

Article

Identification of Spoofing Ships from Automatic Identification System Data via Trajectory Segmentation and Isolation Forest

Hailin Zheng ^{1,2} , Qinyou Hu ^{1,*} , Chun Yang ¹, Qiang Mei ^{1,3}, Peng Wang ^{1,4} and Kelong Li ²

¹ Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China; hlzhzjou@126.com (H.Z.); chunyang@shmtu.edu.cn (C.Y.); meiqiang@jmu.edu.cn (Q.M.); wangp@ict.ac.cn (P.W.)

² School of Naval Architecture and Maritime, Zhejiang Ocean University, Zhoushan 316022, China; 18158057677@163.com

³ Navigation Institute, Jimei University, Xiamen 361021, China

⁴ Institute of Computing Technology, Chinese Academy of Sciences, Beijing 101408, China

* Correspondence: qyhu@shmtu.edu.cn

Abstract: Outliers of ship trajectory from the Automatic Identification System (AIS) onboard a ship will affect the accuracy of maritime situation awareness, especially for a regular ship trajectory mixed with a spoofing ship, which has an unauthorized Maritime Mobile Service Identification code (MMSI) owned by a regular ship. As has been referred to in the literature, the trajectory of these spoofing ships would simply be removed, and more AIS data would be lost. The pre-processing of AIS data should aim to retain more information, which is more helpful in maritime situation awareness for the Maritime Safety Administration (MSA). Through trajectory feature mining, it has been found that there are obvious differences between the trajectory of a regular ship and that of a regular ship mixed with a spoofing ship, such as in terms of speed and distance between adjacent trajectory points. However, there can be a long update time interval in the results of severe missing trajectories of a ship, bringing challenges in terms of the identification of spoofing ships. In order to accurately divide the regular ship trajectory and spoofing ship trajectory, combined with trajectory segmentation by the update time interval threshold, the isolation forest was adopted in this work to train the labeled trajectory point of a regular ship mixed with a spoofing ship. The experimental results show that the average accuracy of the identification of spoofing ships using isolation forest is 88.4%, 91%, 93.1%, and 93.3%, corresponding to different trajectory segmentation by update time intervals (5 h, 10 h, 15 h, and 20 h). The research conducted in this study can almost eliminate the outliers of ship trajectory, and it also provides help for maritime situation awareness for the MSA.

Keywords: automatic identification system; spoofing ship; missing points; jumping points; trajectory segmentation; isolation forest



Citation: Zheng, H.; Hu, Q.; Yang, C.; Mei, Q.; Wang, P.; Li, K. Identification of Spoofing Ships from Automatic Identification System Data via Trajectory Segmentation and Isolation Forest. *J. Mar. Sci. Eng.* **2023**, *11*, 1516. <https://doi.org/10.3390/jmse11081516>

Academic Editor: Sergei Chernyi

Received: 29 June 2023

Revised: 19 July 2023

Accepted: 28 July 2023

Published: 29 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

On 1 June 2020, the special rectification of national maritime communication order for radio equipment on board ship began, such as AIS, very-high-frequency communication (VHF), and so on. The China MSA at all levels have concentrated on the monitoring of maritime communication order and improving the ability of maritime communication supervision and maritime service support in China. The special rectification focuses on the rectification of outstanding problems, such as irregularly authorized ship MMSI, one MMSI owned by several ships, several MMSI owned by one ship, the illegal occupation of channels, and the violation of communication order.

Among the violations of maritime communication order mentioned above, the irregular use of AIS may have a significant impact on the quality of AIS data [1]. The quality of AIS data is a subject of interest for many researchers [2–5], but published research on pre-processing raw data to improve quality is limited. Shelmerdine took the development of a vessel database as the key to managing AIS data and for quality control [6]. All fields

were checked for obvious outliers. If it was not possible to correct an outlier, it was removed. The common method to filter inaccurate single position points is the threshold of position, speed, and course [7]. To solve the problem of sharing MMSI numbers, a method of elimination was applied by Pallotta et al. [8]. Mazzarella et al. proposed a nearest-neighbor approach to assign AIS messages to the right tracks, but there was no detailed experimental method, performance, or results [9]. Wu et al. created a simple algorithm to calculate the likelihood of an association between an AIS message and each candidate vessel [10]. It is used for processing massive data on a global scale, but it cannot be applied in a small region where AIS messages are sampled at a high rate. The reason is that the algorithm is unable to handle an association in the case where there are at least three consecutive abnormal trajectory points. A similar method with speed threshold was proposed by Greidanus et al. [11]. Given that none of these techniques are universally applicable, it is necessary to propose a method with general applicability.

Wei et al. observed an abnormal 'jumping point' in ship trajectory by calculating the speed between adjacent ship trajectory points and setting the speed threshold according to ship maneuvering characteristics, which meant that abnormal trajectory points were identified [12]. Since the constant velocity threshold method does not consider the change in the motion state of a moving ship at different times, it can only detect some abnormal points whose velocity exceeds the specified threshold, and the robustness of this method is poor. Han et al. proposed a novel trajectory outlier detection algorithm based on the adaptive threshold, designing a local threshold window and mean filter window, and calculated the local speed (acceleration) threshold and global speed (acceleration) threshold, and found three classes of abnormal trajectory points, including isolated outliers, continuous outliers, and obvious outliers [13]. Zhang et al. calculated the bow deflection rate between adjacent trajectory points by counting the speed distribution of ship trajectory points, setting the speed threshold of trajectory points according to the probability of the speed distribution of adjacent trajectory points, setting the threshold of the bow deflection rate according to the characteristics of ship cycle, and identified abnormal trajectory points [14]. Liu et al. converted ship speed, heading, and position from AIS data into evidence reliability and used evidence reasoning rule synthesis to detect three classes of trajectory points, referring to the manual identification method of abnormal AIS data adopted by the MSA [15,16]. Chen et al. and Guo et al. cleaned ship AIS data in these three rules: abnormal ship position (the longitude and latitude of ship are beyond the scope of study area), abnormal speed (the difference between adjacent trajectory points exceeds speed threshold), and abnormal rate of turning (the course difference between adjacent trajectory segments exceeds rate of turning threshold) [17–20]. Data derived from AIS plays a key role in water traffic data mining. However, there are various errors regarding time and space. To improve availability, AIS data quality dimensions are presented by Zhao et al. to detect errors of AIS tracks, including physical integrity, spatial logical integrity, and time accuracy [21–23]. After systematic summary and analysis, algorithms for error pre-processing are proposed. In the aspect of abnormal AIS data identification, combined with the characteristics of adjacent trajectory points in a period of time, an abnormal AIS data identification model based on BP neural network was constructed by Wang et al. [24].

Zhang et al. designed an MMSI spoofing detection algorithm based on the spatiotemporal data provided by AIS and radar. When a ship is monitored by AIS and radar before and after MMSI spoofing, both monitoring processes continue for a period of time, meaning the MMSI spoofing algorithm demonstrates a good performance [25]. Iphar et al. propose a rule-based method for data integrity assessment, with rules built from the system technical specifications and by domain experts, formalized by a logic-based framework, resulting in the triggering of situation-specific alerts [26–28]. Jeong et al. provided an automatic shipping route construction method using functional data analysis (FDA), and the proposed approach includes two steps: outlier detection and shipping route construction [29]. Huang et al. proposed a new method for detecting anomalous vessel dynamics using functional data analysis. Empirical investigations of this approach demonstrate the effective detection

of outlier flows in terms of ship traffic volume [30]. In summary, researchers regard outliers in trajectory points as random error values and delete them in order to clean the trajectory. However, these outliers may be the trajectory point of another ship, that is, the trajectory of a spoofing ship sharing the same MMSI with other regular ship. Moreover, due to a large number of missing trajectories, the speed between adjacent trajectory points cannot be used as a basis for distinguishing the trajectory points of a spoofing ship and regular ship.

In this paper, we aim to propose a novel spoofing ship identification framework with the support of trajectory segmentation. Our main contributions can be summarized as follows: (1) We mined the trajectory feature of a regular ship and spoofing ship and obtained the correlation between the time interval and distance and average sailing speed between the adjacent trajectory points for a regular ship and spoofing ship. (2) We segmented ship trajectory by considering the ratio of missing trajectory points and distribution of the time interval between adjacent trajectory points, and obtained the trajectory segment with a low missing points ratio. (3) Considering the low ratio of jumping trajectory points in the data sample and the higher identification efficiency of isolation forest, we adopted isolation forest to identify a spoofing ship and testified the proposed framework performance on 20 regular ship mixed with spoofing ship trajectories. We aim for this study to be able to help the MSA identify spoofing ship trajectories and thus take early warning measurements to enhance maritime traffic efficiency and safety. The remainder of this paper is organized as follows. We introduce the data source used in our study in Section 2. After that, the methodology details about trajectory feature mining of the AIS data are illustrated in Section 3, and then isolation forest used for identifying spoofing ships is presented, combined with trajectory segmentation. The experimental results are shown in Section 4. Section 5 briefly discusses the study and illustrates future work.

2. Data

Shanghai Meili Shipbuilding Technology Co., Ltd. (Shanghai, China) provides large-scale AIS data, which benefit many AIS-relevant studies due to their public accessibility (<https://www.hifleet.com/>, accessed on 1 June 2023). The hifleet has online access to over 50 AIS satellites and over 3000 AIS base stations, receiving 150 million AIS data per day, as well as purchase Lloyd's ship archives and access global electronic chart data and ocean meteorological data. The original AIS dataset includes both kinematic and static information for a ship, which contains the MMSI, latitude, longitude, speed over ground (SOG), heading, course over ground (COG), timestamp, call sign, port of call, and so on.

When collecting records of the port of call for container ships, it was found that some ships continuously call at ports that are far apart within a short time interval. It was found that there were jumping points in the trajectories of these ships when selecting the real-time trajectories of these container ships within the corresponding statistical time. Due to the presence of jumping points in these trajectories, the trajectories of these container ships also exhibit jumping characteristics, rather than showing continuity like those of regular ships. Consequently, we collected the AIS data of container ships which have trajectory jumping points and labeled them with different colors for different ships. We collected 20 container ships and 52,538 AIS data samples from 1 January 2017 to 31 December 2017 (see Figure 1), and the average time interval for sampling the AIS data was 1 h.

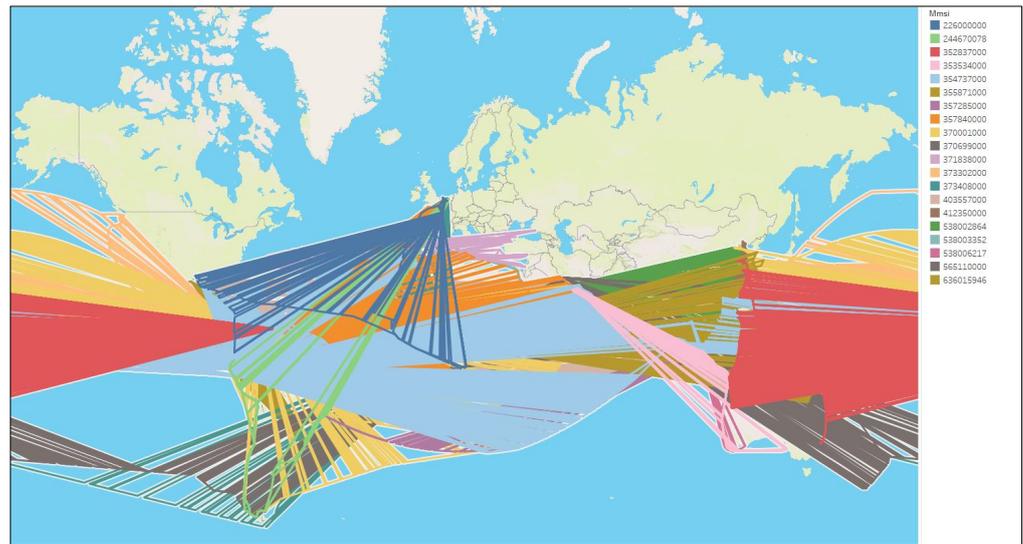


Figure 1. Raw AIS data of regular container ships mixed with spoofing ships.

3. Methodology

We were affected by the limited coverage of AIS base stations, limited AIS communication capacity, and the illegal use of unauthorized ship-borne AIS, such as a ship leaving the coverage area of AIS base stations, the quantity of ships exceeding AIS communication capacity, the deliberate closure of ship-borne AIS, the use of the same MMSI for multiple ships or the use of multiple MMSI for one ship, resulting in large quantities of missing or jumping trajectories. To address this issue, we firstly implemented trajectory characteristics mining and trajectory segmentation to obtain the distribution of speed and distance between adjacent trajectory points and then identified spoofing ships using isolated forest. The schematic overview for the proposed framework is shown in Figure 2. In order to accurately describe the ship motion pattern, ship trajectory is defined as follows:

$$S = \{\{S_1\}, \{S_2\}, \dots, \{S_i\}, \dots, \{S_m\}\} \tag{1}$$

$$S_i = \{s_i^1, s_i^2, \dots, s_i^k, \dots, s_i^n\} \tag{2}$$

$$s_i^k = (m_i^k, t_k, \lambda_i^k, \phi_i^k, v_i^k, c_i^k) \tag{3}$$

In Equation (1), $\{S_i\}$ represents the trajectory of ship i . In Equation (2), s_i^k represents the trajectory point of ship i at time t_k . In Equation (3), m_i^k represents the MMSI of ship i , t_k represents the update time of AIS data, λ_i^k and ϕ_i^k represents the ship longitude and latitude at time t_k , v_i^k represents ship speed at time t_k , and c_i^k represents ship course at time t_k .

3.1. Trajectory Feature Mining

According to the trajectory point distribution for missing and jumping ship trajectory, ship trajectory points are divided into four categories: regular ship trajectory points (Normal_Point, abbreviated as N_P), spoofing ship trajectory points (Spoofing_Point, abbreviated as S_P), and confusion points (No Labeled_Point, abbreviated as NL_P).

In order to accurately detect spoofing ship trajectory points, the distance between adjacent trajectory points and average sailing speed are two important parameters. Generally speaking, average sailing speed between adjacent trajectory points is consistent with the ship maneuvering performance for regular ship. Taking the cargo ship as an example, the speed of this ship would not exceed 50 knots. Therefore, two adjacent trajectory points whose average sailing speed exceeds the speed threshold must not belong to the same ship;

accordingly, the trajectory points of regular ships mixed with a spoofing ship can be effectively identified. The average speed between adjacent trajectory points is closely related to the update time interval of trajectory points, as well as the distance between adjacent trajectory points. The distance between adjacent trajectory points could be calculated by spherical distance (namely, Great Circle distance), as calculated in Equation (4), and speed $v_{avg_i}^{k, k+1}$ between adjacent trajectory points was calculated, as in Equation (5):

$$d_i^{k, k+1} = a \cos\left(\sin(\phi_i^{k+1}) * \sin(\phi_i^k) + \cos(\phi_i^{k+1}) * \cos(\phi_i^k) * \cos(\lambda_i^{k+1} - \lambda_i^k)\right) \quad (4)$$

$$v_{avg_i}^{k, k+1} = \frac{d_i^{k, k+1}}{t_{(k+1)} - t_k} = \frac{d_i^{k, k+1}}{\Delta t} \quad (5)$$

In Equation (5), Δt represents the update time interval of the trajectory points. If the update time interval of the trajectory point is not affected by the working performance of the AIS base station and traffic density, and is only related to ship speed, the trajectory point would update more frequently. At this time, the distance between adjacent trajectory points and the speed of navigation have good discrimination between regular ship trajectory and regular ship trajectory mixed with spoofing ship. The average sailing speed between regular ship trajectories is within the speed threshold, as shown in Equation (6), while the average sailing speed among a regular ship trajectory mixed with a spoofing ship is beyond the speed threshold, as shown in Equation (7).

$$\Delta t < \Delta t_{(th)}, d_i^{k, k+1} < d_{th} \wedge v_{avg_i}^{k, k+1} < v_{avg(th(min))} \quad (6)$$

$$\Delta t < \Delta t_{(th)}, d_i^{k, k+1} \geq d_{th} \wedge v_{avg_i}^{k, k+1} \geq v_{avg(th(max))} \quad (7)$$

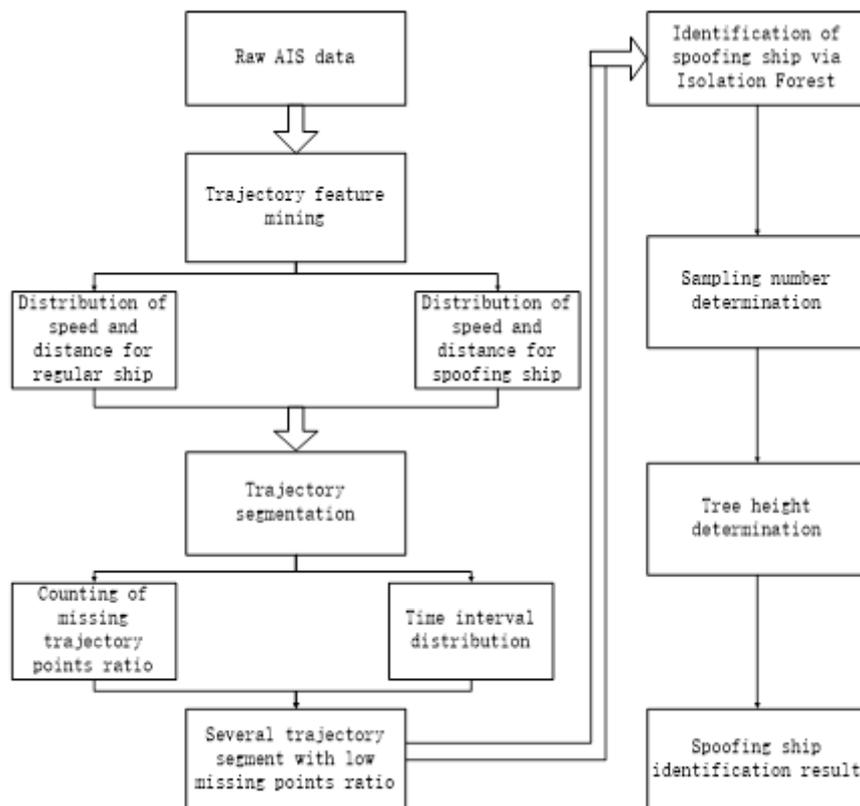


Figure 2. Overview for the proposed spoofing ship identification frame work.

In Equation (6), $\Delta t_{(th)}$ represents the threshold of the update time interval between adjacent trajectory points. $v_{avg(th(min))}$ represents the threshold of minimum speed between adjacent trajectory points, that is, the normal navigation speed between adjacent trajectory points belonging to the same ship, abbreviated as $v_{th(min)}$. d_{th} represents the threshold of minimum distance between adjacent trajectory points, corresponding to $v_{th(min)}$.

In Equation (7), $v_{avg(th(max))}$ represents the threshold of maximum speed between adjacent trajectory points, which is much larger than the normal navigation speed between the adjacent trajectory points of a regular ship, abbreviated as $v_{th(max)}$.

When the continuity of ship trajectories is good, that is, the time interval between trajectory points is short, there is a clear distinction between the distance and average sailing speed between regular ship trajectory points compared to the regular ships mixed with spoofing ships. However, due to the ship trajectory being missing, the time interval between ship trajectory points becomes longer, and the average sailing speed between them will be confused, making it difficult to identify spoofing ships. Therefore, trajectory segmentation is very necessary as it can convert a poorly continuous ship trajectory into several well continuous trajectory segments, which helps to identify the trajectory points of spoofing ships.

3.2. Trajectory Segmentation

For missing ship trajectory, re-emerged regular ship trajectory points may be identified as the trajectory points of a spoofing ship because the distance exceeds the corresponding threshold. At the same time, the re-emerged trajectory points of a spoofing ship may be misjudged as regular ship trajectory points due to the speed between adjacent trajectory points being within a corresponding speed threshold, as shown in Equation (8).

According to Section 3.1 of the paper, with the increase in the time interval between adjacent trajectory points, the average sailing speed between regular ship trajectory points remains unchanged, but the distance between trajectory points will gradually increase. However, the distance between adjacent trajectory points for the regular ship mixed with spoofing ship remains unchanged, but the average sailing speed between adjacent trajectory points will gradually decrease. Therefore, through trajectory feature mining in Section 3.1, it can be observed that when the time interval between adjacent trajectory points increases to a certain value, the distance between regular ship trajectory points is close to that between regular ship trajectory points mixed with spoofing ships, or the average sailing speed between regular ship trajectory points mixed with spoofing ships is close to that between regular ship trajectory points. Consequently, this time interval can be used as the threshold for trajectory segmentation.

Moreover, the time interval threshold for trajectory segmentation varies due to the distance between the trajectory points of the spoofing ship and the regular ship trajectory points. For spoofing ship trajectory points that are close to regular ship trajectory points, or overlapped with regular ship trajectory, when the time interval between adjacent trajectory points is small, trajectory features for this class of a regular ship mixed with a spoofing ship would be similar to regular ships. Therefore, it is necessary to set a small time interval threshold for trajectory segmentation. For spoofing ship trajectory points that are far away from regular ship trajectory points, these two ship trajectories will not overlap. Only when the time interval between adjacent trajectory points is large will the regular ship trajectory characteristics mixed with a spoofing ship be similar to regular ships. Therefore, a larger time interval threshold can be set for trajectory segmentation.

In order to avoid error identification for missing ship trajectory points, ship trajectory could be segmented according to the threshold of the update time interval, as shown in Equations (9)–(12).

$$\Delta t \geq \Delta t_{(th)}, d_i^{k,k+1} \geq d_{th} \wedge v_{avg_i}^{k,k+1} < v_{avg(th(min))} \tag{8}$$

$$S_i = \{TR_i^1, TR_i^2, \dots, TR_i^j, TR_i^k, \dots, TR_i^m\} \tag{9}$$

$$TR_i^j = \{s_i^{j+1}, s_i^{j+2}, \dots, s_i^k \mid t_{k+1} - t_k > \Delta t_{th}\} \tag{10}$$

$$TR_i^k = \{s_i^{k+1}, s_i^{k+2}, \dots, s_i^{k+m} \mid t_{k+(m+1)} - t_{k+m} > \Delta t_{th}\} \tag{11}$$

$$flag(s_i^{k+1}) = \begin{cases} NN_P, \Delta t^1 \geq \Delta t_{(th)} \wedge flag(s_i^k) = N_P \\ NN_P, \Delta t^2 \geq \Delta t_{(th)} \wedge flag(s_i^k) = S_P \end{cases} \tag{12}$$

In Equations (9)–(11), TR_i^j and TR_i^k represent the trajectory segmented by the corresponding time interval threshold. In Equation (12), s_i^k represents the trajectory point of ship i at time t_k , and s_i^{k+1} represents the trajectory point of ship i at time $t_{(k+1)}$. $flag(s_i^k)$ represents the class of trajectory point s_i^k , and $flag(s_i^{k+1})$ represents the class of trajectory point s_i^{k+1} . Δt^1 represents the time difference between the current trajectory point and latest time trajectory point in N_P. Δt^2 represents the time difference between the current trajectory point and latest time trajectory point in S_P. NN_P represents a new class of points derived from the missing ship trajectory, namely the trajectory points of a new regular ship.

3.3. Identification of Spoofing Ship via Isolation Forest

The trajectory segment characteristics between a regular ship and a regular ship mixed with a spoofing ship have an obvious difference, and the proportion of the abnormal trajectory segment is small. In Figure 1, there are only 10 percent of spoofing ship trajectory points included in the overall AIS data sample. Therefore, the isolation forest is applicable for spoofing ship identification for the AIS data sample of the paper. Ship trajectory is divided into a set of trajectory segments composed of adjacent trajectory points. In Equation (13), TR_i^j is a set of trajectory segments composed of the adjacent trajectory points of ship i , defined as follows:

$$TR_i^j = \{tr_i^{j+1,j+2}, tr_i^{j+2,j+3}, \dots, tr_i^{k-1,k} \mid 0 \leq j \leq k, 1 \leq k \leq n\} \tag{13}$$

$$tr_i^{k-1,k} = (d_i^{k-1,k}, v_{avg_i}^{k-1,k}) \tag{14}$$

In Equation (13), n is the number of trajectory points of ship i .

A sample with the number of m is selected from the mother sample. A dimension of the sample is randomly selected, and a segmentation value is also selected. The first isolated tree is constructed according to the binary tree method. Samples less than the segmentation value are divided into the left cross tree, and samples greater than the segmentation value are divided into the right cross tree. Then, the first isolated tree would be constructed until the number of segmentations reaches h . The average path length of isolated trees is calculated as in Equation (15):

$$c(m) = 2 * (\ln(m - 1) + 0.5772156649) - \frac{2(m - 1)}{m} \tag{15}$$

In Equation (15), m represents the number of sub-sampling points.

When the path length of sample tr in j isolated tree is set as h_j^{tr} , the expected path length of sample tr in all isolated trees is calculated as in Equation (16):

$$E(h_j^{tr}) = \frac{\sum_{j=1}^p h_j^{tr}}{p} \tag{16}$$

In Equation (16), p represents the number of isolated trees, and h represents the restricted height of isolated trees.

The abnormal score $s_{(tr)}$ of sample tr is the basis for judging whether the sample is an outlier. The calculation method is as follows:

$$s_{(tr)} = 2^{\frac{-E(h_{tr}^m)}{c(m)}} \tag{17}$$

The threshold of the abnormal score of sample tr is set to $s_{tr(th)}$, and the discriminant method of the outliers is as follows:

$$label_{(tr)} = \begin{cases} 1, s_{(tr)} > s_{(tr)th} \\ 0, s_{(tr)} \leq s_{(tr)th} \end{cases} \tag{18}$$

In Equation (18), $label_{(tr)}$ is the category labeling of ship trajectory segment tr , and 1 means that the ship trajectory segment belongs to an outlier, that is, that the trajectory segment is composed of two types of ship trajectory points (that is the regular ship trajectory mixed with that of the spoofing ship). 0 indicates that the ship trajectory segment is normal, that is, the trajectory segment is composed of only the regular ship trajectory point. Combined with trajectory segmentation, the outliers of ship trajectory points are identified as follows:

$$flag(s_i^{k+1}) = \begin{cases} S_P, label_{(tr)} = 1 \wedge flag(s_i^k) = N_P \\ N_P, label_{(tr)} = 0 \wedge flag(s_i^k) = N_P \\ N_P, label_{(tr)} = 1 \wedge flag(s_i^k) = S_P \\ S_P, label_{(tr)} = 0 \wedge flag(s_i^k) = S_P \end{cases} \tag{19}$$

In Equation (19), if the trajectory segment is labeled as 0, indicating that distance and speed for the trajectory segment tend to be normal, then the trajectory points at the adjacent time have the same category and belong to one ship. Conversely, it shows that the distance and speed for the trajectory segment tend to be abnormal, meaning the trajectory points at adjacent times belong to two different ships.

4. Experiments

According to the trajectory characteristics of regular ships mixed with spoofing ships, these characteristics can be divided into the following four categories: ① the regular ship trajectory is continuous, while the spoofing ship trajectory points are concentrated; both ship trajectories are not overlapped (class I spoofing ship); ② both trajectories are continuous and not overlapped (class II spoofing ship); ③ the regular ship trajectory is continuous, while the spoofing ship trajectory is concentrated, and both ship trajectories are overlapped (class III spoofing ship); and, finally, ④ both trajectories are continuous and overlapped (class IV spoofing ship), as shown in Figure 3. In Figure 3, there appears to be a phenomenon of trajectory point jumping due to the trajectory points of a spoofing ship (labeled with orange dots) mixed in a regular ship trajectory (labeled with blue dots). In Figure 3a,c, there are only some scattered points of the spoofing ship, and they are concentrated in certain areas and taken as some isolated outliers. In Figure 3b,d, there are continuous trajectory points of the spoofing ship, which is obviously the trajectory of another ship that occupies the same MMSI as a regular ship, namely a spoofing ship.

When navigating at sea, the distance between adjacent trajectory points is almost linearly related to the update time interval of trajectory points, as shown in Figure 4a, while the speed between the adjacent trajectory points remains almost unchanged, as shown in Figure 4b. If there are two classes of ships with the same MMSI at sea, that means that the trajectory of a regular ship has been mixed with a spoofing ship. The variation in the trend of average sailing speed and distance between adjacent trajectory points is no longer consistent with Figure 4a,b, as shown in Figure 4c,d. The distance between adjacent trajectory points is large and almost does not change with time, while the speed between trajectory points decreases exponentially with time.

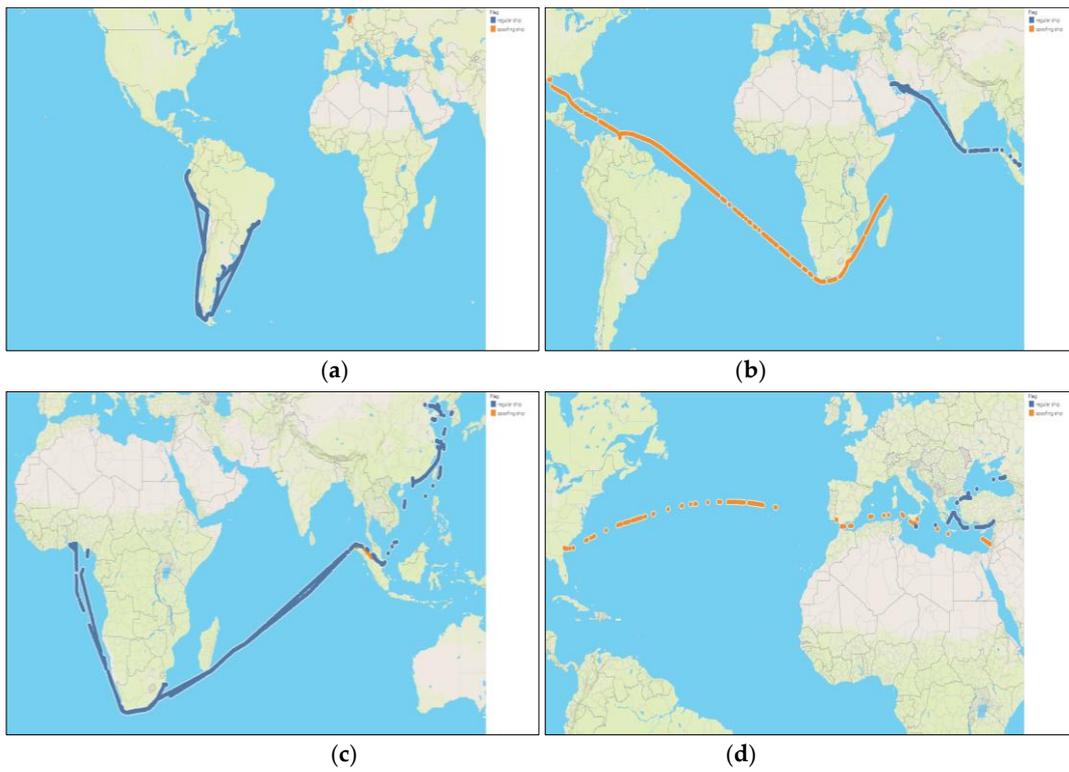


Figure 3. Distribution of regular ship trajectory points mixed with spoofing ship: (a) class I; (b) class II; (c) class III; (d) class IV.

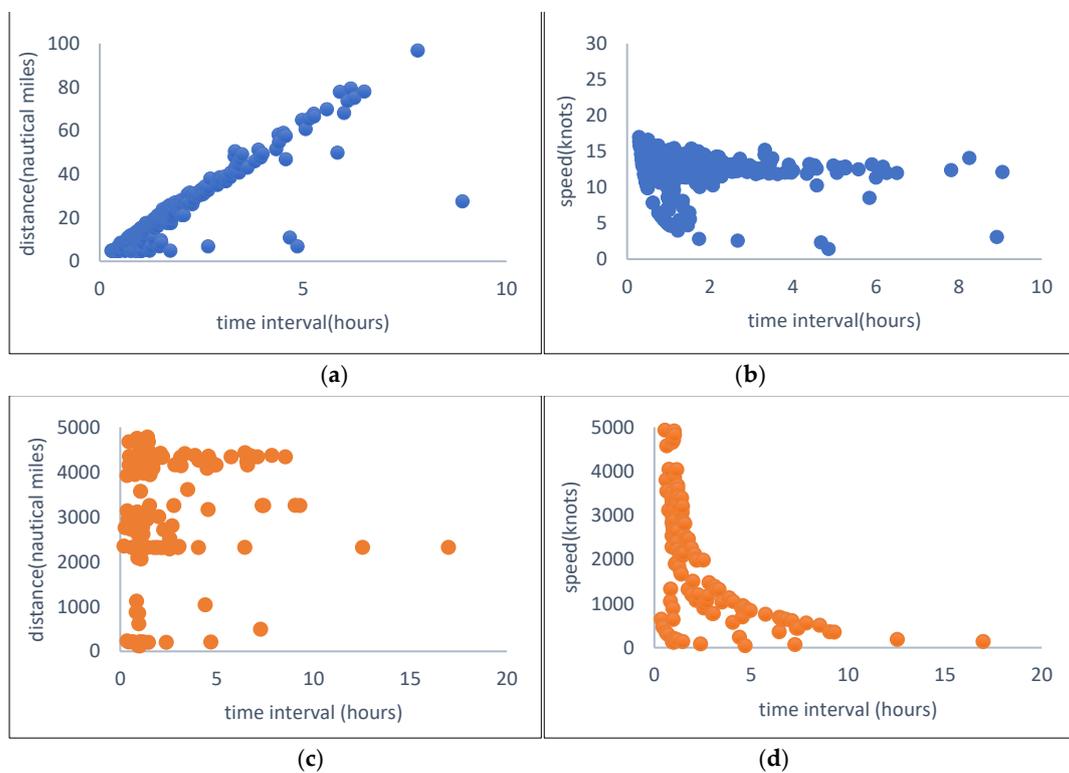


Figure 4. Distribution of distance and speed between adjacent trajectory points: (a) distance distribution of regular ship; (b) speed distribution of regular ship; (c) distance distribution of regular ship mixed with spoofing ships; (d) speed distribution of regular ship mixed with spoofing ships.

4.1. Distribution of Speed and Distance between Adjacent Trajectory Points

In order to set a reasonable threshold of average sailing speed between adjacent trajectory points, it is vital to understand the distribution of the average sailing speed. Through trajectory feature mining, it was found that the average sailing speed among regular ship trajectory points is normally distributed, and the expected value in Figure 5a is 12.5 knots. Average sailing speed between different trajectory points conforms to normal distribution, and expected value in Figure 5b is 2750 knots. Figure 5c shows the probability distribution diagram of average sailing speed among trajectory points, while Figure 5d shows the variation in the trend of the cumulative probability of average sailing speed between trajectory points. Among them, 82.59% of the average sailing speed between trajectory points is less than 16 knots, which can be used as the average sailing speed threshold for identifying spoofing ship trajectory points.

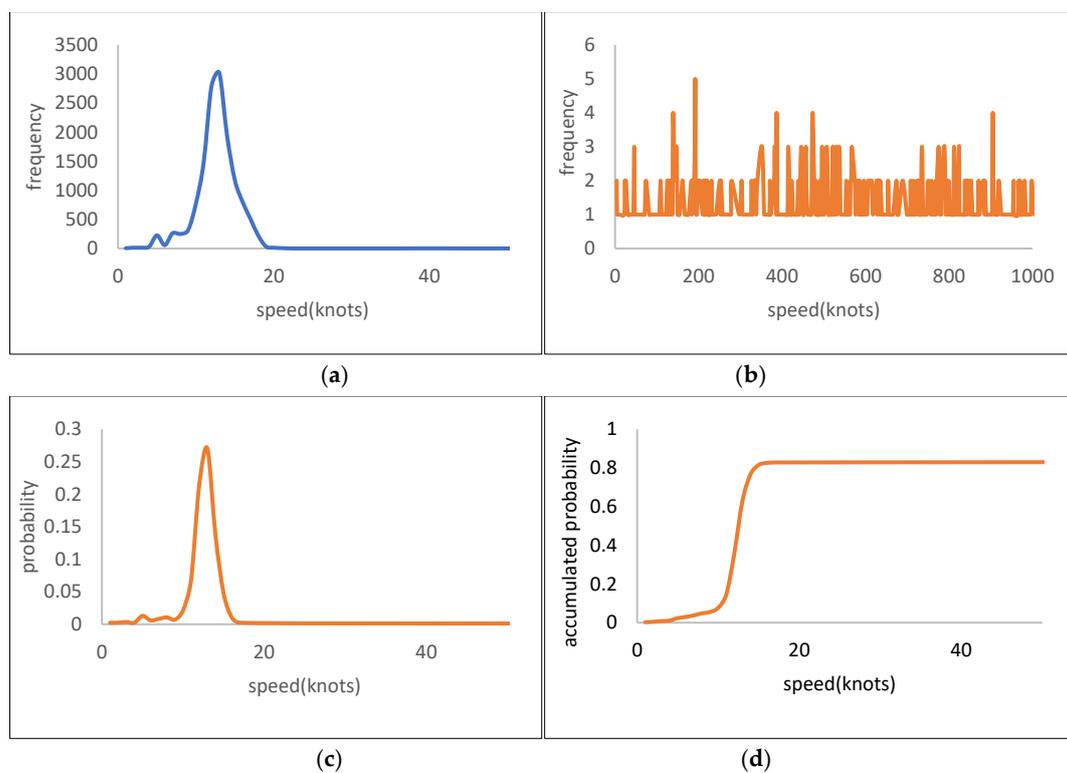


Figure 5. Speed distribution of trajectory segment: (a) speed of trajectory segment for regular ship; (b) speed of trajectory segment for regular ship mixed with spoofing ship; and (c,d) probability distribution of speed among trajectory segments for regular ship mixed with spoofing ship.

Through trajectory feature mining, it was found that the distance between regular ship trajectory points is normally distributed, with the expected value in Figure 6a being 20 nautical miles. The distance between trajectory points mixed with spoofing ships is normally distributed, and the expected value in Figure 6b is 800 nautical miles. Figure 6c shows the probability distribution of the distance between adjacent trajectory points, while Figure 6d shows the variation in the trend of the cumulative probability of the distance between adjacent trajectory points. Among them, 82.42% of the distance between adjacent trajectory points is less than 80 nautical miles, which can be used as the threshold of the distance between adjacent trajectory points for identifying spoofing ship trajectory points.

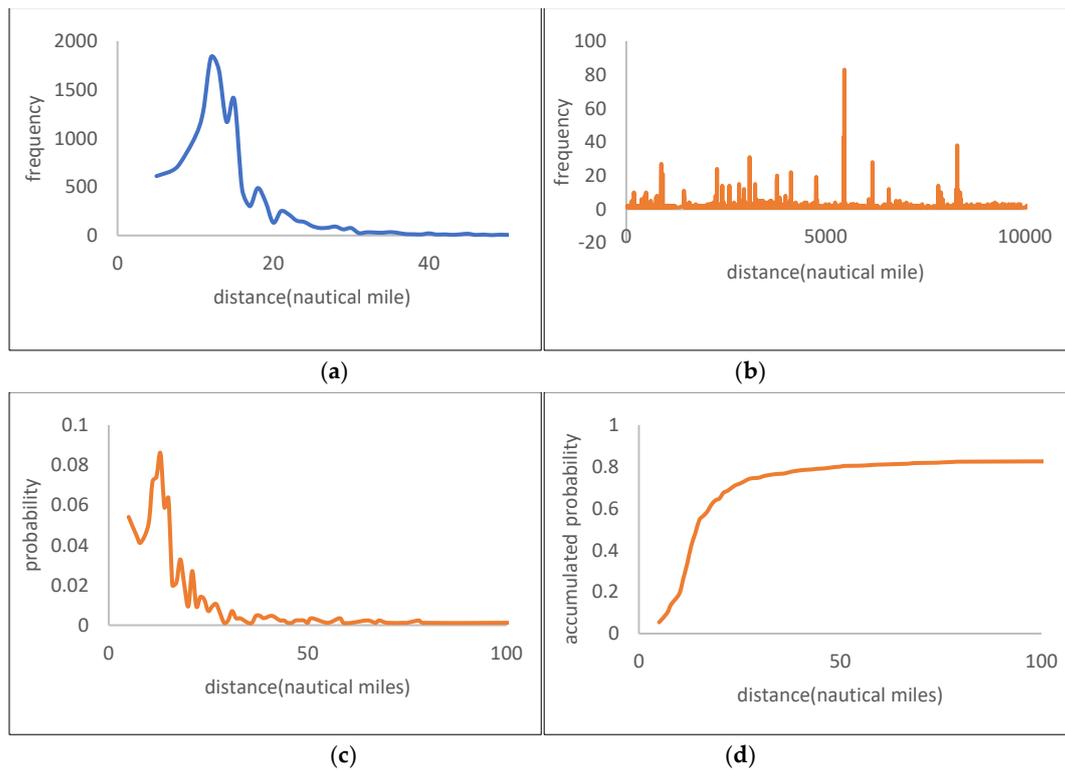


Figure 6. Distance distribution of trajectory segment: (a) trajectory segment distance among regular ship; (b) trajectory segment distance among regular ship mixed with spoofing ship; and (c,d) probability distribution of trajectory segment distance among regular ship mixed with spoofing ship.

With regard to time interval between adjacent trajectory points, it is found that time interval of regular trajectory points is normally distributed through statistical learning, and the expected value is 1.5 h. However, the time interval between some adjacent trajectory points is relatively large, but the ratio of these trajectory segments is relatively small. As the time interval between trajectory points increases, the proportion of the trajectory segment gradually decreases. The distribution pattern of the time intervals between adjacent trajectory points is shown in Table 1.

Table 1. Trajectory segment number distribution for various time intervals.

Time Interval (Hours)	Corresponding Number of Trajectory Segment	The Number of Overall Trajectory Segment	Trajectory Segment Ratio (Percent)
≤1	25,009	52,537	47.6
(1, 2)	24,025	52,537	45.73
(2, 3)	1686	52,537	3.2
(3, 4)	598	52,537	1.15
(4, 5)	337	52,537	0.64
(5, 10)	603	52,537	1.15
(10, 15)	142	52,537	0.27
(15, 20)	53	52,537	0.1
>20	84	52,537	0.16

Figure 7 shows the distribution pattern of the distance and average sailing speed between adjacent trajectory points. The blue dots represent the scatter plots of distance and average sailing speed between regular ship trajectory points, while the orange dots represent the ship trajectory mixed with spoofing ship trajectory points. In Figure 7a, when the time interval between adjacent trajectory points is within 5 h, the continuity of the ship’s

trajectory is good, and the blue and orange points have a good distinguishing ability. When the time interval between ship trajectory points exceeds 5 h, the blue and orange points will overlap, making it difficult to identify the trajectory points of the spoofing ship, as shown in Figure 7b. Moreover, the longer the time interval between adjacent ship trajectory points, the less easily the regular ship trajectory and the ship trajectory mixed with spoofing ships are identified, as shown in Figure 7c,d.

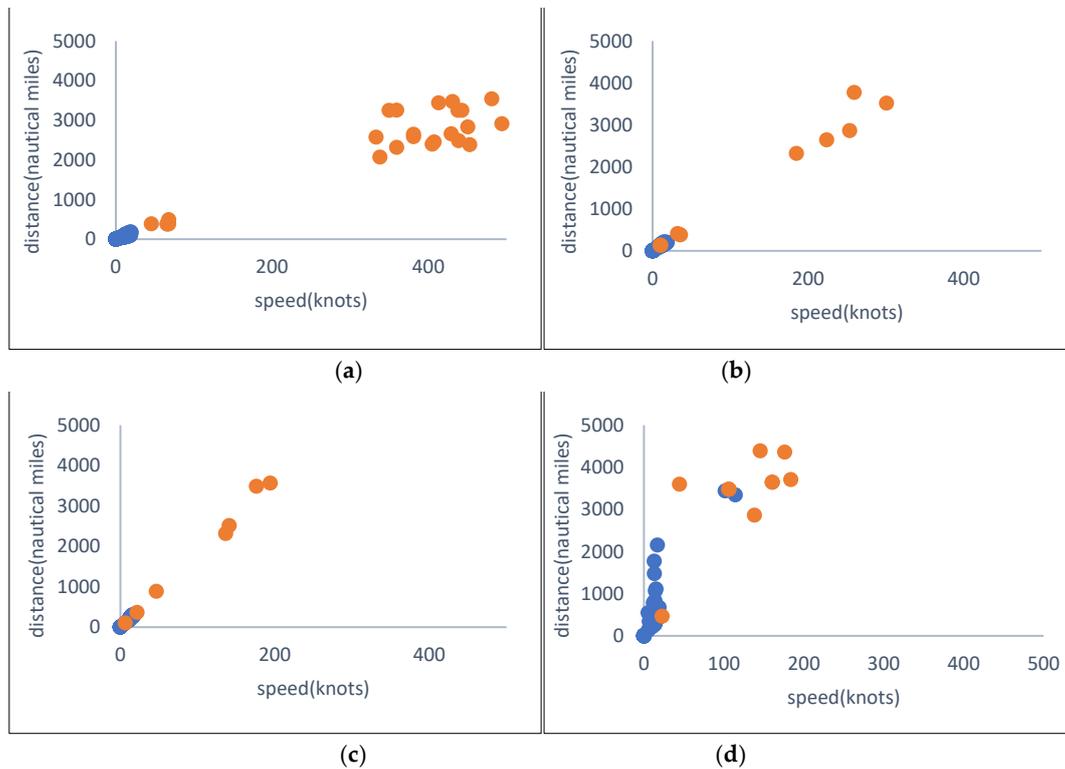


Figure 7. The distance and average sailing speed distribution of trajectory segment corresponding to various time interval between adjacent trajectory points: (a) time interval within 5 h; (b) time interval beyond 5 h and within 10 h; (c) time interval beyond 10 h and within 15 h; and (d) time interval beyond 15 h and within 20 h.

4.2. Identification of Spoofing Ships Based on Trajectory Segmentation and Isolation Forest

For the trajectory of classes I, II, III, and IV of the spoofing ship, the accuracy of identifying outliers of the ship trajectory shows the following trend with the number of sub-sampling points and height of the isolated tree. Figure 8 reflects a correlation between the accuracy of identifying outliers and the number of sub-sampling points. For the trajectory of classes I, II, III, and IV of spoofing ships, the accuracy of identifying outliers of trajectory points gradually decreases, and the error rate within the identification of regular ship trajectory points gradually decreases, with an increase in the number of sub-sampling points. The number of sub-sampling points is one of the important parameters of the isolation forest, which would affect the true positive rate (outliers correctly identified) and false positive rate (trajectory points of regular ship wrongly identified). Generally speaking, the higher the true positive rate is, and the lower the false positive rate is, the more reasonable the number of sub-sampling points is. In Figure 8, when the number of sub-sampling points is about 100, the true positive rate is higher than 0.95, and the false positive rate is lower than 0.05, so the number of sub-sampling points is set as 128.

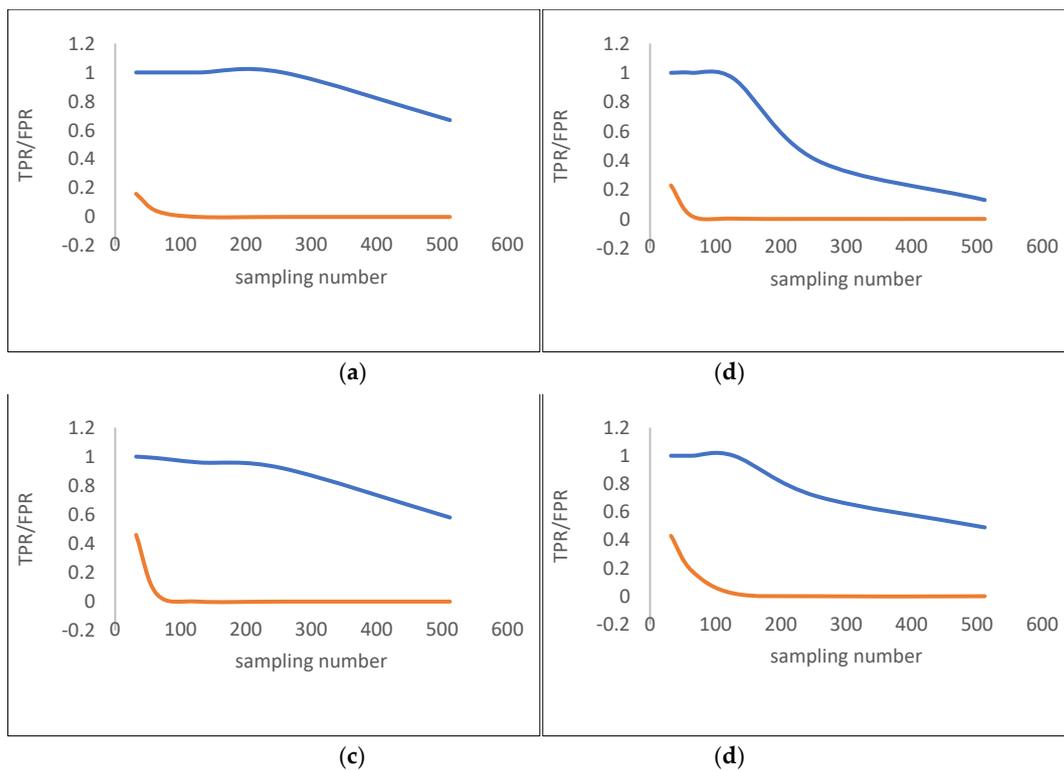


Figure 8. The correlation between True Positive Rate, False Positive Rate, and the number of sub-sampling points for spoofing ship trajectory identification: (a) class I; (b) class II; (c) class III; (d) class IV.

Figure 9 reflects a correlation between the accuracy of identifying outliers and the height of isolated trees. For the trajectory of classes I, II, III, and IV spoofing ships, the accuracy of identifying outliers of trajectory points gradually increases, and the identification error rate of regular ship trajectory points also gradually increases, with an increase in the height of isolated trees. The height of isolated trees is one of the important parameters of isolated forest. The higher the isolated tree height is, the more effectively true positive samples are identified (outliers correctly identified). However, as isolated tree height increases, some false positive samples may also be mistaken for positive samples (regular ship trajectory point wrongly identified). Generally speaking, the higher the true positive rate is, and the lower false positive rate is, the more reasonable the height of the isolated trees is. In Figure 8, when the height of isolated trees is about eight, the true positive rate is higher than 0.95, and the false positive rate is lower than 0.05, so the height of isolated tree is set as eight.

In Figures 10–13, the N_P and S_P of the ship trajectory are labeled with blue and orange dots, and the NL_P of the ship trajectory are labeled with red dots. For outliers of trajectory points that cannot be identified by statistical learning, the number of unidentified trajectory points shows the following trend after adopting the isolated forest algorithm for recognition. The number of unidentified trajectory points for class I spoofing ships gradually decreases with an increase in the trajectory segmentation time; the specific values are listed as 32, 19, 12, and 9, respectively, as shown in Figure 10. The number of unidentified trajectory points for class II spoofing ships gradually decreases with an increase in the trajectory segmentation time, with specific values of 5, 3, 1, and 1, as shown in Figure 11. The number of unrecognized trajectory points for class III spoofing ships gradually decreased with an increase in the trajectory segmentation time; the specific values are listed as 20, 11, 8, and 4, as shown in Figure 12. The number of unrecognized trajectory points has always been 0 for class IV spoofing ships, as shown in Figure 13.

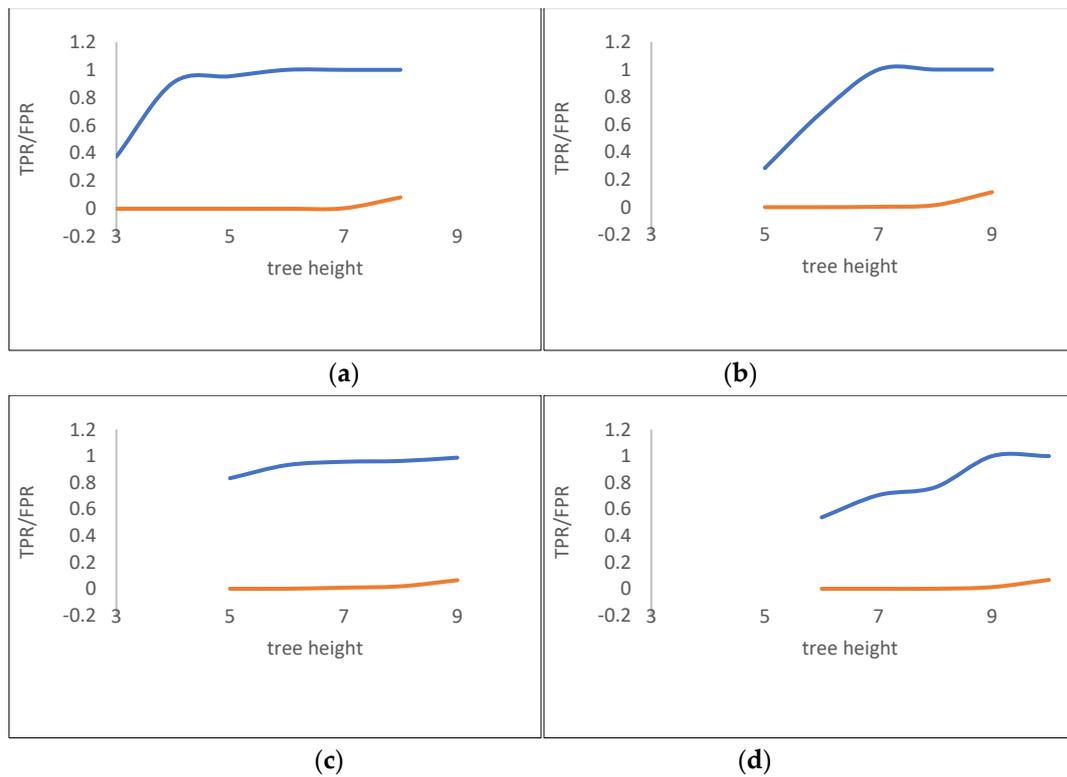


Figure 9. The correlation between true positive rate, false positive rate, and tree height for spoofing ship trajectory identification: (a) class I; (b) class II; (c) class III; and (d) class IV.

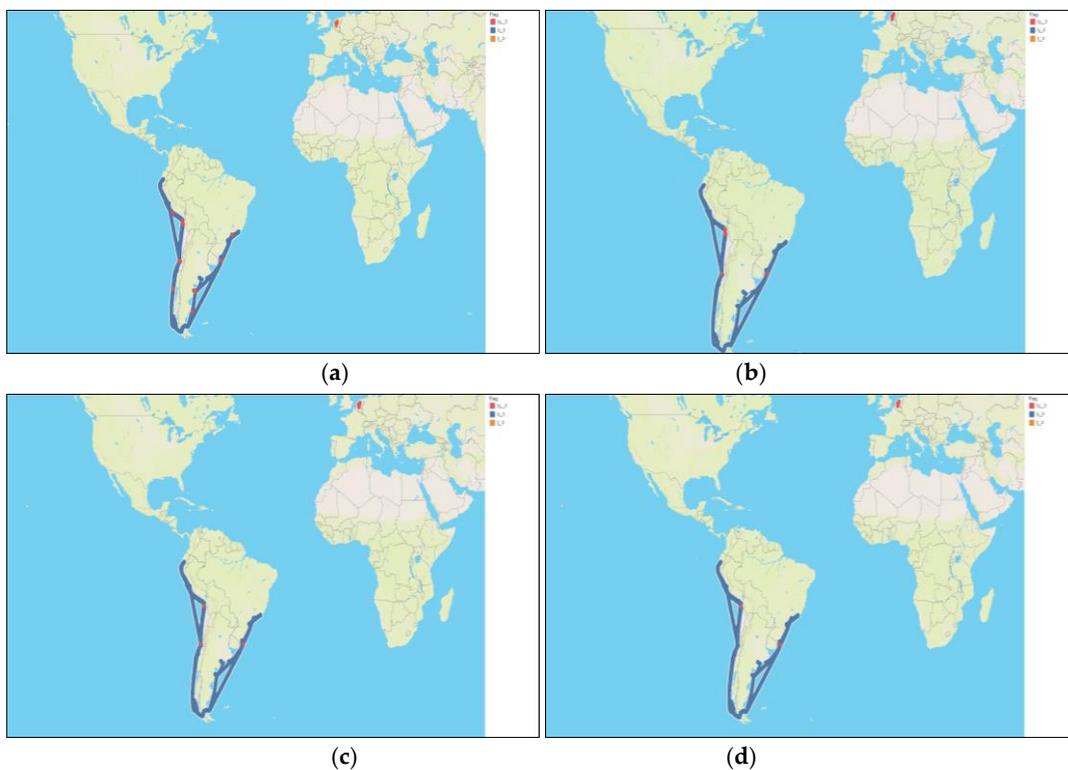


Figure 10. Identification of outliers for class I spoofing ships via isolation forest: (a) trajectory segmented by 5 h, (b) trajectory segmented by 10 h, (c) trajectory segmented by 15 h, and (d) trajectory segmented by 20 h.

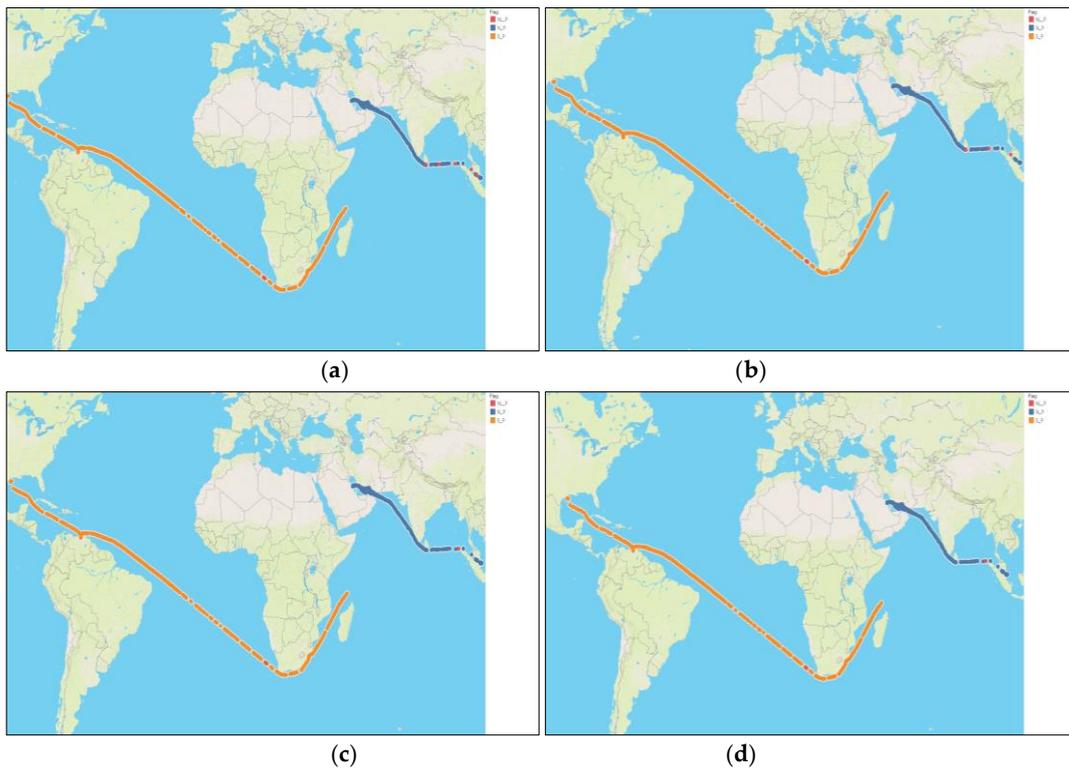


Figure 11. Identification of outliers for class II spoofing ships via isolation forest: (a) trajectory segmented by 5 h, (b) trajectory segmented by 10 h, (c) trajectory segmented by 15 h, and (d) trajectory segmented by 20 h.

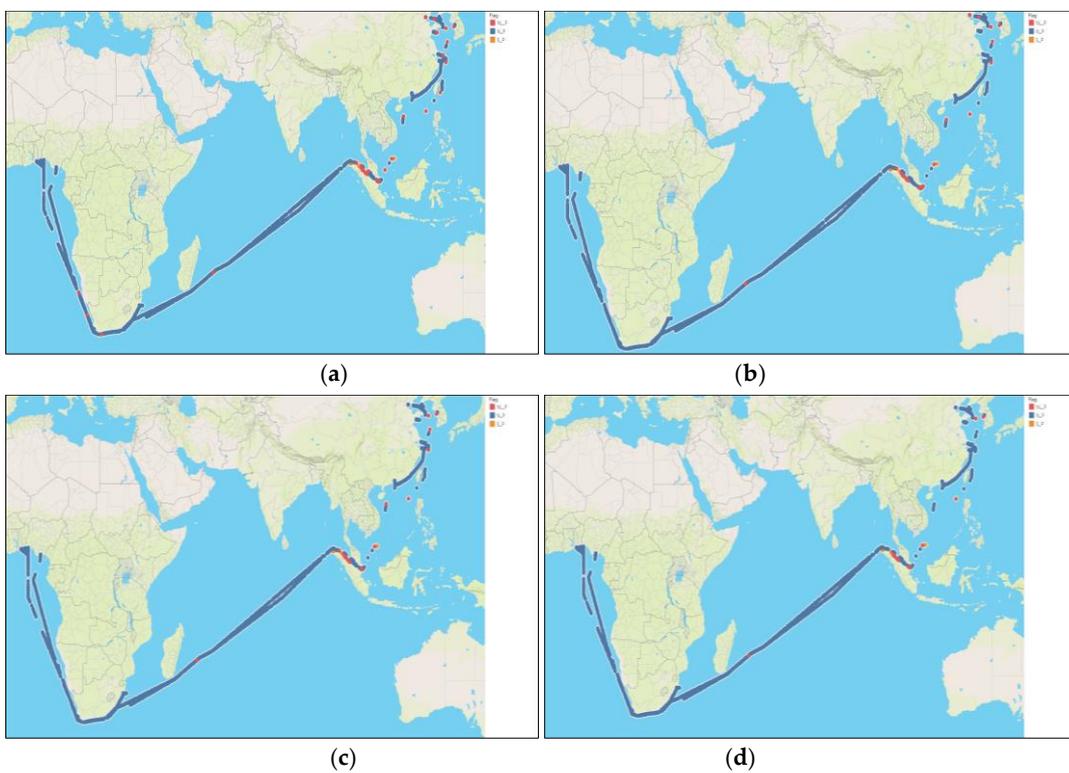


Figure 12. Identification of outliers for class III spoofing ships via isolation forest: (a) trajectory segmented by 5 h, (b) trajectory segmented by 10 h, (c) trajectory segmented by 15 h, and (d) trajectory segmented by 20 h.

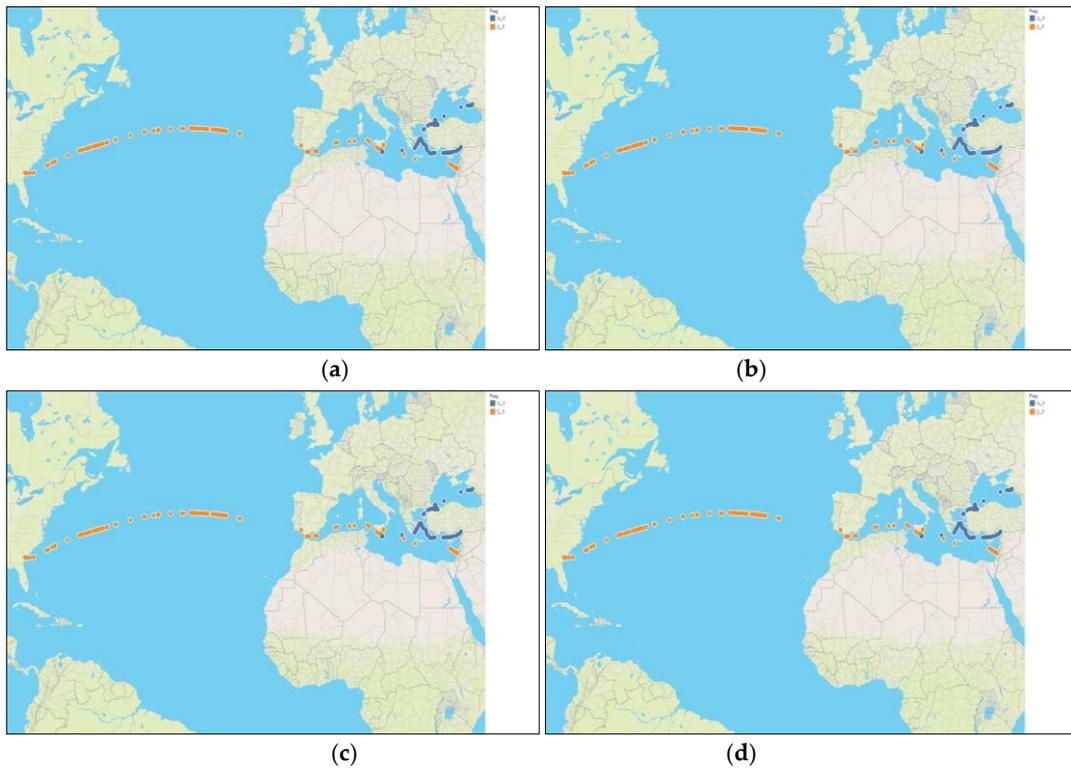


Figure 13. Identification of outliers for class IV spoofing ships via isolation forest: (a) trajectory segmented by 5 h, (b) trajectory segmented by 10 h, (c) trajectory segmented by 15 h, and (d) trajectory segmented by 20 h.

As can be seen in Figure 14, the isolation forest was adopted to continue identifying the outliers of trajectory points that could not be identified through statistical learning. The time interval for trajectory segmentation increased from 5 h to 20 h, and the accuracy of identifying outliers for class I spoofing ships improved first and then decreased, with specific values of 95.7%, 98.3%, 94.5%, and 90.4%, gradually improved specific values of 76.4%, 86.9%, 94.4%, and 94.4% for class II spoofing ships, and gradually improved specific values of 88.2%, 91.9%, 93.4%, and 98.1% for class III spoofing ships. However, the accuracy of identifying outliers for class IV spoofing ships remains high and unchanged, mainly due to the short time interval among trajectory points, with constant values of 100%.

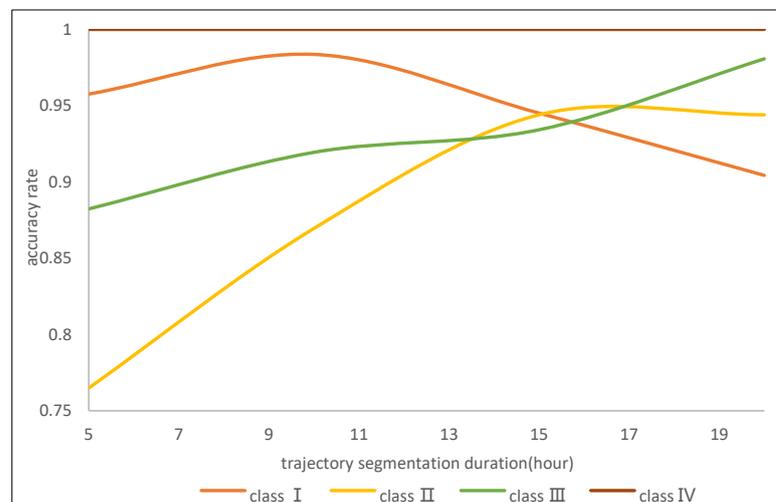


Figure 14. Comparison of the accuracy of identifying outliers for four classes of spoofing ships.

In Figure 14, among the four classes of regular ship trajectories mixed with spoofing ships, the identification accuracy of class IV and class I is relatively high, reaching 100% and 95.7%, respectively, while the identification accuracy of class II and class III is relatively low, at only 76.4% and 88.2%. The segmentation threshold for these four types of ship trajectories is all 5 h, and the reason for the inconsistent identification accuracy is that the ratio of missing trajectories varies. Taking the sampling time of one hour as an example, the complete trajectory points of a ship in one day should be 24. Therefore, the ratio of missing points and jumping points for four classes of ships can be calculated based on the statistical time. The correlation between the ratio of jumping and missing points and identification accuracy for four classes of spoofing ships are shown in Table 2.

Table 2. Correlation between the ratio of jumping and missing points and identification accuracy for four classes of spoofing ships.

Class of Spoofing Ship	Statistical Time (Months)	The Number of Jumping Trajectory Points	The Number of Actual Trajectory Points	Jumping Points Ratio (Percent)	The Number of Complete Trajectory Points	Missing Points Ratio (Percent)	Identification Accuracy (Trajectory Segmented by 5 h)
I	4	18	2197	0.82	2880	23.7	95.7
II	2	528	1085	48.66	1440	24.6	76.4
III	3	58	1629	3.56	2160	24.5	88.2
IV	1	218	690	31.59	720	4.1	100

In Table 2, the ratio of missing trajectories points for class IV spoofing ships is only 4.1%, while the ratio for the other three classes of spoofing ships is close to 25%. Therefore, the accuracy of identifying the trajectories of class IV spoofing ships is much higher than that of the other three classes of spoofing ships. For the other three classes of spoofing ships, class I spoofing ships have the smallest ratio of jumping points, class III spoofing ships have a slightly larger ratio of jumping points, and class II spoofing ships have the largest ratio of jumping points. The accuracy of identifying these three classes of spoofing ships is also consistent with the changes in the ratio of jumping points.

In Figure 14, the identification accuracy of the first three classes of spoofing ships did not reach 100%, and the accuracy did not improve with the increase in the time interval threshold, such as class I and II spoofing ships. The applicability of the three parameters and their related thresholds in this paper varies for each class of spoofing ship. For class I and IV spoofing ships, the small time interval threshold brought the best identification effect due to the low ratio of missing trajectory points. For class II and III spoofing ships, the identification effect is the best when the time interval threshold between trajectory points is large due to the high ratio of missing trajectory points. In addition, in order to avoid mistakenly identifying the trajectory point of a spoofing ship as a regular ship, the distance threshold between trajectory points and the average sailing speed threshold are set to be small, which results in some regular ship trajectory points not being recognized and labeled as confusion points. In Table 3, as the distance threshold and average sailing speed threshold change, there are variations in the trend of the identification accuracy of the four classes of spoofing ships.

Table 3. Correlation between various values of threshold and identification accuracy (True Positive Rate, False Positive Rate) for four classes of spoofing ships’ trajectory segmented by 5 h.

Minimum Speed Threshold (Knots)	Maximum Speed Threshold (Knots)	Minimum Distance Threshold (Nautical Miles)	TPR and FPR for Class I (Percent)	TPR and FPR for Class II (Percent)	TPR and FPR for Class III (Percent)	TPR and FPR for Class IV (Percent)
20	100	100	91.98, 8.02	76.49, 23.51	88.03, 12.54	100, 0
30	100	150	92, 8.57	76.47, 34.45	88.59, 11.93	100, 0
50	100	250	92.28, 8.16	76.49, 49	89.01, 11.46	39.28, 60.72
80	100	400	91.13, 9.2	76.48, 58.28	89.35, 11.09	81.08, 72.97

For the spoofing ship of class I and III, the ship trajectory can be accurately displayed without scatter jumping points, as shown in Figure 15a,c. For the spoofing ship of class II and IV, the trajectories of two ships can be accurately displayed, as shown in Figure 15b,d. The blue lines indicate the trajectories of regular ships, and the orange lines indicate the trajectories of spoofing ships. Through the identification of the spoofing ship, the outliers of the trajectory of a container ship in Figure 1 have been almost removed and classified via trajectory segmentation and isolated forest, and the trajectory of a regular ship and spoofing ship are exhibited, respectively, in Figure 16a,b, which can reflect a ship’s motion pattern accurately.

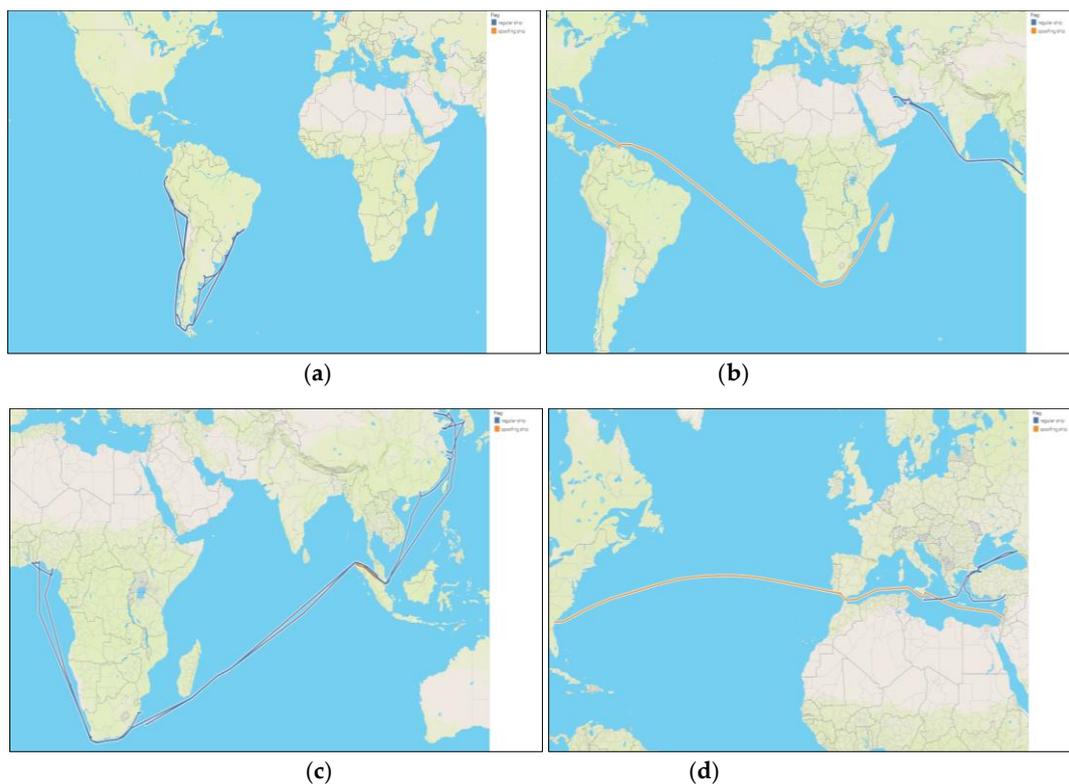


Figure 15. Trajectory of four classes of regular ship trajectory without outlier: (a) trajectory of class I exhibited with regular ship (marked by blue line) and spoofing ship (marked by orange line) separately; (b) trajectory of class II; (c) trajectory of class III; and (d) trajectory of class IV.

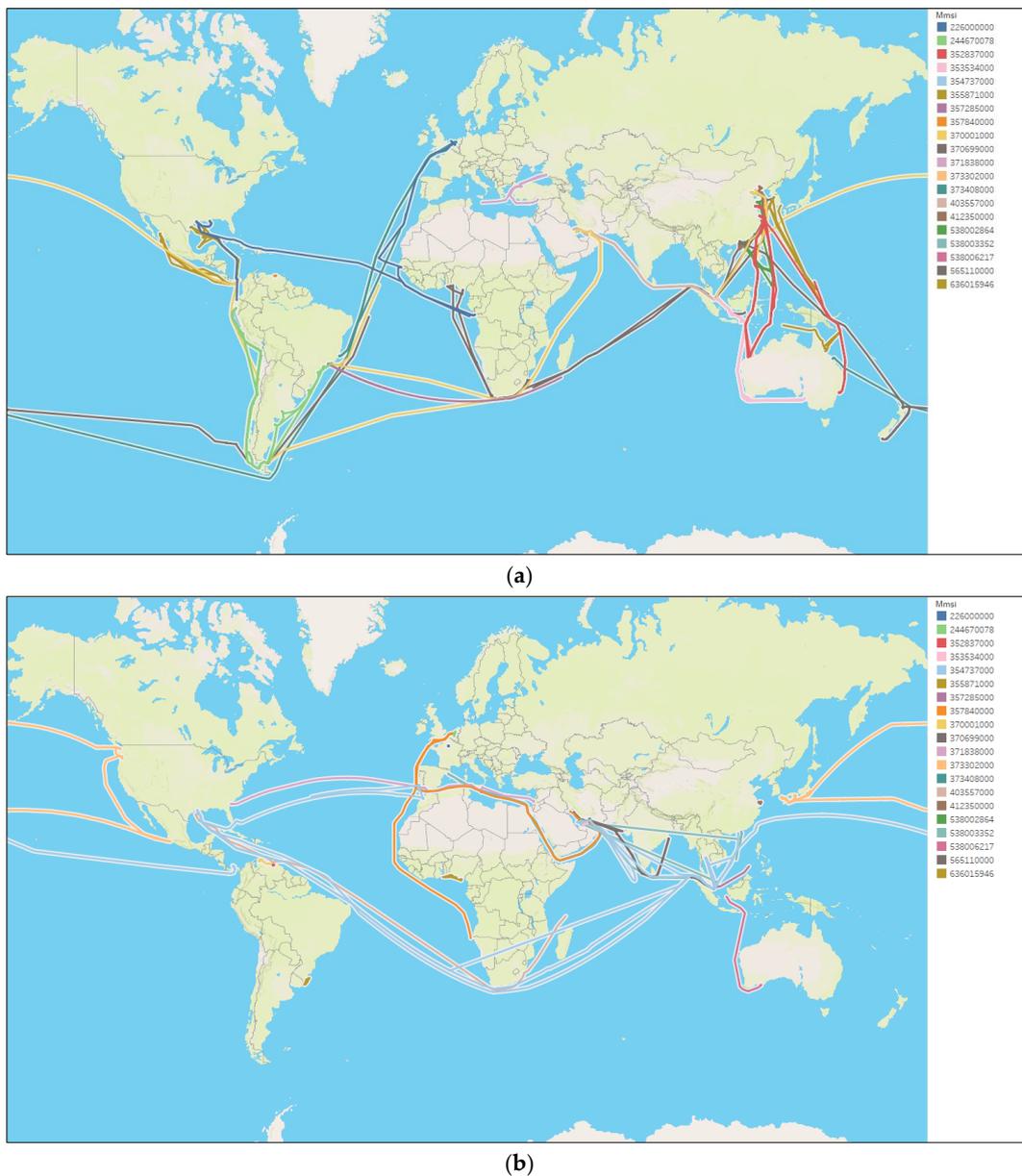


Figure 16. Trajectory of container ships exhibited in Figure 1 without confusion points between regular ship and spoofing ship: (a) regular ship trajectory; (b) spoofing ship trajectory.

The framework was implemented on Windows 10 OS with 8 GB RAM and 2.8 GHz CPU. We employed Matlab (2016 version) to perform trajectory segmentation and the spoofing ship identification procedure on the ship trajectory data. With regard to the runtime test, the paper dealt with 52,538 trajectory points owned by 20 container ships via isolation forest, and runtime was 15.77 s.

5. Discussion

Each ship sailing at sea has a unique MMSI, which can be used to extract the complete trajectory of any ship from AIS data. However, some ships have obtained unauthorized MMSI through illegal approaches, which are duplicated with MMSI owned by existing ships. This leads to a trajectory being extracted from one MMSI that actually belong to multiple ships, which poses serious challenges to ship motion pattern identification based on AIS data.

The paper adopts trajectory feature mining to clarify the distribution patterns of distance and average sailing speed between adjacent trajectory points (see Figures 5 and 6 for details). To address the impact of the missing trajectories of ships, updating the time interval threshold is set to segment ship trajectories. For a segmented ship trajectory, outliers can be identified via isolation trees established based on distance and speed among trajectory segments. By observing the trend of changes in the true positive and false positive rates of trajectory point identification, two important parameters of isolation forest are determined, namely the sampling number and isolated tree height. After adopting isolation forest, the identification accuracy of spoofing ships was improved, as shown in Figure 14. The number of unidentified trajectory points for class I, II, and III spoofing ships has gradually decreased with an increase in the trajectory segmentation time, as shown in Figures 10–12. However, the identification accuracy of class IV spoofing ships was not improved, and the reason may be that class IV spoofing ships have good continuity, as shown in Figure 13.

Due to the identification of spoofing ships by isolation forest, the number of trajectories in Figure 16 is almost half more than that in Figure 1. The trajectory of these spoofing ships would just be removed if using the approaches described in the literature, and more information of the AIS data would be lost. The pre-processing of AIS data should aim to retain more information, which is more helpful to the situational awareness of ship motion for the MSA based on AIS data. For the trajectory of a spoofing ship away from a regular ship, it can be identified according to the serial number of the AIS base station included in the AIS data. A part of the ship trajectory in this paper meets such characteristics and would be identified more efficiently by the serial number of the AIS base station. However, many AIS data do not contain information such as the serial number of the AIS base station, so the spoofing ship identification method supplied by this paper is still necessary.

However, some trajectory points are still falsely identified as a regular ship rather than a spoofing ship. The main reason for this is that trajectory points correlation shows a poor performance among some trajectory segments. Future research should focus on the clustering of trajectory points among various trajectory segments, so as to identify the NN_P listed in Equation (12). In addition, this study only set a constant threshold and established isolated trees via trajectory feature mining from the history trajectory of typical cargo ships so as to identify outliers of ship trajectories. Future research should focus on setting an adaptive threshold for speed and distance between adjacent trajectory points based on differences in the maneuvering performance for various types of ships and the speed difference of ships in different navigation stages.

6. Conclusions

A long time interval between adjacent trajectory points results in severe missing trajectories of a ship, and the identification of spoofing ships are not ideal. In order to eliminate the impact of missing trajectory points on the accuracy of identifying spoofing ships, the trajectory is segmented by the time interval threshold. After trajectory segmentation, the trajectory points of each trajectory segment maintain good continuity, that is, the time interval between adjacent trajectory points for each trajectory segment is relatively short. Combined with trajectory segmentation, the isolation forest is efficient at distinguishing between regular ship trajectory points and spoofing ship trajectory points. Consequently, outliers of ship trajectories were almost removed or classified correctly in this work, and the labeled ship trajectory points can reflect a ship's motion pattern accurately.

Author Contributions: Conceptualization, H.Z.; methodology, C.Y.; software, Q.M.; validation, H.Z. and Q.H.; formal analysis, H.Z.; investigation, H.Z.; resources, C.Y. and P.W.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, Q.H.; visualization, C.Y.; supervision, Q.H.; project administration, Q.H.; funding acquisition, Q.H. and K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Project of Ministry of Transport, grant number 2020MS6162. Project of Zhoushan science and Technology Bureau, grant number 2021C21010; National innovation and entrepreneurship training program for Zhejiang Ocean University, grant number 202210340043.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Androjna, A.; Perkovič, M.; Pavić, I.; Mišković, J. AIS data vulnerability indicated by a spoofing case-study. *Appl. Sci.* **2021**, *11*, 5015. [[CrossRef](#)]
2. Felski, A.; Jaskólski, K.; Banyś, P. Comprehensive assessment of automatic identification system (AIS) data application to anti-collision manoeuvring. *J. Navig.* **2015**, *68*, 697–717. [[CrossRef](#)]
3. Wawruch, R. Ability to test shipboard automatic identification system instability and inaccuracy on simulation devices. *Zesz. Nauk. Akad. Morskiej W Szczecinie* **2017**, *52*, 128–134.
4. Jaskólski, K. Two-dimensional coordinate estimation for missing automatic identification system (AIS) signals based on the discrete Kalman filter algorithm and universal transverse mercator (UTM) projection. *Zesz. Nauk. Akad. Morskiej W Szczecinie* **2017**, *52*, 82–89.
5. Peters, D.J.; Hammond, T.R. Interpolation between AIS reports: Probabilistic inferences over vessel path space. *J. Navig.* **2011**, *64*, 595–607. [[CrossRef](#)]
6. Shelmerdine, R.L. Teasing out the detail: How our understanding of marine AIS data can better inform industries, developments, and planning. *Mar. Policy* **2015**, *54*, 17–25. [[CrossRef](#)]
7. Sang, L.Z.; Wall, A.; Mao, Z.; Yan, X.P.; Wang, J. A novel method for restoring the trajectory of the inland waterway ship by using AIS data. *Ocean. Eng.* **2015**, *110*, 183–194. [[CrossRef](#)]
8. Pallotta, G.; Vespe, M.; Bryan, K. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy* **2013**, *15*, 2218–2245. [[CrossRef](#)]
9. Mazzarella, F.; Vespe, M.; Damalas, D. Discovering vessel activities at sea using AIS data: Mapping of fishing footprints. In Proceedings of the 17th International Conference on Information Fusion (FUSION), Salamanca, Spain, 7–10 July 2014.
10. Wu, L.; Xu, Y.; Wang, Q. Mapping global shipping density from AIS data. *J. Navig.* **2017**, *70*, 67–81. [[CrossRef](#)]
11. Greidanus, H.; Alvarez, M.; Eriksen, T. Completeness and accuracy of a wide-area maritime situational picture based on automatic ship reporting systems. *J. Navig.* **2016**, *69*, 156–168. [[CrossRef](#)]
12. Wei, G.; Yang, C. Detection of AIS Data Error. *Navig. China* **2016**, *39*, 11–14.
13. Han, Z.; Xu, G.; Huang, T. vessel trajectory outlier detection algorithm based on adaptive threshold. *Comput. Mod.* **2018**, *9*, 42–47.
14. Zhang, L.; Meng, Q.; Xiao, Z. A novel ship trajectory reconstruction approach using AIS data. *Ocean. Eng.* **2018**, *159*, 165–174. [[CrossRef](#)]
15. Liu, X.; Chu, X.; Ma, F. Discriminating method of abnormal dynamic information in AIS messages. *J. Traffic Transp. Eng.* **2016**, *16*, 142–150.
16. Liu, X. Study on the Approach of Ensuring AIS Data Availability in Inland Waterway. Ph.D. Thesis, Wuhan University of Technology, Wuhan, China, 2017.
17. Chen, X.; Ling, J.; Yang, Y.; Zheng, H.; Xiong, P.; Postolache, O.; Xiong, Y. Ship Trajectory Reconstruction from AIS Sensory Data via Data Quality Control and Prediction. *Math. Probl. Eng.* **2020**, *2020*, 7191296. [[CrossRef](#)]
18. Chen, X.; Wu, S.; Shi, C.; Huang, Y.; Yang, Y.; Ke, R.; Zhao, J. Sensing Data Supported Traffic Flow Prediction via Denoising Schemes and ANN: A Comparison. *IEEE Sens. J.* **2020**, *20*, 14317–14328. [[CrossRef](#)]
19. Chen, X.; Wang, Z.; Hua, Q.; Shang, W.L.; Luo, Q.; Yu, K. AI-Empowered Speed Extraction via Port-Like Videos for Vehicular Trajectory Analysis. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4541–4552. [[CrossRef](#)]
20. Guo, S.; Mou, J.; Chen, L.; Chen, P. Improved kinematic interpolation for AIS trajectory reconstruction. *Ocean. Eng.* **2021**, *234*, 109256.1–109256.15. [[CrossRef](#)]
21. Zhao, L. Ship Trajectory Outlier Detection Based on AIS Data and Recurrent Neural Network. Ph.D. Thesis, Dalian Maritime University, Dalian, China, 2019.
22. Zhao, L.; Shi, G. Maritime Anomaly Detection using Density-based Clustering and Recurrent Neural Network. *J. Navig.* **2019**, *72*, 894–916. [[CrossRef](#)]
23. Zhao, L.; Shi, G.; Yang, J. Ship trajectories pre-processing based on AIS data. *J. Navig.* **2018**, *71*, 1210–1230. [[CrossRef](#)]
24. Wang, Y. Detection and Early Warning Method of Ship Abnormal Behavior Based on Massive AIS Data. Ph.D. Thesis, Dalian Maritime University, Dalian, China, 2020.
25. Zhang, T.; Zhao, S.A.; Cheng, B. Detection of AIS closing behavior and MMSI spoofing behavior of ships based on spatiotemporal data. *Remote Sens.* **2020**, *12*, 702. [[CrossRef](#)]

26. Iphar, C.; Ray, C.; Napoli, A. Data integrity assessment for maritime anomaly detection. *Expert Syst. Appl.* **2020**, *147*, 113219. [[CrossRef](#)]
27. Feng, C.; Fu, B.; Luo, Y.; Li, H. The design and development of a ship trajectory data management and analysis system based on AIS. *Sensors* **2021**, *22*, 310. [[CrossRef](#)]
28. Varlamis, I.; Kontopoulos, I.; Tserpes, K.; Etemad, M.; Soares, A.; Matwin, S. Building navigation networks from multi-vessel trajectory data. *GeoInformatica* **2021**, *25*, 69–97. [[CrossRef](#)]
29. Jeong, M.H.; Jeon, S.B.; Lee, T.Y.; Youm, M.K.; Lee, D.H. Vessel trajectory reconstruction based on functional data analysis using automatic identification system data. *Appl. Sci.* **2020**, *10*, 881. [[CrossRef](#)]
30. Huang, H.; Qiu, K.; Jeong, M.H.; Jeon, S.B.; Lee, W.P. Detecting anomalous vessel dynamics with functional data analysis. *J. Coast. Res.* **2019**, *91*, 406–410. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.