# Comprehensive Analysis of Automatic Identification System (AIS) Data in Regard to Vessel Movement Prediction

Philipp Last[1,2], Christian Bahlke[1], Martin Hering-Bertram[3]
and Lars Linsen[2]

[1](*Institute for Maritime Simulation, University of Applied Sciences Bremen, Germany*)
[2](*School of Engineering and Science, Jacobs University, Bremen, Germany*)
[3](*Institute of Informatics and Automation, University of Applied Sciences Bremen, Germany*)
(E-mail: philipp.last@hs-bremen.de)

AIS was primarily developed to exchange vessel-related data among vessels or AIS stations by using very-high frequency (VHF) technology to increase safety at sea. This study evaluates the formal integrity, availability, and the reporting intervals of AIS data with a focus on vessel movement prediction. In contrast to former studies, this study is based on a large data collection of over 85 million AIS messages, which were continuously received within a time period of two months. Thus, the evaluated data represent a comprehensive and up-to-date view of the current usage of AIS systems installed on vessels. Results of previous studies concerning the availability of AIS data are confirmed and extended. New aspects such as reporting intervals are additionally evaluated. Received messages are stored in a database, which allows for performing database queries to evaluate the obtained data in an automatic way. This study shows that almost ten years after becoming mandatory for professional operating vessels, AIS still lacks availability for both static and dynamic data and that the reporting intervals are not as reliable as specified within the technical AIS standard.

1.  INTRODUCTION.   The Automatic Identification System (AIS) allows for transmitting data between AIS systems, which can be installed on vessels, base stations like harbour authorities, marks such as buoys, or on search and rescue airplanes. The AIS data that are exchanged are divided into three different types (IMO, 2003):

- Static data (e.g., vessel name and the dimensions of the vessel)
- Dynamic data (e.g., vessel position, course over ground and heading)
- Voyage-related data (e.g., current draught, description of cargo, and destination).

Thus AIS is a useful complement to systems such as radar by providing additional information that would otherwise not be available.

Whereas the static data are set once and cannot be altered by the crew itself, the position report data fields which represent the dynamic data contain measurements which are updated by different sensors directly connected to the AIS system and, therefore, change continuously over time. For example, a Global Positioning System (GPS) receiver provides the current vessel position consisting of latitude and longitude and a time stamp, whereas a compass – when connected – provides the current heading of the vessel. Both static and dynamic AIS data provide useful information for course corrections and collision avoidance, respectively. In addition, AIS data allow users such as a vessel's crew and port authorities to gain a more detailed overview of the current traffic situation at sea.

However, users report that the static and dynamic data provided by AIS systems is sometimes partially missing or simply wrong. Harati-Mokhtari et al. (2007) mentioned that "Poor performance and transmission of erroneous information by AIS are important issues that can affect its usefulness". Since information provided by AIS is also used as an additional overlay in other systems like radar, the examination of the quality is important when it comes to manual collision avoidance manoeuvres. Furthermore, the data provided by AIS can be used in algorithms to implement automatic approaches for collision avoidance. AIS data is used when it comes to legal questions like investigating collisions or identifying vessels that are exceeding speed limits on waterways. Thus, it is important that AIS data are reliable and their integrity and availability are ensured.

Several studies have examined the static and voyage-related data which is considered to be useful for manual collision avoidance. Harati-Moktari et al. (2007) evaluated static AIS data that was recorded at Liverpool Bay with the main focus on identifying invalid Maritime Mobile Service Identity (MMSI) numbers. Furthermore, they evaluated a data set to identify incorrect MMSI and International Maritime Organisation (IMO) numbers, and two dynamic data fields describing course over ground (COG) and speed over ground (SOG). It is mentioned that further research of the dynamic data fields such as heading (HDG) is recommended. However, only a few studies exist which evaluate the dynamic data, although former studies mention the importance of investigating the dynamic data with respect to collision avoidance and data fusion with other systems. Felski and Jaskólski (2012b; 2012c) evaluated the integrity of dynamic AIS data for the Gulf of Gdansk based on weekly analyses with a focus on the availability of the dynamic data fields HDG and rate of turn (ROT). However, these studies did not evaluate all dynamic data fields, e.g., an evaluation of the position accuracy flag is missing. These studies focused on manual collision avoidance, which is performed by crew members using available data from AIS and other means and contacting other vessels where necessary. In conclusion, existing studies have evaluated AIS data by considering either a small data set or short time periods and evaluated AIS data fields referring to human collision avoidance.

In contrast, we present a study that evaluates the usefulness of static and dynamic data fields with respect to a technical view on collision avoidance algorithms for Class A AIS systems, which are used on professional operating vessels. The quality criteria used in our study are integrity and availability. We evaluate which data fields are appropriate for the implementation of motion models or collision avoidance algorithms. Relating to implementing such algorithms, the reporting intervals of AIS

data as described in ITU (2010) play an important role and are therefore evaluated. Felski and Jaskólski (2012a) evaluated the availability of AIS by transmitting specific AIS data. This availability was correlated with the AIS reporting intervals, which have not been evaluated in any study. Moreover, in contrast to former studies such as Felski and Jaskólski (2013), our study assigns received AIS data to vessels. Former studies evaluate the availability of data fields referring to AIS message types, which makes it difficult to give a detailed conclusion about how AIS systems are currently configured and used by vessels.

To achieve reliable evaluations, our study is based on a large data set, which includes AIS data received at the German North Sea coast for a time period of two months. Over 85 million received AIS messages are evaluated. To handle the amount of data, all received AIS messages are stored in a database management system with a schema that allows for performing appropriate database queries over a large data set in a non-error-prone and automatic way.

2.   DATA SET.   Within our study, AIS data from the German part of the North Sea including waterways was collected for a time period of two months starting from 18 June until 18 August 2013. AIS receivers were placed at the following locations:

Table 1.  AIS receiver locations.

| Area | N | E | Receiver Type |
| --- | --- | --- | --- |
| Bremen, Germany | 53°05′51·05″ | 008°46′08·39″ | Dual Channel |
| Bremerhaven, Germany | 53°33′42·01″ | 008°33′15·38″ | Dual Channel |
| Wangerooge, Germany | 53°47′19·44″ | 007°53′57·07″ | Dual Channel |
| Cuxhaven, Germany | 53°51′37·57″ | 008°43′20·97″ | Dual Channel |
| Emden, Germany | 53°20′26·45″ | 007°12′20·68″ | Dual Channel |
| Helgoland, Germany | 54°10′39″ | 007°53′32·00″ | Dual Channel |

Each receiver has coverage of at least 50 km up to 150 km depending on antenna placement and weather conditions. As a result of the coverage, the receiver footprints intersect, meaning that it is possible that one message is received several times by different antennas. To consider only unique AIS messages, duplicates have been removed from the data set. Within the mentioned time period, exactly 85,402,344 unique AIS messages of all 27 message types were received. All AIS receivers use the same true dual-channel technology that allows parallel monitoring of both VHF channels 161·975 MHz and 162·025 MHz, which are used for AIS transmissions since using single channel receivers would result in a data loss (IMO, 2006).

A database management system is used to store the received data into appropriate tables. Table 2 gives a general overview of the distribution of all AIS message types. The static and voyage-related vessel data are sent as message type five named "Static and Voyage-related Data". The dynamic data of Class A systems are sent in the form of position reports, i.e., message types one, two, and three. Percentages are rounded to two decimal places.

AIS message types one, two, and three that have an identical format, provide a total of 80% of all received messages. The percentage of message type 18, which describes the position reports for Class B AIS systems, is approximately 1·69% of all

Table 2. General AIS message distribution.

| Message type | Count | Percentage | Message type | Count | Percentage |
|---|---|---|---|---|---|
| 1 | 58,778,376 | 68.83% | 15 | 1,025,857 | 1.20% |
| 2 | 44,608 | 0.05% | 16 | 8,679 | 0.01% |
| 3 | 9,496,341 | 11.12% | 17 | 4 | 0% |
| 4 | 3,797,737 | 4.45% | 18 | 1,443,227 | 1.69% |
| 5 | 6,166,300 | 7.22% | 19 | 160 | 0% |
| 6 | 2,095 | 0% | 20 | 1,899,894 | 2.22% |
| 7 | 5,889 | 0.01% | 21 | 1,197,622 | 1.40% |
| 8 | 833,619 | 0.98% | 22 | 1,880 | 0% |
| 9 | 8 | 0% | 23 | 1 | 0% |
| 10 | 1,575 | 0% | 24 | 483,154 | 0.57% |
| 11 | 3,344 | 0% | 25 | 6 | 0% |
| 12 | 165 | 0% | 26 | 2 | 0% |
| 13 | 103 | 0% | 27 | 2 | 0% |
| 14 | 144 | 0% | Corrupt | 211,552 | 0.25% |

messages – very small when compared to the position reports of Class A systems. Class B AIS systems provide less information than Class A systems and are usually installed on leisure vessels or buoys. Some message types such as type 23 "Group assignment command" rarely occur. Message types 12 and 14 represent safety-related messages. Interestingly, four messages of types 26 and 27 were received. These message types are supposed to be for satellite AIS and should be transmitted on different VHF channels. The meaning of the remaining message types is beyond the purpose of this paper and can be looked up in literature (ITU, 2010). According to the description above, our study has its focus on evaluating static and dynamic data of Class A systems. Table 2 also shows that a small percentage - approximately 0·25% of all received messages were corrupt. The following section clarifies the meaning of data integrity and, therefore, the meaning of corrupt messages.

3. DATA INTEGRITY. Felski and Jaskólski (2013) state that the data integrity of AIS messages in their study is determined by completeness of data fields. In contrast to their work, in our study the term data integrity describes the formal correctness of AIS messages and AIS data fields. For AIS messages in their entirety, integrity is ensured by appending a checksum to each message. This checksum calculation is done by performing an exclusive disjunction of the message payload. Furthermore, each message type has a specific bit length or an interval with a minimum and maximum bit length. For example, position report Class A messages are supposed to have a fixed message length of 168 bits. Message type five describing static and voyage-related data occupies two time slots when sent and is supposed to have a bit length of 424 bits (ITU, 2010). Besides wrong checksums, AIS messages have been observed where the checksum matches the payload but the bit length for the specified message type is wrong. In addition, incomplete AIS messages have occurred, where at least one message fragment has not been received. For all these cases the data integrity is violated, resulting in messages that cannot be processed. For 211,552 AIS messages, i.e., in approx. 0·25% of all cases, the data integrity has been violated. Table 3 shows the distribution of the different integrity errors. 820 AIS messages of all types have

Table 3. Data integrity errors.

| Error type | Count | Percentage |
|---|---|---|
| CRC error | 237 | 0·11% |
| Bad bit length | 820 | 0·39% |
| Incomplete/fragment | 190,722 | 90·17% |
| Other | 19,733 | 9·33% |

Table 4. Evaluated position report data fields.

| Data field | Unit | Range | n.a. value |
|---|---|---|---|
| Longitude | [°] | ±180 | 181 |
| Latitude | [°] | ±90 | 91 |
| Rate of turn (ROT) | $[\frac{°}{m}]$ | ±127 | 128 |
| Speed over ground (SOG) | [kn] | [0,1022] | 1023 |
| Course over ground (COG) | [°] | [0,3599] | 3600 |
| Heading (HDG) | [°] | [0,359] | 511 |
| Position accuracy (ACC) | − | [*true, false*] | *false* |

been identified, where the checksum matches to payload but the bit length does not accord to the technical characteristics as described in ITU (2010). Neither messages with a wrong checksum nor messages with a wrong bit length nor message fragments are taken into account during our evaluations.

4. EVALUATED DATA FIELDS. The main goal of this study is to evaluate static and dynamic data fields relating to integrity, availability and usability for collision avoidance algorithms. Table 4 shows the data fields of the position reports that are suitable for this and which are, therefore, evaluated within this study. It has to be mentioned that a further dynamic data field exists beside the listed ones in Table 4. "Navigational status" is also a position report data field and considered as dynamic since its value can be changed manually by crew members. However, we only consider dynamic data fields that contain measurements automatically updated by appropriate sensors directly connected to the AIS system. For that reason the dynamic data field "Navigational status" is not listed in Table 4 and each reference to dynamic data in our study is not related to this dynamic data field.

SOG and COG values are specified in 1/10 steps. Each data field in Table 4 has a range of valid values and a special value indicating that the specific information is not available (ITU, 2010). The availability depends on the connected sensors, e.g., the position depends on the current GPS status. Position accuracy is represented as a flag without a unit. It indicates whether the received position of a vessel has an accuracy of $ACC \leqslant 10\,m$. Some previous studies evaluate each position report as a separate statistical unit. This approach does not allow for giving a detailed conclusion about how AIS systems are currently configured and used by vessels. Therefore all data fields are evaluated within our study with respect to the vessel that sent the data, since each data field gives information about a vessel's current state. Matching data to vessels is done using the MMSI numbers. Taking value restrictions into account, logical

Table 5. Reporting intervals of Class A position reports.

| Vessel state | Interval |
|---|---|
| At anchor/moored and not moving faster than 3 kn | 180 s |
| At anchor/moored and moving faster than 3 kn | 10 s |
| $SOG \in [0\,\text{kn}, 14\,\text{kn}]$ | 10 s |
| $SOG \in [0\,\text{kn}, 14\,\text{kn}]$ and changing course | 3 1/3 s |
| $SOG \in (14\,\text{kn}, 23\,\text{kn}]$ | 6 s |
| $SOG \in (14\,\text{kn}, 23\,\text{kn}]$ and changing course | 2 s |
| $SOG > 23\,\text{kn}$ | 2 s |
| $SOG > 23\,\text{kn}$ and changing course | 2 s |

database queries are formulated to allow for an automatic evaluation of the dynamic data listed in Table 4 for the following scenarios:

- Identifying the quantity of unique vessels.
- Evaluating for each data field whether it is always, partially, or never available for each vessel.
- Identifying the quantity of vessels that solely transmit incomplete, partially complete, and complete position reports.

For the last two points only vessels which sent at least 100 position reports are taken into account, since some vessels only appear for short times at the border areas of the AIS receivers. The following scenarios are evaluated for message type five that includes the dimensions and antenna position of a vessel:

- Identifying all vessels that transmit position reports but never transmit static data.
- Identifying the number of vessels that always transmit unavailable values. In this case, all dimension data fields contain zero values.
- Identifying the number of vessels that transmit their dimensions but no GPS antenna position.

As a last step, reporting intervals for both static and dynamic data are evaluated. The transmission frequency of position reports is crucial when it comes to the implementation of collision avoidance algorithms or developing vessel motion models. Until now, this factor has not been examined in any previous study. According to the technical characteristics, the reporting interval changes depending on SOG, ROT, and the current state of a vessel (ITU, 2010). Therefore, position reports should be output periodically by mobile stations (*ibid.*). Table 5 gives an overview of the reporting intervals for position reports according to these technical characteristics.

In contrast, static and voyage-related data should be sent each 360 seconds. According to the reporting intervals the following scenarios are evaluated:

- Evaluating the time between coherent position reports not related to stationary vessel states such as being moored, since the vessel state is entered manually and has been identified as erroneous in former studies such as the one by Harati-Mokhati et al. (2007). Only vessels which sent at least 100 position reports are taken into account.
- Evaluating the static data in the same way.

Table 6. Analysis of the dimension availability.

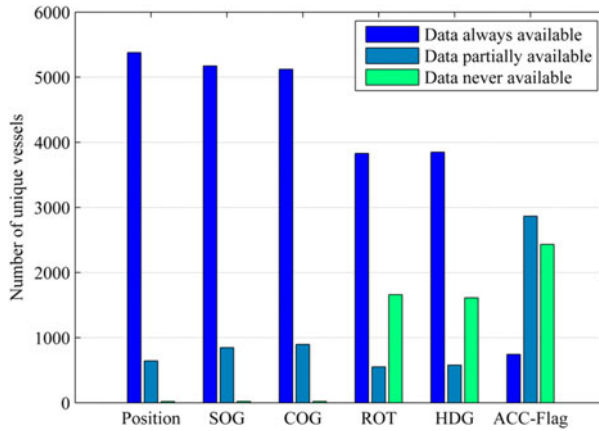| Unique vessels | Complete Reports | No Antenna Position | Rect. Bow | n. a. |
|---|---|---|---|---|
| 8171 | 7846 (96·02%) | 45 (0·55%) | 2 (0·03%) | 278 (3·40%) |



Figure 1. Availability of position report data fields.

## 5. STATISTICAL RESULTS AND DISCUSSION

5.1. *Availability of static data*.   The focus of this section is to evaluate the availability of vessel dimensions and GPS antenna position, since these data fields provide valuable information for collision avoidance algorithms. Each vessel should transmit the distances to the vessel borders based on the position of the GPS receiver that is connected to the AIS system (ITU, 2010). The following data fields are transmitted: distances in metres of the GPS antenna to bow, stern, port, and starboard. All data fields should be zero if no values are available. It can be possible that the antenna position is unknown but the vessel's dimensions are known. In this case the distances to bow and port become zero and the data fields for stern and starboard describe the length and width of the vessel. As a rare case the antenna could be placed at the portside corner of a rectangular bow. In this case, either the value for bow or the value for port has to be set to one (IMO, 2003). Table 6 shows the results of evaluating the dimensions according to vessels.

Even if there is a high availability of approximately 96% for the dimension values, it is still possible that the dimensional data is error-prone, since it was entered manually. Therefore the uncertainty of these data has to be considered in collision avoidance algorithms, especially when vessels are close to each other. Furthermore, 180 vessels have been identified where dynamic data but no static data has been received. Possible reasons for this are listed in Section 5.4.

5.2. *Availability of dynamic data*.   Figure 1 depicts the availability of each data field relating to unique vessels. In total 6,042 unique vessels which sent at least 100 position reports with SOG greater than 3 kn have been identified leading to a total of 65,741,377 evaluated position reports.

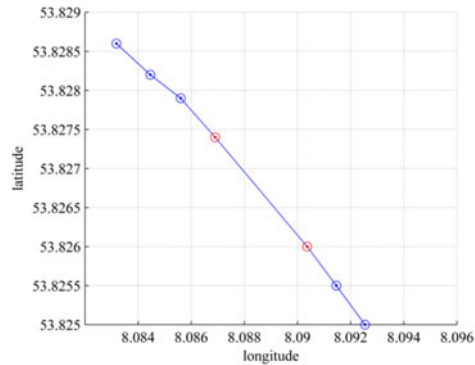| Latitude | Longitude |
|----------|-----------|
| 53.825   | 8.09254   |
| 53.8255  | 8.09146   |
| 53.826   | 8.09036   |
| 91.0     | 181.0     |
| 91.0     | 181.0     |
| 53.8274  | 8.0869    |
| 53.8279  | 8.08560   |
| 53.8282  | 8.08446   |
| 53.8286  | 8.08318   |



Figure 2. Recorded not available values for latitude and longitude of a vessel.

The term "availability" means that a dynamic data field contains a measurement value according to column "Range" of Table 4. "Not available" means that a dynamic data field contains the appropriate value of the "n.a. value" column of Table 4. Hence, the bars for "Data partially available" comprise all identified vessels that deliver both available and not available values for the respective data field. An example vessel belonging to the bar "Data partially available" is a vessel that transmits from time to time not available values for latitude and longitude depending on the current GPS state and the condition of the GPS antennas. Figure 2 shows an example of this by plotting a short history track for a recorded vessel. "Not available" values for the position data fields have been transmitted twice as shown within the table next to the plotted vessel track. For that reason the GPS position of the vessel is unknown between the red marked position reports. The vessel travels a distance of more than 250 metres without sending valid information about the current vessel position. Cases like this lead to problems when it comes to the prediction of vessel movements since prediction algorithms are mostly based on two pieces of data:

1. Data provided by the system. In this case mostly the dynamic AIS data measured by sensors.
2. The differential time $\Delta t$ between two updates, within this context called reporting intervals respectively the time passed between two AIS messages.

Concerning the first statement, Figure 1 shows that even if there is a high availability of dynamic data fields such as COG or SOG it is still likely that from time to time "not available" values are sent beside actual measurements (see "Data partially available" bars). More detailed results concerning Figure 1 are shown in Table 7. Sending "not available" information for a dynamic data field has the same impact on prediction with respect to collision avoidance algorithms as lost AIS messages. These effects are discussed in Section 5.4. Regarding the second statement, Section 5.4 also shows an example prediction algorithm using a vessel motion model derived from the dynamic AIS data. The influence that the reporting intervals have on the example and similar motion models is presented. Furthermore, it is shown why continuously updated measurements are essential to predict vessel movements.

With regard to the dynamic data fields, it is mentioned by Harati-Mokhtari et al. (2007) that vessel sensors may be turned off when a vessel is moored or at anchor

Table 7.  Detailed availability results concerning dynamic data fields.

| Data field | Always available | Partially available | Never available |
|---|---|---|---|
| Position | 5379 (89·03%) | 644 (10·66%) | 19 (0·31%) |
| SOG | 5173 (85·62%) | 847 (14·02%) | 22 (0·36%) |
| COG | 5123 (84·79%) | 897 (14·85%) | 22 (0·36%) |
| ROT | 3830 (63·39%) | 551 (9·12%) | 1661 (27·49%) |
| HDG | 3850 (63·72%) | 579 (9·58%) | 1613 (26·70%) |
| Accuracy flag | 744 (12·31%) | 2866 (47·43%) | 2432 (40·25%) |
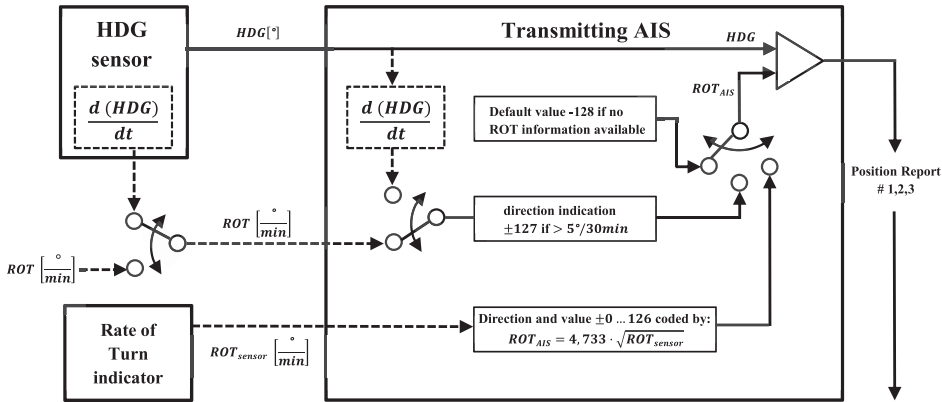


Figure 3.  Rate of Turn sensor input overview based on (IMO, 2003. P10).

while position reports are still being sent. For this reason further assumptions are made, e.g., if a vessel's speed is lower or equal to three knots, it is assumed that a vessel is not moving and therefore sensors delivering, e.g., ROT or HDG data may be inactive. Thus, position reports of anchoring or moored vessels are not part of this evaluation and are filtered out by using the 3 kn criterion. Position, SOG, and COG represent the data fields that have the highest availability. For ROT and HDG similar results concerning the availability exist. Approximately 63% of all vessels always transmit these data fields; about 9% and 27% respectively of the vessels partially or never transmit actual measurements, i.e., values are not available. The obtained results for HDG and ROT differ from studies such as Felski and Jaskólski (2012b) since the percentages of no availability values are lower in their study. Figure 3 gives an explanation for the correlation between ROT and HDG. Figure 3 shows the signal path from the vessel's sensors to a generated AIS position report message that includes the HDG and ROT information.

As can be seen, there are three possibilities for the calculation of ROT, whereas HDG is always obtained from a specific sensor:

1. No ROT information is available. In this case the not available value-128 is used.
2. Another ROT source with restricted information is used or ROT is derived from the HDG sensor. Deriving ROT from the HDG sensor can be done inside or

outside of the AIS system and can give a rough indication if the vessel is turning with more than 5° per 30 minutes.

3. A Rate of Turn indicator is connected. In this case a precise ROT value can be calculated and stored within a position report.

However, a separate Rate of Turn indicator for detailed ROT values is only mandatory for vessels with a gross tonnage of 50,000 and upwards as regulated by the IMO SOLAS Chapter V restrictions. Furthermore, only vessels with a gross tonnage of 500 and upwards are required to be equipped with a gyro compass providing HDG information which can be used to derive ROT. This additionally reduces the possibilities to determine ROT values. For that reason no or restricted ROT information is available in most cases. This is confirmed by our evaluation results. Only 829 (18·92%) of all vessels which transmit ROT values transmit detailed turning information. 3,552 (81·08%) of vessels only transmit restricted ROT values representing a turning indication. 27·49% of all evaluated vessels never transmit ROT values. These results show that only a small percentage of all vessels have a separate ROT indicator and that in most cases ROT is derived from the same sensor which is used to determine HDG. This explains the similar values for ROT and HDG in our study as shown in Figure 1 and Table 7. The absence of separate ROT sensors represents a problem, since less information for vessel prediction and collision avoidance is available. More and more detailed information about ROT would allow a better prediction of a vessel's movement as shown in Section 5.4. 17 (0·28%) of vessels have been identified that solely send position reports that only contain not available values for all data fields. Therefore, the only useful information within these position reports is the MMSI number. 447 (7·40%) of vessels transmit position reports that contain only actual measurements and never "not available" values. Therefore, the majority (5,578 or 92·32%) of vessels occasionally transmit not available values for at least one dynamic data field in addition to actual measurements. Collision avoidance algorithms should be able to handle such cases.

5.3.    *Reporting intervals evaluation.*    Figure 4 shows detailed results of evaluating reporting intervals $\Delta t$ for the dynamic data. $\Delta t$ is measured between two sequenced position reports of a vessel rounded to seconds. Thus, each bar corresponds to a width of one second. In total, 68,225,318 differential times have been evaluated. All $\Delta t$ greater than 1,800 s are not taken into account, since in this case it is likely that a vessel left the coverage of the AIS receivers and returned at a later point. According to Table 4, measured differential times are expected to be within the interval (2 s, 10 s), which is shown in the figure by the two red lines, or around 180 s in cases of non-moving vessels. Figure 4 shows that a high percentage of the measured differential times differ from these expected values.

It is apparent that the largest peak occurs at $\Delta t = 10$ s followed periodically by decreasing peaks at multiples of 10 s. At $\Delta t = 180$ s another larger peak occurs, which corresponds to the technical characteristics for AIS systems since a reporting interval of 180 s is mandated for vessels which are moored. Approximately 56% of all $\Delta t$ are located within the interval (2 s, 10 s), approximately 1·74% of all $\Delta t$ have a value of 180 s, and approximately 42·3% of all $\Delta t$ do not comply with the technical characteristics described in ITU (2010). The clusters around multiples of 10 s repeat while the number of position reports per $\Delta t$ decreases with increasing $\Delta t$ until the threshold of
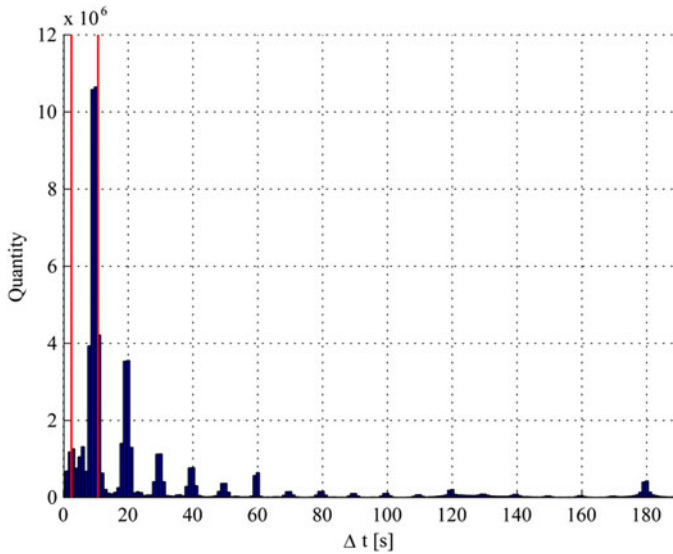
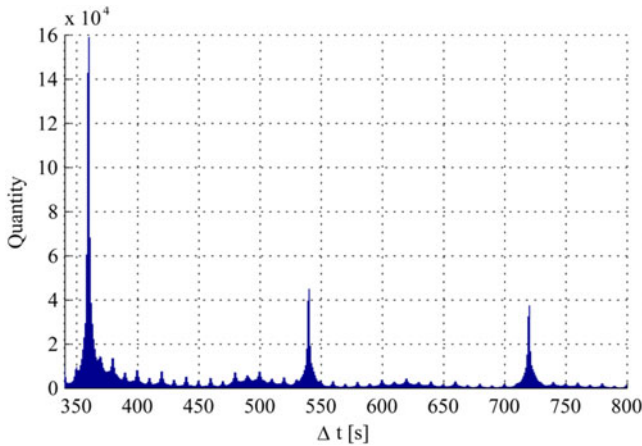Figure 4. Distribution of the differential times for dynamic data.



Figure 5. Periodically decreasing peaks.

1,800 s is reached, see Figure 5. Note that Figures 4 and 5 use different scales to show the quantities.

It is clear that in addition to the smaller peaks that occur every multiple of 10 s additional peaks occur at multiples of 180 s. In Figure 5, these peaks are visible at $\Delta t = 360$ s, $\Delta t = 540$ s, and $\Delta t = 720$ s. This matches to the reporting interval for moored vessels. Concerning the periodically repeating decreasing peaks, it seems that when an AIS message cannot be sent within the specified reporting interval, perhaps due to lack of time slots, the message is skipped or messages are simply lost during VHF transmission. The distribution of $\Delta t$ indicates that the recommended reporting intervals are not complied with in most of these cases.
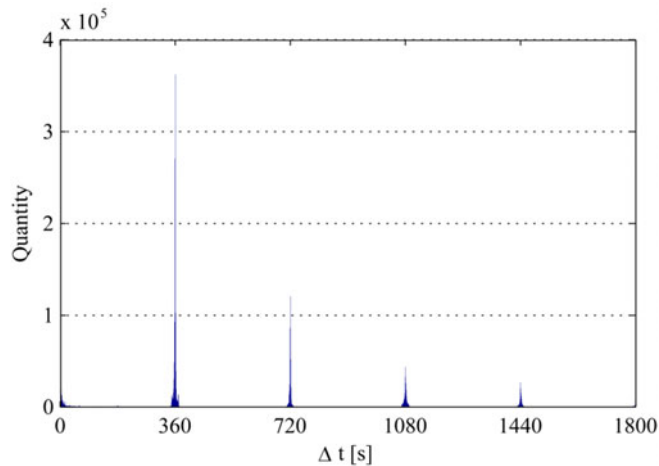
Figure 6. Distribution of the differential times for static data.

Evaluating the reporting intervals of the static and voyage-related data shows similar results. In total 2,962,714 differential times have been calculated and evaluated for message type five. $\Delta t$ is rounded to full seconds. Since the reporting interval for static data is higher, the threshold has been increased to 18,000 s. Figure 4 shows an excerpt of the data.

Figure 6 shows several peaks, where the largest peak occurs at $\Delta t = 360$ s, which complies with the reporting interval for static data. Broadcasting the static data in addition to the fix-reporting interval of 360 s should only occur when a data field of the static data, such as the current destination, has changed. This is a possible explanation for the number of differential times that represent shorter reporting intervals between one second and 360 s. In total approximately 51·32% of all $\Delta t$ lie within the interval (1 s, 380 s), i.e., approx. 48·68% are located outside of this interval. The remaining peaks represent multiples of 360 s that periodically repeat until the threshold of 18,000 s is reached. It should be noted that the chosen thresholds of 1,800 and 18,000 s are already very high. Current systems on board generally remove AIS targets after much less time when no update is received since displaying and predicting outdated data can result in confusion and wrong decisions by crew members. Even though over 50% of all differential times concerning static data match to the reporting intervals, our study shows that it is possible that several hours can pass between two static and voyage-related data reports. Consequently, by exceeding the reporting interval situations can occur where vessels pass without broadcasting their static data even once. This leads to problems for manual collision avoidance as well as for collision avoidance algorithms.

The data for the vessel displayed in Figure 7 are an excerpt of the whole vessel voyage data that has been recorded in an area covered by several AIS receivers.

Figure 7 shows the differential times over the sequential ordered position reports of the vessel track excerpt. This figure is a representative one, since all evaluated vessel tracks show the same effects, such as changing the vessel state from moving to moored or not keeping reporting intervals. Between position report index 1,500 and 1,813, the vessel is sending position reports with reporting intervals of up to 80 s. Between index 1,814 and 2,266, the vessel is moored or at anchor. Therefore the reporting interval of
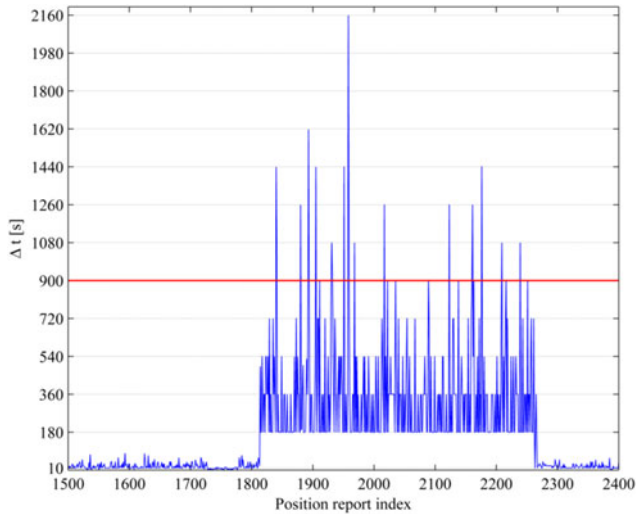
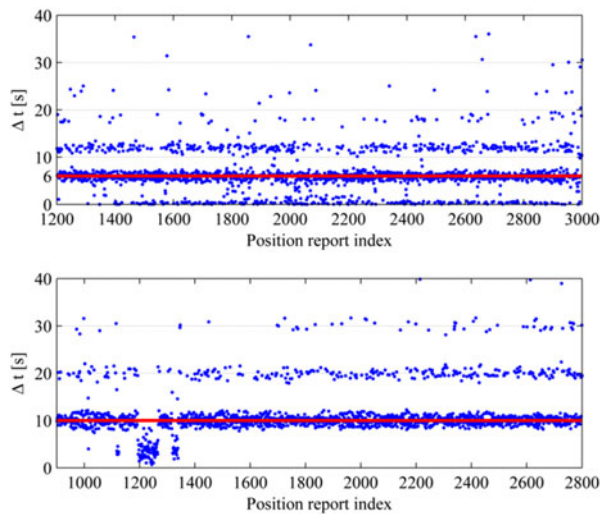Figure 7. Differential times between position reports representing a vessel track excerpt.



Figure 8. Position report differential time evaluation of moving vessels.

the position reports should be equal to 180 s. It can be seen that $\Delta t$ indeed changes and never falls below 180 s, whereas in 22 cases more than 900 s (indicated by the red line) pass until a new position report is sent. Almost all of these violations of the reporting interval for anchored/moored vessels are multiples of 180 s. For example, the biggest peaks within these index numbers have a value of approximately $\Delta t = 2{,}160$ s, which is exactly 18 times the reporting interval of 180 s for moored vessels. Other tracks have been evaluated where the time between two position reports also occasionally drops under 180 s when the vessel is anchored or moored. Figure 8 shows a more detailed evaluation of the differential times when a vessel is moving based on track excerpts of two different vessels.

The vessel, whose track is shown in the upper plot, is traveling with a speed between 14 kn and 23 kn resulting in an expected report rate of 6 s when not turning. When turning the reporting rate should be 2 s. The plot shows that it drops in almost all cases below 2 s when the vessel turns. The plot also shows that the reporting rate is never constant. For example when the vessel is not turning, the reporting rate should be constantly 6 s, but actually varies between 4 and 8 s. Furthermore, the plot shows the same effect as we observed in Figure 7. In many cases, multiples of the expected report rate pass between two AIS messages. Point clusters that are visible at 6, 12, and 18 s show this effect, whereas the number of points decreases with increasing $\Delta t$. The lower plot in Figure 8 shows another vessel that is at a velocity less than 14 kn. As shown in Table 5, the report rate for this velocity should be 10 s. At approximately position report index 1,200, the vessel is turning twice, resulting in a change of the expected report rate which is set to 3 1/3 s. After turning, the vessel continues driving with a velocity smaller than 14 kn. As in the upper plot, there are also several point clusters visible which represent multiples of the report rate of 10 s. All presented figures show that several multiples of the mandatory report rates can pass between two consecutive AIS messages. This is an indicator for message loss, most likely caused by the transportation medium VHF. As a result, the vessel can disappear from AIS systems due to timeouts. If the vessel is also out of range for radar systems of other nearby vessels, perhaps due to occlusions, it could become totally invisible. The following section shows the impacts for vessel prediction and therefore for collision avoidance algorithms.

5.4. *Vessel movement prediction based on AIS data.* Collision avoidance algorithms allow detecting whether two or more entities, in the current case vessels, are on a collision course. For that reason, the future course of a vessel that is unknown needs to be estimated based on available information derived from a vessel state. Making assumptions about the unknown future course of a vessel is called prediction. A reliable prediction of vessel movements is the basis of collision avoidance algorithms. The prediction itself uses a motion model that needs to be defined based on the application context. Current filters that use motion models to estimate system states are the Kalman Filter for linear systems and the Extended Kalman and Particle Filters for non-linear systems. Within this section, a basic motion model based on AIS data is developed. In addition, it is shown that a motion model only based on AIS data lacks in many cases reliability, since the reporting intervals which are evaluated in Section 5.3 are not kept in many cases and since useful information such as ROT is often not available or restricted in availability as shown in Section 5.2.

Since the above filter types perform a state estimation, a formal representation of a valid system state needs to be defined. This procedure is called state definition. The state definition is a mathematical representation of various object properties. Examples for such object properties are size, position, velocity or the shape of an object. These properties are summarised to a vector $x_t$ which is valid for a given time step $t$. Since the dynamic AIS data includes information such as SOG or COG one possibility for a state definition is shown in Equation (1).

$$x_t = \begin{bmatrix} lat \\ lon \\ COG \\ ROT \\ SOG \end{bmatrix} \tag{1}$$

The units are defined according to Table 4. Since the impacts of the reporting intervals and lost messages, respectively, on vessel movement prediction are evaluated, the defined system state in Equation (1) does not make use of the HDG and AIS accuracy information. Therefore, a basic model without acceleration is used. To advance respectively to integrate a given $x_{t-1}$ to $x_t$ an appropriate motion model $f_{Motion}$ has to be defined. $x_{t-1}$ as well as the past time $t$ are arguments for $f_{Motion}$ as shown in Equation (2).

$$x_t = f_{Motion}(x_{t-1}, \ t) \tag{2}$$

To act on the assumption that a vessel travels only small distances on a plane a basic vessel motion model for the defined system state Equation (1) can be described as shown in Equation (3). Equation (3) allows us to advance $x_{t-1}$ to $x_t$.

$$x_t = f_{Motion}(x_{t-1}) = \begin{bmatrix} lat_{x_{t-1}} + \Delta t \ \dfrac{1}{60} \ \cos\left(COG_{x_{t-1}}\right) \cdot SOG_{x_{t-1}} \\[2mm] lon_{x_{t-1}} + \Delta t \ \dfrac{1}{60} \ \dfrac{\sin\left(COG_{x_{t-1}}\right) \cdot SOG_{x_{t-1}}}{\cos\left(lat_{x_{t-1}}\right)} \\[2mm] COG_{x_{t-1}} + \Delta t \cdot \dfrac{ROT_{x_{t-1}}}{60} \\[2mm] ROT_{x_{t-1}} \\[2mm] SOG_{x_{t-1}} \end{bmatrix} \tag{3}$$

$\Delta t$ has to be specified with $[h]$ as a unit since SOG is given in knots and 1 kn equals $1\frac{nm}{h}$. Even though a movement on a planar surface is assumed, one has to consider that arc minutes only stay constant for longitude values. For that reason the calculated movement distance for the longitude value has to be divided by a value based on the current latitude. The factor $\frac{1}{60}$ is used to convert the products from [kn] to [°] since 1° of latitude equals approximately 60 nm. The system state components ROT and SOG are not modified. Therefore only the geographical coordinates as well as COG change continuously over time whereas the ROT and SOG values stay constant.

Equation (3) allows predicting and comparing vessel movements by comparing recorded AIS data with integrated trajectories. Since an initial system state is needed, the first AIS message of a vessel providing dynamic data can be used as $x_0$. $x_0$ is integrated by using time steps of $\Delta t = 1$ s. Integrating $x_0$ is performed until a new AIS position report is received. At this point the current integrated position is compared with the received position. Furthermore all dynamic data from the received message is used as the new system state. Figure 9 shows four possible impacts that might occur if the reporting intervals are exceeded as shown in Section 5.3.

The blue dots which are connected with blue lines represent recorded position reports of a real vessel that provides all dynamic information. Each red dot represents the integrated system state after one second. All plots show an excerpt of a vessel track. The upper left plot shows a vessel driving from east to west. 40 s pass between the two shown position reports. Therefore the estimated vessel trajectory slowly drifts off. Nevertheless the final integrated state has a distance error of approximately 6 m to the received position report. In contrast, the upper right plot shows one of the problems that occur when a system state is not continuously updated or an AIS message gets lost since the vessel has changed its SOG by accelerating. This change has not been
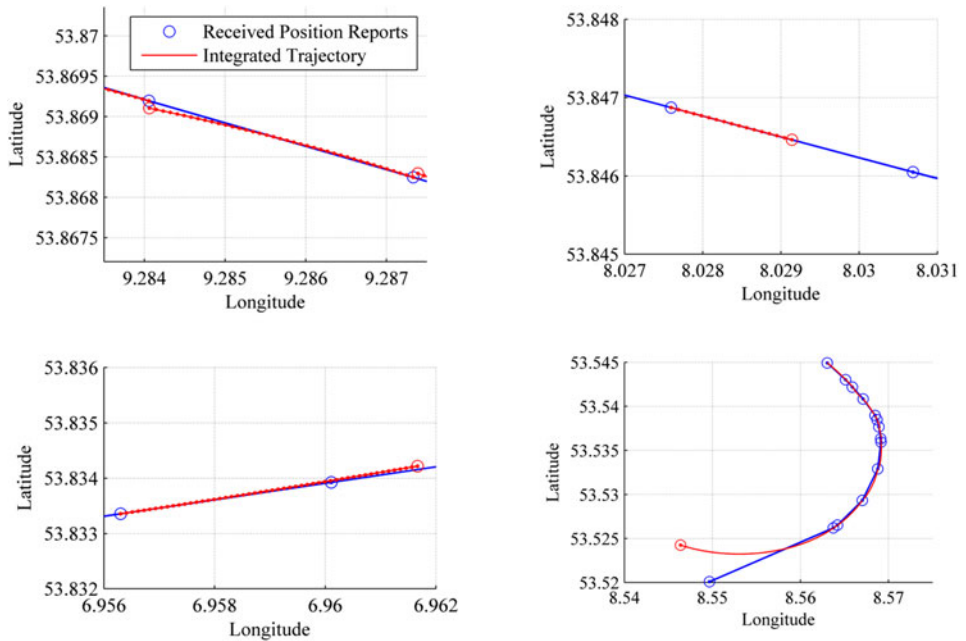
Figure 9. Integrating received dynamic data.

transmitted by AIS, resulting in a distance of approximately 111 m between the final integrated state and the received position report. The lower left plot shows a similar problem. The position report that has been used as $x_t$ is integrated over a time period of 51 s. Again the vessel changed its SOG, this time by slowing down. Since the reporting intervals have not been kept, the predicted vessel state is located ahead of the geographical position provided by the position report, resulting in a distance error of around 107 m. The last plot to the lower right shows the impact of varying reporting intervals in the case of a turning vessel. The scale differs from the former plots to illustrate the manoeuvre by showing the past trajectory that has been integrated. The vessel is driving southwards. The last six measured $\Delta t$ equal to 10 s, 81 s, 89 s, 77 s, 10 s, and 264 s, respectively. Since 264 s pass between the last two position reports, integrating received dynamic data results in a distance of approximately 0·5 km between the measured position report values and the integrated values indicated by the red circle. Even though it seems that the vessel straightened after performing the turning manoeuvre the last AIS position report which has been integrated contained a ROT value of 17·86° per minute. Furthermore one has to consider that a vessel with a separate ROT indicator will almost never deliver exactly 0° because of small course corrections. Varying reporting intervals in combination with small ROT values can also lead to a prediction that diverges. "Not available" values within AIS messages have the same impact as lost messages. Therefore more sophisticated motion models are needed to compensate for an AIS message loss or for the effects shown in Figure 9, respectively. However, varying reporting intervals are just one problem for predicting vessel movements. Another problem is the lack of dynamic information, e.g. ROT, as shown in Table 7. Figure 10 shows three examples to indicate the importance of such information.
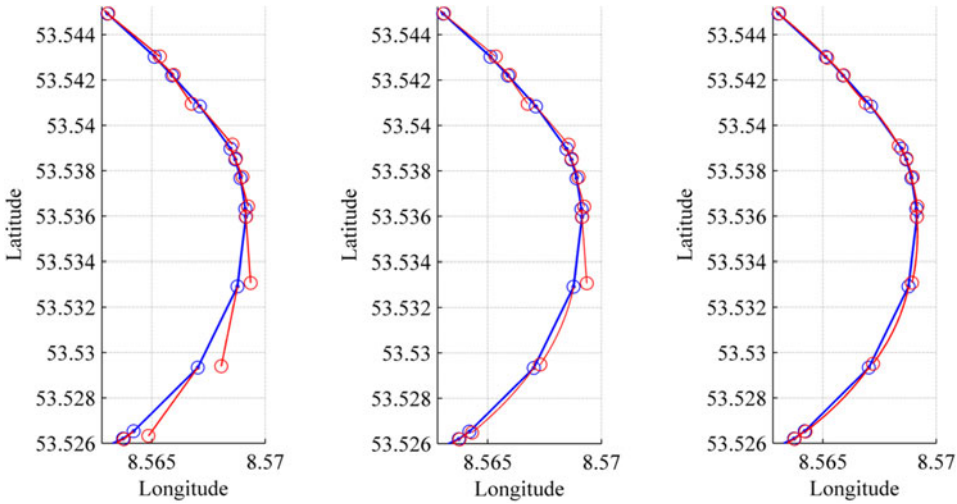
Figure 10.  Impact of absent ROT information.

Again a vessel is chosen which transmits actual measurements for all dynamic data fields such as ROT or SOG. Therefore, the vessel positions displayed as blue points in Figure 10 are integrated based on received dynamic data for each received position report. The integrated trajectory parts are marked by red lines. For two of the three plots the ROT information has been modified. The left plot shows the impact of missing ROT information, since ROT has been set to zero in all cases. It can be seen that the previous path of the vessel is simply continued by integrating the remaining information such as COG. The plot to the middle shows a case where restricted ROT information is available. Restricted means that a vessel transmits only three values for ROT, turning 0° or turning more than $\pm \frac{5°}{30\,s}$. In the present case a value of $ROT = \pm \frac{10°}{min}$ is used if the real information was bigger/smaller than or equal to 10°. It is shown that the integrated trajectory based on restricted ROT data is smoother if directly compared to the left plot. However, the right plot using original ROT values provided by the vessel shows the importance of the ROT information. Figure 10 shows that a prediction close to the actual vessel position is only possible if detailed ROT information is available. In regard to our evaluation results, approx. 27% of the evaluated vessels never transmit ROT values. Approx. 63% of all evaluated vessels always transmit ROT values. However, only 829 (18·92%) of these vessels actually transmit detailed degree turning information, which leads to the results shown in the right plot. The remaining 3,552 (81·08%) vessels provide restricted turning information resulting in less accurate predictions as shown in the plot to the middle.

Although for crew members' decisions, variations concerning the reporting intervals are not crucial, this section has shown that vessel prediction algorithms and collision avoidance algorithms rely on continuous measurements. If no measurements are available, the prediction step is repeatedly performed. Consequently, the uncertainty increases over time as indicated in Figures 9 and 10. One possible explanation for lacking compliance to recommended reporting intervals is that AIS base stations are able to control the reporting time of mobile AIS stations. This is done by sending

message type 23 AIS messages. If mobile AIS stations like vessels receive such a message, the operating mode is switched from autonomous mode to assigned mode. In assigned mode, mobile stations can be forced to be quiet for up to 15 minutes (ITU, 2010). However, within our study only one message of type 23 was received indicating that it is unlikely that variations are caused by switching mobile AIS stations to assigned mode. It is expected that AIS systems are implemented according to the technical recommendations, since these recommendations are national law for states where AIS systems get approved. Therefore AIS systems are tested according to the IMO specifications before they get an approval. It is more likely that too many vessels can cause overload to the AIS network. Redoutey et al. (2008) mention that a high density of vessels can cause overload that results in losses of AIS messages. In IMO (2006), an exemplary calculation about the theoretical capacity of AIS systems according to the Self-Organised Time Division Multiple Access (SOTDMA) method is performed. It is mentioned that when both VHF channels are used 450 AIS systems sending data in one area represent the maximum count. This number is based on ideal assumptions meaning that each vessel is able to reserve time slots whenever needed according to the reporting intervals. Our data set shows that the amount of Class A and Class B AIS systems within the observed area can reach this calculated value. Another reason could be a message loss on the VHF medium caused by weather conditions or a wrongly placed antenna on the vessel. In summary, we have shown that because of exceeding reporting intervals the AIS system is in most cases insufficient as a standalone data source for continuous vessel tracking and collision avoidance in real time applications.

6. CONCLUSIONS AND FUTURE WORK. The availability of both static and dynamic AIS data and the evaluation of the reporting intervals with respect to vessel movement prediction have been performed. Concerning the data fields, it is shown that for almost all dynamic data fields a high availability is achieved except for ROT and HDG. Since ROT and HDG provide useful information when implementing motion models or collision avoidance algorithms, the availability should be increased by connecting appropriate sensors to the AIS systems. Furthermore, concerning the static data, we have shown that almost every vessel transmits its dimensions. Concerning the reporting intervals, it is mentioned that "changes in heading and course are […] immediately apparent." to other vessels (IMO, 2003. P. 9). Our study shows that this is only partially true since AIS reporting intervals mostly do not comply with the reporting intervals described in ITU (2010). Several example vessels have been chosen to show this in detail. Compared to other studies such as Felski and Jaskólski (2012a) a significantly lower availability of AIS data has been observed. We assume that network overload or data loss because of the VHF medium are two possible reasons. Furthermore, it has been shown that dynamic information such as ROT and SOG is not reliable if reporting intervals are too large, since a vessel's state changes continuously by increasing or decreasing speed, performing course corrections, etc. An example vessel has been chosen to show that a reliable prediction is not possible when the reporting intervals are too large, since the uncertainty of an estimated vessel state increases over time. Therefore, implementing motion models or collision avoidance algorithms that are only based on AIS data are hard to implement because current prediction algorithms rely on fixed interval times with short duration.

As a result, this study shows that AIS data needs to be integrated with further data to allow for a reliable state estimation of vessels in matters of collision avoidance algorithms. Our study also shows that even more research concerning AIS data is required to identify the actual reasons for a message loss and the varying reporting intervals, respectively. A loss on the transportation medium represented by VHF is assumed, since AIS systems shall transmit AIS data even in case of a possible network overload. However, further possible factors such as the antenna placement on vessels and land should be evaluated.

## ACKNOWLEDGEMENT

## REFERENCES

Felski, A. and Jaskólski, K. (2013). The Integrity of Information Received by Means of AIS During Anti-collision Manoeuvring. *TransNav the International Journal on Marine Navigation and Safety of Sea Transportation*, **7**, 95–100.

Felski, A. and Jaskólski, K. (2012a). Analysis of AIS availability. *European Journal of Navigation*, **10**(1), 39–43.

Felski, A. and Jaskólski, K, (2012b). Information unfitness as a factor constraining Automatic Identification System (AIS) application to anti-collision manoeuvring. *POLISH MARITIME RESEARCH*, **19**, 60–64.

Felski, A. and Jaskólski, K. (2012c). Information Unfitness of AIS. *Annual Of Navigation*, **19**(1), 17–24.

Harati-Mokhtari, A., Wall, A., Brooks, P. and Wang, J. (2007). Automatic Identification System (AIS): Data Reliability and Human Error Implications. *Journal of Navigation*, **60**(3), 373–389.

IMO. (2006). *Automatic Identification Systems Model course (AIS) 1·34*. London: International Maritime Organization.

IMO. (2003). *Guidelines for the Installation of a Shipborne Automatic Identification System (AIS)* [Homepage Of International Maritime Organization], [Online]. Available: http://www.imo.org/ourwork/safety/navigation/documents/227.pdf. Accessed 15 May 2013.

ITU. (2010). *Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band* [Homepage of International Telecommunication Union], [Online]. Available: http://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-4-201004-I!!PDF-E.pdf. Accessed 12 May 2013.

Redoutey, M., Scotti, E., Jensen, C., Ray, C., and Claramunt, C. (2008). Efficient Vessel Tracking with Accuracy Guarantees. *Lecture Notes in Computer Science*, **5373**, 140–151.