Journal of Marine Science and Technology

Volume 27 | Issue 2 Article 8

STUDY ON COMPLEXITY MODEL AND CLUSTERING METHOD OF SHIP TO SHIP ENCOUNTERING RISK

Yong-Pan Li
School of Navigation, Dalian Maritime University, Dalian, China

Zheng-Jiang Liu
School of Navigation, Dalian Maritime University, Dalian, China., liuzhengjiang@dlmu.edu.cn

Jack Shan Kai
China Energy Ship Management Company, Hong Kong, China.

Recommended Citation
Li, Yong-Pan; Liu, Zheng-Jiang; and Kai, Jack Shan (2019) "STUDY ON COMPLEXITY MODEL AND CLUSTERING METHOD OF SHIP TO SHIP ENCOUNTERING RISK," *Journal of Marine Science and Technology*: Vol. 27: Iss. 2, Article 8.
DOI: 10.6119/JMST.201904_27(2).0008
Available at: https://jmstt.ntou.edu.tw/journal/vol27/iss2/8

Journal of Marine Science and Technology, Vol. 27, No. 2, pp. 153-160 (2019)
DOI: 10.6119/JMST.201904_27(2).0008

STUDY ON COMPLEXITY MODEL AND CLUSTERING METHOD OF SHIP TO SHIP ENCOUNTERING RISK

Yong-Pan Li[1], Zheng-Jiang Liu[1], and Jack Shan Kai[2]

Key words: automatic identification system, complexity, clustering, encountering risk.

ABSTRACT

In terms of waters, vessel density distribution is a significant factor to evaluate the complexity of marine traffic and the collision risk. In previous studies, scholars frequently discovered the high-density vessel clusters according to density-based algorithms. Nevertheless, these algorithms were normally based on Euclidean or Hausdroff distance, etc., in which the encountering situation was prone to be ignored. Apparently, the heavy-density vessels in the traffic separation scheme don't have a high risk due to their well organization, while the micro-traffic relationships such as approaching, receding, head-on and crossing should be crucial factors in clustering. Therefore, this paper majorly focuses on the complexity of vessel couple and it's clustering using data mining technology. The complexity model of vessel couple is improved by taking the following factors into consideration: length overall, distance, movement trend and crossing angle. On the basis of traffic complexity and risk factors analysis, a clustering method of ship to ship encountering risk is presented by proposing a new distance definition, which can more effectively calculate the complexity of a mass of ships in an area.

I. INTRODUCTION

Currently, the large scale, high speed and increasing number of vessels along with busy sea routes have increased the complexity of marine traffic. The traditional intuition-based judgment and analysis cannot satisfy the needs of marine traffic management. Due to it is tough for Vessel Traffic Service (VTS) operators to discern traffic complexity and distinguish the high-risk vessels, they are confronting with a major challenge for guaranteeing the navigation safety of vessels.

Automatic Identification System (AIS), which is to be compulsorily installed on vessels of 500 gross tonnages and above, has collected a great deal of dynamic data of ships. Thus, how to generalize the characteristics of vessel traffic and perceive traffic situation has become research hotspots. At present, the related researches primarily focus on two fields, the statistics and analysis of macro traffic characteristics, and the exploration of micro traffic relationship.

The research on macro traffic characteristics studies vessel density and velocity distribution (Zhen et al., 2014), identifies special areas (Pallotta et al., 2013) and vessel routes (Chen et al., 2015) by mining historical AIS data. Besides, the collective behaviors of vessels are also studied, such as waterway throughput capacity based on the knowledge of fluid dynamics (Zhu and Zhang, 2009). The studies of micro traffic relationship mainly focus on ship domain and movement pattern (Zhou and Zheng, 2016), ships' encounter and collision risk (Pan et al., 2010) using ship domain and probability theory. Based on the experience of road or aviation field, scholars have recently proposed the concepts of marine traffic conflict, vessel near-miss and so on by introducing the microscopic factors of marine traffic (Zhang et al., 2015, 2016; Wu et al., 2016, Van Westrenen and Ellerbroek, 2017).

Vessel density distribution plays an essential role in evaluating traffic complexity and collision risk in waters. In previous AIS-based studies, scholars often used density-based algorithm to discover high-density vessel clusters (Sun et al., 2015, Yan et al., 2016). It can be considered that these algorithms are on account of Euclidean and Hausdroff distance, etc., in which the traffic complexity of the encountering situation is likely to be overlooked. Obviously, well-organized high-density vessels in traffic separation scheme are not highly risky. Factors like approaching, receding, head-on and crossing are vital for clustering in the micro traffic relationship.

Recently, complexity has become a hot topic in the field of transportation. Wen et al. (2014, 2015) and Geng et al. (2016) introduced a marine traffic complexity model to investigate the degree of crowding and risk of collision. However, the proposed model merely assumed the standard Length Overall (LOA) and made simulation analysis in the paper, while its utility in the era of big data is not completely achieved.

Paper submitted 05/10/18; revised 01/11/19; accepted 03/12/19. Author for correspondence: Zheng-Jiang Liu (e-mail: liuzhengjiang@dlmu.edu.cn).
[1] School of Navigation, Dalian Maritime University, Dalian, China.
[2] China Energy Ship Management Company, Hong Kong, China.

In consequence, this paper will mainly analyze the traffic complexity of vessel couple and it's clustering using data mining technology. Compared to the previous studies, the contributions are maily reflected in two dimensions. Firstly, it proposes a complexity model bridging Wen et al. (2014, 2015) and Zhang et al. (2015, 2016), which ultilizes the modeling method by Wen et al. (2015) and the affecting factors by Zhang et al. (2015, 2016). Secondly, it proposes a new distance definition and presents a clustering method of ship-ship encounter risk according to traffic complexity and risk factors analysis to speed up the complexity calculation of a large number of ships in VTS area.

II. COMPLEXITY MODEL

1. Basic Concept

Commonly, two encountering vessels at close range, i and j, constitute the basic marine traffic relationship unit, which is denoted by Vessel Couple(VC_{ij}).

The states of vessels i and j are defined as $i(t, x_i, y_i, v_i, \alpha_i)$ and $j(t, x_j, y_j, v_j, \alpha_j)$ at time t. Specifically, x_i, x_j are regarded as longitudes, y_i, y_j as latitudes, v_i, v_j as Speed Over Ground (SOG), α_i, α_j as Course Over Ground (COG). Moreover, $\overrightarrow{D_{ij}}$ refers to the distance from ship i to j, while $\overrightarrow{v_{ij}}$ indicates the relative speed of ship i to j. If setting the crossing angle of two vessels as φ, $\varphi = \left|\alpha_i - \alpha_j - \pi\right| \in [0, \pi]$.

2. Complexity Analysis

Similar to the definitions proposed by Wen et al., 2015, l_{upper}, l_{middle} and l_{lower} are three relevant parameters. According to the theory of Zhang et al. (2015, 2016), relative distance, movement trend and crossing angle of ships are selected as the most crucial factors on the complexity of vessel couple.

Definition 1 (Complexity of vessel couple): The complexity of VC_{ij} indicates the influence degree of vessel j to vessel i by the factors of relative distance, movement trend and crossing angle, denoted as $complexity_{ij}$.

It follows the hypothesis that the complexity of vessel couple has continuity, that is to say, the complexity values change continuously with the changes of the factors.

1) Distance Factor

Definition 2 (Complexity of distance factor): Taking ship domain of i into consideration, the influence degree of the vessel j by the distance $\left|\overrightarrow{D_{ij}}\right|$ on the vessel i is called the complexity of distance factor, denoted by $distance_{ij}$.

$$distance_{ij} = \begin{cases} \lambda e^{-\alpha\frac{\left|\overrightarrow{D_{ij}}\right|}{R}}, & \left|\overrightarrow{D_{ij}}\right| \geq R \\ \lambda e^{-\alpha}, & \left|\overrightarrow{D_{ij}}\right| < R \end{cases}$$

(Ye and Hu, 2012; Geng et al., 2016)

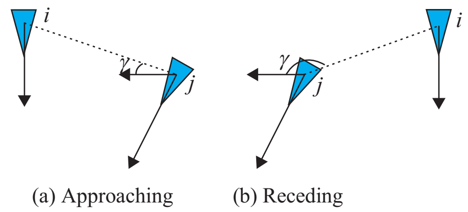

Fig. 1. Movement trend of two vessels.

In the fomula, parameters $\lambda > 0$, $\alpha > 0$, $R = l_{lower}$, which refers to the minimum safety distance between vessels j and i.

When $\left|\overrightarrow{D_{ij}}\right| \geq R$, $distance_{ij}$ increases nonlinearly with the decrease of relative distance. When $\left|\overrightarrow{D_{ij}}\right| < R$, it reflects the other vessel is already in own-ship's domain, undoubtedly, there will be collision risk. Thus, the complexity reach a maximum value.

Due to LOA's difference, the corresponding ship domains have various radiuses. Therefore, $distance_{ij}$ is asymmetric, $distance_{ij} \neq distance_{ji}$.

2) Movement Trend Factor

Definition 3 (Complexity of movement trend factor): The complexity caused by vessels' relative movement trend, denoted by $trend_{ij}$.

In the case of the same distance and crossing angle, there are two opposite kinds of movement tendency, approaching or receding, depending on the relative position and orientation of the vessels. In most cases, there is no risk involved in the receding vessels (Goerlandt et al., 2015). If the vessel couple is receding, $distance_{ij}$ is defined as zero. Otherwise, $distance_{ij}$ is positively related to the relative speed. The relative movement trend of vessel couple can be expressed as:

$$\frac{d\left|\overrightarrow{D_{ij}}\right|}{dt} = \frac{\overrightarrow{D_{ij}} \cdot \overrightarrow{v_{ij}}}{\left|\overrightarrow{D_{ij}}\right|} = \left|\overrightarrow{v_{ij}}\right| \cdot \cos(\overrightarrow{v_{ij}}, \overrightarrow{D_{ij}})$$

(Delahaye and Puechmorel, 2000; Wen et al., 2015)

In the case of approaching shown as Fig. 1(a), the angle γ between $\overrightarrow{v_{ij}}$ and $\overrightarrow{D_{ij}}$ is less than $\pi/2$, $\cos(\overrightarrow{v_{ij}}, \overrightarrow{D_{ij}}) > 0$. While receding shown as Fig. 1(b), the angle γ is larger than $\pi/2$, $\cos(\overrightarrow{v_{ij}}, \overrightarrow{D_{ij}}) < 0$.

$$trend_{ij} = l_{R^+}\left\{\frac{d\left|\overrightarrow{D_{ij}}\right|}{dt}\right\} = \begin{cases} \frac{d\left|\overrightarrow{D_{ij}}\right|}{dt}, \frac{d\left|\overrightarrow{D_{ij}}\right|}{dt} > 0 \\ 0, \frac{d\left|\overrightarrow{D_{ij}}\right|}{dt} < 0 \end{cases}$$

(Zhang et al., 2009)

l_{R^+} is an indicator function to output its value when the variable is positive, otherwise 0.

3) Crossing Angle Factor

Definition 4 (Complexity of crossing angle factor): The complexity caused by crossing angle φ, denoted by $angle_{ij}$.

According to Montewka et al. (2012) and Wen et al. (2015), there is a nonlinear relationship between complexity and crossing angle φ. When $\varphi < 20°$, relative speed of vessel couple is low, the traffic situation changes slowly and vessels have more time to take action to avoid collision with less complexity. When $\varphi \approx 120°$, vessel couple has a higher relative speed, which is more difficult for crews to assess the encounter situation that requires an earlier action. Thus, it has the greatest complexity. When $\varphi \approx 180°$, vessel couple is in a situation of head-on encounter. Although with a high relative speed, they have clear collision-prevention responsibility, which causes the complexity in medium. The function curve of $f(\varphi)$ is shown in Fig. 2.

When $\left|\overrightarrow{D_{ij}}\right| = l_{middle}$, the nonlinear function of the crossing angle φ can be constructed as:

$$f(\varphi) = \frac{1}{2}\cdot\left[1-\cos\left(\frac{180}{67.5}\cdot\frac{\varphi}{2}\cdot\frac{\pi}{180}+\frac{\pi}{10}\right)\right]$$

(Wen et al., 2015)

To ensure the assumption that the complexity of crossing angle factor is continuous, construct the following functions (Ye and Hu, 2012; Wen et al., 2015):

When $\left|\overrightarrow{D_{ij}}\right| \in \left(l_{lower}, l_{middle}\right]$,

$$g_1\left(\left|\overrightarrow{D_{ij}}\right|\right) = \frac{\left|\overrightarrow{D_{ij}}\right| - l_{lower}}{l_{middle} - l_{lower}},$$

$$angle_{ij} = f(\theta)\cdot g_1\left(\left|\overrightarrow{D_{ij}}\right|\right).$$

When $\left|\overrightarrow{D_{ij}}\right| \in \left(l_{middle}, l_{upper}\right]$,

$$g_2\left(\left|\overrightarrow{D_{ij}}\right|\right) = \frac{l_{upper} - \left|\overrightarrow{D_{ij}}\right|}{l_{upper} - l_{middle}},$$

$$angle_{ij} = f(\theta)\cdot g_2\left(\left|\overrightarrow{D_{ij}}\right|\right).$$

3. Complexity of Vessel Couple

As mentioned above, the encounter situation should be an influential factor in the density-based AIS data clustering. Supposing that complexity of vessel couple is sum of the complexity generated by encounter situation and the complexity due to spatial distance, it can be expressed as follows (Wen et al., 2015):

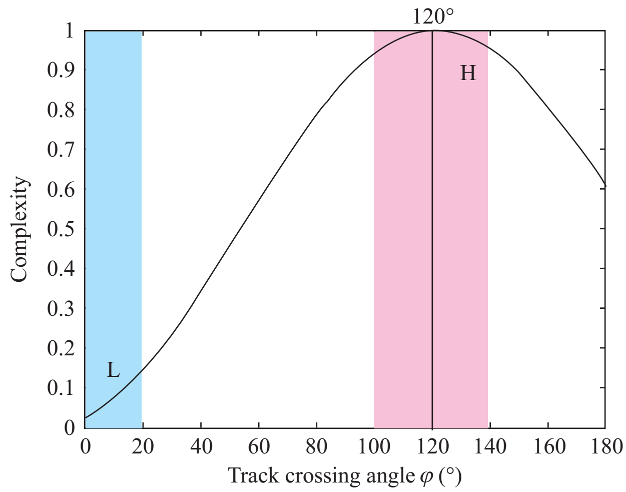

Fig. 2. Function curve of angle complexity and crossing angle. (Montewka et al. (2012) and Wen et al. (2015)).

$$complexity_{ij} = distance_{ij} + angle_{ij}\cdot trend_{ij} = \begin{cases} \lambda e^{-\alpha}, & \left|\overrightarrow{D_{ij}}\right| \in \left(0, l_{lower}\right] \\ \lambda e^{-\alpha\frac{\left|\overrightarrow{D_{ij}}\right|}{R}} + l_{R^+}\left\{\dfrac{d\left|\overrightarrow{D_{ij}}\right|}{dt}\right\}\cdot f(\varphi)\cdot g_1\left(\left|\overrightarrow{D_{ij}}\right|\right), & \left|\overrightarrow{D_{ij}}\right| \in \left(l_{lower}, l_{middle}\right] \\ \lambda e^{-\alpha\frac{\left|\overrightarrow{D_{ij}}\right|}{R}} + l_{R^+}\left\{\dfrac{d\left|\overrightarrow{D_{ij}}\right|}{dt}\right\}\cdot f(\varphi)\cdot g_2\left(\left|\overrightarrow{D_{ij}}\right|\right), & \left|\overrightarrow{D_{ij}}\right| \in \left(l_{middle}, l_{upper}\right] \\ \lambda e^{-\alpha\frac{\left|\overrightarrow{D_{ij}}\right|}{R}}, & \left|\overrightarrow{D_{ij}}\right| \in \left(l_{upper}, +\infty\right] \end{cases}$$

On account of the asymmetry of $distance_{ij}$, $complexity_{ij} \neq complexity_{ji}$. Nonetheless, the accident probability mainly depends on the larger-size party of complexities in the vessel couple. Therefore, the complexity takes the maximum (Goerlandt et al., 2015), that is

$$complexity(i, j) = \max\{complexity_{ij}, complexity_{ji}\}$$

4. Model Parameter Estimation

Admittedly, experts are invited to set parameters. Given the rather extensive scope of the elicitation, it is preferred to select only a limited number of experts who are capable to contribute their expertise over a longer time. The experts include one captain (10 years of experience), two first officers (5 and 3 years of experience) and two VTS operators (8 and 6 years of experience).

Nowadays, motor vessels' LOA are between 20 m and 400 m, $L \in$ (20 m, 400 m). It is generally accepted by experts in the harbor research area that other ships 1 n mile away can be ignored by the ship in a length of 20 m, and that other ships 3 n miles away can be neglected by the ship whose LOA is 400 m.

It can be described as, $\frac{l_{upper}-1}{L-20}=\frac{3-1}{400-20}$, that is, $l_{upper}=\frac{L}{190}+\frac{17}{19}$. $l_{middle}=0.5*l_{upper}=\frac{L}{380}+\frac{17}{38}$.

The shape of vessel domains (circular or elliptical) does not have significant impacts on the locations of hot spots for vessel conflicts (Wu et al., 2016). This paper adopts the ship domain in circular, whose radius is three times the size of LOA, $l_{lower}=3\ L$. For example, if a vessel's LOA, $L=185$ m, then $l_{lower}=0.3$ n mile, $l_{middle}=0.93$ n mile and $l_{upper}=1.87$ n mile.

Furthermore, parameters λ and α are also calculated by experts. $\left|\overrightarrow{D_{ij}}\right|$ should be greater than 3 L. When $L=400$ m, $l_{upper}=$ 3 n mile = 13.9 L, which $\left|\overrightarrow{D_{ij}}\right|$ should be less. In the absence of relative movement trend and crossing-angle impact, the complexity of vessel couple when $\left|\overrightarrow{D_{ij}}\right|=6\ L$ is five times as that when $\left|\overrightarrow{D_{ij}}\right|=12\ L$. In the case of $\left|\overrightarrow{D_{ij}}\right|=9\ L$, when movement trend and crossing angle influence the complexity to a large extent, the total complexity of movement trend factor and crossing angle factor is twice larger than the complexity of distance factor, that is,

$$\lambda e^{-2\alpha}=5*\lambda e^{-4\alpha}$$

$$2*\lambda e^{-\alpha\frac{\left|\overrightarrow{D_{ij}}\right|}{R}}=l_{R^+}\left\{\frac{d\left|\overrightarrow{D_{ij}}\right|}{dt}\right\}\cdot f(\theta)\cdot g_1\left(\left|\overrightarrow{D_{ij}}\right|\right)$$

It can be calculated, $\lambda=110.2$, $\alpha=0.8$, and the maximum complexity, $\lambda e^{-\alpha}=50$.

III. CLUSTERING METHOD

1. Order Distance Definition

Typically, the more complexity, the more dangerous the vessel couples are. This paper proposes a new distance definition of order distance, that is, the narrower the order distance is, the more hazardous the vessel couples are.

Definition 5 (Order distance): The order distance of vessel couple is a nonlinear overlay of the ship's encounter relation to the spatial distance. Its value is equal to the reciprocal complexity of vessel couple (Debnath and Chin, 2009).

$$orderDist(i,\ j)=\frac{1}{complexity(i,\ j)}$$

2. Clustering Method

At one point, there are many vessels in VTS area and every two make up a couple. In order to get the total complexity, the definition and procedure of Ordering Points to Identify the Clustering Structure (OPTICS) (Ankerst et al., 1999) are improved.

1) OPTICS Algorithm Redefinition

OPTICS requires two parameters: ε, which describes the maximum order distance to consider, and *MinPts*, describing the number of AIS points required to form a cluster.

Definition 6 (ε-neighborhood): For database D of AIS points at the same time in a specific area, the ε-neighborhood of an AIS point i, denoted by $N_\varepsilon(i)$, is defined by

$$N_\varepsilon(i)=\left\{j\in D\,\middle|\,orderDist(i,\ j)<=\varepsilon\right\}$$

An AIS point i is a core point if at least *MinPts* AIS points are found within its ε-neighborhood.

Definition 7 (Core distance): The core distance describes the order distance to the *MinPts-th* closest AIS point.

$$coreDist(i)=\begin{cases}UNDIFINED,\ if\ N_\varepsilon(i).size<MinPts\\ MinPts_{th}\ smallest\ order\ distance\ in\ N_\varepsilon(i),\ else\end{cases}$$

Definition 8 (Reachability distance): The reachability distance of another AIS point j from an AIS point i is either the order distance between j and i, or the core distance of i, whichever is bigger:

$$reachDist(i,\ j)=\begin{cases}UNDIFINED,\ if\ N_\varepsilon(i).size<MinPts\\ \max\left\{coreDist(i),\ orderDist(i,\ j)\right\},\ else\end{cases}$$

2) OPTICS Algorithm Re-Description

The basic approach of OPTICS is shown as Fig. 3.

In the Update procedure, the priority queue *Seeds* is updated with the ε-neighborhood of p and q respectively, shown in Fig. 4. OPTICS hence outputs the points in a particular sequence, annotated with their smallest reachability distance.

By the algorithm, the AIS points of the database are linearly ordered so that points which are closest by order distance become neighbors in the ordering. The diagram of reachability distance can visually show the circumstance in low-concave area, indicating a greater risk among vessels, which are more accident-prone.

3. Marine Traffic Complexity

The marine traffic complexity refers to the complexity of marine traffic situation and the efforts required by VTS operators to recognize and manage marine traffic (Zhu and Zhang, 2014). Actually, the higher complexity of marine traffic is reflected that ships are dense and there are various potential conflicts, which is difficult to get rid of in the limited maneuvering space. Although the complexity is affected by multiple factors like meteorology, hydrology, channel status and information support, VTS operators are more concerned about the changes of marine traffic

Table 1. List of AIS data (According to Fig. 5).

Time	MMSI	Longitude/°	Latitude/°	SOG/kn	COG/°	Vessel Name	LOA/m
15:33:00	413456050	122.48360	30.52850	11.6	138.4	FAN AN 166	100
15:33:00	412702640	122.50750	30.51597	14.6	136.5	HUA HANG 1	128
15:33:00	413259000	122.52840	30.52807	9.8	272.5	NING HUA 420	92
15:33:00	413358570	122.52840	30.49823	5.3	332.3	LIN DA	99
15:33:00	413552790	122.62180	30.56325	10	182.3	HONG DA YOU 68	53
15:33:00	413324207	122.61460	30.51823	10.7	182	AK XIN RI QIANG	99
15:33:00	413491910	122.64800	30.57078	12	177.5	JING HAI SHENG	102
15:33:00	477598800	122.64650	30.53972	11.2	252.1	COSCO HOPE	366

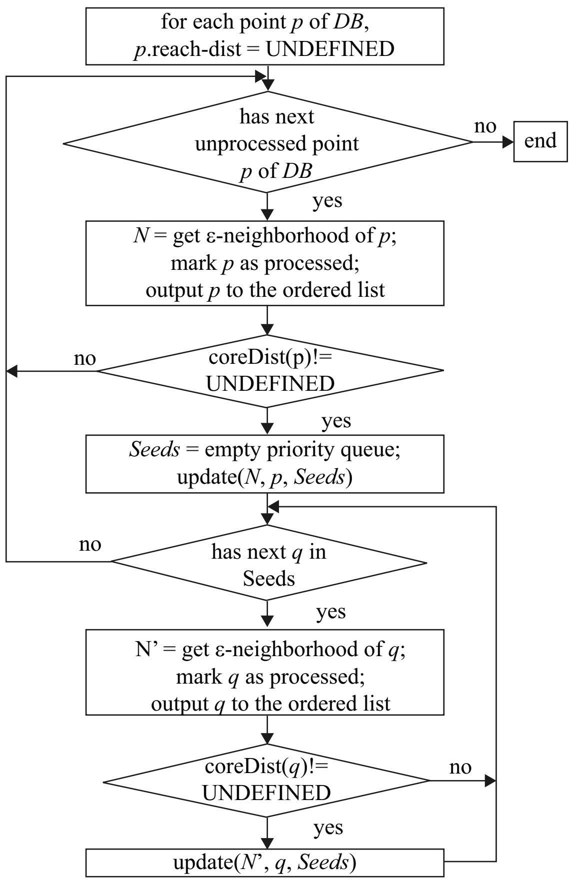

Fig. 3. Diagram of OPTICS procedure.

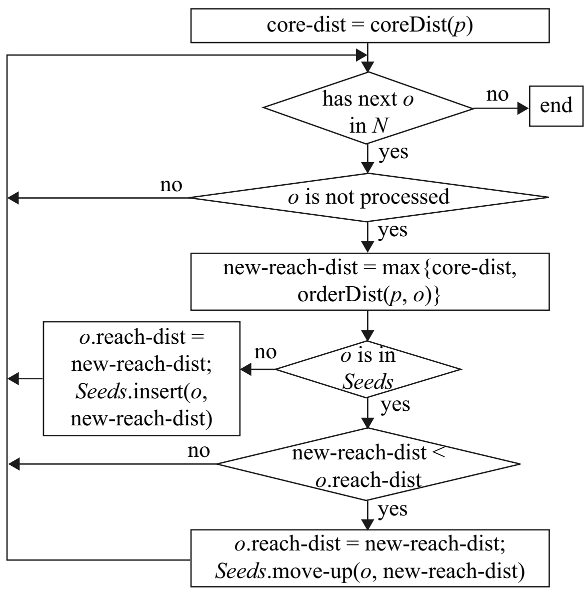

Fig. 4. Diagram of Update procedure.

situation. In this paper, marine traffic complexity is only related to the traffic relationship among ships, that is, the intrinsic attributes of traffic flow (speed, course and position, etc.).

There will be marine traffic complexity among vessel objects if the OPTICS algorithm outputs some vessel objects whose reachability distances are defined.

Definition 9 (Marine traffic complexity): The OPTICS algorithm is run on AIS data at a time slice t in an area. A total of m reachability distances are generated, that is, d_1, d_2, …, d_m. The marine traffic complexity of the area at the time slice t is expressed as C_t, then

$$C_t = 1/d_1 + 1/d_2 + \ldots + 1/d_m .$$

$MinPts \geq 3$ indicates that there are defined reachability distances among at least three vessels, resulting in multi-ship encounter situation.

IV. EXPERIMENT

1. AIS data Collection

A screenshot of the vessel dynamic monitoring at 15:33:00 o'clock on 6 October, 2017 in the Zhoushan port, China is shown in Fig. 5. There were eight vessels sailing in the waters in total, and attributes of Time, Maritime Mobile Service Identity (MMSI), Longitude, Latitude, SOG, COG, Vessel name and LOA are

Table 2. List of complexity and order distance.

Vessel Name	Vessel Name	Complexity	Order Distance
COSCO HOPE	COSCO HOPE	50.000	0.020
COSCO HOPE	JING HAI SHENG	12.835	0.078
COSCO HOPE	FAN AN 166	0.020	50.000
COSCO HOPE	HUA HANG 1	0.020	50.000
COSCO HOPE	NING HUA 420	0.028	36.260
COSCO HOPE	LIN DA	0.020	50.000
COSCO HOPE	HONG DA YOU 68	14.110	0.071
COSCO HOPE	AK XIN RI QIANG	8.268	0.121

Table 3. Result of OPTICS clustering ($\varepsilon = 1$, $MinPts = 3$).

Vessel Name	Core Distance	Reachability Distance
COSCO HOPE	0.078	Undefined
JING HAI SHENG	Undefined	0.078
HONG DA YOU 68	Undefined	0.078
AK XIN RI QIANG	Undefined	0.121
FAN AN 166	Undefined	Undefined
HUA HANG 1	Undefined	Undefined
NING HUA 420	Undefined	Undefined
LIN DA	Undefined	Undefined

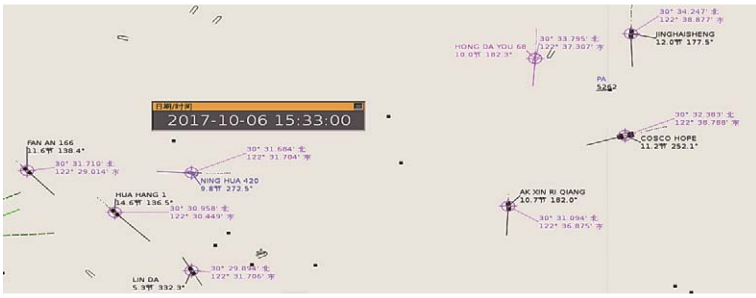

Fig. 5. Screenshot of vessel dynamic monitoring.

shown in Table 1.

2. Complexity Model Verification

According to the formulas above, the calculation results of complexity and order distances between "COSCO HOPE " and other seven vessels are shown in Table 2.

The distances from "COSCO HOPE" to "HONG DA YOU 68" and "JING HAI SHENG" are both 1.9 miles, while the crossing angles between "COSCO HOPE" and "HONG DA YOU 68", "COSCO HOPE" and "JING HAI SHENG" are respectively 70° and 74°, which are subequal. Nevertheless, due to their varied movement trends, the complexities are certainly different. The distance between "COSCO HOPE" and any other four vessels, "FAN AN 166", "HUA HANG 1", "NING HUA 420" and "LIN DA", is beyond the influence scope of complexity, so their complexities are close to zero. The results are basically in accordance with experts' experience.

3. Clustering Analysis and Discussion

Set $\varepsilon = 1$, $MinPts = 3$, indicating that at least three vessels are density-connected within the radius of one by order distance. The results are shown in Table 3 and Fig. 6.

The core distance of "COSCO HOPE" is 0.078. The reachability distances from "COSCO HOPE" to "JING HAI SHENG"

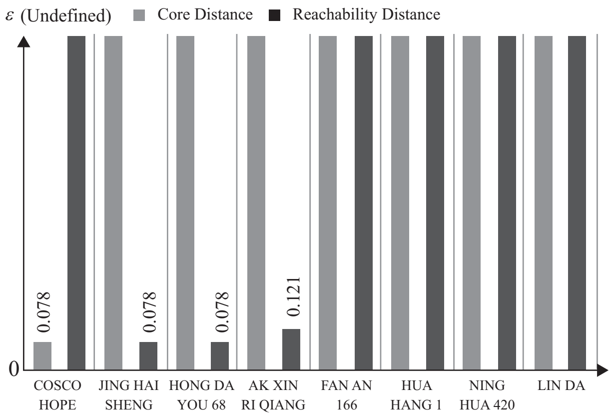

Fig. 6. Illustration of OPTICS clustering result.

and "HONG DA YOU 68" are both 0.078. The grey columns of the reachability distance to three vessels (JING HAI SHENG, HONG DA YOU 68 and AK XIN RI QIANG) form a recessed area, indicating that four vessels (COSCO HOPE, JING HAI SHENG, HONG DA YOU 68 and AK XIN RI QIANG) are density-connected by order distance less than 0.121. Besides, three vessels (COSCO HOPE, JING HAI SHENG and HONG DA YOU 68) are density-connected by order distance less than 0.078, demonstrating that it is more collision-prone with larger complexity. The AIS data is outputted by the algorithm in ordering and the clusters within arbitrary complexity can be drawn by scribing at the output graph. For example, the cluster of three vessels (COSCO HOPE, JING HAI SHENG and HONG DA YOU 68) is gotten by drawing a line at 0.08 in Fig. 6.

The value $C_t = 1/0.078 + 1/0.078 + 1/0.121 = 33.905$ indicates the complexity of marine traffic at 15:33:00 o'clock on 6 October, 2017 in the specific area. The high, medium or low complexity of marine traffic in the area can be delineated on the basis of its statistical distribution, which is obtained by calculating and analyzing the long-term complexity of every minute. Thus, VTS operator should pay extra attention to the traffic if Ct is of high complexity.

V. CONCLUSIONS

In summary, this paper highlights the issue of vessel AIS data mining based on the micro traffic relations. From the microscopic perspective, the complexity model of vessel couple is proposed by fully considering the influence of four significant factors: vessels' distance, movement trend, crossing-angle and LOA, which can describe marine traffic more accurately. Moreover, the OPTICS algorithm is improved based on the complexity and order distance to carry out cluster analysis and to summarize the complexity of marine traffic from the macroscopic standpoint. In the case study, actual AIS data from Zhoushan North Sea in China is employed to demonstrate the model and the algorithm, which can be beneficial to effectively calculate the traffic complexity of numerous ships in VTS area.

Although the complexity of maritime traffic is estimated on account of the existing historical AIS data, this methodology can also be employed to process the received AIS data in real time and to carry out approximate real-time calculation by dividing time slice into every minute. Surely, it is vital to comprehend the real-time complexity in surveillance area for VTS operator who should be cautious about the high complexity. What's more, it is meaningful and valuable to predict the complexity of marine traffic at the next moment, which needs to be thoroughly studied in the future.

REFERENCES

Ankerst, M., M. M. Breunig, H.-P. Kriegel and J. g. Sander (1999). OPTICS: Ordering Points To Identify the Clustering Structure. Proceedings of the 1999 ACM SIGMOD international conference on Management of data.

Chen, J., F. Lu and G. Peng (2015). A quantitative approach for delineating principal fairways of ship passages through a strait. Ocean Engineering 103, 188-197.

Chin, H. C. and A. K Debnath (2009). Modeling perceived collision risk in port water navigation. Safety Science 47, 1410-1416.

Delahaye, D. and S. Puechmorel (2000). Air traffic complexity: towards intrinsic metrics. Proceedings of the 3rd USA/Europe Air Traffic Management R&D Seminar.

Geng, X., L. Huang, Y. Wen and C. Zhou (2016). Modelling of the Marine Traffic Situation Complexity. MATEC Web of Conferences 68, 15002.

Goerlandt F., J. Montewka, V. Kuzmin and P. Kujala (2015). A risk-informed ship collision alert system: Framework and application, Safety Science 77, 182-204.

Montewka, J., F. Goerlandt and P. Kujala (2012). Determination of collision criteria and causation factors appropriate to a model for estimating the probability of maritime accidents. Ocean Engineering 40, 50-61.

Pallotta, G., M. Vespe and K. Bryan (2013). Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. Entropy 15, 2218-2245.

Sun, F., Y. Deng, F. Deng, Q. Zhu and H. Chu (2015). Unsupervised maritime traffic pattern extraction from spatio-temporal data. 11th International Conference on Natural Computation (Icnc), 1218-1223.

Van Westrenen, F. and J. Ellerbroek (2017). The effect of traffic complexity on the development of near misses on the north sea. IEEE Transactions on Systems, Man and Cybernetics: Systems 47, 432-440.

Wen, Y., Y. Huang, C. Zhou, J. Yang, C. Xiao and X. Wu (2015). Modelling of marine traffic flow complexity. Ocean Engineering 104, 500-510.

Wu, X., A. L. Mehta, V. A. Zaloom and B. N. Craig (2016). Analysis of waterway transportation in Southeast Texas waterway based on AIS data. Ocean Engineering 121, 196-209.

Yan, W., R. Wen, A. N. Zhang and D. Yang (2016). Vessel movement analysis and pattern discovery using density-based clustering approach. Proceedings of the IEEE International Conference on Big Data.

Zhang, W., F. Goerlandt, P. Kujala and Y. Wang (2016). An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Engineering 124, 141-156.

Zhang, W., F. Goerlandt, J. Montewka and P. Kujala (2015). A method for detecting possible near miss ship collisions from AIS data. Ocean Engineering 107, 60-69.

Pan, J.-C., Q.-S. Jiang and Z.-P. Shao (2010). Ship encounter data mining algorithm. Navigation of China, 57-60+64. (in Chinese)

Wen, Y.-Q., Y.-M. Huang, J.-L. Yang and C.-S. Xiao (2014). Modeling of marine traffic flow structure complexity. Navigation of China 37, 62-69. (in Chinese)

Ye, B.-J. and M.-H. Hu (2012). Traffic structure-based air traffic complexity modeling. Journal of Transportation Systems Engineering and Information Technology 12, 166-173. (in Chinese)

Zhang, J., M.-H. Hu and C. Zhang (2009). Complexity research in air traffic management. ACTA AERONAUT ICA ET ASTRONAUT ICA SINICA,

30, 2132-2142. (in Chinese)
Zhen, R., Z.-P. Shao, J. C. Pan and Q. Zhao (2014). Statistical analysis of distribution of ship speed within the fairway based on AIS data. Journal of Jimei University (Natural Science), 274-278. (in Chinese)
Zhou, D. and Z. Y. Zheng (2016). Analysis of influencing factors of ship domain based on AIS data. Journal of Shanghai Maritime University, 7-11+52. (in Chinese)
Zhu, J. and W. Zhang (2009). Calculation model of inland waterway transit capacity based on ship-following theory. Journal of Traffic and Transportation Engineering 9, 83-87. (in Chinese)
Zhu, X.-H. and J. Zhang (2014). Air traffic complexity evaluation method based on probabilistic trajectory prediction 36,300-305. (in Chinese)