTOMATIC

IDENTIFICATION SYSTEMS (AIS) (RESOLUTION A.917(22)), y.n. (????).

Investopedia (2018). Baz. https://www.investopedia.com/terms/p/perfectcompetition.asp.

ITU, R. (2014). M. 1371-5-technical characteristics for an automatic identification system using time-division multiple access in the vhf maritime mobile band. *International Telecommunications Union*.

Jia, H., Prakash, V., and Smith, T. (2015). Estimating vessel utilization in the drybulk freight market: The reliability of draught reports in ais data feeds. In *ECONSHIP conference, Chios, June*, 24–27.

Jia, H., Lampe, O.D., Solteszova, V., and Strandenes, S.P. (2017). An automatic algorithm for generating seaborne transport pattern maps based on ais. *Maritime Economics & Logistics*, 19(4), 619–630.

Jugović, A., Komadina, N., and Perić Hadžić, A. (2015). Factors influencing the formation of freight rates on maritime shipping markets. *Pomorstvo*, 29(1), 23–29.

Kaluza, P., Kölzsch, A., Gastner, M.T., and Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48), 1093–1103.

Kavussanos, M.G. (1996). Comparisons of volatility in the dry-cargo ship sector: Spot versus time charters, and smaller versus larger vessels. *Journal of Transport economics and Policy*, 67–82.

Kavussanos, M.G. (2003). Time varying risks among segments of the tanker freight markets. *Maritime Economics & Logistics*, 5(3), 227–250.

Kavussanos, M.G. and Alizadeh-M, A.H. (2001). Seasonality patterns in dry bulk shipping spot and time charter freight rates. *Transportation Research Part E: Logistics and Transportation Review*, 37(6), 443–467.

Kavussanos, M.G. and Alizadeh-M, A.H. (2002). Seasonality patterns in tanker spot freight rate markets. *Economic Modelling*, 19(5), 747–782.

Kavussanos, M.G., Nomikos, N.K., et al. (1999). The forward pricing function of the shipping freight futures market. *Journal of futures markets*, 19(3), 353–376.

Kavussanos, M.G. and Visvikis, I.D. (2006). Shipping freight derivatives: a survey of recent evidence. *Maritime Policy & Management*, 33(3), 233–255.

Kingma, D.P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koekebakker, S., Adland, R., and Sødal, S. (2006). Are spot freight rates stationary? *Journal of Transport Economics and Policy (JTEP)*, 40(3), 449–472.

Koopmans, T.C. (1939). *Tanker Freight Rates and Tankship Building: An Analysis of Cyclical Fluctuations, by Dr. T. Koopmans*. 27. De erven F. Bohn nv.

Leonhardsen, J.H. (2017). *Estimation of Fuel Savings from Rapidly Reconfigurable Bulbous Bows-Exemplifying the Value of Agility in Marine Systems Design*. Master's thesis, NTNU.

Leonov, Y. and Nikolov, V. (2012). A wavelet and neural network model for the prediction of dry bulk shipping indices. *Maritime Economics & Logistics*, 14(3), 319–333.

Lun, Y.V., Lai, K.h., and Cheng, T.E. (2010). *Shipping and logistics management*. Springer.

Lyridis, D., Zacharioudakis, P., Mitrou, P., and Mylonas, A. (2004). Forecasting tanker market using artificial neural networks. *Maritime Economics & Logistics*, 6(2), 93–108.

Magnussen, E., Amdahl, J., and Fuglerud, G. (2014). Marin teknikk grunnlag: Tmr4105: kompendium (5. utg. ed.). *Trondheim: Akademika forlag Kompendieforlaget*.

Mitchell, T.M. et al. (1997). Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45(37), 870–877.

Molvik, E.H. and Stafseng, M.S. (2018). *Forecasting Time Charter Equivalent Oil Tanker Freight Rates - determinant driven, route-specific Markov regime-switching models*. Master's thesis, NTNU.

Mossin, J. (1968). An optimal policy for lay-up decisions. *The Swedish Journal of Economics*, 170–177.

Olah, C. (2015). Understanding lstm networks. URL http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Olsen, M. and da Fonseca, T.R.K. (2017). *Investigating the predictive ability of AIS-data: the case of arabian gulf tanker rates*. Master's thesis.

Poulakidas, A. and Joutz, F. (2009). Exploring the link between oil prices and tanker rates. *Maritime Policy & Management*, 36(3), 215–233.

Randers, J. and Göluke, U. (2007). Forecasting turning points in shipping freight rates: lessons from 30 years of practical effort. *System Dynamics Review*, 23(2-3), 253–284.

Ringheim, J.K. and Stenslet, L.O.D. (2017). *Predicting monthly bulk shipping freight rates*. Master's thesis, NTNU.

Skauen, A.N., Helleren, Ø., Olsen, Ø., and Olsen, R. (2013). Operator and user perspective of fractionated ais satellite systems. In *Proceedings of the AIAA/USU Conference on Small Satellites, Around the Corner, SSC13-XI-5*.

Smestad, B.B. (2015). *A Study of Satellite AIS Data and the Global Ship Traffic Through the Singapore Strait*. Master's thesis, NTNU.

Smith, T., Jalkanen, J., Anderson, B., Corbett, J., Faber, J., Hanayama, S., O'Keeffe, E., Parker, S., Johanasson, L., Aldous, L., et al. (2015). Third imo ghg study.

Smith, T., O'Keeffe, E., Aldous, L., and Agnolucci, P. (2013). Assessment of shipping's efficiency using satellite ais data.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.

Stopford, M. (2009). *Maritime economics*. Routledge.

Strandenes, S. (1984). Price determination in the time charter and second hand markets. *Center for Applied Research, Norwegian School of Economics and Business Administration, working paper MU*, 6.

Tamvakis, M.N. (1995). An investigation into the existence of a two-tier spot freight market for crude oil carriers. *Maritime Policy and Management*, 22(1), 81–90.

Tamvakis, M.N. and Thanopoulou, H.A. (2000). Does quality pay? the case of the dry bulk market. *Transportation Research Part E: Logistics and Transportation Review*, 36(4), 297–307.

TensorFlow (2015). Module implementing rnn cells. URL http://colah.github.io/posts/

2015-08-Understanding-LSTMs/.
Tham, E. (2008). Leading indicators for arabian gulf oil tanker rates. *OPEC Energy Review*, 32(2), 139–149.
USCG (2018). Automatic identification system uscg ais encoding guide. https://www.navcen.uscg.gov/pdf/AIS/AISGuide.pdf.
Veenstra, A.W. and Franses, P.H. (1997). A co-integration approach to forecasting freight rates in the dry bulk shipping sector. *Transportation Research Part A: policy and practice*, 31(6), 447–458.
Zannetos, Z.S. (1966). *The theory of oil tankship rates: an economic analysis of tankship operations*. 4. Cambridge, Mass., MIT Press [1966].