Abstract
The Santos Bay, located in Brazil, holds the largest and busiest container port in Latin America. Many cargo ships pass through this Bay every day and its safety are subject of many researches. Using the Automatic Identification System (AIS) messages sent by those ships it is possible to detect possible near miss collisions events. This study develops an analytical approach to identify the most common cargo ship types involved in near miss collisions in Santos Bay (Brazil) using AIS database. It will also identify the most common type of vessel conflict in this area. The model is applied to rank the severity of an encounter between two vessels based on vessel conflict ranking operator (VCRO). The vessel size and the minimum distance to collision (MDTC) concept are considered in the model. The results show that the proposed methodology is adequate to identify various statistics in near miss collisions. Containers ship is the most common type of cargo ship involved on those situations and the crossing conflict occurs with higher frequency. Understanding the parameters involved in near miss collision around areas of high maritime traffic is important to avoid accidents. For the future, comparisons of data between different areas are suggested.

1. Introduction
Maritime transportation is crucial to the world’s economy. Also, it is important for the development of many countries. In 2017, the maritime transportation was responsible for more than 80% of the global trade between all other transportation modes. As well, cargo ships are the most common type of ship, approximately 90% of the world fleet being cargo ships, (UNCTAD, 2017). Accidents involving ships occasionally occur in the world. The most common types are grounding, collisions and fires (Guedes Soares & Teixeira, 2001). Since the ship itself and the cargo have a high value and there are human lives in risk in those accidents, it is very important to study the conditions that generate those accidents. With the results of those observations it is possible to improve the navigation safety.

Focusing on the accidents involving two or more ships, it can take place in every route where two ships cross each other. The port areas have a highest traffic of ships, so the observation of accidents in those areas has even more importance. Also, passengers’ ships are very present on port areas, which increases even more the interest of improving safety in this matter. Santos Port, located in Santos Bay, is the main Brazilian port and the biggest on the Latin America. It has a high diversity of products and services, including dry bulk, liquid bulk, containers, general cargo and passengers. Even having different types of ships for different purposes, cargos ships are the main type navigating in Santos Bay.
In 2018, there were more than 750 accidents involving navigation in Brazil and 185 people died by reason of those accidents. The accidents
statistics in Brazil are divided by Naval Districts (DN). The eighth DN is formed by the state of São Paulo (where the Santos Port is located) and the state of Paraná. Moreover, most part of deaths occurred by those navigation accidents took place in the eighth DN, being 43 deaths in this DN, (DPC, 2019).

For preventing accidents, ships have the Automatic Identification System (AIS). AIS is a mandatory automatic tracking system used on ships and by Vessel Traffic Services (VTS) for identifying and locating vessels by electronically exchanging data with other nearby vessels throughout the world. AIS system required to be installed on ships by the International Maritime Organization (IMO) and the Maritime Safety Administration in several countries, (IMO, 2003).

Since the AIS send messages automatically, the frequency of messages is very high. An AIS receiver in a port area, where there is a high traffic of ships, can receive more than 100 thousand of messages per day. If this receiver tries to store all those messages or if it tries to compare messages of different periods, it will quickly escalate to a problem of Big Data.

As Big Data is a popular subject in the last 15 years, there are many studies using AIS data with Big Data approaches to extract knowledge from those data. For instance, (Zhou, et al., 2019) use AIS data to identify patterns of routes classifying them by the length and beam of the ship. Also, (Cepeda, et al., 2018) use the AIS data to estimate ship emissions for the Port of Rio de Janeiro. As well, (Cepeda, et al., 2017) use the AIS as input for a simulation of offshore platforms.

The near miss collision is a way to identify ships that are close enough that is possible that an accident occurs. In most part of those events, the accident does not occur. Still, the studies on the detection of the near miss collision make possible the improvement of safety. Studies on this subject started recently with (Debnath & Chin, 2010). The near miss collisions studies in the literature are mostly located in Europe and Asia, creating opportunities for studies in other areas such as South America.

In the previous paper (Cepeda, et al., 2018), the AIS data was used to estimate possible near miss collisions in the area of Rio de Janeiro port. It analyzed different types of ships and comparing them by size. The authors developed a model to rank the severity of an encounter between two vessels based on vessel conflict ranking operator (VCRO). The concept of minimum distance to collision (MDTC) was also considered on the model.

This present paper continues the last work, now applying the methodology in the Santos port area and focus only on cargo ships present on this area. This first section is a general introduction on the subject. The second section explains the data and the methodology used for this search. The third section brings the results from the search. And finally, a conclusion is presented.

2. Data and Methodology

In this paper, AIS data is used to develop an analytical approach to estimate possible near miss ship collisions. In this section, the data and the methodology applied on this data are presented.

2.1 Data

The Automatic Identification System (AIS) is a mandatory collision avoidance system required to be installed on ships by the IMO and the Maritime Safety Administration in several countries. The AIS system makes it possible to locate most vessels throughout the world.

International voyaging ships with a Gross Tonnage (GT) of 300 or more, passenger ships of all sizes, domestic vessels with a GT of 200 or more traveling in coastal waters, and inland ships with a GT of 100 or more, are all required to be equipped with AIS. Special purpose vessels such as military ships, fishery ships, sports ships, and public service ships are exceptions, (Chen, et al., 2018), and (IMO, 2003).

The reported AIS data can be divided into static, dynamic, and voyage-related data categories. Static information includes ship name, ship type, length, breadth, etc. Dynamic data includes ship speed over ground, navigational status (operating mode), heading, rate of turn, position, etc. Voyage-related data includes current draught, description of cargo, and destination, (IMO, 2003). Besides ship information reported by AIS, detailed data for ship type, ship size, date of construction, design speed, gross tonnage and power of the engines can be obtained from the other databases as Marine traffic or IHS Maritime, see Figure 1.

The database used on this paper is the data originating from the AIS messages generate by ships passing by on the region of Santos Bay and Rio de Janeiro Bay. Since the AIS messages gives the longitude and latitude position of the antenna, it is easy to identify ships on the Santos Bay area. This database contained data from the period of January 2018 to April 2018 and from the period of August 2018 to April 2019. Those two periods together give a ten months database.
Furthermore, in this paper we will analyze only the near miss collisions involving cargo ships. This filtering is done after finding all near miss collision on the area. Because there can be cargo ships colliding with non-cargo ships, thus it is not possible to eliminate the messages coming from non-cargo ships before the analyses of near miss collision. Also, since the ship types provide directly from IHS (IHS Markit) database.

The invasion of ellipse is calculated by the Equation(2).

$$D_{ij} = \sqrt{(x_i^t - x_j^t)^2 + (y_i^t - y_j^t)^2}$$  \hspace{1cm} (1)

The approach presented in this paper to estimate the vessel collision risk based on AIS data is based on the domain theory presented by (Weng, et al., 2012) and (Weng , et al., 2012) proposed. At first, the relative speed of vessel i over vessel j denoted as $\vec{v}_{ij}^t$ can be determined according to Equation(3), (Wu, et al., 2016).

$$|\vec{v}_{ij}^t| = \sqrt{(\vec{v}_{ij}^t \cos \alpha_i^t - \vec{v}_{ij}^t \cos \alpha_j^t)^2 + (\vec{v}_{ij}^t \sin \alpha_i^t - \vec{v}_{ij}^t \sin \alpha_j^t)^2} \leq 1$$ \hspace{1cm} (2)

And then the relative angle from $\vec{v}_{ij}^t$ to $\vec{v}_{ij}^t$ is in Equation(4).

$$\theta_{ij} = \cos^{-1} \left[ 1 - \frac{(\vec{v}_{ij}^t \sin \alpha_i^t - \vec{v}_{ij}^t \sin \alpha_j^t)^2 + (\vec{v}_{ij}^t \cos \alpha_i^t - \vec{v}_{ij}^t \cos \alpha_j^t)^2}{|\vec{v}_{ij}^t|^2} \right]$$ \hspace{1cm} (4)

Then the following criteria can be used to categorize the type of vessel conflict:

- **Overtaking conflict** if $\sin \alpha_i^t \cdot \sin \alpha_j^t > 0$ and $(\theta_{ij}^t \leq 10^\circ$ or $\theta_{ij}^t \geq 170^\circ$)

- **Head-on conflict** if $\sin \alpha_i^t \cdot \sin \alpha_j^t < 0$ and $(\theta_{ij}^t \leq 10^\circ$ or $\theta_{ij}^t \geq 170^\circ$)

- **Crossing conflict** if $\theta_{ij}^t > 10^\circ$ and $\theta_{ij}^t < 170^\circ$

Where $\alpha_i^t$ and $\alpha_j^t$ are courses of vessels i and j at time t, respectively, and $\theta_{ij}^t$ is the relative angle at time t. If one vessel is already within the domain of another at time t, a vessel conflict does happen.

The above rules can be also applied to elliptical domains.

Figure 2 shows these two types of domains. It also demonstrates the speed, course, and domain of vessel. It seems that elliptical domains focus more on conflicts between vessels which share similar or opposite courses, (Min Mou, et al., 2010).
2.3 Data Filtering

This paper aims to find situations of near miss collisions for cargo ship. Since for each collision there are two ships involved, a filtering in the data needs to be done for eliminate situations where two non-cargo ships are in situation of near miss collision. Each instance of near miss collision in the database stores MMSI code, type of ship and other information for both ships (Ship1 and Ship2) involved in the near miss collision. It is important to reemphasis that two distinct situations can happen: a cargo ship is near colliding with a non-cargo ship or cargo shipping near colliding with another cargo ship.

In addition, there are lots of types of cargo ships described in AIS. For better understand the results, it is recommended to group certain types of cargo ships in a same type, instead of having lots of different types.

As well as the AIS does not give all the information for every message. Sometimes, the type of the ship is not declared (or declared as NULL) or even the type of ship described is too general for classing. To solve this last problem, anotherdatabase was also consulted to verify if the ship type given by AIS was the true type. This other database is called IHS (IHS Markit).

First step to start filtering is selecting only near miss collisions presented in Santos Bay area. As mentioned, the database contains information about Santos Bay and Rio de Janeiro Bay. Each instance of near miss collision in Rio de Janeiro Bay is excluded.

Second step is eliminating Tug and Pilot boats. Considering that those types of ship are made for contact with another ship, it will be often in a near miss collision situation. So, if an instance of near miss collision that has a tug or pilot as the Ship1 or as the Ship2, this instance is excluded.

After those two steps, it is important to define which ship types are included on the analysis. With this purpose, all the ship types remaining at this point were observed. They are listed in Table 1.

The table separates the types in 3 different categories. The first is the cargo ship types that are well defined. The second is cargo ship types that are not well defined or types that can be cargo or not, and need to be verified with IHS. The third is ship types that are not cargo.

<table>
<thead>
<tr>
<th>Cargo Ships</th>
<th>Possible Cargo or not</th>
<th>Non-cargo Ships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk Carrier</td>
<td>Cargo Hazard A</td>
<td>Fishing</td>
</tr>
<tr>
<td>Chemical Tanker</td>
<td>Cargo Hazard A (major)</td>
<td>Hopper Barge</td>
</tr>
<tr>
<td>Container ship</td>
<td>Inland Unknown</td>
<td>Hopper Dredger</td>
</tr>
<tr>
<td>Crude Oil Tanker</td>
<td>NULL</td>
<td>Military Ops</td>
</tr>
<tr>
<td>Fruit Juice Tanker</td>
<td>Other</td>
<td>Offshore Supply Ship</td>
</tr>
<tr>
<td>General Cargo</td>
<td>Unspecified</td>
<td>Passenger Ship</td>
</tr>
<tr>
<td>LPG Tanker</td>
<td></td>
<td>Research Vessel</td>
</tr>
<tr>
<td>Oil Products Tanker</td>
<td></td>
<td>Research/Survey Vessel</td>
</tr>
<tr>
<td>Oil/Chemical Tanker</td>
<td></td>
<td>Sailing Vessel</td>
</tr>
<tr>
<td>Ro-Ro Cargo</td>
<td></td>
<td>Tanker Barge</td>
</tr>
<tr>
<td>Ro-Ro/Container Carrier</td>
<td></td>
<td>Work Vessel</td>
</tr>
<tr>
<td>Vehicles Carrier</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next step is verifying in IHS database, what it mentions about the ships that are not well defined.
or not specified enough. For that, every ship in this list of types is compared using its MMSI code. Doing this is possible to see that for the type ‘Cargo Hazard A’, the IHS database gives the same answer. So, this type is not changed.

For the type ‘Cargo Hazard A (major)’, it is noted that is actually ‘Container Ship (Fully Cellular)’ according to IHS database. Thus, all ships with this type are changed for the new type.

Also, some ‘Inland Unknown’, ‘NULL’, ‘Other’ and ‘Unspecified’ are Tug ship according to IHS. In those cases, the instance of near miss collision containing a ship that is a Tug ship was excluded. The others ‘Inland Unknown’, ‘NULL’, ‘Other’ and ‘Unspecified’ that are not Tug, are varied types in the corresponding IHS database.

Still, some ships that have the ‘NULL’ type are not able to find in IHS database. Those ships are probably not listed with an IMO number. Since the instances of near miss collisions involving them are less than 5% of the total number of near miss collisions and that those ships could be Tug ships, it was chosen to exclude all instances that remain with the type of Ship1 or Ship2 as ‘NULL’.

After this comparison with IHS data, five new types were added. It is possible to see that the five new types are types that already exist, only with different names. They are:

- Chemical/Products Tanker
- Container Ship (Fully Cellular)
- Crude/ Oil Products Tanker
- General Cargo Ship
- Products Tanker

A total of 27 types of ships were left after this comparison, between cargo and non-cargo ship types. As mentioned, for better comprehension of the results, the types will be put together in few common types. Those common types were chosen based on the number of collisions of each type (considering its relevance for the analysis). The six resulting types chosen, as well as the correspondence with previous types are shown on Table 2.

### Table 2 – Correspondence between AIS/IHS ship types and chosen types

<table>
<thead>
<tr>
<th>Ship Type</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk Carrier</td>
<td>Bulk Carrier</td>
</tr>
<tr>
<td>Container ship</td>
<td>Container</td>
</tr>
<tr>
<td>Container Ship (Fully Cellular)</td>
<td>Container</td>
</tr>
<tr>
<td>General Cargo</td>
<td>General Cargo</td>
</tr>
<tr>
<td>General Cargo Ship</td>
<td>General Cargo</td>
</tr>
<tr>
<td>Chemical/Products Tanker</td>
<td>Tanker</td>
</tr>
</tbody>
</table>

3. Results

This section presents the results from using the data described in section 2.1 as input data on the model described in section 2.2 and after applying the filtering presented in section 2.3. The results only consider data from Santos Bay and for the period of ten months mentioned on section 2.1.

A first result interesting to analyze is the difference between the number of near miss collisions considering the Tug and Pilot ships (all AIS data), the number of near miss collisions after excluding the instances that contains Tug and Pilot ships, and the number of near miss collisions considering only instances involving cargo ships. This result is presented in Table 3.

### Table 3 – Number of near miss collisions in different phases of the analysis

<table>
<thead>
<tr>
<th>Situation</th>
<th># of Near-miss Collisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>All AIS data</td>
<td>16517</td>
</tr>
<tr>
<td>Without Tug and Pilot</td>
<td>1814</td>
</tr>
<tr>
<td>Cargo Ship involved</td>
<td>1344</td>
</tr>
</tbody>
</table>

As expected, the number of near miss collisions is drastically reduced when filtering Tug and Pilot ships. Tug and Pilot ships represent approximately 90% of all near miss collisions. That endorses the fact that those types of ships are constantly in a near miss collision situation given its main purpose. And that is why it is a good choice to exclude those instances.

When comparing the two last results from Table 3, it is possible to see that most part of near miss collisions involves Cargo ships. Almost 75% of near miss collisions are involving cargo ships. This result shows the importance of cargo ships for the area studied. As mentioned, most part of ships in Santos
Bay are cargo ships. Also, since cargo ships tends to be bigger than the other types of ships, it is understandable, following the model of the section 2.2, that they have a bigger Elliptical Domain and are more susceptible to suffer near miss collision situation.

Table 4 – Number and percentage of collisions per ship type

<table>
<thead>
<tr>
<th>Type of Ship</th>
<th>Collisions</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk</td>
<td>520</td>
<td>39%</td>
</tr>
<tr>
<td>Container</td>
<td>781</td>
<td>58%</td>
</tr>
<tr>
<td>General Cargo</td>
<td>102</td>
<td>8%</td>
</tr>
<tr>
<td>Tanker</td>
<td>329</td>
<td>24%</td>
</tr>
<tr>
<td>Other Cargo</td>
<td>106</td>
<td>8%</td>
</tr>
<tr>
<td>Non-Cargo</td>
<td>527</td>
<td>39%</td>
</tr>
</tbody>
</table>

The next result to be presented is the number of collisions per type of cargo ships. This result shows which types of cargo ships are currently more involved in near miss collision situation. Table 4 brings the numbers and Figure 3 represents the percentages.

In this result is important to emphasize that, since there are two ships in each collision (Ship1 and Ship2), it is normal that the sum of all percentage is more than 100% and less than 200%. It is more than 100% because both type of Ship1 and type of Ship2 is counted for each single instance, and that also explain the maximum value of 200%. But it is less than 200% since there are some instances of near miss collision where type of Ship1 and type of Ship2 is the same, so in this case is only counted as one collision for the ships type.

![Figure 3 - Percentage of near miss collisions for each type of ship](image)

As it can be seen in Figure 3, container ship is the most common type of ship involved in near miss collisions. It shows that in more than half of instances studied a container ship is presented. Thus, it can be observed that container ship is the most common type of cargo ship on this area and probably the most relevant cargo ship.

Another observation that can be made by this result is that non-cargo ships are highly involved on near miss collisions with cargo ships. It is an also understandable conclusion, since there a lot of ship types there are aggregated in Non-Cargo type. Thus, it will be many cargo ships in a situation of near miss collision with non-cargo ships.

Table 5 – Number of near miss collision per type of collision and per ship type

<table>
<thead>
<tr>
<th>Type of Ship</th>
<th>Crossing</th>
<th>Overtaking</th>
<th>Head-on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk</td>
<td>509 (98%)</td>
<td>5 (1%)</td>
<td>6 (1%)</td>
</tr>
<tr>
<td>Container</td>
<td>757 (97%)</td>
<td>12 (2%)</td>
<td>12 (2%)</td>
</tr>
<tr>
<td>General Cargo</td>
<td>100 (98%)</td>
<td>2 (2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Tanker</td>
<td>309 (94%)</td>
<td>14 (4%)</td>
<td>6 (2%)</td>
</tr>
<tr>
<td>Other Cargo</td>
<td>104 (98%)</td>
<td>1 (1%)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>Non-cargo</td>
<td>512 (97%)</td>
<td>11 (2%)</td>
<td>4 (1%)</td>
</tr>
</tbody>
</table>

The other types follow the same logic, been Bulk ships more often in Santos Bay than Tankers and so on.

Last result to be presented is the number of types of near miss collisions, Crossing, Overtaking or Head-on as described in section 2.2. In that way it is possible to see which type of near miss collisions should be avoided at most, see Figure 4.

Table 5 brings the number of collisions in each type of collision divided by ship types, and the percentage of the type of collision for each type of ship.

It is clear to see that for all near miss collisions, independent of the ship type, the Crossing type of collision is the most common, being the collision type of at least 94% of collisions for every ship type. The other two collision types are rare and occur on maximum of 5% of collisions.

![Figure 4 - Percentage of Collision Type for each ship type](image)

This is the main information retrieved from this result. Since it does not depend on ship type, it is better to analyze it for all type of ships together, see Figure 5.

As it can be seen, 97% of all near miss collisions involving cargo ships are a crossing type collision. This result is evident if observed with the geography of Santos Bay, illustrated in Figure 1. Santos Bay has a tight channel that is the access to
the Santos Port. Most of ships goes through this channel to access the port and goes out also by this channel.

![Percentage of Collision Type for all ship types summed](image)

Figure 5 - Percentage of Collision Type for all ship types summed

Since the channel is narrow, it is deduced that few routes are possible to make. And the inner routes are probably very similar to outer routes, only changing the sense. In that case, it will be common that ships going to the port cross with ships leaving the port. And if this cross situation is close enough to characterize a near miss collision situation, it will be a crossing type.

4. Conclusions

The focus of the study is the identification of near miss collision for cargo ship in the Santos Bay for 10 months (January 2018 to April 2018 and August 2018 to April 2019).

After analyzing the results, some conclusions can be made. First is that the cargo ships are very important for Santos Port, but as well are very susceptible to near miss collisions situations, since almost 75% of collisions involve a cargo ship.

Seeing the six different ship types described by this paper is possible to determine that container ship is the most relevant, because more than half of near miss collision situations involve a container ship.

And the last analysis shows that the type of collision is independent of the ship type. The crossing type of collision is the most common for every ship type, representing 97% of all collisions. These results can aid navigation authorities to better understand what causes incidents on the Santos Bay and it also can be used by the local pilots to improve their routes.

The quality of the database provided AIS is still to be improved. There are some missing values and values that are not always accurate. For instance the ship type is not always present or accurate. That turns more difficult to make analyses using the type of the ship. The ideal is to compare with another more reliable database, like IHS, and choose the most accurate result.

For future researches on the area, would be interesting to compare those results with other region to see if there is a correlation. Also, would be interesting to see if the geography of the port has indeed a major influence on the type of collision as concluded in Santos Bay.

In an operational perspective, the obtained results could be used to create an onboard system to alert ships about the most common near miss collision situations for the specific region where it is passing by. This system could also be used by port administrations to improve the logistic and avoid the near miss collision situations.

5. Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The present work was realized with the support of CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil.

6. References


DPC, D. d. P. e. C., 2019. *Inquérito Administrativos sobre Acidentes e Fatos da Navegação (IAFNs), s.l.: s.n.*


