Scientific Journals
of the Maritime University of Szczecin

Zeszyty Naukowe
Akademii Morskiej w Szczecinie

2015, 42 (114), 86–92
ISSN 1733-8670 (Printed)
ISSN 2392-0378 (Online)

Vessel classification method based on vessel behavior in the port of Rotterdam

Yang Zhou[1], Winnie Daamen[2], Tiedo Vellinga[1], Serge Hoogendoorn[2]

Delft University of Technology, Faculty of Civil Engineering and Geosciences, Delft, The Netherlands
[1] Department of Hydraulic Engineering, e-mails: {y.zhou-5; t.vellinga}@tudelft.nl
[2] Department of Transport & Planning, e-mails: {w.daamen; s.p.hoogendoorn}@tudelft.nl

Key words: AIS, data analysis, vessel classification, vessel behavior, port, classification method

Abstract

AIS (Automatic Identification System) data have proven to be a valuable source to investigate vessel behavior. The analysis of AIS data provides a possibility to recognize vessel behavior patterns in a waterway area. Furthermore, AIS data can be used to classify vessel behavior into several categories. The analysis results would help the port authority and other equivalent parties in port design and optimization or marine traffic management. For researchers, it provides a systematic way to understand, simulate and predict vessel behavior. This paper focuses on vessel classification in the Botlek area, Rotterdam from the perspective of vessel behavior.

In this paper, the vessel properties, including vessel type, GT (Gross Tonnage), length and beam, have been analyzed to investigate the vessel behavior, which is described by four factors including heading, COG (Course over Ground), SOG (Speed over Ground), and position. In order to discover the behavior patterns in normal situations, several thresholds are set in order to filter the collected AIS data to define such situations. By plotting the AIS data, behavioral changes with the changes of properties have been observed. Hence, the correlations between vessel behavior and different vessel properties are investigated. The results reveal that a vessel's sailing position and COG are both strongly determined by beam, while SOG is affected by GT. For the heading of a vessel, no obvious correlation with any vessel property is found. Each behavioral factor is clustered according to the correlated vessel property. This way, the criteria to classify the vessels are determined. The vessel classification results based on their behavior would likely to lead to more consistency in the analysis, simulation and prediction of the vessel behavior. The reason is that the development of such a simulation model is based on a systematic recognition of the vessel behavior patterns.

Introduction

Waterborne transport is becoming an increasingly important means of cargo transportation and is recognized as the most economical method to move cargo around the world. Due to the large amount of cargo carried by each vessel and the large vessel visit frequency to the hub ports, the safety of vessels and the capacity of ports have been two significant issues for marine traffic. The increasing focus on traffic management, and the predictive waterway and port design are all issues that require a better understanding of vessel behavior, especially their sailing behavior in port areas. Every single vessel sailing in a port will behave differently to some degree. The reason is that, apart from the influence of human factors caused by the officers on board, the sailing situation cannot be exactly the same in two voyages regarding the external conditions, e.g. wind, current, visibility, encounters with other vessels. Still, some general behavior patterns possibly exist in the waterway area. This implies that the vessels can be classified into several categories depending on their revealed behavior. Based on this systematic understanding of marine traffic operations, the possible pattern a vessel will behave in could be identified when she intends to sail in this area. Investigation of vessel behavior requires a set of detailed data with respect to vessels' inherent properties and dynamic paths, which could be provided by AIS (Automatic Identification System).

The requirement to install AIS on all passenger vessels and sea-going vessels larger than 300 GT (Gross Tonnage) was adopted by IMO (International Maritime Organization) and enforced from July 1st 2002. In current practice, almost every sea-going vessel, even under the GT limit, has AIS on board for sailing use. Three types of data are available in an AIS message: static information, dynamic information and voyage-related information. The static information, related to the vessel's properties (e.g. vessel type, GT, length) is input when the AIS system is installed on board. The dynamic information, including sailing position, SOG, COG, etc., is updated automatically during the voyage at a certain time interval. The voyage-related information, e.g. vessel draught, type of cargo, is manually inputted and updated through the voyage.

To the best of our knowledge, there is no specific research on the methodology of vessel classification. The current vessel classification is defined differently by different ship classification societies. The vessel classification in a port, a water area, or a country is determined by the local port authority or other equivalent parties. However, such vessel classifications were provided as guidelines or rules without a detailed classification basis explained.

AIS data have proven to be a valuable source to investigate vessel behavior. Many researchers use the AIS data for general behavior pattern recognition and anomaly detection, without a classification of the involved vessels (Aarsæther and Moan, 2009; Pallotta et al., 2013; Ristic et al., 2008; Zhu, 2011). They regard the vessel behavior in the research area as a whole, ignoring the differences between vessels. However, such behavior differences do exist in any waterway area, which has been revealed in some research. In most studies, a vessel classification is defined before analyzing the AIS data. Goerlandt and Kujala (2011), Silveira et al. (2013), Mascaro et al. (2010) classify the vessels only based on their types, despite the size range of vessels. For the purpose of vessel behavior investigation, this vessel classification method still lacks a detailed description of the behavior differences within the same type of vessels. De Boer (2010) proves the behavior differences of container vessels with different DWT (Deadweight Tonnage). The vessels are classified to ensure approximately the same amount of available tracks in every dataset. This classification method only reveals the behavior changes with the change of DWT, without a clear recognition of the different behavior patterns. The same drawback applies to the research of Shu et al. (2013) and Xiao (2014). They classify the vessels according to their GT. Their classification method is based on the presence frequencies of vessels with different GT. It is a feasible method to classify vessels on the assumption that the dataset is sufficiently large to represent the vessels in the waterway area.

In current studies of vessel behavior, the vessel classification is defined before analyzing the behavior patterns. Or in some cases, vessels are not classified for general behavior recognition. Therefore, the results of the analysis only explain differences in vessel behavior when the classification criterion, e.g. DWT or GT, differs. The criterion is determined subjectively by the researcher or due to the limitation of collected data. The actual behavior patterns of different vessels cannot yet be revealed.

In this paper, a vessel classification method is developed based on the analysis of normal vessel behavior in the Botlek area, Rotterdam, based on AIS data. The vessels' inherent properties, including vessel type, GT, length and beam, are analyzed to classify vessels. In the classification results, vessels in the same group behave in a similar pattern described by heading, COG, SOG and position. In the section *Data description*, the dataset used in this paper is introduced. The section *Data analysis methodology* explains the data analysis methodology for the data filter, correlation investigation, and behavior clustering. The AIS data analysis results are presented in the section *Data analysis results*. We end up with conclusions and recommendations for future research in the section *Conclusions and recommendations*.

Data description

The AIS data from January 2009 to April 2011 was collected from MARIN (Maritime Research Institute, the Netherlands). The research area is located in the Botlek area in the port of Rotterdam, in which the traffic density is high. In this area there is one main waterway, named the Nieuwe Maas, flowing from east to west. The position of the cross-section to collect the AIS data is shown as a red dashed line in Figure 1.

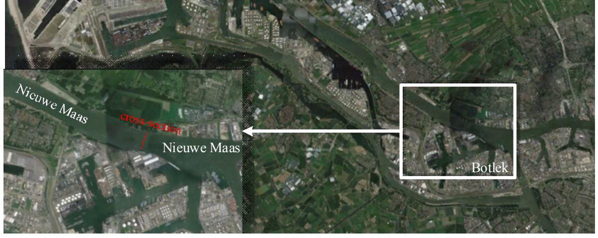

Figure 1. Research area layout (Map source: Google Earth)

If a vessel transmits an AIS signal when passing the cross-section, this data will be stored. If a vessel does not transmit any signal at the time of passing the cross-section, the dynamic sailing information, including position, SOG, COG and heading, is interpolated by the last record before and the first record after the cross-section.

Both endpoints of the cross-section are located at the 5-meter depth contours on two sides of the waterway. The aids to navigation are also set along the contours to indicate the shallow waters in this area. The direction of the cross-section is approximately perpendicular to the traffic flow.

In the collected datasets, the following AIS data information is included.

- Static information: vessel ID (to identify vessels instead of the exact vessel name); vessel type; length; beam; GT.
- Dynamic information: vessel's position with coordinates in decimal degrees; UTC of record; SOG; COG; heading.

Since the coordinates are in decimal degrees, the difference between different vessels is small. Hence, all the coordinates present in this paper have been transferred to one uniform coordinates system.

The collected AIS dataset is combined with the data of wind and visibility to describe the vessel behavior under these external conditions. However, in order to develop the vessel classification method based on the behavior patterns in normal sailing situations, the prevailing conditions need to be defined. The detailed thresholds to define normal situations are described in the section *Data analysis methodology*.

Data analysis methodology

The proposed methodology in this paper starts with the data filter to form a dataset only containing vessel behavior in normal situations. The correlation between each vessel property and behavioral factor is calculated based on the AIS data. The vessel property bearing a strong correlation with the behavioral factor could be used as the criterion to cluster the corresponding behavioral factor by significance test. The behavior with significant differences is clustered. Finally, the thresholds of each classification criterion are determined by the behavior clusters.

Data filter

In order to classify the vessels based on their general behavior pattern in normal situations, the thresholds to eliminate abnormal situations have been determined, including wind, visibility and encounters. In the dataset, the vessels sailing in the situations with at least one of these factors exceeding a threshold value have been removed. In addition, the messages with data error have also been filtered. The principles to filter data are explained as follows.

- Visibility

According to Shu et al. (2013), restricted visibility leads to an impact on vessel speed and path. Thus, a threshold of visibility should be established to identify the normal situation in the research area. The port authority of Rotterdam provides and updates the general hydrological and meteorological information based on statistical data. According to the 'HydroMeteoBundel' by Port Authority (2012), the statistical data of visibility in the research area is shown as 'Botlek' in Figure 2.

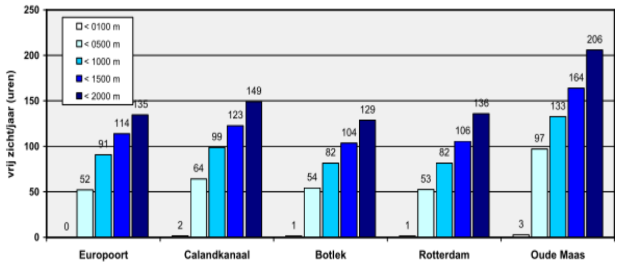

Figure 2. Statistical visibility data per year from 2006 to 2011 (Port Authority, 2012)

Compared to other areas in the port of Rotterdam, the visibility conditions in the Botlek area are slightly better. Visibility distance of less than 2000 meters occurs for about 129 hours per year. Therefore, visibility of more than 2000 meters is considered as the normal situation in the research area. The AIS data of vessels sailing in visibility of less than 2000 meters have been filtered.

- Wind

Wherever a vessel sails, there always exists the impact of the wind on her behavior, i.e. such an impact cannot be eliminated totally. However, each area has its own meteorological characteristics based on statistical data over a long time period. This way, the vessel sailing under the prevailing wind condition could be considered as the normal situation.

According to Port Authority (2012), the information for the research area can be represented by Geulhaven. The wind rose is shown in the left hand side of Figure 3, with the appearance frequency of each Bft (Beaufort) level of wind speed on the right side.

As can be seen in Figure 3, the prevailing wind direction is from the Southwest (ZW in Dutch). The area from South to West (the downward left part of the wind rose) occupies most of the appearance frequency. Therefore, the wind direction from South to West (from 180° to 270°) is set as the normal situation in this area.

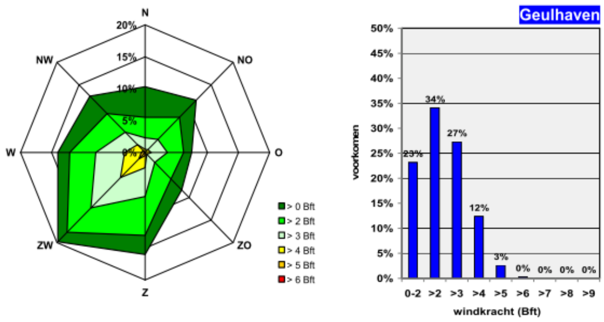

Figure 3. Wind rose in the research area (Port Authority, 2012)

The right hand side of Figure 3 shows the statistical data of the wind speed. A wind speed lower than 4 Bft is recorded 84% of the time. Afterwards, there is a big drop in the frequency of wind speed higher than 4 Bft, from 27% to 12%. Thus a wind speed less than 4 Bft (7.9 m/s) is set as the threshold to represent the normal condition of wind speed in this area.

- Data error

De Boer (2010) presented an overview of the current analysis on AIS data accuracy and error. In the static data, the main source of errors was the vessel type, which was vague or even incorrect. In the voyage-related data, most errors existed in destination and ETA. In the dynamic data, the error of navigational status was the only message analyzed.

Based on our dataset in this research, the error in the collected AIS data cannot be revealed. However, some information of vessel properties was missing, e.g. vessel type, length, beam, etc. The data items with missing information will lead to a decrease in the accuracy of the analysis result. The reason is that if some property datum is null, the correlation between behavioral factors and this property cannot be calculated. This way, the correlation result of this property is incorrect. Therefore, in this paper, the data items with incomplete information are all filtered out from the dataset. The accuracy of the classification results can be guaranteed based on the AIS data accuracy, while the inaccuracy due to data processing during classification can be avoided.

- Encounters

When a vessel is in an encountering situation with another vessel, her behavior will be changed if she is the give-way vessel or due to the good seamanship of the officer on watch. Thus, the vessel behavior during an encounter (from the formation of an encounter to a safe and clear passage) cannot be taken as the normal behavior when analyzing the general behavior pattern in a waterway area.

The impacts of encounters on vessel behavior is not the research objective, thus the different types of encounters are not identified in this paper. As indicated by Fujii and Tanaka (1971), the ship domain in a waterway area in port can be seen as an ellipse. The length of the major axis is 6 times of the vessel length, and the minor axis is 1.6 times of the vessel length. In this paper, when the ship domains of two vessels overlap, it is considered as an encounter. The AIS data during an encounter can be filtered by calculating the distance between two vessels.

Based on the above explained principles, the dataset for vessel classification analysis was filtered and formed.

Correlation investigation

The static data from AIS should provide the vessel's property information including vessel type, length, beam, GT, and location of the electronic position fixing antenna. The voyage-related information includes the draught of the vessel. However, during the data collection, information of draught and position of antenna were not selected. Therefore, the dynamic position data from AIS cannot be revised for better accuracy without information on the antenna, and draught cannot analyzed as the vessel classification criterion. In this paper, the vessel properties of vessel type, length, beam and GT are analyzed to classify vessels. In future research, draught will be further analyzed by the same methodology.

The dynamic information from AIS data can fully describe the vessel behavior in a two-dimensional space, in which the detailed vessel maneuvering behavior in 6 degrees of freedom is not considered.

In the current vessel classification methods, the criterion is different for different researchers, which can be GT, DWT, length, etc. No research has proven the criterion adopted to classify vessels is the most appropriate one for its research purpose. The class of vessels is not necessarily described by a parameter range of a single vessel property, but a combination of parameter ranges of several vessel properties. The revealed behavior analysis results explain that the vessel behavior mostly varies when some vessel property differs, but the vessels in the same group do not mean to behave in a similar pattern. Hence, this paper is inspired by the variation trend of behavior with the change (-s) of vessel property (-ties). The AIS data in this paper also present some kind of trend to some vessel properties. Figures 4 and 5 show the vessel behavior data when GT or beam changes, respectively.

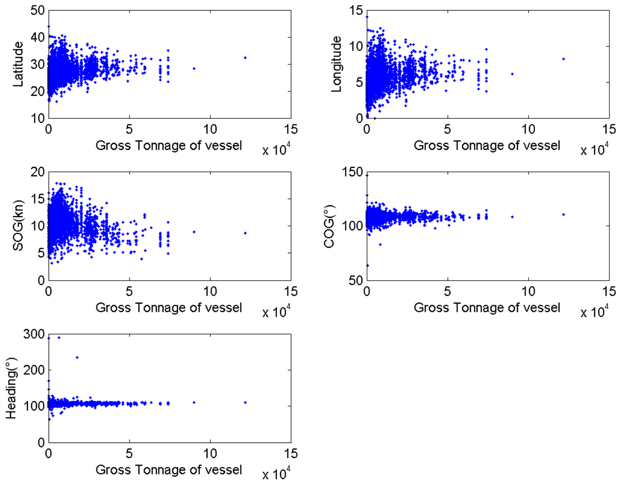

Figure 4. Vessel behavior with change in GT

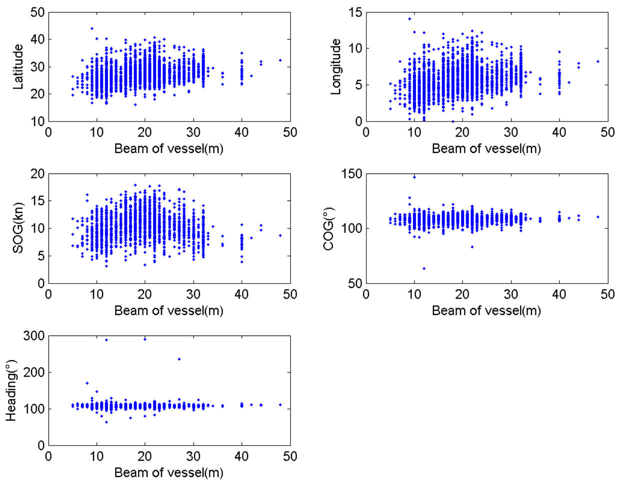

Figure 5. Vessel behavior with change in beam

The plotting of vessel behavior to vessel property cannot provide a clear quantitative relationship between them. Therefore, the correlation between each behavior element and each property should be investigated in a mathematical way in pairs. The pairs with strong correlation indicated by correlation coefficients will be discovered for the further behavior characterization.

In order to increase the practicality of the classification result, the port authority or other equivalent parties should be able to choose the criterion or a combination of criteria that best fits their classification request. Or in some cases, they hope to adopt a simpler classification method with fewer groups as the result, even if there is a decrease of the pattern recognition accuracy. Therefore, it is not only the property with the strongest correlation with behavioral factor that should be identified, a series of vessel properties with adoptable correlation relationship are required, if possible. This way there is more flexibility in the analysis result with indication of the accuracy by correlation coefficient.

Behavior clustering

The properties with correlation to each behavioral factor have been determined in the previous research. In this section, the method to decide the thresholds of vessel classification for each criterion is explained.

If there exists an adoptable correlation between a behavior element and a vessel property, the behavior should be continuously changed with the variation of vessel property. It implies that, for each behavioral factor, if two vessels with one identical or similar property behave in a similar pattern, they should be classified into one group.

The AIS data are sorted according to the value of vessel property. Since the values of properties are not continuous, the dataset is divided into a series of sub-datasets containing an undefined number of data at the same interval of value. The dataset division thresholds are listed in Table 1.

Table 1. Initial dataset division for behavior clustering

Vessel property	Beam (meters)	Length (meters)	GT
1st sub-dataset	(0, 5]	(0, 20]	(0, 500]
Sub-dataset at the same intervals	From 5 to 40 at the interval of 5	From 20 to 300 at the interval of 10	From 500 to 10 000 at the interval of 500
			From 10 000 to 30 000 at the interval of 1000
			From 30 000 to 60 000 at the interval of 5000
Last sub-dataset	(40, ∞)	(300, ∞)	(60 000, ∞)

The statistical test of significance is performed to the adjacent sub-datasets, starting from the 1st sub-dataset and the 2nd sub-dataset. In this paper, p-value at 0.05 is taken as the criterion to decide the significance level.

- If there is no significant difference between these two sets of data, the behavior is considered as in the same pattern. These two datasets will be clustered to be tested with the next sub-dataset.
- If there is significant difference between these two sets of data, the vessels in these two sets should be classified into different groups. Thus the previous cluster of this specific behavioral factor is finished. The latter sub-dataset will be tested with the next sub-dataset.

Through this process, the data in a similar pattern of each behavioral factor are clustered. The thresholds of the corresponding properties in each behavioral factor cluster are set as the criteria to classify vessels. So far, the method of vessel classification based on their behavior is developed.

Since the criteria (including type of property and the thresholds) to cluster vessels in each behavioral factor might be different, the criteria to determine the classification might be one vessel property in a parameter range or a combination of several properties in different parameter ranges. For any vessel, by checking her property (-ties), her behavior pattern in each behavioral factor can be recognized. The proposed vessel classification method is decided by the real-life behavior in the research area. In the classification result, each class of vessels will behave in a similar pattern in the normal situation.

Data analysis results

This section shows the data analysis results corresponding to the steps in the methodology.

Correlation between vessel behavior and vessel properties

The results of correlation coefficients between each behavior element and vessel property are shown in Table 2.

From Table 2, the properties with correlation to each behavioral factor are determined. Vessels classified by beam will sail in a similar position described by latitudes and longitudes. SOG is more likely to be influenced by GT, while COG is influenced by beam. To identify all the possible criteria to cluster vessel behavior, the choices of vessel properties to each behavioral factor are as follows. It needs to be noticed that the criteria are listed in the order of a decrease of correlation, which will result in a decrease of pattern recognition accuracy.

- Position: beam, length, GT;
- SOG: GT, length;
- COG: beam, length, GT.

In the results there is no property in significant correlation to heading. The possible reason is that the heading of a vessel is under the impacts of wind and current. The vessel alters her heading in different leeway and drift angles to keep her COG or path. Since the vessel behavior data analyzed in this paper are from the same cross-section, heading is more determined by the condition, to be more specific the direction, of wind and current in this area, rather than the properties of the vessel. Therefore, no property can be adopted as a criterion to cluster vessels' heading.

Vessel classification according to behavior pattern

According to the correlation results in the research area in this paper, to get the best behavior pattern recognition, beam should be chosen to classify position, while GT should be chosen for SOG. Through the significance tests, the vessel behavior clustering results are shown in Table 3.

For a specific vessel, her behavior pattern can be recognized based on the classification result. The beam decides the pattern of position and COG, and the GT decides the pattern of SOG.

Conclusions and recommendations

This paper presents an AIS data analysis to classify vessels based on their behavior in the Port of Rotterdam, the Netherlands. In this paper, vessel behavior is described by sailing position (with coordinates), SOG, COG and heading. The AIS dataset used for behavior pattern clustering has been filtered to eliminate the abnormal conditions of external factors. The principles and reasons are explained in the paper based on the meteorological conditions in the research area. The analysis results show that there is correlation between vessel behavior and vessel properties. More specifically, SOG is strongly correlated with GT, while COG and sailing position with beam. No correlation between heading and vessel properties is revealed. A possible reason is that the heading of a vessel is more influenced by the wind and current direction in the area than her properties. The behavior cluster results

Table 2. Correlation calculation results

Property	Item of result	Element of behavior				
		Lat	Lon	SOG	COG	heading
Type	Correlation coefficient	–0.111	–0.111	–0.333	–0.111	0.389
	Sig. (2-tailed)	0.677	0.677	0.211	0.677	0.144
Length	Correlation coefficient	0.620**	0.620**	–0.172*	0.136*	0.078
	Sig. (2-tailed)	0.000	0.000	0.011	0.045	0.254
Beam	Correlation coefficient	0.823**	0.823**	–0.235	0.729**	0.245
	Sig. (2-tailed)	0.000	0.000	0.173	0.000	0.156
GT	Correlation coefficient	0.395**	0.395**	–0.189**	0.074*	0.027
	Sig. (2-tailed)	0.000	0.000	0.000	0.037	0.443

*: Correlation is significant at the 0.05 level (2-tailed).
**: Correlation is significant at the 0.01 level (2-tailed).

Table 3. Vesssel classification results

Cluster No.	Position		COG		SOG	
	Thresholds of beam (in meter)	Number of vessels	Thresholds of beam (in meter)	Number of vessels	Thresholds of GT	Number of vessels
1	(0, 15]	916	(0, 15]	916	(0, 500]	203
2	(15, 20]	973	(15, 25]	2173	(500, 1500]	210
3	(20, 30]	1525	(25, 30]	325	(1500, 2000]	170
4	(30, 35]	189	(30, ∞)	219	(2000, 2500]	190
5	(35, ∞)	30			(2500, 3000]	319
6					(3000, 3500]	37
7					(3500, 6000]	431
8					(6000, 8000]	829
9					(8000, 12000]	546
10					(12000, 18000]	219
11					(18000, 20000]	21
12					(20000, 22000]	49
13					(22000, 23000]	23
14					(23000, 24000]	24
15					(24000, 26000]	67
16					(26000, 28000]	67
17					(28000, 40000]	145
18					(40000, ∞)	83

prove that the vessel's behavior can be classified based on a certain criterion of vessel property, respectively. In this paper, the sailing position is classified into 5 groups based on beam; COG is classified into 4 groups based on beam; SOG of vessels is classified into 18 groups based on GT.

In this paper, a new vessel classification method is proposed totally based on the revealed behavior analysis. The classification result helps to recognize the vessel behavior patterns. However, AIS data collected from only one cross-section is analyzed. The results cannot generally represent the behavior patterns in the area. Besides, the pattern of sailing paths has not been investigated. In future research, the vessel behavior in time series will be further analyzed to describe the whole sailing path in a waterway area. Due to the information limitations of the collected data, the properties of DWT and draught have not been analyzed. An AIS data analysis will be performed in a newly collected dataset with complete vessel property information in another area.

The analysis results considering vessel behavior variation in time series based on a dataset with complete information will reveal more correlation relationships between vessel behavior and vessel properties. Thus, the vessel classification result will lead to a better recognition of vessel behavior patterns.

Acknowledgments

This research is financially supported by China Scholarship Council and Delft University of Technology with the Grant number 201306950015. MARIN is acknowledged for the access and usage of the AIS database.

References

1. De Boer, T. (2010) *An analysis of vessel behaviour based on AIS data*. TU Delft, Delft University of Technology.
2. Fujii, Y. & Tanaka, K. (1971) Traffic capacity. *Journal of Navigation*. 24 (04). pp. 543–552.
3. Goerlandt, F. & Kujala, P. (2011) Traffic simulation based ship collision probability modeling. *Reliability Engineering & System Safety*. 96 (1). pp. 91–107.
4. Aarsæther, G.K. & Moan, T. (2009) Estimating navigation patterns from AIS. *Journal of Navigation*. 62 (04). pp. 587–607.
5. Mascaro, S., Korb, K.B. & Nicholson, A.E. (2010) Learning abnormal vessel behaviour from ais data with bayesian networks at two time scales. *Tracks a Journal of Artists Writings*.
6. Pallotta, G., Vespe, M. & Bryan, K. (2013) Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*. 15 (6). pp. 2218–2245.
7. Port Authority, R. (2012) HydroMeteoBundel, No. 4, ed, Rotterdam, the Netherlands, Dutch.
8. Ristic, B., La Scala, B., Morelande, M. & Gordon, N. (2008) Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. *Information Fusion*, 2008. *11th International Conference*, 30 June – 3 July 2008, Cologne. IEEE. pp. 1–7.
9. Shu, Y., Daamen, W., Ligteringen, H. & Hoogendoorn, S. (2013) Vessel Speed, Course, and Path Analysis in the Botlek Area of the Port of Rotterdam, Netherlands. Transportation Research Record: *Journal of the Transportation Research Board*. 2330 (1). pp. 63–72.
10. Silveira, P., Teixeira, A. & Soares, C.G. (2013) Use of AIS data to characterise marine traffic patterns and ship collision risk off the coast of Portugal. *Journal of Navigation*. 66 (06). pp. 879–898.
11. Xiao, F. (2014) *Ships in an Artificial Force Field: A Multi-agent System for Nautical Traffic and Safety*. TU Delft, Delft University of Technology.
12. Zhu, F. (2011) Mining ship spatial trajectory patterns from AIS database for maritime surveillance, Emergency Management and Management Sciences (ICEMMS). *2nd IEEE International Conference*, 8–10 August 2011, Beijing. IEEE. pp. 772–775.