Expanding the Possibilities of AIS Data with Heuristics

B.B. Smestad & B.E. Asbjørnslett
Norwegian University of Science and Technology, Trondheim, Norway

Ø.J. Rodseth
Sintef Ocean, Trondheim, Norway

ABSTRACT: Automatic Identification System (AIS) is primarily used as a tracking system for ships, but with the launch of satellites to collect these data, new and previously untested possibilities are emerging. This paper presents the development of heuristics for establishing the specific ship type using information retrieved from AIS data alone. These heuristics expand the possibilities of AIS data, as the specific ship type is vital for several transportation research cases, such as emission analyses of ship traffic and studies on slow steaming. The presented method for developing heuristics can be used for a wider range of vessels. These heuristics may form the basis of large-scale studies on ship traffic using AIS data when it is not feasible or desirable to use commercial ship data registers.

1 INTRODUCTION

Analyses of ship traffic are important, e.g. to estimate emission of greenhouse gases, monitor fleet efficiency and for conducting studies on ship safety. Automatic Identification System (AIS) data has become an integral part of these studies, as they provide positional and operational information for a large part of the shipping fleet.

AIS is a communication system that uses the maritime Very High Frequency (VHF) bands to transmit ship movement and technical data at specified intervals. This includes static data, such as the ship’s name, draught, destination and Estimated Time of Arrival (ETA), as well as dynamic data from the ships sensors, such as speed and position (ITU 2014). A typical use of AIS is to exchange information between vessels that are in the same area, to automatically identify other ships and avoid high risk situations. It is also used in traffic monitoring, to provide guidance by vessel traffic services (VTS) and by many other shore side users. The development of AIS was a joint project between the International Maritime Organization (IMO) and the International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA). The International Convention for Safety of Life at Sea (SOLAS) states that all ships of 300 gross tonnage and upwards engaged in international voyages, cargo ships of 500 gross tonnage and upwards not engaged in international voyages, as well as all passenger ships built after 2002, or operated after 2008, should have an AIS (IMO 2002). This essentially means that all larger ships engaged in global shipping should have AIS equipment. National requirements will normally also require ships not covered by IMO regulations to carry AIS transmitters. This means that more than 85 000 ships world-wide will transmit AIS data (Mantell 2014).

AIS data is gathered by AIS receivers, which can be found on board ships, on buoys, on land (IALA 2011) and more recently on satellites (hereafter S-AIS).
Land based AIS receivers can detect AIS messages normally up to 40-50 nautical miles offshore (Skauen 2013), ships further off-shore will remain undetected by land based AIS receivers. In 2005, researchers from the Norwegian Defence Research Establishment published the first study investigating whether satellites could be used to gather AIS signals (Wahl 2005). In 2008, a follow up study by Høye et al. (2008), found that AIS signals could be detected by satellite based AIS receivers positioned in altitudes of up to 1000 km. However, since the AIS system was not initially designed for space based receivers, but rather to be a ship-to-ship communication system, there were some problems. A satellite will have a much larger coverage area than AIS receivers were designed for, which could lead to interference problems between the different ships’ AIS signals. According to the study, the result could be that some AIS messages would not be detected by the satellite. In practice this leads to a more reliable satellite coverage in areas with less traffic, while high trafficked areas can have interference problems. In 2010, the Norwegian AIS satellite AISsat-1 was launched. This satellite is in a sun-synchronous polar orbit at 630 km altitude (Eriksen 2010). The satellite transmits the AIS messages it receives to Svalbard Ground Station at each passing. Eriksen et al. (2010) states that over a time span of 24 hours, areas along the equator is covered two to three times, while the High North and South is covered up to 15 times. In 2013, AISSat-2 was launched to give extended coverage. This gave a higher update rate to the Svalbard Ground Station, as well as a higher global detection rate.

The use of AIS data in studies on maritime transportation has become increasingly prevalent. Smith et al. (2014) prepared a report as a part of the World Shipping Efficiency Indices project funded by the International Council on Clean Transportation. The study combined global S-AIS data from 2011 with technical ship data from sources like Clarksons World Fleet Register, and the Second IMO Greenhouse Gas Study (Buhaug 2009). The S-AIS data provided operational characteristics, such as speed and loading condition. In addition, estimates on the distance travelled were derived from the S-AIS data. Data from Clarksons World Fleet Register provided technical specifications, such as the ship type (for instance LNG tanker or crude oil tanker) for each individual ship.

The Third Greenhouse Gas (GHG) study by Smith et al. (2014) had an advantage over the preceding studies, as it could utilize S-AIS data. These data were used to get more precise activity measures and better emissions estimates for each ship. This was aggregated to the total emissions for each ship type. In the previous study, emissions were estimated by using the annual average activity for the different ship types.

Categorizing ships into ship type and size category is vital to perform studies on operational efficiency and greenhouse gas emissions. Knowing the design speed is necessary for developing speed-relative fuel consumption models for ships - where the design speed is the speed giving the “optimal” trade-off between speed and fuel consumption. The design speed is amongst others a factor of the block coefficient of the ship, which in turn is largely given by the ship type. Previous studies, such as Smith et al. (2014), have used commercial vessel databases to retrieve the ship type for each specific ship in the study. However, using external databases to retrieve the ship type can be costly as these databases require a subscription. On the other hand, manual retrieval of the ship type from open databases can be time consuming. The combination of these two factors may inhibit studies on maritime transportations using estimation based on AIS data.

In the SESAME Straits project (SESAME 2017), the challenge was to give guidance to ships headed for and in the Straits of Malacca and Singapore and to estimate possible fuel savings by suggesting more efficient speeds to the ships. A problem, however, is to find enough information about the ships to do a reasonable estimation of fuel use and fuel savings for different speeds. This information can be bought, but in just five days, more than 3000 different ships were recorded by the AIS-stations in the area. As the market for such services are limited and quite cost sensitive, it was not very attractive to buy the information.

The research questions that emerged, when faced with these challenges was: How well can AIS data alone identify the ship type and size? Can heuristics for identifying the ship type for any ship be constructed? The objective of this study was to establish heuristics for identifying the ship type for a large proportion of the world fleet, using S-AIS data.

The method for constructing the heuristics is outlined in Section 2, while the heuristics parameters can be found in Section 3. The performance for these heuristics is provided in Section 4, while the results and the validity of the heuristics are discussed in Section 5. A conclusion is given in the final section.

2 METHOD

Satellite AIS data spanning the time period of May 1st 2014 to September 15th 2014 was retrieved. The S-AIS data had been collected using the two satellites AISsat-1 and AISsat-2, and was provided by the Norwegian Coastal Administration for use in the SESAME Straits research project. AISSat-2 data was only available after its launch in July 2014.

These S-AIS data included static and/or dynamic AIS messages for 85,108 ships, identified by unique MMSI numbers. 43,671 of these ships had both dynamic and static data. Mantell et al. (2014) stated that the total world fleet consisted of 88,483 ships as of May 2014. Approximately 95% of the world fleet is present in our data, and about half of the world fleet is represented with both dynamic and static data. These S-AIS data is shown as group A in Figure 1.

We developed heuristics for a selection of ship types with high relevance to international shipping (Table 1). This selection is in line with the selection in other studies such as Smith et al. (2014).
Table 1. AIS vessel groups, ship types and sizes in this study

<table>
<thead>
<tr>
<th>AIS vessel group</th>
<th>Ship type</th>
<th>Ship size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tankers</td>
<td>LNG and LPG Carriers</td>
<td>General, Q-Flex and Q-max</td>
</tr>
<tr>
<td></td>
<td>Oil Tankers</td>
<td>UL&amp;VLCC</td>
</tr>
<tr>
<td>Cargo ships</td>
<td>Container vessels</td>
<td>Panamax</td>
</tr>
<tr>
<td></td>
<td>Bulk carriers</td>
<td>Panamax</td>
</tr>
</tbody>
</table>

The Clarksons Group provides a database where a selection of vessels of each ship type and size category are listed by the ship’s name (Clarksons 2015). This data is shown in Figure 1 as group B. The ships are only identified by their name, and not by a more unique identifier such as their IMO or MMSI number. Vessel characteristics were also retrieved from the S-AIS data by matching the name of the ship from the vessel database to the name registered in the S-AIS data. The ships that were present in both the S-AIS data and the vessel database are a candidate group, formed by a subset of the two groups, and is shown as group C in Figure 1.

Figure 1. The process used for constructing the heuristics.

The vessels are matched between Clarksons vessel sheets and the S-AIS data based on their name, and not their unique IMO number, so there is a possibility that ships from other ship classes, with the same ship name, are included in the candidate group. To mitigate this source of errors, a data cleaning process was required. In the data cleaning, ships with dimensions outside the expected interval for the ship type in question were removed. For instance, cargo ships and tankers are typically classified into different size categories, which often correspond to the maximum dimensions of important seaways and ports, such as the Panama Canal and the Suez Canal. If a ship was categorized as a Panamax ship in the Clarksons vessel database, but had reported a width or draught exceeding the set of maximum dimensions in the Panama Canal in the S-AIS data, it was not included in the training group.

The initial version of the method used maximum observed speed as one of the parameters for classification. Early testing of this heuristic showed that some ships were misidentified. As an example, a 274 m long and 48 m broad oil tanker had a maximum observed speed of 20 knots. Because of the relatively high speed, this vessel was classified as an LNG carrier. However, speed recordings from AIS data is most commonly speed over ground, and not speed relative to the water. These recordings may thus be a result of particularly favorable wind and current conditions, and not necessarily errors in speed data. To find the frequency of the different speed recordings, all reported speeds were bucketed in one knot intervals. Out of 165 speed recordings for this vessel, there was only one record of the maximum recorded speed of 20 knots. The highest speed, amongst those with the highest frequency, was 14 knots. The data showed that the vessel had this speed at ten occasions. To avoid these rare occurrences of high speed, a new constraint was put in the heuristics; for a maximum speed to be valid, the vessel should have ten or more AIS records of having that speed.

After the data had been cleaned, we used the resulting ships as a training group for the heuristics, shown by group T in Figure 1. Using this training group, common dimensional traits and operational characteristics for each ship type was derived by inspection, and ultimately used to form the heuristics. This was repeated for every ship type in Table 1. The process of making heuristics for panamax bulk carriers is used as an example and outlined below.

The heuristic, which consists of constraints on dimensions, draught, speed and AIS vessel group, were applied on the full set of S-AIS data (group A) as a performance test. The performance of each heuristic was checked by manually confirming the specific ship type of all ships classified by the heuristic, using online ship databases. These are databases where the ship type of a single ship can be found using the IMO or MMSI number. The accuracy of a heuristic was defined as the number of ships correctly identified by ship type, divided by the total number of ships identified.

2.1 Developing heuristics for Panamax Bulk Carriers

The candidate group for the heuristic training group was made by identifying all panamax bulk carriers present in both the Clarksons vessel database and the S-AIS data. Out of the 2459 panamax sized bulk carriers in Clarksons vessel database at the time of retrieval (spring 2015), 2200 ships were also present in the S-AIS data.

2.1.1 Data Cleaning

2.1.1.1 Erroneous ship dimensions

The breadth was required to be less than 34 m, as the maximum width of the Panama Canal is 33.5 m. The extra 0.5 m was allowed, as some panamax bulk carriers seemed to be registered with a width of 34 m in the S-AIS data, probably due to a rounding error. This constraint is illustrated by the top horizontal line in Figure 2. There were a lot of vessels in the candidate group exceeding this breadth. The fact that seemingly panamax vessels could exceed this constraint can be attributed to the lack of a unique identifier in the vessel sheets as earlier described. In other words, these may have been non-panamax vessels having the same name as the panamax vessels in the Clarksons vessel sheets. To ensure that only Panamax vessels were present in the training group, an additional breadth constraint of minimum 30 m was added. Vessels below this breadth would fall into other ship categories. This constraint is illustrated by the bottom horizontal line in Figure 2. These breadth requirements reduced the candidate group to 1668 ships.
especially favorable wind and current conditions. It can also be due to other ships being misidentified as bulk carriers. Because of these findings, the maximum recorded speed allowed in the heuristic was set to 15 knots.

After the breadth constraints was enforced, three vessels in the resulting group had a length over 250 m. The dimensions of these three vessels were manually inspected in an open ship database, to check for any errors. The longest ship, Vishva Anand, actually had a length of 229 m, not the 332 m it was recorded with in the S-AIS data. The second longest ship was a container vessel misidentified as the bulk carrier Santa Regina, as they shared their name. The last vessel was the 259 m long bulk carrier Orissa. This is an exceptionally long bulk carrier, with a breadth of only 32 m. Since these three vessels either were wrongly registered or exceptionally large, they were excluded. After these exclusions, 1665 vessels remained. The rest of the vessels in the candidate group had reasonable sizes, and we had no reason to suspect that their dimensions were erroneous. These ships could now be used as a training group for the heuristics.

2.1.2 Heuristic training

2.1.2.1 Maximum speed constraint

Because of the high utilization of the ship’s volume in bulk carriers, it was expected that the maximum speed as registered by AIS-S is lower compared to other dimensionally similar vessels, such as container vessels. As many as 92% of the container vessels had an observed maximum speed of 15.9 knots or more, while 92% of the Panamax bulk carriers had an observed maximum speed of 15 knots or less. There was a group of ships reporting speeds up to 18 knots, which can be seen in Figure 3. This can be due to
The change in draught ranged from zero to just below 19 m. 99% of the vessels in the training group had a change of draught less than 9.5 meters. Container vessels are most likely the ship type to be misidentified as bulk carriers, because they have the same AIS ship type (cargo ship), but unlike bulk carriers, container vessels have a lower difference between the draught in a fully- or less loaded state. Thus, a lower limit for maximum change of draught was a reasonable boundary to set. When the heuristic for the container vessels was developed, we found that 97% of the container vessels had a change of draught of less than 5.5 m. The lower limit for the maximum change of draught was set to 5.5 m, to exclude these container vessels, and the training group was reduced to 1,210 vessels.

2.1.3 Heuristic testing

When testing the heuristic for the panamax bulk carriers (Table 2) on the S-AIS data, a total of 6,024 vessels matched the dimensional data in the heuristic. After applying the minimum change in draught and the maximum speed constraint, 2,346 vessels remained. 1,210 of these ships were in the training group. This means that the heuristic identified an additional 1,136 ships as panamax bulk carriers. Manual inspection showed that 43 of these vessels were misidentified. Out of the misidentified ships, 38 were general cargo ships and three were offshore support vessels. In addition, a container vessel and a vehicles carrier were misidentified. With 43 vessels misidentified out of 2,346 identified vessels the accuracy of the heuristic was 98%.

A similar exercise was done for each of the ship types listed in Table 1 and corresponding parameters determined.

3 HEURISTICS

The parameter sets for the heuristics for the different ship types can be found in Table 2 to Table 5. The parameters were developed using the method outlined in the previous section. First, a candidate group out of ship registry data was formed, then erroneous data were removed. This subset of the candidate group formed a training group, where common traits was derived. These common traits
were used to establish parameters to identify a ship type and ship size. The heuristics reflect the different ship characteristics, spanning from minimum and maximum draught to maximum and minimum speed, depending on the traits of the ship type and size.

3.1 Bulk carriers

Heuristics were developed for the subgroup of panamax bulk carriers (Table 2). The maximum recorded speed was set to less than or equal to 15 knots, and the length, breadth and minimum change of draught was set according to the parameters of the training group.

3.2 Container vessels

Heuristics for container vessels are provided in Table 3. In the development of this heuristic, the vessels in the panamax training group were split into two groups dependent on length and maximum draught. To accommodate for this, one group is termed Panamax while the longer group with a higher maximum draught is termed Post-panamax. This is not fully consistent with general terminology as Panamax includes lengths up to 280 m.

3.3 Gas carriers

Based on the training group for gas carriers, the characteristic parameters were divided into three main groups: A general group as well as a group for Q-Flex vessels and a group for Q-Max vessels (Table 4). These heuristics shares the AIS vessel group, the maximum recorded speed, the maximum draught and maximum change of draught, while only the breadth and length differ.

3.4 Oil tankers

The parameter set in Table 5 is for two groups of oil tankers: ultra large crude carriers (ULCC) and very large crude carriers (VLCC).

4 RESULTS

The satellite AIS data was stored on a database. Using the heuristics parameters as the retrieval query parameters, we retrieved all the ships that matched the different categories. For each ship that was identified as one of the aforementioned ship types, we did a manual check against public ship databases to check if it was correctly identified. This was done by using the ship’s IMO number. In Table 6, the accuracy of each heuristic, as well as the number of vessels in the world fleet for that ship type and the number of vessels in the training group can be found. The accuracy was quantified as the percentage of the number of vessels correctly identified.

Table 2. Heuristics for panamax bulk carriers.

<table>
<thead>
<tr>
<th>Ship size</th>
<th>Length [m]</th>
<th>Breadth [m]</th>
<th>Min. change of draught [m]</th>
<th>Min. draught [m]</th>
<th>Max. recorded speed [kn]</th>
<th>AIS vessel group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panamax*</td>
<td>180-250</td>
<td>30-30</td>
<td>5.5</td>
<td>5</td>
<td>&lt;= 15</td>
<td>Cargo ship</td>
</tr>
<tr>
<td>*Sub group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Heuristics for panamax container vessels.

<table>
<thead>
<tr>
<th>Ship size</th>
<th>Length [m]</th>
<th>Breadth [m]</th>
<th>Min. change of draught [m]</th>
<th>Max. draught [m]</th>
<th>Max. recorded speed [kn]</th>
<th>AIS vessel group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panamax</td>
<td>210-269.9</td>
<td>31-33</td>
<td>5.5</td>
<td>13</td>
<td>&gt;= 15.9</td>
<td>Cargo ship</td>
</tr>
<tr>
<td>Post-panamax</td>
<td>270-300</td>
<td>31-33</td>
<td>5.5</td>
<td>14</td>
<td>&gt;= 15.9</td>
<td>Cargo ship</td>
</tr>
</tbody>
</table>

Table 4. Heuristics for gas carriers of different ship sizes.

<table>
<thead>
<tr>
<th>Ship size</th>
<th>Length [m]</th>
<th>Breadth [m]</th>
<th>Max. change of draught [m]</th>
<th>Max. draught [m]</th>
<th>Max. recorded speed [kn]</th>
<th>AIS vessel group</th>
</tr>
</thead>
<tbody>
<tr>
<td>General group</td>
<td>270-300</td>
<td>40-52</td>
<td>3.5</td>
<td>13</td>
<td>&gt;= 16</td>
<td>Tanker</td>
</tr>
<tr>
<td>Q-Flex</td>
<td>314-316</td>
<td>48-50</td>
<td>3.5</td>
<td>13</td>
<td>&gt;= 16</td>
<td>Tanker</td>
</tr>
<tr>
<td>Q-Max</td>
<td>344-345</td>
<td>46-54</td>
<td>3.5</td>
<td>13</td>
<td>&gt;= 16</td>
<td>Tanker</td>
</tr>
</tbody>
</table>

Table 5. Heuristics for ULCC and VLCC oil tankers.

<table>
<thead>
<tr>
<th>Ship size</th>
<th>Length [m]</th>
<th>Breadth [m]</th>
<th>Min. change of draught [m]</th>
<th>Min. draught [m]</th>
<th>Max. recorded speed [kn]</th>
<th>AIS vessel group</th>
</tr>
</thead>
<tbody>
<tr>
<td>ULCC &amp; VLCC</td>
<td>320-400</td>
<td>50-70</td>
<td>8</td>
<td>10</td>
<td>&lt;= 16</td>
<td>Tanker</td>
</tr>
</tbody>
</table>
DISCUSSION

Simple heuristics to identify the ship type from satellite AIS data was developed, utilizing a comprehensive set of S-AIS data. The heuristics were developed by combining data from a commercial vessel database with S-AIS data to form a candidate group. The candidate groups were inspected for erroneous data and used as training groups for the heuristics.

The heuristics developed is fully based on S-AIS data from static and dynamic messages, containing information such as AIS-ship type, general dimensions, draught and speed.

It should be noted that the maximum and minimum draught, as well as the maximum speed, are products of the operating conditions of the vessels. These operating conditions can be affected by factors such as seasonal macro-variations, as well as yearly macro-variations. Corbett et al. (2009) has shown that the average speed of the shipping fleet can be influenced by fuel cost. Maximum draught is influenced by the loading condition of the ship, which again is influenced by the market the ship operates in. A strong market means that a large quantity of goods is transported, and the maximum draught recorded can therefore be expected to be higher, compared to in a weak market. In a strong market, ships can be expected to have a different operating speed than in a weak market. As the S-AIS data spans a relatively short time-period, the heuristics could turn out to have lower accuracy with S-AIS data from another time-period.

However, the heuristics in this paper are partly based on static information such as ship dimensions. This combined with a large number of ships can negate some of the expected variation. With data for a longer time-period, the number of vessels identified, compared to the world fleet, is expected to rise. A longer period of time and more data means a higher probability for ships to exceed constraints, such as max/min speed and max/min draught.

Another limitation is that when vessel sheets were used to make training groups to develop heuristics, the only way to match information for each vessel from the S-AIS data was through the ship's name. This was the case because the ships were only identified by their name in the vessel sheets that were used to develop the template groups. As a ship’s name is not a unique property, such as the MMSI or IMO number, it is expected that some of the ships in the template group may belong to another ship type, class or size. To avoid this, the candidate groups themselves were refined through a manual process, where ships not abiding the expected dimensions were sorted out. The development of the heuristic was in all an iterative process, where experience based constraints were put in place to ensure a best possible template group.

Nonetheless, the results shown in Table 6 show the proficiency of this method.

For the gas tankers, especially the Q-flex and Q-max groups, few ships outside of the training group were correctly identified as gas carriers, and thus the accuracy is somewhat misleading. This could either be caused by too strict restrictions, meaning that the training group essentially was not representative for the rest of the gas tankers, or more probable, by a low number of vessels of this type in the world fleet. In other words, it is likely that most of the world fleet was the training group. In a typical statistical, or machine learning, model this could be seen as overfitting, however these heuristics are more descriptive than predictive in nature, and thus this become a lesser issue. However, future studies should be conducted to confirm the accuracy of these heuristics.

<table>
<thead>
<tr>
<th>Ship Type</th>
<th>AIS vessel group</th>
<th>Vessels in the world fleet</th>
<th>Vessels in the training group</th>
<th>Vessels identified in S-AIS data</th>
<th>Vessels correctly identified</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas Carriers</td>
<td>Tankers</td>
<td>249</td>
<td>251</td>
<td>249</td>
<td>249</td>
<td>99%</td>
</tr>
<tr>
<td>General group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q-Flex</td>
<td></td>
<td>26</td>
<td>28</td>
<td>28</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>Q-Max</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>Oil Tankers</td>
<td>Tankers</td>
<td>624</td>
<td>309</td>
<td>374</td>
<td>372</td>
<td>99.4%</td>
</tr>
<tr>
<td>UL&amp;VLCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Container Vessels</td>
<td>Cargo ships</td>
<td>875</td>
<td>665</td>
<td>807</td>
<td>729</td>
<td>90.3%</td>
</tr>
<tr>
<td>Panamax</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulk Carriers</td>
<td>Cargo ships</td>
<td>2,405</td>
<td>1,210</td>
<td>2,346</td>
<td>2,303</td>
<td>98%</td>
</tr>
<tr>
<td>Panamax</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Accuracy of the different heuristics measured by the number of correctly identified vessels out of those identified as a certain ship type and size. The number of vessels in the world fleet is according to Mantell et al. (2014)
Based on global S-AIS data from a period of four months, we developed heuristics to determine the type of a ship from AIS data alone. These heuristics give high confidence identification of ship types of importance to the international shipping industry. This is presented as a proof of concept to show that a simple method of identification can give very good results in the specific cases. We also believe that this method can be extended to many other ship types, possibly by extending the number of identification parameters. Our study showed that the S-AIS data can be erroneous, and require cleaning before reliable identification can be made. However, this cleaning is based on simple and easy to implement criteria. To further improve the accuracy for all ship classes, data from a longer time-period should be used. More refined heuristics can possibly be made using techniques such as advanced cluster analysis. However, as we have shown, acceptable accuracy was reached by using the method outlined in this paper. These heuristics and method are unprecedented in literature and enable studies on emissions where ship type is a factor to be conducted without the use of commercial vessel databases.

REFERENCES


