Ship Operational Efficiency: Performance Models and Uncertainty Analysis

Lucy Gemma Aldous

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

University College London

2015

Supervised by:

Dr Tristan Smith (UCL Energy Institute)

and

Professor Richard Bucknall (UCL Mechanical Engineering)

I, Lucy Gemma Aldous, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

The author claims to offer novel contributions in the following areas:

1. The development of an uncertainty framework specific to ship performance
2. The first rigorous comparison of the relative uncertainty of noon-report and continuous monitoring based methods
3. The extensive testing of performance data and models (using those simulations) to establish in depth knowledge and insight into the generalities of ship performance analysis.
4. A novel, hybrid model for ship operational performance monitoring

Abstract

There are increasing economic and environmental incentives for ship owners and operators to develop tools to optimise operational decisions, particularly with the aim of reducing fuel consumption and/or maximising profit. Examples include real time operational optimisation, maintenance triggers and evaluating technological interventions. Performance monitoring is also relevant to fault analysis, charter party analysis, vessel benchmarking and to better inform policy decisions. The ship onboard systems and systems in which they operate are complex and it's common for data modelling and analysis techniques to be employed to help extract trends. All datasets and modelling procedures have an inherent uncertainty and to aid the decision maker, the uncertainty can be quantified in order to fully understand the economic risk of a decision. An unacceptable risk requires further investment in data quality and data analysis techniques. The data acquisition hardware, processing and modelling techniques together comprise the data acquisition strategy. This thesis presents three models which are deployed to measure the ship's performance. A method is developed to systematically evaluate the relative performance of each model. Model uncertainty is one of four uncertainties identified as being relevant to the ship performance measurement. This thesis details and categorises each source and presents a robust method, based on the framework of the "Guide to Uncertainty in Measurement using Monte Carlo Methods", to quantify the overall uncertainty in the ship performance indicator. The method is validated using a continuous monitoring dataset collected from onboard an in-service ship. This method enables uncertainty to be quantifiably attributed to each source and a sensitivity analysis highlights the relative significance of each. The two major data acquisition strategies, continuous monitoring, CM and noon reported, NR are compared in combination with the other data acquisition parameters to inform the appropriate strategy for the required application and where further investment is required. This work has demonstrated that there is a ten-fold improvement in uncertainty achieved using a continuous monitoring set relative to a noon report dataset. If noon report data were collected perfectly, without the influence of human error, then uncertainties of the 5% level are achievable. The significant data acquisition parameters that improve precision are speed sensor precision and sample size. The equivalent that improve bias are speed sensor trueness and sample averaging frequency.

Acknowledgements

I gratefully acknowledge advice and support from colleagues both at the UCL Energy Institute and Mechanical Engineering department. In particular,

Dr Tristan Smith

Professor Richard Bucknall

Dr David Shipworth

Professor Andreas Schaffer

And personal friend, Dr Alan Murray

Many thanks to you all for your time, knowledge and excellent and invaluable advice throughout.

The research presented here was carried out as part of a UCL Impact studentship with the industrial partner BMT group. I am grateful for the funding that made this work possible and I especially thank John Buckingham at BMT for his continuous support and technical advice.

Contents

List of Figures

List of Tables

Nomenclature and Abbreviations

Symbol	Variable and units
A_F	Area of maximum transverse section exposed to the wind, m^2
AIAA	American Institute of Aeronautics and Astronautics
AIS	Automatic Identification System
A_L	Lateral plane area, m^2
ANN	Artificial Neural Network
ANSI	American National Standards Institute
ASME	The Association for the Study of Medical Education
ASTM	American Society for Testing and Materials
B	Beam, m
BDN	Bunker Delivery Note
BF	Beaufort wind scale measurement
BSFC	Brake Specific Fuel Consumption
C	Resistance coefficient
Cap	Weight of cargo carried
C_b	Block coefficient
CD_L	Longitudinal force coefficient
CD_{lAF}	Longitudinal force coefficient at maximum transverse section exposed to the wind
CD_t	Side force coefficient
C_f	Carbon factor; fuel mass to CO2 conversion factor
CFD	Computational Fluid Dynamics
C_{fr}	Coefficient of friction
CI	Computational Intelligence
CM	Continuous Monitoring
C_m	Midship area coefficient
CO_2I	CO_2 Intensity
Cp	Prismatic coefficient
Cs	Coefficient of in-service roughness allowance
CSC	Clean Shipping Coalition
C_w	Coefficient of wave resistance

Symbol	Variable and units
CWFR	Clarksons World Fleet Register
C_{wind}	Wind resistance coefficient
d	Distance corresponding to cargo carried, nautical miles
D	Propeller diameter, m
δ	Cross force parameter
DAQ	Data Acquisition
ΔP	Difference between P_{meas} and P_{exp}
ΔP_f	ΔP at each time step over the total evaluation period
D_M	Depth moulded, m
DNV	Det Norske Veritas, classification society (now DNV-GL)
DRE	Data reduction equations
ε	Angle of apparent wind, degrees
EEDI	Energy Efficiency Design Index
EEOI	Energy Efficiency Operational Index
EPA	Environmental Protection Agency (US)
EVO	EudraVigilance Organisation
f_{ave}	Sample averaging frequency
FC	Fuel consumption
FCI	Fuel Consumption Intensity
FF	Fully Filtered model
f_{LR}	Light running factor of the propeller
F_n	Froude number
g	Gravitational constant, ms-1
GHG	Greenhouse Gas
GP	Gaussian Process
GUM	Guide to Uncertainty in Measurement
H	Propeller pitch
h	Water depth, m
H&M	Holtrop and Mennen
η_0	Propeller efficiency
H_{accom}	Height of accommodation, m

Symbol	Variable and units
$\eta\Delta$	Quasi-propulsive efficiency
η_{eng}	Engine efficiency
HFO	Heavy Fuel Oil
η_H	Hull efficiency
ηP	Relative rotative efficiency
η_S	Shaft efficiency
H_S	Significant wave height, m
IMO	International Maritime Organisation
ISO	International Organization for Standardization
ITTC	International Towing Tank Conference
J	Advance coefficient
k_1	Form factor
KPI	Key Performance Indicator
KPSS	Kwiatkowski unit root test
K_Q	Coefficient of torque at the propeller
K_T	Coefficient of thrust at the propeller
Lbp	Length between perpendiculars, m
LCB	Longitudinal centre of buoyancy, m
LCV	Lower calorific value
LNG	Liquefied Natural Gas
LSDV	Least Squares Dummy Variable
Lwl	Waterline length, m
MC	Monte Carlo
MCM	Monte Carlo Method
MCR	Maximum continuous rating of the main engine, kW
MEPC	Marine Environmental Protection Committee
N	Sample size; number of observations
n	Exponent, found from logarithmic regression
∇	Hull displacement, m3
NCV	Net Calorific Value
NF	Not Filtered model

Symbol	Variable and units
ν_κ	Kinematic viscosity of sea water, m2/s
NN	Neural networks
NR	Noon Report
OCIMF	Oil Companies International Marine Forum
OLS	Ordinary Least Squares regression method
P	Power, kW
P_B	Brake power, kW
PBA	Probability Bounds Analysis
P_D	Power delivered, kW
PDF	Probability Density Function
Pd_{no}	Dimensionless performance indicator (Deligiannis 2014)
PDr	Propeller diameter ratio
PE	Effective towing power, kW
PEI	Propulsion Energy Intensity
P_{exp}	Expected power (model output) from measured parameters, kW
PF	Partially Filtered model
PI	Performance Indicator
P_{meas}	Measured power from on-board sensor, kW
P_S	Shaft power, kW (equivalent to PD, power delivered to the propeller)
P_{true}	Actual measured power in the simulation, kW
R	Ship total resistance
R_A	Model – ship correlation resistance, kN
ρ_{air}	Density of air
R_{APP}	Resistance of appendages, kN
R_B	Additional pressure resistance of bulbous bow near water surface, kN
R_{eff}	Relative Effects
R_F	Frictional resistance, kN
R_j	Model Variance for each parameter, j
RMSD	Root Mean Squared Deviation

Symbol	Variable and units
RMSE	Root Mean Squared Error
R_n	Reynolds number
RPM	Shaft rpm
RQ	Research Question
RSS	Root Sum of Squares
ρ_{sw}	Density of sea water, kg/m^3
R_T	Total resistance, kN
R_{TR}	Additional pressure resistance due to transom immersion, kN
R_W	Wave making resistance, kN
R_{wind}	Wind resistance, kN
S	Slip ratio, (S = VG / N*H)
SE	Standard Error
SEEMP	Ship Energy Efficiency Management Plan
SEI	Ship Energy Intensity
SEM	Standard Error of the Mean
SFOC	Specific Fuel Oil Consumption, g/kWh
SFOCdes	SFOC at optimum design point, g/kWh
SOG	Ship speed over ground, knots
STW	Ship speed through the water, knots
t	Thrust Coefficient
T_{des}	Design draught, m
t_{eval}	Sample size (evaluation period length), days
T_M	Mean draught, m
t_{tot}	Total time
V	Ship speed, knots
V_{des}	Ship speed through the water, design, knots
V_G	Speed over ground, knots
VIF	Variance Inflation Factor
VLCC	Very Large Crude Carrier
V_{new}	New ship speed, knots
Voy_{bin}	Factor for each voyage

Symbol	Variable and units
V_{waves}	Waves speed, m/s
V_{wind}	Wind speed, m/s
ω	Wake fraction
ω_q	Torque wake fraction
WSA	Wetted Surface Area, m2
Z	Number of propeller blades
$\sigma_{\Delta P}$	The standard error of the mean of the delta P
λ	Wave length, m

Chapter 1. Introduction

1.1 Context

The focus of this study is international shipping, particularly deep sea merchant shipping, for example tankers, bulk carriers and container ships.

Over recent years, fuel reduction incentives have been primarily driven by rising fuel costs and to a lesser, but increasingly significant extent by international regulatory bodies (the International Maritime Organisation, IMO, for example). As fuel costs rise, voyage costs as a percentage of revenue increase and fuel efficiency becomes a defining point in the ability of a ship owner to remain competitive. CO_2 emissions are correlated with the global economic state; global shipping emissions have in recent years stabilised following the 2007/2008 financial downturn (IMO Third GHG Study, (Smith, Jalkanen et al. 2014) however the three decades prior to this, following the oil crisis of the 1980s, saw a sharp increase (348 to 620 million tonnes CO_2 (IEA (2012) indicating that a period of CO_2 level growth is likely to again be on the horizon. In January 2014 the IMO made amendments to MARPOL Annex VI Regulations for the prevention of air pollution from ships. This made mandatory the Energy Efficiency Design Index (EEDI) for new ships and the Ship Energy Efficiency Management Plan (SEEMP) for all ships. Both of these measures quantify the ratio of the environment costs to the transport benefit achieved on a per ship basis (grams CO_2 emitted/tonne nautical mile of goods transported). These measures aim to reduce emissions of greenhouse gases (GHGs) from international shipping, in the case of EEDI by stipulating a maximum allowable CO_2 emission intensity, according to ship type and size, which is gradually reduced over time.

According to a DNV study by Eide and Endresen (2010) (Figure 1) methods of improving fuel efficiency have been categorised into structural, operational, technical and alternative fuels.

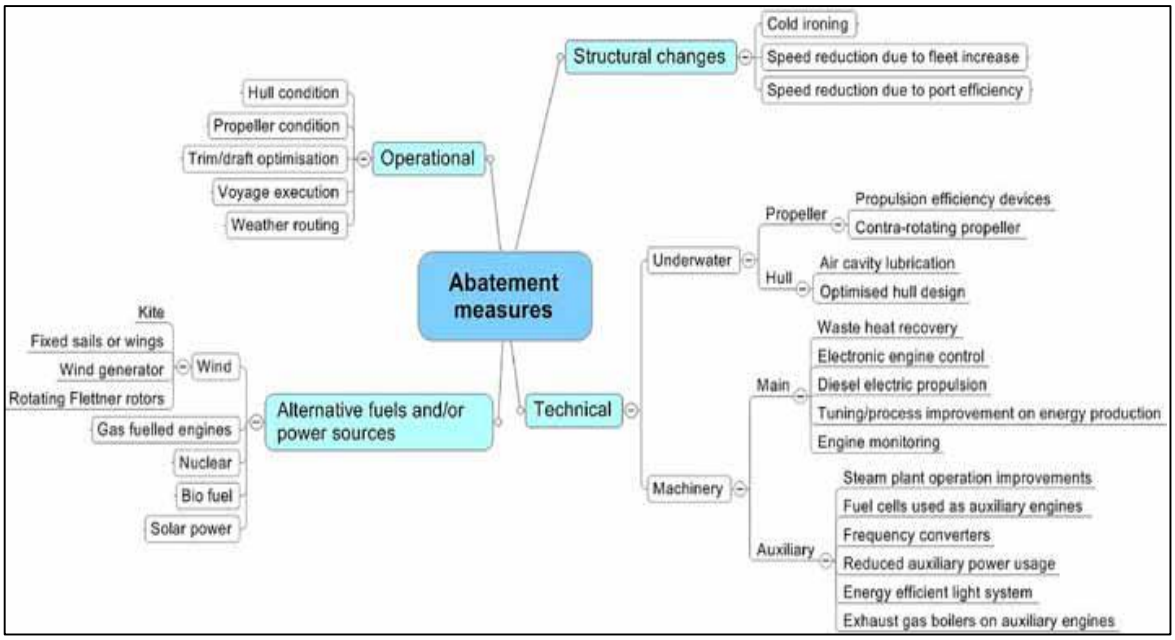


Figure 1: Abatement measures to reduce CO_2 emissions from shipping (Eide and Endresen (2010)

There is a vast amount of research into the advantages and disadvantages associated with each of these areas. The research described in this thesis is constrained to operational measures. In this area there is plenty of potential to make significant fuel savings (10-50% saving in g.CO_2/tonne.nm, IMO (2009)) with relatively low initial investments and a correspondingly low risk. Operational measures are a particularly productive area for research because of the various systems in which a ship operates, according to one literature source operational performance may be classified into four main areas (Reid (1985):

- Power plant and auxiliaries: Engine corrosion and oil deposits
- Propeller efficiency: Affected primarily through propeller blade roughness and damage
- Hull resistance: Affected by mechanical, chemical and biological deterioration of the hull
- Navigation, steering and routing: Speed, displacement, trim, plus dynamic effects of ship motions, steering and weather.

These performances are primarily focussed on the onboard systems, this presentation is an over-simplification because in practice a ship's onboard systems are heavily

influenced by the environmental and economic sphere, as depicted in Figure 2. At the ship level the machinery configurations and efficiencies determine the onboard mechanical, thermal and electrical energy flows which, despite automation being built in to the configuration mode settings at the ship design phase, there is still an appreciable level of human interaction during day to day operations. The environmental conditions (sea state, wind speed, sea/air temperature etc.) are dynamic, unpredictable and complicated to quantify in combination, due in part to the characteristics of the turbulent flow fields by which they are determined.

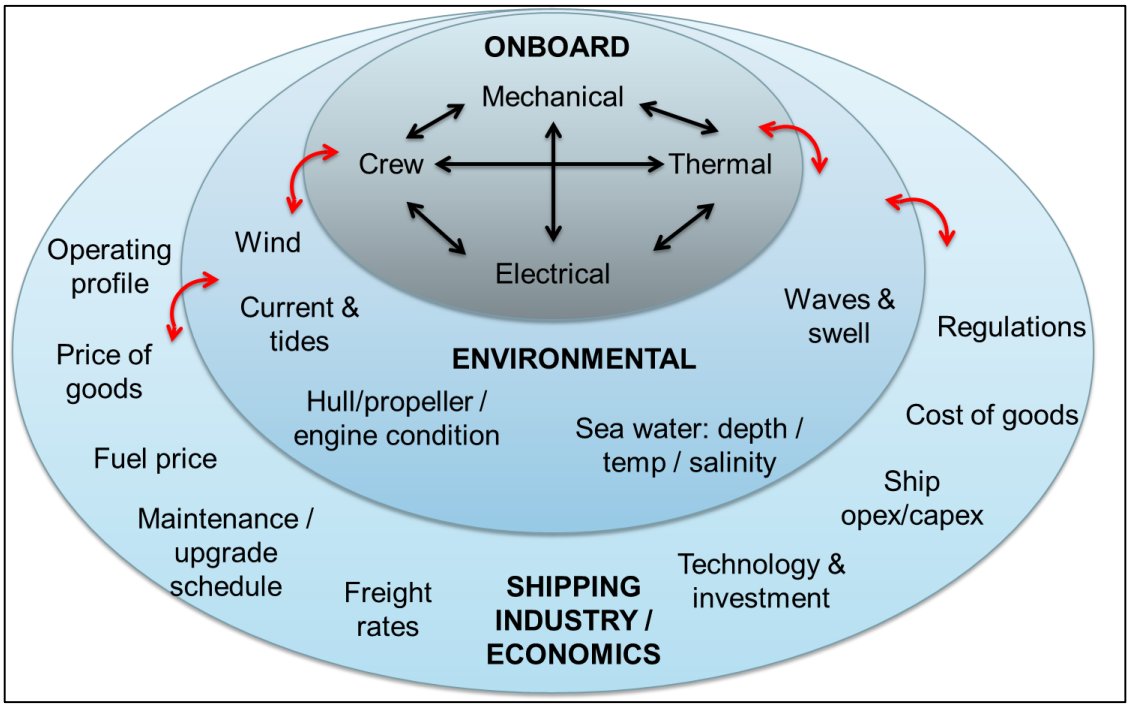


Figure 2: The ship systems and their interactions

These environmental conditions exert an influence on the ship's resistance and therefore on the ship's power requirements in differing quantities. The bunker fuel energy distribution and losses at various components of the propulsion system are described diagrammatically in Figure 3. The environmental and operational conditions that affect specific components have then been superimposed on to this.

The interactions are complex. The ship's fuel consumption, rpm, draught, degradation of systems and the environmental conditions are inextricably linked in a physical manner that is not immediately clear due to the often non-linear relationships between the various elements that make up the propulsion system. The

condition of the hull and it's fouling, for example, depends on the age of the ship, the salinity and temperature of the water in which it operates, the hull cleaning schedule, the hull coating, the ratio of time spent at sea to time in port and the ship's speed. The fouling in turn affects overall ship resistance and is speed dependent because of the effect of the Reynolds number on the coefficient of friction. The subsequent relationship between fuel consumption (FC) and power incorporates deviations from a distinct fuel-speed law due to the varying specific fuel consumption, SFOC (g/kWh) which is theoretically dependent on the combined propeller speed and engine load (adjusted for the shaft efficiency) and where on the engine layout diagram these variables coincide. These are in turn influenced by likely non-linear degradation effects (i.e. soot build-up in the engine) that occur over time. Another example is the Beaufort scale (BF) which is representative of the wind speed, one relationship between shaft power and wind speed, aerodynamic drag, can be approximated by a drag coefficient based on the transverse projected area of the ship perpendicular to the wind direction and the square of the wind speed (N. Hamlin and Sedat 1980). The principal effect of the recorded Beaufort number on fuel consumption however is actually the implicit effect of wind generated surface waves, one simplified relationship is based on the drag coefficient for wave resistance and the square of the wave height that is proportional to the ship's power (Lindstad, Asbjørnslett et al. 2011). However this only represents wind driven surface waves and does not account for swell. It also does not reflect the fact that the ship's speed may have to be reduced in heavy weather so as to not violate the maximum torque/rpm allowance of propulsion system components. These are two examples of the many system interactions that exist.

Furthermore, the shipping industry operates in an economic sphere in which the global consumption of goods and global energy demand, and conditions in the various shipping markets determine operating profiles, costs and prices (Lindstad 2013). In addition, technological investment, fuel efficiency and savings are complicated by the interactions between ship owner-charterer-manager (Agnolucci 2014). The competitive economic climate that exists comprising high fuel prices and overcapacity of ships keeping cargo rates low means that operational decisions by ship owners/operators at the fleet level must then be considered in terms of fuel reductions. There is a need to translate these into a cost-benefit assessment against

potential losses in revenue (i.e. dry docking for hull cleaning or engine maintenance, slow steaming, weather routing; power-speed trade-off). It may be more attractive to look at operational measures in reducing fuel consumption and emissions rather than adopting alternative technologies which may have potentially long pay back periods and increased risk associated with any new investment.

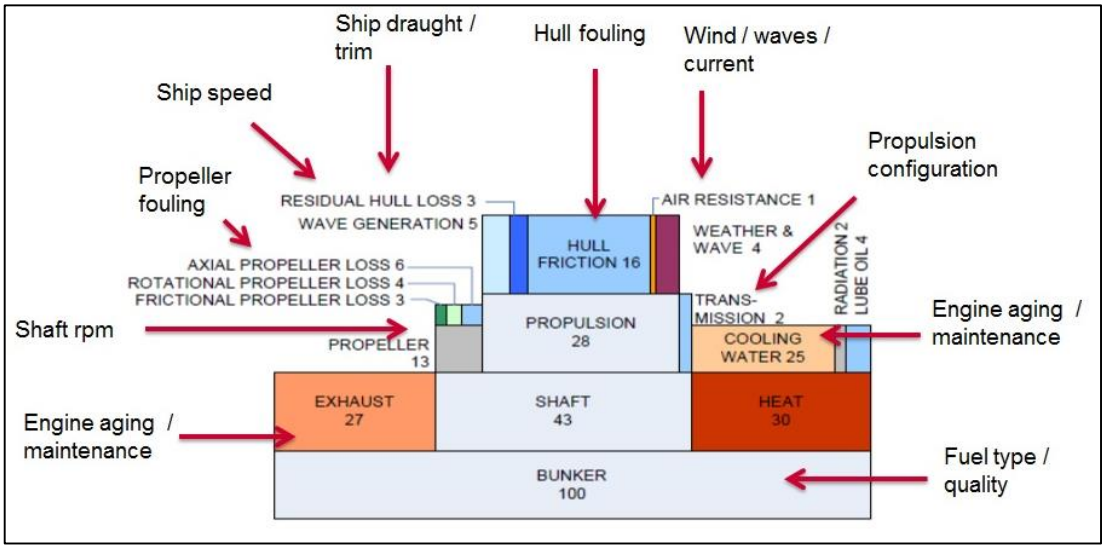


Figure 3: Use of propulsion energy onboard a small cargo ship, head sea BF6, IMO (2009)

To realise savings, assess investment risks and remain competitive in tough financial and regulatory times then changes in performance must be measured for their conversion to a quantifiable economic benefit. In shipping, the financial stakes are high therefore measurements must be made to a known degree of accuracy. As described in the study by Armstrong (2013), quantification is a significant aspect of the development and roll out process of optimisation initiatives and determining margins of accuracy is cited as one particular challenge. Quantification and uncertainty analysis as related to ship performance are the broad topics with which this thesis is concerned.

Chapter 2 introduces the motivations behind investing in the development of ship performance models and the relevance of uncertainty analysis of these models and ship performance monitoring overall. This is followed by a review of the performance indicators and relevant models that are in the literature; the shortfalls in current work are highlighted to identify gaps and formulate the research questions. The second part of chapter 2 discusses the current state of the art with regards to

applied uncertainty analysis, relevant international standards and drawing on findings and applications from other industries. The topics explored and issues raised in chapter 2 culminate in the research questions which are presented in chapter 3 with emphasis on why these are relevant, original and interesting. The structure of the thesis thereafter is described at the end of chapter 3.

Chapter 2. Literature Review

The aim of this chapter is to 1) convey the motivation behind investing resources into ship performance measurement and 2) to illustrate why the uncertainty associated with this measurement is worthy of attention. It will then examine in detail the state-of-the-art in performance measurement as reported in the literature, to critically analyse this work and to identify gaps. Uncertainty analysis has received minimal attention in the context of ship performance so a section is devoted to general uncertainty analysis and identifying and characterising the source of uncertainties as implemented by other industries. Methods of quantification of uncertainty through propagation are reviewed with an emphasis on international standards and well established guidelines.

2.1 Introduction

The context in the previous chapter has very briefly highlighted the complexities of ship performance measurement through a high level description of ship systems and interactions. Before exploring this in more detail, this section looks first at general measurement in shipping, with the aim of conveying the many motivations for ship performance measurement. This is done with reference to specific applications and examples from the literature and relevant international standards. It then looks at general measurement uncertainty and why this is important in other industries and engineering applications. This leads on to what uncertainty means in ship performance measurement and why it should warrant attention.

Ship performance, in this thesis is defined as the combined change in the performance of the hull, propeller and engine over time, assuming no alterations have been made to its design. One simple definition of performance is the rate of fuel consumption required to move the vessel through the water for the given conditions, which may be operational (speed or draught) or environmental (wave height, wind speed, etc.).

2.1.1 Performance Measurement in Shipping

The motivational drivers for measurement in shipping generally reduce to bunker fuel and/or carbon emission auditing for environmental policy, to motivations related to fuel efficiency and economic benefit or to charter party agreement analysis in situations of claims and disputes. This section details some specific examples from the literature.

Data collection, either through daily noon reporting procedures or high frequency, automatic data acquisition systems, and data processing techniques such as filtering and/or normalising have so far proven to be useful tools in capturing and quantifying some of the intricacies and nuances of the interactions described in section 1.1, to better understand the consequences of operational decisions. A ship performance normalisation model may be developed in order to estimate the response in the dependent variable to each operational (trim, ballast, time out of dock) and environmental condition given the ship's fixed design parameters (hull geometry, propeller characteristics, engine configuration). The dependent variable is ship performance which is ultimately either measured in power or fuel consumption, or a change in power / fuel consumption over time. The ship performance model is then deployed in order to normalise each influential variable to a baseline. A simplified model, in terms of a reduced number of variables, is derived if the data is first filtered so that only baseline conditions are included in the dataset.

Some applications of ship performance models are summarised below:

i. *Operational real time optimisation:* Performance can be optimised by altering controllable variables such as trim, ballast, speed or time between maintenance events according to the uncontrollable conditions, either environmental, economic or both. Weather routing takes advantage of weather and currents in order to optimise voyage distance or time travelled and thereby minimise voyage costs (or maximise profits) and maximise safety. Measured operational savings can be up to 3% in fuel, apart from time savings (Armstrong 2013). There has been much focus in the literature of the trade-off between slow steaming to reduce voyage costs and the consequent extended voyage time resulting in lost revenue. In response to the oil

shocks of the 1970s, Ronen (1982) pioneered the analysis of optimum speed sensitivities given different revenue schedules. In recent years, traditional models combine a technical ship performance model (often a cubic relationship between power and speed is assumed, which may be simplistic) with a cost model that includes bunker prices, freight rates and volumes for example (Maloni, Paul et al. 2013) and the number of vessels employed per loop or fleet size (Notteboom and Vernimmen 2009), (Ronen 2010). The objective function is generally to reduce costs or, more recently, to reduce CO_2 emissions. Optimum speed models have been expanded to include sensitivities to sea conditions (Lindstad 2013), to logistical ship routing scenarios (Psaraftis and Kontovas 2014), to other technical efficiency improvements i.e. a more efficient hull (Smith 2012), to other ship types such as LNG carriers where the use of the cargo as a fuel in the form of boil off gas complicates the decision framework (Aldous and Smith 2013) and in response to higher energy costs in Sulphur Emission Control Areas (Doudnikoff and Lacoste 2014). Trim and propeller pitch optimisation algorithms have also revealed cost savings, although trim optimisation by itself has been seen to realise small savings in the order of 1% (Armstrong 2013). Other voyage optimisation methods can improve the punctuality of ships especially in adverse weather conditions or to take advantage of following seas and to aid the "just in time" arrival concept and port operations, for example in assigning berths, crewing and loading planning.

ii. *Maintenance trigger:* Hull and propeller performance monitoring acts as a decision support tool for determining dry dock intervals, hull coating type and quality, the extent of any hull pre-treatments applied, the frequency and method of hull and propeller cleaning, and of hull and propeller modifications (Munk and Kane 2011). For example, for detecting the influence of an accumulation of fouling on the propeller and hull for the purposes of determining when remedial action should be taken, a shaft-power trigger value is derived in a paper by Walker and Atkins (2007). Hull and propeller fouling has been an area of interest not only for ship owners and operators, but it is also recognised by the MEPC as an important factor in reducing the industry's GHG emissions. The Clean Shipping Coalition (CSC) have estimated that the impact of the deterioration in hull and propeller performance is likely to result in a 15 to 20 per cent loss in vessel efficiency on average over approximately 50 months (IMO_MEPC_63/4/8 2011). This is significant in terms of fuel consumption costs and GHG emissions. The CSC has brought to international

attention the need for a transparent and reliable hull and propeller performance standard. They highlight the following drivers to the need for performance monitoring:

a. Enable vessel owners to make informed decisions on which antifouling system to select
b. Define a benchmark for comparison of antifouling systems
c. Differentiate between hull performance over time and step changes in performance
d. Without a transparent method to measure ship performance, the "principal agent" issue[1] makes it difficult for owners/charterers to invest in hull & propeller performance technologies
e. To enable informed investment decisions

iii. *Evaluating technological interventions:* A model quantifies the current ship performance, this then acts as a benchmark from which newly installed technology is evaluated or from which expected performance can be predicted prior to installation in order to quantify cost versus benefit and the economic risk of the investment (for the ship owner and financier). Post analysis is good for learning from investments, proving or disproving manufacturer's claims or proving asset value against which loans can be borrowed and negative net worth avoided. Prior analysis is useful in assessing hull coating efficiency by recording and comparing the rate of change of fuel consumption over time before and after the application of the hull coating or any other new technology. These results may be used to inform financial mechanisms that will apportion savings to the charterer or ship owner/operator (Stulgis 2014).

iv. *Operational delivery plan optimisation:* This is similar to (i) but from a longer term and fleet wide perspective which strives to optimise operational parameters given a possibly heterogeneous fleet, contractual obligations and economic influences with the goal of minimising costs while maximising revenue from spot market transactions. A bulk of literature relates to the ship routing, scheduling and inventory management problems, for example (Rakke, Stålhane et al. 2011). Parameters for optimisation may include optimum speed or optimum maintenance frequencies for main engine/hull/propeller.

[1] The party responsible for the fuel bill (the charterer) is not the one who would have to pay for a performance enhancing technology retrofit (the ship owner)

v. *Fault analysis*: Recognise sudden changes in performance and conduct timely repair.

vi. *Charter party analysis*[2]: Performance monitoring and analysis by comparison of operating speed, draught and fuel consumption with the contract. One of the most commonly encountered disputes between owners and charterers relates the fuel consumption and vessel speed whilst under time charter. Ship performance models can help to settle claims by comparing expected to actual fuel consumption for a given speed.

vii. *Vessel benchmarking*: Collect best practices in operations both at a ship owner's fleet level and also at an industry level and this is supported by the Oil Companies International Marine Forum (OCIMF). A method is also presented in the paper "Ship energy performance monitoring and benchmarking" (Bazari (2007), where an overall scheme for energy performance rating/benchmarking is presented in order to provide a simple method of differentiating ships according to their energy efficiency and to inform a ship owners energy efficiency control program.

viii. *Inform policy*: In the assessment of bottom up modelling of global ship emissions, the starting point is a model of fuel consumption at an individual ship level which is then aggregated for the global fleet. The latest IMO greenhouse gas update study (Smith, Jalkanen et al. 2014) uses extensive AIS data to inform the activity profile (speed and loading) of the global fleet which is related to engine load and fuel consumption by a ship performance model. Studies show a variation of 36% for the global annual emissions for 2007 (J. J. Corbett, V. Eyring et al. 2009), table A1-19). The difference in engine load factor is cited as one of the larger sources of uncertainty in a sensitivity analysis by Corbett and Horst (2003). In this kind of application small inaccuracies at a single ship level multiply and have a large effect at the global level. If the current global emissions cannot be quantified accurately (or if the uncertainty cannot be quantified) then there is no benchmark with which to measure the effectiveness of policies. Eide, Endresen et al. (2009) study a range of operational and design measures that might be implemented to reduce CO_2 emissions in shipping and in a sensitivity analysis they too refer to engine load having an effect on the cost effectiveness of the measures. The cost effectiveness is used as a decision parameter in conjunction with a decision criterion in order to inform investment in

[2] http://www.jeppesenmarine.com/HighSeas-Offshore/Optimization-Solutions/Jeppesen-Fleet-Manager.aspx

emission reduction measures. They also show that their approach provides a viable method for the regulation of shipping emissions and can therefore be used to inform policy decisions. From a wider perspective, measuring any shipped product's carbon footprint may be required for regulation and, where transport by ship is part of the supply chain, then measuring this through models may be necessary. The aforementioned benchmarking/performance rating method (*vii*) may also be used with the added objective of informing a policy tool. For example ship ratings arising from voluntary agreements between flag states and ship owners can be used to differentiate ship's taxes, port dues and charges. Insurance rates, charter rates and other financial conditions could also be related to the ratings (Bazari 2007).

The wide range of applications detailed above identifies why ship performance monitoring is so crucial to industry and government. It is an area which is growing significantly in terms of research and development and in which ship owners are beginning to understand that investment is warranted.

2.1.2 Measurement Uncertainty

Generally in scientific investigation, in order to compare results, measurements need to be compared and conclusions drawn regarding the acceptability of the disparity between measurements and this requires some index of the uncertainty of the measurements (Gleser 1998). Experimental uncertainty and measurement, particularly in engineering, has been increasingly emphasised over the last 20 years. Some professional journals, such as the Journal of Heat Transfer, Journal of Fluids Engineering and all of the journals from the American Institute of Aeronautics and Astronautics (AIAA), have adopted policies requiring some type of adequately presented uncertainty analysis for all of their articles (Wahlin, Wahl et al. 2005). Broadly speaking, experiments are conducted in order to inform a decision, be it related to design, policy or operation, and generally there is some cost related to the decision; cost of raw materials, investment in R&D, time, human life in safety decisions or macro-economic consequences for a nation or globally. Since no measurement can be known exactly then at least its accuracy should be reported in order to fully assess the risk associated with the decision. In climate change research, high priority is placed on communicating uncertainty (M. Mastandrea 2010),

(Mastrandrea, Mach et al. 2011), (Curry and Webster 2011). In this domain, the risk of being incorrect implies serious economic consequences; to ratify the Kyoto agreement on greenhouse gas reduction holds the risk of escalating short term economic recession in some countries and has to be weighed against the certainty of the effect of GHG emissions on climate, the evidence for which comes from scattered data and complex models. Uncertainties relating to metocean data have also been identified as an important topic by the International Ship and Offshore Structures Congress (ISSC) Committees (Bitner-Gregersen, Bhattacharya et al. 2014). Their concern is typically focussed on increasing industry awareness regarding the safety aspect of the design and operations of marine structures. The authors identify that these uncertainties lead to over-design or under-design of marine structures and the consequence of significant economic/risk impact. This again highlights the link between uncertainty and decision making. The same paper details the consequences of metocean data uncertainty for various industries and applications, among many others. They highlight insufficient investigation of uncertainty in the application of detailed sea state data to optimal ship routing.

This risk assessment of decision making applies also to the measurement of ship performance; one possible way of structuring the interaction between these influential factors is depicted in Figure 4. The relative accuracy which ultimately determines the risk, alongside which the cost and benefit of a decision is evaluated, is linked to the amplitude of the noise or scatter in the data relative to the underlying, longer term trends that are to be extracted. The ship system interactions induce the scatter in the data, not only from inherent sensor imprecision but also from unobservable and/or unmeasurable variables. According to the central limit theorem (assuming independent, identical distributions), over time the scatter will tend to a normal distribution with zero mean. The actual time period length is dependent on the data acquisition and processing strategy and influential factors include the temporal resolution of sensors and data collection frequency, the sensor precisions and human interactions in the collection process and the processing method (normalisation or filtering). There are also uncertainties in the data that will introduce a potentially significant bias in the results and these too need to be understood and evaluated.

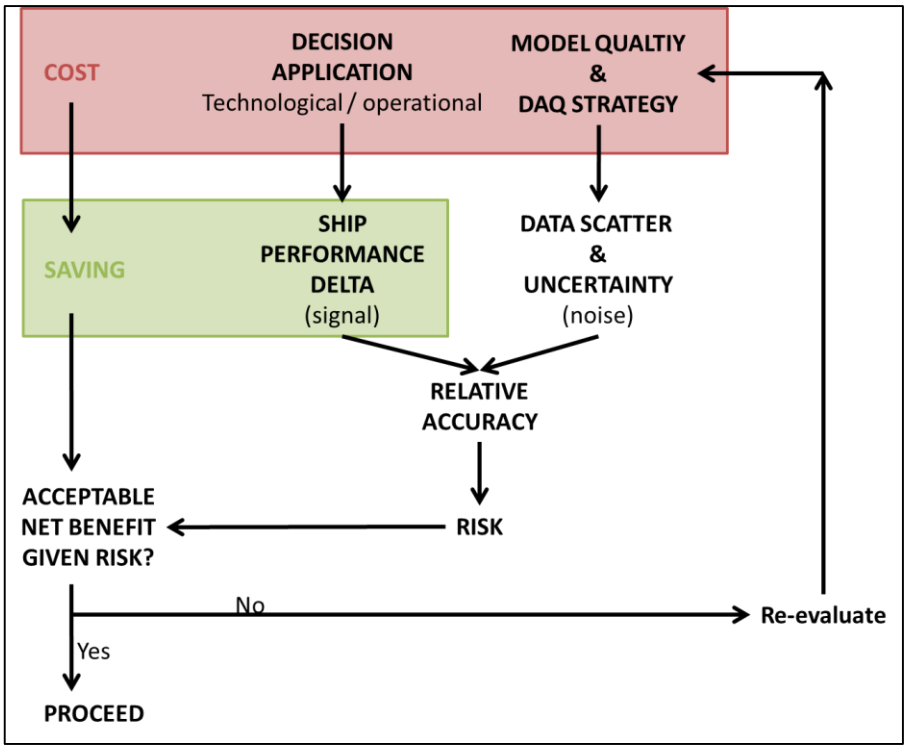


Figure 4: Cost, benefit and risk of ship performance measurement

The magnitude of the underlying trends to be identified are a function of the modelling application, for example, in predicting the expected performance of new technologies the signal delta, i.e. the improvement in ship performance, may be a step change of the order of 1-3% (as in the case of propeller boss cap fins) or up to 10-15% as in the case of hull cleaning or new coating applications (Fathom 2011). In the latter case analysis of trends in the time domain is also necessary. Table 1 presents examples of manufacturer claims from the literature relating to the expected improvement in ship performance. Some of these claims have been made based only on model tests, CFD models and/or sea trials. Generally the actual efficiency gains are specific to the ship's hull geometry, propeller characteristics, its operating profile and the current ship performance as determined by monitoring and analysis. The actual predicted fuel saving therefore has an associated uncertainty.

How these predicted savings fit into the cost-benefit analysis and how this interacts with the risk through uncertainty is depicted in Figure 4. The acquisition strategy and the signal delta determine the relative accuracy which defines the risk, the former has an associated cost; economic, time and resources and the latter has an associated cost and benefit, both economic and environmental. If the risk is deemed unacceptable

given the overall cost and benefit then it makes sense to re-evaluate investment in data quality and data analysis techniques in order to reduce the risk. This is particularly important in the shipping industry for example, measurement and verification is cited as a key barrier to market uptake in fuel efficient technologies and retrofitting. In order to secure capital, investment projects must be expected to yield a return in excess of some pre-defined minimum (Stulgis 2014). Weighing the economic risk of capital investment against the certainty of the effectiveness of a fuel efficient technology is therefore key.

Technology	Saving	Source
Shaft line streamlining	2%	Wartsilla (2008)
Air lubrication	20-30%	Stena Bulk
Advanced propeller blade sections	Nozzles – 5% Winglets – 3%	Fathom (2011)
Propeller boss cap fin	3% - 5%	Armstrong (2013)
Hull surface coatings	up to 9%	Fathom (2011)
Hull cleaning	10%	Fathom (2011)
Flettner rotor	2-20% of main engine power	Traut, Gilbert et al. (2014)

Table 1: Interventions for improvements in ship efficiency

From the perspective of using ship performance modelling as a means of determining and monitoring a 'shaft-power trigger value' as a signal for remedial action (i.e. a hull clean) as in Walker and Atkins (2007), then a business case can be presented. This would include the cost of the hull clean, costs of fuel, the ship specific design and the ship operational profile as part of the ship maintenance trigger. The inclusion of some uncertainty value of the model output would enable the economic risk to be evaluated alongside the information.

The above highlights the significance of uncertainty quantification surrounding ship performance monitoring in being able to quantify the confidence of the model output and thereby understand the risk associated with the conclusions drawn. This is also a subject of particular interest because of the vast variation between input data from

different ships and ship owners/operators. Variation arises between data acquisition hardware, crew recording procedures and data storage/recording intervals. Furthermore, the model applied, the performance indicators extracted and the forecasting methods employed also vary significantly. Subsequently there are a wide range of data acquisition and modelling strategies available to ship owners. A brief introduction to this follows and a more detailed description including associated advantages, disadvantages and inherent uncertainties can be found in Chapter 4.

The full "total solution" approach to data acquisition and performance monitoring is described as requiring the following five components (Ballou 2013):

- Shipboard data acquisition
- Communication method for shore transmission in a timely manner
- Shore based analytical tools for processing the data
- Intuitive "easy-to-use" displays of data and analytical results
- Ongoing user training and awareness programs

There is a vast set of variables that describe a ship's performance for a specific operating condition and at a given point in time. The measurement of each variable is sourced from a variety of different onboard sensor types, each with its own trueness and precision. This leads to a measurement / instrument uncertainty in the model output that is specific to each ship depending on the combination of sensor type employed. Acquisition strategies are broadly separated into two dominant dataset types:

- Noon reports (NR)
- Continuous monitoring (CM)

NR datasets are coarse but cheap to compile and readily available since they are currently in widespread use across the global fleet. The frequency of recording is once every 24 hours (time zone changes allowing), the information is input by the crew and the fields reported are limited. Generally included as a minimum are ship speed and position, fuel consumption, shaft rotational speed, wind speed estimated Beaufort number, date/time and draught. Given the economic and regulatory climate

as well as advances in IT and data acquisition (DAQ) systems, there has been a shift towards more complete automatic measurement systems. These require an array of sensors to be installed throughout the ship (propeller shaft, engine room etc.); these systems in this thesis are referred to as continuous monitoring (CM) system. The uptake of these has been limited by installation costs in service while improved data accuracy, speed of acquisition, high sampling frequency (1 to 5 minutes) and repeatability are cited as the key drivers. It is clear then that further uncertainties arise in the model output according to the sample recording frequency and the sample averaging frequency. Continuous monitoring systems may automatically record environmental conditions from onboard weather sensors (wind vanes and anemometers for example). Alternatively, the ship position and time stamp may be combined with metocean data from satellites and/or wave buoys to provide a complete operational and environmental picture. The metocean data is generally formed from ocean and atmospheric models which also have a degree of uncertainty associated with them. These are the two major types at either end of the spectrum; there are intermittent variations such as higher frequency noon report style measurement which may occur more than once per day (every 6 hours for example). Other data acquisition strategies may also rely on low frequency measurements but the data may be averaged over a shorter time period; 30 minutes worth of data recorded every day for example, this would reduce the errors incurred through sample averaging frequency. An example of this is dedicated speed trials which are discussed in further detail in the next section.

Generally, the majority of ships do not have CM systems installed whilst noon reports are in widespread use across the global fleet therefore a study and comparison of the uncertainty resultant from both is relevant. Further, there is a plethora of performance indicators coming from different techniques that are employed in modelling the ships systems. This leads to a wide range of model parameter and structural uncertainties that also influence the overall uncertainty in the model output measurement. This is a function of the detail and effort employed in the development of the model itself. Given the possible combinations and permutations of collection and analysis then a study of the sensitivities of the uncertainty in the ship performance measurement is pertinent to identify where the most significant uncertainties lie and therefore into which part of the data acquisition and analysis

procedure resources can be invested most effectively in order to improve the confidence of the model response. Then this raises the question of the cost of obtaining additional information and if that is outweighed by the value of the improvement in the model from which the performance estimate is derived (Loucks 2005).

This section has identified the wide range of applications for ship performance measurement and why this is a useful and relevant tool in many aspects of the shipping industry. It has been shown how important the quantification of uncertainty is to other industries and how this is linked closely with cost-benefit analysis and decision making, particularly with respect to making informed investment decisions. The relevance of this to the shipping industry has been stated. Finally, some high level detail of the variability in the data collection and processing methods has been presented. This variability delivers a range of uncertainties in the overall ship performance metric and this therefore brings about questions surrounding trade-offs and the optimum DAQ strategy for minimum uncertainty, or for the level of uncertainty appropriate to the application. The uncertainty analysis of ship performance measurements has not been found in the literature and there is therefore the opportunity to provide detailed and original analysis that is useful to this increasingly relevant area of research.

2.2 Performance Measurement

2.2.1 General Approaches

One aim of ship performance monitoring is to quantify the principal speed/power/fuel losses that result from the in-service deterioration (hull, propeller and engine). A performance indicator may be defined to identify these performance trends; performance being a ratio of input to useful output. For a ship the most overall unit of input is the fuel consumption. Sometimes, it is useful to isolate the performance of the hull and propeller from that of the engine, in which case it is the shaft power that becomes the input to the performance indicator estimation. The aggregate unit of output is transport supply, which for constant deadweight capacity and utilisation, ultimately comes down to the ship's speed. A more detailed

discussion of the appropriate performance indicator metric is found in section 4.2. One of the most basic methods to extract this information is to control all other influential variables; weather, draught, trim, water depth etc. as detailed in 1.1. This may be done in the following ways:

- Data collection in similar conditions: Through dedicated speed trials at a constant and specific speed or set range of speeds and loading condition and with dedicated manoeuvres.
- Model choice: Normalise or correct each influential variable wind condition, wave condition, water depth and loading condition to a baseline by employing a model that quantifies the expected shaft power or fuel consumption for all environmental or operating conditions.
- Data filtering (statistical approach): After filtering to standard reference conditions (draught, speed, fuel type and ambient conditions), then other more minor effects (sea depth, trim, sea temperature, wind, waves) are assumed to follow a normal distribution and will average out over long time periods.

Dedicated speed trials

If executed in accordance with ISO standards and guidelines, dedicated speed trials have the potential to produce accurate results which are straight forward to analyse and interpret, with minimal specialist expertise required. Historically, speed trials are performed by the yard to demonstrate to the owner that the ship's performance meets contractually agreed values. More recently speed/power trial guidelines have been reviewed with a focus on developing a transparent and straight forward procedure on which to base the Energy Efficiency Design Index (EEDI). This was tasked by the IMO to the Specialist Committee on Performance of Ships in Service who submitted a final report and recommendations to the 27th ITTC in 2014. Their remit was, among others, to liaise with relevant bodies to monitor and review the methods relating to the EEDI and EEOI which included reviewing the then existing ITTC trial test procedures (ITTC 2005) and ISO 15016. In their report they presented a review of the speed/power trial methods and their preliminary findings from investigations into the wind, wave, current and load, displacement and trim corrections. Elements

of this are drawn from a paper by MARIN "New Guideline for Speed/Power Trials" (Boom, Huisman et al. 2013). The consequent updated procedures produced are in two parts; part 1 refers to preparation and conduct (ITTC 2012) and part 2 refers to the analysis of speed/power trial data (ITTC 2014). ITTC 2012 is used as a simplified reference for the two procedures. These were submitted to IMO MEPC 64 and were accepted by IMO MEPC 65 when they became the preferred method for deriving the speed/power of ships for the EEDI.

ISO 15016:2015 "Ships and Marine Technology - Guidelines for the assessment of speed and power performance by analysis of speed trial data" is a revision of ISO 15016:2002 which was the first internationally recognised standard relating to sea trials (Strasser, Takagi et al. 2015). This latest ISO development also takes in to account the aforementioned updated ITTC 2012 and the two procedures have been harmonised to an extent, although the ISO 15016:2015 has additional corrections for factors such as resistance due to wind and waves and the effects of water temperature, density and water depth. The paper by Strasser, Takagi et al. (2015) compares the two methods with respect to the critical factor of current measurement and shows that they are fully compatible with each other.

Dedicated speed trials for ship performance measurement during service are advocated by Bazari (2007) and the services provided by hull and propeller performance monitoring companies such as Propulsion Dynamics[3]. Bazari finds that performance (ship energy intensity) based on dedicated, in-service speed trial data is 19% improved relative to that based on operational data with basic filtering applied. This discrepancy is not investigated in detail; the author cites a difference in 'various margins' as a possible source of error and in that case, the difference in SEI using in-service speed trials may be overstated. The negatives of this method include the time and financial cost associated with conducting the trial at sea during normal shipping operations; the ship must be loaded correctly, the weather reasonable and the test area needs to be in deep water and free of immediate traffic; each speed run needs to be repeated in each direction to compensate for the possibility of current (Munk 2006). This may be impractical for the merchant fleet, but possible for others; every ship in the Royal Navy and Royal Fleet Auxiliary that is fitted with a torque meter is

[3] www.propulsiondynamics.com

required to undertake sea trials for maintenance trigger measurements on a monthly basis (Walker and Atkins 2007).

Normalisation Model

The main problem of normalising is that the model used for the corrections may lead to uncertainties that arise from incorrect model functional form (or model parameters, depending on the training / calibration dataset integrity) due to either omitted variables or unknown effects. Instrument uncertainty also becomes important due to the wider array of input variables to the correction algorithm. On the other hand, from a richer dataset trends may be derived that have a more comprehensive physical significance and a larger dataset also reduces the uncertainty making the results applicable to both short and long term analysis (Flikkema 2013).

Data Filtering

The filtering approach is easy to implement and interpret (Flikkema 2013) as few associations and assumptions are made regarding variable relationships (environmental or operational, such as weather or draught); there are no corrections applied and therefore no assumed model form. In that sense uncertainty arising from epistemic model error, both parametric and model functional form, is reduced. However due to their low granularity it is difficult to understand underlying parametric relationships and significant quantities of data may be lost in the filtering step. Longer evaluation periods may be required in order to collect adequate volumes to infer statistically significant results and to reduce the uncertainty to a level appropriate to the application.

There is therefore a trade-off between uncertainty introduced due to the model form, added instrument uncertainty because of the inclusion of additional variables and the uncertainty arising from a reduced sample size. There are also possibilities of performance measurement that are a combination of the above options, for example filtering extreme conditions and then normalising. This approach might reduce the aforementioned 19% discrepancy in performance measurements depending on the processing techniques or data collection procedures, as found by Bazari (2007).

In the same way that there is no ‘standard’ data acquisition approach in terms of frequency of acquisition, type of measurement instrument or variables recorded, there is no standard post-processing or data analysis technique. It is hypothesised that lower fidelity noon reports may be less useful when trying to form and apply statistical models over the short term however they are still used frequently in many shipping companies in the form of filtered datasets for longer term trend analysis. Both techniques of filtering and modelling are therefore explored in this thesis in order to fully explore optimum modelling and data acquisition strategies.

2.2.2 Ship Performance Models

Data which is not filtered to include only baseline conditions (for example design speed, design draught and Beaufort scale less than BF4) needs to be corrected/normalised to this baseline. Because there are trade-offs between normalising and filtering then a study of both is relevant. Normalising requires a model of the ship performance in order to correct the shaft power or ship speed to the baseline, modelling techniques found in the literature are described below.

The literature in ship performance modelling is categorised according to the method. Theoretical models (physical / white box / deterministic) are based on first principles and retain what is known about the physical behaviour of a system. Hydrostatic relationships for example that determine ship trim and stability are, to an extent, based on Archimedes’ principle of buoyancy and the integration of volumes. The specific fuel consumption of the engine given its configuration and loading may also be theoretically modelled with thermodynamic relationships. Hydrodynamic relationships however (and hydrostatic relationships in some respects) that describe the ship’s resistance may be described by semi-empirical relations due to the complexity of the physics that is governed by the unsteady Navier-Stokes formula. Parameters that describe the hull geometry are linked to friction coefficients through regression analysis following systematic model experiments and full scale trials. In this work empirical naval architecture relationships such as these are grouped under the ‘theoretical’ heading because this conforms with the literature and clearly distinguishes them from the purely black box models that are not based on physical reality/behaviour at all.

Black box (statistical) modelling is purely driven by data which determines model parameters and structure. Typical linear regression model coefficients are found from the inputs (i.e. resistance is a function of wave height) while neural networks are based on the statistical model learning relationships between signals from the training data. Hybrid models incorporate both methods whereby black box methods are employed to reduce the residual between the white box predicted result and the desired result (according to the training data set). Semi-mechanistic, semi-physical and grey-box models are all variations on the hybrid modelling theme in that they combine black box and theoretical models in some way to improve the performance of either separately (Leifsson, Sævarsdóttir et al. 2008). A subtle difference between semi-mechanistic and grey box models is described by Braake, Can et al. (1998).

2.2.2.1 Theoretical Models

The elements that combine to describe the overall fuel consumption can be broken down as demonstrated in Figure 5 (Pedersen and Larsen 2009), which describes the major efficiencies that are applied to the effective power; engine, hull and propeller. The effective power is determined by calm water resistance (hull frictional, air, wave making/viscous and residual) and added resistance (due to wind, waves, current, ice, hull fouling and propeller fouling). The details of these calculations or empirical relationships are described later and in this section only ship performance models that are based on these formulations are discussed.

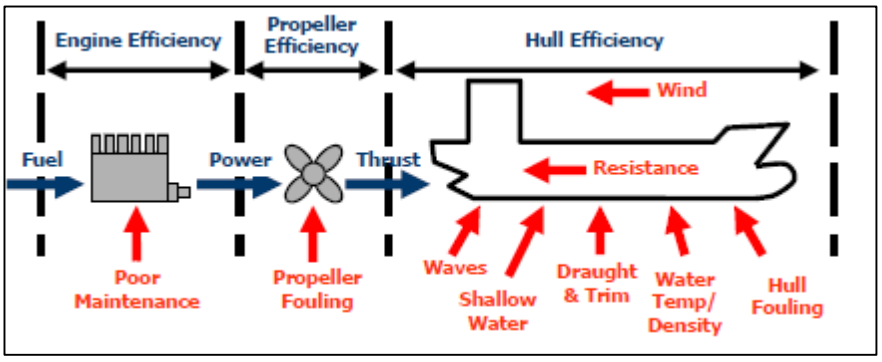


Figure 5: Variables influencing the propulsion performance,(Pedersen and Larsen 2009)

Theoretical models used at the design stage for sizing main engines and determining propeller and hull dimension characteristics are often based on model tests that determine calm water resistance. These are then validated during sea trials that occur

immediately after the ship has been built (see previous section, "dedicated speed trials"). The staple contributions to this area come from the procedures of the International Towing Tank Conference (ITTC) that relate resistance, R_T to hull wetted surface area, WSA and ship speed, V via a total resistance coefficient, C_T as follows:

$$R_T = C_T \, {}^1\!/_2 \, \rho_{sw}(WSA)V^2$$

Equation 1

The water density is denoted by 'ρ_{sw}'. The ship effective power is the product of ship speed and total resistance and effective power is related to shaft power according to equation 3. The resistance is the sum of viscous, residuary, frictional and air resistance, themselves defined by coefficients. A standard *model-ship correlation line* adopted at the 1957 ITTC then accounts for scale effects and enables model results to be applied to the full scale. The exact total resistance calculation method is outlined in the 1978 ITTC Performance Prediction Method. Semi-empirical models have been developed based on extensive towing tank tests followed by ship trials such as those by Holtrop and Mennen (1982), Guldhammer-Harvald (1965,1974), Hollenbach (1997, 1998) and Taylor-Gertler (1910, 1954, 1964). These are valid for different ranges of specific hull geometry parameters and a concise summary can be found in Schneekluth and Bertram (1998). As mentioned previously, including these semi-empirical models in the theoretical model section is not technically correct but conforms to the current literature. The performance of such models is demonstrated in sea trials and is generally considered to be acceptable from an initial design perspective. As noted by Logan (2011), however the correlation allowance that compensates for scaling effects varies according to tank facilities and the empirical model method. The details of this factor are often not made explicit due to commercial sensitivities which makes theoretical models difficult to evaluate and extrapolate to different ship types/sizes with complete confidence.

The underlying formulae in the theoretical models all have assumptions and uncertainties associated with them. Many of the theoretical models that measure the ship's resistance remain un-validated in the scenario in which they are applied,

Logan (2011). One study that presents a detailed uncertainty analysis for added resistance experiment of a KVLCC2 ship finds a resistance of 16% and 9% for a short wavelength case and moderate wavelength case, respectively (Park, Lee et al. 2015). After delivery, the engine wear and hull and propeller fouling create difficulties for validating models that describe these features as each added resistance cannot be very well isolated and attributed to its source, especially as fouling for example, is significantly affected by the ship's operating profile (sea water temperature, time in port, periods of loitering, water salinity). Further, the weather conditions limit opportunities for validation since only calm conditions can be chosen to minimise wave and wind effects to try and isolate as far as possible the fouling effect. Validating models that describe the added resistance due to wind and waves require specific conditions and any full scale validation process requires the trial to be carried out in each direction to minimise the effect of the current which is also difficult given the ship's commercial time constraints. Full validation requires a large dataset that represents a wide range of ship operating conditions and this is costly and may take many years to accumulate, especially if the data is collected only on a daily basis.

In a thesis on the monitoring of hull condition of ships, Hansen (2010) includes theoretical models for added resistance in wind and waves and the effect of steering (rudder angle) and shallow water. Eljardt (2006), in an advance version of a fuel oil consumption monitoring and trim optimisation programme, includes the effect of sea state, wind, course-keeping (rudder angle) and shallow water. It has been found that there are inconsistencies surrounding which specific added resistance factors should be included. Hansen (2010) corrects for the wind/weather condition at the time of measuring (twice every month) in order to calculate equivalent power requirements for calm weather at a reference speed and draught and therefore ultimately to measure the degree of fouling (also known as normalisation). These models are not validated with data for the ships to which they are applied and the uncertainty due to the assumptions on which they are based is not quantified in each case. Also there is no described method that accounts for interaction effects between components of added resistance since separate models are used for each. This is an important disadvantage in these models.

2.2.2.2 Statistical Model

Data derived modelling techniques are heavily dependent on the data quality; black box modelling for the prediction of fuel consumption is relatively recent as sensor quality has improved and data acquisition systems and signal processing techniques have become more mainstream. There are two different approaches in black box modelling; statistics and Computational Intelligence (CI). Statistics involves the analysis of the characteristics of a population from a data sample and is grounded in well-established mathematical principles. CI approaches are based on nature-inspired methods such as fuzzy systems, neural networks (NN) and evolutionary computation which are deployed to understand and structure complex systems from unstructured data.

The major disadvantage of black box models and in particular the CI approach, is that it is difficult to detect the significance of input variables and to understand the actual physical relationships between variables. Statistical methods also frequently fail when dealing with complex and highly non-linear data (Karlaftis and Vlahogianni 2011). Furthermore, the model functional form must be based on 'a priori' and this prior knowledge that informs the definition of the appropriate functional form may be unknown. Statistical models struggle with multicollinearity, outliers or noisy data and there are strict rules regarding the error term. Neural networks are better at dealing with non-linearities however the interactions between variables and implicit assumptions on which the model is based are hidden meaning that they lack transparency and reproducibility. The training process of NN's is slow and the time resource to devleop these models is significant (Karlaftis and Vlahogianni 2011).

With reference to ship performance monitoring, Pedersen and Larsen (2009) describe methods of predicting ship propulsion power and they compare artificial neural networks (ANN) and statistical linear and non-linear regression models, in terms of accuracy of fuel consumption prediction. While they have found that ANNs can be used to successfully predict propulsion power and exhibit improved performance, they also find errors in validating that do not seem to vary when the input variable combination is altered. They note that it is thus difficult to detect which input

variables were most significant. This indicates that a theoretical model would provide more information. Some measurements were supplied through onboard sensors however the weather data had to be taken from hind cast information which is estimated and observed and therefore may be subjective. Noon report data is also subject to human error. Petersen, Jacobsen et al. (2012) compare two statistical modelling approaches used to predict fuel consumption and speed from a set of measured features. They compare gaussian processes (GP) to ANNs. The significance of each of the input signals is assessed from the strength of the correlation between the input and output signals; crosswind was seen to be of least relevance while heading and trim is more important than draft and the propeller pitch (and rpm) is the most relevant to the fuel consumption. They note however that this does not indicate the role of the variables in the underlying physical system. It is described how this instantaneous model is a step towards a full dynamic model which would be more desirable for real time operation optimisation, for example for trim optimisation. A further paper by the same authors (Petersen, Jacobsen et al. 2011) implements a time-delay neural network to predict the response of a range of dynamic variables (speed, trim, draught, heading) to a change in control variable (pitch, rudder angle, current, headwind and crosswind) again, not a full state-space model but speed and fuel consmption are predicted which are two of the variables in a state space model. Finally, the authors implement and compare a Gaussian Mixture Model for the same problem (Petersen, Jacobsen et al. 2011), this they find to be almost as good and less computationally demanding than the GP model.

It is seen that there is room for further research into black box models with regards to ship performance monitoring. There are very few papers focussed on this topic and in those that are published the model measurement and predicition accuracy is unclear. There are only a small number of papers written and often the exact details of the models are not published. The lack of transparency of CI models and the requirement of ‘a priori’ definition for model functional form in statistical models are the most significant barriers to their adoption for ship performance measurement.

The literature review has highlighted the main features of the theoretical and black box modelling methods, these are summarised in Table 2.

	Theoretical Method	Black box Method
Positive	– Does not require historical data – Can evaluate new technologies – Can extrapolate beyond the given data range	– Can be more accurate as based on historical empirical data (i.e. neural networks have high prediction capability)
Negative	– Model structure and parameters are based on prior knowledge. – Require detailed ship information and large amount of input data. – Approximate as based on design condition – Trade-off: detail requires large number of parameters but fewer leads to a more robust model – Accuracy of prediction depends on assumptions and uncertainty implicit in the models – May lack predictability – Many of the theoretical assumptions are not validated with at-sea or sea trial data – Empirical assumptions are often based on old datasets and are not updated for modern ship types/hull forms	– Too high resolution for regression analysis to be applicable for different ships; poor extrapolation/scalability capacity – Co-linearity problems; it is unknown which independent variables impact the EC; improve by further distinguishing between types – Cannot evaluate new technologies – In NN particularly, the inputs have no physical significance – Statistical models cannot easily deal with non-linearities – There are strict rules concerning the error – Require prior knowledge of model functional form

Table 2: Advantages and disadvantages of different modelling methods

2.2.2.3 Hybrid Models

The pioneering work in hybrid modelling of a ship's fuel consumption and power requirement monitoring was conducted by Journée, Rijke et al. (1987), a relatively recent development. The delayed application uptake is attributed to the reliance of the method on sensor quality and robust data acquisition methods; the author describes experimenting with a similar system in 1984 that was unproven due to poor

sensor quality. In this work, measured signals were used to adjust the coefficients of the hydrodynamic model over various draughts, trim and speed combinations (all in calm sea) in order to predict vessel speed, power and fuel consumption. These variables (draught, trim and speed) are then advised by the algorithm in operating conditions input by the user and may be optimised by manual iteration. It is shown mathematically how fuel quality, wind resistance and hull/propeller fouling allowances can be incorporated into the model (by a comparison of modelled and actual power required for a given speed in calm conditions) but experimentation with and the accuracy of these modules are not reported in the paper. The results are validated from the training set but it is not clear that a test data set is reserved for error calculation on a previously unseen data set. It is stated that wind and wave conditions are incorporated in to this model however this requires manual data entry every 20 minutes and these (especially for the weather conditions where wave height is a visual estimation from the bridge) are likely to be subjective and vulnerable to human error. Predictions were found to be poor in bad weather conditions owing to inadequate or incomplete weather measurements.

A commercial model that predicts hull fouling (for optimum hull cleaning scheduling) is described by Munk (2006). This used weekly observed recordings of performance data that were taken during a two hour period with constant navigation conditions. A comparison between observed values and the hybrid model output values yields the speed and added resistance due to fouling (for constant weather/loading conditions). In this case the model is based on first principal formulae and approximation formulae with empirical constants. This is subsequently improved with tank tests and trial data, plus statistical analysis of performance observations. Specific details of the model (input parameters and variables), methods and accuracy of results are not disclosed, except to say that the added resistance may be found with an accuracy of approximately 1%, with deviation from the mean for a single set of observations being around 3%. This work describes a method that is used to measure hull and propeller fouling and as such the data points are taken in calm weather conditions only. Again, the model's reliability depends greatly on the quality of the observed data that is used and there may be significant subjectivity and human error associated with this, particularly when calculating averages over two hours. It does not describe how the crew assess whether the wind and wave

conditions have reached stability for example and how interpretation is kept consistent between crew members.

The continued development of sensor quality, particularly of more accurate weather data availability, portable computer power, servers to cope with large data volumes, high speed networks and satellite data connections, has led to an emerging market for software that can compile measured data, record and analyse trends and model a vessel's performance[4]. Research funded by Marorka led to the next significant work in this area which was published by Leifsson, Sævarsdóttir et al. (2008), they realise the potential market for fuel consumption reduction tools driven by rising fuel costs and likely future regulation. This work incorporates hydrodynamic models, onboard sensors and a feed forward neural network model to predict the fuel consumption and speed of a container vessel. They compare and report the advantage of using a grey-box or black box method over a theoretical-only model for fuel consumption predictions during validation (65% reduction in RMSE). The grey-box method is found to have superior extrapolation abilities over both the other two methods, particularly in more extreme environmental conditions, although it is noted that their theoretical model does not include the effect of added resistance in waves, which would be significant in an extreme condition. The white-box model is superior to the other two over the range of operating values (where the black box particularly fails outside of normal fuel consumption and speed values); this suggests that the performance of the white box model could be improved in the more extreme environmental conditions if wave data and a theoretical wave model were included. They discuss the use of a black box 'sub-model' to work within the framework of the white box model to account for the resistance effect of waves, the inclusion of wave data would be a vast improvement on this model and may affect their conclusions. Also, although the quality and accuracy of the sensor data is significantly improved from the Journée, Rijke et al. (1987) study, enabling superior model predictions, the data is collected over a narrow vessel speed variance which may have limited the network training and have affected the comparisons between white, black and grey box methods.

[4] Sea trend from Force Technology, CASPER from Propulsion Dynamics, Fleet Manager VVOS from Jeppersen and Eniram from Marorka, Voyage Optimisation from NAPA, GlobalTide from TideTech

2.2.3 Summary of Performance Measurement

As discussed (see section 2.2.2.2), there are clear deficiencies in models grounded purely in the theoretical assumptions (including empirical relationships) and those based on purely statistical methods. There are a limited number of up-to-date works into hybrid modelling, the details of which and the accuracy of the results are generally unpublished. Generally, there is a lack of rigorous analysis and evidence that quantifies in a consistent way the accuracy of ship performance monitoring methods. The introduction to this literature review has highlighted in detail the necessity for uncertainty analysis to be more routinely conducted in ship performance measurements and this is again highlighted in this review of ship performance modelling techniques themselves. This section has shown that, in order to describe the ship's performance, there are multiple data processing options; the data may be filtered or normalised and the normalisation model may be formed from a theoretical, statistical or hybrid technique. It was described previously (see section 2.1.2, end) that the different possible combinations of data collection methods are likely to result in a range in the uncertainty of the performance indicator. In the same way, the various processing techniques will also produce different uncertainties. In particular, it has been seen that the models may be based on a different selection and range of different parameters, for example wind and wave models may or may not be incorporated. There is no clear quantification of how these models perform depending on this parameter selection and how this affects the performance of the hybrid / statistical / theoretical models. This is important because it defines how much data must be discarded in the processing and ultimately the time period over which the data must be collected. It also determines the state of our knowledge about how the ship reacts in different operational and environmental conditions and this could be vital information for performance prediction and optimisation.

2.3 Uncertainty Analysis

This section explores literature relevant to the analysis of performance measurement uncertainty. Due to the shortage of literature specific to ship performance measurement uncertainty, approaches from other industries and research areas are used to explore the sources of uncertainty, how it is characterised and the methods

used to quantify it. This helps understand the general principles in order to apply this understanding to addresses uncertainty in ship performance measurement.

No measurement can perfectly determine the value of the quantity being measured (the measurand); imperfections arising from sensor manufacturing process, operator error, environmental fluctuations and misspecifications etc., can lead the measurement to deviate from the measurand value (Gleser 1998). The aim of an uncertainty analysis is to describe the range of potential outputs of the system at some probability level, or to estimate the probability that the output will exceed a specific threshold or performance target value (Loucks 2005).

The quantification of uncertainty is most widely formalised through probability and there are two main schools of thought regarding probability; the classical (frequentist) and the Bayesian (subjectivist). The former relates the probability of an event occurring to the frequency with which it occurs in a given number of trials. The Bayesian approach relates the probability of an event occurring to the degree of belief that a person has that it will occur given the state of information that the person has (Kacker and Jones 2003). These definitions are referred to in this chapter.

Uncertainty analysis methods have evolved in various ways depending on the specific nuances of the field in which they are applied. However, a key document in the area of uncertainty evaluation is the ‘Guide to the expression of uncertainty in measurement’ (GUM) (JCGM100:2008) which provides a procedure adopted by many bodies (Cox 2006). The GUM defines a formulation and calculation stage as follows:

- Formulation
  - Define the output quantity, Y (the measurand)
  - Identify the input quantities on which Y depends
  - Develop the measurement model (relate Y to the inputs), through data reduction equations (DREs) / transfer functions
  - Characterise: Assign a mathematical structure to describe the uncertainty and to determine the numerical values of all required parameters of the structure

- Calculation
  - Propagate the inputs through the measurement model to formulate the output distribution for Y
  - Summarise: Expectation, standard uncertainty and coverage interval containing Y with a specified coverage probability

The first part of the formulation stage (definition of outputs, inputs and measurement model functional form) is concerned with the selected ship performance model that is the focus of chaptersChapter 5,Chapter 6 and Chapter 7. The following sections review the literature concerning uncertainty characterisation and propagation.

2.3.1 Uncertainty Characterisation

The GUM framework is itself derived in part from the work of Coleman (1990) who introduced the balanced treatment of precision and bias errors. They also describe a method to treat correlated errors and small samples sizes. The nomenclature and definitions of Coleman and Steele are consistent with those of the ANSI/ASME[5] standard on measurement uncertainty. Precision error is the random component of the total error, sometimes called the repeatability error, it will have a different value for each measurement, it may arise from unpredictable or stochastic temporal and spatial variations of influence quantities, these being due to limitations in the repeatability of the measurement system and the facility (equipment / laboratory) and limitations due to environmental effects. The bias error does not contribute to scatter in the data but is the fixed, systematic or constant component of the total error which, in a deterministic study is the same for each measurement. The basic premise of the GUM framework is twofold; first, to characterise the quality of the output in terms of the systematic and random errors which are then combined to obtain the overall uncertainty in a probabilistic bases and secondly, it includes the analyst's belief in their knowledge of the true value of the measurand, quantified in terms of probabilities. This is a refinement to traditional error analysis in which the output is a best estimate plus systematic and random error values. The GUM classifies uncertainties according to the method used to evaluate them; Type A evaluation of uncertainties is based on statistical methods or repeated indication values, i.e.

[5] American National Standards Institute / The Association for the Study of Medical Education

Gaussian distributions derived from observed frequency distributions. Type B evaluation of uncertainties is based on scientific judgement (any basis other than statistical), this is a priori distribution based on a degree of belief, a feature of Bayesian inference. In Bayesian inference, a random variable with a corresponding distribution is also assigned to quantities which cannot be treated as random in usual statistics because conventional (frequentist) statistics fail where there is insufficient empirical data (Weise and Woger 1992). The probability distribution provides a statement about incomplete knowledge based on any rational relevant information available. If there is no specific knowledge one can only assume a uniform or rectangular distribution of probabilities should be assigned. Both types of evaluation are based on probability distributions. The philosophy of the GUM in A and B classifications states:

"3.3.4 The purpose of the Type A and Type B classification is to indicate the two different ways of evaluating uncertainty components and is for convenience of discussion only; the classification is not meant to indicate that there is any difference in the nature of the components resulting from the two types of evaluation. Both types of evaluation are based on probability distributions, and the uncertainty components resulting from either type are quantified by variances or standard deviations"

The type A and B classification of inputs is useful in the context of the GUM to determine the method used to compute their expectation and standard uncertainties. However, once the uncertainty of a measurement is given it is used in the same way regardless of whether it is type A or type B. As stated, this type A and B classification is not to indicate differences in the nature of the components of uncertainty but more insightful is the distinction between aleatory and epistemic uncertainty. Aleatory uncertainty (also called randomness, variability, objective uncertainty, or irreducible uncertainty) arises from inherent variation due to natural stochasticity or environmental variation across space or through time. The addition of more information (increased sample size) enables a more complete definition of the form and parameters of the representative probability distribution function (PDF) however the uncertainty is irreducible for a given set of processes. Epistemic uncertainty (also called incertitude, ignorance, subjective uncertainty, or reducible

uncertainty) arises from incomplete knowledge about the world. Epistemic uncertainty can be reduced and in principle eliminated if sufficient knowledge is added (Roy and Oberkampf 2011). As described by Ferson and Ginzburg (1996); while ignorance can be reduced by additional study or improved techniques, variability has an objective reality that is independent of our empirical study of it. For example, if the sensor precision is to reflect the inherent sensor variability that may be described by a PDF representing a large sample then it is a form of aleatory uncertainty. Sensor bias and drift, since it is related to the change in bias over time, reflects the deviation from the truth which is unknown and this therefore is epistemic uncertainty. This aleatory/epistemic classification is prevalent in the risk assessment community (Roy and Oberkampf 2011), see for example Morgan and Henrion (1990) and the US Environmental Protection Agency risk assessment guidance (EPA 2001), and also in structural reliability analysis in which a well-developed recommended practice has been established in the offshore industry in association with DNV (Skjong, E.B.Gregersen et al. 1995).

This aleatory/epistemic classification is useful in determining the approach to the uncertainty propagation through the model of ship performance, either by probability theory (for aleatory uncertainties) or interval analysis (for epistemic uncertainties), see (Roy and Oberkampf 2011) for further discussion. These can then be propagated through the model to form a p-box of the output that is a cumulative distribution function that displays the overall uncertainty in the performance indicator as an interval valued probability.

2.3.2 Uncertainty Calculation: Propagation

The GUM specifies three methods of propagation of distributions:

a. The GUM uncertainty framework, constituting the application of the law of propagation of uncertainty
b. Monte Carlo methods (MCM)
c. Analytical methods

The analytical method gives accurate results involving no approximations however it is only valid in the simplest of cases while methods a. and b. involve approximations. The GUM framework is valid if the model is linearised and the input PDF's are Gaussian. This is the framework followed by the AIAA[6] guidelines (AIAA 1999) and the ITTC guide to uncertainty in hydrodynamic experiments (ITTC 2008), of which relevant examples include applications to water jet propulsion tests (ITTC 2011) and resistance experiments, ITTC (2008), an Uncertainty Assessment for Towing Tank Tests, Longo (2005) and finally an uncertainty analysis for added resistance experiment of KVLCC2 ship by Park, Lee et al. (2015). In these examples, sensor measurement repeatability (same conditions, equipment, operator and location) is identified as precision limits for each variable and are described by a distribution function or simply by a standard deviation (Bitner-Gregersen, Ewans et al. 2014). The bias limit for each elemental input may be present as a fixed (mean value) or as a random variable, in the latter case it would be defined by the band within which you can be 95% confident that the true value lies (AIAA 1999), i.e. the band in which the (biased) mean result, would fall 95 percent of the time if the experiment were repeated many times under the same conditions using the same equipment. The AIAA / ITTC documentation follows this method however use the terminology 'precision limits' and 'bias limits' to represent type A and B methods of evaluation of uncertainties.

If the assumptions of model linearity and Gaussian input PDF's are violated or if these conditions are questionable, then the MCM can generally be expected to lead to a valid uncertainty statement (the use of MC analysis within the GUM framework is documented in a supplement to the standard (JCGM101:2008 2008), see 5.10.1 of this supplement for the specific conditions in which the MCM is valid). Furthermore, the GUM also involves finding partial derivatives of the model to determine the sensitivity coefficients (describing how the estimate of the output would be influenced by small changes in the inputs) which combine to form an estimate of the output and its associated standard uncertainty. These sensitivity coefficients in some cases may be difficult to find. In the probabilistic risk assessment field, Monte Carlo Analysis is perhaps the most widely used probabilistic method (EPA 2001). Relevant

[6] AIAA The American Institute of Aeronautics and Astronautics

examples in the shipping industry include applications in sea trial uncertainty analysis (Insel 2008). A further advantage in the MCM is in the representation of the output uncertainty; the GUM approach is limited to generally simply the mean and standard deviation while the PDF is assumed to be Gaussian. Through the MCM a more insightful numerical representation of the output is obtained and it is not restricted to a Gaussian PDF. This is particularly restrictive in the GUM when the output is used to describe the expanded uncertainty interval whereby a coverage factor is assigned under the assumption that the output is normally distributed. This assumption may be unjustified and the coverage factor so determined may be incorrect (Kacker and Jones 2003). The departures from the GUM in the MCM are (see GUM 5.11.4):

a. The input uncertainties are based on a PDF (rather than associating standard uncertainties with estimates of x_i and X_i) and therefore separation of the inputs into type A and type B is not necessary.
b. Sensitivity coefficients and therefore the partial derivatives of the model are not required
c. A numerical representation of the output is obtained and therefore not restricted to a Gaussian PDF.
d. The output distribution may not be symmetric and therefore the coverage interval for the output may not be centred on Y… The choice of coverage interval may require consideration.

In the probabilistic risk assessment field, Monte Carlo analysis is perhaps the most widely used probabilistic method (EPA 2001), relevant examples in the shipping industry include applications in sea trial uncertainty analysis (Insel 2008). In the marine risk assessment field a MC method has also been applied in a probabilistic analysis of a ship specific decision support system for onboard navigational guidance (Tellkamp, Hansen et al. 2008). In this example the aim is to evaluate the risk associated with the hydrodynamic response of the ship given the ship's loading, environmental and operational conditions. An MCM was utilised in order to assess the relative effect of different input parameter uncertainties.

A further advantage of the MC method is that the input uncertainties are based on probability distributions (rather than associating standard uncertainties with estimates of each input), therefore separation of the inputs into type A and type B is not necessary. Instead, the uncertainties may be classified in to epistemic and aleatory which may be used to determine different methods for the propagation of uncertainty; either by probability theory (for aleatory uncertainties) or interval analysis (for epistemic uncertainties). As previously discussed, aleatory uncertainty is characterised by a frequency PDF with known parameters provided that there are sufficient samples of the stochastic process to be able to fully describe its form and parameters. Otherwise, if the parameters are unknown and cannot be derived experimentally, or if there is insufficient experimental data to fully define the probability distribution, then the uncertainties are classified as epistemic. The traditional approach is one in which epistemic uncertainty is represented as either an interval with no associated PDF, or where the assigned PDF represents the belief of the analyst (as in the GUM). However, it is recommended (in risk analysis for example) not to mix the two types of uncertainties into a single probability distribution in the measurand as the resulting interpretation would be meaningless (Morgan and Henrion 1990), although this very much depends on the application of the analysis. Alternatively, epistemic uncertainty may be represented as an interval valued quantity meaning that the true (but unknown) value can be any value over the range of the interval with no associated likelihood or belief that any value is truer than any other value (Roy and Oberkampf 2011). This probability bounds analysis (PBA) allows probability theory for stochastic input variables to be combined with interval analysis for variables which we are ignorant about their distribution. The output measurand is then represented by a probability box (p-box) that displays the overall uncertainty in the performance indicator as an interval valued probability, so although p-boxes model both uncertainty types, they are never confounded and are clearly distinguishable in the results. Examples and texts on the subject include Ferson and Ginzburg (1996), W. Tucker and Ferson (2003) and Ferson, Oberkampf et al. (2008).

The method of error propagation by PBA is achieved by either MC simulation or by analytical probability bounds. The MC analysis essentially involves nesting one Monte Carlo simulation within another, in order to explore how variability and

uncertainty interact in the calculation of the distribution of the measurand. Typically, the inner simulation represents natural variability of the underlying physical processes, while the outer simulation represents the uncertainty about the particular parameters that should be used to specify inputs to the inner simulation (W. Tucker and Ferson (2003). The outer loop takes random samples N times of the interval input quantities to produce the grey probability distribution in Figure 6. While the probability bounds analyst analytically calculates the bounds on the output distribution (p-box) obtained from all possible outer loop inputs. As the number of Monte Carlo iterations increase the two methods converge on the same answer.

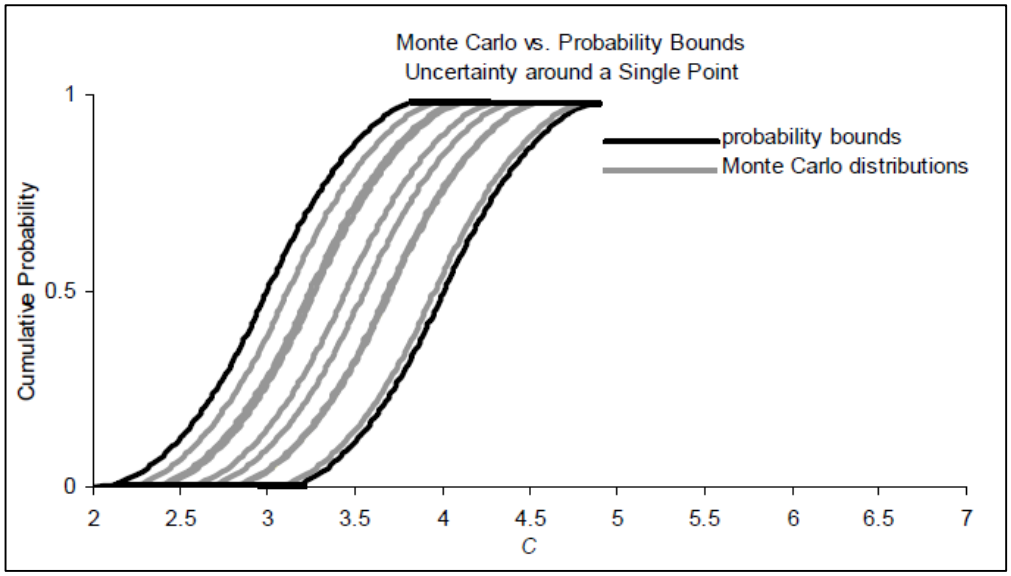


Figure 6: Graphical representation of probability bounds analysis and Monte Carlo methods (ref. W. Tucker and Ferson (2003)

The horizontal span of the probability bounds are a function of the variability in the result, the vertical breadth is a function of our ignorance (Ferson and Ginzburg 1996).

It is recommended in the GUM to assign a uniform or rectangular distribution of possible values in the situation where there is no specific knowledge about the possible values of the input within an interval, i.e. it is equally probable for the measurement to lie anywhere within it (see GUM section 4.3.7). Propagating uncertainty from an assumed uniform probability distribution, where the input quantity is randomly selected, does not yield the same result as elementary interval analysis because it assumes more information than in the original question (Ferson

and Ginzburg (1996). In the GUM approach, simultaneous selection of random variables from the middle of the interval of two input quantities is more likely than simultaneous selection of random variables at the extremes, therefore if, as an example a simple model that represented the product of the two inputs were assumed, then there would be a higher concentration in the centre of the output PDF than at the extremes. So, if the nature of the uncertainty is epistemic (ignorance), then probabilistic methods are inappropriate. However always assuming interval analysis rather than applying some stochastic variability to the input will cause unnecessarily conservative answers. This is relevant to the way instrument and model bias is handled, see section 9.3.1.

2.3.3 Sources of uncertainty on modelling and measurement

In reliability analysis for the offshore environment uncertainties are classified according to the following structure (Bitner-Gregersen and Hagen 1990) :

i. Model uncertainty:
    a. Distribution uncertainty: Refers to the approximation of data to a specific distribution function
    b. Climatic uncertainty: Refers to when the full underlying physical relationship/interaction is not captured by the model
ii. Statistical uncertainty: The uncertainty in the estimation of the distribution parameters and sample size
iii. Data uncertainty, due to instrument imperfections and sampling variability (i.e. sample averaging due to finite instrument temporal resolution). Traditionally, error in a measurement is viewed as having two components (JCGM100:2008 (2008)
    a. Bias; which does not contribute to scatter in the data, fixed systematic and constant (e.g. scale resolution) and, cannot be determined statistically
    b. Precision; random errors, arise from unpredictable or stochastic temporal and spatial variations of influence quantities. They are due to limitations in the repeatability of the measurement system and to

facility and environmental effects and may be estimated by statistical analysis

These authors are only concerned with the uncertainties associated with the data for the offshore environment (and specifically with regards to metocean modelling) and the accuracies are presented in terms of systematic and random error. This classification is also used in the DNV uncertainty modelling section in its guideline for offshore structural reliability analysis (Skjong, E.B.Gregersen et al. 1995) which defines all of these uncertainties as random (stochastic) variables and epistemic by nature.

Sources of uncertainty in scientific computing are specified as model inputs, numerical approximations and model form. In the hydrological systems community, in relation to water resources and planning management (Loucks 2005), uncertainty is classified as knowledge uncertainty (analogous to epistemic uncertainty) and natural variability (analogous to aleatory uncertainty). The former is comprised of parameter value uncertainty (including boundary conditions) and structural uncertainty (due to numerical approximations and residual model error). The latter is the spatial or temporal variability of model input data, particularly relevant for natural, meteorological / hydrological phenomenon. The EPA (2001) guidelines divide risk assessment epistemic uncertainty into parameter, model and scenario uncertainties (missing / incomplete information).

The above shows that in many other fields there are systematic frameworks which have been derived from general frameworks in order to rigorously analyse uncertainty. These frameworks include an ability to separate out different sources of uncertainty. The description in section 2.2.2 of ship performance models shows that there are many overlapping features of these frameworks (e.g. particularly metocean data and offshore structures) and this implies that it should be possible to extend the generalised frameworks and borrow from equivalent sector's frameworks to develop a framework specific to ship performance measurement. Common features, relevant to ship performance measurement are modelling, sampling and instrument uncertainty and these are explored in more detail below. The above frameworks also

all categorise the source according to its nature; aleatory or epistemic and there is therefore a focus on this below.

Sampling uncertainties are generally deemed to be aleatory in nature, arising when only a portion of the population of actual values is measured. This portion may occur either in physical location, population sector or in the time domain; metering occurring only once per day for example. Sample size, bias and sample homogeneity are all significant. Arising due to the sampling of a finite experimental data set from a potentially infinite population, where repeated experimentation will yield more precise results. The effect of sample size is however categorised as epistemic by W. Tucker and Ferson (2003) because a small sample size means there is incomplete knowledge.

Modelling uncertainties are generally epistemic in nature and may be structural or parametric (Loucks 2005). Parametric uncertainty pertains to the correct values of model parameter coefficients. Uncertainty about model structure is related to the correct explanatory variables and their relationships. It is specifically broken down by Chatfield (1995) as arising in different ways such as:

a) Model misspecification (e.g. omission of relevant variables or inclusion of irrelevant variables).
b) Specifying a general class of models of which the true model is special but unknown.
c) Choosing between two or more models of different structures, relationships between variables and using the incorrect functional form for example.

Omitted variables include for example, those that define short term dynamic motions and steering (rudder angle for example). Sources of modelling errors are further expanded in EVO[7] (2012) to include the effects of using out of range data and the effect of using insufficient data / data shortages. These originate as statistical errors due to a finite number of repeated observations (during the sea trials for example) but then remain constant (or moving but predictable) when implemented in the method (for example a diversion from the true V_{ref} at each measured power). The GUM

[7] EudraVigilance Organisation

method also stipulates the inclusion of model corrections as inputs whose uncertainties must be considered. These bias errors (or bias and precision errors if the model estimates involve direct measurement of parameters in order to derive the model) may be subjective. The DNV guidelines specify that model uncertainty (also bias and/or precision errors present) be obtained from tests/measurements or subjective choices. Model uncertainty may be attributable to the inability of the model to perfectly describe the true characteristics of the system, i.e. due to hysteresis effects, omitted variables or secondary effects of some variables (temperature gradients, accelerations etc.).

Measurement uncertainties arise from variations attributable to the basic properties of the measurement system, these properties widely recognised among practitioners are repeatability, reproducibility, linearity, bias, stability, consistency, and resolution (ASTM[8] 2011). The nature of measurement uncertainties may be defined as epistemic or aleatory and defined by a normal distribution with a quantified standard deviation and a mean about zero (Dunn 2004). Roy and Oberkampf (2011) explain further that it depends on the question being asked. If for example the true magnitude of a specific observation is required then the correct answer is that the single true value is unknown (epistemic), however if the question pertains to the measurement of any observation then the answer is a random variable given by the PDF determined using the measurement information from a large sample (aleatory). In error analysis the quantification of the data uncertainty is presented by an estimation of a systematic error (bias) and precision (random error). Sensor measurement repeatabilities (same conditions, equipment, operator and location) are identified as precision limits for each variable and are described by a distribution function or simply by a standard deviation (Bitner-Gregersen, Ewans et al. 2014). In the DNV guidelines for offshore structural reliability analysis (Skjong, E.B.Gregersen et al. 1995), it is suggested that measurement uncertainty be obtained from manufacturer specifications, lab tests or full scale tests, where bias and precision errors are both possible and the relative magnitude of each depends on how well the system is studied and its intrinsic underlying stochasticity. In this thesis the measurement error arising due to imprecision is categorised as aleatory.

[8] American Society for Testing and Materials

This detail and categorisation of the various sources of uncertainty is helpful because in the field of ship performance, quantification of uncertainty is particularly difficult because the truth is unknown. If perfect information is available, uncertainty can be quantified by comparing a benchmark with a given measurement. In the case of ships for which noon report or continuous monitoring data is available, there is no authoritative benchmark. Consequently, this has to be derived from available data, which has its own inherent uncertainties. Disaggregating these presents a challenge which has yet to be investigated in depth and there is no evidence of this kind of investigation found in the literature.

2.4 Chapter Summary

The introduction to this chapter identified the wide range of applications for ship performance measurement and why this is a useful and relevant tool in many aspects of the shipping industry. Examples include operational real time optimisation, maintenance trigger requirements, evaluation of technological interventions, charter party analysis and vessel benchmarking. It has been shown how important the quantification of uncertainty is to other industries and how this is linked closely with cost-benefit analysis and decision making, particularly with respect to making informed investment decisions. The relevance of this to the shipping industry has been stated. High level detail of the variability in the data collection and processing methods has been presented. This variability delivers a range of uncertainties in the overall ship performance metric and this therefore brings about questions surrounding trade-offs and the optimum DAQ strategy for minimum uncertainty, or for the level of uncertainty appropriate to the application.

The performance measurement section has shown that, in order to describe the ship's performance, there are multiple data processing options; the data may be filtered or normalised and the normalisation model may be formed from a theoretical, statistical or hybrid technique. It was described previously (see section 2.1.2, end) that the different possible combinations of data collection methods are likely to result in a range in the uncertainty of the performance indicator. In the same way, the various processing techniques will also produce different uncertainties. There is no clear

quantification of how these models perform depending on this parameter selection and how this affects the performance of the hybrid / statistical / theoretical models.

The uncertainty section has explored literature relevant to the analysis of performance measurement uncertainty. Due to the shortage of literature specific to ship performance measurement uncertainty, approaches from other industries and research areas are used to explore the sources of uncertainty, how it is characterised and the methods used to quantify it. Uncertainty is characterised in terms of precision/accuracy and aleatory/epistemic uncertainty. The GUM method is introduced and three main methods of uncertainty propagation are defined; the law of propagation of uncertainty, Monte Carlo Methods and analytical methods. The advantages and disadvantages of each are described. Sources of uncertainty are categorised as modelling, sampling or measurement.

Chapter 3. Research Questions

The previous section has identified that methods for the rigorous analysis and quantification of ship performance are relevant and have many applications within the shipping industry. It has also highlighted the reasons for the importance of defining the uncertainty associated with this ship performance measurement. An overview of ship performance analysis methods is presented and the two major methods are identified to be filtering or normalising. Normalising procedures require a normalisation model to be established and these are summarised below.

Theoretical models are all based on assumptions and have a varying degree of uncertainty associated with them. The resistance component of the fuel consumption calculation is rarely validated using data in the situation in which it is applied owing to time and cost constraints. There are inconsistencies between models and the effect on fuel consumption of interactions between components is not accounted for. Correlation allowances vary but are ambiguous. Theoretical models tend to be based on 'at design' conditions and relate to specific ship types which may not relate accurately to modern hull forms.

Black box models have shown to have good predictability performance over the conditions in which they are trained, however these may perform badly when extrapolated outside of this parameter range. It is difficult to make inferences from artificial intelligence models and some statistical models deal poorly with non-linearities and require prior knowledge of the model functional form. This is a major hindrance for their application to the complexities of a ship's propulsion system.

Hybrid models appear to have the most potential for development in the field of ship performance analysis. A significant advantage of these is that meaningful information about the ship's propulsion system can be derived that relates to the physics of the system being studied. This knowledge can then become inputs to optimisation studies or prediction algorithms. The problem of unknown functional form is solved by the use of theory to inform the underlying relationships and interactions between variables and then the accuracy of the model is maintained by

adjusting the model parameters to match the actual ship's operational performance using the dataset.

The literature review showed that there are disadvantages to purely theoretical and to purely black box models. It has also showed that the development of hybrid models for ship performance analysis is relatively recent and that little has been published to date. This suggests that a novel and valuable research question would be: "Can hybrid models outperform theoretical/ black box models in the provision of detailed and quantifiable knowledge about a ship's performance?" The efficacy of knowledge is related to two things; accuracy and depth of insight. Accuracy relates to the trueness and precision of the ship performance metric. Depth of insight refers to the degree of quantifiable and reliable evidence that can be derived regarding how the ship performs in its operational condition and, simultaneously, how it interacts with its environment. For this reason, and those discussed in section 2.1.2, the uncertainty associated with this measurement is important and highly relevant to research question 1 because the degree of insight and the model accuracy is determined by the type of model (black box, theoretical, hybrid).

This leads to the research question and two sub-questions:

1. **Can hybrid models outperform theoretical/ black box models in the provision of detailed and quantifiable knowledge about a ship's performance?**
    a. **Can hybrid models deliver a more accurate ship performance metric?**
    b. **Can hybrid models provide deeper insight into the drivers and influences of the ship's performance?**

The literature review provides an understanding of the current methods for uncertainty quantification and the parallels and divergences between industries, however no examples in the literature are found that provide a rigorous analysis of the quantification of the uncertainty of ship performance indicators. This leads to the second research question:

2. **How can the various sources of uncertainty in ship performance measurement be individually quantified?**

This is also a fertile area of research because there are a range of data acquisition and data processing possibilities, which will lead to varying degrees of uncertainty in the ship performance metric and therefore the use of the analysis. The following decision variables are open to ship owners regarding ship performance measurement:

   a. How should the ship performance be represented? Which is the most appropriate KPI?
   b. How should the raw measurements be translated in to the KPI? Either through filtering or normalising.
   c. How should the raw measurements be collected? In terms of frequency, automation and instrumentation type.
   d. What is the required evaluation period to achieve the required accuracy in the performance indicator?

The above variables make up the data acquisition (DAQ) strategy. Given these decision variables and their varying effects on the overall uncertainty and given the lack of defined standard procedure there is space for a rigorous study to help understand these questions. A quantifiable assessment is needed that is based on a constant and transparent metric that defines and compares DAQ strategies. This leads to the third research question:

3. **What can uncertainty analysis tell us about the relative strengths and weaknesses of different DAQ strategies for estimating/characterising ship performance?**

This research question, in combination with research question 2, provides the means to assess the relative influence of each data acquisition strategy variable on the objective function, the ship performance indicator. This includes not just the model uncertainty but also the sampling frequency, the degree of automation and the instrument/sensor characteristics.

3.1 Method and Thesis Layout

To answer these research questions, the industrial sponsor of this work has provided continuous monitoring (CM) data and access to metocean data (wind, waves) for the environmental conditions in which the ship has operated during its voyages over approximately one year. Noon report (NR) data has also been acquired over longer time periods (up to 5 years). This data is applied in a series of models (theoretical, black box and hybrid) in order to understand the relative merits of these models and their ability to accurately and insightfully identify ship performance and ship performance trends.

Before this modelling work can be undertaken, the raw data that is used in the models is investigated to understand more about its quality, see Chapter 4. In conjunction to a description of the data, this section also attempts to build a common language around describing the different sources of data and modelling uncertainty. Chapter 4 also discusses the data in terms of the advantages and disadvantages of the two major data acquisition strategies with an overall view to qualitatively exploring the sources of uncertainty associated with both. A set of pre-processing steps used to remove spurious outliers from the data sources are defined and justified.

The subject of uncertainty, pertinent to both Research Questions 2 and 3, is pervasive throughout the data description, model definition and application and discussion. A basic tool for assessing the relative merits of uncertainty is used in the chapters for each of the models (theoretical, black box, hybrid). The thinking around uncertainty only really crystallised towards the end of the research, once the different modelling attempts had exposed the key components and influences of uncertainty. This evolution of the thinking is also relayed through the write-up of the research, with the most detailed and systematic evaluation of ship performance uncertainty retained until the penultimate chapter.

There are many inter-weaving and interacting threads throughout the work: data acquisition, data manipulation, modelling and uncertainty analysis. This thesis attempts to cover all of these subjects, but does so by focusing on each in turn. In so doing, it sometimes compromises the representation of the interactions between these

subjects for the sake of clarity of discussion of the individual threads. However all threads are then brought back together in the discussion in the concluding chapter. Consequently, the remaining structure of the thesis is as follows:

Chapter 4 describes ship performance monitoring datasets in detail, including their strengths and weaknesses in order to develop an appropriate framework to structure the relevant uncertainties. Performance indicators and data pre-processing steps are also described. The aim of Chapter 5, Chapter 6 and Chapter 7 is to form and evaluate a theoretical, statistical and hybrid normalisation model, the normalisation models enable the actual operating/environmental condition to be reverted back into a reference condition from which the difference between modelled and measured fuel consumption equates to changes in the ship's performance. Chapter 8 discusses and draws conclusions regarding research question 1. Chapter 9 then presents a method to define how the uncertainty of ship performance measurement can be determined in answer to research question 2. This is developed into an uncertainty analysis method for answering research question 3. Chapter 10 concludes with the aim of drawing together the conclusions of each individual research question.

Chapter 4. Data and Uncertainty

This chapter firstly provides a brief overview of CM and NR datasets in order to begin to understand the strengths and weaknesses of onboard ship performance monitoring systems. This is to aid the exploration of sources of uncertainty and the development of a framework of these uncertainties which follows later in the chapter. For similar reasons, some of the performance metrics which have been found in the literature are introduced with a view to deriving the most appropriate for the aims of this thesis. A detailed description of the sources of uncertainty associated with deriving this ship performance metric follows, this deals with both the CM and NR data acquisition strategies and therefore a qualitative discussion of the merits and weaknesses of both ensues. This is structured in the framework of uncertainty in ship performance measurement which is used in this thesis to explore and quantify the various sources. Finally a set of pre-processing steps used to remove spurious outliers from the data sources are defined and justified.

4.1 Overview of CM and NR Data

Continuous monitoring and noon report datasets have been introduced in section 2.1.2; a sample of the fuel consumption – speed and shaft power – speed plots derived from each is shown for one ship in Figure 7 and Figure 8, respectively. The merits and drawbacks of these acquisition systems, in terms of sampling, instrumentation and human influences are discussed in detail throughout this thesis. In particular, the uncertainties associated with CM and NR datasets are described in detail in section 9.2. This section presents examples to serve as a brief introduction.

If the recording period of the noon report data presented in Figure 7 is reduced to a similar time period to that of the continuous monitoring dataset then the reduction in observations and the coarse nature of this dataset relative to that of the CM dataset is immediately apparent (Figure 9).

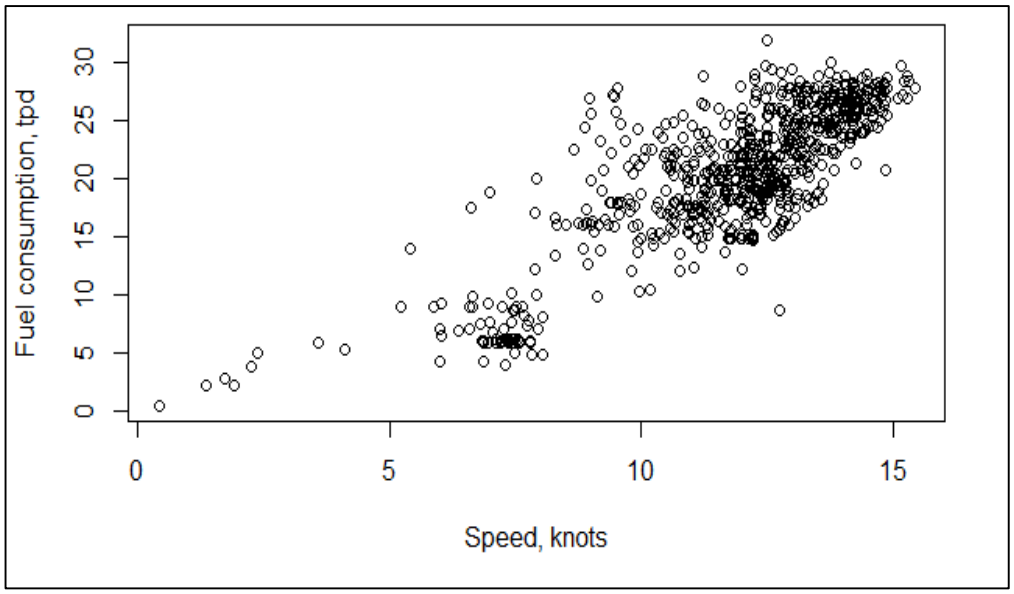


Figure 7: Noon report data over 7 years

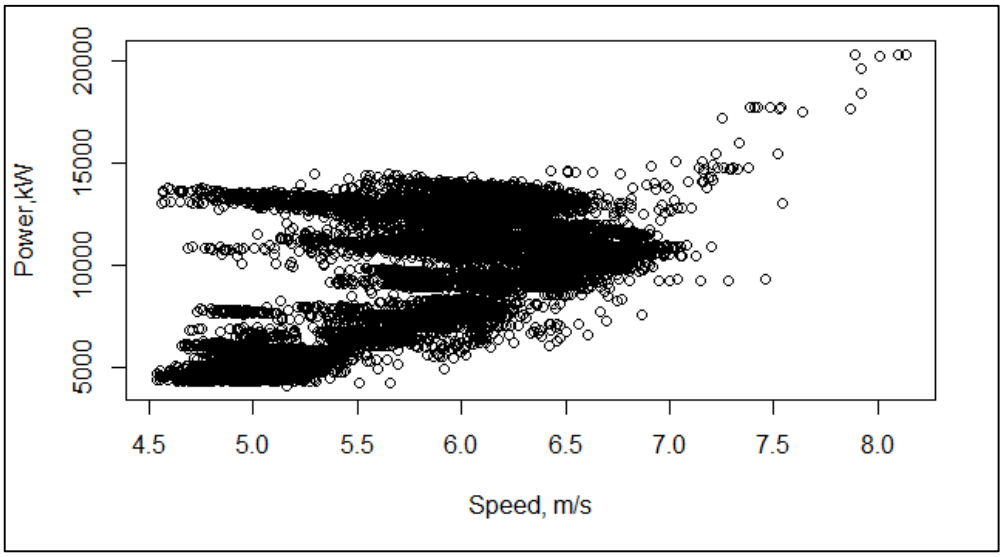


Figure 8: Continuous monitoring data for the shaft power vs ship speed of a bulk carrier over a period of 388 days, sampling rate = 1 sample / 15minutes.

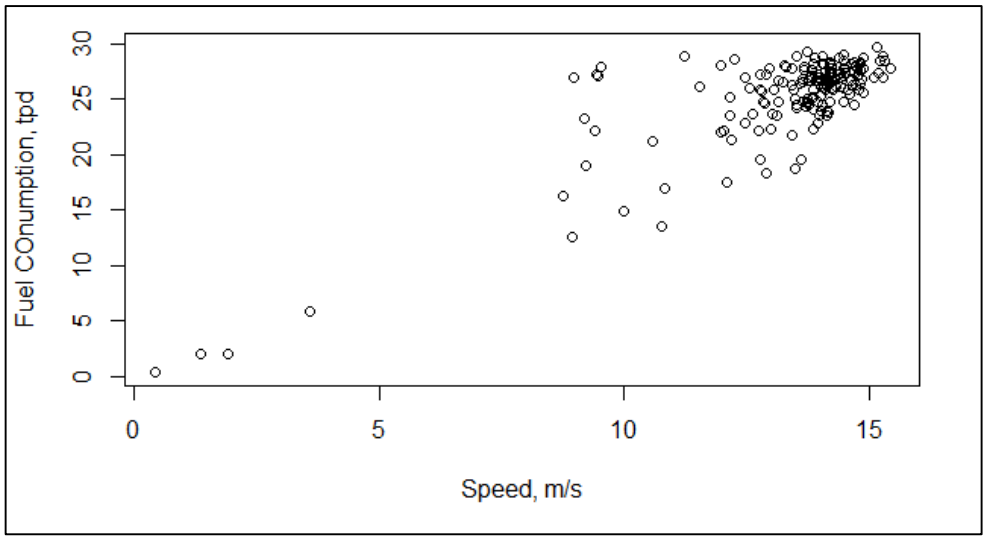


Figure 9: Noon report dataset for 374 days

The data fields that are included in a NR dataset vary between ship owners and ships, some examples are shown in Figure 10 and Figure 11. These, and a number of other NR datasets collected, have shown that there are differences between the types of data collected, variations in data field labels, a lack of units and no field descriptions such as whether wind and wave directions are true or relative and whether speed refers to speed over ground or speed through water, for example. Generally, there is no standardisation among noon report datasets, some have a field for all types of fuel onboard (HFO, LSFO, MDO, lubrication oil) and every power generator and engine (main, auxiliary, cargo, boilers) while others record only the fuel oil used in propulsion.

NR data is manually input and the consequences of this, such as missing observations and inaccuracies in reporting are discussed in section 9.2.4. The data field limitations of the noon report datasets also mean that it is only possible to understand ship performance in terms of fuel consumption rather than the more focussed measurement of shaft power which is reported in CM datasets.

Vessel	Vessel Number	Reading Date Time	Reading Type	Passage Type	Voyage Number	Heading	Wind
		24-Feb-12	Noon in Port	Laden			0
		25-Feb-12	Noon in Port	Laden		252	2
		26-Feb-12	Noon in Port	Laden		180	0
		27-Feb-12	Other	Laden			3
		28-Feb-12	FAOP	Laden			7
		29-Feb-12	Noon at	Laden			7

Wind Direction	Sea	Sea Direction	Swell	Swell Direction	Clock Hours	Observed Distance	HFO Consumption Main Eng 1
0	0	0	0	0	0	0	0
0	2	0	1	0	24	0	0
0	0	0	0	0	24	0	0
0	2	0	1	0	24	0	0
0	4	0	1	0	24	0	0
0	4	0	1	0	24	0	0

LSFO Consumption Main	LSFO Consumption Boiler 1	HFO Consumption Boiler 1	HFO Consumption Aux	LSFO Consumption Aux	Diesel Oil Consumption Aux	Gas Oil Consumption Aux	Obs Speed
0	2	0	0	0	0	0	0
0	0.5	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Main Engine 1	Main Engine 1	Main Engine 1 Torque	Slip	Comments	Main LNG FOE Consumption
0	0	0	0	alongside RH	16.98
0	0	0	0	alongside RH	33.2
0	0	0	0		53.18
0	0	0	0		54.25
0	0	0	0		54.02
0	0	0	0		46.09

Figure 10: Noon report data sample 1

Vslvoy	Vessel	Num Posrep	Port Name	Local Time	Time Zone	Latitude	Longitude	Course
		21		10-May-12 12:00:00	-3.0	X X'S	XX X'W	71
		21		11-May-12 12:00:00	-3.0	X X'S	XX X'W	17
		21		14-May-12 12:00:00	-3.0	X X'S	XX X'W	23
		21		14-May-12 13:33:00	-3.0	X X'S	XX X'W	23
		21		15-May-12 12:00:00	-3.0	X X'S	XX X'W	23
		21		16-May-12 12:00:00	-3.0	X X'S	XX X'W	7
		21		17-May-12 12:00:00	-3.0	X X'S	XX X'W	318
		21		18-May-12 12:00:00	-3.0	X X'S	XX X'W	309

ETA	Time Since Last Report (HRS)	Miles To Go	Miles since last report	Speed	RPM	HFO Main Engine	HFO Auxiliary Engine	Mean Wind Type
11-May-12 15:00:00	:	378	139	13.90	109.200	11.600	.90000	Fresh breeze (17 - 21 knots)
11-May-12 16:45:00	24:00	52	328	13.67	109.800	28.500	2.00000	Moderate breeze (11 -16 knots)
27-May-12 02:00:00	72:00	4236	165	13.77	109.400	13.300	1.00000	Moderate breeze (11 -16 knots)
27-May-12 02:00:00	1:33	4236	165	12.19	109.400	13.300	1.00000	Moderate breeze (11 -16 knots)
27-May-12 02:00:00	22:27	3920	316	13.17	110.200	27.400	2.00000	Moderate breeze (11 -16 knots)
27-May-12 04:00:00	24:00	3598	322	13.42	110.000	27.300	2.00000	Gentle breeze (7 - 10 knots)
27-May-12 04:00:00	24:00	3258	340	14.17	110.100	26.500	2.20000	Gentle breeze (7 - 10 knots)
27-May-12 04:00:00	24:00	2920	338	14.08	110.200	27.000	2.10000	Moderate breeze (11 -16 knots)

Mean Wind Direction	Sea State	Swell Direction	Swell Type	Speed Instruction	Speed Instruction Comments	DWT Remaining on Departure	Draft FWD on Departure	Draft AFT on Departure
1	Moderate 1.25 to 2.5 m	2	Light (short and moderate wave)	Full Speed	Slip = 8.6%	7827	10.150	10.750
8	Moderate 1.25 to 2.5 m	4	Light (short and moderate wave)	Full Speed	Slip = 11.0%	7827	10.150	10.750
1	Moderate 1.25 to 2.5 m	3	Light (short and moderate wave)	Full Speed	Slip: 8.0%	23797	7.150	8.800

Figure 11: Noon report data sample 2

There is also a stark difference between the collected NR and CM environmental variables. For example, in CM datasets the wind speed measurement may be collected from onboard anemometers, or from global metocean data. An example of the wind speed frequencies collected from an onboard wind anemometer is shown in Figure 12. This might be used in combination with the ship's heading to define relative wind speed.

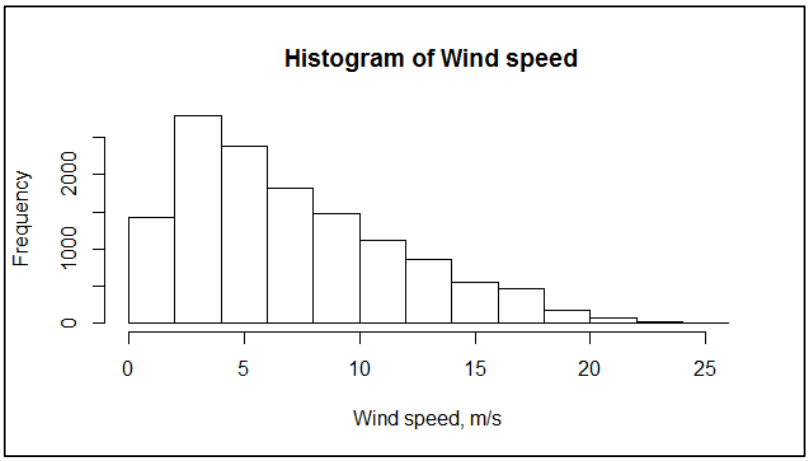


Figure 12: Frequency of wind speed measurements

CM systems offer high frequency sample rates and remove elements of human intervention, however maintenance is required to avoid sensor drift and stuck sensors. Continuous monitoring systems also have a wide array of data fields, an example dataset is shown in Figure 13. This shows some possible additional fields, generally many more metocean data fields may be available, plus engine temperature, fuel flow, shaft power, propeller RPM and propeller pitch, for example.

Year	Month	Day	Hour	Minutes	Second	Longitude	Latitude
2011	11	2	9	55	0	143.224	-11.8701
2011	11	2	10	5	0	143.249	-11.8975
2011	11	2	10	10	0	143.262	-11.9114
2011	11	2	10	15	0	143.274	-11.9253
Average_Beaufort	wind_direction	wave_height	wave_direction	water_depth	salinity	temperature	current_speed
5.16784	0	1.94114	0	42.2089	34.6998	24.75	0
5.16784	0	1.94114	0	47.0177	34.6998	24.75	0
8.4786	0	1.94114	0	47.8759	34.6998	26.65	0
8.4786	0	1.94114	0	46.4355	34.6998	26.65	0
STW	Speed	COG	heading	Draft_aft	Draft_fwd	draught	trim
6.74521	6.79076	137.573	139.08	12.22	12.21	12.215	0.01
6.68546	6.80029	137.322	139.029	12.22	12.21	12.215	0.01
6.71456	6.8231	137.469	138.957	12.22	12.21	12.215	0.01
6.72728	6.84152	137.214	138.85	12.22	12.21	12.215	0.01
current_direction	Shaft_power	RPM	foc	tide direction	tide_speed	tide_level	prop pitch
0	9.71683	97.8155	0.000542043	173.23409	0.088003606	-0.33179718	0.18608393
0	9.68757	97.5806	0.000540753	173.34064	0.090422042	-0.32759726	0.19240947
0	9.69522	98.3257	0.00054082	173.46347	0.092683278	-0.32282311	0.20164914
0	9.68965	98.2097	0.000540321	173.60126	0.094753072	-0.31742325	0.20357499

Figure 13: Continuous monitoring dataset example

The wider array of data fields collected can introduce possibilities of data quality control checks. For example, the onboard wind speed and wind direction sensor measurement can be compared to the wind speed and direction from the global metocean data. Using vector analysis and the ship's speed and heading the absolute global measurement is transformed to a relative measurement and the results shown in Figure 14. In this example the results generally overlay indicating that the calculated relative wind speed and direction matches the measured and therefore confirming the data quality of these sensors. This would be fully clarified by a examining a shorter scale time period at higher resolution.

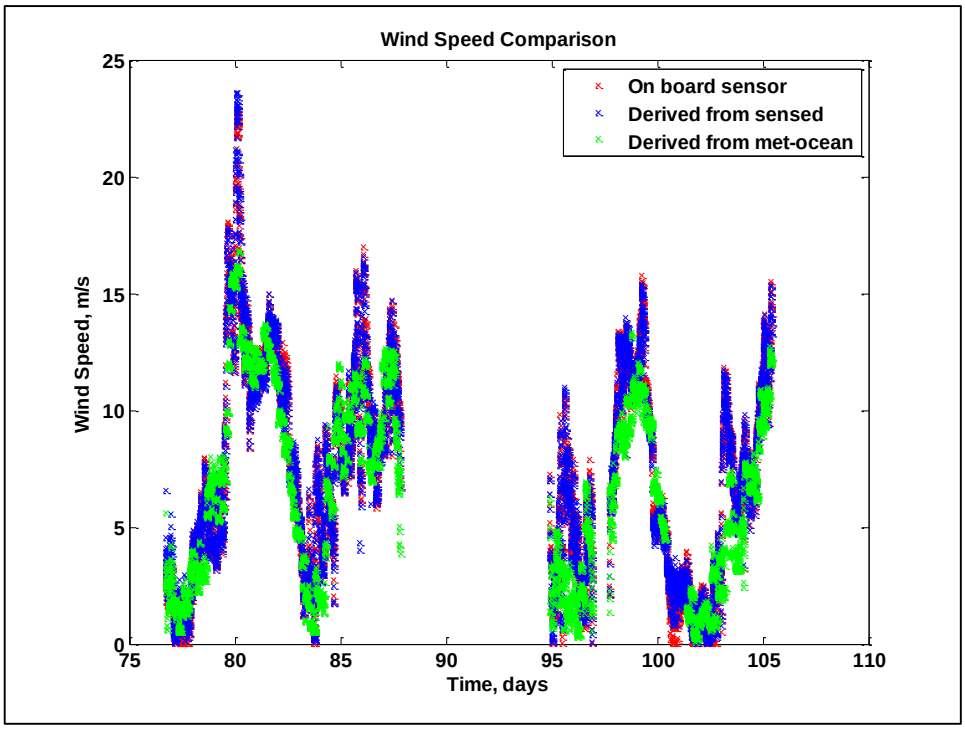


Figure 14: Wind speed comparison between sensed and derived

Conversely the noon report wind speed is only recorded once a day, it is measured visually by a crew member and according to a relatively low resolution scale; the Beaufort wind scale (1 to 12). Similarly the wind direction is measured by the crew and only to the nearest octant.

Another similar quality control indicator may also be derived from CM data by comparing speed over ground (SOG) as measured by GPS and a calculated (modelled) SOG. SOG can be calculated from the measured true current (converted to relative current by vector analysis with ship heading) and reported speed through water (STW). One such comparison is shown in Figure 15, where there is good agreement. The GPS derived SOG measurement is the measurement that is least susceptible to inaccuracies, and therefore this cross-check can highlight possible sources of error in the STW sensors or in the current measurement. As can be seen in the figure, there are also gaps in the SOG calculated data and this may be due to a malfunction in the STW sensor, in that case it may be possible for STW to be back calculated from SOG providing there is a reasonable agreement between the SOG modelled and the SOG observed.

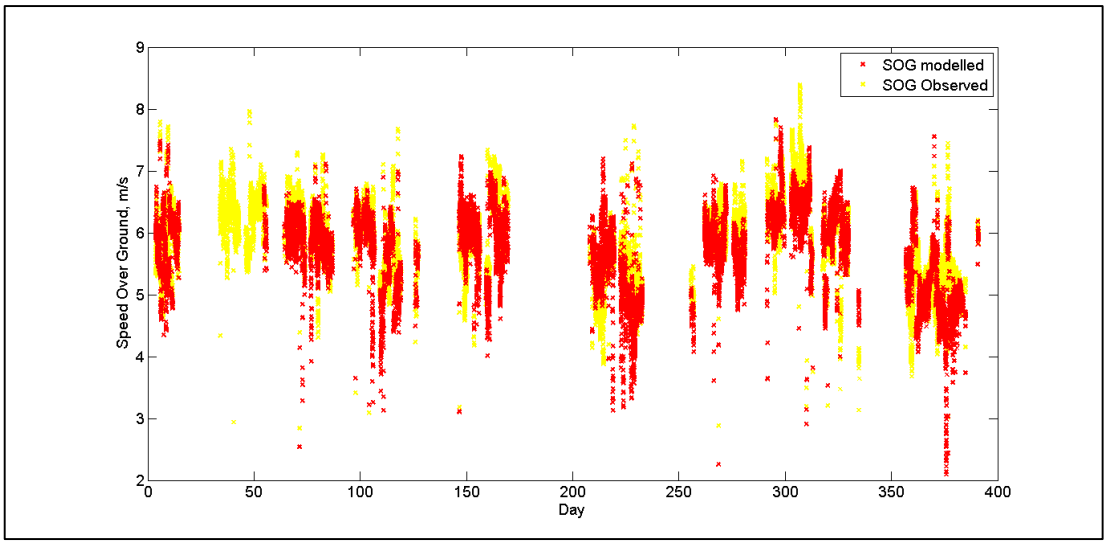


Figure 15: Example of the comparison between SOG measured and modelled

All of the environmental and operational parameters, the differences between the NR and CM collection methods and instrumentation and the impact of these differences are explored in detail throughout this thesis.

In this thesis, the Clarksons World Fleet Register (CWFR) is used to characterise ship's static, technical specifications such as hull geometries and installed engine maximum continuous rating and type.

4.2 Performance Indicators

There are a number of possible metrics presented in the literature by which ship performance can be measured. This section explores these in order to choose the appropriate metric used in the analysis presented in this thesis. The nomenclature used in this section is defined at the beginning of the thesis.

Bazari (2007) asserts that key performance indicators (KPIs) for performance benchmarking/rating should have the following characteristics:

- Be indicative of ship performance
- Show appropriate and consistent variations with ship size
- Require minimal measured data
- Be unambiguous and easy to understand

This is expanded on by Deligiannis (2014) to include also

- Dimensionless number
- Unique for an individual vessel
- Be easy and accurate to determine from common speed trial data
- Inclusive of hull resistance effects (wave, wind, swell, current)
- Statistically constant
- Capable of providing diagnosis on efficiency of main engine and power transmission system
- Specific for crew to grasp
- Measureable, achievable, realistic and timely

These characteristics are meaningful in providing background information although neither author justifies the criteria or explains if they are exhaustive or rigid. For example, it is useful to think about expanding beyond speed trial data given the advent of more complete continuous monitoring datasets. The requirement of a dimensionless indicator and one that shows appropriate and consistent variations with ship size may not be necessary if the data is only used to compare historical trends within one vessel rather than between vessels.

Expanding on the defining ratio of performance that represents unit of input (fuel or power) per useful output (transport work), Bazari (2007) defines the KPI's as presented in Table 3. Possible other performance indicators (PIs) are summarised in Table 4, these are plotted against time to track the trend in ship performance.

KPI	Full Name	Equation	Application	Assumptions Introducing Uncertainty
FCI	Fuel Consumption Index	$\frac{FC}{Cap.d}$	Hull, propeller & engine	None – Fuel consumption measured directly
FCI	Fuel Consumption Index	$\frac{P.BSFC}{Cap.V}$	Hull, propeller & engine	BSFC
SEI	Ship Energy Intensity	$FCI.LHV$	Hull, propeller & engine	LHV
PEI	Propulsion Energy Intensity	$\frac{P}{Cap.V}$	Propeller & engine	None – power measured directly
PEI	Propulsion Energy Intensity	$\frac{PEI}{\eta_S \eta_{eng}}$	Propeller & engine	Engine and shaft efficiencies
CO_2I	CO_2 Intensity	$\frac{CO_2}{Cap.d}$	Environmental	None – CO_2 measured directly
CO_2I	CO_2 Intensity	$3.67.C_f.FCI$	Environmental	Carbon factor

Table 3: KPI's as defined by Bazari (2007)

Performance Indicator	Name	Reference	Application	Filtering / Assumptions / Comments
$100.\frac{R_{new} - R_{baseline}}{R_{baseline}}$	Per cent added resistance	(Munk and Kane 2011)	Hull & propeller	Filtered for ship speed and draught. May require hull geometry assumptions
$PD_{no} = \frac{FC\,.NCV\,.(1-S)}{RPM^2\,.V}$	Propulsion Diagnosis Number	Deligiannis (2014)	Engine, hull & propeller	Assumptions used for NCV
$EEOI = \frac{\sum_j FC_j C_{Fj}}{Cap.d}$	Energy Efficiency Operational Indicator	MEPC1/Circ.684	Environmental	Filtered. Assumption for C_F
$\frac{V}{FC}$	Fuel efficiency	(Petersen, Jacobsen et al. 2012)	Engine, hull & propeller	Filtered for draught and environmental variables
$\frac{\sum(\bar{P}.t_{tot})^n}{d_{tot}} - \left(\frac{P}{V}\right)^n_{ref}$	Increase in standard power	(Townsin R. L., Byrne D. et al. 1981)	Hull & propeller	Filtered for draught and environmental variables
$\frac{V_{measured}}{V_{new}}$	Change in speed	(Townsin, Moss et al. 1974-1975)	Engine, hull & propeller	Filtered for environmental conditions and draught
$P_s - \frac{RPM^3}{f(H)}$	ΔPower	(Logan 2011)	Hull & propeller	The linear pitch function represents the torque-apparent slip relationship under baseline conditions.

$\frac{\Delta\omega_q}{\Delta C_s}$	Change in torque wake fraction vs service roughness allowance	(N. Hamlin and Sedat 1980)	Hull & propeller	Open water propeller tests (and measured torque) required to determine ω_q. Model tests required to derive change in resistance and determine C_S.
$\Delta\omega$	Change in wake fraction	(Reid R. E. 1980)	Hull & propeller	Ship trial reference speed and reference wake fraction required. Open water propeller characteristics required
$\frac{\nabla^{2/3}V^3}{P_s}$	Performance metric	(Aertssen 1966)	Hull & propeller	Displacement may require block coefficient assumptions for different draughts

Table 4: Summary of performance indicators

The difference between performance indictors is in which components of the ship system they represent and the assumptions or filtering required in their formulation; therefore the appropriate PI depends on the application and available data. This is further explored below.

For environmentally focussed applications the CO_2 intensity (CO_2I) is relevant and this is equivalent to the relatively straight forward IMO recommended EEOI calculation. This represents the performance of the engine, hull and propeller combined and their consequent impact on carbon dioxide emissions. The exact data input must be clearly defined as the output (CO_2 per tonne km) and can vary by 18-20% for the same ship depending on operational conditions such as ship speed, days spent in port, unloading /loading, manoeuvring and at anchor (Acomi and Acomi (2014). Since the recommendations dictate that ballast voyages and voyages not used for transport of cargo, such as docking voyages should be included, (MEPC.1_Circ.684 2009), then the EEOI value will depend on the average utilisation of the cargo-carrying capacity that can be achieved in actual operation, therefore it depends on the cyclical 'business climate' for the various trades. The average EEOI, even within a ship category, may vary from one year to the next given changes to demand and competition and among trade routes. Some trade patterns such as return cargo or trade triangle enable higher average utilisation and therefore higher efficiency relative to other trade patterns such as the distribution of smaller cargo parcels that have an inherent low efficiency that is not related to the operation or choice of ship (MEPC55/4/4_Annex 2006). These issues are particularly relevant in benchmarking one ship against another but also will affect a performance trend based on this metric for one ship if its trading patterns are not constant or cyclic.

The FCI by Bazari, or other PIs based on fuel consumption, are broadly equivalent to the EEOI and other carbon indices but preferable in economically conscious applications when fuel consumption is more appropriate as the dependent variable. If the fuel consumption is unknown then it can be approximated from measured shaft power and an assumption of brake specific fuel consumption. Engine degradation and any fluctuations in engine performance are no longer included in the metric. If fuel consumption is the dependent variable (engine, hull *and* propeller performance combined is of interest) then additional uncertainty is consequently introduced as a

consequence of the BSFC assumption. This may be small as the SFOC of a well maintained 2-stroke main engine will normally not change much during its service life (Munk 2006). Additional consequences of this assumption are explored further in section 4.3.1. If the hull and propeller performance is to be isolated from the engine performance then shaft power is the desirable dependent variable, and if it is approximated from fuel consumption then the fluctuations of engine performance add an additional uncertainty.

Generally, indices based on shaft power such as the PEI or that proposed by Townsin R. L., Byrne D. et al. (1981) are relevant in isolating the measurement of hull and propeller performance. It is however proposed by Munk (2006), that added resistance at a specific ship speed and loading condition as a measure of hull degradation is advantageous over added shaft power since the power and added resistance are not linearly proportional. The major components of ship resistance are frictional and wave making effects, other effects include air resistance (from the wind) and pressure resistance. Fouling only affects frictional resistance and the relative proportion of the total depends on speed and draft, so added power also depends on speed and draught. Further, the power measurement includes not just the hull resistance effect but also the boundary layer effect on the propeller efficiency and the boundary layer effect on power through the hull efficiency, so the relationship between power and resistance is complex. In a similar way to the added resistance performance indicator, tracking the wake fraction and coefficient of in-service roughness allowance by N. A. Hamlin and Sedat (1980) is implemented to identify only increases in frictional resistance because of hull fouling. These also have a significant reliance on propeller characteristic curves and several assumptions;

- Ratio of torque wake fraction to thrust wake fraction is the same on the ship as predicted by model tests
- The thrust deduction factor is not significantly altered by hull roughness
- Open water propeller tests are used without adjustment for propeller blade roughness increase in service

These assumptions may enable the PI to reduce uncertainty by focussing on the critical variables in the measurement of the hull condition or, because of the

assumptions (zero propeller blade roughness increase) and the experimental errors in the model tests and characteristic curves, then the overall performance indicator uncertainty may increase. The unavailability of sea trial or model test data may also prove to be prohibitive in the employment of these models.

Other relevant points in the choice of performance indicator include the restrictions in the onboard data collected, in the PD_{no} (see Table 4) calculation for example, the NCV (net calorific value) is rarely measured onboard which may introduce uncertainty surrounding the fuel grade. PD_{no} is a dimensionless number, based on hydrodynamic principles, representing the ratio of effective towing power (which incorporates chemical energy consumed) to the thrust power (the produced propulsion effect). In this form it does not account for the fact that the speed of advance of the propeller is less than the ship's speed due to the wake fraction. Consideration of the wake fraction is important because in principle, it results only from increases in frictional resistance because of roughness. The actual speed of advance into the propeller is extremely difficult to measure without a thrust meter and these are not installed on many ships. Cavitation effects are also not measured in this metric, although depending on the circumstances, these may be considered to be negligible.

Performance Indicator Summary

The above has highlighted that the existing literature is rich in theoretical metrics which may be categorised into one of three according to the application; environmental, whole ship performance (engine, propeller and hull) or hull and propeller performance only. Assumptions may be used (C_f, LCV, BSFC) to move between applications without additional collection of data (direct CO_2 measurements for example) but these assumptions will introduce additional uncertainties. Within each category the methods are broadly comparable in that they rely on the same basic processing (normalising or filtering) in order to strip out the variables that form the metric inputs. Some metrics require additional theory or experiment such as characteristic propeller curves, model tests or assumptions for hull geometries. Generally, the literature is poor in rigorous derivation and justification of these metrics. For this thesis a PI is chosen which is able to create a clear distinction when

theoretical assumptions are used and to provide a rigorous analysis of normalisation versus filtering by ensuring that they are compared from an equal position. The following PI is used:

$$P_{s,measured} - P_{s,modelled}$$

Equation 2

Where P_s refers to shaft power and $P_{modelled}$ would be calculated from a model derived either from theory or during a calibration period which would form the 'training' dataset for a statistical or hybrid model. In this way it is clear where theory is introduced and the metric is constant between different models and between results derived from filtered datasets and those from unfiltered, normalised data. In the absence of deterioration in the performance of the ship's hull or propeller, $P_{modelled}$ should be equal to $P_{measured}$ during the evaluation period. The actual difference therefore quantifies the trend in the ship's performance. This PI also has the advantage of being easy to interpret. If power is not measured, and fuel consumption is used as a proxy (as in the case of noon reported data), then fluctuations in engine performance are an additional uncertainty in the hull and propeller performance measurement.

4.3 Uncertainty Framework

Figure 16 presents the key sources of uncertainty in ship performance monitoring.

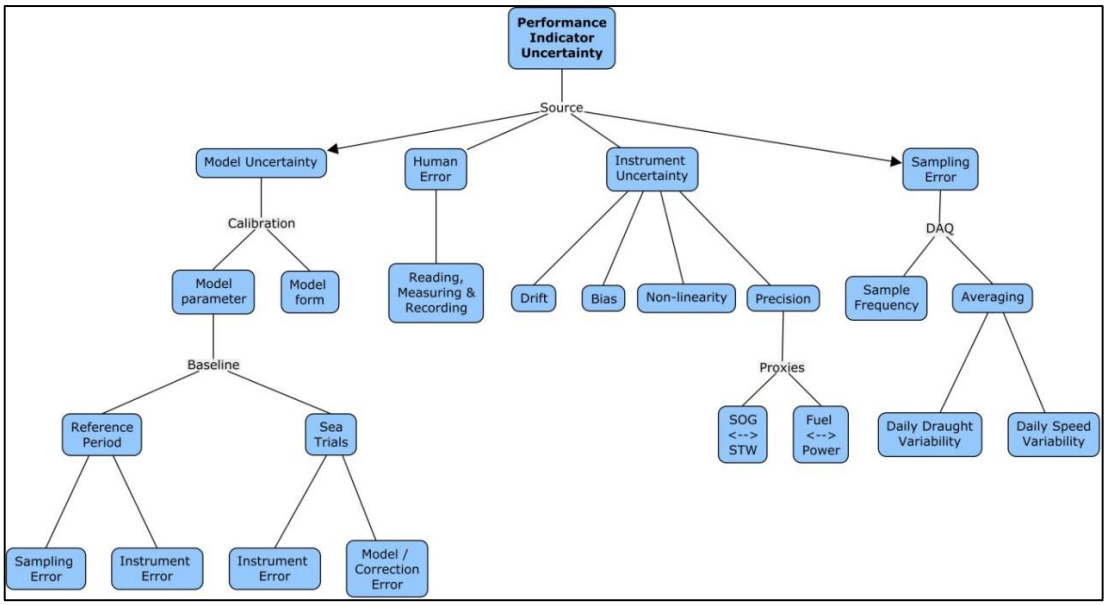


Figure 16: Source of uncertainty in ship performance monitoring

A qualitative discussion of the sources of uncertainty specific to each dataset (noon report and continuous monitoring) is below; the parameters of their probability distributions are quantified in section 9.1.

Which is defined as aleatory and which is epistemic? This is not necessarily black or white and it is not always possible to determine categorically. Bias limits are assigned to all error sources that are not attributed to measurement repeatability; calibration, data acquisition, data reduction equations (correction factors). 'Error analysis' allows both bias and precision uncertainties to be treated essentially by assuming the uncertainty to be a Gaussian random process. However, in the absence of data to define a 95 per cent confidence band for bias limits then instrument bias or drift is categorised in this thesis as epistemic, and given the discussion relating to probabilistic methods for epistemic uncertainties at the end of the previous section (2.3.2) then their effect is propagated through interval analysis rather than assuming a random, Gaussian process.

4.3.1 Instrument Uncertainty

The definitions used in this thesis are from ISO standard 5725-1-1994 "Accuracy (trueness and precision) of measurement methods and results – Part 1: General

principles and definitions" **Trueness** refers to the closeness of agreement between the arithmetic mean of a large number of test results and true or accepted reference value. **Precision** refers to the closeness of agreement between test results, difference occur because of unavoidable random errors inherent in every measurement procedure. This is diagrammatically represented in Figure 17.

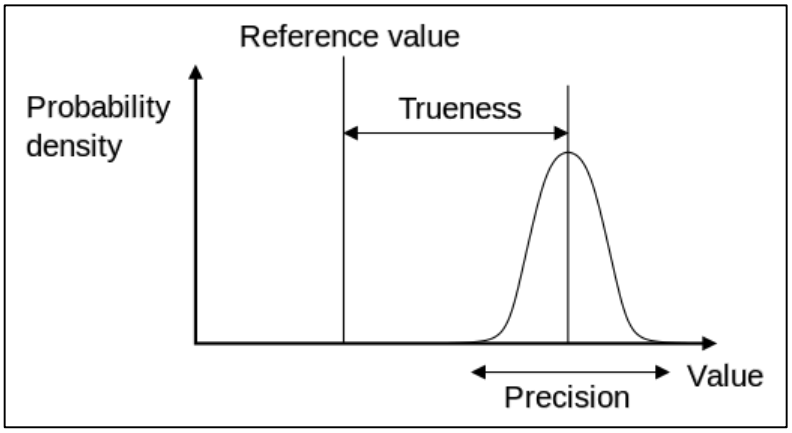


Figure 17: The definition of trueness and precision according to ISO 5725-1-1994

Repeatability and **reproducibility** are two extremes of precision, the first describes the minimum variability in results because the same operator, equipment, calibration, environment and time elapse between measurements are used in the experiment, the second is the maximum variability in results because one or more of these factors are allowed to vary. The general term **accuracy** is used to refer to both trueness and precision. It is however also recognised that a definition of accuracy in common use also refers to the proximity of the measurement to the true value (trueness). This is as defined in (JCGM_200:2008 2008). This is kept in mind in the interpretation of the literature but the terms trueness and precision are used in this thesis. It is also recognised that there may be a degree of quantisation error in the digital conversion of any analogue sensor signal. This is assumed to be negligible relative to the other errors and is not explored in any detail in this thesis.

In continuous monitoring datasets where the acquisition is generally automated there are potential issues of sensor noise or stuck sensors leading to missing data. Each sensor has an associated accuracy in its measurements, as with all sensors, errors also increase if they are poorly maintained or calibrated. Inaccurate sensors will lead to bias and/or imprecision in the measurand.

It was concluded by the Uncertainty Modelling for Ships and Offshore Structures (UMSOS) Workshop in Rostock, 8 September 2012 that the uncertainties associated with metocean data are the most significant and challenging factor in the shipping, offshore and renewable energy industry (Bitner-Gregersen, Bhattacharya et al. 2014).

In continuous monitoring datasets, environmental characteristics may be monitored with onboard sensors (such as anemometers for wind speed and acoustic surface tracking sensors or pressure based sensors for wave data) or by combining ship position data with models that solve environmental characteristics over a wide area. The latter are derived from measured data complemented by numerical models that produce the 10 min average wind speed (and direction) at 10m above sea level and characteristic wave data (wave height, zero crossing point). The measured data in this latter case originates from either in situ sensors such as offshore platforms, wave buoys, onboard wave radars or lasers and/or remote sensors such as from satellites or aircraft. The measured data is then used to validate the numerical models. For a long time, visual wave observations were compiled in a database of ocean wave statistics (Hogben and Lumb 1967) and global wave statistics (Hogben, Dacunha et al. 1986). Studies to compare the observed to actual wave heights and wave periods have been undertaken and have shown the possible presence of systematic bias (of the order of 0.2s for the wave period and 0.2m for wave heights). These studies are summarised in a paper by Gulev, Grigorieva et al. 1999. A detailed description of the types of data sources and associated uncertainties of metocean data can be found (Bitner-Gregersen, Bhattacharya et al. 2014), further details and guidance regarding modelling methods are found here (DNV 2010). The sea state statistics available are usually characterised as follows (Bitner-Gregersen and Hagen 1990):

- Significant wave height, H_S; representing the average over the 1/3 highest waves in the wave record
- Average zero-up (or zero-down) crossing wave period, T_z; representing the average of all wave periods in the wave record
- Main wave direction; the direction from which the waves are coming

In the NR dataset, fields are limited and proxies are often used for example, fuel consumption is generally recorded rather than shaft power. This measurement must then be converted into shaft power using the following measurements/ assumptions:

1. SFOC approximation from test-bed data: to convert the fuel mass to power
2. Fuel characteristics: Lower calorific value, density and temperature, to convert the fuel volume to a mass (if volumetric flow meter is used)
3. Fuel flow meter measuring fuel mass or volume

There are two major consequences of this additional step. Firstly, the performance measurement is now not only reflecting the hull and propeller performance but also the engine performance. If the combined hull and propeller performance is of primary interest then in order to isolate these uncertainties are introduced due to unknown changes to the engine performance because of age deterioration (including maintenance effects) and possible random engine performance fluctuations. Secondly, the additional instruments and the assumptions associated with conversion from fuel consumption to shaft power also introduce error.

The SFOC approximation from test-bed data is often optimistic compared to the SFOC achieved in operations. There is also a change in SFOC with loading due to the shifting propeller curves and this rpm bandwidth should be included in the SFOC power conversion using the SFOC. This is also affected by propeller ageing.

The quality of the flow meters including their calibration and maintenance effects may be significant. The fuel consumption may be measured from tank soundings or automated fuel flow meters or a combination of the two, possibly combined also with information from bunker delivery notes (BDNs). Fuel flow meter accuracies depend on the type, the manufacturer, the flow characteristics and the installation. Approximately 70% of HFO fuelled ships have viscosity sensors and 30% have flow meters. Mass flow meters are founded on the principles of the Coriolis Effect and represent the most expensive and recent technical advance. Alternatively, positive displacement sensors measure the volumetric flow which are cheaper however also require the fuel density for the conversion to mass and a temperature correction, both of which introduce additional uncertainties. The accuracy of fuel consumption by

tank soundings is affected by the heel and trim of the vessel which must be accurately corrected for. The resolution and accuracy of tank gauging tables should consider the tank shape and internal structures, the prevailing weather conditions at the time of data recording may also have an effect.

The fuel consumption measurement accuracy is also susceptible to the accuracy of measurement of the fuel characteristics required for the volume to mass conversion, i.e. the fuel density measurement, often obtained from the BDN, or from the fuel temperature measurement using industry standard conversions. If the density is assumed from fuel grade and calorific value then the random fluctuations in these measurements also increase uncertainty (although this is limited by the fuel standard grading system). The quantity of fuel delivered however may not accurately reflect that consumed as there may be a non-combustible fraction comprised of water or other contaminants such as cat fines. Often fuel characteristics are not recorded and therefore if assumptions are used from ISO standards for example then additional uncertainties are introduced due to natural variations in the fuel (water content / other contaminants). Waste 'sludge' from onboard processing may also be neglected to be accounted for and inaccuracies in ullage measurements due to weather, fuel temperature etc. may arise. The position of the fuel flow meter should also be such that auxiliary engine consumption is not mistakenly included in addition to that of propulsion.

In NR data, speed over ground (SOG) is also used as a proxy for speed through the water (STW) and uncertainty in the speed will therefore increase according to the uncertainty in the current measurement. NR's frequently measure wind speed using the Beaufort scale measure which introduces uncertainty through the resolution as a continuous variable is reduced to a non-linear 1-12 scale causing rounding errors. The same applies to wave height which is rounded according to the Douglas sea-state scale and wind direction which is often estimated to the nearest 45 degree octant. If also the overall weather effect is only represented by a Beaufort scale measurement this considers wind speed only and so only the wind driven waves are represented; swell is not accounted for (Pedersen and Larsen (2009). Possible errors are also associated with the height of the sensor above the deck. Finally, in NR datasets the draught marks at the perpendiculars may be read by eye and noon report entries are

often not altered during the voyage to record variability due to trim and/or the reduction in fuel consumption leading to additional uncertainty. Some NR 'loading' fields are binary in nature (loaded / ballast) and this introduces random rounding errors if the ship is actually partially loaded.

4.3.2 Sampling Uncertainty

The effect of sampling error is two-fold, firstly it is due to taking a finite sample from an infinite temporal population and therefore the exact true value of the overall population is unknown, secondly because variable factors are averaged between samples and thirdly because of the evaluation period length.

In a NR dataset a sample is recorded at noon approximately every 24 hours, depending on time zone changes. The effect of averaging over 24 hours or of recording an instantaneous measurement for a continuously varying parameter, such as wind speed, is possibly significant. The same applies to ship speed which may be calculated from distance and time and therefore time spent at anchor translates a more efficient ship than the reality. Also, averaging a range of speeds will cause power due to higher speeds to be mistakenly attributed to deterioration in ship performance that in reality is not present. The actual influence on uncertainty is of course a function of the daily speed variability and this may be due to ship accelerations, rudder angle alterations, wind speed fluctuations within $0 < BF < 4$ range, and crew behaviour patterns (slow steaming at night for example), which aren't included in the ship performance model or can't be filtered out in the NR data set. Filtering may therefore be advantageous because for example the impact of wind speed on ship power is non-linear and a change in wind speed is less effective at low speeds than at high wind speeds. This however depends on the quality of the wind speed model in capturing the actual effects. The effect of averaging fluctuating variables is less significant in CM datasets whereby the readings are often averaged over 5 minute intervals. In a similar way there is an effect of averaging daily draught variability (i.e. due to intentional changes in trim or due to influences on trim due to fuel consumption).

The actual finite sample size is, realistically, less than the sample averaging frequency because of periods when the ship is at anchor, manoeuvring, waiting or in port and because of the stringent filtering criteria. The proportion of the results that are removed due to filtering depends on the environmental / operational conditions.

The number of observations is also affected by the evaluation period length. There is seen in some datasets the effect of seasonality, whereby weather fluctuations (even within the BF filtering criteria) cause discontinuities in the measured power and therefore a minimum of one year's data collection is required. This effect and the minimum period of data collection is dependent on the cyclic nature of the ship's operating profile and the global location of the ships operations.

4.3.3 Model Uncertainty

Multiple regression theory tells us about the errors resulting from estimating the regression coefficients rather than knowing their true values. These errors are usually much smaller than errors resulting from misspecification due to omitted variables (Chatfield 1995) or from defining incorrect form of the relationships between variables; model form / structure uncertainty. Omitted variables may include those that define ship motions that are not necessarily covered in the regression such as rudder angle, or dynamic effects of swell induced roll or yaw that are not completely defined by head and cross wind which only represent wind driven surface waves and do not account for swell.

In ship performance modelling the model parameter and model form uncertainty depends on whether a filtered or normalised (corrected by modelling of the interactions between variables) dataset is used to calculate the performance indicator. In both cases the model parameters and form may be based on a theoretical, statistical or hybrid model and the data may come from sea trials, model tests, a calibration period during the ships operation or CFD analysis, or a combination of more than one. In the case of filtering, the model will have fewer variables as the majority of effecting conditions are filtered out. Speed is necessary because the ships speed is like to reduce (unless the engine delivers more power further loading the propeller) due to the fouling of the hull and propeller that increases hull resistance

(MAN 2011). Therefore, model uncertainty in the filtered method is likely to be less. The same method of deriving the ship performance model can be applied to both CM and NR data acquisition strategies and the uncertainties associated with any CFD analysis, sea trial data and model tests will be the same for both. However, statistical or hybrid models derived from a calibration period will have different uncertainties depending on the training dataset. In the prediction of full scale propulsion power using artificial neural networks for example the network was able to predict the propulsion power with accuracy between 0.8% and 1.7% using onboard measurement system data and 7% from manually acquired noon reports (Pedersen and Larsen 2009). The uncertainties associated with different data sources and models are described below.

The uncertainties of sea trial data was explored by Insel (2008) who measured the uncertainties due to model uncertainty (due to the corrections applied to trial measurements) and instrument uncertainty for the measurement of torque, shaft rpm, speed, draught and environmental measurements etc. The author identifies the Beaufort scale estimation error as the key measurement error affecting the overall sea trial data uncertainty. However, because reliable methods do not currently exist, the influence of currents, steering and drift are not corrected for which could lead to additional uncertainties that are not represented.

Uncertainty analysis of CFD may be quantified according to ITTC recommended procedures and guidelines (ITTC 2002) this describes sources of uncertainty coming from the simulation error. This is the difference between the simulation and the truth and is composed of additive modelling and numerical errors. Numerical error is decomposed into contributions from iteration number, grid size, time step and other parameters. This can be estimated according to the detailed procedure laid out by Eça (2014). Modelling errors can be decomposed into modelling assumptions and use of previous data. Klaij C. M. (2014) studies 6 scenarios and found numerical uncertainty to be between 1% and 9% according to the application, the minimal being for highly-refined grids and good iterative convergence.

The use of model tests in full scale speed-power predictions have uncertainties relating to the resistance tests and to the extrapolation from the model to full scale.

The uncertainty in the extrapolation is (usually) only a modest contributor to the total uncertainty (Verhulst 2007). Bias error sources in the resistance tests may arise from scale effects, model inaccuracies (from rough model surface, production errors in the shape of model deformations), unsteady carriage speed, errors in test set-up, calibration errors, errors due to environmental modelling such as wave parameters and tank wall effects such as those due to reflected and radiated waves (MARIN 2012). Uncertainty in the extrapolation process arises from the model test results and the assumptions in the method (e.g. the use of Foude's law or the choice of friction line, form factor, wake scaling and correlation allowance) meaning that there is no accurate way to separate the frictional resistance from the total resistance measured during a ship model test (Bose 2009). A method for evaluating these uncertainties can be found in documents of the International Towing Tank Committee covering uncertainty analysis in resistance towing tank tests (ITTC 2008), propulsion performance tests (ITTC 2002) and propulsor open water tests (ITTC 2002). A sensitivity analysis indicated uncertainty in the open water test and, usually in the propeller open water test contributed a greater proportion of the overall uncertainty (Bose 2009). In the extrapolation procedure the same author reported the friction line to have the largest influence on extrapolation uncertainty. However, this study did not investigate the likely magnitude and then the combined effect of each elemental uncertainty on the overall model uncertainty. A master's thesis by Øyan (2012) indicates uncertainty in the brake power from model tests compared to full scale tests of the order of -8.85 % (18 knots) to +7.29 % (23 knots). As previously discussed the sea trial data also has associated uncertainty.

In a statistical or hybrid ship performance model, the model parameter uncertainty is a result of the errors that occur during the calibration/reference period. From this period a training dataset captures the ship performance in terms of a larger array of influential environmental/operational factors and the correction model is defined. Sources of error include:

1. Sampling error because the training dataset is a finite sample taken from an infinite temporal population, this is from
    a. A small sample size which means the range and deviation in the data during the reference period is insufficient to capture the interaction

between variables in the baseline performance, for example, if the ship is operated at only one draught then there will be no way to define the general relationship between draught, shaft power and other variables

   b. Sample averaging
   c. Calibration time period: there is a trade-off between a short period that ensures stationarity and a longer period that increases the number of data points. Non-stationarity in the training period may increase the model error as much, if not more than a shorter period with fewer data points.

2. Instrument uncertainty discussed in the previous section will also be present in the training dataset.
3. Model uncertainty, if the underlying model form is incorrect due to omitted or unobservable variables or incorrect assumed functional form. Omitted variable may be from ship steering / dynamics, sea water temperature / salinity corrections.
4. Human error of the data collection during the calibration period

The various sources of model uncertainty mean that the optimum data acquisition and analysis strategy is ambiguous. On the surface it would seem that CM data together with a normalising approach and a hybrid or statistical model would likely reduce model uncertainty when compared to NR data which is likely to increase model error arising from the reduced fields in the NR dataset and consequent increase in unobserved or unmeasurable variables (such as acceleration and water depth, for example). However NR data that is filtered and based on a hybrid model that combines CFD analysis with operational data and statistical method may reduce instrument uncertainty because there are fewer variables to measure and for the same reason perhaps even model uncertainty. Then the uncertainty in the overall PI also depends on how influential model uncertainty is relative to the other sources.

4.3.4 Human Error

Human error (which is sometimes categorised as instrument uncertainty) may occur in any measurements when operating, reading or recording sensor values if the report completion is not automated.

The automation associated with CM data means that NRs are relatively more likely to be subject to human error; measuring sea state, wind speed or wave height is particularly subjective as is the case with hind cast data, and is very dependent on the training of the observer. It is also anecdotally suggested that crew may be inclined to report higher BF numbers in order to satisfy the ship performance as specified in charter party agreements. Manual inputs are affected if the conditions are unstable during measurements, for example in tank dip readings of fuel consumption. Miss-reporting either by repeating previous days data entries or, when recording continuous variables as daily averages, such as wind speed, the mean or mode or the instantaneous rather than the average may be may reported depending on the crew; differences will arise if the ship has spent some of the period loitering. The noon data entry may not occur at exactly the same time each day as the recording of 'time spent steaming' may not be adjusted to compensate for crossing time zones and it is possible that different sensors are used to populate the same field, for example, some crew may report speed from the propeller rpm and others report speed through water.

4.4 Data Pre-processing (prior to use in any modelling)

The main performance analysis presented in the following chapters is with reference to CM data from a VLCC tanker and compared to and corroborated with the data of a Suezmax tanker, the ship particulars are presented in Table 5.

Outliers that represent physical impossibilities are initially removed such as power equal to zero for nonzero speed, and vice versa. The data was then filtered according to the criterion presented in Table 6. The level of filtering is subjective and requires balance between removing inaccurate data points that will incorrectly skew the results and preserving valuable information about the system physics.

Type	VLCC Tanker	Suezmax Tanker
Built	2000	2005
LOA (m)	335	269
Beam (m)	58	46
Draught (m)	22.7	17.5
Design Speed (knots)	16	15
Power (kW)	27 160	18 660

Table 5: Ship particulars

The filtering criterion here are based on the idea that the theory in the published literature that must be implemented to correct for a specific variable range is not reliable and therefore inaccurate. There are limited studies on engine performance at slow rpm for example and the SFOC in that region of the engine layout diagram is unlikely to be predicted with a degree of accuracy. There are corrections available for water depth, however the purpose of filtering rather than correcting is to help remove data associated with manoeuvring. Around ports for example, modelling ships manoeuvring is complex and not within the scope of this thesis. This filtering criteria is similar to and is based on discussions within the working group of ISO 19030 – Ships and Marine Technology – Measurements of Hull and Propeller Performance.

Water depth	<	$\max\left(h = 3\sqrt{B \cdot T_M}\ , h = 2.75\ \frac{V^2}{g}\right)$
Shaft power	>	110% MCR
	<	ship1: 15% MCR, ship2: 35% MCR
Speed	<	55% V_{des}
||STW-SOG||	>	1 knot
$FC_{residual}$ $= FC_{meas} - FC_{modelled}$	>	$\lvert\overline{FC_{residual}} - 2\sigma_{residual}\rvert$

Table 6: Pre-processing filter, defines the variable ranges for which data is removed

Ideally, vessel acceleration and rate of change of rudder angle would be included in the filtering criterion however the temporal resolution of the available data is 15 minutes which is not sufficiently low to capture the significant dynamic effects.

Fuel mass flow sensors are known particularly to malfunction (stick), also the presence of cat-fines in the fuel may cause large deviations. It is known that the power – fuel consumption relationship should be reasonably linear with a normally distributed scatter due to operational and environmental changes (as discussed in section 5.3). Therefore a linear model of these two variables that predicts the fuel consumption ($FC_{modelled}$) is used to remove outliers, as described by Figure 18. The model errors associated with this linearisation are assumed to be minimal.

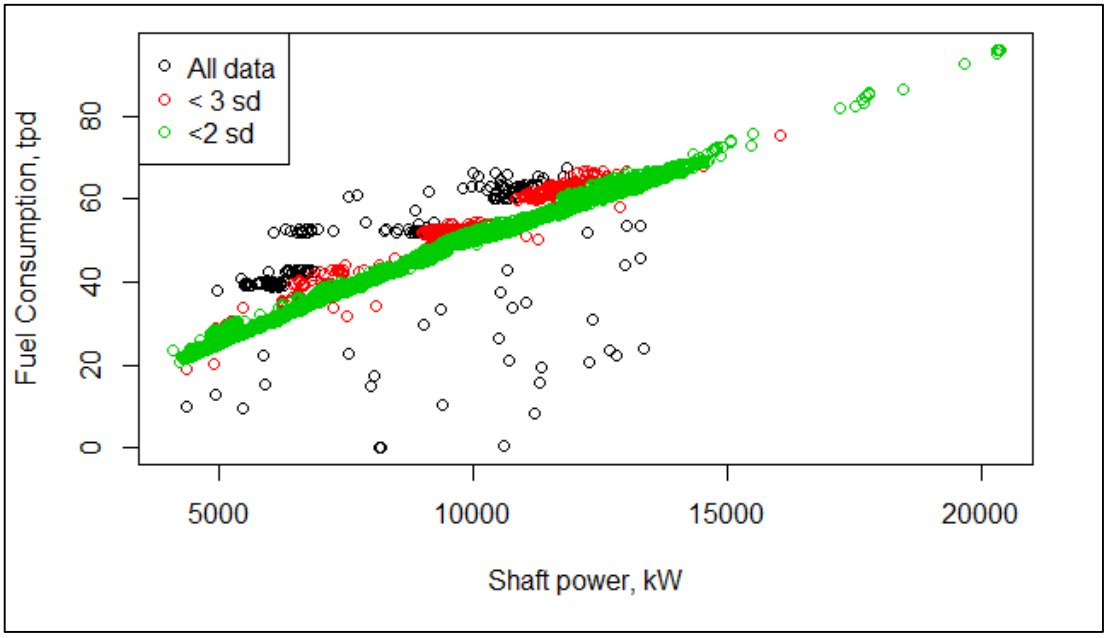


Figure 18: Fuel consumption and Shaft power relationship and the effect of successively removing points that are up to 2sd from the mean of the linear model prediction

Analysis of the VLCC used in this study shows that there were some very high fuel consumption values at the very low engine loads. The filter was therefore set to 0.35% design MCR.

4.5 Chapter Summary

A major difficulty in defining a framework for and subsequently quantifying the uncertainty associated with ship performance monitoring is that the above factors are interlinked. The model parameter uncertainty is dependent on the sample averaging/ recording frequency in the calibration period as is the overall performance indicator uncertainty as calculated in the evaluation period. A longer time period of data gathering will increase the variance of the input dataset and the number of

observations (determined by the sample averaging and recording frequency) thus reducing model parameter uncertainty but, depending on the rate of hull/propeller deterioration, a long evaluation period will increase the model standard error (SE) due to increased non-stationarity. A dataset acquired over a short one month time period may contain sufficient variability among predictor variables and therefore produce small coefficient standard errors relative to that of a three month dataset where the ship is only operated at one speed and loading condition (depends also on ship type and external factors current economic climate /environmental conditions). Sample averaging may improve uncertainty by smoothing out the effects of outliers, or increase uncertainty by reducing granularity of the data and disposing of detailed information about the relationships between the independent and dependent variables. This depends on the rate at which the independent variable fluctuates over time, for example significant draught variations may only occur between each voyage (slow moving variation), while wind speed or wave height may vary on a shorter time period, metocean conditions are estimated to be stationary for approximately 3 hours (fast moving variation). Finally, a high frequency dataset may have a large number of observations and thus report low standard errors/ uncertainties however if the variable measured is slow moving then no additional information is being added and the standard errors reported in the uncertainty reflects only that of the precision of the sensors of the input variables which may be useful or may be to an unnecessarily high level.

This reaffirms the contribution of the work in this thesis in understanding the interactions between and the significance of the data acquisition variables and to develop an understanding of the optimum combination for the desired application.

Chapter 5. Theoretical Model

The aim of the model is to determine the ship's fuel consumption for a range of environmental and operational conditions. The theoretical model is based on the physics of ship resistance and propulsion and the naval architecture and marine engineering relationships that are commonly used to analyse resistance and propulsion, as described in Chapter 2 (section 2.2). This is in the first instance about estimating the performance of a ship in its design condition and then looking at variations to that performance due to the operational and environmental conditions as measured during a calibration period. This then forms the normalisation model that enables the actual operating/environmental condition to be reverted back into a reference condition from which the difference between modelled and measured fuel consumption equates to changes in the ship's performance.

It is important to highlight that ideally, sea trial data, model test data and/or CFD data would be used in combination to help develop the most accurate ship performance model however there was no access to this data for the case study ships (listed in Table 5). It was therefore in the absence of such data that the following models were developed.

The focus of this chapter is to form and evaluate a theoretical normalisation model. It is not possible given the time and resources (computational power and financial) to model every possible interacting physical phenomenon; it is required to set an external boundary surrounding the system and within this to limit the desired level of detail that is required to achieve the solution to a sufficient level of accuracy. This is not only due to the resource limitations but also because the level of knowledge published in the literature is not sufficiently reliable or well documented and therefore there are diminishing returns, in terms of accuracy in the result, on investment in time and effort; particularly in terms of the scope of the overall thesis.

In this theoretical model the physical system boundary is the ship and its immediate environmental weather systems in which it operates (wind, waves, swell and current). This chapter aims to identify the major sources of influence within this boundary and to justify the point at which simplifications are appropriate to execute

the fuel consumption prediction algorithm. In the results part of this chapter (section 5.6), the model is firstly compared to published sea trial data to build confidence in the 'at design' component of the model. Metrics are then identified with which the model performance is evaluated and against which this model can be compared to the alternative models presented in chapters 6 and 7.

The nomenclature and units used in this chapter are summarised at the beginning of this thesis.

The ship's hull, propeller and engine basic characteristics are obtained from the Clarksons World Fleet Register (CWFR) (see details in Chapter 4), these variables are summarised in Table 7 which also summarises the well-established formulae, from which other basic ship characteristics are calculated and the variables which are determined from assumed empirical formulae the details of which follow later in this chapter. Also listed are some generally accepted defined constants.

CWFR inputs	**Calculated**	**Assumed from empirical formulae**	**Constants**
Lbp B T_{des} V_{des} MCR SFOC D_m	$C_p = \frac{\Delta}{C_B Lwl.B.T}$ Lwl = Lbp/0.97 $F_n = \frac{V_{des}}{\sqrt{g.Lwl}}$ $R_n = V_{des}\frac{Lwl}{\nu_k}$ $\nabla = C_b$. Lwl . B. T_{des}	WSA C_F D LCB k_1 C_b C_m	$\rho_{sw} = 1025$ $\rho_{air} = 1.23$ $\nu_k = 1.188x10^{-6}$ g = 9.81

Table 7: Summary of hull, engine and propeller design characteristics and calculated parameters (for units, see the nomenclature at the beginning of this thesis)

Throughout this thesis the signing convention follows that of the ITTC; it is based on the Cartesian right hand co-ordinate system with the x-axis being positive forward (x faces the bow). Angle of apparent wind, ε = 0 in bow wind and the y-axis positive starboard, both are horizontal.

The propeller shaft efficiency depends on the shaft length, number of bearings and the gearbox, without a gearbox this is approximately 0.98 (Kristensen 2012) and 0.98 is the value assumed here. Z is assumed to equal 4.

Condition	**Affecting factors**	**Further disaggregation**	**Included in the theoretical model?**
Calm water – Steady state	Draught		Yes
	Speed		Yes
	Trim	Resistance components	Yes: C_w No: C_{fr}
		Propulsive effects	Yes: w, η_o No: t, η_{RR}
	Appendages		No
	Bulbous bow		No
	Transom stern		No
Calm water – Manoeuvring	Accelerations		Filtered indirectly through low speed and shallow water filters*
	Rudder angle		
Environmental	Wind	Longitudinal	Yes
		Transverse	Yes
		Yaw	No
		Roll	No
	Surface waves		Yes
	Swell		No
	Water depth		Filtered
	Current		Yes (STW is used)
Time	Hull fouling		This effect is the objective function
	Propeller wear		

Table 8: Factors effecting shaft power during operation. *(some may remain in adverse environmental conditions for example)

Table 8 provides a summary of the conditions that affect the ship's required shaft power during operations, the remainder of this chapter presents further details regarding these and justifications for the exclusion of specific factors. From the shaft

power the fuel consumption is derived which is also affected by time due to engine wear, soot build up etc. This is also discussed in further detail in section 5.3.

The temporal effects are not included in the model because, as defined in the closing paragraph of Chapter 3, the objective is to evaluate their combined effect. This is defined by the change in "ship performance indicator" and the ship PI is the difference between the actual measured fuel consumption at the experienced environmental/operating condition and the expected fuel consumption at that condition as determined by a baseline.

5.1 'At-Design' Conditions

The governing equation for ship shaft power is:

$$P_S = \frac{P_E}{\eta_R \eta_O \eta_H}$$

Equation 3

The ship effective power is the product of ship speed and total resistance. The components of ship total resistance in still water are examined in the following section. The efficiency terms in the denominator make up the quasi-propulsive efficiency, η_D and are discussed in the next section (5.1.2).

5.1.1 Resistance

A number of approaches can be used for the estimation of total resistance, R_T of a ship as described in section 2.2. In the theoretical model of this thesis, the method for calculation of total resistance is based on the analysis of Holtrop and Mennen, H&M (Holtrop and Mennen 1982) which is an approximate procedure based on hydrodynamic theory with coefficients obtained from the regression analysis of the results of model tests and full-scale data available at the Netherlands Ship Model Basin. It applies to a wide range of ship types including tankers, general cargo ships, fishing vessels, tugs, container ships and frigates. It was updated to include an even wider range of ship models (334 total), specifically high speed craft with Froude

number greater than 0.5 and including the influence of propeller cavitation and submergence (Holtrop 1984). The formulas established are not technically theoretical since they are regression equations based on the statistical analysis of empirical data however this classification conforms to the literature. The H&M method is one of the most well-known and flexible methods (Kristensen 2012), the predictions are widely used at the initial design stage of a ship (Prpić-Oršić and Faltinsen 2012) and they agree well with other performance prediction methods such as Jalkanen, Johansson et al. (2012) and Hollenbach (1998). The following limitations however are noted:

1. The analysis is based on a relatively small number of experiments and extrapolation to different ship types and sizes may not be accurate; the application is limited to hull forms resembling the average ship described by the main dimensions and form coefficients used in the method.
2. The relationships are developed for a new ship in the at-design condition
3. The relationships were developed 30 years ago and may not be accurate for today's ships

These limitations to the model accuracy are recognised however there has been a trade-off between the number of parameters that are available to describe the ship (hull geometry and propeller characteristics) and the potential fidelity of a theoretical model. The data available is limited by the minimal fields describing these particulars that are included in the CWFR (see Table 7) therefore, whilst higher precision methods are available there is little point investing time and resource into employing these methods if the assumptions that must be made on the missing ship specifications lead to higher inaccuracies than the improvement in accuracy that is hoped to be made.

Details of the H&M method can be found in the referenced paper, a general overview is described below and where more recent assumptions/formulas are found in the literature and their use is justifiable, then these are also presented below.

The ships total resistance, R_T:

$$R_T = R_F(1 + k_1) + R_{APP} + R_W + R_B + R_{TR} + R_A = \frac{1}{2}\rho.WSA.V^2.C_T$$

Equation 4

Many equations exist for WSA, in this thesis the formula of H&M is replaced by that of Kristensen (2012) who analyses more recent ship designs and analysed. The formula is based on that of Mumford, and the formula constants adjusted in order to improve the accuracy given the data.

Bulk carriers and tankers	WSA = $0.99.\left(\frac{\Delta}{T} + 1.9.Lwl.T_{des}\right)$ **Equation 5**
Container vessel (single screw)	WSA = $0.995.\left(\frac{\Delta}{T} + 1.9.Lwl.T_{des}\right)$ **Equation 6**

Table 9: WSA of different ship types

In the absence of detailed data the appendage resistance and additional pressure resistance due to an immersed transom are not included, it is noted though that appendages may increase resistance up to 12% for single screw ships (Schneekluth and Bertram 1998). The magnitude of the effect is also variable in operational conditions depending on trim, speed and draught.

The assumptions behind the quantification of other resistance effects are described below.

Frictional Resistance

According to the ITTC 1957 Model-ship correlation line (ITTC 2002):

$$\frac{R_F}{\frac{1}{2}\rho.WSA.V^2} = C_{Fr} = \frac{0.075}{(logR_n - 2)^2}$$

Equation 7

The form factor, k_1 in equation 4, is calculated from the hull beam, length, LCB and prismatic coefficient, as presented by H&M. LCB is not defined in H&M and the following formula, applicable to ships with a bulbous bow, from (Schneekluth and Bertram 1998) is used:

$$LCB = 8.8 - (38.9.Fn)$$

Equation 8

C_p, the prismatic coefficient is calculated from the block coefficient which is a function of the Froude number and is based on the following assumption (Watson 1998):

$$C_p = 0.7 + \left[\frac{1}{8}\text{atan}\left(\frac{23 - 100Fn}{4}\right)\right]$$

Equation 9

Correlation Resistance

The correlation resistance is added in order to include the effect of the roughness of the ship hull in the at design condition which will differ from the model and the still air resistance. The additional hull roughness is derived from ITTC 1978 performance prediction method (ITTC 2011) and is implemented when the hull roughness of the new ship is higher than the standard 150μm (mean apparent amplitude). In the absence of data the hull roughness of the new ship is assumed to be the standard. The still air resistance is according to H&M. Other assumptions are proposed by Kristensen (2012) however the values of the air resistance coefficient seem small ($0.04x10^{-3}$ for a VLCC) relative to the values proposed by (Schneekluth and Bertram 1998); 0.8 to 1.0 for cargo ships.

Wave Making Resistance

The wave making resistance is established from the H&M method.

The midship section coefficient, C_M is an input to the method and is approximated according to the following formula (Schneekluth and Bertram 1998):

$$C_m = 1.006 - 0.0056{C_b}^{-3.56}$$

Equation 10

Additional pressure resistance of bulbous bow

Predominantly, ships have a bulbous bow installed to improve efficiency by reducing wave making resistance (Schneekluth and Bertram 1998). H&M present a detailed calculation of the additional resistance due to the presence of a bulbous bow which requires inputs of the position of the centre of the transverse area, and the transverse area. The bulb correction is a function also of the Froude number and is also draught and trim dependent, the dependency is complex (Kristensen 2012). The effects of the bulbous bow are numerous including impact on resistance in a seaway (wave making and frictional), sea keeping characteristics, propulsion efficiency (due to changes in the uniformity of the flow and the thrust coefficient), course-keeping ability and the effective drag which depends on draught and trim (Schneekluth and Bertram 1998).

Given this complexity, the formulae detailed in H&M which is useful in the specific case of the 'at design' condition may be of limited value if adapted to the operational conditions, therefore the effect of bulbous bows is not included in the model, the resultant inaccuracy is discussed in relation to the development of the statistical model (see Chapter 6).

5.1.2 Propulsion

The H&M method outputs the hull efficiency and the effective power (the product of resistance and vessel speed), the shaft power is calculated in equation 2, where the denominator, the quasi-propulsive efficiency (η_D) is calculated according to:

$$\eta_D = \eta_R \eta_0 \eta_H = \frac{J}{2\pi}\frac{K_T}{K_Q}\eta_H \eta_R$$

Equation 11

The propulsion model is based on a fixed pitch propeller, where propeller diameter and design shaft rpm in the design conditions are exogenous variables. The following assumption applies for the propeller diameter, D:

$$D = 16.2\left(\frac{0.8MCR^{0.2}}{RPM_{des}{}^{0.6}}\right)$$

Equation 12

It is assumed that the MCR for calm weather and design speed and draught is 75% of the installed engine power which allows for 15% sea margin and 10% engine margin (MAN 2011). The heavy propeller curve (fouled hull, heavy weather) is assumed to be the design basis for the engineering operating curve in service and the light running factor is assumed to be 5% (MAN 2011) where

$$f_{LR} = \frac{n_{light} - n_{heavy}}{n_{heavy}} x100\%$$

Equation 13

In the first iteration the objective function is to maximise propeller efficiency at the design condition (hull geometry, propeller diameter, vessel speed and loading) by evaluating efficiency over a range of propeller pitch using the regression analysis of the open water characteristics of the Wageningen B-Series propellers as investigated by Oosterveld and Oossanen (1975). For each 0.05 increment of pitch diameter ratio between 0.4 and 1.9 the thrust and consequent power delivered by the propeller is calculated, when this meets the power required according to equation 7 then the propeller design pitch diameter ratio is fixed. See example characteristic propeller curves of Figure 19 and Figure 20. The propeller blade thickness (at 0.7R) to chord length ratio is also checked to ensure material strength according to the criterion presented by Gaafary, El-Kilani et al. (2011). The in-and-out-of-water effect and

ventilation of the propeller effects are not taken into account but can be severe (Prpić-Oršić and Faltinsen 2012) however it is thought that the time period over which this may occur is assumed to be a small proportion of the overall time.

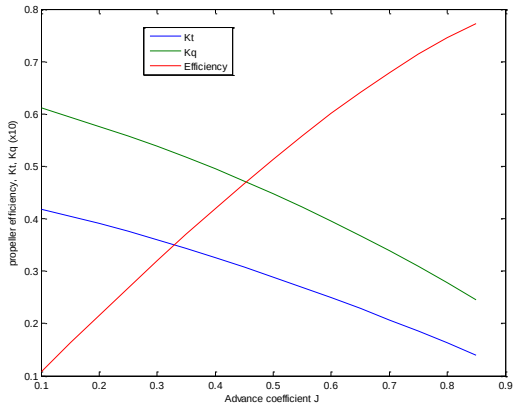


Figure 19: Propeller curves for large crude oil tanker (VLCC)

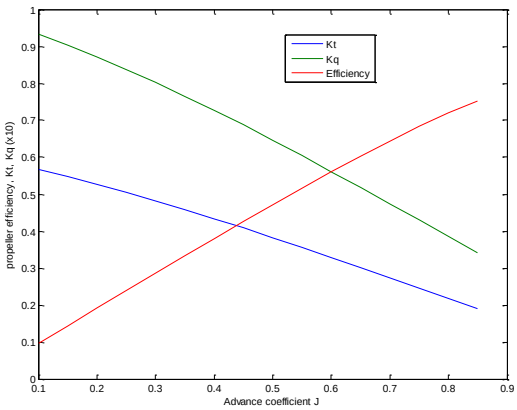


Figure 20: Propeller curves for large containership

5.2 Operational Conditions

The resistance and propulsion characteristics of the 'at design' condition are adjusted according to the operational draught, trim and speed according to the assumptions described below.

5.2.1 Resistance

5.2.1.1 Speed and draught

For the resistance component, at each ship speed and draught, the H&M algorithm is implemented to determine the effective power and hull efficiency.

The effect of a change in draught is to increase the frictional resistance due to a change in surface area. The 'at design' WSA, WSA_1, is adjusted to the operating draught WSA_2, according to the change in draught from T_1 to T_2 and from the following assumptions based on the statistical analysis of Kristensen (2012).

Containerships:	$WSA_2 = WSA_1 - 2.4(T_1-T_2)(Lwl+B)$
Tankers and bulk carriers:	$WSA_2 = WSA_1 - 2.0(T_1-T_2)(Lwl+B)$

Table 10: Wetted surface area assumptions according to Kristensen 2012

When not at the design draught, the length and beam at the waterline are assumed to be constant and the following formula for the block coefficient is used :

$$C_{b2} = 1 - \left[(1 - C_{b1})\left(\frac{T_1}{T_2}\right)^{1/3}\right]$$

Equation 14

The corresponding displacement in the partially loaded condition, Δ_2 can then be approximated from:

$$\Delta_2 = \Delta_1\left(\frac{C_{b2}}{C_{b1}}\right)\left(\frac{T_1}{T_2}\right)$$

Equation 15

The mid ship area coefficient is calculated according to equation 6 and the prismatic coefficient is then found from Δ_2, T_2 and C_{b2}. The new resistance calculations are then made according to H&M. Figure 21 shows the relative magnitudes of the components of resistance as a function of ship draught when the vessel is sailing at its design speed; this is a VLCC. This compares favourably with the data presented in ABS (2014), although they show the VLCC viscous resistance to be slightly higher; approximately 92% as a proportion of total resistance at design speed and draught.

For the propulsion system, the PDr is fixed as described previously and the required propeller speed is obtained from the Wageningen B-Series propeller curve characteristics; the wake faction and speed determines the new speed of advance and then for each $1x10^{-3}$ increment of advance coefficient, between 0.3 and 0.9 the thrust and consequent power delivered by the propeller is calculated, when this meets the

power required according to equation 7 then J is fixed and the propeller speed is found.

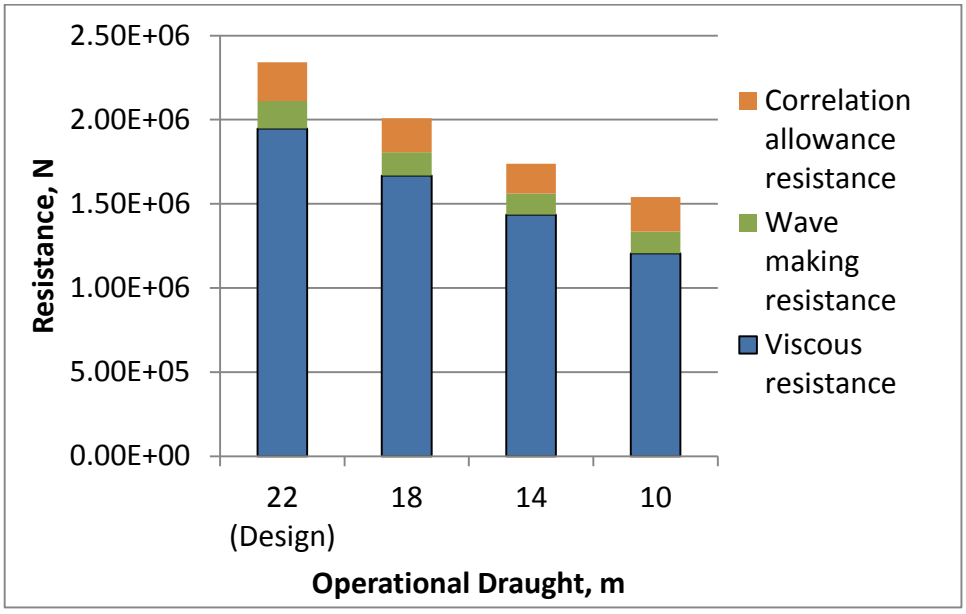


Figure 21: Relative magnitudes of the components of resistance as a function of ship draught (VLCC)

5.2.1.2 Trim

The effect of trim was studied from model tests and CFD analysis by FORCE Technology, N. L. Larsen et al. (2011) who decompose the overall per cent savings in propulsive power into contributions from different elements of the ship resistance and propulsion system. Their results are summarised in Table 11 which indicate that approximately 75% of the effect of trim on the propulsion is due to changes in the ships resistance and of that the wave making coefficient is most significant (they conclude the major contribution to the change in residual resistance is the wave generation around the bulbous bow). The model-correlation coefficient is generally constant over changing draughts, except in the case of a very large vessel, eg a VLCC. The change in propulsion power due to the waterline length is comprised of the effect of the change in Reynolds number and frictional coefficient.

The results of FORCE Technology, N. L. Larsen et al. (2011) are compared to those of the H&M code to study how the model performs for a similar ship type and size (post panama container ship). These are summarised in Table 12. For the FORCE

model tests and for the H&M model the Froude numbers are 0.128 and 0.115 and the form factors are 1.13 and 1.15, respectively.

Trim	**-2.0**	**+2.0**
Resistance Components		
$\Delta P_{D\,Cw}$ (%)	-8.8	16.4
$\Delta P_{D\,Lwl}$ (%)	0.2	-0.3
$\Delta P_{D\,WSA}$ (%)	-0.3	0.1
TOTAL	**-8.9**	**16.2**
Propulsive effect		
$\Delta P_{D\,w}$ (%)	-3.5	1.3
$\Delta P_{D\,t}$ (%)	2.5	0.3
$\Delta P_{D\,\eta 0}$ (%)	0.1	1.5
$\Delta P_{D\,\eta R}$ (%)	-1.7	0.6
TOTAL	**-2.6**	**3.7**

Table 11: Summary of per cent change in delivered power due to trim effects on each ship resistance component and propulsion system component, FORCE Technology, N. L. Larsen et al. (2011)

The H&M code only reflects changes in the wave making coefficient and the wake fraction due to trim, although these are the two largest components of the effects of trim on the propulsion power. The changes according to the H&M code are of a much smaller magnitude than those found in the model tests. This is likely because the H&M code does not include the effect of a bulbous bow. Also, the absolute values at zero trim provide evidence that the theoretical model is broadly correct, a more rigorous validation of the theoretical model is conducted in the final part of this chapter.

Trim	**-2.0m**		**0.0m**		**2.0m**	
	FORCE	H&M	FORCE	H&M	FORCE	H&M
Resistance Components						
WSA (m^2)	16181.4	16225.5	16223.6	16225.5	16241.2	16225.5
ΔWSA (%)	-0.26	0	0	0	0.11	0
ΔLwl (%)	-2.5	0	0	0	1.8	0
R_n ($x10^9$)	1.91	1.95	1.95	1.95	2.00	1.95
C_{fr} ($x10^{-3}$)	1.415	1.411	1.412	1.411	1.407	1.411
ΔC_{fr} (%)	0.21	0	0	0	-0.35	0
C_W ($x10^{-4}$)	0.680	2.53	2.34	2.67	5.41	2.81
ΔC_W (%)	-70.9	-5.2	0	0	131.7	5.3
Propulsive effects						
t	0.166	0.175	0.145	0.175	0.147	0.175
Δt (%)	14.9	0	0	0	1.7	0
w	0.209	0.300	0.181	0.287	0.17	0.278
Δw (%)	15.5	4.5	0	0	-6.1	-3.3
J	0.751	0.689	0.752	0.694	0.729	
η_0	0.638	0.682	0.639	0.686	0.629	0.689
$\Delta\eta_0$ (%)	-0.1	-0.6	0	0	-1.5	0.4
η_{RR}	1.005	0.974	0.988	0.974	0.982	0.974
$\Delta\eta_{RR}$ (%)	1.7	0	0	0	-0.6	-0.05

Table 12: The effect of trim on elements of ship power and propulsion as determined by the H&M output and the FORCE results *346.8041, ship speed=13knots, draught=11.1m

5.2.1.3 Environmental

The environmental variables corrected for in the model are side and longitudinal wind and waves. The environmental data used in this calculation, especially in the case of noon report datasets, is extremely approximate and the need for something robust and pragmatic is a trade-off with fidelity. Therefore the investment of significant time to developing a precise theoretical environmental model may be negated by inputting data of poor quality. Current is not corrected for since speed through the water is measured. The effect of swell is not included since this data is rarely recorded in noon report datasets. The extra resistance created by shallow water

effects, due to the three-dimensional flow of water approaching a two dimensional flow field and thus increasing water pressure on the ship's movement as larger waves are created, are also not included. The magnitude of this effect will depend on a ship's dimensions, speed, and type (Sun, Yan et al. 2013), therefore a water depth filter is applied which prescribes a minimum water depth see Table 6.

Wind resistance

A ships superstructure is a complex, bluff body and therefore the aerodynamic details of its interaction in wind is not amenable to theoretical analysis, empirical analysis of wind tunnel tests are therefore used to predict the resultant forces at full scale (Turk 2009). Wind resistance is particularly significant for ships with a large lateral area above the water level, i.e. containerships. The effect of the wind, as explained by Andersen (2013), is largely comprised of the longitudinal and transverse force due to aerodynamic drag. The longitudinal force is due to the ships forward speed and relative wind direction and according to Andersen constitutes the largest part of the total wind induced resistance. The wind field around the ship is a combination of the wind field from the air flow over the ocean which has a boundary layer and the local wind field caused by the ships velocity which is homogenous and without a boundary layer. The combination of both determines the actual wind field and this is defined by the relative speed and relative direction as calculated by vector analysis. The significant effects of the transverse force, causing yaw, drift and deviation, is twofold; misalignment of the ships heading with the course and increased rudder angle (to account for drift). The significance of rudder angle effect on induced resistance is debatable. The transverse force also causes roll due to the moment around the x-axis, this however does not give rise to added resistance (Andersen 2013), but instead increases wave making resistance. The yawing moment and rolling moment coefficients are not included in this analysis because they have a lesser effect on the overall resistance and because the position of the lateral plane centroid is unknown and may introduce further inaccuracies.

The general relationship between shaft power and wind speed, aerodynamic drag, can be approximated by a drag coefficient based on the transverse projected area of the ship perpendicular to the wind direction and the square of the wind speed (N.

Hamlin and Sedat 1980). The ITTC guidelines recommend correction methods during full scale speed and power performance trials (ITTC 2005), there is stipulated that the resistance increase due to wind be calculated from

$$R_{wind} = C_{wind} \frac{\rho_{air}}{2} (V_{wind,rel})^2 . A_F$$

Equation 16

The wind resistance coefficient (C_{wind}) is a function of the wind resistance coefficient in a head wind and the directional coefficient of the wind resistance. The literature provides a plethora of published wind tunnel test results that may be used to inform ship performance prediction in wind. A simple relationship between relative wind speed, $V_{wind,rel}$, ship length, L and the air resistance, R_{air} is defined by Berlekom (1981) as follows:

$$R_{wind} = 0.615 C_{wind} L^2 V_{wind,rel}^2 \times 10^{-6}$$

Equation 17

C_{wind} is the wind force coefficient which varies according to the ship type and its transverse sectional area. The coefficients of the above expressions may be derived from model tests or summarised by statistical analysis of published results such as those presented in Isherwood (1973), or by analysis of wind tunnel tests as in Gould (1982) and Blendermann (1994). OCIMF investigated wind force coefficients for moored vessels from studies in the 1960's, however these were specifically from VLCC data which may be extrapolated to smaller ship sizes with similar geometry but other ship types may lead to error. These sources are compared in Haddara (1999) who subsequently derive coefficients by training a neural network over a range of ship types in different loading conditions. A qualitative assessment of the accuracy of this method concludes this produces better predictions that those found in the literature however the model is not published. A detailed review of experimental studies may be found in Turk (2009).

The published data of Blendermann is used in this thesis because it covers a wide range of ships in both the loaded and ballast condition and because, in a comparative study of four different methods (Turk 2009), it was found to produce the most accurate results and found to be the most comprehensive and reliable. The overall resistance is decomposed into separate expressions for longitudinal (X axis) and side force (Y axis) coefficients (as well as the coefficients for the yaw and rolling moments, not included) versus the angle of the wind. These are:

$$R_{wind,X} = \frac{\rho_{wind}}{2}(V_{wind,rel})^2 . A_F . CD_l . \frac{A_L}{A_F} . \frac{\cos(\varepsilon)}{1 - \frac{\delta}{2}\left(1 - \frac{CD_l}{CD_t} sin^2 2\varepsilon\right)}$$

Equation 18

$$R_{wind,Y} = \frac{\rho_{wind}}{2}(V_{wind,rel})^2 . A_L . CD_t \frac{\sin(\varepsilon)}{1 - \frac{\delta}{2}\left(1 - \frac{CD_l}{CD_t} sin^2 2\varepsilon\right)}$$

Equation 19

Coefficients are tabulated for ship type and loading condition, for example, for a tanker:

Vessel type	CD_t	$CD_{l,AF}$ $\varepsilon = 0$	$CD_{l,AF}$ $\varepsilon = \pi$	δ
Tanker, loaded	0.70	0.90	0.55	0.40
Tanker, ballast	0.70	0.75	0.55	0.40

Table 13

The area of maximum transverse section exposed to the wind, A_F and the lateral plane area, A_L are estimated as follows:

$$A_F = B.(D_M - T_{des} + H_{accom})$$

Equation 20

Where the accommodation height, H_{accom} is based on an estimate of the number of decks and 3m per deck (Kristensen 2012).

$$A_L = L_{WL}.(D_M - T_{des}) + 30H_{accom}$$

Equation 21

The wind speed, V_{wind} is the apparent wind speed relative to the ship as deduced from vector analysis of the true wind from metocean data and the ships heading, or from an onboard weather vane and anemometer if available. Wind speed in Beaufort number from NR data is converted to m/s by, $V_{wind} = 0.836.BN^{1.5}$ (Schneekluth and Bertram 1998).

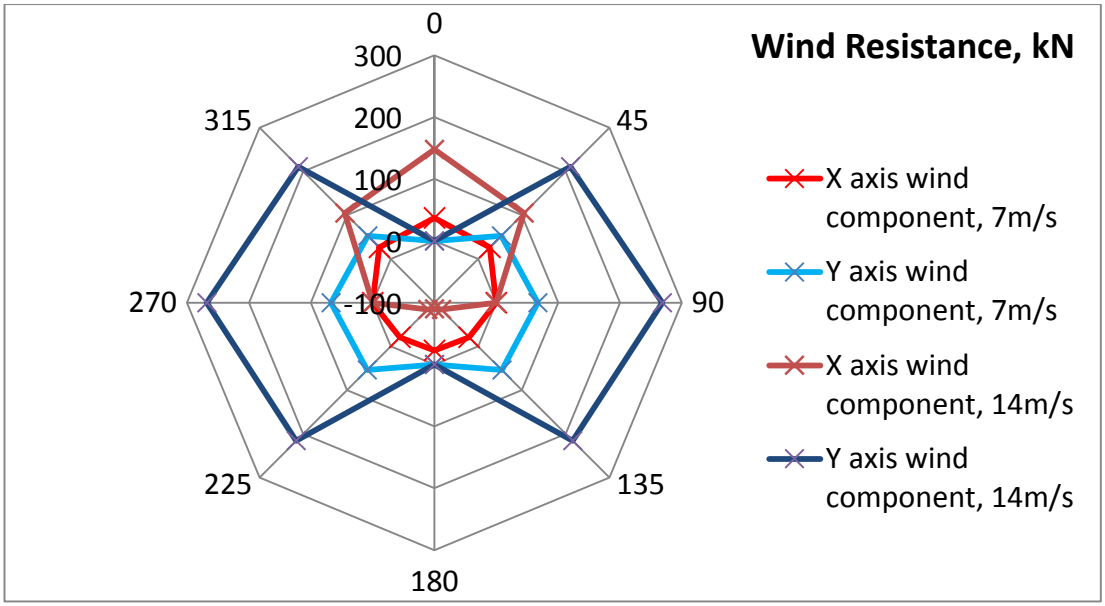


Figure 22: Effect of wind on ship resistance (kN) for a VLCC operating at design draft and speed (8.23m/s)

Figure 22 shows the effect of wind speed and direction on both components of resistance for a VLCC travelling laden and at its design speed, for comparison, its total resistance in the same condition (no wind effects) is 2147 kN. The Y axis resistance is the side force component which has negligible effect for head or aft wind but in winds coming from 90 deg or 270 deg it has a more significant effect on resistance compared to a longitudinal force of any direction at the same magnitude. The longitudinal force, X-axis component, has a negative impact on overall

resistance in a tail wind with increasing impact at higher wind speeds and works to increase ship resistance in a head wind.

Wave Resistance

There is some discussion as to the relative magnitude that wind and wave resistance contribute to total resistance. There is general agreement that under normal operational circumstances the wave resistance contribution is greater, but this depends on the instantaneous environment in which the ship is operating Andersen (2013).

The added resistance in waves is due to the energy dissipated from the ship to the waves both in the form of a diffraction effect of incident waves interacting with ship radiated waves and, to a larger extent, a drifting force due to incident waves interfering with waves diffracted when encountering a ship's hull. A third, "viscous" effect due to vertical motion damping has a less significant effect, (Pérez Arribas 2007). These effects from the drifting force, diffraction effects and viscous motions are non-linearly related to the wave amplitude and are estimated to be additive with the drifting force being the most significant. The drifting force causes heave and pitch in head seas and this effect is more significant on the added resistance relative to the roll, yaw and sway effects on added resistance in beam and quartering seas. There are various methods available for modelling the added resistance from simple approximate methods to more involved theoretical models that examine the wave energy spectrum and the energy transferred from the ship to the water. A logically presented summary and analysis of the major methods can be found in (Pérez Arribas 2007).

In a similar way to the wind tunnel model tests, wave model tests also have limitations when translated to operational environment. The model performance depends on the sea state and in particular the wavelength and so the most accurate method depends on the conditions and there may be no universal 'optimum'. Also, the approximate methods are generally derived from model tests under controlled, experimental conditions, in these tests quantifying the effects of interactions due to exogenous parameters that are of interest in the operational environment (such as

trim or hull fouling) are generally limited. Therefore, these results, while useful in the design process do not translate perfectly to operational situations. For the practical application in this thesis, a simple model is sought, one that is based on the drag coefficient for wave resistance and the square of the wave height which is proportional to the ships power for constant ship speed and wavelength. Various experimental analysis has indicated that this is a sufficient relationship for practical purposes (Pérez Arribas 2007). The following relationship between wave height and added power due to waves is assumed (Lindstad, Jullumstrø et al. 2013):

$$P_{waves} = \frac{C_w \rho g \left(H_s/2\right)^2 B^2}{Lwl} V_{waves}$$

Equation 22

The wave speed, V_{waves}, relative to the ship's speed is found from vector analysis of the true wave speed and direction given the ship's speed and direction. The true wave speed is the product of the wave frequency and wave length (λ) where;

$$\lambda = \frac{(gT_z)^2}{2\pi}$$

Equation 23

The drag coefficient for the waves, C_w is found from the H&M algorithm.

The above wave model only represents wind driven surface waves and does not account for swell. It also does not reflect the fact that the ship's speed may have to be reduced in heavy weather so as to not violate the maximum torque/rpm allowance of propulsion system components.

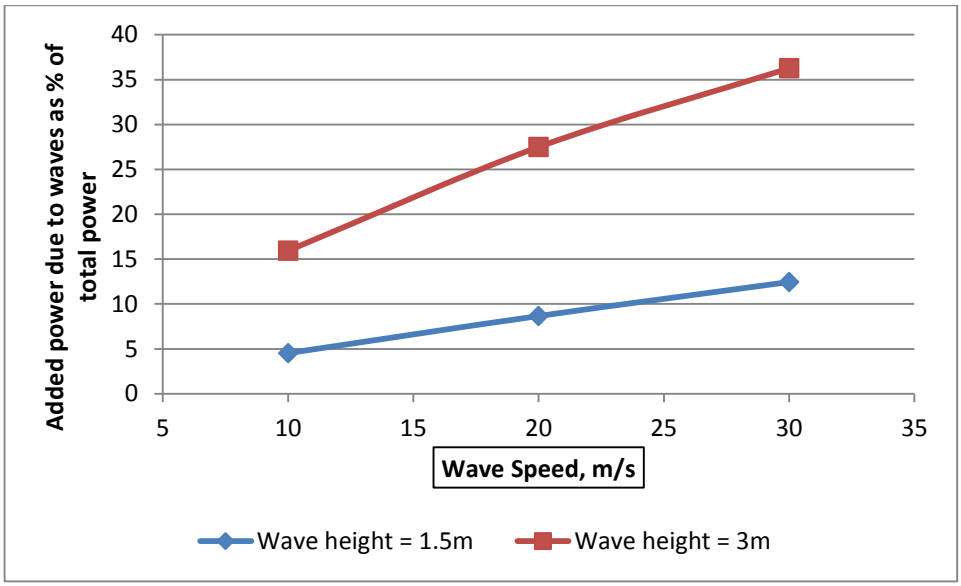


Figure 23: Added power due to waves as a per cent of the total power as a function of wave speed and wave height

5.2.2 Propulsion

As previously discussed in section 5.1.2 the propeller pitch and diameter is defined by the resistance, thrust requirements and the propeller performance in the design conditions. The propeller performance in terms of efficiency and thrust delivered is then deduced at each level of the advance coefficient and so for each combination of operational input parameters the propeller speed, which is endogenous, is determined. The overall chain of dependency is described diagrammatically by Figure 24, as can be seen the input parameters on the far right determine the effective power and the thrust required from which the propeller efficiency is ascertained by calculating the rpm that delivers the required thrust. This is then combined with the hull efficiency and the relative rotative efficiency (not shown), to determine the shaft power delivered which, in conjunction with an assumed shaft and gear efficiency, is transformed into brake power to be output by the prime mover which translates into fuel consumption depending on the SFOC.

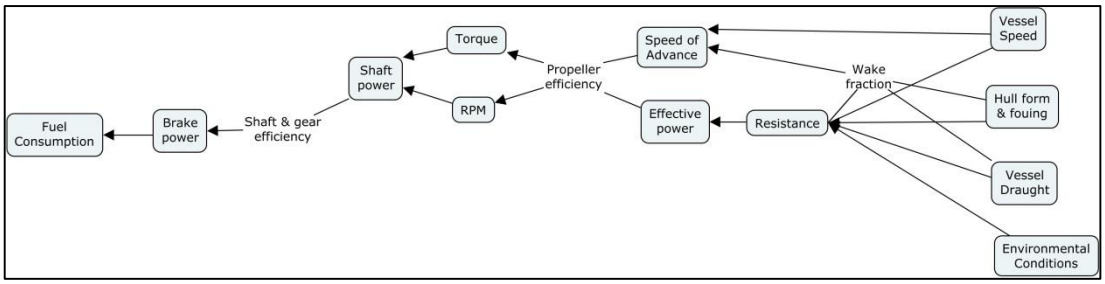


Figure 24: Fuel consumption factors

5.3 Engine

The above analysis is focussed on determining the shaft power requirements given the effective towing power and the hull, propeller open water and relative rotative efficiencies. The brake power delivered by the engine, as determined by the shaft speed (rpm) and torque, is reduced along the shaft due to the shaft efficiency (here assumed to be a constant 0.98, as discussed previously). The fuel consumption required to deliver the required torque at the given rpm is determined by the engine's specific fuel oil consumption (SFOC).

The manufacturer provides data regarding the engine performance in terms of SFOC and engine load relationship, and can be related using the engine layout diagram which identifies the SFOC to the engine speed and power using lines of constant SFOC and isobars. The marine diesel engines are optimised for a specific load in the 'at design' condition and according to the 'at design' propeller curve, deviating from that will cause a change in SFOC; there are three major causes for the engine performance to deviate from the optimum as specified by the manufacturer:

- The engine is operating off design due to change in engine load caused by normal operational decisions and environmental conditions; deviations in speed and draught or bad weather will adjust the ship's resistance, power requirements and propeller efficiency for example.
- The manufacturer specifications often relate to test bed performance, the actual operating performance deviates from these tests which are conducted with the engine burning distillate fuels at ISO conditions (Ozaki Y. 2010). In reality fuel quality and the presence of cat-fines may affect the SFOC in an unpredictable manner.

- Time dependent factors (see next section)

With regards to the off design condition the shape of the function relating engine load and SFOC is defined according to ship type and size, this is illustrated in Figure 25 which defines each of the baseline curves. These curves are the result of the study of manufacturer data originating mostly from Wartsila engine data sheets with the exception of the LNG duel fuel engine and the 2-stroke medium speed and 2-stroke low speed engines which are from MAN engine data.

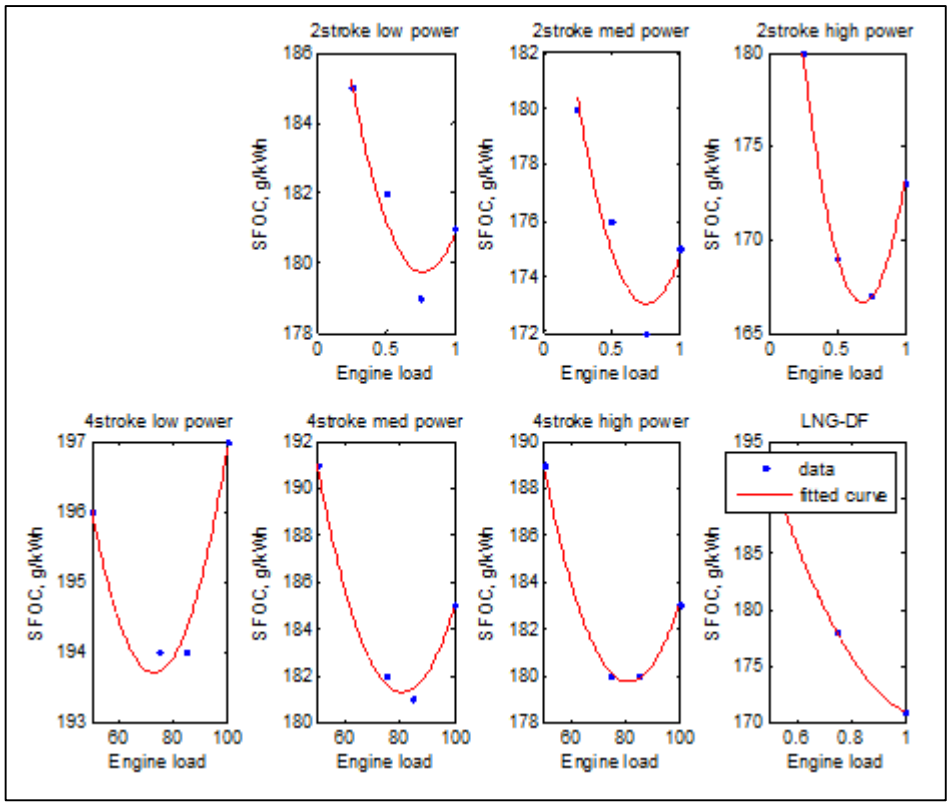


Figure 25: Specific fuel oil consumption as a function of engine load according to engine size and type

The SFOC baseline curves are related to the ship specific data by shifting them by delta SFOC:

$$\Delta SFOC = SFOC_{des} - SFOC_{min}$$

Equation 24

The $SFOC_{des}$ is ship specific and is obtained from Clarksons World Fleet Register, it is assumed to be quoted at the optimum engine efficiency in the 'at new' condition and the $SFOC_{min}$ is the minimum SFOC on the baseline curves (Figure 25). A factor of 1.05 is applied to the $SFOC_{des}$ to account for the deviation between test bed and actual operational data (J. J. Corbett, V. Eyring et al. 2009). This factor is assessed in the results section of this chapter and the hybrid model explores how the factor may be improved by deploying the model on real data.

Fuel quality and the presence of cat-fines and other fuel quality influences such as non-combustible contaminants (e.g. water) are unpredictable and assumed negligible although it is possible that their effects may be significant.

5.4 Time dependent factors

The first research question is concerned with the quantification of ship performance, and the development of a method to measure the trend in the ship condition over time. The trend is defined by the change in the "ship performance indicator"; the ship PI is the difference between the actual measured fuel consumption at the experienced environmental/operating condition and the expected fuel consumption at that condition as determined by a baseline (see section 4.2). The fuel consumption delta represents the deterioration in hull, propeller and engine performance. Therefore, it is not the purpose of the research to model the deterioration but to measure it, however the expected theoretical time effects need to be understood to be clear about what the residual (the delta that is the performance indicator) represents. This is the focus for the next section.

Hull and Propeller Degradation

The effects of hull and propeller degradation are the predominant time dependent factors that are to be measured by the analysis through the variation in the difference between modelled and measured shaft power over time. For datasets that do not include shaft power measurements (NR datasets) then the PI becomes the delta fuel consumption due to hull and propeller degradation. This is therefore a proxy for shaft power under the assumption that the engine performance deterioration is small

relative to the hull and propeller performance changes (the effect on the PI uncertainty of this assumption is studied in Chapter 9).

The effect of hull degradation is predominantly due to two types of roughness (Taylan 2010); permanent/physical (such as corrosion, shell plating deformations or welding seams) and temporary/biological (such as animal, weed or slime fouling). The major effects on ship performance are through the increase in frictional resistance characteristics, discussed in greater detail in Taylan (2010), Townsin (2003) and Schultz (2007). The effect of overall propeller degradation may be due to either fouling (marine growth on the propeller surface such as acorn barnacles or tubeworms) or surface deterioration, the most common cause being corrosion, with impingement attack and cavitation erosion also being influential (M. Atlar 2001). Factors which exacerbate fouling are those linked to sea water temperature, sea salinity, current, ship speed, ship hull roughness and surface coating, ship age, days out of dry dock, maintenance schedule and time spent in port. These factors are the drivers of ship performance and relevant to an alternative approach to hull fouling prediction which is not the focus of this thesis which is determining performance from operational data. This brief description is relevant only in understanding the complexity of the problem and that factors exist which have the potential to cause short term fluctuations in ship performance. These may remain unexplained in the absence of additional data.

The generalised roughness-induced power penalty in ship operation was studied in detail by Townsin (1985). He formulated a linear expression for this by differentiating the log of each side of the expression for delivered shaft power (from Equation 3,Equation 4 and Equation 7), resulting in:

$$\frac{dP_S}{P_S} = \frac{\cancel{d\rho}}{\cancel{\rho}} + \frac{\cancel{dWSA}}{\cancel{WSA}} + \frac{dC_T}{C_T} + \frac{3dV}{dV} + \frac{dK_Q}{K_Q} - \frac{dJ}{J} - \frac{dK_T}{K_T} - \frac{d\eta_H}{\eta_H} - \frac{\cancel{d\eta_r}}{\cancel{\eta_r}}$$

Equation 25

The terms representing changes in WSA and density are small and considered negligible, the same is true for the change in η_r which is justified by the evidence from model tests.

Townsin (1985) shows that Equation 25 may be decoupled into the sum of hull and propeller power penalties

$$\frac{dP_S}{P_S} = \left(\frac{dP_S}{P_S}\right)_H + \left(\frac{dP_S}{P_S}\right)_P$$

Equation 26

The predominant effect on the hull power penalty due to hull roughness $(dP_D/P_D)_H$ is through dC_T (due wholly to the increase in C_F), dJ and $d\eta_H$:

$$\left(\frac{dP_S}{P_S}\right)_H = \left(\frac{dC_T}{C_T}\right)_H - \left(\frac{dJ}{J}\right)_H - \left(\frac{d\eta_H}{\eta_H}\right)_H$$

Equation 27

The changes in K_T and K_Q due to hull roughness cancel out if it is assumed that the propeller is operating in a region where K_T/K_Q is relatively constant, hence the direct effect of the hull on open water efficiency is dictated by dJ only which arises from a change in the propeller operating point through the wake fraction.

The impact on the propeller power penalty due to propeller blade surface roughness $(dP_S/P_S)_P$ is due to the factors adversely affecting the propeller open water efficiency only; K_T, K_Q and J:

$$\left(\frac{dP_S}{P_S}\right)_P = \left(\frac{dK_Q}{K_Q}\right)_P - \left(\frac{dJ}{J}\right)_P - \left(\frac{dK_T}{K_T}\right)_P$$

Equation 28

The relative change in J and K_T are inversely proportional to the relative change in power delivered and the relative change in K_Q is directly proportional. Propeller open water tests were conducted by (M. Atlar 2001) who found that when the propeller blade drag coefficient was increased up to 43.9% the predominant effect is an increase in the propeller torque, the decrease in propeller thrust that accompanies the

increased torque is small. The overall propeller efficiency loss for the 43.9% increase is in blade drag coefficient was 6%.

To summarise, the hull power penalty due to an increase in hull roughness is primarily due to the changes in hull efficiency, the advance coefficient and the total coefficient of resistance. The propeller power penalty due to propeller blade surface roughness is due to changes in K_T, K_Q and J, these are the factors effecting propeller efficiency.

Fuel Consumption

The effect of time on fuel consumption is to increase the engine's SFOC, the reasons for this are twofold:

1. Operating in the engine off design condition due to time dependent factors, i.e. fouled hull/propeller deterioration causing a heavier running propeller and lower ship speed for given power
2. Engine wear, fouling (soot build up etc.) general deterioration in condition, this is significantly affected by engine maintenance procedures and frequency.

A heavier running propeller due to fouling and/or heavier weather will shift the propeller curve on the engine layout diagram (engine speed vs power) to the left. This will cause a change in the SFOC; the exact value is also dependent on the propeller rpm. Therefore the change in SFOC with time is non-linear and difficult to predict because it is itself dependent on the hull and propeller fouling through the thrust and torque coefficients of the propeller which determine the shaft rpm.

5.5 Theoretical Model Summary

The interactions between time and engine, hull and propeller performance are important in the analysis of the calculated performance indicator results. This is because the effect of time is amalgamated in the delta power (modelled power – measured power). Therefore, in the first instance, if the model is unable to effectively capture the interactions in the 'at new' condition then the consequent increase in the

residual (the PI) will be incorrectly attributed to changes in ship performance due to time. This will increase the uncertainty in the PI. In the best case, the effect of this will be to increase scatter around the mean of the PI and not to affect the PI value itself. However, the interactions between factors influential to the ships propulsion system (hull efficiency, SFOC, C_T, etc.) are non-linear, as described in sections 5.1 to 5.3. Therefore, if the model is inadequate then they will manifest not only as an increase in scatter in the performance indicator but also create bias in the PI measure itself.

Secondly, it is important to note that because of the non-linear effects of time highlighted in section 5.4 then there will always be some scatter present in the performance indictor due to changes in the ships operating condition. The aim of the model is to minimise this as much as possible by ensuring it adequately reflects the ships response assuming no influence of time.

5.6 Results

The model is firstly compared to published sea trial data to build confidence in the 'at design' component of the model. Metrics are then identified with which the model performance is evaluated and against which this model can be compared to the alternative models presented in chapters

5.6.1 Sea Trial Comparison

The 'at design' component of the theoretical model was compared to published sea trial data. Published sea trial data with sufficient detail regarding the ship's technical characteristics such as hull geometries and the speed – power results themselves is difficult to find in the literature. This comparison was taken from the paper "Service Performance and Sea Keeping Trials on M.V. Jordaens" (Aertssen 1966). It is notable that M. V. Jordaens is a container ship with length (BP), beam and design draught of 146.15m, 20.10m and 8.84m, respectively. This is a smaller sized and different type of ship to the two case study ships used in the analysis of this thesis (see Table 5: Ship particulars). The sea trial data is compared to the results of the

resistance model based on the H&M resistance model and the propeller power model based on the Wageningen B-series propeller characteristics.

The model inputs were:

Design speed = 16.5knots
Engine design speed = 119rpm
Length between perpendiculars = 146.15m
Design draught = 8.84m
Main engine power =9000*0.745700 kW
Beam = 20.1m

The results for the comparison of model thrust and sea trial thrust and the same comparison for power are plotted in Figure 26 and Figure 27, respectively.

There is a good agreement between the model and the sea trial data for both thrust and power, in both cases there is a small under estimation by the model.

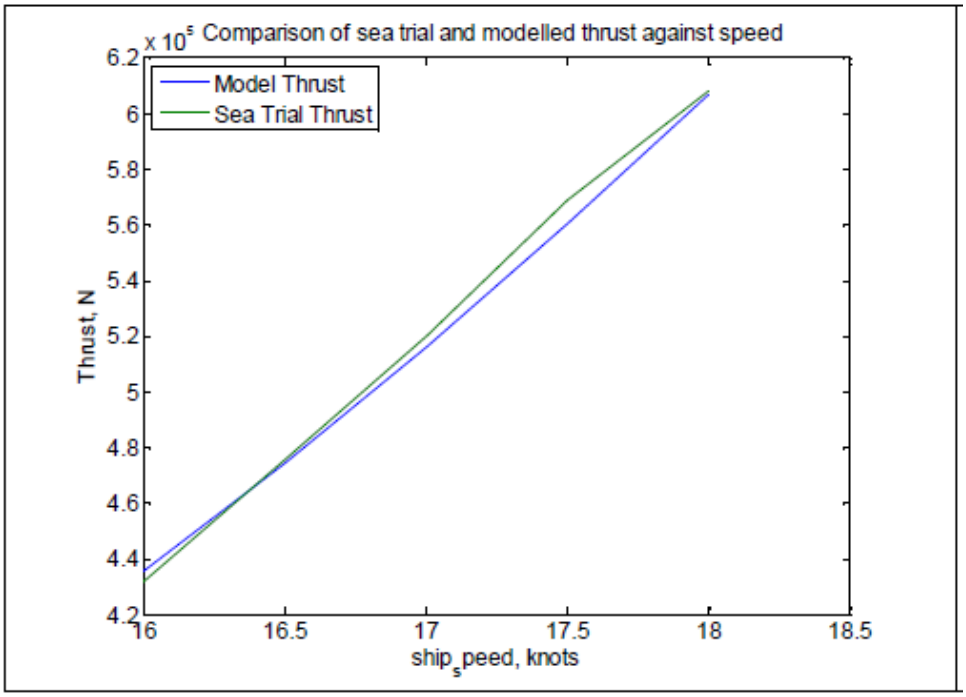


Figure 26: Thrust comparison

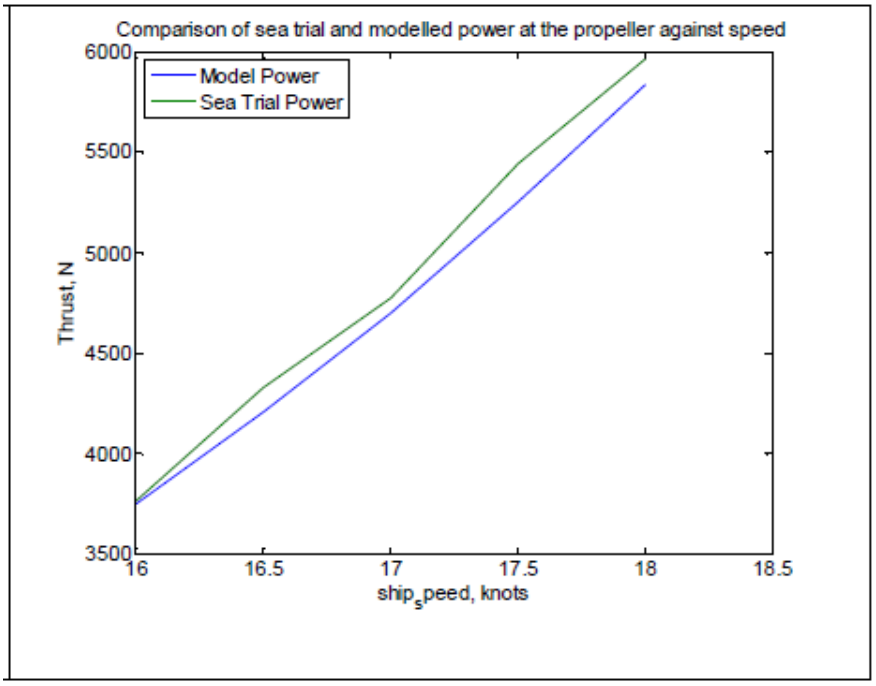


Figure 27: Power comparison

5.6.2 Model Performance Evaluation Results

The models are evaluated in terms of proximity of the modelled answer to that of the measured answer. The evaluation is disaggregated in to three steps;

i. Performance in the shaft power estimation given the operating and environmental variables
ii. Performance in the fuel consumption estimation given the operating and environmental variables and finally
iii. Performance in the fuel consumption estimation given the shaft power.

The third step is used in order to disaggregate the fuel consumption model performance from the shaft power model performance. The conditions for the choice of time period for the model evaluation period are:

- Stationarity
- Include a range within each operating and environmental variable

The stationarity is necessary because there will be some time trend that will cause deviations of the model from the measurement due to hull/propeller/engine

degradation and not because the model is under performing. A 3 month time period is often cited.

Evaluation of model goodness of fitness methods are discussed in E. P. Smith and Rose (1993) and Piñeiro, Perelman et al. (2008). Three possibilities described therein are based on the following methods:

i. Assess linear regression of observed versus predicted values and carry out t-test to determine if the y-intercept coefficient significantly differs from 0 and if the slope differs from 1
    a. H_0: There is consistency (the model adequately captures power across the input range), slope = 1
    b. H_0: There is no bias, intercept = 0
ii. Sum of squared prediction methods, the adjusted R^2 value
iii. Assess Root mean squared deviation (RMSD) which is the square root of the mean squared deviations of the predicted values against the observed: (Piñeiro, Perelman et al. 2008)
    a. $RMSD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$

Equation 29

These metrics are straightforward to compute and intuitive to interpret and describe, the scatter plot of observed vs predicted is still the most frequently used approach for evaluating models (Piñeiro, Perelman et al. 2008). More elaborate methods, such as permutation tests are unlikely to result in a different conclusion, although may be advantageous depending on the quantities being compared, for example scalar values, such as means, rather than a vector representing a group of scalar values (Rose and Smith 1998). Further comparisons exist such as the difference between the modelled and observed means or standard deviations, or using the mean squared deviation and its components (i.e. mean square for non-unity or lack of correlation), however these are regarded as complements to regression parameters, rather than replacements for them (Gauch 2003). A simple and insightfully more detailed evaluation is to partition the variance of the observed values not explained by the predicted values into Theil's partial inequality coefficients; however this level of

detail is not required here in the assessment of an overall goodness of fit for the purpose of comparing two models.

Each model is evaluated from day 34 to day 120 (time period = 86days), a 90 day time period is short enough to be deemed sufficiently stationary in terms of the time effects but long enough to collect a sufficient sample size to carry out meaningful analysis. In each assessment the observed is plotted in the y-axis and the predicted in the x-axis in accordance with the findings of Piñeiro, Perelman et al. (2008), then the slope describes the consistency and the intercept highlights any model bias. The r^2 value indicates the proportion of the total variance explained by the regression model and how much of the linear variation in the observed is explained by the variation in the predicted.

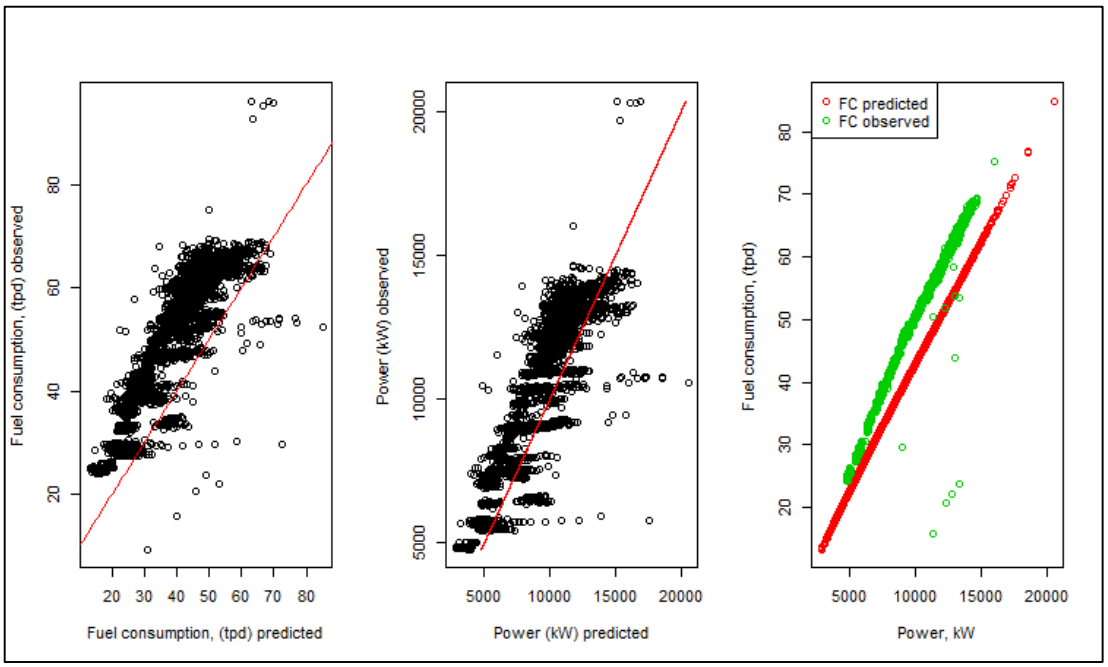


Figure 28: Theoretical model evaluation; observed and predicted power and fuel consumption

As indicated by Table 14 the linear variation in the observed values is reasonably well explained by the variation in the predicted values (adjusted r^2), indicating that the trends and interactions between variables are reasonably well represented by the model. However, the model tends to underestimate the values of power and fuel consumption. This is also indicated by Figure 28; in the power to fuel consumption conversion (fig. 25c) the theoretical model reflects appropriately the linear relationship between the two variables although the engine's performance is

optimistic, and increasingly so at higher engine powers. This might be because the factor of 1.05 that was used to account for deviation between test bed and actual operational data as described in section 5.3 is not appropriate. Or it might be because of the fuel grade and quality, engine wear or age related degradation or other possible omitted variables that may influence the engine SFOC characteristics.

	Fuel Consumption, tpd	**Power, kW**
Adjusted R^2	0.809	0.815
Intercept	10.852	2104.259
Intercept SE	0.319	64.294
Significance of test intercept = 0 (p-value)	0.000	0.000
Coefficient	1.017	0.897
Coefficient SE	0.008	0.007
Significance of test slope = 1 (a) (p-value)	0.036	0.000
RMSD	12.804	1696.424
Nr of observations	4106	4106

Table 14: Theoretical model evaluation summary

The RMSD is the mean deviation of predicted values with respect to the observed ones, in the same unit as the variable under evaluation. The lower the RMSD the closer the model is to the observed variables. These can be directly compared to the results of the statistical and hybrid models. The average deviation of the shaft power model as predicted from the operating and environmental conditions (fig. 25b) is around 1700kW, which demonstrates a reasonable degree of accuracy. Although it is seen that the model does not capture the shaft power response fully as concluded by the rejection of the null hypothesis that the coefficient estimate of the predicted power in the linear model of observed power equals 1. The intercept of the same model was at 2104 kW and null hypothesis that the intercept equals zero was rejected therefore it is possible that there is a degree of bias in the model. This is perhaps due to omitted variables or incorrect model assumptions, although this could also be for

other reasons such as measurement error in the input data. These results are carried through to the fuel consumption model (fig. 17a); the bias is exacerbated by the aforementioned over estimation of the engine performance.

In summary, the null hypotheses that the slope equals unity and the intercept equals zero are rejected in both cases indicating that the theoretical model alone is not sufficient to accurately define and track ship performance.

5.7 Chapter Summary

This chapter has described the underlying theory of ship performance and it is apparent that the ship's systems and their operational arrangements combined with the environmental conditions they are subject to are inextricably linked in a non-linear physical manner and are also subject to the effects of time; consequently predictions of required shaft power and fuel consumption are ambiguous. The theoretical model has attempted to decipher these relationships however there still remains to be limitations due to the following affects that are omitted due to;

I. Data Availability (The effect is known to occur, models exist or can be derived but are excluded due to a lack of data):
   a. Ship specification constants; for example, the presence of (and geometrical details of) a bulbous bow will have a significant effect depending on its surface area and shape which particularly affects the wave making resistance according to trim, speed, draught and fouling however this geometrical detail (and others such as the presence of, size and type of appendages) are generally unavailable in the CWFR. These are constant inputs but have time-varying effects on the shaft power because of their interactions with draught, speed, environment, etc.
   b. Measurements of acceleration, steering and drift, as described in the previous chapter, while it is attempted to filter out the effects of vessel acceleration and changes to rudder angle during manoeuvring by including a minimum threshold for ship speed and water depth, there could remain significant periods where these are influential but not included in the theoretical model. These have direct time varying effects on power.

II. Incomplete Knowledge (The effect is known to occur but accurate quantification is unknown despite data availability):
    a. Variable direct effects and their interactions, also known as the identification problem in econometrics: The exact quantifications of the physical relationships between all parameters are yet to be fully understood and therefore reliable theoretical methods are incomplete. For example, the empirical formulae of the theory section are extracted from model tests which, as well as the inherent uncertainty associated with the test procedure and model correlation coefficients, the model type and shape is also unlikely to match the ship type and shape to which they are extrapolated to, therefore exact quantification of parameters are ill defined.

III. Other unknown omitted variables (the effect is unknown to occur therefore models do not exist and have never been derived); these are sometimes described as the "unknown unknowns" in uncertainty related literature.

This is summarised in the following table:

Type	Knowledge of effect?	Quantification of effect possible? (by model derivation)	Data available?
I	Yes	Yes	No
II	Yes	No	Yes
III	No	No	Unknown

Table 15: Summary of limitations of the theoretical model

A regression model, in combination with knowledge about the physics influencing fuel consumption, aims to capture the physical interactions between parameters. Through statistical control of measured variables the model should describe as much of the variation in the data as possible and therefore enable conclusions about influences and sensitivities of ship performance to the different input parameters to be drawn.

So, although there is underlying theory to ship performance (in which only two examples of the many system interactions that exist are briefly highlighted above), the complexity of the interactions and the low level of detail with which the

independent variables (e.g. environmental conditions) are recorded, means that statistical techniques can provide valuable insights into how changes in input variables independently and collectively effect performance. Furthermore, when applied to a dataset which features both measurement and aleatory uncertainty and a large enough sample, statistical analysis can produce a model which if close enough to the 'true behaviour' of the system can form a baseline against which the magnitude of these uncertainties can be assessed.

In the case of I, a statistical model is advantageous in that even though the time varying theoretically unrepresented factors, such as acceleration and rudder angle changes, are not included explicitly, as long as they don't skew the data, i.e. their effect is normally distributed about the mean, then their average effect is implicitly included in the determination of the coefficient of each represented explanatory variable by the least squares method which endeavours to find the mean effect. The same is also true of the effect of missing fixed design characteristics (bulbous bow, transom stern and appendages) because their effect is trim, ship speed and draught dependent, then again, their effect on shaft power is implicitly included in the coefficient estimation at each level of the varying operational conditions. In order for the statistical model to succeed the following must be true;

1. The functional form of relationships between parameters and their effects on the dependent variable are correctly specified in the underlying statistical model
2. There is sufficient data (collected over a long enough time period) such that the error caused by the omission of time varying factors follow a normal distribution with a mean of zero, $N(0,\sigma^2)$

In the case of type II limitations, although the exact quantification is inaccurate, because the relationship is known to exist then it is possible to use theory (or model test results) to define the correct functional form and then to refine the magnitudes of the parameters using data. Again, it is a prerequisite that the proposed functional form of parameter relationships be correct.

In the case of type III limitations, the adverse consequence of omitted variables and their interactions are difficult to counteract. They manifest as a bias which may skew the results and mean that the advantages of a statistical model being able to counter the effects of type I and type II limitations are negated. Bias is tested for by analysing the residuals, see section 6.2.2.

It is proposed that using historical data and statistical techniques, it may be possible to gain further insight into how changes in input variables, both independently and collectively through interactions effect ship performance. This then forms the normalisation model that enables the actual operating/environmental condition to be reverted back into a reference condition from which the difference between modelled and measured fuel consumption equates to changes in the ship's performance. This is the focus of the following chapter.

Chapter 6. Statistical Model

The focus of this chapter is to form and evaluate a statistical normalisation model; it presents a robust method of determining variable significance and to assess model performance both in terms of optimum explanatory variable combination and functional form. The model is then compared to the theoretical model using the same three month evaluation dataset and the same procedure described in section 5.6.2. A training dataset, also of three months, is used to build the model, over this time period the variables must be stationary (see section 6.1.2) and exhibit a reasonable degree of variance in order to properly capture the variable interactions.

To avoid over fitting a model with variables that are spurious but may happen to enhance significance or raise R^2 then it is important to have tests of robustness, to use prior theory and evidence, to connect theory by a deductive argument on choices of variables and to test for multicollinearties that would result in biased estimates. To ensure that the model reflects as close as possible the 'true behaviour' of the system the following method is applied to the continuous monitoring data set:

1. Explanatory variable identification
    a. Identify all key parameters from theoretical assertions
    b. Examine the statistical evidence of the relationships between variables in conjunction with the theoretical assumptions
    c. Confirm all variables are stationary
    d. Test for multicollinearity between independent variables
    e. Determine which predictor variables to include and which redundant variables to omit from the model
2. Prediction method selection: Compare OLS to alternative methods
3. Model form selection
4. Confirm the regression assumptions are not violated by assessing the residuals for independence, bias and normality

6.1 Explanatory Variable Identification

Excessive parameters in the statistical model will cause the model to be over parameterised and to inaccurately estimate the effects and therefore dependent variable predictions. This section is concerned with refining the variables by identifying and omitting redundant variables and assessing and retaining effective variables. The initial identification of key parameters is limited by availability and with reference to ship theory in conjunction with comparative analysis of statistical evidence of relationships between variables, the statistical analysis also minimises the possibility of bias least square estimates caused by non-linearities between the dependent and independent variables.

6.1.1 Identify key time-varying parameters

After filtering and consideration of data availability limitations, the previous chapter has highlighted the key time varying parameters as follows (with reference to Table 8):

- V, ship speed
- T, draught
- Trim
- H_s, significant wave height and direction
- $V_{wave,app}$ apparent wave speed and direction
- Swell height and direction
- $V_{wind,app}$ Apparent wind speed and direction
- Shaft power (dependent variable)
- Fuel consumption (dependent variable)

6.1.2 Statistical Evidence of Influential Factors

An assumption of multiple regression is that the dependent variable is linearly related to the explanatory variables in order for the least squares method to yield unbiased estimates of the coefficients. A straightforward, initial identification of non-linearity is conducted by visual inspection of scatter plots which present the dependent

variable against each explanatory variable. Of course, this kind of analysis is limited in that the quantification of shaft power is a multidimensional problem and graphically, it is only possible to represent 2 or 3 dimensions simultaneously (which is why multiple regression analysis is required). However, the proceeding analysis aims to slice the data so as to disaggregate relationships into fewer dimensions and, in combination with ship theory, to identify broad trends and to observe and test the appropriate transformations for linearization of the relationships.

Fuel Consumption

Figure 29 shows the relationship between fuel consumption and shaft power (measured from torque and RPM sensors), after filtering (see section 4.2), for a Suezmax tanker for 303 days of data (a) and a VLCC for 387 days of data (b), as can be seen, the variation surrounding the fuel consumption at each power level is reasonably small (~ +/- 3tpd) and is largely due to the changing operational and environmental conditions (including air temperature) that cause the SFOC to fluctuate around the optimum SFOC on the design curve (see section 5.3).

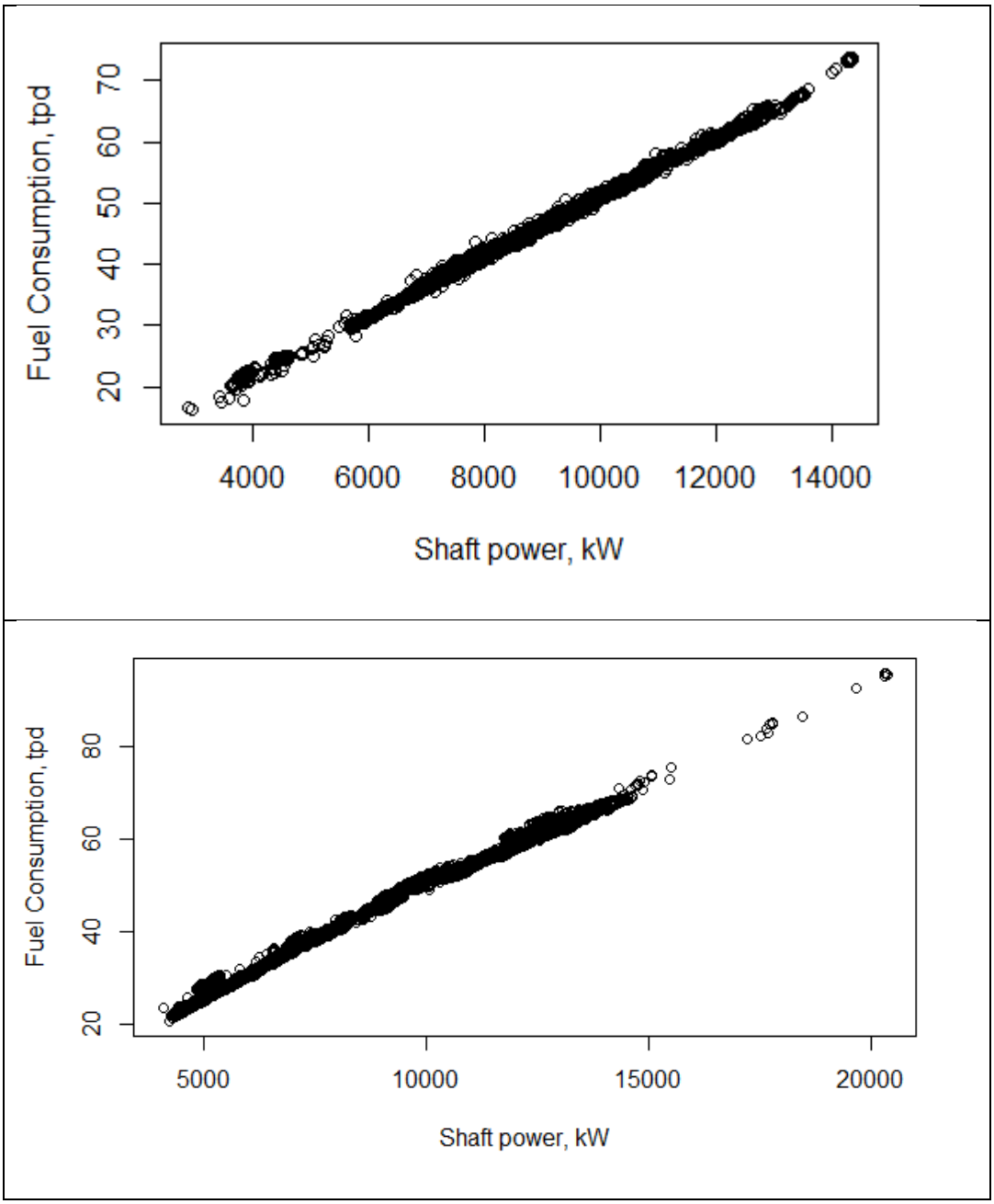


Figure 29: Fuel consumption - shaft power relationship. Top: (a) Suezmax tanker. Bottom: (b) VLCC

Plotting the relationship during the training period demonstrates the clear relationship between these variables in Figure 30. A linear relationship yields the optimum fit in terms of R^2 in the calibration period as demonstrated by the matrix in Table 16.

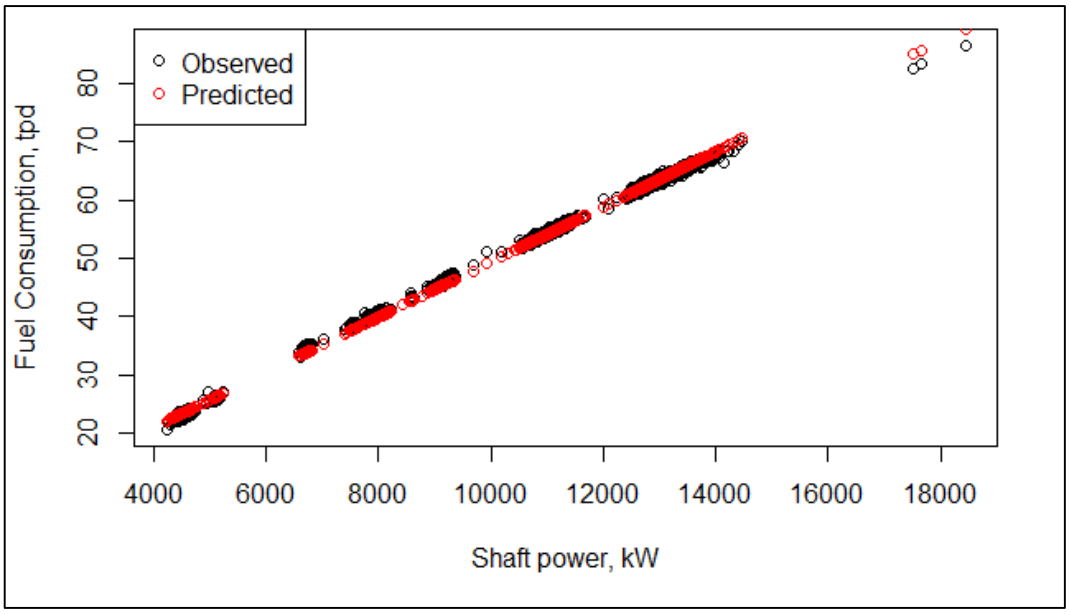


Figure 30: Fuel consumption power relationship during the training period

Adjusted R^2	Log power	Power
Log fuel consumption	0.990	0.967
Fuel consumption	0.977	0.991

Table 16: Comparison of log-log and linear model

Therefore a linear model regressing fuel consumption on shaft power is proposed:

$$FC = \alpha_1 + \beta_1 P_S + \epsilon$$

Equation 30

The remainder of this section focusses on formulating a regression model of the correct functional form with shaft power as the dependent variable and the optimum combination of the other key parameters as the explanatory variables.

Operational Factors: Speed and draught and trim

The theoretical analysis suggests that the speed power relationship is cubic as the resistance coefficient is proportional to the square of speed and effective power is the product of resistance and ship speed. However, realistically the speed exponent is likely to vary from approximately 3.2 for low speed ships (tankers and bulk carriers) to 4.0 for high speed ships like container ships. Figure 31 illustrates this by

progressively isolating the speed-power relationship for the ship in its ballast loading condition, starting with all the data represented in black (after filtering as per section 4.2) then filtering environmental variables (green) and then removing the loaded voyages (blue).

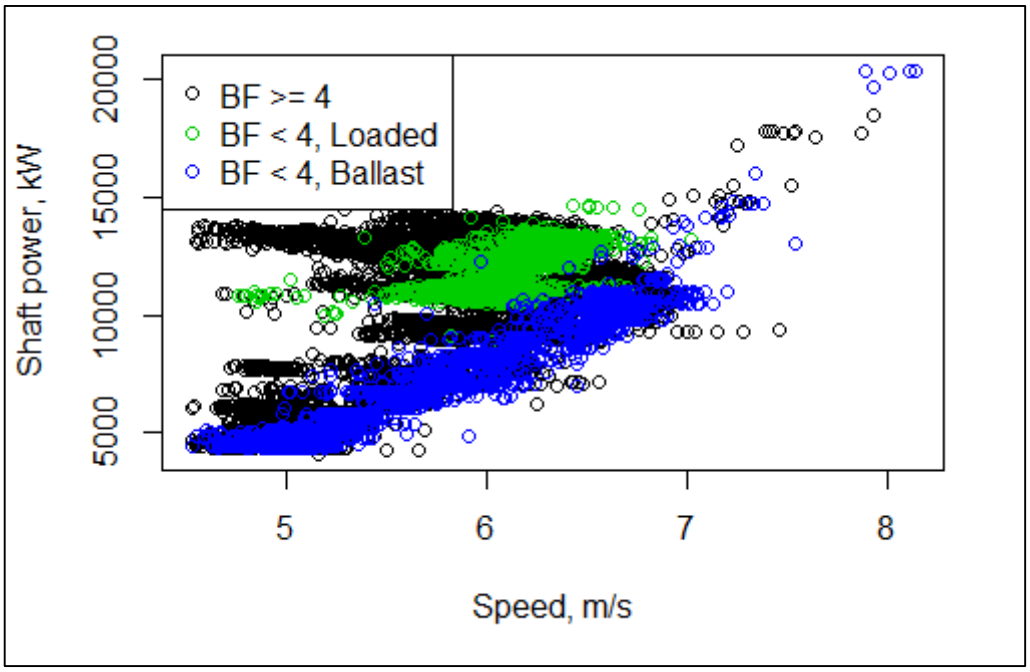


Figure 31: Shaft power – Speed relationship depending on the operational and environmental conditions

Obtaining the correct relationship in the model is vital in order to attribute changes in power to the correct cause, i.e. to avoid the possibility that the effect of time induced degradation is overshadowed by a compensatory speed reduction over time. A log-log model is therefore used to define data derived speed/power relationship.

The box and whisker plots of Figure 32 display the actual distribution of shaft power depending on the ships loading condition; on the x-axis, 1 is ballast and 2 is loaded. The median is indicated by the thick horizontal line in the plot, the box surrounding the median identifies the first and third quartiles and the 'whiskers' above and below the box show the range of the data, excluding the outliers. The points therefore indicate outlying observations which are defined as any value 1.5 times the interquartile range. The plots indicate a consistent power increase when the ship is loaded relative to its ballast state.

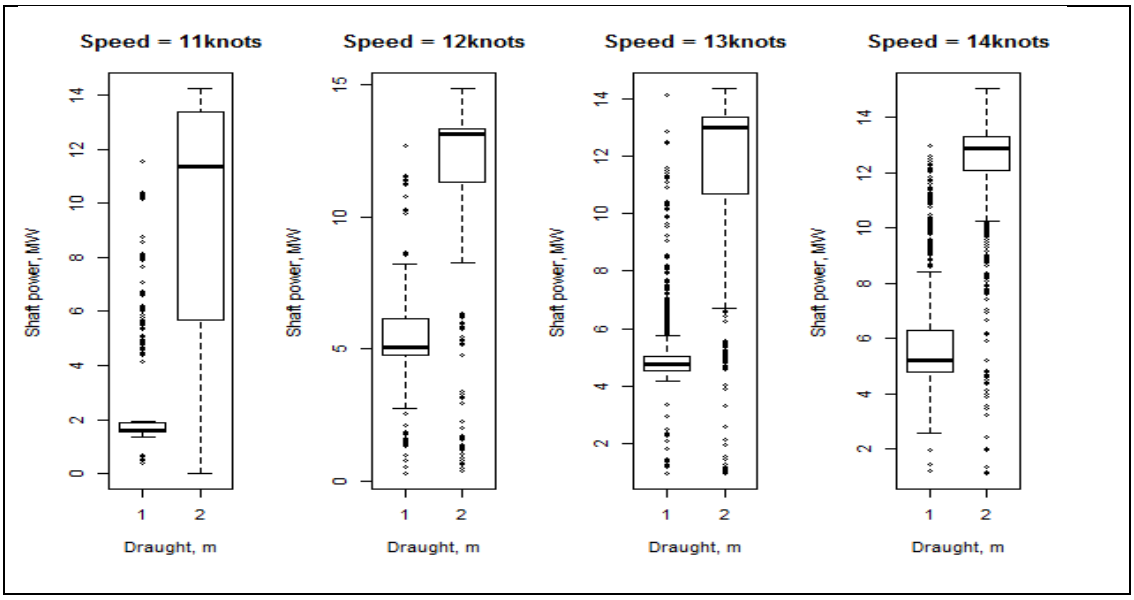


Figure 32: Speed and loading condition as a function of shaft power, draught 1: Ballast, draught 2: Loaded

Trim cannot be explicitly included because it is co-linear with draught (as discussed in section 6.1.3).

Environmental Factors 1: Wind

The theory underpinning the shaft power requirements due to wind is that the wind force depends on the area over which it flows and the drag coefficient. The force required to be overcome by the shaft power depends on the wind direction; the resistance is proportional to the addition of the true wind speed and the vessel's speed squared and the relative direction of the wind. One perspective from which to investigate the wind effect is through the effect on power of the interaction between apparent wind speed and apparent wind direction; an interaction in the regression equation reflects the fact that the effect of the wind speed on the power requirements is conditional on the direction. There is a reasonably clear bimodal distribution to the effect of apparent wind direction on power, (Figure 33) seen particularly clearly in the loaded voyage with peaks at 360^{o} and 180^{o}, for the ballast voyage the peaks are 360^{o} and 90^{o}.

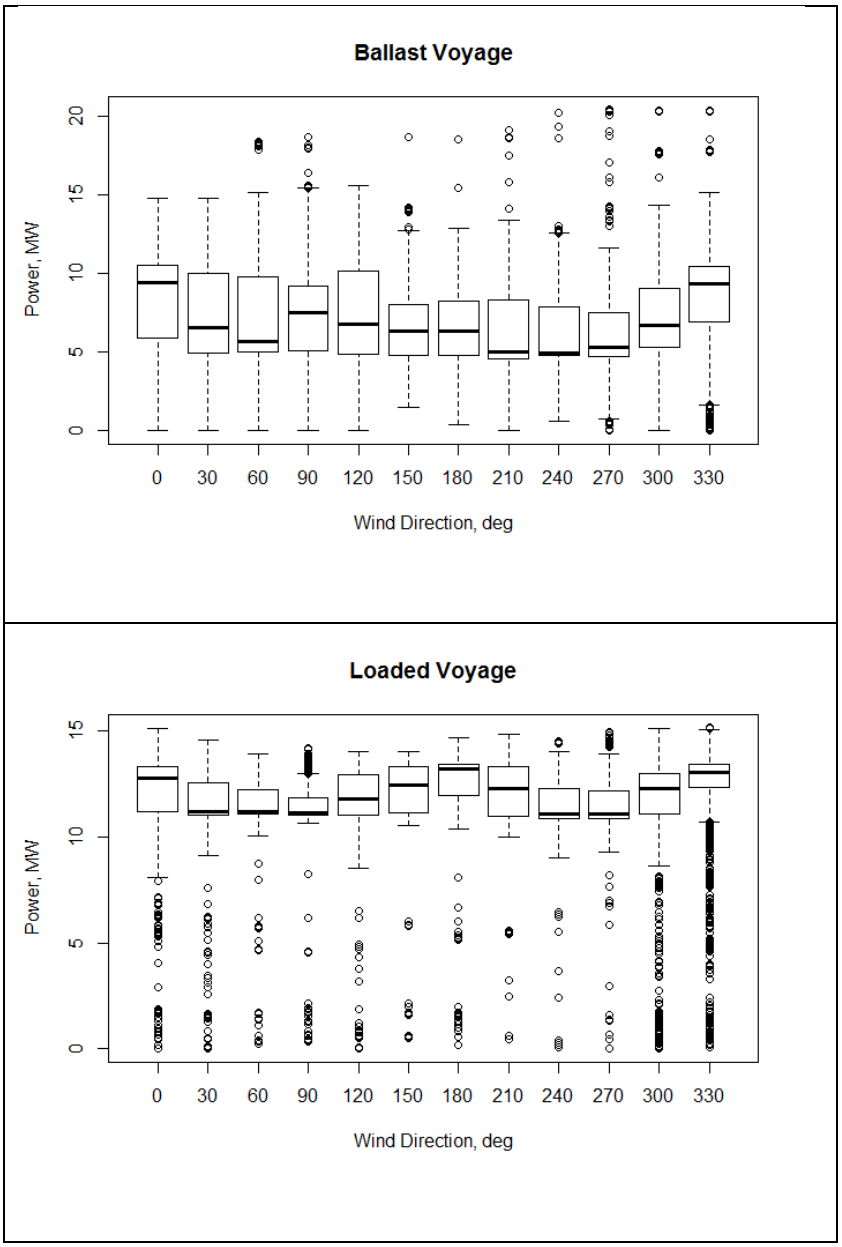


Figure 33: Apparent wind direction and power requirements

A bimodal relationship needs to be linearised for the multiple regression equation. This may be done by resolving the apparent speed vector in to a transverse, x-axis ($V_{wind,app,L}$) and a longitudinal, y axis ($V_{wind,app,T}$) component according to the apparent wind direction and assume both are proportional to the shaft power requirements using an appropriate log or polynomial function

Figure 34 and Figure 35 show longitudinal and transverse wind speed plotted against the median shaft power for each bin, both for loaded and ballast draughts, there is a possible third order polynomial relationship between shaft power and longitudinal wind speed, or a log-log relationship, more clearly observable if plotted on a smaller

scale axis, here loaded and ballast are plotted together for comparison purposes. The wind speed has greater effect when the ship is in ballast, perhaps due to a greater exposed surface area above sea level thus increasing drag (although this effect is likely to be minimal). The transverse wind speed has a more subtle effect on power.

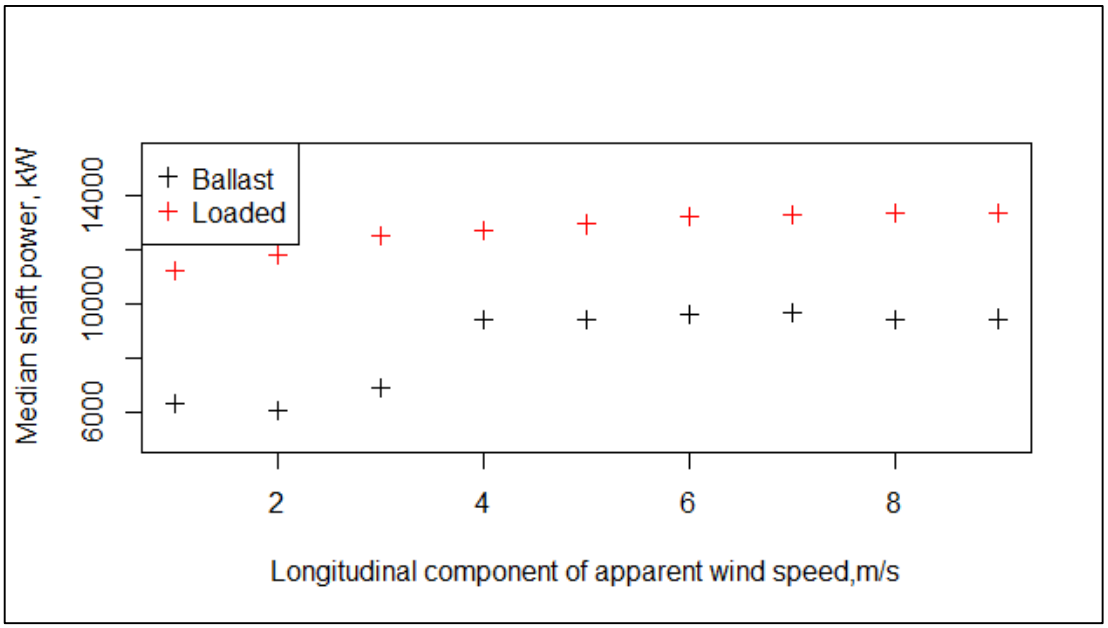


Figure 34: Longitudinal wind speed and mean shaft power

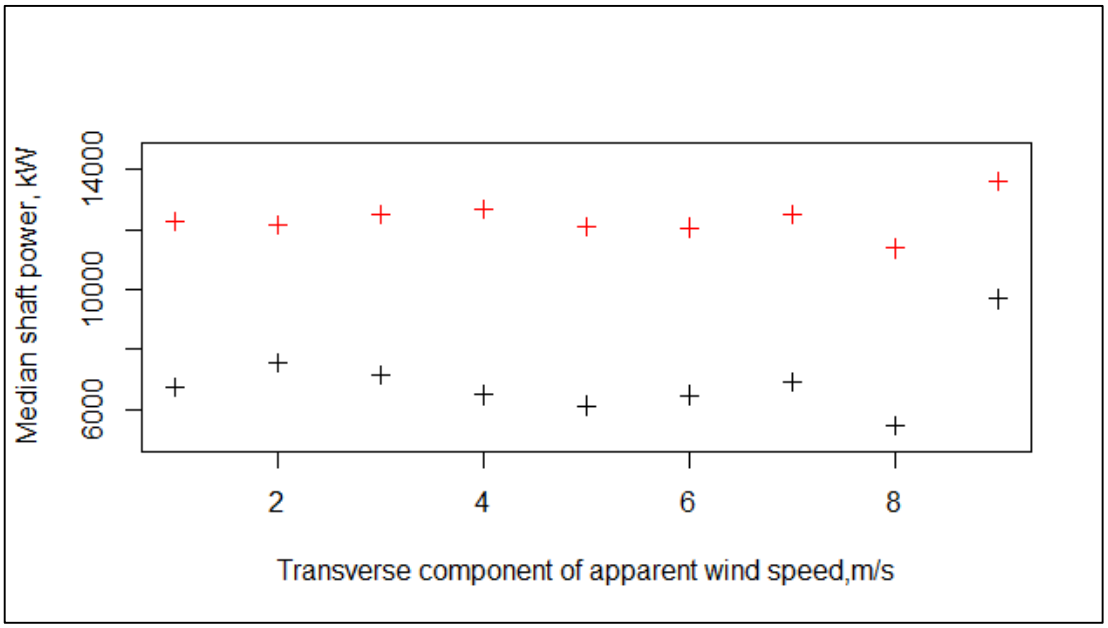


Figure 35; Transverse wind speed and shaft power

To investigate further, and for comparison with the theoretical relationships as shown in Figure 22, a further dimension of apparent wind direction is added to the graphs,

this is shown by the radar plots of Figure 36 and Figure 37. Each series corresponds to an apparent wind speed group (the group median is shown in the legend), and the radii represent shaft power, there is a separate plot for each loading condition and for the longitudinal and transverse component of apparent wind speed.

In Figure 36, in a head wind, the longitudinal component of apparent wind speed is proportional to the power and side winds have a less significant effect. The relationships between these three variables follow those plotted in the theory section in Figure 22.

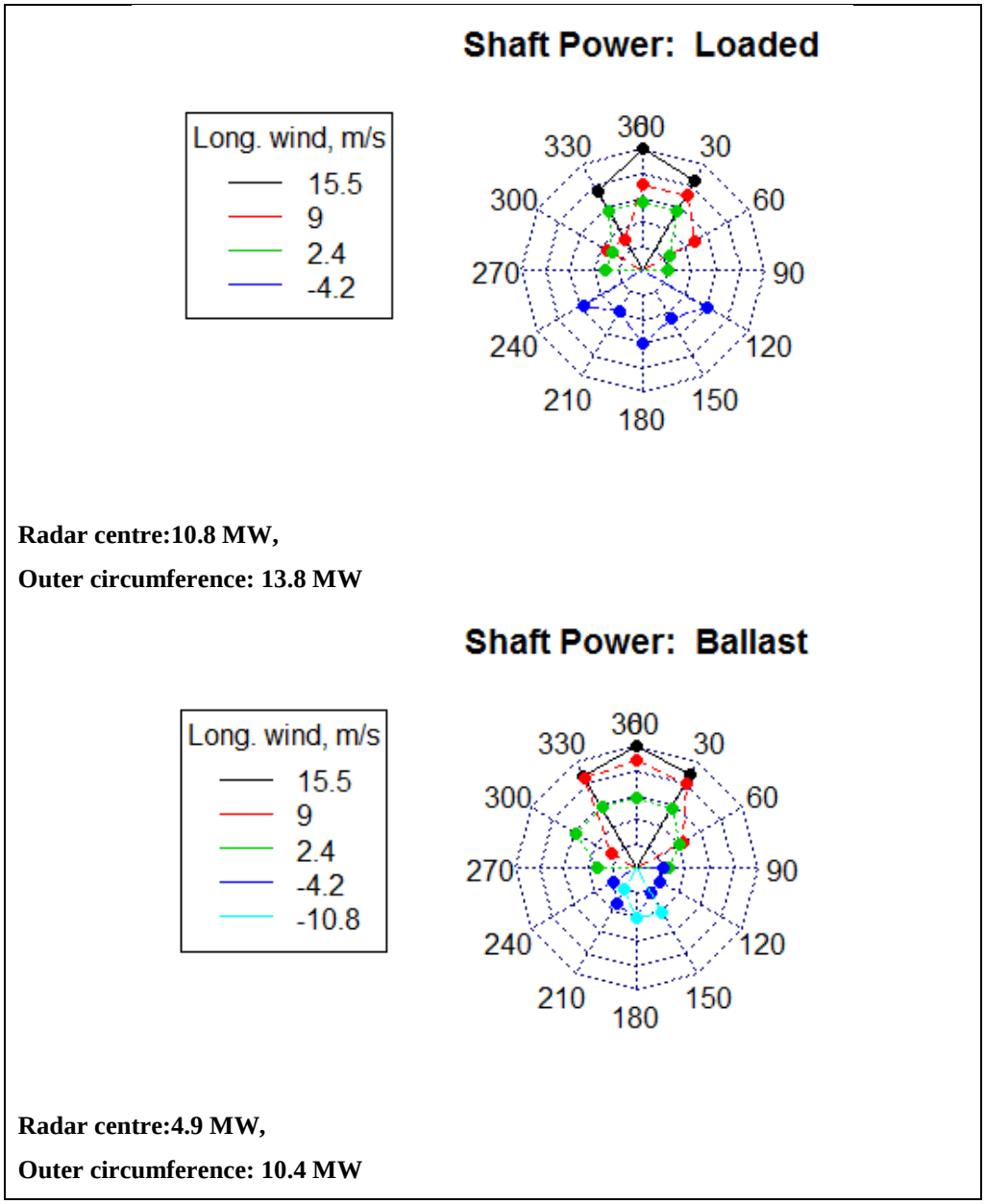
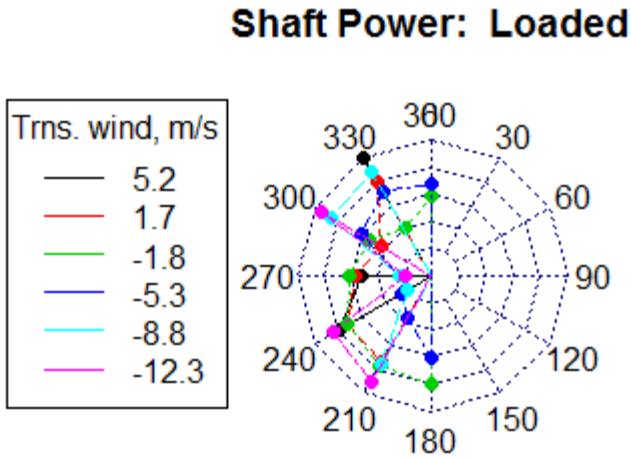


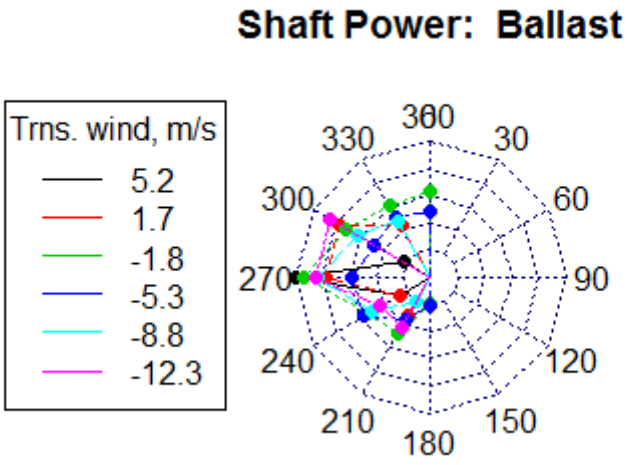
Figure 36: Longitudinal (x-axis) apparent wind speed, direction and shaft power for loaded (top) and ballast (bottom)

As previously discussed, the ships geometry and effect of the wind is assumed to be symmetrical and the wind direction is reflected along the x-axis. The plots of Figure 37 also demonstrate the relationships found in the theory section in that the transverse component of apparent wind speed has an increased effect in a side wind which then reduces as the wind direction moves towards the bow or stern of the ship, this relationship is especially pronounced in the ballast loading condition.



Radar centre:11.0 MW,

Outer circumference: 13.8 MW



Radar centre:4.5 MW,

Outer circumference: 10.9 MW

Figure 37: Transverse (y-axis) apparent wind speed, direction and shaft power for loaded (top) and ballast (bottom)

Generally, when the ship is loaded the results are more obscure with the required power appearing to peak at 300-330^{o} or 210-240^{o}, this could be due to the ships speed or the interaction of other explanatory variables and it is concluded that a 3 dimensional plot is inadequate in determining the complete picture.

Environmental Factors 2: Waves

The plot of Figure 38 demonstrates the sharp increase in power in head waves and due to wave height. The speed is more reactive to wave height in the loaded condition and therefore the relationship between wave height and power is obscured (plot not shown). The theory suggests that there is a quadratic relationship between wave height and shaft power, because of its directional nature, wave speed is also decomposed into longitudinal and transverse components. The two extreme wave heights are shown here for clarity.

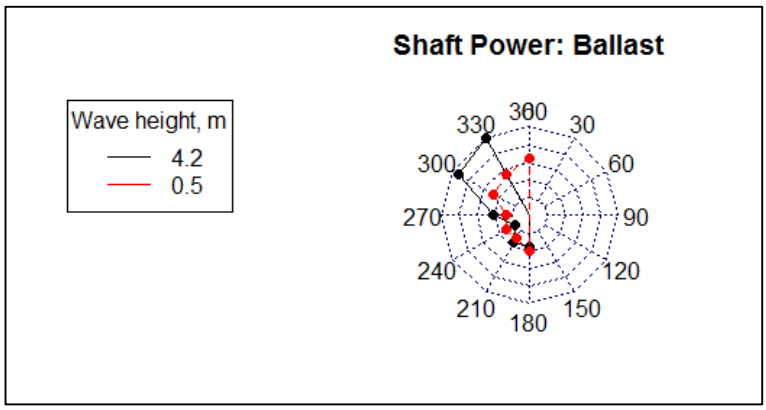


Figure 38: Wave height, shaft power and apparent wave direction in the ballast condition

6.1.3 Test multicollinearity between independent variables

The effect of collinearity is a matter of degree rather than a question of presence or absence (Paul (2008). 'Near linear' dependencies between the explanatory variables are problematic in the interpretation of the estimated parameters and their relative importance on the output variable; there may be a large standard error associated with the parameter estimates. However the adjusted R^2 value is still able to indicate the proportion of total variation in the data described by the model. The relevant adverse effect for this particular analysis is that the coefficients of collinear variables will not be correct with respect to their significance on shaft power and neither will

appear statistically significant when both are included in the model. Figure 39 shows the correlation between the factors relevant to the regression analysis using the Pearson's rank correlation coeffiient factor. The data is filtered as described in section 4.2 and the axis labeling codes are as follows:

- V_ship = ship speed
- Hs_wv = significant wave height
- Vwn_T = transverse component of apparent wind speed
- Vwn_L = longitudinal component of apparent wind speed
- T = Draught
- Vwv_L = longitudinal component of apparent wave speed
- Vwv_T = transverse component of apparent wave speed

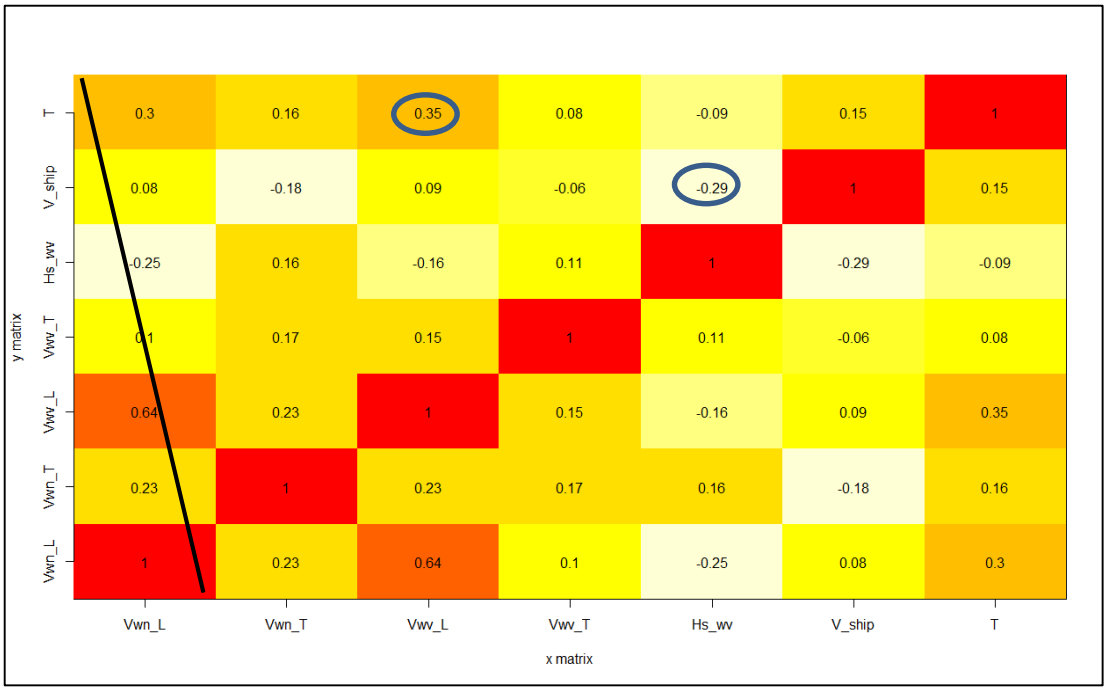


Figure 39: Correlation between the factors relevant to the regression analysis using the Pearsons rank correlation coefficient factor

The colours indicate the magnitude of the Pearsons rank correlation coefficient factor ranging from white to red with increasing factor. The longitudinal components of wind and wave speed are collinear, as one might intuitively expect. The theory of section 5.2.1.3 indicates that wave speed is of greater significance in its effect on

shaft power, therefore the wind speed (longitudinal) is excluded from the explanatory variables.

There is a degree of collinearity between draught and longitudinal wave speed and ship speed and significant wave height. This is expected to a degree and whether this collinearity is causing adverse problems on the results of the regression analysis is quantified by the variance inflation factor (VIF). This estimates how much the variance of a coefficient is "inflated" because of the linear dependence with other predictors. The threshold for presence of collinearity was VIF greater than 10.

$$VIF_j = \frac{1}{1 - {R_j}^2}$$

Equation 31

The VIF is calculated for each variable, j, by regressing it on all other predictor variables (except the jth predictor) and determining the proportion of explained variance by the model, R_j. Another sign of collinearity between variables is an increased standard error in the estimate; this is detected by examining the standard error as a percentage of the coefficient.

$$\log(P_S) = \alpha + \beta_1 \log(V) + \beta_2 T + \beta_6(V_{wind,app,T}) + \beta_7(V_{wind,app,T}^2) + \beta_8(H_s) + \beta_9(H_s^2) + \beta_{10}(V_{wave,app,L}) + \beta_{10}(V_{wave,app,L}^2) + \beta_{11}(V_{wave,app,T}) + \beta_{11}(V_{wave,app,T}^2)$$

Equation 32

6.1.4 Test for stationarity

A stationary dataset requires that the series is without trend, generally second-order stationarity is acceptable which requires the parameters have constant mean and variance over time and exhibit no periodic fluctuations or autocorrelation. Causes of non-stationarity arise from the effects of time gradually degrading the ship performance and it is important that these do not skew the results. Stationarity in each variable is tested for using the R program built in functions that implements the Kwiatkowski (KPSS) unit root test. This tests for the presence of a deterministic trend component, either a linear or constant time trend. All key parameters tested

negative for non-stationarity over the training test period when both constant and linear trends were tested for.

6.2 Model Selection and Diagnostics

6.2.1 Redundant Variables

Redundant variables were identified by a backward elimination procedure beginning with the full model of equation 22 and assessing each coefficient for significance as indicated by a low p-value, its relative δy/δx effect (REff), its variable inflation factor (VIF) which should be <10, the correct sign and magnitude of the explanatory variable's coefficient and the magnitude of its standard error with respect to the coefficient estimate (another sign of multicollinearity between variables).

$$REff = \beta_i \frac{\partial x_i}{\partial y}$$

Equation 33

The p-values which detect the statistical significance of the coefficient should not be relied on fully as reasons to remove variables, especially in large sample sizes when measurement instrument bias may falsely indicate the statistical significance of any coefficient. Examination of the estimates for correct sign and realistic magnitude (substantive for practical importance) builds confidence in the model.

Table 17 shows that both terms representing the transverse component of wave speed, and the quadratic term of the longitudinal component of wave speed were removed as they indicated inverse relationships with shaft power. Also, although not printed, the VIF's of the transverse components were quite high, as were their standard errors represented as a percentage of the coefficient estimate itself. This indicated some degree of multicollinearity present possibly with the longitudinal wave speed component as the VIF of $V_{wave,app,L}$ decreased to below the threshold of 10 when they were removed. The REff of the removed coefficients were low and, as can be seen in the table, the adjusted R^2 of the model is not greatly affected by the

removal of these explanatory variables. Red indicates the variables removed and highlights their low REff.

	Coefficients			
	Model 1	**Model 2**	**Model 3**	**Model 4**
log(Speed)	1.863	1.833	1.838	1.828
Draught	0.047	0.049	0.045	0.046
H_S	0.083	0.083	-0.018	-0.025
H_S^2	0.007	0.007	0.018	0.017
$V_{wave,app,L}$	0.012	0.012	0.006	0.005
$V_{wave,app,L}^2$	-0.001	-0.001	-1.16E-04	
$V_{wave,app,T}$	-0.005	-0.011		
$V_{wave,app,T}^2$	-2.65E-04			
$V_{wind,app,T}$	-0.013	-0.011	-0.013	-0.012
$V_{wind,app,T}^2$	0.001	0.001	0.002	0.002
Adjusted R2	0.936	0.936	0.922	0.921
Number of Observations	3162	3162	3162	3162

Table 17: Effect on model coefficients as explanatory variables of the wrong sign are removed by backwards elimination

The explanatory variables included in the final regression equation (model 4) are therefore as summarised in Table 18.

The variable inflation factors for transverse wind speed and wave height are quite high however it is acceptable to ignore multicollinearities in circumstances when the collinearity is due to the inclusion of higher order terms. The p-values of all of the remaining variables are significant at the 5% level and their relative effects are of an influential order of magnitude.

	VIF	Relative Effect	Adjusted Relative Effect	Estimate	Standard Error	P-Value	$\frac{SE}{Coeff}$ (%)
(Intercept)							
log(Speed)	1.275	0.693	0.543	1.828	0.021	0.000	1.158
Draught	2.205	0.396	0.180	0.046	0.001	0.000	1.097
H_S	9.692	-0.092	-0.009	-0.025	0.005	0.000	-22.220
H_S^2	9.643	0.354	0.037	0.017	0.001	0.000	7.165
$V_{wave,app,L}$	2.133	0.159	0.074	0.005	0.000	0.000	4.484
$V_{wind,app,T}$	10.379	-0.123	-0.012	-0.012	0.002	0.000	-19.183
$V_{wind,app,T}^2$	10.193	0.269	0.026	0.002	0.000	0.000	13.028
Adjusted R^2	0.921						
Sample size	3162						

Table 18: Estimates for the regression of shaft power on operational and environmental explanatory variables

Outliers were identified by looking at how large an effect the removal of a point has on the model coefficients, these are quantified through the Cook's distance, the maximum threshold was 1. Therefore no outliers were removed as indicated by Figure 40.

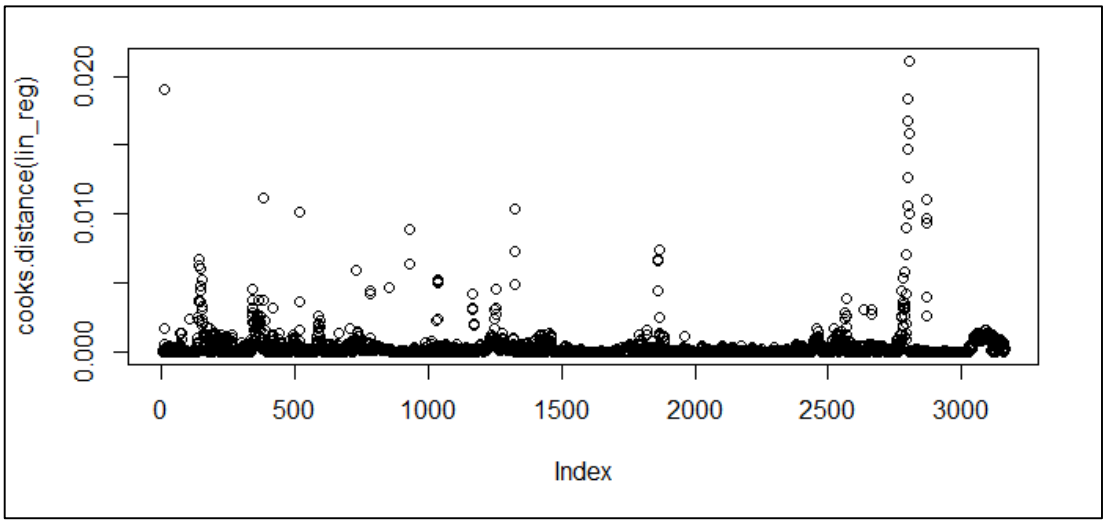


Figure 40: Cooks distance for each data point

6.2.2 Residual analysis

Residual analysis is a critical stage of regression diagnostics and ensures that the fundamental assumptions of regression analysis are not violated and therefore the predictions and confidence intervals are efficient, unbiased and true. Generally, the standardised model residuals must be independent, unbiased and normally distributed to ensure there is little or no information remaining in them. Specifically, the assumptions are:

- Linearity and additivity of the relationship between dependent and independent variables
- Statistical independence of the errors
- Homoscedasticity of the errors versus the predictions and the independent variables
- Normality of the error distribution

Each assumption is discussed with reference to the statistical model.

Linearity and Additivity

The dependent variable must be linearly related to the explanatory variables in order for the least squares method to produce unbiased estimates of the coefficients. The presence of linearity can be tested for by plotting the observed shaft power against the predicted shaft power for the training dataset. Figure 41 provides evidence of this linearity and indicates the efficiency of the least squares method, meaning that the least squares method has found the coefficient estimates with the minimum standard error and the estimates are unbiased.

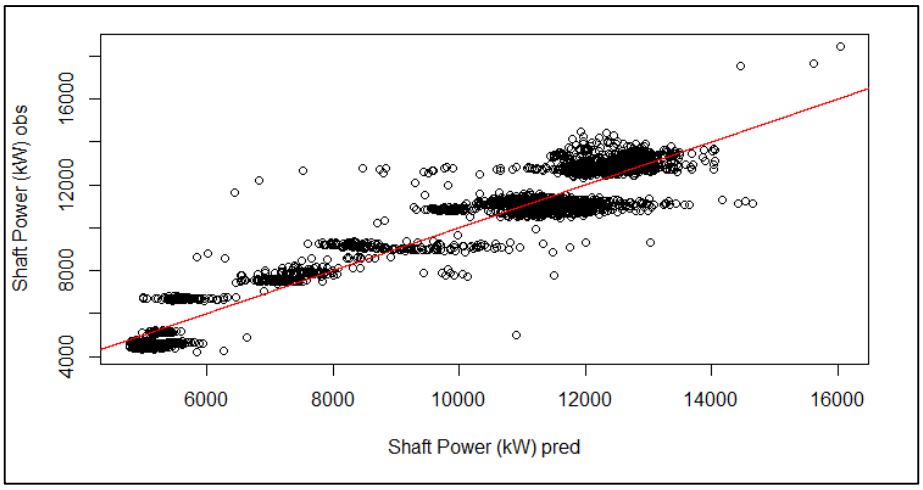


Figure 41: The relationship between the observed shaft power and that predicted by the multiple regression model during the calibration period

The initial iteration of this residual analysis showed that the model predictions exhibited a small drift at high shaft powers, this indicated that the model's functional form was miss-specified at this range, possibly due to the scarcity of data in that space. This problem was investigated by exploring the independence of the errors as detailed in the following section.

Statistical Independence of the Errors

Exploring the possible violation (at high shaft powers) of the previous assumption further, each explanatory variable is considered separately. In order to ensure that the estimates of the model coefficients are unbiased and true, the explanatory variables must not be correlated with the residuals. This is one of the most critical assumptions and can be tested visually by plotting the residuals against each explanatory variable. The residuals should be randomly and symmetrically distributed about zero under all conditions and there should be no correlation.

In doing so, it was found that for high wave heights (> 4m) the model overestimated the shaft power as indicated by a negative correlation between the residuals and the wave height (Figure 42). This indicated that there was a violation of the mean independence assumption possibly caused by model miss-specification, omitted variable bias or measurement error in the explanatory variables (Allison 1999). However, this was not the root cause of the overall model drift which was an

underestimation of the measured shaft power. The cause of this may originate from the log assumption in the speed relationship, likely due to a narrow data range in the calibration period.

The plots of all of the explanatory variables versus the residuals are shown in Figure 42. This indicates that the residuals are generally normally distributed although the scatter is still not always perfectly homoscedastic and there is some evidence of systematic patterns in some variables. The result of this is that the coefficients may be slightly biased and the resulting model precise but possibly inaccurate.

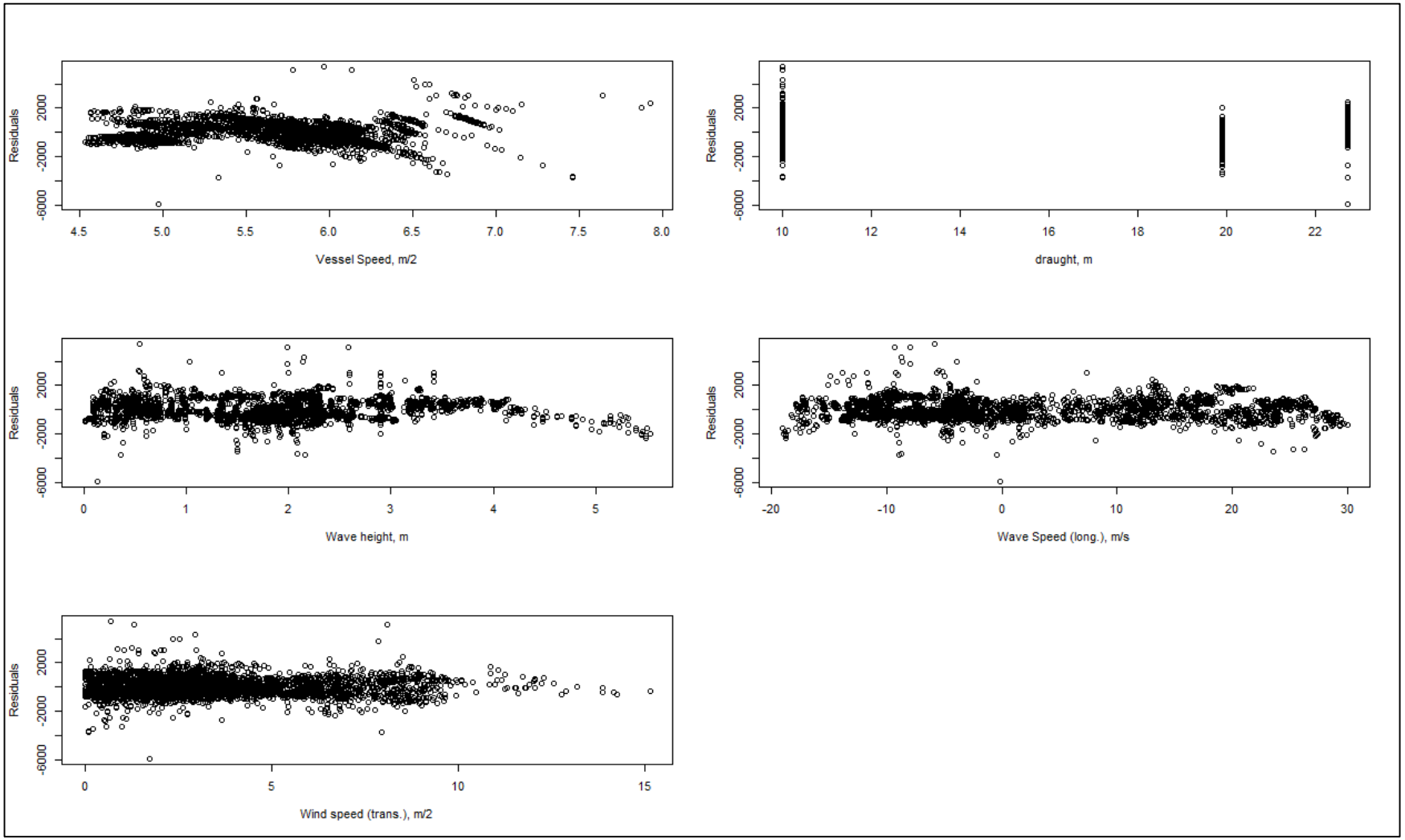


Figure 42: Plots of the explanatory variables versus the residuals

Homoscedasticity

Examination of the residuals against the predicted dependent variable should show a uniform degree of scatter in order to further confirm that the least squares is efficient. A lack of efficiency causes the least squares estimate to not find the minimum standard error when estimating the predictors and the standard errors may be biased and therefore their test statistics and p-values will be incorrect potentially leading to incorrect assessment of statistical significance of variables.

Figure 43 shows the plot of the residuals versus the fitted power for the calibration period; the data points should be randomly scattered with no particular pattern. Visual inspection indicates that, in support of the findings of the previous section (mean independence of the residuals), there is a small increase in the degree of scatter at the higher shaft power range. Evidence of heteroskedasticity may be indicative of inefficiency in the OLS estimation and bias standard errors resulting in bias test statistics and confidence intervals; a remedy to this may be the use of robust standard errors. This was tested by repeating the regression using robust standard errors however this was found to have insignificant influence. Also, heteroskedasticity does not result in biased parameter estimates and therefore its presence will not affect the overall conclusions of this chapter which relate to the predicted and observed variable comparison.

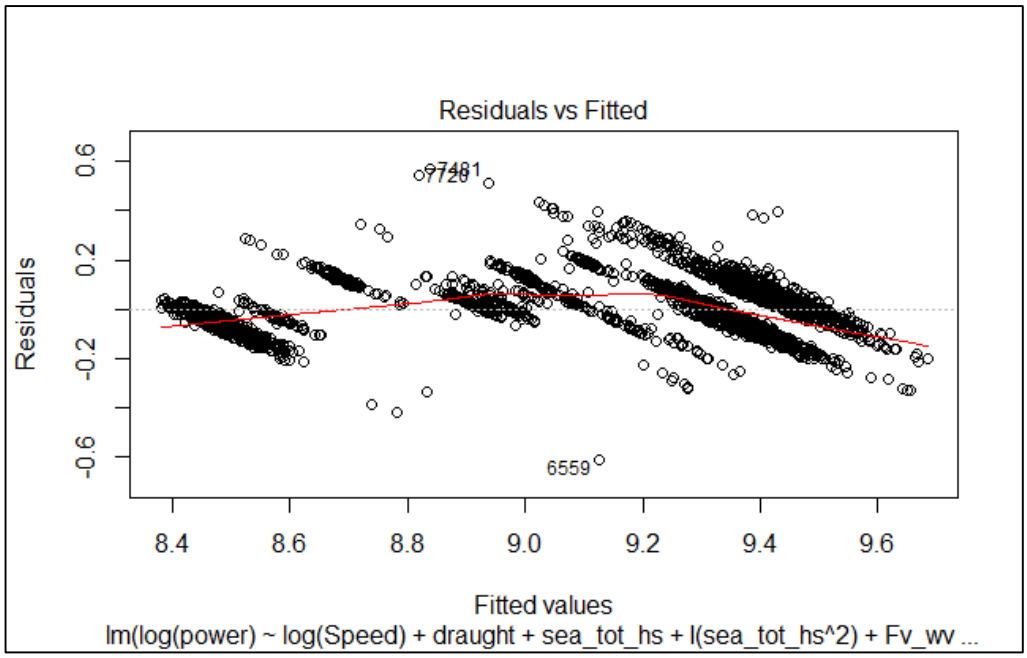


Figure 43: Residual vs fitted plot for the statistical regression

The pattern of the plot in Figure 43 also indicates that there could be a missing variable in the model. Possibly relating to propeller torque and RPM effects, as discussed in this chapter summary and referred to in the following chapter in the development of a hybrid model.

Normality

The residuals are tested for normality by the Jarque-Bera test of the null hypothesis that the residuals come from a normal distribution with unknown mean and variance. The test results reject the null hypothesis and there is a skew present, as described previously. The mean of the residuals was found to be 1.74×10^{-14}. The degree to which this assumption is violated however is small, as indicated in Normal Q-Q plot of Figure 44.

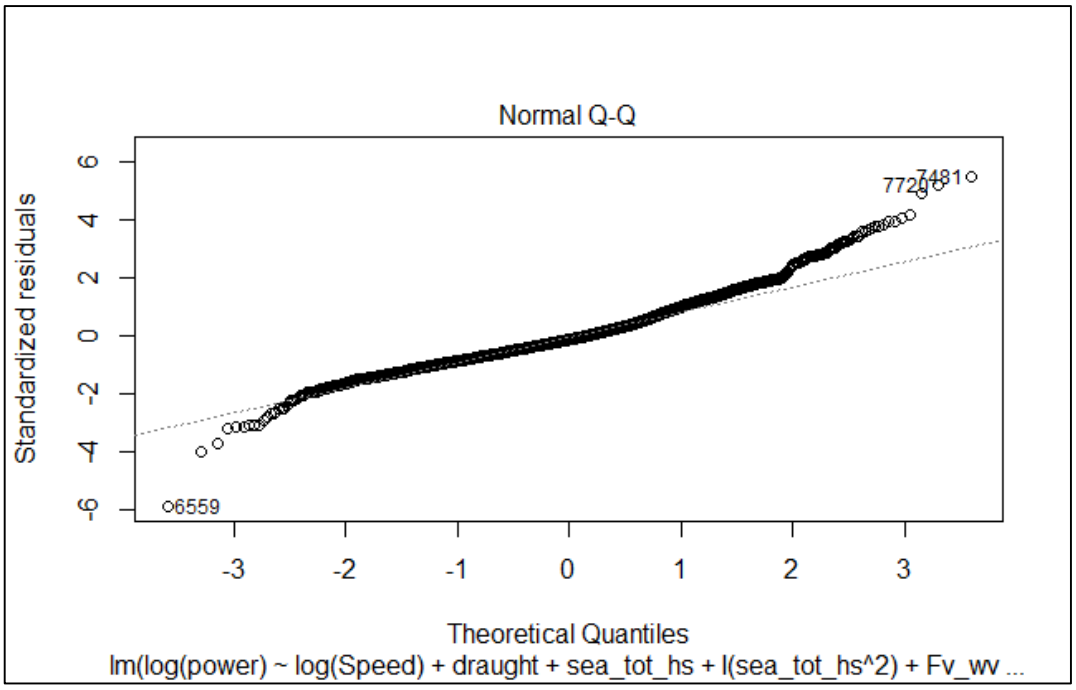


Figure 44: Normal Q-Q plot of theoretical quantiles vs standardised residuals

6.3 Results and discussion

The residual analysis has highlighted that there is a possibility of omitted variables perhaps leading to a degree of bias and inefficiencies in the least squares method due to heteroskedasticity. As shown in the propulsion sections of the theory chapter there are interactions between ship speed, shaft rpm and torque. The way these interact will have an effect on overall shaft power and this is not explicitly defined from a purely statistical model. This may be the source of the omitted variable bias and will lead to greater variability in the residuals and ultimately create greater model uncertainty.

The model was evaluated in the same way as the theoretical model as described in section 5.4. The results are summarised in Figure 45 and Table 19, the results of the statistical model evaluation can be compared to the theoretical model evaluation results by referring to Figure 28.

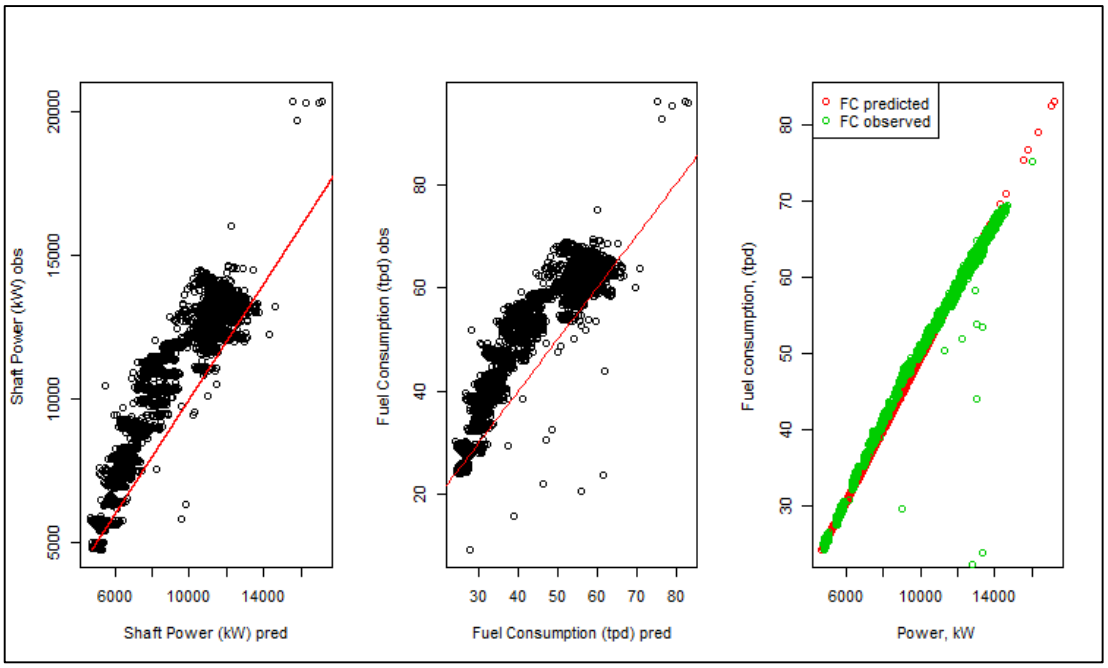


Figure 45: Statistical model evaluation; observed and predicted power and fuel consumption (Left: a, Middle: b, Right: c)

	Fuel Consumption	Power
Adjusted R^2	0.862	0.892
Intercept	7.967	1141.213
Intercept SE	0.281	52.068
Significance of test intercept = 0 (p-value)	0.000	0.000
Coefficient	0.971	1.026
Coefficient SE	0.006	0.006
Significance of test slope = 1 (a) (p-value)	0.000	0.000
RMSD	8.200	1654.022
Nr of observations	4106	4106

Table 19: Statistical model evaluation summary

These results show that generally, when assessed during the evaluation period the statistical model performs better overall, relative to the theoretical model, as indicated by a higher adjusted R^2 value and lower RMSD values for both fuel

consumption and power models. The improvement of 2.5% (relative to the theoretical model) is not stark for the shaft power model however the more marked improvement of 36% for the fuel consumption model is driven by the better engine performance estimation as shown in Figure 45(c). This indicates that the statistical model represents the data well over the training period, which is to be expected, even if the model were to be miss-specified (and including the model linearization errors), the problem of which would become apparent if the data were to be extrapolated over a different range during the evaluation period. In this instance, the evaluation period shaft power requirements are within range of those in the calibration period.

The intercept of the linear model for the regression of the observed on the predicted power is closer to zero, relative to the theoretical models, for both fuel consumption and power, indicating less bias (perhaps due to omitted variables, incorrect model assumptions or measurement error in the input data) in the predictions. The slope coefficient of the same linear regression is marginally closer to 1 for the statistical power model relative to the theoretical power model indicating an improvement in the model in capturing the shaft power response, however the opposite is true for the fuel consumption model, this is possibly because while the statistical model assumes a linear fit, the theoretical model considers the shape of the quadratic relationship between SFOC and engine load.

6.4 Chapter Summary

This chapter has shown that while there are some advantages of the statistical model over the theoretical, there are also some flaws which particularly came to light during the residual analysis. In particular there is some mis-specification of functional form and possible omitted variable bias present, possibly relating to the propeller rpm and torque effects. So, while the theoretical model underperforms due to either lack of data availability (hull specifics and acceleration and steering effects) or due to incomplete knowledge (model test extrapolation errors for example), the statistical model cannot fully compensate for this.

As highlighted in the chapter summary of the previous chapter (section 5.7), the type I and type II limitations to the theoretical model can be negated by the use of a statistical model if two conditions are met. First, that the functional form of

relationships between parameters and their effects on the dependent variable are correctly specified in the underlying statistical model. Second, that there is sufficient data (collected over a long enough time period) such that the error caused by the omission of time varying factors centre on a mean of zero.

While it may be possible to statistically average out the time varying omitted effects, some bias may still remain due to incomplete knowledge since the exact physical relationships between parameters may not be known (rpm and torque for example). The hybrid model presented in the following chapter aims to improve the model specification which will improve the efficiency of the least squares method in estimating coefficients and reduce the possibility of bias.

Chapter 7. Hybrid Model

The focus of this chapter is to form and evaluate a hybrid normalisation model by combining statistical techniques with theoretical knowledge and thereby considering the statistical functional form more carefully through disaggregation into more specific elements of the resistance, propulsion and power generation chain. The normalisation model then enables the actual operating/environmental condition to be reverted back into a reference condition from which the difference between modelled and measured fuel consumption equates to changes in the ship's performance. Two hybrid model versions are proposed and evaluated in this chapter. In the next chapter, the analysis is repeated for a second ship, a Suezmax tanker.

7.1 Shaft Power

7.1.1 Hybrid Model I

One straight forward method of combining the two models and reducing the possibility of miss-specification is to employ the same model as the statistical model in the previous section but with a different dependent variable. Instead of regressing power on to the environmental and operational variables, delta power becomes the dependent variable where delta power (ΔP) is the difference between the theoretically calculated shaft power and the measured shaft power. The advantage is that the theoretical underlying relationships are maintained. For example, in calm water, the theoretical model indicates the power requirements for a new ship with a smooth hull condition (150μm as stipulated by the H&M model) and with no appendages. The power that is above this (ΔP) is then attributed to the resistance arising from unknown appendages, the additional pressure resistance of an immersed transom and the fact that the ship hull roughness may be greater than 150μm. This difference is a function of, and therefore regressed onto the independent variables (speed and draught). When the environmental conditions are also present then these too have implications on shaft power; these are to a certain extent accounted for in the theoretical model. However, if the theory is not perfectly aligned with the reality then there will be a correction applied (which may be either positive or negative) through the regression coefficient estimates. Since the theoretical model of the

baseline already considers the nature of inter-parameter relationships then it is more obviously justifiable to make the assumption of linearity or second order relationships between the relatively small incremental increase (or decrease) that is represented by delta power, thus reducing the likelihood of miss-specification. This is highlighted by Figure 46 which shows a strong linear relationship between speed and delta power for a VLCC in the loaded condition. A larger proportion of the data points fall in the positive ΔP space as one would expect since the theoretical model was found generally to underestimate the shaft power (for the reasons previously mentioned relating to exclusion of appendages and hull fouling effects). There are however also periods where the ΔP is negative, perhaps due to the effects of the bulbous bow and possible over accounting for wind and wave effects in the theoretical model.

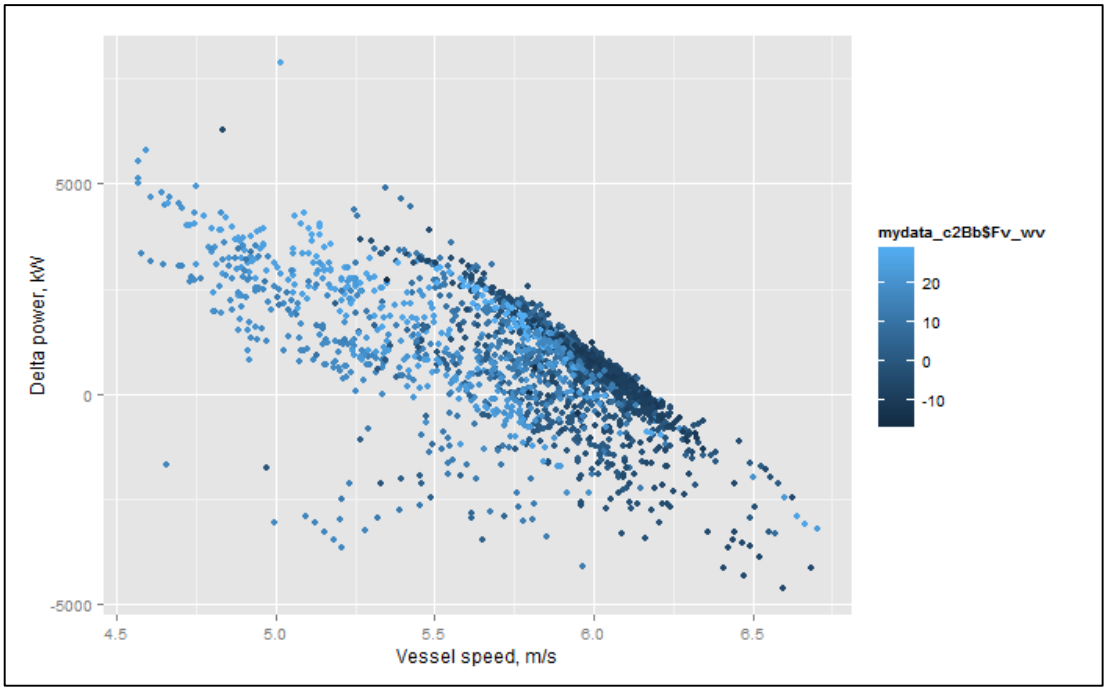


Figure 46: Speed and ΔP relationship for a loaded VLCC, coloured dimension represents the longitudinal component of wave speed

As it is difficult to ascertain the correct size and sign of model variables (as the wind/wave model and the model tests in the theoretical model may under or over-estimate shaft power), then the model originates from the same explanatory variables used in the statistical model formed in the previous chapter and variables are only removed if they are found to be insignificant. Residual analysis is used to qualify the

model and it is assessed in the evaluation period in the same way as the theoretical and statistical models. The full model was:

$$\Delta P_S = \alpha + \beta_1 V + \beta_1 V^2 + \beta_2 T + \beta_6(V_{wind,app,T}) + \beta_7(V_{wind,app,T}^2) + \beta_8(H_s) + \beta_9(H_s^2) + \beta_{10}(V_{wave,app,L})$$

It is noted that a large degree of uncertainty surrounding the theoretical model is likely to originate in the wind and wave effect estimations. As such, the hybrid model was investigated using the theoretical baseline, with and without the inclusion of the wave/wind model, the hypothesis being that the statistical model derived directly from these environmental effects (without first estimating the theoretically derived effects) and will estimate their effects on power more accurately. This was not the case and in all cases the hybrid model derived from the theoretical baseline that included the wind and wave model was preferred. This is because the latter theoretical baseline also includes the underlying functional form of these environmental parameters and therefore reflects the reality with greater trueness, resulting in a reduced ΔP.

The resultant model was:

$$\Delta P_S = \alpha + \beta_1 V + \beta_2 T + \beta_3(V_{wind,app,T}) + \beta_4(H_s) + \beta_5(H_s^2) + \beta_6(V_{wave,app,L})$$

The resultant coefficient and standard error characteristics are summarised in Table 20. The VIFs are all within limits, the coefficient for the transverse component of apparent wind speed is negative and highly significant, indicating that the theoretical model may highly over estimate this factor. P-values are all significant and with low per cent standard errors. This is with the exception of the term for significant wave height, this is the main factor and its higher order term is also in the model; the inclusion of the higher order term necessitates the inclusion of the main term.

	VIFF	Reff	AdjReff	Coeff	SE	P-value	% SE
Intercept				5850.945	202.234	0.000	3.456
V	1.269	-0.275	-0.216	-947.380	37.605	0.000	-3.969
T	2.160	0.056	0.026	54.866	4.984	0.000	9.083
$V_{Wave,app,L}$	2.117	0.065	0.031	16.741	2.133	0.000	12.743
H_S	9.635	0.007	0.001	15.874	54.474	0.771	343.157
H_S^2	9.620	0.125	0.013	51.231	12.243	0.000	23.898
$V_{Wind,app,T}$	1.210	-0.337	-0.278	-279.675	7.874	0.000	-2.815
Adjusted R^2	0.388						
Sample size	3171						

Table 20: Regression of ΔP on to the regression variables, Hybrid model I

In order to ensure the model conforms to the regression assumptions and the least squares method is efficient and unbiased, the residuals are inspected for mean independence between variables, model miss-specification and omitted variables following a similar procedure as for the statistical model. The results are shown in Figure 47. In a similar way to the statistical model correlations are observed between the residuals and wave height and again there is a degree of heteroskedasticity present as evident from the scatter plot of the residuals and power.

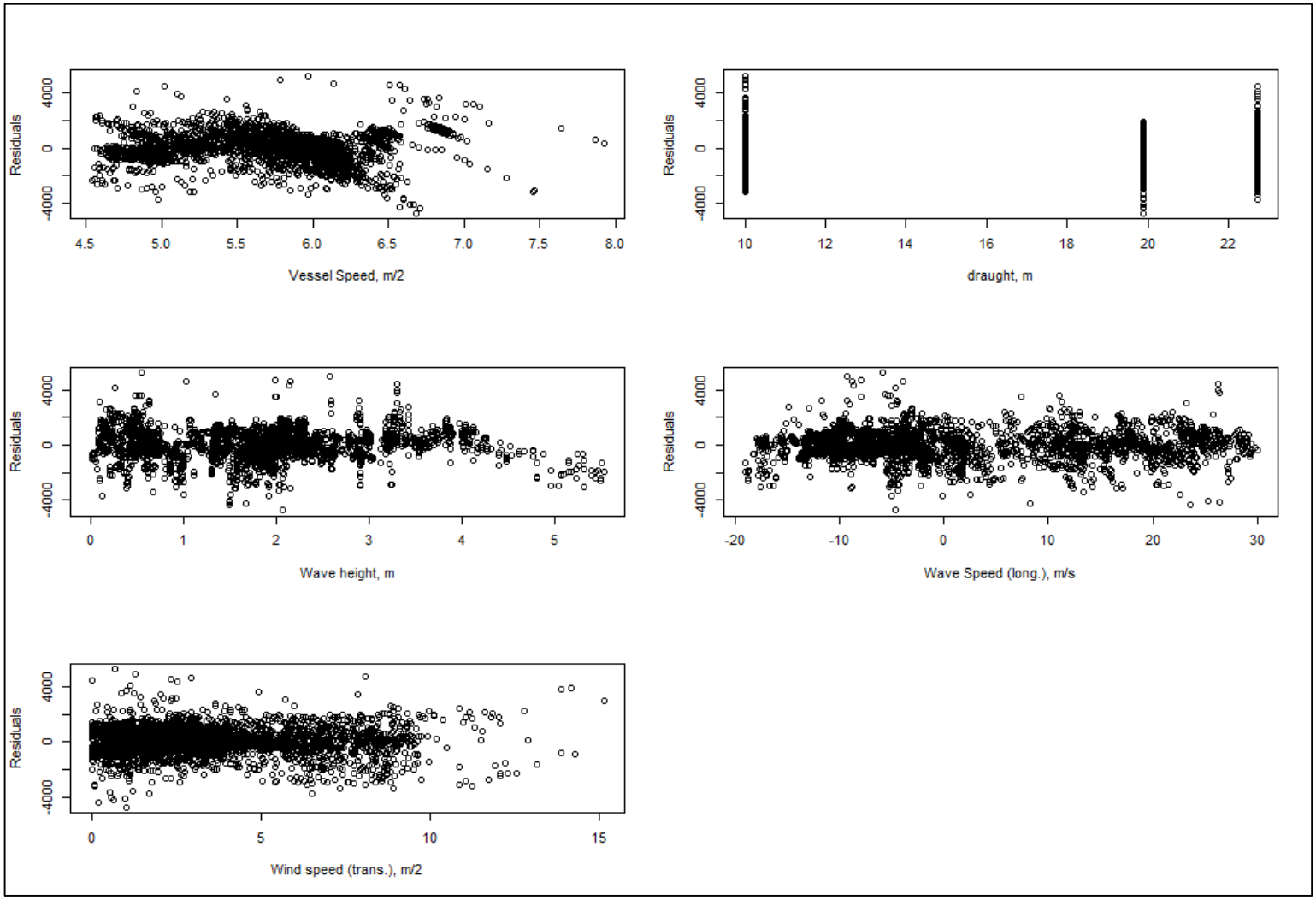


Figure 47: Residual plots relative to each explanatory variable and the dependent variable

The residual verses fitted plot (Figure 48) is an improvement relative to the comparable plot for the statistical model (see Figure 43) as the variables are more convincingly, randomly scattered about the x-axis with no particular pattern.

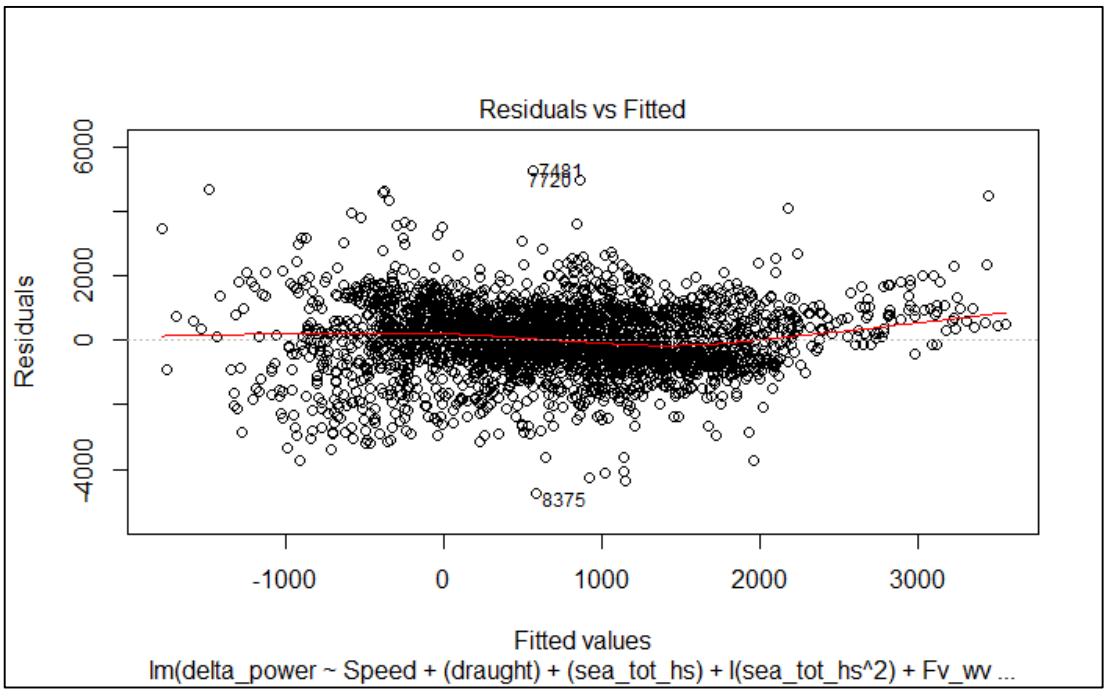


Figure 48: Residual verses fitted plot for hybrid model I

An inspection of the normal Q-Q plot indicates the residuals are not normally distributed, which is corroborated by rejection of the null hypothesis of the Jarque-Bera test for normality.

7.1.2 Hybrid Model II

Background

A disadvantage of the previously described hybrid model is that the delta power may not sufficiently pick up all of the omitted variables of the statistical and theoretical analysis. For example if bulbous bow effects that tend to decrease shaft power cancel out the omitted variable effects that would tend to increase shaft power, such as appendage resistance increase and hull fouling effects, then the delta power will not adequately reflect the nature of either of these changes sufficiently. It is not clear whether the inaccuracies from extrapolation of model tests of different ship type/size (and from a different decade) will cause over or under estimations in the shaft power.

Either way, it may be difficult to disaggregate these from opposing effects. So, it might be that a model relying simply on 'delta power' as the dependent variable lacks the required granularity to be able to decipher variable interactions.

The objective of hybrid model II is to firstly transform the raw effects (draught, speed, trim, waves etc.) into their direct effects on shaft power using the fundamental theoretical relationships, from Equation 3, Equation 4 and Equation 7, then the following equation is derived:

$$P_S = \frac{\frac{1}{2}\rho(WSA)V^2 C_T}{\eta_O \eta_R \eta_H}$$

Equation 34

Taking logs of both sides and eliminating the constant terms leads to the following additive equation:

$$log(P_S) = 3\log(V) + \log(C_T) - \log(\eta_O) - \log(\eta_R) - \log(\eta_H)$$

Equation 35

The proposed regression equation is therefore:

$$log(P_S) = \propto + \beta_1 \log(V) + \beta_2 \log(C_T) - \beta_3 \log(\eta_O) - \beta_4 \log(\eta_R) - \beta_5 \log(\eta_H)$$

Equation 36

By taking logs it is clear to see that the model is truly additive and therefore conforms to the fundamental assumptions of linear regression. Another advantage of this method is that the interactions between variables are accounted for in a more complex manner rather than more simplistic transformations of the raw effects to linear form. For example, the vessel's loading condition and draught affect the wave making resistance, frictional resistance and the added resistance in wind which all contribute in varying quantities to C_T. Draught also affects the hull efficiency through the thrust deduction factor (which is also related to the wake fraction

coefficient and therefore trim), and therefore the required thrust and consequently the propeller efficiency. Significantly, the magnitude of each of these effects is in differing relative quantities depending on the other explanatory variables. As described in sections 5.1 and 5.2, the theoretical model accounts for these interactions by explicitly calculating each and therefore they are carried through in to this hybrid model by separating the factors in this proposed way.

A further advantage of this method is that the rpm is implicitly considered; rpm is related to, but not entirely dependent on ship speed because of other interactions that occur. These additional interactions are due to hull and propeller fouling, ship speed, rpm and the vessel mode of operation. As described by (Krapp 2011), vessels are often operated in one of three modes:

- Constant vessel speed
- Constant main engine fuel supply
- Constant rpm

In the first case the effect of hull fouling is to increase frictional resistance and therefore to increase fuel consumption required to maintain the vessel's speed. In the second case the vessels speed is continuously reduced and therefore the fuel consumption remains constant regardless of the hull and propeller deterioration. In the final case, under constant rpm, counter-intuitively, the vessels average speed decreases and shaft power increases. This is because, even at a constant rpm, the increased hull frictional resistance causes the velocity with which the water enters the propeller to decrease and therefore the forward velocity of the vessel is altered due to the hull state causing the vessel to loose speed for a given thrust. The advance coefficient therefore decreases causing K_T and K_Q to increase and therefore the shaft power must increase (Krapp 2011). Therefore while ship speed and shaft speed are linearly positively correlated, there is a degree of independence between them according to hull and propeller condition. This is significant because in the purely statistical model, rpm effects are not separated from speed. This means that in the comparison of the model predicted shaft power to the measured shaft power for the quantification of hull and propeller performance, if the ship is not being operated in constant rpm mode then variations in rpm due to fouling may be overlooked and the

effect of time underestimated. It is therefore advantageous to somehow include rpm as an endogenous variable within the shaft power model and this is achieved by hybrid model II.

It is therefore hypothesised this model to consider the theoretical, possibly non-linear interactions between explanatory variables will result in a more accurate model relative to one that assumes linearised relationships without interactions. This is because the hybrid model II considers the effect of the operational mode on rpm, speed and shaft power and the magnitude of the interactions between them are then adjusted statistically in order to match the data. This improves on the limitations of the purely theoretical model (data availability, incomplete knowledge and model test extrapolation errors).

The model

The term representing total coefficient (C_T) was excluded because it was found to correlate significantly with speed and therefore the coefficient estimate was unstable and incorrectly indicated an inverse relationship with power. This implies that for this ship the effect of draught is more significant to the propeller, hull and rotative efficiencies than the ship's resistance; this was confirmed by examining the scatter plots. In reality, the inclusion of draught instead of the hull and relative rotative efficiencies yield better results. This also meant it was necessary to include the environmental variables, starting with the full set and removing redundant variables according to the same backward elimination process described in chapter 6; the p-values and REff (Equation 33) indicate their significance, the sign and size of the coefficient is assessed for feasibility and the VIF (Equation 31) and coefficient standard errors are evaluated for multicollinearities. Starting from the full model:

$$log(P_S) = \propto + \beta_1 \log(V) + \beta_2 \log(T) + \beta_3 \log(\eta_O) \quad + \beta_4\left(\mathrm{V_{wind,app,T}}\right) + \beta_5\left(\mathrm{V_{wind,app,T}}^2\right) + \beta_6(\mathrm{H_s}) + \beta_7\left(\mathrm{H_s}^2\right) + \beta_8\left(\mathrm{V_{wave,app,L}}\right) + \beta_9\left(\mathrm{V_{wave,app,L}}^2\right) + \beta_{10}(\mathrm{V_{wave,app,T}}) + \beta_{11}(\mathrm{V_{wave,app,T}}^2)$$

The coefficients of the final regression model are summarised in Table 21. The coefficients of the efficiencies (hull, propeller and rotative) are all negative as

expected, the p-values are all significant although it can be seen that the environmental variables improve the model RMSD overall, although they are not highly significant in reality given the magnitude of their REff. The per cent standard errors are within an acceptable range. The VIF's are quite high but still acceptable because the terms (log(V), log(η_H) log(η_R) and log(η_O)) are actually representing interactions.

	VIF	Reff	AdjReff	Coeff	SE	p-value	% SE
Intercept				3.238	0.043	0.000	1.342
log(V)	1.502	0.777	0.517	2.051	0.021	0.000	1.043
log(η_O)	15.303	-0.699	-0.046	-2.206	0.072	0.000	-3.258
log(T)	10.397	0.141	0.014	0.253	0.015	0.000	6.028
H_S	9.929	-0.063	-0.006	-0.017	0.005	0.001	-30.662
$V_{Wave,app,L}$	2.460	0.092	0.037	0.003	0.000	0.000	7.758
H_S^2	10.311	0.252	0.024	0.012	0.001	0.000	9.678
$V_{Wind,app,T}$	10.984	-0.293	-0.027	-0.029	0.002	0.000	-7.731
$V_{Wind,app,T}^2$	10.272	0.211	0.021	0.001	0.000	0.000	15.482
Adjusted R^2	0.932						
Sample size	3163						

Table 21: Hybrid model II, regression results for the regression of log(power)

Again, in order to ensure the model conforms to the regression assumptions the residuals are inspected for mean independence between variables, model miss-specification and omitted variables following a similar procedure as for the statistical model. The results are shown in Figure 49, in a similar way to the statistical model, correlations are observed between the residuals and wave height and again there is a degree of heteroskedasticity present as evident from the scatter plot of the residuals and log(power), Figure 50 and Figure 51.

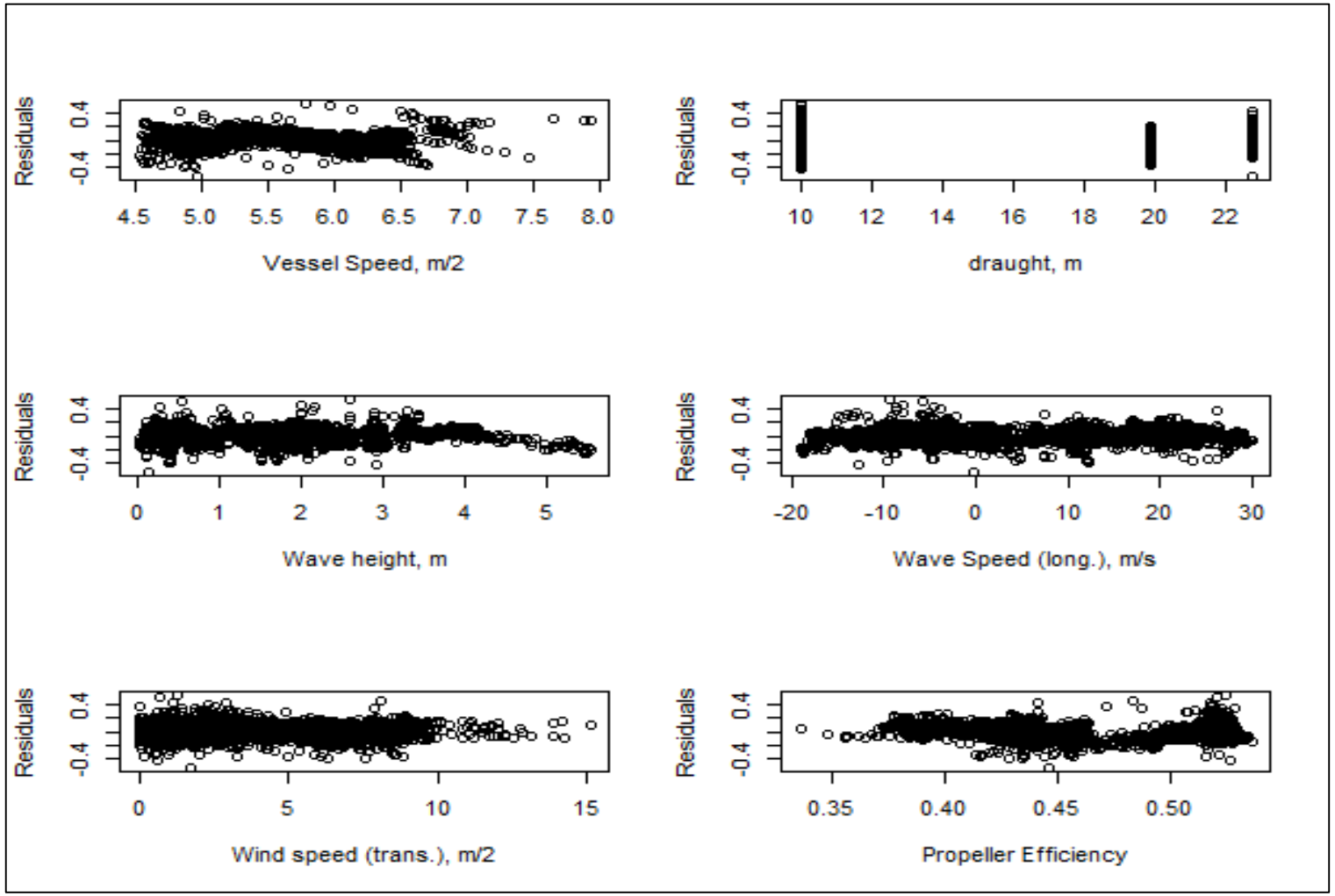


Figure 49: Diagnosis plot of the regression residuals vs the explanatory variables, hybrid II

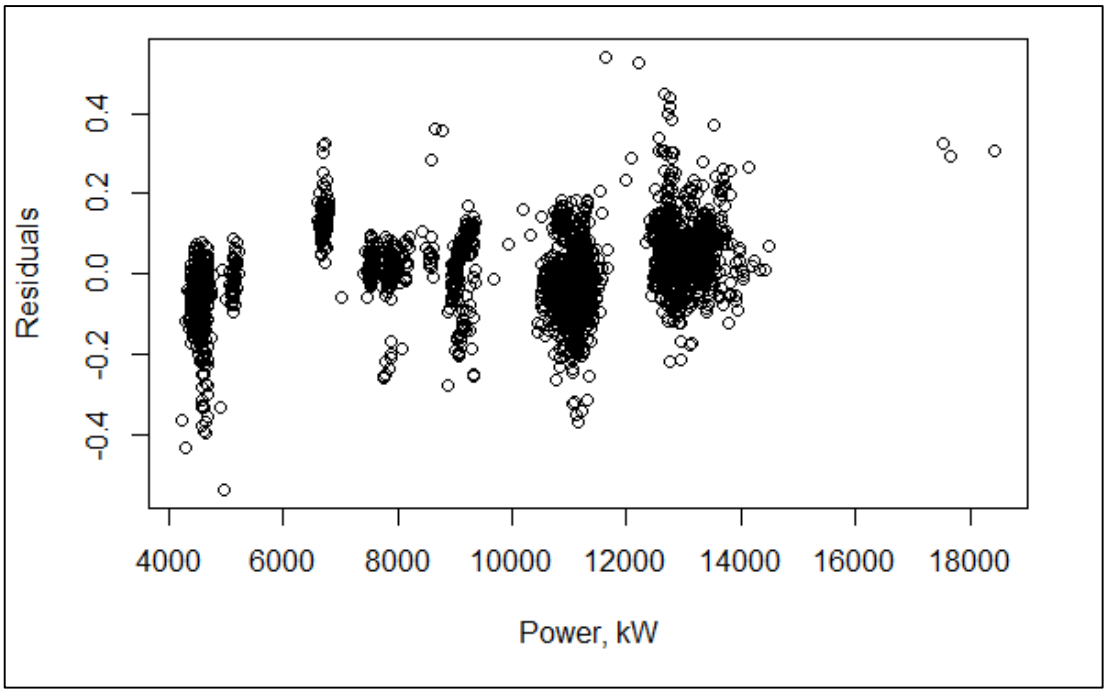


Figure 50: Diagnosis plot of the regression residuals vs measured power, hybrid II

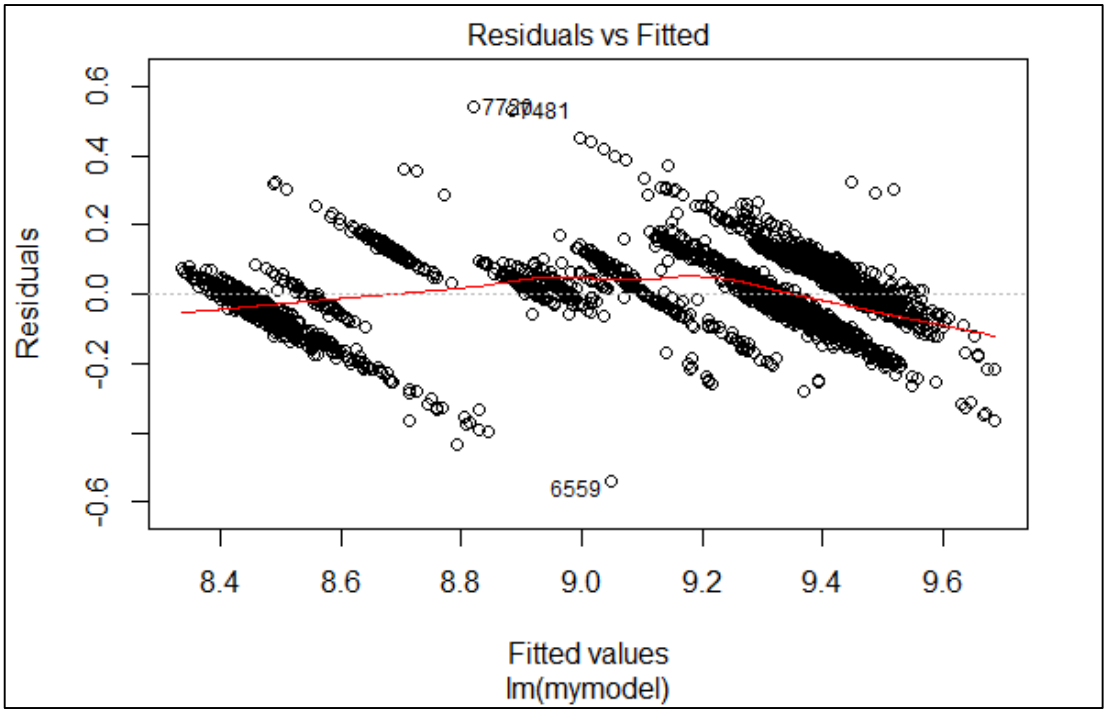


Figure 51: Diagnosis plot of the regression residuals vs predicted power, hybrid II

7.2 Fuel Consumption

The results of chapter 5 highlighted how the factor of 1.05 that was used in the theoretical model to account for deviation between test bed and actual operational

data (as described in section 5.3) is not appropriate and induced a degree of bias. The problem with the statistical fuel consumption model is the possibility of misspecification of its functional form. The data for the calibration period is shown in Figure 52, in the form of SFOC against shaft power.

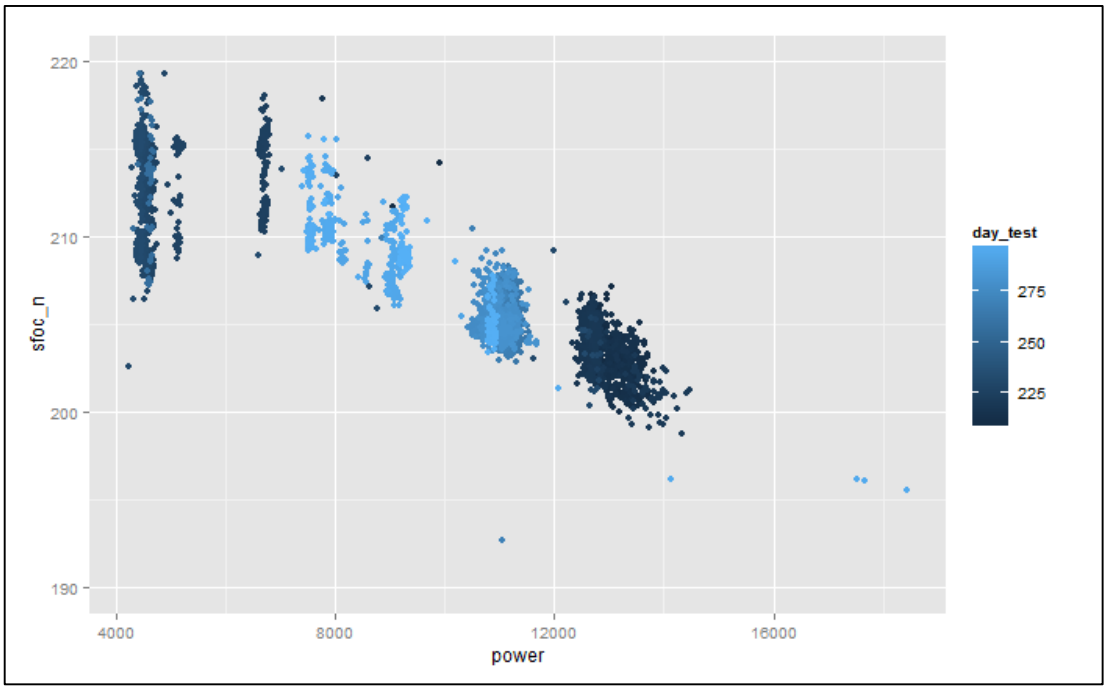


Figure 52: SFOC - shaft power relationship for the VLCC used in this thesis

Fitting the quadratic model over this calibration period yields the results presented in Table 22, which produces an inverse quadratic relationship. This model form is not supported by theory which suggests that a "U" shaped parabola is correct, this is corroborated by the evidence from low-load emissions tests such as those presented by Starcrest Consulting Group, Mitsui Engineering & Shipbuilding Co. et al. (2013).

	Estimate	Std. Error	t-value	P-value
(Intercept)	2.22E+02	2.22E+01	10.008	<2e-16
P_S	-1.40E-03	3.31E-03	-0.422	0.673
P_S^2	-6.04E-09	1.24E-07	-0.049	0.961

Table 22: Results of the SFOC model for the calibration period

The adverse effects of this mis-specification will come to light if it is required that the calibration data set be extrapolated outside the calibration range. Therefore, there may be some advantage in using the detail of the functional form of the engine

relationships proposed by the theoretical model and then adjusting these by employing the model on real data.

The hybrid model version is to compare the measured SFOC and the theoretical SFOC and to shift the theoretical as a function of the mean difference, thus maintaining the correct functional form of the theoretical relationship but refining it to match the data. The following regression model is proposed for the theoretical data (which originates from the theoretical engine curve of the appropriate engine size and type as presented in Figure 25):

$$SFOC_{TH} = \propto + \beta_1 P_S + \beta_2 P_S^2$$

Equation 37

Then $\Delta SFOC = (SFOC_M - SFOC_{TH})$, and

$$FC_{HY} = \frac{(SFOC_{TH} + \overline{\Delta SFOC}) \,.\, P_S \,.\, 24}{10^6}$$

Equation 38

Where subscripts M, TH and HY represent measured data, theoretical model results and hybrid model results, respectively. The results of the model for the calibration period are shown in Figure 53.

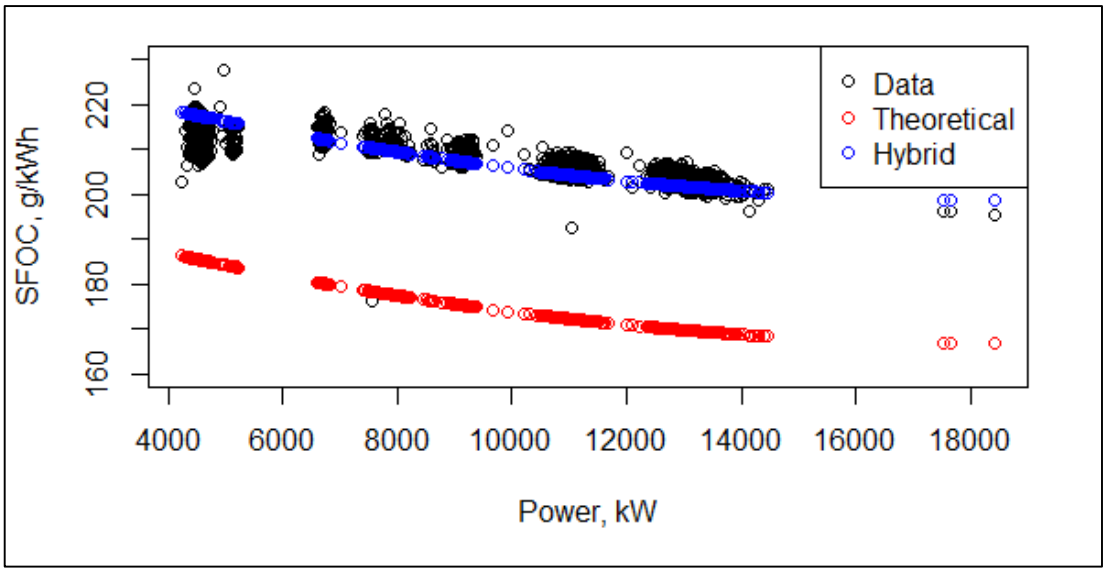


Figure 53: SFOC observed and predicted for the calibration period

7.3 Results

The results of the hybrid model evaluation below can be compared to the theoretical model evaluation results and the statistical model evaluation results by referring to Figure 28 and Figure 45, respectively.

7.3.1 Hybrid I

Figure 54 and Table 23 present the results for hybrid model I. For the engine performance model (predicting fuel consumption from shaft power, Figure 54a), the graphical presentation shows that the hybrid model, whilst being a significant improvement relative to the theoretical model, demonstrates no particular improvement over the statistical model. Between 8 MW and 10 MW the fuel consumption is slightly under predicted by the model, as is the case in the statistical model however to a marginally lesser extent. The results are summarised in Figure 45 and Table 19, the results of the statistical model evaluation can be compared to the theoretical model evaluation results by referring to Figure 28. The graphical presentation also implies that the hybrid power model is an improvement.

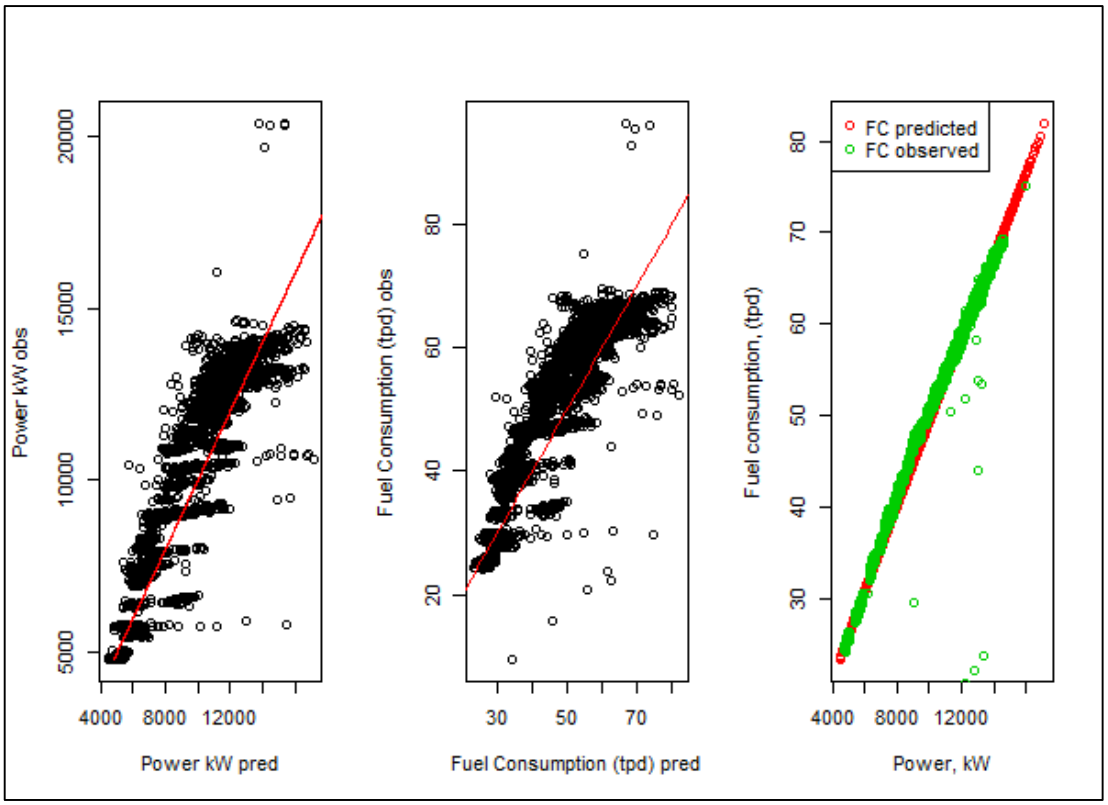


Figure 54: Hybrid model I evaluation; observed and predicted power and fuel consumption

The results are transformed in the same manner as described previously for quantifiable comparison with the statistical and theoretical models and presented in Table 23. In both cases (fuel consumption and power) the RMSD shows an improvement in the hybrid model, although the adjusted R^2 is lower and the intercepts are further from 0. Relative to the purely statistical model, the coefficients are also further from unity in both cases. The null hypotheses that the slope equals unity and the intercept equals zero are rejected in both cases and there is scope for improvement.

	Fuel Consumption	Power
Adjusted R^2	0.820	0.836
Intercept	6.740	1166.141
Intercept SE	0.336	66.009
Significance of test intercept = 0 (p-value)	0.000	0.000
Coefficient	0.933	0.951
Coefficient SE	0.007	0.007
Significance of test slope = 1 (a) (p-value)	0.000	0.000
RMSD	6.502	1335.422
Nr of observations	4106	4106

Table 23: Hybrid model I evaluation summary, ship 1

7.3.2 Hybrid II

Figure 55 and Table 24 present the results for the hybrid model II. The fuel consumption model is identical to that of hybrid model I. The quantification of the results over the evaluation period indicate in some respects an improvement in model accuracy in this model over all other models. For example, when applied in the evaluation period, for both the fuel consumption and power models, the hybrid II method exhibits a lower RMSD and in the linear regression of the observed on the predicted, the intercept is closer to zero for the hybrid method II, and is indicating a reduced bias in this method. However, the coefficient in the linear regression of the observed on the predicted is marginally closer to unity in the purely statistical power model and in both the statistical and theoretical fuel consumption models. As with all the other models, the null hypotheses that the slopes' equal unity and the intercepts equal zero are rejected in both cases.

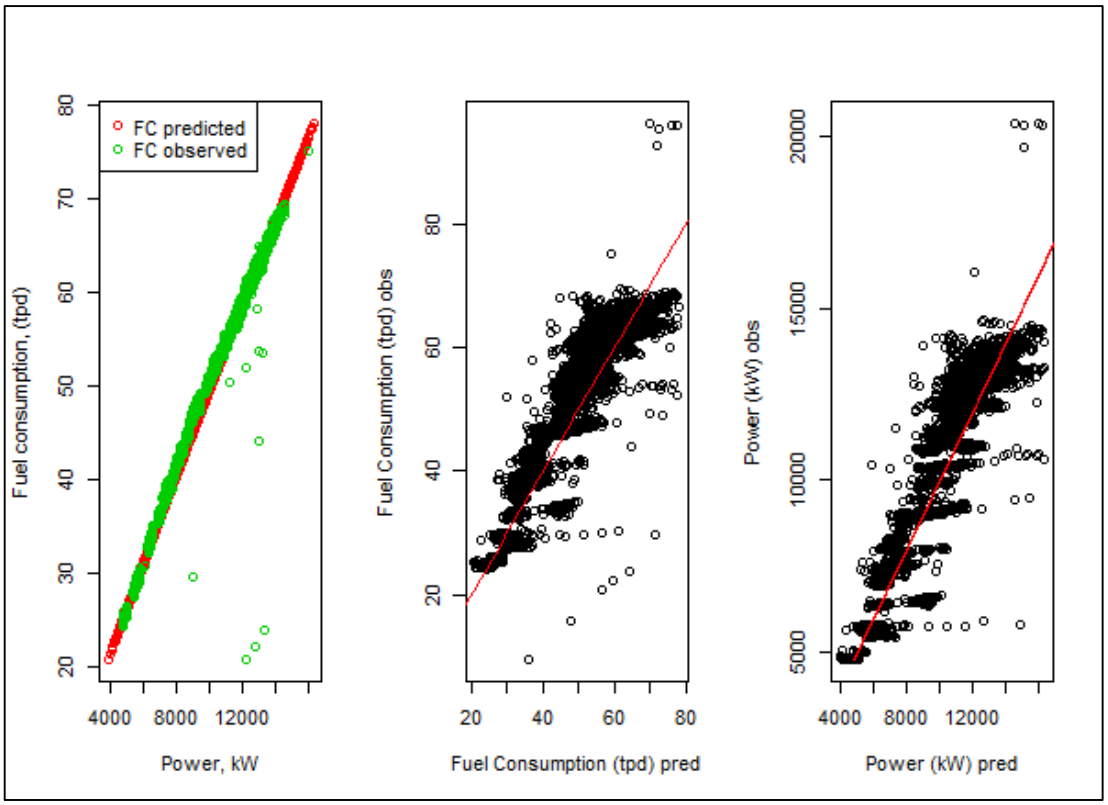


Figure 55: Hybrid model II evaluation; observed and predicted power and fuel consumption

	Fuel Consumption	Power
Adjusted R^2	0.848	0.860
Intercept	4.710	758.110
Intercept SE	0.317	62.659
Significance of test intercept = 0 (p-value)	0.000	0.000
Coefficient	0.953	0.969
Coefficient SE	0.006	0.006
Significance of test slope = 1 (a) (p-value)	0.000	0.000
RMSD	5.540	1141.473
Nr of observations	4106	4106

Table 24: Hybrid model II evaluation summary, ship 1

7.4 Chapter Summary

Hybrid model I is an extension of the statistical model and is an improvement relative to the statistical and theoretical models in terms of improving the accuracy of the ship performance metric. It is categorised as 'hybrid' because it combines statistical knowledge with theory however, after the endogenous parameters are generated in the theoretical model, no statistical methods are used to adjust them to the data and try and reduce the ship specific deviations from reality that these parameters may exhibit. In this way, hybrid model I offers no advantage over hybrid model II in furthering RQ 1.b and providing deeper insight in to the drivers and influences of the ship's performance.

In the hybrid II model, the theoretical model generates the fundamental parameters from the external drivers of ship speed, draught and environmental conditions. Statistical techniques are then used to adjust these to match the ship specific characteristics and therefore improve the accuracy of the power estimation thus reducing the variability in the performance metric (hence the conclusion to RQ 1.a). Furthermore, this method uses statistical techniques to estimate the effects of the fundamental parameters thereby providing deeper insight in to the drivers and influences of the ship's performance relative to the statistical model and hybrid I. In this respect, it is also preferable over the purely theoretical model because this hybrid model allows for adjustment of these fundamental parameters to consider the ship specific characteristics.

The results from this chapter and the preceding two chapters are drawn together and summarised in the following chapter.

Chapter 8. Interpretation, Discussion and Conclusions: RQ1

Research question 1 has two sub questions:

1. **Can hybrid models outperform theoretical/ black box models in the provision of detailed and quantifiable knowledge about a ship's performance?**
    a. **Can hybrid models deliver a more accurate ship performance metric?**
    b. **Can hybrid models provide deeper insight into the drivers and influences of the ship's performance?**

These questions are analysed in turn in this chapter.

8.1 Research Question 1.a.

Chapter 5, Chapter 6 and Chapter 7 have described and evaluated the theoretical, statistical and hybrid normalisation models. The normalisation models enable the actual operating/environmental condition to be reverted back into a reference condition from which the difference between modelled and measured fuel consumption equates to changes in the ship's performance. To summarise, it was found that the theoretical model is limited due to lack of data, incomplete knowledge or unknown unknowns (see chapter summary, section 5.7 for discussion). The statistical model attempts to negate the limitations of data availability and incomplete knowledge however both the correct functional form must be specified and there must be sufficient data (over a sufficiently long time period) in order for the statistical model to provide accurate results. The hybrid model II is shown to fulfil the former condition by using the theoretical model to better specify the correct functional form. This form is derived directly from the theory and by taking logs a truly additive model is produced (Equation 36), thereby fulfilling the requirements of linear regression. This results in the lowest RMSD, and in the linear regression of the observed on the predicted the intercept is closer to zero for the hybrid method II

indicating a reduced bias in this method. However, the coefficient in the linear regression of the observed on the predicted is marginally closer to unity in the purely statistical power model. This may be because of outliers not removed in the pre-processing step or because of characteristics of the specific datasets. Further analysis using data from a larger variety of ships is required to fully investigate this.

It is concluded that the answer to research question 1.a. is positive; the optimum model (of the models investigated) for ship performance estimation in terms of accuracy was found to be based on the hybrid II multiple linear regression model.

The first part of this chapter compares these findings to the analysis of data from a second ship, a Suezmax tanker. The same method and analysis is applied in order to understand if the results are transferable and robust.

8.1.1 Ship 2

The theoretical, statistical and hybrid models for a Suezmax tanker are presented here to establish if the same conclusions are reproduced. The data was filtered as detailed in section 4.6, the theoretical model was formed from the assumptions laid out in chapter 5 and the variable selection strategy and model analysis of chapters 6 and 7 was repeated for the statistical and hybrid models. The evaluation period for ship 2 was 94 days.

Statistical model

The statistical model was based on a calibration period of 100 days and the same procedure applied to ship 1 was applied to ship 2 in terms of identifying redundant variables (through the relative effects and p-values) and determining the presence of multicollinearities (through standard errors as a percentage of the coefficient estimate and VIF's). As in the case for ship 1, coefficient estimates were assessed for feasibility in terms of the correct sign and magnitude. A number of variants were attempted, the best fit for the data is summarised in Table 25.

	VIF	Reff	Adj Reff	Coeff	SE	P-value	% SE
Intercept				20013.411	1812.025	0.000	9.054
V	274.676	-1.974	-0.007	-3870.747	657.024	0.000	-16.974
V^2	279.029	2.213	0.008	346.863	51.737	0.000	14.916
T	2.513	-0.129	-0.051	-102.573	9.863	0.000	-9.616
H_S	2.789	0.351	0.126	274.693	17.380	0.000	6.327
$V_{Wave,app,L}$	1.773	0.086	0.048	10.837	1.762	0.000	16.256
$V_{Wave,app,L}^2$	1.708	0.080	0.047	1.037	0.128	0.000	12.302
$V_{Wind,app,T}$	1.302	0.133	0.102	33.265	5.378	0.000	16.167
Adjusted R^2	0.623						
Sample Size	1745						

Table 25: Statistical model evaluation summary, ship 2

The variable selection procedure indicated a similar model to that of ship 1 although for ship 2, the quadratic term for wave height is removed as it was the wrong sign and of a lower level of significance. Instead it was found that the quadratic term for wave speed was of the correct sign and deemed to be more significant. Secondly, the quadratic transverse wind speed term was found to be of only borderline significance and with a very large standard error (139% of the coefficient estimate) therefore a linear relationship was defined.

Hybrid model I

The variable selection for the hybrid models proceeded in identical format. This was again performed for the case where the theoretical baseline includes the theoretical wind and wave model and where it does not. The model was substantially improved when the wind/wave model was excluded; the results presented are therefore based on this iteration. The hybrid model is summarised in Table 26.

	VIF	Reff	AdjReff	Coeff	SE	p-value	% SE
Intercept				28097.772	347.096	0.000	1.235
V	1.175	-0.825	-0.702	-2969.021	42.453	0.000	-1.430
T	2.232	-0.381	-0.171	-588.919	9.458	0.000	-1.606
H_S	2.807	0.148	0.053	224.403	17.779	0.000	7.923
H_S^2	1.766	-0.002	-0.001	-0.436	1.799	0.809	-412.976
$V_{Wave,app,L}$	1.684	0.056	0.033	1.424	0.130	0.000	9.101
$V_{Wave,app,L}^2$	1.307	0.048	0.037	23.534	5.505	0.000	23.390
$V_{Wind,app,T}$	1.175	-0.825	-0.702	-2969.021	42.453	0.000	-1.430
Adjusted R^2	0.907						
Sample size	1745						

Table 26: Regression of ΔP on to the regression variables, Hybrid model I, ship 2

Hybrid model II

A number of variants were attempted, the best fit for the data is summarised in Table 27.

	VIF	Reff	AdjReff	Coeff	SE	p-value	% SE
Intercept							
log(V)	1.374	0.546	0.397	0.666	0.028	0.000	4.236
$\log(\eta_R)$	2.786	-0.307	-0.110	-29.668	2.419	0.000	-8.154
$\log(\eta_O)$	1.721	-0.622	-0.361	-0.502	0.027	0.000	-5.452
H_S	5.538	0.142	0.026	0.011	0.002	0.000	20.693
$V_{Wave,app,L}$	2.356	-0.036	-0.015	0.000	0.000	0.017	-41.953
$V_{Wave,app,L}^2$	2.340	0.123	0.052	0.000	0.000	0.000	8.803
$V_{Wave,app,T}$	2.149	0.060	0.028	0.001	0.000	0.000	27.028
Adjusted R^2	0.632						
Sample size	1745						

Table 27: Hybrid model II evaluation summary, ship 2

Results

The power and fuel consumption model evaluation summaries are shown in Table 28 and Table 29, respectively, for all models. Generally the model performance is poorer for ship 2 with low overall adjusted R^2 values, higher model intercepts and coefficients being further from unity. This could be because the data for this ship is more sparsely populated; the two data sets of ~90 days in length (one for the evaluation and one for the calibration period) were only available with a large time gap between them. During the period of missing data significant changes could have been made to the hull, propeller or engine.

	Theoretical	Statistical	Hybrid I	Hybrid II
Adjusted R^2	0.303	0.220	0.097	0.323
Intercept	6865.604	-3001.385	2494.823	-7195.367
Intercept SE	127.665	646.125	598.486	659.427
Significance of test intercept = 0 (p-value)	0.000	0.000	0.000	0.000
Coefficient	0.264	1.222	0.692	1.759
Coefficient SE	0.011	0.061	0.056	0.068
Significance of test slope = 1 (a) (p-value)	0.000	0.001	0.000	0.000
RMSD	3652.230	1828.147	1995.954	1653.816
Nr of observations	1407	1407	1407	1407

Table 28: Shaft power model evaluation summary, ship 2

Some comparisons could be drawn; in a similar way to ship 1, the RMSD is lowest for hybrid model II, and, for the power model, the adjusted R^2 is highest for hybrid model II. If it is accepted that reducing the intercept and coefficient to levels where null hypotheses that the slope equals unity and the intercept equals zero are accepted is practically very difficult, then the RMSD and the adjusted R^2 are the most important measures of model performance. In that case, it might be possible to conclude that the hybrid model outperforms the others. It is pertinent however to consider that using the RMSD between the predicted and measured results as the

ultimate measure of model performance is subjective in that it depends on the data. In order to enforce the conclusion of the hybrid model II being the optimum model (of the models investigated), more experiments need to be conducted using data from different ships.

The statistical model for the engine again out performs that of the hybrid model for this ship and so if this is combined with the hybrid power model, then the hybrid fuel consumption models are improved and the adjusted R^2 for the fuel consumption of hybrid model II is 0.326 is the highest of the four models and the RMSD is even further reduced to 8.959 tpd.

	Theoretical	Statistical	Hybrid I	Hybrid II
Adjusted R^2	0.262	0.243	0.037	0.190
Intercept	32.628	-19.470	23.254	-21.478
Intercept SE	0.554	3.004	2.843	3.617
Significance of test intercept = 0 (p-value)	0.000	0.000	0.000	0.000
Coefficient	0.229	1.192	0.386	1.326
Coefficient SE	0.010	0.056	0.052	0.073
Significance of test slope = 1 (a) (p-value)	0.000	0.001	0.000	0.000
RMSD	17.585	11.855	13.329	9.391
Nr of observations	1407	1407	1407	1407

Table 29: Fuel consumption model evaluation summary, ship 2

This has confirmed that for ship 2 also, the optimum model (of the models investigated) for ship performance estimation in terms of accuracy was found to be based on the hybrid II multiple linear regression model.

8.2 Research Question 1.b.

The second part of this chapter is to explore research question 1.b.; to establish if hybrid models can provide deeper insight in to the drivers and influences of the ship's performance.

One relevant advantage of a purely theoretical model is that in constructing the model, many parameters are generated; for any combination of speed, draught, wave height etc. it is possible to output the hull or propeller efficiencies, and the total resistance as well as the dependent variable, be it shaft power or fuel consumption. The major disadvantage however is that the theoretical model does not adjust any of the theoretical assumptions or the interactions between variables that would allow for the influence of ship specific characteristics to be correctly quantified. All deviations from the theory are amalgamated and attributed to the effect of time; this therefore increases the variability in the PI. This also means that no additional information, above the theory, about the ship's response to the environmental and operational state can be ascertained. For the reasons discussed in the first part of chapter 6, theory alone does not fully represent the reality of the underlying parameter interactions or of the resultant fuel consumption.

The statistical model goes a step further by estimating the individual effects of speed, draught and other independent variables. These estimations consider ship specific variations; however the gain in knowledge is limited to the quantification of the ship's response to each of these input conditions. Additional detail regarding possibly more insightful, endogenous parameter interactions, such as the hull efficiency or the total resistance is not output. Hence the term 'black box'. The ship-specific nature of the statistical model has helped deliver a more accurate ship performance metric and therefore is an improvement over the theoretical model in terms of RQ 1.a. (as discussed in chapter 6). However, in terms of providing deeper insight, the statistical model does not have an advantage.

Hybrid model I is an extension of the statistical model and is an improvement relative to the statistical and theoretical models in terms of improving the accuracy of the ship performance metric. It is categorised as 'hybrid' because it combines

statistical knowledge with theory however, after the endogenous parameters are generated in the theoretical model, no statistical methods are used to adjust them to the data and try and reduce the ship specific deviations from reality that these parameters may exhibit. In this way, hybrid model I offers no advantage over hybrid model II in furthering RQ 1.b and providing deeper insight in to the drivers and influences of the ship's performance.

In the hybrid II model, the theoretical model generates the fundamental parameters from the external drivers of ship speed, draught and environmental conditions. Statistical techniques are then used to adjust these to match the ship specific characteristics and therefore improve the accuracy of the power estimation thus reducing the variability in the performance metric (hence the conclusion to RQ 1.a). Furthermore, this method uses statistical techniques to estimate the effects of the fundamental parameters thereby providing deeper insight into the drivers and influences of the ship's performance relative to the statistical model and hybrid I. In this respect, it is also preferable over the purely theoretical model because this hybrid model allows for adjustment of these fundamental parameters to consider the ship specific characteristics.

This is the principal which suggests a positive answer to RQ 1.b. However, when considering data from the two ships, the correct parameter variables to use (because of sign, magnitude and significance of the estimated coefficient) are inconsistent. This is unexpected because theory dictates that they should be relatively consistent between ships; according to Equation 36 the speed and total coefficient of resistance should be positive while the hull, propeller and relative rotative efficiencies should be negative. In order to improve the accuracy of these potentially revealing coefficients and to explore further RQ 1.b., a least squares dummy variable method is explored.

8.2.1 LSDV Method for Time Effects

The least squares dummy variable (LSDV) method considers voyage specific effects. A significant advantage of this is that all of the available data over the entire PI calculation time period can be input to the model while still allowing for effect of

time to be evaluated as an independent variable. This means that the results are less conditional on the range and variance of the 90 day calibration dataset. The sample size is also significantly increased; a larger sample size enables the mean of the time varying, omitted parameters to converge to zero and the least squares to produce unbiased and efficient estimates, which results in an improved understanding of the parameter effects and interactions.

The time effect is included as a nominal-level variable which is assigned to each voyage and assumed to cause a voyage specific shift in the power and fuel consumption. This is represented by the 'voy_{bin}' factor and the regression model is therefore:

$$log(P_S) = \propto + \beta_1 \log(V) + \beta_2 \log(C_T) - \beta_3 \log(\eta_O) - \beta_4 \log(\eta_R) - \beta_5 \log(\eta_H) + FACTOR(Voy_{bin})$$

Where one voyage is a round trip, Voy_{bin}, ensuring that there is sufficient variation in all the explanatory variables within each voyage bin.

Fundamentally, a coefficient is assigned to each voyage and this is interpreted as the effect that that voyage has on the shaft power while controlling for the other independent variables. The coefficients of these continuous variables now represent the effects of those variables while controlling for time. For example, this now estimates the hull efficiency at the baseline (voyage 1) and the effect of time on each of the continuous variables is grouped together in the coefficient for subsequent voyages.

There is a high degree of multi-collinearity between speed and the total coefficient of friction; therefore it is proposed that C_{tot} is removed and the main environmental variables that also effect C_{tot} (wave height, apparent wave speed and wind speed) are included explicitly resulting in the following regression model:

$$log(P_S) = \propto + \beta_1 \log(V) + \beta_3 \log(\eta_O) + \beta_4 \log(\eta_R) + \beta_5 \log(\eta_H) + \beta_5 H_S + \beta_5 V_{wave,app,L} + \beta_5 V_{wind,app,T} + FACTOR(Voy_{bin})$$

The exclusion of draught may be rationalised because its main effect is through the thrust deduction factor which consequently causes a decrease in the hull and propeller efficiencies. However it is also noted that a change in draught affects the wetted surface area and therefore the frictional resistance (and the wave resistance to a small degree) and that this effect is obscured by the exclusion of C_{tot} without compensating explicitly for draught. The overall shaft power model prediction is unlikely to be affected since the coefficient estimates for the other variables will compensate, however it may mean that individual effects are slightly overestimated. The justification for this is that the effect that draught exerts on the efficiencies is large relative to its effect on the overall resistance.

The model is first tested using theoretical power as the dependent variable, the estimate of the regression coefficients are presented in Table 30.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	1.5E+00	1.8E-02	8.3E+01	< 2e-16
log(V)	3.1E+00	6.5E-03	4.8E+02	< 2e-16
$\log(\eta_O)$	-3.8E+00	3.5E-03	-1.1E+03	< 2e-16
$\log(\eta_R)$	-1.8E+01	3.3E-01	-5.5E+01	< 2e-16
$\log(\eta_H)$	-1.3E+00	6.7E-02	-2.0E+01	< 2e-16
H_S	1.1E-02	1.4E-04	8.1E+01	< 2e-16
$V_{wave,app,L}$	2.1E-04	1.1E-05	1.9E+01	< 2e-16
$V_{wave,app,T}$	7.0E-04	6.5E-05	1.1E+01	< 2e-16
factor(Voy_{bin})2	-2.8E-03	4.6E-04	-6.1E+00	1.4E-09
factor(Voy_{bin})3	6.2E-04	4.6E-04	1.4E+00	1.7E-01
factor(Voy_{bin})4	-9.9E-04	4.5E-04	-2.2E+00	2.9E-02
factor(Voy_{bin})5	-1.1E-02	4.7E-04	-2.4E+01	< 2e-16
factor(Voy_{bin})6	8.6E-05	4.4E-04	1.9E-01	8.5E-01
factor(Voy_{bin})7	7.0E-03	4.3E-04	1.6E+01	< 2e-16
factor(Voy_{bin})8	-5.1E-03	5.0E-04	-1.0E+01	< 2e-16

Table 30: Regression of theoretical shaft power on the theoretically calculated fundamental variables

The signs are correct; efficiencies are inversely proportional to power and all others are positive. The orders of magnitude are as expected; speed is close to 3 for example. The voyage factors are approximately equal to unity (green crosses in Figure 56) which is as expected since the theoretical model makes no assumption for the effect of time. This builds confidence in the ability of the OLS method to correctly estimate these coefficients from the data and shaft power. The actual measured power is then input as the dependent variable and the regression results are presented in Table 31.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	9.02E+00	1.19E-01	76.042	< 2e-16
log(V)	3.97E+00	4.26E-02	93.128	< 2e-16
$\log(\eta_O)$	-1.76E+00	2.33E-02	-75.343	< 2e-16
$\log(\eta_R)$	-3.46E+01	2.16E+00	-16.044	< 2e-16
$\log(\eta_H)$	-2.13E+01	4.40E-01	-48.42	< 2e-16
H_S	2.30E-02	8.99E-04	25.585	< 2e-16
$V_{wave,app,L}$	1.49E-03	7.32E-05	20.307	< 2e-16
$V_{wave,app,T}$	-1.63E-02	4.30E-04	-37.77	< 2e-16
factor(Voy_{bin})2	8.11E-02	3.00E-03	27.07	< 2e-16
factor(Voy_{bin})3	8.21E-02	3.02E-03	27.221	< 2e-16
factor(Voy_{bin})4	2.47E-02	2.96E-03	8.344	< 2e-16
factor(Voy_{bin})5	1.38E-01	3.08E-03	44.95	< 2e-16
factor(Voy_{bin})6	-2.36E-02	2.93E-03	-8.054	8.68E-16
factor(Voy_{bin})7	-2.32E-02	2.85E-03	-8.134	4.54E-16
factor(Voy_{bin})8	-2.34E-02	3.28E-03	-7.114	1.19E-12

Table 31 Regression of measured shaft power on the theoretically calculated fundamental variables

Again the size and sign of each coefficient is as expected and each coefficient is significant. The coefficients of the LSDV method represent how the theoretical variables affect the shaft power in reality while controlling for all other variables, including the effect of time through the voyage factor. This model therefore matches

the theory to the data while maintaining the theoretical functional form as defined by the various components of efficiencies.

The effect of speed and the hull and relative rotative efficiencies on shaft power is increased indicating that the theoretical model underestimates these effects. The effect on performance of wave height and speed is also found to be greater in this analysis than the theory suggests. The transverse component of wave speed is now negative indicating that it has a lower positive influence on shaft power than the theory suggests, perhaps because of interaction of beam or head winds. Other signs are the same.

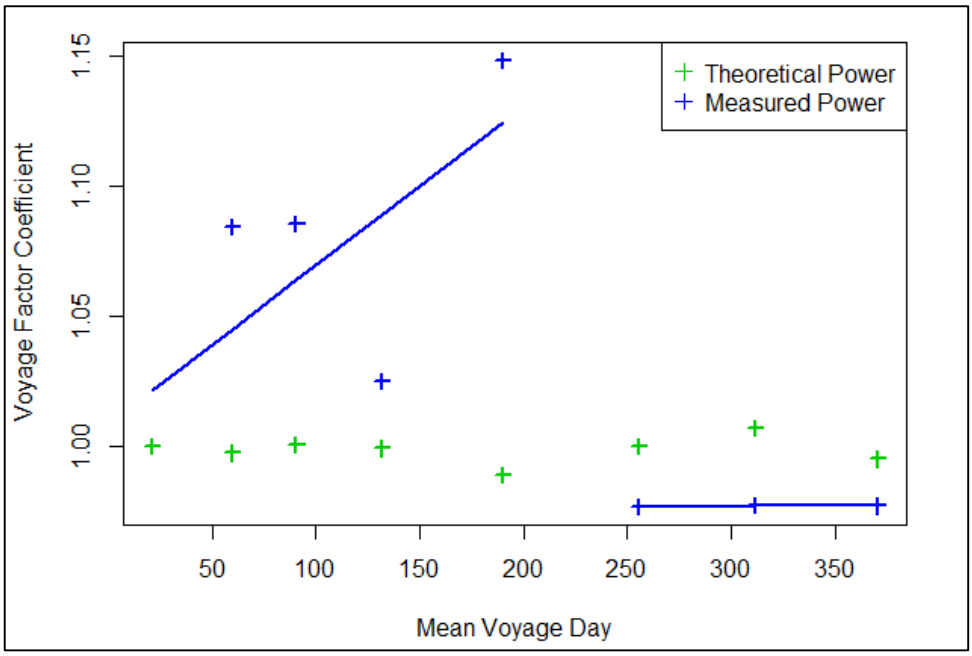


Figure 56: Dummy variable coefficients for factor Voy_{bin} in the statistical power model

Another advantage of this method is that the effect of each voyage on power may be plotted to represent the average voyage (time) effect on the shaft power while controlling for the other variables, as plotted in Figure 56. This is an alternative way to output the trend in the overall ship performance; an increasing trend indicates a deterioration in ship performance. The effect of subsequent voyages is positively increasing over the first 200 days and then plateaus thereafter (as represented by the blue lines in Figure 56. Meanwhile, for the regression of theoretical power on to the same regression model (the results displayed in Table 30), the voyage factor is

approximately constant which is as expected because the theoretical model does not include any time effects.

This translates to the speed – power relationships for each voyage as plotted in Figure 57 where the red lines shows the speed – power relationships for the voyages after day 200 (possible hull clean / maintenance event) and the black lines represent the speed power relationships pre day 200 representing an approximate 20% increase in shaft power at the extreme of the difference. The plots of Figure 56 and Figure 57 demonstrate that some maintenance event that affected the performance of the hull or propeller (hull scrub for example) occurred at approximately day 200.

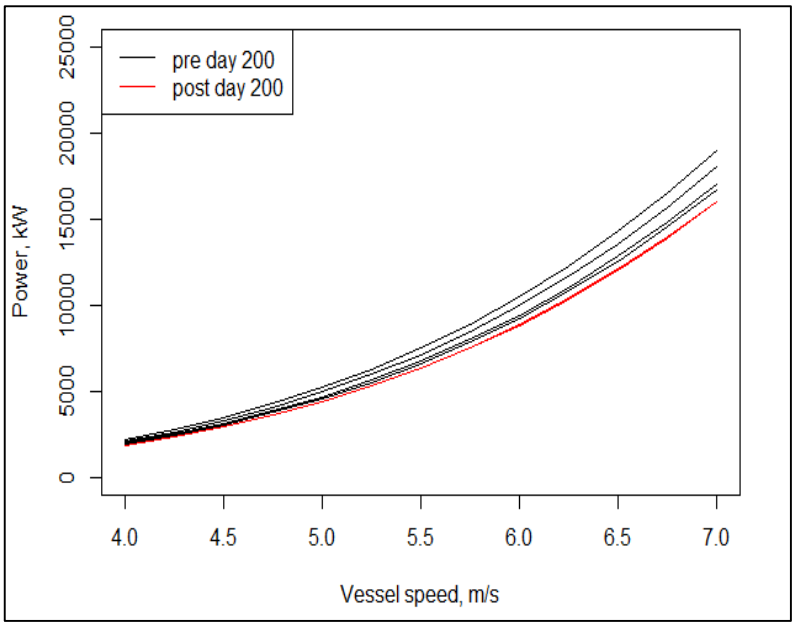


Figure 57: Speed - power relationship for each voyage

The above methods (LSDV and performance indicator method of chapter 7) have both estimated coefficients that represent the environmental and operational conditions on the shaft power. These are summarised in Table 32 and their standard errors and standard errors relative to the coefficients are summarised in Table 33 and Table 34, respectively.

Coefficients	PI calculation method	LSDV method
(Intercept)	14.590	10.310
log(Speed)	5.191	4.204
$\log(\eta_H)$	-39.790	-26.840
$\log(\eta_R)$	-118.800	-10.590
$\log(\eta_O)$	-2.002	-1.768
H_S	-0.049	0.011
I(sea_tot_hs^2)	0.013	0.002
Fv_wv	0.001	0.002
Fh_wn	-0.023	-0.009
I(Fh_wn^2)	0.001	-0.001

Table 32: Coefficients for different methods

Theoretically, the LSDV model should have the most accurate coefficients since more data is used in their estimation, this is perhaps corroborated by the suspiciously high coefficient on log(speed) which should be around 3.2 for low speed tankers such as this VLCC (MAN 2011).

Standard Errors	PI calculation method	LSDV method
(Intercept)	2.69E-01	1.33E-01
log(Speed)	9.01E-02	5.99E-02
$\log(\eta_H)$	9.65E-01	5.00E-01
$\log(\eta_R)$	5.12E+00	6.40E+00
$\log(\eta_O)$	5.88E-02	2.26E-02
H_S	4.24E-03	2.67E-03
H_S^2	9.59E-04	6.27E-04
Fv_wv	1.80E-04	7.31E-05
Fh_wn	1.79E-03	8.08E-04
I(Fh_wn^2)	1.71E-04	7.85E-05

Table 33: Standard errors for different methods

% Standard Errors	PI calculation method	LSDV method
(Intercept)	1.85	1.29
log(Speed)	1.73	1.42
$\log(\eta_H)$	-2.43	-1.86
$\log(\eta_R)$	-4.31	-60.47
$\log(\eta_O)$	-2.94	-1.28
H_S	-8.61	23.38
H_S^2	7.25	25.12
Fv_wv	26.58	4.57
Fh_wn	-7.75	-8.67
I(Fh_wn^2)	21.40	-11.39

Table 34: Standard errors for different methods relative to the coefficient estimate

The propeller efficiency coefficient contains data giving insight into the propeller condition at the baseline voyage (the first voyage). If this regression is run repeatedly on a longer term dataset, for consecutive years over a 5 year dataset for example, then logging the coefficient of the propeller efficiency effect on a yearly basis may reveal relative changes in the propeller efficiency over that time period. The same is true for each of the continuous independent variables and for example, the hull efficiency describes the long term hull condition as fouling causes an increase in wake fraction and a decrease in hull efficiency.

8.3 Conclusions: Research Question 1

The conclusions to the investigations in to research question 1 are summarised in the following paragraphs.

The hybrid II model can deliver a more accurate ship performance metric because it allows the conditions for the type I and type II limitations to the theoretical model to be negated by the use of statistical techniques to allow for ship specific deviations from the theory. This is an advance from the purely statistical model because the functional form or parameter relationships and their effects on the dependent variable are maintained and therefore an improvement in the efficiency of the least squares

method in estimating coefficients and reducing the possibility of bias. The bias in a purely statistical model is thought to stem from the propeller characteristics and the effect on shaft power of the interactions between shaft torque and rpm that cannot be explicitly defined by a statistical model alone. In the hybrid II model, the functional form is derived directly from the theory which therefore allows for such interactions to be accounted for (such as the propeller efficiency). By taking logs of the underlying significant variables (see Equation 36) a truly additive model is produced. This model produces the lowest RMSD and an intercept closer to zero (for the linear regression of the observed on the predicted) indicating a reduced bias in this method. This corroborates the suspicion that the purely statistical model suffers from omitted variable bias. The statistical model does however exhibit a coefficient marginally closer to unity for the same linear regression of the observed on the predicted. Further work is required to validate these result with a wider variety of ship types and sizes.

Furthermore, the hybrid model II provides deeper insight into the drivers and influences of the ship's performance by evaluating the sensitivity of fundamental parameters. The theoretical model disaggregates the input operational and environmental parameters into more specific elements of the resistance, propulsion and power generation chain (the hull and propeller efficiencies for example). Statistical techniques (least squares regression) then adjust these to match the dataset, thereby allowing for ship specific effects. Therefore, the hybrid model provides greater in depth knowledge than the statistical model alone and these parameter estimates are also more reliable than the parameters output by the theoretical model alone.

In order to improve the accuracy of these potentially revealing coefficients and to further establish that greater, in-depth knowledge can be ascertained from a hybrid model, an LSDV method has been deployed. The initial results are encouraging.

An alternative way is suggested to output the trend in the performance indicator which requires plotting the voyage bin coefficients. Further work is required to validate this method, and in particular, datasets which include the dates of known

interventions that affect the hull, propeller or engine performance must be used. This is discussed in the further work section of the conclusion chapter.

Over a longer time period it may be possible to demonstrate how hull, propeller and relative rotative efficiencies evolve by plotting their coefficients on an annual basis by running separate regressions for consecutive years. This would be a further advancement on insight gained from a purely statistical or theoretical model.

Chapter 9. Uncertainty Analysis and Simulation

The aim of an uncertainty analysis is to describe the range of potential outputs of a system at some probability level, or to estimate the probability that the output will exceed a predefined threshold value or performance measure target value (Loucks 2005). The latter case is particularly related to reliability and risk-based assessment. The aim of the analysis in this thesis is the former; to estimate the mean expected measurand (the ship performance index) and the confidence in this result at the 95% level. Further, the method is then developed so that, for the first time, the relative significance of the various sources of uncertainty may be deciphered.

The initial part of this chapter builds on the detailed discussion in Chapter 4 which characterises the relevant sources of uncertainty related to ship performance monitoring. A primary objective is to disaggregate the effect of each source and to understand relative influences on overall uncertainty. With this in mind, a method is presented to quantify the uncertainties in ship performance analysis. If perfect information is available, uncertainty can be quantified by comparing a benchmark with a given measurement. (See the representation and discussion of trueness and precision in chapter 4 and Figure 17). In the case of ships for which noon report or continuous monitoring data is available, there is no authoritative benchmark of ship performance available, even for practical purposes. Knowing the actual true performance is not practically possible, firstly because of the previously discussed aleatory/measurement uncertainty in the dataset and secondly because the model on which the ship performance is estimated may not perfectly represent reality. Consequently, the performance model has to be derived from available data, which has its own inherent uncertainties. Confidence limits assume that the model is true and the error represents the ability of the model to estimate the true response of the system. If the model cannot be confirmed to be true then the confidence limits not only represent the sensor and sampling uncertainties but also the model uncertainty. Since the model uncertainty is unknown then it is not possible to disaggregate and apportion the overall uncertainty to different sources using real life data alone.

Through chapters 5 to 7, the RMSD was used as an indicator of the accuracy of a model/method when applied to a sample dataset (the VLCC and Suezmax data). To provide a more in-depth analysis of the uncertainty of ship performance measurement, this existing RMSD analysis can be complimented by a study of uncertainty carried out using a different and independent approach.

It is proposed that a valuable alternative approach is a simulation of a ship's performance and the process of performance measurement. In this approach, a baseline ship performance simulation represents 'the truth' and a Monte Carlo (MC) method to combine the elemental sources of uncertainty looks at the consequence of those sources on performance measurement uncertainty. This can be perceived as a 'bottom up' approach. The results of this are then compared with uncertainty estimates from experimental data derived using RMSD ('top down approach') in order to build mutual confidence in the two uncertainty analyses. In this chapter, the quantification of each source of uncertainty is presented and the input parameters defined for the MC experiments to be undertaken. The results of these are then presented and finally, a sensitivity analysis is conducted to estimate the magnitude and relative impact of input uncertainties. This enables the analysis of the merits and drawbacks and quantification of the overall accuracy of the two different approaches (NR and CM) which are assumed to be representative of industry 'standard' data acquisition strategies. It also enables insight to be gained into which are the most significant alterations to be made to these strategies in order to reduce the overall uncertainty.

9.1 Uncertainty Quantification: General Method

A simulation of ship performance is constructed in the time domain considering an assumed operational profile, an assumed constant rate of degradation and some randomly fluctuating environmental conditions. The uncertainties highlighted in chapter 4 are all incorporated in to the simulation which takes into account the different methods that can be deployed in the propagation of uncertainties (see section 2.3.2 for a review), the nature of the data available and the uncertainties associated with that data (both aleatory and epistemic). This method follows the GUM approach and the errors are propagated through the model using the Monte

Carlo Method. A similar approach was adopted in Kamal, Binns et al. 2013, there are also other possible approaches to uncertainty quantification as discussed in the literature review.

The aims of this simulation are to:

1. Identify the most significant contributory factors to the uncertainty through a sensitivity analysis which considers both elemental bias and precision limits in the overall resultant uncertainty.
2. Compare the overall theoretical uncertainty in a noon report dataset with that of the continuous monitoring dataset.
3. Assess the statistical model using its associated model parameter uncertainties as inputs to the MC method for both the NR and CM datasets and compare these to the uncertainty achieved from a filtered dataset and simplified model.

This section firstly describes the details of the general method and then outlines the three experiments designed to satisfy the aims. An overview of the method is described in the following 5 steps:

1. Identify each elemental uncertainty source, classify and define probability distribution parameters.
2. Simulate the ships operating profile and performance trend and represent realistic sampling and filtering algorithms.
3. Propagate the errors through the data reduction equations that form the normalisation model using the Monte Carlo method, simulate n_1.
4. Formulate the output distribution of the resultant performance indicator, report precision as the 95% confidence interval of the ship performance linear trend line at the final evaluation period time step.
5. Validate the method by comparison with data derived ‘top down’ analysis.
6. For the sensitivity analysis of bias effects, repeat steps 3 and 4 at the bias limits of the input uncertainties. Report the bias as the change in the mean of the MC iterations.

This is summarised diagrammatically in Figure 58.

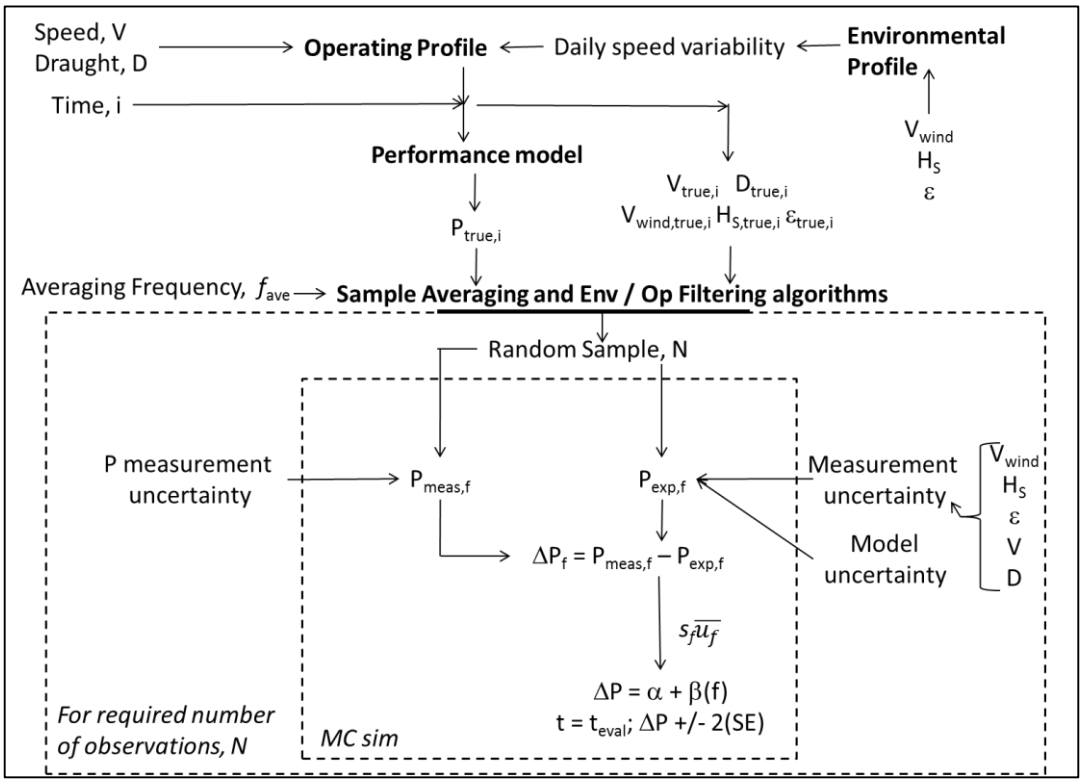


Figure 58: Diagrammatic representation of the Monte Carlo Method. Subscript, i: maximum temporal resolution 1/15minutes, f: sample averaged according to f_{ave}

9.1.1 Simulated ship operational profile

The ship is assumed to be operating with a loaded and ballast speed of 14 knots and a loaded and ballast draught of 18.0 m and 10.0 m respectively. Each voyage is 15 days and the ship alternates between loaded and ballast states spending 50% of its time in each. These values are representative of large wet or dry bulk cargo ships.

A simulated environmental profile is input (wind speed, wave height and prevailing direction which is assumed to be coincident for both), a cyclic weather pattern which is formed from a Weibull distribution repeats over 5 days and this induces some daily speed variability. There is a randomness introduced into the simulation due to natural fluctuations (either from the effects of drift, changes in rudder angle, etc.) through the addition of normally distributed noise to the underlying 'true' ship speed. The effect of averaging daily draught variability is included through its assumed measurement uncertainty.

The ship is assumed to be operated in a constant power/fuel consumption mode therefore this speed pattern is superimposed on a long term speed reduction which is enforced to offset an assumed rate of power increase due to hull and propeller deterioration. The time dependent degradation is assumed to be independent and increasing at a constant rate of 7.8% per year (extrapolated from 2.6% average speed loss (IMO_MEPC_63/4/8 2011)). This is an assumption, the actual degradation, especially with regards to hull and propeller fouling is known to depend on many factors as discussed in section 5.4. The effect of this assumption of linearity on the uncertainty analysis depends on the uncertainty metric used; this is discussed in detail in the 'Model Parameter Uncertainty Comparison' of section 9.2.3.

The underlying ship operational/environmental profile inputs are presented in Figure 59. The effect of these speed, draught and rate of deterioration and environmental profile assumptions on the distribution of the output performance value are discussed in section 9.4.1. The presence of seasonality is dependent on the cyclic nature and the global location of the ship's operations, it is difficult to generalise and it is not included in this analysis.

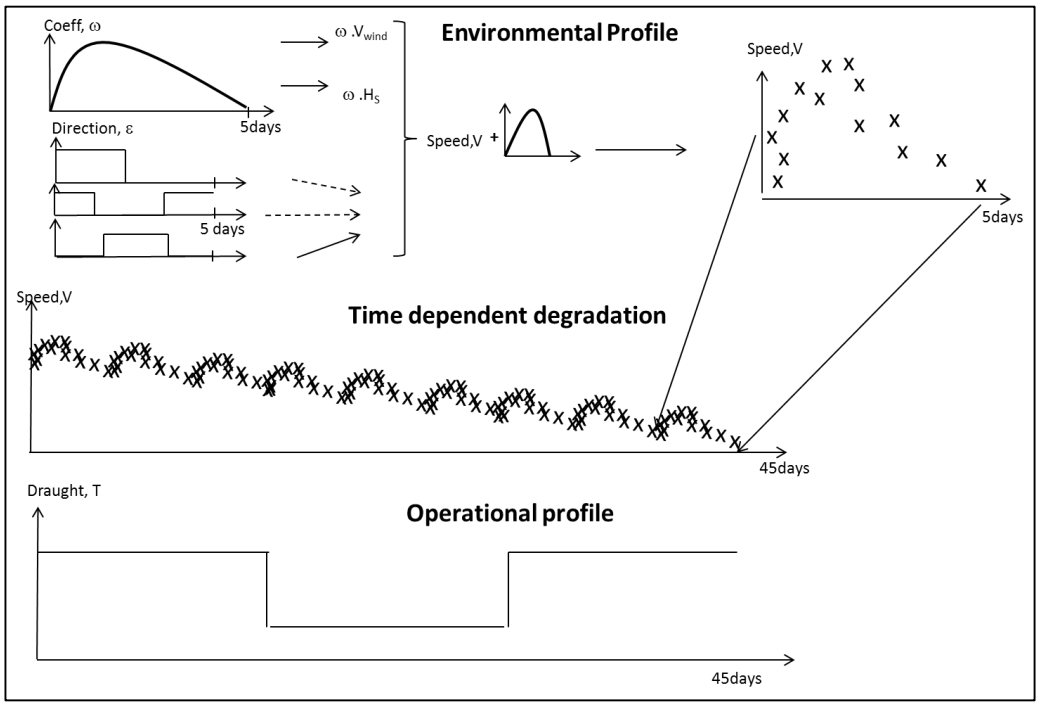


Figure 59: Underlying ship profile

9.1.2 Simulated Ship Performance and MC Sampling

The underlying ship operational and performance profile is established for each independent variable at the maximum temporal resolution (i), 1/15mins. The ship's true performance is assumed to be defined by one of three models, for the full normalisation model for example (the statistical model of chapter 6), all of the independent variables; ship speed, draught, wind and waves are therefore input at each time step, i (1/15mins), and, with the inclusion of some effect of degradation on both the speed and power, the true, underlying power, $P_{true,i}$ is output at each time step using the full normalisation model.

The sampling algorithm averages each variable according to the input sample averaging frequency, f_{ave}. After passing through the sample averaging algorithm, all of the 'true' independent variables have some measurement uncertainty superimposed by random sampling of the PDF representative of each sensor precision. These sampled explanatory variables, defined at each time step, f, are input to evaluate an expected power using the same normalisation model used for the $P_{true,i}$ definition. However, the model now also includes model parameter uncertainties by randomly sampling from a PDF representative of the precision with which the model parameters are defined. This results in the $P_{exp,i}$ measurement which is basically equivalent to $P_{true,i}$ minus the degradation plus the measurement uncertainty. The fact that the same model is used in both the $P_{true,i}$ and the $P_{exp,i}$ simulation means that the ΔP reflects exactly the input degradation, with the uncertainty incorporated which reflects the sampling and averaging uncertainties, the measurement uncertainties and the model parameter uncertainties. These representative model and sensor PDFs are described in section 9.2. Finally, an assumed additive and independent correction factor for noon reports is included in the $P_{exp,f}$ parameter when appropriate.

Hence, the following user-specified, variable drivers are changeable in order to investigate the uncertainties and fulfil the three aims laid out in the initial part of this section;

- Sensor uncertainties

– Sample averaging frequency, f_{ave}
– Sample size (number of observations) due to filtering, N
– Model parameter uncertainties
– Human error
– Sample size (evaluation period length), t_{eval}

This enables the Monte Carlo simulation to establish and measure each of these variable influences on the overall uncertainty of the performance indicator. Each is described in more detail and their distribution parameters are defined in section 9.2.

In order to fulfil the second aim of this chapter, there are three types of model investigated depending on the level of filtering applied. The very basic is simply an assumed cubic relationship between ship speed and power/fuel consumption, which therefore requires a significant degree of filtering on draught and environmental conditions in the pre-processing step. Next includes a relationship with draught, which has less filtering and a slightly more detailed model. Finally, there is a full model that relates the power/fuel consumption to all operational and environmental variables with no filtering applied. These three scenarios are represented diagrammatically in Figure 60, labelled 3,2,1, respectively.

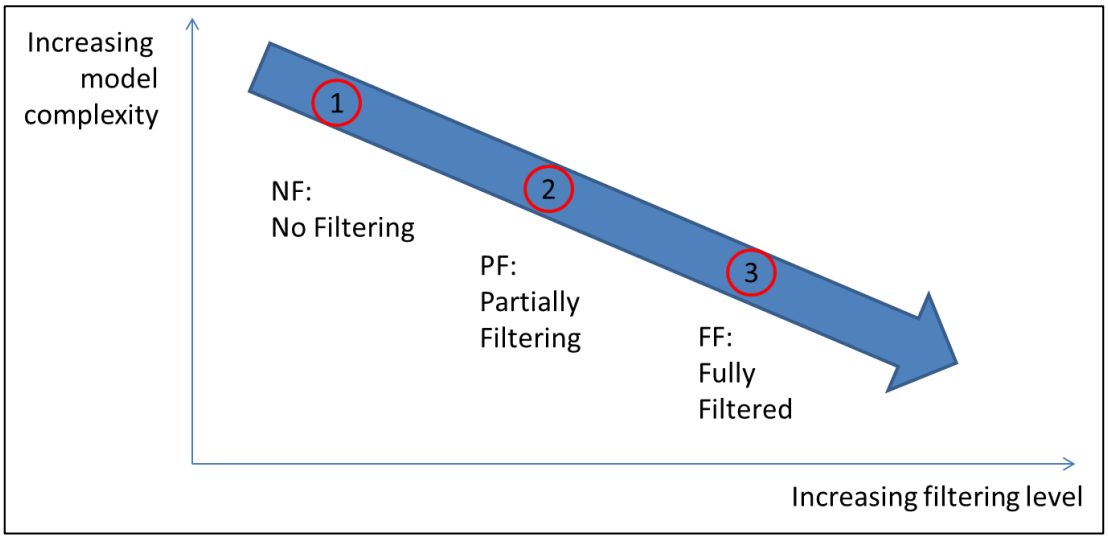


Figure 60: Relationship between model complexity and filtering level for each of the trial models

9.1.3 Overall Uncertainty Measurement

The performance indicator is the delta of the expected and measured power, ΔP_f (the measured power only deviating from the true power by the randomly sampled PDF representing the power sensor precision). A filtering algorithm removes samples according to any filtering criterion imposed (such as sea state for the purposes of the model assessment investigation as described in section 9.3.3) and the number of observations, N, that are realistically likely to remain due to days spent in port or stuck sensors, etc (the value of N is discussed in 9.2.2). In order to evaluate the unbiased effect of N, then the sample size must be repeatedly sampled at random from ΔP_f to produce the final output distribution for ΔP at each time step over the total evaluation period. The total evaluation period, t_{eval}, is also variable and the effect of changing this is investigated in the sensitivity analysis. A random selection from the PDFs at each time step forms a trend of delta P. The performance indicator, β_1 is then established from the regression of the linear trend over time:

$$\Delta P_f = \propto + \beta_1(t)$$

Equation 39

This is repeated and the MC outputs the distribution of β_1. From the distribution the mean and standard deviation indicate the performance indicator and its associated uncertainty which is recorded at the 95% confidence level; 2σ.

The confidence interval of the trend in the performance indicator (β_1) or the standard error of the mean are both possible uncertainty metrics. The standard error of the mean of the delta P values over the time period is calculated from:

$$SEM = \frac{2\sigma_{\Delta P}}{\sqrt{n}}$$

Equation 40

The SE of the mean should be used when comparing the simulation to actual data for validation purposes; the reasons for this are explained in more detail in section 9.2.3.

9.1.4 Method Validation

The method is validated by a comparison of the above 'bottom up' method which combines elemental sources of uncertainty and a 'top down' method which calculates the overall uncertainty from an actual dataset. As the truth is unknown in the actual dataset then only precision, and not bias, can be compared in this way. The validation is deployed using CM inputs to the simulation and a CM dataset, in this way the human error and sample averaging effects are assumed to be excluded and therefore the model parameter and measurement uncertainties can be observed in isolation. Model and measurement uncertainties cannot be disaggregated in the CM dataset however the overall trends and order of magnitude of uncertainties are used to give an indication of the reliability of the method. The relative trends with sample size and model type (filtering/normalisation) are also compared. Further details and results are presented in the section that refers to model uncertainty quantification, 9.2.3, in the model parameter precision comparison part.

9.2 Quantification of Elemental Sources of Uncertainty

This section presents the quantification of the parameters for the uncertainties associated with the input PDFs, a qualitative discussion is presented in Chapter 4.

9.2.1 Instrument Uncertainty

The following discusses realistic quantifications of the instrument uncertainties (bias & precision combined) as described in chapter 4. This Monte Carlo simulation deals only with the random errors associated with the sensors, the effects of the sensor bias that are also presented below are investigated in the sensitivity analysis.

Some manufacturers quote sensor accuracies which are assumed to reflect manufacturing tolerances; however the details of the quantifications (experimental set-up, numbers of trials, etc.) are often not well documented or publicly available. Manufacturers may also quote the sensor repeatability as calculated under laboratory conditions (same measurement system, operator, equipment) or reproducibility (different operator). The actual precision in service conditions is likely to be inferior

as different operators, equipment/facilities and measurement systems work to reduce the overall precision. This is described, for example, in the determination of torque measurement uncertainty in ITTC (2002).

Discussions with industry experts (such as those who are actively involved in the ISO 19030 working group; 'Measurement of Changes in Hull and Propeller Performance') and experience with ship performance datasets suggest that the instrument precisions quoted below are appropriate. However these are estimates as there are many ship / crew specific influential factors that create differences (such as maintenance procedures, calibration, sensor location onboard). The sensitivity of the performance indicator to changes in these is explored in the sensitivity analysis of section 9.4.1. All sensors are assumed to be consistent during the simulation (no change in repeatability over time) and linear (absence of change in bias over the operating range of the measurement instrument). Sensor resolution is assumed to be reflected in the quoted sensor precision and not a restricting factor in the uncertainty of the ship performance measurement.

Wind and Wave Sensors

As discussed in section 4.1 the environmental data may be acquired from onboard sensors or from models based on datasets that cover large areas which are subsequently combined using the ship's position and timestamp data. The range of dataset accuracies, particularly those that involve numerical models, is quite vast.

In a recent study, Jan Tellkamp, Peter Friis Hansen et al. (2008) compare onboard wind and wave measurement sensor accuracies to measurements derived from models using input data from larger area coverage data sets. The onboard wave data comes from a wave radar which was identified as one of the most accurate wave measurement devices. The modelled data is an advanced, 3rd generation model based hind cast data and successfully validated in numerous applications. Their results are summarised in Table 35. It is evident that the onboard measurements are more accurate in this particular comparison. It is also noted in the same study however that wind measurements from onboard wind sensors are influenced by their position and

the error may be up to 100%. In the absence of a clear definition, these accuracies are assumed to relate to the common use of the word accuracy, meaning bias.

Parameter	**Accuracy**
Model results:	
Wind speed	+/- 20%
Wind direction	+/- 10%
Significant wave height	+/- 15% or +/-0.5m
Peak period	+/- 10%
Onboard Measurements	
Wind speed	+/- 5% (approx.)
Wind direction	+/- 3^0
Significant wave height	+/- 10% or +/- 0.5m
Peak period	+/- 0.5 s

Table 35: Summary of the accuracy of modelled and measured environmental data (Jan Tellkamp, Peter Friis Hansen et al. 2008)

In an assessment and review of current literature on the bias and precision of various instrumental measurements of the wave characteristics (direction, T_Z and H_S), Bitner-Gregersen and Hagen (1990) 'tentatively' describe the uncertainties associated with buoys and ship borne radar as negligible for properly calibrated instruments. This of course does not include the model error in the numerical model required to accompany this data for finer resolution spatial-temporal models of the sea state variability.

A specialist committee, tasked to recommend updates to procedures for conducting full scale trials and long term performance monitoring (ITTC 1999), present example relative wind speed precision limits and bias limits as reproduced here in Table 36. These values reflect calibrated, onboard anemometer readings; the wind direction precision limits are similar to those of Table 35. The wind speed precision is on average slightly higher and as the 2008 study is more recent, those precision limits are used. In the absence of alternative bias limits the values of Table 36 are used. These tally well with worst case measurement uncertainties as calculated by Klitsch (1993).

Parameter	Units	Bias Limit	Precision Limit	Uncertainty
Relative wind speed	kn	±0.46	±1.24	±1.32
Relative wind direction	deg	±5.02	±3.26	±5.98

Table 36: Example of bias and precision limit values for surface ship trial data, uncertainties are calculated by the RSS method for a 95% confidence interval

For a NR dataset, crew interpretation and recording of the wind and waves is the measurement instrument and therefore there is an element of associated human error. Insel (2008) suggests that a bias error limit of ±1 BF is representative of the human error on the sea state BF scale measurement; this is then translated to wind speed and wave height by regression formulae presented by the same author:

$$V_{wind} = -0.00532x^3 + 0.23797x^2 + 0.80594x - 0.01113$$

Equation 41

$$\bar{H} = 0.02556x^3 - 0.11520x^2 + 0.17985x$$

Equation 42

The additional model errors on these regressions are found from the SEE of the curve fit and are 0.075 m/s for wind speed and 0.0845 m for wave height, these are assumed to be negligible. To obtain the significant wave height, the approximation $H_S = \bar{H}/0.64$ is used (Ainsworth 2008). Insel (2008) estimates bias error limits as $\pm 10^0$ for the wind/wave directions measurements.

In the study of Bitner-Gregersen and Hagen (1990), the accuracy (defined in this study to be trueness) of observer readings are required to be from +/- 20% for significant wave height, 10^0 for wave direction and +/- 1.0 s for the average wave period (95% confidence interval limits). However, the author also states that this is

very much dependent on the training of the observer and is difficult to generalise. As discussed in section 2.3.3, the classification of epistemic and aleatory uncertainty depends on the question that is being asked and in this instance, because the observer (crew) changes throughout the dataset then this bias should be modelled as random uncertainty (precision). This is in contrast to the characterisation of sensor uncertainties which have a clear bias and precision because it is assumed that the same sensor is used throughout the dataset.

The ship's heading data is required for the vector transformation of absolute into apparent prevailing winds however the uncertainty associated with this measurement is of the order of 0.27% (Klitsch 1993). It is therefore considered negligible.

It is evident that there is a reasonably large spread of the wind and wave sensor uncertainties depending on the data acquisition strategy (NR or CM), data source (onboard sensors or metocean data) and the subsequent sensor type and position onboard or the fidelity of the numerical model selected.

Ship Speed Sensor (STW)

The ITTC (1999) specialist committee, recommending procedure for full scale trials and long term performance monitoring, present example vessel speed precision and bias limits as ±0.10 knots and ±0.05 knots respectively, leading to a calculated uncertainty of ±0.11 knots. Seatechnik quote speed sensors accuracies of 0.01%. The IMO resolution "Performance Standards for Devices to Indicate Speed and Distance" (IMO 1995) stipulate that errors in the indicated speed, when the ship is operating free from shallow water effect and from the effects of wind, current and tide, should not exceed 2% of the speed of the ship, or 0.2 knots, whichever is greater. As this latter example is a guideline then a conservative estimate of 1% is used.

Draught

The draught uncertainty associated with draught gauges is of the order ± 0.1m, in noon report entries the draught marks at the perpendiculars may be read by eye and, depending on sea conditions, a reading error of ±2 cm can be assumed (Insel 2008).

However, noon report entries are often not altered during the voyage to record variability due to trim and/or the reduction in fuel consumption and therefore a greater ±1.0 m (1σ) uncertainty is assumed. Often the draught field in CM datasets is manually input and therefore the same uncertainty applies.

Shaft Power

The shaft power sensor uncertainty arises from the addition by quadrature of the rpm sensor error and the torque sensor error. Industry quoted errors for continuous monitoring systems are in the order of 0.1% (1σ) and 0.5% (1σ) respectively. These result in an overall combined uncertainty of 0.51%. HBM (http://www.hbm.com/) quote an uncertainty of 'well under' 1% for their torque meter. VAF quote an overall error of less than 0.25% for their optical torque measuring system (http://www.vaf.nl/category/id/5/show/brochures). Quoted manufacturer shaft power sensor accuracies range from 0.25% (Seatechnik, http://www.seatechnik.com/) to 0.1% Datum Electronics (http://www.datum-electronics.co.uk/). In the absence however of well documented evidence supporting these claims alternative references are sought. The ITTC (1999) specialist committee recommended procedure for full scale trials and long term performance monitoring, present typical shaft power precision limits and bias limits as ±3% and ±1.08% respectively. A more recent ITTC (2002) specialist committee on speed and powering trials calculate the combined uncertainty of shaft power to be to be 1.38%. The 0.5% and the 1.0% bias for the shaft speed and torque sensors, respectively, as quoted in (ITTC 2002) suggest 1.1% possible sensor bias. A shaft power precision of 1% is used in the MC simulation for the CM results.

Fuel Consumption

As described in chapter 4, in the NR dataset, fuel consumption is generally recorded rather than shaft power. The ship performance metric that is derived from fuel consumption measurements reflects engine, hull and propeller performance. The deterioration with engine performance over time is assumed to be small relative to the changes in the hull and propeller performance and therefore this is viewed as an additional uncertainty in the hull and propeller performance measurement, rather

than an extra part of the ship performance measurement. It is assumed to be negligible in this study. There may also be human error introduced however this is not included here (see section 9.2.4). Each part of the measurement and conversion to shaft power introduces additional uncertainties. This is what is quantified in this section and it is included in the simulation as an uncertainty on the shaft power.

The fuel consumption may be measured from tank soundings or automated fuel flow meters or a combination of the two, possibly augmented with information from bunker delivery notes (BDNs).

Fuel flow meter accuracy ranges between 0.05 per cent and 3 per cent depending on the type, the manufacturer, the flow characteristics and the installation. The following information is based on the information from VAF instruments (http://www.vaf.nl/products/). The mass flow measurement accuracy is quoted as 0.1% to 0.4% depending on the maximum flow measured. Positive displacement sensors measure the volumetric flow and deliver 0.1% trueness and 0.05% precision in the volume measurement over the entire range however, in the subsequent conversion to mass; the density measurement includes an error of the order of 0.1% and this also decreases the overall fuel consumption measurement accuracy. A temperature correction is also required which may introduce an additional error. If there is no temperature correction, the inaccuracy in the calculated volume increases by almost 1% for each 10^{o}C temperature difference which increases proportionally to the circulation flow rate relative to the maximum fuel consumption. The accuracy of VAF viscosity sensors are of the order of 2% (http://www.vaf.nl/category/id/2/show/brochures).

SFOC deviations due to the difference between engine test bed performance and actual installed engine performance is cited as being between 1% and 3% by industry experts such as those who are actively involved in the ISO 19030 working group; ‘Measurement of Changes in Hull and Propeller Performance’.

The accuracy of fuel consumption measurement by tank soundings is difficult to estimate for the reasons described in more detail in chapter 4, Faber (2013) have suggested 2-5%.

The fluctuations in fuel quality such as the presence of non-combustible components (such as water or air) or deviations in LCV and density, particularly due to temperature deviations are difficult to quantify. For this reason the upper end of the range of possible accuracies, 5% is assumed for NR measurement uncertainties.

9.2.2 Sampling uncertainty

The sample should be homogenous and of a sufficient size for the desired precision and confidence levels, both the fact that a finite sample is taken from an infinite temporal population and that variable factors may be averaged between samples cause uncertainty in the result.

The latter influence is investigated through a sampling algorithm that averages the true values from the underlying ship performance simulation according to the parameter f_{ave}. The underlying performance is defined by a maximum temporal resolution of 4 samples per hour. The value, f_{ave} is set to 96/day when simulating the CM dataset and 1/day to simulate a NR dataset.

The former influence is affected by the actual number of observations that will remain in the dataset following any data pre-processing such as the removal of outliers, time spent in port, data acquired in shallow water, data outside the normal speed range and when the absolute difference between STW and SOG is greater than 1. This standard filtering criteria is defined in section 4.2. Table 37 indicates how the number of observations that remain, N, reduces with progressively more stringent filtering applied and therefore a progressively more simplified normalisation model is formulated. This is to understand the relative significance of filtering versus normalisation on the overall uncertainty, see section 9.4.3.

The days spent at sea depends on ship type and size. Figures from the 2014 IMO GHG study (Smith, Jalkanen et al. 2014) show that for 2012 the average number of days at sea for all sizes of tankers, bulk carriers, containers and general cargo ships was 53% of total time. This is similar to the results found from the 397 day CM dataset analysed in this study and shown in Table 37, (51.7% at sea days).

	Simulation		Fully Filtered		Partially Filtered Dataset		Full Normalisation	
	N	%	N	%	N	% of all	N	% of all
All data (397 days, 1 sample / 15mins)	8640 (90 days)	100	38112	100	38112	100	38112	100
At sea data (after exclusion of outliers/missing data)	4320	50	19717	51.7	19717	51.7	19717	51.7
Wind speed < BF4	2916	33.8	12881	33.8	12881	33.8	NA	NA
Loaded draught	1458	16.9	5830	15.3	NA	NA	NA	NA
Power > 0.3 MCR			5660	14.9	8643	22.7	12983	34.1
Abs(SOG-STW)<1knot			5125	13.5	7344	19.3	10714	28.1
Sea depth > criteria			3622	9.5	5636	14.8	8240	21.6
Speed within limits			3617	9.5	5625	14.8	8228	21.6

Table 37: Effect of filtering on the number of observations of a CM dataset

The simulation column of Table 37 builds confidence that the underlying ship operational and environmental model is realistic as the proportion of values that are filtered in the simulation are very similar to those of the actual CM dataset. The final four filtering criterion are not explicitly included in the simulation (their effects are included in the assumed daily speed variation). In the simulation, in order to accommodate for this in terms of sample size and to achieve realistic values for N, after the environmental and loading condition filter is applied, then a sample size is selected at random from the averaged data. The exact sample size is defined by a proportion of the remaining values. This is 62%, 44% and 42% for the fully filtered,

partially filtered and not filtered datasets, respectively. This reduces the datasets to the final size of the CM dataset used in this study; in which 9.5%, 14.8% and 21.6% of the observations for use in the ship performance quantification remained. This proportion is assumed to be equal for NR and CM datasets.

N is a proportion that is ideally representative of a 'normal' ship operation, however it is recognised that N is route dependent and may vary between periods according to a ship's trading pattern. For example, the ship from which this dataset is sampled is operating inside the Persian Gulf and as it is a VLCC with a relatively large draught then the sea depth filter may be more significant than it otherwise would be. The sensitivity of the output uncertainty to N is investigated in section 9.4.1. N is also affected by the evaluation period length, t_{eval} and is scaled up proportionally to this time period. The effect of the evaluation period length and the sensitivities of the overall uncertainty to this is investigated by evaluating both 90 day and 270 day periods.

9.2.3 Model Uncertainty

The model parameter uncertainty, as opposed to model structural uncertainty, is the main source of uncertainty investigated in this section. Model form/structural uncertainty is notoriously difficult to quantify because the truth of the parameter interactions is not exactly known. In the simulation, the expected power and the measured power are based on models of the same independent variables with the same log-linear or log-log relationships, therefore the model form is assumed to be correct and no model form error is included or investigated.

The model parameter uncertainty arises from a divergence between the model parameter estimate and reality; this is comprised of model parameter bias and precision. Bias may originate when the speed coefficient is for example estimated to be 3.0, but in reality it is 3.5 or 4.0. Model parameter bias may also arise in the case of an entirely statistical model, if the calibration period is too short and does not contain the required range and variation between variables (although this possibility is minimised in the pre-processing step) or because the effect of time is not independent of the other model independent variables and therefore comparing an

evaluation period to a baseline calibration period produces errors. Model parameter precision originates from the dynamic nature of the ship's propulsion system and the likely fluctuation around the estimated coefficient due to effects of ship motions, drift or rudder angle. For example, this will mean that the coefficient of speed will not have the exact effect on power as identified by the OLS estimate; instead it will fluctuate around this depending on the environmental conditions. These effects are not adequately represented by the error term in the regression equation because they correlate with the independent variables themselves. The effects of model parameter precision and bias on the magnitude and uncertainty of the overall PI is examined in the sensitivity analysis.

	Continuous Monitoring
Sensor uncertainties: - Speed (2σ) - Draught (2σ) - Power (2σ)	 $\pm$ 1% $\pm$ 2 m $\pm$ 3%
Sample averaging frequency, f_{ave}	Quarter hourly
Sample size, evaluation period length, t_{eval}	90
Sample size, due to filtering, N	Varying (200:1400)
Model parameter uncertainties	Model dependent, see next sections
Human error, (2σ)	NA

Table 38: Simulation input parameters for the top down vs bottom up comparison

This section details how the model parameter uncertainties are calculated and compares the results to a continuous monitoring dataset. This is a good stage to attempt to validate the method because the continuous monitoring dataset has limited complicating and highly variable factors of human error and sample averaging uncertainties which may induce differences between data derived uncertainty and the simulation uncertainty. A comparison here therefore makes it possible to isolate the model parameter and sensor uncertainties. The input parameters for the simulation

are summarised in Table 38. The model parameter uncertainty inputs form the basis of the discussion in the next sections.

As described by section 9.1.2, in order to fulfil the second aim of this chapter, there are three types of model investigated depending on the level of filtering applied. The fully filtered model (FF) is simply an assumed cubic relationship between ship speed and power/fuel consumption. The partially filtered model (PF) includes a relationship with draught, and the non-filtered model (NF) relates the power/fuel consumption to all operational and environmental variables. For the latter two models, the model coefficients come from the statistical model of the calibration period.

The precisions to which the model estimates the coefficients are defined by the confidence intervals of the model coefficients as detailed in the proceeding section. The bias is assumed to be zero because of the pre-processing checks performed on the data as detailed in chapter 4; the sensitivity of the final result on this assumption is quantified in the sensitivity analysis. For the fully filtered model, the calculation of model parameter bias is detailed in the next section.

Model Parameter Bias

For the FF model, it is difficult to properly estimate the speed-power relationship for a specific loading condition in the calibration period because the operational speed range during that period is likely to be low and there is insufficient data to estimate the speed coefficient to a degree of accuracy. Instead, to estimate the speed-power relationship, it is assumed that either sea trial data is used or a straightforward cubic assumption made. In this scenario, model parameter bias and precision are both significant. To understand this further, a sample of 340 ships of different types was analysed using the theoretical model proposed in chapter 5. The relationship between speed and power in the loaded, at-design condition was defined and therefore the constants in the following expression were evaluated for each ship:

$$P = aV^n + e$$

Equation 43

Where *a* and *n* are constants to be evaluated and *e* is the error term. This produced the plot of Figure 61 which indicates how the speed exponent varies between ship types. These results are specific to these 340 ship types studied, however it is worth noting that this tallies with the literature which also indicates an exponent of 4 for container vessels, and 3.2 for bulk carriers and tankers (MAN 2011). This also builds confidence in the theoretical model of chapter 5.

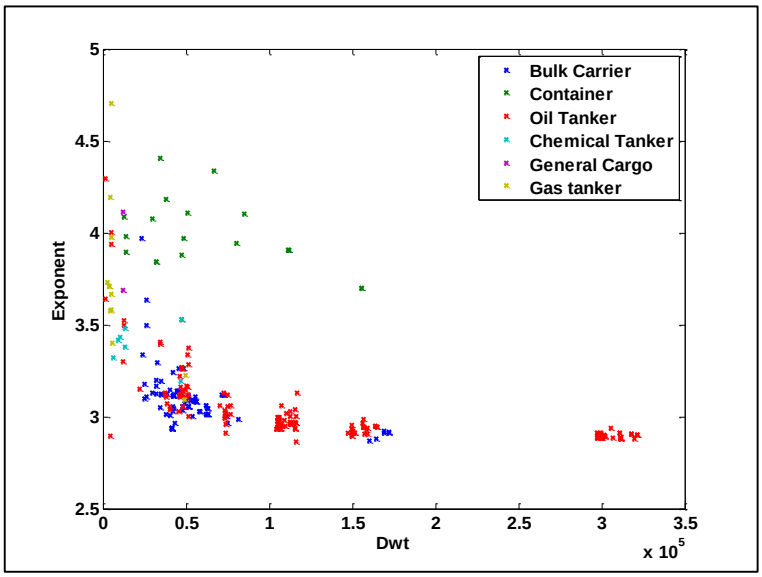


Figure 61: Variation in speed exponent between ship types and with deadweight (Dwt)

If the exponent of speed was assumed to be cubic then, depending on the ship type, this may induce an error of between 3% and 30%.

Ship type	Theoretical Speed Exponent	% Error
'Bulk Carrier'	3.1	3.0
'Container'	3.9	30.0
'Oil Tanker'	3.0	0.9
'Chemical Tanker'	3.4	13.2
'General Cargo'	3.9	30.1
'Gas tanker'	3.7	23.0

Table 39: Average theoretical speed exponent for the different ship types and percent error relative to the cubic assumption

For the comparison with the continuous monitoring dataset the model parameter bias was set to 6.7%. This was ascertained by comparing the cubic with the exponent as found by the theoretical model, which was found to be 2.80. If sea trial data were to be used to inform the power speed relationship then it is likely that the bias would be lower because the data input is specific to the ship being analysed.

Model Parameter Precision

The parameter estimates are formed by the OLS on the calibration period data. It is possible to understand the extent to which the model parameters fluctuate by examining their confidence intervals. The confidence interval describes the confidence with which we can say that the coefficient represents the true effect of the explanatory variable on the dependent variable, given the data. Therefore these confidence intervals represent model parameter precision; they can be automatically output from R, usually at the 95% level.

These are transformed to represent one standard deviation as a percentage of the coefficient itself. This is repeated for each model and the results printed in Table 40.

FF	**Speed**					
% SE	0.74					
PF	**Speed**	**draught**				
% SE	0.81	0.90				
NF	**Speed**	**draught**	**H_S**	**$V_{wave,L}$**	**$V_{wind,T}$**	**$V_{wind,T}^2$**
% SE	1.20	1.13	3.99	4.74	-22.21	13.78

Table 40: Standard deviation of the coefficients as a percentage of the coefficient for each model

It can be seen that the parameter fluctuations increase as the number of parameters of the model increase; this is because the level of filtering reduces and the additional parameters of the model do not fully represent the increasingly complex environmental effects. Therefore the dynamic nature of the ships' propulsion system causes more severe fluctuations in the model parameters. Also, the number of observations increases, however the number of explanatory variables also increases

and this uses up degrees of freedom meaning that the OLS estimator becomes less efficient.

Model Parameter Uncertainty Comparison

The method validation experiment was executed by comparing the bottom up simulation results to the top down, continuous monitoring dataset overall uncertainty results. The simulation input parameters are as defined in Table 38. An initial comparison between the data derived confidence intervals and those of the simulation are presented in Figure 62.

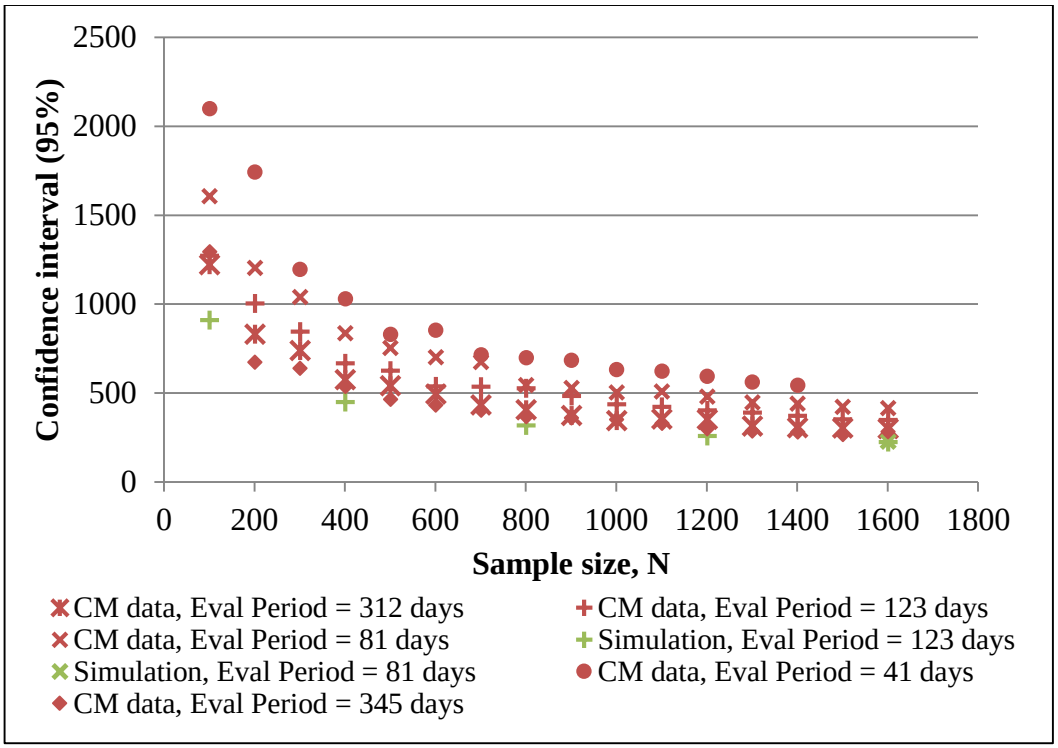


Figure 62: Confidence interval of the linear trend for the CM data and the simulation

The confidence interval of the actual data reduces as the evaluation period increases for the same sample size (x-axis). This was not the case for the simulation where the evaluation period did not influence the confidence interval. This has highlighted that in the CM data, the trend in the performance indicator is not a perfectly linear trend and therefore the overall uncertainty result (uncertainty in the performance indicator) from the CM data includes an uncertainty due to this deviation from assumed linearity. The top down uncertainty results therefore not only represent the model parameter uncertainty and the measurement sensor uncertainties but also the

deviation in the performance trend from linear. Consequently this is not an appropriate comparison metric.

It was found that, in this case, a sufficient evaluation period was required of approximately 312 days before the linear trend is assumed; that is, the top down and bottom up results match and therefore the top down results do not include an uncertainty associated with the non-linearity in the performance indictor trend. Or, it could simply be by coincidence that this is the time period that causes the uncertainty due to the non-linearity in the performance trend to exactly match the overall simulated uncertainty attributed to the instrument and model parameter uncertainties as input to the simulation. It is not possible to tell and so, again, the confidence interval of the trend is not a good metric for comparing the top down to the bottom up uncertainty. Further, this also highlights the necessity of using the proposed simulation in order to understand and decipher the relative significance of the various sources of uncertainty which cannot be accomplished using data alone when the actual performance trend is unknown.

The index used to represent the uncertainty of the performance indicator must be altered in order to compare the data to the simulation; one must be used which is not dependent on the unknown trend of the ship performance. It was found that the standard error of the mean (SEM) fulfils this requirement because the evaluation period does not influence the result, for the same sample size. The SEM is standardised to a percentage of the average power to remove the influence of the intercept. A comparison of the SEM between the simulation and the actual data derived SEM are shown in Figure 63.

As can be seen from the below plot, there is a negative correlation between sample size and the standard error of the performance indicator, as one would expect. The top down and bottom up method exhibit this same general trend which builds confidence in the use of the MC simulation.

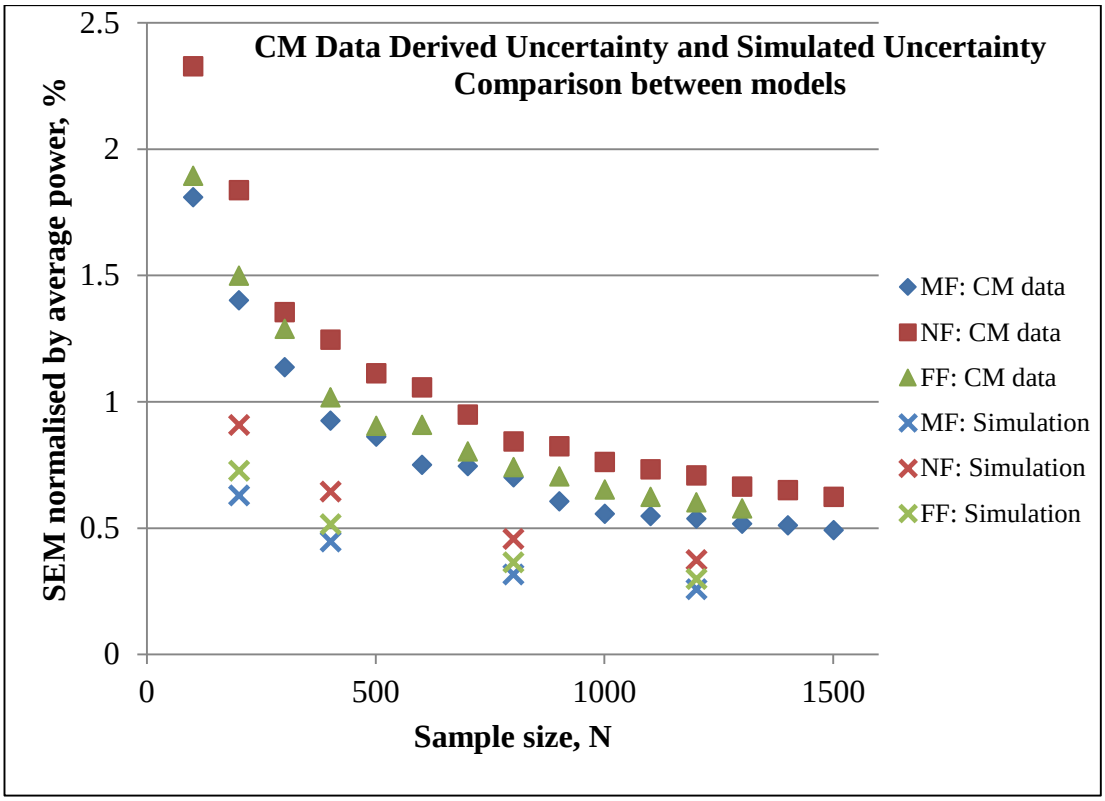


Figure 63: Comparison between simulation and actual data derived SEM and their variation with sample size.

Generally, the simulation exhibits lower SEM and therefore a conservative uncertainty estimate. The discrepancy may be due to the assumed sensor precisions in the simulation which may diverge from the actual precisions of the onboard sensors; the actual onboard sensors may exhibit some bias or drift which is assumed to be negligible in the simulation. The model parameter uncertainties may also be underestimated. Model bias for example is assumed to be negligible in the NF and PF model but may in reality be present due perhaps to some operating or environmental condition present in the calibration period that is absent in the evaluation period, or vice-versa. Spurious factors or random events such as seaweed or other debris attaching to the propeller will be impossible to include in the model (although this will probably have minimal impact). The sensitivity of the overall uncertainty to the model parameter or instrument bias is investigated in further detail in the sensitivity analysis. Finally, although standardising the SEM by presenting it as a percentage removes the effect of the intercept, the SEM itself will still be affected by the absolute power which is greater in the simulation relative to the

dataset. However, so long as the trends are comparable then it will be possible to use the simulation to discuss relative scenario characteristics.

The order of increasing SEM according to model type; PF-FF-NF that was exhibited from the results of the CM dataset is also reflected in the simulation, further validating the use of this method. This trend is discussed in further detail in the results section of this chapter.

9.2.4 Human Error

Human error is introduced in the measurement of wind and waves, in the measurement of fuel consumption, such as through using tank dip measurements and at any other stage as described in section 4.3.4. For the environmental measurement uncertainties, an assumption is made according to literature references as discussed in section 9.2.1. The human error associated with the fuel consumption measurement and the general human error introductions are not included; they are briefly discussed qualitatively when drawing conclusions.

It is extremely difficult to quantify the effect of human error on the overall uncertainty in the performance measurement. Probably, a continuous monitoring dataset would need to be obtained alongside a noon report dataset for the same ship over a reasonable time period. Even in that case however it would be difficult to separate the fluctuation in engine performance and the engine degradation itself from human error in the fuel consumption measurement because noon reports typically only collect fuel consumption data (as a tonnes per day average value) and not shaft power measurements. In order to convert to shaft power from the fuel consumption data, an assumption regarding fuel quality and SFOC would have to be made, which would also have to consider how the engine performance varies with engine load. There would be a degree of unknown, dynamic uncertainty associated with this.

9.3 Uncertainty Quantification Experiments

Having demonstrated the credibility of the MC simulation method for estimating the uncertainty of the ship performance metric, the same MC simulation is deployed in

this section to carry out a series of experiments. These aim to fulfil the three aims highlighted at the beginning of 9.1. The input data acquisition parameters are defined in Table 41.

Sensor precision and/or bias
Sample averaging frequency, f_{ave}
Sample size (number of observations) due to filtering, N
Sample size (evaluation period length), t_{eval}
Sample size (sample frequency)
Model parameter precision and/or bias

Table 41: Data acquisition parameters to be defined as inputs to the MC simulation

Therefore each investigation below defines the specific parameter combination for each run in order to fulfil each aim.

First the experiments are described (Section 9.3.1 to 9.3.3), before the results of the experiments are presented and discussed (Section 9.4)

9.3.1 Input Parameter Sensitivity Analysis

Aim: Identify the most significant contributory factors to the uncertainty through a sensitivity analysis which considers both elemental bias and precision limits in the overall resultant uncertainty.

A local sensitivity analysis was conducted by altering the data acquisition parameters one-at-a-time and measuring the change in the overall uncertainty by computing partial derivatives of the output function (y) with respect to the input factors (x_i). The output function, y is the uncertainty of the measurand at time = t_{eval} (see section 9.1.3), the SEM is used for the sensitivity analysis. The input factor is one of the data acquisition parameters as defined in Table 41. Each factor is altered one-factor-at-a-time (OAT) and a dimensionless expression of sensitivity, the sensitivity index, SI_i is used to measure changes relative to the baseline.

$$SI_i = \frac{x_i}{y} . \frac{\partial y}{\partial x_i}$$

OAT analysis ignores interactions between input factors which may be relevant however it is a quick and straight forward method which makes it possible to find out which factors are important without making simplifying assumptions (A. Saltelli, K. Chan et al. 2008). The sensitivity index provides an efficient method to examine the slope of the output function in the parameter space for the baseline values.

The NR dataset is the baseline in order to highlight which are the most significant data acquisition parameters, and therefore how the highest investment returns can be achieved. The simulation for the sensitivity analysis is based on the partially filtered model.

The parameters of the sensitivity analysis are summarised in Table 42. Each row in the table indicates a separate simulation run with the baseline column indicating the 'normal' state of the variable. The parameter limit column defines the extent to which that variable is adjusted in measuring the sensitivity of the performance indicator to changes in this particular variable.

The sensor bias' (trials 1.1 to 1.3) are assumed to be zero in the baseline with upper limits as defined in section 9.2.1. There is not always the data or work in the literature available to define exactly what the potential bias might be, therefore for transparency, the same change in bias is assumed for each sensor. Trials 2.1 to 2.3 are concerned with the sensitivity to the sensor precision; it is assumed that it is possible to improve instrument and model precisions by 50% and that the human error correction can be reduced by 50%.

The sample size as calculated from the proportion of data filtered is a fixed variable as it is dependent on the operating profile of the ship and the prevailing weather conditions which are, for the purpose of this analysis, uncontrollable factors. It is possible to alter the sample size in three ways:

i. Increase the sample averaging frequency: increasing the frequency with which readings are taken from once daily to quarter daily, which might be

a possible expectation of increase in crew workload. An assumed constant proportion is filtered over the same time period and therefore the sample size increases.

ii. Increase the evaluation period length: Sampling frequency and proportion filtered remain constant

iii. Increase the proportion of filtered data: Evaluation period and sampling frequency remain constant

In each case the data acquisition parameter is altered by a magnitude that generates the same final sample size of 53 observations. The last trial (3.3) is difficult to achieve in reality because the ship's operational and environmental profile are significant to the proportion of observations that can be filtered out, which is therefore fixed. It may be possible to improve crew diligence to reduce the removal of outliers due to erroneous data entries. In 3.1 the sample frequency is increased however the proportion of observations that are assumed to be filtered out is kept constant (hence sample size automatically increases) so this test does not isolate the effect of sample averaging from sample size. Therefore, a separate trial to understand sample averaging significance is carried out (3.4) where the sample size remains constant but the averaging time period is increased from 15 minutes to 6 hourly. The proportion of filtered observations is then forced to decrease by randomly selecting a sample of the same size as in the baseline.

Model parameter precision is assumed to halve in the same way as sensor precision and bias is assumed to increase to 1% each.

As well as the above factors, the sensitivity analysis was also used to confirm that some of the underlying assumptions, particularly regarding operation and environmental profile, which are likely to be highly ship specific, do not alter the overall results by a considerable amount. Those parameters investigated are:

- Operational profile: Speed, draught, days spent loaded/ballast.
- Environmental profile: Daily speed variation, deterioration rate.

The effect of averaging daily draught variability is not included but assumed to be negligible.

	Baseline, Noon Report	**Parameter limit**
Sensor bias:		
1.1. Speed, %	±0	± 1.0
1.2. Draught, %	±0	± 1.0
1.3. Fuel consumption, %	±0	± 5.0
Sensor precisions:		
2.1. Speed (2σ) %	± 5.0	± 2.5
2.2. Draught (2σ) %	± 5.0	± 2.5
2.3. Fuel consumption (2σ) %	± 5.0	± 2.5
3.1 Sample size, sample frequency (samples/day)	Daily ss* = 13	6 hourly ss = 53
3.2 Sample size, evaluation period length, days (t_{eval})	90 ss = 13	360 ss = 53
3.3 Sample size, proportion filtered, %	14.8 ss = 13	59.0 ss = 53
3.4 Sample averaging frequency, f_{ave}	Daily ss* = 13	6 hourly ss = 13
Model parameter Precision (2σ):		
4.1 Speed, %	± 1.6	± 0.8
4.2 Draught, %	± 1.8	± 0.9
Model parameter bias		
4.3 Speed, %	0.0	1.0
4.4 Draught, %	0.0	1.0

Table 42: Data acquisition parameter input matrix for sensitivity analysis (*ss = sample size)

In the above sensitivity analysis the bias errors are treated as epistemic in nature and dealt with by interval analysis, as discussed in section 2.3.2 this may overestimate

the actual uncertainty and therefore the results are indicative and possibly conservative (over estimating uncertainty).

9.3.2 Noon Report and Continuous Monitoring Comparison

Aim: Compare and quantify the overall theoretical uncertainty of a noon report dataset with that of the continuous monitoring dataset.

All sensors are assumed to be calibrated and therefore exhibit zero bias, the sensor uncertainties in Table 43 refer to sensor precisions. The effect of this assumption is discussed in the results of the sensitivity analysis. The wind data for the continuous monitoring data set is assumed to come from onboard anemometer readings while the wave data is assumed to be model derived from hind cast data. This corresponds to the origination of the data in the continuous monitoring dataset that was used for this study. Also in accordance with this dataset, the ship speed in both the NR and CM dataset is assumed to be from the same speed through the water sensor and the draught data is assumed to be manually input for both CM and NR. The quoted precisions correspond to the 95% confidence interval (2σ) and their origin is explained in section 9.2.1.

The model parameters uncertainties are assumed to be the same for the CM and the NR dataset so as not to double account for other differences between the dataset (such as sample frequency) which are included elsewhere in the simulation.

This investigation is based on a moderately filtered dataset and a model of medium number of parameters (environmental conditions are filtered for but not draught). Therefore N represents 14.8% of all possible observations (see Table 37), which is 90 and 270 for NR and 8 640 and 25 920 for CM for 3 and 9 month evaluation periods, respectively.

	Noon Report		Continuous Monitoring	
Sensor uncertainties: - Wind speed (2σ) - Wave height (2σ) - Wind direction (2σ) - Speed (2σ) - Draught (2σ) - Power (2σ) - Fuel consumption (2σ)	 ± 1 BF ± 20% $\pm 10^0$ ± 1% ± 5 % NA ± 5%		 ± 1.24 knots ± 5% $\pm 3.26^0$ ± 1% ± 5 % ± 3% NA	
Sample averaging frequency, f_{ave}	Daily		Quarter hourly	
Sample size, evaluation period length, t_{eval}	90	270	90	270
Sample size, due to filtering, N	13	40	1279	3836
Model parameter uncertainties, (2σ): Speed, % Draught, %	 ± 1.6 ± 1.8		 ± 1.6 ± 1.8	

Table 43: Data acquisition parameter input matrix for NR and CM comparison

9.3.3 Model Assessment

Aim: Assess the statistical model using its associated model parameter uncertainties as inputs to the MC method for both the NR and CM datasets and compare these to the uncertainty achieved from a filtered dataset and simplified model.

The sensor uncertainties are representative of CM and NR datasets as in the previous experiment and consistent for all 4 of the experiments. The core differences between these three studies is the number of observations, N which successively increases from left to right in Table 44 and the model parameter uncertainties.

	Fully filtered, P-V model		Part filtered dataset, P-V-T model		Statistical model	
	NR	CM	NR	CM	NR	CM
Sensor uncertainties: - Wind speed (2σ) - Wave height (2σ) % - Wind direction $(2\sigma)^0$ - Speed (2σ) % - Draught (2σ) % - Power (2σ) % - Fuel consumption (2σ) %	NA NA NA ± 1 NA NA ± 5	NA NA NA ± 1 NA ± 3 NA	NA NA NA ± 1 ± 5 NA ± 5	NA NA NA ± 1 ± 5 ± 3 NA	±1 BF ± 20 ± 10 ± 1 ± 5 NA ± 5	±1.2 knots ± 5 ± 3.3 ± 1 ± 5 ± 3 NA
Sample averaging frequency, f_{ave} (/day)	1	96	1	96	1	96
Sample size, due to filtering, N	9	821	13	1279	19	1866
Model parameter precision, (2σ) Speed, % Draught, % Wave height, % Wave speed, % Wind speed, % Wind $speed^2$, %	± 1.5 NA NA NA NA NA		± 1.6 ± 1.8 NA NA NA NA		± 2.4 ± 2.3 ± 8.0 ± 9.5 ± 44.4 ± 27.6	
Model parameter bias Speed, %	1.0		NA		NA	
Sample size, evaluation period length, t_{eval}	90	90	90	90	90	90

Table 44: Data acquisition parameter input matrix for filtering and normalisation comparison

This is because fewer filtering levels are applied as the model used to normalise the data to the reference value gradually becomes more complex and involves an increasing number of parameters.

The hypothesis is that the greater number of observations will improve the overall accuracy in the performance indicator. Conversely however it might be that this increase in accuracy due to the reduction in sampling uncertainty is offset by an increased number of sensors which increase the measurement uncertainties and the increase in model uncertainty.

9.4 Results and Discussion

This section discusses the results with reference to the three investigations defined above. Further examinations are made in the next section to understand the practical implications of these.

9.4.1 Sensitivity Analysis

The results of the sensitivity analysis are presented in Figure 64; this displays the sensitivity indices for the precision of the uncertainty in the overall performance indicator. The bias in the result (the performance indicator) is investigated also and this is graphically represented in Figure 65. Overall it can be seen that sensor and model parameter precisions have a more significant impact on overall precision than sensor /model bias'. Sample size has a lower but significant impact, the meaning of these results for practical purposes are discussed through the results of the proceeding analyses and in the discussion of RQ2.

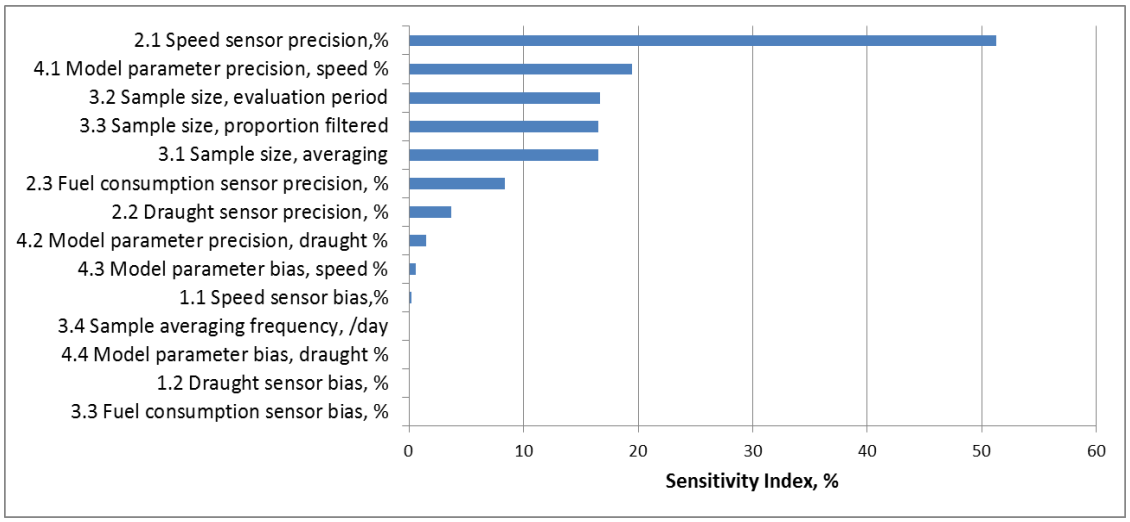


Figure 64: Sensitivity analysis results; the sensitivity of the performance indicator uncertainty according to changes to the data acquisition parameters

Clearly, the uncertainty of the speed measurement (both model parameter and instrument precision) is the most significant. This is as one might expect because this parameter has the single largest influence on the vessel power requirements; in the log-log model the coefficients of speed indicates that a 1% increase in speed will increase power by 3%, while the same percentage increase in draught will increase power by 0.04%. Of the sensors the order of increasing value to the ship performance measurement uncertainty is the draught gauge, fuel consumption and then the vessel speed sensor. This highlights the importance of investment in high precision speed sensor, and the criticality of using speed through the water sensor rather than speed over ground (which may cause a precision of 5%.) which will have dramatic consequences for obtaining meaningful information from the data. CM systems that augment core data acquisition with additional measurement of GPS, tides and currents would provide a means to independently calculate and verify the speed through water measurement from speed over ground. The results also show an influence of draught and shaft power measurement precisions. The draught measurement may be manually input, even in continuous monitoring systems, and therefore the overall uncertainty would benefit from investment in a draught gauge which would also record potentially significant variations during the voyage (including in trim) rather than recording the draught at the last port call. If fuel consumption is used as a proxy for shaft power, which may cause an inaccuracy of 5% (due to uncertainties in the fuel consumption measurement and conversion to

shaft power assumptions) then there could be a significant effect also on the overall uncertainty.

Sample size clearly has a significant impact; the overall uncertainty is inversely proportional to the square root of the sample size. This is irrespective of the method by which a larger sample is achieved which may either be by increasing the evaluation period, increasing the sample averaging frequency or decreasing the number of observations removed by the filtering criteria. A longer evaluation period may be inappropriate depending on the motivation for measuring ship performance. Increasing sampling frequency may be possible however it could be a significant burden on the time of the crew. Addressing the root causes of outliers (i.e. due to stuck sensors or human reporting errors) or data removal due to filtering may be possible. However, the latter is route dependent and may vary between periods according to a ship's trading pattern unless the filtering algorithm is changed because the normalisation model is able to incorporate a wider range of operating/environmental conditions. This is the root of the hypothesis that there is a trade-off to be made between the sample size and the model error which reflects the advantages of filtering over normalising. The evidence here corroborates that possibility and this is further explored in the next section.

The other sampling effect is sample averaging frequency; gradually increasing the sample frequency from 1 per day to 12 per day (while controlling for sample size) shows a very gradual improvement in parameter uncertainty. However, the difference is so small as to be practically negligible to the overall uncertainty, as demonstrated by the sensitivity index.

As one would expect, bias (model parameter and sensor) is of lesser concern to the precision of the overall PI. This is because the bias is assumed to be present in the calibration period and in the evaluation period and because the performance indicator is a reflection of the difference between the two. There is a small change identified because the increased V_{meas} is compounded by the cubic relationship which causes ΔP to drift and therefore decreases the precision in the linear trend line of the PI (see next section). However this effect is of relatively negligible significance. The model

parameter bias is more significant than the sensor bias for the same parameter, which is the converse to the model and sensor precisions.

Generally, the precision of the data acquisition factors (model and sensor precisions) studied have a small effect on the absolute mean of the performance indicator (the bias component of the overall uncertainty). Bias however, be it model parameter or instrument bias, does affect the PI itself and is more significant to the bias component than precision; this is observed from Figure 65. The reason for this is that uncertainty due to variance grows much more slowly than uncertainty due to ignorance (bias) because random variations partially cancel while bias accumulates linearly. Systematic errors in one variable are unlikely to cancel out because of systematic errors in another variable (Ferson and Ginzburg 1996). Of particular significance are the speed measurement bias and the sample averaging frequency.

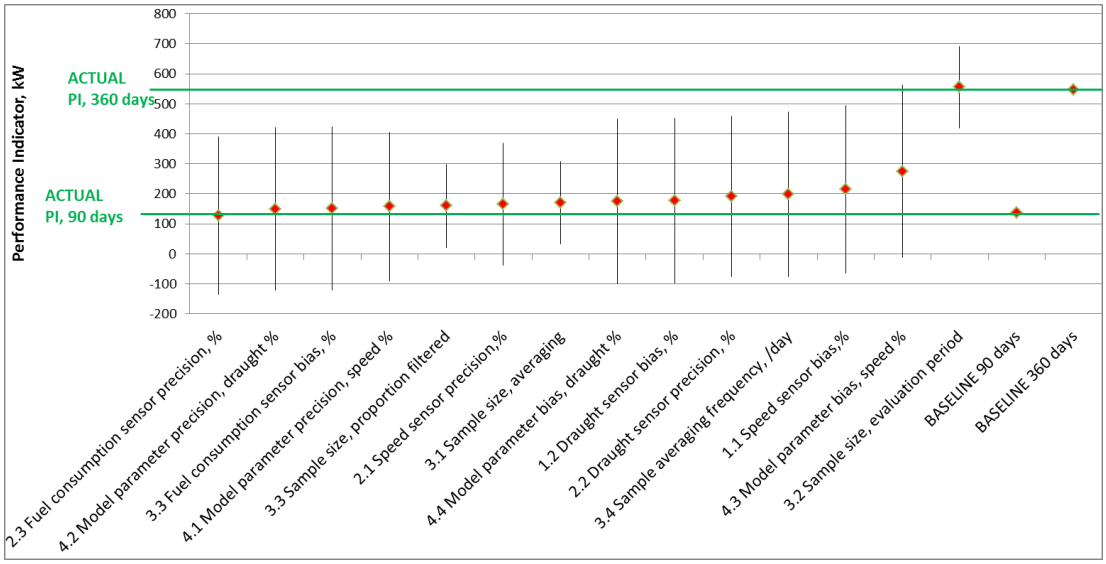


Figure 65: Effect of changes to various data acquisition parameters on the performance indicator

The effect of increasing the STW sensor bias from 0 to 1% is to increase the average performance indicator value from 164 kW to 216 kW, an increase of 32%. The graphic is indicative only since the exact magnitude of the effect is to some extent dependent on the operational profile but it is clear that in cases of sensor bias and model parameter bias, the actual underlying deterioration trend is difficult to identify. Clear also from the error bars in Figure 65 is how results based on daily averaging for a short 90 day time period may be inconclusive; if for example, speed sensor

precision is 1% then it might not be possible to conclude (with 95% confidence) if the ship's performance has improved or deteriorated.

Interestingly, the sample averaging frequency, when controlling for sample size, is significant to the calculated PI value. This is because the daily environmental fluctuations are not captured at low sample averaging frequencies; averaging a range of non-linear speeds will cause power due to higher speeds to be mistakenly attributed to deterioration in ship performance that in reality is not present and therefore an increase in the PI for lower sampling frequencies. This is shown in Figure 66 where the PI decreases as sample frequency increases. The actual influence on uncertainty is of course a function of the input underlying operational profile; the greater the assumed operational speeds, the more significant the cubic effect and the greater the bias. This is also shown in Figure 66 for two different operational profiles (vessel speed).

The fact that the sample averaging frequency is more significant and appears to exert a greater bias than the draught sensor bias or model parameter bias may have implications on the most beneficial route of investment (time and resource). Investment in more frequent measurements (4 time per day for example rather than once) instead of investment in high performance sensors may be preferable; this may be more relevant for higher speed ships. However, this depends on how the results of the sensitivity analysis translate to reality and ultimately how these uncertainties propagate through to the conclusions that can be drawn regarding ship performance and fuel consumption. This sets the scene for the proceeding analyses in the next two sections which look at what these mean in reality and take a more holistic view of the collective data acquisition strategy

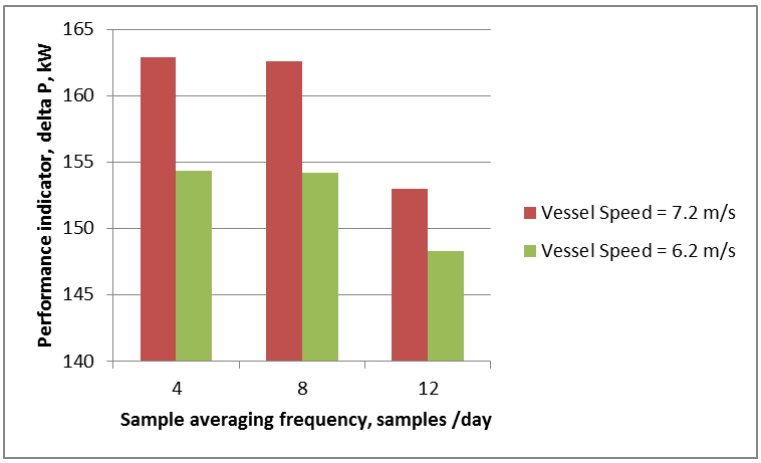


Figure 66: Uncertainty in the performance indicator as a function of the operational profile (ship speed)

The second part of the sensitivity analysis is to look at how the assumed underlying operational profile influences the absolute uncertainties. The previous example has shown how a higher vessel speed magnifies the effect of sample averaging bias at higher speed. The operational speed also affects the absolute value of the SEM (see Figure 67), this was highlighted when making a comparison with the top down method in the model parameter uncertainty quantification section. In reality, although to a lesser extent, the assumed draught, the assumed number of days in loaded and ballast condition and the assumed rate of performance degradation will also affect the overall uncertainty. This is kept in mind when drawing conclusions; however it is assumed to be of little overall significance because, apart from the input vessel speed, the effects of the other operational and environmental profile are very small. The speed profile will affect all model types and both CM and NR DAQ strategies in the same way and therefore should not affect the overall conclusions because they are based on relative differences between strategies rather than absolute values. When translating the results into fuel consumption uncertainties the results must be taken to be indicative rather than definitive.

Sometimes a combined uncertainty is calculated from the sum of squares of the precision and bias limits, which seems like a logical representation however there is no mathematical basis for this formula. As discussed in section 2.3.2, it is sometimes argued that the two should not be mixed together as the result would be meaningless (Morgan and Henrion 1990), although this depends on the application of the analysis.

It is assumed in the remaining analysis that the sensors are calibrated and the bias is negligible.

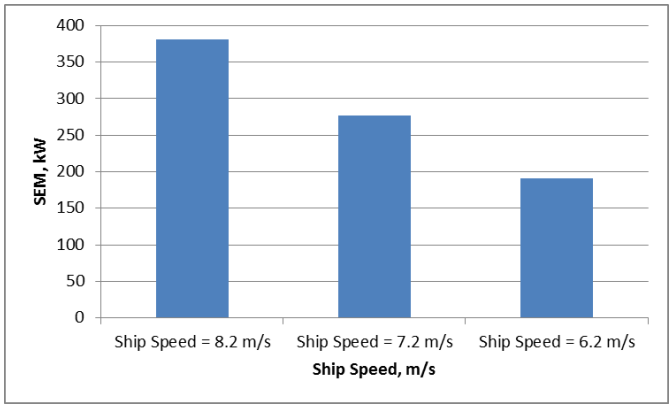


Figure 67: Effect of the operational speed on the SEM

9.4.2 Noon Report and Continuous Monitoring Comparison

The results from the comparison of the uncertainties from the two data acquisition systems are presented in Figure 68.

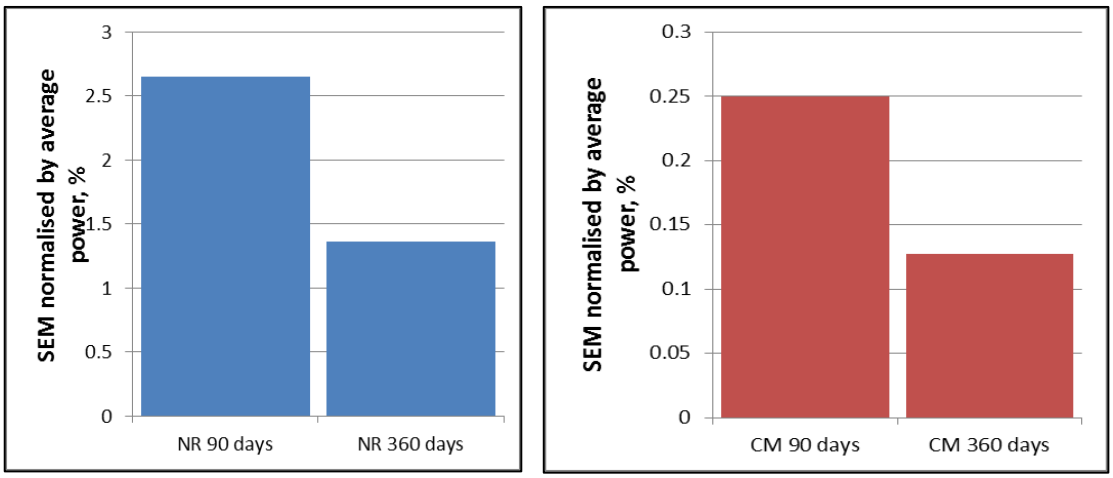


Figure 68: Noon report and continuous monitoring comparison results

The axes identify the ten-fold improvement in uncertainty achieved using a continuous monitoring set relative to a noon report dataset, even without including the likely significant effects of human error. Some of this improvement is due to the improved precision of the power measurement which, as identified in the previous section, has a high sensitivity index. However, if the simulation is repeated using the NR input parameters over 90 days but assuming a power sensor precision equal to

that of the CM input parameters then the SEM (normalised by average power) is calculated as 2.4%. This improvement is still not sufficient to match the uncertainty of the CM dataset. Most of the improvement in the CM dataset is due to sample size; unsurprisingly the 1279 observations of the CM dataset achieve a significantly improved SEM relative to the uncertainty that yields from only 13 samples over 3 months of the NR dataset. Due to the low sampling frequency of the NR dataset even over almost a year (360 days), only 53 samples are likely to be collected and the resultant uncertainty is still approximately 5 times that of the CM dataset.

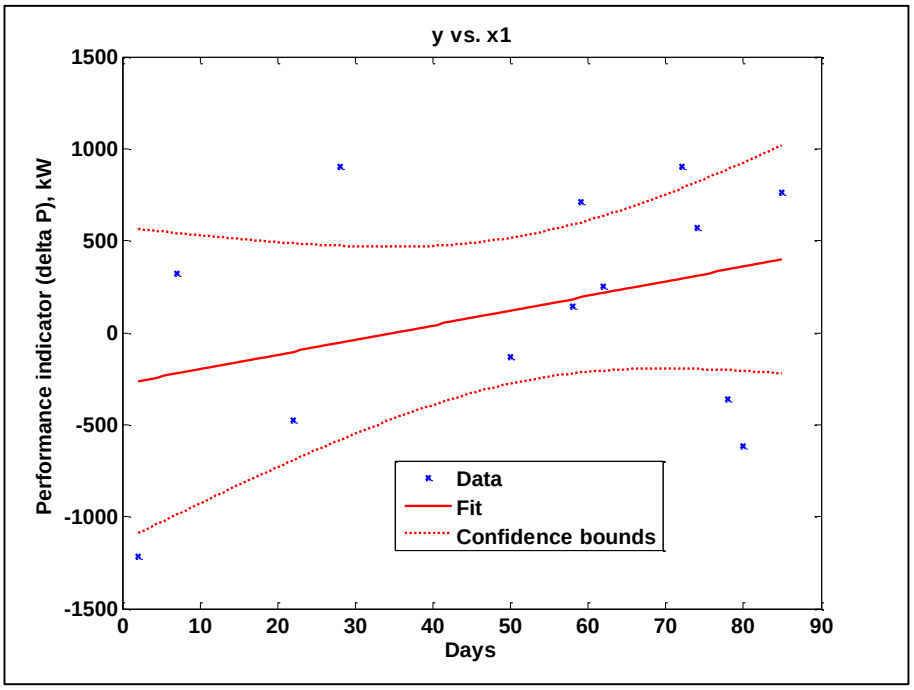


Figure 69: Performance Indicator trend over 90 days from the simulation, NR DAQ inputs

What do these results mean in reality? As discussed in chapter 2 the required level of uncertainty is dependent on the motivation for measuring ship performance in the first instance. Most vessel improvement technologies claim to deliver a percentage fuel consumption saving. Therefore the next section converts the uncertainty associated with the PI to the fuel consumption uncertainty, because of the relationship between the uncertainty and the underlying vessel profiles and the ship size and fuel consumption the following estimates should be taken to be indicative rather than definitive.

In the simulation the ship performance is set to deteriorate at a rate of 7.8%/annum, the PI trend for one iteration over 90 days is presented in Figure 69.

The performance indicator is defined as the coefficient of the linear trend and in the example above it equals 8 kW/day; a different value is calculated at each MC iteration. The output PDF for the PI over all MC iterations is normally distributed about 1.57 kW/day and the confidence interval of the PI is 32 kW/day. Therefore, given the uncertainty relating to the data collection and processing inputs, the actual normalised power, after 90 days, could have been measured as increasing to 13,624 kW or decreasing to 12,202 kW (see Figure 70).

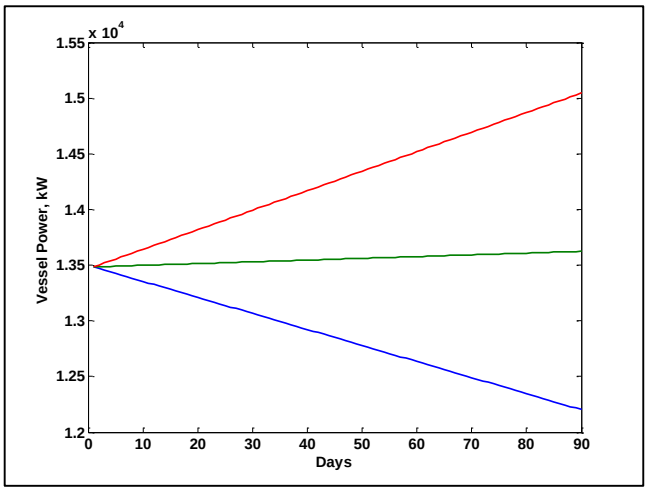


Figure 70: Normalised vessel power and 90 day confidence interval at the 95% level, simulation with NR inputs

From the confidence of the PI measured in kW/day, a crude fuel consumption transformation which assumes SFOC to be a constant (200 g/kWh) leads to a possible 90 day total fuel consumption of 5,855 ± 307 tonnes, or 5,855 ± 5.24 %. On this basis, it would be difficult to identify and confirm that a fuel saving of 3% had been made, however if attempting to prove that a 10% fuel saving had been achieved would be possible. This is an interesting finding because it shows that noon reports may be useful in some instances (such as when the effect of a retrofit technology capable of a large saving in fuel wanted to be observed). Although it is worth remembering that the potentially significant uncertainty due to human error and any severe engine degradation or performance fluctuation is not included in the uncertainty inputs in the simulation.

Figure 71 presents the performance indicator trend over 90 days for a simulation based on the CM DAQ inputs.

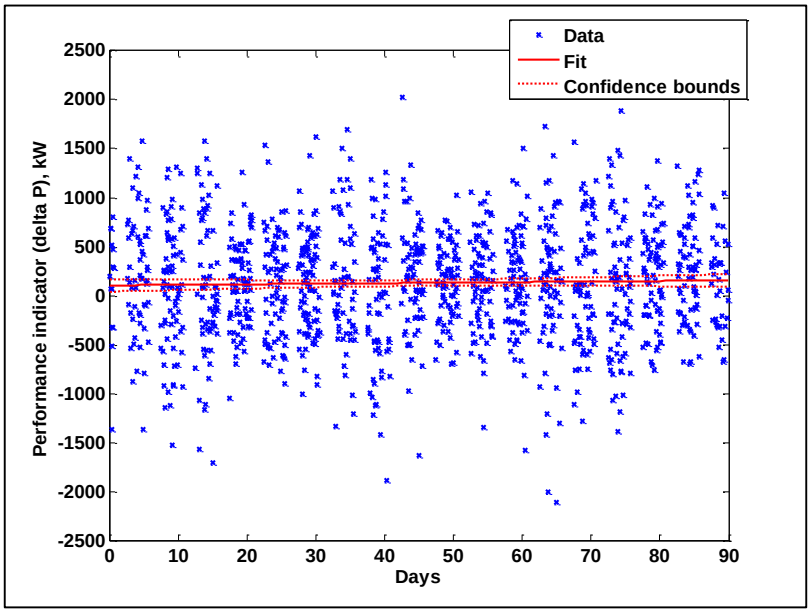


Figure 71: Performance Indicator trend over 90 days from the simulation, CM DAQ inputs

The output PDF for the PI over all MC iterations is normally distributed about 1.51 kW/day and the confidence interval of the PI is 2.52 kW/day. Therefore, given the uncertainty relating to the data collection and processing inputs, the actual normalised power, after 90 days, could have been measured as increasing to between 13,507 kW and 13,731 kW, see Figure 72.

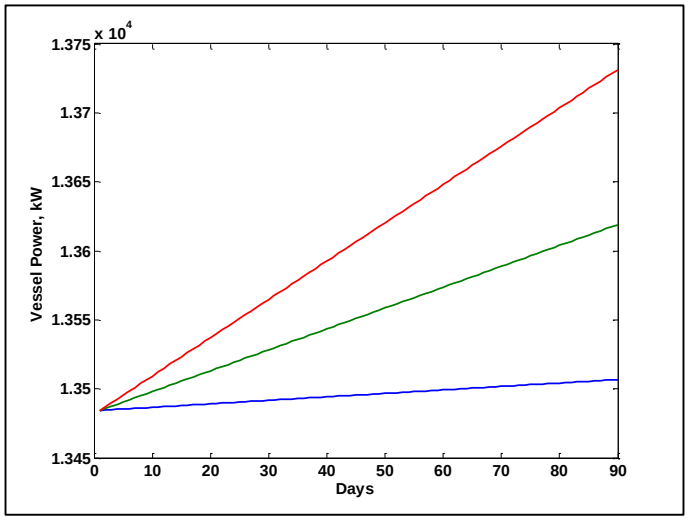


Figure 72: Normalised vessel power and 90 day confidence interval at the 95% level, simulation with CM inputs

In terms of fuel consumption, over the 90 days this equates to 5,854 ± 24.26 tonnes or ± 0.41%. This demonstrates the advantage of the CM acquisition system and its ability to identify fuel consumption savings of a very low order.

9.4.3 Model Assessment

The impact of the level of filtering and the normalisation model complexity on the overall uncertainty was studied both for a simulation based on a CM dataset and one based on the NR dataset. The results are presented in Figure 73.

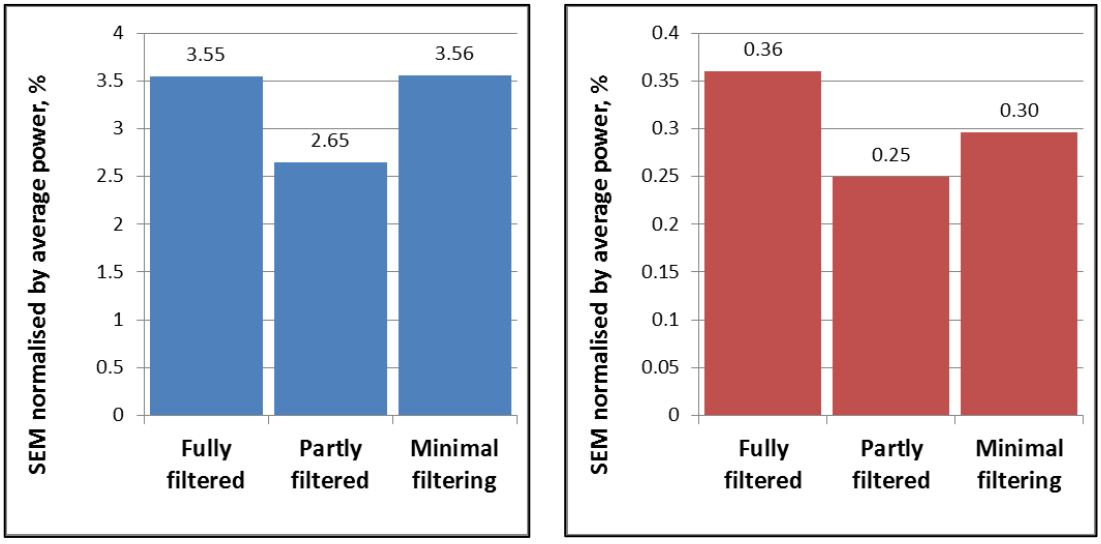


Figure 73: Comparison between model complexities on ship performance measurement uncertainty for CM (right) and NR (left) based simulations

The charts above show that for both data acquisition strategies (NR and CM), the optimum uncertainty is achieved for a model based on speed and draught only (partly filtered). This means that the improvement in uncertainty as a result of an increased sample size due to minimal filtering is more than offset by the combined, negative effects of model parameter uncertainty associated with the extra model variables and the sensor uncertainties associated with collecting the data for the extra model variables. However, this is not the case when comparing the speed-power model to the partially filtered model when the inclusion of draught is found to be preferable over a fully filtered model that only considers the speed-power relationship, despite the additional measurement and model uncertainties that comes with the addition of the extra variable. This is firstly because of the inverse square law of the effect of

sample size on uncertainty, meaning that at lower absolute sample sizes (in the case of going from the FF to PF model) the improvement to uncertainty as a function of sample size is more significant than the effect on uncertainty of changes in larger samples (i.e. going from PF to NF) and therefore outweighs the model parameter uncertainties, especially since, as seen in the sensitivity analysis, speed has the highest uncertainty associated with it relative to draught. Secondly, the fully filtered model is the only one to have a model parameter bias effecting one of the variables in the baseline. As seen by the sensitivity analysis the effect of a model parameter bias on the speed variable is small but not completely insignificant. It was found that the removal of this bias in the simulation improves the fully filtered model uncertainty and so the partially filtered model remains to be the optimum in terms of overall uncertainty reduction.

This effect of sample size is confirmed by running the simulation for the fully filtered model but assuming the same sample size as a partially filtered model which, for the NR dataset reduces the SEM to 2.99% which is closer to the SEM of the PF model.

This is particularly the case for the NR simulation where the uncertainty due to a minimally filtered model is very similar to the uncertainty due to a fully filtered model. This is quite significant because there is significantly more resource (time, money and effort) that would be required to a) collect the data required to inform a more complex model and b) to collect the knowledge required to develop such a model. Although, it should be noted that this may be negated by using an improved model to reduce the model uncertainty such as the hybrid model presented in Chapter 7 (see the discussion of this chapter). Alternatively installing more accurate, better maintained or more regularly calibrated sensors to reduce the measurement uncertainty would be beneficial. This is highlighted by the improvement in the CM, minimally filtered model's uncertainty compare to that from the same model based on a NR DAQ simulation. Deterioration in uncertainty of the NF model relative to the PF model is larger for the NR dataset relative to the CM dataset and this can largely be apportioned to the improvement in sensor accuracy for the CM sensors and in particular the precision of the environmental measurements. For example, the NR dataset is based on the resolution of the Beaufort scale instead of m/s as recorded

by automatic CM sensors. This is corroborated in Figure 74 which shows the NR baseline and the NR baseline with the instrument precisions overwritten with CM sensor precisions and the significant improvement in the minimal filtering model. The main source of sensor imprecision for the minimally filtered model is the wave height, wind speed and wind direction, this demonstrates that for certain applications noon reports may be sufficient for the required level of application if simple improvements were made to the environmental measurements, which may be less costly than installing an entire continuous monitoring system. This kind of trade-off is studied in terms of fuel consumption uncertainty in the next section.

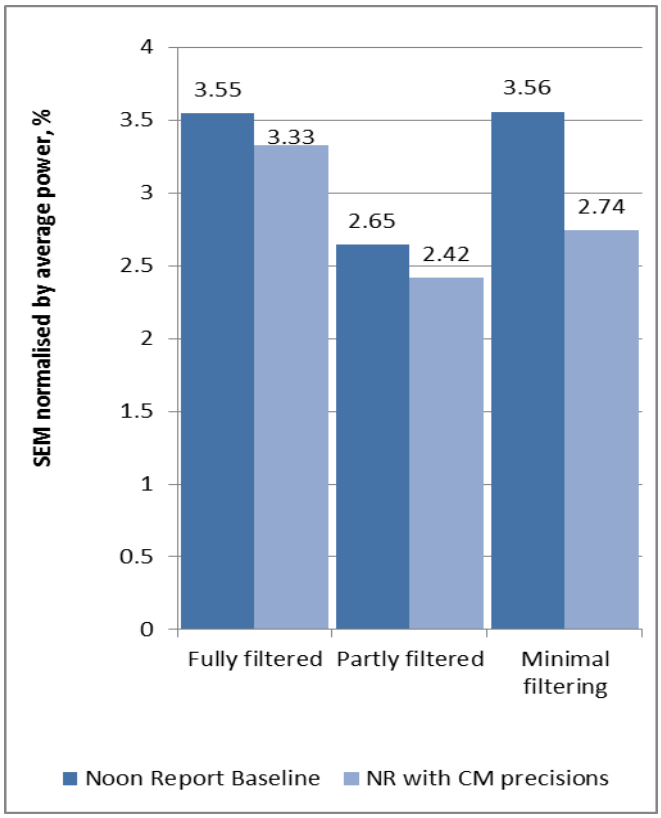


Figure 74: Effect of sensor precision improvements on the NR baseline for different models

9.5 RQ2: Discussion

The analysis in this chapter has raised a number of questions associated with the optimum model (of the models investigated), the sample size and the parameter uncertainties. There appears to be trade-offs possible between these factors and while a NR, low sampling frequency method appears to exhibit much higher uncertainties, when relating this to a fuel consumption uncertainty metric there is some useful

information to be gleaned from the NR dataset. This section brings these results together into a practically meaningful demonstration of how these DAQ parameters and models interact and, depending on the application and required accuracy, then what might be the optimum configuration. The absolute quantifications of uncertainty are indicative because of the dependence on the assumptions of the operational profile, however the conclusions generally relate to relative uncertainties which provide valuable comparisons. Human error is also not included but discussed in qualitative terms.

9.5.1 Sample Size and Sensor Precision

Figure 75 shows that, for a 90 day sample, the error in the fuel consumption is up to 35% for a very small sample size and a high model parameter uncertainty. This rapidly decreases as the sample size increases and at the sample sizes that are likely for a continuous monitoring set (~1000+ observations) then the overall uncertainty in the fuel consumption measurement is very low. At this stage the improvement in the overall uncertainty as the sensor accuracies increase is marginal; there are diminishing returns in sensor quality investment. This is shown in the right hand plot of Figure 76 where the uncertainty increases from 1% to 0.4% over the range of sensor uncertainties. At small sample sizes however the overall uncertainty can be improved to as low as 20% from 35% by reducing the sensor uncertainty (left hand plot of Figure 76). 20% however is still not practically useful for many purposes; if the aim is to detect a more subtle change in ship performance such as the application of a new coating which may only introduce a 3% improvement then this DAQ configuration will not provide conclusive results.

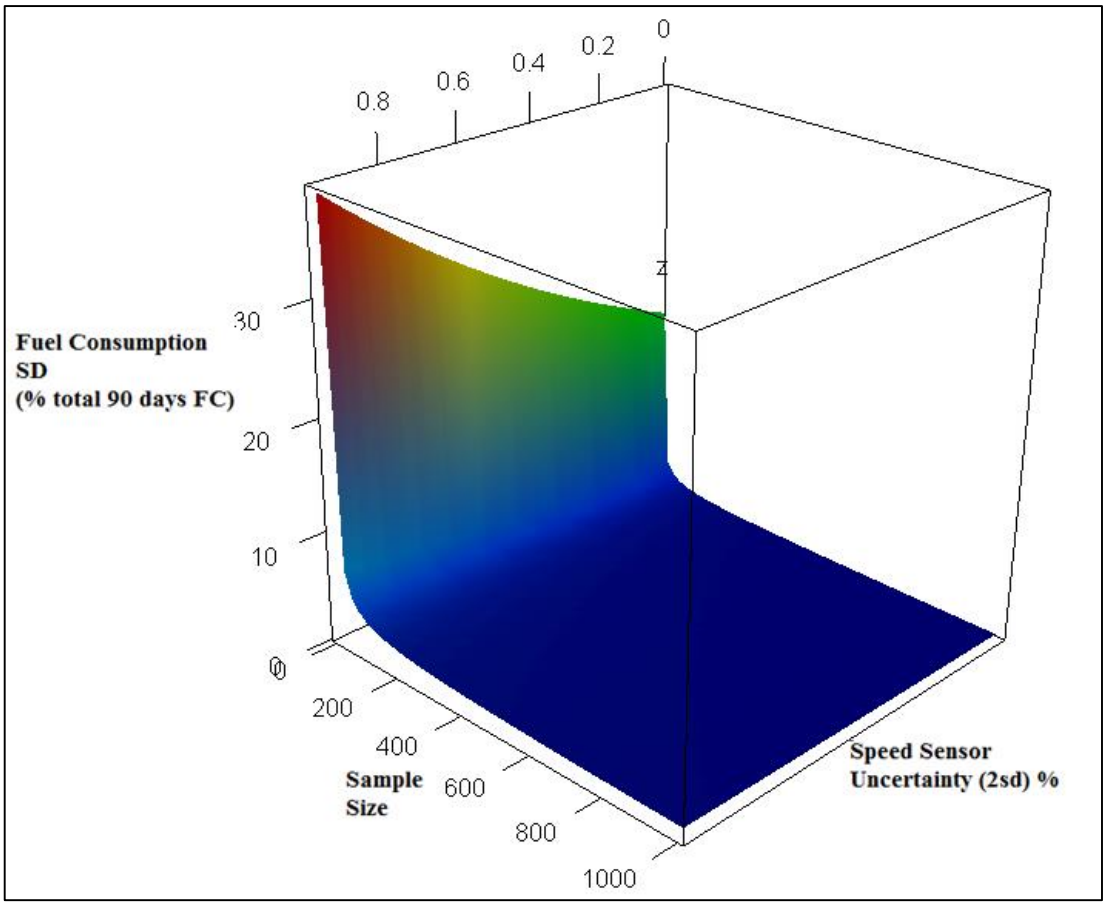


Figure 75: Sample size, sensor uncertainty (represented by the speed sensor uncertainty) and fuel consumption uncertainty (as a % of the 90 day total). Colour scale is proportional to fuel consumption uncertainty where the colour scale is calibrated to the overall range of the resultant fuel consumption uncertainty.

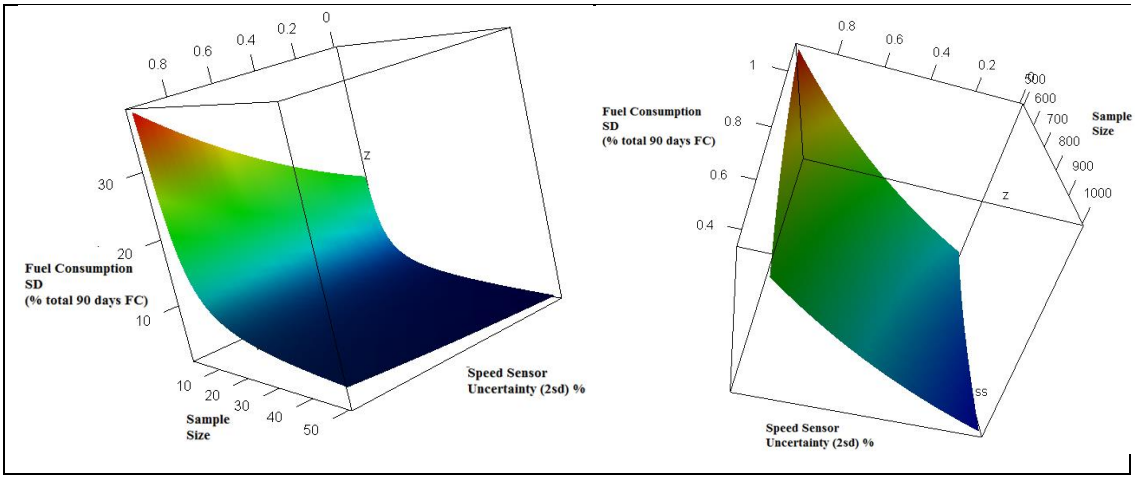


Figure 76: Sample size, sensor uncertainty (represented by the speed sensor uncertainty) and fuel consumption uncertainty (as a % of the 90 day total). Left: Sample sizes 10 to 50. Right: Sample sizes 500 to 1000. Colour scale is proportional to fuel consumption uncertainty where the colour scale is calibrated to the overall range of the resultant fuel consumption uncertainty.

As can be seen from Figure 77, for the partly filtered model, even if the sensor uncertainty is zero it would not be possible to be reasonably confident of detecting a 3% fuel consumption change over 90 days. This is assuming a sample size of 13 which would be a standard expectation from a 90 day noon report dataset.

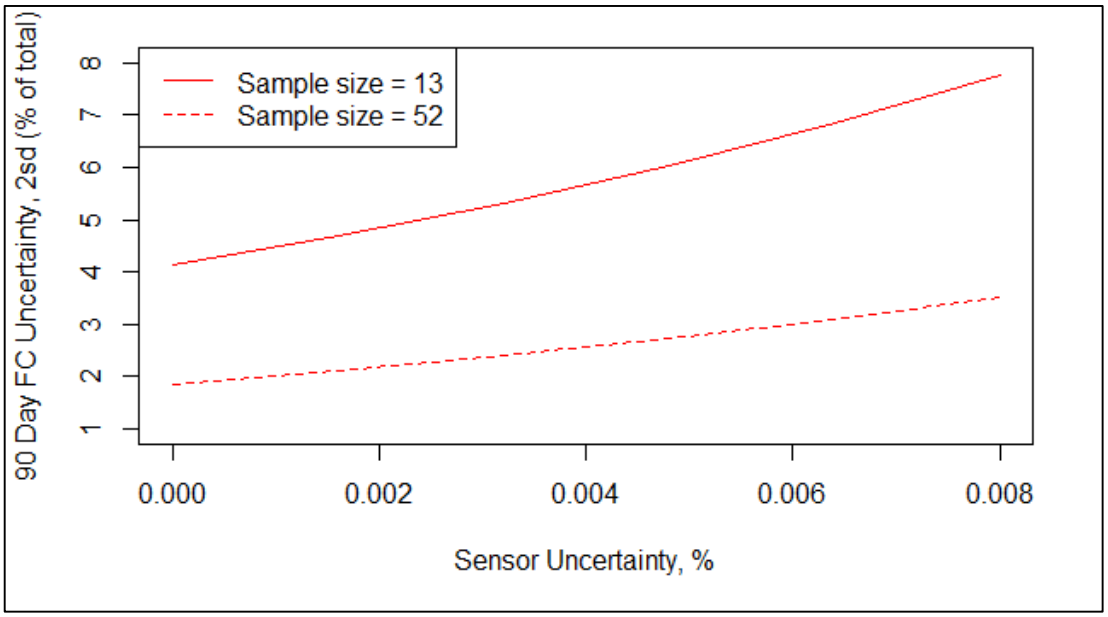


Figure 77: Sensor uncertainty and overall uncertainty for sample size of 13 over 90 days

As shown in the sensitivity analysis, collecting data over a longer time period would be one method to improve the uncertainty to the required level. Alternatively, increasing the sample averaging frequency to 4 per day would also achieve a 3% overall uncertainty with an average sensor uncertainty as shown in Figure 77. However, there would have to be zero human error in the dataset also.

9.5.2 Model type and Sensor Precision

It was seen in the model assessment that the non-filtered model which involved a higher level of model complexity produced greater uncertainty in the overall measure of ship performance (this case the increase in fuel consumption). This is reiterated in Figure 78 and Figure 79. The overall uncertainty in the ship performance measurement for the partly filtered model, in terms of the delta FC as a percent of the 90 day fuel consumption, is generally offset from that of the non-filtered model. This finding is more significant for low sample sizes and high model uncertainties, probably because the NF model introduces a greater range of model parameters.

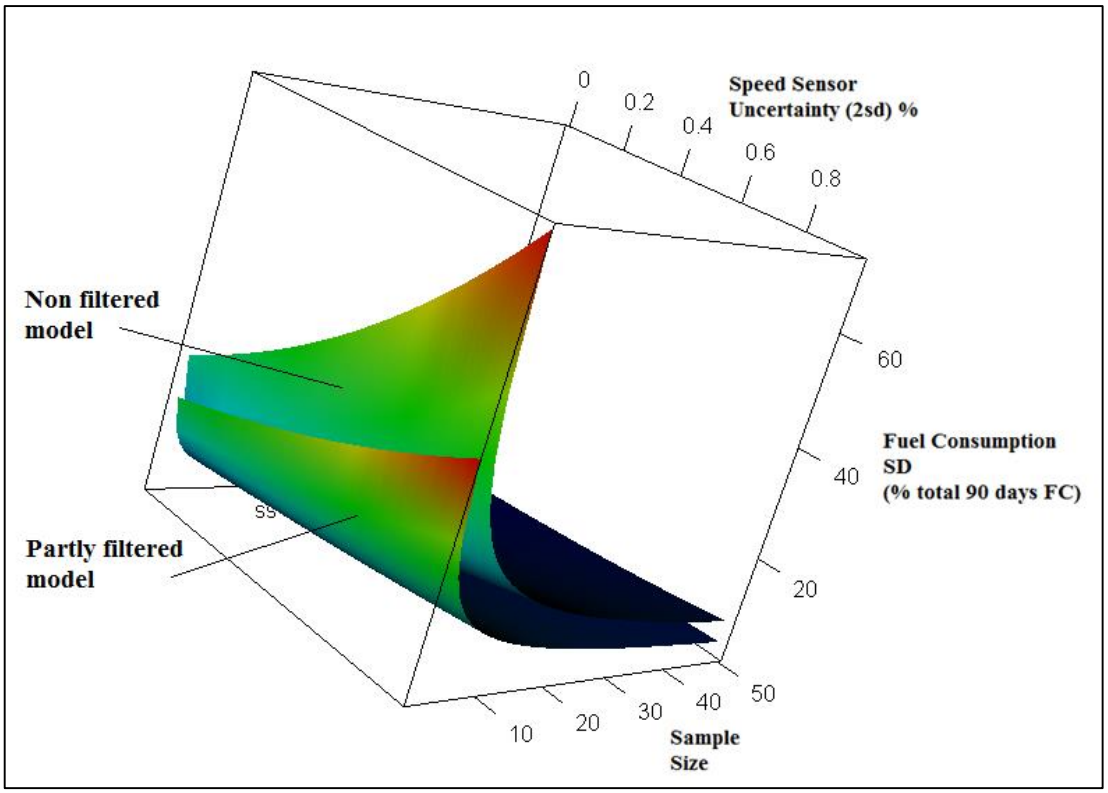


Figure 78: Sample size and sensor uncertainty effect on overall fuel consumption for different model types. Colour scale is proportional to fuel consumption uncertainty where the colour scale is calibrated to the overall range of the resultant fuel consumption uncertainty for each model type.

This also indicates however that if the sensor uncertainties are adjusted then it would be possible for the non-filtered model to achieve similar overall uncertainties to the partly filtered model. This is a significant result because it indicates that there would be no cost to the level of overall accuracy even though there is a significant amount of additional knowledge gained about the ship, such as the ship's interaction with the environmental effects.

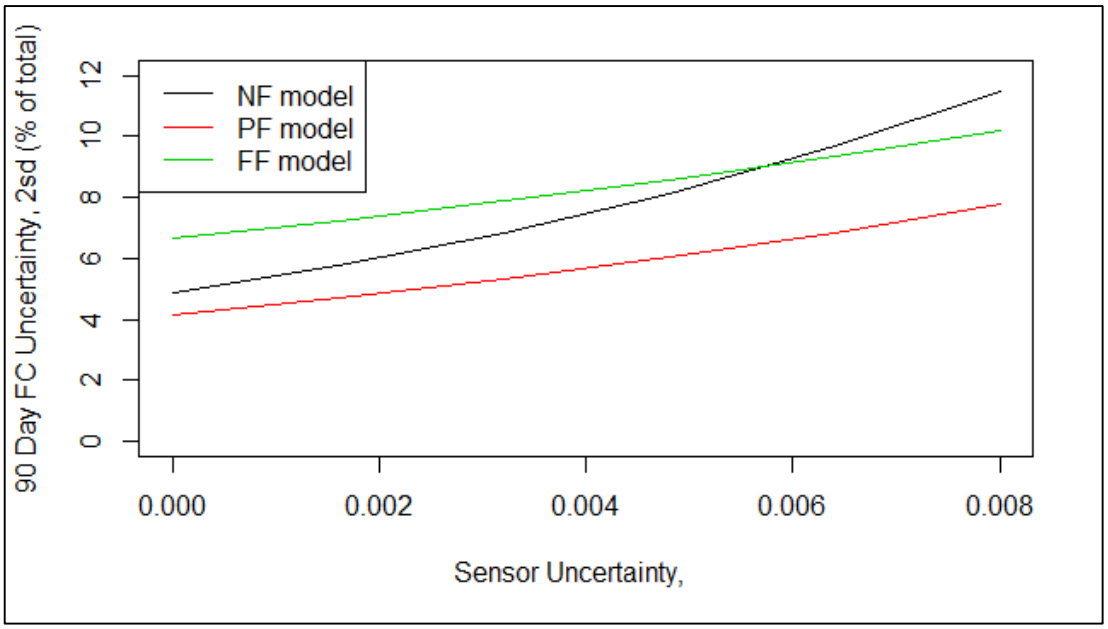


Figure 79: Sensor precision and overall uncertainty for different levels of model complexity and NR sampling frequencies.

This is the maximum theoretical results that could be achieved from a noon report style data acquisition strategy. In reality, a significant amount of human error would be introduced and it is not unrealistic to imagine that the consequence of this would be to increase the overall uncertainty to an impractically high magnitude. This is highlighted in Figure 80, which demonstrates how the power sensor uncertainty affects the overall uncertainty for each model type. A higher power sensor uncertainty is to represent the increased uncertainty as a result of the combination of any additional human error and engine performance fluctuations. These would be the consequence of measuring fuel consumption as a proxy for the shaft power. The overall uncertainty reaches about 10% when the power sensor uncertainty increases to 3.5%, 4.0% and 5.5% for the NF, FF and PF models, respectively.

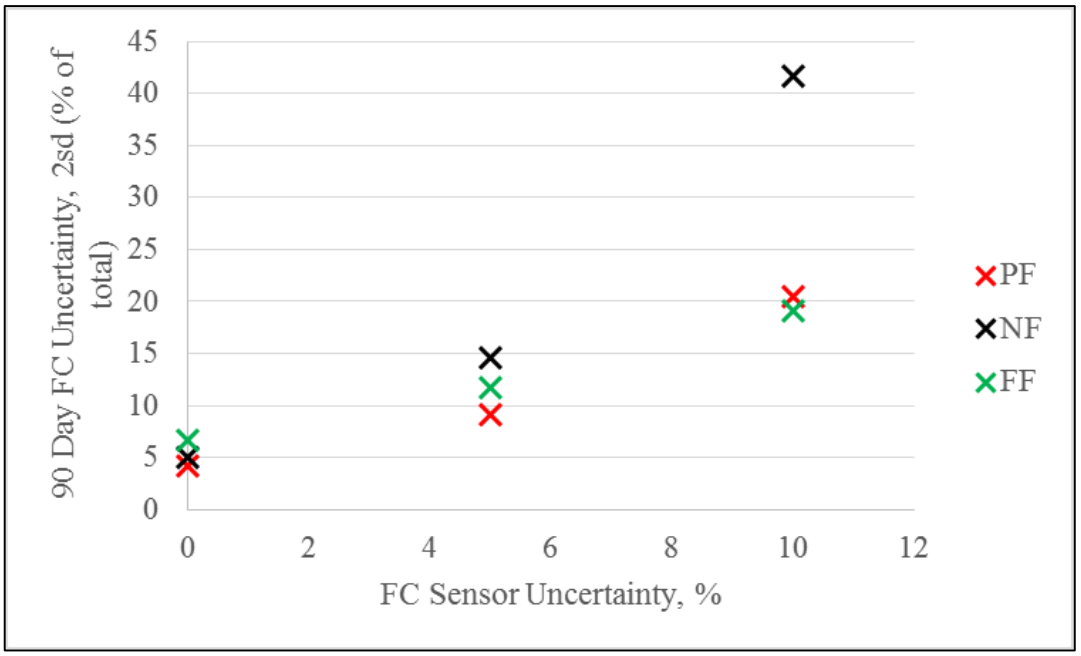


Figure 80: Fuel consumption sensor uncertainty and overall uncertainty for different model types

9.5.3 Model parameter uncertainty and model type

The alternative way to improve the non-filtered model would be to improve the model parameter uncertainty. As can be seen by Figure 81, the non-filtered model will achieve the same overall uncertainty as the partly filtered model for the same sample size if the model parameter uncertainty is improved.

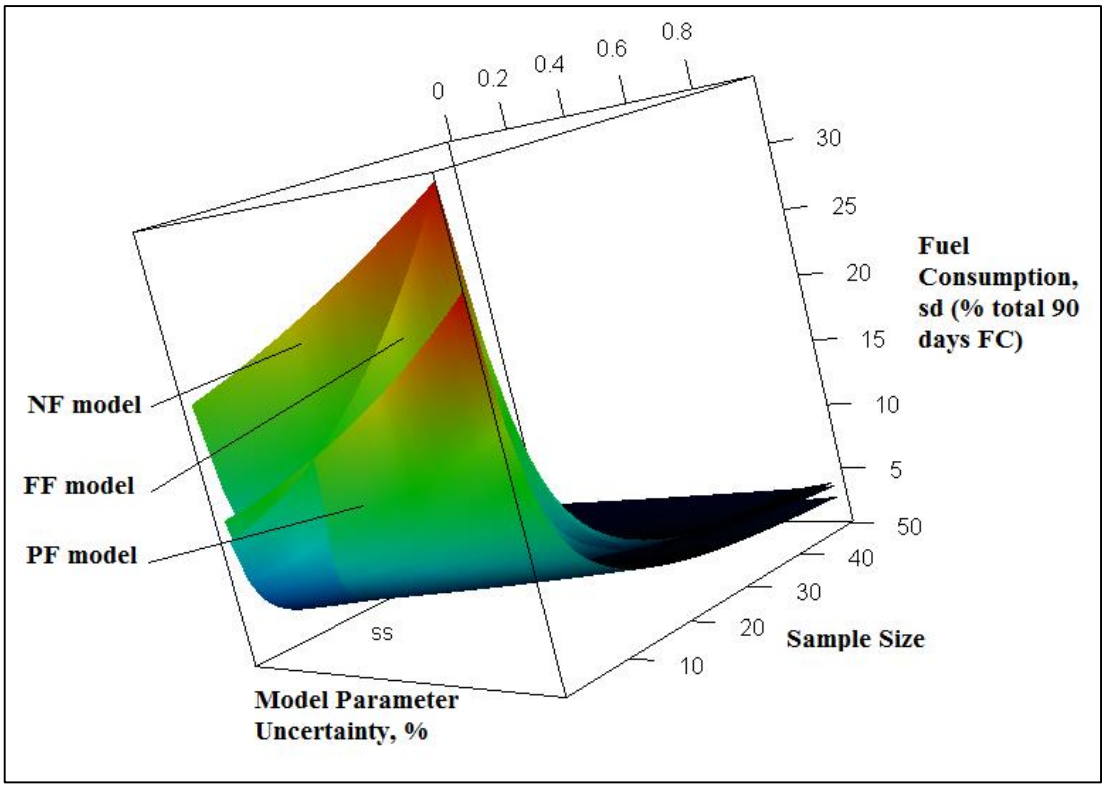


Figure 81: Model parameter uncertainty, sample size and model type. Colour scale is proportional to fuel consumption uncertainty where the colour scale is calibrated to the overall range of the resultant fuel consumption uncertainty for each model type.

This is also reflected in the 2D figure (Figure 82), which represents the model parameter and overall uncertainty interaction for indicative sample sizes of a daily reporting style DAQ strategy. Approximately, for 5% overall uncertainty, the model parameter uncertainty would need to be reduced from about 0.8% to 0.4% if moving from the partly filtered model (PF) to the unfiltered model (NF).

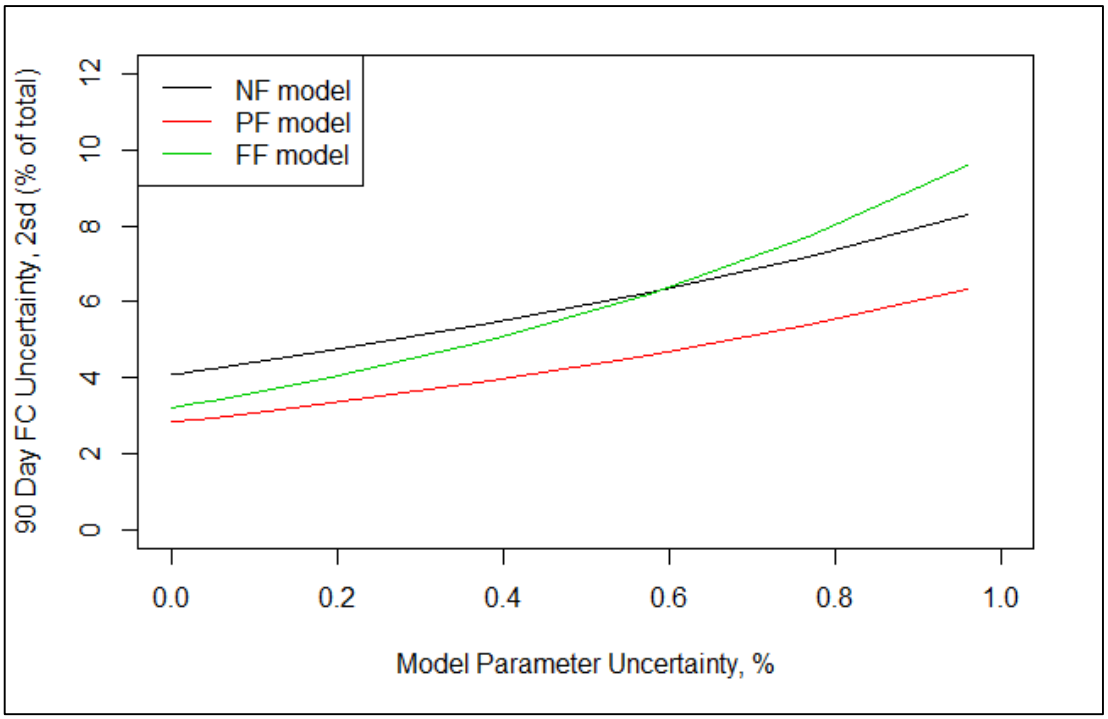


Figure 82: Model parameter uncertainty for different model types. Sample size =19, 13, 9 for NF, PF and FF models, respectively

Improving the model parameter uncertainty for the non-filtered model requires returning to the model structure and functional form assumptions that were introduced in chapters 5 to 7. It would be a complicated and time consuming task to assess the hybrid model by incorporation in to the Monte Carlo simulation in the same way as the statistical model. It is also not deemed necessary because the model standard errors can be examined to understand the model parameter uncertainties in the same way that they were derived for the statistical model. Also, the trend in model performance between the different models as found by the MC simulation reflects that of the top down comparison and therefore this can give a rough indication of the hybrid model performance relative to the others.

	log(Speed)	**log(η_O)**	**log(η_R)**	**log(η_H)**	**H_S^2**	**H_S**
% SE	1.79	-3.27	-5.01	-2.31	6.26	-7.04

Table 45: Standard deviation of the coefficients as a percentage of the coefficient for hybrid model II

For the model parameters, the average model parameter uncertainty is less for the hybrid model II than it is for the statistical model, 4.3% versus 7.8% respectively. The per cent standard errors of the parameters are shown in Table 45.

As seen in Figure 83, the hybrid model (series 7) exhibit a SEM as low as the fully filtered dataset and, at lower sample sizes an SEM as low as the moderately filtered dataset.

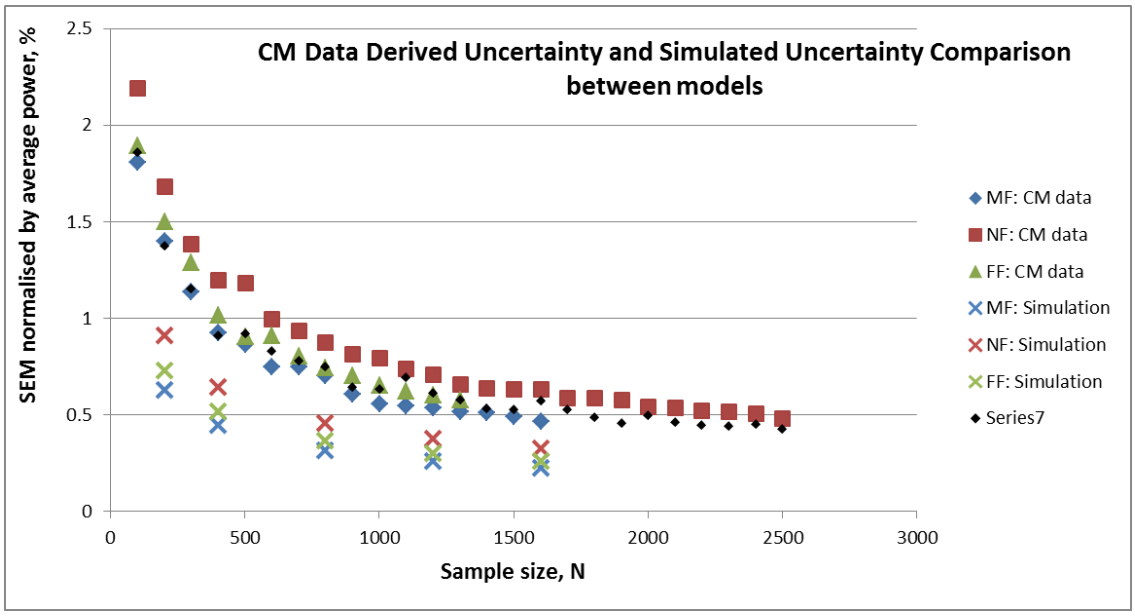


Figure 83: The hybrid model performance (series 7) from the top down comparison

For the continuous monitoring dataset, a low level of uncertainty is achieved regardless of the model type. The intercept represents the residual uncertainty for each model if the model parameter uncertainty was zero. In this high frequency DAQ example, the relatively negligible effects of sample size, and in the absence of sample averaging effects, the offset is largely comprised of sensor uncertainty. The offset exhibited by the full normalisation model indicates that significant gains can be made by improvement in sensor quality. The steeper gradient of the fully filtered model is because of the model parameter bias which is only relevant to this model type. Sea trial data would help to reduce the sensor bias significantly and thus lead to a partly filtered model of very similar accuracy to the fully filtered model.

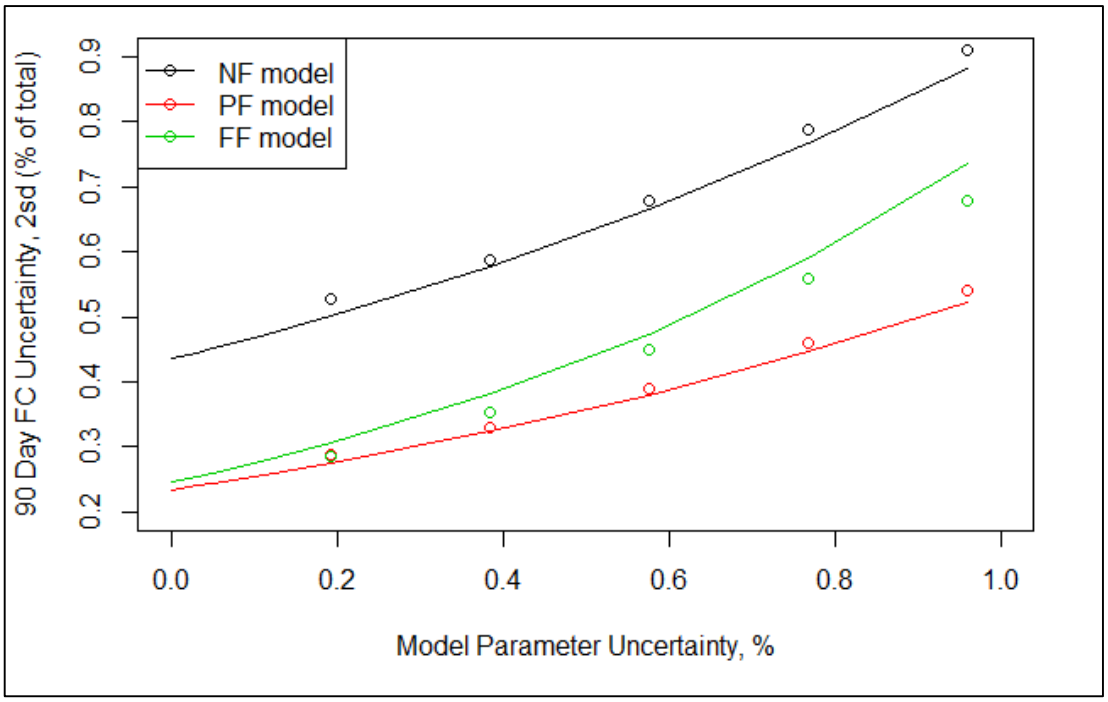


Figure 84: Sample size = 1000

The conclusions of this chapter are brought together and summarised in the following chapter.

Chapter 10. Conclusions and Further Work

This research set out to improve the state of the art in the comparatively under-studied field of ship performance analysis. This was done by addressing the following three questions:

1. Can hybrid models outperform theoretical/ black box models in the provision of detailed and quantifiable knowledge about a ship's performance?
    a. Can hybrid models deliver a more accurate ship performance metric?
    b. Can hybrid models provide deeper insight into the drivers and influences of the ship's performance?
2. How can the various sources of uncertainty in ship performance measurement be individually quantified?
3. What can uncertainty analysis tell us about the relative strengths and weaknesses of different DAQ strategies for estimating/characterising ship performance?

In order to answer these questions, the following work was undertaken:

- Rigorously develop and deploy three separate models (theoretical, statistical, hybrid) for the conversion of performance data into performance quantification and trends
- Propose a method for the statistical analysis of the relative accuracy of different models
- Establish a framework and terminology for the characterisation of uncertainty in ship performance quantification
- Develop a simulation tool (Monte Carlo based) for the forensic investigation of uncertainty in ship performance quantification
- Deploy the simulation tool in a series of investigations to test the significance of different parameters and the requirements of data quality

The key novel contributions being: the development of an uncertainty framework specific to ship performance, the development of a suite of Monte Carlo simulations, the first rigorous comparison of the relative uncertainty of noon-report and

continuous monitoring based methods, and the extensive testing of performance data and models (using those simulations) to establish in depth knowledge and insight into the generalities of ship performance analysis. A novel, hybrid model for ship operational performance monitoring is also introduced.

In a global industry for which analysis of ship performance and detailed understanding of the drivers of fuel consumption, these new knowledge contributions are of great relevance and timeliness. A fact evidenced by the inclusion of much of this work in the development of ISO 19030 a new standard for the measurement of hull and propeller performance.

Specific to the research questions outlined in Chapter 3, the findings are as follows:

RQ1: A unique, hybrid ship performance model is presented which is found to be of superior performance to either a theoretical model or statistical model. The problems of unknown functional form create a problem for a purely statistical model which is also unable to deal with the non-linearities that are characteristic to the complexities of the ship's propulsion system and the environment in which it operates. Additional model parameter uncertainty is introduced in a purely statistical model due to omitted variable bias. This stems from the propeller characteristics and the effect on shaft power of the interactions between shaft torque and rpm that cannot be explicitly defined by a statistical model alone. In the hybrid II model, the functional form is derived directly from the theory which therefore allows for such interactions to be accounted for (such as the propeller efficiency). Theoretical relationships therefore define the underlying functional model form and then by adjusting the model parameters to fit the actual data it has been possible to improve the accuracy of the model. This was demonstrated in Chapter 7 and Chapter 8. Crucially, the model enables the extraction of detailed information relating to the ship's long term hull and propeller efficiencies and the ship's response to wave effects. The effect of time may be extracted either through the trend in the performance indicator or LSDV methods for time effects. There is scope for further work to establish robustness across a wider range of ship types and data acquisition strategies. This may be of significant benefit to industry in that accurate measurements of how the ship responds in different environments can be easily obtained from only a handful of high frequency

sensors. This could be used to help predict ship performance, aid financial forecasts, determine the possible effect of a retrofit technology and measure accurately its impact. It may also be useful to policy work such as helping to determine what the minimum power requirements for new builds should be. For improved accuracy a combination of the theoretical model with CFD results, model test results or sea trial data should be used.

RQ2: A detailed study of the various sources of uncertainty in the ship performance measurement has been presented. A method was then detailed which has made it possible, for the first time, to apportion the various sources of uncertainty to the overall uncertainty in the ship performance measurement. This has been validated to some extent using a continuous monitoring dataset although further work needs to be done to validate the noon report uncertainties. The validation shows a conservative estimate and it may therefore be viewed as the lower bound of an uncertainty estimate. Human error for example has not been included however it is possible to draw conclusions surrounding the other sources of uncertainty and to understand what might be achieved assuming a high quality (human error reduced as much as possible) noon report dataset.

RQ3: The uncertainty quantification method enabled the relative strength and weaknesses of different data acquisition strategies for characterising ship performance to be assessed. Firstly a sensitivity analysis demonstrated the following;

- The model precision uncertainty causes a change in the overall uncertainty however has limited effect on the actual performance indicator itself.
- A model bias uncertainty however does affect the magnitude of the performance indicator as well as the overall uncertainty; random variations partially cancel while bias accumulates linearly
- Important parameters for precision:
  - Speed sensor precision
  - Speed model parameter precision
  - Sample size in any form; altering the length of the evaluation period, altering the sampling frequency or altering the proportion filtered
- Important parameters for bias:

  - Speed sensor bias
  - Sample averaging frequency
- A highly precise, regularly calibrated speed sensor is worth investment

The noon report and continuous monitoring comparison has demonstrated that there is a ten-fold improvement in uncertainty achieved using a continuous monitoring set relative to a noon report dataset, even without including the likely significant effects of human error. This is not a surprising finding however if noon report data were to be collected perfectly, without the influence of human error, then uncertainties of the 5% level are achievable, particularly in the presence of high quality sensors. A significant improvement can also be achieved if the frequency of sampling is increased. If fuel consumption is used as a proxy for shaft power, which may cause an inaccuracy of 5% (due to uncertainties in the fuel consumption measurement and conversion to shaft power assumptions) then there could be a significant effect also on the overall uncertainty.

General Comments

To recap, the data acquisition strategy is comprised of data collection and processing, this includes pre-processing steps such as the removal of outliers, filtering and any subsequent normalisation techniques. The strategy variants are defined in Table 46.

Uncertainty Category	Variants	Variants
Instrument uncertainty	Precision Bias	
Sampling uncertainty	Sample size	Averaging Proportion filtered Evaluation period length
	Sample averaging frequency	
Model uncertainty	Non filtered model (NF)	Bias Precision
	Partially filtered model (PF)	
	Fully filtered model (FF)	
Human error		

Table 46: The variants of the data acquisition strategy

The optimum data acquisition strategy depends on the application, and in this respect there are two elements to consider:

1. The detail of information concerning the ships performance that it is required to derive, for example quantifiable measurements of the ships response in different environmental and operational conditions or in-depth knowledge about the response of the ship's fundamental parameters (hull or propeller efficiencies for example)
2. The level of accuracy required in the overall ship performance measurement in terms of delta fuel consumption

If the requirement is only to measure the overall ship performance to the highest level of accuracy then it has been shown that a partly filtered model, based only on draught and vessel speed provides the optimum result. This is true for both NR and CM acquisition strategies. However this does not tell us a great deal about how the ship reacts in different environments. Even when the likely sample size is increased to that of a non-filtered sample (an increase from 14.8% for a PF model to 21.6% for an NF model) then there are still no gains to be made in the accuracy of the overall performance indicator by normalisation. This is despite the significance of sample size, and this is due to greater significance of the sensor uncertainty. The hypothesis that there is a trade-off to be made between the sample size and the model error has not been accepted when comparing a PF to NF model. However this hypothesis is accepted when comparing the FF to the PF model. This is because the effect of sample size is non-linear.

Sea trial data would help to reduce the sensor bias in the fully filtered model significantly and may lead to a partly filtered model of very similar accuracy to the fully filtered model.

If a greater level of knowledge is required, for example to inspect in greater detail the interactions between significant components of the ship's propulsion system (hull or propeller efficiencies for example), then a hybrid model is advantageous. In the uncertainty analysis the NF model is significantly hindered by sensor uncertainties because of the number of parameters that are required in its estimation and because

of the increase in model parameter uncertainties. However, the NF model in the uncertainty analysis is based on a purely statistical model and it is possible that the accuracy of the hybrid model may approach that of the partly filtered model because of the inclusion of theory which works to reduce the model parameter uncertainties, although this hasn't been explicitly proven.

The simulated differences in overall uncertainties between PF, NF and FF models based on CM datasets are so small that they become insignificant, therefore the CM dataset lends itself to detailed analysis of the ship's performance in a more complex model.

Reflecting on the uncertainty in the performance indicator as a percentage of the total 90 day fuel consumption, the uncertainty in the performance indicator as calculated from a simulation based on a NR DAQ strategy and using the NF model is of the order of 8%, which may be too high for many applications. This is likely to be higher given the influence of human error. Over a longer time period the increased sample size may provide sufficient evidence for reliable and detailed information to be gained even from NR dataset, particularly if the LSDV method is employed.

This is a multi-dimension problem between sample size, model parameter precision and instrument precision as demonstrated by the 3D plots of section 9.5.

10.1 Further work

Firstly, some specific analysis is presented which is relevant to the output of this thesis. Secondly, general further work is suggested.

10.1.1 Time Effects Performance Indicator

In order to demonstrate the use of outputting the performance indicator as a continuous variable, structural break analysis was performed on the time series. The aim of the analysis is to extract step changes in the ship's performance which may be attributed to interventions aimed to improve the hull, engine or propeller, such as the application of a new coating or retrofit propeller technology. This analysis is applied to the performance indicator output from the hybrid II model presented in chapter 7.

The normalisation model calibration period was day 200 to 300 (the relative dates are arbitrary and do not affect the results as the PI analysis is concerned with relative changes across the period); the performance indicator was calculated over the entire 391 day period (the evaluation period). The performance indicator trends are plotted in Figure 85 which shows the resultant PI as calculated according to the model presented in section 7.1.2. A structural break indicates a shift or violent change in the time series, in this analysis a break is likely to occur if the hull, propeller or engine is modified in some way that effects how the power or fuel consumption responds to the changes in the operational or environmental conditions, such as a hull clean, propeller retrofit or engine maintenance. A structural break analysis was performed on each by implementing the method proposed by (Bai 1998) and (Bai 2003) which searches for unknown multiple breaks. The general principle is to compare the F-statistic of models assuming all possible break points.

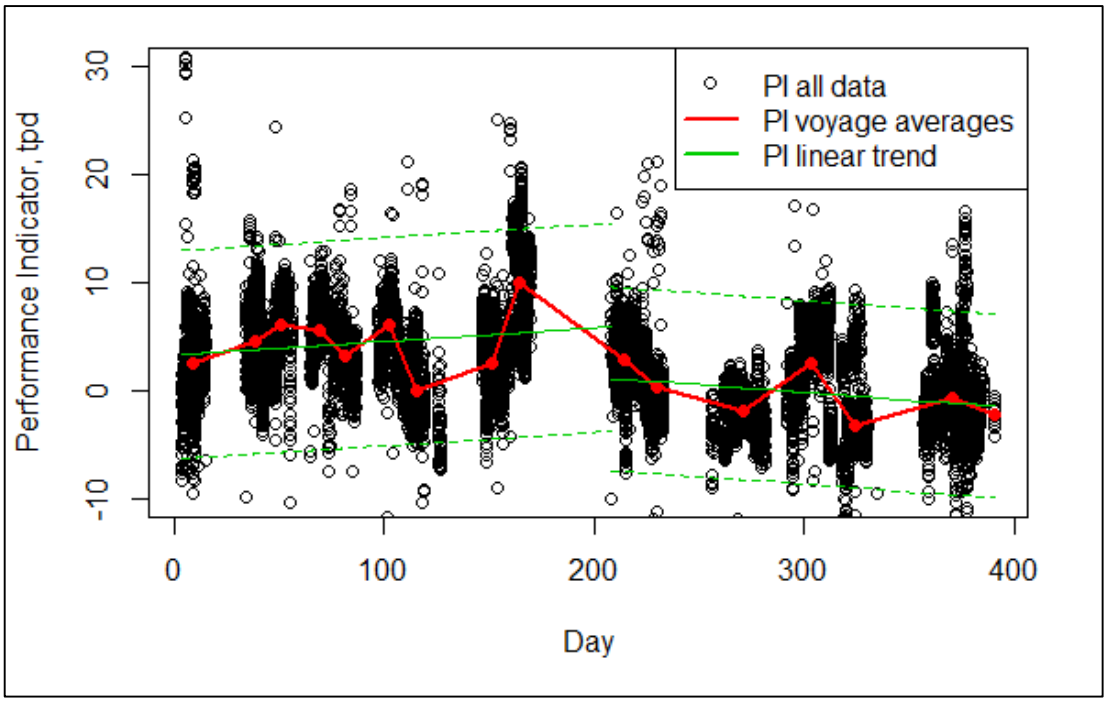


Figure 85: Structural break analysis of the performance indicator, structural break found at day 208.13

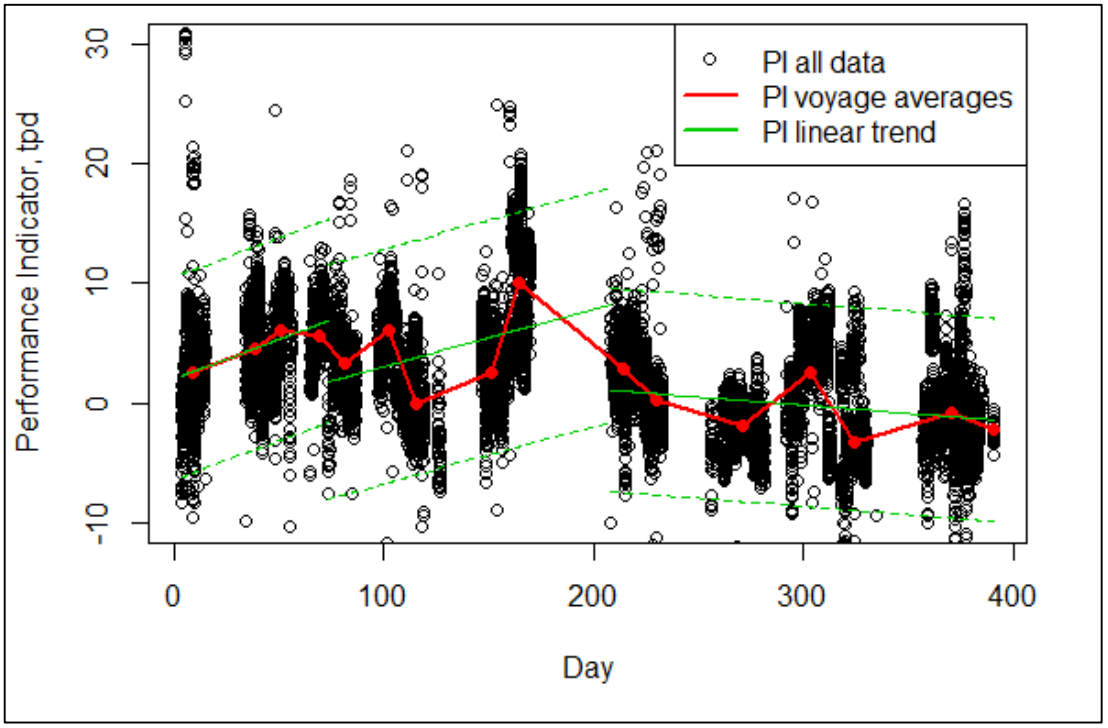


Figure 86: Structural break analysis of the performance indicator, structural break found at day 208.13(first) and 73.63(second)

Figure 85 and Figure 86 show the different gradients when 1 structural break is defined versus 2 when the same model is used in the PI calculation. It is not possible

to say definitively from which model the structural breaks are correctly identified because the ships maintenance schedule is unknown. It seems more likely that a break would fall directly after a period of missing data, during which time the ship may be in port or dry dock. This is the scenario depicted in Figure 85 where the break is identified following a 38 day gap in the data coverage. Immediately following this is the calibration period, and some maintenance event at that time would also explain the positive voyage average performance indicators running up to that period. During that time (days 0 to 200) the linear trend in the performance indicator (PI) shows a positive correlation with time (t), resulting in the following coefficients from a single variable linear regression of performance indicator on day:

$$PI = 5.37 + 2.222\text{x}10^{-3}\,(t)$$
$$(0.119)\ (1.219\text{x}10^{-3})$$

Where the SE of the regression estimates are shown in brackets and the units for PI are in tpd, therefore over the 200 day period, the increase in tpd is a rather exiguous 0.44 (0.8%).

As can be seen from the plots and the prediction intervals around the linear trends, there is a reasonably large degree of scatter in the performance indicator results. This could be due to any one of the reasons outlined in the uncertainty characterisation framework. It has also been demonstrated in section 9.2.3 that a linear trend may not be representative of the underlying performance and therefore this will increase the uncertainty. Sudden fluctuations that cause a deviation from linearity might be due to fluctuating fuel grades, the presence of cat fines in the fuel or the vessel moving through particular areas where conditions are conducive of temporary increased fouling (due to collecting debris on the hull or propeller). Relatively cheap fuels during times of high fuel costs especially may exhibit poor fuel quality and contribute to fluctuation in fuel grade in the global market. Conversely, the removal of hull debris and fouling will have an impact. Local conditions such as warm, saline waters perhaps in shallow, coastal seas are more likely to encourage marine biological growth, while high vessel speeds across the open ocean have the opposite effect as particular hull coatings require a minimum speed to work effectively.

This analysis is simply to present an example of how the data can be manipulated and other statistical techniques applied. In this form the performance indicator may form an input to other environmental or operational models with the aim of optimising ship efficiency or understanding the impact of a wider range of environmental variables on ship performance.

10.1.2 Other Further Work

Building on the work of this thesis, there is scope to establish robustness of the hybrid model across a wider range of ship types and data acquisition strategies. This is both in terms of the accuracy of the performance indicator and the accuracy of the quantification of the effects of the endogenous parameters that provide significant detail and in-depth knowledge of the ship's response.

Datasets that include details of maintenance schedules will be particularly valuable in validating the LSDV method. Without this it is impossible to know whether the output has correctly highlighted an actual increase in performance that corresponds to a change in real life, i.e. a hull or propeller clean.

There remains the rigorous examination of the uncertainty analysis of the hybrid model, and specifically how this compares to the uncertainty achieved from a partly filtered model which is currently deemed to deliver the lowest uncertainty in the performance indicator.

This thesis has also highlighted the sensitivity of the performance indicator uncertainty to the uncertainty of the speed parameter measurement. It is therefore proposed that a significant overall uncertainty improvement could be achieved if the vessel speed was derived from AIS data and combined with the NR data. This is especially true as the coverage and accuracy of AIS data becomes increasingly reliable producing more complete datasets with full global coverage.

The uncertainty in the performance measurement achieved from the hybrid model may also be improved by using a more sophisticated 'theoretical model' where ship

specific coefficients are obtained from CFD, sea trial data or model tests for the specific ship studied.

There is also scope for further work in quantifying the cost of data uncertainty. This is relevant to the application; for example performance monitoring may be applied to measuring hull performance in order to determine when a hull clean should take place. In that case, the cost of the data uncertainty may be attributed to the fuel cost due to the reduction in ship performance during the time that the vessel operates with a fouled hull from the point that a higher fidelity dataset would have revealed that the hull was in sufficiently poor condition to warrant a hull clean. The case for cost effective measures for data collection to obtain the data required for a model of sufficient fidelity would be a worthwhile assessment.

An interesting and extremely valuable study would be to further investigate and attempt to quantify the effect of human error in the noon report dataset, brief details of how this might be achieved and potential difficulties are provided in section 9.2.4. This would further improve the noon report uncertainty quantifications and also help extricate the additional uncertainty caused by the introduction of the assumptions that transform one metric to another, such as the uncertainty surrounding SFOC when using fuel consumption to understand hull and propeller performance.

References

A. Saltelli, K. Chan, et al. (2008). Sensitivity Analysis, John Wiley & Sons, LTD.

ABS (2014). Ship Energy Efficiency Measures - Status and guidance.

Acomi, N. and O. C. Acomi (2014). "Improving the Voyage Energy Efficiency by Using EEOI." Procedia - Social and Behavioral Sciences **138**: 531-536.

Aertssen, G. (1966). "Service Performance and Sea Keeping Trials on M.V. Jordaens." Transactions of RINA **108**: 305-343.

Agnolucci, P., Smith, T. W. P., 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.

AIAA (1999). S-071A-1999 Assessment of Experimental Uncertainty With Application to Wind Tunnel Testing, American Institute of Astronautics and Aeronautics.

Ainsworth, T. (2008). 'Significant Wave Height' A closer look at wave forecasts N. National Oceanic and Atmospheric Association. NWS Juneau, Alaska.

Aldous, L. and T. Smith (2013). Speed Optimisation for Liquefied Natural Gas Carriers A Techno Economic Model. Low Carbon Shipping. Newcastle.

Allison, P. (1999). Multiple Regression: A Primer. California, Pine Forge Press.

Andersen, I. M. V. (2013). "Wind loads on post-panamax container ship." Ocean Engineering **58**: 115-134.

Armstrong, V. N. (2013). "Vessel optimisation for low carbon shipping." Ocean Engineering.

ASTM (2011). American National Standard Standard E2782 - 11: Guide for Measurement Systems Analysis (MSA). ASTM Committee on Quality and Statistics, ASTM. **E2782**.

Bai, J., P. Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes." Econometrica **vol 66**: 47-78.

Bai, J. a. P. P. (2003). "Computation and Analysis of Multiple Structural Change Models." Journal of Applied Econometrics **18**: 1-22.

Ballou, P. (2013). "Ship Energy Efficiency Management Requires a Total Solution Approach." Marine Technology Society Journal **47**(1).

Bazari, Z. (2007). "Ship energy performance monitoring and benchmarking." Journal of Marine Engineering and Technology **A9**.

Berlekom, V. (1981). "Wind force on modern ship forms-effects on performance." Transactions of the North East Coast Institute of Engineers and Shipbuilders **97**: 123-134.

Bitner-Gregersen, E. M., S. K. Bhattacharya, et al. (2014). "Recent developments of ocean environmental description with focus on uncertainties." Ocean Engineering **86**: 26-46.

Bitner-Gregersen, E. M., K. C. Ewans, et al. (2014). "Some uncertainties associated with wind and wave description and their importance for engineering applications." Ocean Engineering **86**: 11-25.

Bitner-Gregersen, E. M. and O. Hagen (1990). "Uncertaintes in Data for The Offshore Environment." Structural Safety(7): 11-34.

Blendermann, W. (1994). "Parameter identification of wind loads on ships." Journal of Wind Engineering Industrial Aerodynamics **51**: 339-351.

Bose, N., Molloy, S. (2009). Reliability and accuracy of ship powering performance extrapolation. First International Symposium on Marine Propulsors. Trondheim, Norway.

Boom, H., Huisman, H. and Mennen, F. (2013). "New Guidelines for Speed/Power Trials, Level playing field established for IMO EEDI", SWZ Maritime, January & February 2013.

Braake, H. A. B., H. J. L. Can, et al. (1998). "Semi-mechanistic modeling of chemical processes with neural networks." Engineering Applications of Artificial Intelligence(11): 507-515.

Chatfield, C. (1995). "Chatfield Model Uncertainty, Data Mining and Statistical Inference." Journal of the Royal Statistical Society. Series A (Statistics in Society) **158**(3): 419 - 466.

Coleman, H. W., Steele, W. G. (1990). Experimentation and Uncertainty Analysis for Engineers. US, John Wiley & Sons.

Corbett, J. J. and K. W. Horst (2003). "Updated emissions from ocean shipping." Journal of Geophysical Research **108**(D20).

Cox, M. G., Harris, P. M. (2006). Software Support for Metrology Best Practice Guide No. 6: Uncertainty Evaluation. UK, National Physics Laboratory.

Curry, J. A. and P. J. Webster (2011). "Climate Science and the Uncertainty Monster." Bulletin of the American Meteorological Society **92**(12): 1667-1682.

Deligiannis, P. (2014). Ship Performance Indicator. Shipping in Changing Climates, Tyndall, 2014. Liverpool, UK.

DNV (2010). ENVIRONMENTAL CONDITIONS AND ENVIRONMENTAL LOADS. Recommended Practice, Det Norske Veritas.

Doudnikoff, M. and R. Lacoste (2014). "Effect of a speed reduction of containerships in response to higher energy cost in SECA." Transportation Research Part D: Transport and Environment **28**: 51-61.

Dunn, G. (2004). Statistical Evaluation of Measurement Errors. US, Hodder Arnold.

E. P. Smith and K. A. Rose (1993). "Model goodness-of-fit analysis using regression and related techniques." Ecological Modelling **77**: 49-64.

Eça, L., Hoekstra, M. (2014). "A procedure for the estimation of the numerical uncertainty of CFD calculations based on grid refinement studies." Journal of Computational Physics(262): 104-130.

Eide, M. S. and Ø. Endresen (2010). "Assessment of measures to reduce future CO2 emissions from shipping." Research and Innovation, position paper 05 - 2010.

Eide, M. S., Ø. Endresen, et al. (2009). "Cost-effectiveness assessment of CO2 reducing measures in shipping." Maritime Policy & Management **36**(4): 367-384.

Eljardt, G. (2006). Development of a fuel oil consumption monitoring system. Institut fur Entwerfen von Schiffen und Schiffssicherheit, Technische Universitat Hamburg-Harburg.

EPA (2001). Risk Assessment Guidance for Superfund. Washington DC, US Environmental Protection Agency. **Vol III - Part A, Process for Conducting Probabilistic Risk Assessment**.

EVO (2012). International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings. E. V. Organization. **EVO 10000 – 1:2012**.

Faber, J., Nelissen, D., Smit, M. (2013). Monitoring of bunker fuel consumption, Delft, CE Delft.

Fathom (2011). Ship Efficiency: The Guide. UK, Fathom.

Ferson, S. and L. R. Ginzburg (1996). "Different methods are Needed to Propagate Ignorance and Variability." Reliability Engineering and System Safety **54**: 134-144.

Ferson, S., W. L. Oberkampf, et al. (2008). "Model validation and predictive capability for the thermal challenge problem." Computer Methods in Applied Mechanics and Engineering **197**(29-32): 2408-2430.

Flikkema, M. (2013). Approaches To Measuring Hull and Propeller Performance. Hull and Propeller Performance Measurement Standard. MARIN. Oslo.

FORCE Technology, N. L. Larsen, et al. (2011). "Understanding the physics of trim."

Gaafary, M. M., H. S. El-Kilani, et al. (2011). "Optimum design of B-series marine propellers." Alexandria Engineering Journal **50**(1): 13-18.

Gauch, H. G., Hwang J. T., Fick, G.W. (2003). "Model Evaluation by Comparison of Model-Based Predictions and Measured Values." Agron. J.(95): 1442-1446.

Gleser, L. J. (1998). "Assessing Uncertainty in Measurement." Statistical Science **13**(3): 277-290.

Gould, R. W. F. (1982). "The estimation of wind loads on ship superstructures." The Royal Institution of Naval Architects, monograph **No. 8**.

Gulev, S. K., Grigorieva, V., Selemenov, K., Zolina, O., "Evaluation of Ocean Winds and Waves from Voluntary Observing Ship Data." Advnces in the Applications of Marine Climatology 53-67

Haddara, M. R., Soares, C. G. (1999). "Wind Loads on Marine Structures." Marine Structures: 199-209.

Hansen, A. (2010). "Monitoring of hull condition of ships." MSc Thesis Norwegian University of Science and Technology.

Hogben, N. and Lumb, (1967): Ocean Wave Statistics. The Ministry of Technology, HMSO, London, 263.

Hogben, N., N.M.C. Dacunha and G.F. Oliver, (1986): Global Wave Statistics. Unwin Brothers, London, 661.

Hollenbach, K. U. (1998). "Estimating resistance and propulsion for singlescrew and twin screw ships." Ship Technology Research **45/2**.

Holtrop, J. (1984). "A statistical re-analysis of resistance and propulsion data." International Shipbuilding Progress **31**(363): 5.

Holtrop, J. and G. G. J. Mennen (1982). "An Approximate Power Prediction Method." International Shipbuilding Progress **29**: 166 - 170.

IEA (2012). CO_2 Emissions From Fuel Combustion. I. E. Agency, IEA Statistics.

IMO (2009). "Prevention of Air Pollution From Shps. <2nd IMO GHG Study 2009.pdf>. Update of the 2000 IMO GHG Study. Final Report Covering Phase 1 and Phase 2." I:\MEPC\59\INF-10.doc.

IMO, A. R. (1995). Performance Standards for Devices to Indicate Speed and Distance.

IMO_MEPC_63/4/8 (2011). "MEPC 63/4/8 A transparent and reliable hull and propeller performance standard."

Insel, M. (2008). "Uncertainty in the analysis of speed and powering trials." Ocean Engineering **35**(11-12): 1183-1193.

Isherwood, R. M. (1973). "WIND RESISTANCE OF MERCHANT SHIPS." R.I.N.A. Supplementary Papers **115**: 11.

ITTC (1999). The Specialist Committee on Trials and Monitoring Final Report and Recommendations to the 22nd ITTC.

ITTC (2002). "7.5 0.4 01 01.4 Specialist Committee of 23rd ITTC on Speed and Powering - Full Scale Measurements Speed and Power Trials Instrumentation Installation and Calibration."

ITTC (2002). CFD General Uncertainty Analysis in CFD Verification and Validation Methodology and Procedures. ITTC - Recommended Procedures and Guidelines. **7.5-03-01-01**.

ITTC (2002). Propulsion, Performance Uncertainty Analysis, Example for Propulsion Test. ITTC - Recommended Procedures and Guidelines. **7.5-02-03-01.2**.

ITTC (2002). Propulsion, Propulsor Uncertainty Analysis, Example for Open Water Test. ITTC - Recommended Procedures. **7.5-02-03-02.2**.

ITTC (2002). The Specialist Committee on Speed and Powering Trials Final Report and Recommendations to the 23rd ITTC. 23rd International Towing Tank Conference

ITTC (2002). Testing and Extrapolation Methods Propulsion, Performance Propulsion Test. ITTC - Recommended Procedures. **7.5-02-03-01.1**.

ITTC (2005). Full Scale Measurements Speed and Power Trials Analysis of Speed/Power Trial Data. ITTC - Recommended Procedures and Guidelines. **7.5-04-01-01.2**.

ITTC (2008). Guide to the Expression of Uncertainty in Experimental Hydrodynamics. ITTC - Recommended Procedures and Guidelines. **7.5-02-01-01**

ITTC (2008). Testing and Extrapolation Methods, General Guidelines for Uncertainty Analysis in Resistance Towing Tank Tests. ITTC - Recommended Procedures and Guidelines. **7.5-02-02-02**.

ITTC (2011). Performance, Propulsion 1978 ITTC Performance Prediction Method. ITTC - Recommended Procedures. **7.5-02-03-01.4**.

ITTC (2011). Uncertainty Analysis - Example for Waterjet Propulsion Test. ITTC - Recommended Procedures and Guidelines. **7.5-02-05-03.3**.

ITTC (2012) Speed and Power Trials Part 1 Preparation and Conduct. ITTC - Recommended Procedures and Guidelines. **7.5-04-01-01.1.**

ITTC (2014) Speed and Power Trials Part 2 Analysis of Speed/Power Trial Data. ITTC - Recommended Procedures and Guidelines. **7.5-04-01-01.2.**

J. J. Corbett, V. Eyring, et al. (2009). PREVENTION OF AIR POLLUTION FROM SHIPS, Second IMO GHG Study 2009, Update of the 2000 IMO GHG Study**:** 289.

Jalkanen, J. P., L. Johansson, et al. (2012). "Extension of an assessment model of ship traffic exhaust emissions for particulate matter and carbon monoxide." Atmospheric Chemistry and Physics **12**(5): 2641-2659.

Jan Tellkamp, A. B., Thomas Gosch, Heinz Günther,, U. D. N. Peter Friis Hansen, Apostolos Papanikolaou,, et al. (2008). ADVANCED DECISION SUPPORT SYSTEM FOR SHIP DESIGN, OPERATION AND TRAINING (ADOPT)

JCGM100:2008 (2008). "Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM 1995 with minor corrections)."

JCGM101:2008 (2008). "Evaluation of measurement data - Supplement 1 to the "Guide to the expression of uncertainty in measurement" - propagation of distributions using a MCM."

JCGM_200:2008 (2008). International Vocalbulary of Metrology - Basic and general concepts and associated terms (VIM).

Journée, J. M. J., R. J. Rijke, et al. (1987). "Marine Performance Surveillance With a Personal Computer."

Kacker, R. and A. Jones (2003). "On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent." Metrologia **40**: 235-248.

Kamal, I. M., Binns, K., Bose, N., Thomas, G., (2013) "Reliability Assessment of Ship Powering Performance Extrapolations using Monte Carlo Methods.", Third International Symposium on Marine Propulsors

Karlaftis, M. G. and E. I. Vlahogianni (2011). "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights." Transportation Research Part C: Emerging Technologies **19**(3): 387-399.

Klaij C. M., V. G., Eca L. (2014). NUMERICAL UNCERTAINTY ESTIMATION IN MARITIME CFD APPLICATIONS. 11th World Congress on Computational Mechanics (WCCM XI).

Klitsch, M. L. (1993). USNS HAYES (T-AG 195) Results of Standardization Trials. Ship Hydrodynamics Department, Research and Development Report. C. D. N. S. W. Center.

Krapp, A. (2011). Fixed rpm and Increased Frictional Resistance. Jotun.

Kristensen, H. O. (2012). Resistance and Propulsion Power.

Leifsson, L. Þ., H. Sævarsdóttir, et al. (2008). "Grey-box modeling of an ocean vessel for operational optimization." Simulation Modelling Practice and Theory **16**(8): 923-932.

Lindstad, H., B. E. Asbjørnslett, et al. (2011). "Reductions in greenhouse gas emissions and cost by shipping at lower speeds." Energy Policy **39**(6): 3456-3464.

Lindstad, H., Asbjørnslett, B. E., Jullumstrø, E. (2013). "Assessment of profit, cost and emissions by varying speed as a function of sea conditions and freight market." Transportation Research Part D: Transport and Environment **19**: 5.

Lindstad, H., E. Jullumstrø, et al. (2013). "Reductions in cost and greenhouse gas emissions with new bulk ship designs enabled by the Panama Canal expansion."

Logan, K. P. (2011). Using a Ships Propeller for Hull Condition Monitoring. ASNE Intelligent Ships Symposium IX. Philadelphia, USA**:** 20.

Longo, J., Stern, F. (2005). "Uncertainty Assessment for Towing Tank Tests With Example for Surface Combatant DTMB Model 5415." Journal of Ship Research **49**(1): 55-68.

Loucks, D. P., Van Beek, E., Stedinger, J. R., Dijkman, J. P.M., Villars, M. T. (2005). Water Resources Systems Planning and Management: CH 9 Model Sensitivity and Uncertainty Analysis. Paris, UNESCO.

M. Atlar, E. J. G., M. Candries, R.J. Mutton, C.D. Anderson (2001). The effect of a foul release coating on propeller performance.

M. Mastandrea, a. C. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. I. C.-W. G. M. o. C. T. o. Uncertainties. Jasper Ridge, CA, USA.

Maloni, M., J. A. Paul, et al. (2013). "Slow steaming impacts on ocean carriers and shippers." Maritime Economics & Logistics **15**(2): 151-171.

MAN (2011). Basic Principles of Ship Design. Denmark.

MARIN (2012). Error and Uncertainty Analysis: Experimental Methods in Marine Hydrodynamics. N. I. f. M. Teknikk.

Mastrandrea, M. D., K. J. Mach, et al. (2011). "The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups." Climatic Change **108**(4): 675-691.

MEPC55/4/4_Annex (2006). Experiences from voluntary ship CO2 emission indexing and suggestions to further work. MARINE ENVIRONMENT PROTECTION COMMITTEE 55th Session. P. O. A. P. F. SHIPS.

MEPC.1_Circ.684 (2009). GUIDELINES FOR VOLUNTARY USE OF THE SHIP ENERGY EFFICIENCY OPERATIONAL INDICATOR (EEOI). I. M. Organisation.

Morgan, M. G. and M. Henrion (1990). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge, UK, Cambridge University Press.

Munk, T. (2006). Fuel Conservation Through Managing Hull Resistance. Motorship Propulsion Conference. Copenhagen.

Munk, T. and D. Kane (2011). Technical Fuel Conservation Policy and Hull and Propeller Performance. Design and Operation of Tankers Conference. R. I. o. N. Architects.

N. Hamlin and R. Sedat (1980). The In-Service Roughness Allowance: Effects of Drydocking, recoating and the passing of time. Proceedings of Shipboard Energy Conservation Conference, New York.

Notteboom, T. E. and B. Vernimmen (2009). "The effect of high fuel costs on liner service configuration in container shipping." Journal of Transport Geography **17**(5): 325-337.

Oosterveld, M. W. C. and P. v. Oossanen (1975). "Further_computer_analyzed_data_of_the_wageningen_b-screw_series." International Shipbuilding Progress **22**(251).

Øyan, E. (2012). Speed and powering prediction for ships based on model testing. Department of Marine Technology, Norwegian University of Science and Technology.

Ozaki Y., L. J., Tikka K., Michel K. (2010). An Evaluation of the EEDI baseline for Tankers Containerships and LNG carriers. Climate Change and Ships: Increasing Energy Efficiency.

Park, D.-M., J. Lee, et al. (2015). "Uncertainty analysis for added resistance experiment of KVLCC2 ship." Ocean Engineering **95**: 143-156.

Paul, R. K. (2008). Multicollinearity: Causes,Effects and Remedies, Indian Agricultural Statistics Research Institute.

Pedersen and Larsen (2009). "Prediction of Full-Scale Propulsion Power using Artificial Neural Networks."

Pérez Arribas, F. (2007). "Some methods to obtain the added resistance of a ship advancing in waves." Ocean Engineering **34**(7): 946-955.

Petersen, J. P., D. J. Jacobsen, et al. (2011). "Gaussian Mixture Models for Analysing Operational Ship Data."

Petersen, J. P., D. J. Jacobsen, et al. (2011). "A Machine-Learning Approach to Predict Main Energy Consumption."

Petersen, J. P., D. J. Jacobsen, et al. (2012). "Statistical modelling for ship propulsion efficiency." Journal of Marine Science and Technology **17**(1): 30-39.

Piñeiro, G., S. Perelman, et al. (2008). "How to evaluate models: Observed vs. predicted or predicted vs. observed?" Ecological Modelling **216**(3-4): 316-322.

Prpić-Oršić, J. and O. M. Faltinsen (2012). "Estimation of ship speed loss and associated CO2 emissions in a seaway." Ocean Engineering **44**: 1-10.

Psaraftis, H. N. and C. A. Kontovas (2014). "Ship speed optimization: Concepts, models and combined speed-routing scenarios." Transportation Research Part C: Emerging Technologies **44**: 52-69.

Rakke, J. G., M. Stålhane, et al. (2011). "A rolling horizon heuristic for creating a liquefied natural gas annual delivery program." Transportation Research Part C: Emerging Technologies **19**(5): 896-911.

Reid R. E., L. K. P., Williams V. E. (1980). Considerations in establishing a speed performance monitoring system for merchant ships: Part 2, techniques based on propeller relationships. Shipboard Energy Conservation '80 SNAME, New York.

Reid, R. E. (1985). "A condition and performance monitoring system with application to US Navy ship operations."

Ronen, D. (1982). "The Effect of Oil Price on Optimal Speed of Ships." The Journal of Operational Research Society **33**: 1035 - 1040.

Ronen, D. (2010). "The effect of oil price on containership speed and fleet size." Journal of the Operational Research Society **62**(1): 211-216.

Rose, K. A. and E. P. Smith (1998). "Statistical assessment of model goodness-of-fit using permutation tests." Ecological Modelling **106**: 129-139.

Roy, C. J. and W. L. Oberkampf (2011). "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing." Computer Methods in Applied Mechanics and Engineering **200**(25-28): 2131-2144.

Schneekluth, H. and V. Bertram (1998). Ship Design for Efficiency and Economy. Oxford, Butterworth-Heinemann.

Schultz, M. P. (2007). "Effects of coating roughness and biofouling on ship resistance and powering." Biofouling **23**(5-6): 331-341.

Skjong, R., E.B.Gregersen, et al. (1995). Guideline for Offshore Structural Reliability Analysis - General, DNV:95-2018.

Smith, T. W. P. (2012). "Technical energy efficiency its interaction with optimal operating speeds and implications for the management of shipping carbon emissions." Carbon Management **3**(6): 589-600.

Smith, T. W. P., J. P. Jalkanen, et al. (2014). Reduction of GHG Emissions from Ships: Third IMO GHG Study 2014. I. M. O. (IMO). London, UK.

Starcrest Consulting Group, L., L. Mitsui Engineering & Shipbuilding Co., et al. (2013). MAN Slide Valve Low Load Emissions Test Final Report.

Strasser, G., Takagi, K., Werner, S., Hollenbach, U., Tanaka, T., Yamamoto, K., Hirota, K., (2015) "A Verification of the ITTC/ISO Speed/Power Trials Analysis", Journal of Marine Science Technology, 20:2-13

Stulgis, V., Smith, T. W. P., Rehmatulla, N., Powers, J., Hoppe, J. (2014). Hidden Treasure: Financial Models for Retrofits. T. L. H. MCMAHON.

Sun, X., X. Yan, et al. (2013). "Analysis of the operational energy efficiency for inland river ships." Transportation Research Part D: Transport and Environment **22**: 34-39.

Taylan, M. (2010). An Overview Effect of Roughness and Coatings on Ship Resistance. INTERNATIONAL CONFERENCE ON SHIP DRAG REDUCTION (SMOOTH-Ships). M. Insel, Helvacioglu, I., Helvacioglu, S. Istanbul Technical University, ISTANBUL-TURKEY**:** 10.

Townsin R. L., Byrne D., et al. (1981). "Estimating the Technical and Economic Penalties of hull and propeller roughness measurement." Transactions SNAME **89**: 225-318.

Townsin, R. L. (1985). "Rough Propeller Penalties." Transactions of the Society of Naval Architects and Marine Engineers **93**: 165-187.

Townsin, R. L. (2003). "The Ship Hull Fouling Penalty." Biofouling **19**(sup1): 9-15.

Townsin, R. L., B. Moss, et al. (1974-1975). "Monitoring the Speed Performance of Ships." Transactions of NE Coast Institution of Engineers and Ship builders **91**: pp159-179.

Traut, M., P. Gilbert, et al. (2014). "Propulsive power contribution of a kite and a Flettner rotor on selected shipping routes." Applied Energy **113**: 362-372.

Turk, A., Prpic-orsic, J (2009). "Estimation of Extreme Wind Loads on Marine Objects." Brodo Gradnja **60**(2): 147-156.

Verhulst, M. (2007). Correlation of speed-power predictions by model tests. MARIN.

W. Tucker and S. Ferson (2003). "Probability Bounds Analysis In Environmental Risk Assessments." Applied Biomathematics.

Wahlin, B., T. Wahl, et al. (2005). Task Committee on Experimental Uncertainty and Measurement Errors in Hydraulic Engineering: An Update. World Water & Environmental Resources Congress. Anchorage, Alaska, Environmental and Water Resources Institute of the American Society of Civil Engineers.

Walker, M. and I. Atkins (2007). Surface Ship Hull and Propeller Fouling Management. Warship 2007 - The Affordable Warship.

Watson, D. G. M. (1998). Practical Ship Design. Oxford, UK, Elsevier.

Weise, K. and W. Woger (1992). "Bayesian theory of measurement uncertainty." Measurement Science and Technology(3): 1-11.