Plausibility analysis of navigation related AIS parameter based on time series

Frank Heymann, Thoralf Noack, Paweł Banyś
Deutsches Zentrum für Luft und Raumfahrt eV, Neustrelitz, Germany
email: frank.heymann@dlr.de

In this paper we study the plausibility of navigation relevant AIS parameter based on time series analysis of HELCOM AIS data, obtained during September 2011. Previous analysis[1] of AIS data sets were mainly based on studies of not known or not existent (default) values in the AIS messages. The results of these studies have shown that especially true heading (THDG) and rate of turn (ROT) parameters were strongly affected by default / unknown values. We discuss algorithms to evaluate the plausibility of data contained in the AIS message for dynamic data, like time, position, speed, and course. The contained information is checked against the plausibility of their values according to other parameters in the time series. For example the speed can be calculated from the two position information from the AIS message and the time difference between these positions. The dynamic data is consistent in this case when both the calculated and the measured value are similar within an appropriate limit. For static data the determination of plausibility is based on checks against the AIS specification. The obtained results will serve as error model for the development of a maritime traffic situation assessment facility, where AIS, radar and specific PNT data shall be fused together to create a reliable traffic situation image.

Biographies

Frank Heymann received a PhD in physics from the University of Bochum in collaboration with the European Southern Observatories (ESO). From 2010 until 2012 he has worked in the field of active galactic nuclei in astrophysics as a postdoctoral researcher at the University of Kentucky. In 2012 he joined the DLR Institute of Communications and Navigation as a Research Associate in the working group “Nautical systems”.

Thoralf Noack holds a Dipl. Ing. in Electrical and Electronic Engineering from Technical University of Dresden. From 1993 until 1998 he has worked as an application engineer and consultant in acoustical measurement engineering for different companies. In 1998 he started his work as a Research Associate at the DLR Institute of Communications and Navigation. Since 2012 he is the group leader of the working group "Nautical systems". Furthermore he is involved in the IALA e-Navigation PNT working group and represents the Research Port Rostock as its spokesman.

Paweł Banyś holds a master's degree in finance and banking. Between 2001 and 2010 he was employed at different IT companies as network and Linux administrator. He also cooperated with Maritime Academy in Szczecin on a vessel traffic safety project. Since 2010 he has been working at the DLR Institute of Communication and Navigation in the field of AIS and maritime traffic systems. Currently he is studying the engineer's degree in geodesy and cartography at Maritime Academy in Szczecin.

I. Introduction (this is ‘Header 1’)

One of the important carriers of the worldwide economy is the transport of goods and persons realized by vessels. A rapid development of new technologies for the maritime traffic system occurred in the last decades to enable the handling of increased transport volume and to improve the safety. To harmonize the developments of electronic aids to navigation and dedicated systems and services aboard and ashore the International Maritime Organization (IMO) has initiated the e-Navigation strategy to integrate existing and new navigational tools, in particular electronic tools, in an all-embracing system.

With the introduction of the Automatic Identification System (AIS) an important element was established 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)[2] 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 an AIS transponder[2]. 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. [5].

The risk reduction of accidents between ships as well as ships and obstacles is the social goal associated with safe shipping from berth to berth. The technological goal covers the development of new tools and methods to avoid safety-critical events and to support ship-side and shore-side crews during decision making in complicated and complex navigational situations. The avoidance of collisions and groundings is only possible, if reliable and comprehensive information of the maritime traffic situation is available.

II. CONCEPT

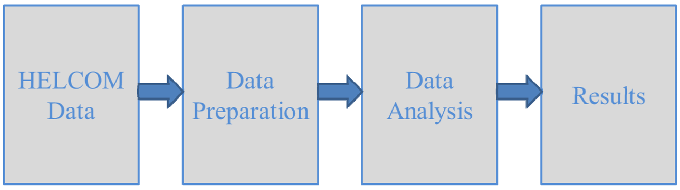

Figure 1: This figure illustrates the concept of the paper

The general quality of AIS data depends on performance of the navigational instruments of a target vessel and radio wave propagation during data broadcast. While the quality of the in the AIS message contained data depends on the performance of on-board devices like the gyro compass or the GNSS receiver and the input of the nautical staff. As the analysis of electromagnetic interference is beyond the scope of this work, the quality of the on-board data sources is to be examined closely. However, it is not possible to gather full information on the state of the equipment, which is responsible for data provision on a vessel, because this kind of data is not contained in the AIS message. This also makes it impossible to fully identify possible errors of telemetry computations or to detect software bugs affecting them. The only chance to figure out, how plausible the data is, seems to be the analysis at the AIS receiver side. The challenge is about finding reasonable proof of faulty data received from an AIS transmitter and computing the impact it could have on proper decision making.

The strategy of this paper to estimate the plausibility of the in the AIS message contained data is illustrated in Figure 1. We use the HELCOM (Helsinki Commission) data from September 2011 with excellent coverage of the Baltic Sea. The data is pre-processed which includes the decoding, sorting and storage. The next step after the pre-processing of the data is the analysis of the received AIS messages according to the algorithms described in section 3. The last part of the paper gives the results of this analysis in section 4 and the summary as well as an outlook in section 5.

III. ANALYZED PLAUSIBILITY OF DYNAMIC AIS DATA

This section explains the possible zero level plausibility checks for dynamic AIS data of COG, SOG, position, heading, navigational status, update rate and the static AIS data of the IMO-number.

A. Course over ground (COG)

The plausibility of the course over ground is checked with the information of two position reports. Assuming that the positions reports are accurate to a certain level it is possible to check the consistency of the course over ground (COG) information. Let's

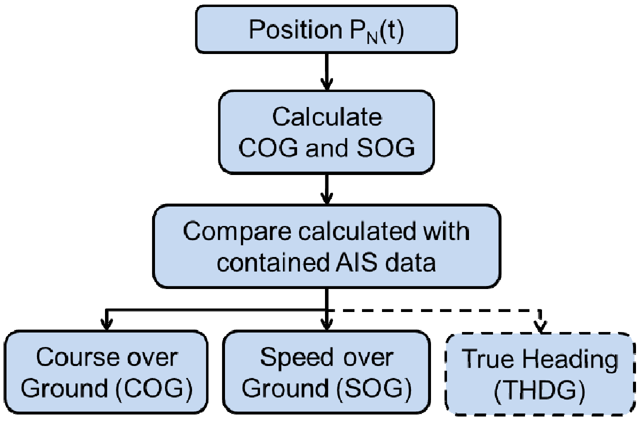

Figure 2: This figure illustrates the strategy of the plausibility analysis

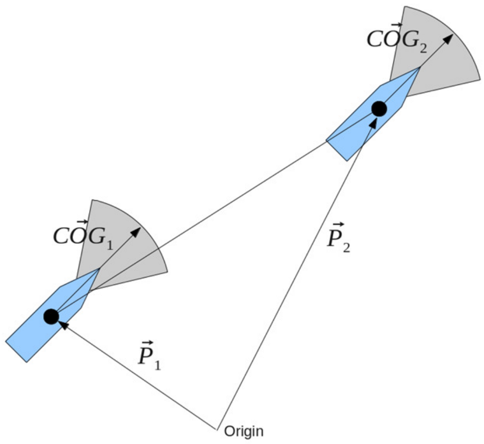

Figure 3: This figure illustrates the situation of a single vessel for at during two times.

consider a case, where the COG information has not changed between two AIS messages. A plausible AIS message would exist, if the angle calculated from the two position reports is comparable to the reported COG (see Figure 3).

B. Speed over ground (SOG)

The integrity of the speed over ground information is checked similarly to the course over ground information. Assuming that the position information is correct it is possible to check the consistency of the speed over ground. The two positions and the time passed during these two reports are used to calculate the average speed. This average speed is compared to the AIS information and the AIS message is considered to be correct if the following equation is true

$$SOG \approx \frac{|\vec{P}_2 - \vec{P}_1|}{\Delta t}, \quad (1)$$

where $\vec{P}_1$ and $\vec{P}_2$ are the positions and Δt is the time passed between the two AIS messages, respectively (see Figure 3). An additional check of the SOG plausibility could be the rate at which the SOG value changes, meaning that the acceleration should be within reasonable values. This check against plausible accelerations is not done in this paper.

C. Position

The position report is crucial to check the integrity of the AIS messages of SOG and COG. It is therefore important to quantify the position report accuracy. The complication is that it would not make much sense to use the information of SOG or COG since we assume that the position report is correct to check for the consistency of these two messages and therefore the absolute error estimate would identical with opposite sign. In this paper we use the continuity of the trajectory of ship travel routes. This means that the position reports should be free of jumps bigger than a given threshold which is motivated by the average speed (5 m/s) and average update rate (6 - 10 seconds) of vessels and AIS messages, respectively. This results in a maximal difference between two points of 30 to 60 meters. With the assumption that not more than 10 AIS messages are missing reasonable position differences are between 300 and 600 meters. In this paper a position report is considered critical if the difference between two points is larger than 500 meters.

The position report consistency can be improved by adding additional external sensor information, like radar data. Note this paper focuses on determination of plausibility information without additional sensor data and possible improvements with data fusion algorithms.

D. Heading

The heading plausibility information can be divided into two categories. Mainly the ones with nonphysical data outside the range of 0 to 359 degrees, most probably the default value of 511. And the other, more challenging heading information, where a value inside the possible range of heading is sent, but it does not correspond to the physical condition of the vessel. In this paper we consider heading information as correct, if the difference to the COG data is smaller than the empirical chosen value of 20 degrees. The reasons for the threshold value instead of the exact value of COG are water currents or ship drifts resulting from wind or other possible sources.

E. Navigational status

The navigational status of a moored or anchored ship can be checked against the position, speed, COG and SOG information and should be consistent with a nonmoving, nonrotating vessel. On the other hand non anchored or moored ships should show some movement, which is implied by changing information of either position, rate of turn or speed.

Table 1: AIS specification[4] of update rates depending on SOG and ROT

Ships Dynamic condition	Speed [knots]	Specified update rate [s]
Ship at anchor or moored	< 3	180
Ship at anchor or moored	> 3	10
Moving	0 – 14	10
Moving and changing course	0 – 14	3⅓
Moving	14 – 23	6
Moving and changing course	14 – 23	2
Moving	> 23	2
Moving and changing course	> 23	2

An additional way of checking the navigational status of anchored or moored ships is the cross correlation with a database of all ships anchored in a harbor, which is beyond the scope of this paper.

F. Update rate

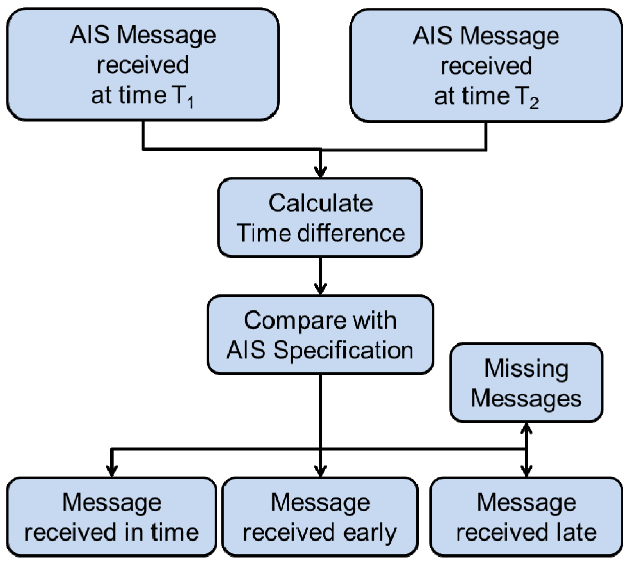

Figure 4 This figure illustrates the strategy of the update rate plausibility analysis

The update rate is defined in the specifications and depends on the rate of turn, the speed and the navigational status of the vessel. The specified update rates for all possible ship conditions are shown in Table 1. The strategy to check the plausibility of the update rate is illustrated in Figure 4.

As the HELCOM database contains the times when the AIS messages were received. In addition contain the dynamic AIS messages the necessary data of SOG and COG in order to check the performance of the AIS update rate by comparing the time difference of the received message with the specified update rate (Table 1). As illustrated in Figure 4. the comparison with the specification may lead to four possible states. The first state is that the AIS message was received in time. The second possible state is that the AIS messages was received prior to the from the specification expected update rate. Third the message was received late and last the message was not received at all.

G. International Maritime Organization (IMO) – Number

Table 2: Illustration the IMO checksum calculation

IMO	9 – 0 – 7 – 4 – 7 – 2 – 9
Weight	7 – 6 – 5 – 4 – 3 – 2
Verification	7*9 + 6*0 + 5*7 + 4*4 + 3*7 + 2*2 = 139 --> 139 mod 10 = 9

IMO numbers can be verified by their checksum. According to Fluit, 2011[6] the IMO-number contains a check digit at the end of the number. The check digit is calculated by the preceding digits in the following way. Weight the digit with the position and sum all weighted digits. The last digit of this sum equals the last digit of the IMO number if the IMO number is valid. The following example explains the verification scheme.

IV. Default and Critical Values

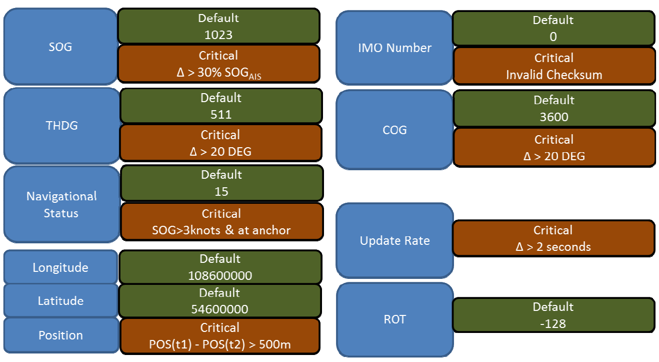

Figure 5: This figure show the definitions of AIS default and critical values as used throughout this paper

This section describes the default value as defined by the AIS standard (). Furthermore values are signed as critical, if applied plausibility of consistency tests are not passed.

We want to note that test criteria are empirical and might be incomplete. An overview of all analyzed parameters and their definition of critical and default values are given in Figure 5. We use the term of contained values as the average value between the two values of the used AIS messages.

The SOG value has the default value of 1023[3]. We consider the SOG as critical if the contained value in the AIS message deviates more than 30% from the calculated values.

The THDG value has the default value of 511[3]. We consider the THDG as critical if the difference between the contained THDG and the calculated COG is larger than 20DEG. This holds only under the assumption that current velocities of water are smaller than the vessel speed. But the AIS system does not contain any information about the water current. It should take care in the conclusion that it is impossible to estimate its influence.

The IMO number has a default value of 0[3]. We consider the IMO number as critical in the case where the CHECKSUM, which is calculated as described in section 3, of the IMO number is not correct.

The navigational status has the default value of 15[3]. We consider the navigational status as critical if the contained values states that the vessel is at anchor but the calculated speed over ground is larger than 3 knots. It should be stated that there are possible situations in which a vessel is at anchor and moves faster than 3 knots but these situation should be rare over the total of all vessels at anchor.

The position is an important parameter since the most calculations are based on the position. The default values of the position are 108600000 and 54600000 for the longitude and latitude, respectively[3]. We consider a position report as critical if the difference between two positions reports is larger than 500m. This is motivated by the average speed of vessels and the specified time difference between two AIS messages as discussed in the previous section.

The update rate is given in the AIS specifications and is based on the speed and rate of turn of the vessel. We consider the update rate as critical if the difference between the received update rate and the specified update rate is larger than an empirical chosen value of 2 seconds.

The default value of the ROT is defined as -128[3]. The calculation of the ROT value is not straight forward for the position reports possible and the analysis of the time dependent behavior of the ROT value is beyond the scope of this paper. Therefore we do not provide a definition for a critical ROT value.

V. Results

This section describes the results of the analysis for the HELCOM data of September 2013 and is structured into three parts describing the analysis for a single day in the first part followed by the analysis of the month and finally estimating the amount of critical values for navigation.

A. Single day analysis

This section describes the analysis of the COG and SOG and update rates as of 13th of September 2011. The differentiation between harbor and sea areas was not applied, to provide a general overview of a single day AIS data set.

1) SOG

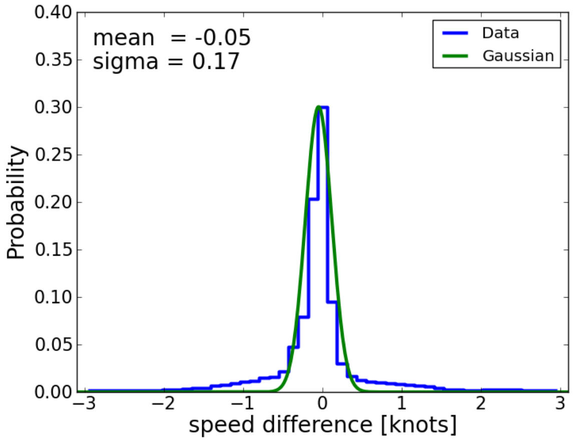

Figure 6: This figure shows the histogram the difference between calculated and reported SOG data.

The single day analysis of the 13th of September 2011 for the speed over ground AIS values shows a small lobe with a peak around zero. The histogram (blue solid line) in Figure 6 shows the probability on the y-axis versus the speed difference between calculated and reported values on the x-axis. In addition we plot a Gaussian curve (green solid line) with a mean value of -0.05 kn and a standard deviation of 0.17 kn. With the assumption that the Gaussian distribution represents the data, the reported speed over ground corresponds to the calculated SOG from two successive positions within 0.34 kn (95% confidence level).

2) COG

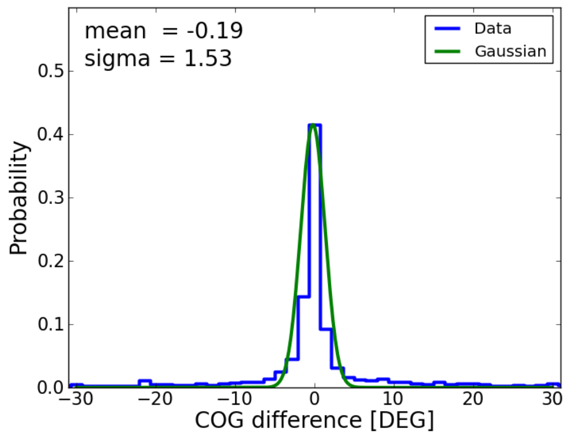

Figure 7: This figure shows the histogram the difference between calculated and reported COG data.

The single day analysis of the 13th of September 2011 for the course over ground AIS values shows a small lobe with a peak around zero. The histogram (blue solid line) in Figure 7 shows the probability on the y-axis versus the course difference between calculated and reported values on the x-axis. In addition we plot a Gaussian curve (green solid line) with a mean value of -0.19 degrees and a standard deviation of 1.53 degrees. With the assumption that the Gaussian distribution represents the data, the reported course over ground corresponds to the calculated COG from two successive positions within 3.06 degrees (95% confidence level).

3) Update rate

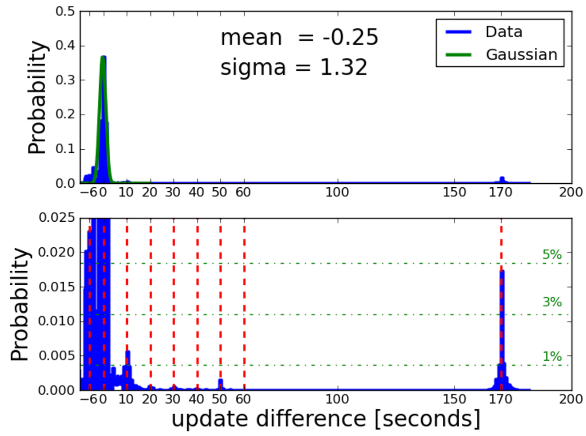

Figure 8: This figure shows the histogram the difference between calculated and reported update rate

The single day analysis of the update rate shows similar behaviour with three interesting areas. The first part is the largest peak around 0. These are the messages which are received at the time expected from the specification. The green solid curve plots a Gaussian with a mean of 0 and a standard deviation of 1.32s to guide the eye of the reader. The analysis ends up with about 87% of the data being within the -6 to 6 second bin. With the assumption of the Gaussian distribution the 95% confidence level is reached for the data with 2.6 seconds difference between observed and expected update rate. We can conclude that 87% x 95% = 82% of the received messages are sent within 2.6 seconds of the specified update rate.

The peak around -6 seconds (around 5% of all data) is most likely caused by wrong ROT values. According to table 1 a vessel at the speed between 0 kn and 14 kn updates every 10 seconds if not turning and every 3.333 seconds if turning. This results in a difference of 10 - 3.33 = 6.67 seconds which is seen in this figure.

The second largest high at 170s (5% of the data) this might be caused by two different effects. First a slow moving vessel below 3 knots which is not moored but flagged as moored will report every 180s instead of 10s. Or a moored vessel which should be slower than 3 knots but in fact is faster than three knots. The latter case would indicate a faulty AIS transponder, while the former one could be the result of a human error or negligence.

The additional peaks (less than 1% of the data) around the points 10, 20 ,30 ,40 and 50 are most likely caused by missed AIS messages.

B. Monthly analysis

This section investigates the day to day variations of the SOG and COG values during September 2011.

1) COG:

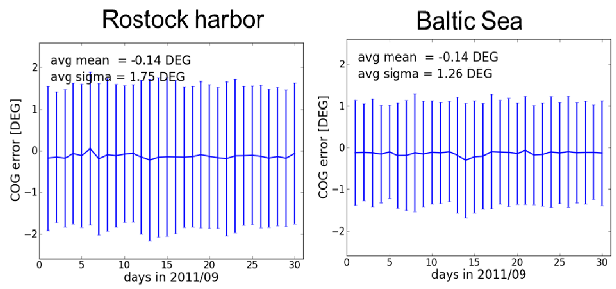

Figure 9: Analysis of the difference between the reported and calculated COG values in September 2011

Figure 9 shows the difference between the calculated and reported COG values during September 2011 in two areas. The left side shows the Rostock harbor while the right side the Baltic Sea. In the graph there is no significant difference between the days respectively during September. Between the different areas there is a small difference in the performance visible. The monthly average between calculated and reported values is around zero in both cases but the calculated standard deviation is slightly lower in the Baltic Sea when compared to the harbor. This might be explained by the fact that the dynamics, read average speed, in the harbor is limited and therefore the position accuracy is more important. A distance travelled at sea compared to a distance made in the harbor during the same period of time is longer.

2) SOG:

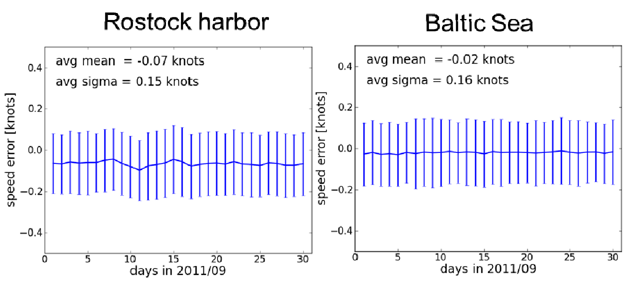

Figure 10: Analysis of the difference between the reported and calculated SOG values in September 2011

Figure 10 shows the difference between the calculated and reported SOG values in September 2011 in two areas. The left side shows the Rostock harbor while the right side the Baltic Sea. In the graph there is no significant difference between the days respectively in September. In comparison to the COG value there is no difference in the performance visible.

The monthly average between calculated and reported value is around zero for both cases and the calculated standard deviation is very similar in both the Baltic Sea and the Rostock harbor.

C. Critical or default?

This part analyzes the performance of navigational relevant AIS parameters focusing on default or critical values as defined in section 4. We study the two areas, the Baltic Sea and Rostock harbor.

1) Rostock harbor

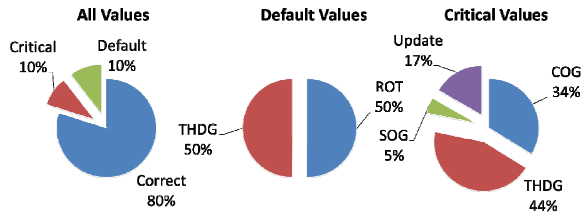

Figure 11: Statistical analysis of the AIS messages received in the Rostock harbor.

The statistical analysis in the Rostock harbor shows that out of all received AIS messages 10% are default and another 10% are critical. Figure 11 shows that the 10% default values are caused by the reported THDG (50%) and ROT (50%) data. The critical values on the other side show also large contributions from the COG (34%) and THDG (44%) data but have additional SOG (5%) and Update rate (17%) contributions.

2) Baltic Sea

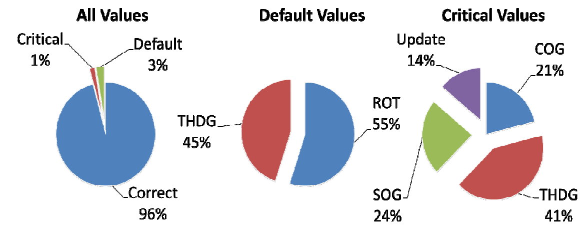

Figure 12: Statistical analysis of the AIS messages received in the Baltic Sea.

The statistical analysis critical and default in the traffic separation scheme in the Baltic Sea shows a much smaller amount of critical (1%) and default (3%) in comparison to the Rostock harbor. Analysing the default values in more details reveals the same behaviour as in the Rostock harbor. The default values are caused by unknown THDG (45%) and ROT (55%) values. As shown in Figure 12 the distribution of critical values shows that in the Baltic Sea the values Update rate (14%), COG (21%), THDG (41%) and SOG (24%) contribute equally to the total critical values budget. The smaller amount of critical values might be explained by the fact that the vessel dynamic is larger in the traffic separation area in comparison to the harbor and therefore position errors have a smaller effect.

VI. SUMMARY AND OUTLOOK

We can summarize that the AIS messages contain on average less than 11% of default values (2.6% in the Baltic sea; 10.5% in the Rostock harbor). Additional we conclude that AIS message contain on average less than 10% of critical values (1.5 in the Baltic sea and 9.6% in the Rostock harbor).

The detailed analysis of the performance of the COG value show that 95% of the reported values are within 3 degree difference to the calculated ones.

The same analysis is done for the performance of the SOG values and we conclude that 95 % of the reported values are consistent with the calculated ones within 0.3 knots.

The analysis of the update rate shows four interesting results. First 82% of the update rate are within 2.6 seconds of the specified update rate (see table 1). Second there are messages which are received around 6 seconds early as expected from the specification (table 1). Third it exists update rates which indicate that there are missing AIS reports. And fourth there are messages which have a delay of 170 seconds which could be explained by wrong navigational status.

What is planned in the future?

The in this paper performed study analysis the total statistics of all vessels and does not investigate the behavior of single vessels. To answer the question if the critical values are mainly caused by single vessels with faulty equipment we are interested to analyze the data on a vessel by vessel base.

Another interesting analysis is to benchmark the performance gain of IALA beacon. The AIS message contains a flag if the IALA beacon service is used. Assuming that this flag is set correctly it is possible to separate the database into vessels with IALA beacon and analyze the performance of SOC and COG.

The final goal of this studies focuses on the final fusion of Radar and AIS data to improve reliability and continuity and generate integrity information for the navigational relevant AIS values.

REFERENCES

[1] P. Banyś., T. Noack, S. Gewies, 'ASSESSMENT OF AIS VESSEL POSITION REPORT UNDER THE ASPECT OF DATA RELIABILITY', Proceedings of the European Navigation Conference 2012, Gdynia

[2] SOLAS. 1974. 'Safety of Life at Sea Conventions ', SOLAS

[3] NMEA 0183. 'Standard for Interfacing Marine Electronic Devices', National Marine Electronics Association

[4] Rec. ITU-R M.1371-3 'Technical Characteristics for an Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Band.'

[5] lloydslistintelligence.com, status of April 2012

[6] A. Fluit, 'AIS Information Quality Report', efficiensea.org 2011