The Journal of Navigation (2022), **75**:4 966–983
doi:10.1017/S0373463322000182

CrossMark

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Detection of abnormal ship trajectory based on the complex polygon

Jinxian Weng,* Guorong Li, and Yahui Zhao

College of Transport and Communications, Shanghai Maritime University, Shanghai 201306, China.
*Corresponding author. E-mail: jxweng@bjtu.edu.cn; wjx206@gmail.com.

Received: 26 June 2021; **Revised:** 23 March 2022; **Accepted:** 27 March 2022; **First published online:** 18 April 2022

Keywords: trajectory; ship behaviour; automatic identification systems

Abstract

Ship anomaly detection is a vital aspect for monitoring navigational safety in specific water areas. Considering the effect of water channel boundaries, we propose the detection of an abnormal ship trajectory based on the complex polygon (DATCP) method to detect ship anomalies in this study. With the automatic identification systems (AIS) data from the Yangtze River estuary, a case study is created to verify the effectiveness of the proposed DATCP method. The case study results reveal that the proposed DATCP method can provide higher detection accuracy than the conventional A* algorithm. The feature analysis results indicate that ship anomalies are significantly influenced by ship type, time period, weather conditions and ship traffic characteristics.

1. Introduction

As an important role in global commerce, maritime transportation accounts for approximately 90% of the global trade volume (Maria et al., 2018). With the rapid increase in the number of ships, frequent shipping activities can lead to more complex and higher-risk environments. In recent years, many researchers have placed their focus on investigating a ship's motion benefitting from the massive amount of automatic identification systems (AIS) data, which is a vital aspect for improving navigational safety. Ship trajectory prediction and abnormal trajectory detection are the two major perspectives for investigating the motion patterns of ships. For ship motion prediction, different methodologies like the Gaussian process model and the BP neural network algorithm have been proposed in previous studies (Rhodes et al., 2007; Zhou and Shi, 2010; Perera et al., 2012; Xu et al., 2012; Zhang et al., 2018; Rong et al., 2019). An accurate ship motion prediction can help maritime authorities to predict the possible activities of target ships.

To figure out abnormal or risky ship trajectories, some other studies have mainly focused on proposing effective anomaly detection methods to improve maritime safety and security systems (Tun et al., 2007; Pallotta et al., 2013a, b; Shahir et al., 2014). Specifically, the capability of identifying abnormal ship trajectories could help maritime authorities to comprehend the regularity of ship anomalies. However, considering the limited resources of maritime traffic surveillance operators, one big challenge is that it is difficult to fully monitor a large number of ship tracks at the same time (Zhen et al., 2017a). In reality, the precise extraction of ship trajectories is a prerequisite to detect abnormal ship manoeuvers. According to previous studies, the ship trajectory extraction techniques could be classified into three main groups including parametric methods, nonparametric methods and clustering methods, respectively.

Many statistical generative parametric models have been applied to extract trajectories of ships/cars based on historical trajectories data. Gaussian mixture models (GMMs) are very popular for approximating the (unknown) multivariate probability density function of ship traffic (Maria et al., 2018). Actually, the GMMs are a combination of multivariate Gaussian distributions. These distributions aim to summarise how the training data cluster and spread in the multi-dimensional spaces (Laxhammar et al., 2009; Anneken et al., 2015). Specifically, Laxhammar (2008) applied GMMs to cluster ship sailing vectors in each surveillance grid. Afterwards, anomaly detection in new data was performed by calculating the occurrence likelihood of abnormal trajectories. The segment was considered anomalous if the anomaly measure was below a certain alarm threshold. However, one disadvantage of GMMs is that it might be difficult for non-experts to understand.

For nonparametric methods, kernel density estimation (KDE) is a popular method that derives a nonparametric model of traffic normalcy. Laxhammar et al. (2009) used an adaptive KDE to develop a normal traffic model. Particle filters were applied to predict the positions of ships based on the derived density. Furthermore, an adaptive bandwidth selection method was also proposed by Laxhammar et al. (2009) and applied by Pallotta et al. (2013b). However, density estimation methods are computationally intensive so that they may not be practical in real-time systems. Furthermore, methods based on kernel density are applicable only for specific smaller regions. Clustering methods for trajectory extraction mainly consist of K-means clustering and density-based spatial clustering of applications with noise (DBSCAN) (Pallotta et al., 2013a, b; Pan et al., 2014; Li et al., 2017; Zhen et al., 2017b; Zhao and Shi, 2019). For instance, Zhen et al. (2017b) used the K-means algorithm to cluster ship trajectories. The best number of clusters was determined based on combinations of different distances. Zhao and Shi (2019) further determined the input parameter selection for the DBSCAN algorithm.

Normally, the identification of theoretical ship trajectories is a prerequisite for anomaly detection. Most previous studies used collected ship trajectory data to obtain the theoretical ship trajectory. However, channel characteristics were rarely considered for anomaly detection. For example, Soleimani et al. (2015) detected the ship trajectory only from a geometrical perspective. The actual ship trajectory is compared with a near-optimal path generated by a graph search A* algorithm. One disadvantage is that the reliability of the optimal path depends too much on the accuracy of the A* algorithm. In addition, ship trajectories should be closely related to the channel boundaries. Therefore, this study proposes the detection of abnormal ship trajectory based on complex polygons (DATCP) method by taking into account the effects of the specific channel boundaries. Meanwhile, the proposed method is proved to be more accurate than other algorithms (i.e., A* algorithm) by presenting a case study in the Yangtze River estuary.

2. Objectives and contributions

This study aims to improve the accuracy of ship anomaly detection by proposing an abnormal ship trajectory detection model based on complex polygons (i.e., the DATCP method). The proposed DATCP method determines the theoretical trajectory by dividing the channel boundary into many polygons according to the channel bend. In contrast, by dividing the water area into excessive grids, there is a significant overfitting problem for the A* algorithm proposed by previous studies. From this point of view, the A* algorithm may not be able to distinguish all the abnormal trajectories successfully. The contributions of this study are two-fold. First, it takes the initiative to propose a new way to solve the overfitting problem in the process of anomaly detection. Second, the proposed DATCP method could identify more abnormal trajectories so that the detected results are more reliable and convincing. In addition, this study also contributes to investigating ship anomaly features for a specific water area (i.e., the Yangtze River estuary), which is beneficial for maritime authorities to infer the safety level of the target waters. Corresponding strategies can thus be proposed to reduce navigational risk.

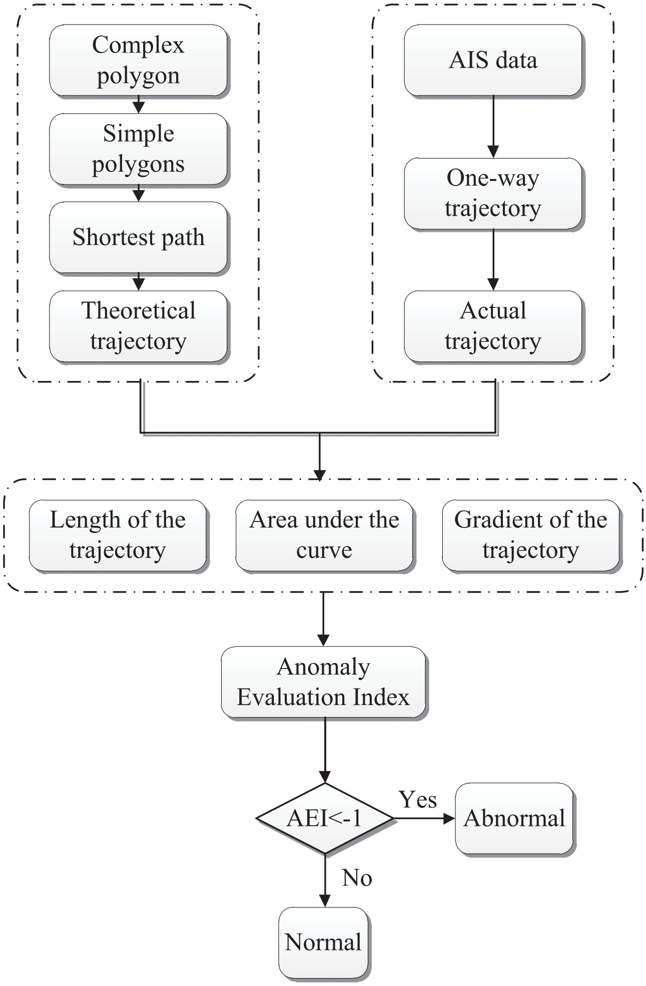

***Figure 1.** A flowchart for the DATCP method.*

3. Anomaly detection

3.1. Definition of abnormal ship trajectory

The abnormal ship movement can be defined as an unreasonable movement deviation from the channel range, optimal path, normal speed or other corresponding parameters (Lane et al., 2011). From the perspective of navigation data, ship movement anomalies can be regarded as observations that are significantly inconsistent with the remainder of the dataset (Hodge and Austin, 2004). Specifically, the abnormalities in navigation data can be reflected depending on the anomaly navigation speed (e.g., too high, too low or wandering) and unusual changes in ship coordinates (e.g., deviating from the normal course, illegally occupying the wrong waterway). Therefore, anomaly detection refers to the identification of trajectory data that do not conform to expected behaviours (Varun et al., 2009). The anomaly type mainly discussed in this study is that the ship trajectory extracted from the collected data deviates from the normal route. This study is dedicated to proposing an anomaly detection method (i.e., DATCP method) to identify the majority of abnormal trajectories that might occur in a specific water area. The detailed procedure of our proposed DATCP method can be seen in Figure 1.

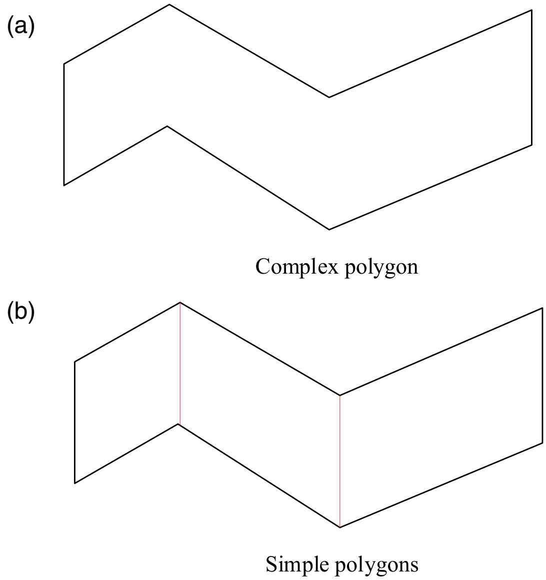

Figure 2. *Transforming a one-way channel into simple polygons.*

3.2. AIS data

Automatic identification systems (AIS) can provide critical information on ship trajectories. In this study, AIS data were used to detect anomalous ship trajectories. AIS data include static information and dynamic information. Generally, static information is sent every six minutes, which contains IMO number, maritime mobile service identity (MMSI), ship name, ship length and so on. Dynamic information is sent every 2 s to 3 min depending on the shipping speed, which contains position, time (UTC), speed over ground (SOG) and so on. High-precision AIS data with short transmission intervals can ensure the accuracy of anomaly detection. In reality, it is also common for AIS data to be sent at long intervals. In this situation, the existing AIS data can be used to fill the data gaps. The extrapolation and interpolation methods are thus applied to supplement the data gaps at an interval of 1 min when the two AIS records have a larger time interval. The trajectory of a ship can be drawn by connecting its consecutive coordinates in the AIS data accordingly. Generally, a complete ship trajectory is defined as a ship trip from the departure port to the destination port. However, the information about the actual departure and destination of the ships is not available from the raw AIS data. Therefore, the ship trajectory in this study refers to unidirectional trajectory segments.

3.3. Theoretical ship trajectory

This study introduces the concept of theoretical ship trajectory as a benchmark to detect abnormal ship trajectories. Namely, a ship trajectory is considered abnormal if the difference between the actual trajectory and the theoretical trajectory exceeds a certain range. Unlike with previous studies (e.g., Soleimani et al., 2015; Maria et al., 2018), the theoretical trajectory is not simply equal to the shortest navigation path in this study. The theoretical trajectory could not only be affected by the shortest path but also by the geometry of the channel boundary. In addition, the waterway hydrodynamic characteristics, such as the depth of waterway, should be also taken into account in identifying the theoretical ship trajectory. To identify the theoretical ship trajectory more easily, it is recommended to simplify the geometric boundary of the channel as a complex polygon (Kaluđer et al., 2011), as shown in Figure 2(a). Subsequently, the complex polygon can be divided into multiple simple polygons as shown in Figure 2(b). Thus, the shortest distance between any two sides of each simple polygon can be calculated separately.

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

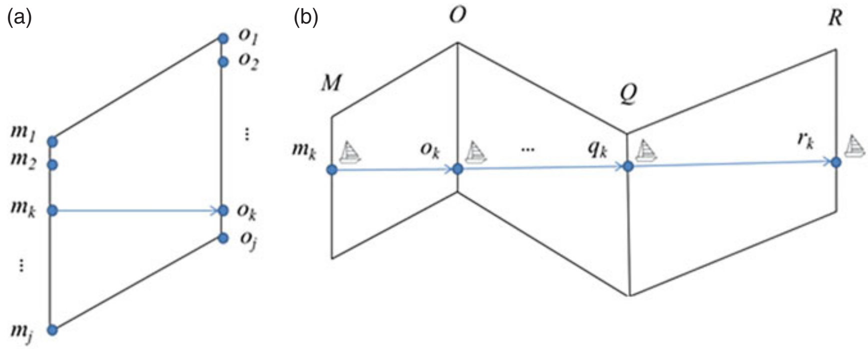

Figure 3. *Theoretical ship trajectories in a one-way channel.*

The theoretical trajectory of a ship can be considered as the optimal one when the ship trajectory in each simple polygon is the shortest.

Figure 3 provides a diagram to show the process of determining the theoretical ship trajectory based on complex polygons. We first treat the boundaries of the target one-way channel as a complex polygon and then divide the complex polygon into multiple simple polygons. The key step to obtain the theoretical ship trajectory is to determine the cross-section of departure and destination for each divided simple polygon. It should be noted that a ship may be started at any position of the departure cross-section. For a simple polygon, Figure 3(a) presents the detailed expression of the starting and the ending positions. More specifically, the starting positions on the departure cross-section can be expressed as m_1, m_2, m_3, $\ldots$ m_j, and the ending positions on the destination cross-section can be expressed as o_1, o_2, $o_3, \ldots$ o_j. Each m_j corresponds to a unique n_j that meets the requirements of the shortest path. It is assumed that there is a ship k starting from m_k on the departure cross-section M. The minimum distance from the departure cross-section M to the destination cross-section O can be obtained based on the starting point m_k. The corresponding ending point o_k is thus determined. The above steps are then repeated with o_k as the new starting point until the entire shortest path from m_k to r_k is determined, as shown in Figure 3(b).

3.4. Anomaly detection

With the obtained theoretical ship trajectory, abnormal ship trajectories can be detected by comparing the theoretical trajectory with the actual ship trajectory extracted from the AIS data. Specifically, anomaly detection is mainly achieved by introducing indicators such as trajectory length, area under the curve and trajectory gradient. Abnormal ship trajectories can be identified with the anomaly evaluation index (AEI) comprised of the three indicators mentioned above.

3.4.1. Trajectory length

Since both the theoretical ship trajectory and the actual ship trajectory are represented by a series of consecutive position coordinates (i.e., latitude and longitude coordinates), the trajectory length can be calculated by adding the distance between adjacent coordinates. In this study, the Haversine formula is applied to calculate the spherical distance between adjacent trajectory points because the latitude and longitude coordinates belong to spherical coordinates. The expressions of the Haversine formula are shown as follows:

$$h(p_1, p_2) = \sin^2\left(\frac{\Delta x}{2}\right) + \cos x_1 \cos x_2 \sin^2\left(\frac{\Delta y}{2}\right) \tag{1}$$

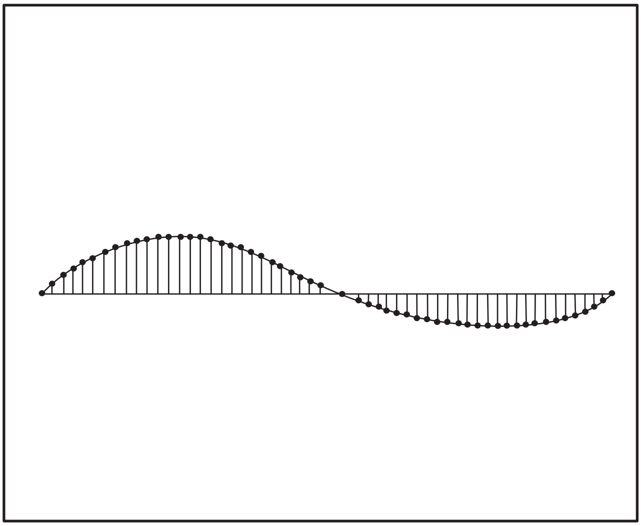

Figure 4. *Area under the curve.*

$$d(p_1, p_2) = 2R \ \arctan\left(\frac{\sqrt{h(p_1, p_2)}}{\sqrt{1 - h(p_1, p_2)}}\right) \tag{2}$$

$$L = \sum_{i=1}^{N-1} d(p_i, p_{i+1}) \tag{3}$$

where R represents the radius of the earth in metres ($R = 6{,}378{,}137$); $p_1 : (x_1, \ y_1)$ and $p_2 : (x_2, \ y_2)$ are the coordinates of two adjacent trajectory points; $\Delta x = x_2 - x_1$ and $\Delta y = y_2 - y_1$; d is the distance between two adjacent points; L represents the total length of the trajectory; N is the number of points in the trajectory.

3.4.2. Area under the curve

The area enclosed by the actual ship trajectory and the theoretical ship trajectory in Figure 4 can be used as an important parameter to detect abnormal trajectories. The area could indicate the extent of the actual trajectory deviating from the theoretical trajectory. In other words, a larger the area signifies a longer distance that the actual trajectory deviates from the theoretical trajectory. The area under the curve can be calculated by dividing the area into multiple trapezoids, as follows:

$$A = \sum_{i=1}^{N-1} |y_{i+1} - y_i| \frac{|x_i + x_{i+1}|}{2} \tag{4}$$

where A represents the area under the curve. Note that the final equation has transformed the coordinate system through the angle between the horizon and the straight line connecting the start and end points. Detailed mathematical derivation could be found in Soleimani et al. (2015).

3.4.3. Trajectory gradient

Gradient could be regarded as another critical indicator for the detection of abnormal trajectories, which could be defined as the partial derivative of distances at longitude and latitude dimensions. Note that the partial derivative only takes into account the position change. Here, G^x and G^y are the gradients of longitude and latitude dimensions, respectively. The calculation formula of trajectory gradient is shown as follows:

$$G^x = \sum_{i=1}^{N-1} |x_{i+1} - x_i|$$
$$G^y = \sum_{i=1}^{N-1} |y_{i+1} - y_i| \tag{5}$$

3.4.4. Anomaly evaluation index

The characteristics of the theoretical trajectory are compared with those of the actual ship trajectory using an AEI, which could be expressed by

$$\mathrm{AEI} = \frac{L_{th} - L_{ac}}{L_{th}} + \eta \left(\frac{A_{th} - A_{ac}}{L_{th}} + \frac{G^x_{th} - G^x_{ac}}{L_{th}} + \frac{G^y_{th} - G^y_{ac}}{L_{th}} \right) \tag{6}$$

where η is the length of one degree of latitude ($\eta \approx 111{,}319$). In addition, *th* represents the theoretical trajectory and *ac* represents the actual trajectory.

Generally, if the value of the AEI is greater than or equal to 0, it means that the actual trajectory is equivalent to or better than the theoretical trajectory. On the contrary, the actual trajectory is longer than the theoretical path if the AEI is less than 0. Considering the fact that the actual situation exhibits a reasonable fault tolerance interval caused by a variety of complex reasons, a threshold value of the AEI should be given if there is a need to identify the abnormal ship trajectories. Specifically, the trajectory will be considered to be abnormal if the score is less than the threshold and vice versa. Furthermore, the threshold of abnormal detection might be different for different types of ships since the role of ships (e.g., cargo/passenger transport, dredge, rescue) could affect the patterns of ship activities.

4. Case study

A case study is established to explore the advantages of our proposed DATCP method for detecting ship abnormal trajectories in the Yangtze River estuary. Meanwhile, the distributions of the abnormal ship trajectories are also explored among different ship types, time periods, weather conditions (i.e., sunny, rainy/snowy, cloudy), and ship traffic conditions (i.e., ship traffic flow, ship traffic density). The ship types include cargo ships, tankers, tug ships, carriers, container ships, dredger ships, passenger ships, fishing ships, supply ships and other ships.

4.1. Water area

The Yangtze River is the longest river in China and the third-longest river in the world, and also has the highest freight volume among the inland rivers globally. As the throat of the Yangtze River, the Yangtze River estuary is one of the most important gateways in China. All ships from the Yangtze River inner port have to pass through this channel. Shanghai Port, located in the Yangtze River estuary, is the largest container throughput port in the world. In 2014, the container throughput reached 35·285 million twenty-foot equivalent units (TEU) and the cargo throughput was 538·624 million tons. However, the hydrological environment of the Yangtze River estuary is complex and uncertain, which means it provides a great challenge to ensure safe navigation. Therefore, AIS data from the Yangtze River estuary in the year 2014 were extracted to detect abnormal ship trajectories.

4.2. Anomaly detection

According to our proposed DATCP method, an example of the anomaly detection process is shown based on an abnormal trajectory identified in the Yangtze River estuary. The detailed steps are as follows.

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

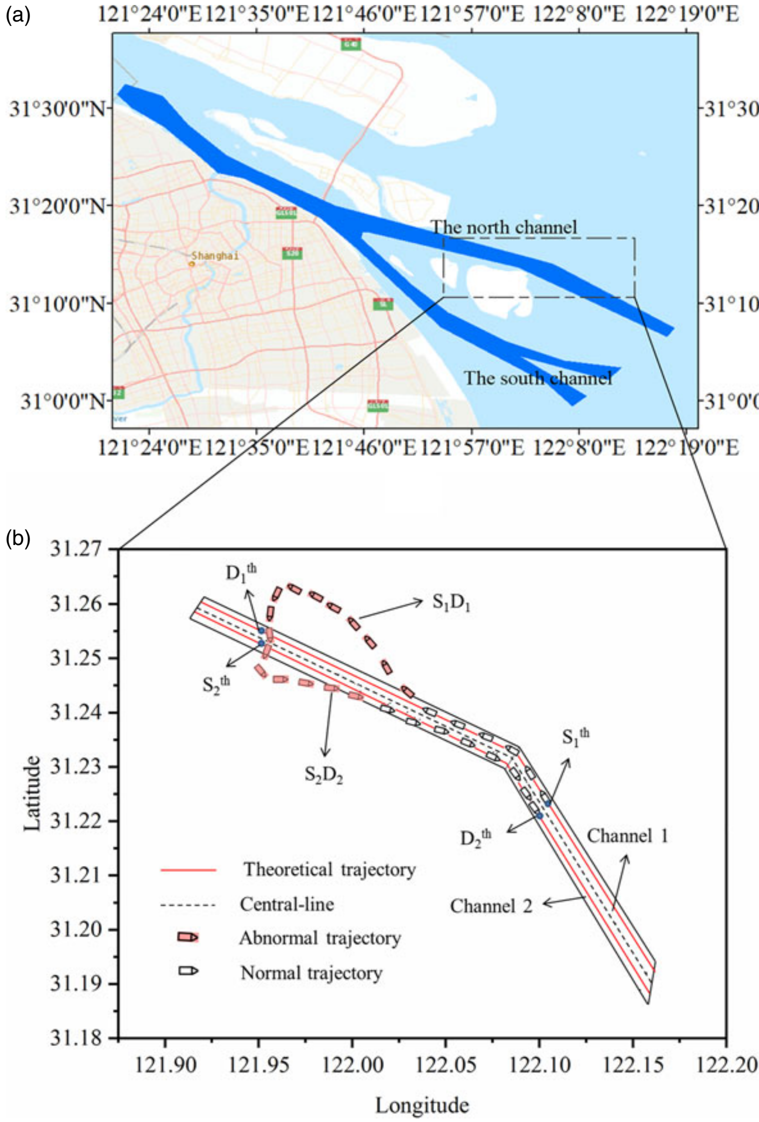

Figure 5. *An example of detected abnormal ship trajectory in the Yangtze River estuary.*

Step 1: Define the polygons
The characteristics of the north channel in the Yangtze River estuary are suitable for displaying the process of anomaly detection, as shown in Figure 5(a). Figure 5(b) presents that the complex polygon of the channel boundary is divided into simple polygons.

Step 2: Determine the theoretical trajectory
The two-way north channel can be divided by the central line. Therefore, two theoretical trajectories in opposite directions can be determined. The channel width is approximately 170 metres in each direction and the average ship width is 19 metres in the Yangtze River estuary. Therefore, ten initial starting points are set for the departure cross-section according to the channel width and the average ship width, suggesting that ten shortest paths can be obtained for each simple polygon. Subsequently, ten complete shortest paths (i.e., theoretical trajectories) can be obtained by connecting the corresponding shortest paths from each simple polygon.

Step 3: Anomaly detection

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

Table 1. *Basic information for the target ship.*

Variable	Information
MMSI	412375290
Ship type	Dredger ship
Ship length	161 metres
Ship width	27 metres
Navigating time	14:06–17:52 May 3, 2014
Weather condition	Rainy/snowy
Average speed	14·55 knots

The actual trajectories in the north channel are compared with the corresponding theoretical trajectories with the closest starting points. We select an available general threshold of the AEI for all the ship types (i.e., −1, Soleimani et al., 2015) to detect the abnormal ship trajectories for the convenience of assessing the performance of our developed anomaly detection method. An example of a dredger ship is proposed to demonstrate the remaining process of ship anomaly detection. The detailed information for the target dredger ship is shown in Table 1.

The actual trajectories and the theoretical trajectories of the dredger ship are displayed in Figure 5(b). It should be pointed out that the actual trajectory of the dredger ship is discussed separately according to the different channel directions. The two-way channel in Figure 5(b) is named channel 1 and channel 2 separately. Moreover, for theoretical trajectories, S_1^{th}, D_1^{th}, S_2^{th} and D_2^{th} represent the starting point of channel 1, the destination point of channel 1, the starting point of channel 2 and the destination point of channel 2, respectively. Here, S_1D_1 and S_2D_2 represent actual trajectories of the dredger ship in channel 1 and channel 2, respectively. Accordingly, Table 2 shows the calculation results of the trajectory length, the area under the curve and the trajectory gradient. With the calculation results in Table 2, the AEI of the dredger ship is −1·303 in channel 1 and is −1·222 in channel 2. It can be seen that both AEIs in the two directions are lower than −1, indicating that the trajectory of the dredger ship is abnormal. The detection result is consistent with the visualised actual trajectory shown in Figure 5(b).

4.3. Feature analysis of abnormal trajectories

A large amount of computation time is required for the feature analysis of the abnormal trajectories in the whole Yangtze River estuary water area because of the large number of AIS records. Therefore, a target water area for feature analysis is chosen from a smaller part of the Yangtze River estuary, with a range of longitudes from 121·49°E to 121·59°E and latitudes from 31·38°N to 31·43°N.

4.3.1. Overall abnormal ship trajectories

According to the extracted AIS data, a total of 169,589 effective actual trajectories in the target water area were obtained during 2014. For comparison, our proposed DATCP method and the A* algorithm were both applied to detect abnormal ship trajectories. Specifically, the number of abnormal trajectories identified by the DATCP method was 4249 while the number of abnormal trajectories detected by the A* algorithm was only 2981. The proportion of detected abnormal trajectories was 2·50% for the DATCP and 1·76% for the A* algorithm. With different anomaly detection methods, some invalid AEI values might come out (e.g., highly negative values) occasionally. After filtering the calculated AEI results, the DATCP method produced only 37 outliers, while the A* algorithm outputted as many as 461 outliers based on the same input data. This phenomenon indicates that our proposed method presents a higher accuracy with fewer outliers.

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

Table 2. *Calculated anomaly detection indicators for the target dredger ship.*

Channel	L_{ac}	L_{ac}	A_{th}	A_{ac}	G^x_{th}	G^x_{ac}	G^y_{th}	G^y_{ac}
1	734,723	759,884	3·20E−07	0·213	0·115	3·215	0·019	5·018
2	728,965	742,685	3·24E−07	0·206	0·114	2·962	0·020	4·850

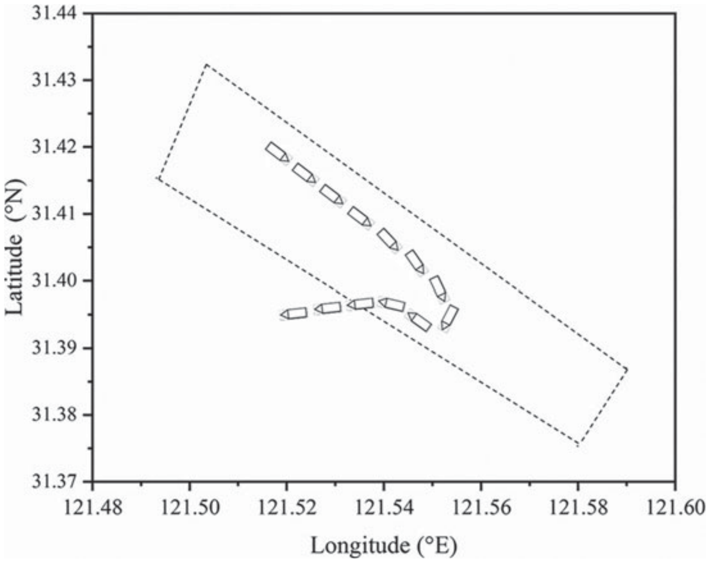

Figure 6. *An abnormal ship trajectory detected by the DATCP method.*

In addition, abnormal trajectories detected by the two methods were compared by matching the MMSI of the ships, time and location of the trajectories. The comparison results show that the abnormal trajectories identified by the A* algorithm were all included in the anomaly identified by DATCP. For the other 1268 abnormal trajectories identified by DATCP, we also verified the accuracy of anomaly detection by visualising these trajectories. The visualisation results indicate that 1048 of the trajectories were abnormal trajectories while 208 of the trajectories were gathered points with ship speeds of less than 3 knots. Only 12 normal trajectories were wrongly regarded as abnormal trajectories by the DATCP method. In other words, 80% of the additional detected abnormal trajectories were precisely detected by the DATCP method, suggesting that the DATCP proposed in this study is more accurate than the A* algorithm. Figure 6 displays an example of detected abnormal trajectory in the target water area based on the DATCP method, which was not identified by the A* algorithm. The rectangular area is the channel boundary of the target water area. The trajectory is from a bulk carrier sailing at 23:00 on January 25, 2014. As can be seen from the figure, the ship had a significant directional shift and sailed out of the channel boundary. The reason for this abnormal trajectory may have been that the helmsman changed its heading privately owing to uncertain causes.

4.3.2. Trajectory anomaly distribution among different ship types

Ship type is an important factor that could affect the behaviour of ships (Weng et al., 2020). Due to the significant characteristic difference among different ship types, the situations of trajectory anomalies may also vary among different ship types. Figure 7 presents a comparison of the anomaly detection results among different ship types using the DATCP and the A* algorithm. As seen in Figure 7, dredger ships, other ships and tankers have large differences of anomaly proportions between the two methods, namely,1·78%, 1·4% and 1·17%. However, for carriers, container ships and passenger ships,

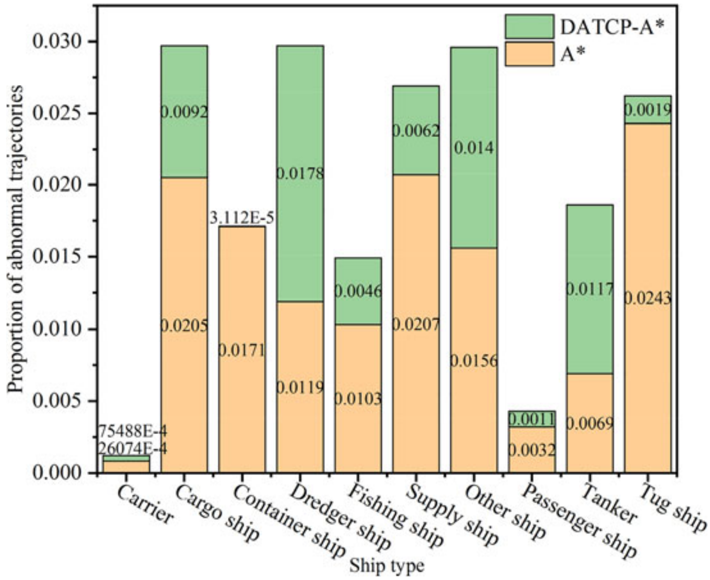

Figure 7. *Anomaly detection results for different ship types.*

there is no significant difference between the proportions of abnormal trajectories detected by the two methods. For ship types with irregular navigation behaviour like dredger ships, the A* algorithm can omit many expected abnormal trajectories based on only one theoretical trajectory (i.e., optimal path). Nevertheless, the proposed DATCP method can identify more abnormal trajectories because different theoretical trajectories are considered according to different starting points.

Figure 8 shows the number of abnormal ships and the proportion of abnormal trajectories for different ship types. It can be seen from Figure 8(a) that the abnormal cargo ships account for the largest proportion (73·12%), followed by tankers (10·76%), tug ships (8·28%), other ships (4·38%), container ships (1·67%), dredger ships (0·71%), carriers (0·38%), fishing ships (0·31%), supply ships (0·31%) and passenger ships (0·09%). In reality, the proportion of abnormal ships depends on the situation of whether the ships frequently sail in the target water area or not. The larger proportion of abnormal cargo ships may be explained by the fact that the cargo ship is the most active ship type in the Yangtze River estuary. It is also found that the proportion of abnormal trajectories varies with different ship types. Figure 8(b) shows that the two ship types associated with the highest proportion of abnormal trajectories are dredger ships (15·00%) and cargo ships (14·98%), followed by supply ships (13·58%), tug ships (13·19%), tankers (9·41%), container ships (8·64%), fishing ships (7·49%) and passenger ships (2·15%) in the Yangtze River estuary water area.

Note that the engineering ships (e.g., dredger ships, supply ships and tug ships) usually do not have specific sailing paths. Similarly, fishing ships are a little more likely to produce abnormal trajectories when sailing in fishing areas. From this point of view, ships sailing through the Yangtze River estuary should place more attention on maintaining a safe distance from engineering ships or fishing ships as they are associated with the high trajectory anomaly rates.

4.3.3. Trajectory anomaly distribution at different time periods

Figure 9 graphically depicts the temporal variation of detected trajectory anomalies in the Yangtze River estuary water area. As can be seen from the deviation lines, there are more abnormal trajectories detected by the DATCP method than the A* algorithm throughout the year. Specifically, the highest proportion of abnormal trajectories appears in May (3·27%), as shown in Figure 9(a). One possible reason may be that the ship traffic density is the largest in this month, as evidenced by Table 3. Obviously, a large number of ships may increase the occurrence frequency of abnormal trajectories. Similarly, the proportion of

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

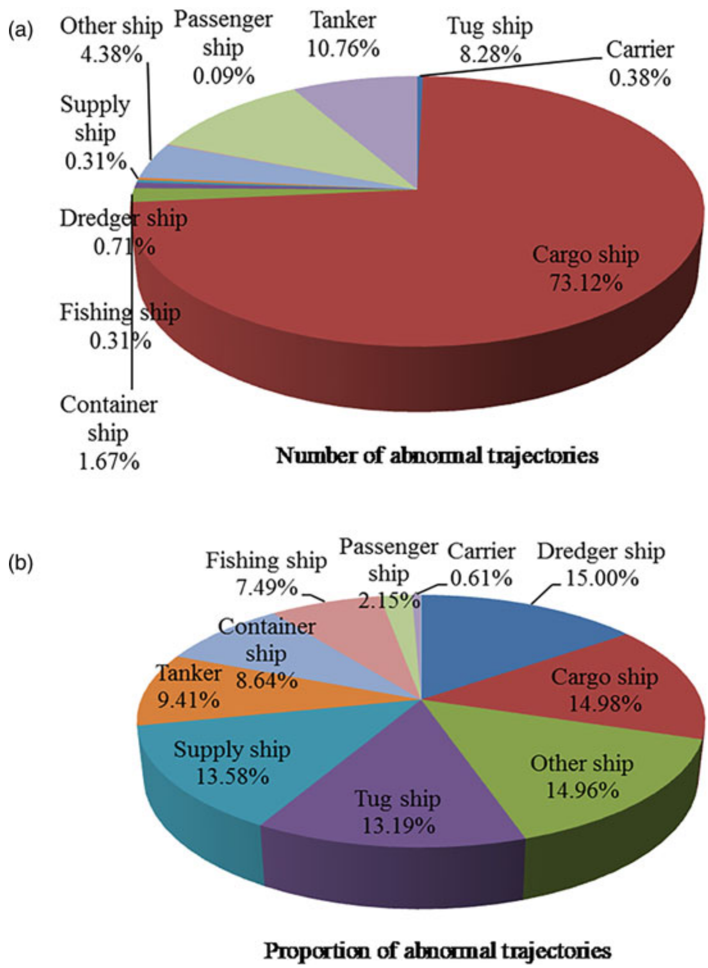

Figure 8. *Abnormal trajectory distributions for different ship types.*

abnormal trajectories is the lowest in March (1·38%) due to fewer total trajectories in this month. It is worth noting that anomaly rates are relatively higher in February and July. However, there are a relatively lower number of ship activities in these two months. Figure 9(b) shows that the occurrence probability of abnormal trajectories is the largest (1·58%) for the time period from 6:00 to 12:00, followed by the time period of 18:00–24:00 (1·24%). Note that ships have the lowest probability (0·80%) of producing abnormal trajectories during the time period from 0:00 to 6:00.

Figure 10 presents the monthly variation of anomaly rates for each ship type. As shown in Figure 10(a), the proportion of abnormal trajectories from dredger ships is the highest in July (12·5%), which is the flood season of the Yangtze River estuary. As mentioned above, the irregular paths of dredger ships are highly associated with their particular activity patterns. A large amount of sediment might accumulate in the Yangtze River estuary during the flood season, which could affect the navigability of the channel directly. Dredger ships are thus necessary to maintain the navigational safety of the Yangtze River estuary. Similar to dredger ships, the anomaly rate of supply ships shows an overall upward trend and reaches the peak in July (8·5%).

Because of the more frequent ship activities in May and June, three ship types including cargo ships, container ships and tug ships are found to produce higher anomaly rates in these two months, as shown in Figure 10(b). For instance, Table 3 shows that there are 10,119 cargo ship trajectories extracted from May, which is the highest among the whole year. Another critical finding is that tankers produce the

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

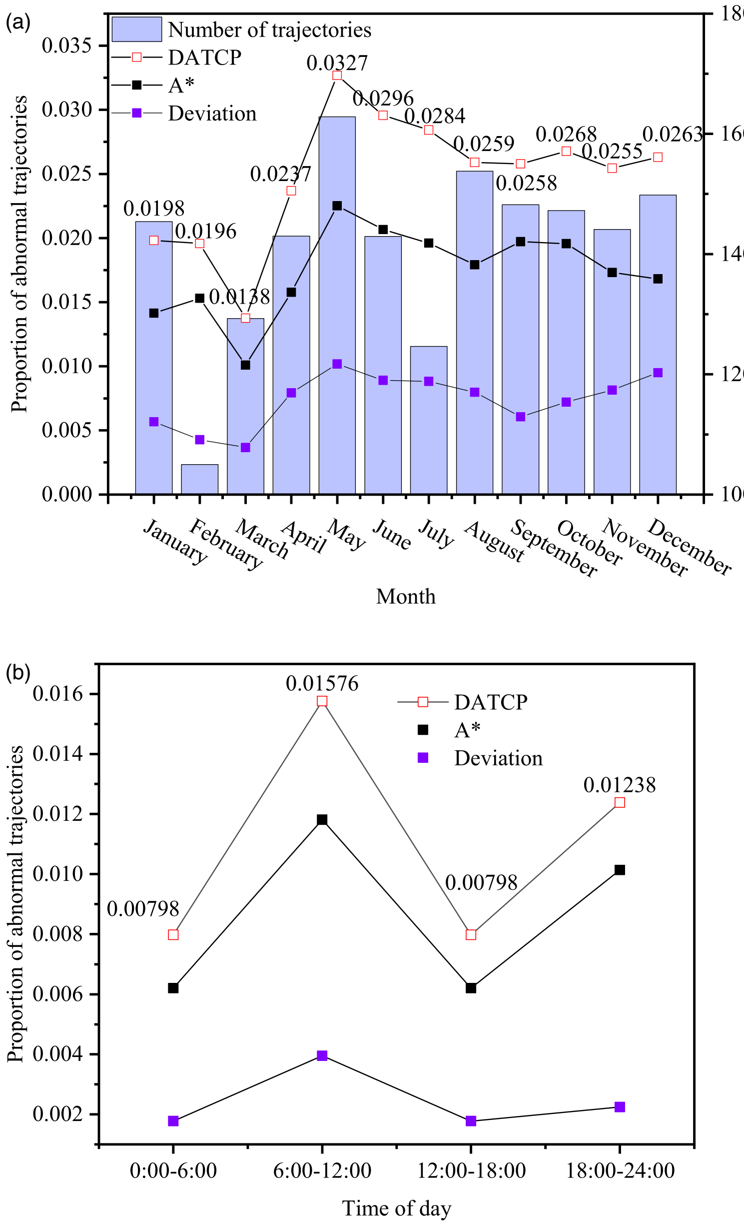

Figure 9. *Temporal effects on detected anomaly rates.*

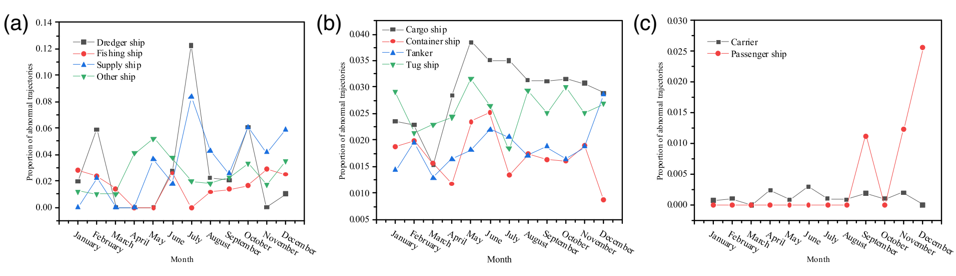

Figure 10. *Monthly anomaly rates for different ship types.*

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

Table 3. *Statistics of actual trajectories in the target water area.*

Ship type	January	February	March	April	May	June	July	August	September	October	November	December
Carrier	1358	995	1353	1254	1194	1025	919	1111	1038	957	975	1137
Cargo ship	8892	5803	8115	8575	10,119	8760	7635	9649	9261	9273	9078	9444
Container ship	373	249	385	340	383	317	298	346	366	375	368	342
Dredger ship	154	17	43	22	9	109	98	134	45	92	181	95
Fishing ship	71	42	70	82	103	73	64	85	73	60	69	80
Supply ship	58	45	42	44	55	55	24	47	39	33	24	17
Other ship	246	291	379	462	1098	427	402	439	986	391	404	746
Passenger ship	47	40	80	97	83	93	79	94	89	75	81	78
Tanker	2586	1596	1780	1818	2093	2041	1800	2109	2075	2204	2130	2275
Tug ship	1205	980	1123	1067	1170	1133	923	1161	1033	1101	1189	1374
Total	14,990	10,058	13,370	13,761	16,307	14,033	12,242	15,175	15,005	14,561	14,499	15,588

Table 4. *Distribution of abnormal ship trajectory under different weather conditions.*

Rainy/snowy	Cloudy	Sunny
40·58%	30·70%	28·72%

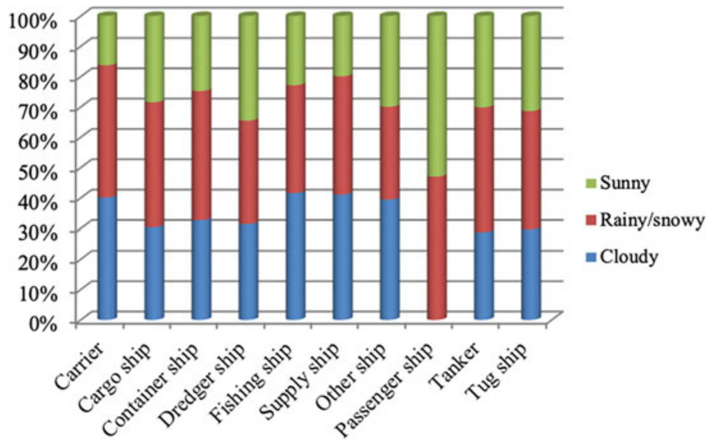

Figure 11. *Ship anomaly distributions under different weather conditions.*

highest proportion of anomalies in December. Since tankers are one of the highly hazardous ship types, special attention shall be paid to the anomaly navigation of tankers. Furthermore, Figure 10(c) reveals that the anomaly rates of passenger ships are always close to zero, suggesting that passenger ships could strictly comply with their fixed routes.

4·3.4. Trajectory anomaly distribution under different weather conditions

Weather conditions at sea have always been a critical factor in maritime safety-related works. According to previous studies, adverse weather conditions can increase the risk of collisions (Vettor and Soares, 2017; Weng et al., 2020). Similarly, there may be a high likelihood of a ship anomaly under adverse weather conditions. In this study, we collected the weather information of the Yangtze River estuary from the National Maritime Data and Information Service (NMDIS). The collected weather conditions mainly consist of three weather categories, including sunny, rainy/snowy and cloudy.

Table 4 tabulates trajectory anomaly distributions under different weather conditions. It can be seen from Table 4 that the anomaly proportion under rain/snow weather conditions is the highest (40·58%), followed by cloudy days (30·7%) and sunny days (28·72%). Note that adverse weather conditions are usually associated with poor visibility and a short visible range for ships, which might cause the ships to deviate from their normal routes. Therefore, more attention should be paid to the ships navigating under severe weather conditions. Moreover, anomaly distributions under different weather conditions are also presented for each ship type, as shown in Figure 11. In general, it is found that approximately 40% of abnormal trajectories have occurred under rainy/snowy weather conditions.

4.3.5. Trajectory anomaly distribution under different ship traffic conditions

Theoretically, ship traffic flow and density might influence the likelihood of abnormal trajectories. Figure 12 shows the effects of ship traffic conditions characterised by ship traffic flow and density on the abnormal trajectory distributions. Figure 12(a) demonstrates that the number of abnormal trajectories generally increases with the ship traffic flow. The median quartile of the abnormal trajectories reaches a peak (i.e., approximately 6 times/day) when the ship traffic flow comes to 500–600 ships/day. However, the number of abnormal trajectories is less than 2·5 times/day when the ship traffic flow is below

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

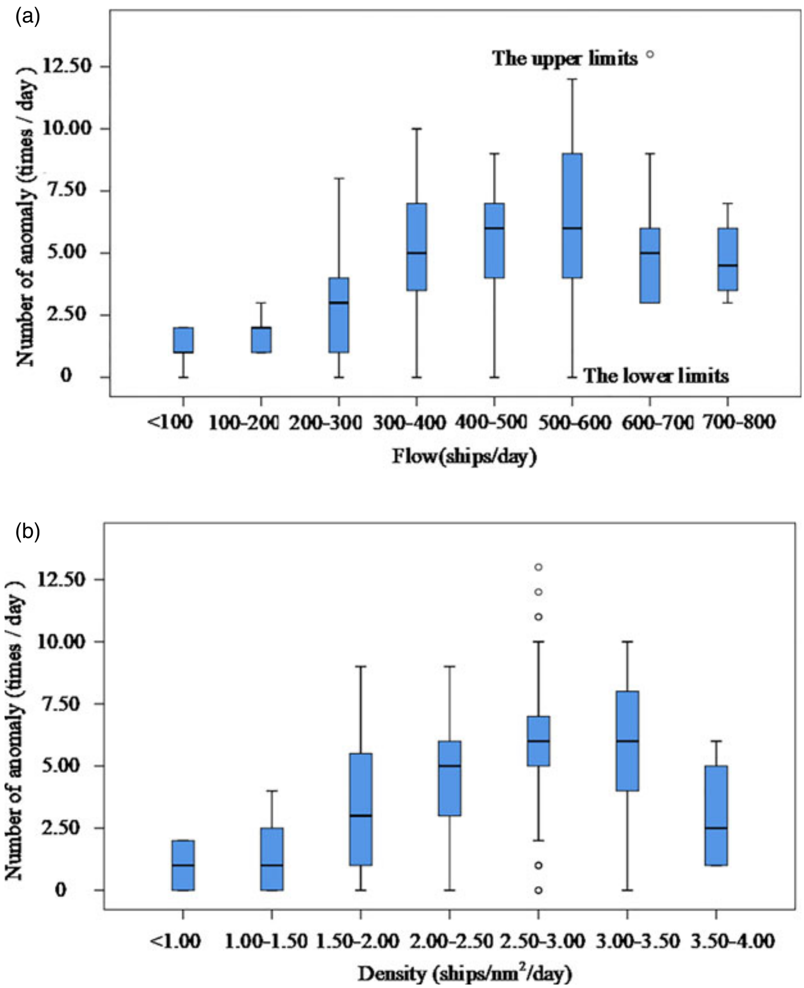

Figure 12. *Ship anomaly distribution under different ship traffic flow and ship traffic density.*

100 ships/day. Similarly, the number of abnormal trajectories also increases with ship traffic density, as shown in Figure 12(b). The upper limit of ship anomalies reaches 10 times/day when the shipping density remains between 2·5 and 3·5 ships/nm^2/day. An interesting finding is that the number of abnormal ship trajectories decreases slightly when the traffic flow is greater than 600 ships/day or the shipping density is greater than 3·5 ships/nm^2/day. This suggests that the majority of ships would like to keep on their normal routes under congested traffic conditions.

5. Conclusions

The ship anomaly detection in maritime surveillance is a key step towards the application of navigation safety enhancement strategies for maritime authorities. However, the effect of channel boundaries has rarely been considered in the past, which might lead to biased results for anomaly detection. Therefore, this study endeavours to fill in the gaps by proposing the DATCP method to detect abnormal behaviours of ships. The abnormal trajectories could be identified by comparing the actual trajectories against the calculated theoretical trajectories through the AEI score. Complex polygons simplified by the channel boundaries are applied to calculate theoretical trajectories. Based on the AIS data from the Yangtze

https://doi.org/10.1017/S0373463322000182 Published online by Cambridge University Press

River estuary in 2014, this study finally creates one case study to validate the capability of the proposed DATCP method in detecting abnormal ship trajectories.

Case study results reveal that the proposed DACTP method performs better than the traditional A* algorithm because the former could detect more abnormal ship trajectories that were missed by the A* algorithm. A total of 4249 abnormal trajectories are detected by the DACTP method from the 169,589 actual trajectories in the target waters. Dredger ships are found to generate a higher proportion of abnormal trajectories, as compared with other ship types. Especially, the highest anomaly rate of dredger ships occurs in the Yangtze River estuary during the flood season. Since engineering ships (e.g., dredger ships, supply ships and tug ships) or fishing ships usually do not have specific sailing paths, ships sailing through the Yangtze River estuary should place more attention on maintaining a safe distance from these ships that are associated with the high trajectory anomaly rates. The temporal analysis results reveal that the high proportion of abnormal trajectories is closely associated with the flood season, the typhoon season and the Chinese Spring Festival. Moreover, the anomaly rate is the highest during the time period from 6:00 to 12:00 in this water area. A large number of abnormal trajectories are detected on rainy/snowy days. This implies that special concern should be paid to avoid the occurrence of abnormal behaviour under adverse weather conditions. In addition, the results also show that ships are most likely to generate abnormal trajectories when the ship traffic flow reaches 500 ships/day or the ship traffic density is higher than 2·5 ships/nm^2/day in this water area.

Case study results demonstrate that the proposed DACTP method could accurately detect the abnormal trajectories, which is beneficial for maritime authorities to put forward more effective countermeasures to reduce the anomaly rate. However, due to data limits, the AEI thresholds of abnormal trajectory for different ship types have not been investigated in this study. We are now working on collecting real case data of abnormal ship trajectories for different types of ships. Our future work will focus on determining the AEI thresholds for different ship types based on our developed abnormal detection methods and the obtained real ship anomaly cases. In addition, the detection performance of our proposed DATCP method may be highly affected by the polygon segmentation scheme (e.g., polygons or grids). Theoretically, the optimal polygon segmentation scheme could be determined by minimising the difference between the number of detected abnormal trajectories against the observed data. Therefore, we will further investigate the detection accuracy under different polygon segmentation schemes for a specific waterway after collecting the observed abnormal ship trajectories in the future.

Acknowledgements. The authors sincerely thank the anonymous referees for their helpful comments and valuable suggestions, which considerably improved the exposition of this work. This study is supported by the National Natural Science Foundation of China (Grant No. 52072237).

References

Anneken, M., Fischer, Y. and Beyerer, J. (2015). Evaluation and Comparison of Anomaly Detection Algorithms in Annotated Datasets From the Maritime Domain. *SAI Intelligent Systems Conference*. IEEE.

Hodge, V. J. and Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, **22**(2), 85–126.

Kaluđer, H., Brezak, M. and Petrović, I. (2011). A Visibility Graph Based Method for Path Planning in Dynamic Environments. *International Convention MIPRO*. IEEE.

Lane, R. O., Nevell, D. A., Hayward, S. D. and Beaney, T. W. (2011). Maritime Anomaly Detection and Threat Assessment. *International Conference on Information Fusion*. IEEE.

Laxhammar, R. (2008). Anomaly Detection for sea Surveillance. *International Conference on Information Fusion*. IEEE.

Laxhammar, R., Falkman, G. and Sviestins, E. (2009). Anomaly Detection in sea Traffic - A Comparison of the Gaussian Mixture Model and the Kernel Density Estimator. *International Conference on Information Fusion*. IEEE.

Li, H., Liu, J., Liu, R. W., Xiong, N., Wu, K. and Kim, T. (2017). A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis. *Sensors*, **17**(8), 1792.

Maria, R., Giuliana, P. and Michele, V. (2018). Maritime anomaly detection: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8**, e1266.

Pallotta, G., Vespe, M. and Bryan, K. (2013a). Traffic Knowledge Discovery From AIS Data. *International Conference on Information Fusion*. IEEE.

Pallotta, G., Vespe, M. and Bryan, K. (2013b). Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*, **15**(6), 2218–2245.

Pan, J., Jiang, Q. and Shao, Z. (2014). Trajectory clustering by sampling and density. *Marine Technology Society Journal*, **48**(6), 74–85.

Perera, L. P., Oliveira, P. and Soares, C. G. (2012). Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction. *IEEE Transactions on Intelligent Transportation Systems*, **13**(3), 1188–1200.

Rhodes, B. J., Bomberger, N. A. and Zandipour, M. (2007). Probabilistic Associative Learning of Vessel Motion Patterns at Multiple Spatial Scales for Maritime Situation Awareness. *International Conference on Information Fusion*. IEEE.

Rong, H., Teixeira, A. P. and Soares, C. G. (2019). Ship trajectory uncertainty prediction based on a Gaussian process model. *Ocean Engineering*, **182**, 499–511.

Shahir, H. Y., Glsser, U., Nalbandyan, N. and Wehn, H. (2014). Maritime Situation Analysis: A Multi-Vessel Interaction and Anomaly Detection Framework. *Intelligence & Security Informatics Conference*. IEEE.

Soleimani, B. H., Souza, E., Hilliard, C. and Matwin, S. (2015). Anomaly Detection in Maritime Data Based on Geometrical Analysis of Trajectories. *International Conference on Information Fusion*. IEEE.

Tun, M. H., Chambers, G. S. and Tan, T. (2007). Maritime Port Intelligence Using AIS Data. *RNSA Security Technology Conference*, 33–43.

Varun, C., Arindam, B. and Vipin, K. (2009). *Anomaly Detection: A Survey*. Minneapolis, USA: Acm Computing Surveys.

Vettor, R. and Soares, C. G. (2017). Characterisation of the expected weather conditions in the main European coastal traffic routes. *Ocean Engineering*, **140**, 244–257.

Weng, J., Liao, S., Wu, B. and Yang, D. (2020). Exploring effects of ship traffic characteristics and environmental conditions on ship collision frequency. *Maritime Policy & Management*, **47**(4), 523–543.

Xu, T., Liu, X. and Xin, Y. (2012). A novel approach for ship trajectory online prediction using bp neural network algorithm. *Advances in Information Sciences & Service Sciences*, **4**(11), 271–277.

Zhang, S., Shi, G., Liu, Z., Zhao, Z. and Wu, Z. (2018). Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity. *Ocean Engineering*, **155**, 240–250.

Zhao, L. and Shi, G. (2019). Maritime anomaly detection using density-based clustering and recurrent neural network. *Journal of Navigation*, **72**(4), 894–916.

Zhen, R., Jin, Y., Hu, Q., Shao, Z. and Nikitakos, N. (2017a). Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naive Bayes classifier. *Journal of Navigation*, **70**(3), 1–23.

Zhen, R., Riveiro, M. and Jin, Y. (2017b). A novel analytic framework of real-time multi-vessel collision risk assessment for maritime traffic surveillance. *Ocean Engineering*, **145**(15), 492–501.

Zhou, B. and Shi, A. (2010). LSSVM and Hybrid Particle Swarm Optimization for Ship Motion Prediction. *Intelligent Control and Information Processing (ICICIP)*, 13–15 August 2010, Dalian, China, 183–186.

Cite this article: Weng J, Li G, Zhao Y (2022). Detection of abnormal ship trajectory based on the complex polygon. *The Journal of Navigation* **75**: 4, 966–983. https://doi.org/10.1017/S0373463322000182