Can We Trust the AIS Destination Port Information for Bulk Ships?– Implications for Shipping Policy and Practice

Abstract:

The Automatic Identification System (AIS) is a shipping database that includes the physical characteristics and real-time dynamics of ships. It has attracted great attention from academia recently and has been broadly applied in solving various problems in different fields. The voyage destination report is a piece of information recorded in AIS that indicates the heading port in a ship's voyage. This information is widely referred to by port operators for traffic estimation, and by shipping traders for supply forecasting, etc. However, we find that a considerable proportion (nearly 40%) of this information has been erroneously entered, both intentionally and unintentionally.

In this paper, we aim to propose targeted policies to correct the inaccurate reports based on assessing the probability of observing wrong destination port reports of ships in AIS. To this end, we first of all conduct extensive interviews with relevant shipping stakeholders to understand the reasons behind the wrong destination reports. Second, based on the interviews and relevant literature we propose the influence factors. Third, we generate a data sample set based on the voyages performed by Capesize and Panamax bulk ships around the globe in a year. To generate this sample set, we leverage data mining techniques to extract the information from an AIS database and other databases. Finally, a discrete choice model is built to achieve the proposed objective.

The results demonstrate that our model has an 84.1% accuracy rate in ascertaining the correctness of destination reports observed in AIS. We also find that, for a voyage, the speed of the ship, the historical accuracy rate of destination reports made by the ship, and the distance between the recognized origin and the reported destination of the voyage, have the most significant impacts on the accuracy of the destination report. Based on the findings, we provide managerial and policy suggestions to ship operators, port authorities, and regulators.

Keywords: Shipping management; Discrete choice model; AIS data; Data mining

1. Introduction

Maritime transportation plays a crucial role in international trade. There are more than 50,000 merchant ships trading internationally on the sea. As required by the International Maritime Organization (IMO), all these ships need to be equipped with an AIS transmitter (IALA, 2004). AIS records the static, dynamic, and voyage-related ship information, including ship identity, timestamp, real-time location, speed, rate of turn, navigation direction, navigation status, destination, estimated time of arrival (ETA), and draught. The initial objective of introducing AIS was to prevent ship collisions and enhance navigation safety. Since 2014, the quality and accessibility of AIS data have been significantly improved, mainly due to the usage of satellites that can transmit real-time AIS data globally. Consequently, its application has been greatly expanded from maritime safety to other fields such as trade and environmental research. Yang et al. (2019) conducted a comprehensive review of the studies regarding AIS applications. They ascertained the quick expansion of AIS applications and identified seven areas into which the AIS studies fall, these being AIS data mining, navigation safety, ship behaviour analyses, environmental analyses, trade analyses, Arctic shipping, and ship and port performance analyses. Among these reviewed studies, many different features in AIS data have been leveraged. For example, ship size, speed, and navigation direction are used in shipping safety analyses, and ship position, speed, and timestamp are used for trade forecasts.

When commencing a shipping voyage, every ship is required to enter its destination port and ETA in the AIS. However, this information hasn't been at all well referred to by the relevant stakeholders. This is because the information is manually inputted by the crew members, and we have found that a considerable proportion of this information is mistakenly entered, either intentionally or unintentionally. In fact, by comparing ships' filled destinations before their next voyages in AIS with the actual destinations observed from the ship trajectories after the completion of these voyages, nearly 40% of the destination reports entered into the AIS by Capesize and Panamax ships were found to be wrong.

In shipping studies, forecasting the destinations of ships is a very important and valuable task, one that can benefit various stakeholders by improving their operations and management. For example, Zhang et al. (2020) stated that accurate information regarding ships' destinations could help port operators make timely and efficient decisions. They proposed a method to predict destinations of voyages using ship trajectories. The improvement of destination report quality in AIS can be a simple and more efficient solution of destination report forecasting. Bulk carriers can also improve their operating efficiency by improving their repositioning strategy with accurate destination information of ships (Yang et al., 2019). They can particularly send ships in ballast to destination ports with relatively more cargoes and fewer competitive ships. Knowing ships' accurate destinations can also increase the bulk shipping market's transparency, thus raising the efficiency throughout the whole bulk shipping industry. The wrong destination reports have hindered the expansion of AIS data applications in these areas. Therefore, studies that can help assess the reliability of destination reports in AIS would be of great value for shipping practitioners.

In view of the hurdles caused by wrong destination reports in AIS, some authorities have begun formulating policies to regulate destination reports in the AIS reports made by ships.[1] For

[1] https://www.maritime-executive.com/article/panama-threatens-sanctions-for-ships-disabling-positioning-signals.

regulators, knowing the underlying factors behind wrong destination reports is valuable. In particular, based on these reasons and associated influence factors, they can set up more effective policies for avoiding wrong destination reports, and devise more efficient methods of detecting violations of relevant regulations.

This study aims to propose a targeted policy to correct the inaccurate reports based on assessing the probability of observing wrong destination port reports of ships in AIS. In this study, we focus on Capesize and Panamax ships in the bulk shipping market, which are two major types of bulk ships having a deadweight ton (dwt) ranging from 60,000 to 210,000. There are three features of Capesize and Panamax ships that drive our investigation on the behaviours of these ships. First, unlike container ships that provide scheduled services like buses do, bulk ships are more like taxis, whose destinations are random, and thus their behaviours merit greater interest for investigation. Second, Capesize and Panamax ships are two relatively large bulk ship types, with more regular trading routes, so it is easier for us to validate the underlying reasons for wrong destination port reports. Third, the dry bulk shipping market is a relatively transparent perfect competition market compared with other segments of the shipping market (Peng et al., 2016). Therefore, for a bulk ship operator, knowing other ships' movements generates competitive advantages (Prochazka et al., 2019a).

The contribution of this paper is threefold. First, this is one of the preliminary studies investigating manually input data quality in AIS, and it proposes new factors to help understand the accuracy of destination reports in AIS. These concepts and factors can be referred to by subsequent studies. Second, cross-disciplinary methods are adopted in this study. In particular, we propose various ways of analysing data in AIS and we also develop a discrete choice model to explain the inaccuracies of destination reports in AIS. The methods of data processing can also shed light on other AIS-related studies. Third, this paper also provides important practical insights. It can help various shipping stakeholders improve their operation and management in daily work.

This paper is divided into eight sections. The second section presents the literature review. We then explain the influence factors and propose the research hypotheses in Section 3. The model and data processing method are introduced in Sections 4 and 5, respectively. In Section 6, we illustrate the results. In Sections 7 and 8, we discuss the implications to practice and draw conclusions, respectively.

2. Literature review

This study is motivated by practical problems faced by bulk shipping operators, and we employ multiple methods to solve the proposed problem. It doesn't fit exactly into any existing study strands, but there are however some areas that it is relevant to.

The first related area is the detection of anomalies in AIS data. Before 2014, the quality of AIS data is poor which limited the studies in this field. Many papers have discussed the quality of AIS data in the early stage. Harati-Mokhtari et al. (2007) first discussed the reliability of AIS data. They indicated that the destination identification in AIS could improve navigation safety and suggested that ships maintain it accurately. In their study, they identified the inconsistencies of data, e.g., vague or incorrect AIS entries for destinations. Baldauf et al. (2011) found that the settings for AIS-data-based collision alarms, which are entered manually, are often found empirically wrong. Felski and Jaskolski (2013) discussed the completeness and integrity of AIS data in collision avoidance with data from 2006 to 2007, and 2010 to 2012. The findings suggest a high level of dynamic data integrity of AIS. With additional satellites launched in 2014, the AIS data quality is significantly improved. However, there is still debate on the correctness of static data in AIS as they are manually input. Lensu and Goerlandt (2019) stated that the AIS stream is vulnerable and may contain considerably erroneous data. Therefore, it should be crossly checked against other information apart from just the checksum. Tu et al. (2018) indicated that there are many data anomalies in AIS, thus various anomaly detection algorithms have been proposed to identify ship trajectories with anomalous characteristics. In the considered problem, we detect errors in destination reports in AIS data. However, this is essentially different from other anomaly detection problems in the literature. For instance, anomalies in AIS data, such as a ship found in a restricted and unexpected region, ship speed going significantly above or below the regular speed, or the unexpected visiting time of a ship, are mainly caused by system problems/errors (Tu et al., 2018). In contrast, wrong destination reports are manually entered by the crew members on a ship, and therefore result from human errors. The algorithms for anomaly detection, including the normalcy box method (Rhodes et al., 2009), the fuzzy ARTMAP (Bomberger et al., 2006), and the Holst Model (Laxhammar, 2008) are hardly applicable for our problem. Conversely, we need to understand the ship operator's decision-making process so as to recognize the factors which may lead to the decision to wrongly enter the destination ports. We find there is no prior research focusing on this kind of anomaly detection problem.

The second related area is the decision analyses of carriers/ship owners, namely, understanding factors that affect the shipping market dynamics and behaviours of carriers. In these studies, causality analysis has been broadly applied. One such case is the ship positioning problem, for example, where carriers need to decide the heading destinations of their ships in order to maximize the potentials of obtaining a cargo contract in the spot market. Bai and Lam (2019) modelled the energy shipping destination choice behaviours and identified their associations with various market factors from the charterers' perspective. They found that freight rate, propane price spread, bunker costs, and the number of ships in the destination areas affect the choice of charterers. Regli and Nomikos (2019) found that geographical routes and ship speed can be used to explain freight rate evolution. The other frequently used factors in determining market dynamics and the behaviours of carriers include the laycan period (Prochazka et al., 2019b), ship size, lead time and charter length (Köhn and Thanopoulou, 2011), fuel consumption (Abadie et al., 2017), perception of market psychology (Papapostolou et al., 2017), buyer and seller heterogeneity and their relationship (Adland et al., 2016), and energy efficiency design (Adland et al., 2017). Different from these studies, this paper aims to analyse the factors that lead to the mismatch between entered destination ports and actual destination ports in AIS. This is a new problem and we need to propose appropriate factors that can not only reflect the decision-making process of carriers in filling in the destination reports, but that are also available to access or derive.

The third related area is AIS application. The AIS is broadly believed to have kickstarted the era of digitization in the maritime industry. An expansion of AIS can be evidenced by rapid growth in publications and increasingly diverse topics covering its applications (Yang et al., 2019). Fournier et al. (2018) reviewed the publications related to the uses and applications of the AIS during 2004-2016. They summarize that over the years, AIS has become a widely used tool for developing applications such as marine environment, safety and security, and many more. Collision avoidance is the most fundamental application of AIS data. Wu et al. (2019) developed a fuzzy logic based approach for ship-bridge collision alert by utilizing AIS data. Recent studies in this area include Bakdi et al. (2020), Gao and Shi (2020), and Greig et al. (2020).

Another important application area of AIS data is trajectory analysis. Trajectory analysis includes trajectory extraction and trajectory prediction. Trajectory extraction indicates the construction of a ship's trajectory based on the reported spatiotemporal sequence data (Yang et al., 2019). Trajectory extraction has attracted increasing attention over recent years. Wang et al. (2017) proposed a methodology to leverage AIS data in order to discover spatiotemporal co-occurrence patterns of ships, which distinguish ship behaviours simultaneously in terms of space, time, and other dimensions. Zhang et al. (2017) developed a tangible analytical approach to analyse ship traffic demand and the spatial-temporal dynamics of ship traffic in port waters using AIS data. Li et al. (2018) proposed a novel approach that combines trajectory mapping and clustering for extracting trajectories from AIS data. The approach was shown to have higher accuracy than other commonly used methods. Alizadeh et al. (2021) proposed a method to predict a ship's trajectory in 10 to 40 mins based on its historical AIS data. Using ship trajectories derived from AIS data, Wu et al. (2020) proposed a method to identity ports visited by ships based on uncertain reasoning. This study belongs to ship trajectory analysis. We aim to compare the entered destinations with the actual destinations for voyages recorded in AIS. Whereas the entered destinations can be directly extracted from AIS, the actual destinations need to be derived from the historical trajectories of ships in AIS.

In addition to navigational safety and trajectory analyses, the applications of AIS data have been extended to many other areas. Lensu and Goerlandt (2019) presented an accumulating multi-purpose database for the northern Baltic Sea that combines AIS data with marine environmental data and emphasized the importance and necessity regarding the integration of AIS data with other databases. Jia et al. (2017) provided an automatic algorithm for deriving the micro information of ships, such as ship type and cargo volume, between any two recognized transportation nodes. They categorized relevant research in this regard into four categories: i. looking at trajectories (e.g. Willems et al., 2010); ii. looking inside trajectories (e.g. Zhang et al., 2017); iii. bird's-eye view of movement (e.g. Wood and Dykes, 2008; Jia et al., 2017); and iv. investigating the movement. Our study falls into category iv as indicated by Jia et al. (2017). In this study, we need to examine the various characteristics of ship movement and build their relationships with the occurrence probabilities of the wrong destination reports.

3. Influence factors and hypotheses

This is a pioneering study, as there exists no previous study discussing the influence factors on the correctness of destination reports in AIS. Therefore, we need assistances from industry practitioners. We conducted an open-ended interview to understand the reasons behind wrong destination reports. Unlike the structured questions, the open-ended interview with open/simple questions and close interactions with interviewers gives interviewees more freedom to comment on specific phenomena. In our interview, the five experts cover the most relevant industry professionals, including two captains, a ship operator, a senior manager, and a consultant, from major local bulk carriers and brokers. The interviewees were requested to provide us with information regarding two open questions:

1. The reasons and factors leading up to entering wrong destination reports into AIS;
2. Any comments or suggestions on the correction of wrong reports.

In December 2019, we presented the preliminary results of this research at a workshop on AIS applications in Tokyo, where we also received many helpful comments from shipping practitioners. Table 1 shows the details of our interviewees.

Table 1. Information of interviewees

Interviewee Code	Position	Working experience	Job description	Affiliation
A	Captain	15 years	Responsible for ship operation, safety, and crew management	China Merchants Energy Shipping
B	Captain	20 years	Responsible for ship operation, safety, and crew management	Ocean Longevity Shipping & Management
C	Ship operator	10 years	Monitor ship's daily movement	COSCO Shipping Bulk
D	Senior Manager	20 years	Oversee the fleet operation & movement	An anonymous Hong Kong based bulk shipping carrier
E	Consultant	10 years	Transmit information between owner and charterer	An anonymous international shipping broker

The information provided by the experts is consistent. The reasons for giving wrong destination information are complicated. Basically, we can summarize them into five reasons within the two categories of unintentional mistakes and intentional mistakes. Table 2 illustrates the reasons we collected from the interviewees.

Table 2. Reasons for giving wrong destination information

No.	Reason description	Category
1	Human errors: lack of training, negligence, fatigue, etc.	Unintentional mistakes
2	No specific destination when the voyage starts	
3	Security: to avoid pirate attacks, stowaway onboard	Intentional Mistakes
4	Commercial consideration: hide the destination	
5	Others: smuggle, etc.	

Among all, human errors serve as the most common reason mentioned by the interviewees. As destination information needs to be input manually into AIS, lack of training, negligence, and fatigue can all lead to mistakes being made by the ship's crew, especially junior seafarers. The second most common reason is speculative behaviour. For some voyages of bulk ships, the captains don't know the ships' destinations when the voyage starts. It is because the shipowner may not have a clear idea of where to place an empty ship. Also, in some cases, the shippers may purchase the cargo for speculation, such that they will look for contracts during a voyage from the loaded port to a demand region with many potential purchasers. Under this circumstance, they normally don't enter a destination, or they randomly fill one in the AIS, and then change it when the destination port is confirmed. Therefore, for a considerable part of the voyage, the destination report is wrong.

Reasons No. 3 to 5 belong to intentional mistakes. Sometimes, captains will not enter the destination, or will simply fill in a wrong one when the ship is in a certain sea area where the risk of pirate attacks exists, or if the ship is departing from a port where stowaways often board. We have also learned from two interviewees that some carriers prefer to hide the destinations of ships during the voyages, because they don't want their competitors to know where they will place the unloaded ships. Besides this, there are also other reasons, for example where smugglers or certain illegal ships will not input the right destinations into the system.

Based on relevant academic research literature, and our discussions with the interviewees, we summarize the factors explaining why ships choose to provide wrong destination reports in AIS data, and then make hypotheses regarding their impacts on the correctness of destination reports as follows:

Ship behaviour pattern: A ship's behaviour pattern indicates whether the historic trajectories of the ship display any regularities, that is, whether the ship sails along certain shipping routes regularly. Regli and Nomikos (2019) found that the geographical trading patterns of VLCCs (very large crude carriers) can be used to explain the market dynamics of bulk shipping. In this study, we believe that human errors (Reason No.1 in Table 2) in filling in destination reports can be avoided to some extent if a ship regularly visits certain ports, as the seafarers are familiar with the operation. In addition, the speculative behaviours (Reason No. 2) can also be less, as the ship rarely visits new ports. Therefore, we propose the first hypothesis (H1) as follows:

H1: *Ships with relatively regular travel patterns are assumed to have less intention of making wrong destination reports.*

Operator size: Large shipping companies (operators) generally have a relatively complete management mechanism and standardized operation procedures for operating ships, thus there is less possibility of lack of training (Reason No. 1), intentional fraud (Reason No. 4), and illegal activities (Reason No. 5). Therefore, we have the following hypothesis:

H2: *Ships under large ship operators have less intention of making wrong destination reports*.

Ship size: Ship size is broadly applied to study shipping investment and energy choices (Lindstad et al., 2015) as well as charter rate changes (Köhn and Thanopoulou, 2011). Because the high operating costs of large ships cause ship operators to be more prudent in their operation of large ships (Reason No. 2), in this study, we make the following hypothesis:

H3: *Larger ships have less intention of making wrong destination reports.*

Ship flag: The ship's flag denotes the country where the ship registers. Different countries have different regulations on the ships that register under them. In particular, flag of convenience (FOC) states have relatively lower requirements for registered ships. Alderton and Winchester (2002) found that ships under FOC states are more likely to have inferior records compared to ships under non-FOC states (Reasons No. 1 and 5). This leads to our fourth hypothesis:

H4: *FOC ships will have more mistakes in their destination reports.*

Historical accuracy rate of destination reports: Due to human habits, ships that have often previously reported wrong destinations are more likely to continue the same behaviour pattern (Reasons No. 1 to 5). Therefore, we have the following hypothesis:

H5: *A ship with a higher rate of providing wrong reports in its previous voyages tends to have a higher probability of generating a wrong report again*.

Operating time (utilization of ship): A ship that records a long operating time within a certain period (high utilization) implies that the ship is in good condition and that the ship's crew is more familiar with the operation of ships and also with the related regulations (Reason No.1). Therefore, we propose a hypothesis as follows:

H6: *A lower probability of giving wrong destination reports is assumed to be a natural by-product of a ship's longer operating time (higher utilization).*

Voyage distance: Long-distance voyages lead to more uncertainties (Regli and Nomikos, 2019) (Reason No.2). In addition, regulatory authorities can hardly monitor ships in the deep sea, the cost of making mistake is low (Reason No. 1), so the seventh hypothesis is as follows.

H7: *Long voyages are associated with an increase in wrong destination reports.*

Ship speed: Ship speed can indicate both market dynamics and carriers' behaviours. Regli and Nomikos (2019) empirically proved that the speed of ships sailing in ballast partly explains part of the freight rate evolution. In the tramp market, where ships often do not know the destination yet when leaving port, it is likely that ships will sail slower to save fuel until they know the destination port. In our study, if a ship travels fast in a voyage, then it tends to have an urgent time constraint and a clear trip destination (Reasons No. 2 and 4). Under this circumstance, the ship is less likely to provide wrong destination reports. Hence, we propose hypothesis H8:

H8: *A higher ship speed in a voyage is associated with a lower probability of making wrong destination reports.*

Loading status: This factor is related to the commercial consideration of ship operators. A loaded ship is more likely to have a predetermined destination before starting the voyage. In comparison, when a bulk ship leaves an unload port, in many cases the shipowner may not have a clear idea of where to place this ship. A ship's destination is sometimes determined after sailing on a common voyage to various possible destination ports, during which more information is collected. For example, right after a Panamax iron ore carrier unloads in Qingdao, China, it has to be repositioned to Australia or Brazil for re-loading. Because the voyages from Qingdao to Australia or Brazil share a common trip from Qingdao to the northwest corner of Luzon Island, Philippines, the shipowner can decide on the final destination after the ship completes this common trip (Reason No.2). Considering this, we propose the last hypothesis as follows:

H9: *Ships in ballast tend to have more wrong destination reports in their voyage records.*

In this paper, we have proposed nine factors related to voyage status and ship conditions that may affect the behaviour of ships when entering the destination ports into the AIS. To the best of our knowledge, some of them have not previously been considered in the literature. Such factors include operator size, the historical accuracy rate of destination reports, and operating time.

4. Model formulation

A multinomial logit model is constructed to control all these factors (explanatory variables) in order to explain the occurrence probability of a wrong destination entered in the AIS report. The underlying assumption is that a carrier will choose the option resulting in the highest utility. Multi-criteria decision-making methods (MCDMs) and discrete choice models (DCMs) (including multinomial logit models) are two widely applied tools to evaluate transport solutions from multiple objectives, while they are used in different situations (Le Pira et al., 2017). MCDMs provide a comparative assessment of alternatives evaluating their contribution to different evaluation criteria and stakeholders, which is more adopted in the procedure of consultation/participation. DCMs are more used in the procedure of stakeholder analysis. DCMs aim at analysing the behaviour of a decision-maker when choosing among different (discrete) alternatives, assuming that she maximizes her utility (Le Pira et al., 2017). The DCM is more suitable for our proposed research question, as we aim to understand the destination choice behaviour of carriers. In addition, the massive AIS data provides us a perfect chance to perform a revealed preferences analysis.

Suppose a ship is about to start a voyage which is denoted by i. The ship would obtain a certain level of utility, denoted by U_{ji}, through choosing option j, where $j = 1$ denotes the option that the ship reports a true destination, and $j = 0$ denotes the option that the ship reports a false destination. U_{ji} is defined as a function as follows:

$$U_{ji} = V_{ji} + \zeta_{ji}, j \in \{0, 1\} \tag{1}$$

where the V_{ji} is the deterministic utility and ζ_{ji} is an error term following the Type 1 Extreme Value Distribution.

The probability that the ship will choose option j in voyage i, denoted by P_{ji}, takes the following form:

$$P_{ji} = \frac{exp(V_{ji})}{exp(V_{0i}) + exp(V_{1i})} \tag{2}$$

With the influence factors we discussed before, the model for analysing the correctness of the destination report in a given voyage made by a given ship can be presented as:

$$\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 Apro + \beta_2 Bpro + \beta_3 os + \beta_3 dwt + \beta_4 flag + \beta_5 freport + \beta_6 lreport + \beta_7 wtpro + \beta_8 dist + \beta_9 speed + \beta_{10} load \tag{4}$$

where p denotes the probability of reporting a correct destination in AIS, β is a vector of parameters. *Apro, Bpro, os, dwt, flag, freport, lreport, wtpro, dist, speed,* and *load* are explanatory variables that may affect the ship's choice as to the correctness of the destination report. We explain each variable in detail as follows.

Apro and *Bpro* denote the ratios (between 0 and 1) of voyages of the ship to/from Australia and Brazil over all of its voyages in a certain period. These two variables represent the behaviour patterns of ships. Our analyses focus on the behaviours of Capesize and Panamax ships, and a large proportion of such ships are used for iron ore transportation (UNCTAD, 2019). Australia and Brazil are the two major exporters of iron ore, having enjoyed an approximately 80% market share of global iron ore exports. A considerable number of Capesize and Panamax ships regularly visit ports in these two countries. Hence, we use *Apro* and *Bpro* to measure the regularity of voyages conducted by the ship.

os indicates whether the ship belongs to a large ship operator. In this study, if the voyage is conducted by a ship operated by one of the top 50 ship operators (in terms of the size of controlled fleets) we code it as *o*s=1, otherwise, it will equal 0.

dwt represents the size of the ship. In this study, we use the deadweight tonnage of the ship to represent its size.

flag denotes whether the ship is a Flag of Convenience (FOC). If it is, we code *flag* as 1. Otherwise, it is 0.

freport and *lreport* represent the historical accuracies of a ship's destination port reports. For a voyage, *freport* corresponds to the historical accuracy rate (between 0 and 1) of the first destination report made by the ship for its former voyages and *lreport* denotes that of the historical accuracy rate (between 0 and 1) of the last destination reports of the ship in its former voyages, respectively.

wtpro is set as the ratio (between 0 and 1) of the total sailing time of the ship to the entire period under consideration. This reflects the utilization level of the ship.

dist. In this study, we don't use the actual navigation distance of a voyage to represent the *dist*, because the actual voyage distance can only be obtained after the voyage is completed, and thus the model can't be used for forecasting if the actual distance is used. Instead, we will represent the *dist* as being the navigation distance between the recognized origin and the reported destination of the voyage. This distance may deviate from the actual distance of the voyage, but it can reveal the intention of the ship operator.

speed denotes the average speed of the ship during a given voyage. We calculate it by dividing the distance of the voyage by the total time spent.

load is a binary variable, which equals 1 if the associated ship is fully-loaded (laden) and 0 if it is unloaded (in ballast) (a bulk ship is basically either fully-loaded or in ballast in practice).

5. Data processing

To validate our assumptions, we conducted an empirical analysis based on a set of records generated by leveraging the information from several databases, including an AIS database, ship fleet databases (from *Lloyd's list*[2] and *Clarksons*[3]), a database of ports (which was generated based on *Google Maps*[4]), and a port distance database (which was generated using *Netpas Distance*[5]). The data in the AIS database was collected from satellite-based AIS

[2] https://lloydslist.maritimeintelligence.informa.com/
[3] https://sin.clarksons.net/
[4] https://www.google.com/maps/
[5] https://www.netpas.net/

receivers. The database has a global coverage of AIS messages transmitted by ships. Our port database includes all ports visited in the voyages that were derived from the AIS data (see Section 5.2). A list of all ports in the database can be found from the link in this footnote[6].

The data mining process is illustrated in Figure 1. This process starts with the raw AIS data, and in the end we will obtain a sample set for applying to the proposed model. The data mining process includes six steps. In what follows, we explain the details of each processing step.

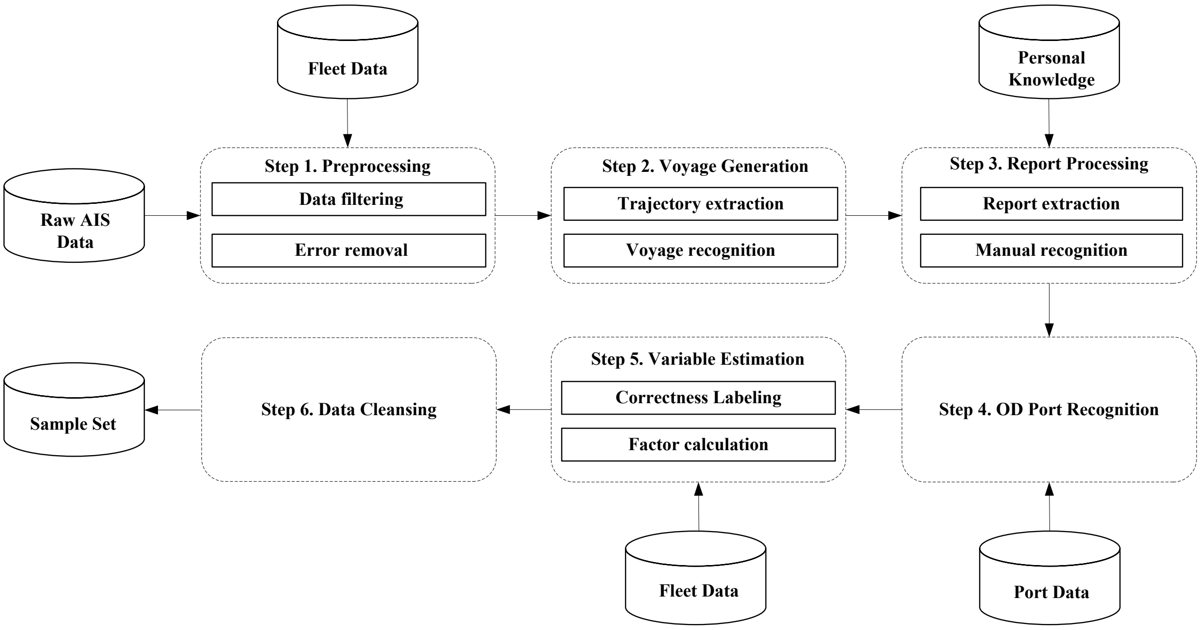

Figure 1. Data processing procedures

5.1. Preprocessing

In order to understand the behaviours of ships, we collected a large volume of raw AIS data (around 1T) retrieved from satellite-based AIS receivers. The data recorded the movements of all AIS-equipped ships in the period from 01/01/2017 to 31/12/2017. In the first step, we filter out AIS messages that are irrelevant to Capesize and Panamax ships as well as messages with incomplete information. This is achieved mainly by comparing the IMO numbers of all Capesize and Panamax ships extracted from the fleet databases with the IMO numbers in AIS data. AIS messages are classified as dynamic and static AIS messages. For dynamic AIS messages, we removed the ones that do not contain timestamps, MMSI numbers, speeds, or ship positions. For static AIS messages, we removed the ones that do not contain timestamps, MMSI numbers, destination reports, or draughts.

5.2. Voyage generation

In the second step, we derive voyages based on the data obtained in the first step. We first extract the trajectory of each ship. In our application, a ship's trajectory reports the position (in longitude and latitude), speed, and draught of the ship at each reporting time point. Here a reporting time point corresponds to the time when the AIS sends out a message. Note that the

[6] https://drive.google.com/file/d/1g_mKtTLPGxxFwJmvQf7VUYrocgXs8wVL/view?usp=sharing

average time interval between two AIS messages from the same ship in our database is 1,050 seconds. Then, based on the extracted trajectory of a ship, we construct its voyages by using the following rules:

- All ships are assumed to be sailing in voyages at the beginning of their trajectories.
- A ship starts mooring if the speed of an originally sailing ship is less than one knot in three consecutive messages sent by the ship; we record this time as the mooring start time.
- A ship ends its mooring if the speed of an originally mooring ship exceeds one knot in five consecutive messages sent by the ship; we record this time as the mooring end time.
- A mooring position is recorded for a ship if its speed is lower than 0.5 knots in all messages sent by the ship during any consecutive five hours between a mooring start and the corresponding mooring end time.
- The trajectory of a ship between any two consecutive mooring positions is recognized as a voyage; we refer to the first mooring position as the origin mooring position and the second one as the destination mooring position for this voyage.
- Voyages with abnormal characteristics identified are removed from the database. In particular, for each recognized voyage, we record its total sailing time as the difference between the time point when the ship ends mooring at the origin position and the time point when the ship starts mooring at the destination mooring position. We estimate the distance of a voyage by using the navigation distance between the two mooring positions. The average speed of this voyage is estimated by dividing the sailing distance by the sailing time. Then, voyages with sailing distances less than 1,600 nm or with average speeds less than 5 knots or higher than 18 knots are excluded from the database.

In this step, we have identified 26,131 valid voyages made by 3,291 different ships, and 4,244 invalid voyages were deleted due to the last rule.

Notably, we chose 1 knot as the threshold value for determining a ship's mooring behaviour because when sailing in the open sea, a ship's speed rarely goes lower than 1 knot. We do not use 0 knot as the threshold value because the speed reported in the AIS of a moored ship may still be larger than 0 because of the ship's movement caused by currents, but the recorded speed in the AIS of a moored ship rarely goes higher than 0.5 knots. As a matter of fact, one can also use other threshold values for detecting the mooring start time as long as they are significantly lower than the normal sailing speeds of ships. The 1 knot threshold is also used in other studies, for example, Jia et al. (2020).

In addition, we have learned that a ship may stop temporarily during its voyage. Our voyage construction procedure can identify and filter out short temporary stops that are no longer than five hours. By doing so, we can eliminate the impacts of most unexpected stops during their voyages. In the meantime, we admit that our method cannot identify long (longer than five hours) intermediate stops of ships during their voyages. One possible reason for long intermediate stops is to avoid sailing in the sea under extremely adverse weather. However, such temporary stops are very costly as they generate delays in cargo dispatch and damage the productivity of ships. Therefore, long temporary stops during a ship's voyage between two ports are very rare. Considering the large sample size (with 7,564 valid samples for training the model) and the very low frequencies of long intermediate stops, we believe that our analyses and the results should still be valid.

The following Figure 2 illustrates a density map for the 26,131 voyages. It can be observed that most of the voyages are between Australia, Brazil, and China.

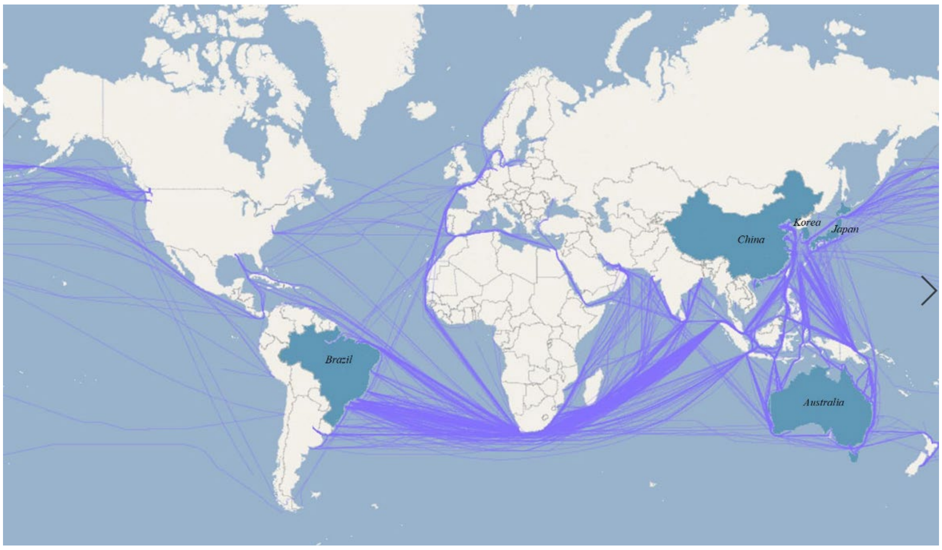

Figure 2. Density map for the ship trajectories

5.3. Report processing

A ship is required to report its destination port before starting a voyage. Ideally, the correctness of the destination report can be obtained by comparing this report with the destination port we derived for the voyage in the previous step. However, the reports extracted from AIS cannot be directly compared with the true destination ports. This is because there are no uniform standards for the destination reports made by ships. Some destination reports may refer to a country or a region but not a particular port. Some reports are actually meaningless. In addition, although most of the reports indeed indicate a specific port, there may be multiple entries that refer to the same port. For example, entries like "HK", "Hong Kong", "Port of HK", and "Port of Hong Kong" have all been found in destination reports in AIS, and they all refer to the Port of Hong Kong. Moreover, many reports contain typos. For example, we have found "HONGKONG", "HONGKONG CN", "HON KONG", "HONG K0NG" (the second "O" is "zero"), "HONG KOND", "HONG KONK", etc., in the reports. When facing these inputs, it is easy for a human to recognize that the correct underlying destination is "The Port of Hong Kong", while it can be difficult to train a computer to do so.

In view of the difficulties in processing destination reports, in the third step we manually examine each destination report extracted from AIS messages sent by ships in the voyages recognized in step 2. Note that a ship may change its destination reports multiple times during a voyage. In total, we have collected 13,638 different entries from the destination reports in AIS data. These entries are classified into three groups: meaningless reports (e.g., "0", "anchorage", "pilotage"), unspecific reports (e.g., "China", "Australia"), and port reports (where a specific port can be recognized). Among the 13,638 entries, there are 904 meaningless reports, 334 unspecific reports, and 12,390 port reports, respectively. For ease of comparison, we standardize the 12,390 port reports. Specifically, all reports such as "HK", "Hong Kong", "Port of HK", "Port of Hong Kong", "HONGKONG", and "HON KONG" are standardized as "The Port of Hong Kong". Notably, 1,234 different ports were recognized from the 12,390 reports after standardization.

5.4. OD port recognition

In the fourth step, we identify the actual origin and destination ports of each voyage. The identification is completed in three steps. In the first step, we construct a port database that contains the location information for all ports identified in the "report processing" step. This is achieved by manually searching in Google Maps. Then, in the second step, we first obtain the coordinates of the origin and destination mooring positions of each voyage and then recognize the origin and destination ports of the voyage by matching the location information from AIS with the location information in the port database. The origin or destination positions of some voyages cannot match any port in the port database. Therefore, in the third step, for such voyages, we search in Google Maps using the coordinates of their origin or destination mooring positions to identify the ports corresponding to these coordinates.

5.5. Variable estimation

In the fifth step, we derive the values of the independent and dependent variables (used in the multinomial logit model) associated with each voyage generated in the previous steps.

We start by labelling the correctness of each destination report made by a ship in a voyage. This is achieved by comparing the (standardized) destination reports made by the ship in the voyage (obtained in step 3) with its actual destination port (recognized in step 4).

Some variables describe the historical performances of a ship (e.g., *Apro*, *Bpro*, *freport*, *lreport*, and *wtpro*). To derive such variables, we divide the voyages into two parts: voyages that started before or on July 31, 2017, and voyages that started after or on August 1, 2017. Among the 26,131 voyages, 15,115 voyages and 11,016 voyages fall into the first and second parts, respectively. Voyages in the first part are used for evaluating the historical performances of ships. We derive the values of independent variables in different ways, as shown in Figure 3.

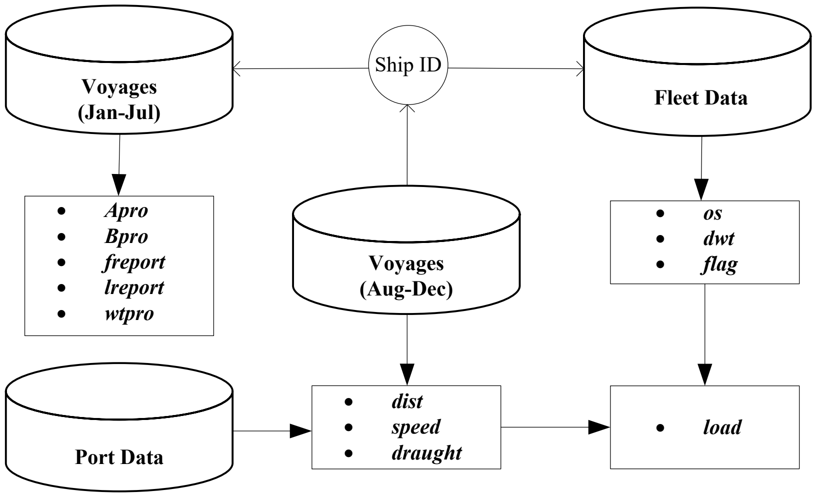

Figure 3. Variable estimation

To begin with, some variables are solely voyage-dependent and, given a voyage, they can be derived directly, based on the voyage information. These variables are *speed* and *draught*. In particular, *speed* is set equal to the average speed of the ship during the voyage. We also derive the *draught* of the ship sailing in this voyage from AIS. This variable will be further used to derive the *load* (loading status of the ship), as explained below.

In addition to *speed* and *draught*, we estimate the distance of the voyage (*dist*) by using the navigation distance between the recognized origin port and the first destination port reported in this voyage. Here the navigation distance between two ports is obtained from a port distance database. To generate the port distance database, we first derived 3,591 different OD pairs from the samples, and then obtained the distance of each OD pair by searching it from a software called "*Netpas Distance*", which is widely used by ship operators to calculate navigation distances.

Then, given a voyage, we derive *os* (operator size), *dwt* (deadweight ton), and *flag* (flag of the ship) from the fleet databases. The information from the fleet databases is matched with the voyage using the ship ID (i.e., IMO no. and MMSI no.: Maritime Mobile Service Identity Number) contained in AIS messages. In addition, one can determine the *load* (loading status of a ship) by comparing the draught of the ship in the voyage and the draught when it is in ballast (which can be estimated from the deadweight tonnage).

By utilizing the destination ports of voyages (which is recognized in step 4), we derive *Apro* (resp., *Bpro*) and *freport* (resp., *lreport*) for a ship from its voyages between January and July 2017. We calculate the value of *wtpro* as the ratio of the total voyage sailing time of the ship over the entire period from January to July 2017.

The dependent variable of a voyage is the correctness of the destination report in this voyage. The destination reports in AIS data may change more than once during a voyage. It is not difficult to understand that the accuracy rate of the destination report grows as the ship sails closer to the destination port and becomes more certain of its final destination. Furthermore, the destination port will monitor the correctness of a ship's destination report when it sails into its control area. We believe that it is more interesting and meaningful to investigate the correctness of the first report. Therefore, we choose the correctness of the **first destination report** to be the dependent variable in our study.

5.6. Data cleansing

We can now generate a set of records corresponding to each voyage made by a Capesize or Panamax ship during the period from August 1 to December 31, 2017. Finally, we need to exclude any abnormal records with incomplete or unrealistic information.

It is found that among the 11,016 voyages, some voyages were made by ships that did not sail in any voyage during the first seven months of 2017 or for which we do not have the information of its operators. As a result, we cannot obtain the operator sizes or the variables representing historical performances of these ships for these voyages, so the records associated with such voyages are deleted. Second, we also delete from our sample set the records of voyages that contain meaningless or unspecific reports in their first destination reports. A total of 1,560 voyages were deleted in this step. Finally, we obtain a sample set of 9,456 valid records from 2,683 ships.

6. Results

The 9,456 records generated from the data are further divided into two subsets, one subset with 80% of them (7,564 voyages) is used for estimating parameters of the model, and the other subset with 20% of them (1,892 voyages) is used for validating the model. We term the first set as the "training set" and the second set as the "validation set", respectively. In this section, we first report in Section 6.1 the descriptive statistics and the correlation matrix of the variables in the sample set. Then, we report the results of the proposed model in Section 6.2. Finally, the performances of the model are presented in Section 6.3.

6.1. Variables and their correlation test

In this study, we have proposed a total of eleven explanatory (independent) variables, some of which have never been considered in previous literature. The descriptive statistics for all independent variables of records in the training set are summarized in Table 3.

Table 3. Descriptive statistics of independent variables

Variable	Definition	Mean	Standard Deviation	Sample size
Apro	Ratio (between 0 and 1) of voyages of a ship to/from Australia in all of its voyages from January to July 2017.	0.389	0.376	7564
Bpro	Ratio (between 0 and 1) of voyages of a ship to/from Brazil in all of its voyages from January to July 2017.	0.189	0.263	7564
os	Whether a ship is owned by a top 50 company in terms of the number of ships operated; 1 for yes and 0 for no.	0.534[1]	0.499	7564
dwt	Deadweight tonnage of a ship (in tons).	138258	55043	7564
flag	Whether the flag of a ship is FOC; 1 denotes yes and 0 denotes no.	0.637[2]	0.481	7564
freport	Historical accuracy rate (between 0 and 1) of the first destination reports in the AIS report of a ship in its voyages from January to July 2017.	0.472	0.252	7564
lreport	Historical accuracy rate (between 0 and 1) of the last destination reports in the AIS report of a ship in its voyages from January to July 2017.	0.899	0.175	7564
wtpro	Ratio (between 0 and 1) of the sailing time of a ship over the period from January to July 2017.	0.533	0.010	7564
dist	Navigation distance between the recognized origin and the reported destination of the voyage	3336	1942	7564
speed	Average speed of the ship in a voyage (in knots).	11.3	1.022	7564
load	Whether the ship sailing in a voyage is loaded/unloaded; 1 denotes that the ship is loaded and 0 denotes it is empty.	0.503[3]	0.500	7564

Note[1]: The data indicates that 53.4% of ships were owned by the top 50 companies.
Note[2]: The data indicates that 63.7% of ships were under FOCs.
Note[3]: The data indicates that 50.3% of voyages were made by loaded ships.

We have also performed the Pearson correlation among these variables with EViews10.0, as shown in Table 4. In general, all the correlation values are within the range [-0.5, 0.5]. Therefore, multicollinearity would not significantly affect the regression results.

Table 4. Correlations of variables

Correlation	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) ***Apro***	1										
(2) ***Bpro***	-0.480	1									
(3) ***os***	0.069	-0.011	1								
(4) ***dwt***	0.176	0.102	0.227	1							
(5) ***flag***	-0.115	0.055	-0.042	-0.104	1						
(6) ***freport***	0.279	-0.135	0.104	0.265	-0.012	1					
(7) ***lreport***	0.108	-0.041	0.064	0.008	0.068	0.344	1				
(8) ***wtpro***	0.034	0.031	0.023	-0.007	-0.019	-0.076	-0.027	1			
(9) ***dist***	-0.098	0.167	0.008	0.136	0.009	0.026	0.050	0.032	1		
(10) ***speed***	0.034	0.010	0.013	0.115	-0.005	0.059	-0.010	0.027	0.288	1	
(11) ***load***	-0.103	0.058	-0.046	-0.077	0.026	-0.070	-0.003	-0.002	0.102	-0.261	1

Finally, as for the dependent variable, among the 7,564 records in the training set, 4,750 are associated with correct destination reports, and 2,814 are associated with incorrect reports.

6.2. Results of the model

We trained the model using the training set, where all variables are normalized into [0, 1] for the purpose of scaling. We performed the logit regression in EViews10.0 and the results are presented in Table 5. Note that the effect size of each independent variable is measured as a factor change in the odds ratio of the dependent variable for a standard error increase in the independent variable. This indicates the sensitivity of dependent variables to the changes in independent variables.

Table 5. Regression results

Variable	*β*	Standard error	*Prob.*	Effect size
Apro	0.347***	0.095	0.000	1.033
Bpro	-0.587***	0.134	0.000	1.082
os	0.160***	0.061	0.001	1.009
dwt	-0.577***	0.192	0.002	1.117
flag	0.004	0.062	0.950	1.000
freport	0.406***	0.135	0.002	1.056
lreport	0.920***	0.182	0.000	1.182
wtpro	1.020***	0.300	0.001	1.357
dist	6.390***	0.204	0.000	3.692
speed	4.130***	0.206	0.000	2.338
load	0.383***	0.060	0.000	1.023
constant	-4.052***	0.248	0.000	2.737

Note: ***, ** and * indicate significance at the 1%, 5% and 10%, respectively.

We can see from the table that 10 out of the 11 coefficients are significant at the 1% significance level. The only insignificant variable is *flag*, which indicates a ship's registration nationality. At first glance, it seems strange that ships registered under an FOC have the same probability of entering a wrong report as non-FOC ships. However, we have discussed this with industry practitioners and understand that, unlike the container shipping industry, both shippers and ship charterers are not concerned about the flag of Capesize and Panamax ships, as they do not have to frequently visit ports that only accommodate non-FOC ships.

As for the behaviour patterns of ships, we found that *Apro* (the regularity of a ship's visits to Australia ports) is positively related to the correctness of destination reports, whereas *Bpro* (the regularity of a ship's visits to Brazil ports) has a negative impact on the correctness of destination reports. This is somewhat inconsistent with H1 but can be explained by two facts. First, along the shipping route from Northeast Asia (where most iron ore is consumed) to Australia, Australia is the only destination for Capesize and Panamax ships. In contrast, along the shipping route to Brazil, other countries such as South Africa and India are also important iron ore exporters. For a ship in the spot market, in most cases she will not fix the destination when she sails toward Brazil, as she can call at any other exporting ports along the route when there is demand. Second, we understand from our interviewees that the giant iron ore shippers in Australia, e.g., BHP and FMG, request that ships fix their destinations in the AIS 10 days before they reach the port, and the trip to Australia is as short as 10 to 20 days, whereas Brazil iron ore shippers have no similar requirement.

The significant positive impact of operator size (*os*) suggests that ships of larger operators have higher probabilities of entering correct destination reports for the voyages. This confirms our hypothesis H2.

The result of ship size (*dwt*) is also counter-intuitive, as it suggests that larger ships are more likely to report wrong destinations, which contradicts our hypothesis H3. From the discussions with our interviewees, we understand that this is because large ships are more often used for long trips. Therefore, they have more opportunities to change the destination in the middle of the trip. Besides this, small ships can accommodate more types of cargoes and thus have fewer unloaded long trips. Loaded ships are more likely to have clear destinations.

The historical performance in destination reports of ships also plays a critical role in determining the correctness of destination reports. *freport* and *lreport* both have positive effects on the correctness of destination reports. These results confirm that a ship that has made wrong reports before has a higher probability of entering a fake destination in the AIS again (hypothesis H5). In particular, the historical correctness rates of the final destination report have a higher positive effect on the probability of giving correct reports than those of the first destination report. This can be explained by the fact that the last report is subject to monitoring by the port state control, and thus the ship operator's motive is more likely to be deliberate if entering a wrong destination in the last report. Therefore, if the last report of a ship is frequently wrong, then the ship must have very little concern about making wrong reports.

The ship utilization (*wtpro*) has a significant positive effect on the correctness of destination reports. This suggests that a higher utilization rate (longer sailing time) of a ship leads to a higher probability of correct destination reports, which confirms our hypothesis H6.

Surprisingly, the distance from the departure port to the entered destination of a ship has the highest positive effect at a 1% significance level. It is estimated that the odds ratio increases by a factor of 3.692 for every additional standard deviation of voyage distance. This suggests that when a ship enters a distant destination, normally she is sure where she will sail. Conversely, when she is not sure, she will enter a temporary closer destination along a certain shipping route, e.g., Singapore when a ship travels from Asia to America.

Among all the variables, *speed* has the second largest significant impact on the choice of filling wrong destination reports, with an effect size of 2.338. This shows that a ship with a lower speed has a higher probability of inputting a wrong destination into the AIS system. This confirms our hypothesis H8.

Finally, we can observe from the table that *load* has a significant positive effect at a 1% significance level. This suggests that loaded ships have a higher probability of filling correct reports, which is in line with our hypothesis H9. This also indicates that when the ship operator places their unloaded ship at an export port, wrong reports are more likely to appear.

6.2. Performances of the model

To evaluate the performance of our model, we first measure the McFadden ρ^2, which measures the overall model fit:

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(0)} \tag{3}$$

where $LL(\beta)$ is the log-likelihood at convergence with parameter vector, and $LL(0)$ is the initial log-likelihood with all parameters set to zero. This measure reflects the improvement of the log-likelihood of the estimated model compared with the intercept-only model. For our model, the Mcfadden ρ^2 value is 0.29. This indicates a fine model fit.

We also calculate the classification accuracy rate (in percentage) of our model, which is obtained by applying the model on the records in the validation set and comparing the estimated correctness of destination reports delivered by our model for the records with their actual correctness. The result is presented in Table 6. We can observe from this table that the overall accuracy rate of our model in prediction reaches 84.1%. In particular, it successfully detects 531 wrong reports out of 732 wrong reports (the detection rate of wrong reports is 72.5%). The model can also accurately recognize the majority of correct reports, and the accuracy rate for recognizing correct reports is 91.5%.

Table 6. Classification results

	Predicted		
Observed	**Correct**	**Total**	**Correct (%)**
RC=0	531	732	72.5
RC=1	1061	1160	91.4
	Overall		84.1

Note: RC represents the correctness of the destination report.

7. Discussions

In this paper, we use a discrete choice model to analyse the behaviours of carriers for making destination reports in AIS. It should be noticed that under the assumption of the discrete choice model, the carrier chooses the option of maximal utility when confronted with a discrete set of options. The unintentional mistakes (e.g., human errors) of carriers as choices are seemingly contradictory to the utility theory. However, though some carriers know that the wrong information caused by mistakes and human errors is against the IMO regulation, they still allow it to happen but do not manage to correct it as it may be costly to do so. This implies that they can obtain higher utility by making this choice (ignoring the mistakes intentionally).

The empirical results suggest that the navigation distance from the departure port to the proposed destination port in its first report has the most significant positive relation with the probability of providing a correct report for a voyage in the AIS, such that a longer proposed voyage indicates a higher probability of giving a wrong report. The average sailing speed is also a very significant factor affecting the correctness of a destination report in AIS, with an effect size of 2.338. When a ship sails faster, there is a much lower probability of observing a wrong destination report in its AIS. Longer operating time (higher utilization) of a ship also suggests a higher probability of reporting correct destinations. The historical accuracy rates of a ship's destination reports can also help to determine the correctness of destination reports. Ships that have records of previously making wrong destination reports are associated with a higher probability of entering wrong destination reports into the AIS again. We have found that ships with a larger deadweight tonnage are more prone to fill in incorrect destination reports in AIS. Although the effect sizes are relatively smaller compared to the above factors, operator size, historical behaviour patterns, and loading status are also proved to have significant positive relations with higher accuracies of destination reports in AIS. We also found that whether or not the ship has an FOC has no significant impact on the correctness of the destination reports it makes.

The findings of this study can provide many important implications for industry practitioners, and can help them to improve their operations and management. First, our model has an 84.1% overall accuracy rate in predicting the correctness of destination reports in AIS. This indicates that the model can well explain the impact of influence factors on the correctness of destination reports. It can also be applied in real applications to estimate the correctness of an observed destination report from AIS. The value of knowing the correctness of a destination report is tremendous. Note that when applying the model to predict the correctness of the destination reported by a ship for its on-going voyage, its average speed in the whole voyage is not available. In this case, one can estimate the variable *speed* by using the ship's average speed in the current voyage.

Second, port operators can refer to the information of destination port combined with ETAs to make timely and efficient decisions for maritime traffic management. In addition, knowing the destinations, ETAs, loading statuses, and DWTs of ships, carriers, and ship operators can better predict the number of ships sailing to certain regions so as to avoid an oversupply of ships.

Third, this study also enables the IMO (International Maritime Organization), PSC (Port State Control) authorities, and other shipping regulators to detect any misconduct of ships in terms of destination reports. As a matter of fact, some authorities have started formulating policies to sanction ships that deliberately report wrong destination ports.[7] In particular, when detecting erroneous destination reports, shipping regulators should pay more attention to ships with the following features: (i) ships in ballast, (ii) ships that have made wrong reports before, (iii) ships that have relatively lower utilization rates, (iv) ships entering a destination that is only a short distance from their departure port, and (v) ships that sail in voyages at relatively lower speeds.

Fourth, this study has identified the reasons behind the wrong destination reports in the AIS and the relative size of their effects. These findings can help the policymakers to improve the regulations by tailoring rules to prevent fake destination reports caused by some specific reasons. For example, policymakers could impose compulsory training to reduce human errors, provide guidance for ships that have no specific destination during a voyage, and set down punishment for deliberately hiding the destination.

8. Conclusions

In this study, we evaluate the impacts of different influence factors on the correctness of the destination reports in AIS by building up a discrete choice model. The variables in the model are obtained from extensive and deep interviews and investigation. AIS-based data mining is adopted to make it possible to quantitatively analyse the influence factors. This study addresses the previously untouched problem of manually input data correction, and adopts multiple approaches to solve the proposed problem, which can be referred to by subsequent research.

[7] https://www.maritime-executive.com/article/panama-threatens-sanctions-for-ships-disabling-positioning-signals.

In the long term, we believe that this study will reduce the odds of ships inputting wrong destination reports. The value of knowing the destinations of ships is of tremendous help to the bulk shipping industry. First, with more trustable destination reports from the AIS, the port authority/terminal operators can better improve their efficiency in managing the ship traffic and berthing resource allocation for incoming ships. Second, when the destination reports in AIS become more trustable for shipping practitioners, the efficiency of the bulk spot market can be improved through better matching of demand with supply, as the carriers can optimize their repositioning strategies when knowing the destination information of ships in the market. Large bulk ships have high operational costs and they are also the main sources of various pollutants. Therefore, a more efficient bulk shipping market contributes to further easing the burdens on shipping companies, as well as on society as a whole. Third, AIS data has been increasingly applied in solving various problems, not only in shipping but also in international trade and economic studies. The untrusted data of ship destination reports in the AIS can heavily affect the expansion of the applications of the AIS in practice. We firmly believe that studies based on AIS data will be further expanded if the destination reports in the system are more reliable.

Although our model has fine estimation accuracy and most of the identified factors are proved to be statistically related to the correctness of destination report in AIS, there exist some deficiencies which can be improved in the future. First, as this study is preliminary work in identifying the factors for wrong destination reports in AIS, there must be some other reasons and factors we missed, for example, the shipping market condition, political reasons, etc. In addition, the measurements of some factors could be reconsidered and improved. Second, some interesting findings are obtained from the empirical results, and we interpreted these findings, e.g., why speed has a positive impact on the correctness of destination reports, based on discussions with our interviewees and our own understanding. We expect that these interpretations can be verified with evidence in the future. Finally, it will be interesting to develop various methods to forecast ships' true destinations in the long term, for which knowing the correctness of the destination report from AIS is helpful.

References

Abadie, L.M., Goicoechea, N., Galarraga, I., 2017. Adapting the shipping sector to stricter emissions regulations: Fuel switching or installing a scrubber? Transportation Research Part D: Transport and Environment 57, 237–250. https://doi.org/10.1016/j.trd.2017.09.017.

Adland, R., Alger, H., Banyte, J., Jia, H., 2017. Does fuel efficiency pay? Empirical evidence from the drybulk timecharter market revisited. Transportation Research Part A: Policy and Practice 95, 1–12. https://doi.org/10.1016/j.tra.2016.11.007.

Adland, R., Cariou, P., Wolff, F.-C., 2016. The influence of charterers and owners on bulk shipping freight rates. Transportation Research Part E: Logistics and Transportation Review 86, 69–82. https://doi.org/10.1016/j.tre.2015.11.014.

Alderton, T., Winchester, N., 2002. Flag states and safety: 1997-1999. Maritime Policy & Management, 29, 151–162. https://doi.org/10.1080/03088830110090586.

Alizadeh, D., Alesheikh, A. A., Sharif, M., 2021. Vessel trajectory prediction using historical automatic identification system data. The Journal of Navigation, 74(1), 156-174. https://doi.org/10.1017/S0373463320000442.

Bai, X., Lam, J.S.L., 2019. A copula-GARCH approach for analyzing dynamic conditional dependency structure between liquefied petroleum gas freight rate, product price arbitrage and crude oil price. Energy Economics, 78, 412–427. https://doi.org/10.1016/j.eneco.2018.10.032.

Baldauf, M., Benedict, K., Fischer, S., Motz, F., Schröder-Hinrichs, J.U., 2011. Collision avoidance systems in air and maritime traffic. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 225(3), 333–343. https://doi.org/10.1177/1748006X11408973.

Bakdi, A., Glad, I. K., Vanem, E., Engelhardtsen, Ø., 2020. AIS-based multiple vessel collision and grounding risk identification based on adaptive safety domain. Journal of Marine Science and Engineering, 8(1), 5. https://doi.org/10.3390/jmse8010005.

Bomberger, N., Rhodes, B., Seibert, M., Waxman, A., 2006. Associative learning of vessel motion patterns for maritime situation awareness, in: 2006 9th International Conference on Information Fusion, IEEE, pp. 1–8. https://doi.org/10.1109/ICIF.2006.301661.

Felski, A., Jaskólski, K., 2013. The integrity of information received by means of AIS during anti-collision manoeuvring. TransNav: International Journal on Marine Navigation and Safety of Sea Transportation, 7(1), 95–100. https://doi.org/10.12716/1001.07.01.12.

Fournier, M., Hilliard, R.C., Rezaee, S., Pelot, R., 2018. Past, present, and future of the satellite-based automatic identification system: areas of applications (2004–2016). WMU Journal of Maritime Affairs, 17(3), 311–345. https://doi.org/10.1007/s13437-018- 0151-6.

Gao, M., Shi, G. Y., 2020. Ship collision avoidance anthropomorphic decision-making for structured learning based on AIS with Seq-CGAN. Ocean Engineering, 217, 107922. https://doi.org/10.1016/j.oceaneng.2020.107922.

Greig, N. C., Hines, E. M., Cope, S., Liu, X., 2020. Using Satellite AIS to Analyze Vessel Speeds Off the Coast of Washington State, US, as a Risk Analysis for Cetacean-Vessel Collisions. Frontiers in Marine Science, 7, 109. https://doi.org/10.3389/fmars.2020.00109.

Harati-Mokhtari, A., Wall, A., Brooks, P., Wang, J., 2007. Automatic Identification System (AIS): data reliability and human error implications. The Journal of Navigation, 60(3), 373–389. https://doi.org/10.1017/S0373463307004298.

IALA., 2004. IALA guideline no. 1028 on the Automatic Identification System (AIS) volume 1, part I operational issues edition 1.3. https://www.e-navigation.nl/sites/default/files/universal-automatic-identification-ais-volume-1-part-1-operational-issues-1028.pdf.

Jia, H., Lampe, O.D., Solteszova, V., Strandenes, S.P., 2017. An automatic algorithm for generating seaborne transport pattern maps based on AIS. Maritime Economic & Logistics, 19, 619–630. https://doi.org/10.1057/s41278-017-0075-7.

Jia, H., Lam, J.S.L., Tran, N.K., 2020. Spatial variation of travel time uncertainty in container shipping. Transportation Research Procedia, 48, 1740–1749. https://doi.org/10.1016/j.trpro.2020.08.210.

Köhn, S., Thanopoulou, H., 2011. A gam assessment of quality premia in the dry bulk time–charter market. Transportation Research Part E: Logistics and Transportation Review, 47, 709–721. https://doi.org/10.1016/j.tre.2011.01.003.

Laxhammar, R., 2008. Anomaly detection for sea surveillance, in: 2008 11th international conference on information fusion, IEEE, pp. 1–8.

Le Pira, M., Marcucci, E., Gatta, V., Ignaccolo, M., Inturri, G., Pluchino, A., 2017. Towards a decision-support procedure to foster stakeholder involvement and acceptability of urban freight transport policies. European Transport Research Review, 9(4), 1-14. https://doi.org/10.1007/s12544-017-0268-2.

Lensu, M., Goerlandt, F., 2019. Big maritime data for the Baltic Sea with a focus on the winter navigation system. Marine Policy, 104, 53–65. https://doi.org/10.1016/j.marpol.2019.02.038.

Li, H., Liu, J., Wu, K., Yang, Z., Liu, R. W., Xiong, N., 2018. Spatio-temporal vessel trajectory clustering based on data mapping and density. IEEE Access, 6, 58939-58954. https://doi.org/10.1109/ACCESS.2018.2866364.

Lindstad, H., Sandaas, I., Strømman, A.H., 2015. Assessment of cost as a function of abatement options in maritime emission control areas. Transportation Research Part D: Transport and Environment, 38, 41–48. https://doi.org/10.1016/j.trd.2015.04.018.

Papapostolou, N.C., Pouliasis, P.K., Kyriakou, I., 2017. Herd behavior in the drybulk market: an empirical analysis of the decision to invest in new and retire existing fleet capacity. Transportation Research Part E: Logistics and Transportation Review, 104, 36–51. https://doi.org/10.1016/j.tre.2017.05.007.

Peng, Z., Shan, W., Guan, F., Yu, B., 2016. Stable vessel-cargo matching in dry bulk shipping market with price game mechanism. Transportation Research Part E: Logistics and Transportation Review, 95, 76–94. https://doi.org/10.1016/j.tre.2016.08.007.

Prochazka, V., Adland, R., Wallace, S. W., 2019a. The value of foresight in the drybulk freight market. Transportation Research Part A: Policy and Practice, 129, 232–245. https://doi.org/10.1016/j.tra.2019.07.003.

Prochazka, V., Adland, R., Wolff, F.-C., 2019b. Contracting decisions in the crude oil transportation market: Evidence from fixtures matched with AIS data. Transportation Research Part A: Policy and Practice, 130, 37–53. https://doi.org/10.1016/j.tra.2019.09.009.

Rhodes, B. J., Garagic, D., Dankert, J. R., Stolzar, L. H., Zandipour, M., Seibert, M., Bomberger, N. A., 2009. Anomaly Detection & Behavior Prediction: Higher-Level Fusion Based on Computational Neuroscientific Principles, in: Sensor and Data Fusion. I-Tech Education and Publishing. https://doi.org/10.5772/6585.

Regli, F., Nomikos, N.K., 2019. The eye in the sky – Freight rate effects of tanker supply. Transportation Research Part E: Logistics and Transportation Review, 125, 402–424. https://doi.org/10.1016/j.tre.2019.03.015.

Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., Huang, G.-B., 2018. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey from Data to Methodology. IEEE

Transactions on Intelligent Transportation Systems, 19, 1559–1582. https://doi.org/10.1109/TITS.2017.2724551.

UNCTAD, 2019. Review of maritime transportation 2019. Paper presented at the United Nations Conference on Trade and Development (New York and Geneva). http://unctad.org/en/PublicationsLibrary/rmt2019_en.pdf.

Wang, J., Zhu, C., Zhou, Y., Zhang, W., 2017. Vessel Spatio-temporal Knowledge Discovery with AIS Trajectories Using Co-clustering. Journal of Navigation, 70, 1383–1400. https://doi.org/10.1017/S0373463317000406.

Willems, N., van Hage, W. R., de Vries, G., Janssens, J.H.M., Malaisé, V., 2010. An integrated approach for visual analysis of a multisource moving objects knowledge base. International Journal of Geographical Information Science, 24, 1543–1558. https://doi.org/10.1080/13658816.2010.515029.

Wood, J., Dykes, J., 2008. Spatially Ordered Treemaps. IEEE Transactions on Visualization and Computer Graphics, 14, 1348–1355. https://doi.org/10.1109/TVCG.2008.165.

Wu, B., Yip, T. L., Yan, X., Soares, C. G., 2019. Fuzzy logic based approach for ship-bridge collision alert system. Ocean Engineering, 187, 106152. https://doi.org/10.1016/j.oceaneng.2019.106152.

Wu, L., Xu, Y., Wang, F., 2020. Identifying Port Calls of Ships by Uncertain Reasoning with Trajectory Data. ISPRS International Journal of Geo-Information, 9(12), 756. https://doi.org/10.3390/ijgi9120756.

Yang, D., Wu, L., Wang, S., Jia, H., Li, K. X., 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. Transport Reviews, 39, 755–773. https://doi.org/10.1080/01441647.2019.1649315.

Zhang, C., Bin, J., Wang, W., Peng, X., Wang, R., Halldearn, R., Liu, Z., 2020. AIS data driven general vessel destination prediction: A random forest based approach. Transportation Research Part C: Emerging Technologies, 118, 102729. https://doi.org/10.1016/j.trc.2020.102729.

Zhang, L., Meng, Q., Fwa, T. F., 2017. Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. Transportation Research Part E: Logistics and Transportation Review, 129, 287–304. https://doi.org/10.1016/j.tre.2017.07.011.