PAWEŁ BANYŚ, THORALF NOACK, STEFAN GEWIES
German Aerospace Center (DLR), Institute of Communications and Navigation (IKN)

ASSESSMENT OF AIS VESSEL POSITION REPORT UNDER THE ASPECT OF DATA RELIABILITY

ABSTRACT

Since its introduction the Automatic Identification System (AIS) has played an important part in improving safety at sea, making bridge watchkeeping duties more comfortable and enhancing vessel traffic management ashore. However the analysis of a AIS data set describing the vessel traffic of the Baltic Sea came to conclusion, that specific parameters with relevance to navigation seemed to be defective or implausible. Essentially, it concerned the true heading (THDG) and the rate of turn (ROT) parameters. With the paper we are trying to clarify, which parameters of the AIS position report and to what extent, are affected. The detailed data analysis gives answers on how reliable the AIS data in different traffic areas is.

INTRODUCTION

The monitoring and assessment of vessel traffic is an important element of safe, secure and efficient shipping and the protection of environment. The collision and grounding avoidance at sea requires a reliable and comprehensive picture of the maritime traffic situation to provide an error-free decision making for the seafarers. After the implementation of AIS in 2004 an important step was done to deploy an improved feature as a basis for decision support directly on the bridge or for the VTS operators on shore. Besides radar as the primary device supporting the collision avoidance, AIS has gained more significance within the scope of the automatic identification of conflict situations and the generation of warnings and alerts.

Like almost every technology, AIS is subject to specific restrictions and limitations, too. Because of the dual character of AIS data (disengageable, dependent on the human initiated processes) and the dependency on other onboard devices (for example the GPS receiver) there is still a margin for errors in both the static as well as the dynamic data. Insofar, a possibility cannot be ruled out, that AIS data is wrong or not meaningful during important maneuvers of a vessel.

The first analysis of a comprehensive two month AIS data set, done by our Institute, describing the vessel traffic of the whole Baltic Sea as of January and February 2010, came to conclusion, that specific parameters with relevance to navigation seemed to be defective or implausible. Essentially, it concerned the true heading (THDG) and the rate of turn (ROT) parameters. With the paper we are trying to clarify, which parameters of the AIS position report (message types 1, 2 and 3), and to what extent, are affected. Using the areas of open sea and the Port of Rostock, we are showing similarities and differences in AIS data patterns between the areas of steady traffic and the areas of frequent tight maneuvers. The detailed data analysis gives answers on how reliable the AIS data in different traffic areas is. The results will serve as a basis for the development of a maritime traffic situation assessment facility, where AIS, ARPA (radar) and specific PNT data shall be fused together to create a reliable traffic situation image.

BASICS

Since its introduction in 2004, the Automatic Identification System (AIS) has played an important part in improving safety at sea, making bridge watchkeeping duties more comfortable and enhancing vessel traffic management ashore. Its usage worldwide is widespread. As the Safety of Life at Sea Conventions (SOLAS) state, all vessels of 300 gross tonnage and upwards engaged on international voyages, cargo vessels of 500 gross tonnage and upwards not engaged on international voyages and both passenger vessels and vessels carrying dangerous cargo irrespective of size shall be fitted with the AIS transponder [SOLAS]. According to the Lloyd's List Intelligence, which is running the world's largest land and satellite based AIS monitoring network, there are currently about 72 000 vessels worldwide equipped with active AIS transponders [LLI].

The main function of AIS is to broadcast dynamic and static navigational data of a vessel to other vessels located within a range of about 20 NM. The data are transmitted on two marine VHF channels reserved for the system. Every transmission contains the information about the time of its next broadcast, too. This way all the vessels within a certain area can organize the AIS radio broadcasts themselves and share vital navigational data with one another.

The AIS transponder is typically connected to the bridge equipment. It is gathering data from a Global Navigation Satellite System (GNSS) receiver, which is in most cases a (differential) Global Positioning Receiver (GPS) receiver, a heading indicator device — be it gyrocompass, fluxgate or GNSS compass — and a rate of turn indicator. The input data from the sensors are split into the AIS variables,

as specified in International Telecommunication Union (ITU) Recommendation M.1371-3 [ITU-R M.1371-3], and then converted into a bit stream of radio packets, which conforms to the National Marine Electronics Association (NMEA) encapsulation standard [NMEA 0183]. If any data from the bridge devices, which AIS transponder depends on, becomes unavailable prior to the upcoming transmission, it has to be internally replaced with a special default value, which explicitly marks that sensor data unknown. These special values documented in the ITU Recommendation M.1371-3 are listed in table 1.

Table 1. Special AIS values which indicate unknown AIS variables [ITU-R M.1371-3]

AIS variable		Raw indicator of an unknown value
Speed over ground	(SOG)	1023
True heading	(THDG)	511
Course over ground	(COG)	3600
Geographic longitude	(LON)	108600000
Geographic latitude	(LAT)	54600000
Rate of turn	(ROT)	−128
Navigational status	(NAV)	15

The occurrence of such values is of significant importance, if the data is not only used for monitoring aspects, but for machine-aided traffic situation assessment as well.

The point on the AIS data flow path, at which the unknown values may be stored into the AIS variables, if decided so by the AIS transponder software, is shown on figure 1.
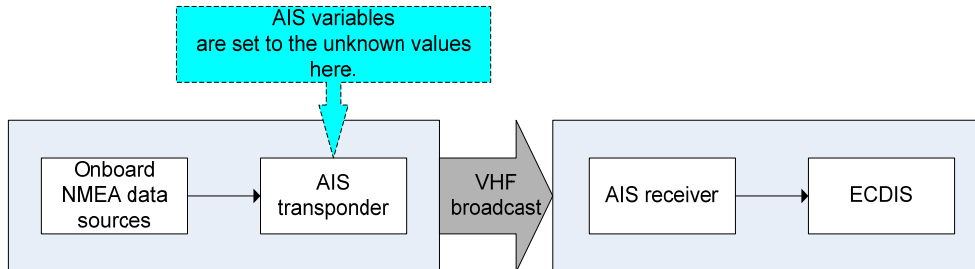

Fig. 1. Schematic diagram of the AIS data flow between two vessels [own study]

In this paper the assessment of data is only focused on the analysis of those unknown values. The closer examination of time series obtained from the AIS dataset or technical reasons, why the unknown values are generated, will be the subject of future work.

DATASETS AND INVESTIGATION AREAS

The AIS data used in this analysis was recorded in September 2011 and covers the vessel traffic in the Baltic Sea nearby the north coast of Germany. All AIS data is provided by the German Federal Waterway Authority as a query from the Helsinki Commission (HELCOM) AIS database server.

Two mutually exclusive zones of interest were defined (fig. 2): 'harbour', to capture moving and made fast vessels inside and nearby a harbour, and 'sea', to capture vessels underway within a specific traffic separation scheme. The aim of such classification was to pinpoint and highlight the differences between the areas of steady traffic and the areas of frequent tight maneuvers.

The complete one month AIS dataset includes Class A as well as Class B data messages. Every AIS transmission of message type 1, 2, 3 (dynamic position report Class A), 5 (static and voyage related data Class A), 18 (standard Class B equipment position report) and 19 (extended Class B equipment position report) was counted and placed into corresponding geographic category. Additionally, a number of unique transmitters identified by their Maritime Mobile Service Identity (MMSI) were extracted, too (table 2).

Table 2. Number of AIS datasets and transmitters in both zones of interest [own study]

Area	Class A messages			Class B messages		
	Dynamic data (ID 1,2,3)	Static data (ID 5)	Transmitters	Dynamic data (ID 18)	Static data (ID 19)	Transmitters
Harbour	6,158,313	380,578	287	80,915	0	62
Sea	6,564,123	158,257	1,700	17,218	0	49

The complete database comprises about 13 million datasets for Class A messages and 98 thousand datasets for Class B messages. The percentage of Class B messages in relation to Class A is below 1 percent and therefore of marginal relevance. Comparing the two different investigation areas, the harbour to sea ratio is around 1 to 1 for Class A receivers and around 5 to 1 for Class B receivers.

In contrast to the volume of AIS messages, nearly 6 times more different AIS Class A transmitters could be found in the harbour area, as compared to the sea area. The number of Class B transmitters is 1.2 times higher in harbour than at sea. The total number of transmitters has minor importance in case of Class B equipped vessels operating at sea whereas in harbour AIS Class B has a rate of 18 % of all equipped vessels.

Figure 2 shows the two investigation zones marked by red rectangles. The harbour area (left) covers the Port of Rostock and its approaches. The sea area, as drawn on the right side of figure 2, is located in the Baltic Sea east of Bornholm between Trelleborg and Arkona.
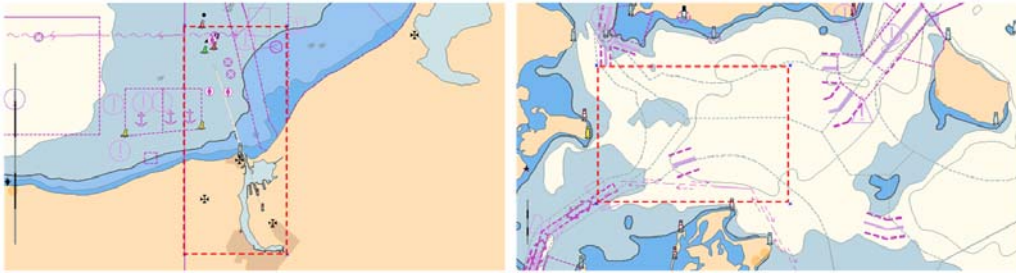
Fig. 2. Investigation areas 'harbour' (left) and 'sea' (right) [own study]

The geographic boundaries of both areas are shown in the table 3.

Table 3. Coordinates of the research areas in WGS-84 [own study]

Area	Upper left corner	Lower right corner
Harbour	54°18'00"N , 012°00'00"E	54°05'24"N , 012°09'36"E
Sea	55°12'00"N , 012°36'00"E	54°42'00"N , 013°48'00"E

APPROACH OF THE DATA ANALYSIS

The whole AIS dataset was converted into plaintext raw variables, in order to avoid any floating point ambiguities, which could easily mask important patterns of unknown values. Based on the information shown in table 1, seven AIS variables were chosen for the analysis. Each variable was counted, when an AIS message with unknown value was detected. The recurrence frequencies of the AIS variables set to the unknown value are shown in the table 4.

In the second step the database was scanned for misleading content generated by improper use of AIS transponder equipment. In each area at least one transponder had to be declared deceptive, because it was sending misleading AIS messages 1, 2 and 3, that they were not authorized to broadcast. The exclusion of the messages from those wrong transponders has refined results shown in table 5. Although the significance of the influence of AIS transponders transmitting false message ID can be neglected, the messages of the affected AIS transponders were removed from the database to allow a focus on real AIS transponders.

In the third step all combinations of the variables having the unknown values were counted together with transmitters responsible for each combination. The transmitters were reckoned based on their MMSI numbers (tables 6 and 7).

To describe the dependency of significant variables on the movement of the vessels, the frequency of occurrence of these variables in relation to the SOG was analyzed in the final step (figures 3 to 6).

RESULTS

Overall results

The first part of AIS message analysis is summarised in table 4. Here the frequencies of unknown values of ROT, THDG, NAV, COG, SOG, LON, and LAT are listed for harbour and sea area, as well as both areas together, for Class A and B transponders separately. Although the NAV variable reaches 7.6% of unknown values in all datasets of Class A data, it will not be included, because its value is set manually by the crew. This study is focused on the automatically generated values of ROT, THDG, COG, SOG, LON, and LAT, which may point out systematic problems of vessel equipment.

Table 4. Frequency of occurrence of unknown values of specific AIS variables
(n/a — no data available) [own study]

AIS variable	Frequency of occurrence in Class A messages [%]			Frequency of occurrence in Class B messages [%]		
	Harbour area	Sea area	Harbour and sea area	Harbour area	Sea area	Harbour and sea area
ROT	30.2	8.8	19.2	n/a	n/a	n/a
THDG	30.1	7.2	18.3	100	100	100
NAV	7.6	6.1	6.8	n/a	n/a	n/a
COG	2.2	< 0.1	1.1	< 0.1	< 0.1	< 0.1
SOG	2.2	< 0.1	1.1	< 0.1	< 0.1	< 0.1
LON	< 0.1	< 0.1	< 0.1	0	0	0
LAT	< 0.1	< 0.1	< 0.1	0	0	0

In case of the Class B messages, it can be noticed, that specific variables are completely unavailable. The true heading transmitted by this equipment is entirely unknown. Although the data output format supports heading information, Class B units are generally not connected to a heading indicating device like compass. That is why the THDG is seldom transmitted. The occurrence frequency of unknown values of SOG and COG is smaller than 1‰. The geographic position data is complete. Because Class B transponders are designed for vessels outside the SOLAS regulations, a deeper analysis of the Class B messages is beyond the scope of this paper.

The removal of the two misleading AIS transponders has a different effect on the frequency of unknown values in both investigation areas (table 5). The frequencies in harbour are affected a little only, whereas the frequency of unknown values at sea drops by about 50%. For the following study it is assumed, that table 5 is the best image of the average of AIS messages in both areas.

Table 5. Frequency of occurrence of unknown values of specific AIS variables after exclusion of improper AIS equipment [own study]

AIS variable	Frequency of occurrence in Class A messages [%]		
	Harbour area	Sea area	Harbour and Sea area
ROT	28.4	5.2	16.5
THDG	28.3	3.6	15.6
NAV	7.8	2.4	5.0
SOG	2.3	< 0.1	1.1
COG	2.3	< 0.1	1.1
LON	< 0.1	< 0.1	< 0.1
LAT	< 0.1	< 0.1	< 0.1

Discrete harbour area (Class A messages)

In harbour area around 65% of the messages are unaffected. This means, that around 35% of all AIS messages have one or more unknown values of the seven analysed variables. The highest occurrence of unknown values (see table 5) could be observed for ROT and THDG, with around 28% for each of them. In case of parameters like navigation status (around 10%), SOG, and COG (about 2% each) the occurrence of unknown values is lower but still of significance. For the geographic position values, which are used to plot the vessel movements on ECDIS, the occurrence rate is less than 1‰.

Table 6 lists different combinations of unknown values, having significant occurrence within the AIS dataset (second column). In the third and fourth column the number of affected transmitters, as well as their proportion to the total count of transmitters in harbour, is presented. As the NAV value is ignored, table 6 shows the combinations of ROT and THDG at 28.4 %, SOG and COG at 2.2 %, and only ROT at 1‰. So there is always a tight union of unknown values of ROT and THDG, as well as SOG and COG, but seldom (<1 ‰) of cross combinations of these four values.

Table 6. Significant occurrence (frequency > 1‰) of unknown values of AIS variables in harbour (Class A) in combinations of the variables [own study]

Combination of the variables	Frequency of occurrence [%]	Number of affected transmitters	Affected transmitters [%]
ROT+THDG	24.9	68	23
NAV	4.4	5	2
NAV+ROT+THDG	3.4	8	2
SOG+COG	2.2	24	8
ROT	0.1	11	4

Within a quarter of all Class A transmitters in harbour, a significant amount of vessels reveal problems in determining ROT and THDG or transmitting that

information via AIS. At 8 % a significantly smaller part of transmitters has trouble sending meaningful SOG and COG data.

An important question concerning the results in tables 5 and 6 is the influence of the speed of the vessel movement. Therefore the frequency of occurrence of unknown values of ROT and THDG in relation to SOG was analysed (fig. 3), as well as the number of affected transmitters (fig. 4). As the table 6 indicates, there are only slight differences between ROT and THDG and their dependency of unknown value frequency and number of affected vessels on SOG.
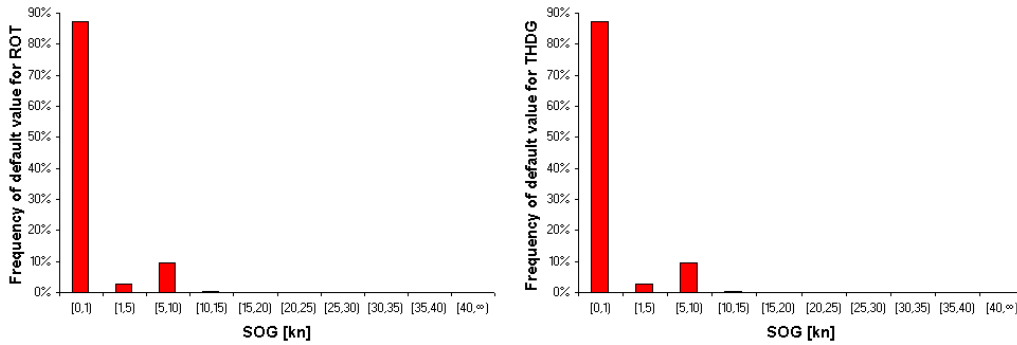
Fig. 3. Histogram of the distribution of unknown values of ROT (left) and THDG (right) in relation to discrete arrays of SOG for Class A messages in harbour area [own study]

Figure 3 shows, that the most part (around 88%) of the AIS messages with unknown values of ROT and THDG occurs, when the SOG is less than 1 kn. Twelve percent of the unknown values are set by vessels moving at the speed between 1 kn and 10 kn. Considering the overall results for ROT and THDG (28% of affected messages) as shown in table 5, about 3% of the data can be declared crucial for the assessment of traffic situation, because of the movement of the vessels.
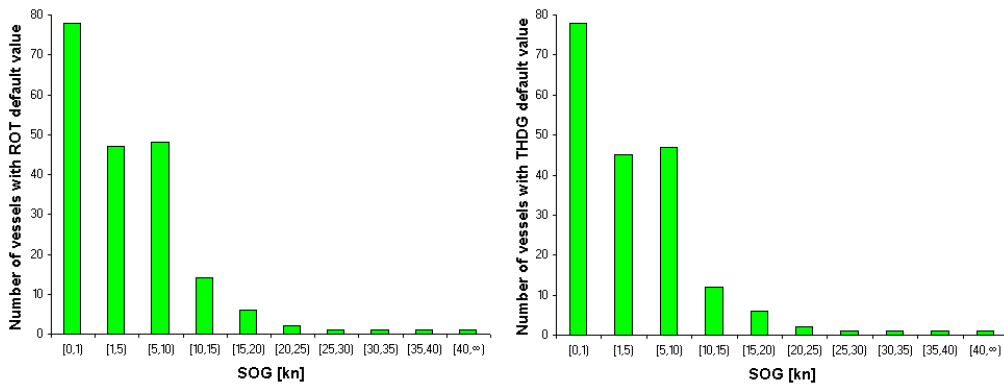

Fig. 4. Number of affected transmitters sending unknown values of ROT (left) and THDG (right) in relation to discrete arrays of SOG for Class A messages in harbour area [own study]

The amount of 45 to 48 vessels, a sixth of all vessels in harbour, contribute to the transmissions of unknown values of ROT and THDG in the two SOG intervals between 1 and 10 kn (fig. 4). Although the frequency of unknown values at higher SOG is nearly zero (fig. 3), a significant number of vessels experience this problem. For SOG lower than 1 kn, nearly 27% of all vessels contribute to the broadcasts of incomplete AIS messages. For interpretation of figure 4 one has to keep in mind that a single vessel can be counted multiple times in each SOG interval.

Discrete sea area (Class A Messages)

In sea area around 92% of the messages are unaffected. With other words around 8% of all messages have one or more unknown values out of the seven analyzed variables. The highest occurrence of unknown values could be observed for ROT and THDG at about 5% for ROT and 4% for THDG (table 5). The navigation status is affected by about 2%. For all other variables (SOG, COG and geographic position) the occurrence is smaller than 1‰.

Like table 6 for harbour, table 7 lists the significant frequencies of AIS messages, number of affected transmitters, and their ratio for different combinations of unknown values for the sea area. The highest occurrence of two or more unknown values at the same time (a single AIS message) takes place for the combination of ROT and THDG. This is followed by unknown NAV and ROT. THDG as well as the combination of unknown SOG and COG are seldom (0.1‰). Speaking of 2 to 4% of vessels, a significant number of them in sea area are affected by problems, that make the AIS transponders set ROT and THDG, only ROT or SOG and COG to the unknown value. This reveals some weakness of the onboard equipment used nowadays.

Table 7. Significant occurrence (frequency > 1‰) of unknown values of AIS variables at sea (Class A) in combinations of the variables [own study]

Combination of unknown variables	Frequency [%]	Number of affected transmitters	Affected transmitters [%]
ROT+THDG	3.4	69	4
NAV	2.4	4	< 1 ‰
ROT	1.7	39	2
THDG	0.1	7	< 1 ‰
SOG+COG	0.1	40	2

Figures 5 and 6 shows the frequency of unknown values of ROT and THDG as well as the number of affected vessels at various SOG intervals. At the speed lower than 1 kn only a small number of vessels (less than 1‰ of all) are responsible for 25% of unknown values of ROT and 36% of unknown values of THDG. The

most part of affected vessels with unknown values of ROT and THDG can be observed at typical speeds at sea: 5 kn to 15 kn. Four percent of all vessels sent unknown values of ROT at least once and 3% of all vessels did it in case of THDG. The frequencies of unknown values sum up for ROT and THDG at speeds lower than 1 kn and speeds in range between 5 kn to 15 kn, reaching over 90%.
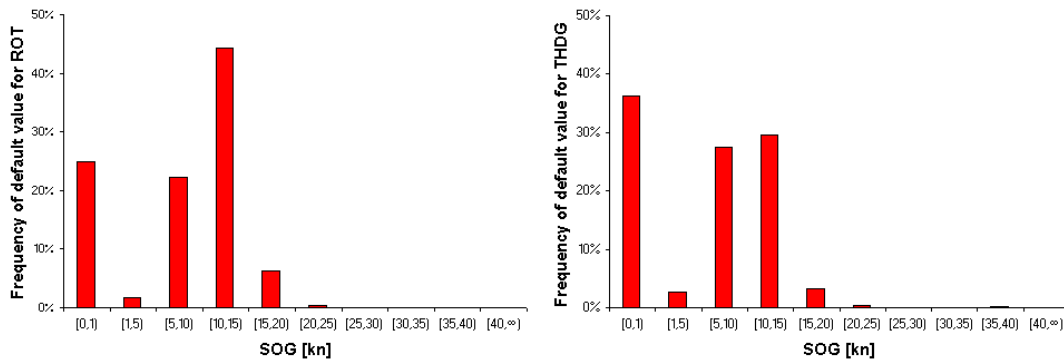

Fig. 5. Histogram of the distribution of unknown values of ROT (left) and THDG (right) in relation to discrete arrays of SOG for Class A messages in sea area [own study]
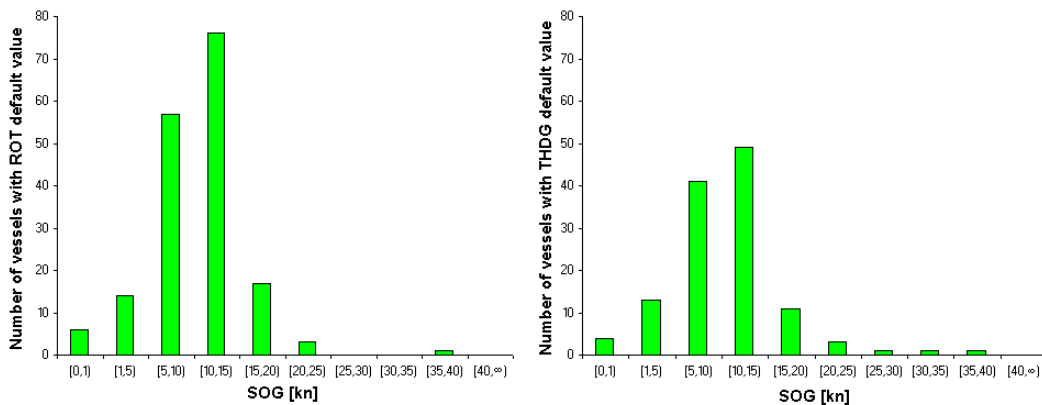

Fig. 6. Number of affected transmitters for unknown values for ROT (left) and THDG (right) in relation to discrete arrays of SOG for Class A messages in sea area [own study]

Comparison of harbour and sea area (Class A Messages)

The harbour and sea areas are two completely different traffic zones. They differ in the type of manoeuvres, average speeds, as well as the model of watchkeeping. The differences can also be found in AIS data.

Based on table 5, generally the number of unknown values broadcast in harbour is several times higher than its counterpart in sea area. That affects mainly ROT and THDG but also NAV, which has to be set manually. Moreover, in both areas a strong connection of unknown THDG and ROT could be identified (tables 6 and 7). As the number of AIS messages is higher, the relative number of affected vessels in harbour is 7 times higher than those in the sea area. Another difference occurs in frequencies of messages which have only unknown values of ROT or of SOG and COG together. They are nearly 20 times higher at sea or in harbour area respectively.

In both areas an important part of received AIS messages with unknown values of ROT or THDG was sent at the speed lower than 1 kn. This is done by 25% of vessels in harbour area and by less than 1‰ at sea. At sea there is a second important interval of unknown values of SOG, visible between 5 kn and 10 kn (fig. 5).

SUMMARY AND CONCLUSIONS

The results of the AIS data analysis presented in this paper show, that the AIS system is prone to have incomplete contents of data broadcast by AIS transmitters onboard the vessels. It can be noticed, that the AIS variables set to the unknown values and having the highest occurrence have their sources outside the AIS transponder. The two most significant examples are the rate of turn (ROT) and the true heading (THDG) variables. Both of them are calculated externally and have an NMEA connection with the AIS device. The rest of the AIS variables, like the geographic position or the course over ground, are generally computed internally by the AIS transponder. Their cases of being unknown are rare, as compared to ROT and THDG.

It was found that the number of AIS messages with at least one unknown value as well as the relative number of affected vessels is several times higher in the Port of Rostock than in the sea area between Trelleborg and Arkona. In both areas a significant number of unknown values were sent by vessels at the speed higher than 1 kn. It is worth noticing, that SOG and COG is more absent in harbour area, and ROT, leaving behind THDG, is missing in the sea area.

As the importance of the AIS in everyday navigation increases, it becomes clear, that the AIS data may not contain all the information necessary to properly support decision making on the bridge. Although the radar still remains the major source of traffic information and assessment of safe way to avoid collision, its usage is nowadays intuitively coupled with AIS data feed. If a seafarer is caught off guard

with incomplete AIS data, which they are used to relying on, their ability to make safety critical decisions may be impaired.

Therefore it is also important to improve the delivery of the traffic situation data by providing additional assessment of data integrity at the receiver's side. This could help detect unusable information and warn the seafarer. It is beyond the scope of this paper to investigate, why an AIS transponder is deciding to set some AIS variables to the unknown values. However, this might be a good idea for the manufacturers to examine the matter and to learn, what technical conditions, be it stress or malfunction, may lead to such operation of AIS transponders.

A c k n o w l e d g m e n t s

REFERENCES

[1]   [ITU-R M.1371-3] ITU Recommendation M.1371-3: Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile band.

[2]   [LLI] lloydslistintelligence.com, status of April 2012.

[3]   [NMEA 0183] Standard for Interfacing Marine Electronic Devices, AIS VHF Data Link Message VDM Sentence Encapsulation.

[4]   [SOLAS] Safety of Life at Sea Conventions (SOLAS), Chapter V, Regulation 19.