



# Navigational risk analysis based on GIS spatiotemporal trajectory mining: a case study in Nanjing Yangtze River Bridge waters

Huiyu Xia<sup>1</sup>

Received: 21 October 2020 / Accepted: 23 January 2021 / Published online: 3 February 2021  
© Saudi Society for Geosciences 2021

## Abstract

With the continuous growth of the quantity, scale, and speed of vessels in recent years, maritime accidents are posing increasing risks to societies and individuals, especially in narrow inland waterways. Therefore, it is of great significance to analyze navigational risks to ensure the safety of waterborne transportation. In this paper, the navigational risks of Nanjing Yangtze River Bridge (NYRB) waters are investigated based on spatiotemporal mining on massive automatic identification system (AIS) trajectories by using geographic information system (GIS) techniques. A time-series-oriented trajectory processing method is proposed to deal with the historical AIS data in the whole year of 2019. The method adopts a periodic processing strategy to produce traffic density estimation products in multiple temporal scales for supporting spatiotemporal analysis. The proposed method greatly improves the data-processing efficiency and provides a flexible way to deeply understand the vessel behavior patterns in NYRB waters. Then the complete characteristics of the spatial distribution and temporal variation of AIS trajectories are revealed. Based on that, three types of critical navigational risks are discovered, which include the safety distance risk, the pier collision risk, and the traffic congestion risk. Moreover, we find that the greatest risk is existed in small vessels in the flood season, which is worth the most concern.

**Keywords** Navigational risk · GIS · Spatiotemporal mining · Trajectory

## Introduction

During the last few decades, we have witnessed a huge growth of traffic volume as well as a rapid increase of capacity of vessels, which in turn pose incrementing risks to water transportation safety (Svanberg et al. 2019). Therefore, great effort is required to prevent accidents and improve navigational safety and traffic efficiency. A deep insight into the navigational risks is of great importance for authorities to understand the influence of navigation environment on vessels and uncover the potential threats to navigation (Chen et al. 2019b).

Navigational risks are hidden in a large amount of regular behaviors of vessels. Detection of potential risks from

accumulated vessel behaviors poses significant challenges to researchers and maritime authorities. In literatures, plenty of methods and models have been proposed to investigate the navigational risks from historical data (Li et al. 2012). For example, Pak et al. (2015) leveraged a fuzzy analytical hierarchy approach to analyze the navigational risks in port waters by using the navigation data collected from several captains. Faghih-Roohi et al. (2014) proposed a simulated accident model based on Markov Chain Monte Carlo simulation for assessing accident risk based on limited accident data. Merrick et al. (2003) created a simulation model of ship navigation based on the Bayesian theory to estimate the navigational risks and the degree of ship congestion. However, the conventional methods for analyzing navigational risks are of limited practical use because they generally depend on subjective knowledge from experts, or modeling based on small amount of data (Zhang et al. 2020).

Nowadays with the continuous improvement of maritime technology, the self-reporting system has been introduced to surveil the maritime safety and collect vessel traffic data. The most widely used self-reporting system is the automatic identification system (AIS) (Zhang et al. 2018). The AIS adopts a

---

This article is part of the Topical Collection on *Geological Modeling and Geospatial Data Analysis*

Responsible Editor: Zakaria Hamimi

✉ Huiyu Xia  
xiahuiyu@whu.edu.cn

<sup>1</sup> Yangtze River Waterway Bureau, Nanjing 210011, China

modern manner of crowdsourcing to collect real-time vessel position data. The quality of AIS data is more reliable than vehicle position data on road transportation since it uses a dedicated Very high frequency (VHF) channel to communicate. As its highly reliability and availability, in recent years, AIS data has been widely leveraged in analyzing navigational risks and maritime safety (Kulkarni et al. 2020).

Many efforts have been made to integrate the information provided by the AIS into different risk analysis models (Bye and Almklov 2019). For example, Chen et al. (2019c) proposed a causal probability model for ship collision accident with the individual encounter information obtained from AIS data. Wang et al. (2013) proposed a phased decision and maneuvering model to study a two-vessel collision accident scenario by leveraging the AIS as input. Silveira et al. (2013) proposed a method to estimate the risk of collision through evaluating the number of collision candidates calculated from the AIS data. Li et al. (2018) established a multi-objective and multi-layer fuzzy optimization model to analyze navigational risk in different sea areas with some key parameters extracted from the AIS data. Current literatures on AIS-based navigational risk analysis mainly depend on the expert risk model which emphasizes on the perspective of specific vessels, for instance, the collision probability between vessels, the vessel maneuvering, or the vessel domain, with limited AIS information integrated (Chen et al. 2019a). However, the applicability of these methods is quite limited since the expert risk models rarely consider the macroscopic behavior pattern of vessels.

Another kind of AIS-based risk analysis is realized by statistical methods. For instance, Bye and Aalberg (2018) conducted statistical analyses between AIS and maritime accident database in Norwegian waters. Correspondence analysis and logistic regression were used to discover the associations between vessels' and accidents' probability. Sormunen et al. (2016) calculated different types of accident frequencies relative to the traffic volume through numerical analysis of AIS and accident data. Although the statistical methods can reveal the internal correlation between risks and vessels to a certain extent, they mainly focus on numerical analysis, ignoring the important spatial features of AIS data.

More recently, with the advent of big data, there is an increasing number of researches use data mining techniques to extract useful information from AIS big data in the fields of route recognition (Lee et al. 2020), ship emission (Li et al. 2016) and fishery activities (Kroodsma et al. 2018). Also, several scholars begin to use machine learning or GIS intelligent algorithms to process AIS big data for analyzing navigational risks (Pallotta et al. 2013; Li et al. 2019; Zhao et al. 2018). Different from previous researches, in their works, the AIS served as the main data source, rather than assisted data in risk analysis. Although such manner of big data can fully explore the hidden information from AIS data and has gradually become popular in community, there are still

research gaps in this field. The existing literatures mainly focus on the static spatial characteristics of AIS trajectories, while the spatiotemporal dynamic characteristics are generally ignored (Zhang et al. 2019). For navigational risk analysis, however, it is vital to have a deep understanding of both the spatial and the temporal characteristics of the vessel trajectories due to that some potential risk factors can only be detected in the context of dynamic changes.

To address the above research gaps, in this paper, we provide a novel spatiotemporal dynamic perspective to investigate the navigational risks hidden in massive AIS trajectories. The contribution of this study is twofold. First, we propose a time-series-oriented trajectory processing method to produce traffic density maps with different time periods, and served as the basic products for the follow-up spatiotemporal mining. Second, the proposed method has been successfully applied in NYRB waters, and three types of navigational risks are successfully revealed. The findings can help the maritime authorities to improve their safety management and optimize navigation rules. In fact, based on our findings, the navigation aids in the NYRB waters have already been adjusted pertinently by the Yangtze River Waterway Bureau.

## Study area

The NYRB is located in the Nanjing section of the lower reaches of the Yangtze River, which has been the world's busiest inland waterways since 2010 (Gan et al. 2017). The bridge is China's first double-deck railway and highway bridge across the Yangtze River, connecting the Nanjing City and the Pukou District (Huang et al. 2019). As shown in Fig. 1, the study area covers around 8-km waters from northeast to southwest of the NYRB. The shape of the water area is generally straight and slightly curved. The average width of the water is 1.5 km, and the average depth is 24 m. Due to the high density of vessels and frequent traffic accidents, the NYRB waters have always been the hot spot of maritime supervision.

The main bridge has nine piers and ten spans, as illustrated in top left of Fig. 1; the width of the first span is 128 m and the other nine is 160 m. According to the rule of traffic separation scheme (TSS), there are three spans that are opened for navigation. The fourth span is open for upstream vessels, the sixth span and the eighth span are open for downstream vessels. For each navigation span, the designed maximum navigable clearance width is 120 m, and the navigable clearance height is 24 m above the designed maximum navigation water level. There is an approach channel set up for each navigation span by using pairwise navigation aids, in order to guide the vessels traveling through the bridge safely.

The general direction of the water flow is from southwest to northeast. The water flow in the middle of the river is faster

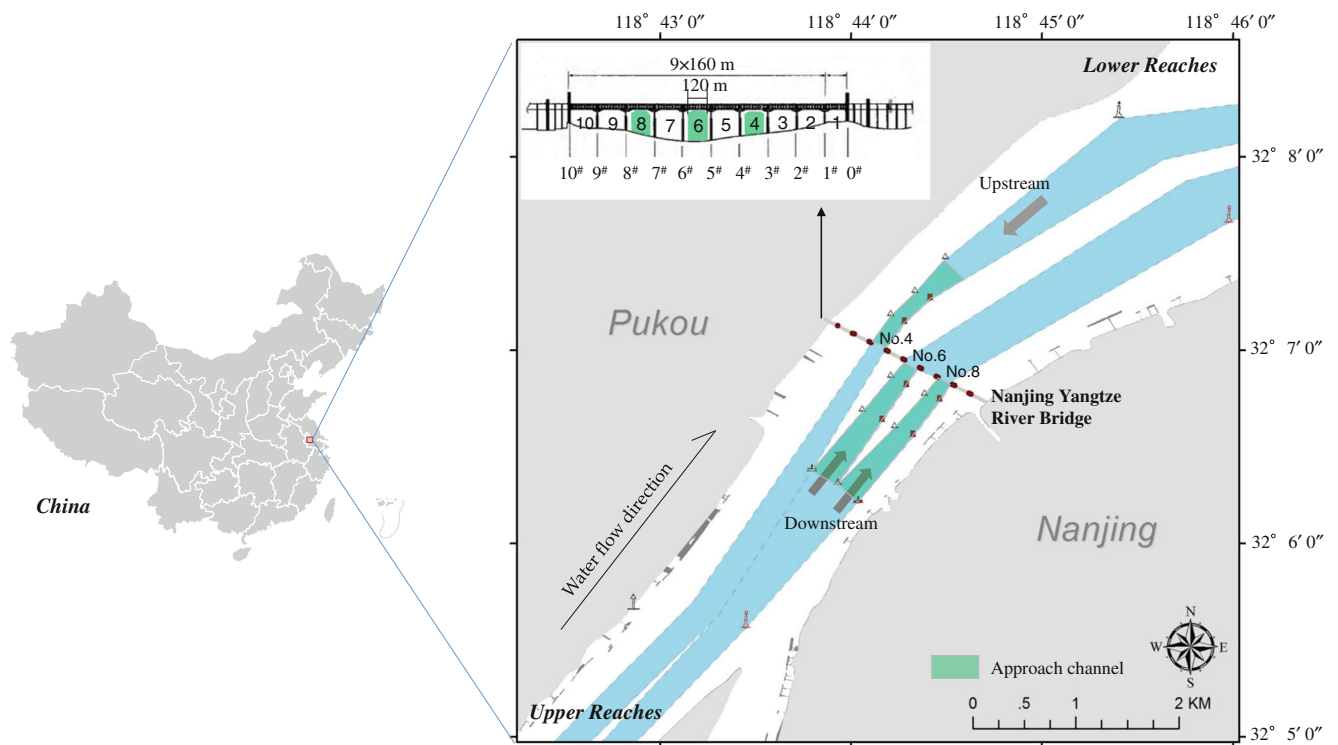


Fig. 1 Overview of the study area

than that in the two banks. In the flood season, normally comes from May to September; the velocity of the water flow increases along with the rising of the water level. Moreover, there is an angle between the water flow direction and the normal direction of the bridge axis. The historical observation records show that the angle at the sixth span is larger than the fourth and eighth span.

The AIS data used covers the entire NYRB waters in the whole year of 2019, including more than 55 million AIS dynamic messages. According to the official navigation rules of the Yangtze River, in the NYRB waters, the vessel length of 80 m is used as the criterion for distinguishing large vessels and small vessels.

## Method

The framework of the time-series-oriented trajectory processing method is shown in Fig. 2. The real-time dynamic AIS data streams are received from the maritime agency of the Yangtze River. Each record of AIS data contains the information of Maritime Mobile Service Identity (MMSI) code, ship name, latitude and longitude, speed over ground (SOG), course over ground (COG), and timestamp. The static vessel attribute data, such as length and breadth, draft-depth, and type, are also received from the maritime agency; however, such information is generally incomplete and unreliable due to management issues. To complete that, we make use of a web

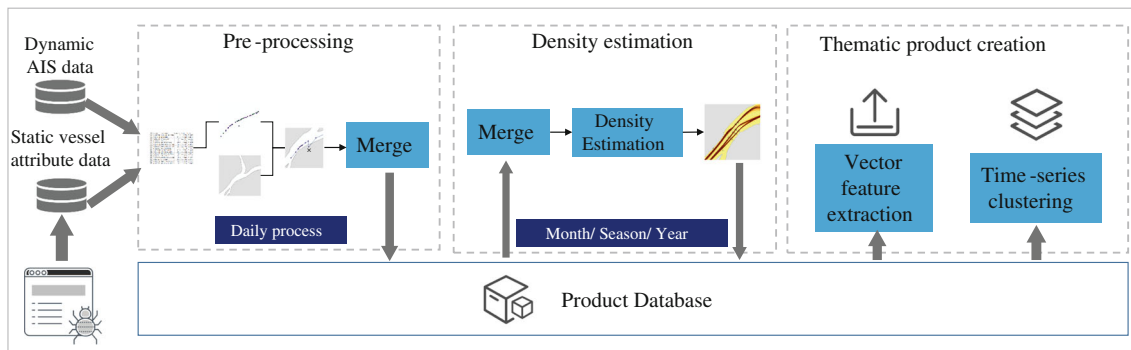
crawler to obtain the attribute information of vessels, in accordance with the global unique MMSI codes, from online AIS data providers.

The first step of the method is daily pre-processing. The workflow starts automatically every day at 0:30 a.m. to pre-process AIS data received yesterday. The pre-processing includes trajectory point reconstruction, error records removal, and abnormal shape elimination. After pre-processing, all corrected AIS trajectories are merged into a dataset labeling with daily timestamp and then stored in the product database.

The second step is density estimation with different time periods. The daily datasets are fetched from the product database according to a user-defined time period such as a month, a season, or even a whole year. Then AIS traffic density maps are created by a density estimation algorithm based on the daily datasets.

The third step is thematic product creation. There are two types of thematic products for spatiotemporal analysis. One is the vector feature products including lane boundary and lane centerline extracted from the traffic density maps. The other one is time-series clustering products. We will give a detailed description for them later.

The method was implemented in Python 3.6, with some GIS functions implemented by using arcpy. Microsoft SQL Server 2018 was used as the products' database. Compare with the existing AIS data-processing methods in literatures, the main advantage of our method is that it adopts a periodic processing strategy to produce AIS trajectory products, which



**Fig. 2** Framework of time-series-oriented trajectory processing method

enables the subsequent spatiotemporal analysis more flexible in the choice of time granularity.

### Pre-processing

AIS data are generally suffered from noise information due to the influence of signal interference, device failure, or transmission loss (Qu et al. 2011). The main aim of the pre-processing is to make AIS data reliable for traffic analysis. The pre-processing takes three steps to eliminate data noises.

The first step is trajectory point reconstruction. The AIS trajectory points belonging to the same vessel are grouped according to the same MMSI code. Then the trajectory points in the same group are reorganized by chronological order and connected to form a trajectory line. Note that if the time interval of two consecutive points in a trajectory line exceeds a certain threshold, it should be segmented into two different trajectory lines. Here, the predefined threshold of the time interval is set to 10 min.

The second step is error record removal. Three types of errors are widespread in the original AIS data and should be removed in pre-processing, including the record with (1) the length of MMSI code is not equal to nine; (2) the latitude value is not in the range of  $(-90^\circ, 90^\circ)$ , or the longitude value is not in the range of  $(-180^\circ, 180^\circ)$ ; (3) the trajectory points of the same vessel are too sparse; here, sparse means the number of points is less than ten.

The third step is abnormal shape elimination. If the distances between a trajectory point and its preceding and subsequent nodes exceed certain values, and their connecting lines form a sharp angle, this trajectory point is regarded as the “jumping” point and should be removed as the outlier. In addition, if a trajectory line intersects with land areas, it should be excluded since it is impossible for vessels to sail on land.

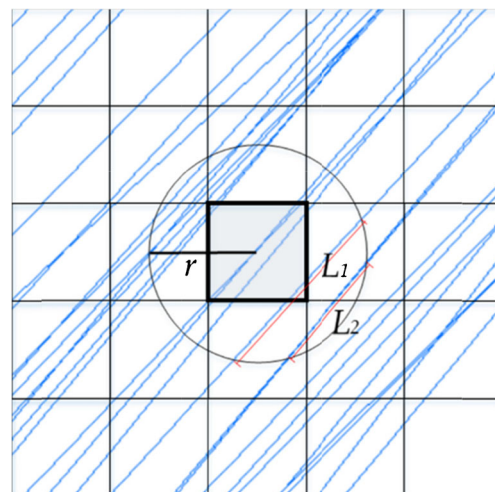
### Density estimation

Density estimation is the most straightforward and effective manner for highlighting the distribution of the vessel trajectories. It is the most commonly used GIS function that maps the

vector points or lines to continuous regular grids. Density estimation quantifies vessel navigation behaviors with probability density values; the larger the grid value means the more vessels have traveled across the area.

There are various ways to realize density estimation. An intuitive and direct way is the point-based method that overlays a very detailed grid upon AIS trajectory points and counts the number of points falling in each grid. However, the point-based method has an obvious flaw that if a vessel goes slowly, the density of points is much greater than when a vessel goes fast. This is determined by the AIS working principle that the slower the vessel’s speed, the more trajectory points it will send out per second, and vice versa.

Instead of the point-based method, we adopted the line-based density estimation (LDE) method to obtain traffic density. The basic principle of LDE is to calculate the density value of linear elements within the neighborhood of each output grid. As shown in Fig. 3, the blues lines represent the AIS trajectory lines. The circle is plotted around the grid center using a predefined radius parameter  $r$ . The weighted sum of the length of each trajectory line falling into the circle is calculated, then divided by the circle’s area. For example, as shown in Fig. 3,  $L_1$  and  $L_2$  denote the length of the portion of two trajectory lines falling in the circle



**Fig. 3** The principle of line-based density estimation



area. The corresponding weights are  $w_1$  and  $w_2$  (in practice, the recommended value of the weight is one). For a given grid, if there are  $n$  trajectory lines that go across its searching circle area, the density value of the grid is defined as Eq. (1):

$$D = \frac{\sum_{i=1}^n (L_i \times w_i)}{\pi r^2}, \quad (1)$$

After all grids are calculated, the AIS traffic density map of the whole region is obtained. It is worth noting the distinction between the concepts of the grid and the circle here. The grid refers to an independent calculation unit based on the regular partitioning of the study area, while the circle is an abstract description of the search area for calculating the density value of each grid cell. The radius  $r$  of the circle is a flexible parameter that reflects the searching scope of the grid. It is an empirical value adjusted dynamically according to the performance of the density estimation results. In our application, the parameter  $r$  is set to 10 m.

### Vector feature extraction

In this step, such vector features that outline the shape of traffic lanes, including lane boundaries and lane centerlines, are extracted from the traffic density maps. These two vector features are key elements for the spatiotemporal analysis in investigating delicate variations of the traffic lanes and quantitative spatial relationships between vessels and navigation facilities.

The lane boundary refers to the edge of the main traffic lane with the density value in the probability range of 95% of the total density distribution (Breithaupt et al. 2017). It is the representative of principle waterways that is normally used by most vessels and have the largest traffic volume (Chen et al. 2015). The lane boundary is a two-dimensional polygon which can be derived from a given traffic density map by leveraging map algebra and converting functions of the GIS software such as ArcGIS 10.2.

The lane centerline depicts the core skeleton line formed by the highest densities in the traffic lane (Davies et al. 2006). It describes the basic form and direction of the traffic lane. The lane centerline can be extracted from the binarized result of the density map using vector generation tools provided by the ArcScan module.

### Time-series clustering

Clustering is a widely used data mining technique to group a number of objects into homogeneous clusters, where objects have the maximum similarity with other objects in the same cluster (Kaufman and Rousseeuw 2009). A special form of clustering is time-series clustering (Esling and Agon 2012). The time-series clustering focuses on analyzing a collection of values obtained from sequential measurements over time. It is

an effective way of finding the spatiotemporal characteristics of geospatial data since it not only measures the spatial similarity but also measures the similarity of the dynamic change pattern over time. A representative case is the time-series clustering of normalized difference vegetation index (NDVI) data (Xia et al. 2019). Monthly NDVI data of the year were retrieved from remote sensing images and composited into a 12-layer dataset. After a certain clustering algorithm performed on the dataset, the regions belonging to the same category could be interpreted as the area with the same vegetation cover, since these regions have similar phenological fluctuation characteristics within the whole year.

Inspired by the NDVI case, we apply clustering on time-series traffic density maps. Specifically, monthly traffic density maps are organized in chronological order and composited into a 12-layer raster dataset. All the density values are normalized before clustering. As for the specific clustering algorithm, K-Means is chosen due to its simplicity and efficiently (MacQueen 1967). K-Means aims to partition  $n$  objects into  $k$  clusters in which each object belongs to its nearest centroid. Given  $n$  objects  $(x_1, x_2, \dots, x_n)$  in  $d$ -dimensional space  $R^d$ , the problem is to determine  $k$  centroids  $(C_1, C_2, \dots, C_k)$  for disjoint clusters  $S_1, S_2, \dots, S_k$ , to minimize the mean squared distance norm in the partitioning metric:

$$J = \sum_{i=1}^k \sum_{x_j \in S_i} D^2(x_j, C_i), \quad (2)$$

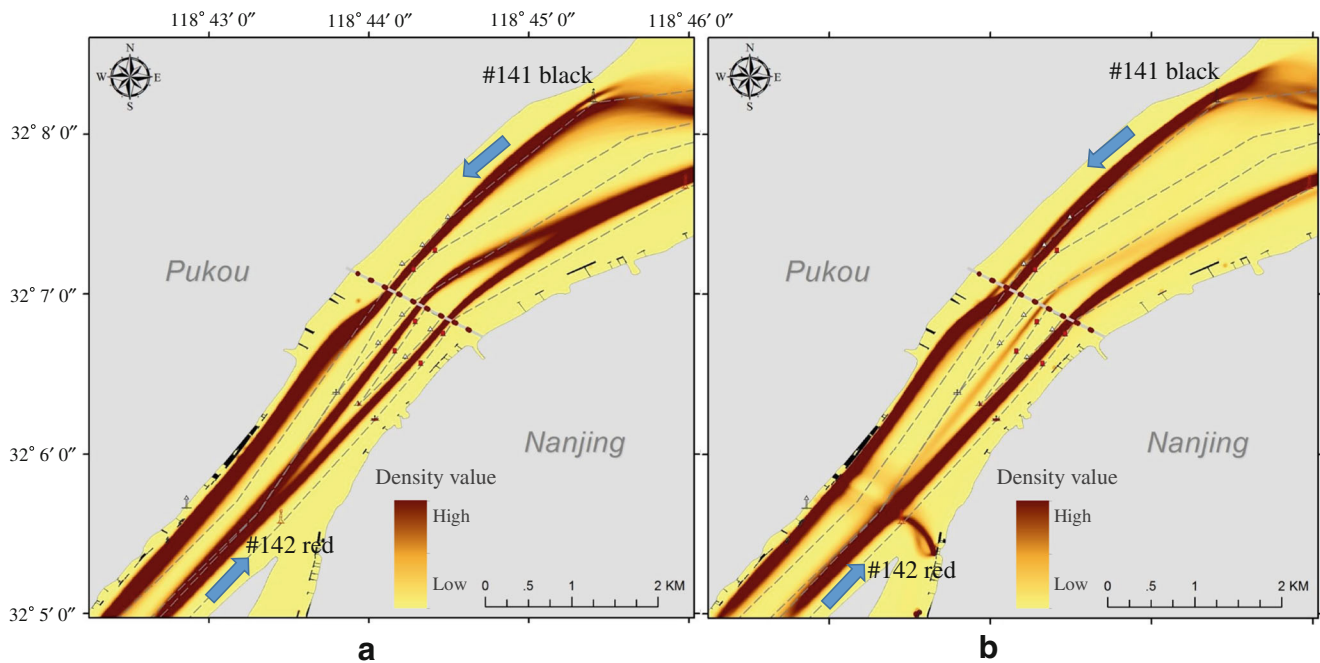
where  $D^2(x_j, C_i)$  denotes the Euclidean distance of object  $x_j$  from centroid  $C_i$ .

Through the time-series clustering, we can identify those water areas with the similar spatiotemporal variation characteristics. Particularly, we can recognize which parts of the waters are in stably high navigation density throughout the year in the context of monthly dynamic changes. These results can provide reliable references to analyze navigational risks.

## Results

### Yearly distribution of AIS trajectories

Figure 4 shows the overall distribution of the vessel trajectories in the whole year. Figure 4a and Figure 4b represent the large vessels and the small vessels respectively. As we can see from the figure, in general, the distribution of the vessel trajectories highlighted three distinct traffic lanes. The upstream vessels enter the approach channel of the fourth span after passing the #141 black navigation aid. After traveling through the bridge, the vessels take an obvious turn to the north bank of Pukou and continued to travel closely to the shore. The downstream vessels enter the approach channel of the sixth



**Fig. 4.** Overall distribution of AIS trajectories in the NYRB waters

span or the eighth span to go through the bridge after passing the #142 red navigation aid. After that, the two traffic streams are converged.

On the other hand, we can also observe the differences of the trajectory distributions between the large vessels and the small vessels.

1. For upstream vessels, Fig. 4a shows that the large vessels generally choose the fourth span to go through the bridge, while Fig. 4b indicates that there are a certain number of small vessels which choose the third span. Furthermore, by comparing the two figures, we can find that after traveling through the bridge, the average distance between the vessel trajectory of the small vessels and the north shore is smaller than that of the large vessels.
2. For downstream vessels, Fig. 4a shows that the traffic densities of the large vessels in the approach channel of the sixth span and the eighth span are basically the same; while in Figure 4b, the traffic density of the small vessels in the approach channels of the sixth span is much lower than that of the eighth span. This indicates that the downstream small vessels are more inclined to choose the eighth span to go through the bridge.
3. Figure 4b shows that near the #142 red navigation aid, there exists an obvious crossing track, which is mainly formed by the small city ferries (40 m–50 m) connecting the Nanjing City and the Pukou District.

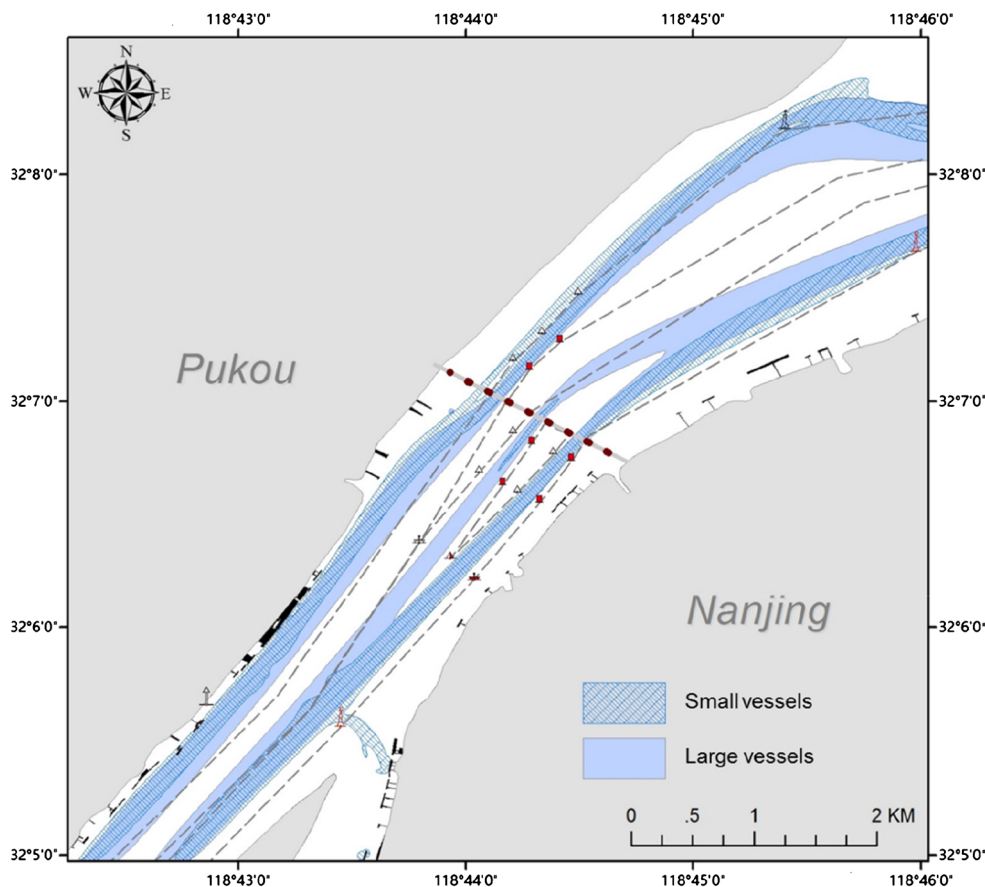
Figure 5 is the comparison result of the lane boundaries. It can be clearly seen that the lane boundary of the upstream small vessels is closer to the north shore than that of the large

vessels. Besides, the overlap between the lane boundary of the upstream vessels and the designed waterway, denoted with dotted lines, is much lower than that of the downstream vessels.

### Seasonal variation of AIS trajectories

Figure 6 shows the variations of the AIS trajectories in different seasons. Here, for vessels traveling in the Yangtze River, the seasons refer in particular to the flood season (May to September) and the dry season (January to April and November to December). In this case, center-lines were used to represent the traffic lines for a clear contrast. The solid lines represent the large vessels and the dotted lines represent the small vessels; the red lines represent the flood season and the blue lines represent the dry season. Note that the lane centerlines of the small vessels in the approach channel of the sixth span could not be explicitly extracted due to the traffic lanes that were not salient. To demonstrate the differences in detail, we further took two cross sections at the upstream and downstream of the bridge respectively, and drew traffic density curves of the cross sections, as shown in Fig. 7 and Fig. 8. It can be found that the red lines are always on the north side of the blue lines, whether they are dotted lines or solid lines. This indicates that vessels are always closer to the north shore in the flood season than in the dry season whether they travel upstream or downstream. This trend can be observed more clearly from the variations of the density curves in Fig. 7 and Fig. 8.

**Fig. 5** Comparison of the lane boundaries between large vessels and small vessels



Next, the spatial relationship between vessels and the bridge axis was quantitatively investigated. This relationship provides an important reference for analyzing navigational risks in NYRB waters since it reflects the attitude and position of the vessels when passing through the bridge span. Figure 9 describes two indices for measuring the spatial relationship between the bridge axis and a sailing vessel. The  $\alpha$  means the included angle between the vessel's bow direction and the normal direction of the bridge axis (dotted line). The  $d$  means the distance between the central axis of the hull and the center of the bridge span. Here, the extracted lane centerlines were used to represent the central axis of the hull for calculating the indices of  $\alpha$  and  $d$ .

Table 1 and Table 2 show the different values of  $\alpha$  and  $d$  at different bridge spans in different seasons. From Table 1, we can find that the angle  $\alpha$  is maximum at the eighth span, followed by the fourth span and the sixth span. In addition, we can also find that the angle  $\alpha$  in the flood season is larger than that in the dry season, and the angle of the small vessels is larger than large vessels. From Table 2, we can observe that the absolute value of the distance  $d$  of small vessels is generally larger than that of large vessels. For small vessels, the absolute value of  $d$  is maximum at the fourth span, followed by the eighth span and the sixth

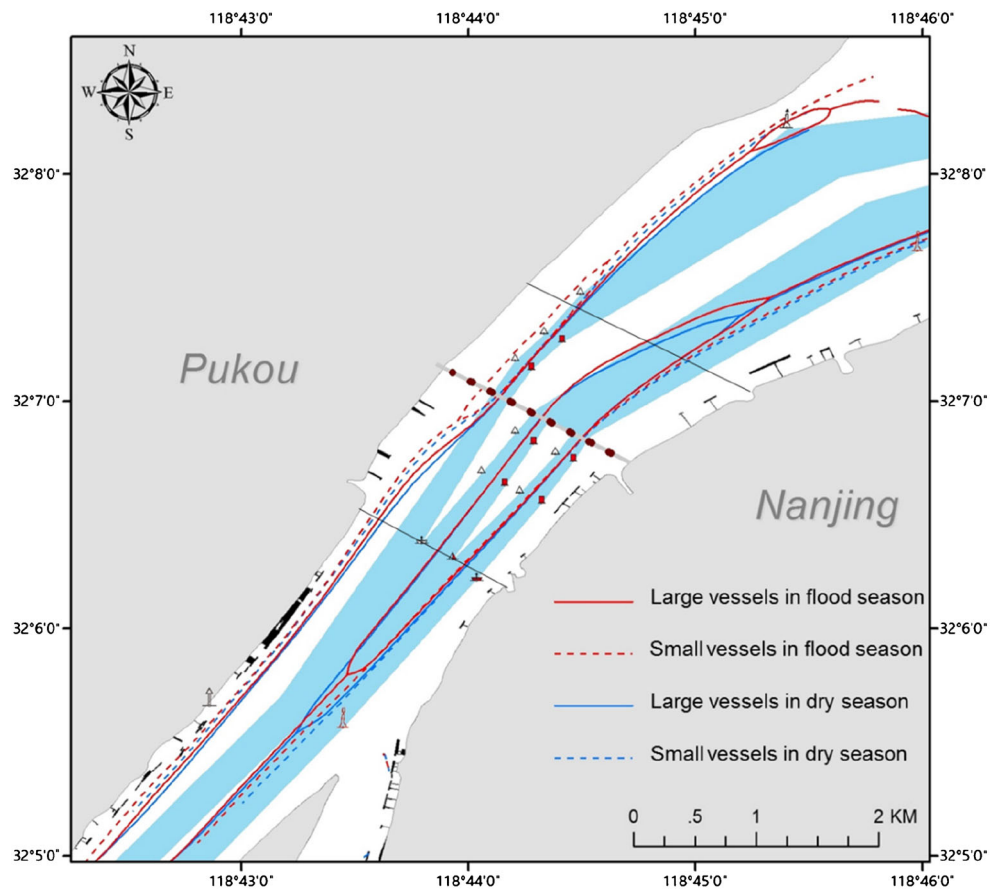
span. For large vessels, the absolute value of  $d$  is maximum at the eighth span, followed by the fourth span and the sixth span.

### Monthly evolution of AIS trajectories

Figure 10 depicts the monthly evolution of the AIS trajectories. For clarity, the representative centerlines are drawn in the first and second half of the year, and different colors are used to represent different months. As depicted in Fig. 10a, the lane centerlines of the upstream large vessels show a distinct changing trend from southeast to northwest over time. The downstream large vessels show the same trend, while the changing extend is relatively smaller. Meanwhile, the change of the centerlines in the approach channel of the sixth span is not obvious. In Fig. 10b, we can find the same changing trend for the small vessels. In Fig. 10c and Fig. 10d, we can observe that in the second half of the year, the centerlines show a reverse changing trend from northwest to southeast, and the changing extend is even greater than the first half of the year.

We infer that there is a correlation between the monthly evolution of the AIS trajectories and the change of water level. To verify our speculation, we measured the monthly change extent of the AIS trajectories by calculating the average distance between the centerline and the right-side boundary line of the

**Fig. 6** Spatiotemporal variations of the AIS trajectories in different seasons

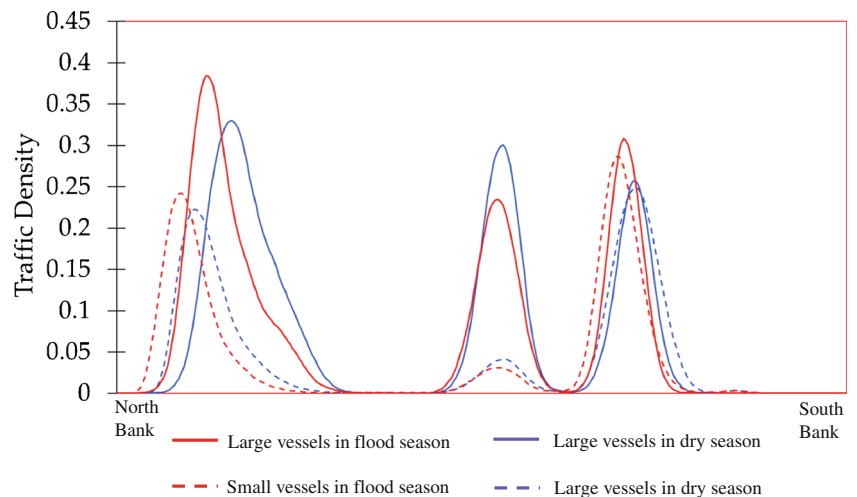


waterway; here, the right-side is relative to the waterway’s direction. As shown in Fig. 11, the bar charts represent the distances between the centerlines and the boundary lines of the waterway, the line charts represent the monthly average water level in the Nanjing section of the Yangtze River from January to December in 2019. We can find that the monthly change extent is consistent with the change of the water level. Such consistency is particularly marked as the water level rises suddenly from February to March (which is called the spring flood).

**Time-series clustering of traffic density maps**

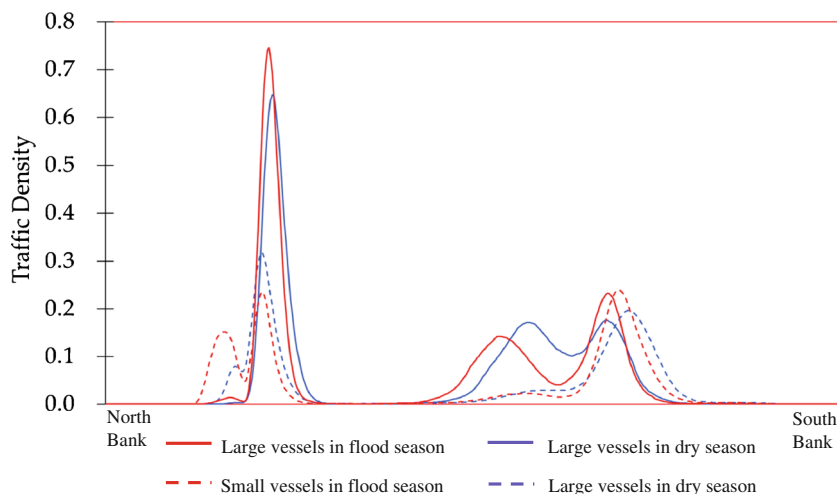
Figure 12 shows the results of the time-series clustering, where Fig. 12a represents the large vessels and Fig. 12b represents the small vessels. As an unsupervised learning algorithm, the output classes of K-Means have no labels. We identified each class manually based on the prior knowledge of the actual traffic distribution. Here, in the two figures, class 1 with red represents those water areas with the highest density value

**Fig. 7** Traffic density curves of the cross section at the upstream of the bridge





**Fig. 8** Traffic density curves of the cross section at the downstream of the bridge

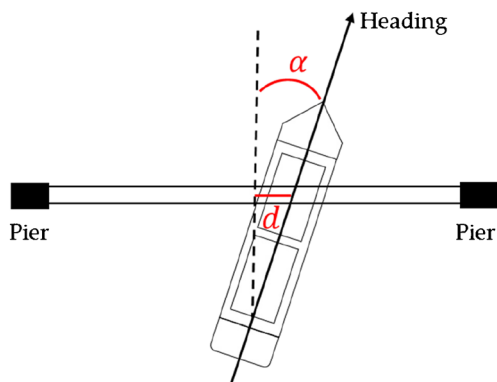


and the most similar variation pattern, which means the navigation pressure is stably high. The representativeness gradually decreases from class 2 to class 4 due to that the inner similarities of the density values descend. From Fig. 12a, we can observe that the class 1 mainly distributes in the three approach channels of the navigable spans. The red area in the approach channel of the fourth span is distinctly longer than that of the sixth and the eighth span. This indicates that the navigation pressure of the waterway for the upstream large vessels is stably high throughout the year. Figure 12b shows the same high pressure in the approach channel of the fourth span for small vessels. Besides, the navigation pressure also focuses on the approach channel of the eighth span, mainly because the small vessels always choose the eighth span to go through the bridge.

**Discussions**

**Safety distance risk**

As we discovered from the overall distribution of AIS trajectories, the upstream vessels prefer to travel closely to the north



**Fig. 9** The spatial relationship between the bridge axis and a vessel

shore of Pukou after passing through the bridge, and this trend is even more pronounced for small vessels in the flood season. Moreover, based on the monthly results, the distance between vessels and the north shore has become closer as the water level rises.

These can be explained by the vessels' choice for the most economic traveling in considering the water flow condition. As we introduced in the "Study area" section, the water flow in the middle of the river is faster than that of the two banks. Therefore, the upstream vessels prefer to sail in nearshore waters with lower resistance for reducing the oil consumption, increasing the speed, and promoting the economic benefits. Especially in the flood season, as the flow velocity increases and the water surface widens, the vessels are closer to the north shore.

However, such behavior presents great risks to navigational safety. If the safety distances between the vessels and the coastal wharves are too small, it may cause serious collisions, particularly when there are several vessels berthing side by side at the wharves. We highlight such safety distance risks by capturing snapshots from remote sensing images of the NYRB waters. As we marked in Fig. 13a, the distances between the starboard side of the sailing vessels with wakes and the portside of the berthing vessels are merely 14 m and 16 m respectively. Therefore, if a sailing vessel is not maneuvered

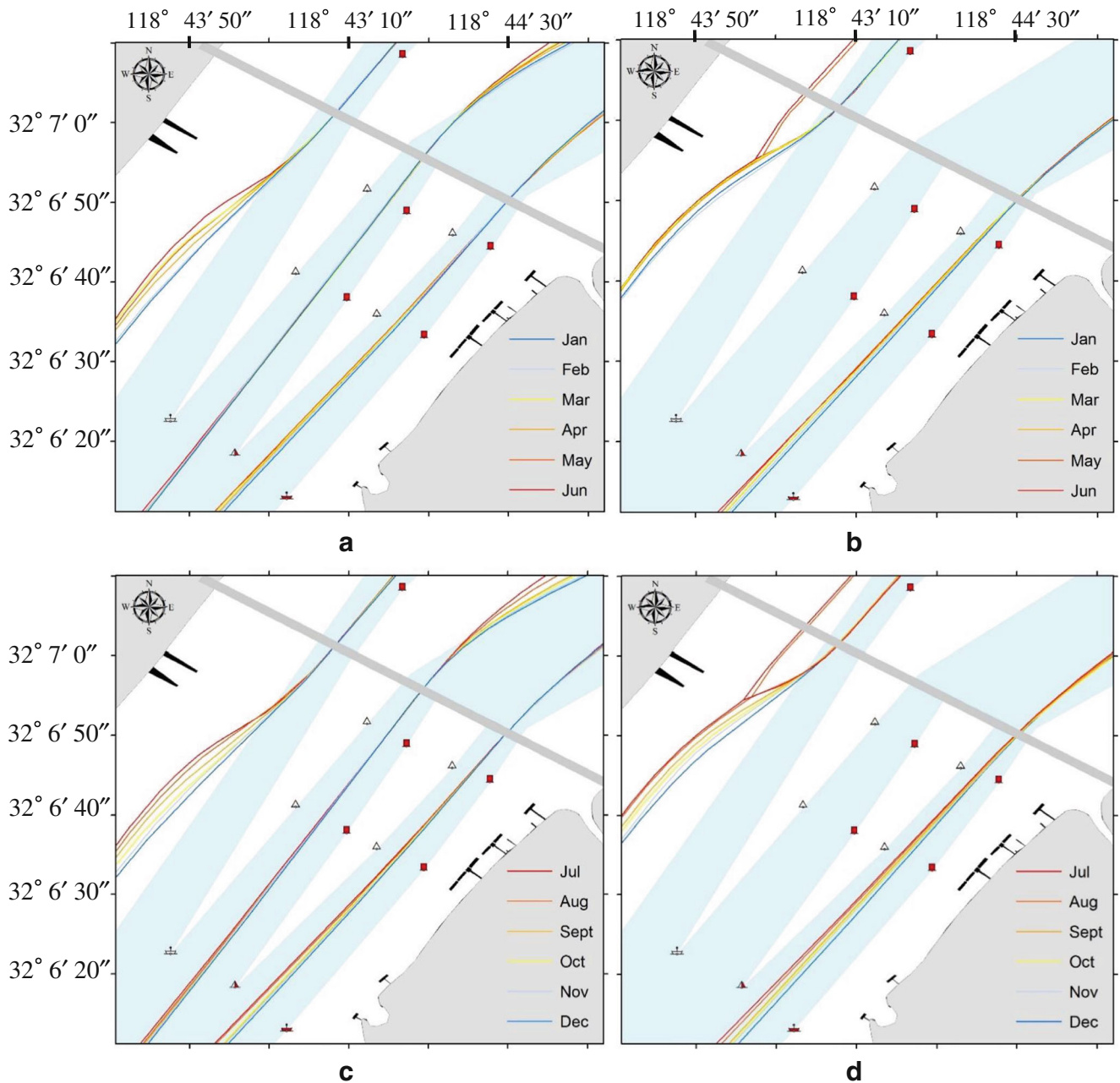
**Table 1** The variation of angel  $\alpha$  at different bridge spans in different seasons

		Span 4	Span 6	Span 8
Large vessels	Flood season	13.58°	11.46°	17.7°
	Dry season	13.54°	11.11°	16.24°
Small vessels	Flood season	18.9°	-	19.95°
	Dry season	16.47°	-	19.05°

**Table 2** The variation of distance  $d$  at different bridge spans in different seasons (south is positive)

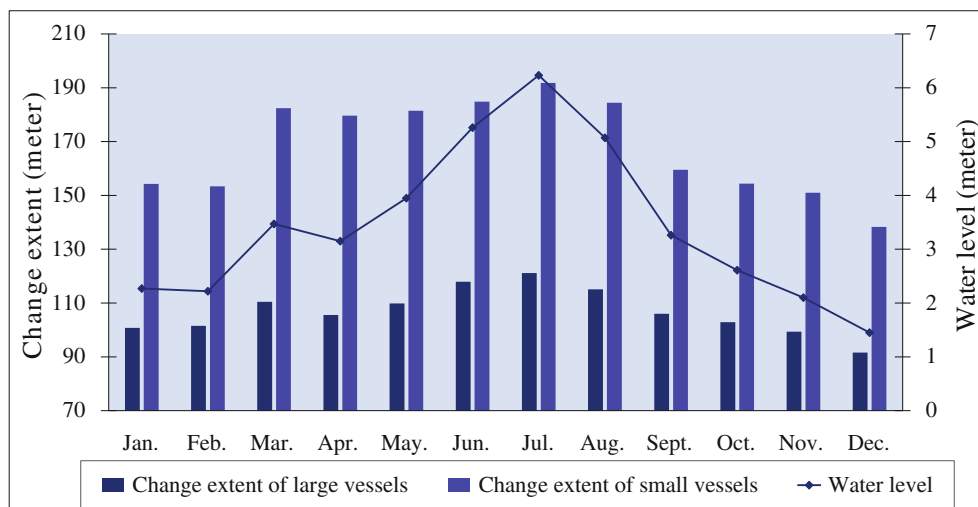
		Span 4	Span 6	Span 8
Large vessels	Flood season	-8.02 m	5.21 m	8.55 m
	Dry season	-6.56 m	4.44 m	8.9 m
Small vessels	Flood season	-26.54 m	-	13.73 m
	Dry season	-25.93 m	-	17.93 m

properly, it is very likely to touch with berthing vessels. The situation would be even worse when berthing vessels begin to move. As shown in the white box of the Fig. 13b, the large container is moving, while the small cargo is passing by, serious collision may occur if either of them made a mistake. Hence, it is very important for maritime authorities to enhance the supervision for the safety distance between vessels and the north shore. In addition, special attention should be paid to the behavior of small vessels since in the Yangtze River a lot of small vessels are owned by family businesses, who usually lack formal navigation education.



**Fig. 10** Monthly evolution of the AIS trajectories. **a** Large vessels (Jan.–Jun.). **b** Small vessels (Jan.–Jun.). **c** Large vessels (Jul.–Dec.). **d** Small vessels (Jul.–Dec.)

**Fig. 11** Relationship between the monthly change of AIS trajectories and the water levels



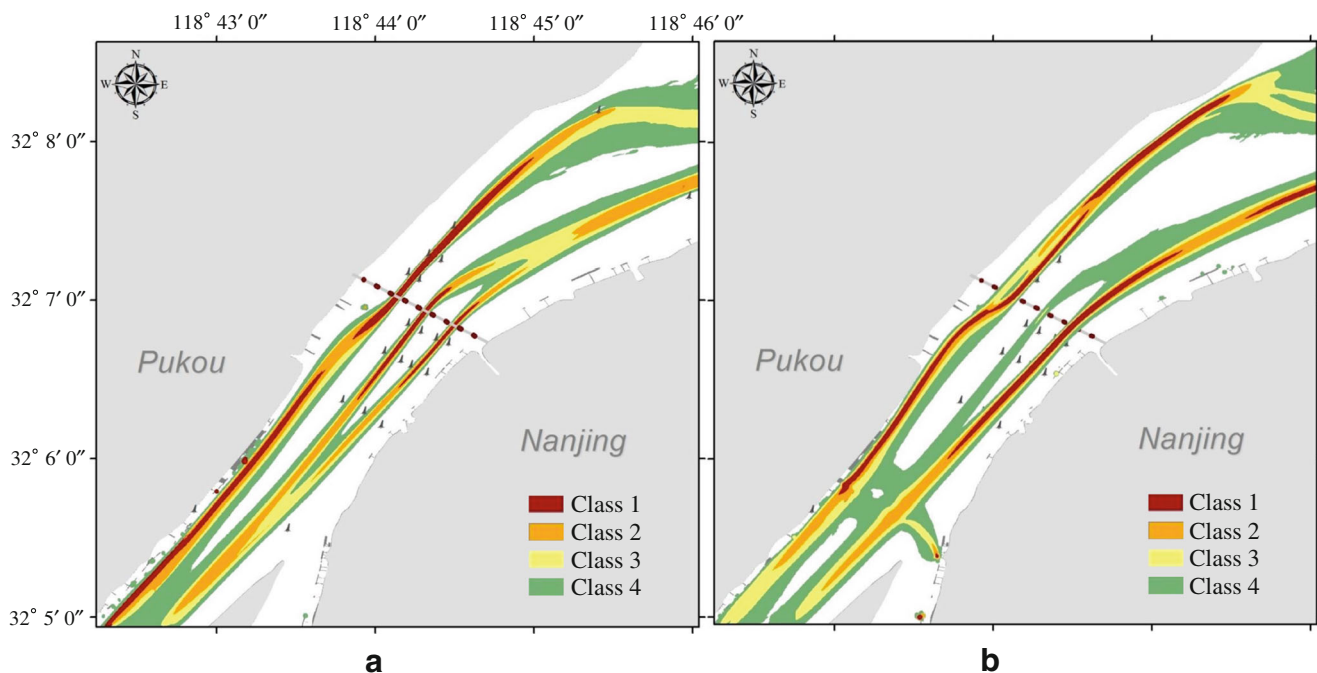
**Pier collision risk**

In investigating the spatial relationship between vessels and the bridge, we find that the bow directions of vessels are not perpendicular to the bridge axis as they passed through the bridge span, and the positions of vessels are not within the middle of the bridge span either. The abovementioned angle  $\alpha$  and the distance  $d$  in the flood season are distinctly larger than those in the dry season. Besides, we also find that the  $\alpha$  and  $d$  of small vessels are larger than large vessels.

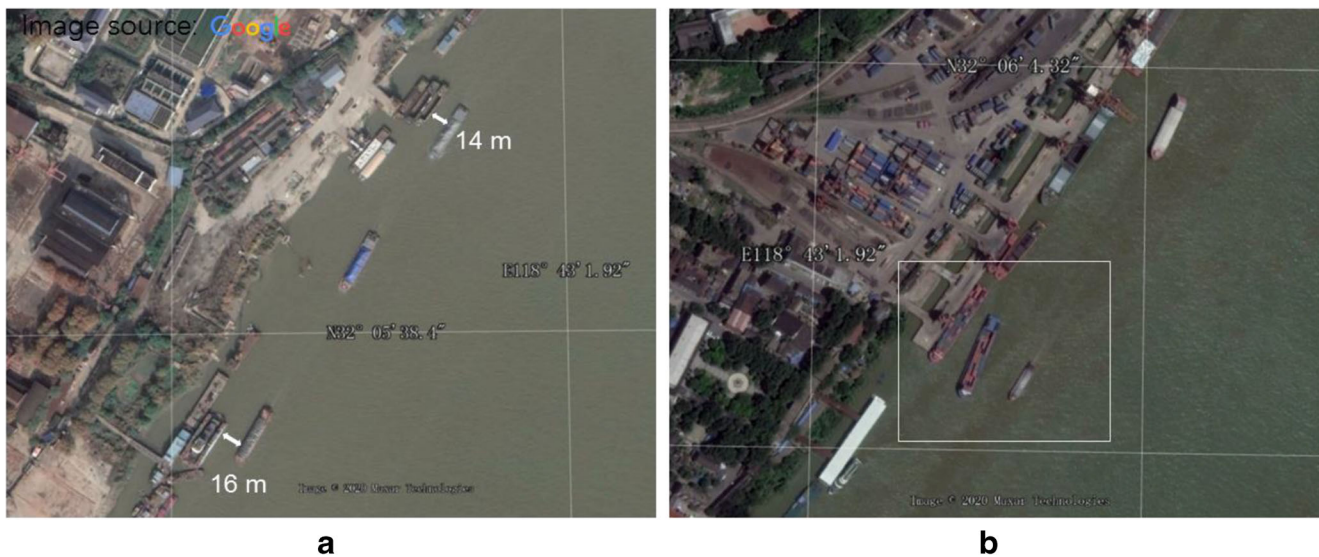
This phenomenon is caused by the water flow conditions under the NYRB. As we mentioned, there is an angle between the water flow direction and the normal direction of the bridge

axis. Hence, the unfavorable force of the water flow would easily carry the vessels to collide to the bridge piers. In fact, the reason why there is a high degree of consistency between the monthly evolution of AIS trajectories and the water level change is that vessels have taken maneuvers to avoid such collision. Vessels need to adjust the hull to north before they sailed into the narrow approach channel, in order to obtain more space to rectify their position to aim at the span. The higher the water level and the faster the water flow, the further north the vessel would take.

As shown in Fig. 14, we highlight such pier collision risk by overlaying the lane centerlines on a remote sensing image. The bridge piers were numbered with tags 3# to 9#. As we can



**Fig. 12** Time-series clustering results of traffic density maps



**Fig. 13** Examples of the safety distance risk in the NYRB waters. **a** The dangerous distances between the sailing vessels and the berthing vessels. **b** Possible collision between the two vessels

see from Fig. 14, the distribution of the extracted vessels from the image is generally consistent with the lane centerlines. The vessels near the fourth span (between the 3# pier and the 4# pier) and the eighth span (between the 7# pier and the 8# pier) are not perpendicular to the bridge, and they are relatively closer to the 3# pier and the 8# pier respectively. The collisions may occur between the vessels and the 3# pier and 8# pier if they are not properly operated, especially when they are in high speeds. Therefore, it is of great importance for maritime authorities to remind vessels to slow down before they passed through the fourth and the eighth span. Furthermore, collision-preventing devices are also necessary to protect these piers of the NYRB.

### Traffic congestion risk

Based on the time-series clustering results, we discover that the navigation pressure in the approach channels of the fourth and the eighth span are stably high throughout the year. It is not difficult to understand the high pressure of the fourth span because it is the only navigable span for vessels to travel upstream. While for the eighth span, the stably high pressure is mainly caused by the traffic flow of small vessels. Although small vessels can choose the sixth span to route, they prefer to choose the eighth span due to that the water flow near the sixth span is irregular and the velocity is faster.

However, such long-term persistent high pressure will bring great traffic congestion risk to safety navigation. The narrow bridge span is the primary cause of accidents. The literature from Park et al. (2008) reported that nearly 90% of the accidents associated with bridges across waterways happened when the width of the navigable span is less than 500 m. However, the clearance width of the navigable span of the

NYRB is merely 120 m. Due to this limitation, only one vessel can pass at a time, and other vessels must slow down and queue up to wait in the approach channel. A certain number of small vessels even choose the unofficial third span to travel in order to catch up with time. Figure 15 demonstrates the traffic congestion risk through snapshots from remote sensing images. In Fig. 15a, a crowd of sailing vessels congests in the narrow approach channel of the fourth span; among them, there is even a fleet of 14 vessels. It is easy to rub against each other in the narrow space. Moreover, the distance between the front and rear vessels may be sometimes very close when speed is fast. As shown in Fig. 15b, the distance between the two small vessels is less than 23 m, nearly colliding with each other. Note that the remote image was captured in April 2020, which means it was still in the period of Coronavirus 2019, the traffic flow was much lower than usual. Once an accident occurs near the bridge span, more serious secondary accidents will happen due to the large inertia of subsequent vessels.

In order to reduce the traffic congestion risk in the NYRB waters, we suggest optimizing the navigation rules that opens the third span or the fifth span for navigation, to release the navigation pressure of the fourth span. Meanwhile, we suggest opening the ninth span for navigation to divert the traffic flow of the eighth span as well.

Compared with the existing researches of AIS-based navigational risk analysis, the contribution of our method is that it can find the potential navigational risks hidden in the spatial and temporal variation of massive AIS trajectories. For safety distance risk analysis, the traditional static mining method can only find the phenomenon that the safety distance is too small. However, by using our method, we can not only find out the phenomenon but also identify the relationship between the distances and water levels. For traffic congestion risk analysis,



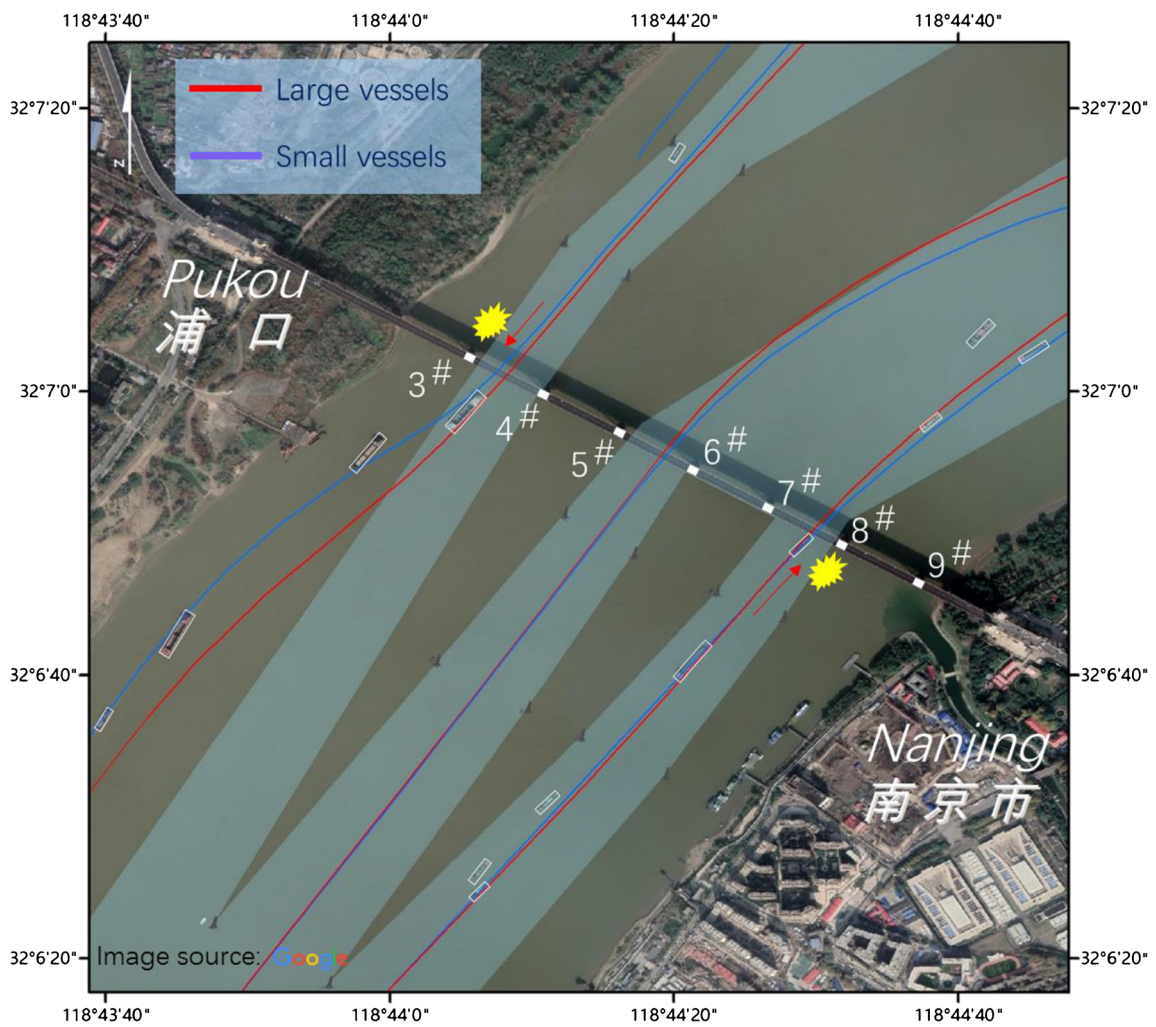


Fig. 14 An example of the pier collision risk in the NYRB waters



Fig. 15 Examples of the traffic congestion risk in the NYRB waters. **a** A crowd of sailing vessels congests in the narrow approach channel. **b** Possible rear-end collision between the two vessels

the traditional static mining method can only detect which places are congested roughly, while by using our time-series-oriented method, we can gain a deep insight on which routes are under high navigation pressure all year round.

## Conclusions

In this paper, navigational risks in the NYRB waters are analyzed based on spatiotemporal mining on massive AIS trajectories by using GIS techniques. A time-series-oriented trajectory processing method was proposed to deal with the AIS data in the whole year of 2019. The method leverages the line density estimation to produce vessel traffic products and provides temporal dynamic traffic information for the spatiotemporal analysis. Supported by the flexible framework of the method, complete pictures of the spatiotemporal distribution and variation characteristics of AIS trajectories were painted. Based on that, a comprehensive analysis for navigational risks was conducted, and several critical potential risks in the NYRB waters were revealed.

First, we identify that as the water level rises, the safety distances between upstream vessels and the north shore of Pukou become smaller, which increases the probability of the collision between sailing vessels and the vessels berthed at wharves. Second, when vessels pass through the narrow spans of the NYRB, the angle between the vessel and the normal direction of the bridge axis increase with the rise of water level, which may bring serious collision to the 3# and the 8# bridge piers. Third, we detect that the approach channels of the fourth and the eighth span are under high navigation pressure all year round. This may bring the traffic congestion risk and cause severe rear-end collision accidents. Furthermore, we suggest the maritime authorities should pay special attention to the supervision for small vessels in the flood season based on our findings.

The research on leveraging AIS big data mining to analyze navigational risk is still in its infancy. Although our work has shown its promise for precisely capturing vessel behaviors and uncovering potential navigational risks, there are still many challenges in future works. Due to the data limits, in this paper, the vessel size was the only criterion for classifying the spatiotemporal characteristics of the AIS trajectories, while in future works, more criteria including the vessel type and the vessel draft would be used for distinguishing risk analysis objects more precisely. In addition, other factors which may impact navigational risks such as the speed and the course of vessels will also be considered. The exploration of adding the meteorological information to risk analysis will be another interesting future research topic.

**Declarations** The author declares that he has no competing interests.

## References

- Breithaupt SA, Copping A, Tagestad J, Whiting J (2017) Maritime route delineation using AIS data from the atlantic coast of the US. *J Navigat* 70(2):379–394
- Bye RJ, Aalberg AL (2018) Maritime navigation accidents and risk indicators: an exploratory statistical analysis using AIS data and accident reports. *Reliab Eng Syst Saf* 176:174–186
- Bye RJ, Almklov PG (2019) Normalization of maritime accident data using AIS. *Mar Policy* 109:103675
- Chen J, Lu F, Peng G (2015) A quantitative approach for delineating principal fairways of ship passages through a strait. *Ocean Eng* 103:188–197
- Chen P, Huang Y, Mou J, van Gelder PHAJM (2019a) Probabilistic risk analysis for ship-ship collision: state-of-the-art. *Saf Sci* 117:108–122
- Chen YJ, Liu Q, Wan CP (2019b) Risk causal analysis of traffic-intensive waters based on infectious disease dynamics. *J Mar Sci Eng* 7(8): 277
- Chen P, Mou J, Van Gelder PHAJM (2019c) Integration of individual encounter information into causation probability modelling of ship collision accidents. *Saf Sci* 120:636–651
- Davies JJ, Beresford AR, Hopper A (2006) Scalable, distributed, real-time map generation. *IEEE Pervasive Comput* 5(4):47–54
- Esling P, Agon C (2012) Time-series data mining. *ACM Computing Surveys (CSUR)* 45(1):1–34
- Faghih-Roohi S, Xie M, Ng KM (2014) Accident risk assessment in marine transportation via Markov modelling and Markov Chain Monte Carlo simulation. *Ocean Eng* 91:363–370
- Gan S, Liang S, Li K, Deng J, Cheng T (2017) Trajectory length prediction for intelligent traffic signaling: a data-driven approach. *IEEE Trans Intell Transp Syst* 19(2):426–435
- Huang L, Yu L, Zhang H, Yang Z (2019) Composition and microstructure of 50-year lightweight aggregate concrete (LWAC) from Nanjing Yangtze River bridge (NYRB). *Constr Build Mater* 216: 390–404
- Kulkarni K, Goerlandt F, Li J, Banda OV, Kujala P (2020) Preventing shipping accidents: past, present, and future of waterway risk management with Baltic Sea focus. *Saf Sci* 129:104798
- Kroodsma DA, Mayorga J, Hochberg T, Miller NA, Boerder K, Ferretti F, Woods P (2018) Tracking the global footprint of fisheries. *Science* 359(6378):904–908
- Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis, vol 344. Wiley
- Li L, Lu W, Niu J, Liu J, Liu D (2018) AIS data-based decision model for navigation risk in sea areas. *J Navigat* 71(3):664–678
- Li M, Mou J, Liu RR, Chen P, Dong Z, He Y (2019) Relational model of accidents and vessel traffic using AIS Data and GIS: a case study of the Western port of Shenzhen City. *J Mar Sci Eng* 7(6):163
- Li S, Meng Q, Qu X (2012) An overview of maritime waterway quantitative risk assessment models. *Risk Analy* 32(3):496–512
- Lee JS, Son WJ, Lee HT, Cho IS (2020) Verification of novel maritime route extraction using kernel density estimation analysis with automatic identification system data. *J Mar Sci Eng* 8(5):375
- Li C, Yuan Z, Ou J, Fan X, Ye S, Xiao T, Zheng J (2016) An AIS-based high-resolution ship emission inventory and its uncertainty in Pearl River Delta region, China. *Sci Total Environ* 573:1–10
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 281–297). University of California Press.
- Merrick JR, Van Dorp JR, Blackford JP, Shaw GL, Harrald J, Mazzuchi TA (2003) A traffic density analysis of proposed ferry service expansion in San Francisco Bay using a maritime simulation model. *Reliab Eng Syst Saf* 81(2):119–132

- Park YS, Lee YS, Park JS, Cho IS, Lee U (2008) A proposal of bridge design guideline by analysis of marine accident parameters occurred at bridges crossing navigable waterways. *J Navigat Port Res* 32(10): 743–750
- Pallotta G, Vespe M, Bryan K (2013) Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. *Entropy* 15(6):2218–2245
- Pak JY, Yeo GT, Oh SW, Yang Z (2015) Port safety evaluation from a captain's perspective: the Korean experience. *Saf Sci* 72:172–181
- Qu X, Meng Q, Suyi L (2011) Ship collision risk assessment for the Singapore Strait. *Accid Anal Prev* 43(6):2030–2036
- Sormunen OV, Hänninen M, Kujala P (2016) Marine traffic, accidents, and underreporting in the Baltic Sea. *Sci J Maritime Univ Szczecin* 46(118):163–177
- Svanberg M, Santen V, Horteborn A, Holm H, Finnsgard C (2019) AIS in maritime research. *Mar Policy* 106(AUG) 103520.1-103520.10
- Silveira PAM, Teixeira AP, Soares CG (2013) Use of AIS data to characterize marine traffic patterns and ship collision risk off the coast of Portugal. *J Navigat* 66(6):879–898
- Wang Y, Zhang J, Chen X, Chu X, Yan X (2013) A spatial-temporal forensic analysis for inland-water ship collisions using AIS data. *Saf Sci* 57:187–202
- Xia H, Huang W, Li N, Zhou J, Zhang D (2019) PARSUC: A parallel subsampling-based method for clustering remote sensing big data. *Sensors* 19(15):3438
- Zhang W, Feng X, Goerlandt F, Liu Q (2020) Towards a convolutional neural network model for classifying regional ship collision risk levels for waterway risk analysis. *Reliab Eng Syst Saf* 204:107127
- Zhang L, Meng Q, Fwa TF (2019) Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. *Transp Res Part E* 129:287–304
- Zhang L, Meng Q, Xiao Z, Fu X (2018) A novel ship trajectory reconstruction approach using AIS data. *Ocean Eng* 159:165–174
- Zhao M, Yao X, Sun J, Zhang S, Bai J (2018) GIS-based simulation methodology for evaluating ship encounters probability to improve maritime traffic safety. *IEEE Trans Intell Transp Syst* 20(1):323–337