Aalto University
School of Business

VIRTUAL FUEL FLOW ESTIMATION IN SHIPPING

How to formulate and validate a model for time-dependent momentary fuel flow estimation by combining high-frequency GPS and meteorological data with low-frequency noon reports – a case study

Master's Thesis
Philipp Martin Back
Aalto University School of Business
Information & Service Management
Fall 2017

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author	Philipp Martin Back
Title of thesis	Virtual Fuel Flow Estimation in Shipping
Degree	Master of Science in Economics and Business Administration
Degree programme	Information & Service Management
Thesis advisors	Pekka Malo and Juuso Liesiö
Year of approval	2017
Number of pages	98
Language	English

Abstract

This thesis validates a novel approach for virtual fuel flow estimation on commercial vessels, such as container ships, tankers, and bulkers. A model is fitted onto high-frequency GPS and meteorological data, and onto low-frequency crew-generated noon reports. The research design is a quantitative case study and utilizes two months of real-life vessel data from a single containership. Based on publicly available Python packages, the validation study examines the benefits of the modelling approach presented by Antola, Solonen, and Staboulis (2017).

Stakeholders in the marine shipping industry are under increasing economic, environmental, and regulatory pressure to improve the energy efficiency of commercial vessels. A growing number of 3rd-party providers has entered the market and offers data-driven solutions for vessel performance monitoring and optimization. Most of them require tight integration to onboard systems to collect the necessary data. In a study on barriers to energy efficiency in shipping, Rehmatulla and Smith (2015) found that principal-agent-problems prevent a widespread diffusion of energy-efficient measures, even if they are cost-effective. Especially for small and chartered vessels, costly integration-based systems remain unfeasible. Consequently, a lightweight system is needed that can monitor vessel and crew performance, without being subject to the barriers to energy efficiency. Antola, Solonen, and Staboulis (2017) proposed a novel approach that uses virtual sensing techniques to model some of the quantities that require costly onboard integration in traditional solutions: vessel speed-through-water and instantaneous fuel flow rates.

My validation study confirms that the proposed modelling approach provides reasonably accurate speed and fuel flow estimates that deviate on average by ca. 10% from the reference data. However, the modelled output is noisier and can be affected by prolonged periods of erroneous crew reporting. Yet, model-based speed-fuel-curves allow stakeholders to demonstrate a vessel's operational profile for various speed and draft levels.

The theoretical main contribution of my study is a transparent validation that highlights the benefits of the novel modeling technique by Antola, Solonen, and Staboulis (2017). For most customers, the alternative to the (slightly inaccurate) lightweight system is not a more accurate integration-based solution, but a sole reliance on crew-generated noon reports. The managerial implication of this study is therefore that the lightweight system offers reasonable accuracy, significant customer value, and a way to reduce the information asymmetry between charterer and ship owner; at a fraction of the price of established integration-based solutions.

Keywords virtual sensor, marine shipping, vessel performance monitoring

Acknowledgements

This Master's thesis was carried out between March and November 2017 as part of my M.Sc. (econ) studies in Information & Service Management at Aalto University School of Business.

I would like to express my sincere gratitude to my thesis advisors, Pekka Malo, Ph.D., and Juuso Liesiö, D.Sc. (Tech.), for their guidance and valuable input throughout the thesis process. Pekka helped me to shape some rough ideas around "ship data" and "predictive modeling" into a coherent topic and showed a lot of patience when I asked (yet again) to develop the topic into a new direction. Moreover, Pekka encouraged me to not settle for an easy topic, but to seize my Master's thesis as an opportunity to develop my skillset as a data analyst further. During the past months, I finished my thesis, but more important, I learned how to perform data analysis in Python and executed a data analysis project from start to finish. Pekka's advice gave a deeper meaning to the (tedious) thesis process. It made me develop my professional skillset, and turned my Master's thesis into a true culmination of my studies – for that I am truly grateful! Special thanks also go to Juuso, who joined the thesis project at the final stage. Despite a tight schedule, he provided invaluable feedback when it was most needed! Juuso's input and suggestions added the final touch to my work.

Moreover, I would like to thank Stratos Staboulis, Ph.D. and Matti Antola, Ph.D. from Eniram who have made their support available in numerous ways. They helped me to access a topic that was certainly over my head when I first encountered it at the beginning of the thesis process, offered their expertise to find a feasible and worthwhile research approach, and guided me throughout the process. I am grateful for their time commitment when proof-reading my work and for the numerous inputs on how to improve the quality and research contribution of my study.

Finally, I would like to thank the whole Eniram team, for welcoming me into their family and for offering me their guidance and support.

Table of Contents

List of Tables

List of Figures

1 Introduction

The Internet of Things (IoT) describes the ongoing trend of augmenting physical objects and devices with sensing, computing, and communication capabilities that enable these objects to collect and exchange data (Guo et al., 2013). Thereby, the IoT also transforms our understanding of sensors, as humans are no longer the only users of data. Sensor data is increasingly used by automated algorithms, such as virtual sensors (also known as soft or inferential sensors). According to Fortuna, Graziani, Rizzo, and Xibilia (2007), "soft sensors are mathematical models that allow us to infer relevant variables on the basis of their dependence on a set of influential variables." They can "estimate a difficult to measure or expensive quantity using one or more mathematical models along with lower cost physical sensors." (Li, Yu, & Braun, 2011). With the IoT spreading into more and more areas of our lives, we become surrounded by products that are only possible thanks to advances in virtual sensing. For example, fitness trackers such as FitBit can replace expensive metabolic measurement devices by combining readings from inexpensive accelerometers in the wrist band with modelling algorithms to estimate e.g. the calories burnt during exercise.

In the past decade, virtual sensors have become increasingly widespread in domains such as automobiles, industrial process control, and avionics, yet they remain underutilized and under-researched in the marine shipping industry. In this paper, I present a novel approach for virtual fuel flow estimation that can help to overcome shipping-specific barriers to energy efficiency, and make vessel performance monitoring tools available for small ships and chartered vessels.

The marine shipping industry enables global trade by connecting the world's raw materials with labor markets and consumers. In 2012, commercial vessels, such as container ships, tankers, and bulkers, carried 90% of the global trade by volume (Smith et al., 2015). Thus, shipping plays a vital role in the world's economic development and enables globalization. The fuel onboard ships (commonly referred to as "bunkers") represents the largest single operational cost item of a vessel and accounts for 50% to 70% of the total voyage costs (Branch & Stopford, 2013). Additionally, environmental protection and global warming are of growing concern. The marine industry accounted for about 4% of total worldwide carbon dioxide (CO_2) emission in 2007, leading authorities such as the International Maritime Organization (IMO) to drastically limit the ecological footprint of ships (Reynolds, 2009).

A growing number of 3rd-party providers has entered the market and offers data-driven solutions that can help marine stakeholders to reduced fuel expenses, lower CO_2 emissions, and improve the asset management. In this thesis, I will use the term **"integration-based system"** to describe all commercially available tools that collect data by tapping into a ship's onboard automation and navigation systems (the term also includes systems that complement the data with readings from additional sensors along the hull). The tight onboard integration of these integration-based systems offers high-fidelity data streams on various performance quantities (e.g. propulsion power, fuel flow, and vessel speed-through-water), and allows for real-time data processing. However, the project-based installation of these systems is expensive and requires significant upfront investments from the ship owner. Despite their energy- and cost-efficiency (positive net present value), the implementation rate of data-driven integration-based vessel performance systems remains low across the industry (Eide, Endresen, Skjong, Longva, & Alvik, 2009).

Rehmatulla and Smith (2015) conducted the first scientific study on barriers to energy efficiency in shipping and found principal-agent problems in the form of split incentives and information asymmetry to make costly integration-based systems unfeasible for a large customer segment. Especially on chartered vessels, the ship-owner is responsible for the technological and operational efficiency of the ship, while the charterer (who hires the vessel for a limited time) has to cover the actual fuel bill of a voyage. Hence, neither the ship owner nor the charterer have an incentive to invest in expensive integration-based fuel saving solutions. To overcome this dilemma, a lightweight system is needed that can track e.g. ship fuel consumption without being subject to the barriers identified by Rehmatulla and Smith (2015). Such a system would enable the charterer (principal) to independently monitor vessel and crew (agent) performance and reduce the information asymmetry in their agency relationship. As chartered-vessels represent around 80% of the market, a lightweight vessel performance monitoring tool would yield tremendous business opportunities; if implemented on a wide scale, it could also significantly reduce shipping-related CO_2 emissions.

In practice, many quantities that are essential for evaluating vessel performance (e.g. rounds-per-minute at the engine shaft, fuel flow) cannot be directly measured without access to a ship's onboard systems. However, Antola, Solonen, and Staboulis (2017) claim that some of these variables can be obtained indirectly by using virtual sensing techniques. The Finnish company Eniram has developed a lightweight device that measures vessel speed-over-ground (SOG) via GPS and combines the readings with high-frequency meteorological

data and low-frequency crew-generated reports to estimate vessel speed-through-water (STW) and instantaneous fuel flow rates through modeling. In other words, the system uses virtual sensing to estimate some not-directly-available variables that are measured directly via physical sensors (e.g. a fuel flow meter) in most commercially available integration-based systems. The lightweight approach drastically reduces the costs for performance monitoring and overcomes the barriers to energy efficiency that were identified by Rehmatulla and Smith (2015).

1.1 Research Gap

Despite these promising findings, some research gaps remain in the work of Antola, Solonen, and Staboulis (2017). Although the authors discuss the mathematical framework, accuracy, and potential shortcomings of the lightweight system, their work is a rather brief validation of the underlying model. A transparent, more extensive validation study could therefore be beneficial to highlight the advantages of the modeling approach by Antola, Solonen, and Staboulis (2017).

The need for such a study can be highlighted in a simple example: assuming that fuel is the main expense for the customer of a chartered-vessel, considering 280 yearly running days, an average fuel consumption of 150 tons/day, and an average fuel price of 300 USD/ton, a mere 5% error in the fuel calculations would accumulate to over 630,000 USD per year (Bialystocki & Konovessis, 2016). Hence, the daily hiring rate that the customer has to pay to the ship-owner would be too high or too low by roughly USD 1,700 per day.

Consequently, a transparent validation study would complement the work by Antola, Solonen, and Staboulis (2017) in two ways. From an academic point-of-view, the study would contribute a second verification that might help to establish the lightweight model as a concrete method on how to overcome the barriers to energy efficiency in shipping (Rehmatulla & Smith, 2015). From a managerial point-of-view, a transparent validation study could increase customer confidence in a novel and seemingly complex product based on modern data analytics – especially in an industry that is notoriously reluctant to implement new technologies. Since the lightweight system is marketed as an independent monitoring tool that enables the charterer to monitor crew and vessel performance, a sensitivity analysis with regards to the crew-generated noon reports would be especially interesting. It would study the extent to which the crew can influence the model predictions, either by carelessly or even willfully reporting flawed values. The below research questions and methods describe how the research gap is addressed in this thesis.

1.2 Research Questions and Method

In this Master's thesis, I validate the novel modeling approach by Antola, Solonen, and Staboulis (2017) exemplarily. A practically applicable model is built from publically available tools and applied to real-life vessel data to answer the following **research questions**:

1. How to formulate a virtual sensor for time-dependent momentary fuel flow estimation on commercial vessels?
2. How well does the model perform in comparison to integration-based systems?

To answer these questions, I have structured my **research method** into four distinct stages:

1. Fit a model for weight parameter estimation onto low-frequency crew reports and onto aggregated high-frequency GPS and meteorological data.
2. Predict instantaneous fuel flow by applying the estimated weight parameters from (1) to unaggregated high-frequency GPS and meteorological data.
3. Assess the model performance by comparing the fuel flow predictions from (2) against measurement-based reference data.
4. Assess the effect of human reporting error by simulating error-free reports from reference data.

1.3 Structure of the Thesis

My Master's thesis is structured in the following way: the literature review in Chapter two discusses the economic, environmental, and regulatory forces that drive the need for data-driven vessel performance monitoring tools in the marine industry. The chapter presents commercially available integration-based solutions and outlines the reasons why these systems are unfeasible for a large market segment. Chapter two further presents a solution in the form of a virtual-sensing-based approach that can overcome this dilemma and potentially allow for an industry-wide diffusion of performance monitoring tools. After the literature review, Chapter three proposes my research methodology. More precisely, the methodology outlines the steps for formulating a model for virtual fuel flow estimation and how it can be validated. Chapter four discusses the data that is used in my study. The fifth chapter presents my research results and also discusses their theoretical and managerial implications. Finally, the conclusion in Chapter six summarizes my study.

2 The Need for a Lightweight Vessel Performance Monitoring System

In order to understand why systems for vessel performance monitoring and optimization (lightweight and integration-based) are needed, one has to study the economic, environmental, and regulatory forces that stakeholders in the maritime industry face.

2.1 Economic, Environmental, and Regulatory Forces in Shipping

With the exception of cruise ships, maritime shipping is considered derived demand: it exists in response to the demand for freight transportation. Historically, there has been a strong correlation between global Gross Domestic Product (GDP) and the demand for shipping, thus GDP can be used to some degree to estimate the future demand for shipping. It should be noted, however, that this relationship has shown signs of decoupling in recent years. Assuming an annual GDP growth rate of 3% to 4%, the International Maritime Organization (IMO) nonetheless expects shipping activity to increase between 200% and 300% by the year 2050 (Buhaug, 2009; Smith et al., 2015).

In light of the IMO's estimations, one must wonder how this development will impact the environment and contribute to global warming. In 2012, 13% of the global CO_2 emissions were caused by the transport sector, with international shipping emitting around 2.2%; that is 796 million tons of CO_2 (Smith et al., 2015). To put these numbers into context: if the shipping industry was a country, it would rank as the sixth largest producer of greenhouse gas and exceed the emissions of Germany (Harrould-Kolieb, 2008). The shipping industry contributes significantly to the overall emission of greenhouse gas (GHG) and is facing growing pressure to implement measures that reduce emissions (Eide et al., 2009). It has been noted that the contribution to total emissions in relation to the volume of transported cargo makes shipping the most energy efficient form of transport when compared to e.g. road, rail, or air trafficking (Buhaug, 2009). The expected growth in shipping activity will nevertheless cause CO_2 emissions from international shipping to grow between 50% to 250% by the year 2050 (Smith et al., 2015). In other words, while shipping is comparable energy efficient, the predicted growth in activity will still bring a significant increase in absolute shipping-related CO_2 emissions.

In addition to environmental concerns, there are strong economic reasons to increase energy efficiency in ships. Fuel costs comprise 50% to 70% of a vessel's operating costs, surpass the crew wages, and are likely to increase as the price of Heavy Fuel Oil (HFO; a

type of residual fuel oil that is the predominantly used type of fuel for ships) is rising (Bialystocki & Konovessis, 2016). Consequently, there exists a linear relationship between the economic interest in fuel efficient vessels and fuel prices. In the 1970's and 1980's, HFO prices increased nearly ten-fold, leading inefficient ships to be decommissioned. When the oil price fell between 1985 and 2000, research on energy efficient measures did not receive much attention in the maritime industry. However, when the crude oil price climbed in the 2000's, engine manufacturers and shipyard designers started again to investigate solutions to reduce fuel consumption and increase energy efficiency in ships (Bialystocki & Konovessis, 2016). While fuel-efficient measures will directly reduce operational costs, the question is whether (considerable) investments in new technologies will amortize over time. This can be expressed by a measure's cost-effectiveness, which is achieved when it is both economically efficient (positive net present value) and energy efficient (Golove & Eto, 1996).

The 2nd IMO GHG Study (Buhaug, 2009) states six "Principal options for improving energy efficiency" and splits them into **technical measures** (e.g. silicon-based hull coating, propeller polishing; some applicable to new vessels, some to existing ones) and **operational measures** (e.g. trim optimization and speed reduction). Researches have further evaluated the saving potential of these methods by plotting shipping specific marginal abatement cost curves (MACC's) (Buhaug, 2009; Faber, Behrends, & Nelissen, 2011; Wang, Faber, Nelissen, Russell, & Amand, 2011). MACC's indicate how the marginal cost-effectiveness depends on the amount of emission being reduced, relative to a baseline. The curves help to identify technologies that can be used to reach a certain emission target in the most cost-effective way.

A study by Eide et al. (2009) evaluated over 50 available measures for energy efficiency in shipping and found several of them to be also cost-effective. For example, through weather routing, trim optimization, silicon-based hull coating, and other methods, bulk carriers and container ships could achieve emission reductions of 30% and 40% respectively. If these methods were combined with general speed reduction, the emission levels could further be halved.

In summary, the literature makes a strong case for making vessels as fuel efficient as possible, both from a business point-of-view as well as out of environmental concerns. Many fuel-saving technologies yield a positive net present value and will pay back over time. Yet, despite their savings potential, the implementation rate of cost-effective measures remains

low across the shipping industry. So why are technologies that lower CO_2 emissions, reduce operating costs, and pay for themselves not widely adopted and installed on every vessel?

In the following two chapters, I will examine the design of existing, integration-based systems for vessel performance optimization, and outline why the design of these systems is inappropriate for a large segment of the maritime industry. This step will clarify why a lightweight tool for vessel performance monitoring is needed.

2.2 Integration-Based Solutions for Vessel Performance Monitoring

In 2008, the Annex VI "Regulation for the prevention of air pollution from ships", published by the International Convention for the Prevention of Pollution from Ships (MARPOL), was revised, allowing ship owners to freely choose the methods to achieve an increase in operational efficiency, as long as they "are at least as effective in terms of emission reductions as that required by this Annex…" (International Maritime Organization, 2008). In order to comply with Annex VI and other regulations, it is therefore vital for ship owners and operators to be able to demonstrate the fuel consumption of their vessels across their entire operational profile, e.g. for different speeds, draft, and trim levels (Trodden, Murphy, Pazouki, & Sargeant, 2015).

A complete characterization could be achieved via repeatedly testing the fuel consumption in controlled sea-trials (Bertram, 2000). This traditional approach would provide high-fidelity data; however, it is prohibitively expensive and would need to be repeated as the vessel properties change over time (e.g. hull fouling or engine wear and tear). In comparison, data-driven performance monitoring tools are relatively inexpensive, allow for the assessment of vessel's base performance, and can further capture the dynamic effect of changes in trim, draft, hull condition, weather, and operating procedures (Hideyuki, 2011). Furthermore, these tools enable continuous condition monitoring, predictive maintenance and can therefore be used as powerful, fleet-wide asset management tools (Simon & Litt, 2011).

Davenport, Harris, and Morison (2010) provide a framework for the classification of different analytics approaches, based on their time frame ("are we looking at the past, present, or future?") and their degree of innovation ("are we working with known information or gaining new insights?"). Figure 1 shows six key questions that can be addressed with data analytics techniques. Past information describes traditional business reporting ("what happened?"). By applying rules of thumb, past data can be used to also

generate alerts about the present ("what is happening now?"). By extrapolating past patterns, one can formulate future forecasts ("what will happen?"). While these questions can reveal valuable information, they cannot tell why something has happened or how likely it is to happen. To produce these insights, different tools are needed: statistical modeling can reveal why and how things happened in the past; insight into the present comes in form of (real-time) recommendations that can suggest the next best course of action; predictive modeling and simulation can offer insights into the future. (Davenport et al., 2010)

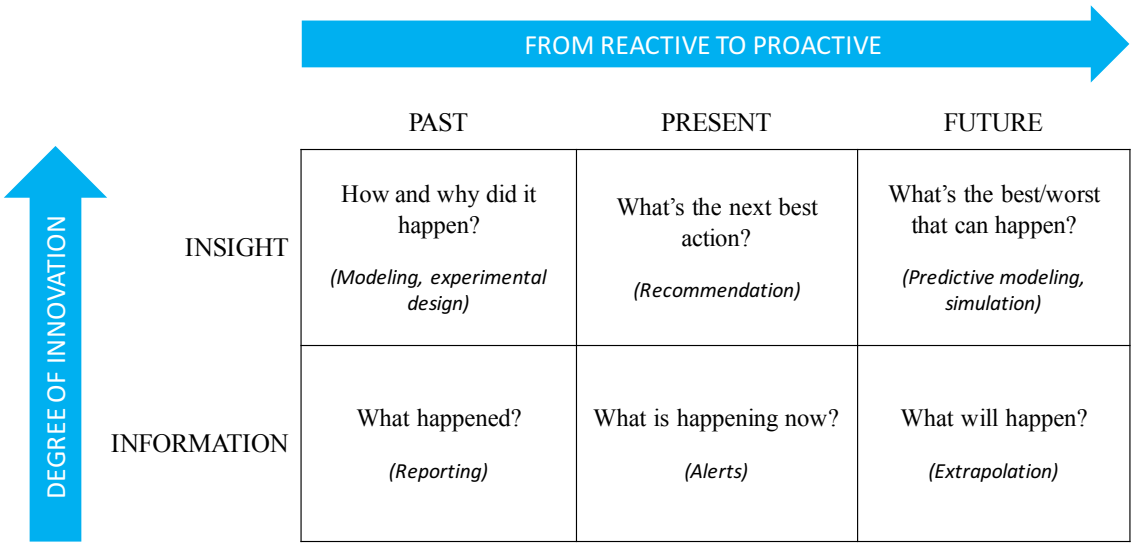


Figure 1. Degrees of analytics (Davenport et al., 2010): the matrix visualizes the time frame and degree of innovation of various analytics approaches. Most data-driven integration-based systems are located at the top row and offer thorough vessel insights. Traditional reliance on crew-generated noon reports is described by the bottom row and mainly answers "what happened?".

In a study on the fuel efficiency of a harbor tugboat, Trodden et al. (2015) combined data from a GPS unit with readings from the tugboat's onboard fuel flow meter and found the result to be in close agreement with the sea-trial experiments. Moreover, it became possible to validate that a recently installed "eco-button" did lead to an increase in fuel efficiency, as the vessel now out-performed the sea-trials. This past-insight example already hints at the underlying approach of most commercially available data-driven monitoring systems. Usually, data is collected by tapping into a ship's onboard automation and navigation systems and further combined with readings from additional sensors along the hull. The high level of integration enables the direct measurement of various quantities, such as propulsion power, fuel mass flow, and speed-through-water. Thus integration-based solutions offer unmatched accuracy in capturing data and provide thorough vessel performance insights. For example, they can answer "how and why" (past-insight) a vessel's

fuel consumption has been increasing over time, by modeling the gradual hull fouling and the resulting increase in water resistance. Furthermore, the tight onboard integration allows for data processing in real time. Integration-based solutions can therefore also show "what is the next best action" (present-insight) by providing the crew with recommendations on how to operate the vessel most efficiently. In the future, predictive modeling and simulations will further allow these high-end solutions to answer "what is the best/worst that can happen", by predicting for example the weather conditions that a vessel *will* face on its current route.

Coming back to the 2nd IMO GHG Study (Buhaug, 2009), where the "Principal options for improving energy efficiency" were split into technical and operational measures, data-driven performance monitoring solutions can help to improve both sides: For example, voyage optimization can improve operational performance by aiding the crew in choosing a fuel-efficient speed profile; on the technical side, unprecedented insight into a vessel's true performance profile at different speeds can help engineers to optimize the propulsion system (Trodden et al., 2015).

Eniram, a Finnish provider of data-driven vessel performance tools, presents the savings potential of an integration-based solution that enables dynamic trim optimization. A vessel's trim level defines its floating position in length direction, or whether the bow (front) or the aft (back) is deeper submerged in the water. Figure 2 visualizes the difference in trim level. The trim has significant impact on a ship's propulsion power requirement and fuel consumption. To increase hydrodynamic efficiency, the crew can adjust the trim by pumping water into ballast tanks along the hull. The optimal trim level will change depending on the cargo type, how and where the cargo is loaded on the ship, the sea state, and various other factors. Traditionally, the crew is using static trim tables to define and adjust the optimal trim level before leaving port. However, the changing conditions along the voyage lead to different optimal trim levels and thus require constant, dynamic trim adjustment, which is impossible with static tables. With a network of sensors, access to the onboard systems, and suitable algorithms, an integration-based system can constantly monitor the trim level and prompt the crew to make adjustments when needed. A Panamax-sized cruise vessel that is sailing 30-40cm off the optimum trim could save 700 tons of fuel over a period of 12 months. Assuming fuel prices of USD 600 / ton, this translates into ca. USD 420,000 annual savings in operating costs (Eniram, 2017).

Figure 2. Visualization of different trim levels: stern trim (top) vs. bow trim (bottom). A slight bow trim will generally decrease water resistance and increase hydrodynamic efficiency (MAN Diesel & Turbo, 2011).

While data-driven integration-based monitoring tools are cheaper than repeated, extensive sea-trials, the tight integration between a ship's onboard systems and additional sensors still makes these systems expensive. The project-based installation of multiple sensors along the hull of a vessel of up to 400 meters in length requires a team of technicians, and significant upfront investments from the ship owner. However, if a integration-based system offers insights that lead to significant savings, high setup costs are economically justified; the system is cost-effective.

In the following, I will discuss the reasons why integration-based solutions are nevertheless not widely adopted by the industry (Eide et al., 2009). Despite their savings potential, most systems are inherently subject to a phenomenon that is known as barriers to energy efficiency in past literature (Jaffe & Stavins, 1994).

2.3 Barriers to Energy Efficiency

Numerous empirical studies indicate that cost-effective measures are not being implemented in the shipping industry, despite their savings potential (Crist, 2009; Faber et al., 2009). This phenomenon is by no means unique to the shipping industry, but has been observed across industries and geographical regions. For example by Velthuijsen (1993) on Dutch firms, Harris, Anderson, and Shafron (2000) on Australian firms, Zilahy (2004) on Hungarian industrial companies, and Rohdin, Thollander, and Solding (2007) on Swedish foundry companies. The studies identified a range of barriers that cause a suboptimal level of implementation of cost-effective measures. The barriers are commonly defined as

mechanisms that prevent investment in technologies. The difference between the actual (lower) level of implementation and the potential, most cost-effective (upper) level of implementation, is thereby referred to as the **energy efficiency gap** (Jaffe & Stavins, 1994).

2.3.1 The Energy Efficiency Gap

Figure 3 illustrates the energy efficiency gap for "X", an example cost-effective measure for energy efficiency. In reality, X could be a system that assists the crew in setting the vessel speed for a just-in-time arrival at the destination port, thereby enabling an even speed profile across the voyage and reducing fuel consumption. Figure 3 assumes a current implementation level of 50%. Various reasons create a gap between this value and the potential, optimal level (100%). Note that the percentages serve simply an illustrative purpose and do not allow for conclusions such as "20% of the gap are always caused by non-market failures".

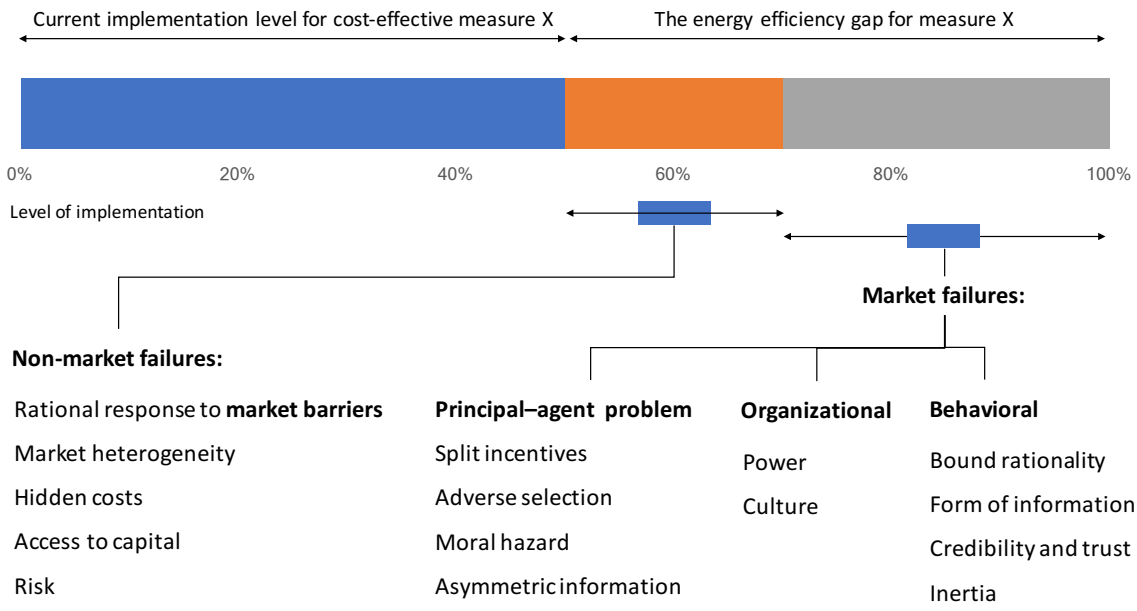


Figure 3. The energy efficiency gap for an example measure X, adopted from Sorrell et. All (2004): Principal-agent problems in the form of split incentives and asymmetric information drive the need for a lightweight system, espically on small and chartered vessels.

Part of the gap occurs due to the organizations' rational response when facing market barriers. Brown (2001) defines market barriers as obstacles that are not caused by market failures, but which nevertheless lead to a slow diffusion and implementation of energy-efficient measures. These non-market failures occur in situations in which an organization is behaving rationally, given the risk-adjusted rate of return on investment, unavoidable

hidden costs, and constraints on capital and energy (Sorrell, O'Malley, Schleich, & Scott, 2004). For example, in a study on Swedish foundry companies, Rohdin et al. (2007) found that limited access to capital was the largest market barrier, followed by concerns about a disruption in production and a lack of budget funding. In other words, if management refrains from implementing energy-efficient measures (even if they are cost-effective) because they suspect hidden costs, or because the organization lacks the required capital, no systematic market failure has occurred; the behavior is optimal from the organization's specific point-of-view. While the percentages in Figure 3 are only illustrative, it can be said that non-market failures usually only explain a small part of the energy-efficiency gap.

The bulk of the gap is typically the result of market failures that occur when the requirements for efficient or optimal allocation of resources are missing; it leads to incomplete markets, imperfect competition, and asymmetric information (Brown, 2001). While market failures can be rectified with appropriate policy intervention, market barriers (non-market failures) cannot (Sorrell et al., 2004; Thollander & Palm, 2012). Generally, all forms of barriers are intangible and the reasons behind them are often linked and follow a causal chain (Blumstein, Krieg, Schipper, & York, 1980). Some researchers conclude that is virtually impossible to find empirical evidence for what truly causes a lack of action (Weber, 1997). After this broad introduction on barrier to energy efficiency, I will focus more deeply on the specific barriers that inhibit the diffusion of energy efficiency measures in the shipping industry.

2.3.2 Marine Industry-Specific Barriers

While barriers generally exist across geographical regions and industries, Rohdin et al. (2007) found that they are actually different for each region and industry. Thus, research on barriers to energy efficiency in shipping must be both industry- and region-specific. The previously presented studies, however, were largely focused on industrial firms, not on the transport sector (let alone shipping). Rehmatulla and Smith (2015) noted that there is only scarce literature on shipping-specific barriers, mostly in the form of industry reports, e.g. by Faber et al. (2011) or Maddox Consulting (2012). These reports have weak (or no) scientific methodologies and utilize neither established frameworks nor theories.

Rehmatulla and Smith (2015) closed this research gap by conducting the first scientific and empirical study on the energy efficiency gap in shipping. They investigated the industry-specific barriers to energy efficiency and found **principal-agent problems** to be the primary reason why cost-effective measures are not implemented. Going back to Figure 3, we see

that principal-agent problems are defined as a form of market failure and linked to split incentives, adverse selection (also known as hidden knowledge), moral hazard, and asymmetric information.

Economists have long recognized the problems that arise whenever a principal delegates a task to an agent, who has some private information that is unknown to the principal. Already in 1955, (Marschak) noted that "by definition the agent has been selected for his specialized knowledge and the principal can never hope to completely check the agent's performance." According to the agency theory, the delegation of a task to an agent who has different objectives than the principal who delegates said task becomes a problem whenever the information about the agent is imperfect. In contrast, if the agent had different objectives than the principal, *but no private information*, the principal could impose a contract that perfectly controls the agent; such contract would align the agent's actions exactly with what the principal would do if the task was not delegated; all incentive issues would disappear. Thus, the two main ingredients of principal-agent problems are **conflicting objectives** and **asymmetric information.** (Laffont & Martimort, 2009) Agency theory has been used in numerous studies to explain some of the market failures that lead to barriers to energy efficiency (Graus & Worrell, 2008; Levinson & Niemann, 2004; Murtishaw & Sathaye, 2006; Vernon & Meier, 2012). Scholars of the agency theory aim to define the most efficient form of contracts for a broad range of agency relationships (Eisenhardt, 1989; Ross, 1973). In the context of the maritime industry, the ship owner and the charterer enter an agency relationship, where the charterer (principal) hires the ship owner (agent) to transport some goods from place A to B (Murtishaw & Sathaye, 2006).

In order to evaluate the effect of principal-agent problems on the energy efficiency in shipping, one has to identify the situations in which these problems occur (Vernon & Meier, 2012). In the following, I will take a closer look at the contracts for the transportation of goods that are commonly used in the shipping industry, and show how the contractual design leads to principal-agent problems that prevents the wide diffusion of energy efficient technologies.

2.3.3 Owner-Operated vs. Chartered Vessels

In shipping, the contracts between the ship owner, the operators (the crew), and the charterer are commonly referred to as charterparties, or charterparty agreements. There exist two major types of contracts for the transportation of goods: voyage charter and **time charter**. The charterparties define the division of responsibility for capital and operating costs

(including fuel costs) between the ship owner and the charterer. Consequently, both parties can have conflicting interests and might try to reduce their share of the costs at different phases throughout the voyage. (Rehmatulla & Smith, 2015)

In addition to the voyage and time charter, there exist also vessels that are not chartered, but owner-operated. In this case, a single entity owns, runs, and reaps the economic benefits of the vessel. It can be argued that an agency relationship also exists between the ship owner and the crew. However, compared to relationship between the owner and a charterer, the resulting principal-agent problem is low. For the purpose of this study, we will assume that the objectives of the owner and the crew are aligned. Note also that Rehmatulla and Smith (2015) did not discuss owner-operated vessels; the researches also largely omitted the voyage charter, and focused on the time charter instead. Within the context of this thesis, the focus will be on owner-operated and time-chartered vessels.

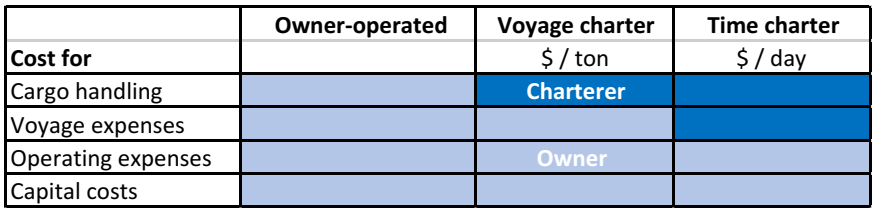


Figure 4. Cost division in shipping, adopted from Rehmatulla and Smith (2015): On (time-) chartered vessels, the ship owner determines a vessel's operational and technological energy efficiency, while the charter has to cover the actual fuel bill of a voyage.

Figure 4 visualizes the cost division for owner-operated and chartered vessels. It indicates that the contractual framework commonly used by the shipping industry, more specifically the charterparty design, might create market barriers that prevent the implementation of cost-effective measures for energy efficiency on chartered vessels (Rehmatulla & Smith, 2015). For a time charter, the ship owner (agent) provides the vessel and thus determines its level of technological energy efficiency. Furthermore, the ship owner often also hires a crew to operate the vessel and is therefore (indirectly) responsible for the ship's operational energy efficiency. As can be seen from Figure 4, the charter (principal) is nevertheless responsible for the costs associated with the level of energy efficiency, as he/she has to pay the actual fuel bill (Agnolucci, Smith, & Rehmatulla, 2014). Because of this cost division, neither the ship owner nor the charterer might have an incentive to invest into expensive fuel saving technology.

A commonly used analogy to this phenomenon is the relationship between a landlord (agent) and a tenant (principal): if the tenant covers the utility costs (e.g. heating), he/she is interested in cost reducing measures (e.g. better insulated windows). However, there is no incentive for the tenant to invest own capital into a property that belongs to the landlord. Meanwhile, the landlord is responsible for renovations, however all future savings from a more energy efficiency property would be retained by the tenant. However, if the landlord invests into renovations that lead to savings for the tenant, the landlord can justify a higher rent and recuperate parts of the investment costs.

Consequently, the magnitude of the principal-agent problem in the shipping industry is directly related to how well the charter rates reflect the vessel's energy efficiency (Rehmatulla & Smith, 2015). If the ship owner invests into a higher energy efficiency of his/her vessel, and if the charterer experiences lower voyage costs as a direct result of these investments, the question becomes to what extent the ship owner can recuperate the investment costs through higher charter rates from the charterer. Agnolucci et al. (2014) investigated this question in a case study on drybulk panamax ships and found that 40% of the fuel savings could be recuperated by the ship owner through higher charter rates.

In contrast to the time charter, owner-operated vessels are not subject to complex agency relationships and the resulting principal-agent problems are largely removed. The ship owner is in charge of the vessel, has to raise the investment capital for energy-efficient measures, but can also retain all future savings that result from lowered fuel consumption. An energy-efficient measure that is also cost-effective (positive net present value) is therefore more likely to be implemented on owner-operated vessels.

In light of the findings by Rehmatulla and Smith (2015) and Agnolucci et al. (2014), it becomes clear why the implementation rate of measures for energy efficiency remains low across the shipping industry. Although Eide et al. (2009) identified more than 50 cost-effective technical and operational measures, principal-agent problems in the form of split incentives and information asymmetry between the ship owner and the charterer prevent a higher rate of diffusion on (time) chartered vessels. Many energy efficient measures are only cost-efficient on owner-operated vessels; on chartered vessels, the ship owner cannot recoup sufficient fuel savings from the charterer to achieve a positive return on investment. Existing high-end systems appear to be primarily designed for owner-operated vessels and do not match the specific needs and limitations of chartered vessels.

In the following, I will discuss the implications of these findings for the manufacturers of vessel performance systems. Specifically, I will outline what is needed to overcome the

barriers to energy efficiency on chartered vessels: a lightweight system for charterparty monitoring.

2.3.4 Implications for Manufacturers

In their study, Rehmatulla and Smith (2015) concluded that the implementation rate of many energy efficient measures does not correspond to their high energy saving potential and positive net present value. Consequently, the development of increasingly powerful (and costly) tools will not guarantee commercial success across all market segments.

For chartered-vessels, principal-agent problems will prevent a widespread diffusion of costly high-end systems. The ship owner, responsible for the technological and operational efficiency of the vessel, is unlikely to recover a sufficient part of the investment costs through higher charter rates to make the investment worthwhile. Meanwhile, the charterer, who would benefit from the increased efficiency, will not invest into expensive technology either, as the vessel belongs to a different party. Nevertheless, the charterer can still be a promising customer for providers of vessel performance tools, if they manage to adapt their products to the complex agency relationship.

According to Antola, Solonen, and Staboulis (2017), a **lightweight system** is needed that can monitor vessel performance, without being subject to the barriers identified by Rehmatulla and Smith (2015). Such system has to be significantly cheaper than the established high-end systems; ideally, it would be offered as a service, to remove any need for upfront investments. The system would be marketed to the charterer as an independent monitoring tool that would enable the tracking of vessel and crew performance. The charterer could identify deviations from the levels that were agreed in the charterparty, demand compensation from the ship-owner, or renegotiate the terms for the next voyage. For example, the charter would be able to identify poor operational practices that lead to an unreasonably high fuel consumption (and fuel bill). While high-end systems achieve savings for the ship owner by directly optimizing the technical and operational efficiency of the vessel, a lightweight system would *indirectly* increase efficiency by reducing the information asymmetry between the charterer and the ship-owner. Coming back to Marschak (1955), the principal (charterer) could finally hope to check (at least partly) the agent's performance. The two parties will still have conflicting interests, however for the first time, the charter would be able to obtain some of the ship-owner's private information and reduce transaction costs from the principal-agent relationship.

In practice, designing a lightweight system that is not subject to barriers to energy efficiency is a complex task. Many quantities that are essential for evaluating vessel performance (e.g. speed-through-water, RPM, fuel flow) cannot be directly measured without access to a ship's onboard system; thus the costly integration of high-end systems. Bialystocki and Konovessis (2016) have presented a statistical approach in which a vessel's speed-fuel curve can be estimated from crew-reported data. While this approach can serve as a decision-making tool for stakeholders, any approach that relies entirely on human input will always be subject to reporting errors. Moreover, it questionable whether it will satisfy the strict regulations of Annex VI (International Maritime Organization, 2008).

Antola, Solonen, and Staboulis (2017) claim that some performance variables can be obtained indirectly by using virtual sensing techniques. Virtual sensors can drastically reduce hardware costs by replacing physical sensors with mathematical models that allow the inference of some desired variables based on their dependencies on a set of influential variables (Fortuna et al., 2007). Virtual sensors are therefore the key for the development of a lightweight system for monitoring vessel performance that is not subject to the barriers identified by Rehmatulla and Smith (2015).

In the following, I will present the design procedure for virtual sensors in the context of industrial process control. Based on this framework, I will then present the novel approach of Antola, Solonen, and Staboulis (2017) in Chapter 2.5.

2.4 Theoretical Design Procedure for Virtual Sensors

As stated previously, virtual sensors are mathematical models that allow the inference of some desired variables based on their dependencies on a set of influential variables (Fortuna et al., 2007). The theoretical and practical aspects of virtual sensors have been extensively researched. Refer for example to Haykin (1999) for an early work on neural networks, or Fortuna et al. (2001) and Sinha and Gupta (2000) on the theory and application of soft computing. However, to my knowledge, no scientific study exists on the development of data-driven virtual sensors in the context of maritime shipping.

In order to embed the novel approach of Antola, Solonen, and Staboulis (2017) within a framework of established literature, one has to study related fields and industries instead. In the following, I will discuss the design procedure of virtual sensors in industrial environments, such as refineries, power plants, and industrial pollution monitoring, based on the work of Fortuna et al. (2007). While a container ship is certainly no industrial plant, both are constantly seeking ways to improve (production) efficiency, but also face

increasingly tight environmental regulations and pollution restrictions. Monitoring a large set of process variables is the key to ensure (production) efficiency and compliance with environmental laws. The implementation of a plant-wide monitoring with physical on-line measurement devices is, however, a highly costly venture; much like the integration-based monitoring of vessel performance.

An alternative to these expensive devices comes in the form of virtual sensors: based on experimental data, it becomes possible to formulate mathematical models that map industrial processes via system identification procedures; these models can greatly reduce the need for physical measurement devices and ensure tight process control. Virtual sensors are used for a variety of tasks, such as real-time predictions, back-up systems for measuring devices, and sensor validation. (Fortuna et al., 2007) Within the context of this thesis, virtual sensing techniques will be used to eliminate hardware requirements, or more specifically, to overcome a lack of hardware (physical fuel flow meter). Figure 5 illustrates the typical design procedure of a data-driven virtual sensor for industrial process control.

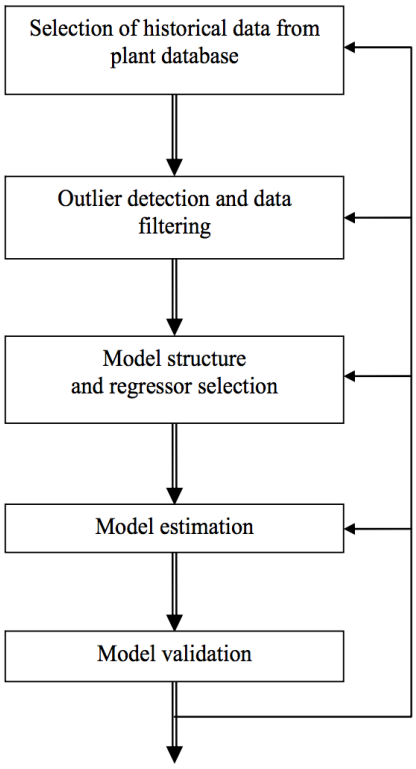


Figure 5. The design procedure of a virtual sensor for industrial process control Fortuna et al. (2007): the five stages of data selection, outlier detection and filtering, model and variable selection, estimation, and validation will provide the rough structure for the fuel flow model as presented in Chapter 3.

Generally, a model can never be better than the data used for its predictions. The first step towards developing a soft sensor is therefore the **data collection**. Most industrial companies have amassed vast historical records on their processes that can be used to identify relevant features (Albazzaz & Wang, 2006). However, the sheer volume of (potentially unstructured) data can also hinder the identification of dependencies among variables (Flynn, Ritchie, & Cregan, 2005). While data mining is a crucial step in the soft sensor development, it is also subject to many obstacles: noise, low accuracy, high dimensionality, redundancy, incorrect values, and non-uniformity in the sampling can all negatively affect the collection of suitable model input data (Fortuna et al., 2007). The lightweight model for fuel flow estimation faces noisy weather forecasts and crew-generated reports that are inherently prone to human errors.

The second step is therefore **outlier detection** and **data filtering**. Hardware failure or transmission problems can create outliers that can potentially compromise the data. Different techniques for outlier detection and removal exist, for example as presented by Chiang, Pell, and Seasholtz (2003) in the context of industrial process control. After outliers have been detected, the data might still require pre-filtering before it can be used to train a model. Refer to Guidorzi (2003) for data filtering techniques in the context of multivariable systems and linear models (as used in the practical part of this Master's thesis).

Once the data has been collected and pre-processed, the third step is the **selection of variables and model structures**. Depending on the level of *a priori* knowledge of the real system, different procedures for virtual sensor development have been proposed: physical modeling (also known as first principles approach, or white-box models), as well as gray- or black-box identification approaches in the form of multivariate statistics or artificial intelligence modeling, such as fuzzy logic or neural networks (Nørgård, Ravn, Poulsen, & Hansen, 2000). The high complexity and dynamics of industrial processes make physical models often unsuitable, as critical parameters are unknown. Instead, empirical data-driven gray- and black-box models are used to provide real-time estimates of process quantities based on their correlation with other system variables. (Fortuna et al., 2007) Especially the gray-box approach can lead to highly accurate models as it combines all available information: known physical properties of a system, as well as historical data. For example, Zahedi, Elkamel, Lohi, Jahanmiri, and Rahimpor (2005) found that a hybrid gray-box model for the simulation of chemical reactions outperformed its purely mechanistic or neural network counterparts. The lightweight model by Antola, Solonen, and Staboulis (2017), as

well as the model based on Equation (10), can both be classified as grey-box models as they combine some prior knowledge with empirical data.

Therefore, a set of explanatory variables (also known as independent variables or predictors) has to be chosen that influences the model output. Among the most popular methods are the graphical inspection of dispersion plots and correlation analysis. See for example Warne, Prasad, Rezvani, and Maguire (2004) for a review of variable selection techniques for linear and nonlinear models. In the practical part of this Master's thesis, literature on hydrodynamic laws (Bertram, 2000; MAN Diesel & Turbo, 2011) will be used to identify a set of explanatory variables that describe the underlying system well.

A widely reported problem during the identification stage is the collinearity of variables; a problem that occurs frequently for variables measured from industrial processes. It can be addressed by techniques, such as partial least squares regression (PLS). While (multi-) collinearity amongst the explanatory variables will be a topic in the practical part of this thesis, we will also see that no measures are necessary to mitigate the problem. Thus, I will not discuss this issue further; for more information, refer to Baffi, Martin, and Morris (1999) for an overview of linear and nonlinear PLS algorithms.

Once the model input and output variables are selected, the fourth step is the **model estimation** (also known as model identification). The goal is to obtain an estimate for the output variable(s) based on predictors (Fortuna et al., 2007). For the practical part in Chapter 3, I will approximate a nonlinear system by using a linear, polynomial model (Chen & Billings, 1989). For an overview of nonlinear black-box modeling techniques see for example Juditsky et al. (1995). All modeling techniques are sensitive to the size of the training data set and Yan, Shao, and Wang (2004) have presented some techniques to mitigate this problem. However, the study presented in Chapter 3 is based on a large data set, thus I will not discuss the effects of insufficient training data further.

The final step involves the **model validation**. This topic has not been solved in a definite way. For linear models, the standard approach is to compute the autocorrelation function of the residuals and the cross-correlation functions between the residuals and the input variables over a set of (unseen) testing data (Söderström & Stoica, 1988). In Chapter 3, we will see how the model validation requires reference data from an integration-based system.

In summary, this chapter covered the design procedure of virtual sensors: data collection and filtering, variable and model selection, model identification, and validation. Given the lack of scientific literature on the application of virtual sensors in the maritime

industry, the context of industrial process control was chosen to provide a theoretical framework for further discussions. In the following, I will show how a virtual sensor is designed in practice by presenting the lightweight model as described by Antola, Solonen, and Staboulis (2017).

2.5 The Lightweight System by Antola et. al. (2017)

The Finnish company Eniram has productized the model of Antola, Solonen, and Staboulis (2017) and offers a lightweight system for vessel performance monitoring. It is sold as a service (no upfront costs) and marketed primarily at small and chartered vessels. The hardware consists of a single shoebox-sized unit that features a GPS transponder, an accelerometer, and a satellite transponder. The installation can be carried out by the crew without the need for a trained technician. Note that the system is already installed on multiple vessels. My research methodology, presented in Chapter 3, relies on availability of real-life vessel data from a lightweight system.

In contrast to integration-based systems, the lightweight system neither has access to a vessel's onboard systems, nor can it rely on a network of sensors. The leading idea behind the lightweight system is that despite the lack of onboard integration, useful information on the vessel performance can be obtained *indirectly* by utilizing virtual sensing techniques. According to Antola, Solonen, and Staboulis (2017), "the model combines the vessel speed with meteorological forecasts and a propulsion power model to estimate instantaneous fuel flow rates based on the daily total fuel consumption reported by the crew."

While the onboard integration of integration-based systems provides access to numerous measured quantities, the lack of onboard integration creates a short list of **input variables** for the lightweight system:

1. **Vessel speed:** the GPS unit provides high-frequency data on the vessel speed-over-ground (SOG) with a five-minute sampling period. Numerous factors, such as sky blockage, atmospheric conditions, and receiver quality make it difficult to determine the accuracy of a GPS signal (Trodden et al., 2015). Most modern GPS units show a horizontal accuracy of at least 3.5 meters (Federal Aviation Administration, 2014); a distance that will not affect model accuracy in the context of shipping so that the GPS signal can be regarded reliable.
2. **Meteorological data:** is obtained from a 3rd-party provider and based on the forecasts of the National Oceanic and Atmospheric Administration (NOAA), an agency of the U.S. Department of Commerce. The high-frequency meteorological data contains

forecast on e.g. ocean currents, wind strength, and wave heights. Note that these forecasts are inherently prone to errors; nevertheless, they are (among) the best available estimates of meteorological values.

3. **Crew-generated noon reports on changes in tank levels:** roughly every 24 hours, the crew is required to submit *noon reports* (usually submitted around noon). These manually filled-out Excel sheets contain - among others - information on how the tank levels have changed since the previous report. The reports often contain crude errors and are sometimes entirely missing. Even if the crew is willing to carefully submit the reports, they might encounter technical difficulties. It is not uncommon for older vessels to lack an automated system for monitoring tank levels. The crew is then required to literally take a long stick, insert it into the tank and estimate the changes to the previous day. Despite their shortcomings the crew-generated noon reports are crucial for obtain fuel flow rates with a virtual sensor and serve as a response variable.

The desired **output variable** of the lightweight system is *instantaneous fuel flow* (in HFO tons per day). It is important to understand that a vessel's fuel consumption can be expressed in different ways. The noon reports state the total consumption over a certain reporting period. This low-frequency value gives no information on any fluctuation in consumption throughout the period, e.g. due to variance in speed. A related value is the gas consumption per 100 kilometers used in automobiles: whether the driver kept an even speed or was first speeding and then going slowly cannot be seen from a single consumption value. In contrast, a physical meter provides measured data on the instantaneous fuel flow. A virtual fuel flow sensor would therefore aim to mimic its physical counterpart as closely as possible. Hence, the desired output variable of the lightweight model is instantaneous fuel flow, a high-frequency measure.

Antola, Solonen, and Staboulis (2017) evaluated their model's performance by compared the modeled fuel flow rates from a lightweight system against reference data from an integration-based system; they summarized their findings as following:

> "The results indicate that the proposed technique can yield speed-fuel models that correlate well with the high-fidelity reference data. Moreover, the method gives a significantly more accurate view on the vessel's performance compared to an analysis that solely rely on the crew-reported aggregates. […] On the other hand,

> […] the temporal resolution of the fuel flow estimate is considerably lower as expected." - Antola, Solonen, and Staboulis (2017)

Despite these promising findings, Antola, Solonen, and Staboulis (2017) also discuss three challenges that come with their method: firstly, the lack of physically measured reference values in combination with noisy high-frequency data sources (mainly weather forecasts) could lead to error aggregation and severely impact **model accuracy**. Secondly, the small number of available input variables can cause **systematic modeling errors**. Thirdly, uncertainty is caused by the **reliance on human input** as the model is fitted on daily changes in tank levels that have to be manually observed and reported by the crew (who does not necessary has an incentive to support a performance monitoring tool).

In summary, this chapter outlined the high-level design, data sources, performance, and potential shortcomings of the lightweight model, as presented by Antola, Solonen, and Staboulis (2017).

2.6 Summary of Theoretical Framework

All in all, Chapter 2 set the theoretical framework for this Master's thesis. Economic, ecological, and regulatory requirements lead ship owners and operators to seek ways to improve fuel efficiency of commercial vessels (Buhaug, 2009). While over 50 methods exist that are energy- and cost-efficient (Eide et al., 2009), principal-agent problems prevent their industry-wide implementation (Rehmatulla & Smith, 2015). Especially on small and chartered vessels, expensive integration-based performance monitoring tools (Trodden et al., 2015) are unfeasible. Alternative approaches that rely solely on crew-reported data are inexpensive, but fail to offer sufficient insight and accuracy to comply with regulatory standards (Bialystocki & Konovessis, 2016). A novel approach based on virtual sensing techniques (Fortuna et al., 2007) has been proposed that can overcome the barriers to energy efficiency and offer some of the insights that require onboard integration in most commercially available solutions (Antola, Solonen, & Staboulis, 2017). With this theoretical framework, I will proceed to the research methodology of my study.

3 Research Methodology

In this chapter, I will present my research methodology and elaborate on the four-stage implementation process as outlined in Chapter 1.2. To recap, my research questions are defined as:

1. How to formulate a virtual sensor for time-dependent momentary fuel flow estimation on commercial vessels?
2. How well does the model perform in comparison to integration-based systems?

These questions are addressed by conducting a **quantitative in-depth case study** that aims to validate the approach by Antola, Solonen, and Staboulis (2017) exemplarily. Real-life data from a single case vessel is collected over a two-month period from May to June 2017. The case vessel was chosen based on the unique situation that it was equipped simultaneously with a lightweight system and an integration-based system. This setup not only provides sufficient lightweight data for training a predictive model, it also offers reference data for assessing the model performance, and for quantifying the effect of human reporting errors. Based on the information presented by Antola, Solonen, and Staboulis (2017), I formulate a (simplified) version of the underlying lightweight model from publically available tools. Specifically, I will use a Jupyter notebook and various Python packages for data analysis and visualization, such as Pandas, Numpy, Matplotlib, Seaborn, and StatsOLS.

As discussed in Chapter 2.4, the prior knowledge of a real system affects which model type is favoured: white-, grey-, or black-box modelling. While white-box models exist that could in theory calculate the fuel flow (Bertram, 2000), these mechanistic models require the availability of many quantities that are unknown in practice; especially without onboard integration. Therefore, an empirical, data-driven grey-box model will be used to estimate instantaneous fuel flow rates based on their dependencies on other system variables.

As a first step, I will discuss how the lightweight model of Antola, Solonen, and Staboulis (2017) combines physical laws on marine hydrodynamics with statistical modelling to approximate fuel flow rates via a parameter estimation problem. This step will provide the necessary theoretical framework for the variable selection and model design in Chapters 3.2 and 3.3.

3.1 From Marine Hydrodynamics to an Explicit Optimization Problem

In this chapter, I will show how a vessel's fuel flow can be approximated by a set of resistance parameters that influence the vessel's need for propulsion power. In general, it makes sense to express a vessel's fuel flow Φ at time t as a function of the propulsion power P_t and a set of (typically estimated) parameters α (MAN Diesel & Turbo, 2011). In particular, the fuel flow is given by

$$\Phi_t = g(P_t, \alpha) = \alpha_1 + \alpha_2 P_t. \quad (1)$$

The key to evaluating Equation (1) is to define a suitable propulsion power model that can quantify P_t. A ship's propulsion power requirement depends primarily on the vessel's speed; hence we need to first define a speed variable.

For the lightweight model, an artificial proxy of the vessel's speed-through-water (STW) is formed by combining speed-over-ground (SOG via GPS) with forecasts on ocean currents. For example, if a vessel has a STW of 15 knots but sails *against* a two knots current, its SOG will only be 13 knots. Likewise, if the vessel sails *with* a two knots current, its SOG will be 17 knots. Thus, the stronger the current, the less accurate it would be to utilize an unadjusted SOG to approximate the STW. The lightweight system refers to this adjusted speed value as **forecast-STW** (since it reflects the SOG adjusted for the ocean current forecasts). The relationship between the forecast-STW v_t, the SOG u_t, and the water current velocity u_t^w is express as

$$v_t = u_t + u_t^w. \quad (2)$$

Note that the water current u_t^w can be positive or negative, depending its relative direction to the vessel. Further note that in most integration-based systems, the vessel speed is obtained directly from the onboard navigation system and calibrated with the help of a broad range of (measured) variables and modeling techniques (Antola, Solonen, & Pyörre, 2017). The result is the most accurate expression of a vessel's actual speed-trough-water, referred to as **reference-STW** within the context of this thesis. In Chapter 4.3, I compare the accuracy of the forecast-STW against the reference-STW.

Once a speed proxy is found, a simplified version of a more general propulsion power model is formulated. It combines the forecast-STW with meteorological data, such as wind speed and direction, and wave heights. More precise, the model quantifies the time-dependent propulsion power P_t consumed by the vessel when moving at forecast-STW v_t while facing the wind force u_t^a. Additionally, a multidimensional resistance parameter θ is used to account for the object geometry and the properties of the media. By multiplying the resistance coefficient R with the forecast-STW v_t we obtain the propulsion power P_t (Antola, Solonen, & Staboulis, 2017). In particular, the propulsion power is given by

$$P_t = R(v_t, u_t^a, \theta)\, v_t. \tag{3}$$

In Equation (3), the forecast-STW v_t is derived based on Equation (2), while the wind force u_t^a is taken from meteorological databases. However, with θ unknown, the propulsion power P_t cannot be evaluated. Antola, Solonen, and Staboulis (2017) claim that this limitation can be overcome by utilizing the crew-generated noon reports.

The crew onboard a ship is reporting changes in the tank levels roughly every 24 hours. This value can be regarded as the integral of the instantaneous fuel flow rates over the respective reporting period ΔT. By summing the instantaneous fuel flow Φ_t over ΔT and plugging Equation (3) in (1), we obtain a formula for the **total fuel consumption $\Phi_{\Delta T}$ over time period ΔT** in the form of

$$\Phi_{\Delta \mathrm{T}} = \int_{\Delta \mathrm{T}} \Phi_t \, dt = \int_{\Delta \mathrm{T}} g(P_t, \alpha)\, dt = \int_{\Delta \mathrm{T}} g(R(v_t,\, u_t^a, \theta)\, v_t, \alpha)\, dt. \tag{4}$$

Equation (4) presents us with the numerical problem to estimate the parameters α and θ. In contrast to Equation (3), the crew is providing a value for $\Phi_{\Delta T}$ in (4). Hence, the equation can be evaluated and the parameters α and θ can be estimated. Antola, Solonen, and Staboulis (2017) have provided a worked-out example of this parameter estimation problem. For simplicity, this example assumes a vessel's fuel flow to be only dependent on the water- and wind-induced resistances; the real lightweight system (e.g. as installed on the case vessel) uses a larger set of explanatory variables. As an explicit form, Equation (3) becomes

$$P_t = \theta_1 v_t^3 + \theta_2 |u_t^a - u_t|^2 v_t \cos \gamma_t. \tag{5}$$

As in Equation (2), v_t describes the forecast-STW, while the wind speed is given by u_t^a. Furthermore, γ_t denotes the angle between the relative wind (u_t^a - u_t) and the direction in which the ship is heading. The resistance is modeled by θ_1 and θ_2. Hence, the propulsion power is described as a weighted sum of the cubed STW and the weighted wind effect.

By plugging Equation (5) in (4), we express the total fuel consumption $\Phi_{\Delta T}$ over time period ΔT in terms of the propulsion power and the constants α_1 and α_2 in the form of

$$\Phi_{\Delta T} = \Delta T \alpha_1 + \alpha_2 \theta_1 \int_{\Delta T} v_t^3 \, dt + \alpha_2 \theta_2 \int_{\Delta T} |u_t^a - u_t|^2 v_t \cos \gamma_t \, dt. \tag{6}$$

crew-reported consumption — water resistance — wind resistance

All the summands in Equation (6) are available, either from the GPS unit, the meteorological forecasts, or the noon reports. While the multipliers $\alpha_1, \alpha_2, \theta_1, \theta_2$ cannot be estimated individually, they can be combined to weight parameters

$$\beta_1 = \alpha_1 \Delta T, \quad \beta_2 = \alpha_2 \theta_1, \quad \text{and} \quad \beta_3 = \alpha_2 \theta_2. \tag{7}$$

Given a sufficient number of noon reports, a model can be trained to estimate these weight parameters. Note that the example in Equation (6) assumed the fuel flow to be only dependent on two variables. While this is sufficient to illustrate the modeling approach, the actual system is more complex and requires a larger set of explanatory variables to approximate fuel flow sufficiently accurately. In the following, I will discuss the variable selection for this study.

3.2 Variable Selection

To recap, it makes sense to express fuel flow as a function of propulsion power (Equation (1)). For a ship to move, it first has to overcome resistance, the force working against the propulsion. The resistance is mainly driven by a vessel's speed, water displacement, and hull form (MAN Diesel & Turbo, 2011). Hence, we can ask what factors increase the resistance and the need for propulsion power? Based on Bertram (2000) and Antola, Solonen, and Staboulis (2017), I have chosen four sources of resistance that mainly affect a vessel's

propulsion power requirement and fuel consumption: frictional (water) resistance, draft, wave height, and wind force.

Frictional resistance

The frictional resistance describes the water-induced resistance. It accounts for 70% to 90% of the total resistance and grows quadratically with a vessel's speed (MAN Diesel & Turbo, 2011). Anybody who has attempted to run through a swimming pool can confirm that it becomes increasingly harder the faster one moves. In order to approximate the effect on fuel consumption, we need to express the power that is lost (or additionally required) due to the water resistance. This **water-power-loss** is approximated by STW^3, the cubed vessel speed (MAN Diesel & Turbo, 2011). For simplicity, it is denoted by: $\boldsymbol{X_1 = water}$.

Draft-induced resistance

The second source of resistance is a vessel's draft, which describes how much of the ship's hull is submerged and which affects the water displacement: the heavier the vessel, the deeper it lays in the water and the higher the frictional resistance becomes. Especially for containerships, the draft can vary by several meters depending on whether all containers are loaded with stones or pillows. Since the frictional resistance accounts for a large portion of the overall power loss, it is important to include a variable that captures the dynamic nature of this resistance. Without onboard integration, we have to rely on the crew-reported draft. However, we can mitigate potential reporting errors, by including a mean vessel draft (provided by the ship owner). The **draft-power-loss** is approximated by $\left(draft_{crew\ reported} - draft_{mean}\right) \times STW^3$ (MAN Diesel & Turbo, 2011). For simplicity, it is denoted by: $\boldsymbol{X_2 = draft}$.

Wave-induced resistance

The wave-induced resistance captures the energy loss due to waves. At low speeds, the wave-induced resistance is proportional to the squared vessel speed, however it increases much faster at higher speeds (MAN Diesel & Turbo, 2011). Meteorological forecast on the wave heights are combined with the vessel speed to describe the **wave-power-loss**, denoted by: $\boldsymbol{X_3 = waves}$.

Wind-induced resistance

The wind-induced resistance captures the energy loss (or gain) due to wind. In calm water, the wind resistance is roughly proportional to the squared vessel speed. It is dependent on the cross-sectional area of the ship above the water and the relative angle of the wind to the ship (Equation (6)). The wind accounts for 2% to 10% of the total resistance. In full head wind, the extra resistance will be high; if the wind blows from the back, it will actually reduce the power requirement and take a negative value. The **wind-power-loss** is denoted by: $\boldsymbol{X_4 = wind}$.

It would go beyond the scope of this business-oriented Master's thesis to discuss the exact physical laws from which the four predictor variables are derived. Bertram (2000) provides an exhaustive work on practical ship hydrodynamics; for an easier read refer to MAN Diesel & Turbo (2011). The level of information presented in this chapter is sufficient so that a customer who ordered a validation study can understand how the four predictors are chosen and how the physical phenomena (e.g. waves and wind) are approximated in the lightweight model. Given access to reliable vessel speed data, a customer could acquire the remaning values from a provider of meteorological forecasts (e.g. NOAA) and replicate my research.

Before we move on, note that the literature on ship hydrodynamics (Bertram, 2000) supports the idea that all four predictors are dependent on the vessel speed. In Chapter 5, we will discuss the implications of this setup on the regression results, especially regarding multicollinearity. In conclusion, the four predictors used in this study are defined as

$$X_1 = water, \quad X_2 = draft, \quad X_3 = waves, \quad X_4 = wind,$$

with the response variable being $Y = fuel\ flow$, denoted by Φ.

3.3 Model Design

After discussing the theoretical framework by Antola, Solonen, and Staboulis (2017) and outlining the variable selection process, we proceed to the first stage of my research method's implementation process: fit a model for weight parameter estimation onto low-frequency crew reports and onto aggregated high-frequency GPS and meteorological data. With $p = 4$ predictors, we need to predict a quantitative response Y based on multiple

independent variables $X_1, X_2, \dots, X_p$. Hence, a **multiple linear regression model** will be used in the form of:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon. \quad (8)$$

In general, X_j denotes the *j*th predictor and β_j quantifies the relation between that variable and the response *Y*. In Equation (8), $\beta_0, \beta_1, \dots, \beta_p$ are unknown constants that represent the intercept and slope terms in the linear model. Together, $\beta_0, \beta_1, \dots, \beta_p$are known as the model coefficients or weight parameters. We interpret β_j as the average effect of a unit increase in X_j on *Y*, holding all other predictors constant. (James, Witten, Hastie, & Tibshirani, 2013) Following the general form of Equation (8), we can formulate a relationship between the total fuel flow $\Phi_{\Delta T}$ over a known time period ΔT and aggregates of the predictors (water, draft, waves, wind) in the form of

$$\Phi_{\Delta T} = \beta_0 + \beta_1 \int_{\Delta T} water_t dt + \beta_2 \int_{\Delta T} draft_t dt + \beta_3 \int_{\Delta T} waves_t dt + \beta_4 \int_{\Delta T} wind_t dt + \epsilon. \quad (9)$$

If ΔT described a 24-hour period, we would receive only one value for the total fuel flow $\Phi_{\Delta T}$ from the low-frequency noon reports. However, the high-frequency data sources (GPS, meteorological forecasts) would provide over 300 data points (based on the five-minute sampling frequency of the GPS unit). In order to match the different sampling periods, we have to aggregate the high-frequency data to the sampling periods of the low-frequency noon reports. For example, in Equation (9), $\int_{\Delta T} water_t \, dt$ represents one single value for the *aggregated* or *time-averaged* frictional resistance over period ΔT. It is of utmost importance to understand the difference between the aggregated and unaggregated values: both stem from the same source and time period, but have different sampling periods.

Based on Equation (9), we obtain estimates for the weight parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ by fitting a model in the form of

$$\Phi_{\Delta T} \approx \beta_0 + \beta_1 \int_{\Delta T} water_t dt + \beta_2 \int_{\Delta T} draft_t dt + \beta_3 \int_{\Delta T} waves_t dt + \beta_4 \int_{\Delta T} wind_t dt. \quad (10)$$

The weight parameter estimation is done using a **least squares** approach. In general, we choose $\beta_0, \beta_1, \dots, \beta_p$ to minimize the sum of squared residuals

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2, \quad (11)$$

where y_i and $x_{i1}, \dots, x_{ip}$ are observed and $\hat{y}_i$ denotes the predicted value of the response. The values $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ that minimize Equation (11) are the multiple least squares regression coefficient estimates or weight parameters (James et al., 2013).

By applying the least squares approach to Equation (10), we obtain estimates for the weight parameters $\hat{\beta}_0, \dots, \hat{\beta}_4$. Thus, I have completed the first stage of my research method. In accordance with the design procedure for virtual sensors (Fortuna et al., 2007), I have proposed the variable selection and model design of my study. In the following, I will proceed to the second phase of my research method and utilize the estimated weight parameters to predict instantaneous fuel flow rates.

3.4 Making Predictions

In the second stage of my research method, we predict instantaneous fuel flow by applying the estimated weight parameters to unaggregated high-frequency GPS and meteorological data. Once the weight parameters $\hat{\beta}_0, \dots, \hat{\beta}_4$ are estimated, we can make predictions using an equation in the general form of

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p, \quad (12)$$

where $\hat{y}$ indicates a prediction of *Y* on the basis of $X = x$. We use Equation (9), but replace the crew-reported total fuel flow $\Phi_{\Delta T}$ with the *instantaneous* fuel flow rate Φ_t. Where ΔT denoted a time span (ca. 24 hours), *t* denotes a single point in time. Likewise, the aggregated predictors (water, draft, wave, and wind) are replaced with their unaggregated values. In particular, the instantaneous fuel flow is given by

$$\widehat{\Phi}_t = \hat{\beta}_0 + \hat{\beta}_1 \times water_t + \hat{\beta}_2 \times draft_t + \hat{\beta}_3 \times waves_t + \hat{\beta}_4 \times wind_t. \quad (13)$$

With the entire right side of the equation known, we can evaluate the formula and predict instantaneous fuel flow values.

This concludes the second stage of my research method. Note that this step corresponds to the model estimation part in theoretical framework of Fortuna et al. (2007). The next step is to assess the accurate the fuel flow predictions, for which I will introduce reference data from an integration-based system.

3.5 Model Validation

The final step in the soft sensor design process is model validation (Fortuna et al., 2007), usually done by comparing some predictions against reference data. However, the reason for developing a *virtual* fuel flow sensor is that the vessel is not equipped with a physical fuel flow meter; no reference fuel flow exists. Otherwise we could split the data, train a model on the reference values, and use cross-validation to assess the goodness of fit. Following the method of Antola, Solonen, and Staboulis (2017), it became possible to estimate instantaneous fuel flow without the need of (high-frequency) reference data, by training a model on the aggregated sample data. Nevertheless, for validating the accuracy of the model predictions, reference data is indispensable.

The case vessel was simultaneously equipped with a lightweight and an integration-based system. This setup allows for model validation as reference data on the vessel speed and fuel flow rates can be taken from the integration-based system.

In accordance with the third stage of my research method, I will assess the model performance and check for systematic modelling errors (two of the three main problems that were identified by Antola, Solonen, and Staboulis (2017)), by comparing the fuel flow predictions against measurement-based reference data. I will visually compare the fuel flow curves over the two-month study period, analyze the residuals, and compute different error terms.

Subsequently, I will address the fourth phase of my research method and assess the effect of human reporting error by simulating error-free reports from the reference data. As the model is trained on crew-generated reports, and marketed as an independent monitoring tool, the question is if and how the crew can influence the model output; either by accidently or even willfully reporting incorrect values. It would cast doubt on the whole concept if a supposing independent monitoring tool can be manipulated by the very people it is devised to monitor. In a third step, I will introduce speed-fuel-curves, a standard analytical tool used in the marine industry to assess fuel consumption at different speeds.

In summary, my research methodology follows the design procedure of virtual sensors (Fortuna et al., 2007): variable selection, model design, model predictions, and model validation. In the following, I will examine the data used for this study.

4 Data

The sample size and sample period for this study were determined by the simultaneous availability of data from a lightweight system and an integration-based system, installed on the same vessel. From May 1st until June 30th 2017, the case vessel was equipped with both systems. Hence, the **sample period** of my study is this particular **two-month period**.

The sample size is more difficult to determine, since the sampling frequency is different for all three main data sources: 30 seconds for the integration-based system, five minutes for the lightweight system, and roughly (but varying) 24 hours for the crew-generated noon reports. Over the two-months period, we obtain 175,679 observations from the integration-based system (reference), 17,151 observations from the lightweight system, and 61 observations from the noon reports. Despite their discrepancies, the lightweight system and the integration-based system are both considered high-frequency data sources, while the noon reports are considered low-frequency. The term *observation* refers to one entry in the database, meaning one instance with a unique time stamp (or span) and corresponding values for a certain number of variables (e.g. speed and draft). The reference data is resampled to match the five-minute period of the lightweight system. In practice, we are therefore left with a **sample size** of 17,151 observations for the high-frequency data and 61 observations on the low-frequency side.

The **data analysis** is implemented in Python, using a Jupyter notebook. The Pandas data analysis library is used as it allows for the creation of powerful data frames and data analysis tools. Additionally, the NumPy, StatsModels, Seaborn, and Matplotlib libraries are used for statistical computations, linear regression models, and scientific plotting.

4.1 Unit of Analysis

All vessel data used for this study is taken from a single case vessel with the unique attribute that it has been equipped simultaneously with a lightweight system and an integration-based system. This case vessel is therefore the **unit of analysis** of my study. The vessel's true name has to remain anonymous so I will refer to it as *"Boaty McBoatface"*. See Figure 6 for a picture of the vessel class; Table 1 provides the main specifications and principal characteristics of the case vessel.

Figure 6. The case vessel "Boaty McBoatface": it is equipped with a lightweight system and an integration-based solution for study purposes. Hence, it provides data for formulating a model for virtual fuel flow estimation and also offers reference data for model validation and for the simulation of error-free noon reports.

Table 1: Main specifications of the case vessel "Boaty McBoatface"

Year built:	2009
Type:	Container ship
Gross tonnage:	73,899 t
Deadweight:	80,293 t
Length x breadth:	275m x 40m
Draft (mean):	8.8 m (provided by customer)
Engine:	Sulzer 8RTA96CB 2-stroke
Power (nom.):	45,760 kW
Speed avg./max.:	14.3 kn / 26 kn

It would go beyond the scope of this Master's thesis to validate the approach of Antola, Solonen, and Staboulis (2017) across multiple vessels and vessel classes. In Chapter 6.3, I will discuss the limitations that arise from the case study approach.

4.2 Data Exploration

The data can be classified as observational, nonexperimental time series data with no time trend. It is observational, since no experiments were conducted to gather data specifically for this study; we simply observe the data as it is collected by the two systems during their normal operations. The data can be classified as time series data, as each entry is indexed by

a time stamp (or time span) that represents the value(s) for one or more quantities at a particular time (e.g. fuel flow on 01.05.2017 at 00:00:05 o'clock). The data shows no time trend and does not require adjustments for autoregression. There is no reason to assume that the crew operates a vessel differently tomorrow because of what has happened today. Every day is a new case (new cargo, changing weather, etc.) and there is no up- or down-trend in the data.

The data is stored in three separate Pandas data frames: 'reports', 'data', and 'data_ref'. In the following, I will discuss the three data frames and visualize their structures. Refer to Appendix A for an overview of all variables and their definitions.

Low-frequency

The data frame **'reports'** holds the low-frequency data. It is indexed with time spans that capture the start- and end-times of each reporting period. The regression model for weight parameter estimation is fitted to the 'reports' data set, therefore it has to contain the crew-reported quantities, as well as aggregates of the speed and meteorological data. The '_spanned' suffix denotes a quantity that has been aggregated from its native sampling period to a time span. There are eight variables in 'reports':

- `draft_manual`: the crew-reported vessel draft (in meters)
- `stw_manual`: the crew-reported vessel speed (in knots)
- `fuelflow_manual`: the crew-reported fuel consumption (in tons/day)
- `stw_spanned`: the aggregated forecast-STW (in knots)
- `water_spanned`: the aggregated water-induced power loss ($\int X_1 dt$)
- `draft_spanned`: the aggregated draft-induced power loss ($\int X_2 dt$)
- `wave_spanned`: the aggregated wave-induced power loss ($\int X_3 dt$)
- `wind_spanned`: the aggregated wind-induced power loss ($\int X_4 dt$)

The data frame contains 61 rows and has no missing values, except for the crew-reported speed (27/61). The reporting time spans in Figure 7 reveal the irregular reporting intervals.

		fuelflow_manual	stw_manual	draft_manual	stw_spanned	water_spanned	draft_spanned	wave_spanned	wind_spanned
time	end								
2017-04-30 10:00:00	2017-05-01 10:00:00	41.00	NaN	11.25	10.62	1196.70	333.17	15.25	2689.70
2017-05-01 10:00:00	2017-05-02 08:12:00	34.60	10.71	11.25	10.47	1146.80	319.28	13.90	-1015.60
2017-05-02 08:12:00	2017-05-02 14:30:00	12.14	10.50	11.25	0.57	0.18	0.05	0.70	1.12
2017-05-02 14:30:00	2017-05-03 10:00:00	114.46	NaN	10.40	18.89	6741.10	1225.65	58.55	-2713.70
2017-05-03 10:00:00	2017-05-04 10:00:00	122.25	18.82	10.40	18.46	6287.30	1143.15	103.98	6127.90

Figure 7. The 'reports' data frame (head): it stores all crew-reported quantities, as well as aggregates of the high-freuqency GPS and meteorological data. A model is fitted onto 'reports' to obtain the weight parameters $\hat{\beta}_0, \dots, \hat{\beta}_4$. The data frame has no missing values, except for the manually-reported vessel speed (27/61). Note the irregualar reporting intervals, captured by the timespan index. Further note how the wind effect can be positive or negative.

High-frequency

The data frame **'data'** contains the high-frequency data from the lightweight system and has a regular five-minute sampling period. The instantaneous fuel flow is calculated from 'data', thus it contains all variables in their *un*aggregated form.

- `sog`: the speed-over-ground (in knots)
- `stw`: the forecast-STW (in knots)
- `draft_manual`: the crew-reported vessel draft (in meters)
- `water`: the water-induced power loss (X_1)
- `draft`: the draft-induced power loss (X_2)
- `wave`: the wave-induced power loss (X_3)
- `wind`: the wind-induced power loss (X_4)
- `Lightweight_pred`: the fuel flow predictions from the real lightweight system that is installed on the case vessel (in tons/day)
- (`fuelflow_pred`: the modeled fuel flow (in tons/day); added to 'data' once fuel flow predictions are obtained)

The 'data' data frame has eight variables and 17,151 rows. We observe 36 NULL values in all variables except for the `sog`. However, the missing values appear to be the same for all observations. Given the large data set, 36 NULL values will not significantly influence the model performance.

	sog	stw	draft_manual	water	draft	wave	wind	Lightweight_pred
time								
2017-05-01 00:00:32	9.80	10.08	11.25	1023.89	285.06	14.51	3904.90	33.44
2017-05-01 00:05:32	9.80	10.08	11.25	1025.11	285.40	14.52	4704.70	34.11
2017-05-01 00:10:32	9.90	10.19	11.25	1056.84	294.24	14.67	4732.70	34.62
2017-05-01 00:15:32	9.80	10.09	11.25	1026.94	285.91	14.53	4635.50	34.08
2017-05-01 00:20:32	9.90	10.19	11.25	1058.40	294.67	14.68	4685.80	34.61

Figure 8. The 'data' data frame (head): it stores the unaggregated values of the GPS and meteorological data. The estimated weight parameters are multiplied with the values from 'data' to obtain instantaneous fuel flow estimates. Note the regular five-minute sampling period.

Reference data

The data frame **'data_ref'** contains the high-frequency reference data obtained from the integration-based system. It has a 30-second sampling period and provides reference values for the model validation and for the simulation of error-free reports. There are three variables:

- `draft_ref`: a reference draft value (in meters)
- `stw_ref`: a reference STW value (in knots)
- `fuelflow_ref`: a reference fuel flow value (in tons/day)

The data frame has 175,679 rows. A few NULL values occur, however, given the absolute size of the data set, they are irrelevant.

	draft_ref	stw_ref	fuelflow_ref
time			
2017-05-01 00:00:00	11.16	10.72	39.62
2017-05-01 00:00:30	11.16	10.72	39.68
2017-05-01 00:01:00	11.16	10.72	39.58
2017-05-01 00:01:30	11.16	10.72	39.58
2017-05-01 00:02:00	11.16	10.72	39.57

Figure 9. The 'data_ref' data frame (head): It contains reference data from the integration-based system.

After this overall exploration of the available data, I will examine one input further that is included (in different forms) in all three data frames, and which could arguably be the most import input variable for the lightweight model: vessel speed.

4.3 Speed Proxy

The relationship between fuel flow and a vessel's speed-through-water is roughly cubic (e.g. $X_1 = water = STW^3$), thus any error in the speed input variable will aggregated throughout the model and significantly affect the estimated fuel flow output. In the following, I will therefore examine the accuracy of the speed value used as input for the lightweight model. Figure 10 plots different speed measures over the two-month sampling period:

- **SOG:** from the GPS unit
- **Forecast-STW**: SOG adjusted for the relative effect of the ocean currents (Equation (2)); speed input variable for the lightweight model
- **Reference STW:** most accurate approximation of a vessel's speed; measurement-based and only available on integration-based system; used as reference.

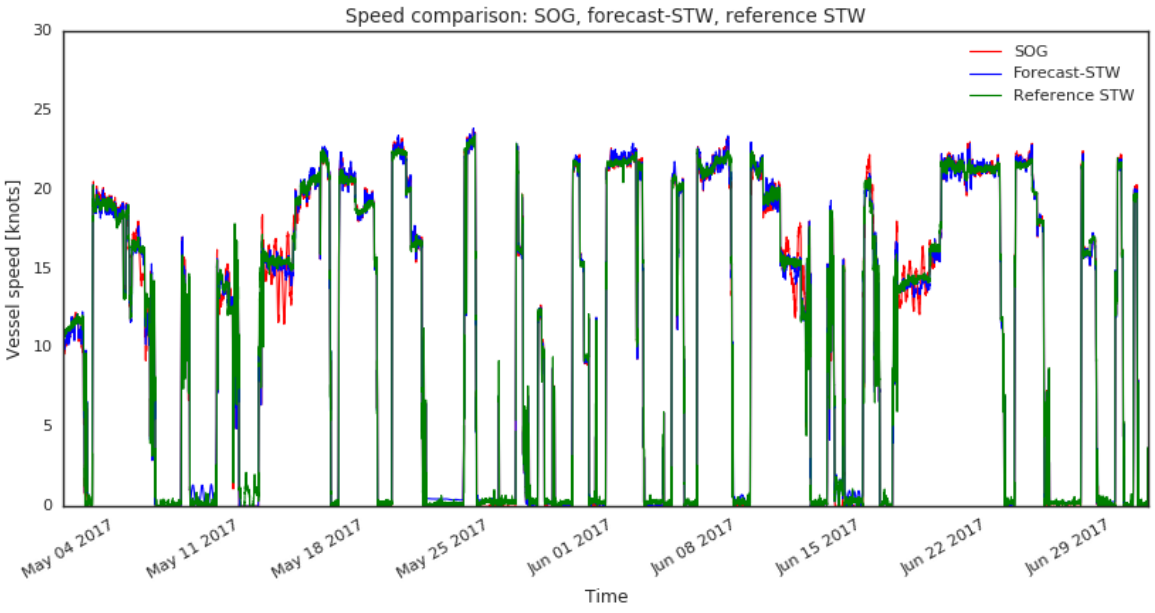


Figure 10. Speed proxy comparison: The SOG (red) represents the unadjusted vessel speed-over-ground from the GPS; the forecast-STW (blue) shows the speed input variable for the lightweight model; the reference STW (green) provides a highly accurate speed reference value from the integration-based system.

Overall, the SOG and forecast-STW correlate nicely with the reference STW. Zooming into the graph provides a more detailed view.

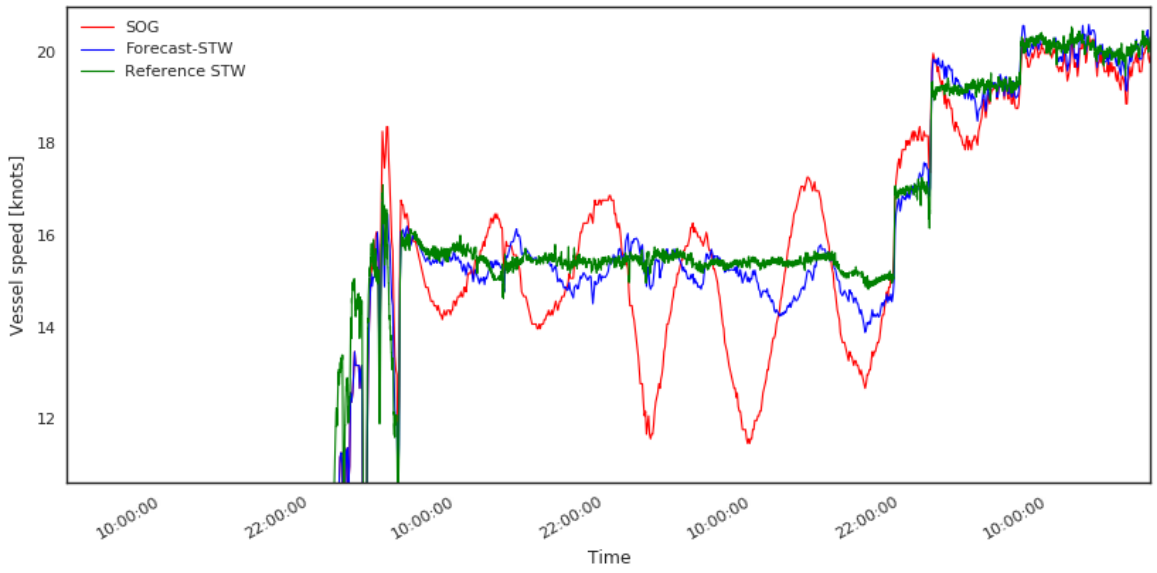


Figure 11. Detailed speed proxy comparison: The SOG is highly erratic at times and should therefore not be used as a speed proxy. The forecast-STW is smoother and correlates nicely with the reference STW.

While the SOG is accurate in what it measures (speed-over-ground), it is highly erratic if used as a speed-through-water proxy. Around May 13th, the SOG shows a high oscillation while the green reference curve remains smooth. The blue forecast-STW is overall slightly noisier than the reference speed, but not nearly as volatile as the SOG. Adjusting the vessel speed-over-ground with meteorological data on ocean currents, improves the accuracy and variance of the speed input variable significantly. The visual inspection of the graphs suggests that the forecast-STW can be regarded a sufficiently accurate input variable for the following fuel flow model. Given the inexpensive way in which the forecast-STW is obtained (GPS unit plus external weather forecasts), this finding further suggests that researchers could in general improve the accuracy of their studies by adjusting a raw GPS signal. For example, the approach by Trodden et al. (2015), in which SOG from GPS was combined with an onboard fuel flow meter might further be improved if a forecast-STW-like speed value was used.

Moreover, these findings concur with the results of Antola, Solonen, and Staboulis (2017), who also presented a speed proxy comparison as illustrated by Figure 12. The relation between SOG (grey), forecast-STW (green), and reference STW (pink) concurs with my own findings as presented in Figure 10 and Figure 11. Especially in the presence of significant currents, the forecast-STW is more in accordance with the reference STW than the SOG.

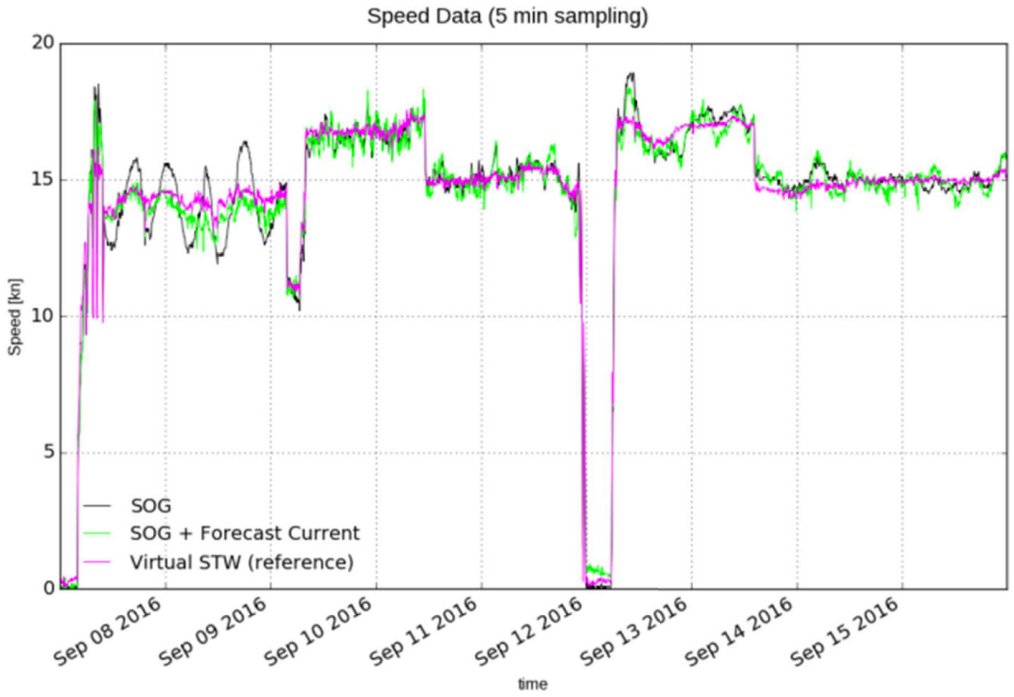


Figure 12. Speed proxy comparison by Antola et. al. (2017): a visual inspection confirms the previously discussed findings from Figure 10 and Figure 11.

4.4 Limitations of the Data

The first data limitation concerns the weather data. Antola, Solonen, and Staboulis (2017) state that "although the GPS-based SOG estimate is fairly accurate, the [meteorological] forecasts are often not (they have low bias, but momentary errors can be high)." Figure 13 shows how the meteorological forecasts often contain data gaps. For example, the wave forecasts are generally unavailable in close proximity to the shorelines and are further missing for certain areas, e.g. the Mediterranean Sea. However, given the large sample size of the 'data' data frame, a few missing meteorological values are of little concern when training the model. A robust fuel flow model has to be able to handle gaps in e.g. the wind force and the wave height. I will therefore neither remove entries with missing weather data, nor try to fill them.

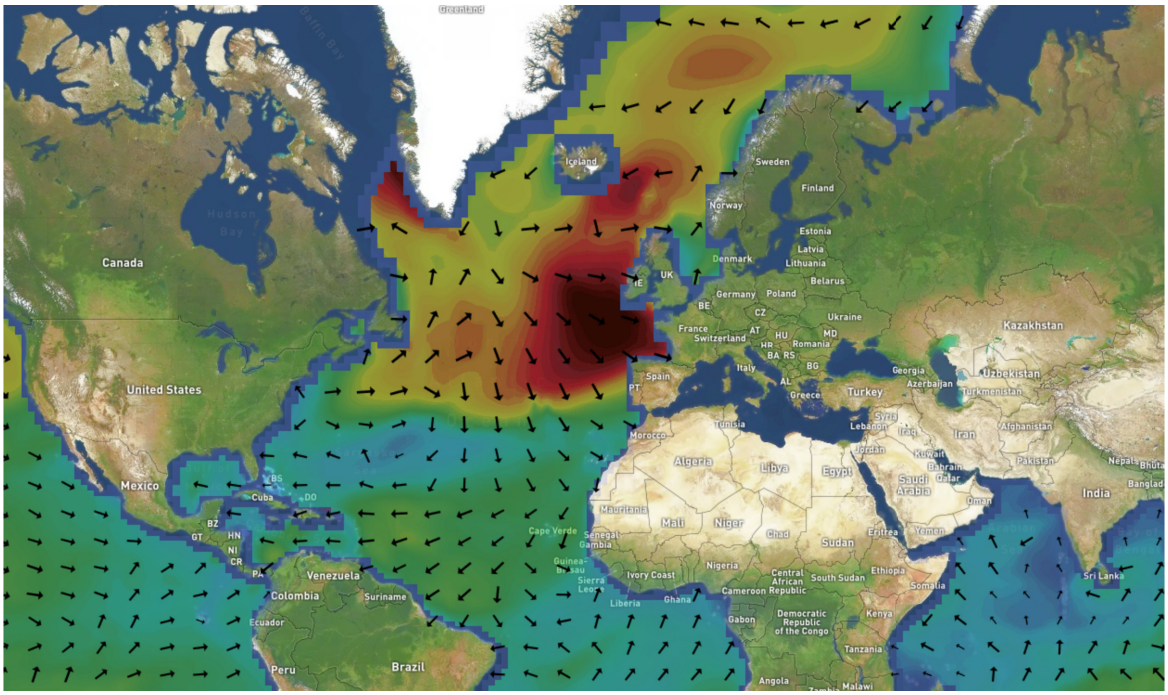


Figure 13. Global wave forecasts from Tidetech (based on data from NOAA): the map shows forecasts for the height and primary direction of waves. Note how the data is generally unavailable in proximity to the shorelines and also for certain regions, e.g. the Mediterranean Sea.

Second, the data is not raw, but often pre-processed. In Equation (2) (Chapter 3.1), an artificial measurement of the vessel's speed-through-water is formed by combining speed-over-ground with forecasts on ocean currents. In practice, this is not a simple addition/subtraction of two speed values. Instead, one has to look at the course of the vessel relative to the currents, expressed as two vectors in a certain angle to one another. It would be cumbersome to query the raw data in this (and similar) case(s) when the preprocessed data is available. As a general rule, I queried only preprocessed data when it was subject to mathematical conversions that could have been done in theory by myself (given considerable time and effort).

A third limitation concerns the reference data from the integration based system. The case vessel is not equipped with a physical fuel flow measurement device. Therefore, the reference data is not measured fuel flow, but a **scaled propulsion power value.** Physical mass (or volume) fuel flow meters are costly and error-prone devices that are mostly found on cruise ships. As the lightweight system is marketed only at commercial vessels (e.g. container ships, tankers, and bulkers), it would be nearly impossible to find a case vessel that is suitable for the lightweight system and equipped with a physical fuel flow meter. Moreover, vessels that are sophisticated enough to be equipped with a physical fuel flow meter usually have not one, but multiple main engines (and fuel meters) that are run in

various combinations and on various fuel types, depending on the required power output. Hence, if we obtained physical fuel flow measurements, we would need to find a way to combine their readings into a single fuel flow value. This non-trivial step would significantly complicate the validation and add more room for errors. In contrast, the scaled propulsion power can be expressed in a single value. On commercial vessels equipped with integration-based systems, the fuel flow can be accurately modeled with a sophisticated engine model that utilizes a broad range of measured quantities. So, while the reference value is technically not a measured fuel flow, it describes the fuel flow satisfactory well and will therefore be used as the reference value within the context of my study. Nevertheless, there is a probability that the scaled propulsion power is biased.

5 Results and Discussion

In this chapter, I will present the findings and implications of my research.

5.1 Model Fit to Aggregated Sample Data

In Chapter 3.3, I proposed how to "fit a model for weight parameter estimation onto low-frequency crew reports and onto aggregates of high-frequency data". In Equation (10), the crew-reported fuel consumption was regressed onto aggregates of the resistance variables water, draft, waves, and wind. In practice, this is done using the StatsModels OLS Python package. Figure 14 shows the OLS regression results.

OLS Regression Results

Dep. Variable:	fuelflow_manual	R-squared:	0.882
Model:	OLS	Adj. R-squared:	0.873
Method:	Least Squares	F-statistic:	104.4
Date:	Mon, 09 Oct 2017	Prob (F-statistic):	2.82e-25
Time:	07:47:06	Log-Likelihood:	-266.09
No. Observations:	61	AIC:	542.2
Df Residuals:	56	BIC:	552.7
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
Intercept	17.5017	3.895	4.493	0.000	9.699	25.305
water_spanned	0.0145	0.002	8.757	0.000	0.011	0.018
draft_spanned	-0.0033	0.006	-0.534	0.595	-0.016	0.009
wave_spanned	0.0985	0.147	0.670	0.506	-0.196	0.393
wind_spanned	0.0017	0.001	2.503	0.015	0.000	0.003

Omnibus:	32.694	Durbin-Watson:	2.817
Prob(Omnibus):	0.000	Jarque-Bera (JB):	714.473
Skew:	-0.432	Prob(JB):	7.15e-156
Kurtosis:	19.744	Cond. No.	9.17e+03

Figure 14. OLS regression results: the model fits the data well (R^2 = 0.882) and provides estimates for the weight parameters $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$. Note also the high multicollinearity, given by the Cond.No.

We observe the weight parameter estimates $\hat{\beta}_1 = 0.015, \hat{\beta}_2 = -0.003, \hat{\beta}_3 = 0.099$, and $\hat{\beta}_4 = 0.002$. The estimated intercept value $\hat{\beta}_0 = 17.502$ can be interpreted as the base fuel consumption of the auxiliary engines. When the ship is anchored at port, and the main engine(s) is/are shut off, the auxiliary engines are still running to provide electricity for the onboard systems. Hence, the total fuel consumption never drops to zero, but stays at around 17.5 tons per day. Next, we run some regression diagnostics to assess the model fit and to ensure the reliability of the estimates. As a first step, we plot the fitted values ($\hat{y}$) against the crew-reported ones (`fuelflow_manual`, y) in a regression plot.

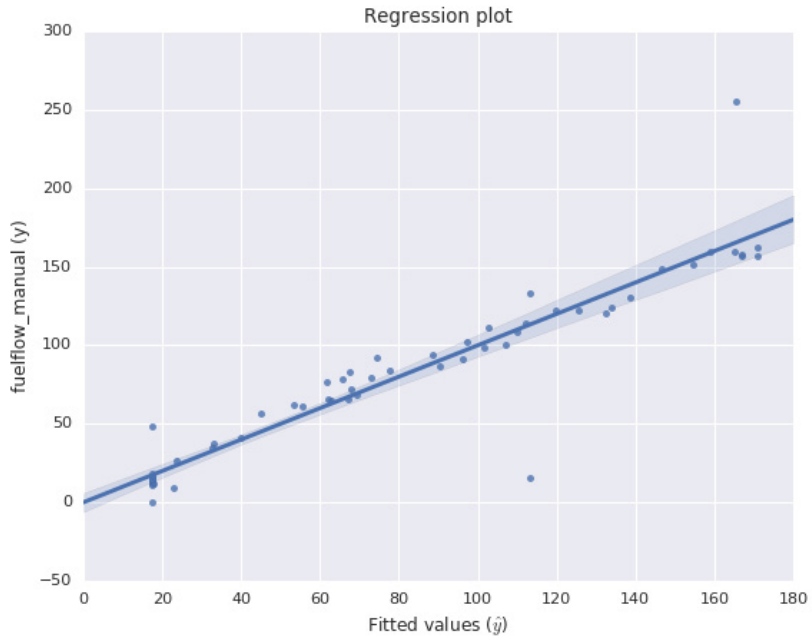


Figure 15. Regression plot ($\hat{y}$ vs. y): overall, the model fits the noon report data well. The two outliers will be examined in the subsequent chapters. They hint at the sometimes significant human reporting errors.

Overall, the data points in Figure 15 form a linear pattern close to the regression line. Two outliers add some noise at higher consumption levels and increase the spread. Nevertheless, the model fits the noon report data well.

Thus, we proceed to quantify the extent to which the model fits the data. The quality of a linear regression fit is typically evaluated using two related quantities: the residual standard error (RSE) and the R^2 statistic. The RSE provides an absolute measure of the lack of fit of the model to the data and is measured in units of *Y*. The model shows an RSE of 19.8, meaning that the fitted fuel flow deviates on average by 19.8 tons per day from the crew-reported value. With a mean reported value of 75 tons per day, this RSE translates into a percentage error of 26%.

The R^2 provides an alternative measure of fit. It describes the fraction of variance explained (James et al., 2013). In multiple linear regression, R^2 equals the square of the correlation between the response and the fitted linear model, express as $Cor(Y, \hat{Y})^2$. Essentially, the fitted linear model maximizes this correlation among all possible linear models. Generally, an R^2 value close to one indicates that the model explains a large portion of the variance in the response variable (James et al., 2013). The model shows an R^2 value of 0.882 which indicates a good model fit to the noon reports. The model is able to explain roughly 88% of the variance in the crew-reported consumption, despite the small number of input variables. Note that we only assess the model fit in isolation, that is we compare the

crew-reported data (y) against the fitted values ($\hat{y}$). The R^2 and RSE say nothing about the accuracy of the instantaneous fuel flow predictions, but only quantify how well the regression model fits the training data.

Response-predictors-relationship

Next, we assess the relationship between the response fuel flow variable and the predictors. The question is whether at least one of the predictors X_1, X_2, X_3, X_4 is useful in predicting the response (James et al., 2013). Since we are dealing with $p = 4$ predictors, we need to check whether all of the regression coefficients are zero. We test the null hypothesis,

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

against the alternative

$$H_a : at\ least\ one\ \beta_j\ \ is\ non - zero.$$

The hypothesis test is done using the F-statistic. F takes a value close to one when there is no relationship between the response and the predictors and a value greater than one if H_a is true. In Figure 14, we observe a F-statistic of 104, far greater than one. This provides convincing evidence to reject the null hypothesis H_0. At least one of the predictors must be related to fuel flow.

Moreover, we can check whether a particular subset q of the coefficients is zero. The corresponding null-hypothesis is expressed as

$$H_0: \beta_{p-q+1} = \beta_{p-q+2} = \cdots \beta_p\ = 0.$$

Figure 14 also provides the t-statistic and a p-value for each predictor. Thus, we can check whether each predictor is related to the response, after adjusting for the remaining predictors. The p-values indicate that the water (p=0.000) and wind (p=0.015) are significantly related to the fuel consumption. This is not surprising as the frictional resistance (approximated by STW^3) accounts for 70-90% of the total resistance. However, it is difficult to judge whether the draft (p=0.604) and the wave effect (p=0.509) are significantly associated with fuel flow, in the presence of the water predictor.

We investigate the response-predictor-relationship further, by analyzing several nested model variations of the full model. Table 2 shows the coefficient estimates, significance

levels, R^2, and F-statistics for different reduced model variations that exclude more than one variable. Note that all nested models include the water, since it is clearly significantly related to the response variable.

Table 2: Statistics of nested model variations: the dependent variable is "fuelflow_manual", the crew-reported fuel consumption.

	Full model	Water + draft	Water + waves	Water + wind	Water
β_1	.0145 ***	.0141 ***	.0143 ***	.0142 ***	.0148 ***
β_2	-.0033	.0033	-	-	-
β_3	.0985	-	.1365	-	-
β_4	.0017 *	-	-	.0016 *	-
R^2	88.2%	86.6%	86.7%	87.9%	86.5%
s^2	392	428	424	384	123
F-statistic	104	187	189	212	379

Signif. codes: 0 '***' p<0.001 '**' p<0.01 '*' p<0.05 '.' p<0.1 ' ' p<1

While the R^2 is naturally highest in the full model, it does not decrease significantly in the reduced models. Even a model based only on the water predictor fits the data almost as well as the full model.

Water remains the single most significant predictor in all reduced models. As in the full model, the wind variable is also somewhat significant in the reduced model water + wind. Neither the draft, nor the waves gain in significance in the reduced models. The residual variance s^2 shows 392 for the full model. Reducing the model to water + draft or water + waves increases the variance to 428 and 424 respectively. With 384, the variance of the water + wind variation is close the full model. The water-only model shows the lowest residual variance of 123. To clarify whether certain variables could be excluded from the model, we compute F-tests for the nested model variations. The results are summarized in Table 3.

Table 3: Comparison of nested model variations: the full model has four predictors in the form of $\Phi_{\Delta T} \approx \beta_0 + \beta_1 \int_{\Delta T} water_t dt + \beta_2 \int_{\Delta T} draft_t dt + \beta_3 \int_{\Delta T} waves_t dt + \beta_4 \int_{\Delta T} wind_t dt$ *and an* R^2 *of 88.2%.*

Nested model variations	Hypothesis	R^2	R^2 change compared to full model	F-test	Pr(>F)	Conclusion
Water + draft	$H_0: \beta_3 = \beta_4 = 0$	86.6%	-1.6%	3.666	0.032	Reject H_0
Water + waves	$H_0: \beta_2 = \beta_4 = 0$	86.7%	-1.5%	3.403	0.040	Reject H_0
Water + wind	$H_0: \beta_2 = \beta_3 = 0$	87.9%	-0.3%	0.447	0.642	Accept H_0
Water	$H_0: \beta_2 = \beta_3 = \beta_4 = 0$	86.5%	-1.7%	2.566	0.064	Accept H_0

The first reduced model uses only the water and draft predictors. We test $H_0: \beta_3 = \beta_4 = 0$ against the alternative $H_a: either\ \beta_3\ or\ \beta_4 \neq 0$. The F-test shows F = 3.666 and a p-value 0.032 < 0.05. Thus, we reject H_0 and conclude that the constraint should be relaxed in favor of the full model. That is, at least one of the excluded predictors (waves, wind) adds significant additional information. The second reduced model uses only the water and waves predictors. We test $H_0: \beta_2 = \beta_4 = 0$ against the alternative $H_a: either\ \beta_2\ or\ \beta_4 \neq 0$. With F = 3.404 and a p-value 0.040 < 0.05, we reject H_0 and conclude that at least one of the excluded predictors (draft, wind) adds significant additional information. The third reduced model uses only the water and wind predictors. We test $H_0: \beta_2 = \beta_3 = 0$ against the alternative $H_a: either\ \beta_2\ or\ \beta_3 \neq 0$. With F = 0.447 and a p-value 0.642 > 0.05, we accept the null hypothesis. A model based only on water and wind appears to be equivalent to the full model; including the draft and wave effect does not improve the fit of the model to the data significantly. The final nested model uses only the water predictor. We test $H_0: \beta_2 = \beta_3 = \beta_4 = 0$ against the alternative $H_a: either\ \beta_2\ or\ \beta_3\ or\ \beta_4 \neq 0$. With F = 2.566 and a p-value 0.063 > 0.05, we accept H_0. The remaining predictors are not significant in explaining the variation in the fuel flow. It appears that a model that uses only the water-induced resistance is equivalent to the full model.

In summary, these findings confirm the impression from Figure 14 that speed is the single most important factor in explaining fuel flow. The water predictor captures a large portion of the speed-induced effect on fuel flow and the remaining speed-dependent predictors are less influential on the response variable. The draft and wave predictor appear to be not significantly associated with the fuel flow. However, the F-test could not conclusively show whether to include the wind variable, as the reduced models "water + wind" and "water" appear to be both equivalent to the full model.

In light of these ambiguous findings, no variables will be excluded. From the literature on marine hydrodynamics, e.g. by Bertram (2000), we know that there is a physical relation between the fuel consumption and the four predictors. Instead of labeling some variables insignificant based on the statistics of Table 2 and Table 3, we will apply a more intuitive approach and take a closer look at the case vessel Boaty McBoatface; a containership. Compared to tankers or bulkers, its cargo can vary significantly in weight. However, it is possible that the weight and draft remained relatively equal throughout the two-months sampling period, causing the draft parameter to appear statistically insignificant. Moreover, the wave resistance is proportional to the speed-squared at low speeds; but at higher speeds, it increases nearly exponentially. The wave-predictor might appear insignificant because the vessel is mostly sailing at low speeds or in good weather, when the relative wave effect is low and already captured by the water predictor. However, if we applied the model to a vessel with a higher draft variance or observed a time of heavy storms (high waves), the draft and wave predictors might grow in significance.

We conclude this section on the response-predictors-relationship that some predictors might not be strongly related to the response in the underlying model, yet we cannot treat them as irrelevant. The nature of the vessel might influence the relevance of some predictors. As stated in Chapter 3.2, the water-induced resistance accounts for 70% to 90% of the total resistance. It appears that the remaining predictors mainly fine-tune the model output. Since we seek to generally validating the approach of Antola, Solonen, and Staboulis (2017), we need to keep the model applicable to as many vessels as possible. I will therefore keep all four predictors.

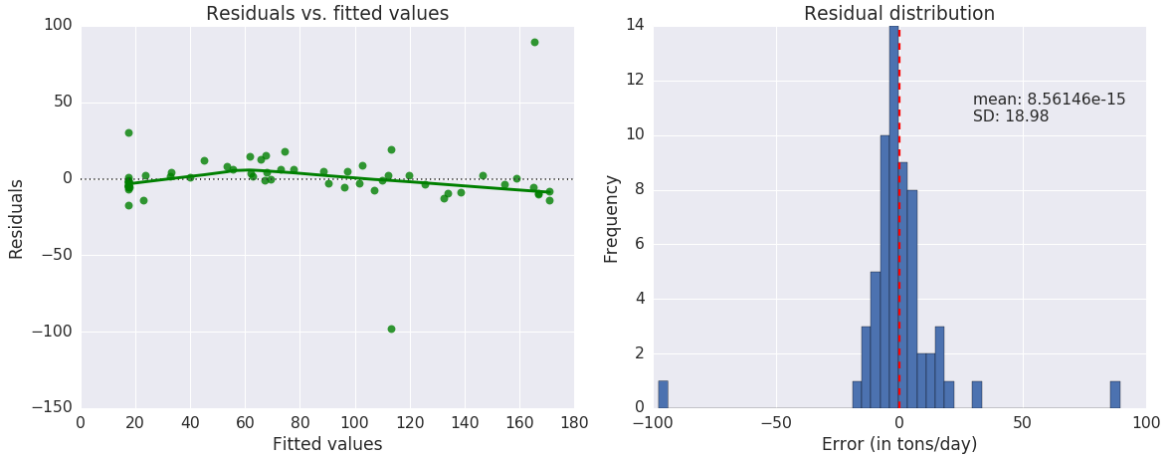
Collinearity

Collinearity refers to the situation in which two or more predictors are closely related to one another. In general, the presence of collinearity can create problems in regression models, as it can be difficult to separate the individual effects of collinear variables on the response (James et al., 2013). Collinearity among three or more variables is known as multicollinearity. Figure 14 shows a cond. No. of 9170, indicating a high multicollinearity.

However, within the context of fitting the model to the crew-reported consumption, we are not interested in the statistical significance, but only the model fit; multicollinearity can be ignored. Given how each predictor is including the vessel speed (squared or cubed), we would actually expect them to be closely related to one another.

Non-linearity of the data

In a linear regression model, we assume a straight-line relationship between the predictors and the response. If the true relationship is far from linear, then basically all conclusions that we draw from the fit are questionable. The model based on Equation (10), is non-linear w.r.t. to the predictors (e.g. $X_1 = water = STW^3$), but linear w.r.t. to the weight parameters $\beta_0, \dots, \beta_4$. I assume that a non-linear relationship can be approximated using a model that is linear in parameters, but non-linear in variables. Thus, a linear regression model (Figure 14) can be used to obtain estimates for the weight parameters $\hat{\beta}_0, \dots, \hat{\beta}_4$. Residual plots are a useful tool for identifying non-linearity. We plot the residuals versus the fitted values $\hat{y}_i$. The presence of a pattern may indicate an issue with some aspect of the model, hence we want the plot to show no discernable pattern.



a) Residual vs. fitted values b) Residual distribution

Figure 16. Residual plot: the data points show no discernable pattern.

Figure 16 panel a) suggests that there is no significant pattern in the residuals; the cubic term of the speed seems to improve the fit of the data. The model is indeed linear w.r.t. the weight parameters. Moreover, panel b) indicates that the residuals are distributed around zero. However, both plots also reveal the presence of outliers in the data.

Outliers

An outlier is a point for which y_i is far away from the value predicted by the model. Generally, outliers can arise for a variety of reasons, such as incorrect recording of an observation during the data collection (James et al., 2013). The regression plot in Figure 15 and the residual histogram in Figure 16 b) revealed two outliers. The fitted values deviate

from the crew-reported consumption by nearly 100 tons. In light of these outliers, we should not automatically assume that the predictions are incorrect. We cannot tell whether the crew reported correctly and our model failed, or whether the crew reported incorrectly and the combination of speed and weather reveals the true fuel consumption. The plots only reveal the fit of the model to the crew-reported data. In Chapter 5.4 we will learn who is right and who is wrong by introducing true reference data for comparison.

Nevertheless, there is a chance that the outliers affect the RSE and R^2. However, the objective of this study is not to build a model that perfectly fits the crew-reported training data, but one that can predict instantaneous fuel flow. Thus, it would not be reasonable to exclude the outliers. If we are to build a virtual fuel flow sensor, it has to be able to handle outliers. Moreover, the outliers might help us to show that a product based on data analytics can generate significant value to the customer, in contrast to relying purely on erroneous, manually-reported values.

In summary, we have obtained the estimated coefficients $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ and assessed the goodness of fit of the model to the data. The linear regression model fits the crew-reported data well. In the next step, we will combine low-frequency and high-frequency data to predict *instantaneous* fuel flow.

5.2 Instantaneous Fuel Flow Prediction from Unaggregated Data

Once we have regressed total fuel consumption onto the predictors, we proceed to predict instantaneous fuel flow by applying the estimated weight parameters to unaggregated high-frequency GPS and meteorological data. To recap, we make predictions based on Equation (13) in the form of

$$\widehat{\Phi}_t = \hat{\beta}_0 + \hat{\beta}_1 \times water_t + \hat{\beta}_2 \times draft_t + \hat{\beta}_3 \times waves_t + \hat{\beta}_4 \times wind_t. \quad (13)$$

The estimated instantaneous fuel flow for every five minutes is stored as `fuelflow_pred` in the 'data' data frame.

	sog	stw	draft_manual	water	draft	wave	wind	Lightweight_pred	fuelflow_pred
time									
2017-05-01 00:00:32	9.80	10.08	11.25	1023.89	285.06	14.51	3904.90	33.44	39.54
2017-05-01 00:05:32	9.80	10.08	11.25	1025.11	285.40	14.52	4704.70	34.11	40.94
2017-05-01 00:10:32	9.90	10.19	11.25	1056.84	294.24	14.67	4732.70	34.62	41.43
2017-05-01 00:15:32	9.80	10.09	11.25	1026.94	285.91	14.53	4635.50	34.08	40.85
2017-05-01 00:20:32	9.90	10.19	11.25	1058.40	294.67	14.68	4685.80	34.61	41.38

Figure 17. The 'data' data frame including the fuel flow predictions: the column 'fuelflow_pred' has been added and contains momentary fuel flow predictions for every five minutes.

The low-frequency data frame "reports" contained only the total crew-reported fuel consumption over roughly (but varying) 24 hours. Over the two-month sampling period, we collected 61 observations. In contrast, 'data' now contains 17,151 fuel flow estimates for the same two-month period. By applying the estimated weight parameters $\hat{\beta}_0, \ldots, \hat{\beta}_4$ to the unaggregated high-frequency data, we are able to estimate instantaneous fuel flow rates and conclude the second stage of my research method.

5.3 Accuracy of Fuel Flow Predictions

So far, we have fitted a model by linear regression to aggregated sample data to estimate the weight parameters $\hat{\beta}_0, \ldots, \hat{\beta}_4$. These were combined with unaggregated high-frequency data to predict instantaneous fuel flow. The next step is to assess the accuracy of predictions by comparing them against measurement-based reference data. As discussed in Chapter 3.5, the hardware limitations of the lightweight system require us to obtain true reference data from an integration-based system. Recall that the reference values are technically not measured fuel flow rates but scaled propulsion power values (see Chapter 4.4). In this section, I will visually compare the fuel flow curves over the two-month sampling period, analyze the residuals, and compute different error terms.

As a first step, we plot the crew-reported values, the modeled fuel flow predictions, and the reference data over the study period. Figure 18 shows the entire period, while Figure 19 and Figure 20 are zoomed-in to provide a more detailed impression.

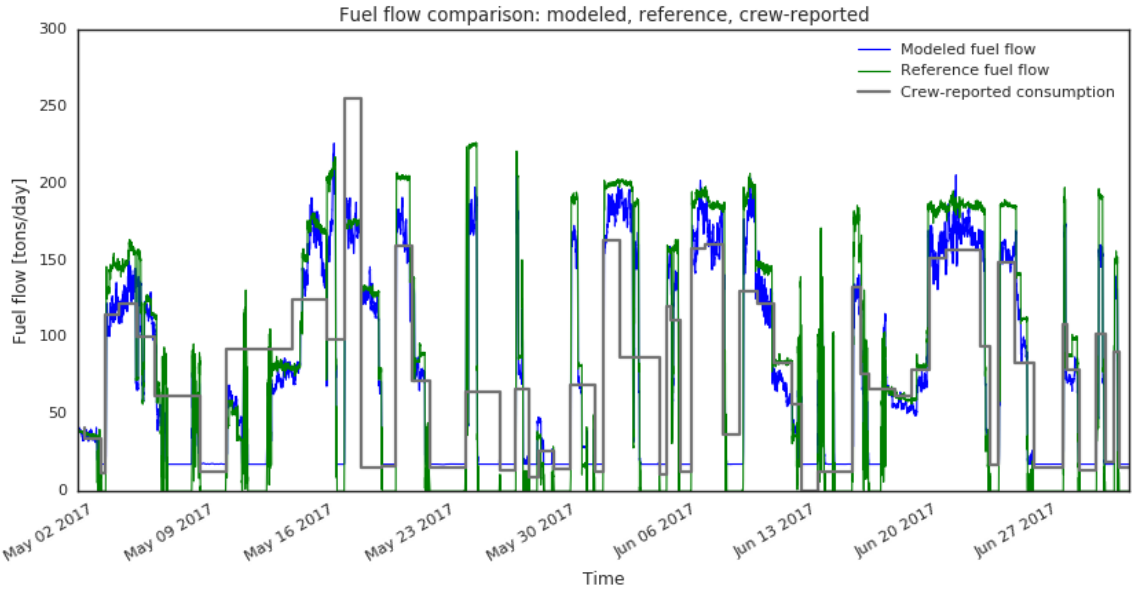


Figure 18. Fuel flow comparison: modelled vs. reference fuel flow vs. crew-reported consumption. The irregular and infrequent reporting of the crew creates a crude stepwise line. In contrast, the high-frequency modeled fuel flow offers a higher temporal resolution.

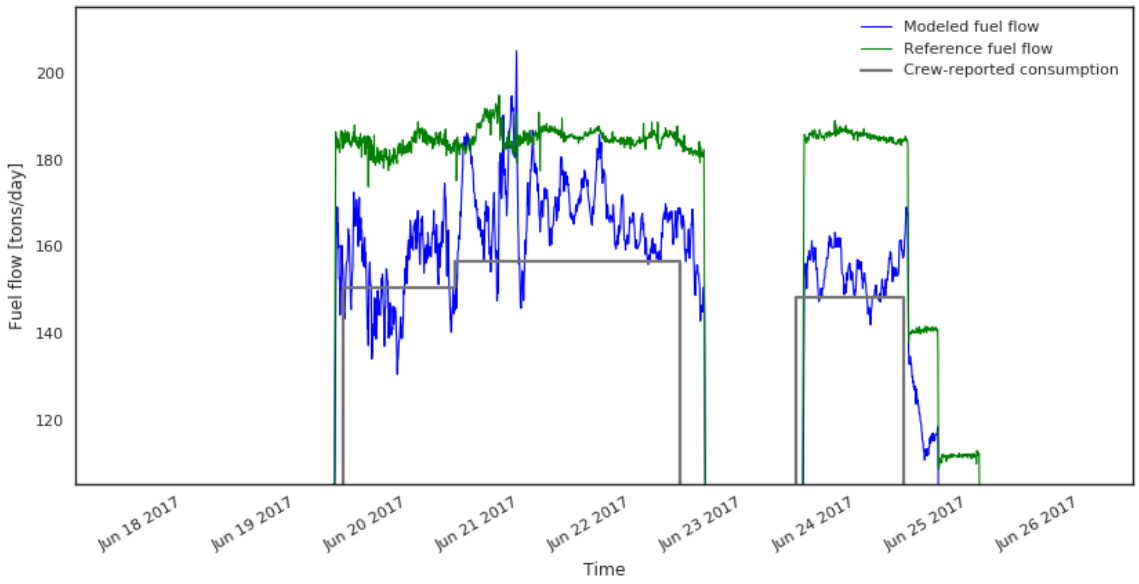


Figure 19. Detailed fuel flow comparison (1): the modeled fuel flow often underestimates the true consumption and is more erratic than the reference data.

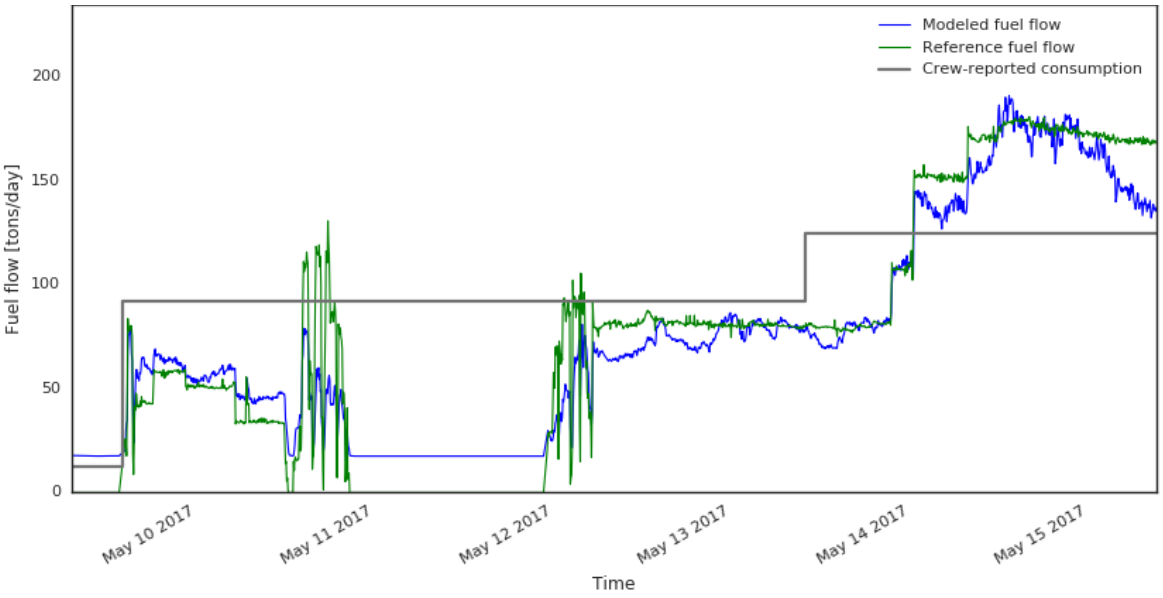


Figure 20. Detailed fuel flow comparison (2): the crew-reported consumption remains unchanged at ca. 90 tons/day for three consecutive days. Meanwhile, the modeled fuel flow provides stakeholders with a higher level of insights and reveals varying consumption levels between ca. 20 tons/day and 130 tons/day.

Figure 18, Figure 19 and Figure 20 illustrate the different sampling periods. The low-frequency noon reports (61 observations) are shown in grey as a crude step-wise line. Especially Figure 20 visualizes how much information is lost by utilizing only one value per day. The crew is reporting a constant consumption of roughly 95 tons/day over three consecutive days and fails to report what appears to be a port stay between May 11th and 12th. In contrast, the high-frequency modeled fuel flow in blue (17,151 observations)

captures the temporal variance in the fuel flow accurately. Overall, the modeled fuel flow (blue) correlates nicely with the reference data (green), however it often underestimates the reference consumption. Nevertheless, the modeled estimates provide a much more accurate and detailed view of the true consumption than the crew-reported values.

In a second step, we compare the modeled fuel flow based on Equation (13) with the fuel flow predictions from the lightweight system on the case vessel. The latter applies Bayesian regression techniques and is regularized with priors (known properties of the vessel, such as the mean draft or maximum possible speed) (Antola, Solonen, & Staboulis, 2017). The priors prevent overfitting, counter the effects of poor crew reporting, and prevent the model estimates from taking unrealistic values; a problem that arises especially when the system is new and the amount of crew-reported data available for training is low. Over time, the amount (and quality) of the crew reports tends to increase and the relative weight of the priors decreases. The exact logic behind the priors and the regularization technique cannot be disclosed in this thesis. However, the key take-away is that the regularization effect is low when a sufficient number of noon reports is available for training. In the context of this study, a sufficiently large training data set was ensured (see Chapter 4.2). Hence, I introduced an unregularized regression model in Chapter 3.3, which is expected to produce similar results than the more complex, regularized Bayesian regression model of Antola, Solonen, and Staboulis (2017). Nevertheless, a visual inspection of the fuel flow predictions from the (simplified) model based on Equation (13), and the real lightweight system as installed on the case vessel is necessary to show that the former can be used to validate the latter. Figure 21 and Figure 22 plot the modeled fuel flow, the values from the lightweight system on the case vessel, and the crew-reported values over the sampling period. The estimates from the (simplified) model based on Equation (13) (blue) correlate well with the lightweight system on the case vessel (yellow). Equation (13) appears to approximate the predictions of the real system (as installed on the case vessel) with reasonable accuracy.

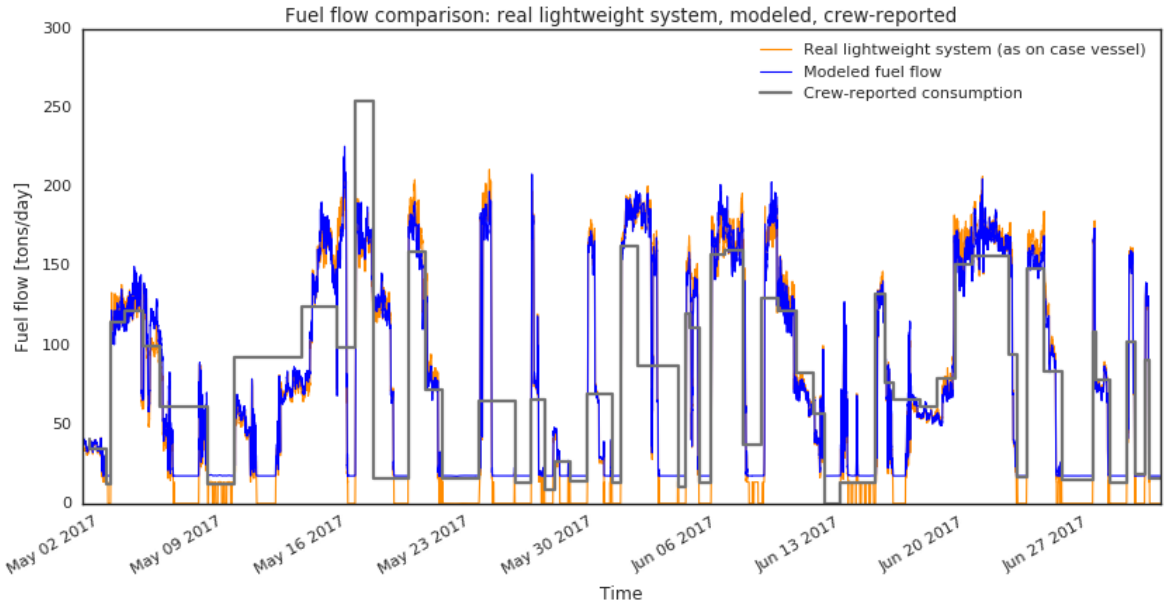


Figure 21. Fuel flow comparison: modelled fuel flow vs. readings from case vessel: The (simplified) model based on Equation (13) correlates well with the real system.

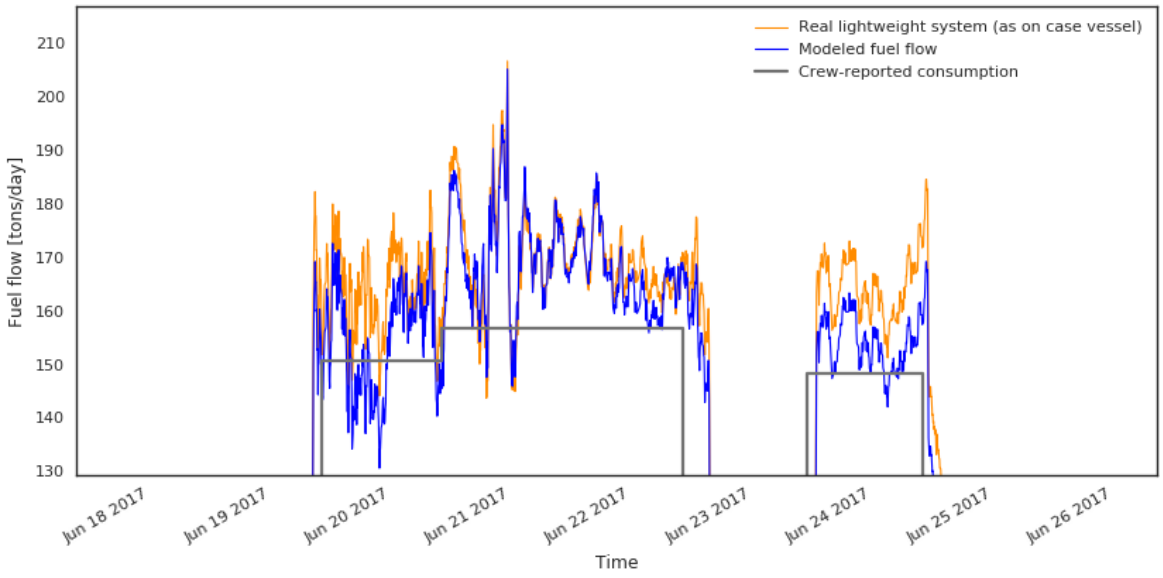


Figure 22. Detailed fuel flow comparison (3): modeled fuel flow vs. estimates from the lightweight system on the case vessel.

As a third step, we compare the previously discussed findings with the results of Antola, Solonen, and Staboulis (2017). Figure 23 plots the fuel flow predictions (green) against the reference data (pink) and the crew-reported consumption (black). In accordance with my own findings from Figure 18, the modeled output correlates well with the reference data. As expected, the former shows more random fluctuation, but still offers significantly more insight than the mere noon reports. My own findings are in accordance with the results

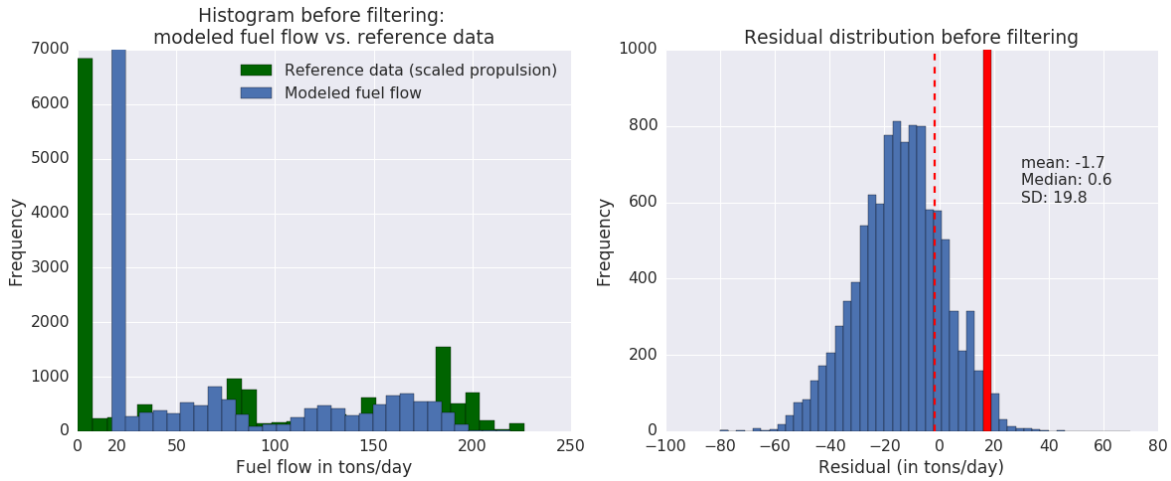
of Antola, Solonen, and Staboulis (2017); the model based on Equations (10) and (13) can be used to validate the underlying modeling approach of the lightweight system.

Figure 23. Fuel flow comparison by Antola et. al. (2017): the findings concur with the fuel flow model based on Equations (10) and (13).

In summary, a visual inspection of the fuel flow values suggests that the model based on Equations (10) and (13) provides reasonably accurate fuel flow estimates that correlate well with the reference data. The model output is in accordance with the readings from the real lightweight system (as installed on the case vessel) and with the findings of Antola, Solonen, and Staboulis (2017). We conclude that the model used in this study approximates the (more complex) real lightweight system well enough to be used for a transparent validation study.

5.4 Adjusting the Reference Data for Further Comparison

Recall from Chapter 4.4 that the reference data is technically not measured fuel flow, but a scaled propulsion power value. Despite the previous promising findings, Figure 18 also reveals an issue that arises from comparing modeled fuel flow with propulsion power: when the vessel is anchored at port, the propulsion power reference value (green) drops to zero, since the ship is not moving. The auxiliary engines, however, keep running, thus the fuel flow (crew-reported and modeled) never drops below ca. 18 tons/day. Both curves are correct, but they can only be used for direct comparison when the ship is at sea; not during port stay. Figure 24 visualizes the issue by plotting the distribution of the predicted and the reference values (and their residuals) in a histogram.



a) Histogram before filtering: modeled fuel flow vs. reference data

b) Residual distribution before filtering

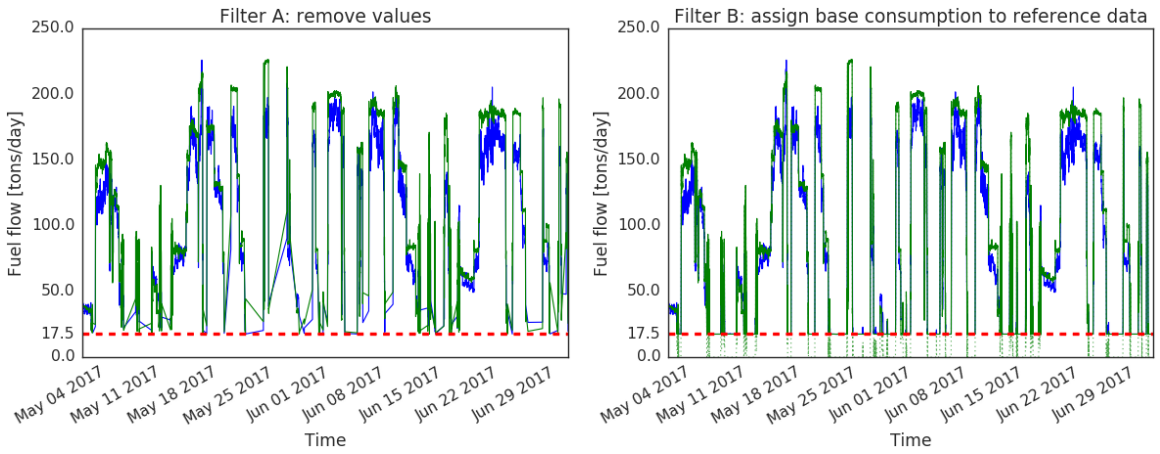
Figure 24. Visualizing the need for preprocessing the reference data: in panel a), the modeled fuel flow never drops below ca. 18 tons/day. The reference data is technically a scaled propulsion power measure and therefore shows (correctly) zero when the ship is anchored and not moving. The error distribution in panel b) shows that there is a large error bin at around 18 tons/day. Hence, we need to adjust the data to allow for a fair comparison.

In panel a) of Figure 24, we see how out of the 175,679 observations in the reference data set (green), roughly 7,000 are (close to) zero, meaning zero propulsion power is generated. At the same time, the predicted fuel flow (blue) shows around 7000 values at ca. 18 tons/day and *none* below. Even if we built a model that could predict the fuel flow perfectly (at sea), it would be off every time the ship is at a port; roughly 40% of time. By plotting the error distribution of the predicted values (compared to the reference data) in Figure 24 panel b), we observe a large bin around 18 tons/day at the right end of the distribution. This bin is

equivalent to the large (blue) bin in panel a) and together, they capture the systematic difference in what values the predicted fuel flow and the reference data show during a port stay. The histogram in panel b) also visualizes how the large bin distorts the residual: despite the distribution being centered at around ca. -18, we observe a mean of only -1.7. The large bin of systematic errors (ca. 7000 as seen in panel a) of Figure 24) incorrectly improves the error mean. The large standard deviation of 19.8 confirms the existence of a substantial error in the comparison.

If we ignored these signs and moved on to quantifying the error between the modeled fuel flow and the reference data, the model would perform badly. It would show a Mean Absolute Error (MAE) of 17.3, a Mean Squared Error (MSE) of 395.9, and a Root Mean Squared Error (RMSE) of 19.9; the Mean Absolute Percentage Error (MAPE) could not be calculated due to NULL values. For a fair assessment, we need to adjust the data for the systematic error that occurs during every port stay. More specifically, we need to treat the base consumption of the auxiliary engines as a threshold: the modeled fuel consumption will never drop below ca. 18 tons/day and the reference data cannot be used for fair comparison below this value. Hence, this value can be used to define a filter.

My first approach was to remove all entries from the reference data for which `fuelflow_ref` was below 18 tons/day. Figure 25 panel a) visualizes the approach. A cut-off point was defined and all values below were removed. This approach drastically shrank the data set and created data gaps. The model performance compared to the reference data was actually *decreased*.



a) Remove values below filter

b) Assign filter value to 'fuelflow_ref'

Figure 25. Comparison of different filtering approaches: in panel a), every entry below the filter was removed from the reference data. In panel b), every entry below the filter was assigned the filter value.

Consequently, I changed my approach: instead of removing values, I assigned the base consumption of ca. 18 tons/day to every entry in the reference value below this threshold. The result is visualized in panel b). The dotted green values were replaced with the filter value (red). This approach leaves the data set intact, creates no gaps, and ensures that a reference value is available for each predicted fuel flow value. One concern might be that the fit of the curves was artificially improved, by assigning values to `fuelflow_ref` that perfectly match `fuelflow_pred`. However, if the case vessel Boaty McBoatface was equipped with a physical fuel flow meter (so that we did not need to use scaled propulsion power) the graph would look similarly to panel b); and probably not like panel a). The reference curve from a physical fuel flow meter would show values around the base consumption during a port stay. Thus, I believe that my approach is a reasonable way of ensuring a fair comparison between the predicted and the reference values. With the systematic error removed, we plot the error distribution again.

Figure 26 visualizes the positive effect of the data filtering on the residual distribution. The bin containing the base consumption values is now at zero, since there is no difference between the predictions and the reference values whenever the ship is at a port. The overall distribution remains and is now more accurately captured by a mean value of -8.7; the standard deviation has been reduced from 19.8 to 13.9. We compute the error measures again and notice a significant improvement in the model performance: the MAE improved from 17.3 to 10.2, the MSE dropped from 395.9 to 270.2, the RMSE improved from 19.9 to 16.4, and the previously undefined MAPE is now around 9.6%.

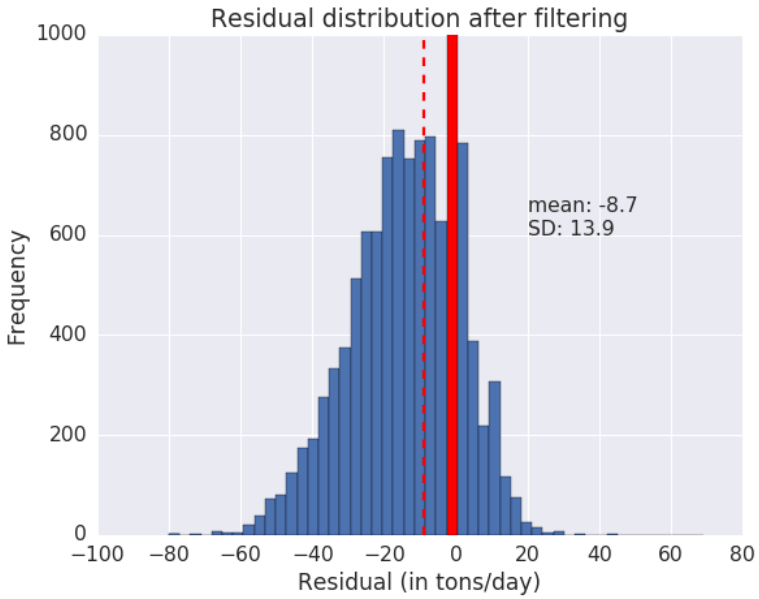


Figure 26. Residual distribution after filter: note the large red bin which is now centered around zero.

The regression plot in Figure 27 visualizes the predicted fuel flow (`fuelflow_pred`) regressed onto the filtered reference data (`fuelflow_ref_filtered`). As Figure 18 already suggested, the values correlate nearly perfectly (r = 0.99). The data points show a clear linear trend and a reasonable level of dispersion. However, the data points, and with them the regression line, is slight shifted upwards of the identity line. This shift confirms our previous impression that the modeled fuel flow generally underestimates the true consumption. Since the model is trained on the aggregated sample data (Equation (10)), more specifically on the crew-reported total consumption over a known time span, an underestimated fuel flow might occur due to an underreported consumption in the noon reports. Note, however, that the offset might also (partially) occur due to an unknown bias in the reference data (see Chapter 4.4).

Figure 27. Regression plot: predicted fuel flow vs. reference data. We observe a clear linear trend and a reasonable dispersion of the data points. The regression line is slightly shifted upwards of the identity line, indicating that the model (Equation (13)) is underestimating the true fuel consumption.

5.5 Out-of-Sample Forecasting

In addition to the regression plot in Figure 27, we can analyze the effect on the accuracy of the modeled fuel flow when performing **out-of-sample forecasting**. In Chapter 5.1 we discussed different nested model variations in isolation, that is the fit of reduced models to the training data. Figure 28 visualizes the effect of using these reduced models to predict fuel flow. In accordance with the regression plot in Figure 27, the fuel flow predictions for different model variations are regressed onto the (filtered) reference data.

Overall, the regression plots do not differ significantly from each other. The full model (orange) exhibits the best predictive performance, measured by the MAE, MSE, and RMSE. The “water only” variation that was shown to be equivalent to the full model based on the F-test shows the highest error when compared to the reference data. Figure 28 confirms the previous impression that the vessel speed, captured by the water predictor, is by far the most important factor when predicating fuel flow. The remaining fine-tune the model and improve the performance slightly. Note also that all model variations tend to underestimate the consumption as well.

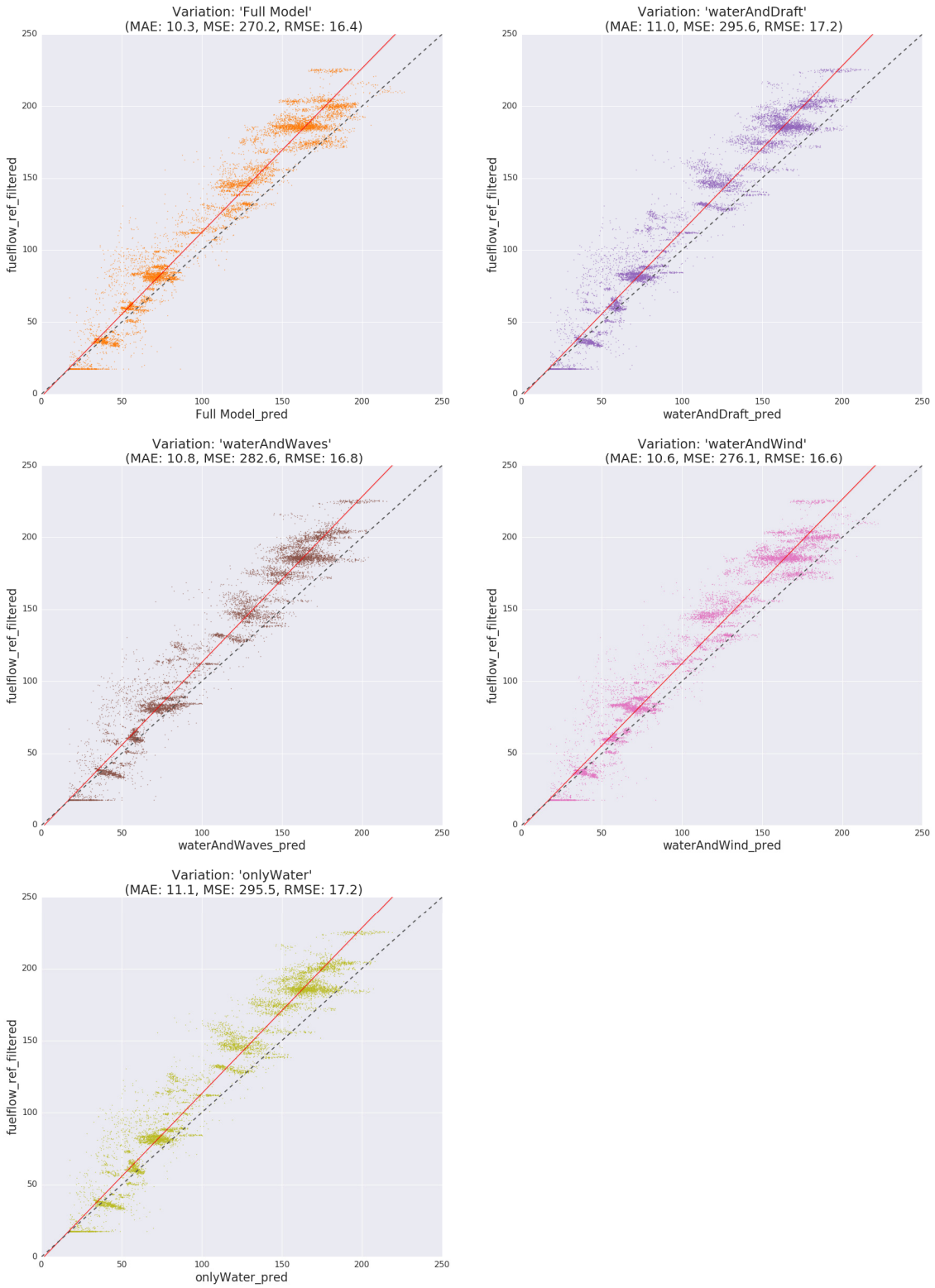


Figure 28. Regression plots: out-of-sample forecasting. The full model (orange) is visually compared to the nested variations.

In summary, the original model based on Equation (13) is able to predict instantaneous fuel flow rates with reasonable accuracy, by combining the vessel speed with meteorological forecasts and crew-reported changes in tank levels. **Despite a lack of onboard integration, the model deviates on average by only 9.6% (MAPE) from the integration-based system; about 10 tons per day (MAE). The modeled instantaneous fuel flow correlates nearly perfectly with the reference data, but tends to underestimate the fuel consumption.** These findings conclude the third stage of my research method. I have shown how to assess model performance by comparing the fuel flow predictions against measured reference data. The next logical step is to study the reason for the offset in the predictions. We will therefore continue our analysis by taking a closer look at the human input in form of the noon reports.

5.6 Effects of Human Reporting Errors

Aside from model accuracy and systematic modelling errors, the third potential problem that Antola, Solonen, and Staboulis (2017) identified in their approach is the reliance on human input. In Chapter 3.3 I discussed how the model is fitted to the total daily fuel consumption (Equation (10)); a quantity that has to be manually checked and reported by the crew. Since the lightweight system is marketed as a monitoring tool that enables the charter to monitor crew and vessel performance, it is essential to understand the extent to which the crew can influence the model predictions; either by carelessly or even wilfully reporting incorrect values. The fourth and final stage of my research method is therefore to assess the effect of human reporting error by simulating error-free noon reports from reference data.

My approach will be similar to a sensitivity analysis: the study of how uncertainty in the model output can be attributed to different sources of uncertainty in the inputs (Saltelli et al., 2008). In order to evaluate how much uncertainty in the modelled fuel flow can be attributed to human errors, one has to alter the crew-reported consumption, while keeping the other variables constant. Any change in the modelled fuel flow can then be attributed to the human errors.

In practice, the noon reports will always contain an unknown level of errors. Long (irregular) reporting periods, the manual taking of fuel flow readings, and general human imperfection make the noon reports inherently prone to errors. Moreover, we cannot determine how much of the model variance can be apportioned to human errors, when fitting a model to the crew-reported consumption (Equation (10)). A bad model fit (low R^2) might also occur because the explanatory variables fail to approximate the system adequately.

We overcome this problem by simulating error-free noon reports from the reference data. First, we resample the (filtered) reference fuel flow to the irregular timespans of the noon reports. The Pandas `DataFrame.resample()` function offers a convenient tool to resample time series with *regular* intervals (McKinney, 2012). The function is, however, not suitable for resampling data to irregular timespans. Instead, a helper function `simulate_noon_report()` is defined that sums up some input time series (time-stamp-indexed) of one data frame with respect to the irregular timespans of another data frame.

```
def simulate_noon_reports(df, spanned_ref, column='fuelflow'):
    """
    This function simulates noon-report fuel flow data by summing up the input series
    with respect to the time windows present in the reference time span data.
    """
    out = spanned_ref.copy()
    l = []
    index = spanned_ref.index
    begin_times = index.levels[0]
    end_times = index.levels[1]
    for start_time, end_time in zip(begin_times, end_times):
        mask = (df.index > start_time) & (df.index <= end_time)
        delta = end_time - start_time
        days = delta.total_seconds() / 86400.
        s = df.loc[mask]
        l.append(s.sum() / 1.2e05 / days / 24 * 10000)  # unit conversion: kg / h -> kg / (30 s) -> t / d
    out[column] = l
    return out
```

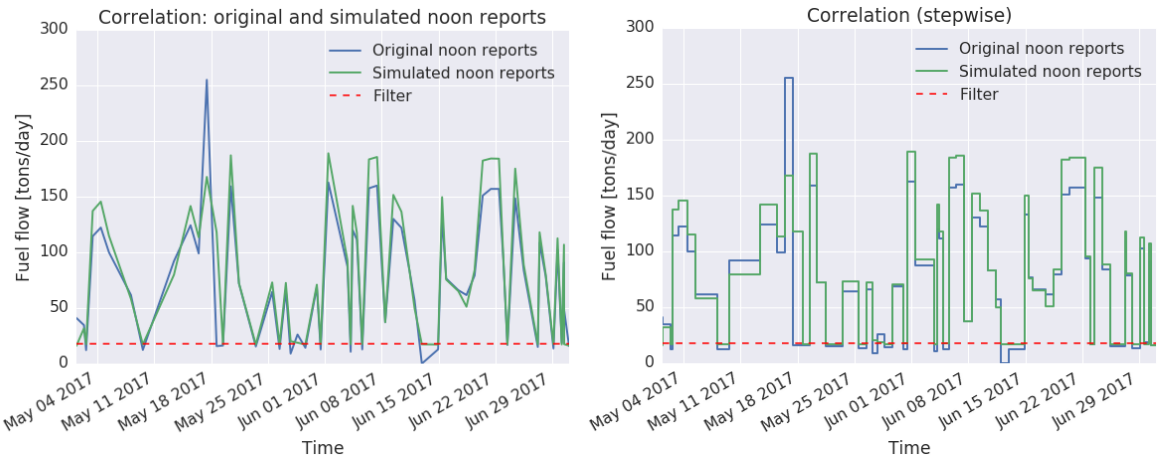
Figure 29. Function for resampling time series data to irregular time spans.

The filtered fuel flow reference data is resampled to the irregular time spans of the 'reports' data frame; the outcome is stored in a new data frame 'reports_sim'.

	fuelflow_ manual	fuelflow_ref_ spanned	stw_ manual	stw_ spanned	draft_ manual	water_ spanned	draft_ spanned	wave_ spanned	wind_ spanned
end									
2017-05-01 10:00:00	41.00	16.09	NaN	10.62	11.25	1196.70	333.17	15.25	2689.70
2017-05-02 08:12:00	34.60	31.70	10.71	10.47	11.25	1146.80	319.28	13.90	-1015.60
2017-05-02 14:30:00	12.14	17.36	10.50	0.57	11.25	0.18	0.05	0.70	1.12
2017-05-03 10:00:00	114.46	137.20	NaN	18.89	10.40	6741.10	1225.65	58.55	-2713.70
2017-05-04 10:00:00	122.25	145.58	18.82	18.46	10.40	6287.30	1143.15	103.98	6127.90

Figure 30. The 'reports_sim' data frame: the error-free noon reports that were simulated from the reference data are stored in a new column 'fuelflow_ref_spanned'. Note how the '_spanned' suffix indicates that this former high-frequency variable has now been aggregated over a certain reporting period. Note further how the irregular timespans from the original noon reports were preserved.

The simulated, error-free noon reports are labeled as `fuelflow_ref_spanned`, to indicate what they actually are: reference time series data aggregated to a time span format. Already in the first five rows we can observe a discrepancy in the values of the original crew-reported consumption (`fuelflow_manual`) and the simulated one (`fuelflow_ref_spanned`). To ensure that the approach of simulating error-free reports from the reference data is trustworthy, we visualize the correlation between the original and the simulated values over the sampling period.



a) Smooth draw style b) Stepwise draw style

Figure 31. Correlation of crew-reported and simulated reports: the stepwise draw style in panel b) visualizes the irregular timespans.

Figure 31 shows a nearly perfect correlation between the two curves. While they deviate from each other at times, this discrepancy reflects the type of human error we wish to analyze further. We conclude that the function `simulate_noon_reports()` reasonably converts the high-frequency reference data into a form that corresponds to the low-frequency noon report data.

The fourth stage of my research method was to assess if, when, and how the crew fails to report the true consumption. In Figure 32 we address this objective by plotting the errors between the crew-reported and the simulated reports over the sampling period. The simulated values provide an error-free baseline.

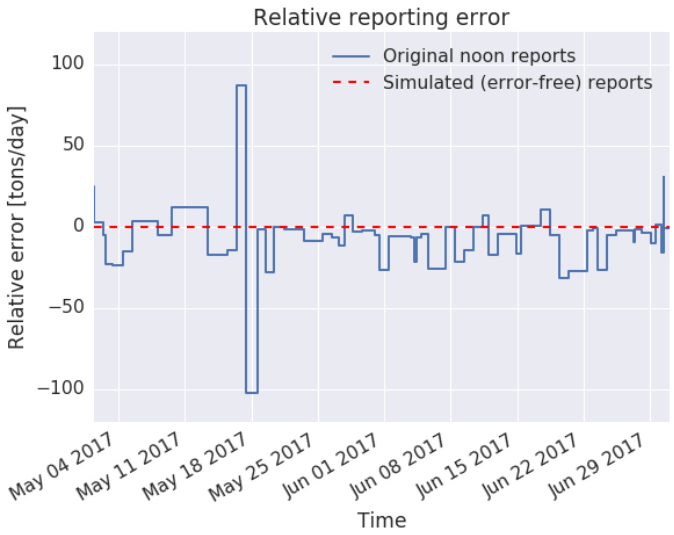


Figure 32. Errors between the crew-reported and the simulated reports: the error-free simulations provide a baseline that enables us to see how much the crew over- or underreports the fuel consumption.

On a first glance, we observe two consecutive instances around May 18th, when the crew was drastically over- and then underreporting the consumption. More important, however, is the ongoing trend of *underreporting* the true consumption over the entire period; the blue line runs mostly below the dotted reference line. While the two drastic deviations could be regarded as outliers that seem to balance each other out, the less drastic but continuous trend of reporting a too low consumption has a more severe impact. It causes the model to also predict too low instantaneous fuel flow rates as we saw in Figure 18 and Figure 27. These findings might be troublesome for the customer as they clearly indicate that the crew can influence the model output.

We proceed by fitting a model based on Equation (10) onto the simulated reports (the dependent variable is `fuelfow_ref_spanned`). Figure 33 shows the regression result.

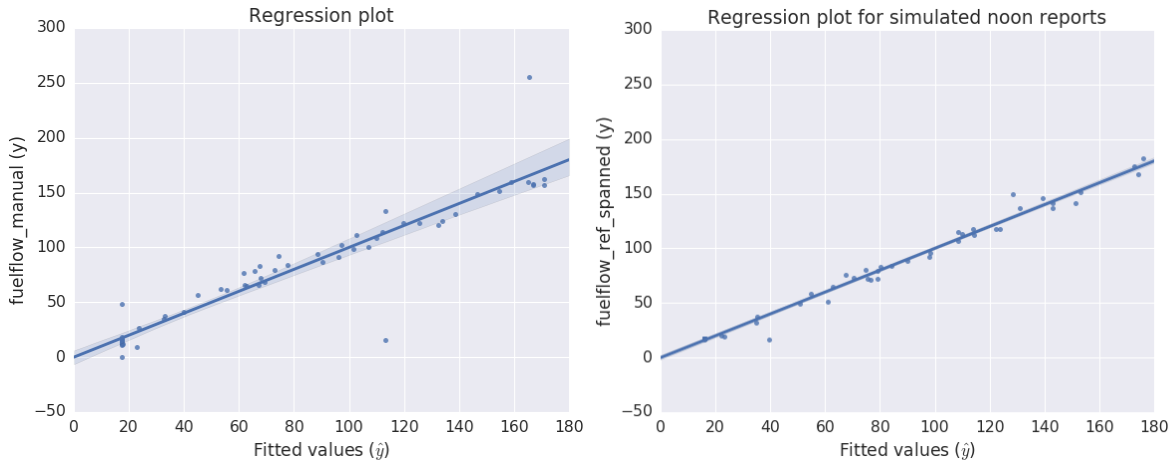
OLS Regression Results

Dep. Variable:	fuelflow_ref_spanned	R-squared:	0.991
Model:	OLS	Adj. R-squared:	0.991
Method:	Least Squares	F-statistic:	1581.
Date:	Mon, 09 Oct 2017	Prob (F-statistic):	7.43e-57
Time:	18:14:38	Log-Likelihood:	-191.40
No. Observations:	61	AIC:	392.8
Df Residuals:	56	BIC:	403.3
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
Intercept	15.7931	1.145	13.795	0.000	13.500	18.086
water_spanned	0.0169	0.000	34.865	0.000	0.016	0.018
draft_spanned	-0.0040	0.002	-2.182	0.033	-0.008	-0.000
wave_spanned	0.1468	0.043	3.395	0.001	0.060	0.233
wind_spanned	0.0010	0.000	5.023	0.000	0.001	0.001

Omnibus:	20.125	Durbin-Watson:	1.495
Prob(Omnibus):	0.000	Jarque-Bera (JB):	104.143
Skew:	-0.504	Prob(JB):	2.43e-23
Kurtosis:	9.321	Cond. No.	9.17e+03

Figure 33. OLS regression results based on simulated noon reports: the dependent variable is now 'fuelflow_ref_spanned'. Note the perfect model fit (R^2=99%).



a) Model fitted to original noon reports (Figure 15). b) Model fitted to simulated noon reports.

Figure 34. Comparison of regression plots: after removing the human errors from the noon reports, no more outliers occur in the regression plot in panel b).

Compared to the regression results from the first stage of my research method (see Figure 15), the RSE has improved from 19.8 to 5.8; the absolute lack of fit of the model to the data has been reduced. Moreover, R^2 has improved from 0.882 to 0.991, indicating that the model can now explain 99% of the variance in the (simulated) noon reports. Removing the human reporting errors has significantly increased the model fit and the four predictors (water, draft, waves, wind) explain now nearly perfectly the fuel consumption. The improved model fit is visualized in the regression plot in Figure 34 b). **We conclude that virtually all of the unexplained variance in the original model was due to the human errors.**

We proceed by predicting instantaneous fuel flow from the estimated weight parameters based on Equation (13). In Figure 35 and Figure 36, the simulation-based modeled fuel flow is plotted against the original predictions and the reference data.

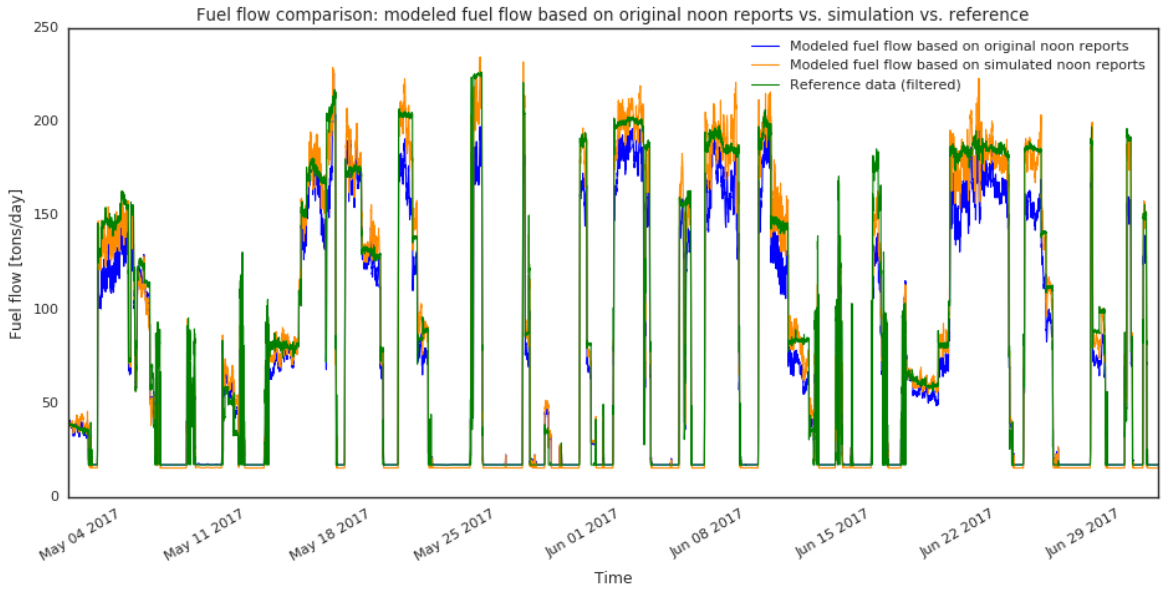


Figure 35. Fuel flow comparison: output based on noon reports vs. simulation. In contrast to the original model, the fuel flow estimates based on the simulated, error-free noon reports are centered around the reference data. Nevertheless, they still show considerable noise.

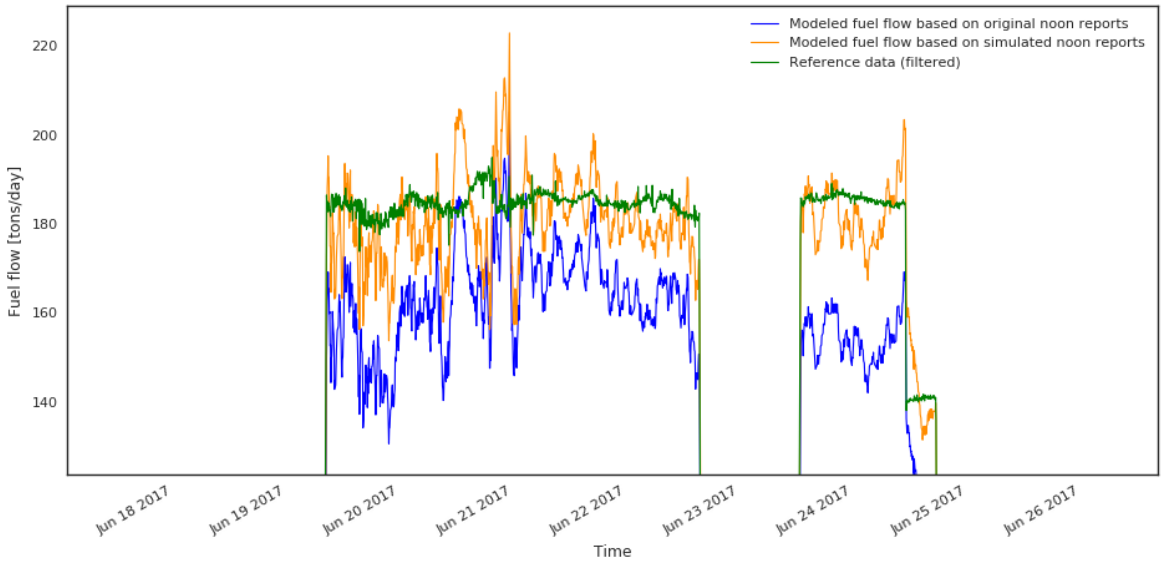


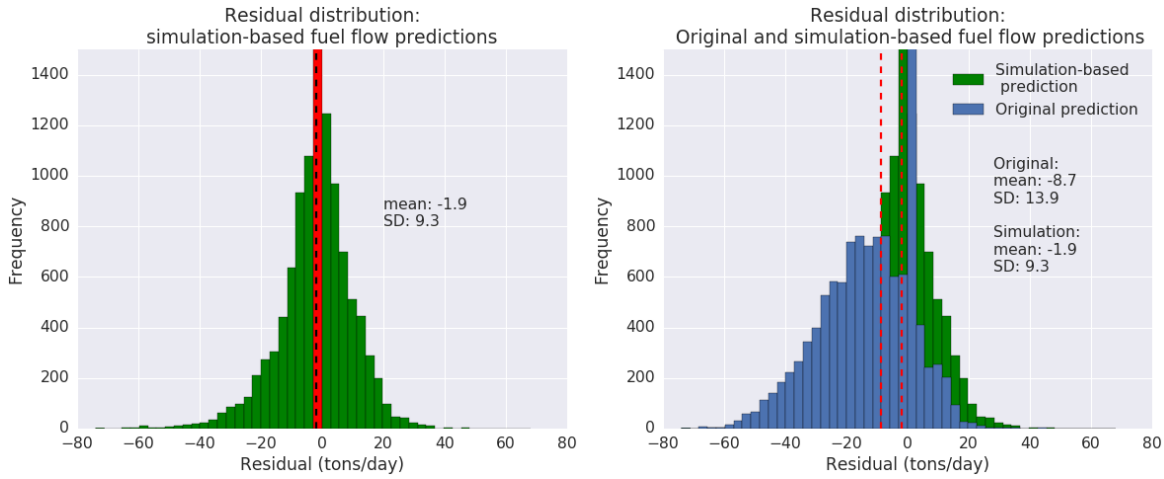
Figure 36. Detailed fuel flow comparison with simulation.

The modeled fuel flow based on the simulated noon reports (yellow) is overall more centered around the reference data (green) than the predictions based on the original noon reports (blue). This is of little surprise, given that the model behind the yellow curve was trained on fuel flow aggregates from the reference data. We also observe that the simulation curve exhibits nearly as much – if not more – erratic behavior than the one based on the crew input. **We conclude that removing the human errors seems to have no effect on the variance in the predictions.** The volatility comes most likely from noisy forecast-STW values that

propagate throughout the model, due to their cubic relationship to the fuel flow. Recall that we saw in Chapter 4.3 how the forecast-STW (Equation (2)) provides a better speed proxy than the unadjusted SOG. However, the forecast-STW exhibits still more variance than the reference STW (measurement-based, obtained from integration-based system).

As in Chapter 5.4, we compute several error terms: by using the error-free reports, the MAE has further dropped from 10.3 to 6.0, the MSE went from 10.2 to 6.0, and the RMSE decreased from 16.4 to 9.5. As expected, the simulation-based model performs better.

We combine the implications from Figure 32 and Figure 35 to derive another interesting observation: the overall trend of underreporting the true consumption appears to be reason why the modeled fuel flow (blue) runs mostly below the reference curve. Removing the human errors removes this offset and centers the predictions around the true values. **This finding strongly indicates that the crew can indeed influence the predicted fuel flow**. Figure 37 panel a) shows the residual distribution for the simulation-based predictions; panel b) compares the residuals from the simulation-based model against the model that was trained on the original (crew-generated) noon reports (see Figure 26).



a) Residual distribution: simulation

b) Residual distribution: original vs. simulation

Figure 37. Residual distributions after simulation: the simulation-based error distribution (green) is nearly perfectly centered around zero. The original predictions (blue) are shifted to the left, indicating an overall trend of underreporting and underestimating the true consumption.

We observe an error distribution for the simulation-based predictions that is almost perfectly centered around zero, while the predictions based on the crew-reported values are shifted to the left. Removing the human error reduces also the standard deviation.

In summary, the model appears to be somewhat sensitive to human reporting errors. **Occasional reporting errors, even severe ones, do not significantly reduce the model**

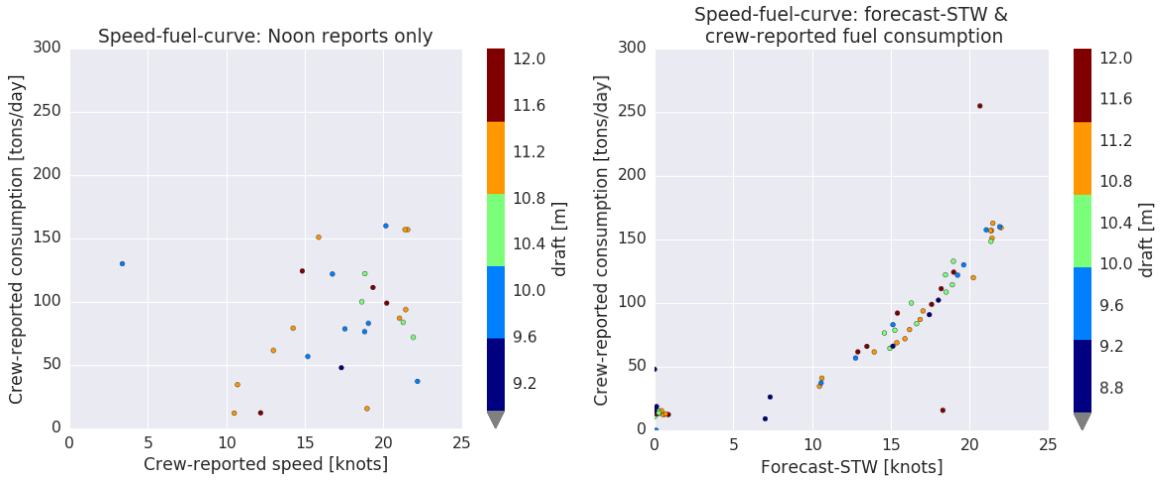
performance. On the other hand, a less severe but continuous trend of underreporting the true consumption causes the model to also underestimate the instantaneous fuel flow rates. The human errors seem to create a bias in the predictions, in the form of a general (non-constant) offset. In contrast, the variance or erratic behavior of the predictions are not significantly affected. Moreover, the four predictors (water, draft, waves, wind) are in able to approximate the noon reports (and also instantaneous fuel flowrate) surprisingly accurately. If the human reporting errors are removed, the model captures the reports' variance perfectly (R^2 = 99%), despite using a simplified propulsion power model and only four independent variables. The model fit decreases once human reporting errors occur in the training data.

5.7 Assessing Customer Value with Speed-Fuel-Curves

Speed-fuel curves are used in the marine industry as a standard tool to assess a vessel's fuel consumption at different speeds. Since the power requirement and fuel consumption are roughly equal to the speed-cubed ($X_1 = water = STW^3$), a slight change in the speed will have a large effect on the fuel consumption. As we saw in Chapter 3.2, the consumption is further influenced by the draft. Therefore, speed-fuel curves usually have different curves to show the consumption at different speeds and at different draft levels. In the following, I will present the results of my study in the form of speed-fuel curves.

Figure 38 a) plots the crew-reported speed (`stw_manual`) against the crew-reported consumption (`fuelflow_manual`). The colors indicate the crew-reported draft levels (`draft_manual`). We observe a loose scatterplot of points without any clear trend. Over two months, the crew reported a meager 27 (out of potentially 61) values for the vessel speed. A stakeholder would find it impossible to draw any meaningful conclusion based on this data. Furthermore, these findings cast doubt on any approach that relies solely on data from the noon reports, for example the statistical approach "on the estimation of a ship's fuel consumption and speed curve" by Bialystocki and Konovessis (2016). A longer time horizon (one year in the aforementioned study) might mitigate the crew's poor reporting practices; however, at least within the two-month sampling period of my study, poor reporting obstructs any meaningful insights for the stakeholders.

In Figure 38 panel b), the crew-reported speed has been replaced with the forecast-STW (`stw_spanned`). With a measurement-based speed variable, the points already show expected behavior as we start to observe an upwards-sloping trend. Yet, especially for lower speeds around 10 knots, few data points from the aggregated sample data (61 over the two-months period) make it difficult to clearly determine the consumption. Moreover, it is impossible to determine any draft effect as a result of differences in the cargo. The installation of a GPS unit, even if its readings are adjusted for the ocean currents (Equation (2)), is not enough to provide a level of insight that could lower the information asymmetry between the charterer and the ship-owner.



a) Manual data only

b) Crew-reported consumption and STW

Figure 38. Speed-fuel-curves for manual data: the data from the noon reports alone (a) forms a loose scatterplot of points that provide virtually no information to stakeholders. Combining the crew-reported consumption with the forecast-STW (b) improves the plot and the points start to exhibit a pattern. However, the scarce data points from the low-frequency data are still an issue.

In contrast, a virtual fuel flow model provides significantly more customer value. Figure 39 visualizes the speed-fuel curve that is based on the predictions from the model based on Equation (13) (trained on the crew-generated noon reports). The 17,151 observations form a neat curve that highlights the cubic increase in fuel consumption at higher speeds. We observe consumption values for the entire range of speeds. The points begin to disperse more at higher speeds, however, this effect might be explained by the draft: the draft-induced resistance (X_2) affects the power requirement more at higher speeds. Thus, the consumption will differ especially at higher speeds depending on the weight of the cargo. This modeling-based fact was not available in the previous figure; it requires high-frequency values. Moreover, Figure 39 gives some indication on the design draft of the vessel, which is the draft (in meters) for which the vessel is designed and at which it will show optimal performance. While we would expect the consumption to increase at higher draft levels, we observe the lowest consumption at a draft of 10.0 meters to 10.6 (at speeds higher than 20 knots). Higher and lower draft levels will both increase consumption. This information, based on empirical data on the vessel's actual consumption, might help stakeholders to choose an optimal draft level for their vessels.

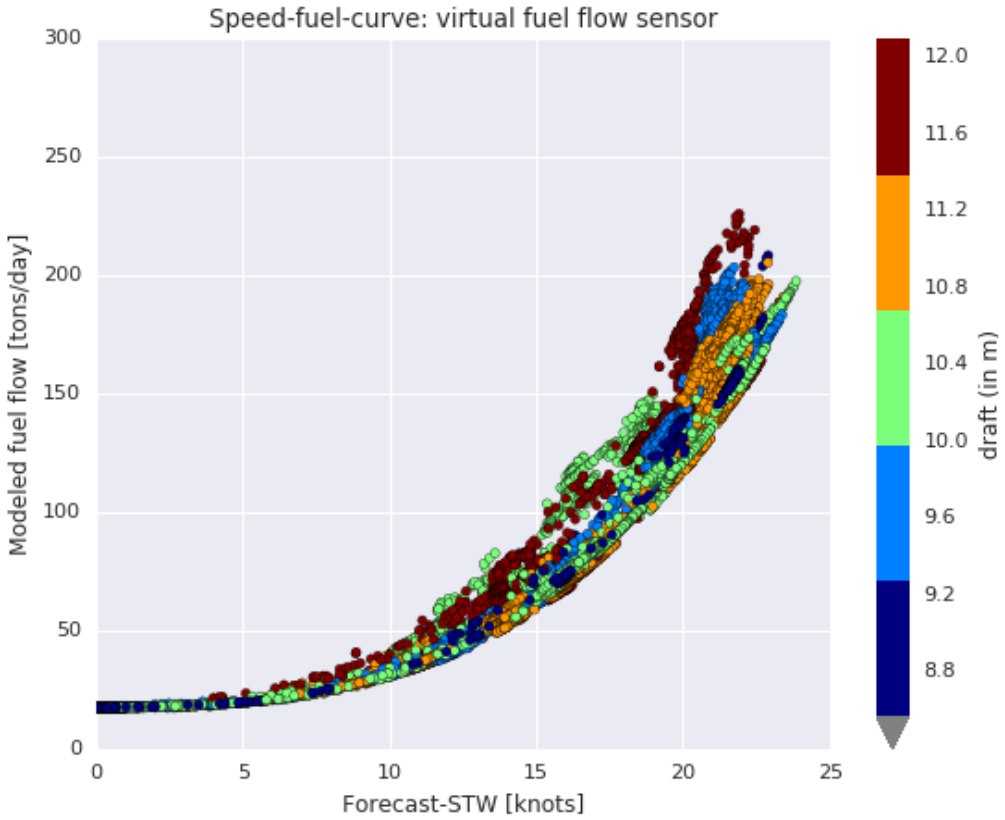


Figure 39. Speed-fuel-curve based on original noon reports (Equation (10)): the high-frequency data forms a detailed and neat curve. The model allows stakeholders to demonstrate the fuel consumption of their vessels across their entire operational profile.

As a final step, we plot the speed-fuel curve for the modeled fuel flow, based on the simulated error-free reports. The curve in Figure 40 represents the optimal model output, given that all human errors are removed from the training data. Especially at higher speeds, the data points are less dispersed than in Figure 39. More important, however, is that the curve in Figure 40 does not differ significantly from the one in Figure 39. While the latter shows a curve that can actually be achieved by fitting a model to real life data, the former visualizes an unrealistic case that is based on reference data. The model that was fitted onto the real noon reports provides sufficient insights into a vessel's operational profile to aid stakeholders in demonstrating the operational profile of their ships.

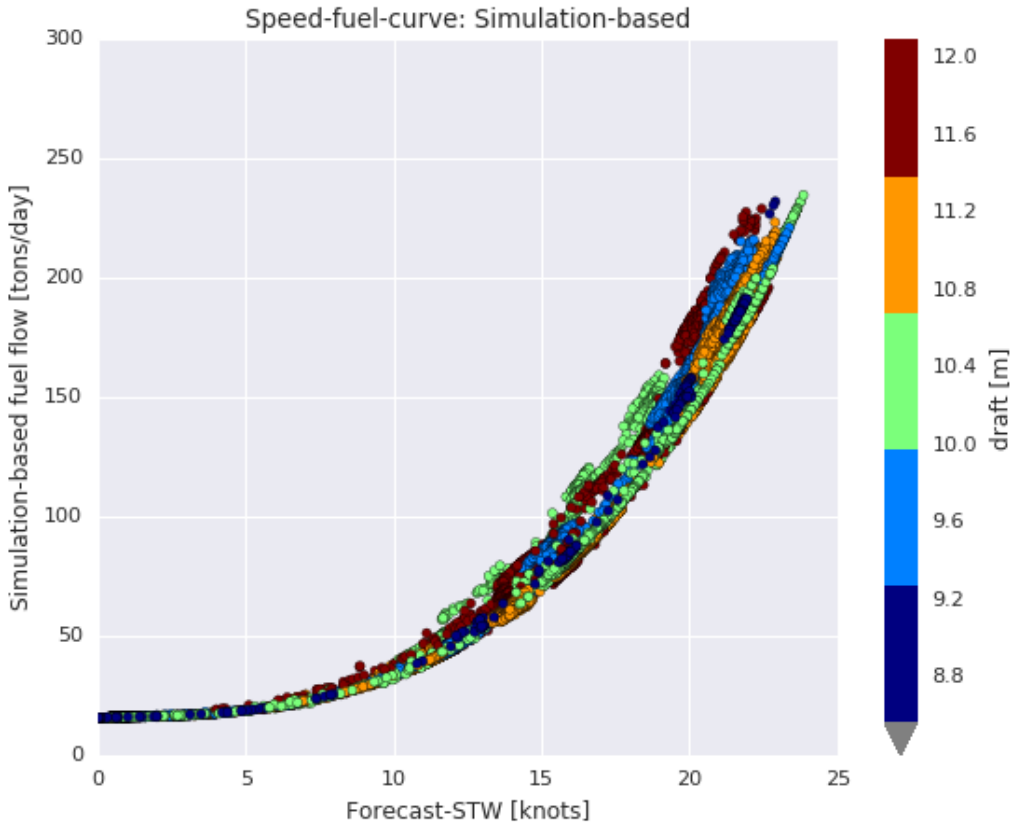


Figure 40. Speed-fuel curve based on simulated noon reports: the points show less dispersion at higher speeds. Yet, the plot does not differ significantly from the previous one (which was based on the crew-generated noon reports).

5.8 Summary of Results

In summary, the lightweight system provides reasonably accurate speed and fuel flow estimates, without the need for expensive onboard integration. The system allows ship owners and operators to demonstrate the fuel consumption of their vessels across their entire operational profile, especially for different speeds and draft levels (Trodden et al., 2015). Furthermore, it allows stakeholders to comply with increasingly stringent regulations, e.g the Annex IV regulations of the IMO (International Maritime Organization, 2008). The system provides sufficient insight to reduces the information asymmetry between the ship owner and the charter, and allows the latter to identify unfavorable operational practices. Furthermore, the lightweight system could be used by ship owners as inexpensive tool to validate the effect of certain fuel-efficient measures, e.g. a hull cleaning. The inexpensive hardware requirements of the lightweight system make it suitable for chartered vessels and prevent it from being subject to barriers to energy efficiency (Rehmatulla & Smith, 2015). While the lightweight system is overall less accurate and can be affected by erroneous crew-reporting, it is important to remember that for most customers, the alternative to the

lightweight system is not an integration-based solution, but a sole reliance on manual reports (due to the barriers to energy efficiency). The lightweight system offers significant customer value, at a fraction of the price of most commercially available, integration-based systems.

6 Conclusion

In this Master's thesis, I discussed the economic, environmental, and regulatory forces that lead stakeholders in the marine industry to seek ways to improve fuel efficiency of commercial vessels (Buhaug, 2009). Fuel-related expenses comprise the single largest operational cost item of a voyage. Moreover, commercial vessels contribute significantly to the world-wide CO_2 emission and face increasingly strict regulations. Eide et al. (2009) presents over 50 operational and technical methods that can lead to substantial fuel savings, and reduce CO_2 emission by 30% to 40%. Many of these measures are not only energy-efficient, but also cost-effective, meaning that they offer a positive net present value and will amortize over time. Yet, despite these savings potential, their implementation rates remain low across the industry. Rehmatulla and Smith (2015) conducted the first scientific study on shipping-specific barriers to energy efficiency, and found principal-agent-problems in the form of split incentives and information asymmetry to be the primary reason for the low diffusion rate of integration-based systems. Especially on chartered vessels, but also an small and medium-sized ships, these costly monitoring tools (Trodden et al., 2015) are unfeasible. Alternative approaches that rely solely on crew-reported data are inexpensive (Bialystocki & Konovessis, 2016), but fail to offer sufficient insight and accuracy to comply with regulatory standards. A novel approach based on virtual sensing techniques (Fortuna et al., 2007) has been proposed by Antola, Solonen, and Staboulis (2017). Their method can overcome the barriers to energy efficiency and offer some of the insights that require onboard integration in most commercially available systems. In this Master's thesis, I conducted a transparent validation study to highlight the advantages of the proposed approach by Antola, Solonen, and Staboulis (2017).

6.1 Research Summary

The research questions of my Master's thesis were "how to formulate a virtual sensor for time-dependent momentary fuel flow estimation on commercial vessels?" and "how well does the model perform in comparison to integration-based system?". Based on the design procedure for virtual sensors (Fortuna et al., 2007) and the work of Antola, Solonen, and Staboulis (2017), I divided my research method into four steps and structured my study accordingly.

From the literature on ship hydrodynamics, we saw that it makes sense to express fuel flow as a function of propulsion power, and that a ship's propulsion power requirement

depends on the resistances that the vessel faces while moving through water (Bertram, 2000; MAN Diesel & Turbo, 2011). In the variable selection stage, I therefore chose four explanatory variables that are known to account for a large proportion of a vessel's power requirement: frictional (water) resistance, draft-induced resistance, and the extra resistances due to wind and waves. I then formulated a regression model and fitted the low-frequency noon reports onto aggregates of the predictor variables to estimate their weight parameters. The model explained a large portion of the noon reports' variance (R^2 = 88%). It showed strong multicollinearity, however, this was expected as all predictors include the squared or cubed vessel speed.

In a second step, I applied the estimated weight parameters to high-frequency data in form of the unaggregated predictors (using the native five-minute sampling period of the GPS unit), in order to predict instantaneous fuel flow. While the crew had only reported 61 values for the fuel consumption, the instantaneous fuel flow model produced over 17,000 values over the same two-month sampling period.

In a third step, I introduced reference data from an integration-based system to assess the accuracy of the predictions. As no physical fuel flow meter was available on the case ship, I used a scaled propulsion power measurement instead. After some preprocessing, the scaled propulsion power provided a solid reference value for the subsequent comparison. Overall, the modeled fuel flow correlated nicely with the reference data. Yet, the modeled output was noisier and tended to underestimate the true consumption. Nevertheless, the model was able to capture much more of the system variance than the noon reports.

In a fourth step, I simulated error-free noon reports from the reference data in order to assess the level of human reporting error. In a few instances, the crew-reported values deviated significantly from the reference data. More importantly, the simulation revealed an overall trend of underreporting the true consumption. This observation could, however, also be partly caused by a bias in the reference data. After training the model on the simulated error-free reports and producing new fuel flow estimates, it became apparent that the model output is not majorly affected by occasional reporting errors; an ongoing (but less severe) trend of underreporting, however, causes the model to also underestimate the instantaneous fuel flow rates.

In summary, my research methodology provided a transparent way of validating the lightweight model by Antola, Solonen, and Staboulis (2017). Not only did my study confirm their findings, it also added more insight into the model performance by quantifying the absolute and relative deviation from the reference data. Hence, it became possible to state

that the lightweight model deviates on average by ca. 10% (MAPE) from the integration-based system (on the case vessel). Moreover, my study attributed the lack of fit in the model to the human reporting errors and found the model to be affected by on ongoing trend of over/underreporting. In the following, I will summarize the practical implications of my findings.

6.2 Practical Implications

My research provides several implications for stakeholders in the maritime industry. First of all, my validation study showed exemplarily that the novel approach by Antola, Solonen, and Staboulis (2017) provides reasonably accurate fuel flow estimates. For most customers, the alternative to the lightweight system is not an integration-based solution, but purely manual reports, due to the barriers identified by Rehmatulla and Smith (2015). At a fraction of the price of an integration-based system, a slightly worse model-based result is still significantly better than anything the customer has seen before or could see by solely relying on the noon reports.

Secondly, the model-based speed-fuel-curve shows a reasonable level of data point dispersion and allows the customer to observe a vessel's operational profile at different speeds and draft levels. The system could therefore be used to comply with increasingly strict regulatory requirements.

Furthermore, the lightweight system provides sufficient insight into the vessel and crew performance to serve the charterer as an independent monitoring tool. While the crew can to some extend influence the model output, the charterer would nevertheless be able to identify unfavorable operational practices that lead to unnecessarily high fuel bills, e.g. not keeping an even speed profile throughout the voyage. The lightweight system can reduce the information asymmetry between the ship-owner and the charterer and help the latter to (re-) negotiate more favorable charter chartering fares that reflect the true operational efficiency of a vessel.

6.3 Limitations of the Study

The limitations of this study are a logical result of the methodological choices I purposefully made to best suit the nature, the motivations, and the resources for this research. The limitations mainly concern the case study approach and the sample period.

It would have been beyond the scope of this Master's thesis to validate the approach of Antola, Solonen, and Staboulis (2017) across multiple vessels and vessel classes. Instead,

I chose a case study approach and utilized a single case vessel to exemplarily validate the lightweight model. The case vessel is a containership, with characteristics listed in Table 1. While my methodology can be applied to other vessel types, and could even been used for a fleet-wide study, the findings of this particular study are not universally applicable without further validation. The bigger the difference between a vessel and the case vessel, the higher the need to repeat the study. This holds especially for vessels other than containerships, such as bunkers, tankers, and LNG carriers, which show vastly different operational profiles.

In addition, a second limitation arises from the use of a scaled propulsion power measure for reference. With the scalar unknown, it is impossible to rule out a bias in the reference data.

Furthermore, the way in which the level of reporting error was determined poses a third limitation of this study. With only one case vessel, error-free reports had to be simulated from reference data, in order to obtain a value for comparison. However, it is unlikely that a crew could ever achieve this level of reporting accuracy. The single-case-vessel approach made it impossible to determine a realistic optimal (and achievable) reporting level.

Finally, a fourth limitation concerns the study's two-month sample period. While the high sampling frequency of the integration-based system (30 seconds) and of the lightweight system (five minutes) provides ample data, the regression model is still fitted onto the low-frequency noon reports; which only yield 61 observations for the two-month period. There is a probability that the reporting errors, which reduce the model fit and increase the spread in the fuel flow estimates, become less weighty over a larger time horizon. While the real lightweight system uses Bayesian regression and is regularized with priors to mitigate this effect, the (simplified) model used in this study is more prone to a lack of training data.

Nevertheless, I believe that the promising findings of my study are scientifically sound and offer significant value to the customer.

6.4 Suggestions for Further Research

Following to the limitations of this study, some suggestions for further research include the application of my methodology onto other vessel classes and across entire fleets. By comparing the results from multiple vessels, it would be possible to estimate statistics of the model performance and of the estimation error, and to reduce the effect of outliers. Moreover, by comparing the reporting accuracy of different crews on different vessels, it would become possible to determine what a "good", realistic level of reporting accuracy constitutes; in contrast to the simulated, error-free reports that are known to be unrealistically

accurate. Additionally, one could further study the effects of human reporting errors by adding (random) noise the simulated error-free reports and examine how the modeled output changes.

Beyond these suggestions that arose from the limitations of this study, a qualitative study on the crew contracts and their incentive schemes could yield interesting results. Intuitively, one could image that a crew which is well compensated and which receives additional bonuses for operational excellence would have a higher incentive to provide accurate reports for a monitoring system than a crew that perceives the tool as a threat. A qualitative study on commonly used incentive schemes in combination with a measurement for the noon report quality could help to reveal factors that improve or hinder the accuracy of the crew-generated input data.

This Master's thesis offered an exemplarily validation of a novel method for fuel flow estimation, based on modern data analytics and virtual sensing techniques. In light of the growing economic, environmental, and regulatory pressure on the stakeholders in the maritime industry, innovative data-driven systems are likely to become more widespread in the future. It will be exciting to see how modern data analytics will transform a longstanding and traditional industry.

References

Agnolucci, P., Smith, T., & Rehmatulla, N. (2014). Energy efficiency and time charter rates: Energy efficiency savings recovered by ship owners in the Panamax market. *Transportation Research Part A: Policy and Practice, 66*, 173-184.

Albazzaz, H., & Wang, X. Z. (2006). Historical data analysis based on plots of independent and parallel coordinates and statistical control limits. *Journal of Process Control, 16*(2), 103-114.

Antola, M., Solonen, A., & Pyörre, J. (2017). *Notorious speed through water*. Paper presented at the 2nd Hull Performance & Insight Conference (HullPIC), Ulrichshusen, Germany (pp. 156-165). Retrieved from: http://data.hullpic.info/hullpic2017_ulrichshusen.pdf [04.03.2017].

Antola, M., Solonen, A., & Staboulis, S. (2017). *The art of scarcity: Combining high-frequency data with noon reports in ship modeling*. Paper presented at the 2nd Hull Performance & Insight Conference (HullPIC), Ulrichshusen, Germany (pp. 118-123). Retrieved from: http://data.hullpic.info/hullpic2017_ulrichshusen.pdf [04.03.2017].

Baffi, G., Martin, E. B., & Morris, A. J. (1999). Non-linear projection to latent structures revisited (the neural network PLS algorithm). *Computers & Chemical Engineering, 23*(9), 1293-1307.

Bertram, V. (2000). *Practical Ship Hydrodynamics*. Oxford, UK: Butterworth-Heinemann.

Bialystocki, N., & Konovessis, D. (2016). On the estimation of ship's fuel consumption and speed curve: A statistical approach. *Journal of Ocean Engineering and Science, 1*(2), 157-166.

Blumstein, C., Krieg, B., Schipper, L., & York, C. (1980). Overcoming social and institutional barriers to energy conservation. *Energy, 5*(4), 355-371.

Branch, A., & Stopford, M. (2013). *Maritime economics*. Oxon, UK: Routledge.

Brown, M. A. (2001). Market failures and barriers as a basis for clean energy policies. *Energy Policy, 29*(14), 1197-1207.

Buhaug, Ø., Corbett, J.J., Endresen, Ø., Eyring, V., Faber, J., Hanayama, S., Lee, D.S., Lee, D., Lindstad, H., Markowska, A.Z., Mjelde, A., Nelissen, D., Nilsen, J., Pålsson, C., Winebrake, J.J., Wu, W., Yoshida, K. (2009). *Second IMO greenhouse gas study 2009*. London, UK: International Maritime Organization (IMO). Retrieved from:

http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPollution/Documents/SecondIMOGHGStudy2009.pdf [04.04.2017].

Chen, S., & Billings, S. A. (1989). Representations of non-linear systems: the NARMAX model. *International Journal of Control, 49*(3), 1013-1032.

Chiang, L. H., Pell, R. J., & Seasholtz, M. B. (2003). Exploring process data with the use of robust outlier detection algorithms. *Journal of Process Control, 13*(5), 437-449.

Crist, P. (2009). *Greenhouse gas emissions reduction potential from international shipping*. Paper presented at the Joint Transport Research Centre of the OECD and the International Transport Forum, Paris, France.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Boston, USA: Harvard Business Press.

Eide, M. S., Endresen, Ø., Skjong, R., Longva, T., & Alvik, S. (2009). Cost-effectiveness assessment of CO2 reducing measures in shipping. *Maritime Policy & Management, 36*(4), 367-384.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review, 14*(1), 57-74.

Eniram. (2017). Retrieved from: https://www.eniram.fi/services/trim/ [01.09.2017]

Faber, J., Behrends, B., & Nelissen, D. (2011). *Analysis of GHG marginal abatement cost curves*. Delft, CE: Ocean Policy Research Foundation. Retrieved from: http://www.cedelft.eu/art/uploads/file/CE_Delft_7525_The_Fuel_Efficiency_of_Maritime_Transport_def.pdf [21.05.2017].

Faber, J., Markowska, A., Nelissen, D., Davidson, M., Eyring, V., & Buhaug, Ø. (2009). *Technical support for European action to reducing greenhouse gas emissions from international maritime transport*. Delft, CE: European Commission. Retrieved from: http://www.cedelft.eu/?go=home.downloadPub&id=1005&file=7731_finalreportJF.pdf [24.06.2017].

Federal Aviation Administration. (2014). *Global positioning system (GPS) standard positioning service (SPS) performance analysis report*. Retrieved from: http://www.nstb.tc.faa.gov/reports/PAN86_0714.pdf [28.06.2017].

Flynn, D., Ritchie, J., & Cregan, M. (2005). *Data mining techniques applied to power plant performance monitoring*. Paper presented at the IFAC World Congress, Prague, Czech Republic (pp. 369-374). Retrieved from: https://pdfs.semanticscholar.org/8455/2af64bded0ce37acc50c12132f8fe4a36784.pdf [09.09.2017].

Fortuna, L., Graziani, S., Rizzo, A., & Xibilia, M. G. (2007). *Soft sensors for monitoring and control of industrial processes*. London, UK: Springer.

Fortuna, L., Rizzotto, G., Lavorgna, M., Nunnari, G., Xibilia, M., & Caponetto, R. (2001). *Soft computing: new trends and applications*. Berlin: Springer.

Golove, W. H., & Eto, J. H. (1996). *Market barriers to energy efficiency: a critical reappraisal of the rationale for public policies to promote energy efficiency*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Graus, W., & Worrell, E. (2008). The principal-agent problem and transport energy use: Case study of company lease cars in the Netherlands. *Energy Policy, 36*(10), 3745-3753.

Guidorzi, R. (2003). *Multivariable system identification: from observations to models*. Bologna, Italy: Bononia University Press.

Guo, B., Zhang, D., Yu, Z., Liang, Y., Wang, Z., & Zhou, X. (2013). From the internet of things to embedded intelligence. *World Wide Web, 16*(4), 399-420.

Harris, J., Anderson, J., & Shafron, W. (2000). Investment in energy efficiency: a survey of Australian firms. *Energy Policy, 28*(12), 867-876.

Harrould-Kolieb, E. (2008). *Shipping impacts on climate: a source with solutions*. Oceana. Retrieved from: http://www.cleanshipping.org/download/Oceana_Shipping_Report1.pdf [18.04.2017].

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (Vol. 2): Prentice Hall.

Hideyuki, A. (2011). *Performance monitoring and analysis for operational improvements*. Paper presented at the International Conference on Ship Efficiency, Hamburg, Germany. Retrieved from: http://www.ship-efficiency.org/onTEAM/pdf/PPTAndo.pdf [17.05.2017].

International Maritime Organization. (2008). Amendments to the annex of the protocol of 1997 to amend the international convention for the prevention of pollution from ships, 1973 (Revised MARPOL Annex VI) *MEPC.176(58)*.

Jaffe, A. B., & Stavins, R. N. (1994). The energy-efficiency gap: What does it mean? *Energy Policy, 22*(10), 804-810.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York, USA: Springer.

Juditsky, A., Hjalmarsson, H., Benveniste, A., Delyon, B., Ljung, L., SjÖberg, J., & Zhang, Q. (1995). Nonlinear black-box models in system identification: Mathematical foundations. *Automatica, 31*(12), 1725-1750.

Laffont, J.-J., & Martimort, D. (2009). *The theory of incentives: the principal-agent model*. Princeton, NJ: Princeton University Press.

Levinson, A., & Niemann, S. (2004). Energy use by apartment tenants when landlords pay for utilities. *Resource and Energy Economics, 26*(1), 51-75.

Li, H., Yu, D., & Braun, J. E. (2011). A review of virtual sensing technology and application in building systems. *HVAC&R Research, 17*(5), 619-645.

Maddox Consulting. (2012). *Analysis of market barriers to cost effective GHG emission reductions in the maritime transport sector*. European Commission. Retrieved from: https://ec.europa.eu/clima/sites/clima/files/transport/shipping/docs/market_barriers_2012_en.pdf [21.04.2017].

MAN Diesel & Turbo. (2011). *Basic principles of ship propulsion*. Copenhagen: MAN. Retrieved from: https://marine.man.eu/docs/librariesprovider6/propeller-aftship/basic-principles-of-propulsion.pdf?sfvrsn=0 [26.05.2017]

Marschak, J. (1955). Elements for a theory of teams. *Management Science, 1*(2), 127-137.

McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. Sebastopol, CA: O'Reilly Media, Inc.

Murtishaw, S., & Sathaye, J. (2006). Quantifying the effect of the principal-agent problem on US residential energy use. *Lawrence Berkeley National Laboratory*.

Nørgård, P. M., Ravn, O., Poulsen, N. K., & Hansen, L. K. (2000). *Neural networks for modelling and control of dynamic systems*. London: Springer.

Rehmatulla, N., & Smith, T. (2015). Barriers to energy efficiency in shipping: A triangulated approach to investigate the principal agent problem. *Energy Policy, 84*, 44-57.

Reynolds, G. (2009). *The reduction of GHG emissions from shipping - a key challenge for the industry*. Paper presented at the World Maritime Technology Conference WMRC. Retrieved from: http://imare.in/media/30635/paper-no-3b-1-msgl-reynolds.pdf [10.08.2017].

Rohdin, P., Thollander, P., & Solding, P. (2007). Barriers to and drivers for energy efficiency in the Swedish foundry industry. *Energy Policy, 35*(1), 672-677.

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *The American Economic Review, 63*(2), 134-139.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., & Tarantola, S. (2008). *Introduction to sensitivity analysis*. Chichester, West Sussex: John Wiley & Sons.

Simon, D. L., & Litt, J. S. (2011). A data filter for identifying steady-state operating points in engine flight data for condition monitoring applications. *Journal of Engineering for Gas Turbines and power, 133*(7).

Sinha, N., & Gupta, M. M. (2000). Introduction to soft computing and intelligent control systems. In *Soft Computing and Intelligent Systems: Theory and Applications* (pp. 23-38). San Diego, CA: Academic Press.

Smith, T. W. P., Jalkanen, J. P., Anderson, B. A., Corbett, J. J., Faber, J., Hanayama, S., . . . Pandey, A. (2015). *Third IMO greenhouse gas study 2014*. London, UK: International Maritime Organization (IMO). Retrieved from: http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPollution/Documents/Third%20Greenhouse%20Gas%20Study/GHG3%20Executive%20Summary%20and%20Report.pdf [05.05.2017]

Söderström, T., & Stoica, P. (1988). *System identification*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Sorrell, S., O'Malley, E., Schleich, J., & Scott, S. (2004). *The economics of energy efficiency: Barriers to cost-effective investment*. UK: Edward Elgar.

Thollander, P., & Palm, J. (2012). *Improving energy efficiency in industrial energy systems: An interdisciplinary perspective on barriers, energy audits, energy management, policies, and programs*. London, UK: Springer Science & Business Media.

Trodden, D. G., Murphy, A. J., Pazouki, K., & Sargeant, J. (2015). Fuel usage data analysis for efficient shipping operations. *Ocean Engineering, 110*, 75-84.

Velthuijsen, J. W. (1993). Incentives for investment in energy efficiency: an econometric evaluation and policy implications. *Environmental and resource economics, 3*(2), 153-169.

Vernon, D., & Meier, A. (2012). Identification and quantification of principal–agent problems affecting energy efficiency investments and use decisions in the trucking industry. *Energy Policy, 49*, 266-273.

Wang, H., Faber, J., Nelissen, D., Russell, B., & Amand, D. (2011). *Marginal abatement costs and cost effectiveness of energy-efficiency measures*. Institute of Marine Engineering Science and Technology (IMarEST). Retrieved from: http://www.cedelft.eu/?go=home.downloadPub&id=1090&file=Marginal_abatement_costs.pdf [09.08.2017]

Warne, K., Prasad, G., Rezvani, S., & Maguire, L. (2004). Statistical and computational intelligence techniques for inferential model development: a comparative evaluation

and a novel proposition for fusion. *Engineering Applications of Artificial Intelligence, 17*(8), 871-885.

Weber, L. (1997). Some reflections on barriers to the efficient use of energy. *Energy Policy, 25*(10), 833-835.

Yan, W., Shao, H., & Wang, X. (2004). Soft sensing modeling based on support vector machine and Bayesian model selection. *Computers & Chemical Engineering, 28*(8), 1489-1498.

Zahedi, G., Elkamel, A., Lohi, A., Jahanmiri, A., & Rahimpor, M. R. (2005). Hybrid artificial neural network - First principle model formulation for the unsteady state simulation and analysis of a packed bed reactor for CO2 hydrogenation to methanol. *Chemical Engineering Journal, 115*(1), 113-120.

Zilahy, G. (2004). Organisational factors determining the implementation of cleaner production measures in the corporate sector. *Journal of Cleaner Production, 12*(4), 311-319.

Appendix A: Variable Overview

Variable	Definition
_spanned	Suffix that indicates a variable that has been aggregated over a certain (irregular) reporting time span
sog	Vessel speed-over-ground (in knots); obtained from GPS
stw	Vessel speed-through-water; forecast-STW equals SOG adjusted for ocean currents
stw_ref	Reference STW (in knots); best estimate for a vessel's true STW; only available from an integration-based system
stw_manual	Crew-reported vessel speed (in knots)
draft_manual	Crew-reported vessel draft (in meters)
draft_ref	Reference draft (in meters); from integration-based system
water	Model input that expresses the additional power requirement (or power loss) due to frictional (water) resistance
draft	Model input that expresses the additional power requirement (or power loss) due to changes in the draft level
wind	Model input that expresses the additional power requirement (or power loss) due to winds; can be positive or negative
wave	Model input that expresses the additional power requirement (or power loss) due to waves
fuelflow_manual	Crew-reported fuel consumption (in tons/day)
fuelflow_pred	Predicted instantaneous fuel flow rate (in tons/day)
fuelflow_ref	Reference fuel flow (in tons/day); from integration-based system; used for model validation and simulation of error-free noon reports; technically a scaled propulsion power measure
Lightweight_pred	Fuel flow prediction (in tons/day) from the real lightweight system as installed on the case vessel