MASTER'S THESIS | LUND UNIVERSITY 2017

Evaluating weather routing

Ludvik Brodl

Department of Computer Science
Faculty of Engineering LTH

ISSN 1650-2884
LU-CS-EX 2017-07

Evaluating weather routing

Ludvik Brodl
`ludvik.brodl@gmail.com`

June 9, 2017

Master's thesis work carried out at SMHI(Swedish Meteorological and Hydrological Institute), Shipping department.

Supervisors:
Jonas Thoursie - `jonas.thoursie@smhi.se`
Anders Rosen - `anders.rosen@smhi.se`
Jonas Skeppstedt - `jonas.skeppstedt@cs.lth.se`

Examiner: Krzysztof Kuchcinski, `krzysztof.kuchcinski@cs.lth.se`

Abstract

Today it is an accepted truth that consultation and evaluation of weather models by meteorologists can help a ship navigate the sea efficiently, using the weather and sea currents in the ship's favor. However, there exists little to no studies to support such a claim. In this report, we will first validate a simulation model by statistical means. This simulation is then used on different routes to compare them to each other. The most significant route comparison is comparing the initial route planned for a voyage, with the factual sailed route by the vessel. This report concludes that there is *no* statistical significant improvement to fuel and time economy when using land based meteorologists for the limited and specific dataset used.

Keywords: weather routing, SMHI, voyage analysis, voyage optimization, route evaluation, meteorological advice, weather factor

Acknowledgements

I would like to thank Jonas Thoursie for his counsel and guidance through the jungle of data that was used. But also for his welcoming nature which made me feel like part of the team and workplace.

I would also like to thank Anders Rosen for having faith in me and being endlessly supportive, positive and motivating.

Further thanks are also directed towards Marcus Sandbacka for his very knowledgeable input and discussions regarding the values in the dataset.

A thanks is also due to my wife, Fabiola Poblete, with whom I had quite a few discussions to spawn new ideas and angles of attack for the problems at hand.

Contents

Chapter 1

Introduction

To help the reader with some technical and oceanic terms in this report, a dictionary has been created and it is strongly recommended to use it during reading, see Appendix A.

Shipping companies' ships spend many hours in open water, transporting goods from one port to another. The scale of the transoceanic voyages in January 2016, was 90 thousand commercial ships, with a combined dead weight tonnage[1] of 1.8 billion [20]. To put this into the perspective of world trade, more than 90 percent of all international trade includes transport by sea [15].

1.1 Background

The total dead weight tonnage, available from commercial ships, is growing faster than the demand. This causes competition amongst shipping companies because there is *less* cargo than the ships can handle. The result of this supply-demand misbalance is that ships tend to slow steam (operating at below maximum speed), thus decreasing dead weight tonnage over time. However, cruising at lower speeds means that the time spent on each voyage is longer, which increases upkeep costs. On the other hand, slow steaming also has the positive effect of lowering fuel consumption.

Ship manufacturing and design technology advancements have led to larger ships, with increasing dead weight tonnage per ship and reduced emission per tonne cargo. The result is that fewer transport ships are built [15]. However, a larger ship implies a slower ship, and a slower ship implies more time spent at sea. This increase of time spent at sea makes weather routing even more crucial since every hour spent in unfavorable weather and sea currents burn fuel.

Although transportation via ship is very effective in terms of fuel used per transported tonne, it is not optimal for all situations. The largest drawback is the time it takes for the

---

[1] DWT(dead weight tonnage) is a term used for the maximum loading capacity of a ship.

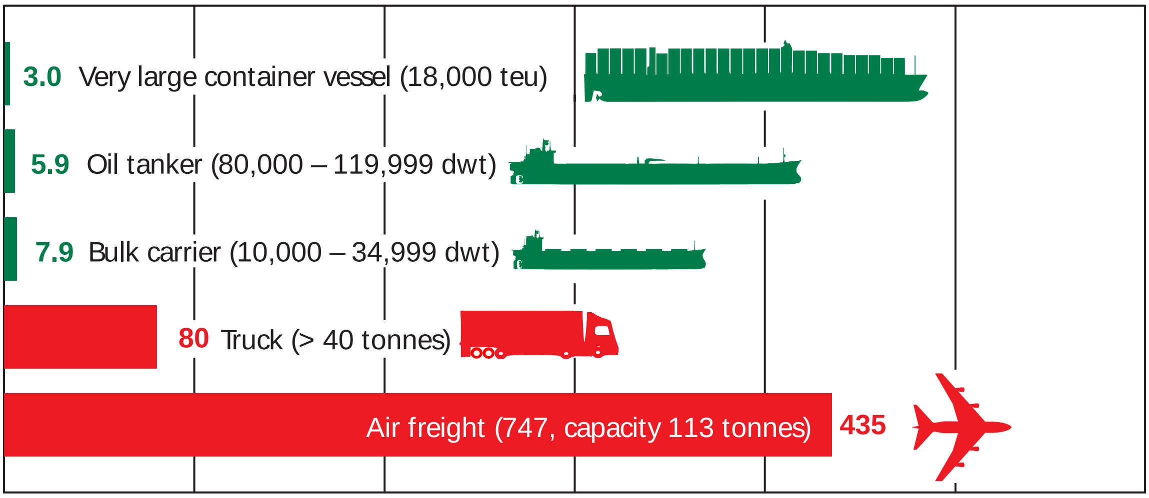

Figure 1.1: Grams of CO_2 emission for different transportation methods per tonne/kilometer. A lower value implies more weight transported for less emission [18].

transport. A ship going from China to England can take up to a month, but an airplane covers the same route in less than a day. Thankfully the fuel used is representative for the cost of the transportation mode and thus 90% of all goods, measured by volume, are transported via ocean. An overview of carbon emission by different transportation modes can be seen in Figure 1.1.

1.2 Weather routing

Weather routing is the term used for navigating at sea while taking both currents and forecasted weather into consideration to get the most favorable route. In order to make good decisions about routes, one must first have a forecast of the weather. There are many services today that provide 'very short'-(12 hours) up to 'long'-range(30 days to 1 year) forecasts. The methods to create these forecasts differ and have varying strengths and weaknesses in terms of accuracy, resolution, and complexity. Since transoceanic voyages generally last somewhere between one to 30 days, medium to extended range forecasts are used. The weather data, for which the results of this report are based on, are supplied by ECMWF[2] [13].

There can be many different routes for a given ocean, see for example the Atlantic ocean in Figure 1.2. Choosing the best route from A to B when there are n number of nodes between A and B is solvable in polynomial time, given that the cost to go from one node to another is constant [12]. However, add another dimension, variable costs, and it becomes an NP-hard problem which is solvable at best in exponential time. However, even if we manage to calculate the best route from A to B with the cost being the weather factor as a function of time, the weather factor function would still be a construct of a heuristic weather model. This is where experienced meteorologist can be very helpful, their experience allows them to make dynamic decisions with the help of prior experiences.

---

[2] European Centre for Medium-Range Weather Forecasts

Meteorologists can also take more dynamic constraints into consideration, such as: routing through waves of a certain height for a particular vessel with a particular cargo.

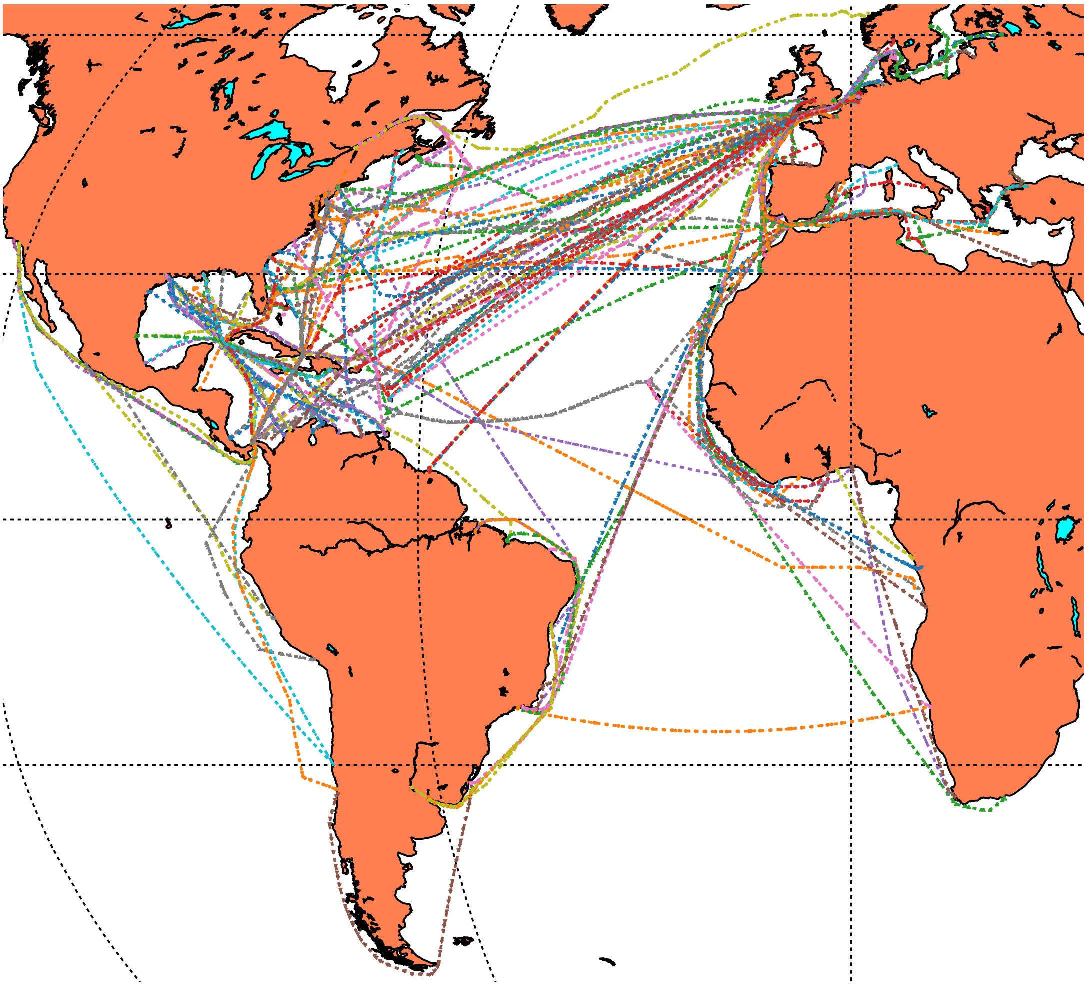

Figure 1.2: There are many different routes across the Atlantic Ocean.

1.3 Evaluating weather routing

The avoidance of extreme weathers and the increase of voyage safety is something that is considered when using weather routing. There is not much analysis needed to conclude that weather routing leads to safer routes. One need only consider the accuracy of extreme weather forecasts, which is readily available by most forecast providers. With knowledge of position and severity of weather anomalies for the next six days, with 80% accuracy, one can make better decisions to avoid these anomalies, see Figure 1.3 for an overlook of the development of weather models accuracy since 1998.

Today, many methods exists for weather routing and voyage optimisation. But there is little to no data of how well weather routing works in real cases. There are many businesses that claim to give a certain percentage of improvement in fuel economy through weather routing, but they are never backed up by real data. The reason for the lack of proof can

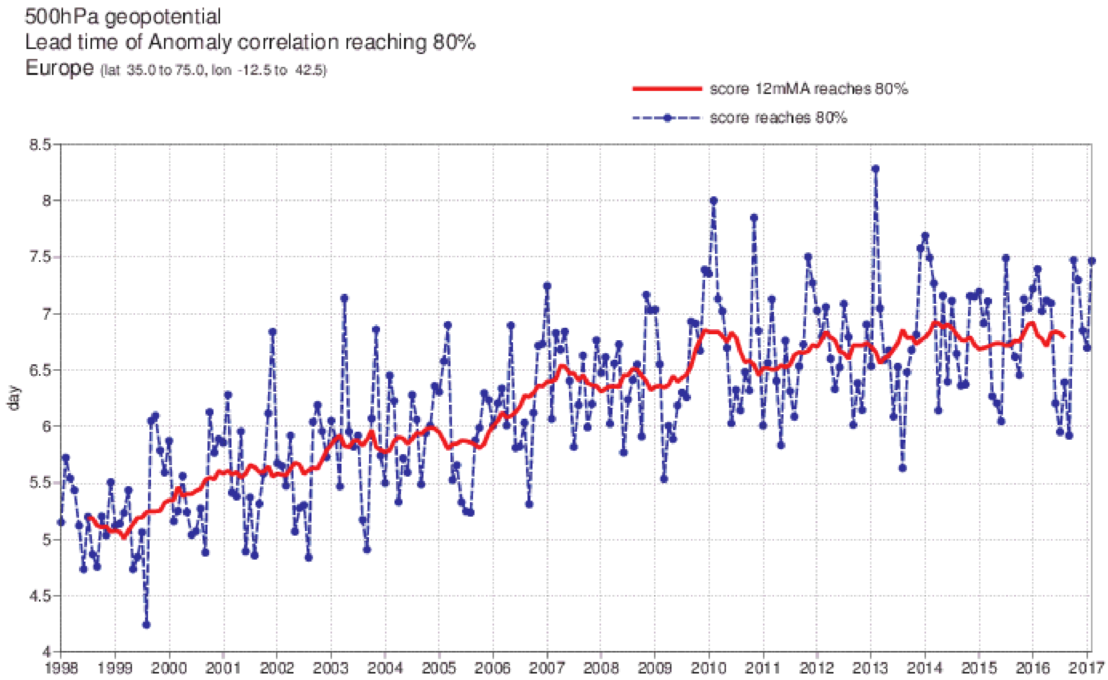

Figure 1.3: The plot shows for each month the number of days until forecast anomaly correlation dropped below 80%. The blue measurements represents the monthly mean mean while the red function represents a 12-month mean. This graph is based on ECMWF's medium range forecasts model in Europe (latitudes 35 to 75 and longitudes -12.5 to 42.5). The graph is created and supplied by ECMWF [14].

be due to the fact that the methods used to achieve claims of fuel savings are not very scientific and require some bold assumptions.

In this report we will attempt to fill this gap of uncertainty of how both fuel and time economy can be evaluated for voyages. The goal is to end up with a percentage of how much can be gained by using weather routing as opposed to just following the standard route.

Giving a captain route recommendation does not imply that the route will be followed. The route which the ship will take is the captains decisision. Also, weather routing can only be interpreted as advice, and not as in fact the best route, since it is a projection made by extrapolation of current weather data and past experiences by the one giving the advice. However, conclusions that forecasts help captains make better decisions can be made [10].

The main aspects that will be considered are fuel economy and time constrain in form of an ETA[3]. The reason for looking at fuel economy and time is because it is relevant to almost all form of transportation. Other aspects such as slamming, extreme weather conditions, energy used to heat cargo, etc. will be mentioned but will not be taken into account, simply because it would make this report too large.

---

[3] Estimated Time of Arrival

1.4 Contribution

Hopefully this project will lead to information regarding the actual effects of weather routing. It seems that it is given, that weather routing is something that every shipping company should use. But there exists no empirical evidence that it does work and will net an improvement in fuel and time economy. The main issue with this analysis is that something that has not happened, must be assumed would have happened based on a simulation of the weather and ocean current effects on a vessel.

We will show that there exists models that model weather and current factors very well. But the model used will not be described in detail since it is outside the scope of this report. The verification of the weather and current factor model is done with statistical comparison to ship performance baselines. This analysis assumes that the noon reports by ships are factual. At first glance someone might not understand why the noon reports might be tampered with. But in the world of shipping, where there is a lot of money involved, individual ships are chartered. In this charter party agreement certain things are agreed upon, such as the parameters; maximum RPM[4] of the main engine, minimum or maximum speed over ground, etc. If these parameters are not followed by the vessel, meaning a breach of terms, there can be consequences. However erroneous measurements and reports can not be assumed to be consensual but rather a result of the human factor of reporting manually. Nonetheless, the validity of noon reports is not in the scope of this report.

1.5 Related work

There are many methods for finding the optimal route across oceans. However since the problem is NP-hard[5] there is no perfect applicable method, only compromises. Optimally you would want to have an algorithm that provides:

Variable Speed Function of the speed over time for the voyage. Along with a:

Variable Route Set of points ((longitude, latitude) coordinates). Since the weather is not constant and changes with time, finding the optimal solution has exponential time complexity with each added possible point on the route [21].

Even though solving for the optimal route regarding time and fuel efficiency deterministically is not applicable, there are heuristic solutions. The weather routing commonly used today uses either a fixed speed or a fixed route. There are claims regarding the benefits of this type of heuristic weather routing and they are 2-4% net profit for CO_2 emission, assuming the fuel consumption is proportional to CO_2 emission [19]. This has later been estimated to be around 1.7% [22] and it is mentioned that it is hard to justify paying a premium for weather routing since there are no fuel-saving guarantees. If analysing with variable speed and route, some studies have found a reduction of emission by 52%, but at the cost of doubling the voyage time. The emission reduction is due to sailing at slower speeds and avoiding weather by alternating speed [16]. But thus far this is only estimations

---

[4] Revolutions per minute

[5] See section A.2 for a simplified explanation of NP-hard.

of the potential of weather routing with no hard data, showing empirical evidence of the results of weather routing in real cases.

1.6 Scope

In this report we will consider reported arrival times along with reported required arrival times, weather and current effects on vessels. Baselines supplied by the operators of the vessel will be used. Parameters that are taken into account when evaluating the benefits of weather routing are weather factor, current factor and timeliness. Because vessels differ a lot in size, shape, weight, dead weight tonnage, etc. it would be best to isolate to a certain type of vessel, but that is *not* done in this analysis, simply because there exists to few data points for such a filtering to give accurate results.

We feel obligated to repeat, in this report we only consider weather factor, current factor and timeliness of a voyage. A non-conclusive result in these very limited parameters does *not* conclude weather routing to be a waste of resources.

The sailed routes, which are compared to the initial route, are created with data gathered with the vessels public AIS (Automatic Identification System), which give accurate positioning of the vessel on a much higher time resolution than the noon reports. The baselines supplied by the operators will be assumed to be fact and are only loosely validated by the trained staff at the company for which the data has been collected from.

Chapter 2
Approach

The method used in this report is to first confirm the weather and current factors received from simulation. The verification will be done by statistically comparing data from completed voyages with simulation data. The weather factors and current factors are compared to a reported fuel consumption, RPM and engine load along with the voyage's baseline, as can be seen in Figure 2.2.

The weather and current factors are then used to perform comparisons between advised and non-advised voyages.

Finally the development of the characteristics of the initial (first suggested) route of a voyage, versus the actually sailed route, is examined. A simplified view of the dataflow of our implementation for comparing the routes can be seen in Figure 2.5.

2.1 Voyages

During a voyage, vessels make noon-reports with an interval of generally 24 hours. These reports include the coordinates gathered using GPS technology, on board of the vessel. There is also a time stamp for when they were on this coordinate. Alongside these two parameters there are many measurements included in the report, but there is no hard standard for what is included, nor how it is to be interpreted.

Along the voyage, waypoints are placed, either by the captain of the ship or a meteorologist. These waypoints consists of only coordinates. A combination of waypoints make up a route of a voyage.

Theoretically the weather affecting the vessel en route is a continuous function. However it is not feasible to regard it as such, since we do not have the required resolution of weather data. Thus, between the noon reports and waypoints; weather waypoints are created. The distance of the weather waypoints are determined by the speed over ground of the vessel. The weather waypoint contains only information regarding the weather affecting the vessel based on the ship characteristics and the weather provided by ECMWF.

The task of creating a route from a set of waypoints might seem trivial, but since they are two dimensional in the sense that they both span geographical location and time, it can sometimes be difficult to make a conclusion as to what route is appropriate to choose as the valid route. As can be seen in Figure 2.1 the route that would be created using only noon-reports could yield results which are difficult to trust.

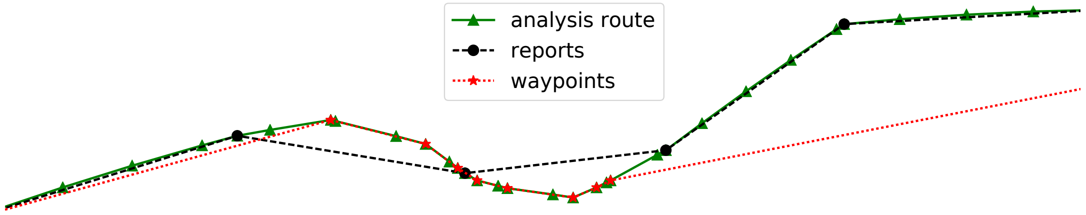

Figure 2.1: The analysis route is a composition of waypoints and reports. The analysis route is not always trivial to compute, in some cases the report is positioned before a way point in coordinates, but not in the time space. A simple interpolation using time would then create an analysis route going backwards.

2.1.1 Normalising measurements

Even though reports are more or less consistently produced daily, this does not represent the amount of time the vessel is steaming(engine operating). The fuel consumption, x, is normalised to 24 hours with the reported time spent steaming resulting in $\tilde{x}$.

$$\tilde{x} = x * \frac{24}{t_1 - t_0} * \frac{24}{t_s} \tag{2.1}$$

Where t_0 is the first report (in the time domain), t_1 is the following report and t_s is the time spent steaming between the two reports.

2.2 Speed estimation

When a vessel is propelling itself across water, there is a difference between the speed over ground and the speed at which the vessel would have moved if weather conditions were ideal and there were no losses caused by weather or sea currents going in unfavourable directions.

$$V_{sog} = D/t \tag{2.2}$$

where V_{sog} is the speed over ground, D is the displacement between the two reports, and t is the time between two reported geographical coordinates.

2.2.1 Weather and current factor

Weather factor, w_f, is a value representing the number of knots reduced from the vessel's performance speed due to weather. Current factor, c_f, works similar to weather factor, but for sea currents.

Producing an average weather factor, $\bar{w}_{f_i}$ requires weighted measurements, for the time spent in the calculated weather factor.

$$\bar{w}_{f_i} = w_{f_i} * \frac{t_{1_i} - t_{0_i}}{T} \tag{2.3}$$

Where T is the total time spent on the voyage. The arithmetic mean for $\bar{w}_f$ is then trivial to produce. The same concept is applied to current factor, performance speed and other measurements where there are several measurements during the course of the voyage.

To calculate the accuracy of the weather and current factor: performance speed, V_p, is compared to the baseline's inversion functions $V_b'(f)$. See section 2.3 for explanation of baselines.

$$V\Delta = V_b'(f_c) - V_p \tag{2.4}$$

2.2.2 Performance speed

The speed a vessel would have if the conditions were ideal (no weather or currents) is called performance speed and is calculated using the speed over ground V_{sog} and the weather and current factors:

$$V_p = V_{sog} - w_f - c_f \tag{2.5}$$

where V_p is the calculated performance speed.

The performance speed is the result of a simulation and needs to be validated. In order asses the correlation to a factual performance speed, a baseline speed is produced, see section 2.3 for an explanation of baselines. The baseline speed is the result of a baseline function, and a reported value by the vessel, and in this report we interpret the baseline speed as being heuristic. In the following chapter we will describe the baseline, see Figure 2.2 for an overview of the simulated performance speed validation.

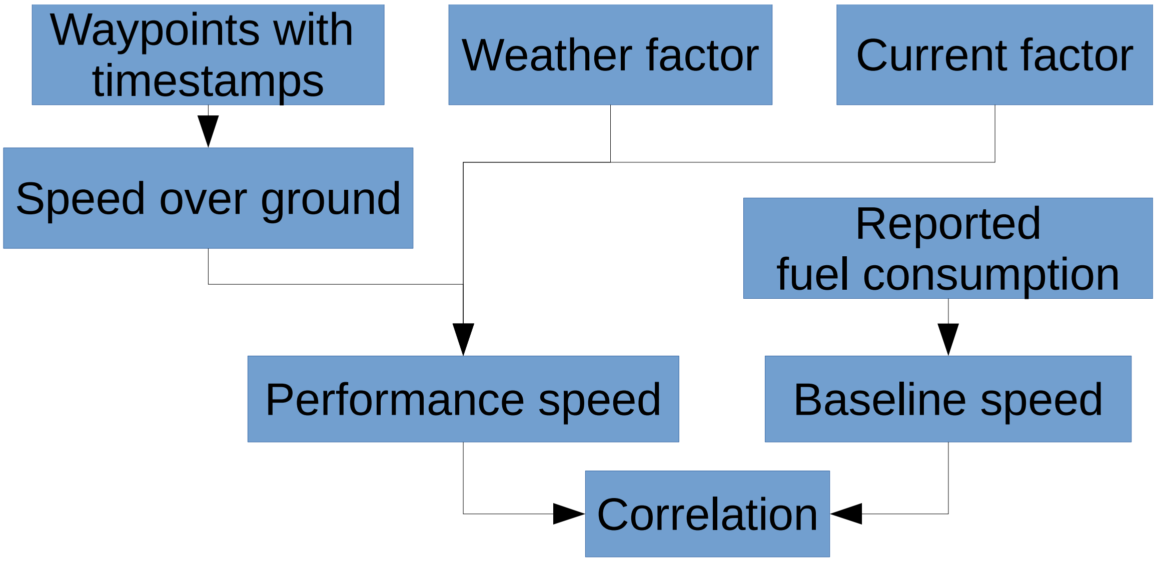

Figure 2.2: The accuracy of weather and current factor is controlled by comparing them to the loss of knots expressed with the baselines.

2.3 Baselines

To estimate the behaviour of a vessel, several baselines are produced. The baselines are a function of speed and give information about;

Fuel consumption The fuel consumption, of the main engine, reported in metric tons per day. There are other parts of a vessel that consume energy such as to maintain the temperature of cargo. The consumption unrelated to the main engine is reported as auxiliary engine consumption and is not part of the fuel consumption of the main engine. A typical fuel consumption baseline can be seen in Figure 2.3.

RPM The revolutions per minute of the main shaft connected to the propulsion system. This value is most likely measured using a tachometer and then averaged over the time difference between the two reports.

$$\text{RPM} = \frac{\text{Revolutions}}{t_1 - t_0} \tag{2.6}$$

However, the validity of the reported RPM can be discussed since we have no way of ensuring that it is not just a sample taken at the time of the report.

Engine load The amount of power used to propel the vessel forward. The unit is kilo watts.

The baseline is a function achieved by performing non-linear least square for datasets with the reported fuel consumption, RPM or engine output. To eliminate weather and current factors, only values when almost no weather was present are used. Degree of 4 is used for the fit. The baselines are defined as

$$V_b(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 \tag{2.7}$$

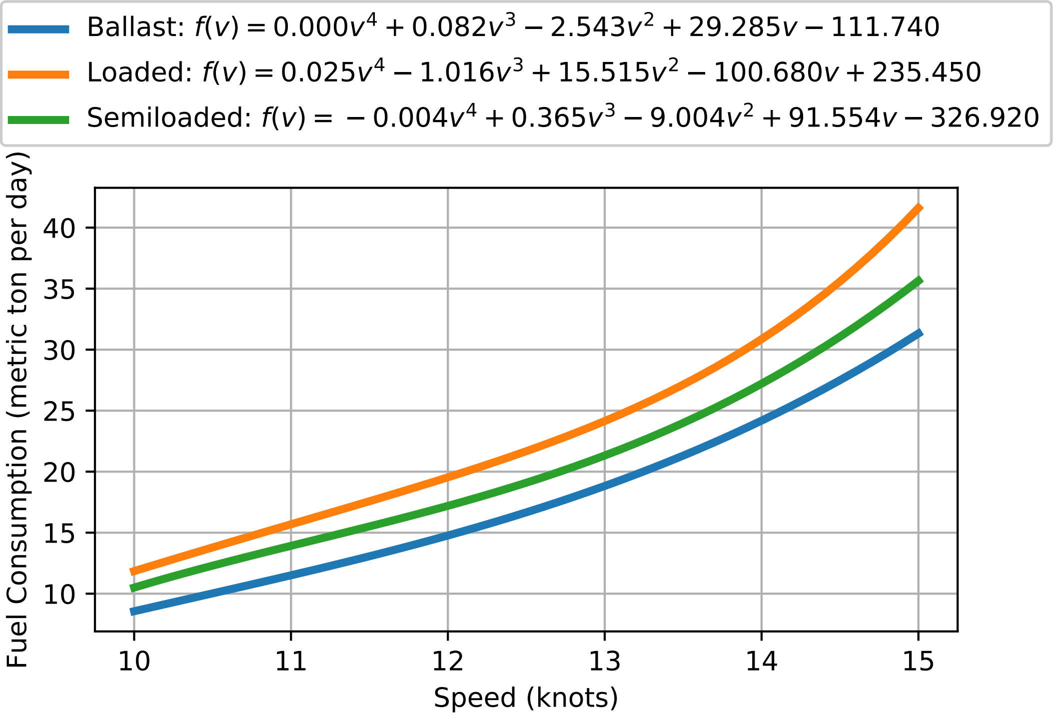

Figure 2.3: Example fuel consumption baseline for a vessel with three different loading conditions, see subsection 2.3.1 for an explanation of loading conditions.

Where x is the fuel consumption, RPM or engine load and c_i are the constants that give the characteristics of the baseline.

Since the inverted function generally can not be found algebraically for a polynomial function, the inversion is numerically found for a given interval for which the baseline is valid. The baselines and their valid interval are supplied by the ship operator. Further pitfalls and issues with baselines have been researched before [8], but we feel because the method for creating baselines are so simple there should be little issues with the validity of baselines supplied by ship operators.

2.3.1 Loading conditions

Because the force required to propel forward is heavily influenced by the draft of the vessel, and the draft is dependent on the number of tonnes loaded on the vessel, different baselines are created for different loading conditions. There is not a function for every tonnage value, instead it is common to have three different baselines;

Ballast Implies an empty cargo hold, because the vessel is designed to be used when loaded, no cargo gives the vessel a centre of gravity which is dangerously high above the water. Large waves could potentially make the vessel roll. Because of this, the vessel fills its ballast with sea water to lower the center of gravity.

Loaded A fully loaded vessel.

Semi loaded Because there is such a large gap between *ballast* and *loaded*, a baseline is also created for when the vessel is partially loaded.

However, there are shipping companies that have more loading conditions than these which should in theory further increase accuracy.

2.4 Non-advised vs. advised

The voyages in the dataset that is used in this report can be divided into two sets. One set where meteorological advice was given and another where the voyages were only monitored and data was collected. In order to evaluate if there is a difference in the two sets, several comparisons are made. These comparisons are in form of expected value and variance of the parameters;

Weather factor A value, in knots, representing the impact of the weather on the vessel. The weather factor depends on the speed of the ship[1]. Both maxima and minima are also interesting in this case because a high weather factor implies encountering significant weather anomalies.

Current Factor A value representing the number of knots that the vessel is slowed or sped up by the sea currents. Just like for weather factor; maxima and minima are also studied.

Timeliness On the commencing of a voyage the ship reports a required time arrival. The difference of the factual arrival and the required time arrival is defined as the timeliness of a voyage. According to a study by SeaIntel Maritime Analysis timeliness of voyages are not precise. During April 2015 only 42 % of voyages were within 24 hours of their first estimated ETA [9].

What is defined as late depends on interpretation of the reported required ETA. Along a voyage it is not uncommon for the required ETA to change. Is the vessel late if it arrives on time with its last modified required ETA or should it be held accountable for the first required ETA made during the commencing of the voyage? For this analysis, arrival time will be compared to the last reported required ETA since we made the qualified guess that it is the difference of the latest required ETA that costs them money in terms of port fees and other delays. But it should be mentioned that this choice does not give a fair result of how well weather routing services are in regards to assisting timeliness.

2.5 Voyage development

A voyage with a departure port and arrival port can take many different routes to avoid and exploit weather and sea current conditions. Generally there is an initial route when

---

[1] Identical vessel set ups, where the only variable is the speed over ground, results in different weather factors.

the vessel leaves the departure port. However when a voyage is advised, the route can change many times during the period of the voyage.

In order to assess the improvement of changing route in accordance to supplied advice, the development of the route is studied. Since we need to simulate the route versions, we can not rely on reports for accurate fuel consumption, coordinate and time values on which we can calculate a speed over ground and thus get a performance speed. This is the reason why we evaluate weather and current factor in previous steps, to ensure the simulation results are trustworthy.

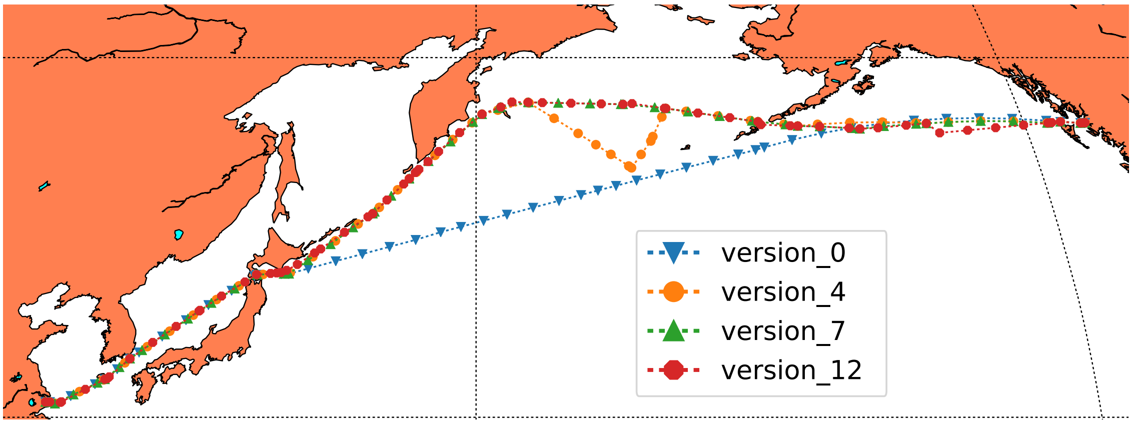

Figure 2.4: Route versions can affect the fuel and time consumption of a voyage. Most of the times though, the initial route differ very slightly from the final version. This specific route is from Prince Rupert in Canada to Shanghai in China.

In Figure 2.4, the voyage development dataset information for a single voyage can be seen. For all routes (version 0-11 in Figure 2.4), except for the sailed route (version 12), there are only waypoints which were created in collaboration between the vessel's captain and meteorologists and no GPS reported coordinates. The reason for the lack of GPS coordinates is that the routes are theoretical routes. Moreover, since weather and current factors requires a *speed in calm condition* parameter, one must be assumed. In this case, the *speed in calm condition* is constant for the entire route, this constant is taken from the vessel's commence of sea passage report at the start of the voyage.

2.6 Implementation

The data we have access to is:

Noon-reports The daily reports created and received from ongoing voyages.

Potential routes For ongoing voyages, route suggestions and recommendations are created and saved.

AIS data For a finished voyage, this gives a precise, in terms of GPS coordinates, route.

The data is used in different combinations to produce the results. Noon-reports are used for validating the performance speed. AIS data is used in both advised vs. non-advised and initial vs. sailed route analysis to find a finished sailed route. The routes data is used in the initial vs. sailed route analysis where we use the initial route as the control group.

As can bee seen in Figure 2.5 a number of modules were created to produce our results. The data from the three different data sources are pulled down with a downloader module, which uses HTTP GETs to request data. The downloader module orders and compiles the answers from the data sources for easier traceability in later steps. Individual routes are then sent one at a time, to Performa[2] with HTTP POSTs. Along with the simulation results; metadata, that was also received in the downloader module, are compiled into dataframes for easier handling. The compiled dataframes are analysed and sent through a pruning and filtering module that detects abnormal measurements from the databases[3]. An example of a typical route that would be filtered is a route with a section of travel speed of more than 30 knots. These pruning and filtering thresholds are heavily dependent on the input data and has manually been found in collaboration with meteorologists, the main purpose of this step is to remove dirty data from the results. The routes with dirty data are marked to enable distinction from clean data. After filtering the dataframes, they are sent to a summary module that produces statistical values. Summarizations are then piped into the statistical analysis and graph generation module to allow for human interpretation. The results from the last module is observed and analysed manually and compiled into the results of this report.

The modules are python scripts which were written to complete this report. In these python scripts, several python libraries were used:

python 2.7 The language the modules were written in. [4]

pandas 0.19.2 Creates dataframes, which are large matrices of data, used for it's simplicity and fast learning curve. [5]

seaborn 0.7.1 Creates plots with very little prior knowledge of plotting required, great for fast iterations of plotting data. [6]

statsmodels 0.6.1 Produces pearson correlation coefficients, perform Z-tests and examine correlations. [7]

numpy 1.12.1 Performs numerical calculations. [3]

matplotlib 2.0.0 Used for producing most of the graphs in this report. [2]

basemap 1.0.7 An extension of matplotlib, which allows for plotting of routes on a map. [1]

---

[2] Performa is the simulation that calculates amongst others: weather factor, current factor and arrival time.

[3] Abnormal data can also be detected by analysing the results from Performa.

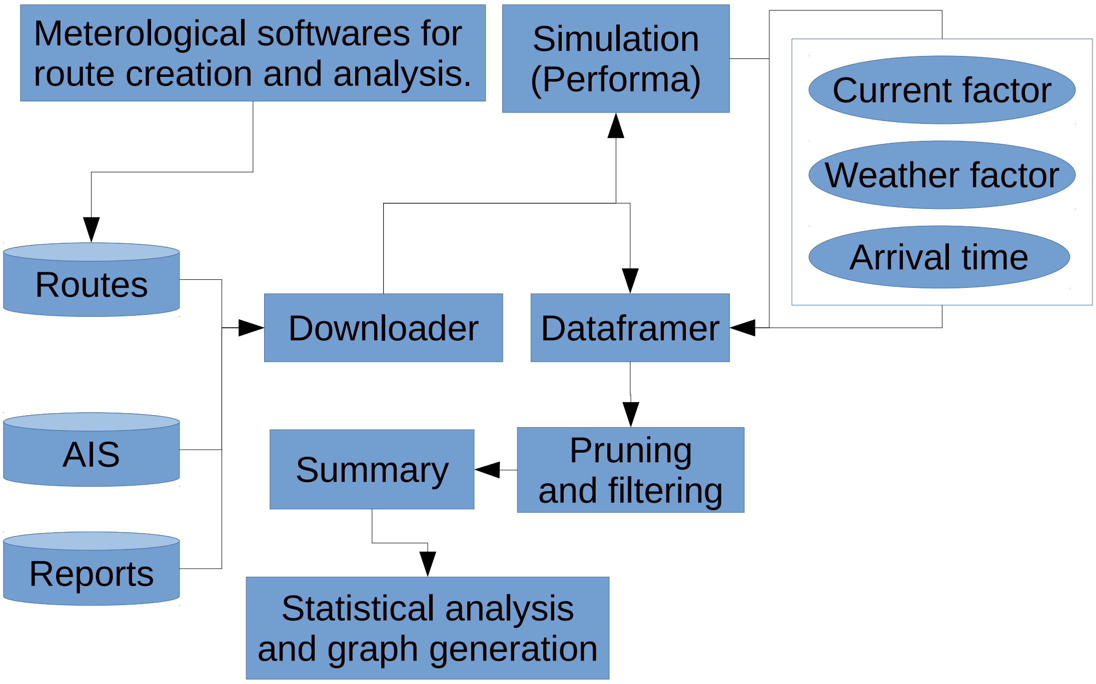

Figure 2.5: Dataflow of our implementation.

Chapter 3

Evaluation

First some properties and numbers of the dataset used for the different comparisons are declared. Then we proceed to show the results obtained by analysing the data, and finishing the chapter with a section for discussion of the results.

3.1 Experimental Setup

The dataset for validating baselines is limited to the year 2016. It includes 28 000 voyages, of which around 20 000 were deemed to be accurate enough to be part of the analysis dataset. The filtering and pruning of the dataset was not done to alter the results, but to remove extreme outliers that were no realistic, such as a speed over ground of 30 knots or a fuel tank status implying negative consumption. The main culprit for these outliers is believed to be due to the fact that there is no automatized way of making noon reports, they are performed manually. Not only is it manual work, but there is no real standard for what must be included and in what way. Thankfully due to recent efforts in decreasing emission a new standard has emerged regarding how and what to measure and report [11]. In the dataset used for this report, the human factor is assumed to be the sole cause for unrealistic measurements.

Unfortunately, not all of the 20 000 voyages had a baseline. The dataset in this report consists of 8905 voyages with the *loaded* loading condition, 159 *semi loaded*, 2541 *ballast* and 9143 *other*. Of these 8905 *loaded* voyages, only 3831 have a fuel baseline, 553 a RPM baseline and 164 an engine load baseline.

For the advised vs. non-advised voyage comparison the distribution is 4885 advised and 14656 non-advised voyages.

And the dataset for which the historic development of the routes are studied is limited to the time frame of 2016-10 up to and including 2017-4. It includes 151 completed voyages.

3.2 Results

This section is divided into the results of weather and current factor validity, non-advised vs. advised routes, and finishing with the results for initial versus sailed route.

3.2.1 Weather and current factor validity

The comparison between performance speed and the speed calculated using the baselines show a normal distribution for the error when using fuel or RPM baselines. For the power baseline the results are almost normally distributed, but contains some questionable distortion in the distribution's right tail.

To study the validity, of the calculated performance speed, the performance speed is compared to the baseline speed, see Figure 2.2 for an overview of the relation between the two parameters.

Looking at the three; fuel, RPM and engine load baseline speed comparisons, it is clear to see that the weather and current factor is heavily correlated to the actual resistance caused due to weather and current, see Table 3.1. This allows us to use the weather and current factors as a measuring tool for route analysis. In this section, the derived speeds from baselines are studied a bit further.

Table 3.1: Speed differences of performance speed calculated using Equation 2.5. Resulting in the difference between baseline speed and performance speed achieved using; calculated weather and current factors, and reported waypoints.

	Fuel	RPM	Engine Load
error mean	0.00±1.32	0.18±0.83	-0.22±1.32
# voyages	3831	553	164
# measures	33603	5061	2114
PCC	0.84	0.94	0.85
p-value	<0.00	<0.00	<0.00

Fuel

Examination of the distribution of the error for the fuel based baseline speed, give results that have normal distribution, with a mean of 0.00 and a standard deviation of 1.32, see Figure 3.1. The normal distribution is a good sign that the errors are random and there is no systematic error which would most likely result in uneven distribution.

Correlation between the performance speed and the fuel based baseline speed are present in the dataset, see Figure 3.2. The plot shows promising correlation with a PCC[1] of 0.84. There are some very rare outliers in the plot, since they are seemingly random, no attempt to isolate their cause and filter them is made, which could in theory increase the PCC even further.

---

[1] Pearson correlation coefficient.

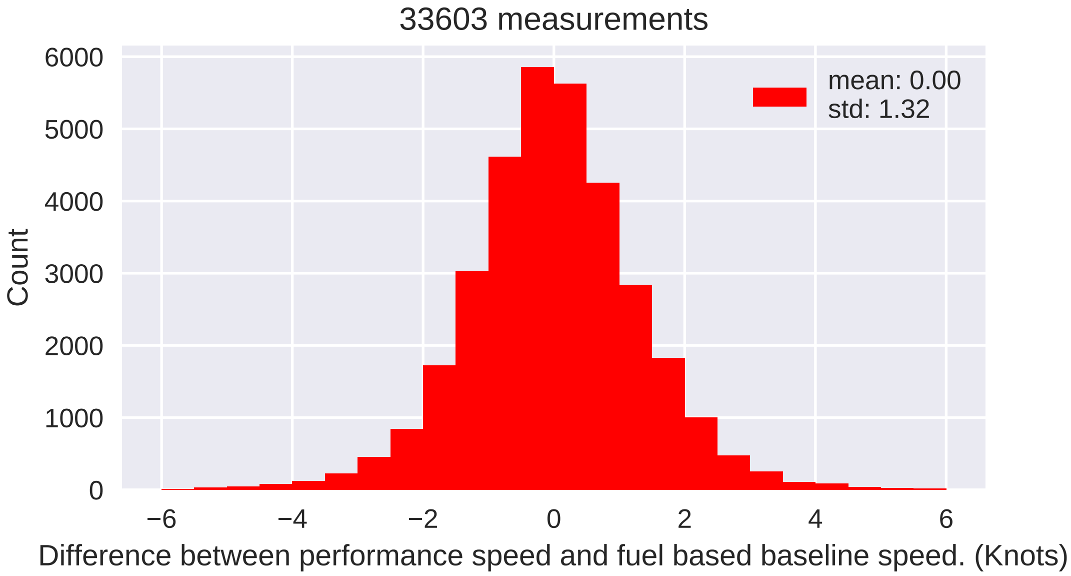

Figure 3.1: A histogram showing the error of the performance speed compared to the fuel based baseline speed, see Equation 2.4.

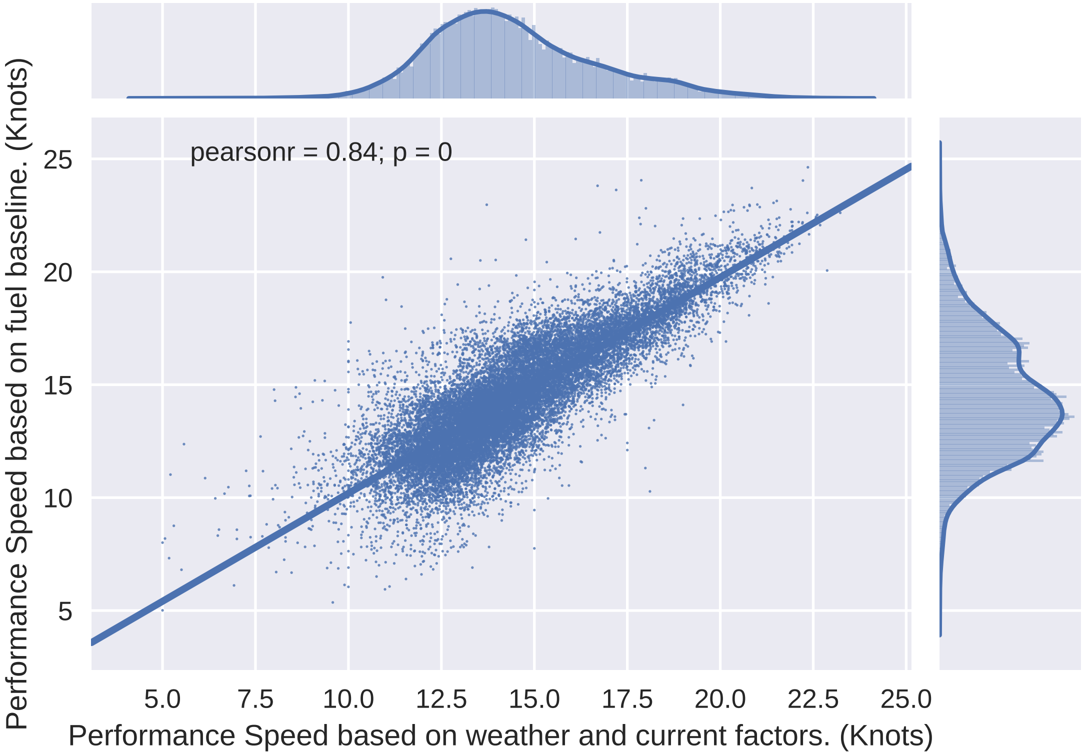

Figure 3.2: Plotting the performance speed calculated using weather and current factor along with the fuel based baseline speed. They are heavily correlated with a PCC of 0.84.

RPM

The same comparison of error distribution and correlation is studied for the RPM-based baseline speed, see Figure 3.3. We have roughly one-fifth as many measurements as for the fuel based baseline speed, resulting in a cruder error distribution. However, the linear correlation is a lot smother and the PCC is an impressing 0.94, see Figure 3.4.

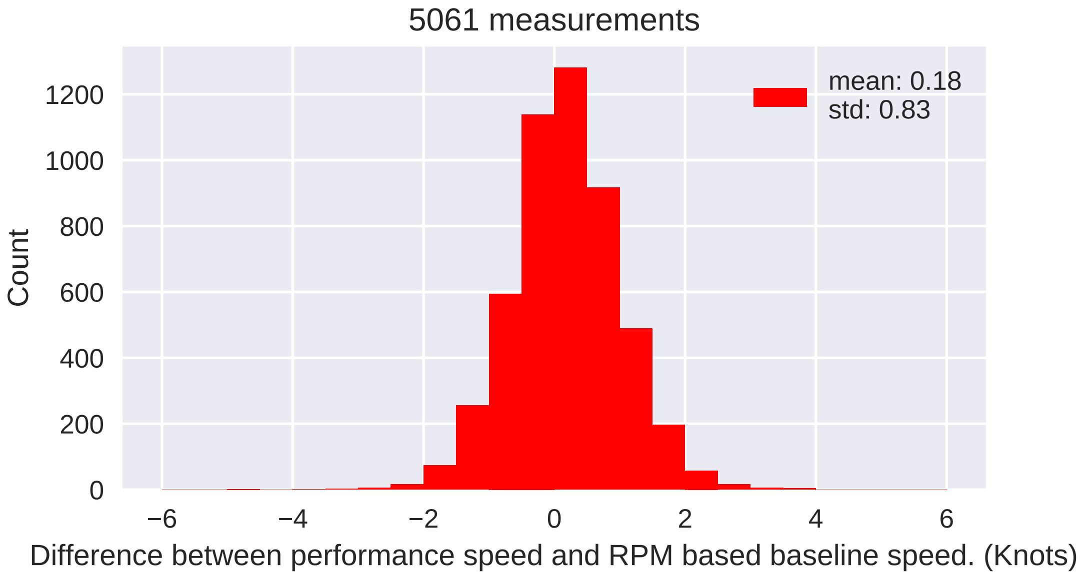

Figure 3.3: A histogram showing the error of the performance speed compared to the RPM based baseline speed, see Equation 2.4.

Figure 3.4: Plotting the performance speed calculated using weather and current factor along with the fuel based baseline speed. They are heavily correlated with a PCC of 0.94.

Engine load

Finally, the engine load based baseline speed is studied. Here we only have one tenth as many measurements as for the fuel based baseline speed. The distribution is even cruder and it seems to be weighted in the negative domain, see Figure 3.5. When looking at the correlation between the performance speed and the engine load baseline speed it is easy to spot a problem with the reported power. For several different measurements that main engine load has stayed constant, it is seen by the seemingly connected horizontal data points, see Figure 3.6. The baselines have all been studied manually and they are never constant, meaning the input engine load into the baseline is the only way to receive constant speed. This has lead us to believe there is some systematic error in how the engine load is reported or interpreted.

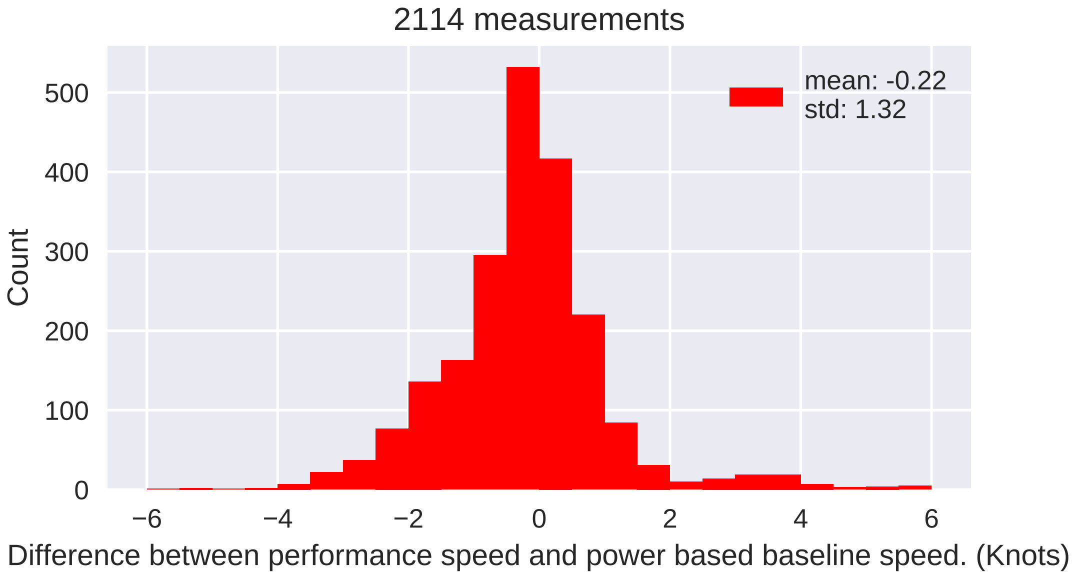

Figure 3.5: A histogram showing the error of the performance speed compared to the engine load based baseline speed, see Equation 2.4.

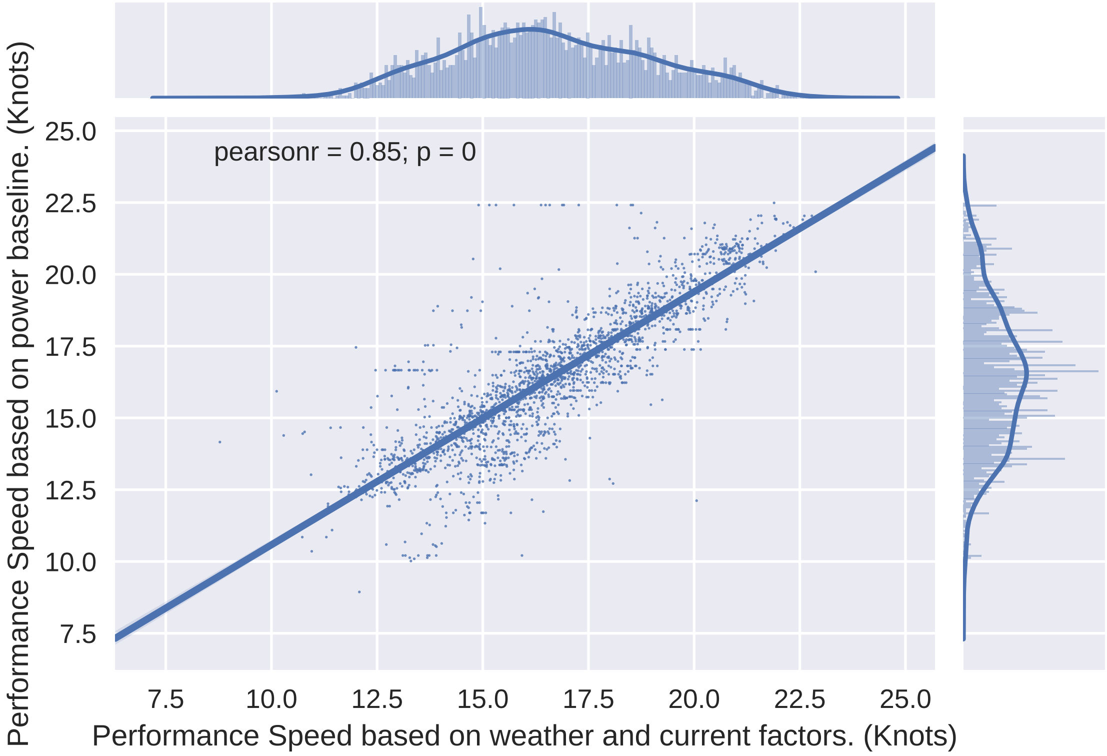

Figure 3.6: The correlation between the true speed calculated using main engine power and performance speed calculated using weather and current factor are heavily linearly correlated with a PCC of 0.85.

3.2.2 Non-advised versus advised

The dataset of non-advised routes is compared to the routes that were given advice en route by land-based meteorologists. There are some issues with this dataset, limiting it's validity, which will be discussed later. First, we will look at the weather and current factor parameters of the voyages, and then we will examine the timeliness of the voyages.

Weather factor

Comparing the weather factor of advised voyages to the weather factor of non-advised voyages show that the average weather factor is more favourable and more predictable for non-advised routes, see Figure 3.7 and Figure 3.9. It is also clear that the worst weather encountered during the voyage is less dramatic for non-advised routes since the lowest simulated weather factor is lower for advised routes, see Figure 3.8. Just looking at the results, Table 3.2, it is tempting to conclude that it is worse to get advice than to sail on your own, but keep in mind that correlation does not imply causation, more on this in section 3.3.

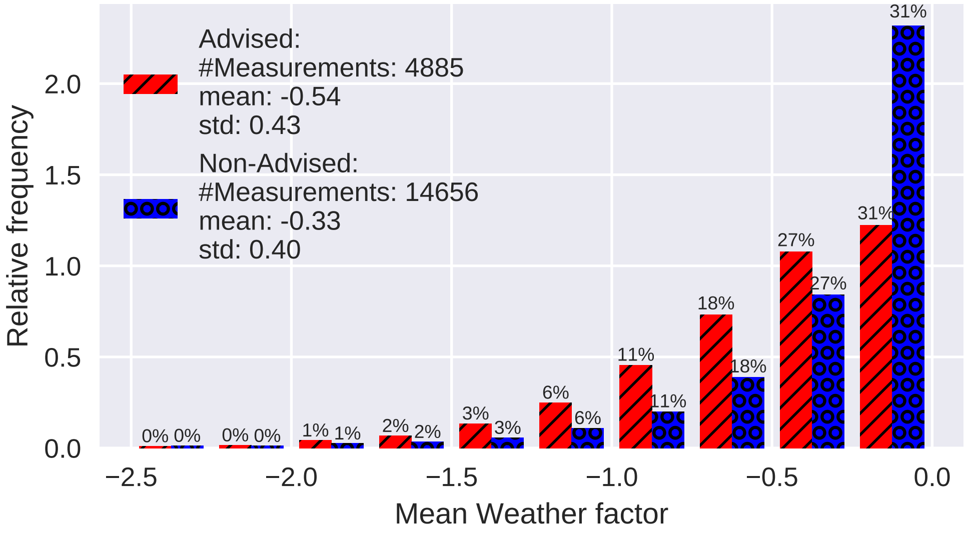

Figure 3.7: Distribution diagram of the average weather factor for advised(red) vs. non-advised(blue) voyages. A lower value implies more severe weather.

Table 3.2: Weather factor measurements of non-advised and advised voyages.

	weather factor(knots)		
	mean	std	min
advised	-0.54±0.43	0.39±0.32	−1.58±1.22
non-advised	-0.33±0.40	0.16±0.24	-0.68±0.82

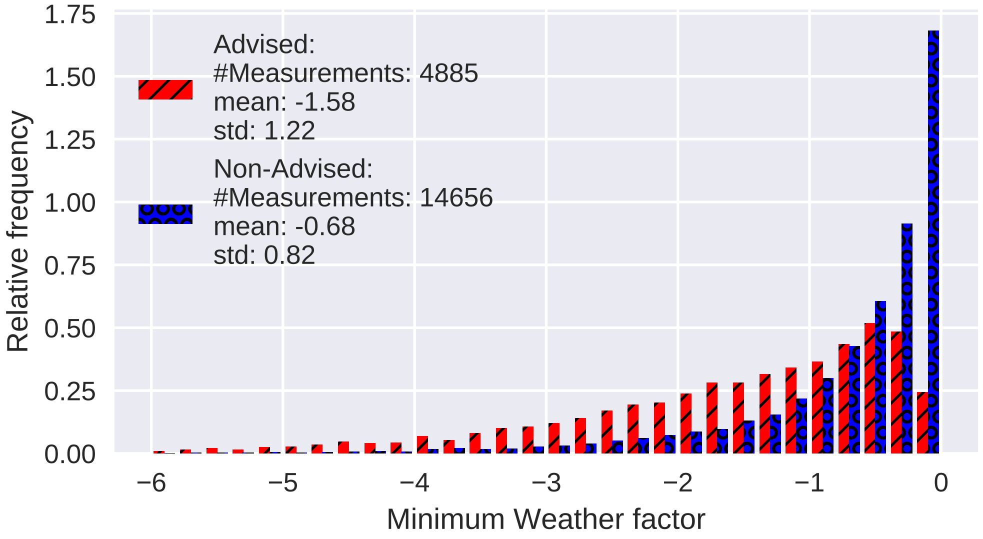

Figure 3.8: Distribution diagram of the lowest simulated weather factors during a voyage. A lower value implies more severe weather.

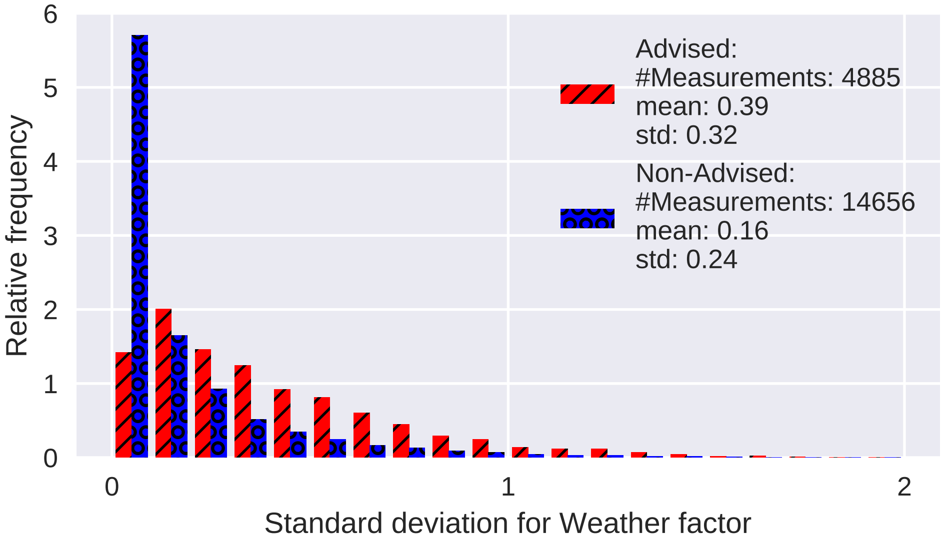

Figure 3.9: Distribution diagram of the standard deviation of simulated weather factors during a voyage. A large value implies large variations in simulated weather factors.

Current factor

When comparing current factor of advised voyages to the current factor of non-advised voyages we get similar results as for the weather factor, see Figure 3.11 and Figure 3.12. It should be noted that, unlike the weather factor, the current factor can be positive, implying a current that is going in the same direction as the vessel, increasing its speed over ground for the same amount of fuel. Just like for weather factor, the results imply it's better to have no advice en route Table 3.6. But once again, recall that correlation does not imply causation.

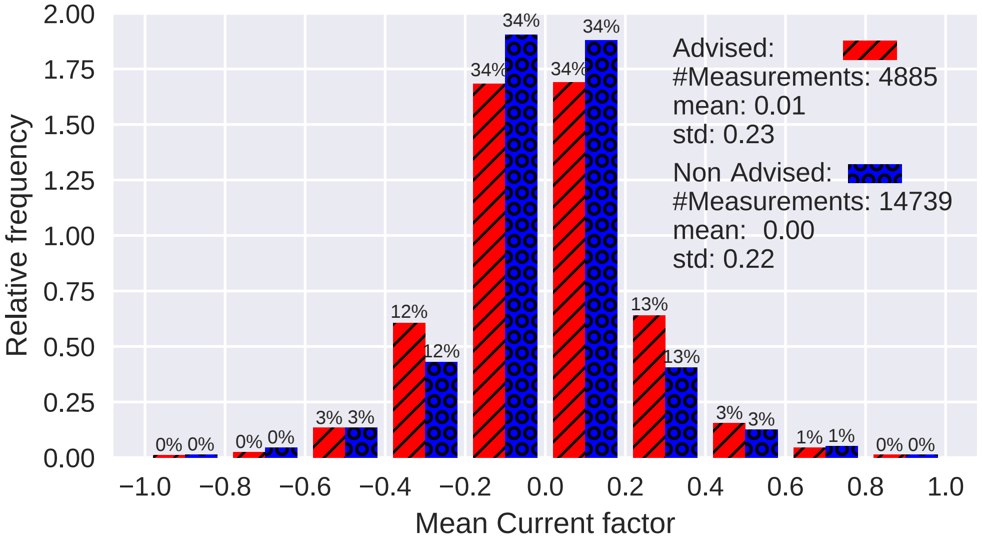

Figure 3.10: Distribution diagram of the lowest simulated current factors during a voyage. A low value implies current going in unfavourable direction relative the vessel's bearing.

Table 3.3: Current factor measurements of non-advised and advised voyages.

	current factor(knots)			
	mean	std	min	max
advised	0.01±0.23	0.38±0.17	-0.88±0.56	0.90±0.59
non-advised	-0.00±0.22	0.15±0.17	-0.28±0.42	0.27±0.41

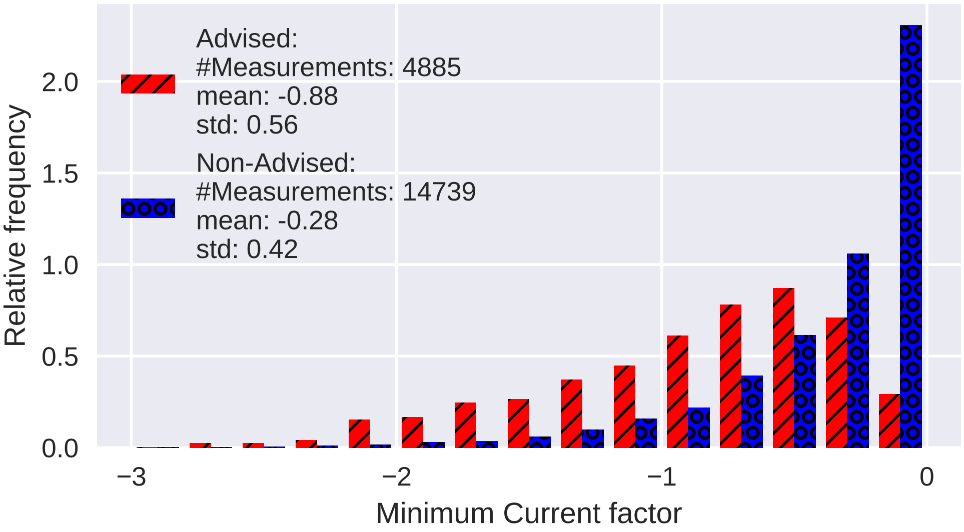

Figure 3.11: Distribution diagram of the lowest simulated current factors during a voyage. A low value implies current going in unfavourable direction relative the vessel's bearing.

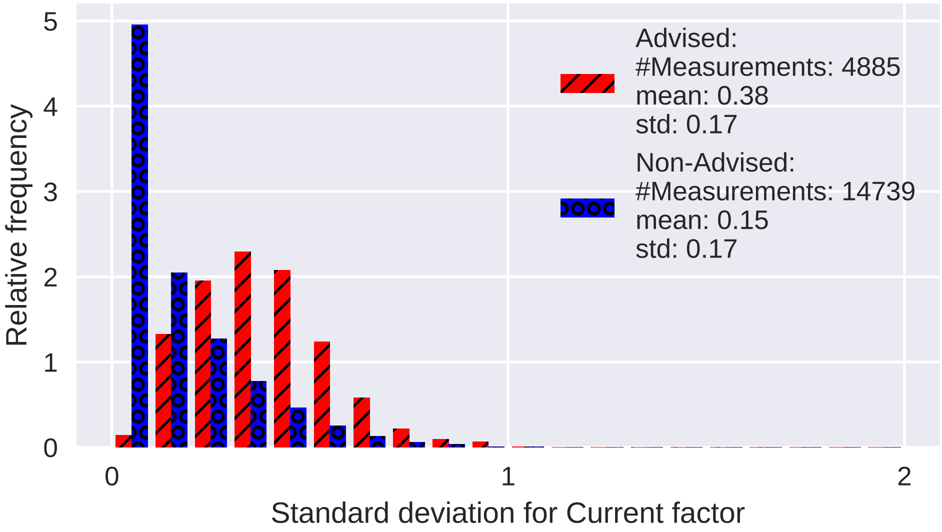

Figure 3.12: Distribution diagram of the standard deviation of simulated current factors during a voyage. A large value implies large variations of the weather factor during the voyage.

Timeliness

When evaluating a voyage, it is common to look at the timeliness of the voyage. Meaning the required ETA minus the arrival time. Recall that the required ETA can change during the course of a voyage. We compare the *last* reported required ETA with the arrival time, see Figure 3.13 and Table 3.4. The timeliness definition used here differ from the one in subsection 3.2.3, where we compare to the *initial* required ETA and not the *last* reported.

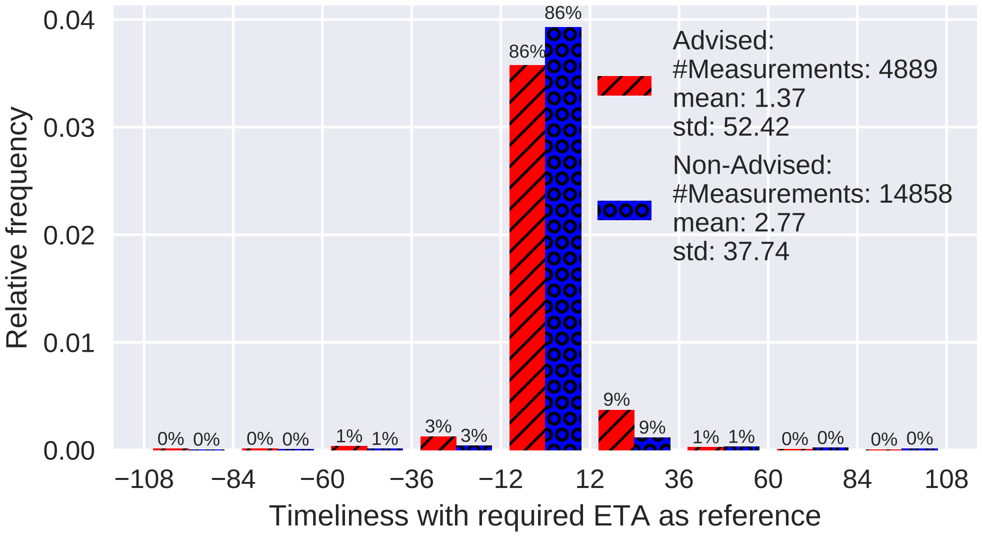

Figure 3.13: Distribution diagram of the timeliness of a voyage. A negative value implies being later than the *last reported* required ETA.

Table 3.4: The time difference between arrival and the last reported required ETA of the vessel. A negative value represents being later than the required ETA.

	timeliness(hours)
advised	1±52
non-advised	3±38

3.2.3 Voyage development

The initial route, defined as the first one made, by either the vessel's captain or as a standard route by the meteorologists is compared to the sailed route. For this comparison to work, we make the assumption that the vessel's sailed route is affected by the advised routes given en route.

Weather and current factor

For 151 voyages we compare the initial route to the sailed route. The results show that there is little difference to the distribution of the *mean* and *minimum* weather factor for the two different route sets, see Figure 3.14 and Figure 3.15. The expected value and variance of weather and current factor are almost identical for the two route sets, see Table 3.5 and Table 3.6. Z-tests confirm that the distributions are not statistically different with p-values close to 1[2].

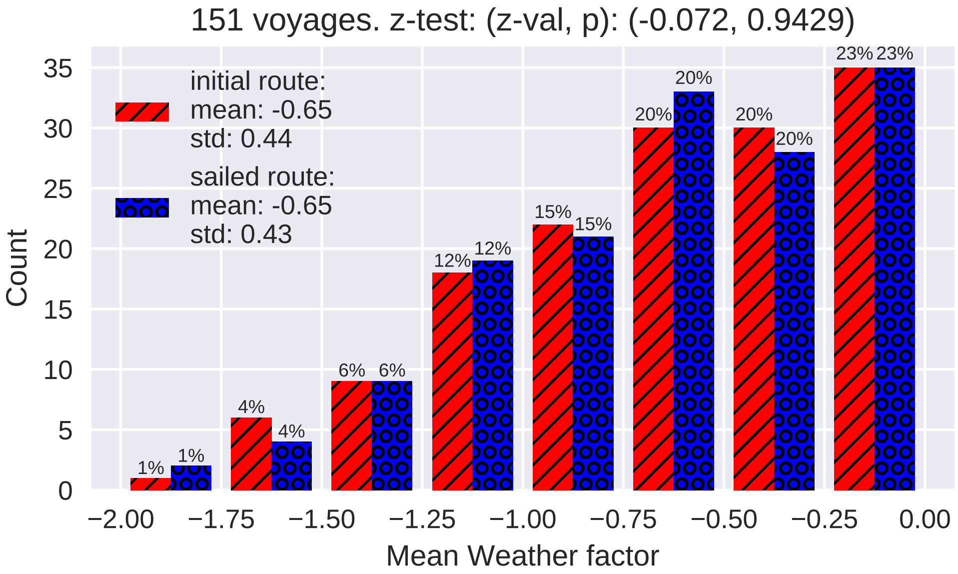

Figure 3.14: The initial and sailed routes' simulated mean weather factor are very similar.

Table 3.5: The initial and sailed routes' simulated weather factors are very similar in characteristics.

	weather factor(knots)			
	mean	std	min	max
initial route	-0.65±-0.44	0.58±0.43	-2.44±1.84	-0.02±0.05
sailed route	-0.65±0.43	0.58±0.40	-2.43±1.71	-0.02±0.04

---

[2] A Z-test is a statistical tool that takes two distributions and examines them in order to find a measurable difference between the two. A z-test p-value of 1 implies zero chance of a difference between the distribution, while a p-value of 0 implies 100% chance that there is a difference between the two distributions. $1 - p$ is a fraction representing the chance that the distributions are different. For example: $p = 0.90 \rightarrow 1 - p = 1 - 0.90 = 0.10 = 10\%$ chance that the distributions are different.

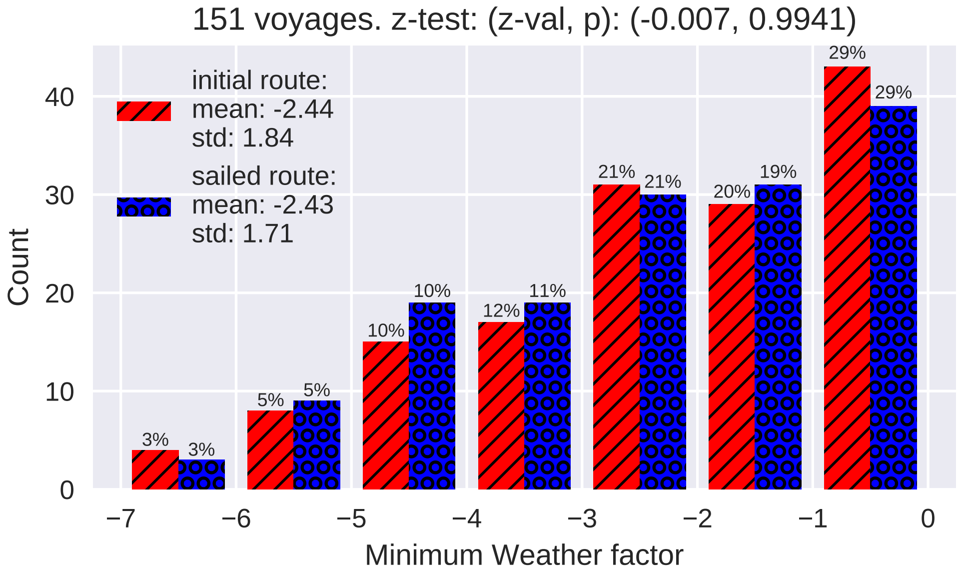

Figure 3.15: The initial and sailed routes' lowest simulated weather factor are very similar.

Table 3.6: The initial and sailed routes' simulated current factors are very similar in characteristics.

	current factor(knots)			
	mean	std	min	max
initial route	0.03±0.17	0.31±0.13	-0.72±0.37	0.83±0.49
sailed route	0.03±0.18	0.31±0.13	-0.71±0.33	0.85±0.48

Timeliness

Examining arrival times of a simulation run with the same starting time for both the initial and sailed route show that there is no real improvement to timeliness, see Figure 3.18 and Table 3.7. Even though there is a small difference in timeliness and the relatively small sample size of 151 results in z-test giving very high p-values. In an attempt to find any significant differences between the initial and sailed route we also examined the total route distance, and the fraction of the total distance traveled due to weather.[3] The parameters; *sailing time*, *route distance* and *fraction of distance due to weather* also do not show any difference in distribution that is noteworthy, see Table 3.8, Figure 3.18, Figure 3.17 and Figure 3.16.

---

[3] The fraction represents the percentage of the traveled route that is created due to weather and currents. For example: Sailing from A to B with the performance speed of 8 knots. And weather and current factor of 1 knot each, the result would be that 0.20 of the distance covered is due to weather and currents.

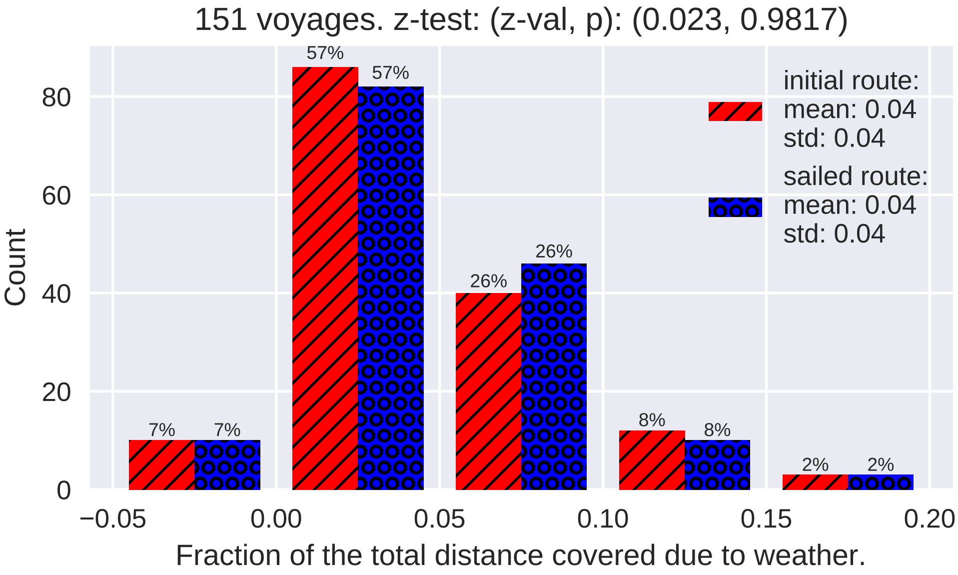

Figure 3.16: Normalising the distance propelled for a route, this value represents the percentage of the route-distance travelled due to resistance in weather and current.

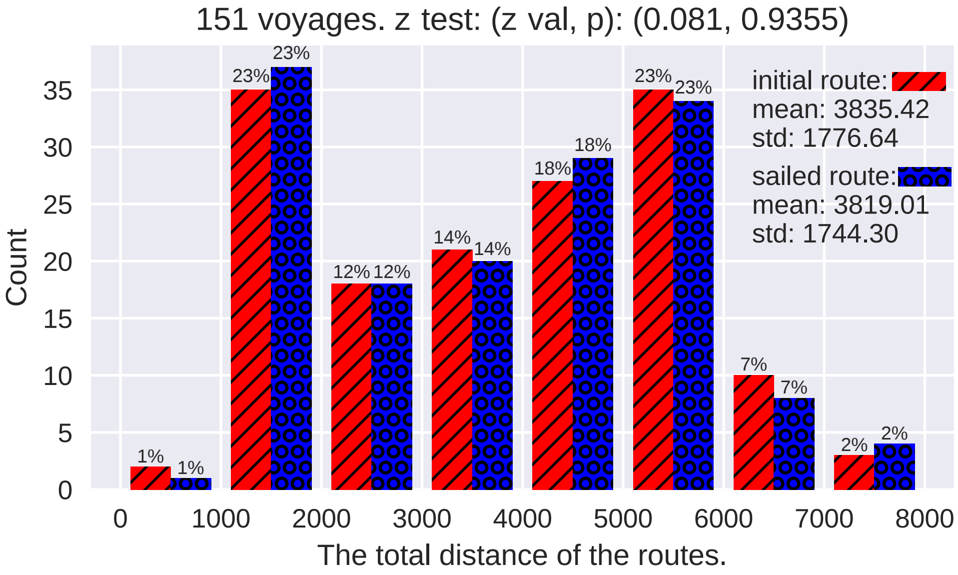

Figure 3.17: The total distance between the waypoints that are input to the simulation, this histogram is not part of the output of the simulation but a way to control that the input route is not simply shorter or longer for the initial or sailed route.

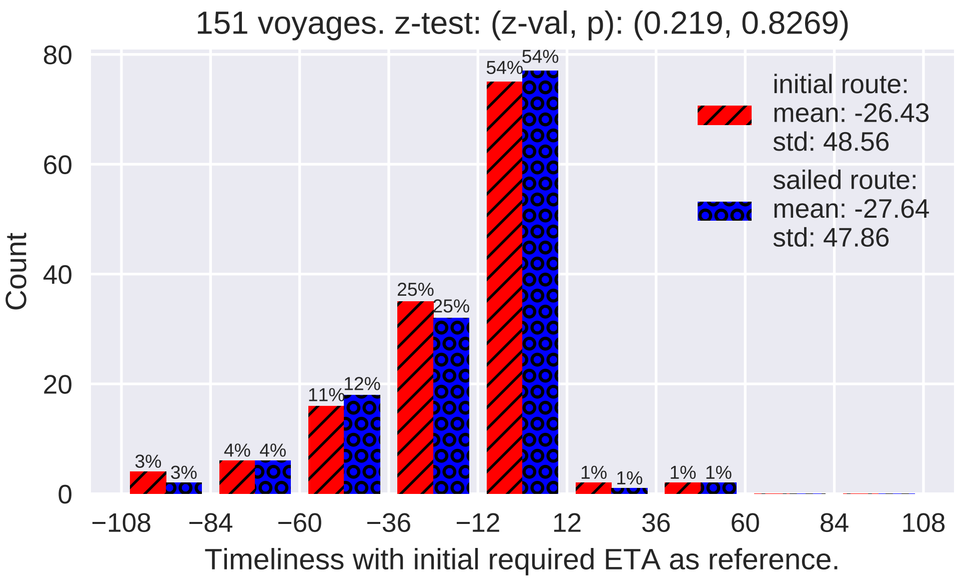

Figure 3.18: The time difference between arrival and the initially reported required ETA for the voyage. A negative value represents being later than the required ETA.

Table 3.7: The time difference between arrival and the initially reported required ETA of the vessel. A negative value represents being later than the required ETA.

	timeliness(hours)
intial route	-26±49
sailed route	-27±48

Table 3.8: The initial and sailed routes' sailing time and fraction of distance covered due to weather are very similar. As a control, the distance of the routes was also noted, however, these are not part of the output of the simulation but rather a fact of the total distance between the input waypoints.

	sailing time(hours)	distance(nautical mile)	fraction of distance due to weather
initial route	254±123	3835±1777	0.04±0.04
sailed route	254±121	3819±1744	0.04±0.04

3.3 Discussion

The results received when validating the weather and current factors are promising. Very high correlation is found between the baseline performance speeds and the performance speed based on the factual speed over ground along with weather and current factor. The Pearson correlation factor of 0.84 - 0.94 is impressive and we feel convinced that weather and current factors are good tools for measuring the lost knots of a route due to weather and currents.

There are a lot of question marks regarding the results of comparing advised and non-advised voyages. The results are questionable since it is not possible today to get a control group for which to compare weather routing against. This is because there is not a single vessel today that crosses an ocean without looking at some form of meteorological data and basing their route on the information received from said meteorological data. Thus the results can not be trusted for analysis of weather routing's benefits or shortcomings since both groups are in fact using weather routing in some way or another.

The attempted approach of this report was to use a large dataset, with thousands of voyages. However, since there is such large variance in the characteristics of the vessels, no good result can be derived from using all vessels in the same analysis. It is a fact that it is very hard to compare two vessel's weather factors fairly, because it is largely affected by a large number of parameters, such as time of year, the age of the vessel or time since hull cleaning. We hoped that the large dataset we had would allow for a selection of voyages with similar characteristics for comparison, but in the end that was not applicable due to the presence of too many parameters. Simply put it is hard to establish control voyages for which to compare against.

There is also a very large issue with the dataset regarding the initial route. It is not guaranteed to be the route provided by the vessel, it can very well be a route created by experienced meteorologists. This is due to the fact that not all vessels provide a route suggestion but leave the task entirely up to the weather routing service. Unfortunately there is no data available to distinguish if the initial route was suggested by the captain or a meteorologist.

An interesting observation is that 86% and 94% for advised and non-advised respectively has a timeliness within 24 hours of their last reported required ETA, see Figure 3.13. But when observing the timeliness of the initial required ETA, the timeliness results in 53% and 55% for the initial route and sailed route respectively, see Figure 3.18. This leads us to assume that it is common praxis to be late and update the required ETA en route.

Chapter 4

Conclusions

We can clearly see that the weather and current factor is heavily correlated to performance speed derived from the vessel supplied baselines. The one giving the best correlation is the RPM baseline with a PCC of 0.94. The power baseline showed least correlations with a PCC of 0.84. Very likely this is due to the complexity of the power measurement. It is reported as a percentage of the vessel's MCR(Maximum Continuous Rating) which is a value in kilo watt of what power output they can have without risking damage of the vessel. However the MCR can change without being reported properly, this can give a skewered mapping between main engine load percentage and power output.

The results, if looked at without consideration of the issues with the dataset, give an impression of using weather routing being worse than no weather routing. But considering the fact that a vessel chooses to *not* pay for the premium service of weather routing, if most of the vessel's routes are in areas where the weather is statistically less severe, we realize that we can *not* conclude that weather routing is worse than no weather routing. The possibility of easier voyages not using weather routing can be observed when looking at the mean, minimum and standard deviation of weather and current factors for non-advised routes and advised route, see Figure 3.7, Figure 3.9 and Figure 3.8. The main reason, that our opinion is that the results are not representable of the benefits of weather routing is because of the differences, between the datasets are extreme, and it is not likely to be caused by the sole factor of weather routing. Also, the weather experienced is very mild for non-advised routes.

There is also the issue of a non-advised route being converted to an advised route if the vessel's captain can see with the help of simple weather forecasts that there is tough weather ahead. Meaning that the results would not be representative because if there is a large chance of trouble en route, the captain asks for help and the voyage is converted to an advised voyage. This more or less eliminates the validity of a non-advised dataset as a control dataset for comparison against advised voyages.

We believe one of the largest reasons that weather routing services struggle to give guarantees in terms of fuel and time economy savings is because there simply is no control group for which to compare the sailed routes against. And vessels have a very large variation of set ups of machinery, hull and loading capabilities making a trip being made between the same ports and in the same direction with similar vessel characteristics very unlikely to occur. Thus we must turn to simulations which in themselves carry a certain amount of uncertainty. It is from simulations that previously mentioned 1.7% improvement to fuel and time economy have been estimated. In previous studies, finished/standard routes have been examined and analyzed, and the optimal route has been found, post voyage, leading to a potential improvement compared to the sailed route [22][19].

It would be very interesting to make a comparison between a simulation of a "standard route" and a simulation of the sailed route in large scale. Given the same departure time and post voyage analysis weather data. If there is an increase in performance, for the sailed route, in an empirical manner, one could make direct conclusions about the benefits of diverging from the "standard route". Because there is a large difference of weather characteristics over seasons and geographical location, we believe, weather routing as a whole would most likely net a very small percentage of improvement, but for certain troublesome routes where it is common with heavy weathers and currents, more could be gained from weather routing.

Further improvements to the value of this report's analysis, regarding CTA[1], could be made if a simulated versus reported arrival time was compared. This would close the gap between uncertainties of the CTA and real arrival time. However, this type of comparison would not change the results of the weather and current factors since they are already confirmed as heavily correlated to the factual loss due to weather and currents.

An interesting future project is analyzing a vessel's compliance to the supplied route. The shortest route between two geographical points on Earth is always the great circle route because it is the shortest path it is often the fastest, which is why many vessels choose to follow a great circle route. But because vessels can be operated manually, the continuous function of the bearing creating a great circle is not always followed precisely. Such an analysis of the vessel compliance could yield results regarding the effects of using a sophisticated auto pilot versus more crude steering where the vessel take a bearing and sail in that direction until, the vessel, maybe hours later change the bearing to a new angle. Of course one would also have to consider the fact that modern gyro compasses have an error of $\pm 2^\circ$ [17]. Moving along with the analysis: simulation of the initial route is compared to the final version taken by the vessel. One would assume because vessels often struggle to follow a great circle perfectly along a voyage [17], that the sailed route should generally be longer than the first route which consists of waypoints with perfect great circles. But the results do not show this clearly, see Figure 3.17.

It would be interesting to go further back in time, several years, and research the development of weather and current factor over time, to see if there are improvements since more advanced vessels and sophisticated weather routing methods have emerged.

Although this report was an attempt to find empirical evidence of time and fuel economy changes due to weather routing, such an analysis is close to impossible to conclude with the data set available for this report. The fact that vessels are not only interested in time and fuel economy but rather have dynamic interests that change over time makes conclusions regarding the limited parameters time and fuel close to impossible to make. The improvement in weather factor for one voyage can easily be offset by the loss of weather

---

[1] Calculated Time of Arrival.

factor for another voyage where time was more crucial. To make conclusions regarding weather routing more data on route level needs to be collected. Some sort of point system could be used, where you state how important time compared to fuel economy was when the route was created. This is most likely an over simplification since there are many other factors to consider, such as wave height or period, or the temperature of the ocean if you have temperature sensitive cargo. There are even cases when you can not have the wind coming in from a certain direction because the fumes from the cargo in fore must not reach the cargo in the aft. A system this complex is not very likely to ever be developed because it serves no other purpose than to evaluate weather routing, and introduces a lot of complexity to the routing service.

The final conclusion of this report is that it is irrelevant to prove fuel and time savings in retrospective view of a route, without heavy consideration to the demands and characteristics of each individual voyage. We conclude that there was no measurable gain to fuel and time economy as a whole for the initial route and the sailed route of the vessels. However, we accept the fact that it is fully possible to make improvements to voyages regarding a selection of different parameters en route. But since these parameters are dynamic it is hard to find empirical evidence in their favor without manual selection and filtering. During conversations with shipping companies they reported that they estimated that they could save a few tenths of a knot for an entire voyage, simply because they have better look ahead and trust regarding the weather models, allowing them to go a little slower at times because they do not feel the same need to have a large time marginal.

Bibliography

[1] Matplotlib basemap toolkit. `https://matplotlib.org/basemap/`. (Accessed on 06/09/2017).

[2] Matplotlib: Python plotting. `https://matplotlib.org/`. (Accessed on 06/09/2017).

[3] Numpy. `http://www.numpy.org/`. (Accessed on 06/09/2017).

[4] Python. `https://www.python.org/`. (Accessed on 06/09/2017).

[5] Python data analysis library — pandas. `http://pandas.pydata.org/`. (Accessed on 06/09/2017).

[6] Seaborn: statistical data visualization. `https://seaborn.pydata.org/`. (Accessed on 06/09/2017).

[7] Statsmodels: Statistics in python. `http://www.statsmodels.org/stable/index.html`. (Accessed on 06/09/2017).

[8] Nicolas Bialystocki and Dimitris Konovessis. On the estimation of ship's fuel consumption and speed curve: A statistical approach. *Journal of Ocean Engineering and Science*, 1:157 – 166, 2016.

[9] Joseph Bonney. Trans-atlantic ship schedule reliability hits three-year low. *JoC Online*, page 1, 2015.

[10] Henry Chen, Vincent Cardone, and Peter Lacey. Use of operation support information technology to increase ship safety and efficiency. *SNAME transactions*, 106:105–127, 1998. `http://www.sname.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=7d1c9fba-a588-4d0a-8232-c67a6c237acc`.

[11] European Comission. 70th session of the marine environment protection committee (mepc 70) at the international maritime organization (imo) - european commission.

https://ec.europa.eu/transport/media/media-corner/70th-session-marine-environment-protection-committee-mepc-70-international_en, 10 2016. (Accessed on 04/05/2017).

[12] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[13] ECMWF. European centre for medium-range weather forecasts. http://www.ecmwf.int/.

[14] ECMWF. Lead time of anomaly correlation reaching a threshold | ecmwf. http://www.ecmwf.int/en/forecasts/charts/medium/lead-time-anomaly-correlation-reaching-threshold?time=2017031100,0,2017031100&threshold=80%&area=Europe, 04 2017. (Accessed on 04/05/2017).

[15] IMO - International Maritime Organization. Imo's contribution to sustainable maritime development, 2016. http://www.imo.org/en/OurWork/TechnicalCooperation/Documents/Brochure/English.pdf.

[16] Haakon Lindstad, Bjørn E. Asbjørnslett, and Egil Jullumstrø. Assessment of profit, cost and emissions by varying speed as a function of sea conditions and freight market. *Transportation Research: Part D*, 19:5 – 12, 2013.

[17] Evgeny Lushnikov and Krzysztof Pleskacz. An analysis of practices of monitoring the accuracy and reliability of compasses on modern merchant ships. *Scientific Journals of The Maritime University of Szczecin, Zeszyty Naukowe Akademii Morskiej w Szczecinie*, 117(45):174 – 180, 2016.

[18] International Chamber of Shipping (ICS). Delivering CO_2 emission reductions. http://www.ics-shipping.org/docs/default-source/resources/environmental-protection/shipsandco2-cop21.pdf?sfvrsn=16, 2015.

[19] IMO International Maritime Organization. Prevention of air pollution from ships, mepc 58/inf.21, 2008. http://www.catf.us/resources/filings/vessels/MEPC_58-INF.21-FOEI.pdf.

[20] UNCTAD - United Nations Conference On Trade and Development. *Review of Maritime Transport 2016*. United Nations Publication, 2016. http://unctad.org/en/PublicationsLibrary/rmt2016_en.pdf.

[21] Laura Walther, Anisa Rizvanolli, Mareike Wendebourg, and Carlos Jahn. Original article: Modeling and optimization algorithms in ship weather routing. *International Journal of e-Navigation and Maritime Economy*, 4:31 – 45, 2016.

[22] Haifeng Wang and Galen Hon. Reducing greenhouse gas emissions from ships, 2011. http://www.theicct.org/reducing-ghg-emissions-ships.

Appendices

Appendix A

Dictionary

A.1 Abbreviations

AIS Automatic Identification System. Automatic GPS system onboard vessels.

CTA Calculated Time of Arrival.

DWT Dead Weight Tonnage.

ECMWF European Centre for Medium-Range Weather Forecasts.

ETA Estimated Time of Arrival.

GPS Global Positioning System.

PCC Pearson Correlation Coefficient.

MCR Maximum Continuous Rating.

NP-hard Non-deterministic Polynomial-time hard, see section A.2.

A.2 Glossary

Many of the definitions in this appendix are taken directly from Wikipedia and thus do not represent my own work. This glossary is intended for the user to have somewhere to look when a word has slipped from the mind.

advised voyage voyage that has had land-based meteorological advice en route.

ballast ballast is a cargo area which can be filled with sea water in order to weigh down the vessel. It is common to call the loading condition simply: ballast.

baseline function describing the speed relative; fuel, RPM or engine load.

current factor number representing the number of knots lost due to ocean currents. This value can be both negative and positive.

dead weight tonnage measurement unit for the loading capacity of a ship.

extreme weather unexpectable, unusual, unpredictable severe or unseasonal weather; weather at the extremes of the historical distribution.

heuristic algorithm that does not find the absolute optimal solution but deemed good enough for it's application.

initial route the first created route for a voyage in the dataset.

loaded loading condition of a vessel. Implies fully loaded cargo.

non-advised voyage voyage that has *not* been advised by land based meteorologists en route.

NP-hard non-deterministic polynomial-time hard. Simply put it means that the time complexity of an algorithm is at best 2^n where n is the number of inputs. For example, a route with 5 waypoints and 10 different speed settings would take $2^{15} = 3.2 * 10^4$ time units to solve. While a more realistic example is a route with 50 waypoints and 10 different speed settings, resulting in $2^{60} = 1.1 * 10^{18}$. Even if the first relatively easy problem took 1 nano second to solve(unlikely). The more realistic input would take 407 days to solve.

operator the vessel has a captain on board the ship that has final say on how to navigate and control the vessel, however, on main land they have operators that command the vessel to ensure it follows schedules and contracts. A captain is only captain of a single vessel at a time. An operator can be the operator of several vessels at the same time.

performance speed speed of a vessel if there was no weather or currents.

Pearson correlation coefficient a value representing the linear correlation between two variables. The value interval is between -1 and 1 where:

- **-1** complete negative linear correlation.
- **0** no linear correlation at all.
- **1** complete positive linear correlation.

required ETA reported, by the vessel, required estimated time of arrival.

route set of waypoints.

sailed route route sailed by a vessel if the waypoints are defined as the AIS data.

semi loaded loading condition of a vessel. Implies more load than ballast but less than loaded.

slamming impact of the bottom structure of a ship onto the sea surface. It is mainly observed while sailing in waves, when the bow raises from the water and subsequently impacts on it.

slow steaming operating a ship at lower speeds than it is designed for.

speed in calm condition reported speed setting of a vessel. In theory, this value should the exact same as the calculated performance speed.

speed over ground speed if measuring relative geographical coordinates.

standard route definition of standard route is vague, it has a different meaning in different context, loosely defined it is the route that is most common.

steaming when the engine is operating.

timeliness measurement of how late a vessel is according to the required ETA. The unit is hours.

vessel ship used for transportation across water.

voyage journey from one port to another.

voyage optimization similar to weather routing in the sense that it's main goal is to minimize the negative effects of weather, but instead of a constant route/speed, both parameters are variable.

weather factor number representing the number of knots lost due to weather. This value can only be negative.

weather routing concept of navigating open waters to minimize the negative effects of weather on a vessel.

Z-test statistical test to see how likely it is that two distributions are different. The result of a Z-test is, amongst other, a p-value that represents the likelihood of the two distributions being different from each other. A p-value of 1(maximum) implies there is no difference between the two distributions, while a p-value of 0(minimum) implies the distributions are not related at all.