Expert Systems With Applications 178 (2021) 114975

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

The promises and perils of Automatic Identification System data

Ties Emmens [a], Chintan Amrit [a,*], Asad Abdi [b], Mayukh Ghosh [a]

[a] Department of Operations Management, Amsterdam Business School, University of Amsterdam, Netherlands
[b] Department of Industrial Engineering and Business Information Systems, University of Twente, Netherlands

ARTICLE INFO

Keywords:
AIS data
Data mining
Navigation safety
Ship behavior analysis
Environmental evaluation
Advanced applications of AIS data

ABSTRACT

Automatic Identification System (AIS) is used to identify vessels in maritime navigation. Currently, it is used for various commercial purposes. However, the abundance and lack of quality of AIS data make it difficult to capitalize on its value. Therefore, an understanding of both the limitations of AIS data and the opportunities is important to maximize its value, but these have not been clearly stated in the existing literature. This study aims to help researchers and practitioners understand AIS data by identifying both the promises and perils of AIS data. We identify the different applications and limitations of AIS data in the literature and build upon them in a sequential mixed-design study. We first identify the promises and perils that exist in the literature. We then analyze AIS data from the port of Amsterdam quantitatively to detect noise and to find the perils researchers and practitioners could encounter. Our results incorporate quantitative findings with qualitative insights obtained from interviewing domain experts. This study extends the literature by considering multiple limitations of AIS data across different domains at the same time. Our results show that the amount of noise in AIS data depends on factors such as the equipment used, external factors, humans, dense traffic etc. The contribution that our paper makes is in combining and making a comprehensive list of both the promises and perils of AIS data. Consequently, this study helps researchers and practitioners to (i) identify the sources of noise, (ii) to reduce the noise in AIS data and (iii) use it for the benefits of their research or the optimization of their operations.

1. Introduction

The Automatic Identification System (AIS) is a system that makes the tracking of vessels possible, originally introduced as a tool for the identification of vessels in maritime navigation (Mazzarella et al., 2015). AIS data[1] is used for a variety of applications such as protection of the environment, the management of vessels in waterways, and overall surveillance to improve safety. However, although the development of AIS data seems promising, it has its limitations which should be considered (Šakan et al., 2018). The quality of data is essential to make productive use of AIS data. Both the huge volume that is collected and the lack of quality make it difficult to capitalize on the potential of using the data (Tsou, 2010). Therefore, the usage of AIS data involves both promises and perils which need to be understood to maximize the advantages and minimize the disadvantages of AIS data. Although the need for acknowledging the limitations of AIS data has been made explicit, little research has been done and these limitations have not been stated clearly (Šakan et al., 2018). Previous studies have mentioned problems related to AIS data, like "data redundancy" (Pallotta et al., 2013; Tsou, 2010), "noise" (Tsou, 2010), or the technical failure of equipment, but in almost all the studies, these issues are not considered in depth. A literature review has been conducted for the limitations of AIS data in the context of conservation (Robards et al., 2016), an overview of limitations of S-AIS has been presented (Carson-Jackson, 2012) and even several studies have been combined to identify AIS data-related issues (Harati-Mokhtari et al., 2007). However, either these studies focus on one particular area (e.g. conservation) (Robards et al., 2016), predominantly on the human factors or mainly on S-AIS (Carson-Jackson, 2012). Regardless of the specific purpose that AIS data is used for, the problem with AIS is that the data is getting more complex because of the increasing amount that is collected from many professional vessels. One of the key problems of AIS data that emerges from the literature is that it is hard to collect reliable information. For instance, factors like the range of AIS and the equipment can have an impact on AIS data. Therefore, AIS data can be noisy and inconsistent. Also, Harati-Mokhtari et al.,(2007) researched the impact of humans on the accuracy of AIS data, found that

* Corresponding author.
E-mail addresses: c.amrit@uva.nl (C. Amrit), s.abdiesfandani@utwente.nl (A. Abdi), m.ghosh@uva.nl (M. Ghosh).

[1] Sample AIS Data can be seen in Fig. 4 in Appendix A.

https://doi.org/10.1016/j.eswa.2021.114975
Received 15 October 2020; Received in revised form 16 March 2021; Accepted 27 March 2021
Available online 2 April 2021

AIS data is very unreliable.

When all the above-mentioned factors are taken into account, it is clear that there is a need to study both the opportunities and limitations of AIS data. While the process of handling the issues (perils) of AIS has been discussed by many researchers (focusing on the "how"), little is known about "why" and "where" the issues of AIS arise. The findings of this study help to better understand the use of AIS data for different domains. Eliciting the different promises of AIS data can help practitioners and researchers to either capitalize on the opportunities or can help them to identify new ones. More importantly, this study shows the drawbacks of AIS data that could help in avoiding pitfalls and make more effective use of the data by reducing noise. At last, giving a clear oversight of the promises and perils of AIS data provides academics with the opportunities and limitations of AIS data which could be further analyzed in future research. Therefore, the research question that this study aims to answer is: *What are the promises and perils of the Automatic Identification System data?*

In this paper, we first explain the concept of AIS data. Next, we review the literature to find the issues that researchers have run into when conducting their research. We then explain the method for our empirical study. Subsequently, we discuss the results of the quantitative analysis of the AIS data from the area of the port of Amsterdam, to validate and elaborate on the potential promises and perils, and the follow-up interviews conducted with both domain experts and practitioners in the shipping industry to validate the data. We present the promises and perils of AIS data by incorporating the results from the literature review and the mixed-design empirical study, that combines both exploratory quantitative and qualitative research. We present our research design in Section 3 and Appendix B.

2. Literature review

2.1. The concept of Automatic Identification System technology

AIS is a system that automatically identifies vessels, mainly used for improving navigational safety, ship-to-ship communication, and ship reporting. Although earlier regulations were in place, the use of AIS became mandatory for all vessels through the Safety of Life at Sea (SOLAS) agreement in 2004[2]. AIS data was processed through VHF signals between base stations at the coast and vessels previously. The problem with the first implementation was that AIS signals could not be received when vessels were too far away from the base stations at the coast, which is still one of the common limitations of AIS data (Carson-Jackson, 2012; Le Guyader et al., 2017; Šakan et al., 2018; Zhao et al., 2018) Hence, there was an interest to increase the range to track vessels further away from the coast. Nowadays, technology has been developed significantly. Therefore, AIS data can now be collected through receivers at base stations on land and by satellite-based receivers. However, the use of S-AIS brings other challenges, like message collisions (Alessandrini et al., 2019; Carson-Jackson, 2012; Høye et al., 2008). Several commercial satellite data services providers sell their data, claiming that their sources provide "maximum coverage" of AIS data[3]. The wide variety of applications of AIS data makes it worthwhile to study its potential. On the other hand, AIS data has many flaws, which should be considered before making use of the data.

2.2. The promises and perils of Automatic Identification System (AIS) data

Most studies either ignore that there are issues with the AIS data, or they state briefly that there are limitations related to AIS data and try to find ways to deal with it. For instance, Kang et al. (2019), developed methods for de-noising AIS data with respect to noise in positional and navigational information. There are only a few studies that indicated different in-depth opportunities or limitations of AIS data. For instance, Robards et al. (2016) published a literature review of the limitations of AIS data in the context of conservation. Carson-Jackson (2012) presented an overview of the limitations of S-AIS. Harati-Mokhtari et al. (2007) have conducted research that indicates AIS data-related issues. However, the current studies focus on one particular area (e.g. (Robards et al., 2016)), namely, on the human factors (Harati-Mokhtari et al., 2007) or mainly on S-AIS (Carson-Jackson, 2012). Our literature review shows that either the studies do not explicitly study AIS related limitations and opportunities, or they do, but only to a certain extent. Thus, we have collected several limitations and opportunities that have been mentioned in different studies. The following section lists the promises and perils that have been identified and classified based on all the limitations and opportunities related to AIS data found in the literature.

2.2.1. The promises of Automatic Identification System data.

Promise 1: *AIS data can be used to validate or give context to other data*

AIS data can be used to facilitate the validation of other data. For instance, AIS could serve as "ground truth" to validate or complete information of satellite images (Mazzarella et al., 2015), radars, more specific synthetic aperture radars (SAR) (Pelich et al., 2015), or tracking radars (Last et al., 2015; Papi et al., 2015) AIS data could also complement other data sources such as data sources related to geography (Ou & Zhu, 2008) or noise underwater (Hatch et al., 2008). Furthermore, AIS data could help to understand weather-related data. For instance, the impact of weather conditions like the wind, visibility, currents, waves, ice, hurricanes, and/or season (Kim & Lee, 2018) on vessel behavior. Consequently, this could greatly benefit the management of vessels and the effects on the environment.

Promise 2: *AIS data can help to enhance safety.*

AIS has originally been introduced for navigational safety and can help to prevent vessel collisions, which is quite similar to using GPS for the same purpose (Ngai et al., 2012). A great number of studies mention this original risk and safety-related purpose (Alessandrini et al., 2019; Goerlandt & Kujala, 2011; Greidanus et al., 2015; Harati-Mokhtari et al., 2007; Last et al., 2015; Ou & Zhu, 2008; Pallotta et al., 2013; Papi et al., 2015; Šakan et al., 2018). AIS data could also provide legal evidence in case of a collision (Ou & Zhu, 2008; Zhao et al., 2018). Moreover, AIS data might not only be handy for the prevention of collision of vessels (Ngai et al., 2012) but also the prevention of other types of collisions (e.g. *vessel into land*) (Silveira et al., 2013). Furthermore, AIS data can help to gain insight into the steering behavior of the captains of different vessels. Additionally, AIS data can also help in identifying piracy (Carson-Jackson, 2012). In some of these cases, AIS data can help to identify "anomalies" of tracks (Mazzarella et al., 2015, 2017; Pallotta et al., 2013; Papi et al., 2015) to find criminal activities.

Promise 3: *AIS data as a frame of reference to protect the environment.*

AIS data can function as a tool to regulate the environment. In general, AIS data can be used to evaluate adherence to environment-related regulation (van der Hoop et al., 2012). For instance, the violation of specific protected or forbidden areas can be monitored (Papi et al., 2015). The quality of air can be also protected by measuring, estimating, and predicting emissions of vessels using AIS data (Carson-Jackson, 2012; Jalkanen et al., 2014, 2012; Johansson et al., 2013, 2017; Kim & Lee, 2018; Yau et al., 2012). Furthermore, the spillage of oil can be traced. Moreover, AIS data can help to protect animals by finding areas where animal territory and vessel routes overlap (Hatch et al.,

[2] http://www.imo.org/en/OurWork/Safety/Navigation/Pages/AIS.aspx.

[3] https://www.orbcomm.com/eu/networks/satellite-ais (accessed January 25, 2019).

2008; van der Hoop et al., 2012).

Promise 4: *AIS data can be used to map routes and improve efficiency.*

One of the great benefits of AIS data is that it can be used as a strategic planning tool when it is combined with Geographic Information Systems, other databases, as well as data mining (Harati-Mokhtari et al., 2007; Zhao et al., 2018). In general, AIS data can be used to map routes of vessels, find route patterns, manage the traffic efficiently (Pallotta et al., 2013), recover incomplete tracks (Pallotta et al., 2013), make more accurate predictions (Greidanus et al., 2015), and estimate the fundamental diagram of ship traffic by mining ship speed-density relationships (Kang et al., 2018).

Promise 4b: *Predictive analytics and AIS data for the management of supply chains.*

More broadly, there is a need for more exploratory research in predictive analytics. Some researchers have investigated the use of AIS data in the context of logistics. For instance, Ben Ayed et al. (2015) investigated the use of big data analytics for logistics and found that big data analytics software can have many advantages and can be used to "enhance routing for planes, trains and trucks, etc.". Di Ciccio et al. (2016) predict flight diversions based on the automated detection of anomalous behavior. More specifically, several studies researched the prediction of travel time in logistics and developed a predictive model to estimate the arrival time of vessels.

2.2.2. The perils of Automatic Identification System

Peril 1: *AIS data contains noise in the presented information.*

AIS data contains noise in static, dynamic, and voyage related information. For example, speed over ground, timestamps, position, etc. are sometimes wrongly communicated. There may be duplicated data or the data could be missing. Harati-Mokhtari et al. (2007) have claimed that 80% of the data contains errors. However, it should be noted that the number of errors could differ between ship type (Greidanus et al., 2015). Furthermore, the process of cleaning AIS data is rather difficult (Zhao et al., 2018). For these reasons, AIS data might be unreliable, which can be a problem when it is used for data analysis and the mapping of routes.

Peril 2: *The quality of equipment for AIS is lacking.*

Generally, AIS can be distinguished into terrestrial AIS and S-AIS. Both have advantages and disadvantages. Terrestrial AIS has a limited range (Carson-Jackson, 2012; Le Guyader et al., 2017; Šakan et al., 2018; Zhao et al., 2018) but is more accurate than S-AIS (Greidanus et al., 2015). S-AIS has better coverage, but its key problem is the collision of messages (Alessandrini et al., 2019; Carson-Jackson, 2012). Another significant problem is the time between the communicated messages. Due to temporal resolution between messages (Carson-Jackson, 2012; Johansson et al., 2013), many messages might not be received (Greidanus et al., 2015). More technically, the quality of AIS data relies on the equipment for communicating these messages. Sometimes the different equipment, consisting of a transmitter, receiver, etc., does not function properly (Kazimierski & Stateczny, 2015; Ou & Zhu, 2008). Furthermore, different items can affect the quality, such as the medium/VHF transmission (Last et al., 2015), AIS receivers being unavailable (Pelich et al., 2015), faulty installation or placement of the equipment (Carson-Jackson, 2012), and the high frequency of transmissions making the human effort and receivers unable to cope with high traffic zones.

Peril 3: *AIS data contains too much information, more than is necessary for proper analysis.*

AIS data contains large amounts of data and parts of it are redundant. Several researchers have acknowledged this issue as "data redundancy" (Pallotta et al., 2013) or "noise" (Tsou, 2010), while others just mention the enormous quantity of data (Johansson et al., 2013). As a result, the problem that comes to light is that the data size can become too big to go through manually (Creswell, 1998).

Peril 4: *AIS data shows incomplete or unrealistic tracks.*

One of the key problems in using AIS data is that it can lead to incomplete or infrequent tracks (Pallotta et al., 2013; Šakan et al., 2018). This is often associated with a lack of signal due to limited receiver range or intermittent signals because of the revisit time and latency of satellites (Kim & Lee, 2018; Šakan et al., 2018).

Peril 5: *AIS does not function well in dense traffic areas.*

Another common problem of AIS data is that the tracking of vessels in dense and complex traffic areas is difficult (Carson-Jackson, 2012; Greidanus et al., 2015; Høye et al., 2008; Last et al., 2015; Pelich et al., 2015; Tsou, 2010) There's an interaction between density and vessel detection (Kazimierski & Stateczny, 2015) and the problem is related to equipment. For instance, many transmissions in the same area could lead to the collision of messages (Carson-Jackson, 2012), and the network could be inflated resulting in transmission failure (Greidanus et al., 2015; Last et al., 2015). As a result, dense traffic can lead to errors in modeling (Pelich et al., 2015).

Peril 6: *AIS data is vulnerable to external conditions.*

External conditions have an impact on AIS data. This can be due to weather conditions/atmospheric refraction (Last et al., 2015; Pelich et al., 2015; Tsou, 2010; van der Hoop et al., 2012) intercepting transmissions on land, or other vessels jamming the VHF (Last et al., 2015) decreasing the range of the signal. Interestingly, some researchers have argued that AIS can complement radars in such conditions (Cervera et al., 2011). Several studies (Zhao et al., 2013) mention S-AIS as a potential solution for overcoming the problem of weather conditions, implying that this is generally a problem of AIS data.

Peril 7: *Human input leads to (un)intentional inaccuracies in AIS data*

The quality of AIS data is dependent on the correct use of AIS by humans. Mistakes made by humans can lead to errors in data (Harati-Mokhtari et al., 2007; Johansson et al., 2013; Tsou, 2010; Zhao et al., 2018). There are several ways in which humans can negatively affect AIS data: faulty installation or handling of the equipment (Harati-Mokhtari et al., 2007; Johansson et al., 2013), the manually entered information being incorrect (Alessandrini et al., 2019; Harati-Mokhtari et al., 2007), and the incorrect processing of information (Harati-Mokhtari et al., 2007; Johansson et al., 2013). AIS can also be switched off leading to errors or making it possible to engage in illegal activities (Mazzarella et al., 2015; Zhao et al., 2013) Therefore, there's a need to validate AIS data to improve maritime safety.

Peril 7a: *Falsification and Attacks*

AIS system is used in the marine industry for vessel traffic monitoring. Due to its importance in collision detection, search and rescue operations, Balduzzi et al. (2014) conducted a security evaluation of AIS. To this aim, the authors identified several threats that affect both the current implementation and the protocol specification of AIS. In summary, the threats can be categorized as Spoofing which indicates crafting a valid non-existent ship. This process consists of assigning to the fictitious ship static information, such as the vessel name, MMSI, type of ship, cargo type, and dynamic information like the ship's status, position, speed, course and destination. Hijacking involves altering any information about existing AIS stations, e.g. about the cargo, speed, location. Availability disruption threats include three attacks that can be performed only in radiofrequency such as Slot Starvation: it includes impersonating the maritime authority to reserve the entire AIS transmission "address space", to prevent all stations within coverage in communicating; this includes ships and aids-to-navigation, as well as AIS gateways used in traffic monitoring. As a result, the attacker can disable AIS systems on a large scale. Frequency Hopping: the attacker impersonates the maritime authority to instruct one or more AIS transponders to change their frequencies of operation. Timing Attack: the malicious user instructs the AIS transponder to delay its transmission time. However, the authors designed and implemented a novel software based AIS transmitter (called AISTX) to tackle the aforementioned issues. Using AISTX, authors discovered and experimentally verified that identified threats affect all transponders deployed globally on vessels, vessels tracking and other maritime stations such as buoys, vessel traffic services. AISTX is designed and implemented as software-defined radio (SDR). An SDR consists of a software application, which implements the signal elaboration chain, and hardware peripheral, which converts binary data to radio-frequency signals for over-the-air transmission.

Iphar et al. (2015) present several data quality dimensions to assess truthfulness in an AIS message, to detect erroneous, false or spoofed messages. To this aim, this study considers various factors to evaluate data quality including accuracy, precision, reliability, currentness, completeness, consistency and integrity.

Peril 7b: *Coverage and deliberate switch off*

Salmon et al. (2016) introduced the concept of "Black Holes" where the aim is to identify regions respectively covered and non-covered (i.e., regions from where AIS positioning signals are either received or not received). In fact, there still exists many areas that signals emitted from ships cannot be received as no antennas cover these regions, this being typically an issue for AIS. Authors present a hybrid approach to detect Black Hole which includes three steps. The idea behind this hybrid approach has been first described in (Salmon et al., 2015). The computational steps include: Offline process for Black Hole detection: the process to identify the less populated cells is applied to historical data. The process is similar to the one used in (Hadjieleftheriou et al., 2003) for dense cells finding. Online process for identifying Black Holes: the objective is to categorize the different cells of the grid and identify the formally identified Black Holes cells, as well as the formally identified covered cells and the empty cells. This then gives a picture of the candidate Black Hole set concerning the online part similarly to the offline part. Using this step, the cells that recorded a few positions for a short time period are considered as not covered. Black Hole extraction from both offline and online data: it merges information from steps 1 and 2 to find not covered regions.

AIS switch-off refers to the fact that many vessels turn off their AIS transponder in order to hide their whereabouts when travelling in waters with frequent piracy attacks or potential illegal activity, thus deceiving either the authorities or other piracy vessels. In this paper, Kontopoulos, et al. (2020) developed a system that notifies the user for communication gaps in real-time and distinguishes AIS switch-off from network coverage issues. In other words, the algorithm uses the network coverage to infer whether communication gaps are due to AIS switch-off caused by the vessels themselves or due to the lack of AIS messages. The main components of the system are the consumer node, coordinator actors and worker actors. The proposed system extends the method proposed in (Kontopoulos et al., 2018). Consumer Node: is responsible for consuming messages from a Kafka topic, which contains decoded AIS messages, and distributing the messages to the coordinators. Coordinator: each coordinator receives messages from the consumer node and is responsible for creating new worker actors and routing the messages to the existing ones. Worker: each worker is responsible for one ship id. This node receives messages from a ship. When the current step detects an AIS switch-off, it checks that it is due to the network coverage or not.

3. Research design

3.1. Purpose of empirical study

The purpose of the empirical study is to arrive at a more in-depth understanding of the promises and perils of AIS data. Our focus is mainly on the perils of AIS data. Researchers have discussed the process of handling the limitations (perils) of AIS data focusing on the "how". But little is known about "why" and "where" the issues of AIS data arises. This study aims to understand AIS data and its perils more thoroughly by exploring AIS data, visualizing it quantitatively, and, thereafter, getting expert feedback on found issues captured through semi-structured interviews. Furthermore, we discuss the different applications of AIS data that can be used for various purposes in the '*results*' section.

3.2. Research method

We apply a mixed research design, combining both quantitative and qualitative methods, to understand the perils of AIS more thoroughly. A mixed method is useful when a topic under study cannot be captured by either quantitative or qualitative research alone (Venkatesh et al., 2013). For this study, we apply the mixed method for "confirmation", "completeness" and "expansion" of the perils of AIS data (Venkatesh et al., 2013). Fig. 5 in Appendix B provides a detailed description of our research design.

We conduct our research sequentially, where quantitative research precedes the qualitative research. Such a research design is recommended for topics where a solid framework of theories exists, but the context of the research is novel or previous findings have been fragmented and/or inconclusive (Venkatesh et al., 2013). In this case, the qualitative research conducted through interviews helps in understanding how and why the perils, identified in the quantitative study, occur in AIS data.

For the qualitative research, semi-structured interviews with domain experts handling AIS data for the Port of Amsterdam, and with practitioners of the professional shipping industry were conducted. The reason for this is that there was no quantitative source for the identification of noise and other perils in the AIS data. Harati-Mokhtari et al. (2007) were not able to assess the quality of positional AIS data and, therefore, recommend in their suggested future research that the position of vessels provided in the AIS data needs to be compared with another source to assess its quality. Hence, interviews have been used as a second source in this empirical study to gain more insights into AIS data in general, but also more related to the position of vessels.

The benefit of using qualitative research is understanding local causality and extracting insightful explanations (Miles & Huberman, 1994). It provides an extensive description, and it helps in understanding how and why something, in this case, a peril, occurs (Miles & Huberman, 1984). The adoption of interviews as means of validation by Amrit et al. (2012) and the means of interviews for validation as explained by Young et al. (2018), although in a different discipline, indicates that interviews can be used as a measure to validate and to provide feedback on the AIS data analyses. This is in line with Kaplan and Duchon (1988) who find that when "collecting different kinds of data by different methods from different sources", data can be used for crosschecking, and can provide a better context. Consequently, the adopted mixed method design improves the quality of this study, since the AIS data was analyzed extensively by the first author before conducting the interviews.

At first, the interviews were conducted and transcribed. At the same time, initial ideas, codes and patterns were identified (pre-coding) based on the gained knowledge during the literature review. The transcribed interviews were coded with the help of NVivo software, which enabled us to select sentences and label them with codes that captured the meaning of the sentence (descriptive coding), for instance, "time of the day" of Table 4 in Appendix C. Afterwards, the category names were grouped in a smaller number of subset categories, grouping "time of the day" under "human input", as the time of the day had an impact on human behaviour. Thereafter, the subsets were grouped again in themes that function as an overarching subject that captures and summarizes the smaller subsets and category names. Accordingly, "human input" in this example, was grouped under the impact of different factors that cause "noise", the overarching theme.

The right column of Table 4 (Appendix C) indicates how many times the codes, categories or themes were mentioned by the interviewees. It can be the case that the references column indicates 0. This means that the overarching theme that we identified, was not mentioned specifically or was mentioned in other words for example. The complete coding process is described in Fig. 6 (Appendix C).

3.3. Quantitative research

Data collection & sample — the data source for this study is AIS data from AIShub[4] collected by the CWI[5] national research institute. Sample AIS data can be seen in Fig. 4 of Appendix A. Using AIS data from the area of the Port of Amsterdam instead of using available AIS data from other areas in the world provides us with the opportunity to make use of domain knowledge by interviewing AIS experts handling AIS data for this specific area. Also, the Port of Amsterdam is a high traffic zone with a lot of data from many different vessels, which is important for gaining insights into the impact of dense traffic areas, the differences between vessel types, and for the consistent availability of AIS data for a specific area over time in general. It is assumed that most of the issues with AIS data found in this dataset are more widely applicable to AIS data in general since AIS standards/equipment are the same worldwide. Since the equipment is the same, it is reasonable to assume that they are influenced by similar conditions. However, it should be noted that the intensity of the different factors could differ between areas. Another reason for the aforementioned assumption is that multiple studies have used AIS data from areas in the Netherlands, which were running into the same perils as mentioned in our literature review (Dobrkovic et al., 2015, 2018; Hatch et al., 2008; Shu et al., 2017; Zhou & van der Veen, 2014)

Our final data sample is the historical (thus non-updated) AIS data from April 1st until April 29th, 2018. Our total dataset consists of 21,178,375 rows and 19 columns representing the transmissions of all the vessels in the area of the Port of Amsterdam. The data was then split into two datasets. One dataset contained vessels that are assumed to be lying still/mooring, i.e. speed over ground (SOG) = 0. The other dataset consisting of vessels that are assumed to be moving, defined as 'sailing vessels' with a transmitted SOG $\geq$ 0.1. The splitting of the data has been done to improve the later analyses since moored vessels and sailing vessels required different approaches for the analyses. The data was unequally distributed between the two datasets, the moored vessels contained 16,667,245 rows, while the sailing vessels contained 4,511,130 data rows.

Variables — the AIS data include different variables such as Vessel Name, Latitude, ETA, etc. (Harati-Mokhtari et al., 2007) as defined by International Telecommunication Union (2014). It should be noted that the variables '*length*' and '*beam*' were missing in the dataset, which could be part of AIS data according to the International Telecommunication Union (2014). Table 1 lists the variables of the AIS dataset.

Table 1
Variables of AIS Data.

Variable	Description
Static Data	
IMO (International Maritime Organization)	The unique identification number of a vessel ("IMO" + 7 digits)
MMSI (Maritime Mobile Service Identity)	An alternative for IMO, more often used as an identification number (9 digits) of the vessel
Vessel Name	Name of the vessel, consisting of a maximum of twenty characters
Callsign	The alphabetic number for vessel assigned by country of registry
Vessel Type	The type of vessel in categorical numbers
Navigational Status	Status of the vessel (e.g. 'at anchor')
A	Position of GPS related to bow in meters
B	Position of GPS related to stern in meters
C	Position of GPS related to starboard in meters
D	Position of GPS related to port side in meters
Dynamic Data	
Timestamp	Unix timestamp, counting the second from January 1st, 1970.
Latitude	Geographical position/coordinate of a vessel ranging from −90° to 90°
Longitude	Geographical position/coordinate of a vessel ranging from −180° to 180°
SOG (Speed over Ground)	Speed of vessel relative to the ground (knots)
COG (Course over Ground)	Course over ground in 1/10 = (0–3599). Position relative to North (0.1°)
Heading	Direction 0°–359°
Voyage Related Data	
Draught	Draft of the vessel ranging from −0.1 to 25.5 m
ETA	Estimated Time of Arrival of the vessel in months, hours, minutes
Destination	The destination of the vessel

Construct data — we transformed the existing attributes and created new attributes programmatically by using Python. For both moored vessels and sailing vessels, the time stamp needed to be transformed from Unix timestamp, to the universal time ('%Y-%m-%d %H:%M:%S'). Also, the vessel types needed to be re-coded from numeric values to specific vessel types. We created the *distance* attribute to reflect the distance between two consecutive data points as transmitted per vessel. We used the Haversine formula to calculate the distance between two positions based on latitude and longitude, both available variables in the dataset. The Haversine formula is an equation important in navigation, giving great-circle distances between two points on a sphere from their longitudes and latitudes (Chopde & Nichat, 2013; Arifin et al., 2016). Haversine distance takes into account the curvature of the earth when calculating the distance between two points and hence is widely used in navigation including in Google maps (Arifin et al., 2016; Basyir et al., 2018) and therefore is common practice when calculating distances using AIS data, given that AIS data contains latitude and longitude information. Moreover, vessels follow the curvature of the Earth and the Haversine formula takes the curvature of the Earth into account.

We created the *distance* attribute to reflect the distance between two consecutive data points as transmitted per vessel. The Haversine distance can be calculated using the following formula (Chopde & Nichat, 2013):

$$d = 2r\sin^{-1}\left(\sqrt{\sin^2\left(\frac{\varnothing_2 - \varnothing_1}{2}\right) + \cos(\varnothing_2)\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right)}\right) \quad (1)$$

[4] https://www.aishub.net.
[5] https://www.cwi.nl.

Table 2
Demographic information participants of the interview.

Participant	Gender	Role	Duration
A	M	Captain	55 min
B	M	Pilot	32 min
C	M	Pilot	47 min
D	M	Traffic engineer/advisor inland shipping	1hr
E	M	Engineering consultant/data scientist	1hr
F	M	Project Manager innovation	44 min

where d = distance, r = radius of the world (=6371 in KM), $)$ = latitude and φ = longitude

We subtracted the timestamp of the selected data point by the timestamp of the previous data point of the same vessel to generate the attribute *time diff*. We generated the attribute *Distance time Diff* for moored vessels by dividing the calculated distance by the time in seconds.

We added two more attributes for sailing vessels. The attribute *SOGcalc* (Speed over Ground calculated) was generated by multiplying the *Distance time Diff* (speed in m/s) attribute with 1.943846, which converted the calculated speed to the speed in knots which is the same unit of measurement as the speed in the AIS data. Finally, the attribute *SOGper* (Speed over Ground percentage/ratio) was generated by dividing the calculated speed (*SOGcalc*) by the given speed (SOG) in the AIS data.

3.3.1. Additional data sources

Weather, Height of buildings, Electronical wires — we accessed other data sources to better understand the AIS data and its context. We downloaded the weather conditions from the official Dutch website (KNMI.nl (2019)). We then downloaded multiple data sources providing geodata in the area of Amsterdam and combined it with our dataset. We downloaded the height of buildings in the area of Amsterdam from two different data sources[6], along with the geodata of electrical wires.

3.4. Qualitative research

Approach — we interviewed experts to collect qualitative information to validate and expand the findings from literature along with the quantitative findings. We conducted individual semi-structured interviews. This type of interview combines both specific and open questions and it provides the opportunity to ask follow-up questions.

Sample and selection — we used the expert sampling method, a subset of purposeful sampling to select the participants for the interview. In total, we interviewed six participants whose demographic information can be seen in Table 2.

Interview and coding process —we recorded as well as transcribed the interviews. Thereafter, we systematically coded the interviews, following the coding process proposed by Miles and Huberman (1994), consisting of descriptive coding and pattern coding. Furthermore, we wrote significant remarks that had great importance for the study during the interview, called "pre-coding" and noted initial codes, ideas, and patterns that came to mind during the collection of data (Saldaña, 2015). We then used NVivo 12 software for coding the data where we used descriptive codes to describe the meaning of the statements of the experts. We then identified the relations between the different statements (pattern coding) to arrive at higher-level concepts (Miles & Huberman, 1994). The results of this study were then described using the final codes, categories, and themes in an organized manner.

[6] https://zakelijk.kadaster.nl/-/top10nl.

4. Results

In this section, we build upon the promises and perils identified in the literature and we present the results of the quantitative analysis of the AIS data from the Port of Amsterdam and combine them with qualitative explanations and feedback. We first present the different promises of AIS, identified through interviews.

4.1. The promises of Automatic Identification System data

Promise 1: *AIS data can be used to validate or give context to other data*

At the most fundamental level, AIS is a source of data that can provide meaning to other data sources. AIS data provides additional information for navigation to the radar which is mainly related to safety (Promise 2). Respondents A, B, and C found that AIS data combined with radar provides more reliable and detailed information. Furthermore, it is not always clear which vessel is displayed on the radar, and sometimes it is difficult to see smaller objects on the radar. AIS data can not only be combined with other non-safety related data sources, to validate or be validated, but also with safety-related data (Promise 2, Promise 3, and Promise 4). However, the type of data source that AIS data can and should be compared with depends on the purpose.

Promise 2: *AIS data can help to enhance safety*

AIS data is a very useful tool for navigational safety. Respondent B explained that AIS data helps to identify other vessels and to communicate with them: *"If there are problems, you immediately know with whom you are dealing. Before, you had to mention the position of the vessel concerning a particular object"* (Respondent B). Respondents A and C substantiated this claim, saying that AIS is not used for collision avoidance, but as an identification system that could result in collision avoidance. Respondents A, B, and C explained that AIS data needs to be corroborated with other sensors, like radar data, time, and a visual sight to be useful.

Promise 3: *AIS data as a frame of reference to protect the environment*

AIS data can be useful for understanding factors that have an impact on the environment. Respondent F remarked that AIS can be used for modelling vehicle emissions and for making better use of the environment. Respondent A explained that fishermen can use AIS data to find back their fishing spots and the places they left their fishing nets. However, Respondent F noted that fishermen sometimes switch off their AIS equipment for competitive reasons: *"They don't want the other vessels to know where the fish is, or crab, or whatever they are fishing for"* (Respondent F).

Promise 4: *AIS data can be used to map routes and improve efficiency*

AIS data can be used for monitoring traffic in ports. According to Respondents A, B, and C, it is a very useful tool for reconnaissance and planning. The port master can track the AIS signals and notice when something is wrong. Respondent F explained that the traffic intensity, the impact of traffic measures, the occupancy of berths, the period of stay of the vessels, the type of vessels, and their berth location in the port can be analysed. AIS data can also help in analysing the behaviour of other vessels. Another example of the potential of AIS data that can benefit both the captains of vessels and the port traffic management is the automatic collection of port taxes, Respondent F explained: "*You enter, you check-in, you check-out, and then the port gets paid for the time the vessel spent in the port, instead of manually declaring the information*" (Respondent F).

Promise 4b: *Predictive analytics and AIS data can be used for supply chain management*

AIS data can be used to predict the behaviour of vessels and to adjust their logistics accordingly. Respondent F explained that they are interested in the different paths and sequences a vessel travels through. Respondent C explained that a tug tow company can identify with the help of AIS data where a certain vessel is approximately located and estimate when it would arrive at a certain point, which helps in the economic planning for the company. Respondent F noted that the port can optimize their facilities to match the demands of the vessels and combine that with the occupation of berths.

In conclusion, the results present support for the proposition that AIS data can be used for the optimization of supply chains. Moreover, concrete examples of applications of AIS data are given, which show the wide range of opportunities of AIS data for supply chain management.

4.2. *The perils of* Automatic Identification System *data*

Peril 1: *AIS data contains noise in the presented information*

The most important and obvious problem is that AIS data contains noise in the presented information. Noise is defined as meaningless data related to uncertainty, precision, and corruption of the data. Although the perils are interrelated, the causes are discussed per peril to show in a structured manner what the problems of AIS data are, how much noise there is in the AIS data, and to what extent the noise can be explained. The noise can be in positional data for moored vessels or the positional data and Speed over Ground (SOG) of sailing vessels. For moored vessels, the noise in positional data is relatively small. For sailing vessels, the speed as transmitted in the signal often does not match the speed as calculated. This means that there is either noise in the positional data or noise in the speed as transmitted in the signal. Also, the results show that more noise outside the 10-second interval exists. For values close to zero, the "noise" implies that a vessel does not transmit its position while a difference in the distance has been found, resulting in incomplete tracks (Peril 4). Finally, there are differences between vessel types in the amount of noise, of which passenger vessels transmit the least reliable AIS data.

Peril 2: *The quality of equipment of AIS can be lacking*

The noise in AIS data can partially be explained through AIS equipment failures. Speed over Ground and Course over Ground as transmitted in the AIS data should be carefully analysed. This was confirmed by Respondent E: "*When a vessel is sailing, I would rather trust the values of the GPS than the Speed over Ground as communicated by the AIS signal.*". "*Sensor data, like the Course over Ground and the Speed over Ground, were for many vessels unstable.*" (Respondent E). Respondent B said the following regarding AIS data when the vessel sails at slow speeds: "*If you sail with a speed of fewer than 3 knots, the system does not stay up-to-date sometimes*" (Respondent B). Furthermore, the AIS signal often does not correctly show the direction of the bow of a vessel when mooring (Respondent C). The noise could depend on the type of AIS responder (Respondent E). Furthermore, it is found that there are discrepancies between the positional data of the AIS signal and the radar. Consequently, positional data and SOG can be incorrect (Peril 1).

Peril 3: *AIS data contains too much information, more than is necessary for proper analysis*

A common problem of AIS data for researchers is that it contains more information than is necessary for analysis. The size of the dataset for the conducted analysis in this study consists of 16,667,245 (moored vessels) + 4,511,130 (sailing vessels) data points for the area of Amsterdam for one month only. How much AIS data is needed for a proper analysis, is dependent on the purpose. This specific study aims to find how much noise there is, where it is, and how it is caused. Therefore, it was necessary to make use of the total AIS dataset.

Added Peril 3b: *the quality of AIS data is dependent on the data source*

One of the biggest challenges for researchers is to get access to reliable AIS data. For specific projects, reliable AIS data often needs to be purchased. The price of the AIS data depends on the number of positions, the size of the area, and the time frame. Data points can appear to be redundant during analysis. What is more, the quality of AIS data can differ between data sources: "*We have compared four different data suppliers and compared the number of missing values in the column as provided by the data supplier. We saw that the same MMSI-number, that normally should be the same vessel, can have different vessel types or sizes.*" (Respondent E). Also, it can be the case that AIS data is aggregated when requested. For instance, Respondent E found that when AIS data was requested for a specific zone in minutes, the timestamp of the transmissions was rounded to the closest minute.

Peril 4: *AIS data shows incomplete or unrealistic tracks*

Sailing vessels often show deviations from the line that is expected. Whenever AIS data is mapped, it becomes clear that many of the resulting tracks are illogical, go overland, or both. One of the reasons for a track going overland can be that the interval between two data points as communicated by the AIS signal is too large, potentially leading to a direct line over land when connected. Incomplete tracks are also noticed by professionals in their work environment and are referred to as "jumping". Because of the positional deviations and jumping of AIS data, it should never be used for collision avoidance according to Respondent C: "*If there's an error in your GPS or else, and you plot that data, then you plot something that doesn't exist*" (Respondent C). The same problem of incomplete tracks applies to practitioners who use AIS data for other purposes. Respondent E noted for one of his projects that: "*For the Port of Amsterdam, we have seen many vessels that showed a gap in its visits*" (Respondent E). Meaning that a vessel comes in with an AIS signal but leaves without being noticed. Consequentially, measures should be taken to deal with the noise. Respondent D and E explained that filters and step-by-step plans can be used to improve the quality of AIS data. An analysis of our dataset shows that the sailing vessels can have very long intervals. For instance, 2.1% of the transmissions of sailing vessels had a greater interval than 180 s while transmitting a SOG > 0, and where the distance between the consecutive data points was substantial. The reason for this could be that the captain of the vessel forgot to switch on their AIS signal (Peril 6) or that the equipment (Peril 2) was not transmitting frequently enough.

Peril 5: *AIS does not function well in dense traffic areas*

The transmission of many AIS signals in the same area could lead to errors. When asked about a specific area in which noise was found, Respondent F explained that many vessels are normally positioned in that area (and that high buildings are present, see Peril 6). However, he remarked that there could be many other explanations. The behaviour of vessels on King's day[7] was analysed to find if traffic density had an impact on the noise. On King's day, many vessels come together on the canals of Amsterdam. It is assumed that there are more transmissions for this area on that day, hence a higher density of vessels. When the data points are filtered on the canal area, the results indicate that relative to the number of total transmissions per day the noise of moored and sailing vessels together is the highest on the 27th of April (Fig. 1).

[7] 27th of April, the day when the King's birthday is celebrated in The Netherlands.

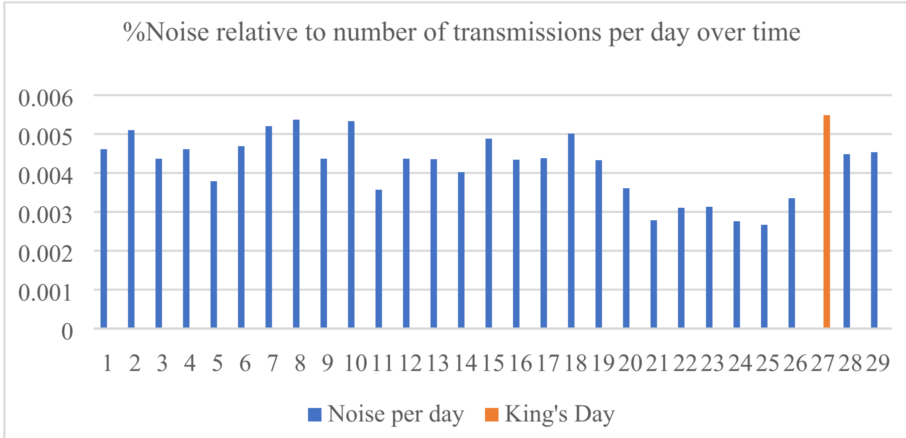

Fig. 1. Percentage of noise for the Amsterdam canal area per day.

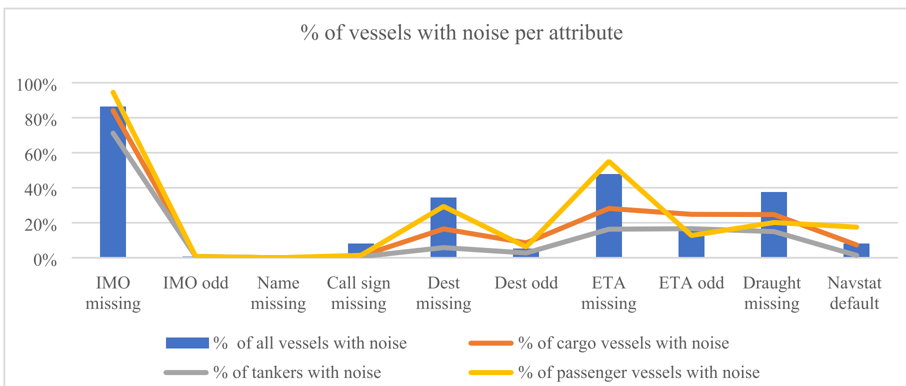

Fig. 2. Total % of vessels with noise vs. % of noise in three different vessel types related to their subgroup.

Peril 6: *AIS data is vulnerable to external conditions*

AIS data is transmitted via VHF, which can be influenced by multiple external factors according to the literature. The factors can be storms, lightning, and other weather conditions such as rain or fog. Furthermore, high buildings could have an impact on VHF and AIS data (Respondents A and D). Metal objects could also have an impact by causing reflection (respondent F). For instance, certain areas in the Port of Amsterdam have many metal storage tanks for fossil fuel.

Peril 7: *Human input leads to (un)intentional inaccuracies in AIS data*

Static and voyage related AIS data are manually entered by humans. However, as shown in Fig. 2 , missing values can constitute 86% of the static data and up to 48% of voyage related data, which is in line with the literature that describes the incompleteness of manually entered AIS data.

The analysis below shows what the human impact on AIS data is:

Noise: missing values and odd values: Since the entered data is static for the IMO, Name, and Call Sign data, it can be assumed that one vessel transmits the same AIS data (and noise) multiple times. Therefore, the number of unique vessels transmitting either missing values or incorrect values was analyzed to identify the number of unique vessels that transmitted noise. The static and voyage related data were analyzed for the sailing vessels only since Draught, ETA, and Destination values might not yet have been entered for moored vessels if it is going to sail at all. So, missing and incorrect values could have been overstated otherwise. Fig. 2 summarizes the most important findings of the noise found, as will be discussed in more detail.

IMO, Names & Call sign: For sailing vessels, more than 86% of the vessels had missing IMO numbers. However, not all vessel types are subject to SOLAS regulation. Therefore, the focus is on the vessel types that are subject to the regulation, i.e. cargo vessels (to some extent), passenger vessels, and tankers. For cargo vessels, 84.1% had a missing IMO number, whereas 94.5% of the passenger vessels and 71.2% of the tankers did not transmit an IMO number. The vessel names were almost in all cases present for every vessel group, except for a few pleasure crafts and unidentified vessel types, which is 0.11% of the vessels in total. Finally, the call sign of 8.12% of the vessels was missing on average, but that was mainly caused by the unidentified vessel group. For cargo vessels, passenger vessels and tankers only 1% of the vessels per group had missing call signs.

ETA, Destination, Draught, Navigational status: The destination for roughly 35% of the sailing vessels was missing. We find that passenger vessels contain the most, and tankers the least noise in their static and voyage related data. 53% of the total transmissions had missing values for ETA. Also, the percentage of transmissions that reported null values for 'draught' was relatively high (30%). More importantly, the results show that a lot of static and voyage related AIS data contains relatively high percentages of noise. Noise in AIS data, more specifically for cargo vessels, passenger vessels, and tankers, show that the guidelines are not being complied with.

Causes and consequences: AIS data entered by humans has improved over the last two years (Respondent F). However, as pointed out, the entered AIS data still has a lot of noise. Especially, for (most) cargo vessels, passenger vessels, and tankers this is noteworthy since these vessel types under the regulation of IMO [19]. For cargo vessels, it is often part of the procedure: *"It is the job for the duty officer. Before departure or arrival, he needs to prepare the bridge for the departure"* (Respondent A). Apart from the commercial purposes of the vessel, there are several other problems related to the human entered AIS data.

The results show that humans have a great impact on the quality of AIS data. For instance, the analysis showed that the voyage related data "ETA" for 47.8% of the vessels were missing and that more than 34% of the vessels transmitted no destination. Respondent A noted, when asked about potential causes for missing values or incorrect input, that *"It is a typical human mistake. For example, if the pilot is very busy"* (Respondent A). Consequentially, Respondent B said: *"I don't look at the destinations. I can see that they are wrong"* (Respondent B). On the other hand, Respondent A did not encounter many problems with AIS. However, he explained at the same time that he is not always able to validate the AIS data of other vessels: *"For example, a vessel is 300 m and it [AIS] says 250 m. How can I tell the difference? AIS data are initial signals, an installation if someone makes a mistake. What can we do? It is only possible to check by the captain of this vessel"* (Respondent A). Moreover, all respondents mentioned the problem of switching off the AIS device. That might be intentional: *"I have heard that they blur their signals sometimes."* (Respondent B); unintentional: *"It might be that you have to hurry up. Sometimes your transponder is off, but nobody knows."* (Respondent A); or part of natural human behaviour, for moored vessels: *"During night-time, we see that vessels disappear."* (Respondent E). Respondent D explained: *"We see a clear decrease around 10 PM …it is not logical that the vessels left the port. It's more reasonable than it's bedtime and that they switch off their transponders"* (Respondent D). Also, it might be the case that the equipment is incorrectly installed or maintained. The IMO, MMSI, Call sign, and vessel name are entered when the equipment is installed, which means that the information was incorrectly entered at the moment of installation. In agreement with the literature, Respondent C explained that the positional AIS data can differ between vessels, dependent on where the AIS transponder is installed on the vessel. Respondent A found that it could be the case that the AIS transponder was not switched on after maintenance. Furthermore, Respondent F explained that due to privacy reasons some static AIS information of smaller vessels cannot be used since it contains personal information. The reason for this is that the law prohibits the use of AIS data for commercial purposes that could lead to the identification of persons, without the approval of the captain/owner[8]. To conclude, noise in static and voyage related data, non-transmitting signals, and unavailability of AIS data can be a difficult problem when doing analyses. For instance, Respondent F noted that non-signalling vessels complicate berth occupation or traffic density analysis.

On the other hand, there is the potential of manipulating AIS data on purpose. Multiple respondents mentioned the possibility of humans consciously influencing AIS data. For instance, Respondent C found that AIS data is unreliable since humans determine what data is put into the system: "*The data you can see is the data that people want you to see*" (Respondent C). However, Respondent D noted that a vessel benefits from AIS data since they can use it for their safety. Especially smaller vessels can benefit since they are difficult to spot on the radar, respondent A explained. Therefore, it is remarkable that some (smaller) vessels do not transmit an AIS signal.

Altogether, we found evidence that human activities lead to (un) intentional noise in AIS data, due to errors in the installation and maintenance of equipment, forgetting to enter information in the system, circumstances beyond one's control, manipulation, and the difficulty to validate AIS data of other vessels. The list of Promises and Perils can, therefore, be summarized as:

5. Discussion

5.1. Discussion results

We adopted a mixed-design approach in which two different data sources were used to identify the promises and perils of AIS data. These sources are AIS data from the CWI research institute and data gathered from interviews with domain experts in the industry. The latter was used as a qualitative source of validation for the identified perils in the AIS data and to identify the promises of AIS data. Moreover, the results of both data sources showed similarities and complemented each other. Therefore, the results validated each other. Also, this study may apply to a wider range of areas than the Port of Amsterdam, since multiple researchers who have used AIS data from areas in The Netherlands faced the same perils as in other areas and The Netherlands, being a part of the international IMO regulation, has the same AIS standards/equipment as other countries.

In this study, four fruitful ways of implementing AIS data were identified. Although most AIS data is far from complete, it can still be used for many different purposes most of which are related to either safety of the environment or supply chains. AIS data is a very important tool for navigational safety, which has been discussed by many researchers. But the qualitative results give more insights into the way AIS can be analyzed for navigation. We find that it only provides a basic level of safety, i.e., identification and communication. Also, it appears that AIS data can help the navigator make better estimations for movements in the port. However, captains should never rely solely on AIS data for their decisions of vessel movements. We found little evidence on the use of AIS data for the environment. This might be explained by the fact that the respondents were not conservation researchers and have not used AIS data for this purpose. The results did show that AIS data can be used for emission models, which was also described or implemented by Carson-Jackson (2012), Jalkanen et al. (2014), Jalkanen et al. (2012), Johansson et al. (2013), Johansson et al.,(2013), Kim and Lee(2018) and Yau et al.,(2012). Although it has been pointed out before that AIS data can be used for the regulation and dispersal of fishing (Le Guyader et al., 2017), the have seen that fishermen can use AIS data exclusively. We also found that AIS data can be used for the optimization of supply

Table 3
The promises and perils of Automatic Identification System data.

Promises	Perils
Promise 1: AIS data can be used to validate or give context to other data	**Peril 1 (general):** AIS data contains noise in the presented information
Promise 2: AIS data can help to enhance safety	**Peril 2:** The quality of equipment of AIS can be lacking
Promise 3: AIS data as a frame of reference to protect the environment	**Peril 3:** AIS data contains too much information, more than is necessary for proper analysis
Promise 4: AIS data can be used to map routes and improve efficiency	**Added Peril 3b:** The quality of AIS data is dependent on the data source
Promise 4b: Predictive analytics and AIS data can be used for supply chain management	**Peril 4:** AIS data shows incomplete or unrealistic tracks
	Peril 5: AIS does not function well in dense traffic areas
	Peril 6: AIS data is vulnerable to external conditions
	Peril 7: Human input leads to (un) intentional inaccuracies in AIS data
	Peril 7a: Falsification and Attacks
	Peril 7b: Coverage and deliberate switch off

[8] http://europa.eu (accessed June 21, 2019).

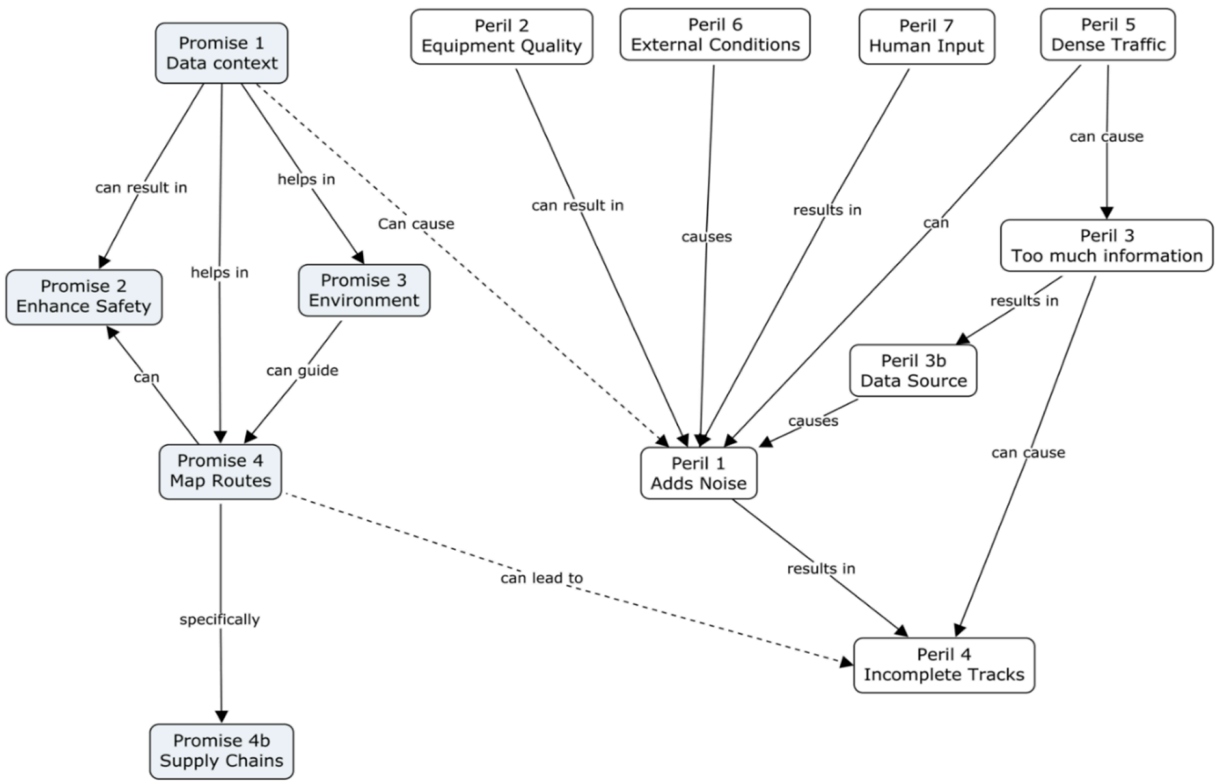

Fig. 3. Concept Map of the relationships among the Promises and Perils of AIS data.

chains. Although no specific supply chain purpose was mentioned (Dong et al., 2018), our results show evidence that AIS data can be used for more efficient operations at ports (Alessandrini et al., 2019). Concrete examples of the applications of AIS data were provided, which show the wide range of opportunities of AIS data for supply chain management. We find that new ways of using AIS data are already (close to) being introduced, like the collection of port taxes with the use of AIS data. However, for all the different purposes, AIS data should be corroborated with other data sources to be more advantageous. AIS data, when complemented with other data, can provide more useful insights.

Moreover, the results of this study showed that noise is present in three different types of AIS information. Static, voyage, and dynamic AIS data all contained noise to a certain degree. Relative to the number of transmissions, the noise found was 5.6% for sailing vessels and 0.31% for moored vessels. The results imply that the quality of equipment, i.e. Peril 2, is the underlying cause of noise in dynamic AIS data. We noticed that AIS equipment can have technical failures, which was also mentioned by Goerlandt and Kujala(2011), Papi et al.,(2015). Consequently, positional data and SOG can be incorrect (Peril 1). The qualitative data shows, in line with the literature (Kazimierski & Stateczny, 2015; Ou & Zhu, 2008), that the quality of AIS data relies on the equipment. The question that arises is under what conditions the equipment would fail, leading to noise in dynamic AIS data.

Although we found evidence for all the perils listed in Table 3, we could not identify all the specific issues outlined in the literature review. For example, different external factors can have an impact on dynamic AIS data by affecting the VHF, but the effects seem to be very minimal. The impact of fog and rain on the transmission of AIS data pointed out by Last, et al. (2015) was not confirmed. Also, the impact of overhead power lines did not seem to have an effect. On the other hand, it was found that jamming the VHF by "land" (Last et al., 2015) in the form of buildings, can be one of the causes for the noise found in the dataset. Therefore, the results imply that AIS data on either side of tall buildings should be carefully managed.

The results showed evidence for statements made in the literature about the incompleteness and irregularity of tracks (Zhao et al., 2018). However, the intermittent signals in this dataset were unlikely to be caused by a limited receiver range as suggested in the literature, since the AIS station is relatively close to the area of the data points. On the other hand, part of the intermittent tracks may be caused due to the boundaries of the focus area of this study. Also, the literature finds that there's an interaction between density and vessel detection (Høye et al., 2008) that can lead to errors in modelling (Pelich et al., 2015). The results of our study also indicate that there is a relationship between density and noise, as the most relative noise was observed at the Amsterdam canal area during King's day. This relation between noise and the number of transmissions might be explained by the phenomenon as described in the literature, which states that the tracking of vessels in dense and complex traffic areas is difficult, which could mean that not all the transmissions at King's day were received. Apart from the difficulty of tracking if a vessel transmitted a signal, finding clear causes of noise in dynamic information appears to be hard in general: "We have never done a quantitative or qualitative study to find where the best reception is. Then again, to what do you compare?" (Respondent F).

For noise in static and voyage related AIS data, different causes were found. Considering the possibilities of noise due to errors in installing and maintaining equipment, forgetting to put information in the system, circumstances beyond one's control, manipulation, and the difficulty of validating AIS data of other vessels, it can be concluded that AIS data is very prone to noise due to human data entry. This is in agreement with Zhao et al. (2018) where the authors focused mostly on the noise of static and voyage data and suggested analyzing the noise of positional data in future research since they were not able to come to conclusions for those variables. Also, given that a significant amount of time has passed since their study was conducted, it is important to assess if their findings for noise in static and voyage AIS still hold. Especially since, as seen in the interviews, the use of AIS has been improved. Therefore, this study, amongst others, partially builds on those findings and provides details about noise in static and voyage data between vessel types. It appears that passenger vessels transmit the least reliable static and voyage related data. For some attributes, the results were similar, like the ETA that was missing in 53% of the transmissions vs. 49% for ETA (including destination) as reported by Harati-Mokhtari et al., (2007). However, overall, it appeared that higher percentages of noise were found, like 2.477% for missing values of vessel names. versus the 0.5% of both vessel names and call signs reported by Harati-Mokhtari et al. (2007). More importantly, it appears that the noise in static and voyage related AIS data has not improved since. Therefore, the results indicate that the ETA and destinations provided in AIS data are unreliable and, therefore, lack valuable information in its current implementation.

The size and noise in AIS data create problems for data analysis, which was also pointed out in the literature. Moreover, the results revealed that the amount of data and noise in the data could differ between different data sources. Consequently, both researchers and practitioners need to know beforehand what requirements the AIS data should meet to solve their specific problem.

The promises and perils as summarized in Table 3, represent a Type I theory according to Gregor (2006). The promises and perils are indeed related. We have outlined the relations as a concept map in Fig. 3. The grey boxes denote the Promises on the left side of the figure, while the white boxes represent the Perils, and the dotted lines represent the indirect relationship between certain Promises and Perils. For example, Promise 1 indirectly leads to Peril 1. We notice from Fig. 3 that many of the Perils lead to Peril 1 (that AIS data contains noise). This is the reason why we call Peril 1 a general AIS data peril in Table 3. Fig. 3 might not cover all the possible relations among the concepts, but we think we cover the ones we predominantly see in the data.

5.2. Theoretical and practical implications

Multiple studies have paid specific attention to AIS related opportunities or limitations (Carson-Jackson, 2012), But, either these studies focused on one particular area (e.g. conservation) (Robards et al., 2016), predominantly on the human factors (Harati-Mokhtari et al., 2007) or mainly on S-AIS (Creswell, 1998). This study extends the literature by considering multiple opportunities and limitations of AIS data across different domains at the same time. Also, the results of this study validate and partially deny the findings of other literature that looked at potential causes of noise in AIS data. The results indicate that static and voyage related data have not improved over time.

This study combines both the promises and the perils researchers and practitioners run into. The created classification of promises and perils contributes to the literature by constituting a type I theory of AIS data. Moreover, the contributions of this study are useful as both an extension to the literature and for practitioners by taking a step back, showing how to detect noise in AIS data, what the potential causes are, and what other limitations of AIS data should be considered before using it. The results of this study provide insights for researchers who can use the findings of this study for their research. Showing the different promises of AIS data provides new directions for researchers to build upon or help them to identify new ones. Moreover, the perils show the drawbacks of AIS data that could help them to avoid the pitfalls and make more effective use of AIS data by reducing noise.

Traffic engineers and other practitioners could also benefit from a better understanding of AIS data for the optimization of their operations. This study shows that AIS data provides a rather incomplete picture of logistics. Being aware of the perils of AIS data can help practitioners in logistics by arranging their input data more effectively. The results show that AIS data is unreliable and that there's a need for more regularly updated voyage related AIS data. Practitioners can take the matter into their own hands by improving the incentives for the captain of the vessels to update their voyage related data, by creating win–win situations in collaboration with vessels' captains. Moreover, practitioners could benefit from the insights of this study when requesting commercial AIS data for the optimization of their operations, since it becomes clear that some AIS data can be redundant or contains noise under certain conditions. At last, the promises of this study provide concrete applications of AIS, which could be implemented by practitioners or build upon.

5.3. Limitations

The most important limitation of this study is that there is no quantitative data source available to triangulate the dynamic AIS data. As a result, assumptions needed to be made to identify noise in dynamic AIS data. To overcome overestimations, conservative assumptions were made which also excluded other potential noise. Next to that, the focus area was filtered on coordinates for the area of Amsterdam and, therefore, deviations in positional data could only be measured within the limits of the area. On the other hand, this means that when researchers request AIS data for their study for a specific area, the positional noise outside a focus area is already filtered. Nevertheless, it might be the case that not all the noise in the data has been found in this study as a result of these limitations. For static and voyage related AIS data, it also might be the case that noise was underestimated since it could not be verified if a particular value was incorrect.

For this study, the different variables from the used datasets matched the most common variables of AIS data as defined by International Telecommunication Union (2014). Therefore, it served the purpose of studying the problems that occur in AIS data.

In this paper, our experimental data has been about terrestrial AIS and does not include S-AIS. We only mention the promises and perils of S-AIS in our literature review (Carson-Jackson, 2012), and future work can look into both S-AIS specific data and/or perform a more exhaustive quantitative and qualitative analysis of terrestrial AIS data.

6. Conclusion

This study aims to answer the research question: *What are the promises and perils of the Automatic Identification System (AIS) data?*

The answer to the research question, consisting of both the promises and the perils of AIS data that have been identified, is summarized in Table 3. The insights from previous studies were insufficient to understand both the promises and the perils of AIS data. For this reason, we expanded the set of individual opportunities and limitations of AIS data identified in the literature.

In our mixed-design study, we have identified both the opportunities and limitations of AIS data. As far as we are aware, this is the first study that systematically analyses all the promises and perils of AIS data. For the promises, we found clear qualitative evidence, especially for Promises 1, 2, and 4. AIS data is widely used for navigation in the professional industry, and it appeared that AIS data can be useful for other purposes. However, it should always be used in corroboration with other data sources and never be considered on its own. When combined with other data, AIS data is useful for navigation, protection, efficient use of the environment, and commercial supply chain optimization purposes.

For the perils, noise in AIS data was detected caused by equipment, humans, and, to some extent, external factors. Accordingly, both quantitative and qualitative evidence was demonstrated for five common perils, that is Perils 1, 4, 5, 6, and 7. The most significant findings were that the amount of positional noise was 5.6% for sailing vessels and 0.31% for moored vessels within their specific intervals, and that missing values can reach up to 86% for static and up to 48% for voyage related AIS data. Moreover, it appeared that high buildings were often present when noise was found. On the other hand, the impact of density on AIS data, that is Peril 5, was hard to confirm, since the consequence of dense traffic areas is the loss of transmission. Although all perils are related and overlap to some extent (Fig. 3), it was not feasible to come up with strong direct quantitative proof for Peril 2 and Peril 3. However, it became clear from qualitative evidence that Peril 2 and Peril 3 are causes for other perils (seen in Fig. 3), and, therefore, its consequences can be seen in other perils. Also, Peril 3b was added since this limitation of AIS data was not frequently discussed in the literature.

For the perils, the noise in both static, voyage related, and dynamic AIS data was assessed. However, it appears to be difficult to find with certainty what dynamic information is incorrect since it cannot be compared with another data source. Future studies could focus on assessing dynamic information of AIS data in particular, by setting up a data source that keeps tracks of the deviations, impractical coordinates, impossible speeds for specific areas, present vessels that do not transmit AIS, etc. In this case, visual sight confirmation or the use of video cameras is recommended. Also, the repetition of this study in other areas for different data sources could provide more insight into the total impact of external factors.

To conclude, as a data source on its own, AIS data provides a rather incomprehensive view, providing both too much irrelevant information and too little relevant information. However, when used effectively it can be useful for many different purposes.

CRediT authorship contribution statement

Ties Emmens: Data curation, Writing - original draft, Validation. **Chintan Amrit:** Conceptualization, Supervision, Writing - review & editing. **Asad Abdi:** Writing - review & editing. **Mayukh Ghosh:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by NWO under Project number 439.16.120 – (Complexity Methods for Predictive Synchromodality (Comet-PS)).

Appendix A

	A	B	C	D	E	F	G	H	I	J
1	MMSI	TSTAMP	LATITUDE	LONGITUDE	COG	SOG	HEADING	NAVSTAT	IMO	NAME
2	256396000	1517329944	52.40167	4.774	175	0	16	5	9318034	ARCTIC BAY
3	244710876	1517330003	52.42422	4.84554	360	0	511	0	0	VOLHARDING 11
4	244660653	1517330023	52.40572	4.7729	355.3	0	511	0	0	MAIN IX
5	244750690	1517330025	52.38074	4.89987	245.9	0	511	0	0	IJVEER 56
6	244870236	1517330001	52.41013	4.90551	360	0	511	0	0	GAST
7	244740843	1517329999	52.36039	4.90394	161.8	4.4	511	0	0	TIJD ZAL HET LEREN
8	244620944	1517328491	52.36801	4.89335	360	0	511	5	0	ALBERT PIETER
9	244780319	1517330025	52.38333	4.89788	269	5.1	511	0	0	J.VERMEER
10	244750687	1517330027	52.37959	4.90242	69.3	0	511	15	0	IJVEER 51
11	244780008	1517330022	52.40064	4.86478	0	0	511	15	0	ENTERPRISE
12	244750692	1517330026	52.38704	4.89728	345.5	6.7	511	0	0	IJVEER 55
13	244690927	1517330022	52.41693	5.16418	57.6	7.7	511	0	0	MIDLIFE.C
14	244820227	1517329794	52.42115	4.85148	88.4	0	511	5	0	MAIN XXI
15	244090957	1517329929	52.33861	4.9285	360	0	511	15	0	APSARA
16	244710484	1517330021	52.40753	4.77046	0	0	511	0	0	ETERNITY
17	244670209	1517330013	52.40953	4.83602	360	0	511	15	0	COMPONIST
18	244110862	1517330022	52.36187	4.88261	360	0	511	0	0	AMSTEL ROBIJN
19	244780326	1517330013	52.3658	4.87866	0	0	511	5	0	DANIEL STALPAERT
20	235071342	1517329847	52.40623	4.88292	54	0	54	5	9419125	SEVEN ATLANTIC
21	244110652	1517329606	52.31161	5.14474	217.3	0	511	0	0	COLUMBUS

Fig. 4. . Sample AIS Data.

Appendix B

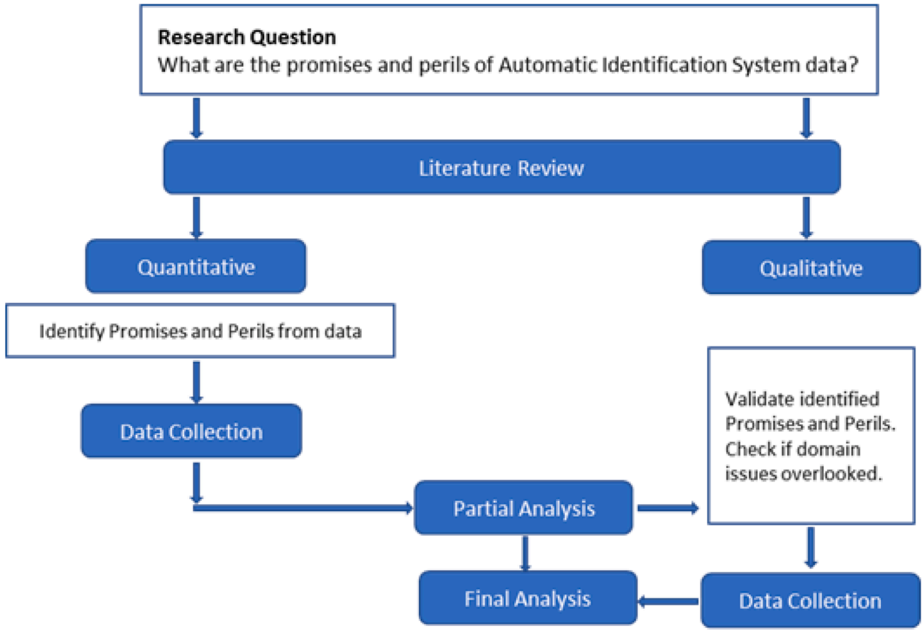

Fig. 5. Research method: representation of mixed-design study of AIS data (visual representation inspired by Santos et al. (2017).

Appendix C

Table 4
Coding: Codes, categories and references.

Name	Description	References
Incomplete story	AIS data provides only a small piece of the total information	13
Work around	Deal with incomplete information related to AIS	2
Negative consequences	Consequences of issues in AIS data	5
Mapping tracks	Consequences for analysts	1
Inconsistent tracks	Consequences for analysts specific	5
Noise	Noise in AIS data	8
Dealing with noise	How to deal with noise	8
Difference between vessel types	How vessel types differ in having noise	16
Impact of different factors	Multiple factors that influence AIS data	0
Data source related problems	Impact of data sources on AIS data	9
Equipment related issues	Impact of equipment on AIS data	17
Human input	Impact of humans on AIS data	41
Improvements over time	Difference in behaviour of humans over time	4
Manipulation	Manipulation of AIS data by humans	3
Problems in validating AIS data	Incapability of humans to validate AIS data	1
Time of the day	Impact of time of the day on human behaviour related to AIS	2
Impact of external factors	Impact of external factors on AIS data	23
Missing values	Values that are missing in AIS data	6
Redundancy	Size of available AIS data	9
Wrong value	Values that are incorrect in AIS data	25

Table 4 (*continued*)

Name	Description	References
Positional data	Positional values that are incorrect	31
Rounded values	Values that are incorrect due to rounding off	3
Speed	Speed over Ground values incorrect	15
Purposes of AIS data	Different purposes of AIS data	38
Catastrophes	AIS data used during catastrophes	2
Emission models	AIS data used for the estimation of emissions	1
Fishing	AIS data used for fishing	1
Logistics	AIS data used for logistics	4
Economical reasons	AIS data related to economical purposes	3
Port traffic monitoring	AIS data used for the monitoring of traffic in ports	11
Harbour Master	AIS data used by the harbour master	3
Back-up	AIS data as back up data for harbour master	1
Own operations	AIS data used by vessels for efficiency purposes	1
Traffic separation scheme	AIS data used by vessels to plan their own route	3
Mutual corroboration	AIS data as a way to validate other data and the other way around	0
Common sense	AIS data validated by common sense	10
Other data sources	Other data sources that can be combined with AIS data to make sense of bigger picture	4
Radar	AIS data combined with radar	15
Visual confirmation	AIS data confirmed by visual confirmation	6
Operational safety system	AIS data used for operational safety	6
Part of integrated system	AIS data combined with multiple systems for safety	3
Collission avoidance	AIS data used to prevent collissions	3
Identification and communication	AIS data used for identification and to communicate	8

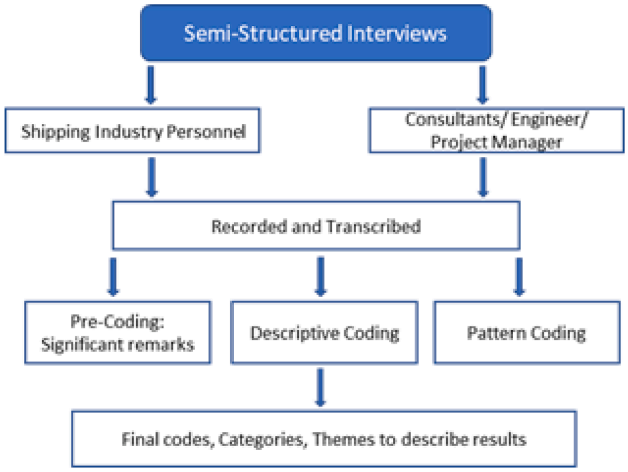

Fig. 6. Flowchart: Coding Process for Interviews conducted.

References

Alessandrini, A., Mazzarella, F., & Vespe, M. (2019). Estimated time of arrival using historical vessel tracking data. *IEEE Transactions on Intelligent Transportation Systems, 20*(1), 7–15. https://doi.org/10.1109/tits.2017.2789279

Amrit, C., van Hillegersberg, J., & Kumar, K. (2012). Identifying coordination problems in software development: finding mismatches between software and project team structures. ArXiv:1201.4142.

Arifin, Z., Ibrahim, M. R., & Hatta, H. R. (2016). Nearest tourism site searching using Haversine method. In *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* (pp. 293–296). https://doi.org/10.1109/ICITACEE.2016.7892458

Balduzzi, M., Pasta, A., & Wilhoit, K. (2014). A security evaluation of AIS automated identification system. In *ACM International Conference Proceeding Series, 2014-December(December)* (pp. 436–445). https://doi.org/10.1145/2664243.2664257

Basyir, M., Nasir, M., Suryati, S., & Mellyssa, W. (2018). Determination of nearest Emergency Service Office using Haversine formula based on android platform. *EMITTER International Journal of Engineering Technology, 5*(2), 270–278. https://doi.org/10.24003/emitter.v5i2.220

Ben Ayed, A., Ben Halima, M., & Alimi, A. M. (2015). Big data analytics for logistics and transportation. In *In 2015 4th International Conference on Advanced Logistics and Transport (ICALT)*. IEEE. https://doi.org/10.1109/icadlt.2015.7136630.

Carson-Jackson, J. (2012). Satellite AIS – developing technology or existing capability? *Journal of Navigation, 65*(2), 303–321. https://doi.org/10.1017/s037346331100066x

Cervera, M. A., Ginesi, A., & Eckstein, K. (2011). Satellite-based vessel Automatic Identification System: A feasibility and performance analysis. *International Journal of Satellite Communications and Networking, 29*(2), 117–142. https://doi.org/10.1002/sat.957

Chopde, N., & Nichat, M. (2013). Landmark based shortest path detection by using A* and Haversine Formula. *International Journal of Innovative Research in Computer and Communication Engineering, 1*(2), 298–302. http://www.ijircce.com/upload/2013/april/17_V1204030_Landmark_H.pdf.

Creswell, J. (1998). *Qualitative Inquiry And Research Design*. Sage.

Di Ciccio, C., van der Aa, H., Cabanillas, C., Mendling, J., & Prescher, J. (2016). Detecting flight trajectory anomalies and predicting diversions in freight transportation. *Decision Support Systems, 88*, 1–17.

Dobrkovic, A., Iacob, M.-E., & van Hillegersberg, J. (2015). Using machine learning for unsupervised maritime waypoint discovery from streaming AIS data. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business - i-KNOW '15*. ACM Press. https://doi.org/10.1145/2809563.2809573.

Dobrkovic, A., Iacob, M.-E., & van Hillegersberg, J. (2018). Maritime pattern extraction and route reconstruction from incomplete AIS data. *International Journal of Data Science and Analytics, 5*(2–3), 111–136. https://doi.org/10.1007/s41060-017-0092-8

Dong, C., Boute, R., McKinnon, A., & Verelst, M. (2018). Investigating synchromodality from a supply chain perspective. *Transportation Research Part D: Transport and Environment, 61*, 42–57. https://doi.org/10.1016/j.trd.2017.05.011

Goerlandt, F., & Kujala, P. (2011). Traffic simulation based ship collision probability modeling. *Reliability Engineering & System Safety, 96*(1), 91–107. https://doi.org/10.1016/j.ress.2010.09.003

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly, 30*(3), 611. https://doi.org/10.2307/25148742

Greidanus, H., Alvarez, M., Eriksen, T., & Gammieri, V. (2015). Completeness and accuracy of a wide-area maritime situational picture based on automatic ship reporting systems. *Journal of Navigation, 69*(1), 156–168. https://doi.org/10.1017/s0373463315000582

Hadjieleftheriou, M., Kollios, G., Gunopulos, D., & Tsotras, V. J. (2003). In *On-Line Discovery of Dense Areas in Spatio-temporal Databases* (pp. 306–324). https://doi.org/10.1007/978-3-540-45072-6_18

Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic Identification System (AIS): data reliability and human error implications. *Journal of Navigation, 60* (3), 373–389. https://doi.org/10.1017/s0373463307004298

Hatch, L., Clark, C., Merrick, R., Van Parijs, S., Ponirakis, D., Schwehr, K., et al. (2008). Characterizing the relative contributions of large vessels to total ocean noise fields: A case study using the Gerry E. Studds Stellwagen Bank National Marine Sanctuary. *Environmental Management, 42*(5), 735–752. https://doi.org/10.1007/s00267-008-9169-4

Høye, G. K., Eriksen, T., Meland, B. J., & Narheim, B. T. (2008). Space-based AIS for global maritime traffic monitoring. *Acta Astronautica, 62*(2–3), 240–245. https://doi.org/10.1016/j.actaastro.2007.07.001

Iphar, C., Napoli, A., & Ray, C. (2015). Detection of false AIS messages for the improvement of maritime situational awareness. In *OCEANS 2015 - MTS/IEEE Washington* (pp. 1–7). https://doi.org/10.23919/OCEANS.2015.7401841

Jalkanen, J.-P., Johansson, L., & Kukkonen, J. (2014). A comprehensive inventory of the ship traffic exhaust emissions in the Baltic Sea from 2006 to 2009. *Ambio, 43*(3), 311–324. https://doi.org/10.1007/s13280-013-0389-3

Jalkanen, J. P., Johansson, L., Kukkonen, J., Brink, A., Kalli, J., & Stipa, T. (2012). Extension of an assessment model of ship traffic exhaust emissions for particulate matter and carbon monoxide. *Atmospheric Chemistry and Physics, 12*(5), 2641–2659. https://doi.org/10.5194/acp-12-2641-2012

Johansson, L., Jalkanen, J. P., Kalli, J., & Kukkonen, J. (2013). The evolution of shipping emissions and the costs of regulation changes in the northern EU area. *Atmospheric Chemistry and Physics, 13*(22), 11375–11389. https://doi.org/10.5194/acp-13-11375-2013

Johansson, L., Jalkanen, J.-P., & Kukkonen, J. (2017). Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. *Atmospheric Environment, 167*, 403–415. https://doi.org/10.1016/j.atmosenv.2017.08.042

Kang, L., Meng, Q., & Liu, Q. (2018). Fundamental diagram of ship traffic in the Singapore Strait. *Ocean Engineering, 147*, 340–354. https://doi.org/10.1016/j.oceaneng.2017.10.051

Kang, L., Meng, Q., Zhou, C., & Gao, S. (2019). How do ships pass through L-shaped turnings in the Singapore strait? *Ocean Engineering, 182*, 329–342. https://doi.org/10.1016/j.oceaneng.2019.04.033

Kaplan, B., & Duchon, D. (1988). Combining qualitative and quantitative methods in information systems research: A case study. *MIS Quarterly, 12*(4), 571. https://doi.org/10.2307/249133

Kazimierski, W., & Stateczny, A. (2015). Radar and automatic identification system track fusion in an electronic chart display and information system. *Journal of Navigation, 68*(6), 1141–1154. https://doi.org/10.1017/s0373463315000405

Kim, K.-I., & Lee, K. (2018). Dynamic programming-based vessel speed adjustment for energy saving and emission reduction. *Energies, 11*(5), 1273. https://doi.org/10.3390/en11051273

Kontopoulos, I., Spiliopoulos, G., Zissis, D., Chatzikokolakis, K., & Artikis, A. (2018). Countering real-time stream poisoning: An architecture for detecting vessel spoofing in streams of AIS data. In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (pp. 981–986). https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00139

Kontopoulos, I., Chatzikokolakis, K., Zissis, D., Tserpes, K., & Spiliopoulos, G. (2020). Real-time maritime anomaly detection: Detecting intentional AIS switch-off. *International Journal of Big Data Intelligence, 7*(2), 85. https://doi.org/10.1504/IJBDI.2020.107375

Last, P., Hering-Bertram, M., & Linsen, L. (2015). How automatic identification system (AIS) antenna setup affects AIS signal quality. *Ocean Engineering, 100*, 83–89. https://doi.org/10.1016/j.oceaneng.2015.03.017

Le Guyader, D., Ray, C., Gourmelon, F., & Brosset, D. (2017). Defining high-resolution dredge fishing grounds with Automatic Identification System (AIS) data. *Aquatic Living Resources, 30*, 39. https://doi.org/10.1051/alr/2017038

Mazzarella, F., Vespe, M., Alessandrini, A., Tarchi, D., Aulicino, G., & Vollero, A. (2017). A novel anomaly detection approach to identify intentional AIS on-off switching. *Expert Systems with Applications, 78*, 110–123. https://doi.org/10.1016/j.eswa.2017.02.011

Mazzarella, F., Vespe, M., & Santamaria, C. (2015). SAR ship detection and self-reporting data fusion based on traffic knowledge. *IEEE Geoscience and Remote Sensing Letters, 12*(8), 1685–1689. https://doi.org/10.1109/lgrs.2015.2419371

Miles, M. B., & Huberman, A. M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher, 13*(5), 20–30. https://doi.org/10.3102/0013189x013005020

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE.

Ngai, E. W. T., Leung, T. K. P., Wong, Y. H., Lee, M. C. M., Chai, P. Y. F., & Choi, Y. S. (2012). Design and development of a context-aware decision support system for real-time accident handling in logistics. *Decision Support Systems, 52*(4), 816–827.

Ou, Z., & Zhu, J. (2008). AIS database powered by GIS technology for maritime safety and security. *Journal of Navigation, 61*(4), 655–665. https://doi.org/10.1017/s0373463308004888

Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy, 15*(12), 2218–2245. https://doi.org/10.3390/e15062218

Papi, F., Tarchi, D., Vespe, M., Oliveri, F., Borghese, F., Aulicino, G., & Vollero, A. (2015). Radiolocation and tracking of automatic identification system signals for maritime

situational awareness. *IET Radar, Sonar & Navigation, 9*(5), 568–580. https://doi.org/10.1049/rsn2.v9.510.1049/iet-rsn.2014.0292
Pelich, R., Longepe, N., Mercier, G., Hajduch, G., & Garello, R. (2015). AIS-based evaluation of target detectors and SAR sensors characteristics for maritime surveillance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(8), 3892–3901. https://doi.org/10.1109/jstars.2014.2319195
Robards, M. D., Silber, G. K., Adams, J. D., Arroyo, J., Lorenzini, D., Schwehr, K., & Amos, J. (2016). Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)—a review. *Bulletin of Marine Science, 92*(1), 75–103. https://doi.org/10.5343/bms.2015.1034
Šakan, D., Rudan, I., Žuškin, S., & Brčić, D. (2018). Near real-time S-AIS. *Pomorstvo, 32*(2), 211–218. https://doi.org/10.31217/p.32.2.6
Saldaña, J. (2015). *The coding manual for qualitative researchers*. SAGE.
Salmon, L., Ray, C., & Claramunt, C. (2015). A hybrid approach combining real-time and archived data for mobility analysis. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on GeoStreaming*. https://doi.org/10.1145/2833165.2833176
Salmon, L., Ray, C., & Claramunt, C. (2016). Continuous detection of Black Holes for moving objects at sea. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoStreaming, IWGS 2016*. https://doi.org/10.1145/3003421.3003423
Santos, J. L. G., Erdmann, A. L., Meirelles, B. H. S., Lanzoni, G. M. D. M., Cunha, V. P. D., & Ross, R. (2017). Integrating quantitative and qualitative data in mixed methods research. *Texto & Contexto-Enfermagem, 26*(3).
Shu, Y., Daamen, W., Ligteringen, H., & Hoogendoorn, S. P. (2017). Influence of external conditions and vessel encounters on vessel behavior in ports and waterways using Automatic Identification System data. *Ocean Engineering, 131*, 1–14. https://doi.org/10.1016/j.oceaneng.2016.12.027
Silveira, P. A. M., Teixeira, A. P., & Soares, C. G. (2013). Use of AIS data to characterise marine traffic patterns and ship collision risk off the coast of Portugal. *Journal of Navigation, 66*(6), 879–898. https://doi.org/10.1017/s0373463313000519
Tsou, M.-C. (2010). Discovering knowledge from AIS database for application in VTS. *Journal of Navigation, 63*(3), 449–469. https://doi.org/10.1017/s0373463310000135
van der Hoop, J. M., Vanderlaan, A. S. M., & Taggart, C. T. (2012). Absolute probability estimates of lethal vessel strikes to North Atlantic right whales in Roseway Basin, Scotian Shelf. *Ecological Applications, 22*(7), 2021–2033. https://doi.org/10.1890/11-1841.1
Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly, 37*(1), 21–54.
Yau, P. S., Lee, S. C., Corbett, J. J., Wang, C., Cheng, Y., & Ho, K. F. (2012). Estimation of exhaust emission from ocean-going vessels in Hong Kong. *Science of The Total Environment, 431*, 299–306. https://doi.org/10.1016/j.scitotenv.2012.03.092
Young, J. C., Rose, D. C., Mumby, H. S., Benitez-Capistros, F., Derrick, C. J., Finch, T., et al. (2018). A methodological guide to using and reporting on interviews in conservation science research. *Methods in Ecology and Evolution, 9*(1), 10–19. https://doi.org/10.1111/2041-210X.12828
Zhao, L., Shi, G., & Yang, J. (2018). Ship trajectories pre-processing based on AIS data. *Journal of Navigation, 71*(5), 1210–1230. https://doi.org/10.1017/s0373463318000188
Zhao, Z., Ji, K., Xing, X., Zou, H., & Zhou, S. (2013). Ship surveillance by integration of space-borne SAR and AIS – Review of current research. *Journal of Navigation, 67*(1), 177–189. https://doi.org/10.1017/s0373463313000659
Zhou, M., & van der Veen, A.-J. (2014). Blind beamforming techniques for automatic identification system using GSVD and tracking. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. https://doi.org/10.1109/icassp.2014.6854153.