



Data-driven Vessel Performance Monitoring

Pedersen, Benjamin Pjedsted

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Pedersen, B. P. (2014). *Data-driven Vessel Performance Monitoring*. DTU Mechanical Engineering. DCAMM Special Report No. S180

General rights

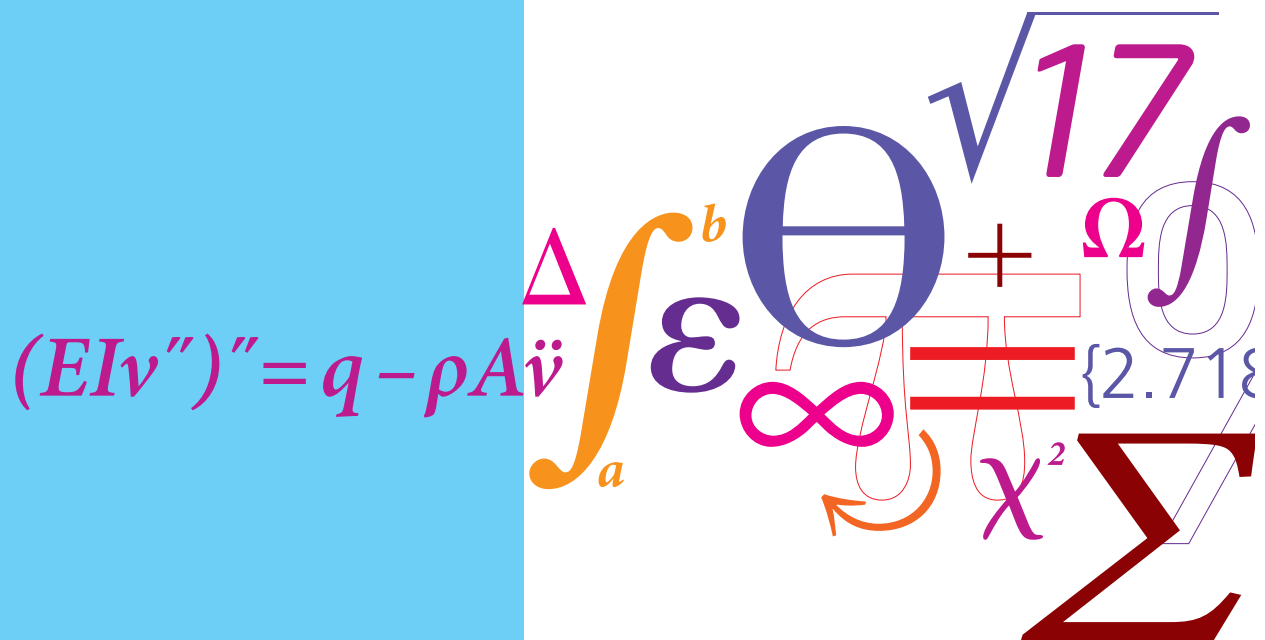
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Data-driven Vessel Performance Monitoring

PhD Thesis



Benjamin Pjedsted Pedersen
 DCAMM Special Report No. S180
 June 2014

Section of Fluid Mechanics, Coastal and Maritime Engineering
Department of Mechanical Engineering
Technical University of Denmark

FORCE Technology

Data-driven Vessel Performance Monitoring

Benjamin Pjedsted Pedersen

Supervisors: Poul Andersen
Jan Larsen
Peter Sinding

Kongens Lyngby, June 2014

Abstract

Due to increasing fuel prices and focus on the environmental impact of shipping, there is greater need for control and evaluation of ships' actual propulsion energy performance. The objective was to develop an improved method for evaluation of ship propulsion performance based on both noon reports and full-scale measurements using machine learning methods.

The fuel efficiency is expected to decrease gradually due to fouling of the hull and propeller. This effect on the propulsion performance can be difficult to determine because the ship is rarely sailing in the same condition or state at two points of time, and the difference in the fuel efficiency cannot be directly compared at two arbitrary points of time.

Empirical ship hydrodynamic methods can estimate a theoretical energy consumption that is comparable with the actual energy consumption, but the methods show significant discrepancies without proper tuning and require several ship-specific data that are often unavailable.

The machine-learning methods Artificial Neural Network (*ANN*) and Gaussian Process Regression (*GPR*) were used to illustrate the possibility of using purely data driven methods, without any other information on the ship, for prediction of the energy consumption in order to generate a trend of the propulsion performance.

ANN and *GPR* showed similar prediction performance, but *GPR* was selected as the primary method due to its ability to predict the variance and determine the influence of the input variables separately.

Data from two different ship types were studied: A product tanker, Torm Marie, and five sister containerships. A customized logging system was installed on board Torm Marie in order to sample measured data for a period of two months. Using these data it was possible predict the energy consumption (*EC*) with a prediction error of less than 1%. Using noon report (*NR*) data from the same vessel over a two-year period, the prediction error increased to 4.4%.

Noon report data from the five container ships were collected over a period of about ten years and included dry-dockings. The relative prediction errors from these data were similar to the ones for the product tanker and are thus not ship type specific.

Data from the container ships showed that 50-60 or one year of valid noon reports are sufficient for determining long-term trends of the performance.

By introducing hindcast weather information data, the prediction errors were significantly reduced for all of the tested data sets.

The prediction error from the regression methods was used to find the long-term performance trend (*VPT*) between external hull events such as dry-dockings and to evaluate the effect of these. The trends were in general in correspondence with what has been found by others, although on the smaller side.

It was attempted to use a similar method to detect events that affected the performance, this was found to be more difficult and resulted in several mis-detections.

Resumé

Et skibs effektforbrug til fremdrivning forventes at forøges gradvist over tid som følge af begroning af skrog og propeller. For at kunne bestemme ændringen i fremdrivningseffekten er det nødvendigt at sammenligne denne effekt ved en bestemt tilstand af skibet, dvs. dybgang, temperaturer, vind, fart, bølger, med effekten for skibet i nøjagtig den samme tilstand, men ved et andet tidspunkt (reference). Dybgange, fart, vind og vejr skal være identiske ved de to forskellige tidspunkter for at givet pålidelige sammenligningsgrundlag; men et skib er sjældent eller måske aldrig i præcis samme tilstand ved to forskellige tidspunkter. Der er derfor nødvendigt at udvikle metoder, der kan korrigere for dette.

En velkendt metode er at sammenligne den aktuelle fremdrivningseffekt med én, der er estimeret ved hjælp af f.eks. empiriske metoder eller ved brug af modelforsøgsresultater. Men empiriske metoder er mest velegnede til bestemmelse af fremdrivningseffekten i projekteringsfasen og for designkonditionen af skibet, og det er ikke altid, at modelforsøgsresultater er til rådighed i fornødent omfang.

I nærværende projekt er der derfor brugt statistiske regressionsmodeller, der forudser fremdrivningseffekten i en given tilstand baseret på tidligere data for skibet.

Tre forskellige sæt af data blev brugt til at evaluere disse metoder: automatisk opsamlede data over en periode på 2 måneder fra en produkttanker, manuelle dagsrapporter fra en produkttanker i periode på ca. 2 år og dagsrapport-data fra fem søstercontainerskibe indsamlet i en periode på omkring 10 år.

Den ikke-parametriske regressionsmetode Gaussisk Proces-Regression viste sig at give de bedste resultater. Datadrevne modeller er afhængige af pålidelige og hyppige data, og datasættet med automatiske opsamlede data viste sig også at give de bedste

forudsigelser.

Langtidseffekter blev undersøgt med data fra de fem containerskibe. Resultaterne viser, at der er stor forskel skibene imellem; men fremdrivningseffekten kan estimeres med en usikkerhed på omkring 5%. Ved hjælp af beregningerne blev det forsøgt at eftervise begivenheder som dokning med skrogpolering, maling og propellerpoleringer.

Preface

This thesis is submitted as a partial fulfilment of the requirements for the Industrial PhD program (Danish: ErhvervsPhD).

The project was conducted by Benjamin Pjedsted Pedersen from 2007 to 2013 in a collaboration between the section of Fluid Mechanics, Coastal and Maritime Engineering, Department of Mechanical Engineering, Technical University of Denmark and FORCE Technology. It was partly funded by the Industrial PhD program and by the Danish Centre for Maritime Technology (*DCMT*).

The project was supervised by Poul Andersen from MEK, DTU and Jan Larsen from IMM, DTU. The company supervisors from FORCE Technology were Thomas Eefsen and Peter Sinding.

Benjamin Pjedsted Pedersen
Kongens Lyngby, June, 2014

Acknowledgement

I would like to thank my supervisors, Poul Andersen for his consistent support throughout the process and Jan Larsen for his guidance within the world of machine learning and mathematical modelling. Peter Sinding has been a great help with his efficient review, unconditional support and always being ready for a fruitful discussion.

Many thanks to Christian Schack and all my colleagues in the Department for Hydro and Aerodynamics at FORCE Technology for their patience, encouragement and understanding. Particularly my former colleagues Thomas Eefsen and Kjeld Roar Jensen for their part of initiating the project and Per Bech Safft for his consistent support throughout the elongated final process.

In the final phase Søren Hattel from FORCE Technology has shown an impressive commitment with his thorough review and comments, which has been a great encouragement.

I would like to thank *D/S Torm* for their cooperation with installation of the on-board measurement systems on *M/S Torm Marie* and especially chief engineer Rasmus Hoffmann.

Many thanks to Kim Henriksen for his support with configuring the on-board logging system.

Finally I am deeply grateful for the help, support and encouragement received from friends and family. In particular Jolene for her immense support, encouragement and help throughout the process and Liva, Mathilde and Torkil for always giving me the opportunity to change focus to more practical issues and give my mind a break from theoretical matters.

Contents

Abstract	iii
Resumé	iv
Preface	v
Acknowledgement	vii
Nomenclature	xix
1 Introduction	1
1.0.1 Motivation	1
1.0.2 Performance evaluation methods	3
1.0.3 Roughness	5
1.0.4 (Bio)-Fouling and the effect on the resistance	7
1.1 Existing methods for vessel performance monitoring	8
1.2 Motivation for an improved vessel performance evaluation method	11
1.3 Data-driven models	12
1.4 Objective of the thesis	13
1.5 Structure of the thesis	15
2 Methods for predicting ship propulsion energy consumption	19
2.1 Empirical methods	22
2.1.1 Resistance	22
2.1.2 Wind resistance	26

2.1.3	Added resistance from waves	27
2.1.4	Propulsion	28
2.2	Data-driven regression models	29
2.2.1	Linear regression	31
2.2.2	Custom non-linear regression	32
2.2.3	Artificial Neural Network, <i>ANN</i>	34
2.2.4	Gaussian Process, <i>GP</i>	36
2.3	Training and testing of the regression models	40
2.3.1	Training/test sets	40
2.4	Training	42
2.4.1	Training ANN	42
2.4.2	Training GPR	43
2.5	Data-driven performance evaluation method	44
2.6	Summary of propulsion prediction and performance methods	45
3	Data collection and evaluation	46
3.1	Desired data	47
3.2	General data sources and influence	48
3.2.1	Draught and trim	48
3.2.2	Ship speed	49
3.2.3	Wind speed and direction	50
3.2.4	Sea waves	51
3.2.5	Water and air temperature	52
3.2.6	<i>time</i> as an input variable	54
3.2.7	Selection of the most significant input variables	54
3.2.8	Hindcasts	55
3.3	Data from Torm Marie	55
3.3.1	On-board measured data	56
3.3.2	Observations during voyage	59
3.3.3	Heading variation	64
3.3.4	Noon report data from Torm Marie	66

3.4	Containership data	68
3.5	Filtering	71
3.5.1	Measured data from Torm Marie	71
3.5.2	Noon report data from Torm Marie	71
3.5.3	Containership noon reports	72
3.6	Input dataset for the regression model analysis	73
3.7	Data summary	76
4	Analyses using the different propulsion prediction methods	78
4.1	Evaluation of empirical methods for the measured data	81
4.2	Comparison between prediction methods based on logged data	81
4.2.1	Results using the linear and the custom non-linear methods with the measured data from Torm Marie	82
4.2.2	Evaluation of ANN regression of the measured dataset from Torm Marie	85
4.2.3	Results from GPR with measured datasets from Torm Marie	85
4.2.4	Comparison of regression models for the measured data from Torm Marie	87
4.3	Analysis of noon report data from Torm Marie using <i>ANN</i> and <i>GPR</i>	90
4.3.1	Evaluating the prediction errors of the <i>NRTM</i> using <i>ANN</i>	91
4.3.2	Evaluating the prediction performance of the <i>NRTM</i> with <i>GPR</i>	93
4.3.3	Comparison of ANN and GPR for the <i>NRTM</i> dataset	95
4.4	<i>ANN</i> and <i>GPR</i> applied to long-term anonymous noon report data	95
4.4.1	Optimization of <i>ANN</i> regression for the noon report dataset from the containerships	97
4.4.2	Comparison of ANN and GPR using the containership data	97
4.4.3	Optimization of <i>GPR</i> regression method for noon report dataset from containership 1	98
4.5	Analysis of long-term data using <i>GPR</i>	101
4.5.1	Evaluation of the effect of the data length	102

4.6	Discussion of the results of the prediction methods	106
4.7	Evaluation of input variables using <i>ANN</i> and <i>GPR</i>	108
4.7.1	Evaluating the input variables from the measured dataset from Torm Marie	109
4.7.2	Evaluating the input variables from the noon report data from Torm Marie	112
4.7.3	Evaluation of the input variables of the containership noon re- port dataset	114
4.8	Discussion of variable analysis	118
5	Propulsion performance trend detection	120
5.1	Performance trend detection method	120
5.2	Performance trend detection results	122
5.2.1	Summary of trend detection results	123
5.2.2	Trend detection of containership 5	125
5.2.3	Discussion of Performance trend detection	127
6	Detection of events	130
6.1	Event detection method	131
6.1.1	Detection tests	133
6.1.2	Event detection for containerships 1-5	135
6.2	Discussion of event detection	139
7	Conclusion	142
7.0.1	Suggestions for further development	147
A	Modeling of Ship Propulsion Performance	1
B	Prediction of Full Scale Propulsion Power Using Artificial Neural Networks	2
C	Input/output variable data	201
C.1	Histograms of measured data measured data	201

C.1.1	Input variable	201
C.1.2	Output variable	201
C.2	Histograms of the noon report data from Torm Marie	210
C.2.1	Input variable	210
C.2.2	Output variable	210
C.3	Input variable	215
C.3.1	Output variable	215
D	Trend detection results of Container ship 1-5	221
E	Prediction results	243

Nomenclature

Symbol	Dimension	Description
Ship quantities		
L	m	Ship length, usually taken as L_{wl}
L_{wl}	m	Ship length in the water line
B	m	Ship breadth
T_m	m	Draught of the ship amidship $((T_A - T_F)/2)$
T_A	m	Draught of the ship on aft perpendicular
T_F	m	Draught of the ship on forward perpendicular
$Trim$	m	Trim of the ship $(T_A - T_F)$
C_B	–	block coefficient
A_F	m^2	Projected frontal area of the ship above the waterline
Δ	kg	The displacement
S_{APP}	m^2	Wetted area of the appendages
A_{BT}	m^2	Transverse area of the bulbous bow at FP

Condition quantities

A_C	-	Admiral coefficient, $\frac{\Delta^{2/3}U^3}{P}$
C_A	-	Allowance resistance coefficient
BF	-	Beaufort wind force
C_{AA}	-	Allowance resistance coefficient of the air resistance
C_{AS}	-	Allowance resistance coefficient of the steering resistance
C_F	-	Frictional resistance coefficient
ΔC_F	-	Change in frictional resistance coefficient
C_R	-	Residual resistance coefficient
C_{tot}	-	Total resistance coefficient
C_v	-	Viscous resistance coefficient
C_x	-	Longitudinal wind resistance coefficient
g	m/s^2	Gravitational acceleration ($g = 9.81m/s^2$)
k	-	The form factor
k_s	μm	Roughness, diameter of equivalent sand grain
K_T	-	Thrust coefficient
K_Q	-	Torque coefficient
U	m/s	speed of ship
P_D	kW	Delivered propulsion power
P_E	kW	Effective propulsion power
P_{MCR}	kW	Max. continuous rate power
R_{APP}	N	Resistance from appendages
$\overline{R_{AW}}$	N	Added resistance from waves
Re	-	The Reynolds number
R_{SW}	N	Still water resistance
R_{tot}	N	Total resistance
R_{wind}	N	Wind resistance
T_{SW}	$^{\circ}C$	Sea water temperature
T_{air}	$^{\circ}C$	Air temperature
t	-	The thrust deduction

V_R	m/s	Relative (apparent) wind speed
w	-	The wake fraction
γ_R	$^\circ$	Relative wind direction
η_D	-	Propulsion efficiency
η_H	-	Hull efficiency
η_0	-	Open water efficiency
η_{RR}	-	relative rotative efficiency
η_S	-	Mechanical shaft efficiency
ν	m^2/s	The kinematic viscosity
$\overline{\sigma_{AW}}$	-	Added resistance coefficient from waves
ρ_{SW}	kg/m^3	Sea water density (1025 sea water and 1000 for fresh water)
ρ_{air}	kg/m^3	Air density

Regression quantities

D	–	Dimension of the input
EC	–	The actual energy consumption, in kW or tons fuel per hour
\hat{EC}	–	The predicted energy consumption, in kW or tons fuel per hour
K	–	Covariance matrix or number of subset in a training/test
l	–	The GPR <i>length-scales</i> for the SE-covariance function
T_p	–	Time step length of the forward-backward test
X	–	The input variable matrix
\mathbf{x}	–	The input variable vector
x_i	–	The input variable vector for each of the i, \dots, D variables
\mathbf{x}_t	–	The input variable vector for the training data
\mathbf{y}	–	The output variable vector
\mathbf{y}_t	–	The output variable vector for the training data
ϵ	–	The training error ($f(\mathbf{x}) - \mathbf{y}$)
ϵ_t	–	The test error ($f(\mathbf{x}_t) - \mathbf{y}_t$)
ϵ_{tr}	–	The training error ($f(\mathbf{x}_t) - \mathbf{y}_t$)
Θ	–	The GPR <i>hyperparameters</i> ($l_1, \dots, l_d, \sigma_f, \sigma_n$)
σ_f^2	–	The GPR signal variance
σ_n^2	–	The GPR noise variance
σ_R	–	The relative predicted standard deviation from GPR
σ_R^2	–	The relative predicted variance from GPR
σ_ω	–	The standard deviation of the relative prediction error ω
ϕ	°	Heading of the ship
ω	–	The relative prediction error of EC $\omega = \frac{\hat{EC} - EC}{EC}$
$\bar{\omega}$	–	The average of the relative prediction error ω
$\overline{\omega_K}$	–	The average of the cross-validated relative prediction errors ω

Abbreviations

AHR		Average Hull Roughness
ANN		Artificial Neural Network
ARD		Automated Relevance Determination
COG		Course Over Ground
CS		Containership
DD		Dry-docking
GP		Gaussian Process
GPR		Gaussian Process Regression
HC		Hindcast
HCL		Hull cleaning
GH		Guldhammer and Harvald Power Prediction Method
HM		Holtrop-Mennen Power Prediction Method
IVC		Input Variable Combination
LT		Long term
LIN		Linear regression method
MDTM		Measured Data from Torm Marie
NH		Number of hidden units
NONLIN		Custom Non-linear regression method
NR		Noon reports
NRCS		Noon report Data from Containership #1-#5
NRTM		Noon report Data from Torm Marie
SFOC	tons/day	Specific Fuel Oil Consumption
SOG	knots	Speed Over Ground
VPI		Vessel Performance Index or ω
VPT		Vessel Performance Trend

Chapter 1

Introduction

The propulsion performance of a ship in service is an expression of the energy consumption to drive the ship through the water at a certain state (speed, loading and weather condition), relative to a previous *state* or a reference *state*.

Over a ship's lifespan, the energy consumption is generally expected to increase as the ship's performance efficiency declines which means that the fuel consumption will increase for a certain speed or the speed will be reduced for a certain fuel consumption. This performance reduction is mainly due to fouling of the hull and propeller, but can also be attributed to other irregularities of the hull and propeller surface caused by damages or erosion. Fouling can be cleaned off the hull during dry-docking, however damages and erosion are almost impossible to repair, so the ship never returns to the same performance condition as, in the original state.

1.0.1 Motivation

Evaluation of ship propulsion performance attracts the attention of ship owners mainly because the bunker expense is becoming an increasingly part of the total service costs of a ship. The performance deterioration or increase in energy consumption can vary from a few percent a year to a dramatic 30-40% in a short time (Carlton (1994)). For example, the fuel consumption of a large containership is about 250 tonnes (metric) per day, and if the price of fuel oil is 534USD/tonne (2008), a modest 5% decrease in

performance will result in a 6,675 USD daily penalty or, if the ship is sailing 200 days per year in a substantial 1,355,000 USD annually.

In addition to fuel consumption, the discussion of global climate change has forced ship owners as well as customers of ship transportation to consider their environmental image. It is estimated that the shipping industry accounted for 3.3% of the global CO_2 emissions in 2007(IMO (2009)). A better understanding of ship performance can show that ship owners care and are actively working to reduce their footprint on the environment.

The International Maritime Organization, IMO, states that more than 90% of global trade is carried by sea, and the total trade in tonnage is expected to increase as evidenced by the increase from 2,500 mil. tonnes in 1970 to 8,400 mil. tonnes in 2010, although with a short stagnation during 2009 (IMO (2012)). In order to transport these goods, it is estimated that 339 mil. tonnes of bunker fuel was consumed in 2008 (IMO (2009)).

A great deal can be saved by optimizing the logistics of shipping operations. The ship speed is a governing factor for energy consumption, and by optimizing the voyage by travelling at optimal speed for the entire duration of the journey, energy consumption can be minimized. Optimal planning of the route due to weather and water depths are also relevant parameters in this optimization.

Another short-term energy consumption factor is the trim of the vessel defined as the draught aft subtracted by the draught forward. The effect of the trim on energy consumption varies among different ship types, where e.g. containerships are rarely loaded to maximum draught and are often able to change the trim by moving ballast water. Other ships such as bulk carriers and oil tankers are often in close to full load or in ballast loading condition. Furthermore the design speed of the vessel has an important effect on how sensitive it is to changes in the trim. Trim is thus of increasing interest to ship owners and operators (Larsen et al. (2012)).

The effect of fouling varies significantly depending on the ship, operation profile, anti-fouling paint, etc. Knowledge of the fouling growth rate makes it possible to determine if or when a propeller or hull cleaning is economically beneficial, or if it is profitable to conduct a premature dry-docking, rendering it very useful information

for a ship operator.

Another trend in the maritime industry is the growth in fuel saving methods and devices that are being installed, and tested in order to reduce the energy consumption by only a few percent. The final stage in the development of these methods is often based on model test experiments or *CFD* studies, but when a device like a newly designed propeller is installed it may be hard to identify small improvements in the energy consumption of the real ship due to the large scatter of existing performance evaluation methods. An improved method that can detect small percentage changes would give the supplier and customer of fuel saving products better confidence in how the ship performs in real life. A good example of this is given in Andersen et al. (2005), where the performances of two sister ships (product tankers) one equipped with the new KAPPEL propeller and the other one with the a conventional propeller, are described. The reduction in power was anticipated to be 6%, and a 4.4% reduction was shown by over two years of sea trials.

Similarly manufacturers of anti-fouling paint are also keen to document the efficiency of a new product in a service condition.

1.0.2 Performance evaluation methods

Generally ship propulsion performance is an expression of how well a ship performs versus how well it should perform. Evaluation by comparison of the energy consumption can be described in the following ways:

1. By comparing the actual measured or reported energy consumption EC_0 at time t_0 with the energy consumption EC_1 at another time t_1 for the exact same *state*.
2. By using a model based on empirical methods, model test or *CFD* to estimate the theoretical energy consumption of the vessel in the present *state* and comparing it with the actual energy consumption.

A ship is sailing at different draughts and speeds, and furthermore it is exposed to many external effects that influence the energy consumption, i.e. wind speed, temperatures, waves. Although the first method is the ideal method, it is difficult to apply,

as the exact state where all factors are identical is unlikely to occur more than once. All the parameters have to be identical at two different times in order to give a reliable picture of the performance of the ship, e.g. the measured difference in the propulsion power.

The second method is based on empirical methods which, apart from the general operational and environmental properties (i.e. speed, draught, wind, waves), also rely on many input parameters of the hull shape and propeller. Empirical propulsion prediction methods are mostly used in the design phase of ship design projects where they are a useful tool for comparing designs and estimating engine and propeller properties. They are partially based on statistics from model tests at the design draught. This makes them vulnerable to other loading conditions. However, it is possible to tune the empirical methods with reliable full scale measurements or model test data.

Instead of evaluating the performance effect on energy consumption, different propulsion related variables can be used. For example by evaluating the measured speed at a certain power compared with a reference speed, a so-called speed index, as described by e.g. Hansen (2011). In Eefsen (1996), many different indices are established by comparing the theoretical values with the actual values.

Instead of comparing the energy consumption at a certain *state*, other measures can be used for performance evaluation such as "Added resistance" by Propulsion Dynamics Inc. Munk (2006) or the average hull roughness by Malone et al. (1980).

The predicted energy consumption \hat{EC} can be compared with the actual reported or measured energy consumption EC , and the relative difference ω as presented in (1.1) is then an expression of how well the ship performs compared to how it "should" perform - also referred to as a *Vessel Performance Index, VPI*. The trend of the development in performance over time t can be defined as the Vessel Performance Trend, *VPT*, which is assumed to be linear with a slope α and a constant β .

Figure 1.1 shows a typical decrease in performance over time where the performance is determined by the relative difference or prediction error ω . The dashed vertical lines are the dates of hull cleanings, and the solid vertical lines indicate dry-dockings.

$$\omega = \frac{\hat{EC} - EC}{EC} \quad (1.1)$$

$$VPT(t) = \alpha t + \beta \quad (1.2)$$

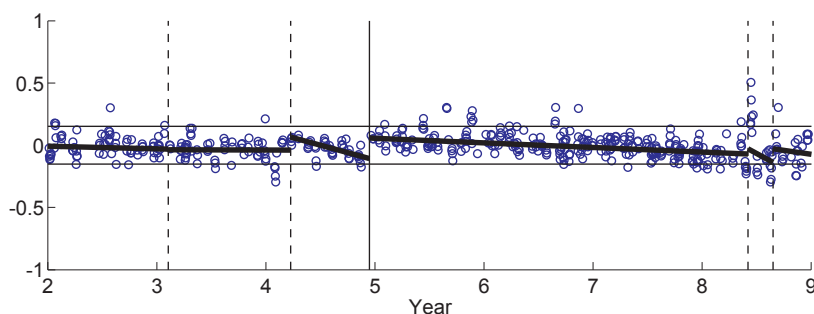


Figure 1.1: Example of performance decrease ($\omega = \frac{\hat{EC} - EC}{EC}$) over time where the solid vertical lines represent dry-dockings and the dotted hull cleanings. The solid trend lines illustrate the VPT

Large abrupt changes in ω indicate the occurrence of an event that affect the performance. Performance relevant events such as a dry-dockings are often known to the shipowner and crew, but unknown events such as propeller damage are also possible.

1.0.3 Roughness

In order to understand the effect of fouling on the propulsion power, the roughness of ships is introduced, and a common approach used is presented in Chapter 2

The roughness of the hull skin has a significant effect on the frictional resistances of a ship because it affects the boundary layer and boundary layer development over the hull. Carlton (1994) suggest to view the roughness in two separate components, so that the hull surface roughness consists of the sum of **permanent roughness** and **temporary roughness**. The permanent roughness relates to the base surface conditions from a newly launched ship where the surface is clean, but still has an uneven

skin due to e.g. welding seams, bowing of plates and "hungry horse" phenomena, which in total is estimated by Carlton (1994) to increase the energy consumption by approximately 1.75%.

Permanent roughness can also include damage, corrosion or old paint as described by Carlton (1994). These "scars" are impossible or difficult to remedy which explains why ships' propulsion performance most likely never will return to the same level as at the launching. The temporary roughness relates to the additional roughness that can be removed such as fouling.

Roughness has traditionally been defined as the diameter of a sand grain k_s , but is in practice substituted by measurements with a Hull Roughness Analyser from e.g. BMT (British Maritime Technology) where the maximum difference in height of the surface is found within a length of 50mm, also defined as $R_t(50)$ in King (1982). International Marine Coatings (International (2004)) recommends performing 10-15 measurements over a length of 750-1000mm. This is performed at about 100 locations on the submerged hull surface, and ultimately the Average Hull Roughness (AHR) can be determined, which represents the overall hull surface roughness. AHR is used to find the frictional resistance coefficient for the actual hull surface condition.

The frictional part of the total resistance varies among different ship types, where the friction fraction (C_F/C_{tot}) is higher for blunt hull forms than slender hulls. Numbers of 0.67 for a product tanker and 0.62 for a containership are suggested in Carlton (1994), where a ULCC can reach a frictional resistance fraction of about 0.85. Furthermore, the local speed of the flow varies significantly from different areas on a ship hull, and consequently the local frictional resistance also varies. Walderhaug (1986) performed a series of measurements on plates with different roughness and a range of speeds and established a more detailed frictional resistance coefficient based on the ship hull form where the hull is divided into local frictional speeds and roughness.

A simple formula for evaluating the increase in propulsion power ($P_2 - P_1$) due to the change in roughness is presented in Townsin et al. (1980) as in (1.3) where $A = 5.8$ is proposed based on a series of roughness measurements, and P_{MCR} is the maximum continuous power rate of the main engine.

$$\frac{P_2 - P_1}{P_{MCR}} \% = A \left(k_2^{1/3} - k_1^{1/3} \right) \quad (1.3)$$

The effect of the propeller roughness can also contribute significantly to the propulsion efficiency and has been evaluated in Townsin et al. (1985) where measurements of the propeller blade roughness were taken at several different locations. This led to an increase of 3.4% over a year compared with a newly polished propeller. It was estimated that the payback period of a propeller polish at 1985 prices would be five weeks assuming the ship is at sea 200 days per year.

1.0.4 (Bio)-Fouling and the effect on the resistance

Hull and propeller fouling are general terms to describe marine growth that attaches to a ship and can include anything from slime to shells. Biologically the fouling can be divided into micro-fouling and a macro-fouling where examples of micro fouling are the algae attachments such as "slime" and macro fouling refers to barnacles and seaweed (Callow and Callow (2002)). The development of fouling starts the moment an object is immersed into seawater with micro fouling, developing to finally include shells and seaweed.

This build-up of fouling varies over time and depends on many factors such as the geographical area and climate of the operation (tropical or arctic), the ship's service speed and type of anti-fouling paints (Carlton (1994)). In Bertram (2000), an added resistance of 3-5% in 40 days due exclusively to slime was reported.

The amount and type of fouling also varies for different locations on the hull surface, and evaluations of the effect from fouling measurements should treat different areas separately as in Townsin et al. (1981).

The increase in roughness from fouling varies not only as an effect of the speed and operation profile of the ship, but also due to use of different anti-fouling paints as described by Carlton (1994) where modern self-polishing paints increase by 10-30 $\mu/year$ and traditional paints 40-60 $\mu m/year$.

Townsin et al. (1980) demonstrates how the rate of roughness increases with higher

initial roughness, so for a relatively high initial roughness the roughness increases with $20 \mu\text{m}/\text{month}$ and for a very smooth surface with an initial roughness of $60 \mu\text{m}$ is argued to have no increase in roughness over time and to be maintenance free.

A thorough study was performed by Townsin et al. (1986) where ship roughness measurements from 47 different ships over ten years were collected resulting in 147 measurements. The data show a trend of $40 \mu\text{m}/\text{year}$, although the latter part of the period had a lower rate due to better paints applied, and it is shown that fouling has dropped to about $20 \mu\text{m}/\text{year}$. Initial roughness from new buildings was reported by three different data sets, one including 13 ships reporting a mean value of $129 \mu\text{m}/\text{year}$ (1976-79), one including 111 datasets reported $113 \mu\text{m}/\text{year}$ (1983-85), and measurements based on 6 new buildings reported a mean value of $107 \mu\text{m}/\text{year}$ (1980-85)

Using (1.3) and knowing the roughness increase, the effect on the propulsion power has been estimated to increase 3.5% over 2.5 years by Eefsen (1996), assuming a growth rate $20 \mu\text{m}/\text{year}$.

Other examples are given by Willsher (2007) where slime is reported to increase the resistance by 1-2%, seaweed up to 10% and shells up to 40%.

1.1 Existing methods for vessel performance monitoring

As previously outlined the ship propulsion performance can be estimated by comparing the actual measured or reported energy consumption EC with the theoretical determined value given the same conditions.

In Carlton (1994), a comprehensive overview and discussion of ship performance monitoring methods are presented. A crude method is to evaluate the *Admiralty coefficient*, A_C ((1.4)), where Δ is the displacement of the ship, U is the ship speed and P is the shaft propulsion power. P can alternatively be replaced by the fuel consumption. The method does not account for wind and sea and should only be used for comparative loading conditions.

$$A_C = \frac{\Delta^{2/3} U^3}{P} \quad (1.4)$$

Carlton (1994) furthermore describes a procedure that finds the theoretical values for the hull and propeller, and by including theoretical effects from the weather and sea, ratios of different performance variables can be evaluated.

An example of empirically based ship performance monitoring is demonstrated by Eefsen (1996) where performance evaluation was carried out on six product tankers and one bulk carrier. The sea trial data was compared with estimates from Holtrop and Mennen (1984) and Harvald (1983) in order to determine which of the two methods was most suitable for the present ship (Product tanker). The Holtrop and Mennen (1984) method showed the best results and was used exclusively. The air and wind resistance coefficients were determined by Blendermann (1990), and the added resistance in waves was estimated by the tabular values in Bhattacharyya (1978). Furthermore, the wave heading correction factors from Townsin et al. (1993) were used. Using these methods, it was found that the influence of the weather on the test ship in *Beaufort* 5 increased the resistance by approximately 17%, with 13% due to wind and 4% to waves. Many other factors were also included in the model, e.g. temperatures and fuel calorific values. With the final model it was possible to establish different ratios between the calculated values and the measured values, e.g. resistance, power, daily fuel consumption. Linear regression was used to evaluate the long-term trend based on the data with a significant scatter. Due to the empirical resistance calculation, the method is very sensitive to changes in the draught and trim, but if the model is supported by e.g. model tests, it can be a robust tool to use without having any previous data.

A detailed study of the effect of the hull roughness was carried out by Malone et al. (1980) where the main focus was on planning hull cleanings and dry-dockings. Empirical methods were used to estimate the effective propulsion power due to the hull roughness, the wind and waves. By isolating the hull roughness, the change in the roughness was used to establish a "Ship Economic Model" based on the change in

propulsion power and fuel consumption. Simulations using the model made comparisons between dry-docking and hull cleanings for a containership, a bulk carrier and a tanker. For the scenarios in the study the optimum docking intervals were respectively 2.1, 2.05 and 2.0 years. Sensitivity analyses were performed to evaluate the effect of changes in oil prices, port fouling severity, time at sea and cargo load.

In Andersen et al. (2005), results from model tests were used to establish a model that found the propulsion power at a reference speed. Linear interpolation between the tested conditions was used to find the still water resistance for the actual loading conditions, and wind and wave effects were accounted for using the methods by Isherwood (1972) and Gerritsma and Beukelman (1972), respectively. The main objective was to detect the performance difference of two sister ships, one fitted with a normal propeller and the other with a KAPPEL propeller^a. The trend due to fouling was found to $34kW/month$ corresponding to $0.47\%/month$ or $5.68\%/year$ relative to the normal continuous power rate of the machinery.

A recent work by Logan (2011) studied the hull performance of two sister tankers from the US Navy with controllable pitch propellers. On-board data collection made it possible to make a propeller model which was compared with later measurements, and the relative comparison is referred to as the Key Performance Index. The comparison was only carried out for two loading conditions but illustrated how the effect of being at berth for a few months can increase the energy consumption of up to 35%.

J.M.J. Journée has used his experience in ship routing and heavy weather guidance (in Journée (1976) and Journée and Meijers (1980)) to use full scale measurements for performance surveillance in Journée et al. (1987). The method is based on traditional ship propulsion theory where the essential coefficients, the residual resistance coefficient C_R , thrust coefficient K_T , torque coefficient K_Q , wake fraction w and thrust deduction t are fitted to polynomial expressions based on the full scale measurements. The wind resistance is based on Isherwood (1972), and wave resistance is calculated by Gerritsma and Beukelman (1972) for head seas and Boese (1970) for beam and following seas. Different variations of the system were also used to post-process full scale

^aThe KAPPEL propeller concept design developed by JJ Kappel

measurements carried out earlier in Journée (2003a) and Journée (2003b). The analyses and measurements show very good agreement and give valuable insight into the fuel consumption depending on different operations, e.g. trimmed conditions or operating a twin screw ship with only one propeller, but no long trends of the performance have been evaluated.

Recently Hansen (2011) showed how detailed simulations could be used for performance monitoring, using automatically sampled data collected from many sources on a containership over a period of a year. Within this period a propeller polish was performed, and through the relative change in roughness, speed and power the effect from this and the linear trends were studied using different filtering. The model demonstrated a 2% reduction in the power due to the propeller polish, and the slopes of the power trend varied from 3.1%/year and 3.8%/year increase in the propulsion power before and after the propeller polish.

A data-driven approach is described in Petersen et al. (2011) where a Neural Network model trained with measured data from one of *DS Norden's* product tankers was used to predict the energy consumption with an accuracy of about 1.5%. Leifur P. Leifsson et al. (2008) explored the difference between empirical models *White-box*, a Neural Network *Black-box* and a combination of an empirical and a Neural Network *Grey-box*. There is no clear indication of a superior model, but it indicates that the *Grey-box* is better in unusual environmental conditions. The data analysed were gathered over 5 months and no long-term effect is expected to be detected.

The recently finished Ph.d. thesis Petersen (2012) also explores different statistical approaches for estimating the energy consumption using measured data.

1.2 Motivation for an improved vessel performance evaluation method

As shown in Eefsen (1996), a scatter of $\pm 10\%$ in added resistance from fouling is not uncommon, using simulations with automatically sampled data the method as described by Hansen (2011) reveals a scatter of only about $\pm 2\%$ although very few

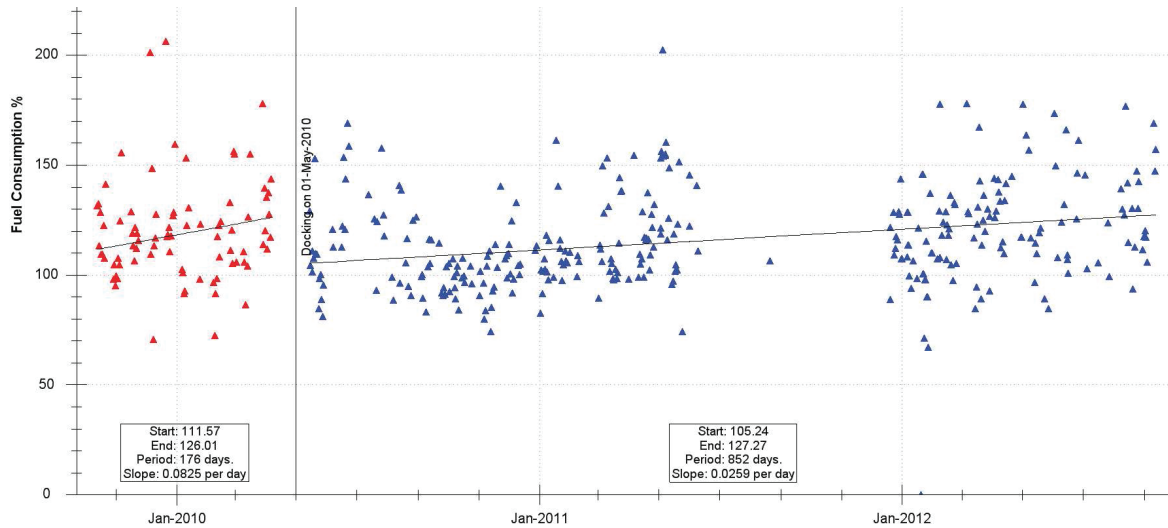


Figure 1.2: Example of the relative change in the fuel consumption of a tanker 300,000 dwt crude oil tanker evaluated by SeaTrend[®]

points are available for the linear trend.

Using noon report data demonstrates significant scatter as also illustrated in Figure 1.2 where an extract from the performance monitoring system Seatrend[®] developed by FORCE Technology is shown. Noon reports are by far the most common method for collecting performance data and hence have a larger market potential, so supporting noon reports data is desirable. Reducing the scatter would result in a more reliable trend and make it possible to use less data to detect a reliable trend and thus give the ship operator an earlier warning of significant changes in the performance.

1.3 Data-driven models

Data-driven models, also referred to as "Machine learning" or "Pattern recognition", are being used in an increasing number of applications in many different fields. Several different machine learning methods exist and can be used for problems such as speech recognition and multivariate regression models for prediction of weather forecasts. Data-driven methods train or learn the data to estimate if parts of a dataset or

grouped allocations of a dataset are related. The methods are often grouped into *supervised learning* and *unsupervised learning*. In *supervised learning* problems, an input vector (\mathbf{x}) has corresponding target outputs (\mathbf{y}). When the model is trained to match certain discrete categories, e.g. recognition of hand written digits, it is called "classification" and when training the model with respect to a continuous output variable is referred to as regression.

In *unsupervised learning*, no targets are available, and it is attempted to find groups of similar data in the input data (\mathbf{x}), called *clustering*, or to find distributions of different *clusters* (*density estimations*).

Until recently, maritime applications of data-driven methods have mostly included manoeuvring analysis and predictions as demonstrated by Hess et al. (2009) and Petersen and Lauridsen (2000). However, they have also been used by Hess et al. (2006) for interpolation of propeller open water coefficients using *Artificial Neural Networks* (*ANN*).

1.4 Objective of the thesis

The objective of this thesis is to establish an improved method for evaluation of propulsion performance. The method should be adaptive and able to adjust to measured and collected vessel performance data. In this context we will investigate how so-called data-driven methods can accommodate this. The data-driven method should be based partly or entirely on data which includes the traditional noon report data supported by automatically collected data that is comparatively of a much higher data density and varies more over time.

The weather information from hindcast data will be evaluated and implemented alone or as a supplement to the existing weather data collected in noon reports or by automatic measurements.

In order to predict the energy consumption (EC) for a given *state* of the ship, a suitable regression model has to be determined. Multivariate non-linear models such as Artificial Neural Networks (*ANN*) which are expected to be appropriate will be tested as well as linear models.

Regression models with different input datasets will be evaluated to clarify their effect on the energy consumption, e.g. the effect of using hindcast data versus observed data and automatically sampled data versus noon report data.

When a suitable prediction method is established, it is possible to compare the predicted energy consumption \hat{EC} with the actual reported or measured energy consumption EC , and the relative difference or prediction error ω of those is then an expression of how well the ship performs compared to how it "should" perform at a given *state* and time as defined in (1.1). This reveals that if the actual energy consumption is higher than the predicted, the relative prediction error ω decreases and indicates a negative change in the ship propulsion performance compared with the period of the training set.

The prediction method can be trained with a dataset representing a sufficient variation of *states*. Ideally it should be independent of time. However, this is not possible using full scale sampled data, so the shortest suitable time for a reliable training should be estimated. Using a training like this, based on training of e.g. the first year after launching of a ship, prediction can be made for any actual *state* of the ship and compared with the actual value, and the development over time can be evaluated. The trend of the relative prediction error ω thus represents the total performance of the ship and can be expressed as a function of the time t and be defined as the Vessel Performance Trend *VPT*. The *VPT* can only be found for periods where a continuous trend of the performance is expected, so for every abrupt change such as a dry-docking or a hull cleaning, an individual *VPT* must be found as illustrated in Figure 1.1. The *VPT* is assumed to be linear based on previous studies such as Carlton (1994) and Munk (2006).

Using a similar methodology as for detection of the *VPT*, detection of events should also be tested.

Summary of sub-objectives includes:

1. Evaluate dataset
2. Find the best regression method
3. Evaluate predictions by using different training sets: automatic, noon reports with or without hindcast weather information.
4. Determine the best combinations of input variables and what variables are most important for the predictions
5. Understanding the effect of the length of input training data
6. Use the initial training to predict the energy consumption of the actual states at a certain time and detect the trend of the relative prediction error ω over time
7. Detection of abrupt changes in the performance trend

1.5 Structure of the thesis

An outline of how the thesis is structured is given below.

Chapter 2 - METHODS FOR PREDICTING SHIP PROPULSION ENERGY CONSUMPTION

This chapter describes the background and basic principles of hydrodynamics and empirical methods for estimating the resistance and propulsion power. The overall assumptions and individual components of these methods are described which introduce the basic background of the physics which will be used in the regression models. A basic multivariate linear regression model is described, and a custom non-linear method is derived which incorporates the fundamental features of the empirical models.

Due to the previous models' limited performance for complex problems, an Artificial Neural Network (*ANN*) is introduced. The structure of *ANN* is described to give an understanding of the models' governing parameters.

Gaussian Process Regression (*GPR*) is introduced due to its flexibility and ability to predict the variance for each prediction, along with the characteristic *length-scales*. A description of *GPR* is given to explain the basic mechanics of GPR as well as the advantages and disadvantages of this method.

The data filtering of the three input data sources and a description of the subdivision of the data sets input training and test sets. Furthermore, special training techniques for *ANN* and *GPR* are described.

Chapter 3 - DATA COLLECTION AND EVALUATION

This chapter presents the three data set sources: measured data from Torm Marie (*MDTM*), noon report data from Torm Marie (*NRTM*) and noon report data from five container ships (*NRCS* 1-5). A short presentation of the desired data for the evaluation and the data collection obstacles encountered for the basic input variables is given.

Installation and data sampling of the on-board logging system is described with channel lists, logging intervals and data sources, together with a statistical presentation of the four measured data sets.

Similarly, the collection and statistical representation of noon report data from Torm Marie are presented.

The data from the five containerships is collected over a longer period of time and also includes dry-dockings.

The quality of hindcast weather information is described and compared with the observed and measured data.

Chapter 4 - ANALYSES USING DIFFERENT PROPULSION PREDICTION METHODS

The chapter shows the prediction performance by the relative prediction errors ω for various combinations of input variables using the methods and procedures described in Chapter 2. The outcome of different lengths of the datasets are evaluated, and the variance of the prediction models are found by *GPR*. Analysis of the relevance of each of the available input variables by *ANN* and *GPR* regression methods are presented.

Chapter 5 - TREND DETECTION OF THE PROPULSION PERFORMANCE

In this chapter, post-processing of the prediction errors from Chapter 4 is used to find the trends of ω over time or Vessel Performance Trend (*VPT*), which is an expression of the performance development. This trend shift is abrupt after events like a dry-docking or hull cleaning, and the effect of this can be evaluated between these events.

Chapter 6 - DETECTION OF PERFORMANCE EVENTS

This chapter attempts to predict abrupt changes in the performance trend as described in the previous chapter using a piecewise trend analysis.

Chapter 7 - CONCLUSIVE DISCUSSION

Chapter 7 presents a summary of the main conclusions and discusses how it is possible to use data-driven models for performance monitoring and the benefits and disadvantages of the methods. Suggestions for further development and exploration of the topic are discussed.

The following publications have been produced during this thesis:

Pedersen and Larsen (2009a) and **Pedersen and Larsen (2009b)**

They can be found in Appendix A and B

Chapter 2

Methods for predicting ship propulsion energy consumption

The objective of this chapter is to describe traditional ship propulsion theory and the background assumptions on which this theory is built. In order to establish a data-driven approach for performance evaluation, four regression models are presented as plausible methods to predict the energy consumption (EC) for an arbitrary "state" of a vessel: linear, custom non-linear, Artificial Neural Networks (ANN) and Gaussian Process Regression (GPR). The prediction errors of the regression models will ultimately be used to evaluate the instantaneous or long-term performance of the vessel.

The energy needed to drive a ship at a certain speed in certain conditions depends on several variables as illustrated in Figure 2.1. The term *state* refers to a certain set of input variables that influence the ship propulsion energy consumption.

Historically, energy consumption has been defined by empirical resistance and propulsion methods that require many ship specific parameters such as the length, breadth, propeller diameter, wetted surface, etc. This information is not always available or assessable. For example, shipyards may be reluctant to hand out details about hull form to the ship owner. If the parameters are unknown, many assumptions must be introduced to the models, and the predictions become uncertain. A different approach is obtained by focusing on the collection of data for different *states*. This is

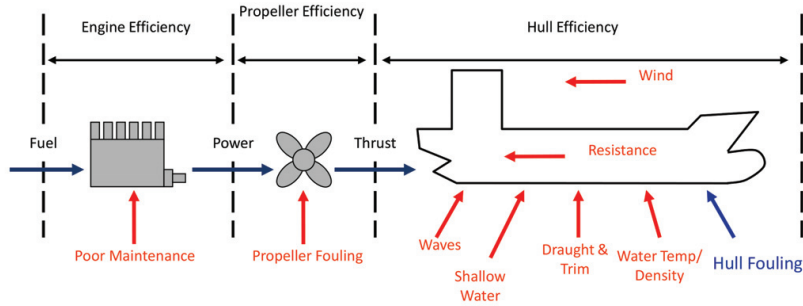


Figure 2.1: Variables of the propulsion performance

referred to as a data-driven approach. Using data-driven predictions, the fixed parameters and information about the ship are irrelevant. The only necessary variables are those that affect the ship propulsion performance.

This chapter first introduces the general principles and assumptions of ship propulsion and hydrodynamics and how it is possible to estimate the resistance and propulsion power by both traditional empirical and data-driven methods.

Two specific empirical methods, Harvald (1983) and Holtrop and Mennen (1984), have been used extensively in the past decades and are regarded as industry standards. These methods are based on ship models and full scale tests, primarily in design loading conditions, or sea trial conditions to give estimates based on a design loading condition. This is useful in the design phase of the shipbuilding process and in situations where the ship is sailing in loading conditions close to the design draught. However, the methods have greater uncertainty in off-design conditions. Both methods offer some flexibility, and experienced users can tune them with experimental data to improve the results, but using the methods without any adjustment can lead to large discrepancies in certain conditions. One of the central issues with the empirical methods is accounting for the effect of the bulbous bow in off-design conditions where the bulbous bow is e.g. slightly out of the water. This creates unpredictable effects on the resistance and can only be found by model tests or *CFD* simulations. In recent years, ship owners have shown an interest in the resistance and power needed in many other loading conditions which has led to several trim model tests at FORCE Technology such as the tests evaluated in Larsen et al. (2012).

The still water ship resistance can be determined reasonably well by Computational Fluid Dynamics, *CFD*, Simonsen et al. (2008). However, it is both time consuming to build a grid of the hull and computationally demanding to solve the numerical problem, although faster hardware and software are seen as an increasing threat to tank model tests. Tank model tests are often still preferable for their reliability and when a ship model is already in the tank, it can also be faster to test another speed or loading condition, where *CFD* only finds the resistance for one speed and draught per simulation.

Data-driven models were introduced because they were believed to be able to model many of the non-linearities of ship hydrodynamics which is difficult when empirical models are used. Naturally, data-driven methods are data based, and the amount and quality of the data is important for proper modeling.

The data consist of a number of different input variables and a single output - the energy consumption represented either by the measured fuel consumption or the propeller propulsion power.

Every data point represents a certain *state* of the input variables and has a corresponding output. The data sets are collected by a number of data points from the same stationary period, i.e. that the variables do not vary with time during that period. The use of stationary datasets gives the opportunity to make predictions for an arbitrary state of the vessel, given all the input variables are within the boundaries of the data. The predicted output can then be compared with the actual measured value for the same state, and the relative difference subsequently represents the change in energy performance between the present time and time of origin of the dataset used for training.

Assuming that the fouling affects the propulsion power linearly over time, the *time* as a linear variable, increase 1 per day, was included for some predictions in order to evaluate the influence of the *time* variable on the prediction results. A lower prediction error by using *time* would then indicate the effect of the linear fouling trend over time. Inclusion of time as a variable makes the dataset non-stationary over the training data period, and it cannot be used for predictions outside the range of the input variables and thus cannot be used for predicting in the future and comparing with actual values

previously described.

Different regression models have been used in an attempt to find the best model to predict the output \mathbf{y} (EC) based on the continuous input variables $\mathbf{x} = (x_1 \cdots, x_D)$ where D is the dimension of the input variables. The methods applied in this thesis are presented in the following sections.

2.1 Empirical methods

The traditional way of predicting the propulsion power is described below in order to give an overview of the main components and assumptions in the empirically based methods. The advantages and disadvantages of empirically based methods are discussed and compared with data-driven methods.

2.1.1 Resistance

Different methods for calculating ship resistance and propulsion power are available. It is generally accepted that the ship resistance in calm water (R_{SW}) can be found based on the ship speed U , the wetted surface S , the seawater density ρ_{SW} and the total resistance coefficients C_{tot} as shown in (2.1) Pedersen et al. (2002). Adding resistance components from the environmental effects such as the wind R_{wind} and waves $\overline{R_{AW}}$ leads to the total resistance R_{tot} (2.3).

$$R_{SW} = C_{tot} \frac{1}{2} \rho_{SW} S U^2 + R_{AAP} \quad (2.1)$$

$$\text{where } C_{tot} = C_v + C_R + \Delta C_F + C_A + C_{AA} + C_{AAS} \quad (2.2)$$

$$R_{tot} = R_{SW} + R_{wind} + \overline{R_{AW}} \quad (2.3)$$

The coefficient C_{tot} consists of the following ITTC (2011):

C_v Viscous resistance coefficient

C_R Residual resistance coefficient

ΔC_F Roughness allowance coefficient

C_A Allowance resistance coefficient

C_{AS} Allowance resistance coefficient for the effect of steering

C_{AA} Air allowance resistance coefficient

R_{AAP} Allowance resistance for appendages

Viscous resistance coefficient, C_v ,

The Viscous resistance coefficient, C_v , represents resistance components related to the friction C_F of the hull surface. The skin friction coefficient for a flat, smooth surface with a turbulent flow can be found by the *ITTC 1957* (ITTC (2011)) procedure: $C_f = \frac{0.075}{(\log_{10}(Re)-2)^2}$, where Re is the Reynolds number defined as: $Re = \frac{L_{wet}U}{\nu}$. In order to account for the different pressure distributions of a 3D flow around a ship hull, the *the form factor* k is introduced to obtain the *viscous frictional resistance*, $C_v = C_f(1 + k)$. k normally has values in the region of 0 – 0.25.

The frictional resistance is dependent on the roughness of the surface, and the calculation of the frictional resistance coefficient above is based on a ship with a smooth surface.

Residual resistance coefficient, C_R

The residual resistance coefficient C_R mainly accounts for the energy radiated by waves made by the ship. The shape of the hull form is governing this energy dissipation, and in order to analyse it, several different coefficients have been established to describe the hull form. As some modern ships have more complex hull forms including bulbous bows and submerged transoms, the shape must be described by simple measures in order to compare ship resistances.

C_R can be estimated empirically by methods described by Harvald (1983) and Holtrop and Mennen (1984). Harvald (1983) method is based on the ship length, displacement and the speed of the vessel, whereas Holtrop and Mennen (1984) uses a

more detailed description of the hull form as e.g. the prismatic and midship section coefficient and waterline angle entrance at the bow. The latter method also includes both bulbous bows and submerged transoms. The bulbous bow contribution is based on three measures: the transverse area of the bulb at the forward perpendicular A_{BT} , the height of the centroid of A_{BT} above the keel and the draught at forward perpendicular (Holtrop and Mennen (1982)). The effect of the submerged transom is described by the submerged transom area. Despite using a detailed hull description, Holtrop and Mennen (1984) lacks reliability in off design conditions.

In Larsen et al. (2012), trim tests of a large cargo ship with a bulbous bow have been investigated, and a reduction in C_R of 70% was found by trimming the vessel 2m forward compared to an even keel trim and an increase of about 30% by 2m trim aft. Summing up other contributions from a trimmed loading condition, i.e. change in waterline length and wetted surface, the change in the effective power is -8.9% and 16.2% for 2m forward trim and 2m aft trim, respectively. This change is assumed to be related to the change of the flow around the bulbous bow. Larsen et al. (2012) also finds that trimmed simulations by RANS (Reynolds Average Navier-Stokes Solver) CFD and model tests show good correlation as opposed to the potential flow theory CFD which deviates considerably from the model tests.

Roughness allowance coefficient, ΔC_F

The *skin roughness* has been investigated by gluing sand grains with the diameter k_s on a plate with a smooth surface and measuring the frictional resistance due to the difference in roughness (Wadskjær (2001)). ITTC (ITTC (2011)) proposed use of the frictional resistance coefficient:

$$\Delta C_F = 0.044 \left[\left(\frac{k_s}{L_{wl}} \right)^{1/3} - 10 \cdot Re^{-1/3} \right] + 0.000125 \quad (2.4)$$

where k_s is the equivalent sand grain diameter, L_{wl} is the waterline length and Re the Reynolds number.

If measured data is not available, $k_s = 150 \cdot 10^{-6}m$ is assumed for a new ship.

In practice the sand grain diameter k_s is substituted by measurements with a Hull Roughness Analyser from e.g. BMT (British Maritime Technology) where the maximum difference in height of the surface is found within a length of 50mm also defined as $R_t(50)$ in King (1982). It is recommended by International Marine Coatings (International (2004)) to do 10-15 measurements over a length of 750-1000mm. This is performed at about 100 locations on the submerged hull surface, and ultimately the Average Hull Roughness (AHR) can be determined which represents the overall hull surface roughness. AHR is thus used to find the frictional resistance coefficient for the actual hull surface condition.

Allowance resistance coefficient, C_A

The allowance resistance coefficient, C_A , accounts for differences in the hull roughness between the model and the ship. In ITTC (2011), the following formula is recommended:

$$C_A = (5.68 - 0.6 \log Re) \cdot 10^{-3} \quad (2.5)$$

Air allowance resistance, C_{AA}

The air resistance is represented by the air allowance resistance coefficient C_{AA} that accounts for the air drag the hull and superstructure induce in windless conditions. The following formulation can be used to describe the air resistance coefficient:

$$C_{AA} = C_x \frac{\rho_{air} A_T}{\rho_{SW} S} \quad (2.6)$$

where C_x is the air resistance coefficient, ρ_{air} is the density of air and A_T is the transverse frontal area above the water line.

Steering resistance allowance, C_{AS}

The allowance coefficient of the steering resistance, C_{AS} , accounts for the resistance due to the continuous rudder motions that are necessary to keep the ship on a steady course. Unless rudder motions are available, it can be estimated to $0.004 \cdot 10^{-3}$ according to Pedersen et al. (2002) and is thus insignificant compared to the other resistance coefficient components.

Allowance resistance for appendages, R_{AAP}

R_{AAP} , represents the resistance contribution of the appendages such as e.g. rudder, open shafts, bilge keel, stabilizing fins, etc.

The resistance from appendages is by Holtrop and Mennen (1982) found by summing up the wetted area of each of the components with a weight factor $(1 + k_2)$ and finding the total resistance from appendages by using the frictional resistance coefficient for the ship C_F as in (2.7):

$$R_{APP} = 0.5\rho V^2 S_{APP}(1 + k_2)C_F \quad (2.7)$$

2.1.2 Wind resistance

In almost all conditions, the presence of hull and superstructure of the ship will result in a resistance component from the relative wind (the resistance can be negative in case of strong following winds!). The wind resistance is calculated by (2.8)

$$R_{wind} = C_x \frac{1}{2} \rho_{air} A_T V_R^2 \quad (2.8)$$

The wind resistance coefficient C_x is determined empirically, by e.g. Isherwood (1972) or by model tests, and vary with the relative wind direction, γ_R . A_T is the transverse projected area above the waterline, and V_R is the relative wind speed. The wind coefficient C_x is based on a wind speed measured at a certain level above the water, often 10m. Corrections of the wind measurement height can be applied by

the method described in Blendermann (1990). Furthermore, it is assumed that the wind speed and direction is of the free wind, i.e. undisturbed by the presence of the ship. Naturally this is impossible to acquire by on-board measurements where the anemometer(s) will never be in the free wind. Corrections for this can be made, but it is ship specific, and no general model has been developed for this, therefore the measured wind is generally used without corrections. Wind force from observations as the *Beaufort scale* which is defined by the condition of the sea are measures of the free wind, but on the other hand the resolution is smaller and can suffer from human error and different interpretations.

2.1.3 Added resistance from waves

The added resistance R_{AW} from the sea waves is difficult to determine, and no accurate empirical method is available. It has been shown by many (e.g. in Pedersen (2007a)) that the added resistance in waves scale with the significant wave height squared (H_S^2) and can be expressed as in (2.9) Faltinsen (1990) for a stochastic sea state.

$$\overline{R_{AW}} = \overline{\sigma_{AW}} \rho g (B^2/L) H_S^2 \quad (2.9)$$

where $\overline{\sigma_{AW}}$ is the average added resistance coefficient from waves, B is the beam of the ship, L is the length of the ship, g is the gravitational acceleration.

The added resistance coefficient $\overline{\sigma_{AW}}$ depends on many parameters, e.g. ship speed, wave period and direction, weight distribution of the ship and hull shape. In Bhattacharyya (1978), the added resistance R_{AW} can be calculated for many variations of the standard ship model (*Serie 60*), and tabular values from this can be used to estimate the added resistance for a specific ship by interpolation. In Moor and Murdey (1968) and Moor and Murdey (1970), a statistical model was developed of the propulsion power in waves based on a number of model tank tests of mostly naval ship hull shapes. A number of numerical methods have been developed and are reviewed in Pedersen (2006) and Jorgen Strom Tejsen (1973).

Due to complicated estimation, the added resistance in waves data from heavy weather observations, usually *Beaufort* 5 or H_S 2m, is often omitted for performance

monitoring evaluation.

2.1.4 Propulsion

The effective power, P_E , needed to drive the vessel through water is defined as:

$$P_E = U \cdot R_{tot} \quad (2.10)$$

When applying a propeller to the ship, a number of efficiency ratios are introduced in order to determine the "delivered power" P_D that is the power needed to be delivered by the main engine to the propeller of the ship at a certain speed.

The total efficiency of the ship power delivered by the main engine η_D is defined as in (2.11) and can be split into the components in (2.12),

$$\eta_D = \frac{P_E}{P_D} \quad (2.11)$$

$$\eta_D = \eta_H \cdot \eta_{RR} \cdot \eta_0 \cdot \eta_S \quad (2.12)$$

where:

η_H is the hull efficiency defined as $\frac{1-t}{1-w}$, t is the thrust deduction factor defined as $t = \frac{T-R_T}{T}$, and w is the wake fraction $w = \frac{V-V_A}{V}$. The thrust deduction factor t is an increased resistance component due to the suction from the propeller on the aft part of the hull. The wake fraction w is an expression of the reduced flow velocity in the aft body of the hull relative to the ship speed.

η_{RR} is the relative rotative efficiency or the propeller efficiency behind the ship relative to the propeller efficiency in open water.

η_0 is the efficiency of the propeller in open water

η_S is the mechanical efficiency of the propeller shaft line.

If resistance and self propulsion model tests are not available, the values can be estimated using Harvald (1983) or Holtrop and Mennen (1984).

According to Pedersen (2007b), it is reasonable to calculate the propulsion in waves by using the still water propulsive efficiencies together with the added resistance.

2.2 Data-driven regression models

In contrast to the traditional empirical model based approach described in Section 2.1, different data-driven approaches that rely exclusively on accumulated ship data and not detailed knowledge about a ship's physics, have been suggested. In the quest to find the relation between the energy consumption of a ship and its related variables, different regression models have been suggested to generate the best prediction model. The input comprises a number of different variables that are expected, according to empirical methods to influence the output variable, energy consumption (EC), which is represented by the propulsion power or fuel consumption. All the regression models used herein are based on the assumptions in (2.13) where the training error ϵ_{tr} is assumed to be Gaussian distributed.

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \epsilon_{tr} \quad (2.13)$$

Common for the regression models tested is the attempt to minimize the error ϵ_{tr} which is the difference between the model value ($f(\mathbf{x})$) and the corresponding output data \mathbf{y} for a given set of training input variables \mathbf{x} and is optimized by adjusting different control parameters or weights \mathbf{w} depending on the type of model. This optimization of the model is also called "learning", or as referred to in this thesis "training", where a set of input/output variables are used to find the best model.

In order to evaluate the model, a test set is used which only includes data that has not been used for training, resulting in a test error ϵ_t . To accommodate an efficient use of the data, the data sets are split into K training and test sets, so that the data in each test set is only used for test once. This leads to K test errors ϵ_t^k , and the average of these is referred to as the cross-validation error. Training and testing is described in depth in Section 2.4.

The following subsections will describe the four different regression models that were tested.

1. Linear
2. Custom Non-linear
3. Artificial Neural Network - ANN
4. Gaussian Process Regression - GPR

2.2.1 Linear regression

The linear model is the simplest method used. Assuming that the output $y(\mathbf{x})$ is a function of weights \mathbf{w} and the input variable vector $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$, the linear combination can be expressed as in (2.14). The optimum combination of weights are found by minimizing the sum-of-squares error in (2.16).

$$y(\mathbf{x}) = w_0 + \sum_{i=1}^d w_i x_i = w_0 + \mathbf{w}^T \mathbf{x}, \quad (2.14)$$

where \mathbf{w} is a weight vector, and d is the dimension of the input vector \mathbf{x} .

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(\mathbf{x}; \mathbf{w}) - y_{tr}^n\}^2 \quad (2.15)$$

$$= \frac{1}{2} \sum_{n=1}^N \{\mathbf{w}^T \mathbf{x}^n - y_{tr}^n\}^2 \quad (2.16)$$

where y_{tr} is the target output value

(2.16) can be transformed into matrix format, and deriving $E(\mathbf{w})$ with respect to \mathbf{w} gives the optimum weight vector \mathbf{w}^T which can be found as in (2.17)

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}_{tr} \equiv \mathbf{X}^\dagger \mathbf{y}_{tr}. \quad (2.17)$$

where $\mathbf{X}^T = (\mathbf{x}^1 \ \mathbf{x}^2 \ \dots \ \mathbf{x}^N)$ and the vector $\mathbf{y}_{tr} = (y_{tr}^1, y_{tr}^2, \dots, y_{tr}^N)^T$ is the target output.

The linear regression method is a very fast method and for some problems, especially simple problems, it shows good results.

2.2.2 Custom non-linear regression

This section describes the non-linear method that was attempted in the initial investigation and was only developed and tested for the four measured data sets from Torm Marie. The data sets are based on four individual voyages and show very little variance in draught, trim and sea water temperature. It originates from the assumption that the propulsion power can be expressed by the *ship speed* U , the *ship resistance* R_{tot} and a total efficiency η_D which includes propeller, shaft and hull efficiency. The total resistance can be divided into the following three parts as described in (2.3). Since there is no simple, reliable approximation for the added resistance in waves, this part has been left out in further analysis. To justify this, only measurements with a negligible sea state ($H_s < 2m$) have been included in the model. Usually the still water resistance is expressed as in 2.8. Applying the wind resistance R_{wind} to the still water resistance, the propulsion power can be written as in (2.18) or rewritten to (2.19)

$$P_D = \eta_t^{-1} U (R_{SW} + R_{wind}) \quad \text{or} \quad (2.18)$$

$$P_t = \eta_D^{-1} U \left(\tilde{K} U^2 + \tilde{L} V_R^2 \right) \quad (2.19)$$

where $\tilde{K} = C_{SW} 1/2 \rho_{SW} S$ and $\tilde{L} = C_x \frac{1}{2} \rho_{air} A_T$.

For this model, the ship speed U and the relative wind speed V_R are the input variables. η_D^{-1} , \tilde{K} and \tilde{L} are constants that can be estimated by empirical methods as described in Section 2.1. The variation between the empirically determined values and the training data can be modelled by adjusting these parameters with an additional variable part: $\eta_D^{-1} + \Delta\eta_D^{-1}$, $\tilde{K} + \Delta\tilde{K}$ and $\tilde{L} + \Delta\tilde{L}$ respectively. (2.19) can now be rewritten as:

$$P_D = (\tilde{\eta}_D^{-1} + \Delta\tilde{\eta}_D^{-1}) \left(\tilde{K} + \Delta\tilde{K} \right) U^3 + (\tilde{\eta}_D^{-1} + \Delta\tilde{\eta}_D^{-1}) \left(\tilde{L} + \Delta\tilde{L} \right) V_R^2 U \quad (2.20)$$

This expression can be divided into two main parts, one related to the empirical estimates model *Basic Power*, P_{bas} , and another part related to the data fitting problem *Residual power*, P_{res} .

$$P_D = \underbrace{\tilde{\eta}_D^{-1} \tilde{K} U^3 + \tilde{\eta}_D^{-1} \tilde{L} V_R^2 U}_{\text{Basic power } P_{bas}} + \underbrace{\Delta \tilde{\eta}_D^{-1} (\tilde{K} U^3 + \tilde{L} V_R^2 U) + \Delta \tilde{\eta}_D^{-1} (\Delta \tilde{K} U^3 + \Delta \tilde{L} V_R^2 U) + \tilde{\eta}_D^{-1} (\Delta \tilde{K} U^3 + \Delta \tilde{L} V_R^2) U}_{\text{Residual power } P_{Res}} \quad (2.21)$$

(2.21) can be rewritten as:

$$\begin{aligned} P_{D,n} &= P_{bas,n} + w_1 (A_n x_{1,n} + B_n x_{1,n}) + w_1 w_2 x_{1,n} + w_1 w_3 x_{2,n} \\ &+ C_n w_2 x_{1,n} + C_n w_3 x_{2,n} \end{aligned} \quad (2.22)$$

where $w_1 = \Delta \tilde{\eta}_D^{-1}$, $w_2 = \Delta \tilde{K}$, $w_3 = \Delta \tilde{L}$ are the "weights" of the model. n is the index of the measurement parameters; $x_{1,n} = U^3$ and $x_{2,n} = V_R^2 U$ are the input variables, and $A = \tilde{K}$, $B = \tilde{L}$ and $C = \tilde{\eta}_D^{-1}$ are derived parameters for the actual condition.

Detailed mathematical derivations are found in Pedersen (2008).

The model described in (2.21) is non-linear due to $\Delta \tilde{\eta}_D^{-1}$ that appears in terms containing the $\Delta \tilde{K}$ and $\Delta \tilde{L}$. The solution to (2.22) can be found by minimizing the sum of the error squared

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \left\{ \underbrace{\hat{P}(\mathbf{x}_n; \mathbf{w}) - P_n}_{\text{Error } e} \right\}^2 \quad (2.23)$$

where N is the total number of measurements, \mathbf{x}_n is a $n \times i$ matrix containing all the input parameters, and \mathbf{w} is a vector with the i weights.

The optimization (minimization) of this problem was done using the "Levenberg-Marquardt" method described by Hans Bruun Nielsen in Nielsen (2004) and Nielsen et al. (2004). This method needs the *error function* and the *Jacobi matrix* J as input in (2.24)

The model was trained using a cross-validation scheme with K training/test sets,

and the weights w_i for the final solution were taken as the average of the weights $w_i(n)$ from the K solutions.

$$\mathbf{J} = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \cdots & \frac{\partial e_1}{\partial w_i} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_n}{\partial w_1} & \cdots & \frac{\partial e_n}{\partial w_i} \end{bmatrix} = \begin{bmatrix} \frac{\partial \hat{P}_1}{\partial w_1} & \cdots & \frac{\partial \hat{P}_1}{\partial w_i} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{P}_n}{\partial w_1} & \cdots & \frac{\partial \hat{P}_n}{\partial w_i} \end{bmatrix} \quad (2.24)$$

$$\frac{\partial \hat{P}_n}{\partial w_1} = A_n x_{1,n} + B_n x_{2,n} + w_2 x_{1,n} + w_3 x_{2,n} \quad (2.25)$$

$$\frac{\partial \hat{P}_n}{\partial w_2} = C_n x_{1,n} + w_1 x_{1,n} \quad (2.26)$$

$$\frac{\partial \hat{P}_n}{\partial w_3} = C_n x_{2,n} + w_1 x_{2,n} \quad (2.27)$$

Linear approximation

The variation of the propulsion efficiency is in many cases limited and can thus be regarded as constant. This implies that $\Delta \tilde{\eta}_D^{-1}$ can be neglected, and the model will become linear in the form of (2.28). This linear model can now be solved by Least Square in the form of $\mathbf{w} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{t} \equiv \mathbf{X}^\dagger \mathbf{t}$.

The linear model was solved by the "Leave One Out" routine (described in Section 2.3.1) where the final weights have been taken as the mean of the N weights.

$$P_D = \underbrace{\tilde{\eta}_D^{-1} \tilde{K} U^3 + \tilde{\eta}_D^{-1} \tilde{L} V_R^2 U}_{\text{Basic power } P_{bas}} + \tilde{\eta}_D^{-1} \Delta \tilde{K} U^3 + \tilde{\eta}_D^{-1} \Delta \tilde{L} V_R^2 U \quad (2.28)$$

2.2.3 Artificial Neural Network, ANN

An Artificial Neural Network ANN is a non-linear mathematical system where the so-called hidden layer(s) with a number of hidden neurons or units is the non-linear link between the inputs and output(s) as described in Larsen (1999) and Bishop (2006). This is illustrated in Figure 2.2 with multiple outputs. A single hidden layer Neural Network as described below is a universal approximator, i.e. it can approximate any continuous function arbitrarily close by increasing the number of hidden units. It can

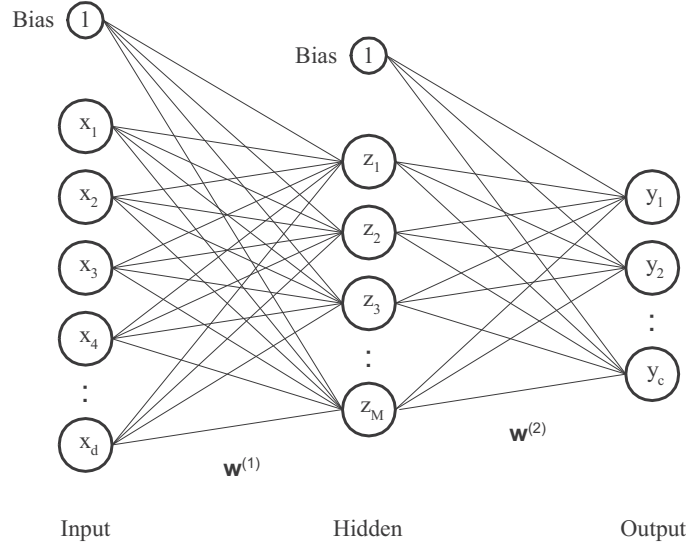


Figure 2.2: Illustration of the structure of artificial neural network

be described mathematically as in (2.29) and (2.30).

$$y(x) = \sum_{j=0}^M w_{kj}^{(2)} z_j \quad (2.29)$$

$$z_j = g \left(\sum_{i=0}^d w_{ji}^{(1)} x_i \right) \quad (2.30)$$

where x are the input variables, $y(x)$ is the output, in this case the propulsion power, z_j are the non-linear basis functions, w_{jj} and w_{kj} are the weights for the hidden units and output respectively, and g is the so-called activation function which is a non-linear function with an output of 0 to 1.

There are no specific rules or guidelines on how to find the optimal number of hidden units and layers for an ANN. However as a rule of thumb, networks for simple problems, e.g. input/output variables that have a linear relation, need less hidden units, and more complicated problems with more input variables not related linearly to the output variable need more hidden units (Hess et al. (2006)).

An ANN that is too simple or too complicated can result in a non-optimal solution,

so a number of different hidden units have to be tested.

The network used for this analysis is a flexible non-linear regression model with additive Gaussian noise and is trained with a Bayesian learning scheme. It has a tangent hyperbolic sigmoidal activation function and is trained using a BFGS (Broyden-Fletcher-Goldfarb-Shanno) optimization algorithm with a soft-line search to determine step lengths. The Hessian matrix is evaluated using the Gauss-Newton approximation.

More details into the specific neural network used here can be found in the following references: DTU-toolbox (2002), Larsen (1993), MacKay (1992), Pedersen (1997) and Svarer et al. (1993). A basic description of neural networks can be found in Bishop (2006) and Larsen (1999).

ANN are well known methods and are regarded as robust for many applications for regression and classification problems. However, they do not give any information about how the input variables influence the model, and the models can not necessarily be reproduced by a new training because they are connected by randomly chosen weights.

2.2.4 Gaussian Process, *GP*

The theory and underlying philosophy which motivates the use of Gaussian processes (*GP*) for regression analysis is an important and interesting topic, but due to the mathematical and statistical complexities, a complete derivation of the mathematical framework is out of scope for this work. For a complete and thorough description and derivation of the framework, we refer to the key founding text by Rasmussen and Williams (2006). In this work we will focus on one of the important results and applications of the use of *GP* as one of the main motivations for this thesis.

A Gaussian process (*GP*) can be described as a generalization of the Gaussian probability distribution. Where the Gaussian probability distribution describes the random variables which are scalars or vectors, the Gaussian process describes a probability distribution over functions $f(x)$ where the values of $f(x)$ evaluated for an arbitrary set of input points $(x_1 \cdots x_N)$ have a jointly Gaussian distribution. A Gaussian stochastic process like this can be defined completely by the mean (m) and the covariance (k)

which leads to the general form which can be described in (2.31) where $m(x)$ is the mean function, and $k(x, x')$ is the covariance function.

$$f(x) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (2.31)$$

Rasmussen and Williams (2006) have defined *GP* as: "a collection of random variables, any finite number of which have a joint Gaussian distribution." *GP* has also been described, by Neal (1994) as the most flexible class of non-parametric function estimators and can be interpreted as an infinite ANN.

GP for regression, GPR

Gaussian process regression is performed by making a joint distribution of the training output \mathbf{y} and the test output \mathbf{y}_t , given a set of training input \mathbf{x} and test input variables \mathbf{x}_t . The test output functions are found by a conditional joint distribution on the training data to find the predictive distribution: $\mathbf{y}_t | \mathbf{x}_t, \mathbf{x}, \mathbf{y} \sim \mathcal{N}(m(\mathbf{x}_t), cov(\mathbf{x}_t))$

Given a single test \mathbf{x}_t point, the predictive mean and variance are derived as in Rasmussen and Williams (2006) to 2.32 and 2.33. Where the mean can be described as a linear combination of n kernel functions $k(\mathbf{x}_i, \mathbf{x}_t)$ for the n training points.

$$\mathbf{y}_t = \sum_{i=1}^n \alpha_i k(x_i, x_t), \text{ where } \boldsymbol{\alpha} = (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \quad (2.32)$$

$$\sigma_t^2 = K(\mathbf{x}_t, \mathbf{x}_t) - K(\mathbf{x}_t, \mathbf{x}) \boldsymbol{\alpha} K(\mathbf{x}, \mathbf{x}_t) + \sigma_n^2 \quad (2.33)$$

where:

\mathbf{K} is the covariance function of the training input

$K(\mathbf{x}_t, \mathbf{x}_t)$ is the covariance of the test input

$K(\mathbf{x}, \mathbf{x}_t) / K(\mathbf{x}_t, \mathbf{x})$ are the covariance matrix between the training and the test input.

σ_n^2 is the noise variance

GPR finds the predictive distribution of the Gaussian distribution functions defined by the mean value \mathbf{y}_t in (2.32), and the predictive variance σ_t^2 in (2.33). σ_t^2 can thus be interpreted as the uncertainty of the prediction of \mathbf{y}_t . This is a unique feature of *GPR* that is not readily available in other regression analysis and this "extra" outcome will be used later in Section 2.5 to assist the long-term performance trend.

A natural development of the prediction variance is that it increases in areas with less information. This is illustrated in Figure 2.3a with a series of predictions where all the variables are kept constant except the significant wave height H_S . The average of the relative predicted propulsion power is plotted together with error bars corresponding to plus and minus two times the predicted standard deviation σ_t , and it is noted how the error bars increase significantly in magnitude for higher wave heights where less or no data of H_S are available as shown by the distribution of the H_S in Figure 2.3b

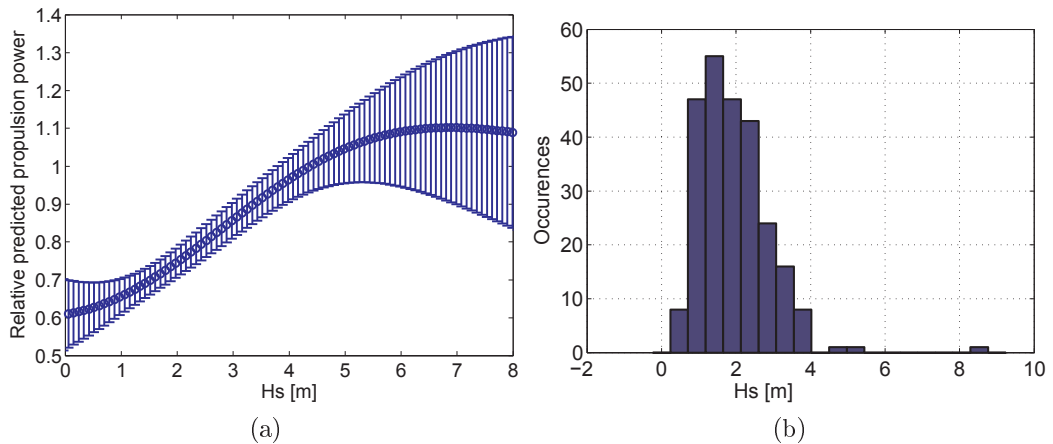


Figure 2.3: Illustration of the variation in the predicted variances due to the number of input.

In order to find the predictive mean \mathbf{y}_t (2.32) and predictive variance (2.33), the algorithm in (2.34) to (2.38) has been applied. First a Cholesky decomposition of (2.34) is performed which is used to determine α , for finding the predictive mean and to calculate the predictive variance in (2.37). For more than one test point, (2.36) to (2.38) are repeated.

$$L = \text{cholesky}(K + \sigma_n^2 I) \quad (2.34)$$

$$\boldsymbol{\alpha} = L^T \backslash (L \backslash \mathbf{y}) \quad (2.35)$$

$$f(\mathbf{x}_t) = \mathbf{k}_t^T \boldsymbol{\alpha} \quad (2.36)$$

$$\mathbf{v} = L \backslash \mathbf{k}_t \quad (2.37)$$

$$\sigma_t^2 = k(x_t, x_t) - \mathbf{v}^T \mathbf{v} \quad (2.38)$$

Determining the covariance functions K is central in GP and different models are available. In this problem the *Squared Exponential* (K_{SE}) (2.39) has been used. K_{SE} is easy to incorporate into an optimization routine and has been found to be useful for a range of regression applications of Gaussian Processes as discussed in Bishop (2006).

$$K_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp \left[-\frac{1}{2} \sum_{d=1}^D \left(\frac{x_d - x'_d}{l_d} \right)^2 \right] + \sigma_n^2 \mathbf{I} \quad (2.39)$$

where σ_f^2 is the predictive variance, σ_n^2 is the noise variance, and l is the length-scale, D is the dimension of the input variables \mathbf{x} .

K_{SE} has three free parameters or *hyperparameters* that can be varied: the characteristic *length-scales*, l , the signal variance σ_f and the noise variance σ_n .

In the present problem, the input \mathbf{x} is multi-dimensional and thus there is one *length-scale* for each of the input variables, and the *hyperparameters* are thus defined as: $\Theta = \{l^1, l^2, \dots, l^d, \sigma_f^2, \sigma_n^2\}$. The estimation of the *hyperparameters* is the crucial point in using the method for prediction and regression.

The model is trained by optimizing the covariance function with respect to the *hyperparameters* by *marginal likelihood* also referred to as Automated Relevance Determination, *ARD* routine Nabney (2002). This leaves one optimum *length-scale* for each of the input variables $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^d\}$

ARD determines the optimal values of the *hyperparameters*, and the variance of the signal or prediction σ_f^2 is thus determined along with variance of the noise level

σ_n^2 . The predicted distribution of the noisy test data is thus found by adding the noise variance σ_n^2 and the predicted variance σ_f^2 to the predictive mean $f(\mathbf{x})$. The characteristic *length-scale* is a measure of the relevance of the *feature* or input variable d and can then be used for evaluation of the input variables.

When performing *ARD*, it is important to be aware of scaling the input variables. Not scaling or incorrect scaling will result in erroneous outputs and values of the *hyperparameters*. This issue is discussed extensively in Neal (1995). The normalization for the *GPR* will be described in Section 2.4.2.

GPR is a fast method for solving the present problem, but the complexity increases with the number of inputs by $\mathcal{O}(n^3)$ for the training and $\mathcal{O}(n^2)$ for each prediction, so it is not appropriate for training of large data sets. This is in contrast to predictions with *ANN* that is independent of the length of training dataset.

One of the strengths of *GPR* is that from the Automatic Relevance Determination, *ARD*, the *length-scale* is found which determines how relevant each input variable is for the regression.

2.3 Training and testing of the regression models

Following is a description of how the training and testing of the regression methods were performed and definitions of the training and test data used for the methods described in the previous part of this chapter.

2.3.1 Training/test sets

The available data was partitioned into training and test/validation sets in order to validate the trained models.

First training was performed using a part of the data to build the regression - the training set - and afterwards the model was used to find the predicted test output \hat{y}_t based on the remaining test data input variables \mathbf{x}_t , the test set. Comparing the predicted output and the actual test output y_t for the test set generates a prediction error $\epsilon_t = \hat{y}_t - y_t$.

Prediction error(s) were calculated for each training/test set, and the mean value of the test error from all the test sets is referred to as the "Cross-validation error" $\overline{\epsilon_K}$ as in (2.41).

$$\epsilon_k = \frac{1}{N_k} \sum_{n=1}^{N_k} |y_{\hat{t}_n} - y_{t_n}| \quad (2.40)$$

$$\overline{\epsilon_K} = \frac{1}{K} \sum_{k=1}^K \epsilon_k \quad (2.41)$$

$$\sigma_k = \sqrt{\frac{1}{N_k} \sum_{n=1}^{N_k} (\epsilon_n - \epsilon_k)^2} \quad (2.42)$$

$$\overline{\sigma_K} = \frac{1}{K} \sum_{k=1}^K \sigma_k \quad (2.43)$$

where:

$y_{\hat{t}_n}$ is the prediction value

y_{t_n} is test value

N_k is the number of data points within each subset ($N_k=1$ for *LOO*, which is described in the following section)

K is the number of subset

ϵ_k is the mean prediction error of the subsets k

$\overline{\epsilon_K}$ is the cross-validate prediction error

In the present study, the relative prediction error of the energy consumption, ω and the cross-validation error of the relative prediction error $\overline{\omega_K}$ is used for evaluation.

In order to utilize the available data most efficiently, it was divided into more training/test sets. The most efficient use is by training with all the data except one,

$N - 1$, and testing with the single set of input variables and the output variable not used for training. This can be done alternately N times so all data points are used. This is referred to as "Leave-One-Out" (*LOO*). Since Leave-One-Out requires N trainings, it is often too computationally slow for practical use, and other subdivisions of training/test sets are used.

The *LOO* was not feasible for training of the Artificial Neural Networks so the data was split into 5 training/test sets based on the recommendation of Rasmussen and Williams (2006). First the complete data set was randomized, and then 80% of the data was used for training and 20% for testing. This ratio has been shuffled alternately in order to create 5 training/test sets per complete data set. The subdivision is illustrated in Table 2.1.

Training/test set	Data set				
1	20%	80%			
2	20%	20%	60%		
3	40%		20%	40%	
4	60%			20%	20%
5	80%			20%	

Table 2.1: Cross-validation scheme of 5 training/test data set, the shaded cells indicates test-part of the data and the clear cells the training part

For training the *GPR*, *LOO* training/test was used, and analogous to the cross validation error $\overline{\epsilon_K}$, cross-validation values were found for the *hyperparameters*.

When direct comparison of the *GPR* and *ANN* methods were needed, training was also performed with the 20/80 test/training ratio for the *GPR* with the same permutations for both methods.

2.4 Training

2.4.1 Training ANN

The structure of the ANN used consisted of a single hidden layer with N_H hidden units. The optimal number of hidden units had to be found by testing, and initial

testing showed that the results did not improve with more than 20 hidden units. In order to determine the best setup of the *ANN* model, the network was trained with 5,10,15 and 20 hidden units to facilitate both complex models (where more hidden units would be best) and more simple problems with few hidden units being optimal.

To avoid the training getting stuck into local minima, multi start (*MS*) was introduced allowing the training to begin with a new random set of initial values 10 times for each analysis, and the network setup with the lowest training error was then selected.

2.4.2 Training GPR

Large variations in the standard deviations between the different input variables could lead to large differences in the *length-scales* and consequently affect the model and predictions. In order to avoid this, the input variables \mathbf{x}^d were linearly rescaled by the mean and standard deviation: $\tilde{x}_i^d = \frac{(x_i^d - \bar{x}^d)}{std(x^d)}$. In the definition of GPR, the output variables are assumed to have a *zero-mean*, thus they were centred around 0 with the following normalization: $\tilde{y}_i = y_i - \bar{y}$.

The initial guess of the *hyperparameters* tended to result in local minima depending on the data length. This was discovered by the prediction variance becoming very small or zero. To overcome this, the training/test set was also split up into parts accumulating to the full dataset (2.44). Each training was restarted with the final *hyperparameters* from the previous run in total *MS*-times as described in (2.44). After a few tests, it was concluded that it was sufficient to restart the training twice depending on the data size, i.e. three trainings of the GPR.

$$\Theta(\mathbf{x}_{ms_{i+1}}) = \Theta(\mathbf{x}_{ms_i}) \quad (2.44)$$

$$\mathbf{x}_{ms_i} \in \mathbf{x} \left(0 : i \frac{|\mathbf{x}|}{MS} \right) \quad (2.45)$$

where:

\mathbf{x} the total input data set

\mathbf{x}_{ms_i} the input data multi start subset ms_i

2.5 Data-driven performance evaluation method

As mentioned in the introduction, the propulsion efficiency is expected to decrease over time, mainly due to fouling of the hull and propeller. This deterioration of performance can be measured by comparing the actual measured energy consumption EC in a certain state with one from a reference state.

The relative difference between the predicted energy consumption (\hat{EC}) and the actual measured energy consumption (EC) is a measure of how well the vessel performs compared to how it should perform. Due to fouling, EC is expected to increase, while \hat{EC} will maintain the same level since it is based on a stationary input for all predictions. This change leads to a decrease in the relative prediction error ω (defined in (1.1)) over time as illustrated in Figure 1.1.

The trend of this development is assumed to be linear between interruptions that affect the performance. Linear trends have been used as “best practice” in the industry although it is acknowledged by Munk (2006) that the slope of the hull resistance penalties due to fouling may be different in the first part of a period after a dry-docking, but will asymptotically reach a certain level with a more restricted change in the hull fouling penalty. It is stated that a linear slope is a good approximation for determining the future development of hull fouling. Munk (2006) uses the ratio F between “the added frictional resistance coefficient” and “the total resistance coefficient”, ($F = \Delta C_F / C_{tot}$), as a measure of the fouling penalty. A linear trend of the performance development is also assumed for the present work.

There may be other variations, but it is very difficult to make a better general assumption since the hull and propeller fouling can differ from ship to ship and is also largely dependent on the operational area and profile.

After the tests and comparison of the different regression models, *GPR* has been used exclusively for post-processing, and the linear trend as a function of time is thus established from the relative prediction errors of the energy consumption ω . According to Faraway (2002), *weighted least square regression (WLS)* can be useful if the errors

are uncorrelated and have unequal prediction variance which is the case for the prediction results. Herein the reciprocal σ_R^2 variance as weight is recommended, so input with low variability should get a higher weight than input with high variability.

Weighted least square regression is similar to normal least square methods, but the residual $(\omega - f(x, \alpha))$ is multiplied by the weights, w_{ii} in the sum of square errors (2.46) which is minimized with respect to α .

$$E(\alpha) = \sum_{i=1}^n w_i (\omega_i - f(t_i, \alpha))^2 \quad (2.46)$$

where:

t is the time

w_i is input weights: $\frac{1}{\sigma_{R_i}^2}$, which are determined from the *GPR* predictions

α is the slope of the function

ω the relative prediction error

To solve this problem, the MatLab[®] function `lscov` was used.

The linear trend of the prediction errors thus follows the form described in (1.2).

2.6 Summary of propulsion prediction and performance methods

The basic principles and methods of ship resistance and propulsion were presented to give an overview of what input variables are important for creating a good propulsion prediction. The empirical methods are expected to be less reliable for off-design conditions. Different regression methods ranging from simple linear models to complex non-linear models such as *ANN* and *GPR* were described in this section. The decision to use the long-term development of the propulsion prediction error has been used to create a performance trend definition by means of a *weighted least square* model to form a linear trend based on the results from the *GPR* results.

Chapter 3

Data collection and evaluation

This section describes the data needed for propulsion performance evaluation of ships with the purpose of identifying the important input variables and how they affect propulsion power. As the aim of this thesis was to explore the feasibility of partly or entirely data-driven propulsion performance monitoring, sufficient and high-quality data from different independent sources over a long period of time was paramount.

Traditionally, performance evaluation is based on noon report (*NR*) input data and/or speed trials which result in one or two data points per day. Furthermore, *NR* often consists of a mix between instantaneous values such as the weather at the report time and the average speed and fuel consumption over the previous 24 hours. Increasing the number of data points per day would give a higher resolution and more points for the regression models. To accommodate this, automatically sampled data was used. The shipowner Torm allowed us to equip the product tanker Torm Marie with a data logging system and also delivered noon reports from the ship for a two-year period beginning with the launch in 2006.

Additionally, in the autumn of 2009, noon report data from five container ships over a ten-year period was acquired from an anonymous shipowner.

The data consisted of three individual sources:

MDTM Measured performance data on board Torm Marie in February and March 2008.

NRTM Noon report data from Torm Marie

NRCS Noon report data from five container ships

3.1 Desired data

In order to predict the propulsive energy consumption of a ship, all the separate components that have an influence must be represented by a variable. Figure 2.1 is an illustration of how the draught and trim, speed and environmental properties affect the resistance and consequently the propulsion power needed.

Based on the basic ship resistance and propulsion theory described in Chapter 2, certain variables were expected to have a higher influence than others. A measure of energy consumption is necessary as it is the output that has to be predicted. Ship speed through the water and the draught/trim are essential input variables that have direct physical effects. According to ship resistance theory, the power increases with the speed cubed, and the draught influences the wetted surface and residual resistances. Wind speed and direction can also have a significant effect on the resistance depending on the ship type, loading conditions wind speed and wind direction. Air and seawater temperature are relatively easy to measure, and the effects are easy to define as described in the methods section. Wave height, period and direction are significant factors for resistance, but they are more difficult to measure and are therefore usually based on estimates. Furthermore it is difficult to quantify the effect of sea resistance. The following list shows the main input variables ordered by expected importance. The variables will be discussed further in the subsequent section.

1. Draught midship [m] and Trim Ta-Tf (Draught aft - draught fore) [m]
2. Ship speed [knots]
3. Relative wind speed [m/s]
4. Relative wind direction [°]
5. Wave height [m]
6. Relative wave direction [°]
7. Water temperature [°C]
8. Air temperature [°C]

3.2 General data sources and influence

This section gives an overview of the origin of the input/output variables and how they affect the propulsion power and resistance.

3.2.1 Draught and trim

The draught fore and aft are usually only observed before departure, and the accuracy of the readings are approximately $\pm 10\text{-}20\text{cm}$. In comparison, the design draught on Torm Marie is 12.0 m, i.e. the accuracy will be $\pm 3\%$. This means that the accuracy of the trim will be twice the inaccuracy of draughts which is $\pm 20\text{-}40\text{cm}$. The accuracy can be improved by reading the draught amidships although this is usually difficult at berth.

The on-board loading computer also writes out the draught which is based on information about the cargo load and fuel and ballast tank levels, but as the draught, is a derived source in this case it can be problematic. Some ships are equipped with pressure gauges fore and aft to measure draught by the hydrostatic pressure. This method requires very sensitive pressure measurements, and yet the same gauges should be able to cope with high-pressure measurements from rapid wave elevations in rough weather.

This combination usually makes the draught determination by pressure measurements unreliable.

The static trim and draught midship are among the main parameters for basic ship resistance in calm water and thus the propulsion power. A change in the draught/trim will change the wetted surface and consequently the viscous resistance, as is well documented in general ship resistance theory. Furthermore the resistance due to wave making (C_R) will change, but this effect is more complicated to predict. The draught and trim will change dynamically due to the ship speed, but this is not measured or reported and cannot be used as input.

Due to fuel consumption during a voyage, the draught and possibly the trim will change, subsequently resulting in a power consumption change. This variation in the ship's draught is limited and will affect the still water resistance as a change in frictional resistance due to the new wetted surface.

The wind resistance will also change slightly due to the change of the projected area above the waterline. A simple estimate has been made in order to evaluate this effect for Torm Marie. Given that her fuel consumption is approximately 50 tons/day, the change in draught and wetted surface is found from interpolating the hydrostatic tables. The residual resistance coefficient C_R and total efficiency η_t is assumed to be constant due to the limited change in draught. Table 3.1 shows an example of a 20-day voyage travelling at a speed of 15 knots which resulted in a reduced power of $29kW$ or about 0.29% of the initial power. The change in power due to fuel consumption is minuscule in this case and can be neglected.

For fast ships, the fuel consumption is higher,, and the hull shape might also be more vulnerable to changes in the displacement and therefore the draught is reported more carefully during voyages in the containership datasets #1-5.

3.2.2 Ship speed

The speed through the water (U) is usually measured by a speed log, based on the Doppler principle, mounted in the bottom of the forward part of the ship. This is a relatively accurate device, but it has to be checked regularly because experience shows

Service speed	15	knots
Initial draught	12	m
Still water shaft power	10029	kW
Fuel consumption per day	50	ton/day
Tons to change draught	1.1210^{-4}	m/tons
Draught change per day	-5.61^{-3}	m/day
Change in wetted surface per day	2.95	m^2/day
Voyage duration	20	days
Total efficiency	0.693	-
Change in power due to friction resistance per day	1.5	kW/day
Change in power due to friction resistance per voyage	29	kW/day

Table 3.1: Changes in the propulsion power due to the fuel consumption

that data collected tend to drift over time. The ship speed over the ground (*SOG*) is measured by the GPS navigator, and together with sea current information, this could lead to another measure of the speed through the water, but sea current information is not readily available and is thus not an option.

Ship speed has an essential effect on propulsion power because resistance defined by basic propulsion theory is proportional to the speed squared, and consequently the power is proportional to the speed cubed.

Speed also has a minor effect on the dynamic draught and trim, but this is difficult to determine empirically or measure.

3.2.3 Wind speed and direction

The relative wind speed (3) and direction (4) are present in almost any condition, since the ship speed will create a relative wind from its own speed. The most common way to report the wind speed is still by the Beaufort Wind Force Scale (BF) presented by e.g. Bhattacharyya (1978) where the wind "Force" is estimated based on visual observations of the seawater surface conditions. The relation between the Beaufort Wind force and the wind speed in m/s 10m above the water surface can be described as: $u_{10} = 0.836(BF)^{2/3}$ according to Blendermann (1990). The wind direction is also estimated by observations and/or a wind vane. Most ships are equipped with cup anemometers that can be used to assess the relative wind speed.

Although the relative wind speed and direction can be measured on board with reasonable accuracy, the relative wind will always be affected by the presence of the ship.

Wind speed and direction can also be found from hindcast weather information and are thus not available in all areas as will be described in Section 3.2.8.

The wind resistance of the ship has been defined in Chapter 2, but aside from giving rise to a resistance directly related to the longitudinal component of the wind, any wind direction off the longitudinal will give rise to a turning moment. This will indirectly result in a response from the rudder and hence steering resistance. This was not directly accounted for due to lack of rudder angle information.

3.2.4 Sea waves

It is difficult to measure wave height, period and direction. Traditionally visual observations are used, but different methods for measuring or estimating the wave properties exist. It is possible to measure the relative wave height based on microwave reflections as described in Nielsen et al. (2006). These are measurements close to the hull and will consequently suffer from hull reflected wave interference. A wave radar gives a more detailed picture of the wave pattern in the area around the ship, but is, as described by Nielsen (2005), a complex and costly system that needs careful calibration and maintenance. Based on ship motions, estimates of the wave height were also performed in Nielsen (2005); this requires a transfer function which is based on the hull shape and the weight distribution and measurements of the motions. We considered it but were unable to add accelerometer measurements to the measured data from Torm Marie. It would have added information about ship motion which would have been a useful variable to represent the effect from the waves.

Wave rider buoys determine the wave height and period on a specific location and are often used for research purposes and in field development for offshore structures. There are various types of wave buoys designed for different purposes, and they generally give good estimates of wave properties. Wave buoys are often used for adjusting meteorological hindcast models as will be described in Section 3.2.8

Sea waves are often described as a combination of swells and wind-driven waves. Swells are so-called "old waves" created in the past or waves that have dissipated from another area with e.g. a storm and can thus be present in calm water. Swells are generally longer waves that have a limited effect on the added resistance (Pedersen (2006)) and contain less energy than the wind-generated waves.

Wind-generated waves are thus the most important factor for the added resistance in waves. Consequently, wind speed and direction are also reasonable variables to represent wave properties.

Ideally the height and period of the swell and wind-generated waves should be reported separately. Although this is a difficult task, it would make it possible to estimate the added resistance from the swell and wind waves individually.

A ship's added resistance from waves cannot be accurately estimated empirically and is difficult to determine numerically. Moreover, the added resistance in waves depends on the weight distribution of the specific loading conditions of the ship which in most cases have to be estimated.

The added resistance from the waves is assumed to be proportional with the wave height squared which is documented by e.g. Gerritsma and Beukelman (1972). The direction has significant influence on wave resistance, and the maximum added resistance is, according to Salvesen (1978), usually found in the region of $\pm 30^\circ$ around the head sea (0°) condition.

Research and experience show that the added resistance from waves less than significant wave heights (H_S) of 2m (Beaufort 5) has an insignificant effect on ship resistance for large cargo ships like the ones in Pedersen (2006). As sea waves less than H_S 2m in e.g. the North Atlantic and North Pacific occur more than 50% of the time (Faltinsen (1990)) and in the Indian Ocean about 60% of the time (Hogben et al. (1986)), the probability of "good" weather is dominant on most common shipping routes.

3.2.5 Water and air temperature

Changes in seawater and air temperatures affect the densities and viscosities and thus the resistance.

It is acknowledged that the ship resistance can be expressed as in (2.3) so that the density of water has a linear effect on the resistance. Furthermore the changes in the viscosity affects the Reynolds number (R_e) and subsequently the frictional resistance coefficient (C_f) as described in Section 2.1.1. The density decreases with increasing seawater temperature and the kinematic viscosity increases, which leads to a higher R_e and a lower C_f .

To illustrate the effect of the changes in sea water temperature an example is carried out for realistic extreme temperatures for Torm Marie. The density of sea water (salinity of 3.5‰) at 0 and 30 °C is 1028.1 and 1021.7 kg/m³ respectively, i.e. a difference of approximately 0.6%. Changes in the viscosity has a significantly higher impact due to the increase from $1.83 \cdot 10^6 \text{m}^2/\text{s}$ to $8.50 \cdot 10^7 \text{m}^2/\text{s}$ for 0 and 30 °C respectively, resulting in C_f of $1.530 \cdot 10^{-3}$ and $1.394 \cdot 10^{-3}$ or a reduction of 8.9%.

The frictional resistance accounts for approximately 65% (for the present ship at 15 knots) of the total resistance and altogether the reduction in propulsion power can be estimated to 7.5% or approximately 0.05% per degree.

The air temperature also has a significant effect on the air density and is affecting the resistance linearly. Furthermore, the range of air temperatures is larger than the range of water temperatures, maybe -30 °C ($1.45 \text{kg}/\text{m}^3$) - +50°C ($1.09 \text{kg}/\text{m}^3$), which results in a difference in the density and thus air resistance of 33%. On Torm Marie, the air resistance was approximately 215kN out of 1168 kN in the total resistance at 10 m/s and 0 °C headwind, and 199kN in the same conditions with 10°C resulting in a total resistance of 1152 kN . The difference in air density thus has a relative influence on total resistance of approximately 1.4% per 10°C or 0.14% per °C for this vessel in the current conditions.

The sea and air temperatures both have well documented influences on the still water resistance. The sea water temperature has a smaller effect per change in°C and is less likely to fluctuate, although entering a sheltered area or a river could lead rapid changes. Furthermore the frictional resistance is a dominant and consistent factor of the total resistance so even small changes will accumulate over time, and is expected to have a noticeable effect on the performance evaluation.

An ocean going vessel is often travelling in a relative head wind condition due to

its own speed, so changes in air temperature is certainly important, but since both the wind speed and temperature can change significantly during one day it is thus difficult to get data of this with a high resolution.

Air and sea water temperatures are easy access data and measurements can be performed by relative simple and reliable devices making automatic sampling relatively easy to setup.

3.2.6 *time* as an input variable

Fouling of the hull and propeller can be due to various marine growths that are expected to increase the resistance and decrease propulsion efficiency over time. The effect of fouling is expected to vary continuously over time, and often the trend is assumed to be linear. The *time* as a linear continuous variable, e.g. 1 per day, can thus be used as a variable to represent the effect of fouling.

3.2.7 Selection of the most significant input variables

After listing all the relevant parameters, we will assess the importance of variables. Ocean-going cargo ships travelling on the most common trade routes are sailing more than 50% of the time in what can be regarded as good weather, i.e. Beaufort 5 or below. The significant wave height in this wind force is $H_s=1.25-2.5$ m, where the added resistance for a ship with a length of ~ 200 m, corresponds to less than 2% of the total resistance. By neglecting all measurements and observations in stronger wind forces, it is possible to neglect the wave parameters.

Since changes in air temperature have a minor effect on total resistance, the temperature can be neglected in datasets with little temperature variance.

The variation in sea water temperature have some influence, but many ships travel in an area where the sea water temperature is relatively stable. Therefore it is possible to neglect it if the seawater temperature has only minor variations.

For conditions in good weather and with minor changes in temperature, the most important variables are the draught and ship speed.

3.2.8 Hindcasts

Hindcast (*HC*) information has been used as a supplement to the reported observed and measured data. Hindcasts are weather information at a certain time and position in the past. The was data collected by a tool developed for Seatrend^{®a} at FORCE Technology based on weather information from the NOAA^b database. This database is based on observations by wave buoys, anemometers and meteorological data. For a given position, the tool returns wind speed and direction, significant wave height, peak period and direction. Some areas are not included in the database, e.g. the Mediterranean, the North Sea and areas close to the coast, thus using hindcast data usually reduces the original dataset.

The hindcast data obtained for the present work were:

Wind speed [m/s]

Wind direction [°]

Significant wave height [m]

Wave peak period [s]

Wave direction [°]

Since hindcasts and observed values are both instantaneous, it is easy to compare the observed and measured data at any given time and available position. A comparison of the hindcast data and the measured tanker data is presented in Figure 3.6. It shows that the wind speed from the ship anemometer has a fair agreement with the hindcast wind speed, but wind direction tends to show larger discrepancies.

3.3 Data from Torm Marie

Torm Marie is a 110,000 dwt product tanker built in 2006 as part of a series of five sister ships. The vessel was mainly operating (in the test period) on voyages longer

^aPerformance Monitoring tool developed at FORCE Technology, www.force.dk

^bNational Oceanic and Atmospheric Administration, United States Department of Commerce <http://www.noaa.gov/>

than a week with a service speed of approximately 14.5-15.0 knots in calm weather conditions. Torm Marie was originally equipped with a torsionmeter and a vessel performance system delivered by KYMA that measured and recorded shaft torque, thrust and power together with other parameters making it an obvious candidate for data collection.

As many tank ships, the most common operational profiles were ballast condition or fully loaded which are further described later in this section.

The vessel was fitted with pneumatic draught gauges fore and aft, but these were not connected to the system and were regarded as unreliable due to the considerable variations in the measured draughts compared with the manual readings from the noon reports which is illustrated in Figure 3.1. The observed draughts are thus the best available data for this input.

Torm Marie's main particulars are listed in Table 3.2

Length overall	244.60	m
Length between perpendiculars	233.0	m
Breadth	42.0	m
Depth	22.2	m
Design draught	12.0	m
Scantling draught	15.5	m
Deadweight	109.637	ton
Cargo capacity	121.655	m ³
Propeller diameter	7.05	m
Number of propeller blades	5	-
Main engine power (85%MCR)	15260	kW@105RPM

Table 3.2: Main particulars of Torm Marie

3.3.1 On-board measured data

Torm Marie was permanently equipped with a torsionmeter from KYMA^c to measure the shaft torque and thrust. KYMA integrated measurements of the specific fuel

^cKYMA a.s. - A Norwegian manufacturer of shaft power measurements
<http://www.kyma.no/products/>

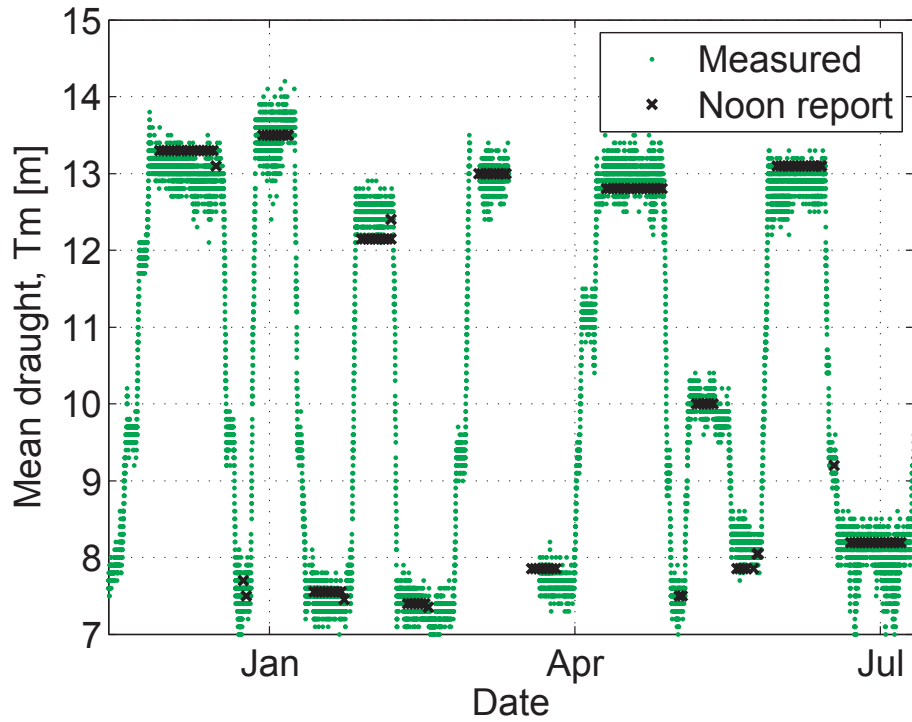


Figure 3.1: Measured mean draught from KYMA compared with the noon report draught

consumption, the speed through the water (logged speed), shaft rate of revolutions together with torque and thrust into a monitoring system.

The logged speed was measured by a Doppler speed log mounted in the bottom of the ship in the bow, and the specific fuel consumption was measured by flow meters on the main engine. This system had a readout-display on the bridge where all the instantaneous values were available together with derivatives such as shaft power. The data were collected on a logging PC in the engine control room, but due to unfortunate data logging methods, the logged speed was sampled with steps of 1.2 knots which was too sparse for this analysis. The data sampling was implemented using a customized logging system (*GPSwin*) based on NMEA 0183^d standard data strings.

Specific data logging equipment was installed that combined the *KYMA* readout

^dNMEA 0183 is a combined electrical and data specification for communication between marine electronic devices controlled by, the U.S.-based National Marine Electronics Association

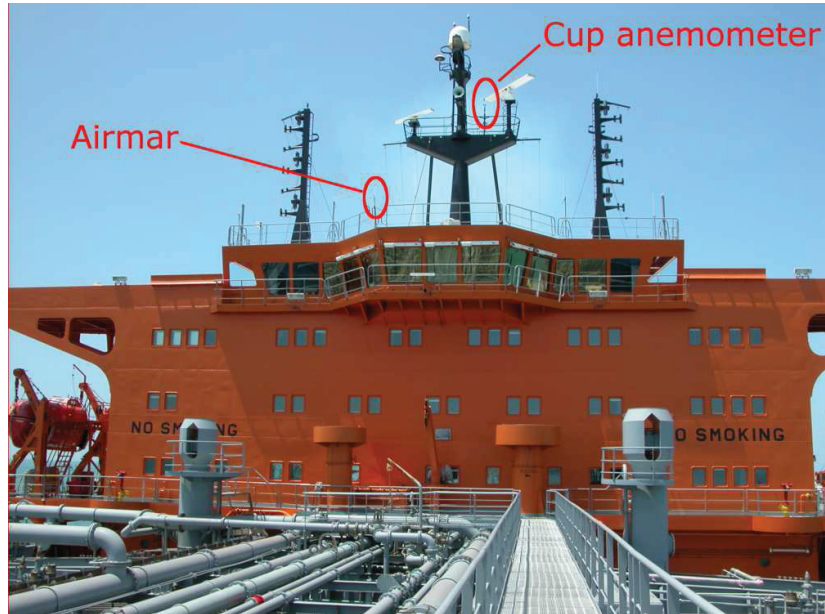


Figure 3.2: Airmar mounted on Monkey Island

with an Airmar® weather station^e. The unit was a combined wind anemometer, compass, GPS receiver, thermometer, barometer, accelerometer and humidity sensor. The main argument for choosing the Airmar® unit was that it was a combined anemometer and GPS tracker, furthermore it was relatively easy to install and had no moving parts. It was mounted on the railing of "Monkey Island" as shown in Figure 3.2.

Specific installation instructions for the *KYMA* system were delivered by *KYMA a.s.* and made it easy to access the data. The customized logging software developed by FORCE Technology *GPSwin* combined the two datasets into one log file. Figure 3.4 shows an example of the logging window during a test run.

An overview of the logged signals is given in Figure 3.3 and Table 3.3. The sampling frequency in *GPSwin* of the time series was one second, but many of the measurements had inconsistent and missing signal values and were not necessarily updated every second. All signals from *KYMA* were sent one string at a time every second, given 13 strings were sent; each signal was updated every 13 seconds. *GPSwin* sampling was split into time series of ten minutes with ten-minute intervals.

^eWeatherStationTM PB100

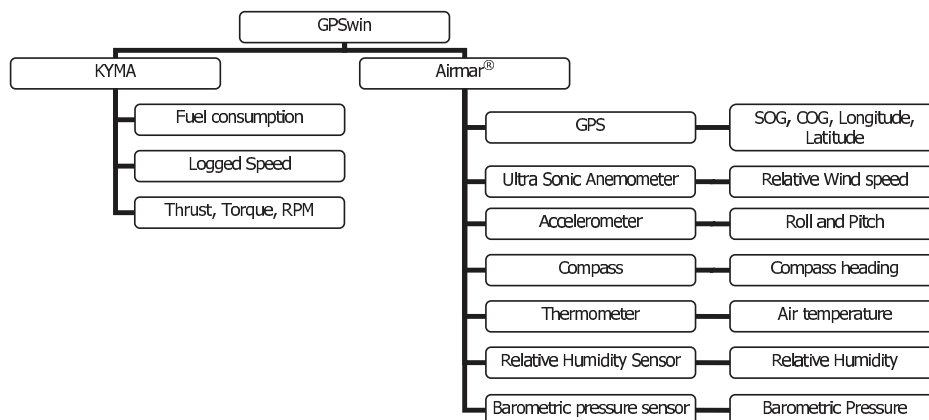


Figure 3.3: GPSwin sensor diagram



Figure 3.4: Screen shot of the GPSwin logging window

After the installation, the system was observed for one voyage, and parallel manual observations were conducted. The system seemed to work flawlessly at departure, but after two months it broke down and was unable to be recovered by the crew on board. It resulted in continuous reliable time series results in the period January 30th - March 27th 2008, which corresponds to four voyages: two in loaded conditions and two in ballast conditions.

3.3.2 Observations during voyage

Manual observations were made on board Torm Marie in the period January 23rd 2008-February 5th 2008 in order to assess the quality of the measured data and the

Description	Unit	Method	Device	Sampling frequency frequency
Ship Speed over ground	knots	GPS	Airmar	1 Hz
Ship Course over ground	deg	GPS	Airmar	1 Hz
Longitude	deg.HHMM	GPS	Airmar	1 Hz
Latitude	deg.HHMM	GPS	Airmar	1 Hz
Relative wind direction	deg	Ultrasonic	Airmar	1 Hz
Relative wind speed	m/s	Ultrasonic	Airmar	1 Hz
Air temperature	°C	Thermometer	Airmar	1 Hz
Pitch	deg	Accelerometer	Airmar	1 Hz
Roll	deg	Accelerometer	Airmar	1 Hz
Heading	deg	Compass	Airmar	1 Hz
Shaft revolutions	Tacho	rpm	KYMA	1/13 Hz
Shaft Torque	Strain Gauge	kNm	KYMA	1/13 Hz
Shaft Power	derived	kW	KYMA	1/13 Hz
Propeller thrust	Strain gauge	kN	KYMA	1/13 Hz
M/E fuel consumption	Flow meter	kg/hr	KYMA	1/13 Hz
Ship speed by Log	Doppler log	knots	KYMA	1/13 Hz
Fuel oil specific gravity at 15.6 Å°C	-	-	KYMA	1/13 Hz
Fuel oil lower calorific value	kJ/kg	KYMA	derived	1/13 Hz
Ship overall efficiency, actual	kg/mile	KYMA	derived	1/13 Hz
Ship overall efficiency, corrected	kg/mile	KYMA	derived	1/13 Hz
M/E specific fuel rate, actual	g/kWhr	KYMA	derived	1/13 Hz
M/E specific fuel rate, corrected	g/kWhr	KYMA	derived	1/13 Hz
Propulsion efficiency	m/kWhr	KYMA	derived	1/13 Hz

Table 3.3: List of measurements *p/nm: pulses per Nautical miles **p/l: pulses per liter

noon reports made by the crew. Naturally some of the measured and observed parameters are more reliable than others. GPS derived parameters, Lat, Long, Speed Over Ground, Course Over Ground, are assumed to be both reliable and precise. Measurements taken in a controlled environment are also reliable, or the error can be estimated. This includes: Shaft torque/thrust, shaft revolutions and fuel flows.

The most difficult measurements to take are of the surrounding environment: wind speed, wind direction, speed through the water and wave height/sea state and direction.

The anemometer was able to measure the horizontal wind speed and direction accurately where the anemometer was mounted, however, how well the wind around the

anemometer represented the general relative wind was uncertain. By manual observation with wool threads, the wind direction was checked, and a significant vertical component was observed as illustrated in Figure 3.5. This indicates, as expected, an increase in the wind speed over the superstructure. The wind direction is not expected to be affected to the same degree, but this is more difficult to assess.

A comparison of the wind speeds during the test period from three different sources is shown in Figure 3.6. The Airmar® wind speed was explained in Section 3.3.1, the "Ship anemometer" is a cup anemometer mounted on the antenna mast illustrated in Figure 3.2, and the "Wind speed from hindcast" originates from hindcast data extracted from the NOAA^f database. The Airmar® wind speed seemed to measure much higher values than the ship anemometer and hindcasts which are rather close to each other and show the same trends. This might be due to the relatively low position of the Airmar® sensor on the deck and exposure to higher wind speeds. The wind direction was analysed in the same manner, also presented in Figure 3.6, and gives a more scattered picture for some periods. This might be caused by a local weather system not captured by the hindcast analysis or due to unfortunate relative wind directions that twist the wind at the positions of the measurements. During the manual observation on board, several simple manual wind direction assessments with a wool thread on a stick were conducted on different locations on the vessel. The general experience was that large variation was found between observations on e.g. starboard side and port side of the bridge wing and near the bow. This confirms the difficulty of making on-board wind measurements.

The speed through the water was measured by a Doppler log mounted in the forward part of the ship and measured the relative water speed ten meters below the keel with an accuracy of 1.0% or 0.1 knots whichever is greatest.

For further analysis, the measured dataset was divided into four datasets each representing one voyage/loading condition. Table 3.4 gives an overview of the period, loading condition, speed range, power range, how many data points and how many

^fNational Oceanic and Atmospheric Administration, United States Department of Commerce <http://www.noaa.gov>

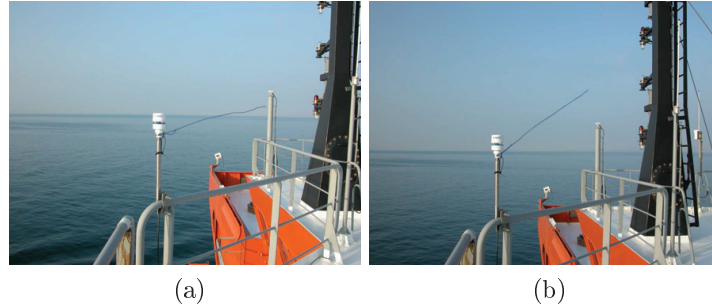


Figure 3.5: AIRMAR vertical wind profiles

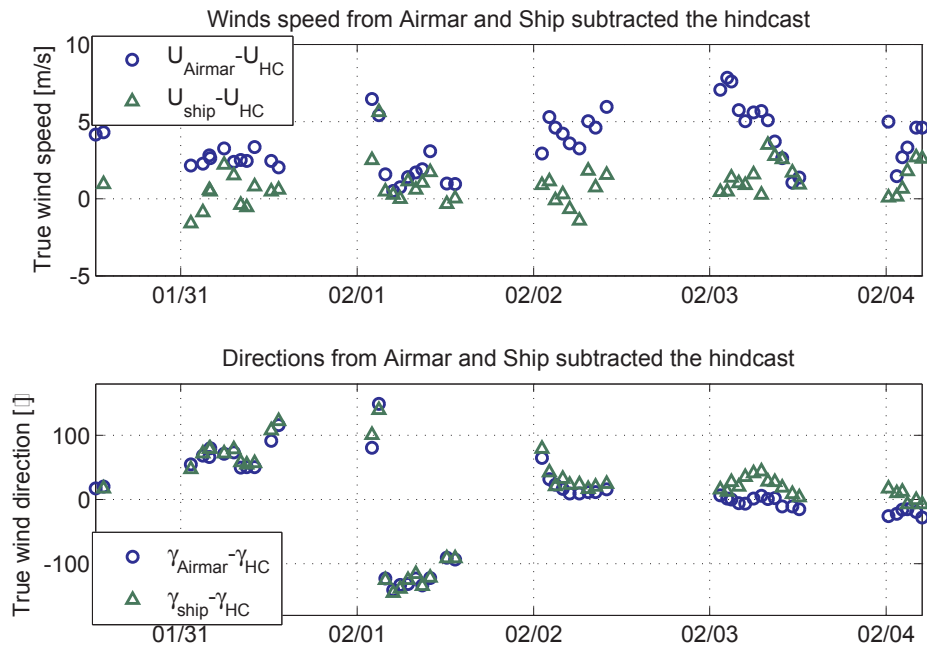


Figure 3.6: Comparison of wind speed and direction during the test period. Airmar and ship anemometer measurements are subtracted from the hindcast values

noon reports are available within the period, of each dataset, and in Figure 3.7 histograms of the ship speed and true wind are shown in order to give an overview of the speed variations and the weather conditions for each of the measured datasets.

To justify using the mean and standard deviation of each time series as input, the governing variables were evaluated, i.e. ship speed U , propulsion power P and the

Data set	$N_{samples}$	Start date- End date	No. of valid NR	T_m [m]	Trim $T_A - T_F$ [m]	U_{min} - U_{max} [knots]	P_{min} - P_{max} [kW]
M		-	-				
1	236	09.02-14.02-08	3	7.4	2.4	14.2-16.2	7573-11283
2	109	22.03-27.03-08	4	7.85	2.7	13.6-15.1	7750-9248
3	301	30.01-06.02-08	7	12.15	0	13.4-16.0	8138-11216
4	555	01.03-11.03-08	9	13.0	0	13.0-15.9	9741-12096

Table 3.4: Measured dataset division

relative wind speed V_R .

The relative average value ($\bar{\mu}_{x,M}$) and the relative standard deviation ($\bar{\sigma}_{x,M}$) which are defined Eq. (3.1) and (3.2), of the value were used to evaluate how much variation is within one dataset (voyage).

As seen in Table 3.5, $\bar{\mu}_{x,M}$ for U and P are fairly small compared to V_R which varies considerably within each time series.

The relative standard deviation was used for filtering the data, and furthermore $\bar{\mu}_P$ gives an estimate of the accuracy of the output variable.

M	N	$\bar{\mu}_U$	$\bar{\mu}_P$	$\bar{\mu}_{V_R}$
-	-	%	%	%
1	236	0.64	0.69	18.0
2	109	0.55	0.49	9.1
3	301	0.56	0.88	9.5
4	555	0.61	0.57	11.4

Table 3.5: Average of the relative standard deviation

$$\bar{\mu}_{x,M} = \frac{1}{N} \sum_{n=1}^N \left| \frac{\sigma_{x,n}}{\mu_{x,n}} \right| \quad (3.1)$$

$$\bar{\sigma}_{x,M} = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{\sigma_{x,n}}{\mu_{x,n}} \right)^2} \quad (3.2)$$

where:

$\sigma_{x,M}$ is the standard deviation of the n' th time series,

$\mu_{x,M}$ is the mean value of the n 'th time series and

x indicates the input/output variable

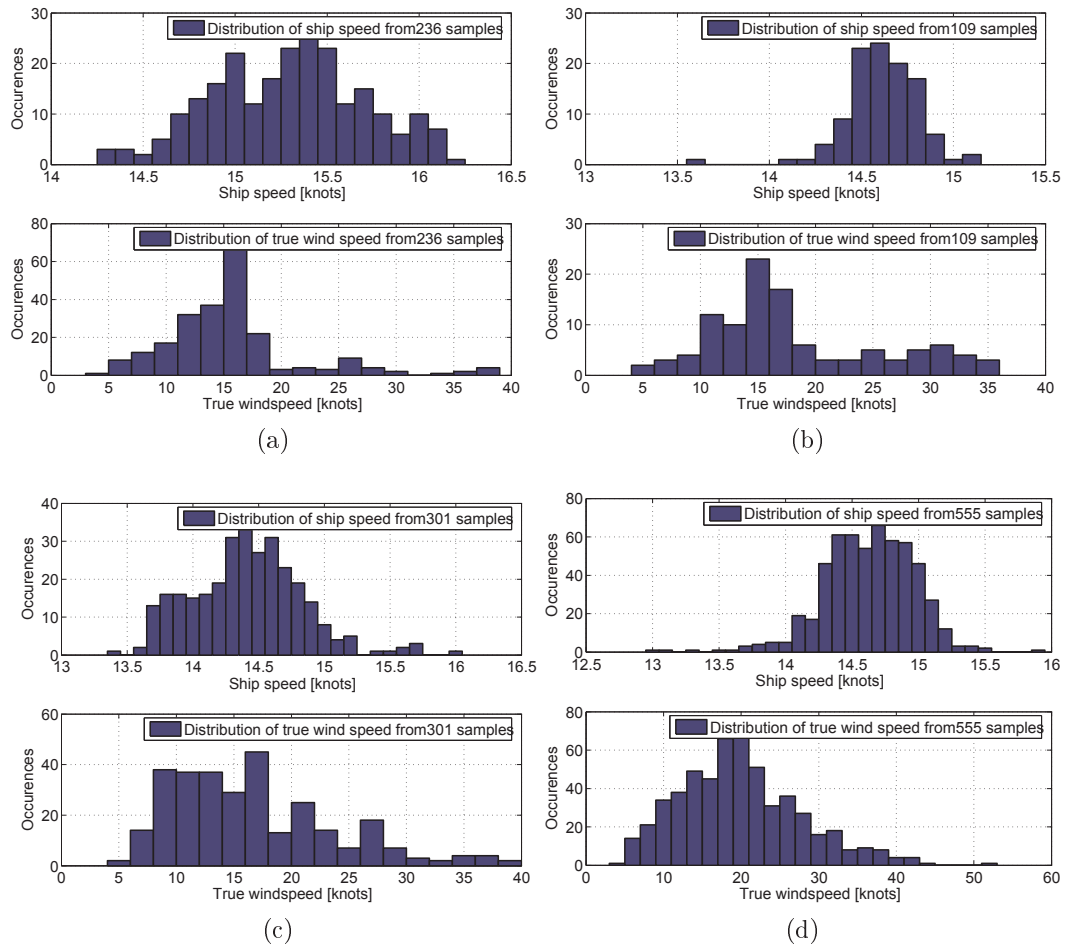


Figure 3.7: Ship speed and true wind speed distribution for dataset #1-#4

3.3.3 Heading variation

The variance of the heading is one of the governing variables on the variation of the propulsion power in *MDTM* data. Even small changes (less than 1°) in the heading have a significant influence on the measured propulsion power. This is illustrated in

Figure 3.8b where a change in heading of 5 deg to starboard results in a shaft power increase of approximately 400kW out of 9900kW or $\sim 4\%$, similarly the port side turn will result in a decrease of the shaft power as illustrated in Figure 3.8a. This asymmetry is assumed to be due to the propeller rotational direction as a right-handed (clockwise rotation looking from astern) propeller tends to turn the vessel to the port side if the rudder is in the center line and no corrections are made. Therefore it will consume more power by turning starboard than to port side. Since even small variations are detectable, measurements with a standard deviation in the heading of more than 0.6° were excluded.

It was not possible to explore this phenomenon in depth due to the lack of rudder motion samples. Data with rudder motions were analysed by Petersen (2012), and relevance determination shows that the rudder motions actually have a higher influence than the draught for that vessel and dataset.

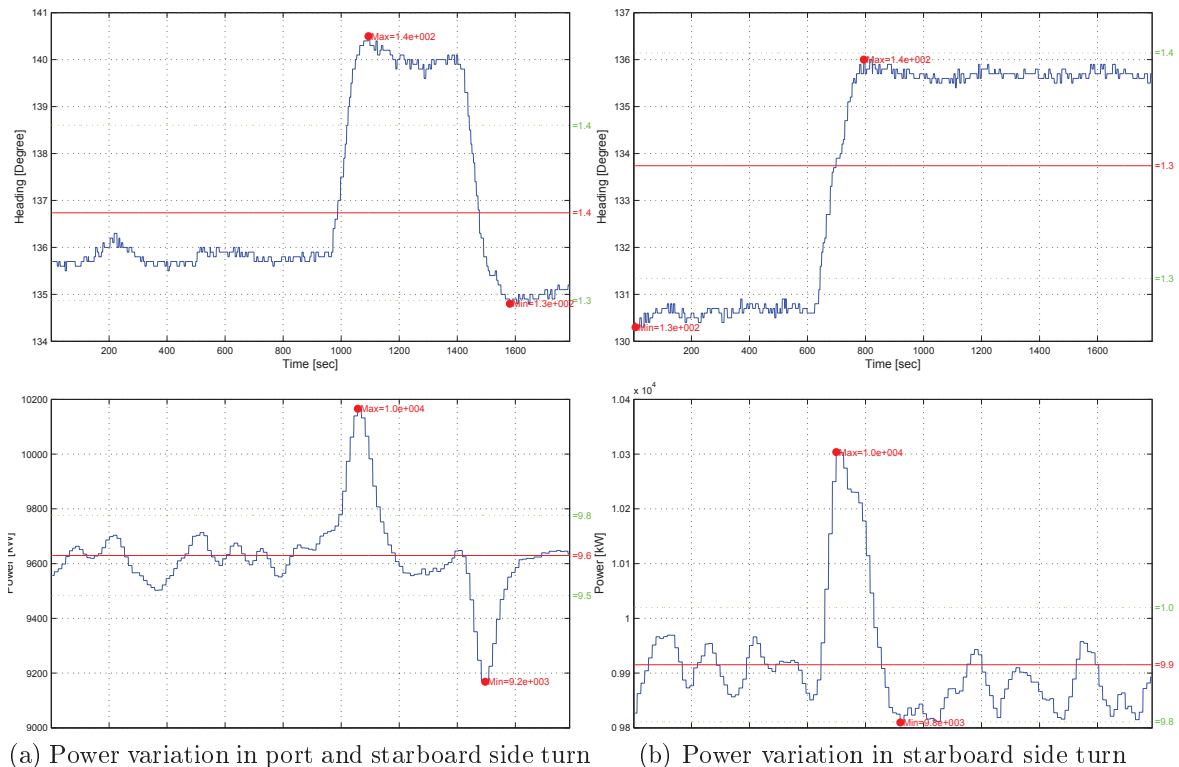


Figure 3.8: Variation in propulsion power due to turns

3.3.4 Noon report data from Torm Marie

Noon reports are part of the daily procedure on board any commercial vessel where all information regarding the vessel's general condition is documented. They contain information regarding the main and auxiliary engines as well as the vessel's position and weather observations. The data in the noon reports are obtained manually once a day, usually around noon, and thus represent the past 23-25 hours, (depending on change in the time zone and reporting time). Part of the data are readings that need to be reported, e.g. temperatures, average speed, and fuel consumption, where others are simply observations, e.g. wave height and direction, wind speed and direction, and also some are in between, e.g. draught fore and aft that has to be taken from the marks on the ship side, which can be difficult to do on board the ship. The draught is consequently only taken at departure and arrival. The noon report data is also a mix of average values, e.g. average speed and fuel consumption, and instant values, e.g. temperatures and weather conditions.

Noon reports are subject to human error due to the manual reporting system. The most common errors are typing errors or plus/minus errors which are usually identifiable, and subsequently deleted or altered. Other issues can be more difficult to identify such as outlying observations. This issue was addressed by purging the dataset of outliers as described in Section 3.5.2.

Table 3.6 shows a list of relevant variables from the noon reports.

Description	Unit
Latitude	deg
Longitude	deg
Report date and time of the day	-
Time zone at report position	-
Heading at report time	deg
Hours of continuous steaming in report period	HR
Distance over ground	nm
Distance through the water	nm
Average speed over ground	knots
Average speed through the water	knots
Wind force Beaufort	BF
Wind direction (absolute)	deg
Sea state Beaufort	BF
Sea direction (absolute)	deg
Air temperature	deg C
Seawater temperature	deg C
Rate of propeller shaft revolutions	rpm
Arrival draught on forward perpendicular	m
Arrival draught on aft perpendicular	m
Main engine fuel consumptions per day	tons/day
Main engine fuel consumptions per hour	tons/hr

Table 3.6: Torm Marie noon report variables

Date UTC	$N_{samples}$	T_m [m]	Trim, Ta-Tf [m]	U [knots]	T_{SW} [°]	SFOC HFO [tons/day]
09-12-2006 - 05-12-2008	323	7.35-	0-3.4	9.9-17.5	12-32	1.1-3.9

Table 3.7: Torm Marie noon report dataset

3.4 Containership data

Noon reports from five sister ships were systematically collected for a period of up to 10 years. These datasets are significantly different than the previous ones for several reasons. The logging period is long and includes both dry-dockings and hull cleanings, which gives interesting insight into how these operations influence performance. The data is well organized, purged of irrelevant data and seems to be very consistent, especially the manual observations of the wave height/direction and wind speed/direction. Since the ship type is different from *Torm Marie*, many of the values differed greatly, the service speed was significantly higher, and consequently the propulsion power was also higher.

Unfortunately the dimensions and ship names had to be kept confidential, and only the performance data was available. This made it impossible to make comparisons with traditional performance evaluation methods.

An overview of the datasets are given in Table 3.8 where the logging period is presented together with the total number of data points, number of dry-dockings and hull cleanings within the period.

Ship ID	Period Year	Total nr. NR	NR w. filter NR	No. DD	No. HCL
1	0-9.3	2337	1010	2	7
2	0-9.1	2283	1092	2	2
3	0-8.8	2268	1014	1	3
4	0-10.5	2679	1311	2	3
5	0-10.0	2564	1249	2	4

Table 3.8: Noon reports from containerships

Figures 3.9a - 3.9e show the distribution of the logged speed and shaft power of containership 1-5, they indicate a narrow speed profile with a mean around 23 knots and with almost no occurrence of speeds out of the range of 20-25 knots. The power distribution was much broader which indicates that the power was adjusted to meet the speed.

A comparison with hindcast data was performed for the container ships, the error of the observed/hindcasted data has been reported. In Table 3.9 the observed wind speed

is generally about 2m/s higher than the hindcast value, or 20-28 % higher. The mean values and standard deviations between the ships are very consistent. Table 3.10 shows the error between observed and hindcast significant wave height where, generally, the observed values are consistently underestimated for all of the ships with the exception of ship 1. The consistency in the estimates between the ships indicates that they operated in the same conditions and most likely on the same routes.

Ship ID	$\overline{V_{obs} - V_{HC}}$	$\sigma_{(V_{obs}-V_{HC})}$	$\overline{V_{obs}}$	$\overline{V_{HC}}$	$\sigma_{(V_{obs})}$	$\sigma_{(V_{HC})}$	$\frac{\overline{V_{obs}-V_{HC}}}{\overline{V_{obs}}}$
1	2.45	4.38	8.84	6.39	4.95	3.44	27.7%
2	2.14	4.45	8.65	6.51	4.89	3.48	24.8%
3	1.68	3.94	8.09	6.41	4.44	3.38	20.7%
4	1.97	4.23	8.66	6.70	4.98	3.60	22.7%
5	1.60	4.08	8.18	6.59	4.65	3.44	19.5%

Table 3.9: Comparison of hindcast and observed wind speed, $V[m/s]$ for the containership data

Ship ID	$\overline{H_{S_{obs}} - H_{S_{HC}}}$	$\sigma_{(H_{S_{obs}}-H_{S_{HC}})}$	$\overline{H_{S_{obs}}}$	$\overline{H_{S_{HC}}}$	$\sigma_{(H_{S_{obs}})}$	$\sigma_{(H_{S_{HC}})}$	$\frac{\overline{H_{S_{obs}}-H_{S_{HC}}}}{\overline{H_{S_{obs}}}}$
1	0.05	1.27	1.72	1.66	1.49	1.20	3.1%
2	-0.23	1.31	1.59	1.81	1.32	1.30	-14.4%
3	-0.25	1.19	1.43	1.68	1.26	1.27	-17.3%
4	-0.20	1.30	1.67	1.87	1.42	1.39	-12.0%
5	-0.25	1.51	1.60	1.85	1.57	1.30	-15.4%

Table 3.10: Comparison of hindcast and observed significant wave height, $H_S[m]$ for the containership data

Since the weather observations are instantaneous values, usually taken around the reporting time, and other important variables are average values from the previous noon report time, it is not correct to use them together. It would have been ideal to have "noon reports" for every hour to increase the accuracy of the hindcast weather information. As this was not possible, hindcasts were made between every noon report, but with one-hour intervals and equivalent positions, and then the average of the values between the present and the previous noon report were found. Similarly, the variance of the hindcasts for every noon report period was determined and made available for input to give more detailed weather information. The goal was to find weather that

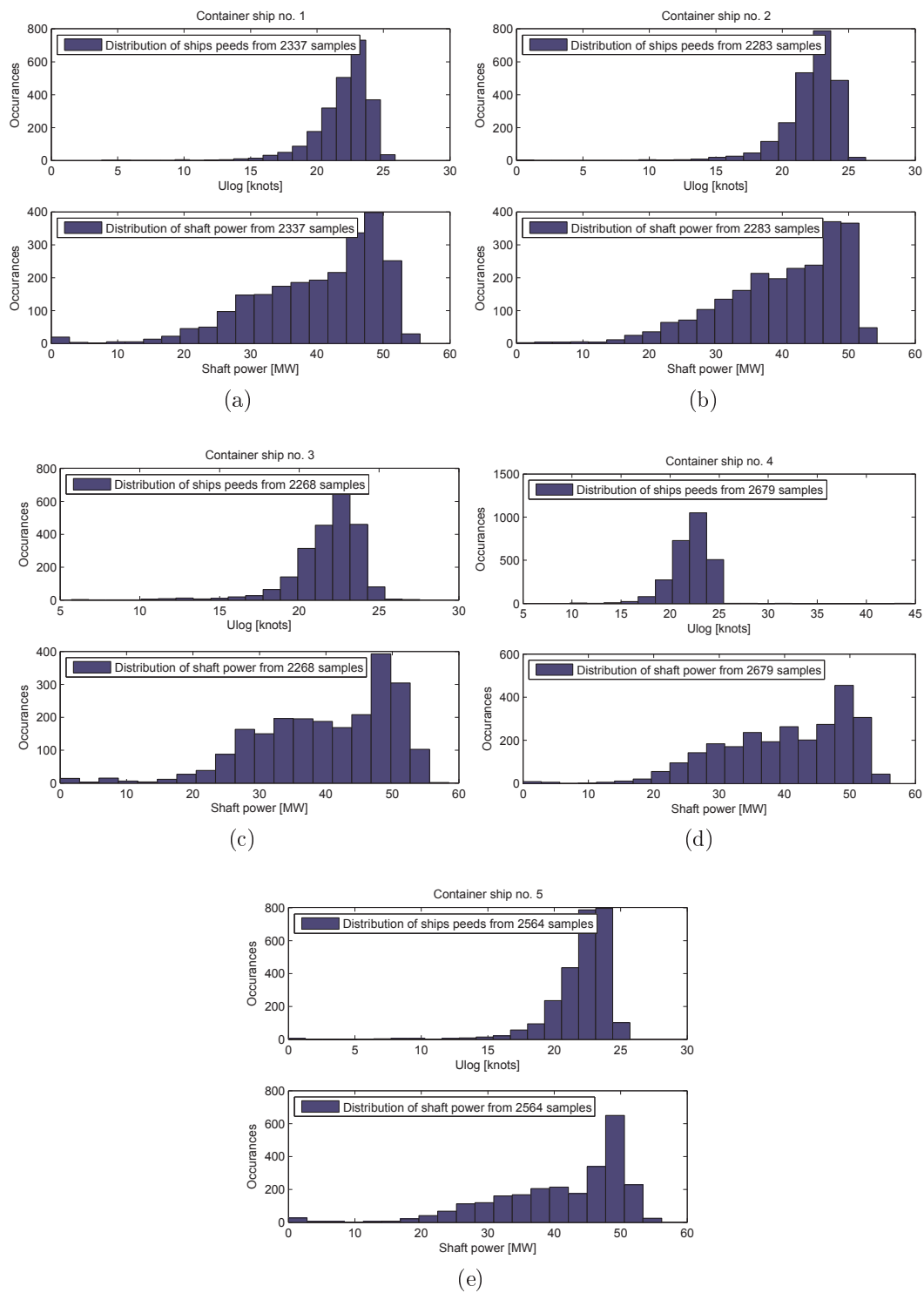


Figure 3.9: Ship speed and shaft power distribution for container ship 1-5

is equivalent to the effect spent in the same period, but since this is not possible, it is believed to be better than using only one observation (noon report time) to represent the past 24 hours.

3.5 Filtering

The three datasets were filtered differently due to the significant differences in data quality. They were filtered first to capture data logging errors, then to ensure a stationary state during the logging period and finally to exclude data with excessive impact from the environment.

The noon report data was generally purged for erroneous notations of the latitude and longitude which seemed to be a common problem.

A summary of the filtering described below is listed in Table 3.11.

3.5.1 Measured data from Torm Marie

Since the measurements were logged as 10-minute time series, most of the data is of a good quality. As mentioned in Chapter 3, the time series with variations in the heading of more than 0.6° were excluded since the effect of this was noticeable in the propulsion power. Furthermore, only measurements with a mean shaft power P of more than $6000kW$ and a relative wind speed V_R between $0 - 50knots$ were included in samples.

3.5.2 Noon report data from Torm Marie

By matching the noon reports from Torm Marie with the hindcast data, much filtering was eliminated since the hindcasts were generally only available on the ocean where the vessel was steaming. Although the length of the datasets were reduced significantly ($\approx 50\%$), hindcast data added a number of new consistent and relevant input variables to the datasets.

To avoid excessive changes in the heading within the noon report period, the heading difference between two noon reports was limited to 30° . Error in speed readings

were purged by only including noon reports within 5-25 knots.

3.5.3 Containership noon reports

Although the container ship data were of good quality, only some of the noon reports were valid for training and testing, usually because the ship was not operating in a stationary condition within the noon report period or e.g. was travelling in shallow water.

Aside from the basic information, the noon reports also included performance code parameters that indicated the operational profile of the ship during the reported time. Ten different codes indicated if the ship was steaming, manoeuvring, in harbour, stopped due to technical problems etc. This is very useful information when filtering the data for the most stationary operation.

Filtering thus only consisted of the performance code criteria ("Performance test", "Normal Cruising", "Power test") and minimum water depth of 100m.

Different parts of the full dataset with variable lengths was used to evaluate the effect of the dataset lengths, and subsequently the minimum logging period needed for reliable predictions could be determined.

Due to the geographical limitations of the hindcast data the training/test sets including hindcast data *HC*, were reduced by approximately 50% (for containership # 1). The training/test sets including only noon report data, and thus about twice the amount of data, were also tested with the same setup.

	Measured dataset	NR from Torm Marie	NR from container ships
Heading	$\text{std}(\phi) < 0.6^\circ$	COG < 30 between NR	COG < 30° between NR
Speed	-	5kn < SOG < 25kn	5kn < SOG < 35kn
Power	$\overline{P} > 6000\text{kW}$	-	-
Wind speed	$0\text{kn} < \overline{V_R} < 50\text{kn}$	-	-
Water depth	-	-	Waterdepth < 100m

Table 3.11: Input data filters

3.6 Input dataset for the regression model analysis

The following section presents input variables for the three data set in Table 3.12 - 3.14 that are used in the regression analyses presented in Chapter 4.

X	Data variable		ID	Unit
1	Speed through water	KYMA	U	knots
2	Relative wind speed	Airmar	Vrel	knots
3	Relative wind direction	Airmar	grel	°
4	Air temperature	Airmar	Tair	°C
5	Sea water temperature	NR	Tsw	°C
6	Air temperature	NR	Tair	°C
7	True wind speed	HC	Ws	m/s
8	True wind direction	HC	g	°
9	Significant wave height	HC	Hs	m
10	Wave period	HC	Tp	s
11	True wave direction	HC	Td	°
12	Relative wind speed	HC	Vrel	m/s
13	Relative wind direction	HC	grel	°
-	Shaft power	KYMA	P	kW

Table 3.12: Data variables for the measured dataset analysis from Torm Marie

X	Data variable		ID	Unit
1	Average ship speed through the water	NR	U	knots
2	Air temperature	NR	Tair	°C
3	Sea water temperature	NR	Tsw	°C
4	Mean draught $(T_a+T_f)/2$	NR	Tm	m
5	Trim, T_a-T_f	NR	Trim	m
6	Average significant wave height during <i>NR</i> period	HC	mean(Hs)	m
7	Average wave period during <i>NR</i> period	HC	mean(Tp)	s
8	Average relative wave direction during <i>NR</i> period	HC	mean(Tdrel)	°
9	Variance of the significant wave height during <i>NR</i> period	HC	var(Hs)	m
10	Variance of the wave period during <i>NR</i> period	HC	var(Tp)	s
11	Variance of the wave direction during <i>NR</i> period	HC	var(Td)	°
12	Average relative wind speed	HC	mean(Vrel)	m/s
13	Average relative wind direction	HC	mean(gammarel)	°
14	Wind force (Beaufort)	NR	WindBF	BF
15	Wind direction	NR	WindDir	°
16	Report date and tim (Matlab numeric value)	NR	UTC	-
-	Average HFO consumption per hour	NR	HFOME	tons/hr

Table 3.13: Data variables for the *NR* dataset analysis from Torm Marie

X	Data variable		ID	Unit
1	Speed through water	NR	Ulog	knots
2	Speed over ground	NR	Uobs	knots
3	Sea water temperature	NR	Tsw	deg
4	Mean draught $(T_a+T_f)/2$	NR	Tm	m
5	Trim, T_a-T_f	NR	Trim	m
6	True wind speed	NR	WindSpeed	m/s
7	Relative wind direction	NR	WindDir	deg
8	Average relative winds speed during <i>NR</i> period	HC	mean(Vrel)	m/s
9	Average relative winds direction during <i>NR</i> period	HC	mean(gammarel)	deg
10	Average significant wave height during <i>NR</i> period	HC	mean(Hs)	m
11	Average wave period during <i>NR</i> period	HC	mean(Tp)	s
12	Variance of the significant wave height during <i>NR</i> period	HC	var(Hs)	m
13	Variance of the wave period during <i>NR</i> period	HC	var(Tp)	s
14	Variance of the wave direction during <i>NR</i> period	HC	var(Td)	deg
15	Variance of the winds speed during <i>NR</i> period	HC	var(Ws)	m/s
16	Variance of the winds direction during <i>NR</i> period	HC	var(gamma)	deg
17	Report date and time (Matlab numeric value)	NR	UTC	-
18	Average winds speed during <i>NR</i> period	HC	mean(Ws)	m/s
19	Average winds direction during <i>NR</i> period	HC	mean(gamma)	deg
27	Sea state	NR	SeaState	m
28	Relative sea direction	NR	TrueRelSeaDir	deg
-	Average shaft power	NR	PropPower	kW

Table 3.14: Data variables for the container ship noon report dataset analysis

3.7 Data summary

This chapter outlined the variables needed for determining the propulsion power of a ship as well as the effect of each of the variables on propulsion power and how they were obtained in practice. Hindcast weather information and the installation of the on-board measurement system on Torm Marie was described. Manual observations made on board Torm Marie were compared with data measured during the test period and a discussion of how the high data resolution of the measured data facilitated a detailed study of how changes to the heading affected instantaneous propeller power, is given.

A detailed account of data sampling and how the measured data was split into four datasets, each one for a single voyage with a constant loading condition, is given.

The acquisition and quality of Noon report data from Torm Marie over a two year period was described.

The noon report data from five sister containerships are presented. These datasets were collected in a period of up to 10 years and include dry-dockings, hull cleanings and propeller polishings. The wind speed and significant wave heights from the noon reports were compared with the corresponding hindcast values and showed surprisingly consistent differences across the five ships with the observed wind speed generally being higher than the *HC* and the observed wave height generally being lower than the *HC* values.

Finally a summary of the filtering used for the three sources of data (*MDTM*, *NRTM* and *NRCS*) is given.

Chapter 4

Analyses using the different propulsion prediction methods

This chapter presents and compares the results from the methods described in Chapter 2. The available data allowed for different types of analysis that enabled us to analyse different aspects of the data-driven methods. The measured data from Torm Marie (*MDTM*) had a very high resolution with many variables although no variation in the loading conditions were included, furthermore the measurements ensured more consistent data and are regarded as the best datasets which were therefore used for comparing the regression models. The noon report data from Torm Marie (*NRTM*) and containerships 1-5 (*NRCS* 1-5) featured a significantly lower data density, but they included many different loading conditions over a long period of time. This offered the possibility of including *time* as a variable and since all weather data was obtained manually, the effect from introducing additional weather information from hindcast data could also be analysed.

First the four measured datasets from Torm Marie were used to evaluate the prediction errors with the empirical methods (*EMP*) and the four regression models: linear (*LIN*), custom non-linear (*NONLIN*), Artificial Neural Network (*ANN*) and Gaussian Process Regression (*GPR*). This analysis allowed us to compare the prediction performance of the different methods.

Generally the non-linear models *ANN* and *GPR* were superior in performance

	<i>time</i> input included	without <i>time</i> input
Including hindcast	H <i>Ctime</i>	HC <i>Notime</i>
Without hindcast	NoHCT	NoHC <i>Notime</i>

Table 4.1: Selecting the best model of Input Variable Combinations (*IVC*) for each of the four given variable

and were used for the predictions of the noon report data from Torm Marie and Containerships 1-5. The prediction errors for *ANN* and *GPR* regression were similar, but due to the dataset sizes and feature availability, the *GPR* was primarily used for further investigation. Initially the comparisons for the containerships were performed on dataset 1 in the period from launching until the first dry-docking, and the best combinations of the input variables were found in this period.

The large amount of data from the containerships allowed more possibilities for training and evaluations. Five sister ships were available, each with a history of approximately 10 years and some with several dry-dockings. The weather data from the noon reports are consistent and of reasonable quality, and as discussed in Section 3.2.8, it was thus possible to use the data without introducing hindcast data.

The input dataset from the noon reports had additional data from hindcast weather information as described in Chapter 3, but including them limited the amount of input data because hindcast data was not available in all areas. The inclusion of hindcast data added weather and sea information. However, it also reduces the input dataset by approximately 50%. This effect was investigated by direct comparison between datasets with only noon report data and datasets including hindcasts.

Four combinations of the input variables (*IVC*) were of particular interest: best *IVC* only including noon report data and best *IVC* including hindcast data, both with and without *time* as a variable. Results defined these optimal *IVC* for containership 1, and the datasets from containerships 2-5 were subsequently trained with these *IVC*. The matrix for testing the optimum *IVC* for containership 1 is presented in Table 4.1.

In order to evaluate the length of a training period needed for obtaining reliable results, containerships 1-5 were trained with data from the first year, then data from the first two years, accumulating one year at a time until the first dry-docking that

occurred after approximately 5 years as illustrated in Table 4.2

Training period	Year 1	Year 2	Year 3	Year 4	Year 5
1	█				
2	█	█			
3	█	█	█		
4	█	█	█	█	
5	█	█	█	█	█

Table 4.2: Accumulation of the training year for containerships 1-5

The *LIN*, *NONLIN* and *ANN* models were evaluated by cross-validated relative prediction errors $\overline{\omega_K}$ as defined in Section 2.3.1. They were based on a dataset subdivision into 5 subsets ($K = 5$), and the *GPR* was trained and tested with *LOO*, but for comparative purposes a subdivision of 5 subsets was used for the *GPR* as well.

An overview of the combinations between method and dataset are given in Table 4.3.

By evaluating all of the datasets with both *ANN* and *GPR* using different combinations of input variables, the prediction errors produced can be used to evaluate the effect of specific variables. The characteristic *length-scales* in the *GPR* model represent the influence of every input variable as described in Section 4.5. Length-scales are a more direct way to evaluate which variables are most important. The *length-scales* reported are average values based on the *LOO* training and test procedure. The magnitudes of the *length-scales* often vary significantly, and for presentation purposes it is given on a logarithmic scale.

Data	Prediction Methods					
	<i>EMP</i>	<i>LIN</i>	<i>NONLIN</i>	<i>ANN</i>	<i>GPR</i>	LT trend
Torm measured data 1-4	X	X	X	X	X	
Torm NR				X	X	
Containership NR 1-5				X	X	X

Table 4.3: Applied methods and dataset combinations tested, including Long-term Trend evaluations (LT Trend)

4.1 Evaluation of empirical methods for the measured data

Results from the empirical methods are mainly included to illustrate how they can result in significant discrepancies if used without any adjustments to the model from experience with similar ships. Furthermore, the lack of fixed ship parameter data made it impossible to test the containership data using the empirical methods.

The propulsion power was estimated using the empirical methods Harvald (1983) (*GH*) and Holtrop and Mennen (1984) (*HM*) described in Chapter 2. The models were used with all the hydrostatic parameters given as well as the propeller open water data from the model test. The wind resistance was based on Isherwood (1972). All other parameters were estimated by the models without modification. As illustrated in Figures 4.1a and 4.1b, both methods show significant error, up to above 50%, compared to the measured values or as found in Pedersen and Larsen (2009a) cross validation errors up to 28%. *GH* generally over-predicts and *HM* under-predicts the measured propulsion power. In Figure 4.1b, the errors from both methods seem to have linear trends, *GH* with a higher slope than *HM*. This indicates that both methods put too much weight on the speed, or that the wind resistance could be overestimated.

Both methods could be improved by adjustments to model test data, sea trial data, or practical experience from sister ships. However, because this thesis is seeking solutions based on a minimum number of manual adjustments, the errors of the empirical methods have mainly been presented for illustration purposes.

4.2 Comparison between prediction methods based on logged data

The four datasets consisted of samples of measurements on board the product tanker Torm Marie (*MDTM*) and have been used as a benchmark between the prediction methods and to show how measured data can improve the prediction performance. As described in Chapter 3, the datasets are measured data from four individual voyages

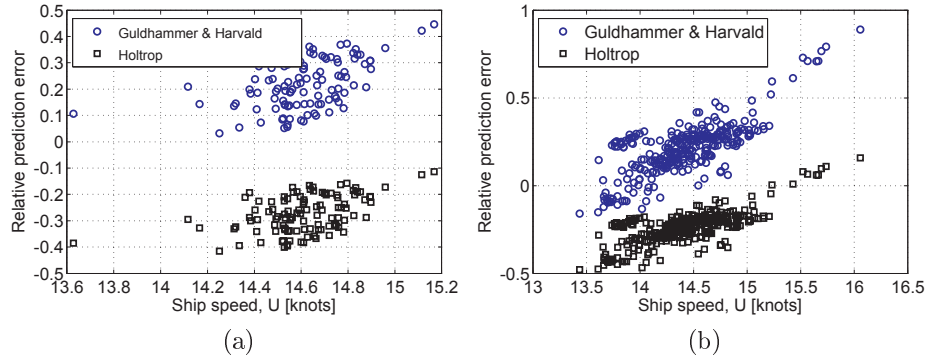


Figure 4.1: Prediction errors for the empirical methods: Guldhammer & Harvald and Holtrop-Mennen

where the loading conditions are assumed to be fixed, and therefore the data cannot be directly compared with the noon report data from Torm Marie. (*NRTM*)

ANN and *GPR* regression models facilitate an infinite number of input variables and thus combinations of input variables. *ANN* can be vulnerable to excessive variables that might mislead the predictions. It is therefore recommended that different combinations of the input variables are tested. *GPR* is less sensitive to excess variables. The input combinations are given in Table 4.4.

4.2.1 Results using the linear and the custom non-linear methods with the measured data from Torm Marie

The linear and custom non-linear methods were developed to predict the trend within the measured data, and since *ANN* and *GPR* showed significantly better results, further development was discarded.

Based on linear and non-linear regression methods described in Section 2.2.1 and 2.2.2, the cross-validated prediction errors were calculated. The linear and non-linear methods were trained like the *ANN* and *GPR* with five cross-validation training/tests (80%/20%)-set for the measured data. For each dataset, the average of the relative prediction errors is shown in Table 4.5.

X	ID	Data variable	Input variable setup - IVC					
			2	8	10	11	12	15
1	U	Speed through water	x	x	x	x	x	x
2	Vrel	Relative wind speed	x	x				x
3	grel	Relative wind direction	x	x				x
4	Tair	Air temperature	x	x				x
5	NR.Tair	Air temperature			x	x	x	x
6	NR.Tsw	Sea water temperature	x	x	x	x	x	x
7	HC.Ws	True wind speed		x	x			x
8	HC.g	True wind direction		x	x			x
9	HC.Hs	Significant wave height		x	x		x	x
10	HC.Tp	Wave period		x	x		x	x
11	HC.Td	True wave direction		x	x		x	x
12	HC.Vrel	Relative wind speed				x	x	x
13	HC.grel	Relative wind direction				x	x	x
Y	P	Shaft power	x	x	x	x	x	x

Table 4.4: Input variable setup combinations (IVC) for the measured data from Torm Marie ("NR" indicates that the data originates from the noon reports and "HC" indicates that the data originates from hindcasts)

The linear model performed best with three simple input variables: the ship speed U , the relative wind speed V_R and the wind direction γ_R . In Table 4.5, it is evident that the linear model is superior to the non-linear model. This indicates that the dynamics of the model are not necessarily related to the propulsion power with the speed cubed and wind speed squared as previously assumed for the present speed range. Table 4.5 also shows that there is a large difference between the results of the datasets, and there is the same tendency for both the linear and non-linear methods. This may be due to the variability in the datasets. In Figure 3.7, datasets 2 and 4 have slightly smaller speed ranges that are more uniformly distributed than in datasets 1 and 3. This is not consistently reflected by the standard deviation of the power and ship and wind speeds in Table 3.5, but it is noted that dataset 2 with the smallest prediction error also has the smallest variations here.

It is emphasized that the measured dataset does not include variations of draught and trim since they are measurements from four different loading conditions.

Figures 4.2a to 4.2d show that the error distribution for the linear regression

models is only normally distributed in dataset 2. In dataset 4, the distribution is a bit skewed, and both 1 and 3 are multiply distributed and significantly skewed.

M	Linear	Non-linear
	$\overline{\omega_K}$ %	$\overline{\omega_K}$ %
1	9.35	11.36
2	1.59	3.58
3	7.85	10.79
4	3.39	5.98

Table 4.5: Cross-validation errors $\overline{\omega_K}$ of the linear and non-linear methods for the measured dataset

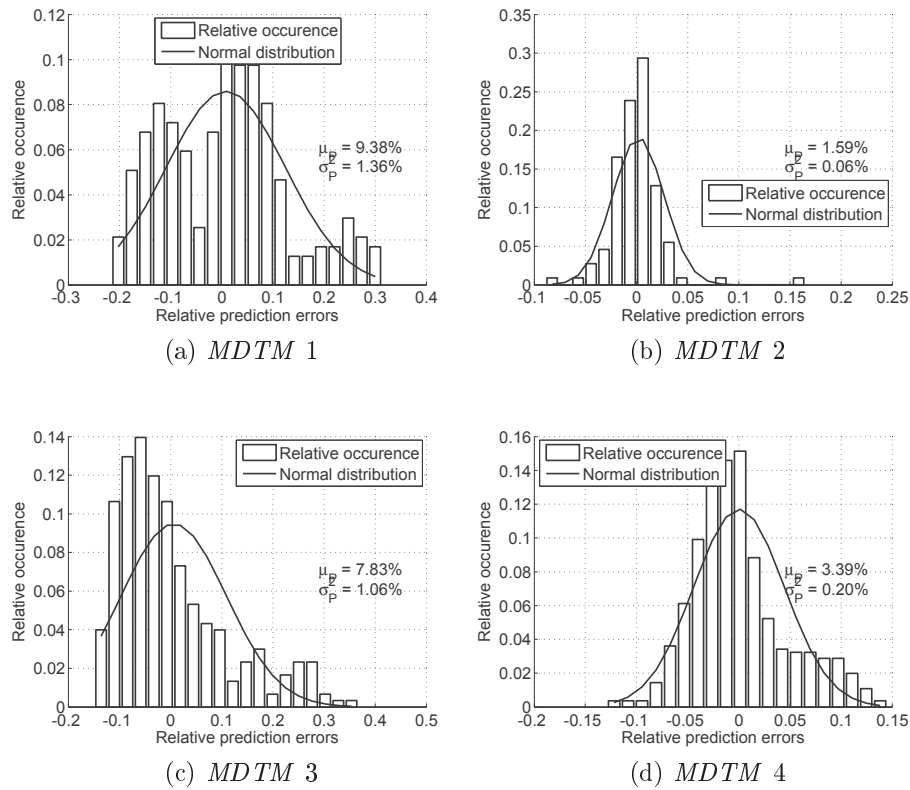


Figure 4.2: Histograms of the relative prediction errors from the measured dataset (*MDTM 1-4*) using the linear method

4.2.2 Evaluation of ANN regression of the measured dataset from Torm Marie

This section presents the results from the ANN method on the measured dataset from Torm Marie (*MDTM*) for comparison with the other regression models. Initial tests in Pedersen and Larsen (2009b) of the measured dataset show that the model with the high number (20) of hidden units had the lowest error, and for computational efficiency, the network was subsequently only trained with 5 and 20 hidden units.

The network was trained with 80% of the dataset and tested with the remaining 20% of the dataset alternately five times so all parts of the dataset were used.

Table 4.6 shows that the network with 20 hidden units is consistently better than the one with 5 hidden units for all the measured datasets. As discussed in Section 4.7, datasets 2 and 4 had very little variance in the input and output variables which makes regression more simple. This is also evident in Table 4.6 where the results of dataset 2 are similar for 5 and 20 hidden units.

For the two remaining datasets, the tendency is that the lowest prediction errors are found with input variable set *12* which consists of the logged ship speed, sea and air temperature and hindcast data for wind and waves, and 20 hidden units.

The values in Table 4.6 are generally very small, but since they are based on the mean values of the time series, it cannot be expected that they can reach values below the relative standard deviation of the measured output variable (the propulsion power) presented in Table 3.5.

Figures 4.3a-4.3d show histograms of the relative prediction errors together with a normal distribution fitted to this. All the histograms have higher peaks than the normal distribution, but all are around zero. For dataset 2 in Figure 4.3b, the histogram has a few outliers; it should be noted that these still have small prediction errors.

4.2.3 Results from GPR with measured datasets from Torm Marie

The measured datasets were trained using the same *IVC* used for testing the *ANN* and all available input variables equivalent to *IVC 15* in Table 4.4. The training and

testing was performed as described in Section 2.4.2 with cross-validation of the same five training/test sets as for the *ANN*.

In Table 4.7, a collection of the cross-validation errors together with the mean of the predicted standard deviations are presented. As in the previous analysis with the *MDTM* data, it is also evident that datasets 2 and 4 both have smaller prediction errors and also smaller prediction standard deviations.

The distributions of the relative prediction errors are presented in Figure 4.4a-4.4d, and the cross-validation error for each dataset is shown in Table 4.7. Comparing with the *ANN* prediction errors in Section 4.2.2, the prediction errors are smaller. Furthermore the average predicted standard deviations are presented and show that the standard deviation tends to increase with the increase of the corresponding cross-validation error.

In the two columns to the right, the average of the relative prediction errors and the relative standard deviation for datasets 1-4 are shown for each *IVC*. *IVC 8, 10, 12* and *15* all have similar prediction errors, but *IVC 15* has a smaller relative prediction standard deviation.

Figures 4.5a- 4.5d show the relation between the prediction errors ω_k and the predicted standard deviation σ_k . Generally the *MDTM 2* and *4* have a lower level of σ_k with a more confined distribution, whereas *MDTM 1* and *3* have more outliers.

<i>NH</i>	Dataset 1		Dataset 2		Dataset 3		Dataset 4		Average $\overline{1-4}$	
	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$	$\overline{\omega_K}$
<i>IVC</i>	%	%	%	%	%	%	%	%	%	%
2	3.93	2.92	1.07	0.81	3.07	2.37	1.72	1.30	2.45	1.85
8	2.63	1.97	0.97	0.89	2.21	1.65	1.49	1.04	1.83	1.39
10	2.14	1.65	0.99	0.95	2.17	1.65	1.52	0.94	1.71	1.30
11	3.77	2.79	1.10	0.90	2.45	1.88	1.42	1.02	2.19	1.65
12	2.25	1.65	0.99	0.94	1.75	1.40	1.28	0.90	1.57	1.22

Table 4.6: Cross-validation errors, $\overline{\omega_K}$, of the *ANN* for the measured data from Torm Marie for different Input Variable Combinations (*IVC*) and Number of Hidden units (*NH*)

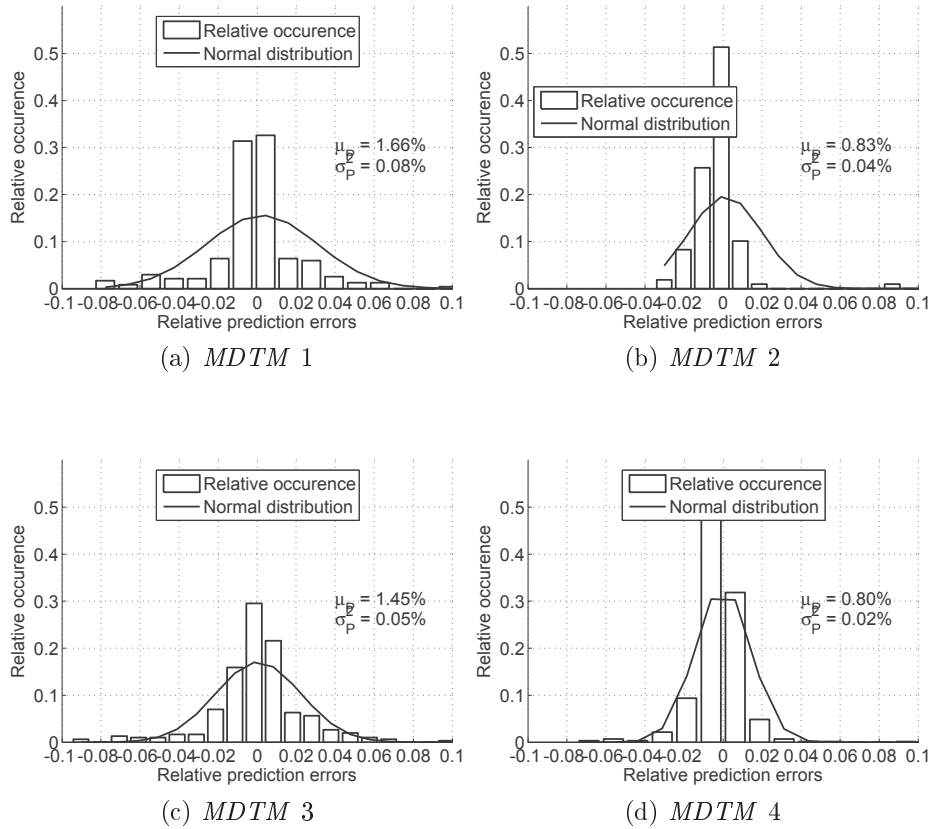


Figure 4.3: Histograms of the relative prediction errors from the measured datasets (*MDTM 1-4*) using ANN, with 20 hidden units and *IVC 12* for all datasets, except dataset 2 where *IVC 2* is used.

4.2.4 Comparison of regression models for the measured data from Torm Marie

Comparison of the results from four different regression methods was conducted with the four measured datasets (*MDTM*) from Torm Marie in order to evaluate the prediction performance. Figure 4.6 gives an overview of the lowest prediction errors, and it clearly shows that *ANN* and *GPR* both reveal significantly smaller prediction errors than the remaining methods. Both *ANN* and *GPR* are very similar in performance, and they are consistent across the datasets. The measured datasets are, as mentioned

IVC	Dataset 1		Dataset 2		Dataset 3		Dataset 4		Average $\overline{1-4}$	
	$\overline{\omega_K}$ %	$\overline{\sigma_K}$ %	$\overline{\omega_K}$ %	$\overline{\sigma_K}$ %	$\overline{\omega_K}$ %	$\overline{\sigma_K}$ %	$\overline{\omega_K}$ %	$\overline{\sigma_K}$ %	$\overline{\omega_K}$ %	$\overline{\sigma_K}$ %
2	2.33	2.99	0.77	0.63	1.89	2.76	0.88	1.24	1.47	1.91
8	1.35	1.75	0.74	0.75	0.91	1.24	0.52	0.73	0.88	1.12
10	0.89	1.28	0.82	0.87	1.16	1.75	0.57	0.82	0.86	1.18
11	1.79	3.06	0.97	1.24	1.34	1.79	2.80	3.65	1.73	2.44
12	0.85	1.28	0.90	1.05	1.01	1.39	0.54	0.79	0.83	1.13
15	1.00	1.37	0.78	0.68	1.10	1.25	0.56	0.74	0.86	1.01

Table 4.7: Crossvalidation errors $\overline{\omega_K}$ of the GPR for the measured dataset for different Input Variable Combinations IVC

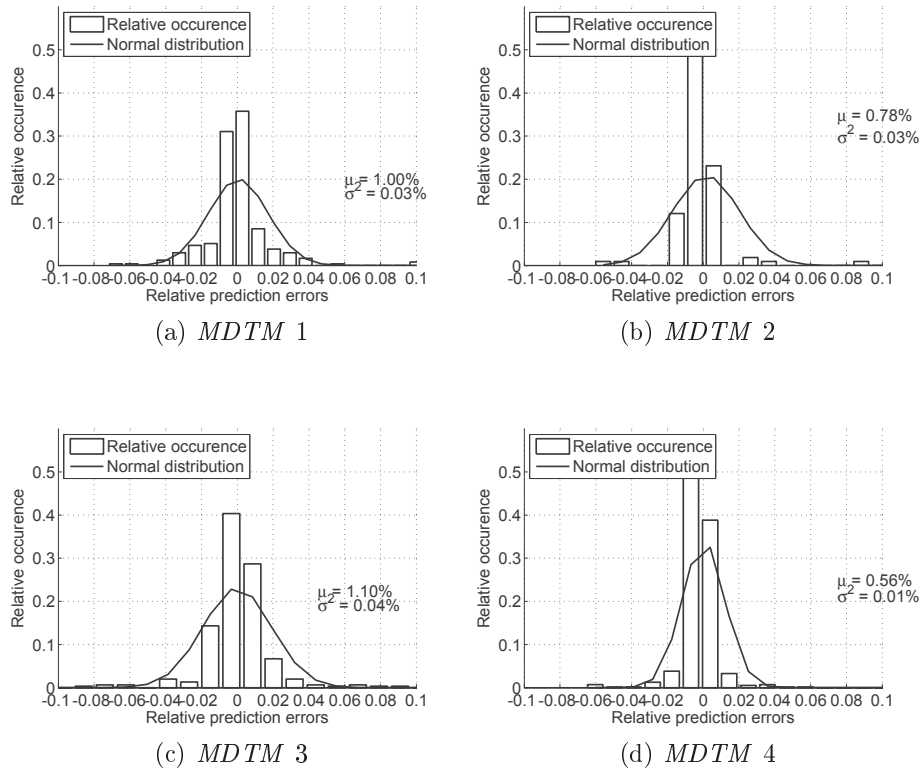


Figure 4.4: Histograms of the relative prediction errors from the measured data ($MDTM 1-4$) sets using GPR with input variable setup 15

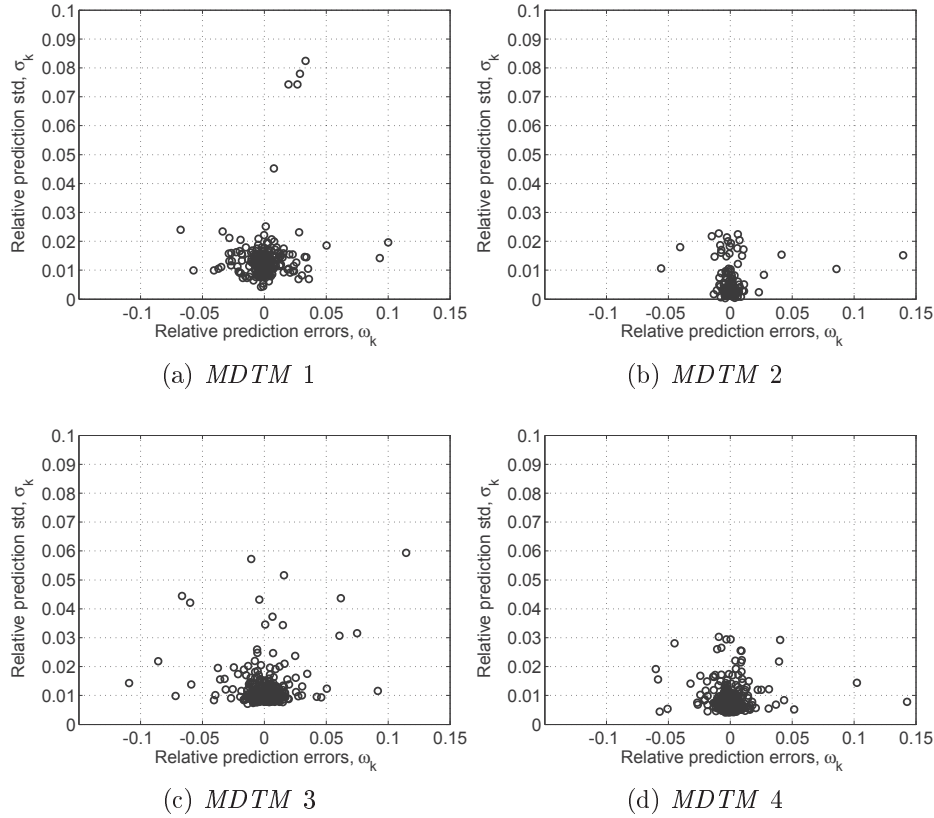


Figure 4.5: The relative prediction errors vs. the predicted standard deviation from the *GPR* analysis with the measured datasets (*MDTM 1-4*)

before, without variation in draught and trim. Introducing the draught and trim as additional variables, as with the noon report data, is not expected to change these trends, and the *ANN* and *GPR* models are expected to perform equally well or better than the linear and non-linear methods for any performance input dataset. This led to the decision to only use *ANN* and *GPR* for analysis of the remaining datasets.

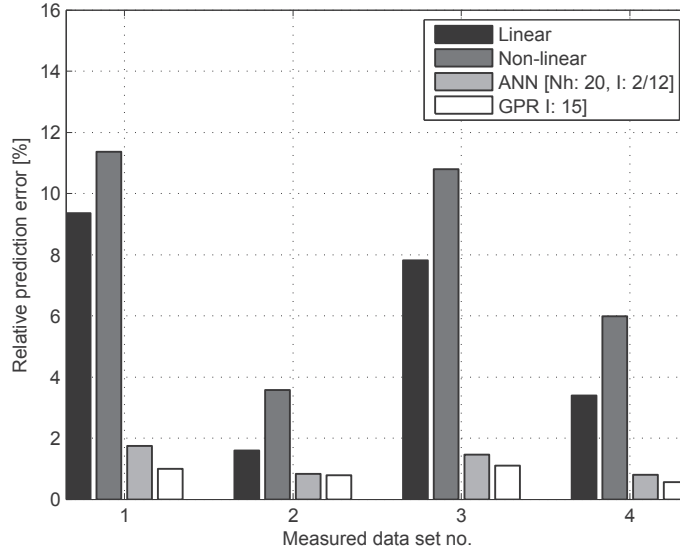


Figure 4.6: Comparison of the relative prediction errors between all the regression methods: linear, non-linear with and without modifications and the best input variable combinations of ANN and GPR methods of measured datasets 1-4

4.3 Analysis of noon report data from Torm Marie using *ANN* and *GPR*

This section collects the results found by training and testing the noon report data from Torm Marie (*NRTM*) with the *ANN* and *GPR* models. Using this data allowed evaluation of prediction methods for the same vessel as for *MDTM*, but for a longer duration which includes changes in loading conditions. Unfortunately the data was based on noon reports and was not directly comparable with *MDTM* due to the sparse data density.

The training and test procedure for the noon report data from Torm Marie was similar to the one for the *MDTM* data, but due to the different input-output data, different input variable combinations (*IVC*) had to be used. Table 4.8 specifies 8 different *IVC* selected in order to evaluate the effect of the different parts of the data, for example with and without the weather data. For this reason four of them included the *time* as a variable.

X	1	4	6	7	8	9	10	13	15	16
1 Avg(U)	x	x	x	x	x	x	x	x	x	x
2 NR.Tair	x	x	x	x	x	x	x	x	x	x
3 NR.Tsw	x	x	x	x	x	x	x	x	x	x
4 Tm	x	x	x	x	x	x	x	x	x	x
5 Trim	x	x	x	x	x	x	x	x	x	x
6 mean(HC.Hs)		x			x			x	x	x
7 mean(HC.Tp)		x			x			x	x	x
8 mean(HC.Td)		x			x			x	x	x
9 var(HC.Hs)						x			x	x
10 var(HC.Tp)						x			x	x
11 var(HC.Td)						x			x	x
12 HC Vrel			x	x	x	x		x	x	x
13 HC gammarel			x	x	x	x		x	x	x
14 NR.WINDCONDITIONATNOONDEG							x		x	x
15 NR.WINDCONDITIONATNOONBF							x		x	x
16 NR.UTC				x		x	x	x	x	

Table 4.8: Input variable combinations (*IVC*) for noon report data from Torm Marie

4.3.1 Evaluating the prediction errors of the *NRTM* using *ANN*

The relative prediction errors ($\overline{\omega_K}$) from the *ANN* regressions in Table 4.9 show how they are related to the number of hidden units (5, 10, 15, 20) in the *ANN* and *IVC*. It is noted that the prediction errors are consistently smaller for the simplest network with 5 hidden units compared with more complex networks with more hidden units. This is in contrast to the findings in Section 4.2.2 where the *ANN* with 20 hidden units performs best. The best result is obtained by input variable combination 7, which is one of the simpler *IVC* with five hidden units giving a relative prediction error of 7.18%, but all the *IVC* that include the time (*IVC* 7, 9, 10 and 13) have considerably lower $\overline{\omega_K}$ than the ones without. A simpler network with only 5 hidden units with few input variables indicate that the problem is less complicated, and a more complicated network, Figure 4.7, shows the error histograms and an estimated normal distribution for this combination which has a reasonable fit regardless of some outliers.

IVC	Number of hidden units, NH			
	5 %	10 %	15 %	20 %
1	8.68	-	-	9.51
4	8.70	10.00	11.22	13.78
6	8.18	8.50	9.07	9.05
7	7.18	7.74	9.28	10.06
8	8.46	9.06	9.60	9.18
9	7.26	8.47	10.58	9.25
10	7.21	8.50	10.89	12.46
13	7.77	-	-	-

Table 4.9: 20/80 cross-validation errors $\overline{\omega_K}$ from ANN for the noon report data

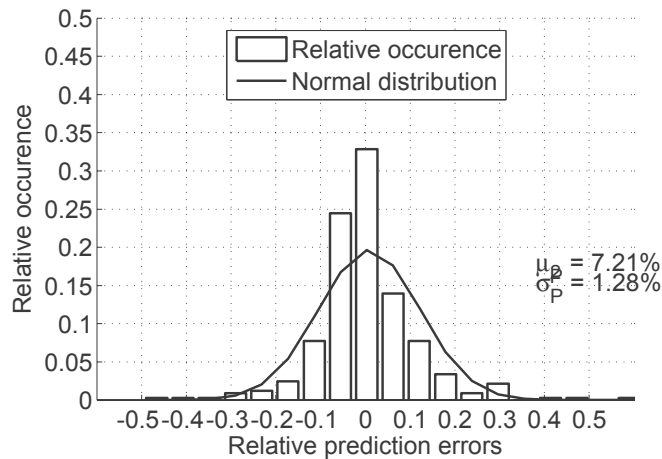


Figure 4.7: Histogram and normal distribution of the relative prediction errors from ANN with 5 hidden units and IVC 7, using the noon report dataset from Torm Marie

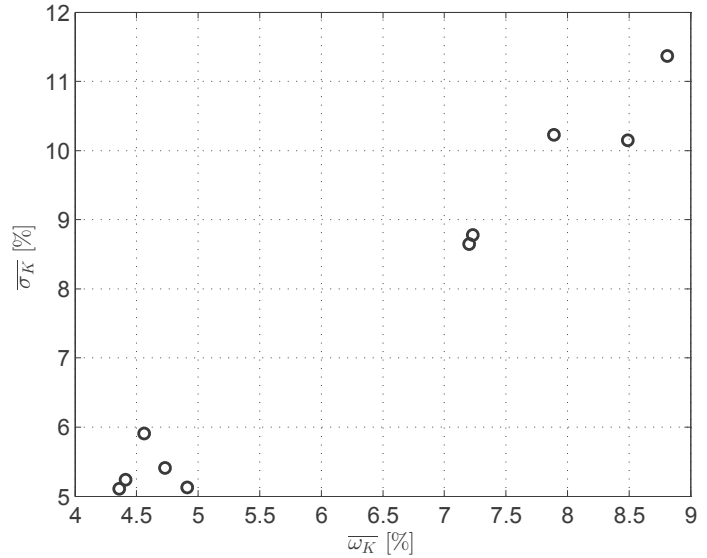
4.3.2 Evaluating the prediction performance of the *NRTM* with *GPR*

In Table 4.8a, the summary of the cross-validation prediction errors $\overline{\omega_K}$ and standard deviations $\overline{\sigma_K}$ shows that there is not much variation between the best input variable combinations (7,9,10,13). Furthermore, the results are shown in Figure 4.8b illustrating that the cross-validation errors and standard deviations are positively correlated.

IVC 15 has the lowest prediction error and a reasonably low prediction standard deviation. As with the measured dataset, the GPR method managed to identify the best results by using all the available input variables, including the *time* input. The distributions of the prediction errors and the predicted standard deviations from *IVC 15* are presented in Figure 4.9a. The prediction error histogram is offset to the left whereas the normal distribution is well centred around zero but does not fit the histograms very well.

Figure 4.9b gives a picture of how the relative prediction errors (ω_k) and the relative predictive variance (σ_k) are related. This shows a narrow distribution of ω_k and that the standard deviation σ_k is only partially correlated to ω_k , meaning that for ω_k until 0.2 it is not given that σ_k increases, but for higher errors, the standard deviation seems to increase, except at one point. High σ_k for low ω_k is not regarded as a problem because it is a conservative estimate, and in post-processing, this is more important than having large errors with small σ_k . The standard deviations are relatively low with the majority of the occurrences below 5%. Both figures suffer from a few outliers that could be purged by the high prediction standard deviation.

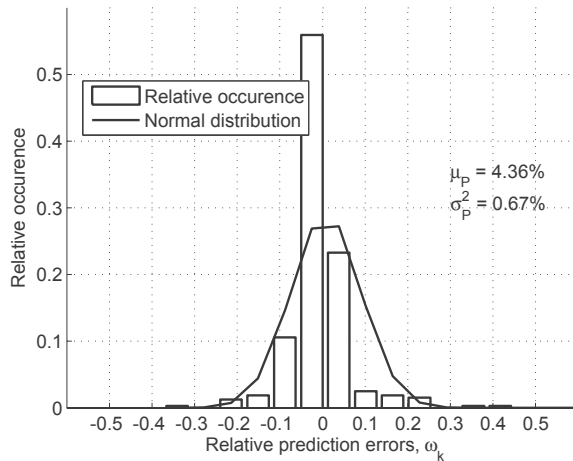
<i>IVC</i>	$\overline{\omega_K}$ %	$\overline{\sigma_K}$ %
1	8.81	11.37
4	8.49	10.15
6	7.89	10.23
7	4.73	5.41
8	7.23	8.78
9	4.91	5.13
10	4.56	5.91
13	4.41	5.24
15	4.36	5.11
16	7.20	8.65



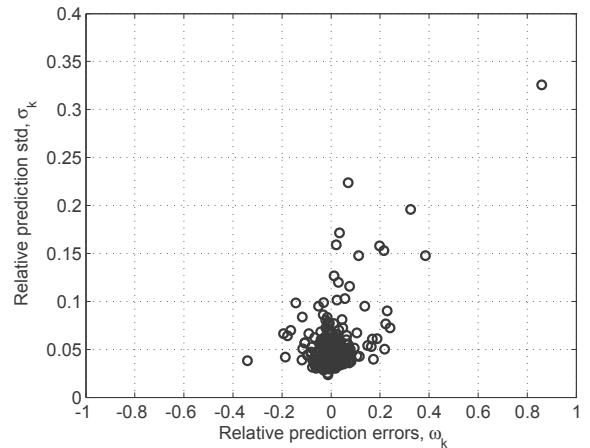
(a) Tabular values of $\overline{\omega_K}$ and $\overline{\sigma_K}$

(b) Cross-validation errors $\overline{\omega_K}$ versus cross-validation std. dev. $\overline{\sigma_K}$

Figure 4.8: Cross-validation errors $\overline{\omega_K}$ and cross-validation of the predicted standard deviations $\overline{\sigma_K}$ of the GPR for the noon report dataset from Torm Marie



(a) Histogram and normal distribution of the relative prediction errors from GPR using the noon report dataset from Torm Marie (*IVC* 15)



(b) Relative prediction errors vs. relative standard deviation of GPR using the noon report dataset from Torm Marie with (*IVC* 15)

4.3.3 Comparison of ANN and GPR for the *NRTM* dataset

The comparisons between the ANN and GPR shown in Figure 4.9 indicate that the GPR method generally performs better than the ANN for the noon report data from Torm Marie.

The prediction error was reduced by approximately 2% with *GPR* for *IVC* that included the *time* (7,9,10,13), which was not captured as well by the ANN method.

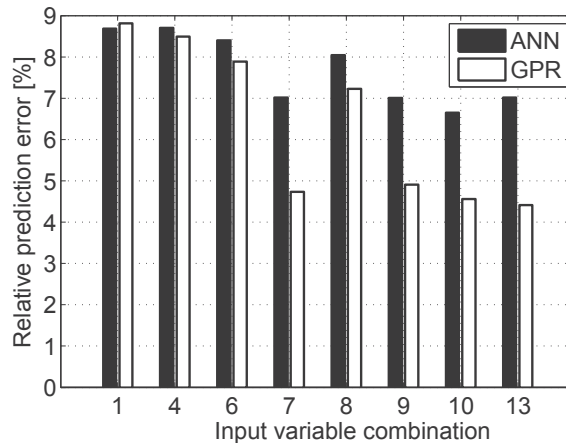


Figure 4.9: Comparison of the relative prediction errors between *ANN* and *GPR* methods of the noon report data from Torm Marie for different *IVC*. *IVC* 7,9,10 and 13 include the *time* as a variable.

4.4 *ANN* and *GPR* applied to long-term anonymous noon report data

The anonymous noon reports from containerhips 1-5 (*NRCS* 1-5) were very interesting for our analysis as they contained consistent data for a very long period and due to the anonymity the data provided an interesting case as it was not possible to apply classical empirical analysis but only allow data-driven analysis as the knowledge of the individual vessels resistance data was not available.

4.4 ANN and GPR applied to long-term anonymous noon report data 95

Initially only the data from containership 1 was trained for the period from the launching to the first dry-docking in order to compare the results from the ANN and GPR regression models. As in the previous section different combinations of the input variables (IVC) as in Table 4.10 were trained and tested to find the best setup.

ANN was only trained with 5 cross-validation training/test whereas the GPR was primarily trained by the Leave One Out (LOO), but in order for direct comparison the training was performed with 20/80 cross-validations.

As discussed in Chapter 3 the introduction of hindcast data limits the quantity of data due to the limited geographical coverage from the NOAAH data. The pros and cons of datasets with and without hindcast information are evaluated.

X	IVC	1	2	3	4	7	8	9	10	11	14	15	17	18	20	21
1	NR.Ulog	x	x	x	x	x	x	x		x	x	x	x	x	x	x
2	NR.Uobs								x		x	x			x	x
3	NR.Tsw	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4	NR.Tm	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
5	NR.Trim	x	x	x	x		x			x	x	x	x	x	x	x
6	NR.True_wind_speed			x						x	x	x	x	x	x	x
7	NR.True_rel_wind_dir			x						x	x	x	x	x	x	x
8	HC.Vrel	x	x		x	x			x						x	x
9	HC.gammarel	x	x		x	x			x						x	x
10	HC.mean.Hs										x	x			x	x
11	HC.mean.Tp										x	x			x	x
12	HC.var.Hs										x	x			x	x
13	HC.var.Tp										x	x			x	x
14	HC.var.Td										x	x			x	x
15	HC.var.Ws				x						x	x			x	x
16	HC.var.gamma				x						x	x			x	x
17	NR.UTC		x	x	x						x			x		x
18	HC.mean.Ws										x	x				
19	HC.gamma										x	x				
27	NR.Sea_state												x	x	x	x
28	NR.True_rel_sea_dir												x	x	x	x

Table 4.10: Input variable setup combinations for Noon Report data from containerships 1-5

4.4.1 Optimization of *ANN* regression for the noon report dataset from the containerships

The containerships were not studied as thoroughly as Torm Marie with *ANN* because it became evident at an early stage that the Gaussian Process Regression (*GPR*) was superior in speed and had a comparable prediction accuracy. Furthermore *GPR* had the advantage of the extra features *length-scale* and *prediction variance*. Consequently only containership 1 was trained with *ANN*, for the period from the launching until the first docking and only including datasets with hindcasts.

The cross-validation errors presented in Table 4.11 show that the lowest error for *ANN* is found by *IVC 2* with five hidden units. The error distribution of this solution is shown in Figure 4.10 and has a fair fit to the normal distribution. *IVC 2* include the time as a variable as seen in the previous section.

<i>IVC</i>	Number of hidden units			
	5	10	15	20
	%	%	%	%
1	6.07	6.82	6.99	7.04
2	5.50	7.12	6.16	8.38
3	5.87	6.97	6.57	7.54
4	6.57	7.24	9.00	12.35
7	6.32	6.66	6.33	7.02
8	6.28	6.57	6.47	6.53
9	5.77	6.96	6.14	5.98
10	5.87	6.13	6.78	6.34
11	6.21	6.73	6.63	8.47

Table 4.11: Relative cross-validation errors from *ANN* for the containership noon report data, $\overline{\omega_K}$ trained with *20/80* cross-validation.

4.4.2 Comparison of *ANN* and *GPR* using the containership data

For comparison between the *ANN* and *GPR* containership 1 was trained with cross-validation *20/80*. This dataset included hindcasts for the period from launching until

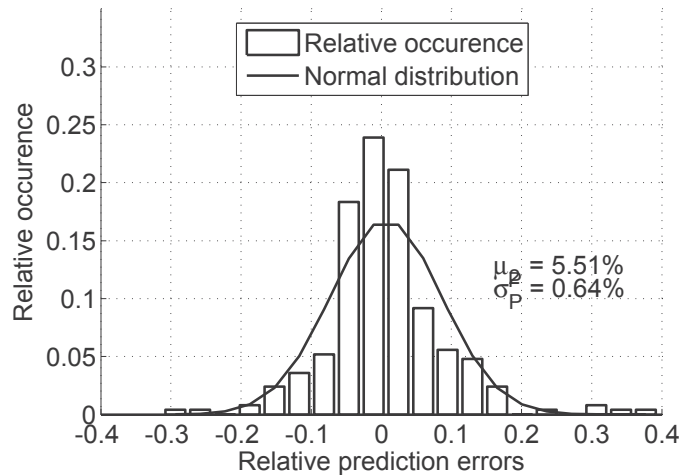


Figure 4.10: Histogram and normal distribution of the relative prediction errors from ANN using the noon report dataset from containership 1 and *IVC 2* (Described in Table 4.10) and five hidden units.

the first dry-docking, approximately four years after, resulting in 251 data points.

Figure 4.11 show that the cross-validation prediction errors are comparable, but for different *IVC ANN* and *GPR* differ from one another. *ANN* seems to be slightly better for *IVC 1,2,3* and 11 and *GPR* performs slightly better for *IVC 7-9*. *IVC 10* results in a very similar solution, but for *IVC 4*, $\overline{\omega_K}$ for *GPR* is approximately 2% higher than for *ANN*. The only difference between *IVC 4* compared to *IVC 2* is that the first includes the variance of the hindcast wind speed and direction. It is surprising to see that adding these two variables that are expected to be of little importance can result in such a change. This confirms that the inclusion of the variation in the hindcast data during the report period is relevant for the model.

4.4.3 Optimization of *GPR* regression method for noon report dataset from containership 1

In the previous Section the training with *GPR* of the *NRCS 1* was performed using the 20/80 cross-validation in order to make direct comparison with the *ANN* analysis.

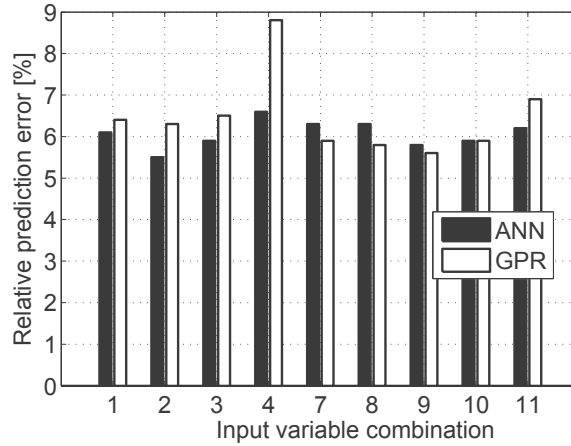


Figure 4.11: Comparison of the relative cross-validation prediction errors ($\overline{\omega_K}$) between *ANN* and *GPR* methods of the noon report data from containership 1 trained with 20/80 cross-validation scheme

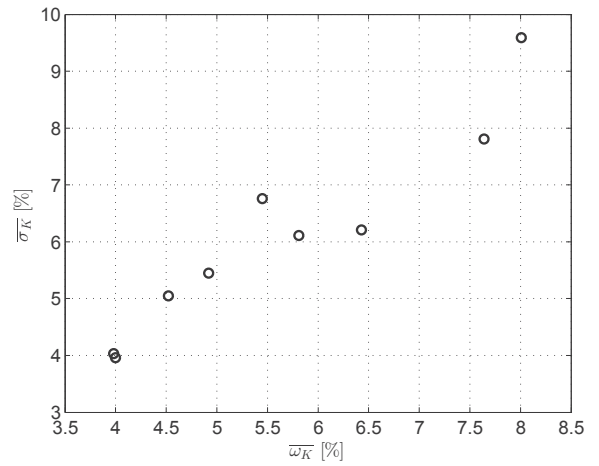
GPR is relatively faster than *ANN* for smaller training sets as the ones analysed in the thesis. This allowed *LOO* training for *GPR* analyses with different input variable combinations *IVC*.

Table 4.12a and Figure 4.12b shows the relative cross-validation errors ($\overline{\omega_K}$) and standard deviations ($\overline{\sigma_K}$) trained for the same period as in Section 4.4.2, but trained with *LOO*. The best *IVC* is 21 that includes all the input variables. Figure 4.12b shows a significant correlation between the $\overline{\omega_K}$ and $\overline{\sigma_K}$ identified by the analysis of *NRTM* data in Section 4.3.2

In Figure 4.13a, the histogram of the relative prediction errors (ω_k) of *IVC* 21 shows a nice fit to the normal distribution which indicates reliable predictions for this input combination and confirms the assumption of the *GPR* that the error is normal distributed.

Figure 4.13b illustrates the predicted standard deviations (σ_k) as a function of ω_k . This shows a similar pattern As the *NRTM* data with the highest ω_k having a high σ_k , but there is less scatter for *NRCS* 1.

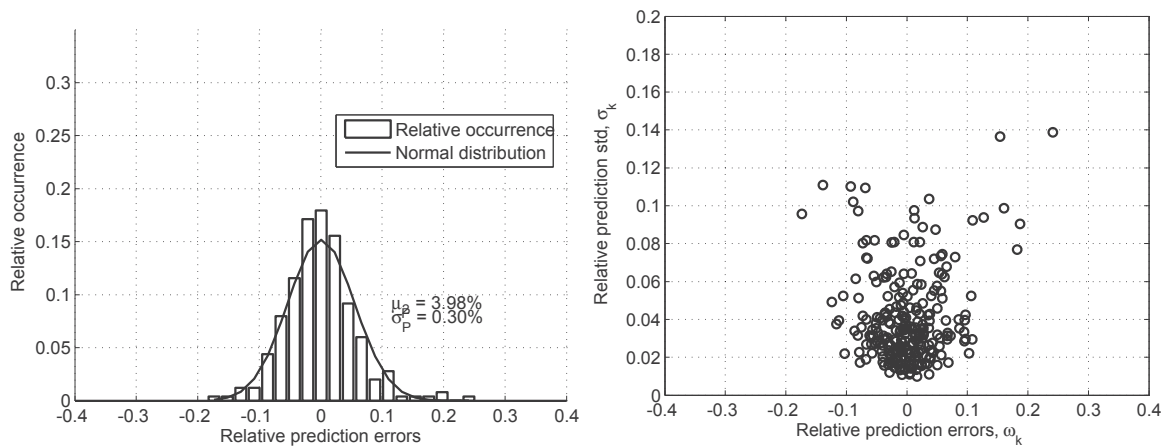
<i>IVC</i>	$\overline{\omega_K}$ %	$\overline{\sigma_K}$ %
3	5.81	6.11
9	5.45	6.76
11	8.01	9.59
14	4.00	3.96
15	6.43	6.21
17	7.64	7.81
18	4.92	5.45
20	4.52	5.05
21	3.98	4.03



(a) Tabular values of $\overline{\omega_K}$ and $\overline{\sigma_K}$

(b) Cross-validation errors versus cross-validation std. dev.

Figure 4.12: Relative cross-validation errors $\overline{\omega_K}$ and predictive standard deviation $\overline{\sigma_K}$ of the *GPR* for the noon report dataset from containership 1, using *LOO* training/test



(a) Histogram and normal distribution of the relative prediction errors from *GPR* using the noon report dataset from containership 1 *IVC* 21)

(b) Relative prediction errors vs. relative standard deviation of *GPR* using the noon report dataset from containership 1 with *IVC* 21

Figure 4.13: *GPR* prediction results for containership 1 using *IVC* 21

4.5 Analysis of long-term data using *GPR*

Since the GPR method's prediction performance was comparable to or better than the results from ANN, as was demonstrated in the previous Section 4.4 it was decided to work exclusively with *GPR*.

Initially, containership 1 was used to evaluate the input variable combinations (*IVC*), trained for the period between the launching and the first dry-docking. This period was approximately 4.5 years and a change in the hull propeller/performance was expected.

The first evaluation of the input dataset with and without the hindcast data was made using containership 1 from launching until the first dry-docking. This resulted in 251 and 515 data points for the hindcast and pure noon reports input respectively. Direct evaluation of the amount of data can only be performed for input variable combinations including the noon report data alone, e.g. input variables 3, 9, 11, 17, 18. A similar analysis was made of the *length-scale* in the variable analysis in Section 4.7.3.

The prediction errors are presented in Figure 4.14, and the definitions of input variable setups are given in Table 4.10. Apart from *IVC 11*, the *IVC* that includes hindcast data has smaller prediction errors than those without. For *IVC 11*, the prediction errors using the hindcast input data are slightly higher than the ones using the pure noon report data, and the general trend is that the use of hindcast data significantly improves the prediction despite limiting the amount of data.

Based on the results from Figures 4.11 and 4.14 the focus is kept on input variable setups 9, 18, 20, 21, for representing the following input variable setup and dataset:

9 has the best performance of the *IVC* without *time* input for the exclusive noon report data.

18 has the best performance for the *IVC* including *time* input for the exclusive noon report data.

20 has the best performance for the *IVC* without the *time* input for the noon report data and hindcast dataset.

21 has a similar performance to input variables *13* and *14* for the *IVC* including the *time* input, the noon report data and hindcast dataset, but *21* also includes the reported sea states and can thus be regarded as more robust.

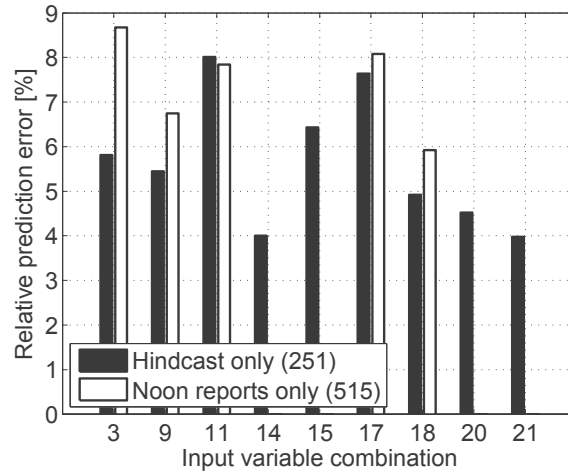


Figure 4.14: Comparison of the relative prediction errors found by *GPR* for datasets of containership 1, including hindcast (251) and without (515).

4.5.1 Evaluation of the effect of the data length

Training and testing of containerships 2-5 was performed in a similar manner to the training and testing of containership 1. The four different *IVC* 9, 18, 20, 21 were used together with datasets including both noon reports and hindcasts. In order to evaluate the influence of the length and period of the datasets, each dataset was trained for different periods of time. All the periods started from the launch and accumulated in one-year increments until the first docking (year 0-1, year 0-2, year 0-3, year 0-4, year 0-first dry docking).

The relative prediction errors are shown in Figures 4.16a- 4.16e and do not indicate dependence on the quantity of data. For containerships 1, 2 and 3, there is a general trend of decreasing errors with time for the average input setup variables, yet the best input variable combination, *21*, is increasing slightly with increasing number of data points in Figure 4.16a. The amount of data varies between the containerships

from 44-71 data points for the one-year period and 234-297 data points for the longest period from the launch to the first dry-docking. This does not seem important as the lowest error for (21) is 3.7% with 234 data points, and the highest error (6.9%) for 21 is found with 297 data points available.

Across the containerships, the lowest prediction errors are generally found by *IVC 21*. Although the errors from input variable setup 20 is in many cases very good or slightly better, it is not as consistent as *IVC 21*.

The predictive standard deviations describe the uncertainty of the predictions. Around well defined areas with high number of consistent input data the predictive standard deviation will be small. The further you get from dense regions of consistent input data the higher the predictive standard deviation is. Assuming that the training set of data is indeed representative of the test dataset then one can expect that the error in predictions for the test set will correlate well with the predictive standard deviation. The ability to make a direct comparison between the GPR methods predicted standard deviation and the observed standard deviation is a unique feature of GPR.

The cross-validated predictive standard deviations ($\overline{\sigma_K}$) for the accumulated training periods are shown as a function of the corresponding cross-validated prediction errors ($\overline{\omega_K}$) in Figure 4.15. Here the containerships are not distinguishable, but only the *IVC* and the training period is marked differently, where the marks "1-5" indicated the training period 0-1 year, 0-2 year and so forth. This plot shows a general correlation between $\overline{\omega_K}$ and $\overline{\sigma_K}$, in particular for the region where $\overline{\omega_K}$ and $\overline{\sigma_K}$ are below 6%. For a group of trainings of *IVC 9* shows slightly higher $\overline{\sigma_K}-\overline{\omega_K}$ ratio than the remaining *IVC*, this might be due to the relatively simple input combination that only includes the ship speed, the mean draught and the sea water temperature. It is difficult to see if and how the results depend on the training period length from this. Looking at $\overline{\omega_K}$ exceeding 8% all training periods are represented.

Generally it can be concluded that when using the right *IVC*, the relative prediction error depends on the quality of the input data. For the best input dataset, the relative prediction error is a consistent 4% independent of the data length (containerships 1 and 2 *IVC 21*). It is confirmed that the input from hindcast data improves the prediction (*IVC 20* and *21*), and the *time* variable (*IVC 18* and *21*) also improves the prediction

for IVC 21.

The relation between the prediction errors and prediction standard deviations are consistent for all the datasets and *IVC* and is thus very useful in the prediction.

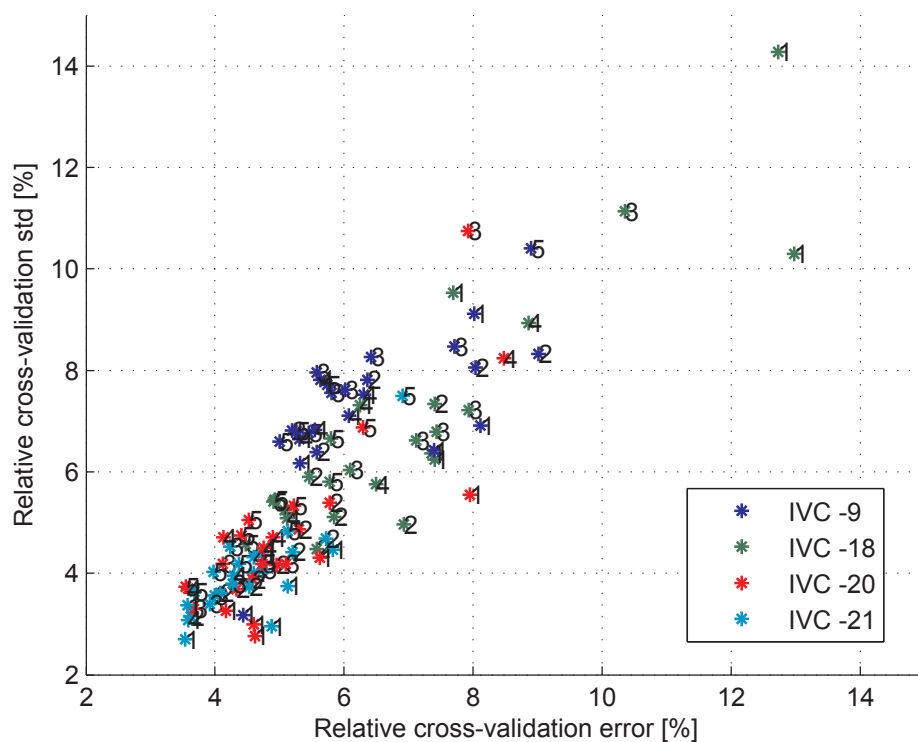


Figure 4.15: Summary of the five containerships together. Trained accumulative for $0-1$, $0-2$,... $0-5$ years, which is indicated by numbered markers. The coloured asterisk indicate the *IVC*.

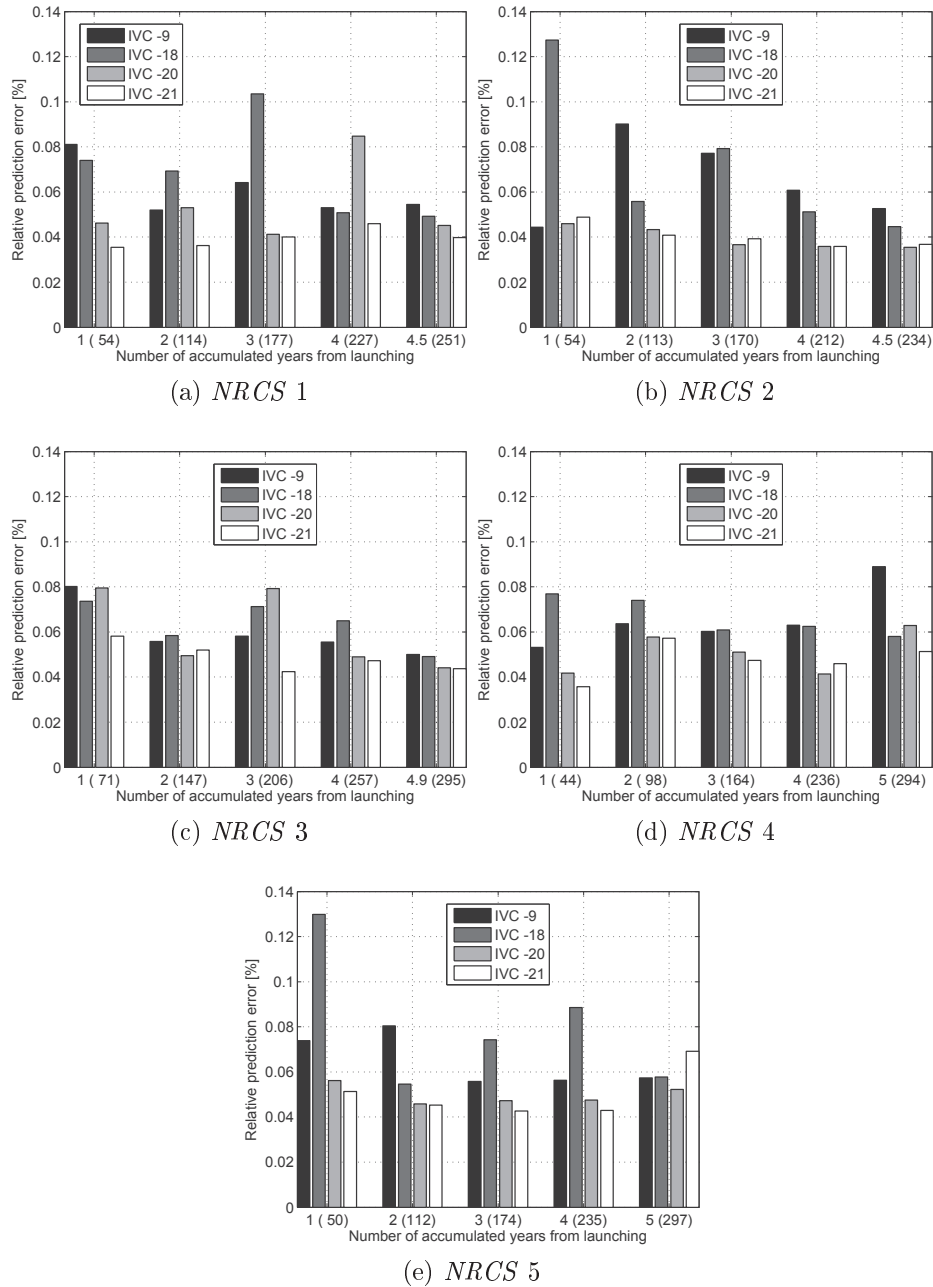


Figure 4.16: Relative prediction errors from containerships 1-5 (*NRCS* 1-5) trained with *IVC* 9,18,20,21 for five different periods of time starting at the launch with one-year intervals accumulating until the first dry-docking

4.6 Discussion of the results of the prediction methods

The previous chapter demonstrated how the non-parametric data-driven regression methods, Artificial Neural Network, *ANN*, and Gaussian Process Regression, *GPR*, improve the prediction performance significantly compared with the custom non-linear and linear regressions. For most combinations of datasets and *IVC*, *GPR* had slightly smaller prediction errors than *ANN*. *GPR* was chosen for further analysis because of its ability to find the predictive variance and for its computational efficiency with dataset sizes used (max ≈ 500 data points). Furthermore, the option of using *length-scale* for evaluation of the input parameter's relevance, is an interesting feature which is not readily available in *ANN* as discussed in Section 4.7.

Four *IVC* have been selected, *9*, *18*, *20* and *21*. Where *20* and *9* are the best *IVC* for datasets with and without hindcast weather information, respectively. *IVC 20* and *21* are similar to *9* and *18*, but they include *time* as an additional variable. Table 4.12 gives a schematic overview of the *IVC*.

The *time* as an input should not be used for predictions beyond the time frame of the input, because the *time* has a constant linear slope and thus will act as an extrapolation out of the region from where it has been trained. It has mainly been used to demonstrate its strong effect on the prediction errors, and it is thus representative of the long-term deterioration of the ship and is assumed to represent the propeller and hull fouling.

As also stated in Section 4.7 when using hindcast input data to replace or support the noon report data, the prediction errors are reduced by approximately 1% even though the hindcast data limits the amount of data by up to 50%.

There was no general trend between the prediction errors and length of the dataset as illustrated in Figures 4.16a- 4.16e, and for most of the datasets for the containerships (1-5) the shortest periods have similar prediction errors as the longest. The relative predictive standard deviations in Figures 4.15 show a similar pattern for different periods of training/test data.

For all the datasets, the average of the relative predictive standard deviation, $\overline{\sigma_K}$ is of the same magnitude as the cross-validated relative prediction errors $\overline{\omega_K}$. This gives

X	Input combination ID	9	18	20	21
1	NR Logged speed	x	x	x	x
2	NR Observed speed			x	x
3	NR Seawater temperature	x	x	x	x
4	NR Draught midship	x	x	x	x
5	NR Trim		x	x	x
6	NR True wind speed		x	x	x
7	NR True relative wind direction		x	x	x
8	HC mean relative wind speed			x	x
9	HC mean relative wind direction			x	x
10	HC mean significant wave height			x	x
11	HC mean peak wave frequency			x	x
12	HC variance of significant wave height			x	x
13	HC variance of peak wave frequency			x	x
14	HC variance of wave direction			x	x
15	HC variance of the true wind speed			x	x
16	HC variance of the wind speed			x	x
17	NR UTC <i>time</i> variable		x		x
27	NR Observed sea state		x	x	x
28	NR Observed true relative wave direction		x	x	x

Table 4.12: The best input variable combinations for Noon Report data from the containerships

a picture of the average predictive confidence interval, CI , for each of the predicted values, but does not indicate how the prediction errors are distributed.

By looking at the estimated normal distributions of the relative prediction errors, the estimated variance can be used to determine a confidence interval, CI . A 95% CI , which is approximately $\mu \pm 2 \cdot \sigma$ (Larsen (2000)) was calculated for the best setups for ANN and GPR in each of the evaluated datasets.

Table 4.13 shows that the CI for the measured datasets predicted with ANN and GPR are similar, and are in the range of $\pm 0.2 - \pm 0.57$ or $\frac{1}{3} - \frac{1}{2}$ relative to the cross-validation error μ_ω . Prediction of the measured datasets using GPR are thus able, with 95% certainty, to predict the energy consumption within 0.76 – 1.5%.

For the Torm Marie noon report data, there is a reduction of 0.5% of the 95% CI (and μ_ω) using GPR compared with ANN . Prediction with GPR makes it possible to

make a prediction within 6.15%. The lowest cross-validation error and 95% *CI* is also found by *GPR* for containership 1 and results in a prediction probability of 5.08%

Vessel	Var.	<i>IVC</i>	μ_ω %	σ_ω^2 %	σ_ω %	95% <i>CI</i> %	$\max(95\% \text{CI} + \mu_\omega)$ %
Torm Marie 1	12	ANN	1.66	0.08	0.28	± 0.57	2.23
Torm Marie 2	12	ANN	0.83	0.04	0.20	± 0.40	1.23
Torm Marie 3	12	ANN	1.45	0.05	0.22	± 0.45	1.90
Torm Marie 4	12	ANN	0.80	0.02	0.14	± 0.28	1.08
Torm Marie 1	15	GPR	1.00	0.03	0.17	± 0.35	1.35
Torm Marie 2	15	GPR	0.78	0.03	0.17	± 0.35	1.13
Torm Marie 3	15	GPR	1.10	0.04	0.20	± 0.40	1.50
Torm Marie 4	15	GPR	0.56	0.01	0.10	± 0.20	0.76
Torm Marie NR	7	ANN	7.21	1.28	1.13	± 2.26	9.47
Torm Marie NR	15	GPR	4.41	0.76	0.87	± 1.74	6.15
Containership 1 NR	2	ANN	5.51	0.64	0.80	± 1.60	7.11
Containership 1 NR	21	GPR	3.98	0.30	0.55	± 1.10	5.08

Table 4.13: A summary of the cross-validated relative prediction error, variance, standard deviation and 95% confidence interval, based on the estimated normal distributions.

4.7 Evaluation of input variables using *ANN* and *GPR*

In this section it is shown how Artificial Neural Networks (*ANN*) and Gaussian Process Regression (*GPR*) can be used to evaluate the influence of separate input variables.

The results from the regression models could be used to quantify the influence of each of the specific input variables. *GPR* was a unique and powerful tool for this because of the characteristic *length-scales* found during the optimisation routine: *Automatic Relevance Determination (ARD)* as described in Chapter 2. *ANN* does not give information about how the input variables influence the output and it was only possible to make more qualitative investigations where the results from different input variable combinations *IVC* were evaluated to give estimates of how the results of a certain *IVC* would be affected by including different variables, for example the trim.

Three different sets of input variable combinations were used to define the three datasets. Table 3.12 presents the input variables for the measured dataset from Torm Marie (*MDTM*), Table 3.13 presents the input variables for the noon report dataset from Torm Marie (*NRTM*), and Table 3.14 presents the input variables for the noon report dataset from the five containerships (*NRCS* 1-5).

In *Gaussian Process Regression GPR*, the hyperparameters from the solution consist of one *length-scale* (l) for each of the input variables, the standard deviation of the process, σ_f and the standard deviation of the learned noise σ_n . Short *length-scale*'s indicate that input x_i is relevant for the solution, and a longer *length-scale* l_i indicates a less relevant or noisy input x_i as described in Section 2.2.4. The *GPR* was trained with Leave-One-Out (*LOO*) (described in Chapter 2), and the prediction errors and the *hyperparameters* presented are average values of the N test results. A detailed description can be found in Chapter 2.

4.7.1 Evaluating the input variables from the measured dataset from Torm Marie

The data was sampled over a short period of time, thus variations of time and draught/trim are not included among the input variables. It is a good example for investigating the difference between the measured data, noon report and hindcasted data. Histograms of all input/output variables are available in Appendix C.1.2.

Evaluating the input variables by *ANN* with the measured datasets from Torm Marie

The selected combinations of input variables are shown in Table 4.4. The prediction errors are presented in Figure 4.17 and show that datasets 2 and 4 have similar errors, because these datasets were mostly stationary, mainly due to good weather. Datasets 1 and 3 showed more variation making them more appropriate for method evaluation. In general the prediction errors are small for *IVC* 8, 10 and 12, with approximately 1% lower error than *IVC* 2 and 11, which indicates that the hindcast wave properties are important variables as they are not included in *IVC* 2 and 11.

The wave properties (wave height, period and direction) are the only differences between input variable combinations *11* and *12*, and the prediction errors indicate that introducing the wave properties improved the model.

The only difference between *10* and *12* is that the latter includes the relative wind opposed to the absolute wind as in *10*. This indicates that the relative wind direction has little influence.

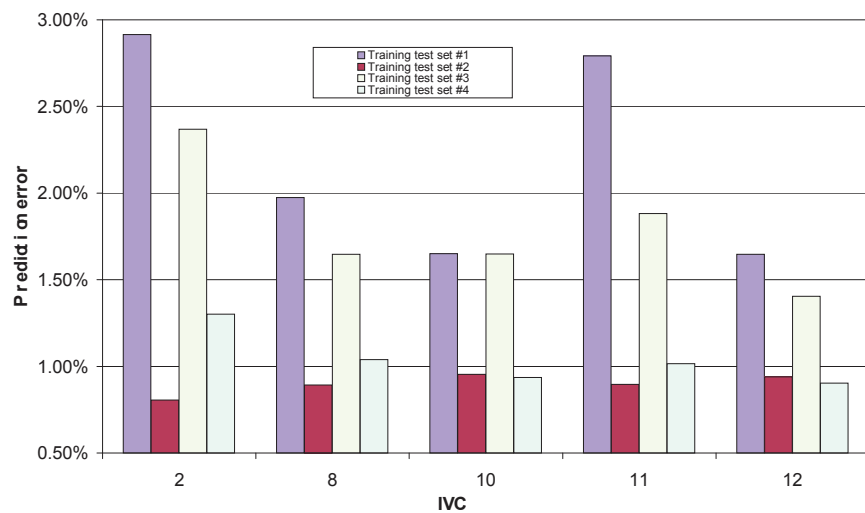


Figure 4.17: Mean errors for the measure datasets (*MDTM*) analysed with *ANN* for different *IVC*

Evaluating the input variables of the measured datasets using *GPR*

Gaussian Process Regression was performed with the measured data using a *LOO* training/test set. The input dataset included all of the available input variables, i.e. *IVC* 15 in Table 4.4.

The *length-scales* are plotted in Figure 4.18 and show little agreements between the datasets. The ship speed *1*, relative measured wind direction *2* and the wave direction *11* all have *length-scales* in the same range.

The sea water temperature *6* is generally very low, but particularly low for dataset 2 and 4 which is almost a calm water condition.

Datasets 1 and 3 which show most variance in wind and waves as described in the previous section have noticeably low *length-scales* for input 9 which is the significant wave height and indicates that the presence of small waves of up to 2m according to Figure C.10a and C.10a in Appendix C.1.2 has a considerable influence on the solution. Even though so small waves are not expected to induce significant resistance (less than %2 Eefsen (1996)) it is in these cases enough in order to be relevant for the solution.

Some input variables represent the same element, e.g. the wind is represented by 2 *Vrel*, 7 *HC.Ws* and 12 *HC.Vrel*, the wind direction by 3 *grel*, 8 *HC.g* and 13 *HC.grel*, and the air temperature by 4 *Tair* and 5 *NR.Tsw*. This is apparent in dataset 1 where the average *length-scales* of the input variables 2,7 and 12 is 2.1, and for dataset 3 it is 2.0, i.e. the wind speed is equally important for the two datasets.

Dataset 2 shows the greatest variance in *length-scales*, with many variables relatively high and some very low such as the measured relative wind speed 2, the measured air temperature 4, the sea water temperature 6 and relative wind direction 13. This can be explained by the input/output variables having little variance and are thus easy to predict without many variables.

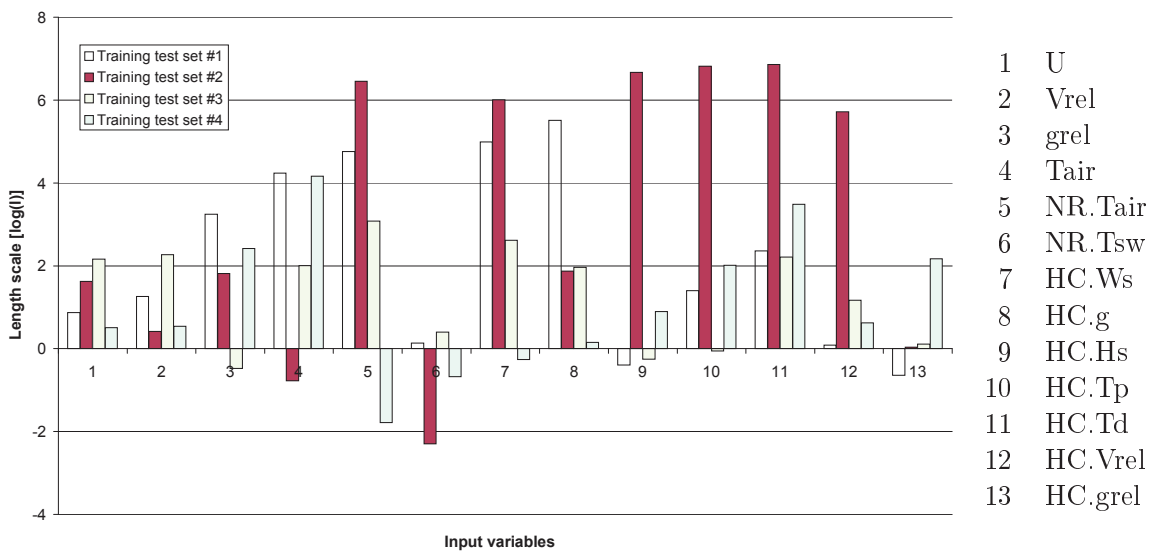


Figure 4.18: Logarithmic *length-scales* of GPR evaluation of the four measured datasets

Discussion on variable analysis of the measured datasets

The input variable analysis of the measured dataset confirms that the wave conditions have a significant influence on the results except in completely calm conditions (input datasets 2 and 4). The ship and wind speed were important factors as was expected based on the propulsion theory, whereas the sea water temperature was found to be the most significant variable. This might be explained by the good weather and relatively steady speed that govern all the measured datasets.

4.7.2 Evaluating the input variables from the noon report data from Torm Marie

The noon report data from Torm Marie (*NRTM*) was collected over a two-year period and under several different loading conditions which introduced the variables' mean draught and trim. Time was also introduced as a variable in numeric format as in *MatLab*®[®], which is an accumulated value of 1 per day starting at year 0. Table 4.8 shows a schematic overview of the input variable combinations.

Evaluating input variables of *NRTM* data using *ANN*

As in Section 4.7.1, the ANN was trained for different selected *IVC* in order to evaluate the influence on the prediction error. All input combinations included ship speed, air temperature, sea water temperature, mean draught and trim, since these variables were assumed to be fundamental to the propulsion performance.

Figure 4.19 shows that the lowest prediction errors are found with *IVC* 7, 9 and 10. The main difference between these inputs (except *IVC* 13) is the inclusion of time (as an input). This can be illustrated by comparing *IVC* 6 and 7 where the only difference is the time input. Here the prediction error is reduced by 1% or relatively 12%. The same phenomenon occurred between 8 and 13, with a reduction of 0.7% and 8% relatively.

The wave properties do not seem to have the same influence on the noon report data as on the measured data. They actually have a negative effect as seen when comparing *IVC* 1 and 4, 6 and 8, 7 and 13.

The relative wind properties, on the other hand, had a significant positive effect which is seen by comparing *IVC 1* and *6*, *4* and *8*. Substituting the hindcast wind properties with the reported wind properties, input variables *19* and *20* as in *IVC 10*, only results with a slightly negative effect compared with *IVC 7* which is the best combination.

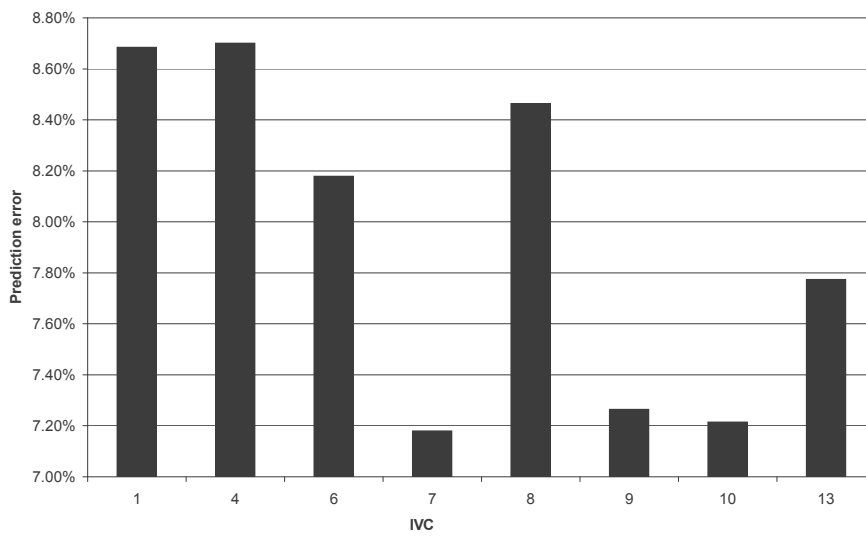


Figure 4.19: Mean error for different *IVC* using *ANN* with NR data from Torm Marie

Evaluating the input variables *NRTM* data using *GPR*

The *length-scales* of the *GPR* hyperparameters from *IVC 15* and *16* are presented in Figure 4.20, where it is confirmed that the ship speed *1*, mean draught *4* and trim *5* were relevant to the solution, due to the short length scales, although the draught *4* show a higher length-scale for *IVC* including time (*15*). The air and sea water temperature had a weaker but consistent relevance. The wave height *6* was moderately relevant which was unexpected since it showed a poorer performance with the *ANN* method. The variation in sea direction *11* is consistently low together with the hindcasted wind speed and direction (*12* and *13*). For *IVC 15* the time input *16* is the governing variable which indicates a strong linear trend in the *EC* over time.

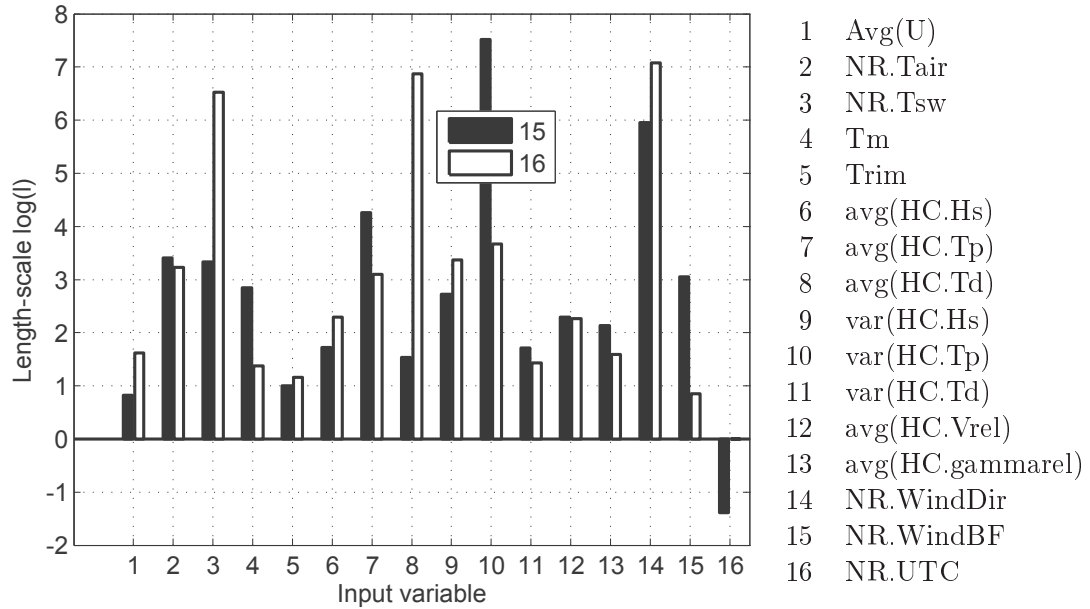


Figure 4.20: Logarithmic *length-scales* of *IVC* 15 and 16, of the noon report data from Torm Marie.

From the variable analysis of the Torm Marie noon report data using both *ANN* and *GPR*, it can be concluded that the most significant input variables are: ship speed, mean draught and trim, wind direction and wind speed, and that the most important variable might be time.

4.7.3 Evaluation of the input variables of the containership noon report dataset

As previously described, training was performed with different combinations of input variables for containership 1 (*NRCS* 1) listed in Table 4.10. Since the datasets span almost 10 years, including events such as docking, hull cleanings and propeller polishing, the main focus of the analysis was the period up to the first docking (29-2-2004). The events within this period were assumed to have only minor effects.

Evaluating the input variables of *NRCS 1* using *ANN*

In Figure 4.21, the relevance of the time input *17* is again confirmed by comparing *IVC 1* and *2*, and *3* and *11*. In both cases the prediction error was reduced noticeably by about 0.4%. Furthermore, it is noted that the second best combination (*9*) has the simplest *IVC*, only the ship speed, mean draught and sea water temperature are included. It is superior to the combination that included the trim (*8*). In general, the variation in prediction errors between the different input combinations is not very large, for the most part less than 1%, and the results will be evaluated together with *GPR* analysis of the same data.

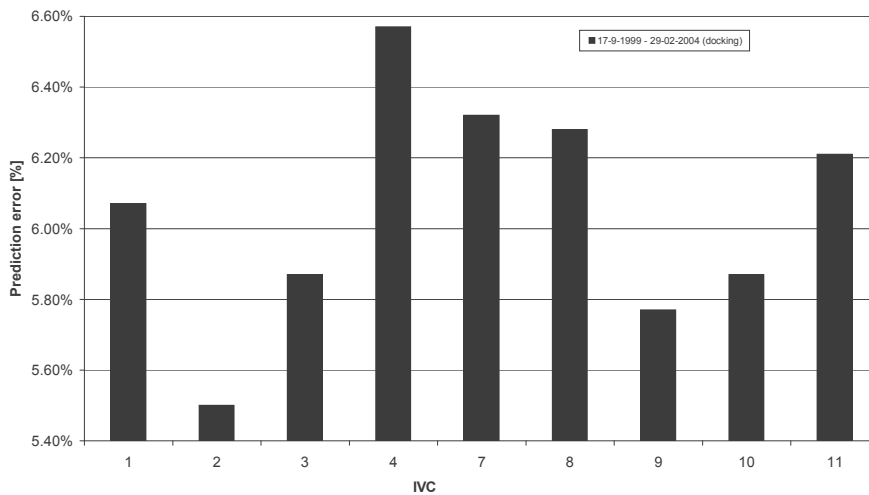


Figure 4.21: Mean error for different *IVC*, given in Table 4.10 from training with NR data from containership 1

Evaluating the input variables of *NRCS 1-5* using *GPR*

IVC 21 and *20* include most of the significant input variables with and without the time (*NR.UTC*), which gives a good picture of the influence of each variable. The hindcast data are included here which explains the high number of variables but lower number of training/test data. The training was performed on containership 1 for a five-year period until the first docking.

Although there are some discrepancies between the two *IVC* in Figure 4.22, there are also common tendencies. The logged speed (1), the GPS speed (2) and the mean draught (4) are relatively short compared to the other variables for both *IVC* and confirm the theory that the speed and draught are the governing input variables. The relative wind speed (8) and direction (9) and significant wave height (10) from the hindcast data also have a consistent relevance. There are some discrepancies between the two input variable sets which are due to the important role of the time input 17. The training is performed over a long period of time, the time input thus represents the state changes over time which is mainly due to fouling of the hull and propeller. The time *length-scale* (17) is relatively small which explains why the *length-scales* of many of the other variables are longer for *IVC 21*, than *20*. The least relevant input variables are input variables 13 – 16 which is no surprise since they are the variance of the hindcast wave and wind properties during the noon report period.

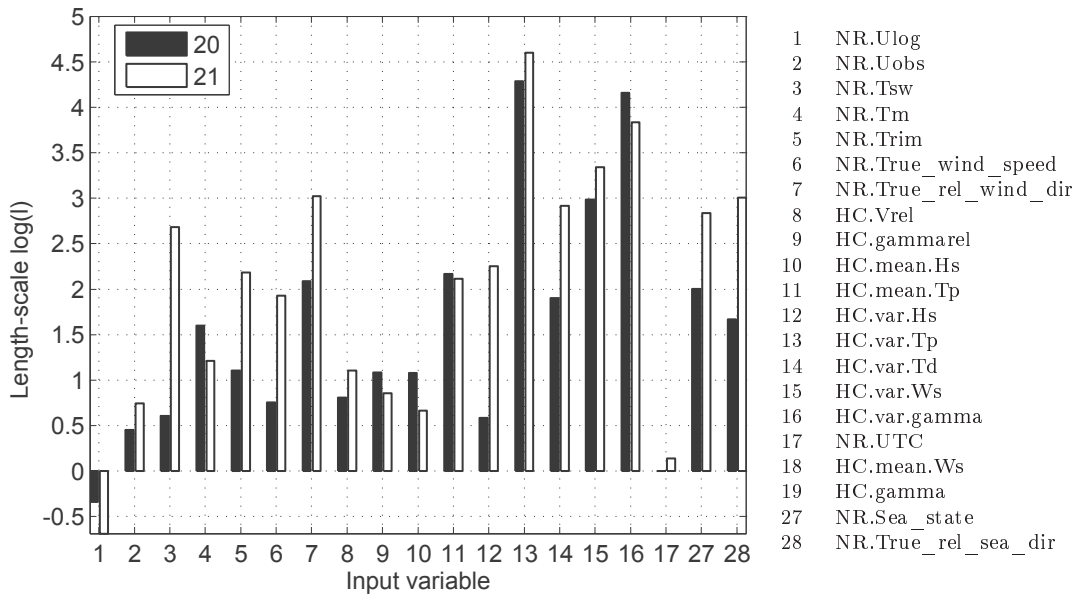


Figure 4.22: Logarithmic *length-scales* of input variable combinations 20 and 21, with and without time as a variable (251 input dataset) for containership 1 (Launching 1st dry-docking)

IVC 17 and 18 only include data from noon reports but have approximately twice

the amount of data (515) for the same period of time. The trained data originates from the same period of time (launch to first docking).

Figure 4.23 again shows that the speed (1) is a ruling variable, and the draught and trim are relevant input variables for the *IVC* not including time as a variable. The seawater temperature is also significant in this setup. In general all the *length-scales* are short and thus relevant to the solution. Note that input variables 1 and 3 for setup 17 and 28 for setup 18 are not noticeable because they are close to zero.

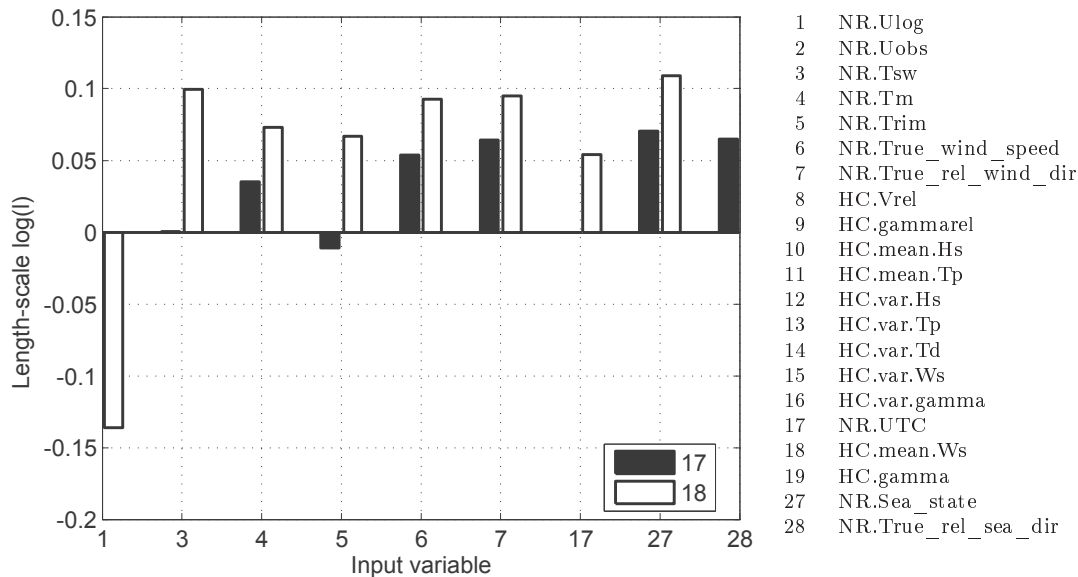


Figure 4.23: Logarithmic *length-scales* of *IVC* 17 and 18, with and without time as a variable. Using only noon report data, with and without time as a variable (515 input dataset) for containership 1 (Launching 1st dry-docking)

Figures 4.24 and 4.25 show the accumulated *length-scales* using three different periods of time all starting at the launch, at one year and two years, until the first docking at year ~ 4.5 after launching, for input setups 20 and 21 respectively.

Both *IVC* 20 and 21 show significant differences between the longest training period (0-4.5 years) and one and two shorter training periods (0-1 year, 0-2 year). This indicates that some input variables are more relevant to shorter periods of time, and that some first become important after a certain period of time.

The accumulated bars give an impression of the general relevance of each variable

which seems to have the same tendency as seen in Figure 4.23 with low *length-scales* for the ship speed(1), draught(4) and significant wave height (10). The time also had a significant relevance. Surprisingly the wind properties, observed and hindcast, both have higher *length-scales* for the one and two year periods, for input setup 20 (without time-input).

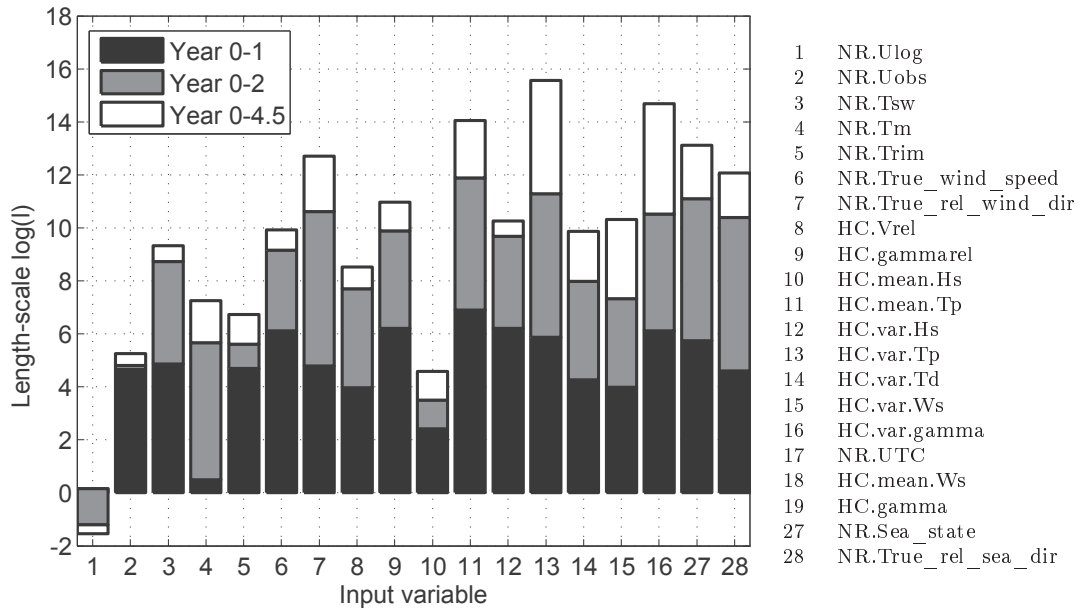


Figure 4.24: The accumulated *length-scales* of IVC 20 (without *time*), trained for three periods of time: 0-1, 0-2, 0-4.5 years (1^{st} dry-docking)

4.8 Discussion of variable analysis

The data driven evaluation of the input variables confirms many of the hypotheses stated in Section 3 about the governing input variables. Even for the datasets with small variations, such as the measured datasets, it is clear how the ship speed, relative wind speed and direction, and also the sea water temperature are the most important variables.

When looking at the noon report data, it can be concluded that the speed, draught and trim are the most relevant input variables for the model. For noon report data

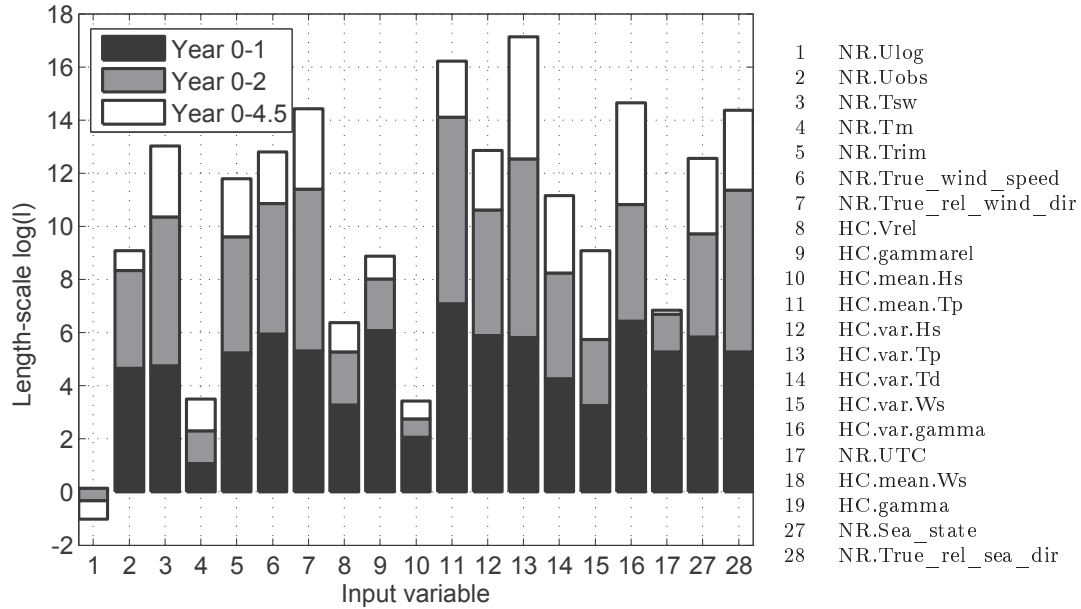


Figure 4.25: The accumulated *length-scales* of *IVC* 21 (including *time*), trained for three periods of time: 0-1, 0-2, 0-4.5 years (*1st* dry-docking)

from containership 1, the relative wind speed and direction from the hindcast data together with the significant wave height are important as well.

For all three data sources, the introduction of hindcast weather data improves the relative prediction errors by up to 1%.

The *time* as an input variable shows a strong relevance to all *IVC* and reduces the relative prediction error also by approximately 1%, which confirms the hypothesis that the *time* is representing the decrease in performance over time or the effect of fouling.

Chapter 5

Propulsion performance trend detection

The ability to estimate trends in the propulsion performance is the key motivation for working with regression methods. In this chapter we explore how the regression methods treated in previous chapters can be used to detect the performance trend.

The analysis in this chapter is only performed using the dataset from the five containerships (*NRCS* 1-5) due to their long data period and known events, which qualifies them for the long-term performance evaluations.

5.1 Performance trend detection method

In the analysis, the relative prediction error ω (Eq.1.1) was used to evaluate the behaviour of the performance. The actual energy consumption EC is expected to increase over time due to the fouling, while ω is expected to decrease since the predicted values are based on the training data which are not assumed to be affected by fouling. It is thus desirable to train on the shortest possible period of time in order to limit the effect of a trend in the training data. Yet the training set should include a reasonable variation of input variables.

As described in Chapter 2 the trend of ω is assumed to be linear and is referred to as the Vessel Performance Trend, VPT and is described in (1.2). The coefficients, α and

β are found by *Weighted Least Squared Regression*, *WLS* as introduced in Section 2.5.

VPT is only continuous in periods without external disturbances that change the propulsion performance. External disturbances can be known or unknown. The known disturbances are e.g. dry-docking and hull and propeller cleaning, and unknown disturbances could e.g. be unknown damage to the propeller or rudder.

All known disturbances or events are available for the containerships, and in Chapter 6, the detection of known events are used to attempt to simulate unknown disturbances.

In the following section, *VPT* was found in between the known events (VPT_i), including dry-docking *DD*, hull cleaning *HCL* and propeller cleaning *PCL* for the containership. All these events are expected to increase the relative prediction error ω , because the energy consumption is expected to drop. But for the hull and propeller cleaning, this effect can sometimes be difficult to detect due to its limited impact on the relatively large data scatter. In the Chapter 4 it was shown that there was limited but noticeable variations in the prediction errors due to training with different length datasets. It has thus been decided to perform the initial training for two periods: The first year and the first two years after the launching.

The *time* variable in input variable combinations (*IVC*) 18 and 21 increases linearly with time. This means that if *time* has a relevance for the regression model, it will be dominant for prediction far into the future. Input combinations including the *time* were thus inapplicable for the trend detection. As found in Section 4, the best *IVCs* without *time* were 9 and 20, with 9 based exclusively on noon report data, and 20 includes the hindcast data. Although omitting the use of hindcast reduced the cross-validation prediction errors for the training period it resulted in about twice the number of input for the predictions. Using data without hindcasts should thus be taken into consideration for the detection of trends due to the larger amounts of data.

Due to the considerations described above, \hat{EC} for the five containerships were predicted based on two training periods, one and two years from launching, two *IVC*, 9 and 20, including all noon report data and only noon report data including hindcasts, respectively. This results in *four* different variations of training and test, as illustrated in Table 5.1. The trends were detected only between all the known events and the

dry-dockings.

	NR	NR+ HC
IVC	9	20
One year initial training	x	x
Two year initial training	x	x

Table 5.1: VPT test variations for containerships 1-5

5.2 Performance trend detection results

In the following, trend detection results are presented for the selected combinations of ship, data length, input variable combinations etc. as described above. Results from all the analyses can be found in Appendix D.

The main results from the analysis is the slopes of the VPT within the intervals (α_i) between the events and the change between the final value of VPT ($VPT_i(max(t))$) before an event and first value ($VPT_{i+1}(min(t))$), ΔVPT_i after an event. ΔVPT_i quantifies the jump in the trend lines across an event, and hence illustrates the immediate effect of the event.

Trend detection for containership 1

Containership 1 has the most complicated pattern of the five containerships with many minor events such as hull and propeller cleanings and weak general trends and will be used as an example. Figure 5.1 gives the trend based on training on the first two years with IVC 20 which is among the best representations. The many events result in short trend regression periods with few data points that make realistic trend detection difficult, but for the longer periods e.g. between year \sim 2-3 and year \sim 4.5-7.5, the trend slopes show the expected tendencies. After the first hull and propeller cleaning, the prediction error increased indicating that these events had the intended effect: to increase the performance level. The trend slopes were positive after the first events, this was unexpected and might be due the relatively low number of input for trend for the short periods. Moreover, as expected there is a noticeable change in prediction

error and trend slope after the first dry-docking. A certain level of instant increase in performance due to the dry-docking is anticipated, because blasting and reapplying anti-fouling paint is expected to lower friction compared to the previous roughness of the hull.

Figure 5.2 shows the trends between the dry-dockings only, with the same prediction errors as seen in Figure 5.1. This shows two noticeably decreasing trend slopes, the last steeper than the first. This may seem surprising at first glance but it is well known that dry-dockings, blastings and paint jobs can be of varying quality and can sometimes increase the hull fouling rate (Townsin et al. (1980)).

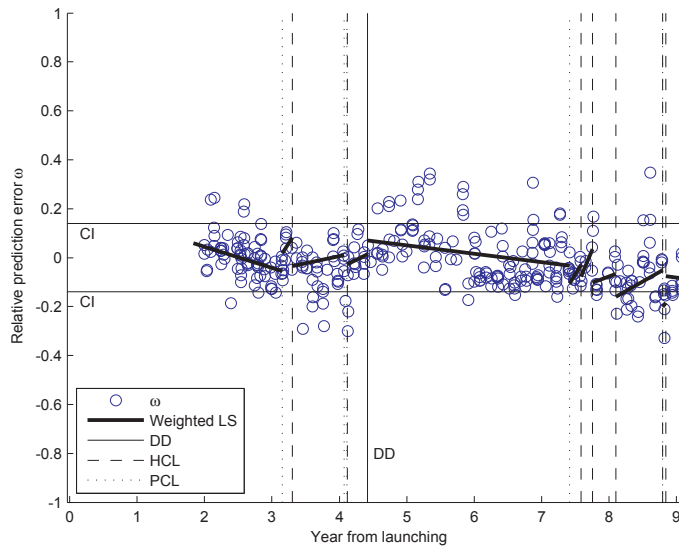


Figure 5.1: Performance trend between all events of containership 1 using *IVC 20* with data including hindcasts, trained with the first two-year period

5.2.1 Summary of trend detection results

Similar analyses were performed for all the containerships. In order to evaluate the results in a comprehensive way a summary of the *VPT* slopes α and the effect from events ΔVPT is given. A summary of the findings from containerships 1-5 is given in Table 5.2 for the *IVC 20* and in Table 5.3 for *IVC 9*.

Figure 5.3 shows the slopes of the *VPT* α which represents the rate of change

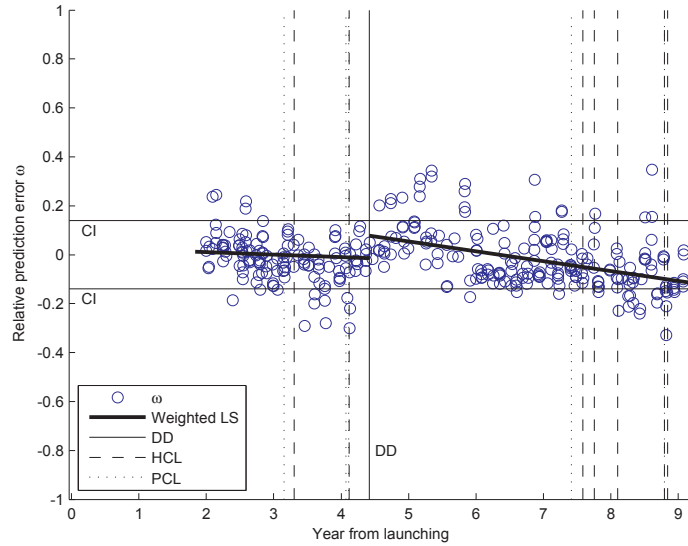


Figure 5.2: Performance trend between dry-dockings of containership 1 using *IVC 20* with data including hindcasts, trained with the first two-year period

in performance, which can also be interpreted as the fouling rate. The left figure shows *VPT* slopes for the first trend interval (α_1), from the launching to the first dry-docking and the right the *VPT* slopes for the second interval (α_2), from the first dry-docking until the second. For the first trend interval α_1 range from -6.6% to 0.5% in relative change in *EC* per year ($\omega/year$). There are significant differences between the four training variations (Table 5.1) used for the detection and no general tendency of what method is the best or the correlation between the methods are found. For the 2nd trend interval, Figure 5.3 right, the methods show smaller discrepancies and all methods show less steep slopes and range from -3.8%/year to 3.7%/year. The latter number originates from containership 4 and all the methods for this vessel show a consistent, although small, positive trend. α_2 for containership 2 is around zero and this indicates together with containership 4 that a better anti fouling paint was applied after the dry-docking that and that no change in roughness was detected as described in Townsin et al. (1980). For containership 4 a third trend interval was analysed because sufficient data was available after the second dry-docking, this reveals α_3 of -8.7% to -2.1% depending on the method used.

Figure 5.4 illustrates the effect of dry-docking in percent change in the relative

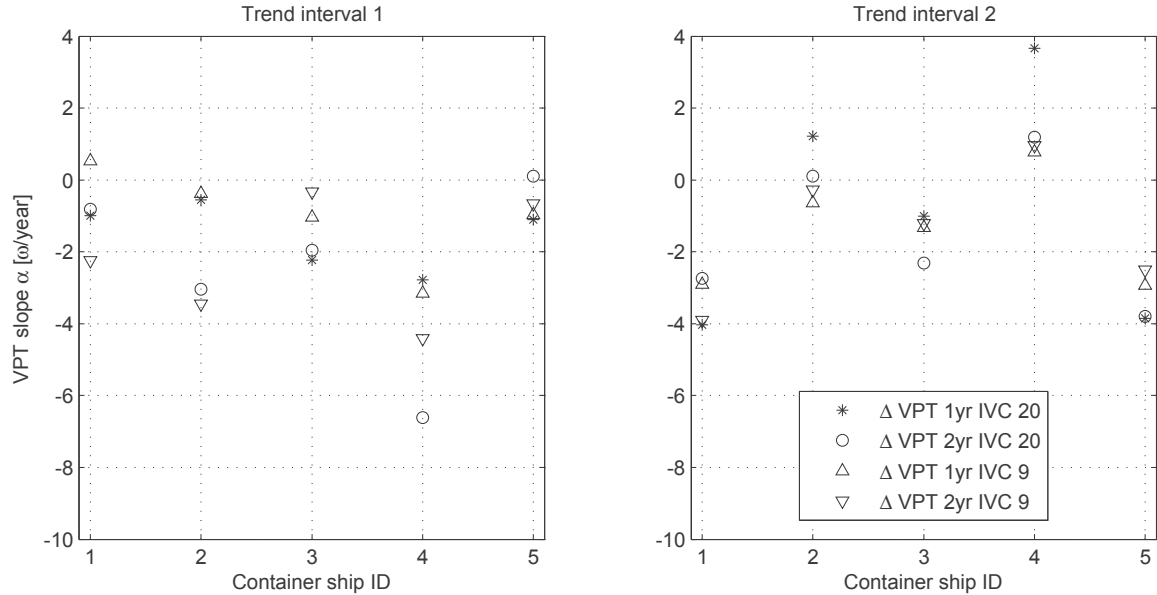


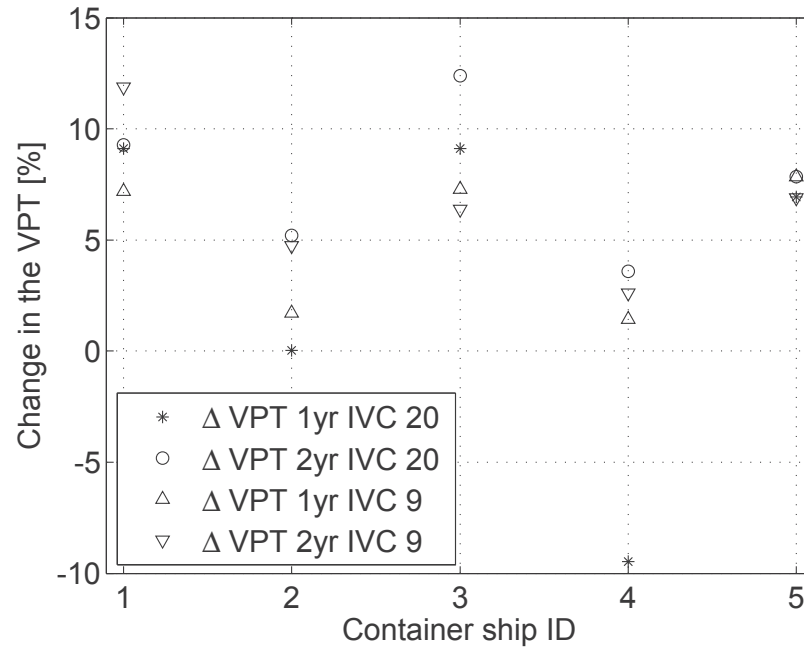
Figure 5.3: VPT slopes, α_1 and α_2 for the *VTP* interval 1 and 2, being trained for 1 and 2 years with *IVC* 20 and 9

energy consumption. The predictions are fairly consistent, ranging from 0-5% change to 7-13% for a dry-docking, except for an outlier point for *IVC* 20 trained for one year with a large negative value. It is noted that containership 2 and 4, both have the lowest increase in *VPT*, but they also both had the smallest decrease in *VPT* (α_2) after the dry-docking. The third dry-docking had a consistent positive effect ranging from 3.7% to 7.4% depending on the method used.

In general minor events such as hull and propeller cleaning had little effect and the intermediate trends relied on too little data for a reliable trend detection, and thus blurred the picture of the general trend between the dry-dockings. For containership 5 the effect from hull cleanings was much more distinct and therefore it was used as an example here and will also be used to detect events in the next chapter (6).

5.2.2 Trend detection of containership 5

The best performance trend representation was found by training on noon report data including hindcasts for the two first years using input variable 20. This data includes

Figure 5.4: The effect of the first dry-docking ΔVPT_1

IVC 20			α_1	ΔVPT_1	α_2	ΔVPT_2	α_3
Ship ID	Tr years	No.Tr data	$\omega/year$ %	%	$\omega/year$ %	%	$\omega/year$ %
1	1	54	-1.0	9.1	-4.0		
2	1	54	-0.6	0.0	1.2		
3	1	71	-2.2	9.1	-1.0		
4	1	44	-2.8	-9.5	3.7	3.4	-2.1
5	1	50	-1.1	6.9	-3.8		
1	2	114	-0.8	9.3	-2.7		
2	2	113	-3.0	5.2	0.1		
3	2	147	-2.0	12.4	-2.3		
4	2	98	-6.6	3.6	1.2	7.4	-2.8
5	2	112	0.1	7.8	-3.8		

Table 5.2: Key figures from VPT analysis of containerships 1-5 using data including hindcast

IVC 9			α_1	ΔVPT_1	α_2	ΔVPT_2	α_3
Ship ID	Tr years	No.Tr data	$\omega/year$ %	%	$\omega/year$ %	%	$\omega/year$ %
1	1	107	0.5	7.2	-2.9		
2	1	94	-0.4	1.7	-0.6		
3	1	135	-1.0	7.3	-1.3		
4	1	117	-3.2	1.4	0.8	6.2	-4.1
5	1	122	-1.0	7.8	-2.9		
1	2	264	-2.3	11.9	-3.9		
2	2	214	-3.5	4.7	-0.3		
3	2	293	-0.3	6.4	-1.2		
4	2	264	-4.4	2.6	1.0	6.1	-8.7
5	2	262	-0.7	6.9	-2.5		

Table 5.3: Key figures from VPT analysis of containerships 1-5 using data including noon report only

four hull cleanings and two dry-dockings of which the last can be neglected because it was at the very end of the dataset.

Figure 5.5 shows the performance trends between all the events. The trend is almost constant until the second hull cleaning at year ~ 4.2 where VPT shifts up and the slope decreases rapidly, resulting in VPT just before dry-docking that is less than if the previous trend continued. This indicates that even if a hull cleaning has an instant positive effect, it might damage the long-term performance which may be caused by a rough cleaning job that has damaged the anti-fouling effect of paint and increases the fouling rate.

After the dry-docking, the shift up in VPI is evident, and the the negative trend is slightly steeper than for the previous period. The last two hull cleanings at year $\sim 8.4 - 8.6$ do not improve the trend or the level of the VPI .

5.2.3 Discussion of Performance trend detection

Using the methods described above, it is possible to detect the general trends in performance change over time.

As illustrated, the dry-docking, hull and propeller cleanings do not always have the intended effect.

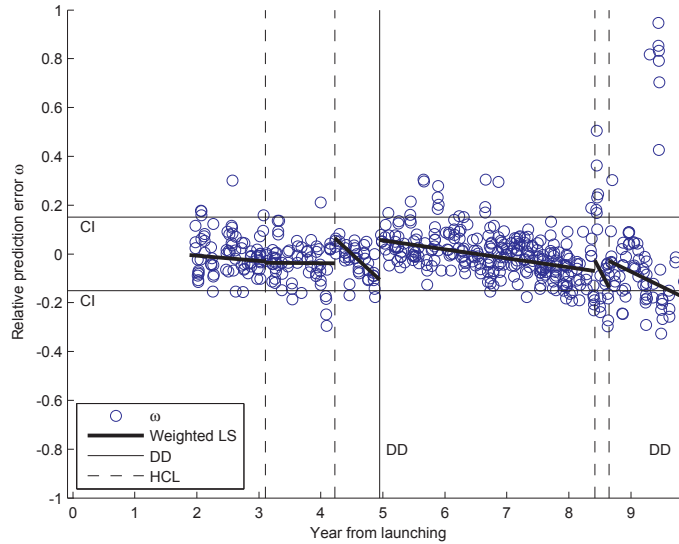


Figure 5.5: Performance trend between all events of containership 5 using *IVC 20* with data including hindcasts, trained with the first two-year period

The decrease in performance ranges from -6.6 to 3.9 and if average values from across the four methods are used the slope ranges from -4.2 to 1.6 %. A negative slope was expected and the ranges found here correspond with what has been found in similar research: 5.7%/year found in Andersen et al. (2005), 2.5-8.7%/year given by Eefsen (1996) both representing a tanker and 3.1-3.8%/year by Hansen (2011) which used containership comparable to the one used in this study.

The data suggested that the hull cleanings may have a positive immediate effect, but the long-term effect can be negative. This could be due to an excessive hull cleaning that has increased the permanent roughness and thus result in a higher fouling rate as discussed by Carlton (1994). It might also be that the event has no immediate effect, but the long-term effect can be beneficial. In order to evaluate this, detailed information about the event is needed such as what part of the ship was cleaned and what equipment was used. Having such information would make it possible to better categorize the events and for example use the events as discrete variables in the training of the methods.

In general, the dry-dockings had a more consistent effect with a positive change in *VPT* after the docking, but the slope afterwards varied from being steeper than

before, seen in containership 1, to being flatter than the previous trend, as seen with containership 4. This is in accordance with findings by Townsin et al. (1980) where the different paint types result in very different fouling rates.

The instantaneous effect from events, primarily dry-dockings, was, except for at one point, positive and using average values of the four methods reveals a range of -0.5 to 9.4%. This is slightly less than what was found in the literature, where Eefsen (1996) shows a reduction of $\sim 10 - 20\%$ and Munk (2006) is reporting numbers in the order of 20-30% reduction. This might be due to vessel type e.g. tankers used in Eefsen (1996) have a relatively higher frictional component of the total resistance and thus are more vulnerable to roughness .

Chapter 6

Detection of events

Whereas in the previous Chapter (5) the objective was to determine the trend of the relative prediction errors ω between known events, primarily dry-dockings, the present chapter attempts to identify the events based on the changes to ω alone.

The main focus of the thesis is not event detection and therefore a rather simple method developed for this purpose has been used.

In most cases, detection exercises are not necessary for the detection of known events as the ship operator is responsible for ordering the external events of the ship. Unknown events that may have an immediate effect on the performance of the ship can be detected and evaluated, such as minor propeller blade damage which will show a negative effect given the method is sufficiently sensitive. Furthermore detection of abrupt changes is also the ultimate test for propulsion performance system because most systems will always suffer from a certain extent of scatter and the challenge is thus detection of "real" changes, such as dry-dockings, in this scatter.

Knowledge of event detection is also fundamental for establishing a performance "alarm" that will trigger and inform the ship operator in case sudden changes in propulsion performance occur.

The data set consists of noon reports and hindcast data from five different "sister" containerships, as described in Chapter 3. The same data was used in the previous Chapter 5 to find the *Vessel Performance Trends (VPT)*. During this time, the ship hulls and propellers have been cleaned and polished as part of the regular maintenance

program. These events naturally influence the propulsion performance although it can be difficult to detect. The purpose of the first test is to detect major events such as dry-dockings.

6.1 Event detection method

Due to the scatter in the prediction errors, any changes can be difficult to detect. The error trend is expected to sustain its value or gradually decrease over time due to hull and propeller fouling. One way to detect this change is by comparing the mean values and the trend of the relative prediction error ω at two different time steps.

For each prediction time step t_j , the average value of the prediction errors was found by looking back on a period of time T_p to $\overline{\omega_j^B}$ (6.8) and by looking forward at the same period of time $\overline{\omega_j^F}$ (6.7) as illustrated in Figure 6.1.

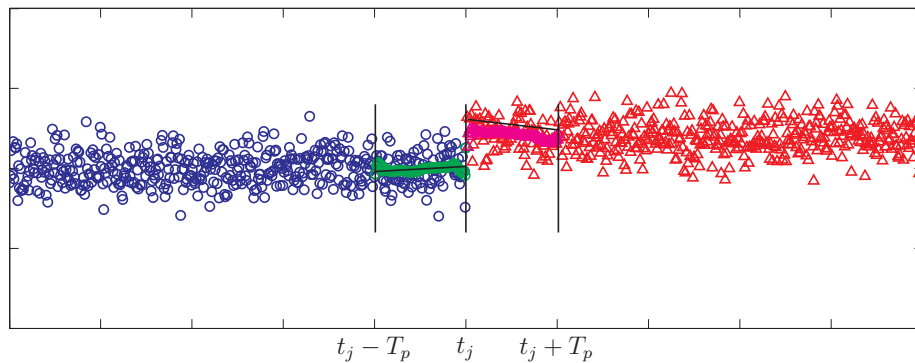


Figure 6.1: Backward-Forward average

Similar to the method described in Chapter 5, a weighted linear square regression can be performed on the same periods: $f_{VPT_j}^B(t_j - T_p)$ and $f_{VPT_j}^F(t_j + T_p)$.

The function values $f_{VPT}(t_j)$ and the slope α_j found by *LWS* (6.10), (6.9) can be used to evaluate the changes in the trend. They can be used as their original values or as the difference in the forward-backward time step.

The following measures were used to derive the forward-backward average:

The forward-backward (*FB*) difference of the average prediction error $\delta\omega_j^{FB}$ (6.11) gives a good and simple picture of the performance level over the time step period T_p ,

and the general shift in this level will be detected by it.

The forward-backward (*FB*) difference of the function values of $\delta f_{VPT_j}^{FB}$ (6.12) compares the first and the last value of each trend line VPT_j and thus includes the trend and mean values.

The forward-backward (*FB*) difference of the trend slopes $\delta\alpha_j^{FB}$ (6.13) shows when general changes in the trend occur, e.g. if an event did not have an instant effect but changes the future trend line. The values will be negative in the case of an upward change in performance level.

The governing control parameters for the detections are: the upper and lower threshold values for the detection (confidence interval *CI*), the time period for the moving average and the length of the training dataset.

$$\mathbf{t}_j^F = t(t_j \leq t_j + T_p) \quad (6.1)$$

$$\mathbf{t}_j^B = t(t_j - T_p < t_j) \quad (6.2)$$

$$\boldsymbol{\omega}_j^F \in \omega(t_j \leq t_j + T_p) \quad (6.3)$$

$$\boldsymbol{\omega}_j^B \in \omega(t_j - T_p > t_j) \quad (6.4)$$

$$N_j^F = |\boldsymbol{\omega}_j^F| \quad (6.5)$$

$$N_j^B = |\boldsymbol{\omega}_j^B| \quad (6.6)$$

$$\overline{\omega}_j^F = \frac{1}{N_j^F} \sum_{i=1}^{N_j^F} \omega_{ji}^F \quad (6.7)$$

$$\overline{\omega}_j^B = \frac{1}{N_j^B} \sum_{i=1}^{N_j^B} \omega_{ji}^B \quad (6.8)$$

$$[\alpha_j^F \beta_j^F] = f_{VPT}(\mathbf{t}_j^F, \boldsymbol{\omega}_j^F, \mathbf{w}_j^F) \quad (6.9)$$

$$[\alpha_j^B \beta_j^B] = f_{VPT}(\mathbf{t}_j^B, \boldsymbol{\omega}_j^B, \mathbf{w}_j^B) \quad (6.10)$$

$$\delta\omega_j^{FB} = \overline{\omega}_j^F - \overline{\omega}_j^B \quad (6.11)$$

$$\delta f_{VPT_j}^{FB} = (\alpha_j^F t(1)_j^F + \beta_j^F) - (\alpha_j^B t(N_j^B) + \beta_j^B) \quad (6.12)$$

$$\delta\alpha_j^{FB} = \alpha_j^F - \alpha_j^B \quad (6.13)$$

where:

N_j^F the number data within one time step forward of t_j

N_j^B the number data within one time step backward of t_j

\mathbf{w}_j vector of input weights corresponding to ω_j

T_p the time step length

j index for each time step

6.1.1 Detection tests

Containership 5 was used to tune the control parameters, *time step length* T_p and threshold levels, because it showed the most consistent changes in prediction errors due to known events as illustrated in Section 5.2.2.

Figure 6.2 shows an example of the three measures noted as: $\delta\omega_j^{FB}$, $\delta f_{VPT_j}^{FB}$ and $\delta\alpha_j^{FB}$, for a selected time window of containership 5 trained for the first two years with input variable combination *IVC 20*. A time step period T_p of 120 days, was found suitable for finding significant shifts and detecting events. The change in the mean error $\delta\omega_j^{FB}$ is shown by blue circles and indicates a shift in average level. Average periods that are too long may hide real changes. $\delta f_{VPT_j}^{FB}$ pictured as triangles is the difference between the last value of the performance trend (*VPT*) for the period looking back and the first value of the (*VPT*) looking forward, and thus takes into account the slope of the trend. It is equivalent to the changes of an event detected in the previous Chapter 5. The combined level $\delta f_{VPT_j}^{FB}$ and $\delta\omega_j^{FB}$ indicates a event right after the first dry-docking.

The difference in the *VPT* slope looking forward and backward $\delta\alpha_j^{FB}$ is evaluating events that do not necessarily have an instantaneous impact but change the fouling rate as seen in Chapter 5. This is evident after the second hull cleaning, where a large negative change is detected and is the only indication of an event. This change is also noted by the VPT_j plotted with dotted lines.

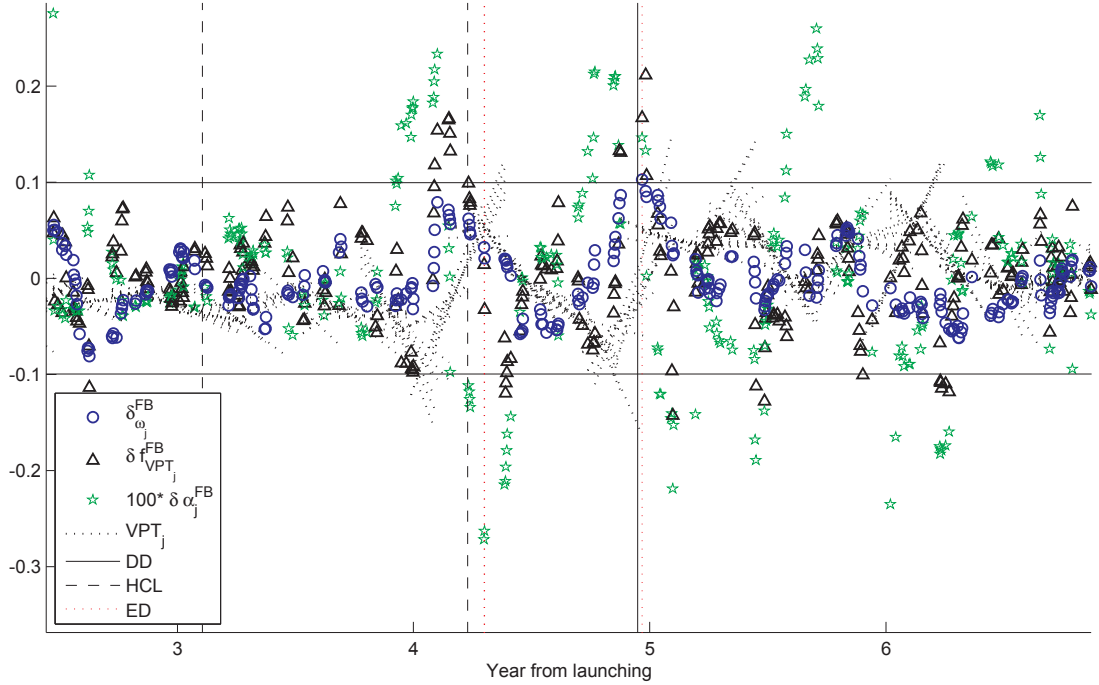


Figure 6.2: A selected view of the event prediction of containership 5, used for determining the detection limits. Initial training was 2 years with *IVC* 20 and the detection time step period

Note that $\delta\omega^{FB}$ increases just after the dry-docking (year ~ 4.9) and a little before the second hull cleaning (year ~ 4.1), which indicates the largest shift in the prediction error level. In both instances, the values of $\delta f_{VPT_j}^{FB}$ are high.

Furthermore, the piece-wise trends *VPT* illustrate (with some scatter) the general trends found in Section 5.2.2. The change in the trend slope is most significant at the second hull cleaning *HCL* which is illustrated by the piece-wise weighted linear regression lines, *VPT*.

An appropriate level for detection of the events was found when: $\delta\omega_j^{FB} > \sigma_R * 0.66 \wedge \delta f_{VPT_j}^{FB} > \sigma_R * 0.66$ or for changes in level $\delta\alpha_j^{FB} \sigma_R * 1.65 \cdot 10^{-3}$, where σ_R is the *GPR* predictive standard deviation.

6.1.2 Event detection for containerships 1-5

The procedure determined above has been applied to the five containerships (*CS 1-5*) trained with the same four training combinations as used for the trend detection in Chapter 5, *IVC 9* and *20* trained for one and two years. This resulted in four different series of "guess" event dates for every *CS*. For every ship they were reported and summarized in Table 6.1 together with the best training period and *IVC*. The number of encounters of mis-predictions and false-positive far from an event varies significantly between the ships. For *CS 4* the first prediction is the nearest and *CS 3* has three mis-predictions ahead of the nearest to the dry-docking. The amount and positions of mis-prediction are best illustrated by the individual ships in Figure 6.3-6.7 and will be discussed separately.

Table 6.1 shows that the first dry-docking can be detected with accuracy \pm a month (-0.8 - 0.9 months). *IVC 20* has the majority of best performances and *IVC 9* is only better for *CS 1*. One year of training is sufficient for *CS 1-4* while *CS 5* performs better with two years of initial training.

In the section below the event detection performed with the best training combinations as described above for *CS 1-5* are presented and discussed.

CS	Training procedure		1 st <i>DD</i> year	<i>ND</i> year	Detection delay months
	<i>IVC</i>	Training years			
1	9	1	4.42	4.46	0.6
2	20	1	4.47	4.50	0.4
3	20	1	4.90	4.98	0.9
4	20	1	4.98	5.04	0.7
5	20	2	5.03	4.97	-0.8

Table 6.1: Summary of the event detections nearest *ND* to the first dry-docking *DD*, both given in *years* from the launching of the vessel.

Event detection from containership 1

As illustrated in Figure 6.3 containership 1 has only one mis-detection prior to the first dry-docking. Then there are a few detections right after and a few at around year 9.

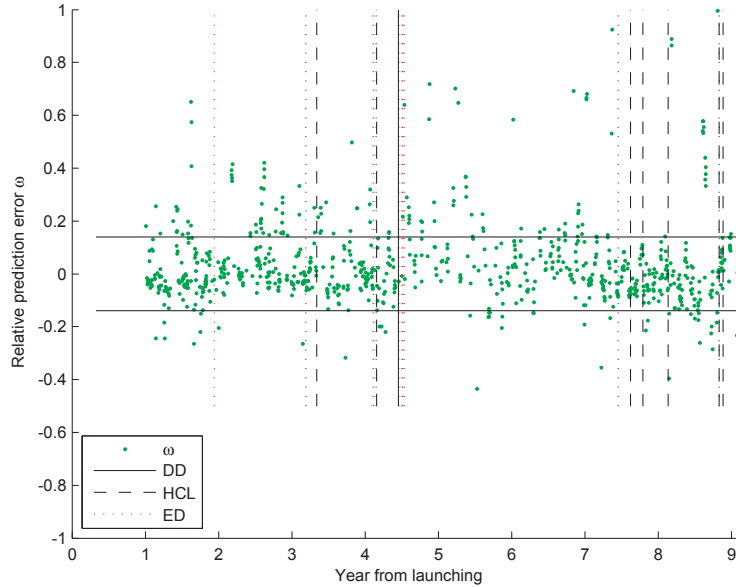


Figure 6.3: Event prediction of containership 1, trained with the first one years of noon report data including hindcasts, with *IVC 9*.

Event detection from containership 2

Figure 6.4 shows that three mis-prediction are encountered before the one at the first dry-docking. After year ~ 5.5 of mis-predictions are found but also one positive for the second hull cleaning which is detected at ~ 8 .

Event detection results from containership 3

Figure 6.5 shows a significant amount of mis-detections prior and after the first dry-docking, which is detected 0.9 month after the occurrence. The following hull cleanings are detected along with some mis-detections.

Event detection results from containership 4

This configuration has no mis-detections before the first dry-docking as illustrated in Figure 6.6. After that a series of mis-detections are detected.

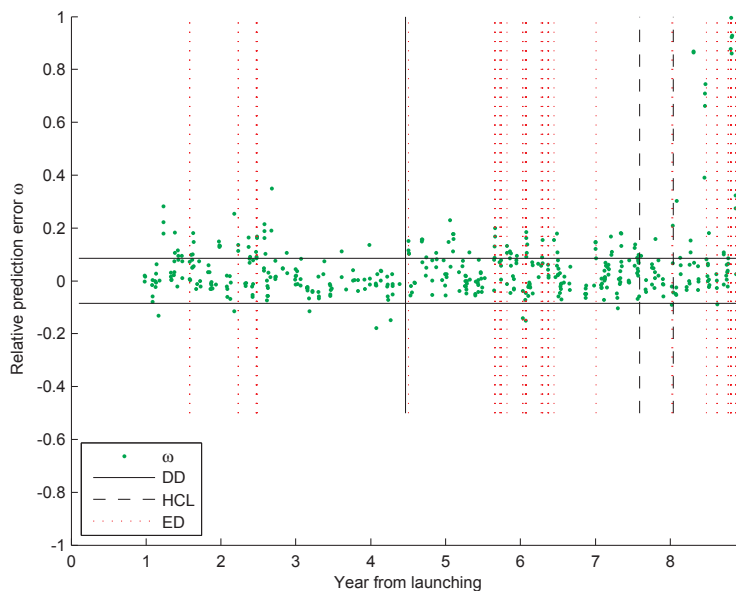


Figure 6.4: Event prediction of containership 2, trained with the first year of data for all noon reports with *IVC 20*.

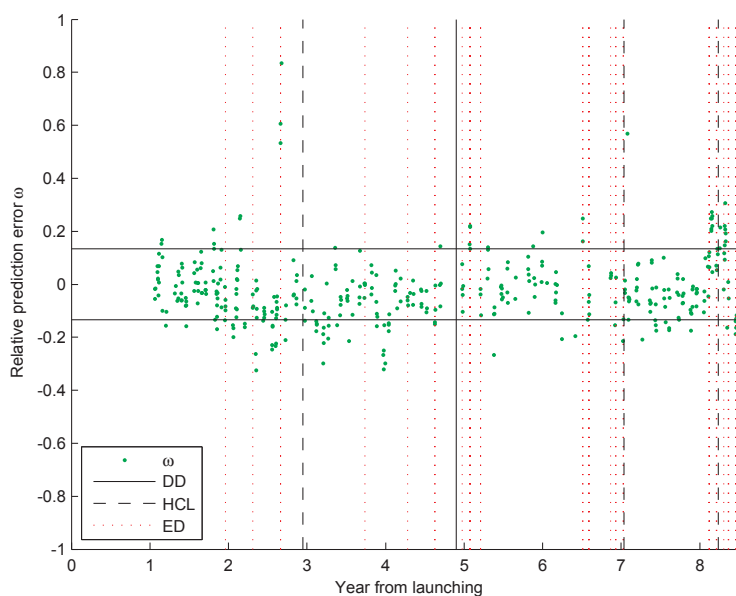


Figure 6.5: Event prediction of containership 3, trained with the first year of data for all noon reports with *IVC 20*.

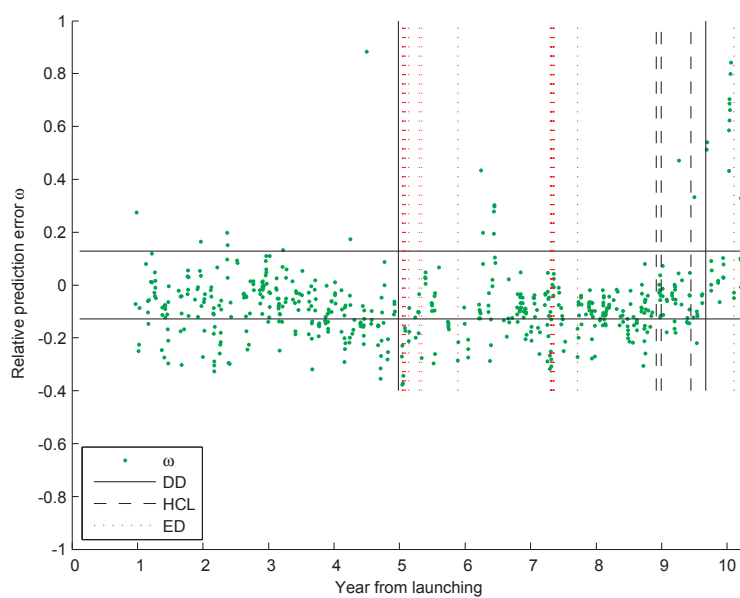


Figure 6.6: Event prediction of containership 4, trained with the first one year of data for all noon reports with *IVC 20*.

Event detection results from containership 5

This was the data that the method was tuned by, and Figure 6.7 shows the same data as Figure 6.2 in its entirety. It is illustrated how both the second hull cleaning and the first dry-docking are captured well.

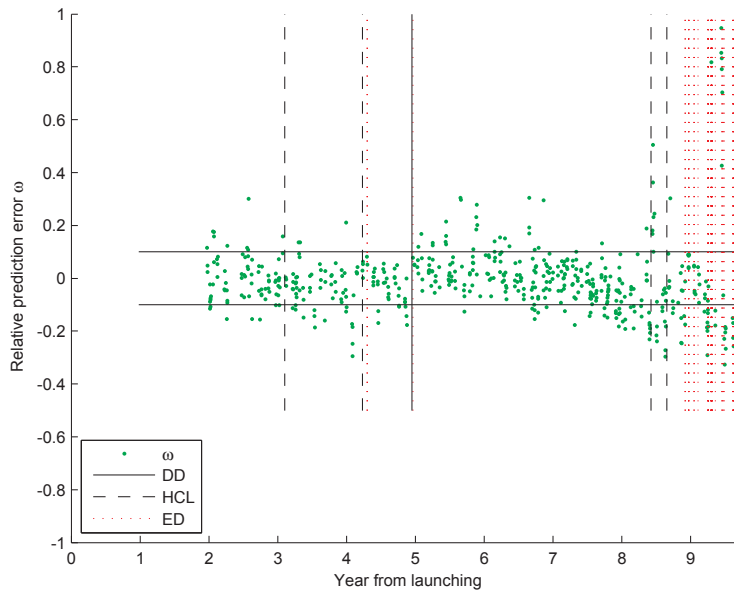


Figure 6.7: Event prediction of containership 5, trained with the first two years of data for noon reports with hindcast data, with *IVC 20*.

6.2 Discussion of event detection

In this chapter it was demonstrated that it is possible to use results from *GPR* predictions to detect abrupt changes in the propulsion performance. However, the method detects a number of mis-predictions or false positive detections and in practice it must only be used as a guide rather than a fixed indicator of an event.

The effect from different events (dry-docking, hull and propeller cleaning) varies considerably and consequently the method had to be tuned to find the events assumed to be most significant, hence the dry-dockings. As illustrated in Table 6.1 it was possible to detect dry-dockings within a month, which is regarded as being adequate.

It should be taken into consideration that only a limited amount of data is available and within this period the data density varies from 0.14 per day (using HC only) to 0.4 per day (using all NR). A month subsequently results in a maximum \sim of 12 data points. Furthermore, using a certain time step period T_p looking forward and backwards, the number of time points in each direction will not necessarily be the same. Due to the varying number of data points within each time step, the scatter in the trend and the mean values can also vary.

The present method requires adjustment of the time step length, detection limits and consistent data quality, and is thus dependent on reliable information of prior events or from a similar ship to tune the method.

In order to improve event detection more sophisticated methods can be used as e.g. rendering it as an *un-supervised* learning problem a *cluster* analysis or by *density estimations* to find distributions of different *clusters* (Bishop (2006)). Another approach is introducing adaptive filtering as investigated by Hansen (2011) for a similar problem where it was attempted to detect the change in propulsion performance due to a propeller cleaning based on measured data. Here a *CUSUM* (Gustafsson (2000)) function was used to identify changes in the performance. The signal to noise ratio was considerably smaller for the problem in Hansen (2011) and thus very different from the data in this thesis. Evaluating e.g. Figure 72 in Hansen (2011) it is not expected that the *CUSUM* function would perform better than the present method.

Chapter 7

Conclusion

In this thesis, two data-driven methods - Artificial Neural Network *ANN* and Gaussian Process Regression *GPR* - were applied to the ever more important question of monitoring and analysing propulsion performance of ships. The data-driven methods have been compared to traditional empirical model methods. It has been demonstrated how data-driven methods can be used for evaluation of the propulsion performance without any ship-specific information, but only using operational data.

Regression models were used to determine energy consumption by first comparing the predicted energy consumption for a certain *state* of the vessel with the actual measured energy consumption for the same *state*. The relative prediction difference or error between these values represents the relative difference of how the ship performs compared with the expected propulsion performance based on the initial training. The linear trend of the relative prediction error over time represents the development of the propulsion performance.

Data evaluation

Three datasets were used to evaluate the regression models:

- 1 A dataset consisting of automatically measured data collected on a product tanker over a two-month period. The data was subsequently split into four subsets.

2 Noon Report data from the same product tanker from a 2-year period

3 Noon Report data from five containerships over a period of up to 10 years

All datasets were used with and without hindcast weather data which reduced the amount of data by approximately 50% due to the limited availability of hindcast in certain areas.

The data quality and density vary significantly between the three datasets, as do the available input variables.

Regression methods

Artificial Neural Networks (*ANN*) and Gaussian Process Regression (*GPR*) were found to be superior to linear and customized non-linear models that were tested with the measured dataset. *ANN* and *GPR* showed similar prediction performance, but *GPR* was chosen for the further analysis due to its additional features of finding the *hyperparameters* that included the characteristic *length-scales* and prediction variances. The *length-scales* gave valuable information about what input variables were most relevant to the output and were used to confirm general assumptions about the input variables. The predictive variance was found for every prediction and was later used to determine the performance trend.

Data-driven evaluation of the datasets

The results from the *GPR* analyses of the three different datasets were compared. The measured dataset showed prediction errors of 0.76-1.5%, the *NR* from the tanker revealed 6.15%, and *NR* from one containership resulted in 5.08%, all errors included a 95% confidence interval. Adding *hindcast* had a beneficial effect by reducing the prediction error by 1% although the amount of data was reduced. The best results from the noon report data were found by including *time* as a variable.

Determination of the input variable combinations

Different combinations of input variables were used to find the best combination. Including the *time* as a linearly increasing variable resulted in a significant and consistent reduction of the prediction errors by approximately 1%. It was suggested that the *time* variable could represent the effect of hull and propeller fouling, because it was the only variable having a linear trend over time.

Analysis of the *length-scales* from the *GPR* models confirmed that the *time* had a significant influence. Furthermore, it was seen, as expected, that the speed and draught had a high relevance, and also the hindcast wind speed, direction and significant wave heights had a high influence.

For the data including hindcasts, the best combination was found using all available variables. However, without hindcast data the best variable combination was simpler. When analysing the data from one containership without the *time* variable, the best combination only included the logged speed, seawater temperature and draught mid-ship.

Understanding the length of the training set

The length of the training set is important for performance evaluation, because it governs when the data-driven predictions can be imitated. The best combinations of input variables were tested using difference length test sets. The results varied between the containerships, however longer datasets did not perform better than shorter datasets. Generally the input combinations that included hindcast data showed more consistent results across the different lengths of training data.

Predicting performance trend

The prediction errors from the regression models were used to form a linear trend to represent the performance decrease (*VPT*). In order to utilize the feature of predictive variance from *GPR* the trend was determined by using "weighted least square" where the reciprocal predictive variances were used as weights.

This is a unique feature of *GPR* which is not feasible or even impossible with

other methods. Fundamentally it enables taking into account that the predictions in certain states of the ship are not as well determined due to scarce data for those states. Consequently the method for extracting trends will rely less on poor data and more on sound data.

The *VPT* was split into sections so that the trend would not be affected by hull "events" such as dry-docking. The average of the linear trend slopes representing the decrease in performance over time varied from -4.2% to 1.6% in relative propulsion power per year. -4.2% is a realistic but small number in comparison with what is found by others, but a positive trend of 1.6% indicates that the ship performance is improving over time which is surprising. No increase in roughness over time has been reported by Townsin et al. (1980) so *VPT* close to zero is not unheard of, but a positive trend is unusual. A physical explanation for this could be that after a dry-docking where self-polishing anti-fouling or fouling release paint was applied, the vessel was at berth for a period, so that fouling would have time to adhere, and after starting service again, it would take some time for this growth to be polished off. No examples of similar trends are found by others, so there might be other explanations to this development.

The effect from external events was determined by evaluating the difference between the *VPT* right before an event and right after. This resulted in improvements of -0.5 to 9.4%. This is less than reported by others (e.g. Eefsen (1996)), but maybe due to the smaller decrease in performance detected over time and thus less is gained by a dry-docking. Having a small negative effect is not alarming if the *VPT* slope after dry-docking has a smaller magnitude than before as this indicates that a dry-docking may have instantaneously made the performance a bit worse, but the performance will not deteriorate as fast, and consequently the performance will be overall better viewed over a longer period of time.

The data-driven method presented for predicting *VPT* using only operational data is thus regarded as comparable with traditional methods using noon report input data. The regression analyses of automatically sampled data showed significantly lower prediction errors than the traditional methods, and long-term performance trend of such data is thus expected to be considerably better.

Predicting abrupt changes

Event detection was not the main objective of the thesis, but it has been used as a test of how well the method can perform. The method used for detection is a simple test of the shift in mean and function values looking forward and backwards. One set of data from a containership was used as a test case to find the best length of the piecewise mean values and the threshold values for detection. Testing on the remaining dataset for containerships showed varying results as the first dry-docking was generally detected within a month, but some datasets had a significant number of false positive detections. It is thus only recommended to use the method as a guide for the ship operator in order to make a qualified judgement regarding need for further investigations.

Economic perspective

When estimating the economic aspects of vessel performance monitoring and control, several factors have to be taken into account. In Townsin (2003), an example of finding the global savings due to better anti-fouling paints is divided into the four groups: *1* related to the fuel cost of increased frictional resistance, *2* related to the cost of changes in docking intervals, *3* savings due to lower cost of dry-docking and *4* indirect savings due to for example lower costs for transport of fuel oil.

The economic benefits of the present work were only evaluated with regard to the direct savings or penalties incurred due to changes in propulsion power and consequently fuel consumption.

The ship performance decreased by up to 4% per year between dry-dockings. If the propulsion efficiency is assumed to be constant; the fuel consumption will increase similarly. The daily fuel consumption of a containership (60,000kWh, 170g/kWh) is estimated to be 250 tonnes/day, and assuming 200 days of operation per year, an annual consumption of 50,000 tonnes of fuel results. The estimated price of bunker fuel (IFO380cSt) in 2013 is 630\$ per tonne^a which gives an annual increase of 1,260,000\$ and the annual accumulated increase in fuel consumption is therefore estimated to

^awww.bunkerworld.com

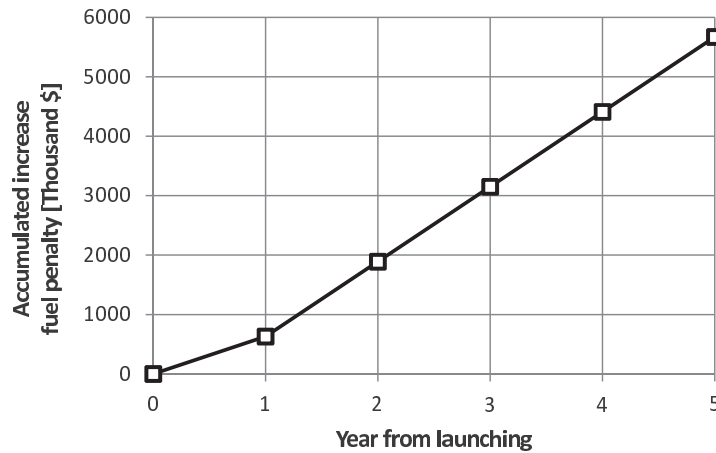


Figure 7.1: Accumulated increase in fuel consumption

630,000\$. So for the first year, the increased fuel consumption will represent 630,000\$ and after that increase by 1,260,000\$ per year as illustrated in Figure 7.1.

7.0.1 Suggestions for further development

The current work can lead to improvements and exploration in several directions.

Data quality improvement

In data-driven models, an increased amount and/or better quality of data is expected to increase the prediction performance. Subsequently, it will be possible to develop a reliable trend of the prediction error (*VPT*) in less time, and the operator can be warned earlier in case of bad performance.

One of the reasons why the introduction of hindcast data reduces the data sets is that they are not available in all areas. Receiving data from more areas would give access to more data and improve the predictions.

More detailed hindcast weather information could also eliminate the need for some of the reported data, such as e.g. logged speed if the sea current was available. Ultimately, the report values could be limited to the speed over ground and the energy consumption as all weather information is covered by the hindcasts.

Since the noon report data are average values over about 24 hours, the variables may vary within this period without the variations being registered which is a major source of uncertainty. Much better data for performance evaluation can be collected on "speed-runs" which are data collections from shorter periods of time where the ship is travelling at a constant speed or power, preferably in fair weather conditions.

Method improvements and exploration

The detection of events could be studied as a problem of classification, and dividing the events into more detailed categories may offer a better solution because the effect of different events, such as dry-dockings, vary significantly. This variation is most likely due to different treatments during the dry-docking, such as different paint qualities or blasting procedures.

A major drawback of the data-driven methods is that the data has to be collected before the initial training can begin. Using synthetic produced data from model tests or sea trials could potentially "speed up" learning of the initial training, so that the active performance monitoring could start closer to the launching of the vessel. It may also be possible to filter out the effect of time in the initial training set to avoid any linear effect expected from the fouling. This may make it possible to use longer initial training periods without being biased by the effect of fouling.

Introduction of an "unsupervised" data-driven approach, such as a *cluster* or *density estimation*, could be tested for detection of events. These methods attempt to find patterns in data without any previous information, and potentially it would be possible to detect the different regimes of the prediction errors due to the events. Adaptive filtering methods as described by Gustafsson (2000) could also be tested as in Hansen (2011) in order identify changes in the data trend.

It should be studied how the *length-scales* can be used more actively. The *length-scale* of the *time* variable could for example be evaluated for many accumulative trainings over time in order to evaluate the trend of the *length-scale* of *time* and subsequently be used as a measure for the fouling.

Artificial prediction, where the development of different variables are evaluated given that all other variables are held constant was briefly presented in Section 2.2.4

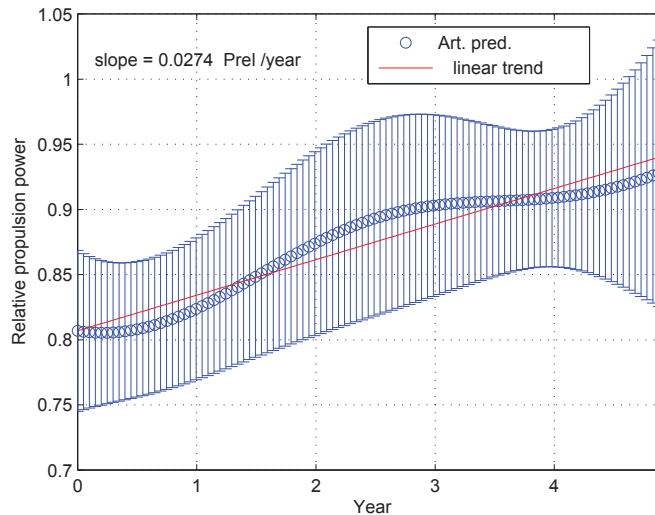


Figure 7.2: Artificial prediction of the relative as a function of time where time is the only changing variable (containership 4)

to illustrate the predictive variance found by *GPR*. This could also be used to evaluate the development of the input variables separately. It would, for example, be possible to develop a speed-power curve by keeping all constant except the speed.

In the trainings that include *time* as a variable, development of the time as a function of the relative predicted *EC* should represent the fouling effect. An example of this is given in Figure 7.2 for a containership, where all the input data is constant except the *time* input that varies from 1 to 5 years.

Bibliography

- Andersen, P., Borrod, A.-S., and Blanchot, H. (2005). Evaluation of service performance of ships. *Marine Technology*, 42:177–183.
- Bertram, V. (2000). Past, Present and Prospects of Antifouling. Technical report, HSVA, Bramfelder Str 164, D 22305 Hamburg.
- Bhattacharyya, R. (1978). *Dynamics of Marine Vehicles*. Ocean Engineering. Wiley-Interscience, 1 edition.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, 1 edition.
- Blendermann, W. (1990). External force, wind forces. *Schiff & Hafen / Seewirtschaft*, 42:Heft 2: 41–43, Heft 3: 55–58, Heft 4: 47–48.
- Boese, P. (1970). Eine Einfache Methode zur Berechnung der Widerstandserhöhung eines Schiffes im Seegang. Technical report, Institut für Schiffbau der Universität Hamburg,. Bericht Nr. 258,.
- Callow, M. E. and Callow, J. A. (2002). Marine biofouling: a sticky problem. *Biologist*, 49:1.
- Carlton, J. S. (1994). *Marine Propellers and Propulsion*. Butterworth-Heinemann Ltd., Oxford.
- DTU-toolbox (2002). Neural regressor with quadratic cost function. <http://isp.imm.dtu.dk/toolbox>.

- Eefsen, T. (1996). Hydrodynamisk fremdrivningskontrol. Master's thesis, Department of Naval Architecture and Offshore Engineering, Technical University of Denmark.
- Faltinsen, O. M. (1990). *Sea Loads on Ships and Offshore Structures*. Cambridge Ocean Technology Series. Cambridge University Press, 1999 edition.
- Faraway, J. J. (2002). *Practical Regression and Anova using R*. <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>.
- Gerritsma, J. and Beukelman, W. (1972). Analysis of the Resistance Increase in Waves of a Fast Cargo Ship. *International Shipbuilding Progress*, 19:285–293.
- Gustafsson, F. (2000). *Adaptive Filtering and Change Detection*. JOHN WILEY & SONS, LTD, Baffins Lane, Chichester.
- Hansen, S. V. (2011). *Performance Monitoring of Ships*. PhD thesis, Section of Coastal, Maritime and Structural Engineering, Department of Mechanical Engineering, Technical University of Denmark.
- Harvald, S. A. (1983). *Resistance and Propulsion of Ships*. John Wiley & Sons.
- Hess, D. E., Faller, W. E., Junghans, K., Lewis, J. L., Lee, J., and Ammeen, E. S. (2009). Neural Network Progress at NSWCCD. In *COMPIT 8*, Budapest, Hungary. COMPIT.
- Hess, D. E., Roddy, R. F., and Faller, W. E. (2006). Neural Networks Predictions of the 4-Quadrant Wageningen B-screw Series. In *COMPIT 5*, pages 315–334.
- Hogben, N., Dancunha, N., and G.F.Olliver (1986). *Global Wave Statistics*. Unwin Brothers Limited.
- Holtrop, J. and Mennen (1982). An Approximate Power Prediction Method. *International Shipbuilding Progress*, 29:166–170.
- Holtrop, J. and Mennen (1984). A statistical re-analysis of resistance and propulsion data. *International Shipbuilding Progress*, 31:272–276.

- IMO (2009). Second IMO GHG Study 2009. Technical report, International Maritime Organization.
- IMO (2012). International Shipping Facts and Figures - Information Resources on Trade, Safety, Security, Environment. International Maritime Organization - Maritime Knowledge Centre.
- International (2004). Hull Roughness Calculator. International Marine Coatings.
- Isherwood, R. (1972). Wind resistance of merchant ships. Technical report, The Royal Institute of Naval Architecture. Issued for Written Discussion.
- ITTC (2011). *ITTC 1978 - Recommended Procedures and Guidelines - Performance Prediction Method*. ITTC. 7.5-02 03-01.4, Rev. 2.
- Jorgen Strom Tejsen, Hugh Y. H. Yeh, D. D. M. (1973). Added resistance in waves. *The Society of Naval Architecture and Marine Engineers, Transactions*, 81:109–143.
- Journée, J. (1976). Prediction of Speed and Behaviour of a Ship in a Seaway. *International Shipbuilding Progress*, 23(265):1–24. Updated September 13th 2001.
- Journée, J. (2003a). Review of the 1979 and 1980 Full-Scale Experiments Onboard Containership m.v. Hollandia. TUD Report 1349, Faculty of Design, Construction and Production, Ship Hydromechanics Laboratory Delft University of Technology.
- Journée, J. (2003b). Review of the 1985 Full-Scale Calm Water Performance Tests Onboard m.v. Mighty Servant 3. TUD Report 1361, Faculty of Design, Construction and Production, Ship Hydromechanics Laboratory Delft University of Technology.
- Journée, J. and Meijers, H. (1980). Ship Routeing for Optimum Performance. In *Transactions IME, 21 February 1980, Conference on Operation of Ships in Rough Weather*. Rapport 0529-P, 1980, Delft University of Technology, Ship Hydromechanics Laboratory,.
- Journée, J., Rijke, R., and Verleg, G. (1987). Marine performance surveillance with a personal computer. In *Automation Days 87*. Finish Society of Automatic Control.

- Updated: 05-12-2000. Report 0753-P, Delft University of Technology, Ship Hydromechanics Laboratory, Appendix I include a description of the mathematical propulsion model.
- King, M. J. (1982). The Measurement of Ship Hull Roughness. *Wear*, 83(2):385–397. Elsevier.
- Larsen, J. (1993). *Design of Neural Network Filters*. PhD thesis, Electronics Institute, Technical University of Denmark.
- Larsen, J. (1999). Introduktion to Artificial Neural Networks. Section for Digital Signal Processing, IMM, DTU. IMM2443.
- Larsen, M. O. (2000). *Statistiske grundbegreber*. Den Private Ingeniør Fond ved Danmarks Teknisk Universitet, 6 edition.
- Larsen, N. L., Simonsen, C. D., Nielsen, C. K., and Holm, C. R. (2012). Understanding The Physics of Trim. In *9th annual Green Ship Technology Conference, Copenhagen*.
- Leifur Þ. Leifsson, Hildur Sævarsdóttir, Sven Þ. Sigurðsson, and Ari Vésteinsson (2008). Grey-box modeling of an ocean vessel for operational optimization. *Simulation Modelling Practice and Theory*, 16:923–932.
- Logan, K. P. (2011). Using a Ship’s Propeller for Hull Condition Monitoring. In *ASNE Intelligent Ships Symposium IX*, Philadelphia, PA, USA. ASNE. Presented May 25th, 2011.
- MacKay, D. J. C. (1992). A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4:448–472.
- Malone, J. A., Little, D. E., and Allman, M. (1980). Effects of hull foulants and cleaning/goating practices on ship performance and economics. *SNAME Transactions*, 88:75–101.
- Moor, D. and Murdey, D. (1968). Motions and Propulsion of Single Screw Models in Head Sea. *Transactions of RINA*, 110:403–446.

- Moor, D. and Murdey, D. (1970). Motions and Propulsion of Single Screw Models in Head Sea, Part II. *Transactions of RINA*, 112:121–164.
- Munk, T. (2006). Fuel Conservation Through Managing Hull Resistance. In *Motor-ship Propulsion Conference, Copenhagen, April 26th, 2006*.
- Nabney, I. (2002). *Netlab: Algorithms for Pattern Recognition*. Springer. ISBN 1-85233-440-1.
- Neal, R. (1994). Priors for infinite networks. Technical report, Technical report CRG-TR-94-1, Department of Computer Science, University of Toronto.
- Neal, R. M. (1995). *Baysian Learning for Neural Networks*. PhD thesis, Graduate Department of Computer Science, University of Toronto.
- Nielsen, H. B. (2004). *A Matlab Toolbox For Optimization And Data Fitting*. DTU Informatics, IMM, Technical University of Denmark, version 2.2. edition.
- Nielsen, H. B., Frandsen, P. E., Jonasson, K., and Tingleff, O. (2004). Unconstrained Optimization. Informatics and Mathematical Modelling, Technical University of Denmark. 3. Edition.
- Nielsen, J. K., Pedersen, N. H., Michelsen, J., Nielsen, U. D., Baatrup, J., Jensen, J. J., and Petersen, E. S. (2006). SeaSense - Real-time Onboard Decision Support. In *World Maritime Technology Conference*, London.
- Nielsen, U. D. (2005). *Estimation of directional wave spectra from measured ship responses*. PhD thesis, Section of Coastal, Maritime and Structural Engineering, Department of Mechanical Engineering, Technical University of Denmark.
- Pedersen, B. P. (2006). A Comparative Study of Methods for Predicting Added Resistance in Waves. Master's thesis, Section of Coastal, Maritime and Structural Engineering, Department of Mechanical Engineering, Technical University of Denmark.

- Pedersen, B. P. (2007a). 183.25 m Handymax Bulk Carrier - Added Resistance in Waves. Technical Report 107-24345.11, FORCE Technology. Danish Centre for Maritime Technology.
- Pedersen, B. P. (2007b). Propulsion in waves analysed with four different methods - 50,000 dwt oil tanker. Technical Report 107-24345.14.20, FORCE Technology.
- Pedersen, B. P. (2008). Modeling of Ship Propulsion Performance. Report for DTU course 02459 Machine Learning.
- Pedersen, B. P. and Larsen, J. (2009a). Modeling of Ship Propulsion Performance. In *World Maritime Technology Conference, Mumbai, 21-24 January 2009*, Mumbai.
- Pedersen, B. P. and Larsen, J. (2009b). Prediction of Full-Scale Propulsion Power using Artificial Neural Networks. In *COMPIT 2009, Budapest 10-12 May 2009*, pages 537–550.
- Pedersen, M. (1997). *Optimization of Recurrent Neural Networks for Time Series Modeling*. PhD thesis, Institute of Mathematical Modeling, Technical University of Denmark.
- Pedersen, P. T., Andersen, P., and Aage, C. (2002). *Grundlæggende Skibs- og Off-shoreteknik*. DTU, MEK, Maritim Teknik, 4 edition.
- Petersen, J. B. and Lauridsen, B. (2000). Prediction of Hydrodynamic Forces From a Database of Manoeuvring Derivatives. In *Presented in MARSIM 2000, Orlando, Florida 8-12 May 2000*.
- Petersen, J. P. (2012). *Mining of Ship Operation Data for Energy Conservation*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark.
- Petersen, J. P., Jakobsen, D. J., and Winther, O. (2011). A Machine-Learning Approach to Predict Main Energy Consumption under Realistic Operational Conditions. In *COMPIT 2011*, pages 305–316.

- Rasmussen, C. and Williams, K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press, London, England, 1 edition.
- Salvesen, N. (1978). Added Resistance of Ships In Waves. *Journal of Hydronautics*, 12(1):24–34.
- Simonsen, C., Otzen, J., and Stern, F. (2008). EFD and CFD for KCS heaving and pitching in regular head waves. In *27th Symposium on Hydrodynamics. Oct. 5-10, 2008, Seoul, Korea*, volume 1, pages 302–320, Seoul, Korea.
- Svarer, C., Hansen, L., and Larsen, J. (1993). On Design and Evaluation of Tapped-Delay Neural Network Architectures. In *Proceedings of the 1993 IEEE International Conference on Neural Networks*, pages 46–51, New York, New York.
- Townsin, R. (2003). The Ship Hull Fouling Penalty. *Biofouling*, 19:9–15.
- Townsin, R., Byrne, D., Svensen, T., and Milne, A. (1981). Estimating the Technical and Economic Penalties of Hull and Propeller Roughness. *SNAME Transactions*, 89:295–318.
- Townsin, R., Byrne, D., Svensen, T., and Milne, A. (1986). Fuel Economy due to Improvements in Ships Hull Surface Condition 1976-1986. *International Shipbuilding Progress*, 33:383.
- Townsin, R., Kwon, Y., and Baree, M. (1993). Estimating the influence of weather on ship performance. *RINA Transactions*, 135:191–209.
- Townsin, R., Spencer, D., Mosaad, M., and Patience, G. (1985). Rough Propeller Penalties. *SNAME Transaction*, 93:165–187.
- Townsin, R. L., Byrne, D., Milne, A., and Svensen, T. (1980). Speed, Power and Roughness: The economics of Outer Bottom Maintenance. *RINA Transactions*, 122:459–483.
- Wadskjær, P. E. (2001). 76720 Skibshydrodynamik 1. Institut for Skibs- og havteknik, DTU.

Walderhaug, H. (1986). Paint Roughness Effects On Skin Friction. *International Shipbuilding Progress*, 33:382.

Willsher, J. (2007). The effect of biocide free foul release systems on vessels performance. In *International Conference on Ship Efficiency*, Hamburg. October 8-9th.

List of Figures

1.1	Example of performance decrease ($\omega = \frac{\hat{EC}-EC}{EC}$) over time where the solid vertical lines represent dry-docking and the dotted hull cleanings. The solid trend lines illustrate the <i>VPT</i>	5
1.2	Example of the relative change in the fuel consumption of a tanker 300,000 dwt crude oil tanker evaluated by SeaTrend [®]	12
2.1	Variables of the propulsion performance	20
2.2	Illustration of the structure of artificial neural network	34
2.3	Illustration of the variation in the predicted variances due to the number of input.	37
3.1	Measured mean draught from KYMA compared with the noon report draught	56
3.2	Airmar mounted on Monkey Island	57
3.3	GPSwin sensor diagram	58
3.4	Screen shot of the GPSwin logging window	58
3.5	AIRMAR vertical wind profiles	61
3.6	Comparison of wind speed and direction during the test period. Airmar and ship anemometer measurements are subtracted from the hindcast values	61
3.7	Ship speed and true wind speed distribution for dataset #1-#4	63
3.8	Variation in propulsion power due to turns	64
3.9	Ship speed and shaft power distribution for container ship 1-5	69

4.1	Prediction errors for the empirical methods: Gulddammer & Harvald and Holtrop-Mennen	81
4.2	Histograms of the relative prediction errors from the measured dataset (<i>MDTM</i> 1-4) using the linear method	83
4.3	Histograms of the relative prediction errors from the measured datasets (<i>MDTM</i> 1-4) using ANN, with 20 hidden units and <i>IVC 12</i> for all datasets, except dataset 2 where <i>IVC 2</i> is used.	86
4.4	Histograms of the relative prediction errors from the measured data (<i>MDTM</i> 1-4) sets using GPR with input variable setup <i>15</i>	87
4.5	The relative prediction errors vs. the predicted standard deviation from the <i>GPR</i> analysis with the measured datasets (<i>MDTM</i> 1-4)	88
4.6	Comparison of the relative prediction errors between all the regression methods: linear, non-linear with and without modifications and the best input variable combinations of ANN and GPR methods of measured datasets 1-4	89
4.7	Histogram and normal distribution of the relative prediction errors from ANN with 5 hidden units and <i>IVC 7</i> , using the noon report dataset from Torm Marie	91
4.8	Cross-validation errors $\overline{\omega_K}$ and cross-validation of the predicted standard deviations $\overline{\sigma_K}$ of the GPR for the noon report dataset from Torm Marie	93
4.9	Comparison of the relative prediction errors between <i>ANN</i> and <i>GPR</i> methods of the noon report data from Torm Marie for different <i>IVC</i> . <i>IVC 7,9,10</i> and <i>13</i> include the <i>time</i> as a variable.	94
4.10	Histogram and normal distribution of the relative prediction errors from ANN using the noon report dataset from containership 1 and <i>IVC 2</i> (Described in Table 4.10) and five hidden units.	97
4.11	Comparison of the relative cross-validation prediction errors ($\overline{\omega_K}$) between <i>ANN</i> and <i>GPR</i> methods of the noon report data from containership 1 trained with <i>20/80</i> cross-validation scheme	98

4.12	Relative cross-validation errors $\overline{\omega_K}$ and predictive standard deviation $\overline{\sigma_K}$ of the <i>GPR</i> for the noon report dataset from containership 1, using <i>LOO</i> training/test	99
4.13	<i>GPR</i> prediction results for containership 1 using <i>IVC</i> 21	99
4.14	Comparison of the relative prediction errors found by <i>GPR</i> for datasets of containership 1, including hindcast (251) and without (515).	101
4.15	Summary of the five containerships together. Trained accumulative for 0-1, 0-2,... 0-5 years, which is indicated by numbered markers. The coloured asterisk indicate the <i>IVC</i>	103
4.16	Relative prediction errors from containerships 1-5 (<i>NRCS</i> 1-5) trained with <i>IVC</i> 9,18,20,21 for five different periods of time starting at the launch with one-year intervals accumulating until the first dry-docking	104
4.17	Mean errors for the measure datasets (<i>MDTM</i>) analysed with <i>ANN</i> for different <i>IVC</i>	109
4.18	Logarithmic <i>length-scales</i> of <i>GPR</i> evaluation of the four measured datasets	110
4.19	Mean error for different <i>IVC</i> using <i>ANN</i> with NR data from Torm Marie	112
4.20	Logarithmic <i>length-scales</i> of <i>IVC</i> 15 and 16, of the noon report data from Torm Marie.	113
4.21	Mean error for different <i>IVC</i> , given in Table 4.10 from training with NR data from containership 1	114
4.22	Logarithmic <i>length-scales</i> of input variable combinations 20 and 21, with and without time as a variable (251 input dataset) for containership 1 (Launching 1 st dry-docking)	115
4.23	Logarithmic <i>length-scales</i> of <i>IVC</i> 17 and 18, with and without time as a variable. Using only noon report data, with and without time as a variable (251 input dataset) for containership 1 (Launching 1 st dry-docking)	116
4.24	The accumulated <i>length-scales</i> of <i>IVC</i> 20 (without <i>time</i>), trained for three periods of time: 0-1, 0-2, 0-4.5 years (1 st dry-docking)	117
4.25	The accumulated <i>length-scales</i> of <i>IVC</i> 21 (including <i>time</i>), trained for three periods of time: 0-1, 0-2, 0-4.5 years (1 st dry-docking)	118

5.1	Performance trend between all events of containership 1 using <i>IVC 20</i> with data including hindcasts, trained with the first two-year period . .	122
5.2	Performance trend between dry-dockings of containership 1 using <i>IVC 20</i> with data including hindcasts, trained with the first two-year period	123
5.3	VPT slopes, α_1 and α_2 for the <i>VTP</i> interval 1 and 2, being trained for 1 and 2 years with <i>IVC 20</i> and 9	124
5.4	The effect of the first dry-docking ΔVPT_1	125
5.5	Performance trend between all events of containership 5 using <i>IVC 20</i> with data including hindcasts, trained with the first two-year period . .	127
6.1	Backward-Forward average	130
6.2	A selected view of the event prediction of containership 5, used for determining the detection limits. Initial training was 2 years with <i>IVC 20</i> and the detection time step period	133
6.3	Event prediction of containership 1, trained with the first one years of noon report data including hindcasts, with <i>IVC 9</i>	135
6.4	Event prediction of containership 2, trained with the first year of data for all noon reports with <i>IVC 20</i>	136
6.5	Event prediction of containership 3, trained with the first year of data for all noon reports with <i>IVC 20</i>	136
6.6	Event prediction of containership 4, trained with the first one year of data for all noon reports with <i>IVC 20</i>	137
6.7	Event prediction of containership 5, trained with the first two years of data for noon reports with hindcast data, with <i>IVC 20</i>	138
7.1	Accumulated increase in fuel consumption	146
7.2	Artificial prediction of the relative as a function of time where time is the only changing variable (containership 4)	148
C.1	Histograms of measured data set 1-4 of out variable <i>Y</i> as defined in Table C.1 as <i>Y</i>	202

C.2	Histograms of measured data set 1-4 of input variable $X1$ as defined in Table C.1	203
C.3	Histograms of measured data set 1-4 of input variable $X2$ as defined in Table C.1	203
C.4	Histograms of measured data set 1-4 of input variable $X3$ as defined in Table C.1	204
C.5	Histograms of measured data set 1-4 of input variable $X4$ as defined in Table C.1	204
C.6	Histograms of measured data set 1-4 of input variable $X5$ as defined in Table C.1	205
C.7	Histograms of measured data set 1-4 of input variable $X6$ as defined in Table C.1	205
C.8	Histograms of measured data set 1-4 of input variable $X7$ as defined in Table C.1	206
C.9	Histograms of measured data set 1-4 of input variable $X8$ as defined in Table C.1	206
C.10	Histograms of measured data set 1-4 of input variable $X9$ as defined in Table C.1	207
C.11	Histograms of measured data set 1-4 of input variable $X10$ as defined in Table C.1	207
C.12	Histograms of measured data set 1-4 of input variable $X11$ as defined in Table C.1	208
C.13	Histograms of measured data set 1-4 of input variable $X12$ as defined in Table C.1	208
C.14	Histograms of measured data set 1-4 of input variable $X13$ as defined in Table C.1	209
C.15	Histograms of noon report data from Torm Marie output variable Y as defined in Table C.2 as Y	210
C.16	Histograms of the noon report data set 1-4 of input variable as defined in Table C.2	211

C.17 Histograms of the noon report data set 5-8 of input variable as defined in Table C.2	211
C.18 Histograms of the noon report data set 9-12 of input variable as defined in Table C.2	212
C.19 Histograms of the noon report data set 13-16 of input variable as defined in Table C.2	212
C.20 Histograms of the noon report data set 17-20 of input variable as defined in Table C.2	213
C.21 Histograms of the noon report data set 21 of input variable as defined in Table C.2	213
C.22 Histograms of noon report data from Torm Marie output variable Y as defined in Table C.3 as Y	215
C.23 Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 1-4 as defined in Table C.3	216
C.24 Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 5-8 as defined in Table C.3	217
C.25 Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 9-12 as defined in Table C.3	218
C.26 Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 13-16 as defined in Table C.3	218
C.27 Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 17-27 as defined in Table C.3	219
C.28 Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 28 as defined in Table C.3	219
D.1 Performance trend between all events of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first one-year period	224
D.2 Performance trend between dry-dockings of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first one-year period	224

D.3	Performance trend between all events of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first two-year period	225
D.4	Performance trend between dry-dockings of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first two-year period	225
D.5	Performance trend between all events of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first one-year period	226
D.6	Performance trend between dry-dockings of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first one-year period	226
D.7	Performance trend between all events of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first two-year period	227
D.8	Performance trend between dry-dockings of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first two-year period	227
D.9	Performance trend between all events of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first one-year period	228
D.10	Performance trend between dry-dockings of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first one-year period	228
D.11	Performance trend between all events of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first two-year period	229
D.12	Performance trend between dry-dockings of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first two-year period	229

D.13 Performance trend between all events of Container ship 2 using input variable setup <i>20</i> with data including hindcasts, trained with the first one-year period	230
D.14 Performance trend between dry-dockings of Container ship 2 using input variable setup <i>20</i> with data including hindcasts, trained with the first one-year period	230
D.15 Performance trend between all events of Container ship 2 using input variable setup <i>20</i> with data including hindcasts, trained with the first two-year period	231
D.16 Performance trend between dry-dockings of Container ship 2 using input variable setup <i>20</i> with data including hindcasts, trained with the first two-year period	231
D.17 Performance trend between all events of Container ship 3 using input variable setup <i>9</i> with all Noon report data, trained with the first one-year period	232
D.18 Performance trend between dry-dockings of Container ship 3 using input variable setup <i>9</i> with all Noon report data, trained with the first one-year period	232
D.19 Performance trend between all events of Container ship 3 using input variable setup <i>9</i> with all Noon report data, trained with the first two-year period	233
D.20 Performance trend between dry-dockings of Container ship 3 using input variable setup <i>9</i> with all Noon report data, trained with the first two-year period	233
D.21 Performance trend between all events of Container ship 3 using input variable setup <i>20</i> with data including hindcasts, trained with the first one-year period	234
D.22 Performance trend between dry-dockings of Container ship 3 using input variable setup <i>20</i> with data including hindcasts, trained with the first one-year period	234

D.23 Performance trend between all events of Container ship 3 using input variable setup 20 with data including hindcasts, trained with the first two-year period	235
D.24 Performance trend between dry-dockings of Container ship 3 using input variable setup 20 with data including hindcasts, trained with the first two-year period	235
D.25 Performance trend between all events of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first one-year period	236
D.26 Performance trend between dry-dockings of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first one-year period	236
D.27 Performance trend between all events of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first two-year period	237
D.28 Performance trend between dry-dockings of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first two-year period	237
D.29 Performance trend between all events of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first one-year period	238
D.30 Performance trend between dry-dockings of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first one-year period	238
D.31 Performance trend between all events of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first two-year period	239
D.32 Performance trend between dry-dockings of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first two-year period	239

D.33	Performance trend between all events of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first one-year period	240
D.34	Performance trend between dry-dockings of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first one-year period	240
D.35	Performance trend between all events of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first two-year period	241
D.36	Performance trend between dry-dockings of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first two-year period	241
D.37	Performance trend between all events of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first one-year period	242
D.38	Performance trend between dry-dockings of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first one-year period	242
D.39	Performance trend between all events of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first two-year period	243
D.40	Performance trend between dry-dockings of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first two-year period	243
E.1	Relative prediction errors from container ships #1-#5 trained with <i>IVC 9,18,20,21</i> for five different periods of time starting at the launch with one-year intervals accumulating until the first dry-docking	247
E.2	The relative standard deviation from container ships #1-#5 trained with <i>IVC 9,18,20,21</i> for five different periods of time starting at the launch with one-year intervals accumulating until the first dry-docking	248

List of Tables

2.1	Cross-validation scheme of 5 training/test data set, the shaded cells indicates test-part of the data and the clear cells the training part . . .	41
3.1	Changes in the propulsion power due to the fuel consumption	49
3.2	Main particulars of Torm Marie	55
3.3	List of measurements *p/nm: pulses per Nautical miles **p/l: pulses per liter . .	59
3.4	Measured dataset division	62
3.5	Average of the relative standard deviation	62
3.6	Torm Marie noon report variables	66
3.7	Torm Marie noon report dataset	66
3.8	Noon reports from containerships	67
3.9	Comparison of hindcast and observed wind speed, $V[m/s]$ for the containership data	68
3.10	Comparison of hindcast and observed significant wave height, $H_S[m]$ for the containership data	68
3.11	Input data filters	71
3.12	Data variables for the measured dataset analysis from Torm Marie . . .	72
3.13	Data variables for the <i>NR</i> dataset analysis from Torm Marie	73
3.14	Data variables for the container ship noon report dataset analysis . . .	74
4.1	Selecting the best model of Input Variable Combinations (<i>IVC</i>) for each of the four given variable	78
4.2	Accumulation of the training year for containerships 1-5	79

4.3	Applied methods and dataset combinations tested, including Long-term Trend evaluations (LT Trend)	79
4.4	Input variable setup combinations (IVC) for the measured data from Torm Marie ("NR" indicates that the data originates from the noon reports and "HC" indicates that the data originates from hindcasts) . .	82
4.5	Cross-validation errors $\overline{\omega_K}$ of the linear and non-linear methods for the measured dataset	83
4.6	Cross-validation errors, $\overline{\omega_K}$, of the ANN for the measured data from Torm Marie for different Input Variable Combinations (IVC) and Number of Hidden units (NH)	85
4.7	Crossvalidation errors $\overline{\omega_K}$ of the GPR for the measured dataset for different Input Variable Combinations IVC	87
4.8	Input variable combinations (IVC) for noon report data from Torm Marie	90
4.9	20/80 cross-validation errors $\overline{\omega_K}$ from ANN for the noon report data . .	91
4.10	Input variable setup combinations for Noon Report data from containerships 1-5	95
4.11	Relative cross-validation errors from ANN for the containership noon report data, $\overline{\omega_K}$ trained with 20/80 cross-validation.	96
4.12	The best input variable combinations for Noon Report data from the containerships	106
4.13	A summary of the cross-validated relative prediction error, variance, standard deviation and 95% confidence interval, based on the estimated normal distributions.	107
5.1	VPT test variations for containerships 1-5	121
5.2	Key figures from VPT analysis of containerships 1-5 using data including hindcast	125
5.3	Key figures from VPT analysis of containerships 1-5 using data including noon report only	126
6.1	Summary of the event detections nearest ND to the first dry-docking DD, both given in years from the launching of the vessel.	134

C.1	Date variables for the measured data set analysis'	202
C.2	Date variables for the noon report data set analysis'	210
C.3	Date variables for the noon report data set analysis'	215

Appendix A

Modeling of Ship Propulsion Performance

Presented at the World Maritime Technology Conference, *WMTC*, January 21-24 2009 in Mumbai, India.

Full scale measurements of the propulsion power, ship speed, wind speed and direction, sea and air temperature, from four different loading conditions has been used to train a neural network for prediction of propulsion power. The network was able to predict the propulsion power with accuracy between 0.8-2.8%, which is about the same accuracy as for the measurements. The methods developed are intended to support the performance monitoring system SeaTrend[®] developed by FORCE Technology (FORCE (2008)).

Modeling of Ship Propulsion Performance

Benjamin Pjedsted Pedersen (FORCE Technology / Dept. of Mechanical Engineering, Technical University of Denmark)

Jan Larsen (Department of Informatics and Mathematical Modeling, Technical University of Denmark)



Full scale measurements of the propulsion power, ship speed, wind speed and direction, sea and air temperature, from four different loading conditions has been used to train a neural network for prediction of propulsion power. The network was able to predict the propulsion power with accuracy between 0.8-2.8%, which is about the same accuracy as for the measurements. The methods developed are intended to support the performance monitoring system SeaTrend® developed by FORCE Technology (FORCE (2008)).

KEY WORDS

Propulsion; performance; monitoring; non-linear neural network regression; neural networks; fuel consumption

INTRODUCTION

As part of the Industrial PhD project "Ship Performance Monitoring" automatic data sampling equipment was installed on the tanker "Torm Marie" in January 2008 and so far data from four different loading conditions are available.

Modeling of these loading conditions are fundamental to achieving accurate prediction of propulsions power. In the future, the variation in draught and trim will be added as variables.

Ship propulsion performance (referred to as the performance) is a measure of the energy consumption at a certain state, i.e. speed, loading condition, weather condition and other factors. During the lifetime of the ship the performance will decrease e.g. the fuel consumption will increase at a certain state or the speed will decrease at a certain power setting. This is mainly due to fouling of the hull and propeller. A typical trend of the speed reduction is illustrated in Figure 1.

Hence, performance evaluation is about comparing the fuel efficiency or propeller power at one time to another time, in other words to compare the ship at one state with another state. Since a ship is subjected to external factors such as wind, waves, shallow water, change in sea water temperature, etc. as illustrated in Figure 2, it is unlikely that the ship will ever be in the exact same situation more than once. Furthermore these external factors can be difficult to measure accurately and thus the detection of a similar situation is problematic.

This deterioration is only a few percent and is therefore difficult to detect with traditional performance monitoring methods.

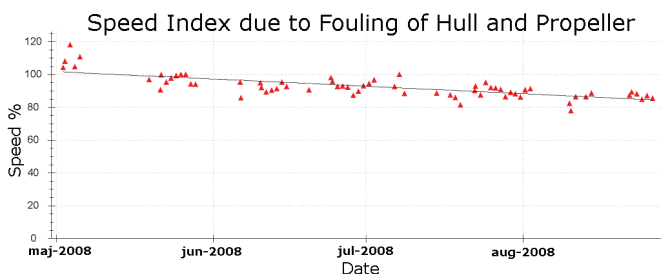


Figure 1: Increase of the fuel consumption as an effect the fouling

Traditionally, the problem has been solved by calculating a theoretical propulsion power for the actual condition using standard empirical resistance and propulsion methods, for example Harvald, S. A. (1983) or Holtrop, J. (1984) methods. For the estimation of the wind resistance a method proposed by Isherwood, R. (1972) can be used if no wind resistance coefficients are available for the ship.

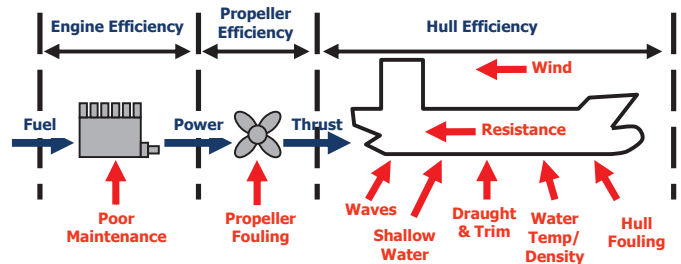


Figure 2: Performance variables

These empirical methods are derived from model tests and sea trials, and since most model test are carried out in a design condition (even keel) and speed, this is the region where it should be applied. In operation the ship will travel in many other conditions i.e., ballast draught and trimmed conditions.

Consequently, these methods give a rough estimate of the propulsion power rather than an accurate reference point. If some measured values from model tests or sea trials are available, they can be used to adjust the empirical data and thus give a more accurate result.

Another part of the problem is to have sufficient input data for the analysis in order to capture the dynamics of the propulsion power. This is relevant for the traditional method and any other method that can be used. A short description of the input is given below:

Draught and trim - usually these fundamental variables for the power estimation are found from visual observation or from the loading computer before departure; sometimes the arrival condition is determined by observations, but usually only from the loading computer. Some ships are equipped with dynamic draught measuring devices, but these are very sensitive devices which deliver a signal with a significant variance. Draught and trim have approximately an accuracy of 0.2m, as that is the usual scale for draughts marks.

Power measurement - the power can be measured in different ways. Measuring the propeller shaft torque with a torsionmeter, and the rate of revolution with a tachometer will give the direct power delivered to the propeller and is thus the preferable method.

The main engine fuel consumption is also a fairly good measurement, but it is necessary to have sufficient information of the fuel quality. A change in the main engine performance will also show a change in the fuel consumption, so it can be difficult to determine the propeller and hull performance from the fuel performance alone.

Speed through the water - is measured by the speed log that is based upon the Doppler principle. Experience shows that the signal from the speed logs has a tendency to drift and hence many ship officers do not trust the speed logs. It is also possible to estimate the speed through the water from the sea current

determined by a meteorological prognosis and from the speed over ground given by the GPS navigator. Although the speed log can drift it is expected to give more reliable results than the one estimated from the sea current and GPS speed.

Relative wind speed and direction - is measured by a doppler anemometer Airmar® Weatherstation™ PB100 mounted on top of the superstructure. At this position the wind speed and direction is altered from the free stream wind due to the presence of the ship. Traditional methods for estimating the wind resistance is based on wind tunnel tests and hence assumes that the wind speed and direction is undisturbed. Since the undisturbed wind speed and direction is unknown from measurements, the measured (disturbed) values are used directly in the empirical model, although this is not correct. Furthermore, the wind resistance coefficients are in this case determined empirically, which induce additional uncertainty.

Air temperature - is also measured by the Airmar® Weatherstation™ PB100 unit. The air temperature can vary significantly within a few hours, which has a direct effect on the air density and consequently on the ship resistance, e.g. for the ship traveling with a speed of 15 knots in 10 m/s and 0°C headwind the air resistance will be 215kN and total resistance 953+215 = 1168 kN, for the ship and wind speed, but with an air temperature of 10°C the wind resistance is 199kN and the total resistance 953+199 = 1152 kN. The difference in air density has a relative influence on total resistance of (1168-1152) / 1168 = 1.4%. For the present measurement the air temperature varies between 24.5-28.8°C and has a variance of ~0.47.

Sea water temperature - is usually measured once daily by the engine crew. The seawater temperature has a significant impact on the sea water density and viscosity, and consequently on the resistance. The difference between sailing in 0°C and 30° C seawater results in a frictional resistance around 7%, and for the present type of ship the frictional resistance accounts for 90% of the total resistance.

Traditional Performance Evaluation

Traditionally the performance has been evaluated by rather simple procedures, where the daily fuel consumption has been reported in the "Noon Report" together with distance traveled over the last 24 hours, the corresponding average observed speed and a single weather observation.

This method gives a limited number of observations since there is a maximum of 365 observations per year. Then the days in port are deducted, together with observations including maneuvering, shallow water and significant changes in speed or heading. This might leaves 200 observations per year, each with only one weather observation which introduces a significant uncertainty, since the weather can change considerably during 24 hours.

SHIP PROPULSION THEORY

Classical ship propulsion procedures can be used as a reference point for the model. In these models the effects of ocean waves have been neglected due to the uncertainty of both the estimate, the added resistance from the waves and measuring the actual wave height and period.

The total resistance thus consists of the still water resistance R_{SW} and the wind resistance R_{wind} .

Resistance

First, the still water resistance is found using the following equation:

$$R_{SW} = \frac{1}{2} C_{tot} \rho_{SW} S U^2 \quad (1)$$

The coefficient C_{tot} is defined as:

$$C_{tot} = C_v + C_R + C_A + C_{AA} \quad (2)$$

where C_v is the viscous resistance coefficient defined as $C_v = C_f (1+k)$. Here the the frictional part C_f is determined from the formula:

$$C_f = \frac{0.075}{(\log_{10}(R_e) - 2)^2} \quad (3)$$

$$R_e = \frac{L_{wl} U}{\nu} \quad (4)$$

where

- R_e is the Reynolds
- k is the form factor that accounts for 3D flow effects around the hull, usually in the region 0-0.25.
- C_R , Residual resistance coefficient is mainly accounting for the energy radiated by waves made by the ship. C_R can be estimated empirically by e.g. Harvald, S. A. (1983) or Holtrop, J. (1984), but the discrepancies can be up to 50% possibly making C_R the most difficult variable to determine in ship resistance.
- C_A , Incremental resistance coefficient, accounts for differences in the hull roughness of the model and the ship. It is usually constant and in the region of $0.15 \cdot 10^{-3}$ to $0.55 \cdot 10^{-3}$
- C_{AA} , Allowance includes air and steering resistance.

Wind Resistance

In almost all conditions the hull and superstructure of the ship will result in a resistance component from the relative wind (the resistance can be negative, in case of strong following winds!). The wind resistance is calculated by

$$R_{wind} = \frac{1}{2} C_x \rho_{air} A_T V_R^2 \quad (5)$$

Where the wind resistance coefficient C_x is determined empirically, by e.g. Isherwood, R. (1972), or by model tests, and vary with the relative wind direction, γ_R . A_T is the transverse projected area above the waterline and V_R is the relative wind speed. The wind coefficient C_x assumes that the wind speed and direction is undisturbed by the ship, which naturally is impossible for the on-board measurements.

Propulsion

The efficiency of the propeller η_D behind the ship is found by combining the results from model tests of the propeller alone, the so called open water test and model tests of the ship, with and without the propeller. If model tests are not available the values, wake fraction, w , thrust deduction, t , and relative rotative efficiency can be estimated by e.g. Harvald, S. A. (1983) or Holtrop, J. (1984). This results in the overall propulsion efficiency η_D .

Using the above described methods with the empirical method by Harvald, S. A. (1983) and Holtrop, J. (1984) the propulsion power has been calculated for each of the measurements described in the section *Test Data Set*. show the results of the two empirical power prediction methods for one of the four conditions measured. It is obvious that the accuracy of both of the empirical methods are insufficient for a reliable performance analysis.

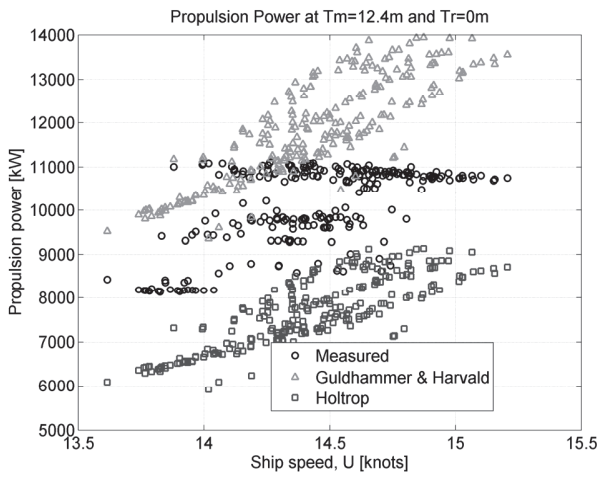


Figure 3: Propulsion calculation by empirical calculations, for data set #1 (Table 1), where T_m is the mean draught and T_r is the trim.

LINEAR AND NON-LINEAR METHOD

Assuming that the ship is sailing in calm and deep water (depth/draught > 8), the propulsion power can thus be written as:

$$P_D = \eta_D^{-1} U (R_{SW} + R_{wind}) \quad (6)$$

A non-linear and a linear method has been developed based on the relation in (6), which can be written as:

$$P_D = \eta_D^{-1} U (K U^2 + L V_R^2) \quad (7)$$

Where:

$$K = C_{SW} \frac{1}{2} \rho_{SW} S \quad (8)$$

$$L = C_X \frac{1}{2} \rho_{air} A_T \quad (9)$$

Is it now possible to adjust the three parameters, η_D^{-1} , K and L , by introducing the additional weights, $\Delta\eta_D^{-1}$, ΔK and ΔL .

$$P_D = (\eta_D^{-1} + \Delta\eta_D^{-1}) U ((K + \Delta K) U^2 + (L + \Delta L) V_R^2) \quad (10)$$

By adjusting the weights a predicted power \hat{P}_D will appear and a minimization of the of the sum of errors between the measured and predicted power gives the optimal set of weights that defines the model. The optimization (minimization) has been done using a “Levenberg-Marquardt” method (Madsen, K.; Nielsen, H.B. & Tingleff, O. (2004) and Nielsen, H. B. (1999)).

The method above is non-linear since $\Delta\eta_D^{-1}$ is represented in the two independent terms ($K + \Delta K$ and $L + \Delta L$). If the overall efficiency is regarded constant ($\Delta\eta_D^{-1} = 0$) a linear model is obtained.

The non-linear model was slightly improved by adding the wind speed squared and the wind direction as additional linear terms. Similarly the linear model was improved by adding the wind direction as a linear term.

Figure 4 show the prediction errors for the linear and non-linear models. The data corresponds to the ones in Figure 3. The summarized cross validation errors for these methods are presented in Table 6.

Both the linear and the non-linear method resulted in a error of approximately 10%, except for data set #4 which is a rather sparse representation of the problem with only 63 data points. The non-linear method showed only slightly better results than the linear method. This rather poor accuracy was the driving factor for discarding these methods in the further development.

The accuracy can be improved by using a “Leave One Out” (LOO) routine for training of the linear and non-linear models and subsequently use the mean of the N weights from the LOO training as the final weights. But in order ensure consistency with the ANN models the data has been split into test and training sets as described in section *Training*.

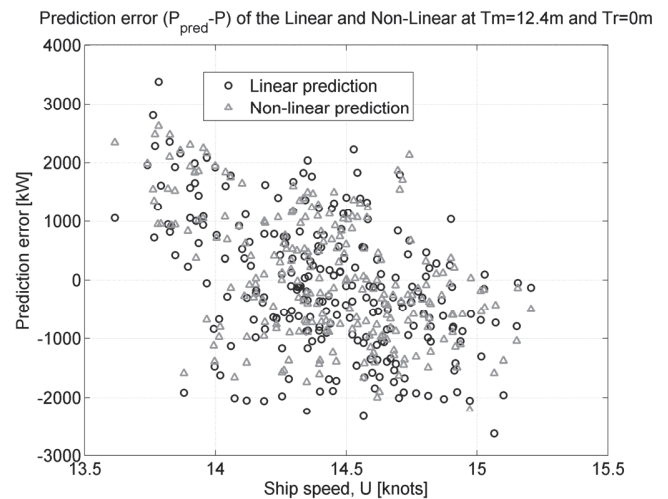


Figure 4: Prediction errors for the linear and non-linear models

ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN) is an advanced form of non-linear regression that can be used to model complex relationships between input and output variables. ANN can be described as linear combinations of nonlinear regression models, with nonlinear basis functions, z_j .

$$y(x) = \sum_{j=0}^M w_{kj}^{(2)} z_{kj} \quad (11)$$

$$z_j = g \left(\sum_{i=0}^d w_{ji}^{(1)} x_i \right) \quad (12)$$

Where:

x are the measured input data.

y are is the output, in this case the propulsion power

The network being used for this problem is a so called one hidden layer (z_1 - z_M). Figure 5 illustrates an equivalent network with multiple output variables, whereas the present method only uses one output (y_1).

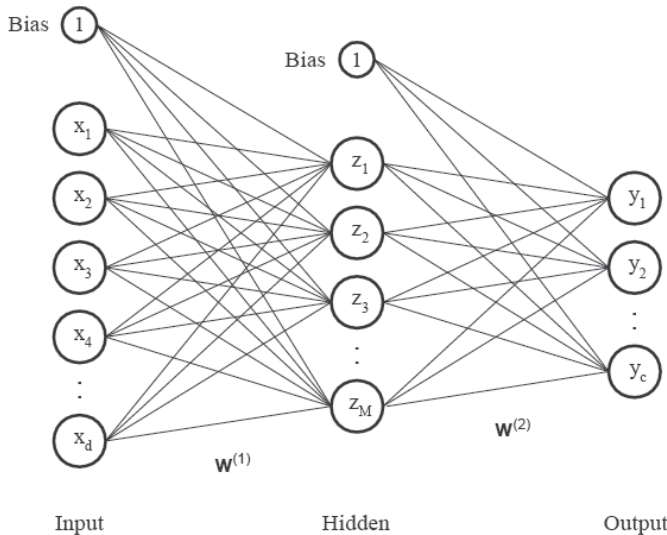


Figure 5: A single hidden layer artificial neural network, with multiple outputs.

The network is a flexible non-linear regression model with additive Gaussian noise and trained with a Bayesian learning scheme. It has a tangent hyperbolic sigmoidal function and is trained using a BFGS (Broyden-Fletcher-Goldfarb-Shanno) optimization algorithm with a soft line search to determine step lengths. The Hessian matrix is evaluated using the Gauss-Newton approximation.

More details in the specific neural network used here is can be found in following references: DTU toolbox (2002), Larsen, J. (1993), MacKay, D. J. C. (1992), Pedersen, M. (1997), Svarer, C.; Hansen, L. & Larsen, J. (1993).

A basic description of neural networks can be found in Bishop, C. M. (2006).

After different attempts of modeling the propulsion power by using the physical and empirical relation, a neural network was tested and immediately showed surprisingly good results. Using a neural network efficiently thus requires sufficient input variables, hidden units, as well as a sufficient amount of data to train with. We do not believe the structural details of the training method are critical to obtain the reported capability.

From the physical relations of the ship propulsion theory the most important variables for the propulsion power, P , can be deducted to: ship speed, U , wind speed, V_R and direction, γ_R , air temperature, T_{air} and seawater temperature, T_{SW} . Consequently the input and output variables are defined as:

$$x = [U \quad V_R \quad \gamma_R \quad T_{air} \quad T_{SW}]$$

$$y = P$$

The relationship between the different variables and the propulsion power is also known to a certain extent, e.g. the power is expected to be proportional to the ship speed cubed.

TEST DATA SET

Four independent data set, with different loading conditions, has been used for the analysis. During each data sampling period factors related to the ship performance including the hull and propeller fouling, were assumed to be constant, consequently this influence will not affect the analysis.

The data has been collected onboard the 110,000 dwt tanker "Torm Marie" where a number of measurement were continuously logged, from where only the relevant data for this problem has been taken. The sampling was split into intervals of 10 minute time series with 10 minutes intervals. The sampling frequency of the times series was 1 second, but many of the measurements had inconsistent signals, i.e., sometime more than 10 second intervals. Power and speed were more or less consistently updated every 13 seconds.

Naturally the recorded data included samples from non-stationary situations as well as situations with zero forward speed. One significant variable to the variations in the samples was the change of heading. Even small changes (less than 1°) of the heading, had significant influence on the measured propulsion power. Samples with excessive variance in the heading have thus been excluded.

The sea state has a significant influence on the ship resistance and hence the propulsion power. No direct measurements of the sea state have been made, but the wind driven waves can be represented by the true wind speed to a certain extent. Making this assumption the swell is not included.

In Table 1 the key figures for each dataset are outlined. It is noted that the ship speed intervals are approximately in the same region for each sample. The distributions of the ship and true wind speed are illustrated in Figure 6-Figure 9. It should be

noted that the Beaufort wind force (BF) 5 starts at approximately 16 knots wind speed. In this condition the wind driven waves are around 2 m high, which is when the sea state starts to influence the added resistance. From Figure 6 and Figure 7 it is noted that only a few occurrences are above this level and thus data sets #1 and #2 can be regarded as calm water conditions. Data set #3 and #4 on the other hand has a more significant contribution of measurements above BF 5 and the added resistance must be regarded as an extra contribution.

Table 1: Trained data sets, where N represents the number of 10 minute recording windows

Data set	Number of Samples	Mean draught, T_m [m]	Trim Ta-Tf [m]	U_{\min} - U_{\max} [knots]	P_{\min} - P_{\max} [kW]
M	N	[m]	[m]	[knots]	[kW]
1	238	12.4	0.0	13.6-15.2	8139- 11111
2	236	7.4	2.4	14.3-16.2	7574- 11283
3	142	7.85	2.7	13.6-15.2	7750-9248.5
4	63	12.15	0.0	13.4-16.1	9764- 11216

The input data are the mean values of the 10 minutes time series. In order to justify this, spreading of the signal has been analyzed, for the ship speed, U , propulsion power, P and apparent wind speed, V_R . The air temperature has been neglected since it is very stable. For every 10 minute period the relative standard deviation, $(\sigma_{x,n}/\mu_{x,n})$ has been found and for every dataset the average of the relative standard deviation, $\bar{\sigma}_M$ has been determined:

$$\bar{\sigma}_{x,M} = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{\sigma_{x,n}}{\mu_{x,n}} \right)^2} \quad (13)$$

Where:

- $\sigma_{x,n}$ is the standard deviation for the n 'th time series
- $\mu_{x,n}$ is the mean value for the n 'th time series
- x indicate the input/output variable (U, P, V_R, γ_R)

Similarly the average of the relative standard deviation $\mu_n, \bar{\mu}_{x,M}$, can be found.

$$\bar{\mu}_{x,M} = \frac{1}{N} \sum_{n=1}^N \left| \frac{\sigma_{x,n}}{\mu_{x,n}} \right| \quad (14)$$

The average of the relative standard deviation $\bar{\mu}_{x,M}$ and the average of the relative standard deviation $\bar{\sigma}_{x,M}$ are shown for every dataset in Table 2 and Table 3. It is noted that both the measured power and ship speed are all less than 1, but for the wind speed there are significant variations.

Table 2: The average of the relative standard deviation

M	N	$\bar{\mu}_U$	$\bar{\mu}_P$	$\bar{\mu}_{V_R}$
1	238	0.6%	0.8%	10.0%
2	236	0.6%	0.7%	18.0%
3	142	0.6%	0.6%	12.4%
4	63	0.6%	1.0%	7.9%

Table 3: The standard deviation of the relative standard deviation

M	N	$\bar{\sigma}_U$	$\bar{\sigma}_P$	$\bar{\sigma}_{V_R}$
1	238	0.2%	0.4%	5.9%
2	236	0.3%	0.2%	13.7%
3	142	0.2%	0.2%	7.2%
4	63	0.3%	0.5%	4.8%

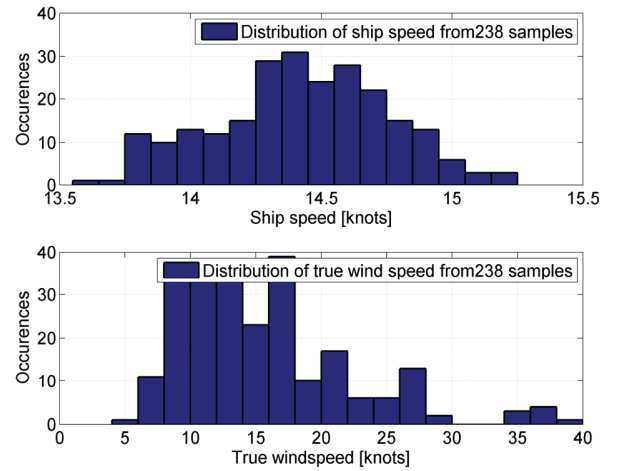


Figure 6: Ship speed and true wind speed distribution of sample #1

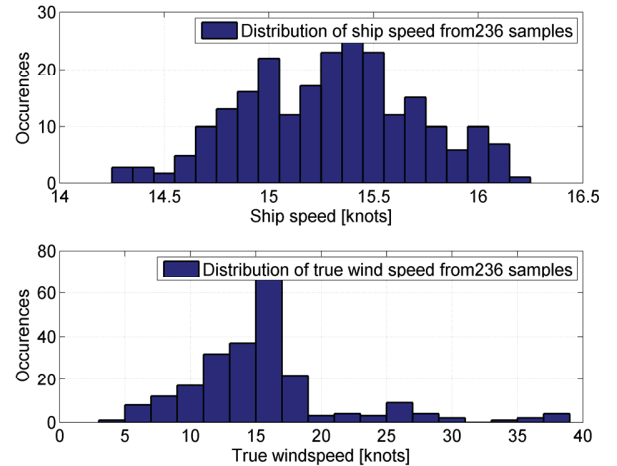


Figure 7: Ship speed and true wind speed distribution of sample #2

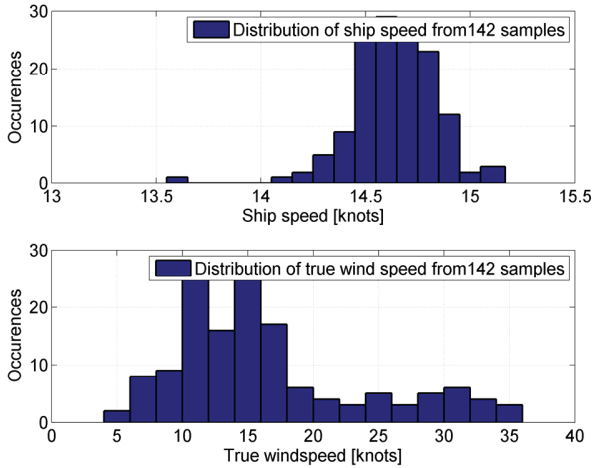


Figure 8: Ship speed and true wind speed distribution of sample #3

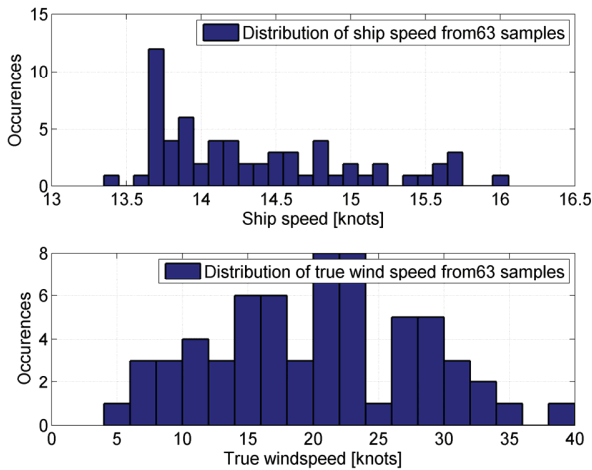


Figure 9: Ship speed and true wind speed distribution of sample #4

TRAINING

The training and test has been performed by a neural network (DTU toolbox (2002) Larsen, J. (1993), MacKay, D. J. C. (1992), Pedersen, M. (1997), Svarer, C.; Hansen, L. & Larsen, J. (1993))

The training procedure has been restarted 10 times in order to ensure that the network found the best possible solution for that particular case.

In order to cross validate, each data test set (1-4 in Table 1) has been divided into 5 training and test subsets, where 20% of the data set has been left only for testing and the remaining part for training. Before the subdivision the data set was permuted randomly.

In order to find the best number of hidden units the network has been trained with respectively 5, 10, 15 and 20 hidden units.

RESULTS

Due to the nature of the input data which is the mean values of the time series of 10 minutes, the resulting network is able to predict the mean propulsion power for a period of 10 minutes.

The results of each network have been evaluated by the relative sum of the errors squared, σ :

$$\tilde{\sigma} = \frac{\sum_{n=1}^N (\hat{P}_{test,n} - P'_{test,n})^2}{\sum_{n=1}^N P'_{test,n}{}^2} \quad (15)$$

The mean of the relative error, ω , has also been found in order to give number of expected accuracy.

$$\omega = \frac{1}{N} \sum_{n=1}^N \left| \frac{\hat{P}_{test,n} - P'_{test,n}}{P'_{test,n}} \right| \quad (16)$$

Where: $\hat{P}_{test,n}$ are the predicted values of the test data

$P'_{test,n}$ are the test samples from the cross validation set

N is number of test set

Every dataset set has been trained with a network with 5, 10, 15 and 20 hidden units. Each of these networks has been trained five times in order to alternately use 20% of the data set for testing. In order to validate the results the cross validation error $\bar{\omega}$ and the cross validation error squared $\bar{\sigma}$:

$$\bar{\omega} = \frac{1}{K} \sum_{k=1}^{K=5} \omega \quad (17)$$

$$\bar{\sigma} = \frac{1}{K} \sum_{k=1}^{K=5} \tilde{\sigma} \quad (18)$$

Where K is the total number training/test set (5).

In Table 4 these two quantities are shown for each of the data sets. It is noted that data set #3 and #4 are much better results than #1 and #2, this is most likely because the limited dataset (142 and 63), are sampled around the same time, and thus have very little variation in the input variables. This is particularly pronounced in Figure 8 where the ship speed has been 14.5-5 knots about 90% of the time.

Taking this into account one should be careful using this network for ship speeds out of this range!

In the error plots of the best solutions, shown in Figure 10-Figure 13, the majority of the predictions are within an error of 500 kW. The prediction error distribution is illustrated in Figure 14-Figure 17, in the same plot a Gaussian distribution (shown as

a blue line) has been generated using the mean value and the variance of the predicted errors. For #1 and #2 the normal distribution fits the histograms very well. For #3 the distribution is skewed due to a few outliers and for #4 the data set is most likely too small to be used, for this purpose, both #3 and #4 have a small spread, thereby justifying their use.

Table 4: Best ANN results and related errors.

M	N	Mean draught	Trim Ta-Tf	No Hidden Units	cross validation error squared	cross validation error
		[m]	[m]		$\overline{\sigma^2}$	$\overline{\omega}$
1	238	12.4	0.0	20	0.13%	2.56%
2	236	7.4	2.4	15	0.15%	2.69%
3	142	7.85	2.7	20	0.03%	0.82%
4	63	12.15	0.0	20	0.04%	1.24%

Furthermore the cross validation error $\overline{\omega}$ and the cross validation error squared $\overline{\sigma^2}$ have been calculated for the two empirical performance evaluation methods, Harvald, S. A. (1983) and Holtrop, J. (1984). The results are shown in Table 5 and as expected these methods gives rather poor results compared with the data driven methods.

Table 5: Cross validation errors for the empirical methods.

M	N	Harvald, S. A. (1983)	Harvald, S. A. (1983)	Holtrop, J. (1984)	Holtrop, J. (1984)
		cross validation error squared	cross validation error	cross validation error squared	cross validation error
		$\overline{\sigma^2}$	$\overline{\omega}$	$\overline{\sigma^2}$	$\overline{\omega}$
1	238	3.63%	17.92%	6.13%	23.74%
2	236	7.47%	26.48%	8.18%	27.78%
3	142	5.68%	22.35%	7.93%	27.41%
4	63	10.75%	23.14%	9.70%	28.14%

Table 6: Cross validation errors for the linear and non-linear models

M	N	Linear	Linear	Non-linear	Non-linear
		cross validation error squared	cross validation error	cross validation error squared	cross validation error
		$\overline{\sigma^2}$	$\overline{\omega}$	$\overline{\sigma^2}$	$\overline{\omega}$
1	238	1.47%	10.09%	1.40%	10.06%
2	236	1.67%	10.66%	2.49%	9.54%
3	142	1.94%	11.39%	1.87%	11.14%
4	63	0.21%	3.46%	0.07%	1.98%

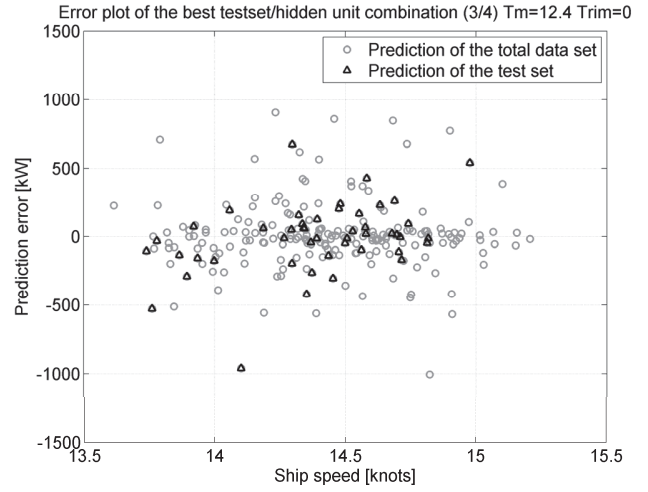


Figure 10: Prediction errors for sample #1

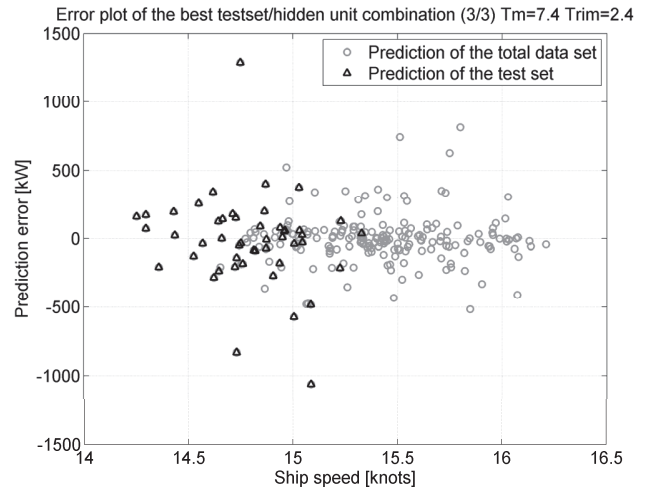


Figure 11: Prediction errors for sample #2

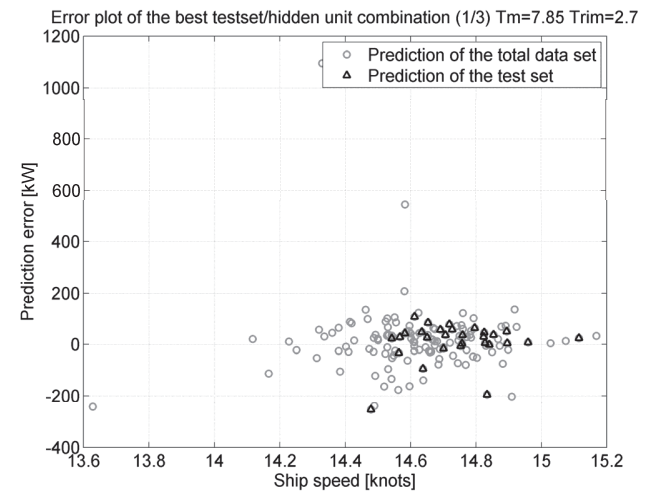


Figure 12: Prediction errors for sample #3

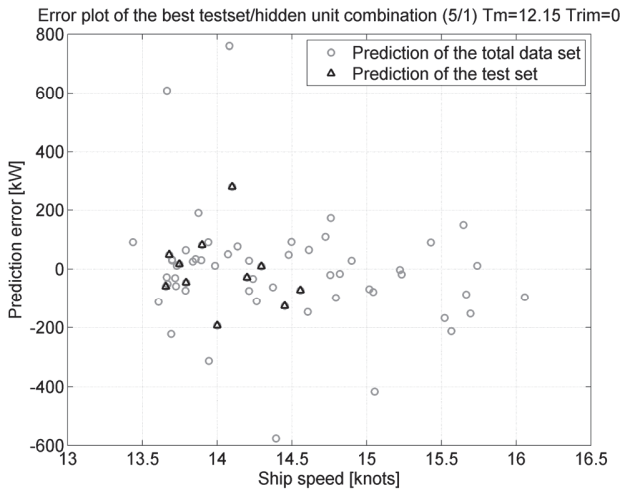


Figure 13: Prediction errors for sample #4

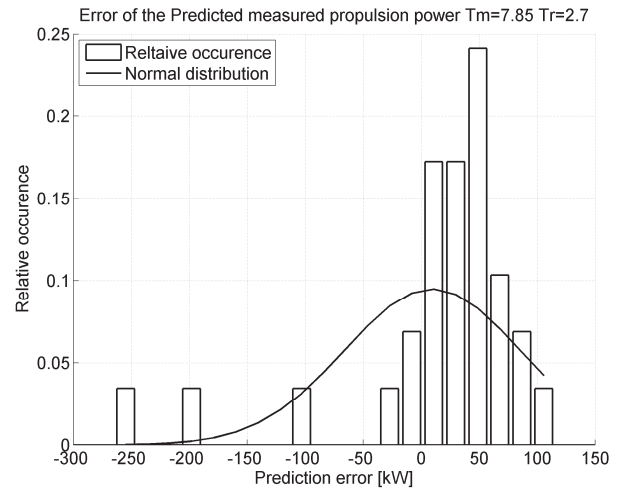


Figure 16: Relative distribution of the predicted errors for sample #3

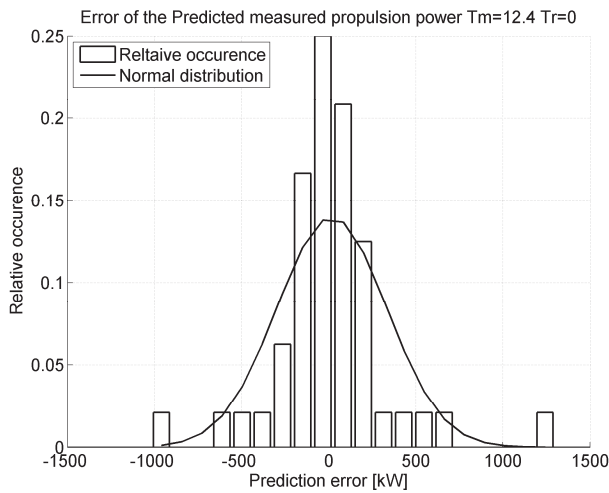


Figure 14: Relative distribution of the predicted errors for sample #1

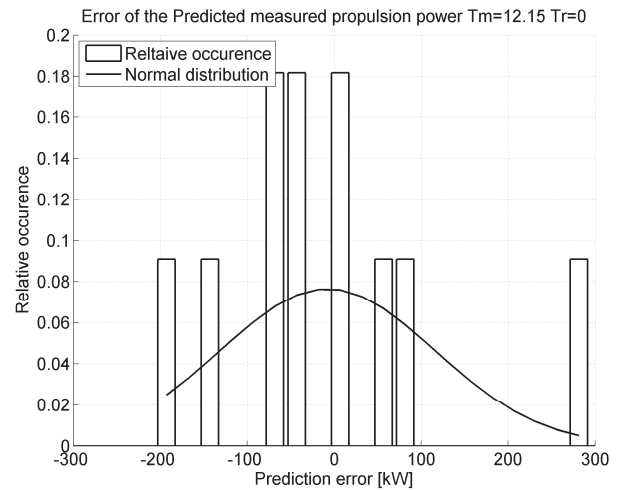


Figure 17: Relative distribution of the predicted errors for sample #4. T_m is the mean draught and T_r is the trim.

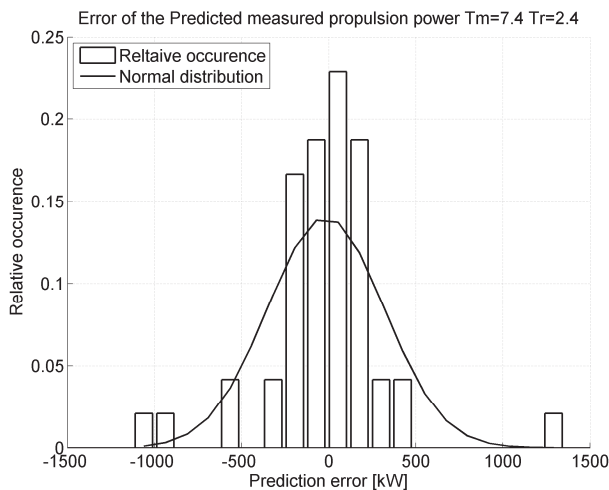


Figure 15: Relative distribution of the predicted errors for sample #2

CONCLUSIONS

It is possible to predict the mean propulsion power during a period of 10 minutes with a relative error of less than 2.7%, using a single hidden layer neural network. This is significantly better than empirical or data-driven methods based on hydro-mechanical relationships.

The prediction was carried out for four different states with the following input variables: ship speed, relative wind speed and direction, air temperature and sea water temperature. It should be emphasized that each of the four different trained neural networks are trained for the specific ship, time and condition only. The network parameters can thus not be used or extrapolated for any other ship or condition. Similarly the variables in the prediction should be within boundaries of the trained variables.

It is believed that the prediction is not critical to the specific structure of the neural network model, provided that proper training is performed.

The current neural network model can be used to determine a performance benchmark of the ship for the specific condition, which will be valuable for the performance evaluation or make easier to compare performance of sister ships.

Although the sea state is not measured directly, the ANN model using the wind speed as a measure of the sea state seems to give reasonable results. It should be noted that most of the wind speed measurements are below ~ 16 *knot* and thus the effect very of the waves is minimal.

As more data is collected onboard the vessel the model will gradually be extended and further work will focus implementing the trim and draught as variables which will lead to a model that predict the propulsion power in any realistic situation. This can e.g. be used to evaluate the hull and propeller performance or determine what trim is the most efficient in a certain condition.

ACKNOWLEDGEMENTS

Many thanks to the ship owner Torm and the crew onboard Torm Marie for allowing me to install the equipment and providing noon report data. In particular Chief Engineer Rasmus Hoffman, who has showed great interest in the project and been very helpful monitoring the onboard system and sending data home.

The work presented in this paper has been carried out during Pedersen's PhD study at FORCE Technology and the Technical University of Denmark, which is sponsored by The Danish Industrial PhD programme and Danish Centre of Maritime Technology (DCMT)/The Danish Maritime Fund.

REFERENCES

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer.

FORCE (2008), SeaTrend® info sheet is available at: http://www.force.dk/en/Menu/Products+and+Concepts/Products/080220_seatrend.htm

Harvald, S. A. (1983), *Resistance and Propulsion of Ships*, John Wiley & Sons.

Holtrop, J. (1984), 'A Statistical re-analysis of resistance and propulsion data', *International Shipbuilding Progress* **31**, 272-276.

Isherwood, R. (1972), 'Wind resistance of merchant ships', *Royal Institute of Naval Architecture*.

ITTC (1978), 'ITTC – Recommended Procedures - Performance, Propulsion 1978 ITTC Performance Prediction Method', ITTC, 7.5-0203-01.4.

DTU toolbox (2002), DTU toolbox: Neural regressor with quadratic cost function, <http://isp.imm.dtu.dk/toolbox>

Larsen, J. (1993), 'Design of Neural Network Filters', PhD thesis, Electronics Institute, Technical University of Denmark.

MacKay, D. J. C. (1992), 'A practical Bayesian framework for backpropagation networks', *Neural Computation* **4**, 448-472.

Madsen, K.; Nielsen, H.B. & Tingleff, O. (2004), 'Methods for Non-Linear Least Squares Problems, 2nd Edition' IMM, DTU

Nielsen, H. B. (1999), 'Damping Parameter in Marquardt's Method' Report IMM-REP-1999-05, IMM, DTU . 31 pages.

Pedersen, M. (1997), 'Optimization of Recurrent Neural Networks for Time Series Modeling', PhD thesis, Institute of Mathematical Modeling, Technical University of Denmark.

Svarer, C.; Hansen, L. & Larsen, J. (1993), On Design and Evaluation of Tapped-Delay Neural Network Architectures, in 'Proceedings of the 1993 IEEE International Conference on Neural Networks', pp. 46-51.

Appendix B

Prediction of Full Scale Propulsion Power Using Artificial Neural Networks

Presented at the "Conference on Computer Applications and Information Technology in the Maritime Industries", *COMPIT*, 10-12 May 2009 in Budapest, Hungary.

Full scale measurements of the propulsion power, ship speed, wind speed and direction, sea and air temperature from four different loading conditions, together with hindcast data of wind and sea properties; and noon report data has been used to train an Artificial Neural Network for prediction of propulsion power. The model was optimized using a double cross validation procedure. The network was able to predict the propulsion power with accuracy between 0.8-1.7% using on-board measurement system data and 7% from manually acquired noon reports.

Prediction of Full Scale Propulsion Power Using Artificial Neural Networks

Benjamin Pjedsted Pedersen, FORCE Technology / Department of Mechanical Engineering, Technical University of Denmark, bpp@force.dk

Jan Larsen, Department of Informatics and Mathematical Modeling, Technical University of Denmark, jl@imm.dtu.dk

Abstract

Full scale measurements of the propulsion power, ship speed, wind speed and direction, sea and air temperature from four different loading conditions, together with hindcast data of wind and sea properties; and noon report data has been used to train an Artificial Neural Network for prediction of propulsion power. The model was optimized using a double cross validation procedure. The network was able to predict the propulsion power with accuracy between 0.8-1.7% using onboard measurement system data and 7% from manually acquired noon reports.

1 Introduction

As part of the Industrial PhD project "Ship Performance Monitoring" automatic data sampling equipment was installed on the tanker "Torm Marie" in January 2008 and presently data from four different loading conditions are available.

By considering the ship as a dynamical system which can be modelled as a general nonlinear state-space model, the ship propulsion performance (referred to as the performance) is a measure of energy consumption which depends on the current *state* of the ship and a large number of external factors/variables such as speed, loading conditions, ship conditions, weather and sea conditions. Figure 1 shows some factors influencing propulsion performance.

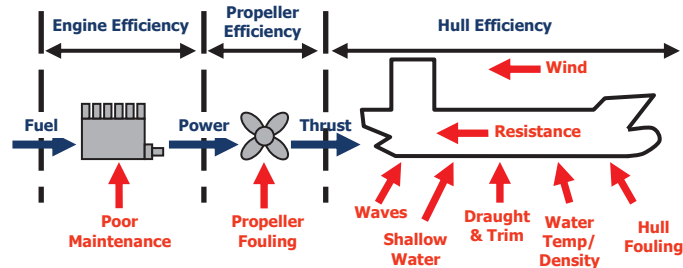


Figure 1: Variables influencing propulsion performance.

The variables have different properties. Some of the variables are observable (and measurable with high reliability) whereas others are difficult to observe, e.g. the fouling. Some variables are largely controllable, whereas others are almost uncontrollable. For instance, heading is controllable whereas wind conditions and fouling are almost uncontrollable. The first goal in performance measurement is to provide a reliable estimation of performance as a function of the state and external variables. The second goal is to optimize performance by manipulating the controllable variables. This paper will focus on the first goal.

During the lifetime of the ship the performance will decrease. As an example the fuel consumption will increase at a certain state, or the ship speed will decrease at a certain power setting. This is mainly due to fouling of the hull and propeller. A typical trend of the speed reduction is illustrated in figure 2.

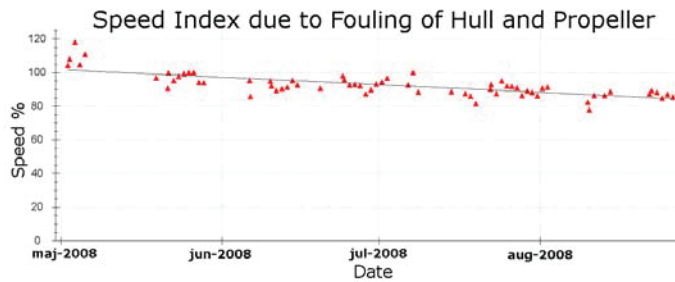


Figure 2: Increase in fuel consumption as consequence of fouling

This work will not consider a full dynamical model of the ship but merely focus on a model which predicts propulsion power in a specific state based on the measurements of a set of significant input variables, which are:

- Ship speed through the water
- Wind speed and direction
- Seawater temperature
- Air temperature
- Water depth
- Wave height and direction

The model can be based on a classical physical/empirical model, e.g. *Harvald, S. A. (1983)* or *Holtrop, J. (1984)* or a datadriven (non-parametric) approach e.g. an artificial neural network. Previous work suggests that a datadriven approach is preferable (see e.g. *Pedersen & Larsen 2009*),

The empirical methods are derived from model tests and sea trials, and since most model tests are carried out in the design condition (even keel) and speed, this is the region where it should be applied. In operation the ship will travel in many other conditions i.e., ballast draught and trimmed conditions. Consequently, these methods give a rough estimate of the propulsion power rather than an accurate reference point. If measured values from model tests or sea trials are available, they can be used to adjust the empirical data and thus give a more accurate result. Figure 3 shows the measured power together with estimated power using respectively *Harvald, S. A. (1983)* and *Holtrop, J. (1984)*, with a standard setup i.e., without any adjustment. It is obvious that a change in a few percent, which is realistic performance deterioration over a year, is impossible to detect.

Furthermore the traditional methods are based on “Noon Reports” data, which are reports containing information of the ship speed, travelled distance, position, heading, and a number of other measurements and readings. One problem with noon reports are that only one *sample* is collected per day, excluding days in harbour and e.g. travelling in areas with limited water depth. This might leave out 200 observations per year. Many noon reports data are mean values over time from the last noon report, e.g. average logged ship speed, and others are observations at the report time, e.g. current wind speed. This makes it difficult to analyse relations e.g., between the average ship speed and the instantaneous wind speed.

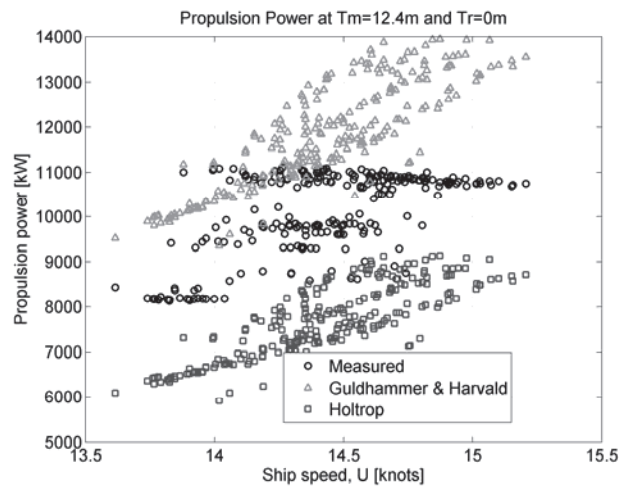


Figure 3: An example of empirical power prediction using the *Harvald, S. A. (1983)* and *Holtrop, J. (1984)* methods, compared with measure values for a single loading condition (Mean draught: 12.0m, even keel).

If sufficient data is available it is possible to make a partial or fully data driven model. The first step of such a model is to capture the dynamics of the fastest changing variables, e.g. ship speed, wind speed, etc. where the slowest changing variables are draught and trim. Initially, this leads to a prediction model for data sets where draught and trim are kept constant.

Adequate quality of input data is fundamental to get reliable results from the prediction model. This is a problem for sea and weather information data due to the difficulty of measuring these quantities, but especially to noon report data which are collected manually hence human factors and errors can play a significant role.

2 Data sources

Three data input sources have been used to train and predict the propulsive power:

1. Onboard measured data (4 conditions)
2. Noon report data
3. Hindcast weather and sea information

An overview of all the relevant data set variables is listed in Table 4.

2.1.1 Onboard measured data

The data was collected onboard the 110,000 dwt tanker “Torm Marie” where a number of measurements were continuously logged. Only the relevant data for this problem has been taken. The sampling was split into time series of 10 minutes with 10 minute intervals. The sampling frequency of the time series was 1 second, but many of the measurements had inconsistent and missing signal values. Power and speed were updated consistently, every 13 seconds. Four independent data sets, with different loading conditions have been sampled so far.

The data include samples from non-stationary situations as well as situations with zero forward speed, which are deleted. The variance of the heading is one of the governing figures on the variation of the propulsion power in particular. Even small changes (less than 1°) in the heading, had significant influence on the measured propulsion power. Samples with excessive variance in the heading have thus been excluded.

During each data sampling period factors related to the ship performance including the hull and propeller fouling, were assumed to be constant, consequently this effect will not be accounted for in the analysis

2.1.2 Noon report data

The noon reports contain a long array of data and basically the same variables as the measured ones, but with differences in quality and resolution. Due to their nature, noon reports are usually only collected once a day, which gives a smaller resolution and a mix of data with different origins, e.g. logged average speed over ~24 hours and one weather observation at the report time. Noon reports are usually filled in manually and are thus also subject to human factors and errors.

In this analysis the noon reports are important for obtaining the draught and seawater temperature.

2.1.3 Hindcast data

Hindcast data has been received from a tool developed for SeaTrend®¹ at FORCE Technology based on weather information from NOAAH². For a given position and time this tool returns wind speed and direction, significant wave height, peak period and direction. Some areas, e.g. the Mediterranean are not included in this database.

2.2 Dataset for training and test

Two different configurations of the dataset were used for the analysis. One based on the measured values and one based on noon report data.

2.2.1 Onboard measured dataset

The dataset based on measured values has a high density of data (approx. 72 per day), but there is only a limited amount of this data available: in total around 27 days (see Table 1).

Table 1: Onboard measured data sets. N represents the number of 10 minute recording windows.

Data set	Number of Samples	Start date End date	Number of valid noon reports	Mean draught, T_m	Trim Ta-Tf	U_{\min} - U_{\max}	P_{\min} - P_{\max}
M	N			[m]	[m]	[knots]	[kW]
1	236	09-02-2008 14-02-2008	3	7.4	2.4	14.2- 16.2	7573- 11283
2	109	22-03-2008 27-03-2008	4	7.85	2.7	13.6- 15.1	7750- 9248
3	301	30-01-2008 06-02-2008	7	12.15	0	13.4- 16.0	8138- 11216
4	555	01-03-2008 11-03-2008	9	13.0	0	13.0- 15.9	9741- 12096

All the measured input data are the mean values over 10 minutes of the time series. In order to justify this, variance of the signal has been analyzed for the ship speed, U , propulsion power, P and apparent wind speed, V_R . The air temperature has been neglected since it is very stable. For every 10 minute period the relative standard deviation, $(\sigma_{x,n}/\mu_{x,n})$ has been found and for every dataset the average of the relative standard deviation $\bar{\mu}_{x,M}$, has been determined, see Eq. (1).

$$\bar{\mu}_{x,M} = \frac{1}{N} \sum_{n=1}^N \left| \frac{\sigma_{x,n}}{\mu_{x,n}} \right|, \quad (1)$$

where,

$\sigma_{x,n}$ is the standard deviation for the n 'th time series, $\mu_{x,n}$ is the mean value for the n 'th time series, and x indicates the input/output variables (U, P, V_R, γ_R)

¹ Performance Monitoring tool developed at FORCE Technology, www.force.dk

² National Oceanic and Atmospheric Administration, United States Department of Commerce
<http://www.noaa.gov/>

It should be noted that the relative standard deviation of both the measured power and ship speeds are all less than 1, whereas the wind speed has a significantly high variance.

Table 2: The average of the relative standard deviation

M	N	$\bar{\mu}_U$	$\bar{\mu}_P$	$\bar{\mu}_{V_R}$
1	236	0.6%	0.7%	18.0%
2	109	0.6%	0.5%	9.1%
3	301	0.6%	0.9%	9.5%
4	555	0.6%	0.6%	11.4%

From Table 2 it is noted that the ship speed intervals are approximately in the same region for each sample. However inspecting the distributions of the ship and true wind speed illustrated in Figure 4-7, it is noted that the actually ship speed range is different. Especially for dataset #2 where most of the ship speeds is in a band of around 14.7 knots.

It should be noted that the Beaufort wind force (BF) 5 starts at approximately 16 knots wind speed. In this condition the wind driven waves are around 2 m high, which is when the sea state starts to influence the power increase in waves. From Figure 4 and 5 it is noted that only a few occurrences are above this level and thus datasets #1 and #2 can be regarded as calm water conditions. Datasets #3 and #4 on the other hand have a more significant contribution of measurements above BF 5 and the power increase in waves must be regarded as an extra contribution.

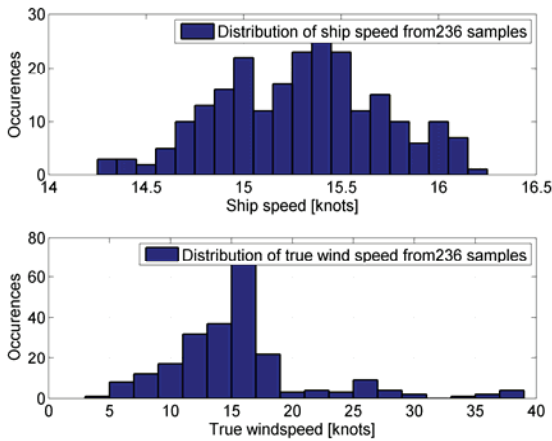


Figure 4: Ship speed and true wind speed distribution of dataset #1

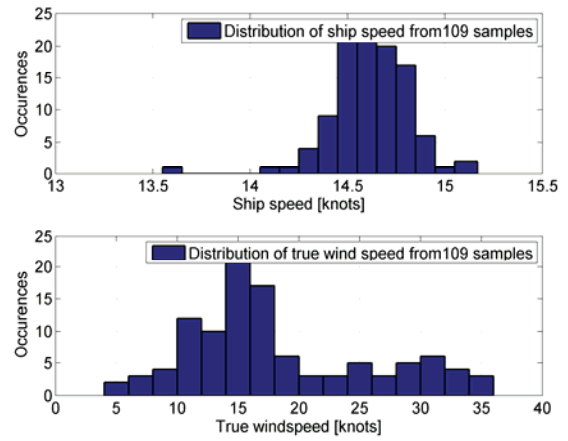


Figure 5: Ship speed and true wind speed distribution of dataset #2

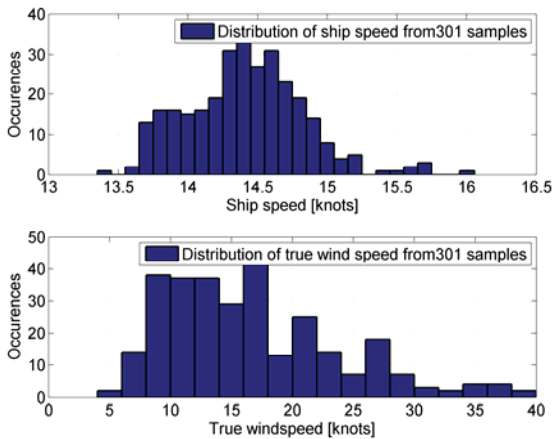


Figure 6: Ship speed and true wind speed distribution of dataset #3

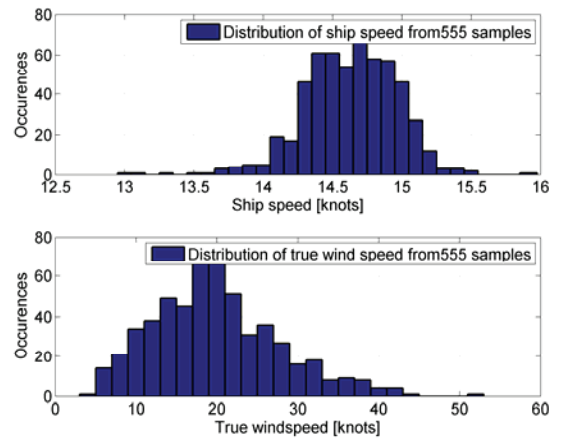


Figure 7: Ship speed and true wind speed distribution of dataset #4

The sea state has a significant influence on the ship resistance and hence the propulsion power. No direct measurements of the sea state have been made, but the wind driven waves can be represented by the true wind speed to a certain extent. Making this assumption the swell is not accounted for.

Hindcast information gives an estimate of the sea state, including significant wave heights, peak period and direction, at the specific position and time, and has been found for all the relevant data. Furthermore the hindcasts also give the true wind speed and direction.

2.2.2 Noon report dataset

The time density of the dataset based on noon reports is much less than for the measure based dataset. There is a maximum of one sample per day and many are invalid due to e.g. anchoring, alongside in harbour. But the time span is much longer, approximately 2 years and the variation in draught and trim has to be taken into account, and possibly also the time (see Table 3).

In order to give a more representative value of the sea state and wind condition for the noon report data, hindcast has been made for every hour in between each noon report. Afterwards the mean value and variance of the time series (approximately 24 hours) prior to the report time, has been found and are thus ready to use for the analysis.

Table 3: Noon report dataset for analysis

Date UTC	Number of valid samples	Mean draught [m]	Trim, Ta-Tf [m]	Ship speed [knots]	Seawater temp [°C]	Specific HFO [tons/day]
09-12-2006 - 05-12-2008	323	7.35-15.35	0-3.4	9.9-17.5	12-32	1.1-3.9

Table 4: Propulsion performance variables

Data		Unit	Data source
Speed through water	U	[knots]	Measured onboard
Relative wind velocity	V_{rel}	[knots]	Measured onboard
Relative wind direction	g_{rel}	[deg]	Measured onboard
Air temperature	T_{air}	[degC]	Measured onboard
Propulsion power	P	[kW]	Measured onboard
Logged mean speed	$NR.U$	[knots]	Noon report
Sea water temperature	$NR.T_{sw}$	[degC]	Noon report
Air temperature	$NR.T_{air}$	[degC]	Noon report
Arrival draught fore	T_f	[m]	Noon report
Arrival draught aft	T_a	[m]	Noon report
Specific fuel consumption	$SpHFO$	[ton/hour]	Noon report
Report time, UTC	$NR.UTC$	[hh:mm:ss]	Noon report
True wind speed	$HC.W_s$	[m/s]	Hindcast
True wind direction	$HC.g$	[deg]	Hindcast
Significant wave height	$HC.H_s$	[m]	Hindcast
Wave period	$HC.T_p$	[s]	Hindcast
True wave direction	$HC.T_d$	[deg]	Hindcast
Mean arrival draught	T_m	[m]	Derived from noon reports
Arrival trim, T_a-T_f	$Trim$	[m]	Derived from noon reports
Relative wind speed	$HC.V_{rel}$	[knots]	Derived from hindcasts
Relative wind direction	$HC.g_{rel}$	[deg]	Derived from hindcasts

3 Regression models for propulsion power prediction

Three different regression models have been tested and evaluated: a linear model, a (custom) non-linear model and a Artificial Neural Network model

3.1 Linear and Non-linear models

Both a linear and non-linear method based on the general assumption of relation between the ship speed, wind speed and power was developed and presented in *Pedersen, B.P. and Larsen J. (2009)*. In short the methods are based on the relation presented (2) which can be developed to the form presented in (3) and (4) where $\Delta\eta_D^{-1}$, ΔK and ΔL . are adjustable parameters that are optimized using a ‘‘Levenberg-Marquardt’’ (*Madsen, K.; Nielsen, H.B. & Tingleff, O. (2004) and Nielsen, H. B. (1999)*) optimization routine. If $\Delta\eta_D^{-1}$ is zero the model is regarded as linear.

$$P_D = \eta_D^{-1} U (R_{SW} + R_{wind}) \quad (2)$$

$$P_D = \eta_D^{-1} U (K U^2 + L V_R^2) \quad (3)$$

Is it now possible to adjust the three parameters, η_D^{-1} K and L , by introducing the additional weights, $\Delta\eta_D^{-1}$, ΔK and ΔL .

$$P_D = (\eta_D^{-1} + \Delta\eta_D^{-1}) U ((K + \Delta K) U^2 + (L + \Delta L) V_R^2), \quad (4)$$

where,

$$K = C_{SW} \frac{1}{2} \rho_{SW} S \quad (5)$$

$$L = C_X \frac{1}{2} \rho_{air} A_T \quad (6)$$

Both the linear and non-linear methods resulted in a cross validation error (9) of 3-12%. This could be improved a bit by using a “Leave One Out” (LOO) routine for training of the linear and non-linear models and by subsequently using the mean of the N weights from the LOO training as the final weights. But in order to ensure consistency with the ANN models the data has been split into test and training sets as described in section 5.

3.2 Artificial Neural Network (ANN)

After a brief test of regression with an ANN this method appeared superior to the previously described methods which lead to a thorough exploration of the ANN methods. An ANN is a non-linear method where the so called hidden layer with hidden units is the non-linear link between input and output, as illustrated in Figure (shown with multiple outputs) or described in (7) and (8).

$$y(x) = \sum_{j=0}^M w_{kj}^{(2)} z_j, \quad (7)$$

$$z_j = g \left(\sum_{i=0}^d w_{ji}^{(1)} x_i \right), \quad (8)$$

where,

x is the measured input data.

y is the output, in this case the propulsion power.

z_j are the nonlinear basis functions.

w are the weights for the hidden units and output.

The network used for this analysis is a flexible non-linear regression model with additive Gaussian noise and is trained with a Bayesian learning scheme. It has a tangent hyperbolic sigmoidal function and is trained using a BFGS (Broyden-Fletcher-Goldfarb-Shanno) optimization algorithm with a soft line search to determine step lengths. The Hessian matrix is evaluated using the Gauss-Newton approximation.

More details into the specific neural network used here can be found in the following references: *DTU toolbox (2002)*, *Larsen, J. (1993)*, *MacKay, D. J. C. (1992)*, *Pedersen, M. (1997)*, *Svarer, C.; Hansen, L. & Larsen, J. (1993)*. A basic description of neural networks can be found in *Bishop, C. M. (2006)*.

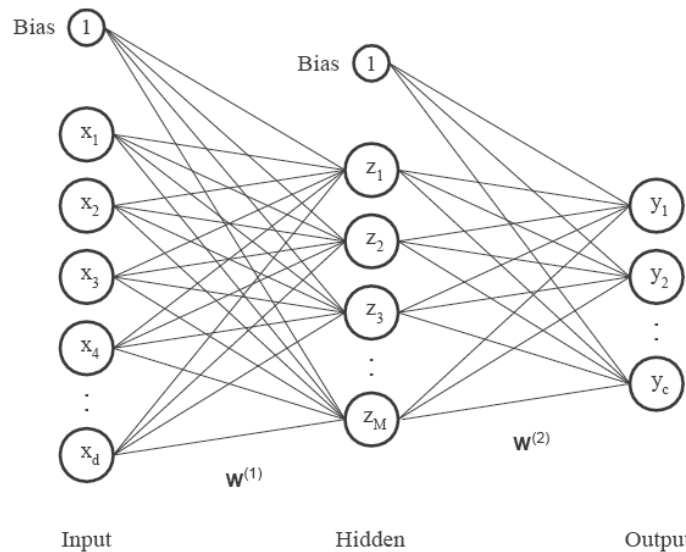


Figure 8: A single hidden layer artificial neural network, with multiple outputs.

2	x		x	x	x	x						x	x	x
3	x		x	x	x	x	x	x	x			x	x	x
4	x		x	x			x	x	x			x	x	x
5*	x		x	x	x	x	x	x	x			x	x	x
6	x		x	x						x	x	x	x	x
7	x	x	x	x						x	x	x	x	x
8	x		x	x			x	x	x	x	x	x	x	x
9	x	x	x	x			x	x	x	x	x	x	x	x

* The variance of HC.Ws, HC.g, HC.Hs, HC.Tp and HC.Td over the steaming time has also been included as input variables.

5 Evaluation

In order to make a consistent evaluation of the ANN training and testing two cross validation errors have been introduced. One for the *inner-loop* of the double cross validation, testing on the validation set, $\bar{\omega}_K$ (9) and one for the *outer-loop* of the double cross validation, testing on the test set, $\bar{\omega}_M$.

$$\bar{\omega}_K = \frac{1}{K} \sum_{k=1}^K \omega_k \quad (9)$$

Where K is the number of cross validation set for the *inner-loop* and ω_k is the mean of the relative error for each of the cross validation sets (10).

$$\omega_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \left| \frac{\hat{P}_{test,n} - P'_{test,n}}{P'_{test,n}} \right|, \quad (10)$$

where $\hat{P}_{test,n}$ are the predicted values of the test data, $P'_{test,n}$ are the test samples from the validation set, and N is number of test data.

The *outer-loop* cross validation $\bar{\omega}_M$ error is equivalent to $\bar{\omega}_K$ (9) except that the test set has been used for the mean relative error opposed to the validation set.

6 Results

6.1 Results from measured input/output data

Due to lack of computation time the ANN was only trained for 5 and 20 hidden as these are the extremes. This can be justified by *Pedersen, B.P and Larsen, J (2009)*, where training/test were performed with 5,10,15 and 20 hidden units, it was concluded that the number of hidden units are not critical to the solution, although in general 5 hidden units were too little.

Table 8 shows the cross validation errors of the *inner-loop* $\bar{\omega}_K$ for the input variable combinations defined in Table 6 and 5 and 20 hidden units.

Looking at all the cross validation errors for each of the datasets in Table 8 it is clear that some datasets in general have smaller errors. Particularly in dataset #2 and to some extent #4 it is noted that the cross validation errors do not vary no matter what input data variables or number of hidden units are used. In these datasets it is thus difficult to detect what input variables have the most influence on the solutions.

The cross validation errors of the *outer-loop* are presented in Table 9, together with respectively the best combination of hidden units and input variable combinations. Datasets #2 and #4 have rather low cross validation errors, which must be due to the nature of the dataset. What is more interesting is to see how the error drops by introducing hindcast sea state information and the best solutions in general are where only the hindcast information has been used for the sea and wind property inputs.

Table 8: Table of *inner-loop* cross validation errors, $\bar{\omega}_K$

Dataset #1 Dataset #2 Dataset #3 Dataset #4

Number of hidden units	5	20	5	20	5	20	5	20
Input variable combination	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$
2	3.93%	2.92%	1.07%	0.81%	3.07%	2.37%	1.72%	1.30%
8	2.63%	1.97%	0.97%	0.89%	2.21%	1.65%	1.49%	1.04%
10	2.14%	1.65%	0.99%	0.95%	2.17%	1.65%	1.52%	0.94%
11	3.77%	2.79%	1.10%	0.90%	2.45%	1.88%	1.42%	1.02%
12	2.25%	1.65%	0.99%	0.94%	1.75%	1.40%	1.28%	0.90%

Table 9: Table of *outer-loop* cross validation errors, $\bar{\omega}_M$

	Dataset #1	Dataset #2	Dataset #3	Dataset #4
Optimum number of hidden units	20	20	20	20
Optimum input variable combination	10 / 12	2	12	12
$\bar{\omega}_M$	1.63%/1.74%	0.83%	1.46%	0.80%

Figures 9 - 16 show the test errors for the best combination of the number of hidden units and input variables. The plots on the left show the test errors as a function of the ship speed; it is noted that there is no apparent correlation between ship speed and the error, which indicates that the ship speed has integrated properly into the model.

The plots on the right are a relative histogram of the errors together with a normal distribution based on the mean and the variance of the test errors. Except for dataset #2 all the error distributions are centered approximately around 0 and have a nice distribution. Dataset #2 is a sparser dataset so each bar represents 1-4 counts, but the errors are relatively small.

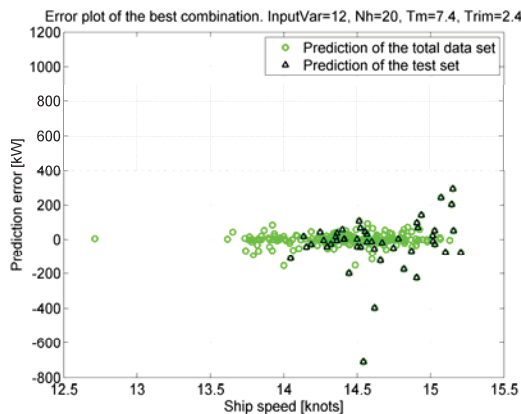


Figure 9: Prediction errors for dataset #1

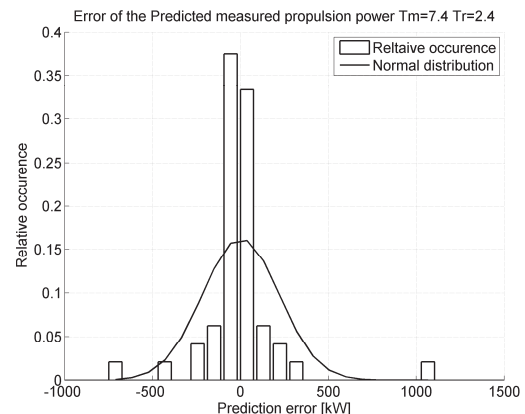


Figure 10: Relative distribution of the predicted errors for dataset #1

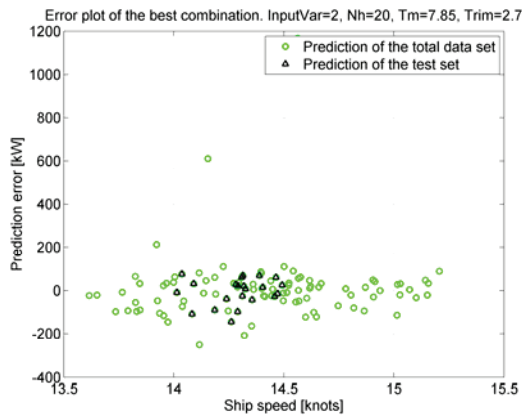


Figure 11: Prediction errors for dataset #2

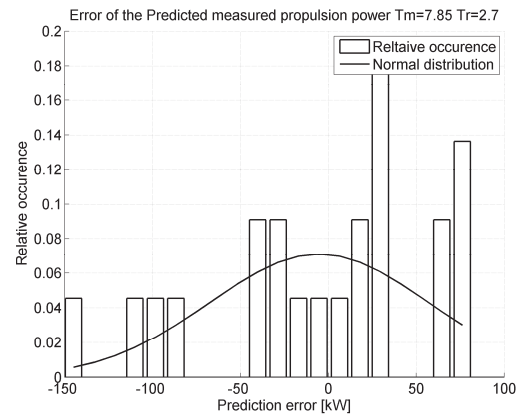


Figure 12: Relative distribution of the predicted errors for dataset #2

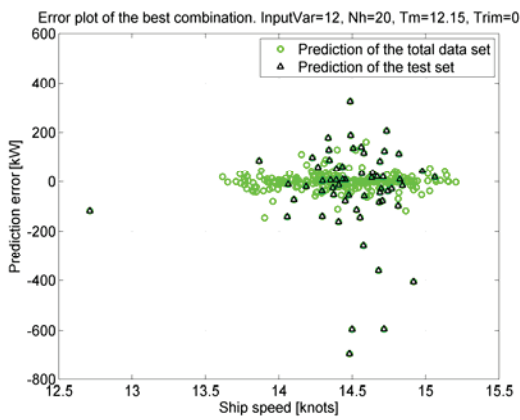


Figure 13: Prediction errors for dataset #3

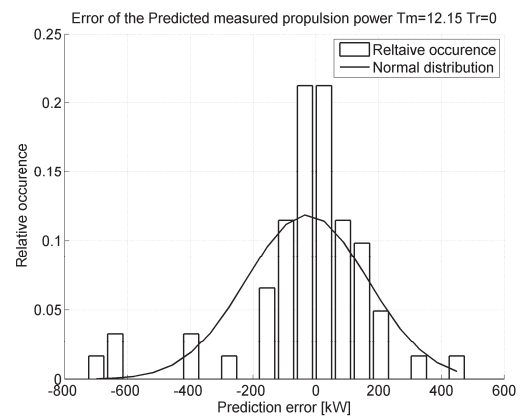


Figure 14: Relative distribution of the predicted errors for dataset #3

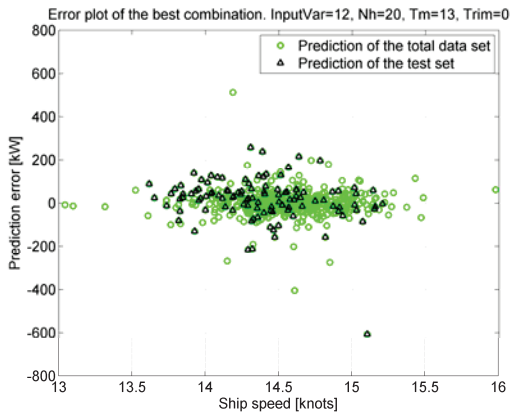


Figure 15: Prediction errors for dataset #4

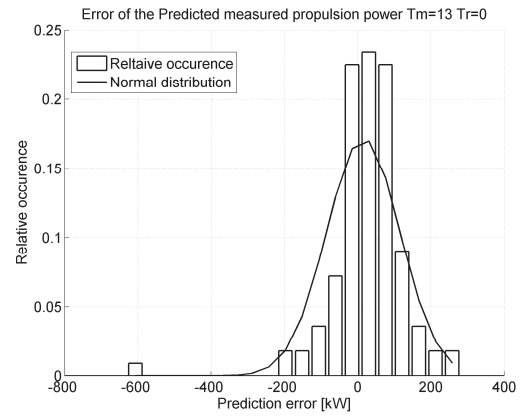


Figure 16: Relative distribution of the predicted errors for dataset #4

6.1.1 Results from the linear and non-linear models

Training and testing was performed as described above using a similar cross validation although not double since the input variables are specified by the model.

Table 10: Table of cross validation errors, $\bar{\omega}$ for the dataset based on measurement using a linear and non-linear method.

	Cross validation error	Dataset #1	Dataset #2	Dataset #3	Dataset #4
Linear method	$\bar{\omega}$	11.58%	3.63%	12.04%	5.98%
Non-linear method	$\bar{\omega}$	11.36%	3.58%	10.79%	5.98%

6.2 Results from noon report data

From the *inner-loop* cross validation errors listed in Table 10 it is noted that the model in general is less sensitive to the number of hidden units. The dependency on certain variables seems not very strong since most errors are in the same region. It is interesting to see the error drop significantly when the time is introduced as a variable for input variable combination 7 and 9. Furthermore combination 5 might have been over trained since it has the highest number of input variables but one of the highest errors.

The *outer-loop* cross validation error is presented in Table 12 and is based on the input variable combination 7 which does not even take into account the sea state.

Table 11: Table of *inner-loop* cross validation errors, $\bar{\omega}_m$ for the noon report dataset

Number of hidden units	5	10	15	20
Input variable combination	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$	$\bar{\omega}_K$
2	9.05%	9.57%	8.91%	9.13%
3	8.84%	9.91%	10.26%	10.82%
4	8.70%	10.00%	11.22%	13.78%
5*	9.79%	11.13%	8.95%	8.38%
6	8.18%	8.50%	9.07%	9.05%
7	7.18%	7.74%	9.28%	10.06%
8	8.46%	9.06%	9.60%	9.18%
9	7.26%	8.47%	10.58%	9.25%

Table 12: Table of *outer-loop* cross validation errors, $\bar{\omega}_M$ for the noon report dataset

Optimum number of hidden units	5
Optimum input variable combination	7
$\bar{\omega}_M$	7.02%

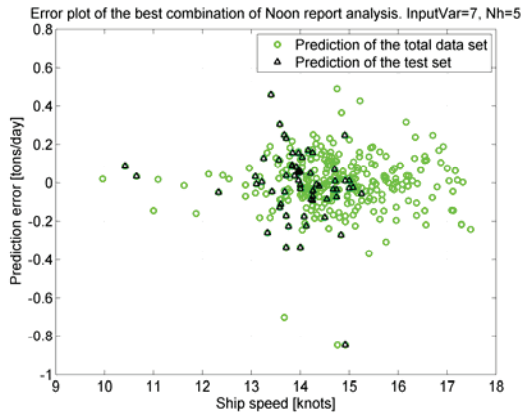


Figure 17: Prediction errors of the noon report analysis

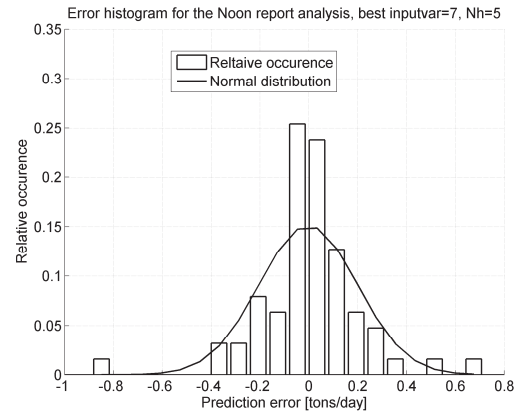


Figure 18: Relative distribution of the predicted errors from the noon report analysis

6.3 Comparison of the results

It is only possible to compare the dataset based on measured values, since it is the only one tested by other methods. In Figure 19 the cross validation error is shown for every condition and predicted by a linear, non-linear or ANN method. As previously mentioned dataset #2 gives significantly lower prediction errors for all methods due to the rather narrow ship speed band, the low wind speed and low number of samples.

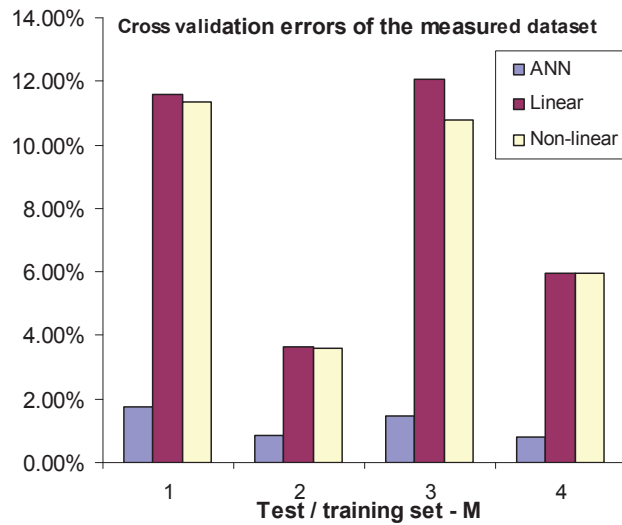


Figure 19: Comparison of different prediction methods.

7 Conclusion

Artificial Neural Networks (ANNs) can successfully be used to predict propulsion power, given that sufficient data are available. They have a significantly better performance than the linear and non-linear models tested. The propulsion power was predicted with an accuracy of less than 2% for the measured dataset. This accuracy is although of the same order of magnitude as the standard deviation of the propulsion power, so if a confidence interval analysis is introduced it is questionable if the method can get better.

ANNs can also be used with noon report data to predict the specific fuel consumption with an accuracy of about 7%, which is a bit surprising considering the rather rough input/output data. It is noted that this accuracy was obtained using “time” as an input variable, this indicates that it is possible to detect a trend of the fuel consumption over time.

It is shown that by introducing sea states and wind property information from the hindcast, the ANN solutions can be improved significantly, in the best case, from 2.97% to 1.65%. This eliminates the need for onboard measured wind speed and direction.

Unfortunately it was not possible to compare the solutions of the *four* different measured datasets with a solution using the noon report data from the same time, simply because of the lack of a sufficient number of noon reports for each dataset (there is only 3,4,7 and 9, see Table 1). Since the ship is not usually sailing in a single loading condition more than three weeks (21 noon reports), it will always be a problem to acquire enough data for making a reliable comparison of manual data acquisition (noon report) and automatic.

If measured data for more loading conditions were available it would be possible to make an analysis similar to the one made for the noon reports.

Acknowledgements

Many thanks to the ship owner Torm and the crew onboard Torm Marie for allowing me to install the equipment and providing noon report data. In particular Chief Engineer Rasmus Hoffman, who has showed great interest in the project and been very helpful monitoring the onboard system and sending data home. Many thanks to Kjeld Roar Jensen from FORCE Technology who has been providing the hindcast data.

The work presented in this paper has been carried out during Pedersen's PhD study at FORCE Technology and the Technical University of Denmark, which is sponsored by The Danish Industrial PhD programme and Danish Centre of Maritime Technology (DCMT)/The Danish Maritime Fund.

References

PEDERSEN, B.P & LARSEN J. (2009), Modeling of Propulsion Power, WMTC 2009, Jan. 21-24, Mumbai, India.

BISHOP, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer.

FORCE (2008), SeaTrend® info sheet is available at:
http://www.force.dk/en/Menu/Products+and+Concepts/Products/080220_seatrend.htm

HARVALD, S. A. (1983), *Resistance and Propulsion of Ships*, John Wiley & Sons.

HOLTROP, J. (1984), 'A Statistical re-analysis of resistance and propulsion data', *International Shipbuilding Progress* **31**, 272-276.

ISHERWOOD, R. (1972), 'Wind resistance of merchant ships', *Royal Institute of Naval Architecture*.

ITTC (1978), 'ITTC – Recommended Procedures - Performance, Propulsion 1978 ITTC Performance Prediction Method', ITTC, 7.5-0203-01.4.

DTU toolbox (2002), DTU toolbox: Neural regressor with quadratic cost function,
<http://isp.imm.dtu.dk/toolbox>

LARSEN, J. (1993), 'Design of Neural Network Filters', PhD thesis, Electronics Institute, Technical University of Denmark.

MACKAY, D. J. C. (1992), 'A practical Bayesian framework for backpropagation networks', *Neural Computation* **4**, 448-472.

MADSEN, K.; NIELSEN, H.B. & TINGLEFF, O. (2004), 'Methods for Non-Linear Least Squares Problems, 2nd Edition' IMM, DTU

NIELSEN, H. B. (1999), 'Damping Parameter in Marquardt's Method' Report IMM-REP-1999-05, IMM, DTU . 31 pages.

PEDERSEN, M. (1997), 'Optimization of Recurrent Neural Networks for Time Series Modeling', PhD thesis, Institute of Mathematical Modeling, Technical University of Denmark.

SVARER, C.; HANSEN, L. & LARSEN, J. (1993), On Design and Evaluation of Tapped-Delay Neural Network Architectures, *in* 'Proceedings of the 1993 IEEE International Conference on Neural Networks', pp. 46-51.

Appendix C

Input/output variable data

C.1 Histograms of measured data measured data

Histogram distribution of the the measured data sets.

C.1.1 Input variable

Histograms of the output variables e.i. measured propulsion power, from the 4 measured dataset.

C.1.2 Output variable

Histograms of the measured power as output variable.

X	Data variable	Source	ID	Unit
1	Speed through water	KYMA	U	knots
2	Relative wind speed	Airmar	Vrel	knots
3	Relative wind direction	Airmar	grel	°
4	Air temperature	Airmar	Tair	°C
5	Air temperature	Noon report	NR.Tair	°C
6	Sea water temperature	Noon report	NR.Tsw	°C
7	True wind speed	Hindcast	HC.Ws	m/s
8	True wind direction	Hindcast	HC.g	°
9	Significant wave height	Hindcast	HC.Hs	m
10	Wave period	Hindcast	HC.Tp	s
11	True wave direction	Hindcast	HC.Td	°
12	Relative wind speed	Hindcast	HC.Vrel	m/s
13	Relative wind direction	Hindcast	HC.grel	°
Y	Shaft power	KYMA	P	kW

Table C.1: Date variables for the measured data set analysis'

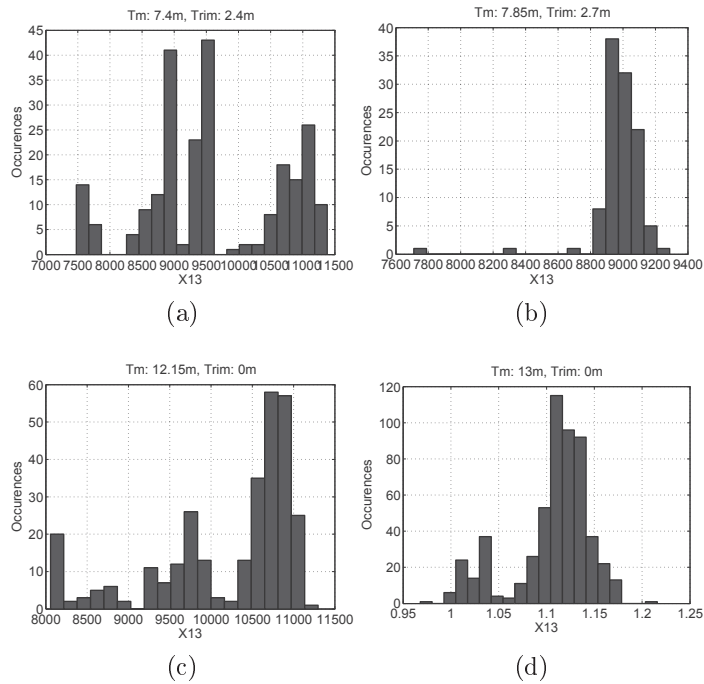


Figure C.1: Histograms of measured data set 1-4 of out variable Y as defined in Table C.1 as Y

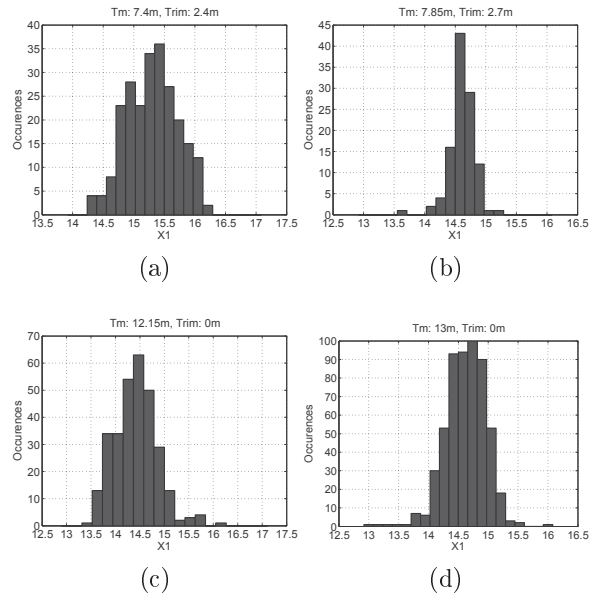


Figure C.2: Histograms of measured data set 1-4 of input variable X_1 as defined in Table C.1

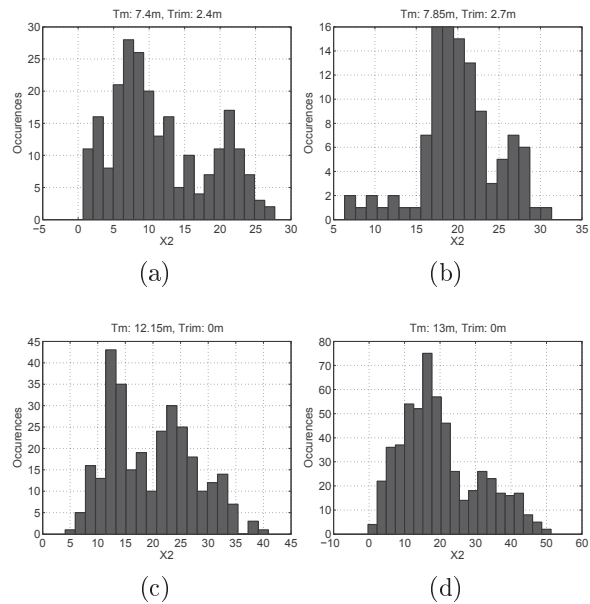


Figure C.3: Histograms of measured data set 1-4 of input variable X_2 as defined in Table C.1

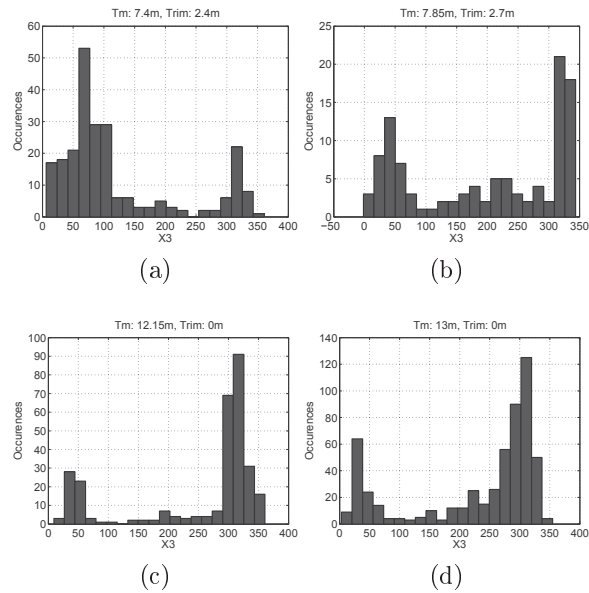


Figure C.4: Histograms of measured data set 1-4 of input variable X_3 as defined in Table C.1

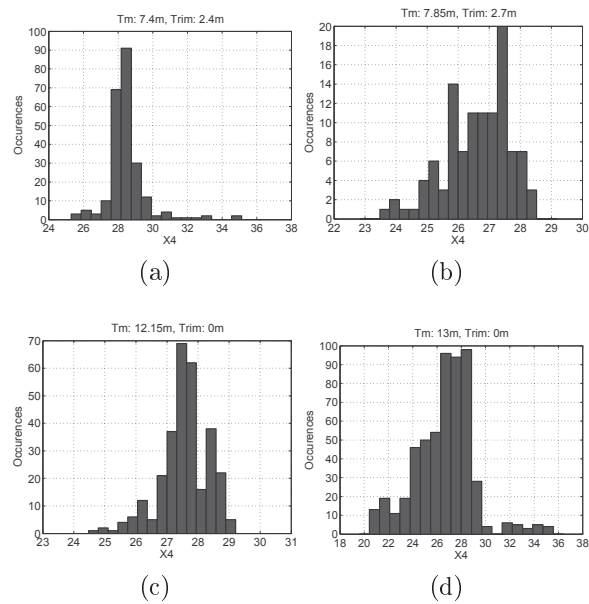


Figure C.5: Histograms of measured data set 1-4 of input variable X_4 as defined in Table C.1

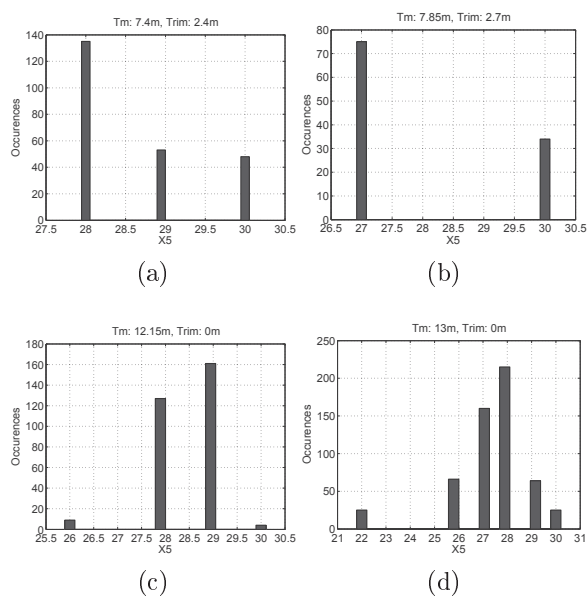


Figure C.6: Histograms of measured data set 1-4 of input variable X_5 as defined in Table C.1

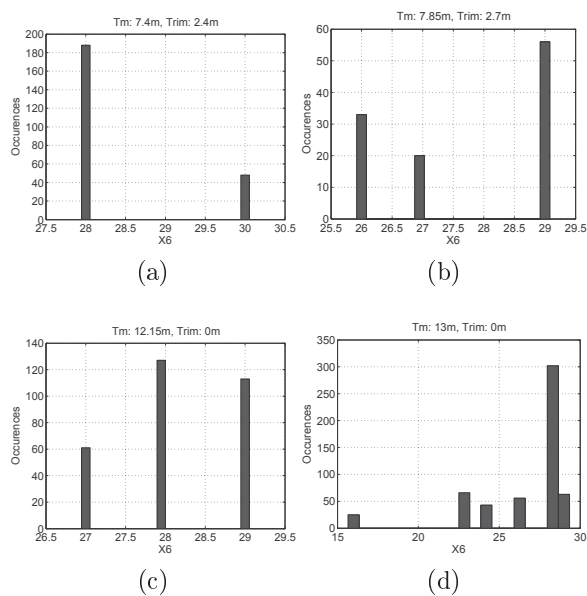


Figure C.7: Histograms of measured data set 1-4 of input variable X_6 as defined in Table C.1

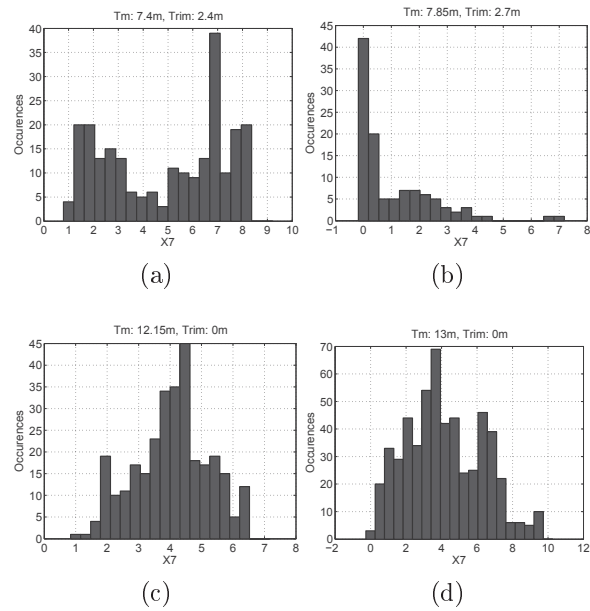


Figure C.8: Histograms of measured data set 1-4 of input variable X_7 as defined in Table C.1

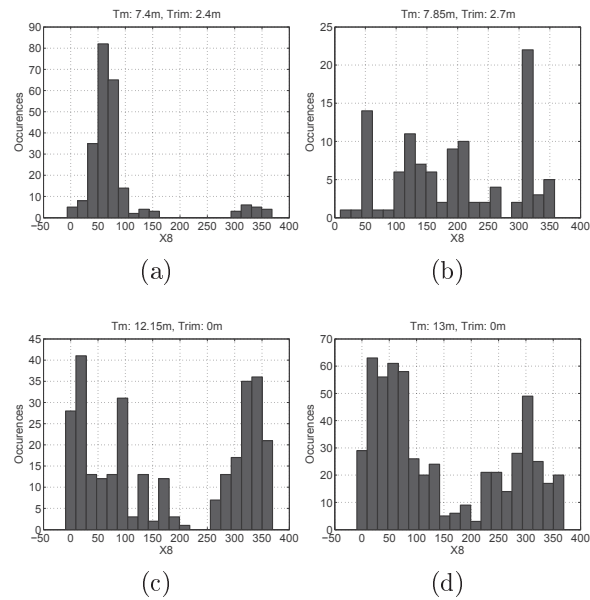


Figure C.9: Histograms of measured data set 1-4 of input variable X_8 as defined in Table C.1

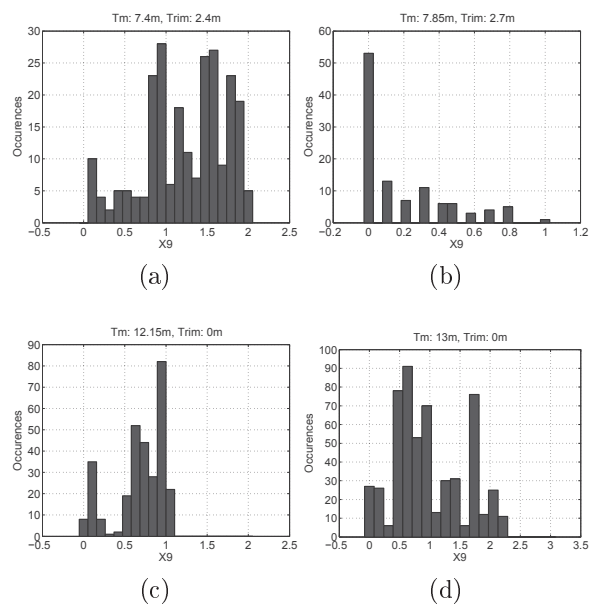


Figure C.10: Histograms of measured data set 1-4 of input variable X_9 as defined in Table C.1

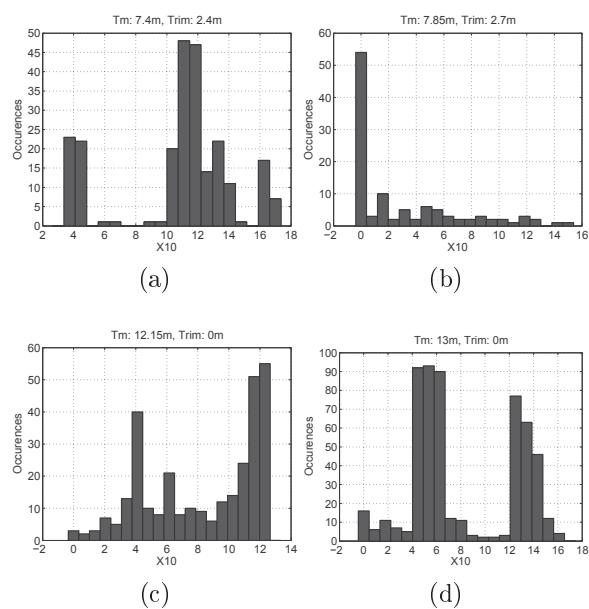


Figure C.11: Histograms of measured data set 1-4 of input variable X_{10} as defined in Table C.1

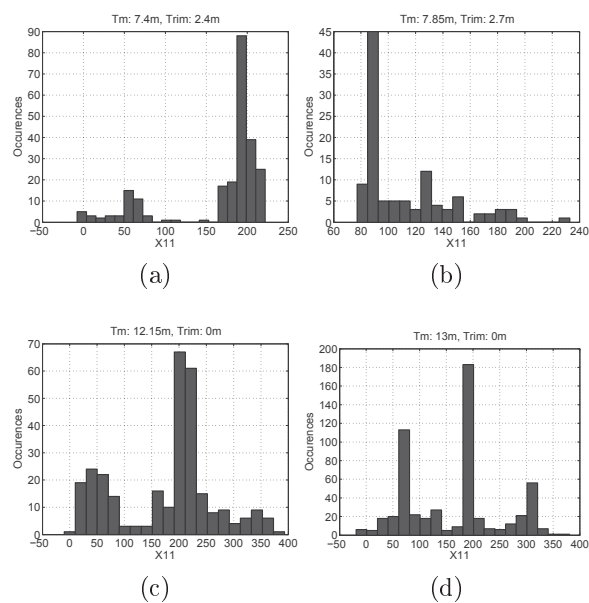


Figure C.12: Histograms of measured data set 1-4 of input variable X_{11} as defined in Table C.1

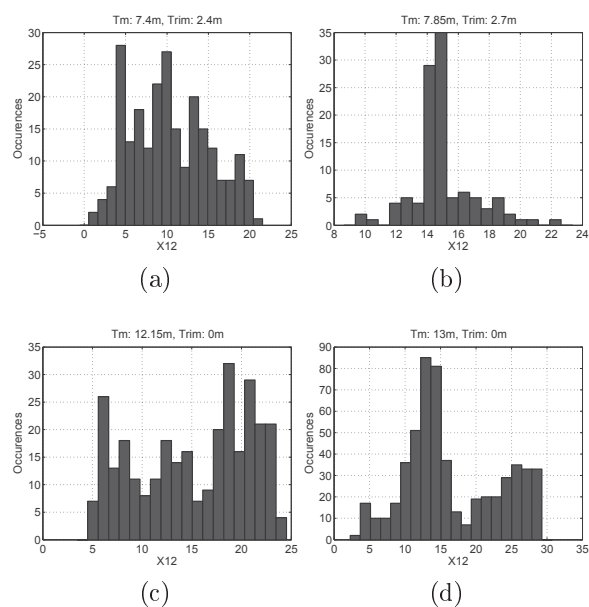


Figure C.13: Histograms of measured data set 1-4 of input variable X_{12} as defined in Table C.1

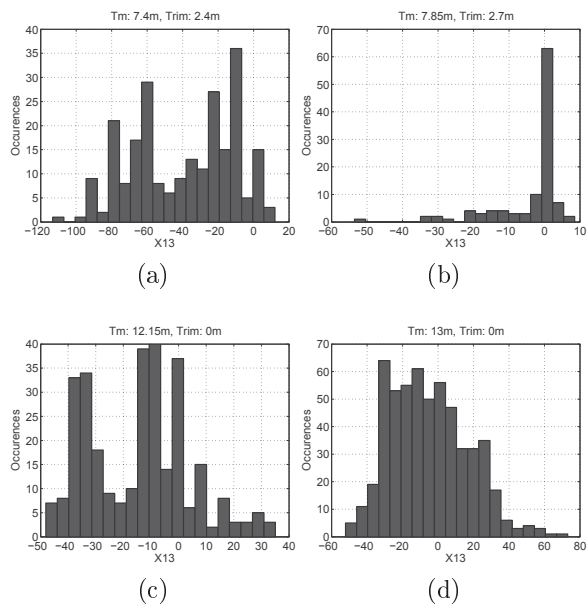


Figure C.14: Histograms of measured data set 1-4 of input variable X_{13} as defined in Table C.1

C.2 Histograms of the noon report data from Torm Marie

C.2.1 Input variable

Histograms of the input variables of the noon report data set.

X	Data variable	Source	ID	Unit
1	Average ship speed through the water	Noon report	Avg(U)	knots
2	Air temperature	Noon report	NRxls.Tair	°C
3	Sea water temperature	Noon report	NRxls.Tsw	°C
4	Mean draught (Ta+Tf)/2	Noon report	Tm	m
5	Trim, Ta-Tf	Noon report	Trim	m
6	Average winds speed during report period	Hindcast	mean(HC.Ws)	m/s
7	Average winds direction during report period	Hindcast	mean(HC.gamma)	°
8	Average significant wave height during report period	Hindcast	mean(HC.Hs)	m
9	Average wave period during report period	Hindcast	mean(HC.Tp)	s
10	Average wave direction during report period	Hindcast	mean(HC.Td)	°
11	Variance of the winds speed during report period	Hindcast	var(HC.Ws)	m/s
12	Variance of the winds direction during report period	Hindcast	var(HC.gamma)	°
13	Variance of the significant wave height during report period	Hindcast	var(HC.Hs)	m
14	Variance of the wave period during report period	Hindcast	var(HC.Tp)	s
15	Variance of the wave direction during report period	Hindcast	var(HC.Td)	°
16	Relative average wind speed	Hindcast	HC Vrel	m/s
17	Relative average wind direction	Hindcast	HC gammarel	°
18	Relative average wave direction	Hindcast	HC.Tdrel	°
19	Wind force (Beaufort)	Noon report	NRxls.WINDCONDITIONATNOONBF(II)	BF
20	Wind direction	Noon report	NRxls.WINDCONDITIONATNOONDEG	°
21	Report date and tim (Matlab numeric value)	Noon report	NRxls.UTC	numeric
Y	Averag HFO consumption per hour	Noon report	CONS_OFHFOME_SPEC	tons/hr

Table C.2: Date variables for the noon report data set analysis'

C.2.2 Output variable

Histograms of the output variable i.e. fuel consumption per hour.

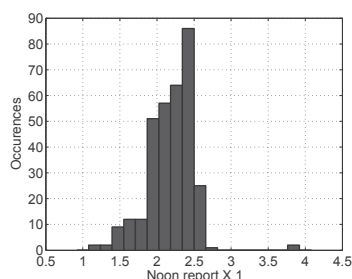


Figure C.15: Histograms of noon report data from Torm Marie output variable Y as defined in Table C.2 as Y

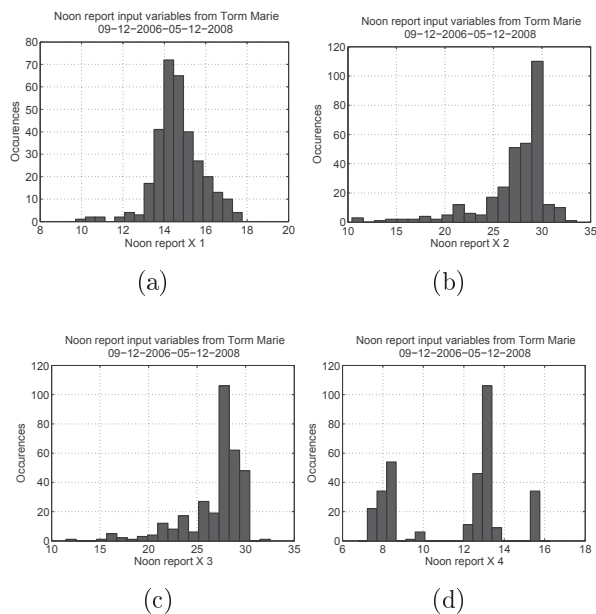


Figure C.16: Histograms of the noon report data set 1-4 of input variable as defined in Table C.2

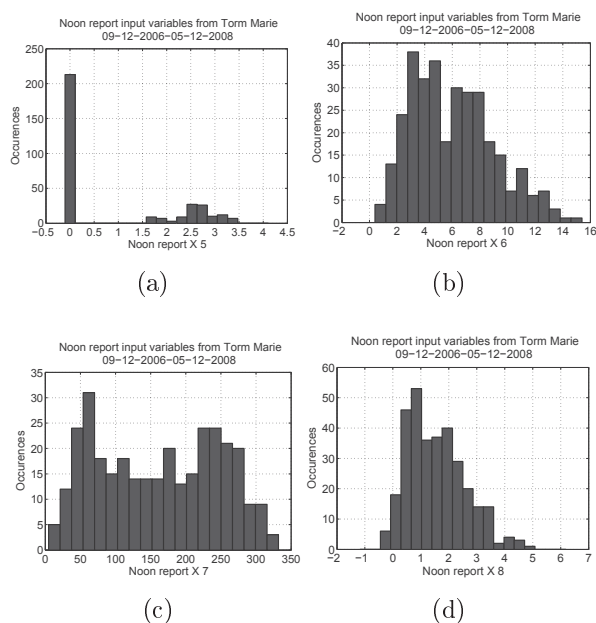


Figure C.17: Histograms of the noon report data set 5-8 of input variable as defined in Table C.2

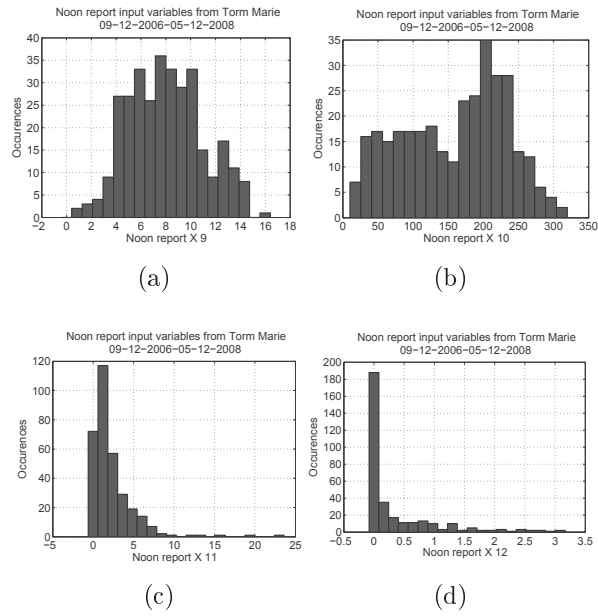


Figure C.18: Histograms of the noon report data set 9-12 of input variable as defined in Table C.2

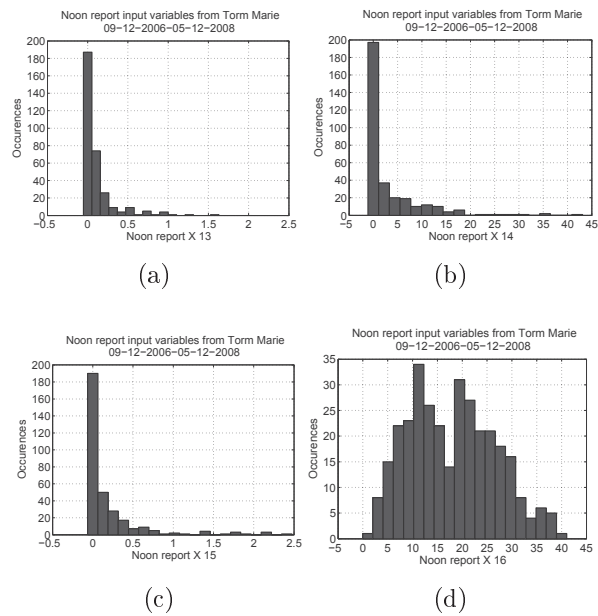


Figure C.19: Histograms of the noon report data set 13-16 of input variable as defined in Table C.2

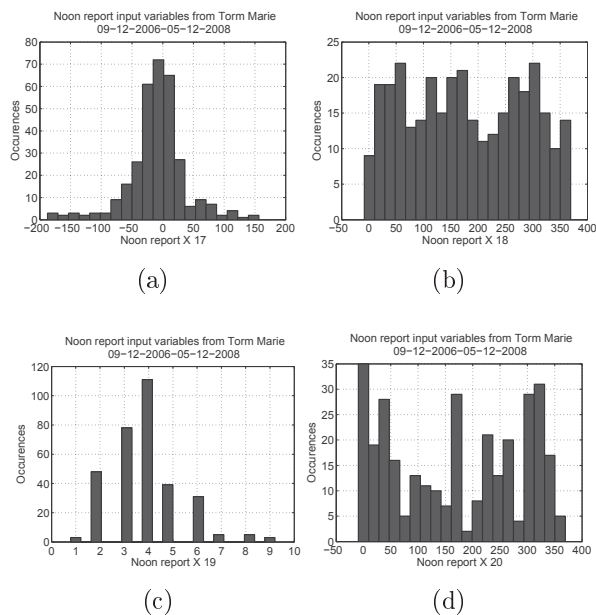


Figure C.20: Histograms of the noon report data set 17-20 of input variable as defined in Table C.2

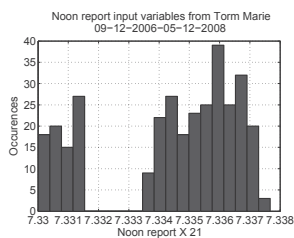


Figure C.21: Histograms of the noon report data set 21 of input variable as defined in Table C.2

Histograms of the input variables from the container ship datasets in the period from the launch until the first dry-docking

C.3 Input variable

X	Data variable	Source	ID	Unit
1	Speed through water	Noon report	NR.Ulog	knots
2	Speed over ground	Noon report	NR.Uobs	knots
3	Sea water temperature	Noon report	NR.Tsw	deg
4	Mean draught $(T_a+T_f)/2$	Noon report	NR.Tm	m
5	Trim, T_a-T_f	Noon report	NR.Trim	m
6	True wind speed	Noon report	NR.WindSpeed	m/s
7	Relative wind direction	Noon report	NR.WindDir	deg
8	Average relative winds speed during report period	Hindcast	HC.mean.Vrel	m/s
9	Average relative winds direction during report period	Hindcast	HC.mean.gammarel	deg
10	Average significant wave height during report period	Hindcast	HC.mean.Hs	m
11	Average wave period during report period	Hindcast	HC.mean.Tp	s
12	Variance of the significant wave height during report period	Hindcast	HC.var.Hs	m
13	Variance of the wave period during report period	Hindcast	HC.var.Tp	s
14	Variance of the wave direction during report period	Hindcast	HC.var.Td	deg
15	Variance of the winds speed during report period	Hindcast	HC.var.Ws	m/s
16	Variance of the winds direction during report period	Hindcast	HC.var.gamma	deg
17	Report date and tim (Matlab numeric value)	Noon report	NR.UTC	numeric
18	Average winds speed during report period	Hindcast	HC.mean.Ws	m/s
19	Average winds direction during report period	Hindcast	HC.mean.gamma	deg
27	Sea state	Noon report	NR.SeaState	m
28	Relative sea direction	Noon report	NR.TrueRelativeSeaDirection	deg
36	Average shaft power	Noon report	NR.PropPower	kW

Table C.3: Date variables for the noon report data set analysis'

C.3.1 Output variable

The output variable is fuel consumption per hour.

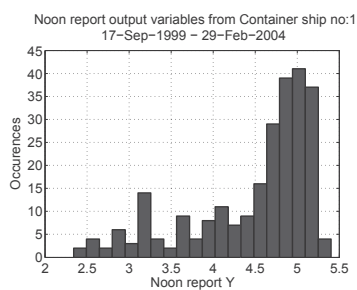


Figure C.22: Histograms of noon report data from Torm Marie output variable Y as defined in Table C.3 as Y

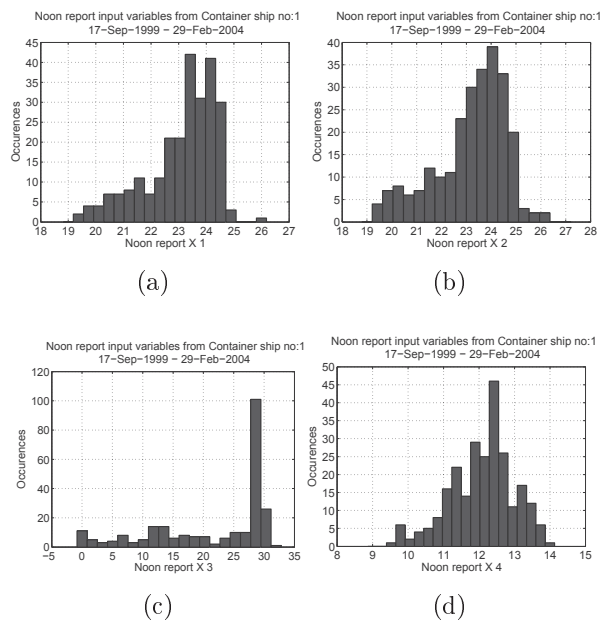


Figure C.23: Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 1-4 as defined in Table C.3

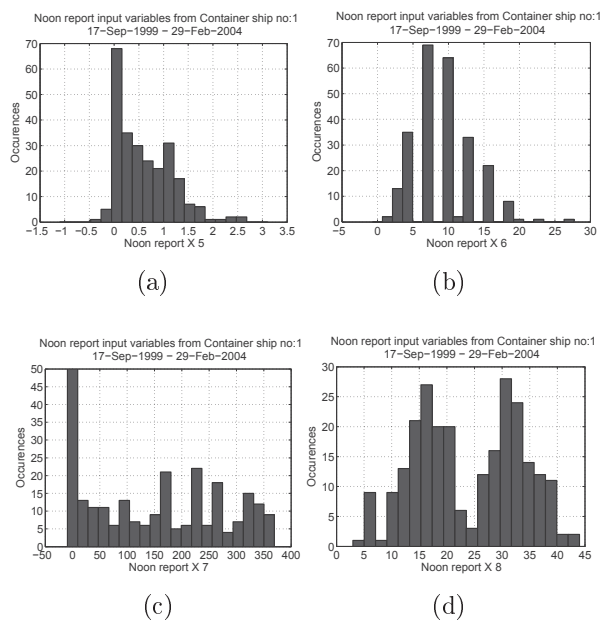


Figure C.24: Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 5-8 as defined in Table C.3

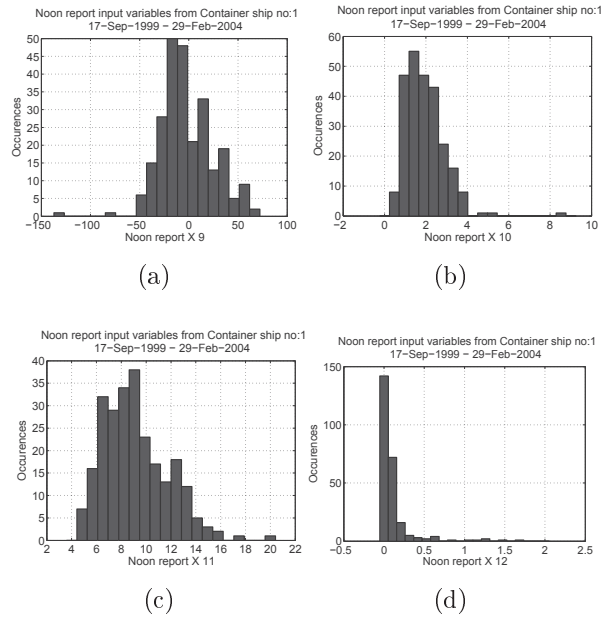


Figure C.25: Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 9-12 as defined in Table C.3

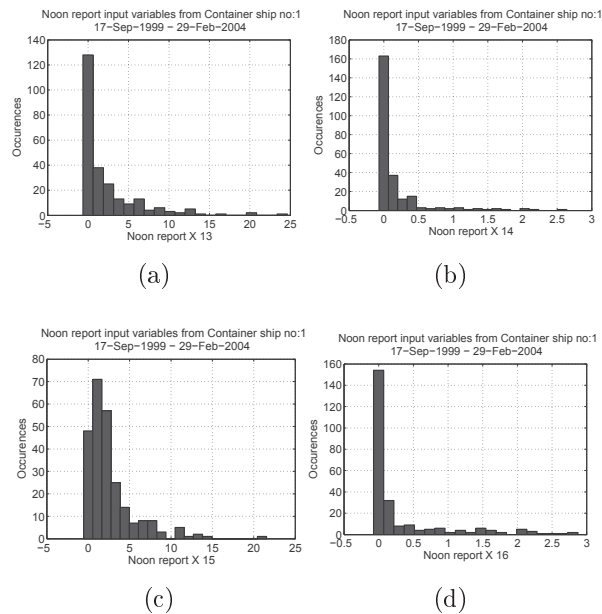


Figure C.26: Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 13-16 as defined in Table C.3

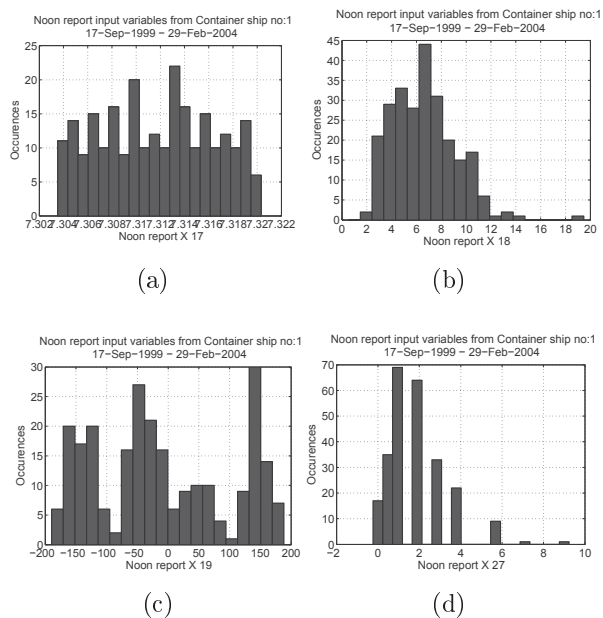


Figure C.27: Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 17-27 as defined in Table C.3

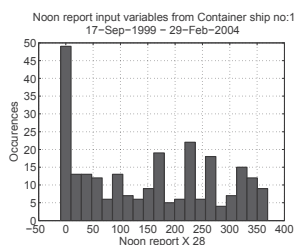


Figure C.28: Histograms of the noon report data set from Container ship #1 (17-9-1999 - 29-2-2004) of input variable 28 as defined in Table C.3

Appendix D

Trend detection results of Container ship 1-5

This appendix presents the full results from Chapter 5.

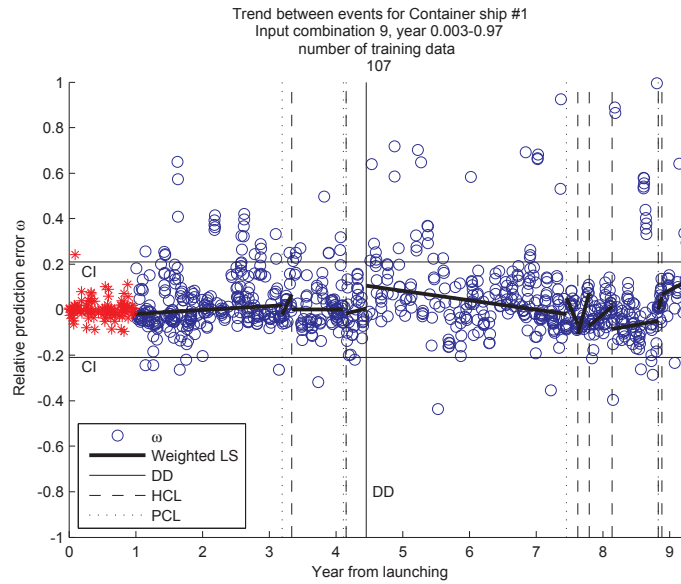


Figure D.1: Performance trend between all events of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first one-year period

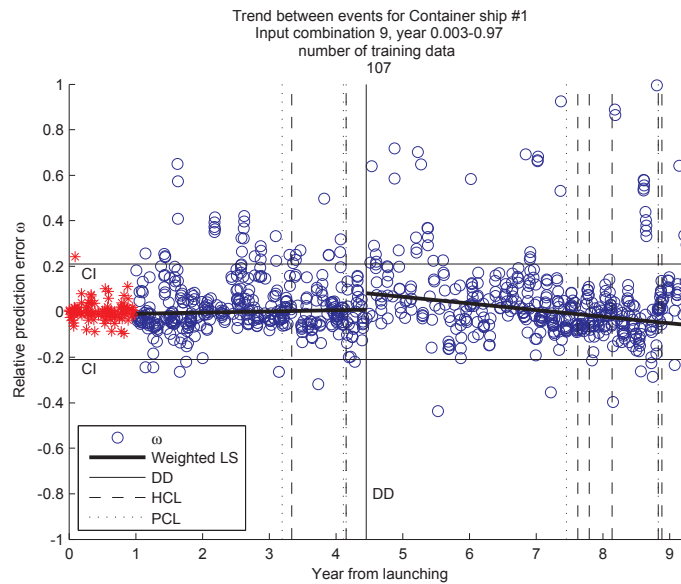


Figure D.2: Performance trend between dry-dockings of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first one-year period

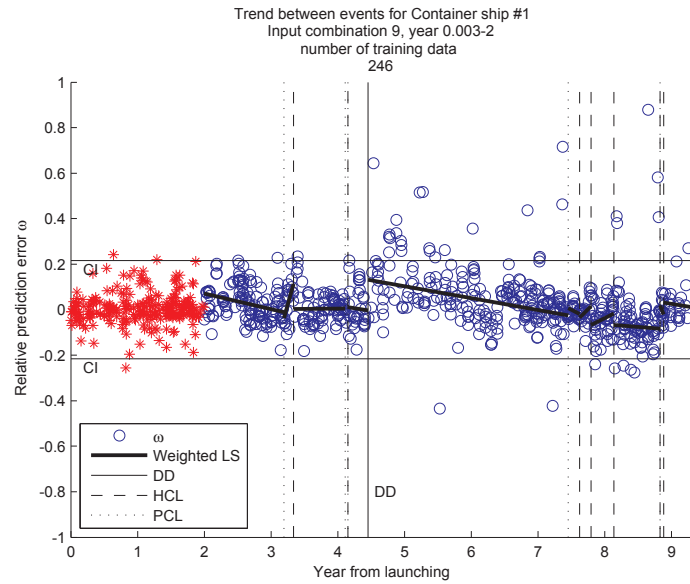


Figure D.3: Performance trend between all events of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first two-year period

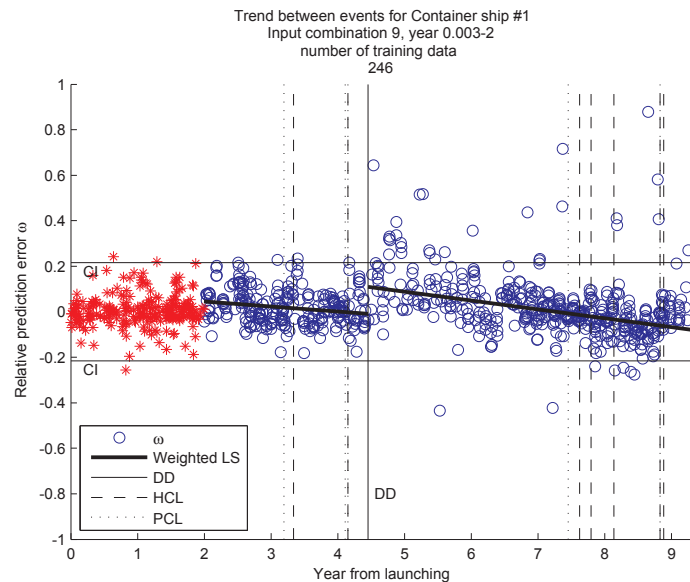


Figure D.4: Performance trend between dry-dockings of Container ship 1 using input variable setup 9 with all Noon report data, trained with the first two-year period

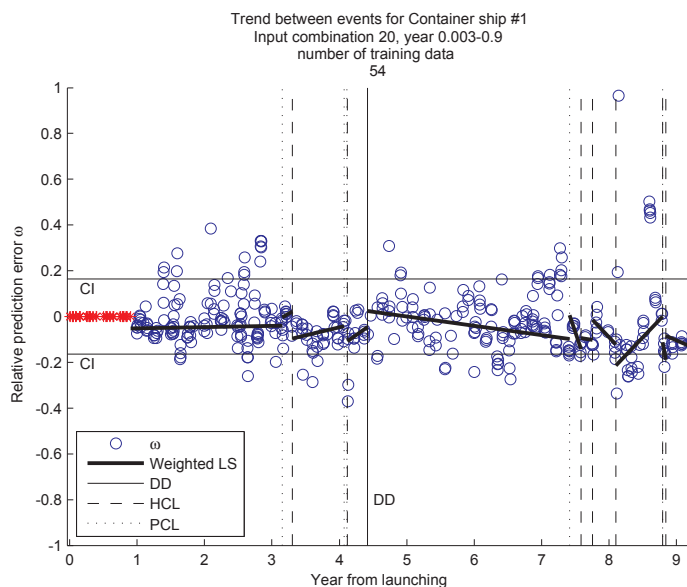


Figure D.5: Performance trend between all events of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first one-year period

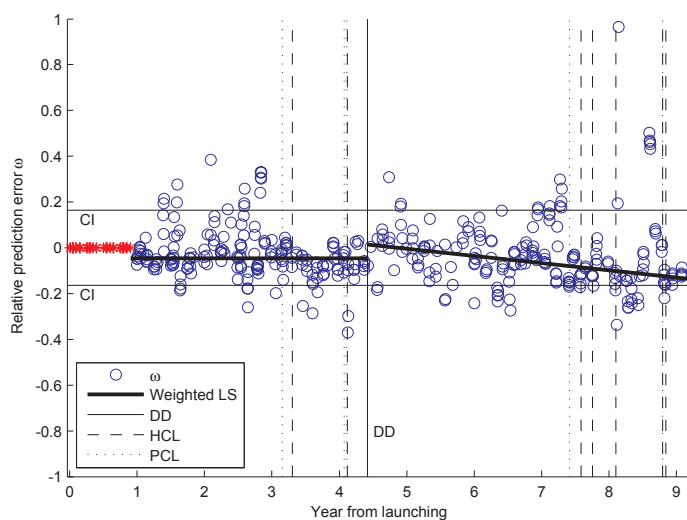


Figure D.6: Performance trend between dry-dockings of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first one-year period

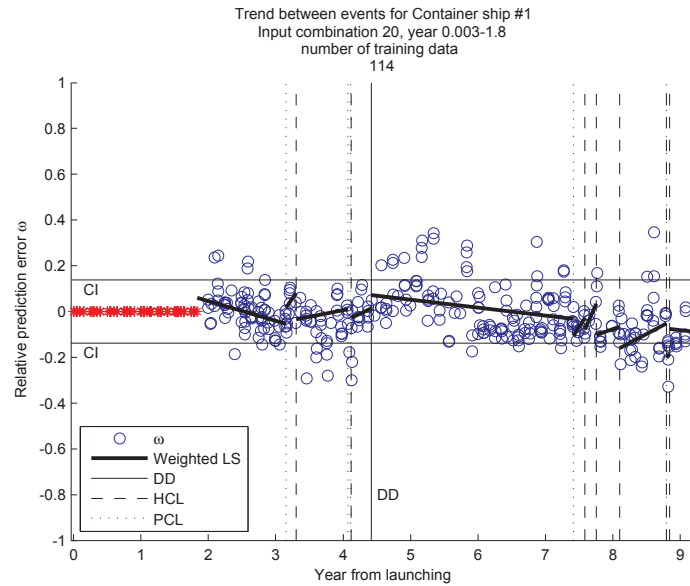


Figure D.7: Performance trend between all events of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first two-year period

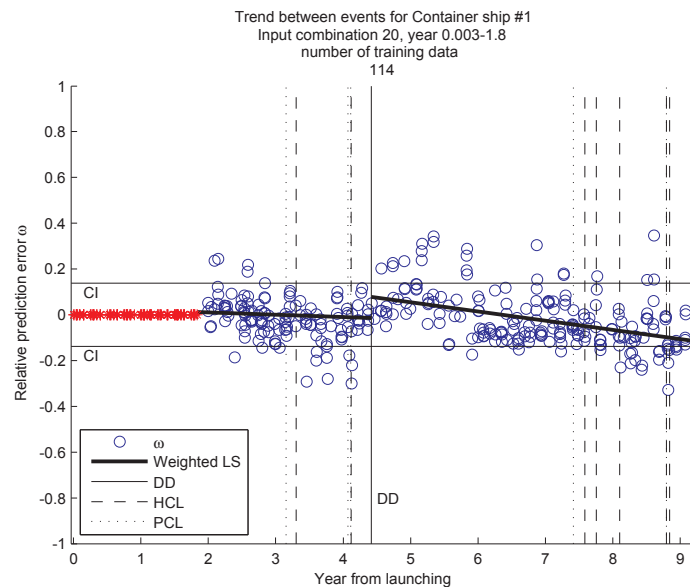


Figure D.8: Performance trend between dry-dockings of Container ship 1 using input variable setup 20 with data including hindcasts, trained with the first two-year period

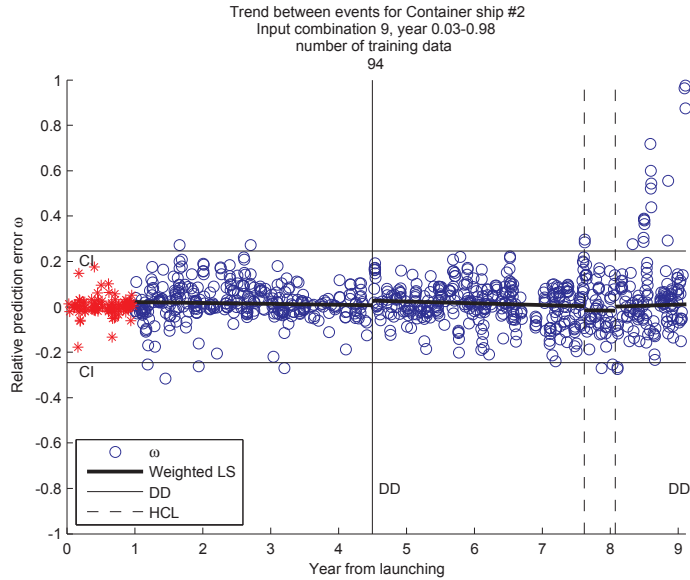


Figure D.9: Performance trend between all events of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first one-year period

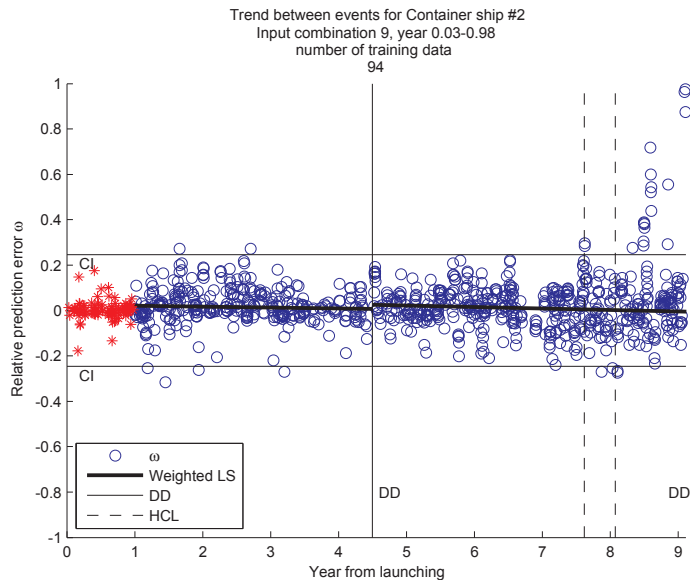


Figure D.10: Performance trend between dry-dockings of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first one-year period

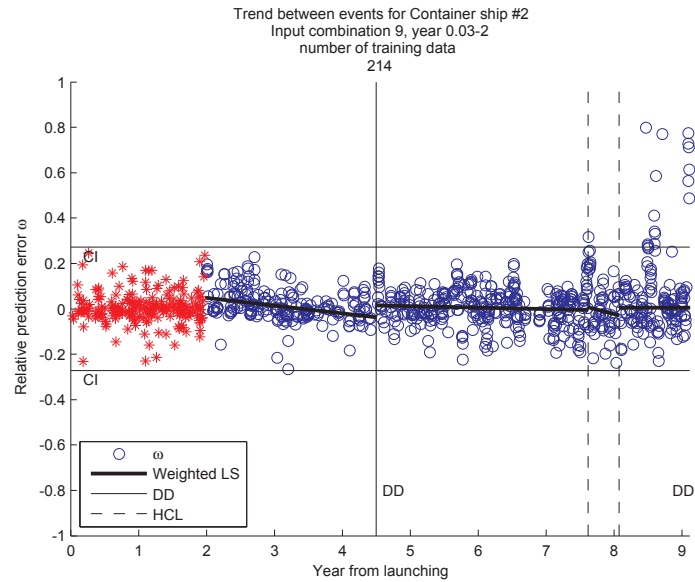


Figure D.11: Performance trend between all events of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first two-year period

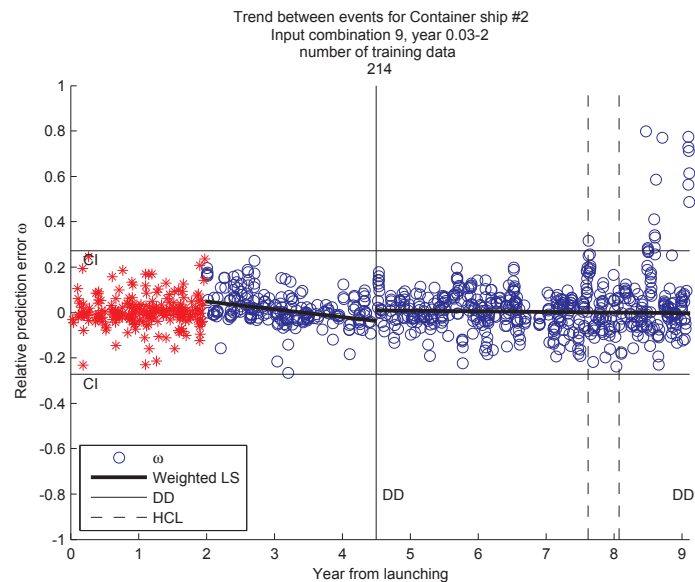


Figure D.12: Performance trend between dry-dockings of Container ship 2 using input variable setup 9 with all Noon report data, trained with the first two-year period

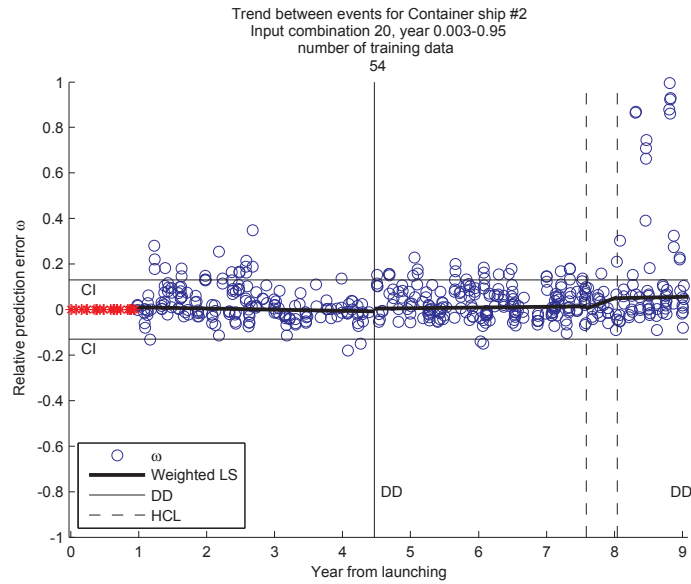


Figure D.13: Performance trend between all events of Container ship 2 using input variable setup 20 with data including hindcasts, trained with the first one-year period

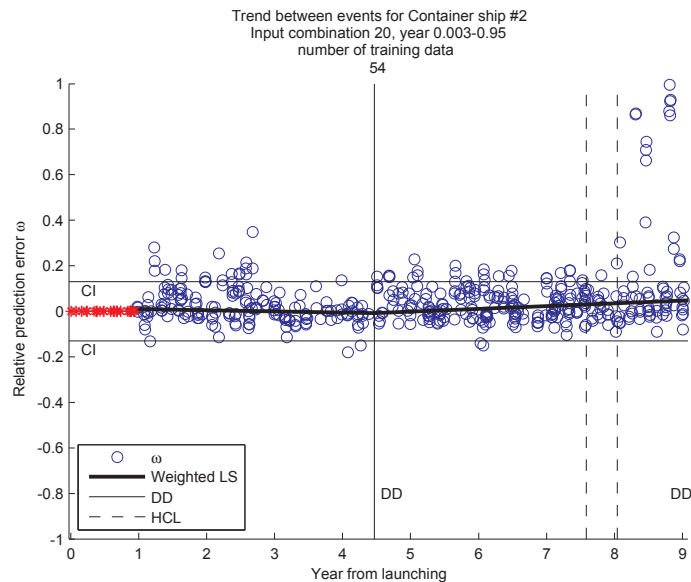


Figure D.14: Performance trend between dry-dockings of Container ship 2 using input variable setup 20 with data including hindcasts, trained with the first one-year period

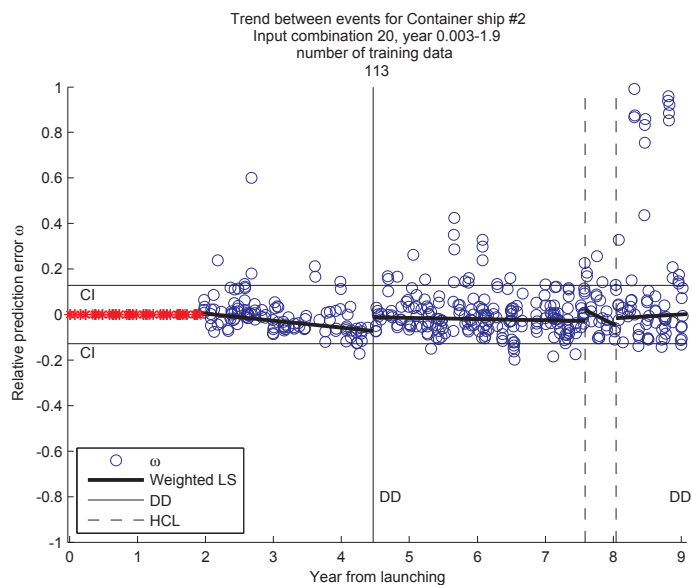


Figure D.15: Performance trend between all events of Container ship 2 using input variable setup 20 with data including hindcasts, trained with the first two-year period

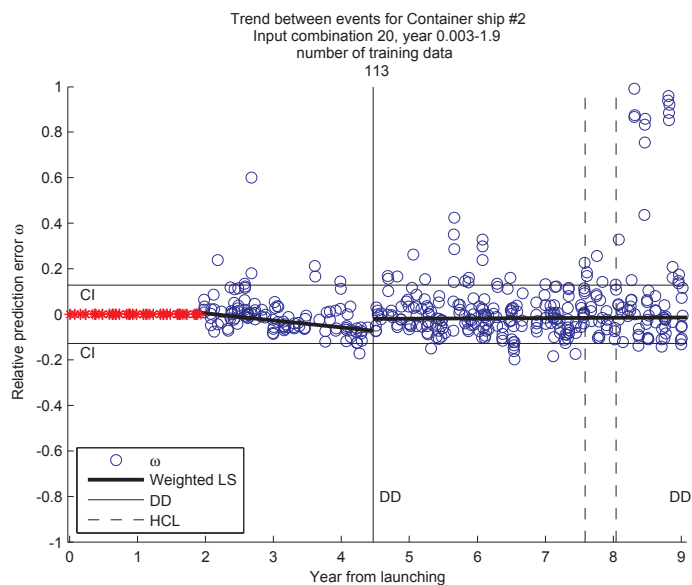


Figure D.16: Performance trend between dry-dockings of Container ship 2 using input variable setup 20 with data including hindcasts, trained with the first two-year period

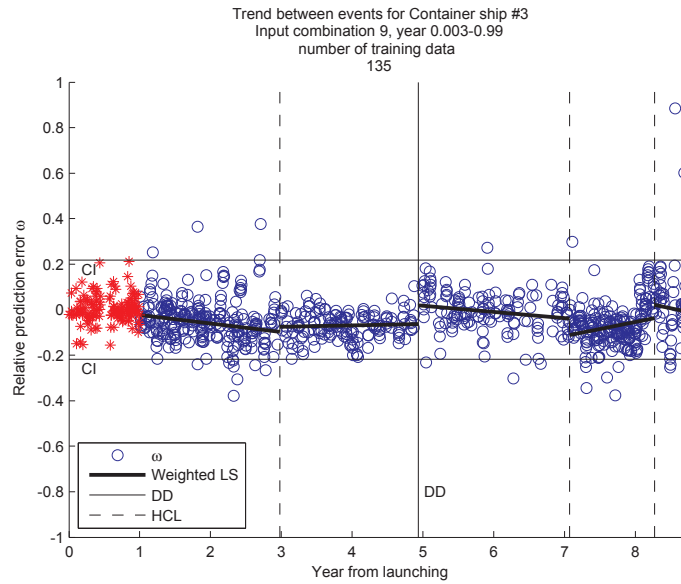


Figure D.17: Performance trend between all events of Container ship 3 using input variable setup 9 with all Noon report data, trained with the first one-year period

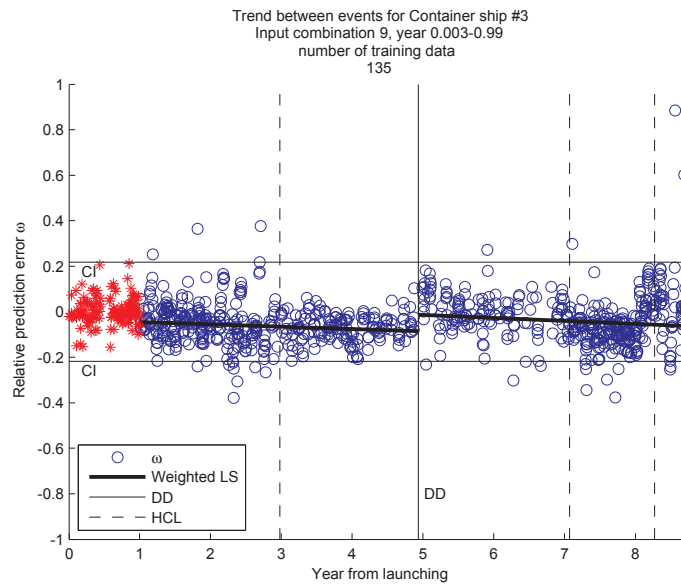


Figure D.18: Performance trend between dry-dockings of Container ship 3 using input variable setup 9 with all Noon report data, trained with the first one-year period

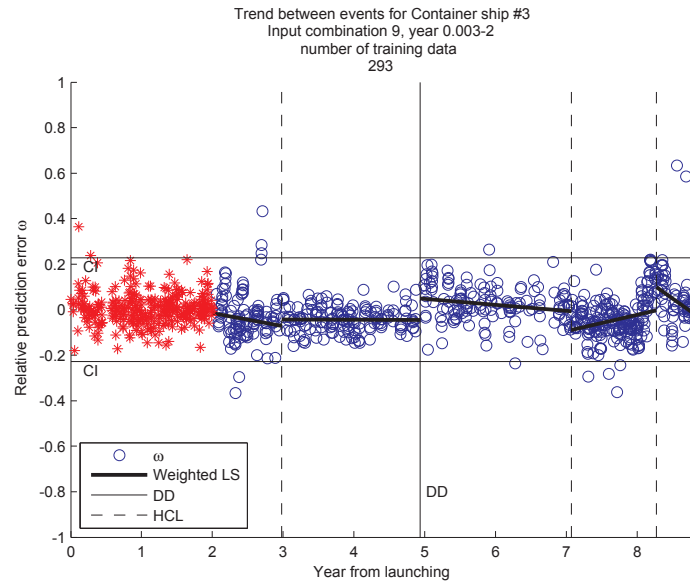


Figure D.19: Performance trend between all events of Container ship 3 using input variable setup 9 with all Noon report data, trained with the first two-year period

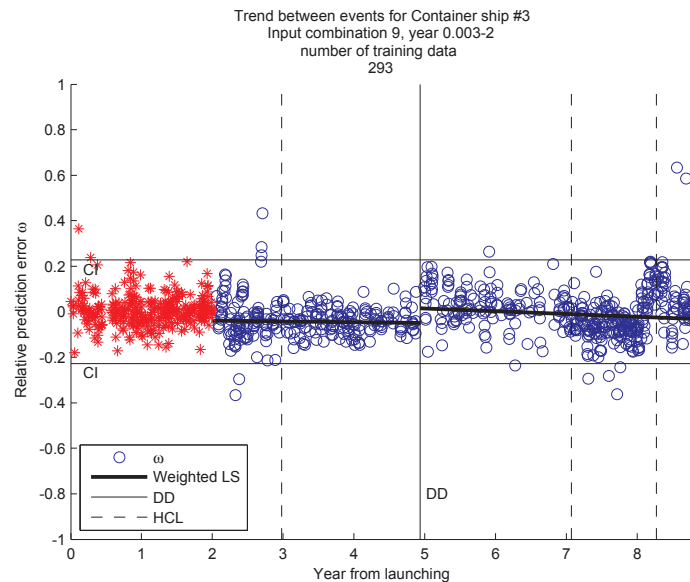


Figure D.20: Performance trend between dry-dockings of Container ship 3 using input variable setup 9 with all Noon report data, trained with the first two-year period

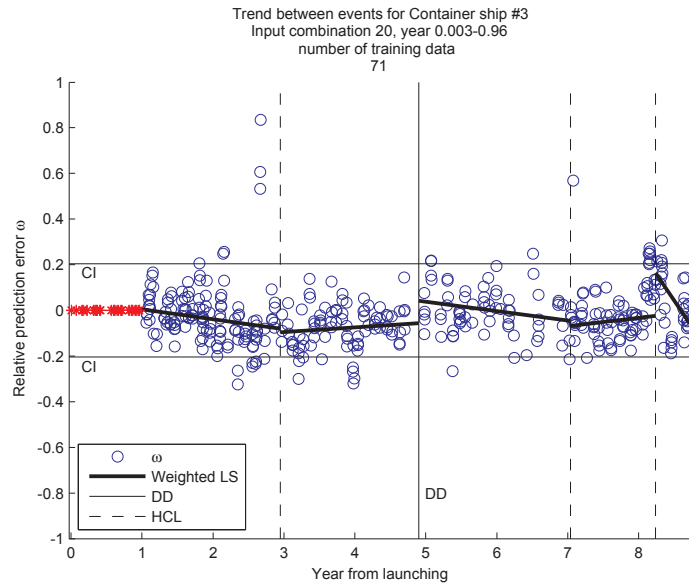


Figure D.21: Performance trend between all events of Container ship 3 using input variable setup 20 with data including hindcasts, trained with the first one-year period

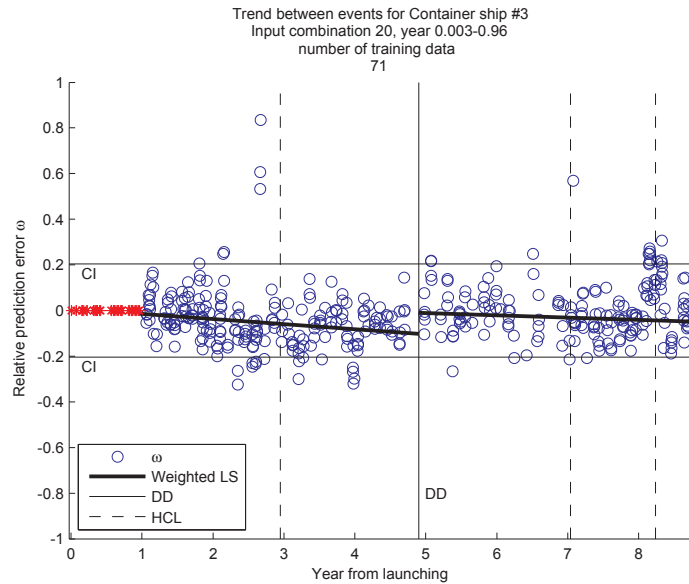


Figure D.22: Performance trend between dry-dockings of Container ship 3 using input variable setup 20 with data including hindcasts, trained with the first one-year period

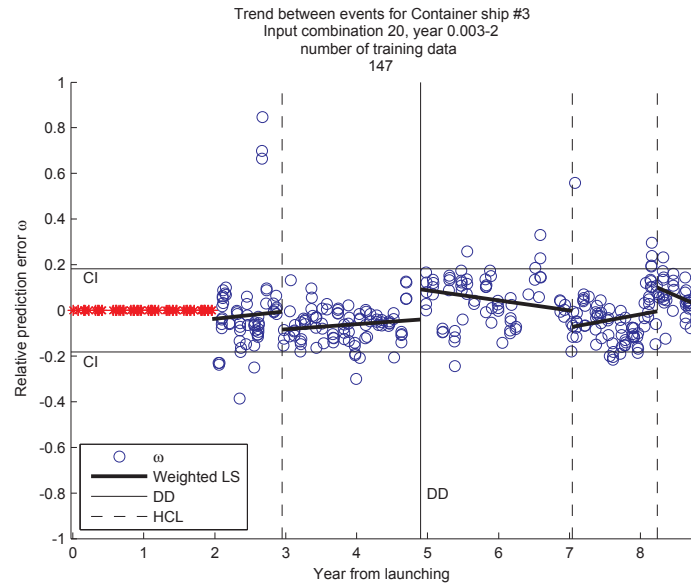


Figure D.23: Performance trend between all events of Container ship 3 using input variable setup 20 with data including hindcasts, trained with the first two-year period

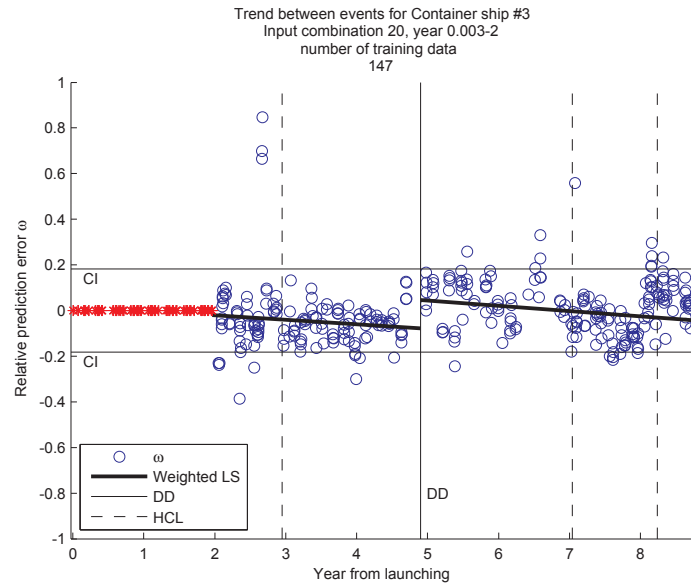


Figure D.24: Performance trend between dry-dockings of Container ship 3 using input variable setup 20 with data including hindcasts, trained with the first two-year period

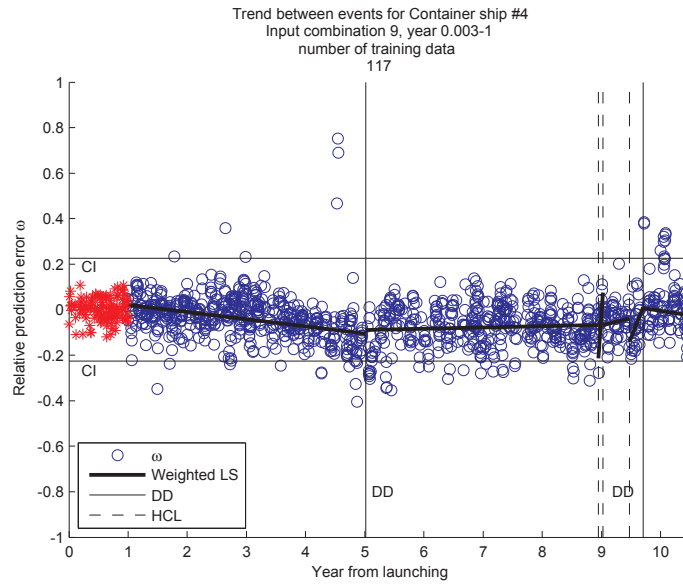


Figure D.25: Performance trend between all events of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first one-year period

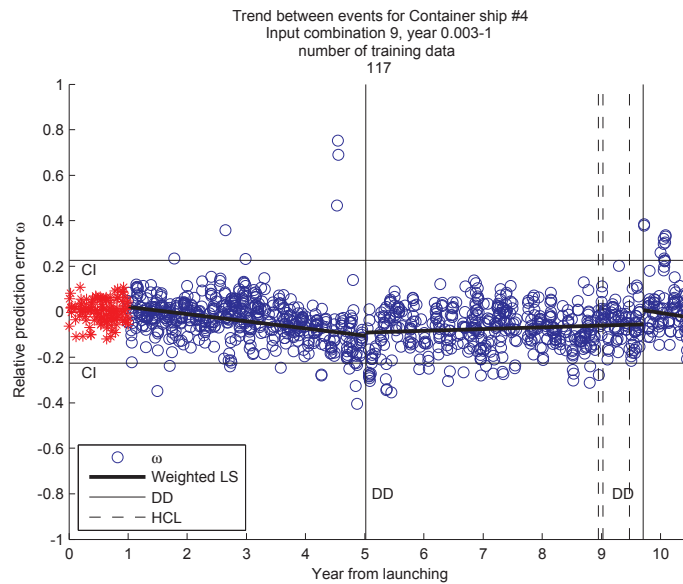


Figure D.26: Performance trend between dry-dockings of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first one-year period

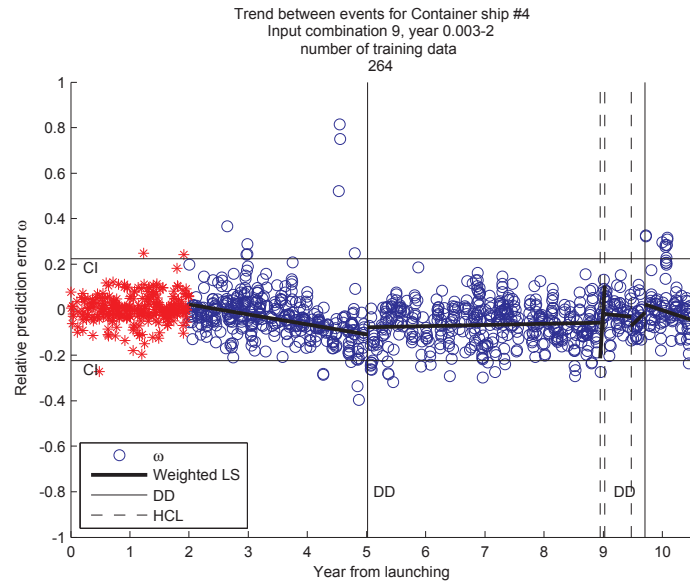


Figure D.27: Performance trend between all events of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first two-year period

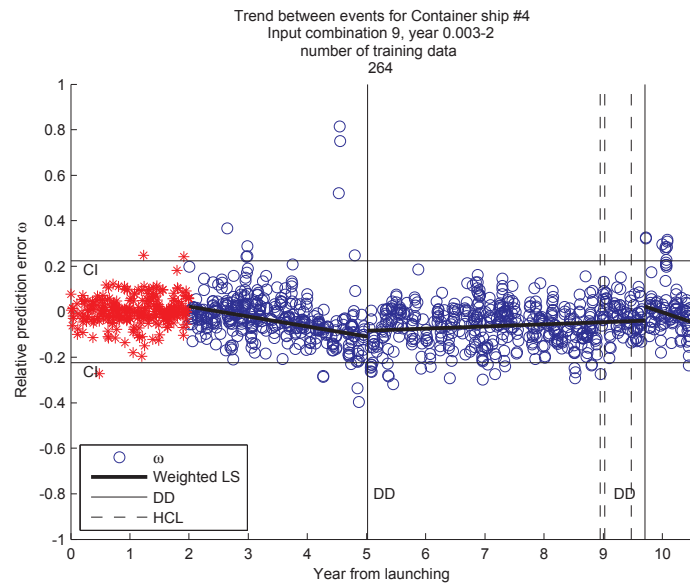


Figure D.28: Performance trend between dry-dockings of Container ship 4 using input variable setup 9 with all Noon report data, trained with the first two-year period

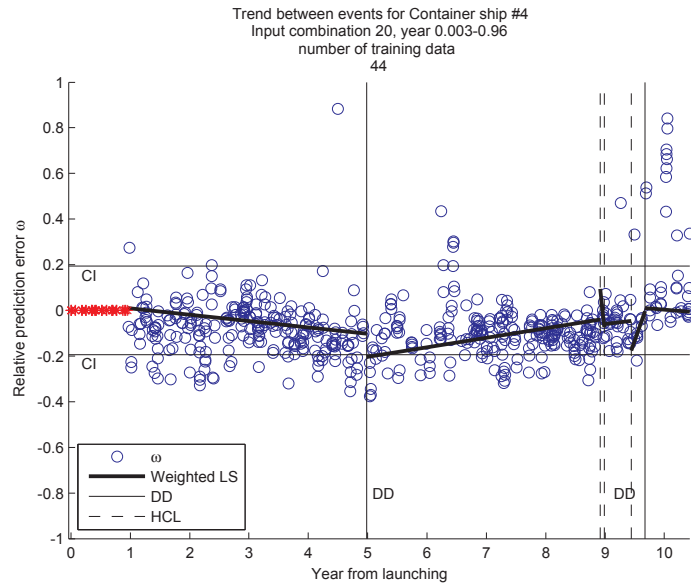


Figure D.29: Performance trend between all events of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first one-year period

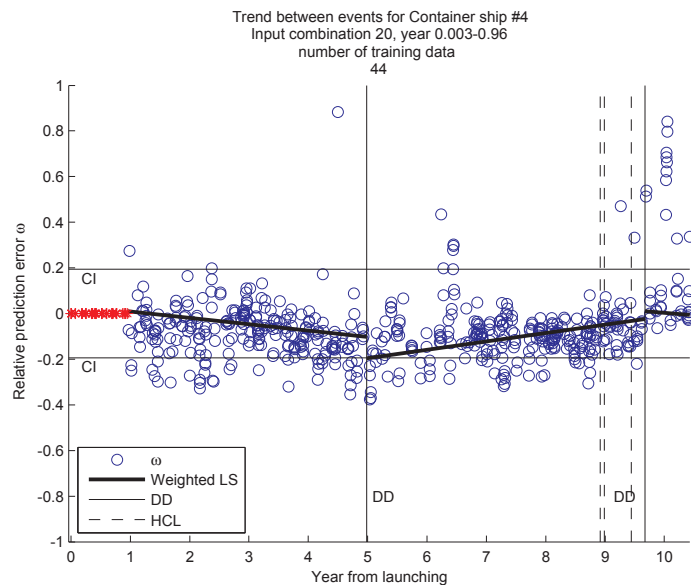


Figure D.30: Performance trend between dry-dockings of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first one-year period

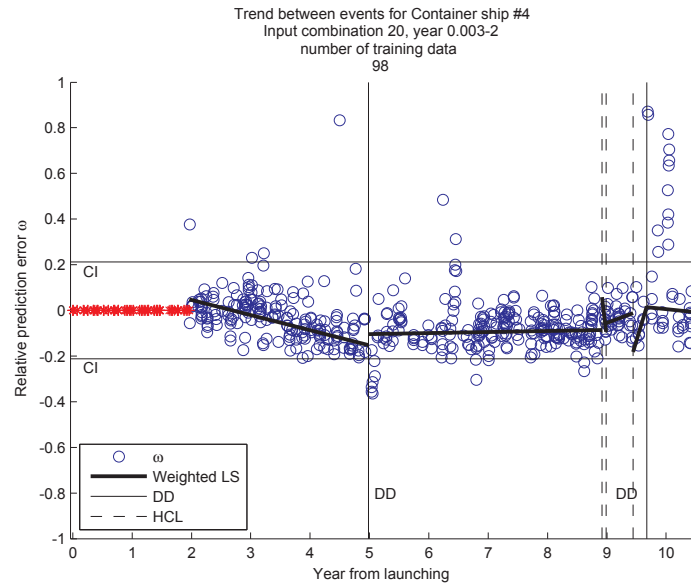


Figure D.31: Performance trend between all events of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first two-year period

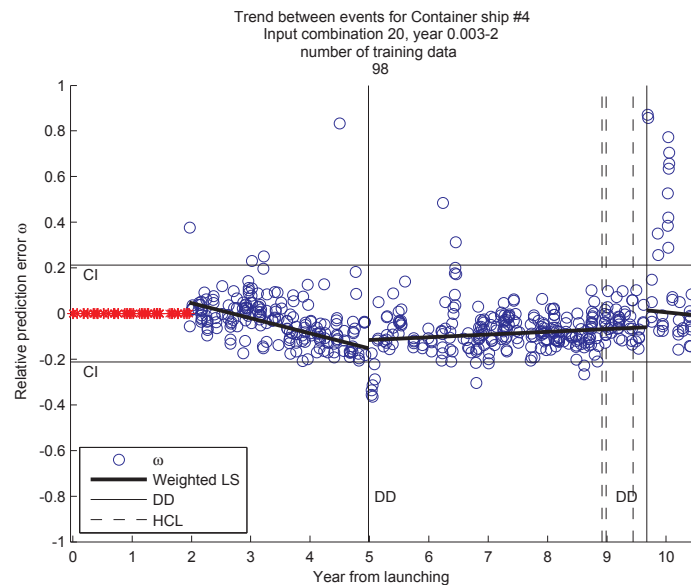


Figure D.32: Performance trend between dry-dockings of Container ship 4 using input variable setup 20 with data including hindcasts, trained with the first two-year period

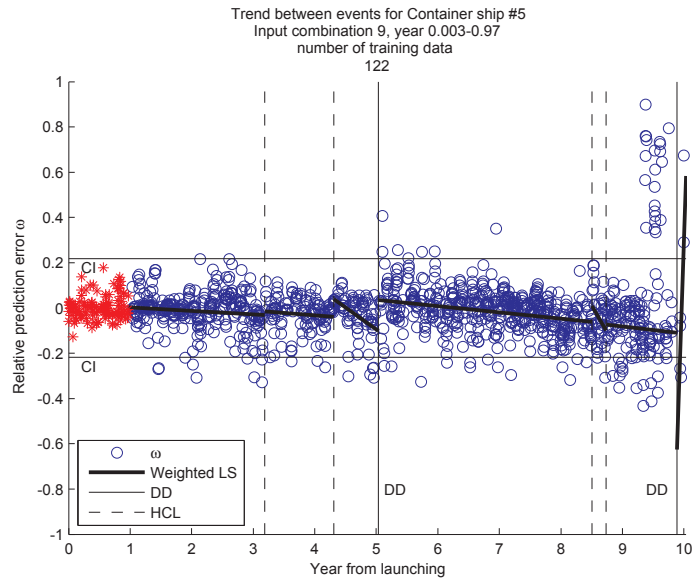


Figure D.33: Performance trend between all events of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first one-year period

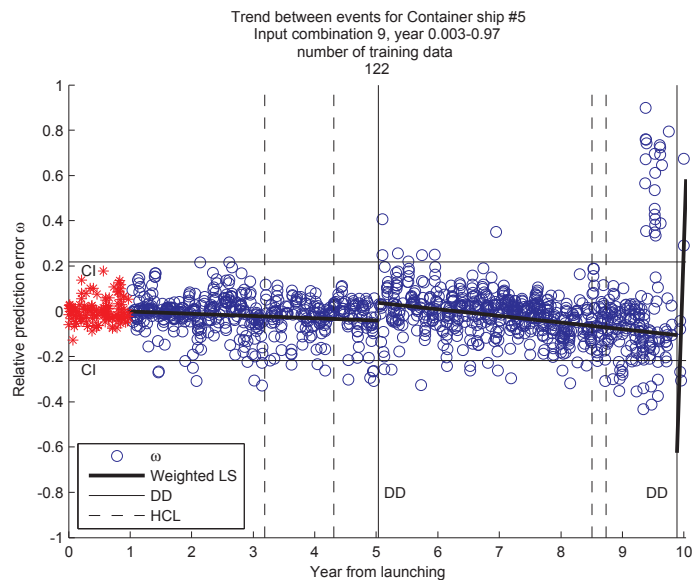


Figure D.34: Performance trend between dry-dockings of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first one-year period

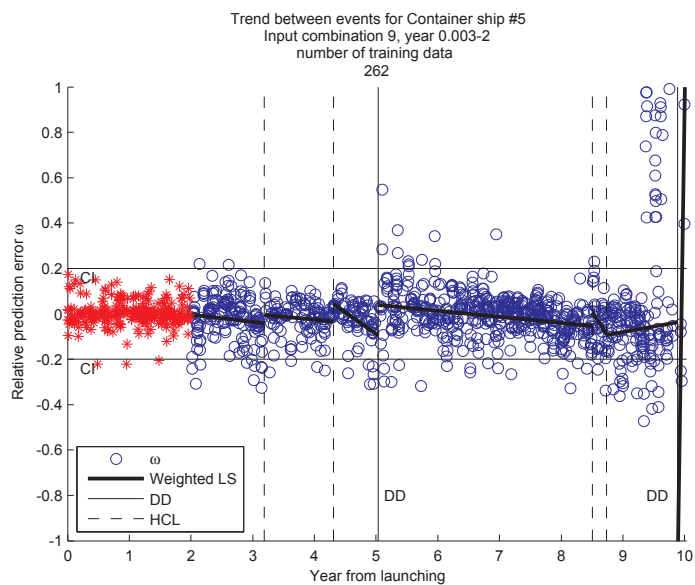


Figure D.35: Performance trend between all events of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first two-year period

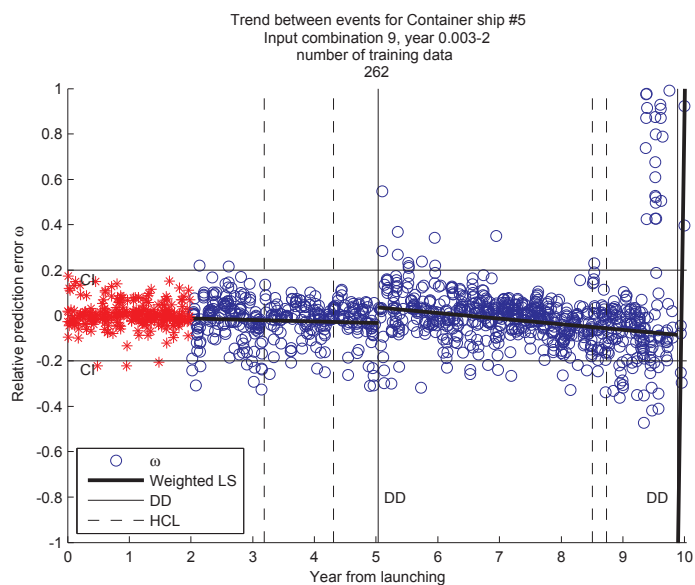


Figure D.36: Performance trend between dry-dockings of Container ship 5 using input variable setup 9 with all Noon report data, trained with the first two-year period

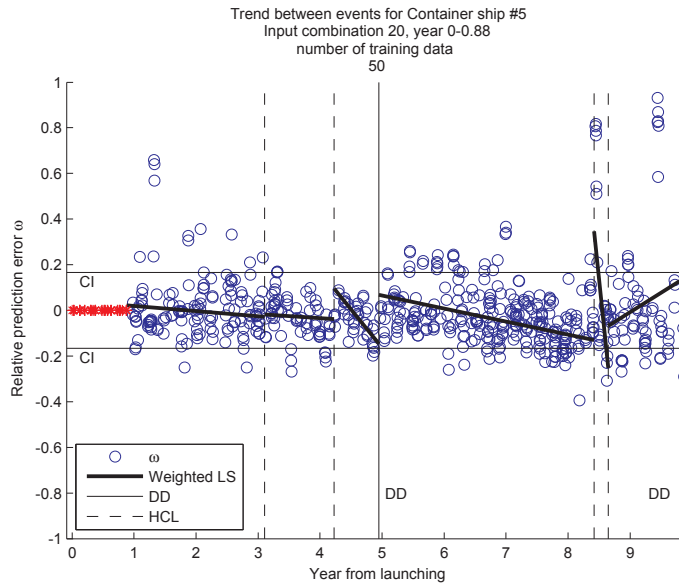


Figure D.37: Performance trend between all events of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first one-year period

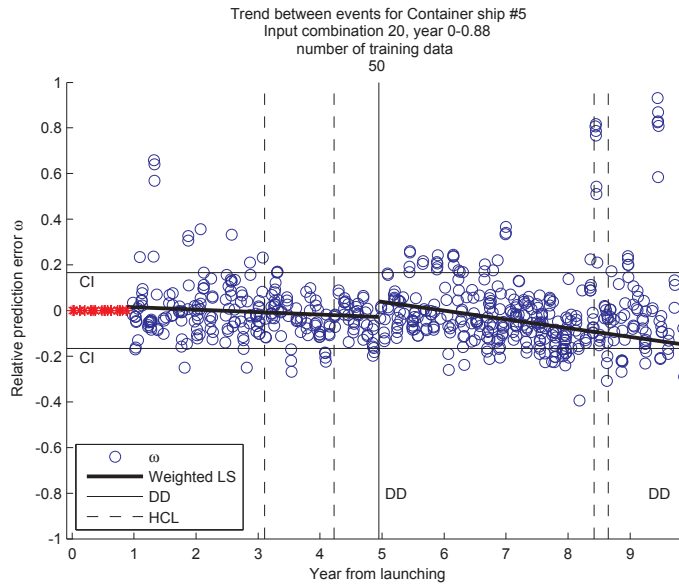


Figure D.38: Performance trend between dry-dockings of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first one-year period

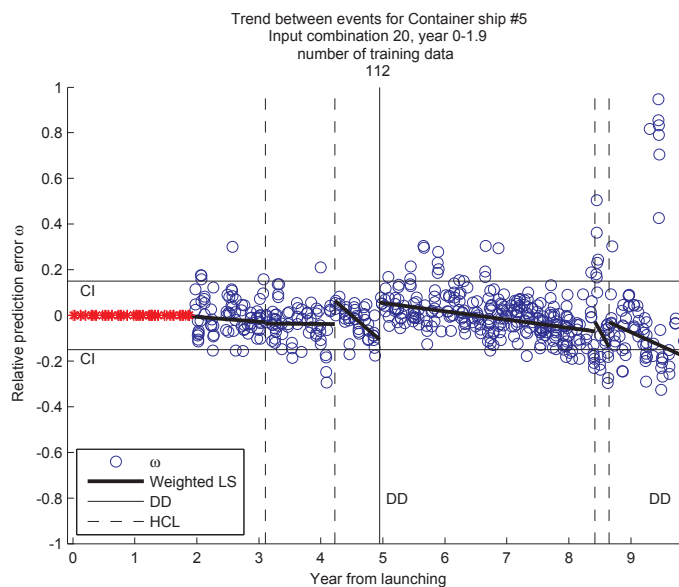


Figure D.39: Performance trend between all events of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first two-year period

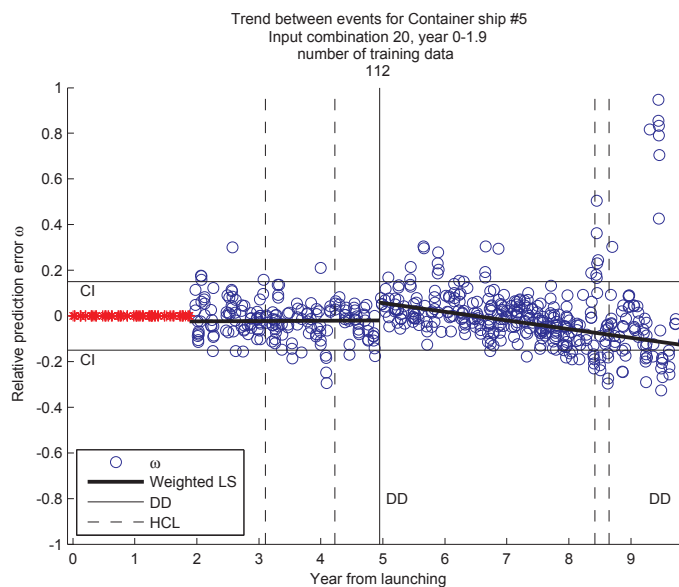


Figure D.40: Performance trend between dry-dockings of Container ship 5 using input variable setup 20 with data including hindcasts, trained with the first two-year period

Appendix E

Prediction results

The full accumulated histograms from the results in Chapter 4 are presented in this appendix.

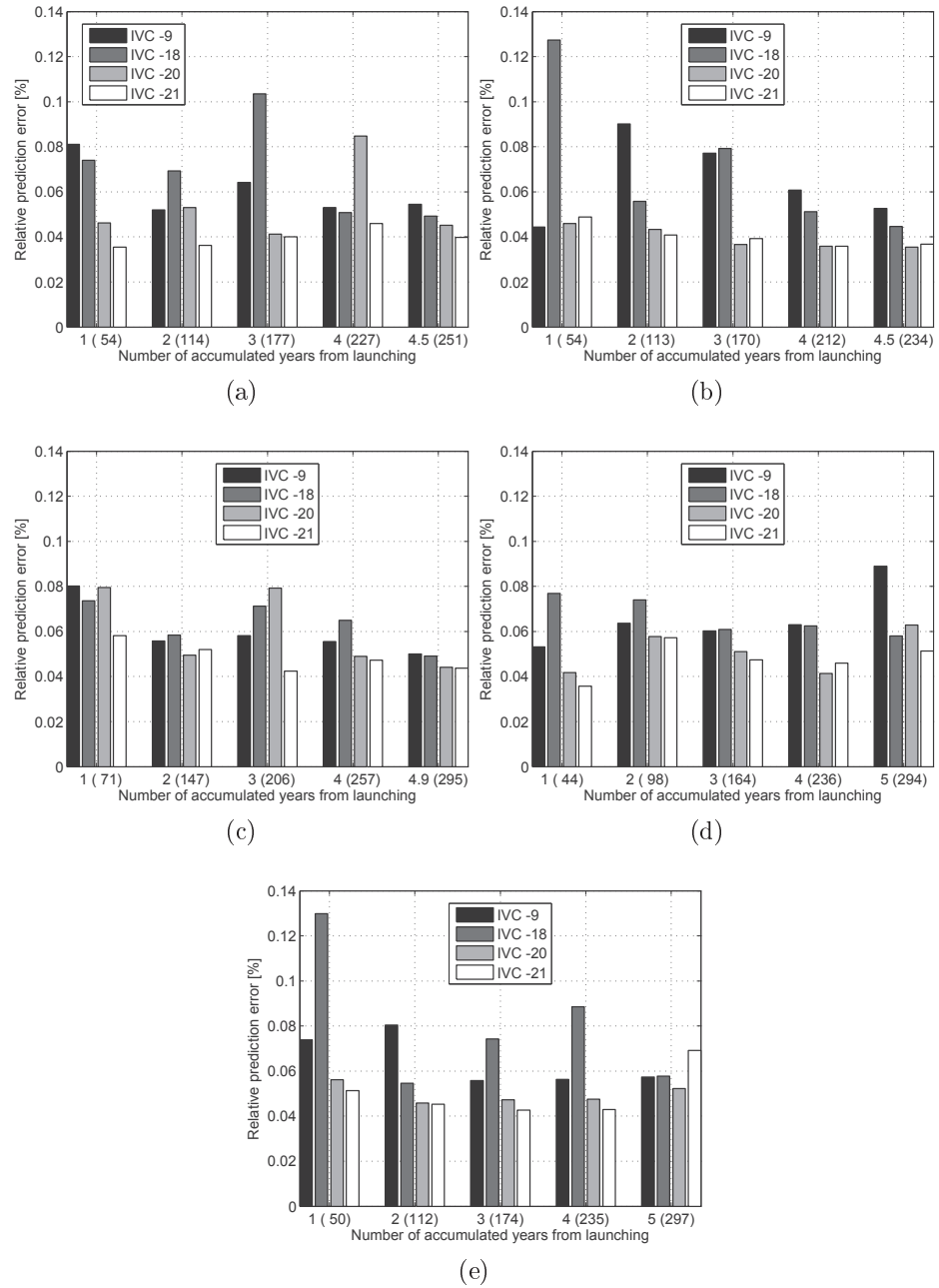


Figure E.1: Relative prediction errors from container ships #1-#5 trained with *IVC* 9,18,20,21 for five different periods of time starting at the launch with one-year intervals accumulating until the first dry-docking

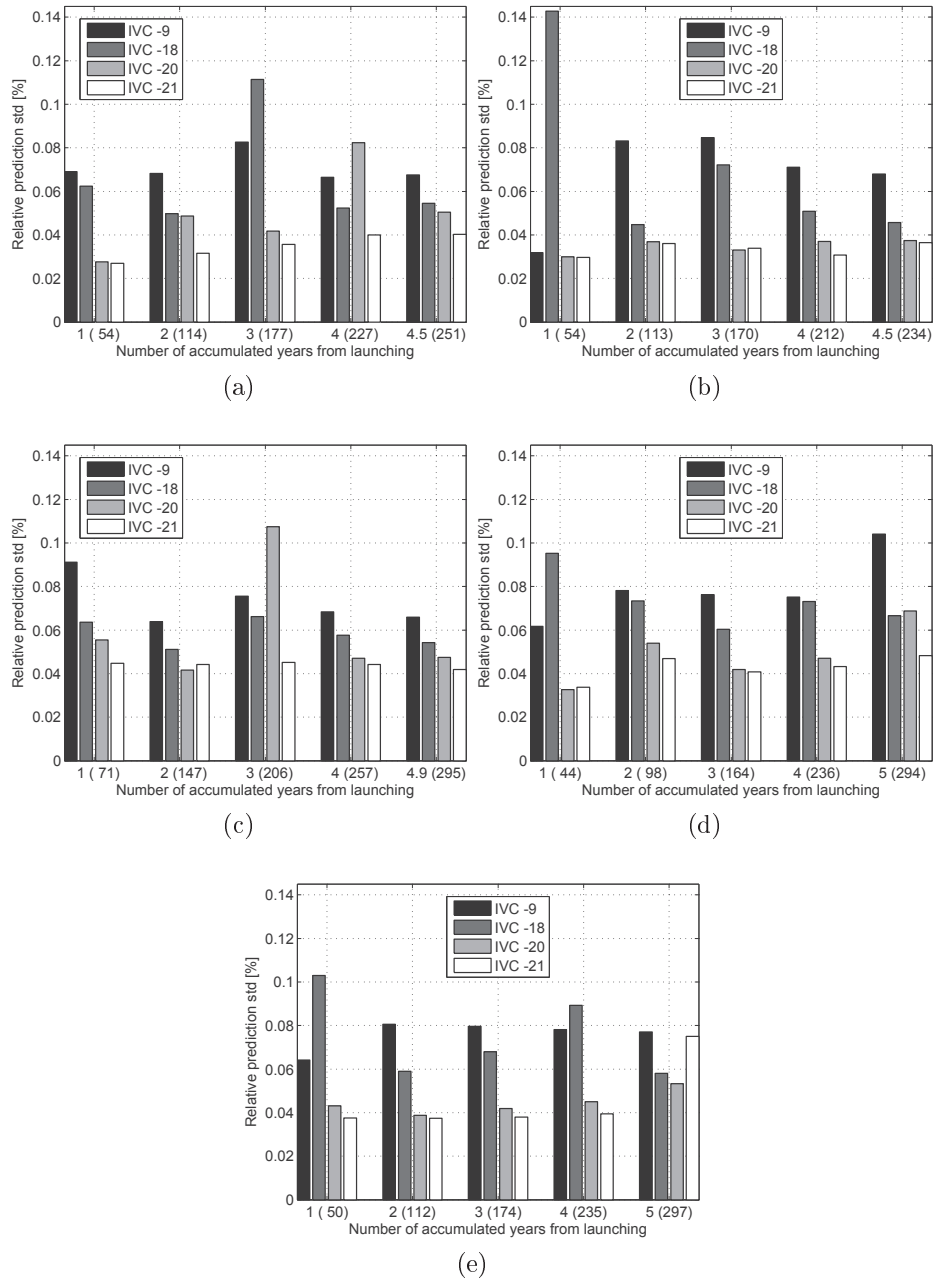


Figure E.2: The relative standard deviation from container ships #1-#5 trained with *IVC 9,18,20,21* for five different periods of time starting at the launch with one-year intervals accumulating until the first dry-docking

DTU Mechanical Engineering
Section of Fluid Mechanics, Coastal and Maritime Engineering
Technical University of Denmark

Nils Koppels Allé, Bld. 403
DK- 2800 Kgs. Lyngby
Denmark
Phone (+45) 4525 1360
Fax (+45) 4588 4325
www.mek.dtu.dk
ISBN: 978-87-7475-408-4

DCAMM
Danish Center for Applied Mathematics and Mechanics

Nils Koppels Allé, Bld. 404
DK-2800 Kgs. Lyngby
Denmark
Phone (+45) 4525 4250
Fax (+45) 4593 1475
www.dcammm.dk
ISSN: 0903-1685