

Mining and Analysis of the Traffic Information Situation in the South China Sea Based on Satellite AIS Data

Tianyu Pu, Hmn Smart Co., Ltd., China*

ABSTRACT

The loading of Automatic Identification System equipment on low-orbiting satellites can adapt to the demand of exchanging data and information with greater “capacity” brought by the AIS data information of ships in deep waters that cannot be covered by land-based stations. The information in the satellite AIS data contains a large number of potential features of ship activities, and by selecting the ship satellite AIS data of typical months in the South China Sea in 2020. Data mining, geographic information system, and traffic flow theory are used to visualize and analyze the ship activities in the South China Sea. The study shows that the distribution of ship routes in the South China Sea is highly compatible with the recommended routes of merchant ships, and the width of the track belt is obviously characterized. The number of ships passing through the southern waters of the Taiwan Strait has increased significantly, and the focus of traffic safety in the South China Sea should also focus on major route belt and important straits.

KEYWORDS

Data Repair Fusion, Satellite AIS, Statistics Lines Analysis, Traffic Information, WLS-Hermite

INTRODUCTION

The South China Sea, as critical waters of the 21st Century Maritime Silk Road (Zhong & White, 2017), is one of the most active waters in the world at present and is also a key area of high concern for countries around the world. Many scholars have been studying the waters of the South China Sea from multiple dimensions; the safety of ship navigation is the most important among these topics (Rosenberg & Chung, 2008). In view of the limitation of data sources in the South China Sea, most of the previous studies have focused on the port waterways along the South China Sea, including Singapore and the Strait of Malacca (Weng et al., 2012), and the Pearl River Delta region of China, etc. (Sasa et al., 2021). Few of them have analyzed the overall traffic conditions of the whole South China Sea and the key waters in depth (Du et al., 2016).

DOI: 10.4018/IJDWM.332864

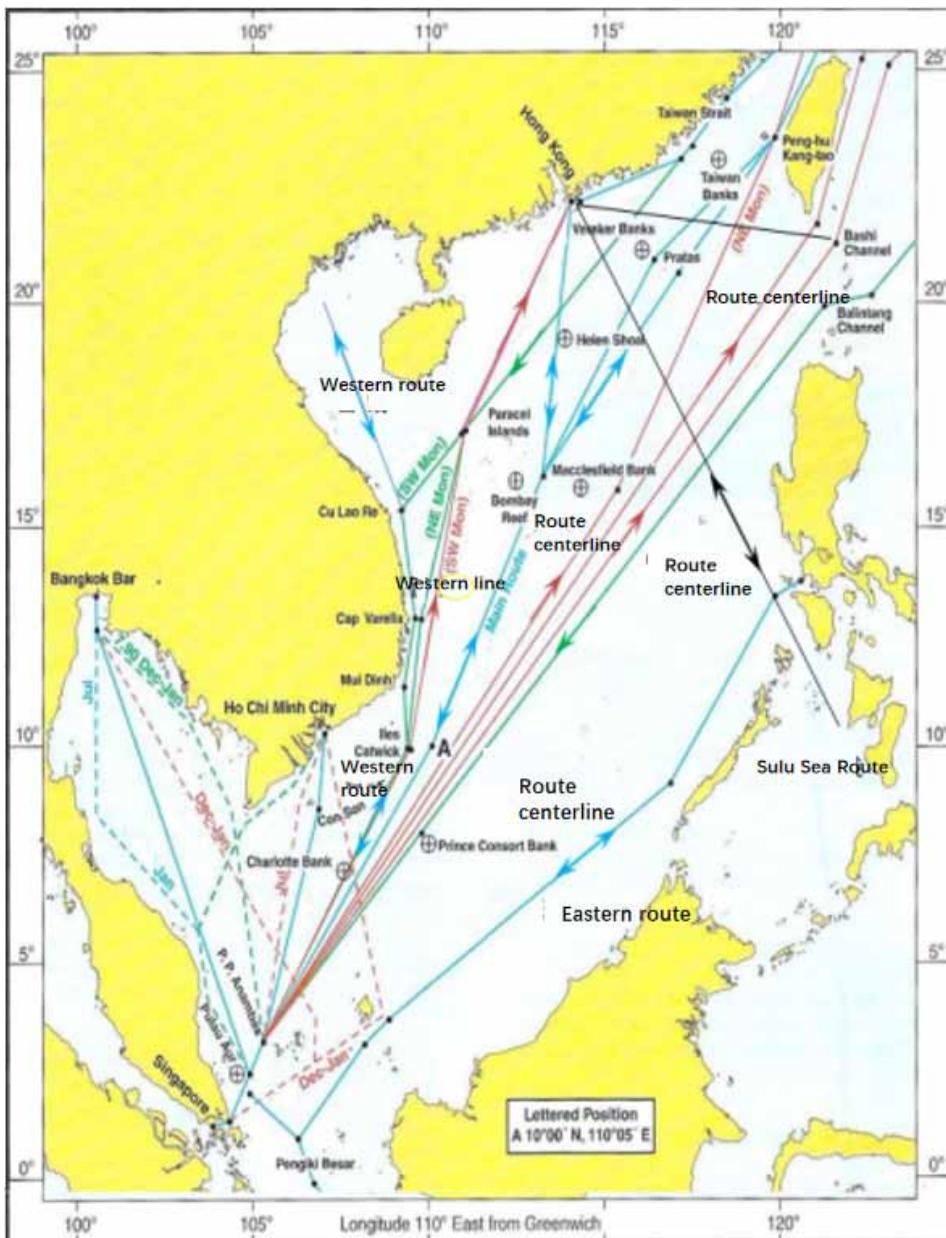
*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

As Figure 1 shows, according to the recommended routes in the authoritative book *World Ocean Routes* (Jenkins, 1973), the recommended routes in the South China Sea are mainly in the southwest and northeast directions and are generally divided into the eastern, central, and western routes, of which the central route is the main route and the two directions of the southern Taiwan Strait and the Bashi Strait through which the route passes are the key waters.

At present, the installation of AIS data transceiver equipment on low-orbiting satellites has served the modern demand for larger “capacity” data exchange (Greidanus et al., 2016). The mining of ship AIS data to build a maritime navigation characteristics map is one of the important means to track

Figure 1. Recommended routes for major merchant ships in South China Sea



the global ship behavior characteristics (Vespe et al., 2015) and regulate the ship order (Yliskylä-Peuralahti & Gritsenko, 2014). By analyzing and visualizing ship movement patterns, traffic density, route preferences, and other relevant navigation information to provide a comprehensive depiction of maritime navigation characteristics in a specific area or region.

There is little research on the spatiotemporal characteristics, key areas, and port spatial information of water transportation in the South China Sea by domestic and foreign scholars. Based on AIS data, using computer technologies and methods such as time-space division, linear density analysis, cluster analysis, and complex network, the paper explores the time-space characteristics of water traffic in the South China Sea waters, identifies the key areas of the South China Sea waters, divides the levels of ports, and provides decision-making reference for the water regulatory authorities to optimize the channel and formulate port management policies.

The technology roadmap of this paper is as follows. Firstly, this article extracts trajectory points from the AIS database and uses Python for voyage recognition, achieving the preprocessing process from trajectory points to trajectory lines. Secondly, the time-space division result map and linear density analysis map are obtained by using the time-space division of route set, linear density analysis, time-space statistical analysis, and other methods to explore the time-space distribution characteristics of water traffic in the South China Sea waters. Then, based on hierarchical clustering, the ship berthing points are clustered to identify key areas. Finally, this paper uses the complex network method to build the port shipping network, analyze the centrality of the port network, divide the port hierarchy, obtain the centrality analysis diagram and port hierarchy result diagram, and mine the characteristics of port spatial information.

METHODS RESEARCH THEORY AND METHODOLOGY

Research Technology

The QGIS (Quantum Geographic Information System) is a powerful open-source geospatial analysis software. It enables users to analyze and interpret geographic data with a user-friendly interface. Its wide range of tools supports tasks like spatial querying, overlay analysis, and spatial statistics. QGIS promotes interoperability and collaboration within the geospatial community. Overall, it is a valuable tool for effective geospatial analysis. QGIS compares favourably to proprietary geospatial analysis software in terms of functionality and usability. It offers a wide range of features and capabilities while being user-friendly. Key features of QGIS that make it a powerful tool for geospatial analysis include its ability to handle various data formats and support for advanced geoprocessing and analysis, robust cartographic capabilities, extensive plug-in ecosystem, multi-platform compatibility, and strong community support. By compiling and analyzing the parameter fields of Maritime Mobile Service Identify (MMSI), time, longitude, latitude, speed, course of the South China Sea ships, to obtain the spatial distribution and temporal characteristics, and then reveal the spatial and temporal distribution patterns of traffic information in the waters of the South China Sea. MMSI data, which stands for Maritime Mobile Service Identity, is compiled and analyzed by collecting AIS (Automatic Identification System) signals emitted by ships. These signals contain MMSI information, which is then processed and analyzed using geospatial analysis techniques to determine ship locations, movements, and patterns over time. Spatial distribution refers to how a phenomenon or variable is arranged across a geographic area, focusing on its location, density, and concentration. It involves analyzing the spread, clustering, or dispersion of the phenomenon in different spatial units. Temporal distribution, on the other hand, refers to how the phenomenon changes or varies over time, examining the frequency, timing, duration, and magnitude of its occurrence or changes. Spatial distribution relates to the arrangement in space, while temporal distribution relates to the pattern over time.

Data Pre-Processing

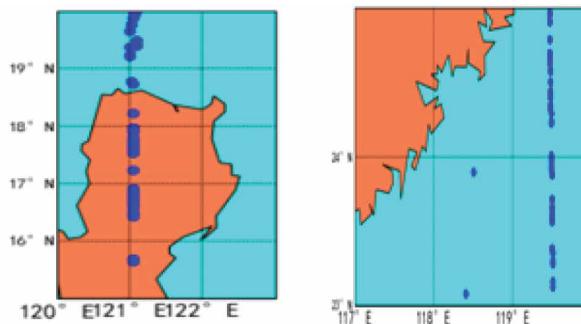
Data Cleaning

Data cleaning (Gunnarsson, 2021) is the first step in the data processing process, and its purpose is mainly to find abnormal information and duplicate information in the data.

Preliminary cleaning of duplicate data. For the satellite AIS data in the South China Sea where the ship's latitude and longitude have not changed for a long time, the satellite AIS data are judged to be at anchor or berthing and only one data point of the same latitude and longitude is retained, while the rest of the data are deleted. Identifying and retaining only one data point for ships at anchor or berthing with the same latitude and longitude is necessary to prevent data duplication and maintain accurate spatial analysis. The removal of redundant data for ships at anchor or berthing improves the analysis of satellite AIS data in the South China Sea by providing a more accurate representation of ship movements and patterns, eliminating data duplication, and reducing noise in the analysis. A ship's latitude and longitude may remain unchanged for an extended period in the South China Sea due to various reasons. It could be anchored or moored, engaged in fishing or aquaculture activities, conducting research or surveys, involved in monitoring or surveillance operations, or experiencing technical or operational issues. Each case is unique, and factors such as the ship's type, purpose, and prevailing conditions can contribute to its stationary position.

Screening of outliers. Calculate the residual error, the arithmetic mean and its standard deviation σ for the ship's latitude and longitude, respectively, and remove data with errors greater than 3σ according to the 3σ rule (Karagiannidis & Themelis, 2021). Illogical ship anomalies in raw satellite AIS data can occur due to various reasons. These anomalies may result from data transmission errors, sensor malfunctions, intentional manipulation, environmental factors, or technical limitations. To address these issues, implementing data quality control measures and collaborating with maritime authorities and vessel operators can help identify and rectify the causes of these anomalies, ensuring accurate and reliable AIS data. There are illogical ship anomalies in the raw satellite AIS data due to AIS malfunction and sensor signal misalignment, which mainly include AIS data point drift, isolated points on the ship's track line, and other anomalies. Sensor signal misalignment can contribute to illogical ship anomalies in satellite AIS data by causing positional inaccuracies or inconsistencies, leading to false or misleading ship observations and data anomalies. AIS malfunctions occur when illogical ship anomalies appear in raw satellite AIS data. These malfunctions can be caused by technical issues, signal interference, intentional interference, human error, or system vulnerabilities. To address these malfunctions, investigate the root cause, enhance system security, provide proper training, and conduct regular maintenance. Continuous monitoring and analysis of AIS data are crucial for detecting and resolving inconsistencies, as seen in Figure 2a for whole segment drift and Figure 2b for isolated points.

Figure 2. Schematic diagram of noise types of ship track data in South China Sea waters



Note: a) Whole Section Drift and b) Isolated Points.

Due to excessive deviation of data outliers from actual values, in this paper the authors use the method of ship dynamic information change regulation for data outlier screening (Negenborn et al., 2021).

The ship dynamic information change supervision method involves monitoring changes in a ship's track by comparing current and previous track points using parameters like heading angle, distance difference, and speed difference. By analyzing these differences, the method detects deviations or anomalies in the ship's trajectory. It ensures the accuracy and reliability of ship tracking data, which is essential for navigation and safety purposes. The ship dynamic information change supervision method is used to judge whether the current ship track point p_i and the previous track point p_{i-1} are within the setting range of standard deviation σ by the difference in heading angle ΔCog , distance difference ΔDis , and speed difference ΔV . If a single-track point is out of range, the point p_i is considered to have drifted and the data repair algorithm is used to repair the data. The data repair algorithm is used to fix data when a single data point, " p_i ," is outside the expected range due to drift. It identifies outliers, estimates the drift, and adjusts the value of " p_i " to bring it back within the acceptable range. This ensures accurate and reliable data for analysis and decision-making.

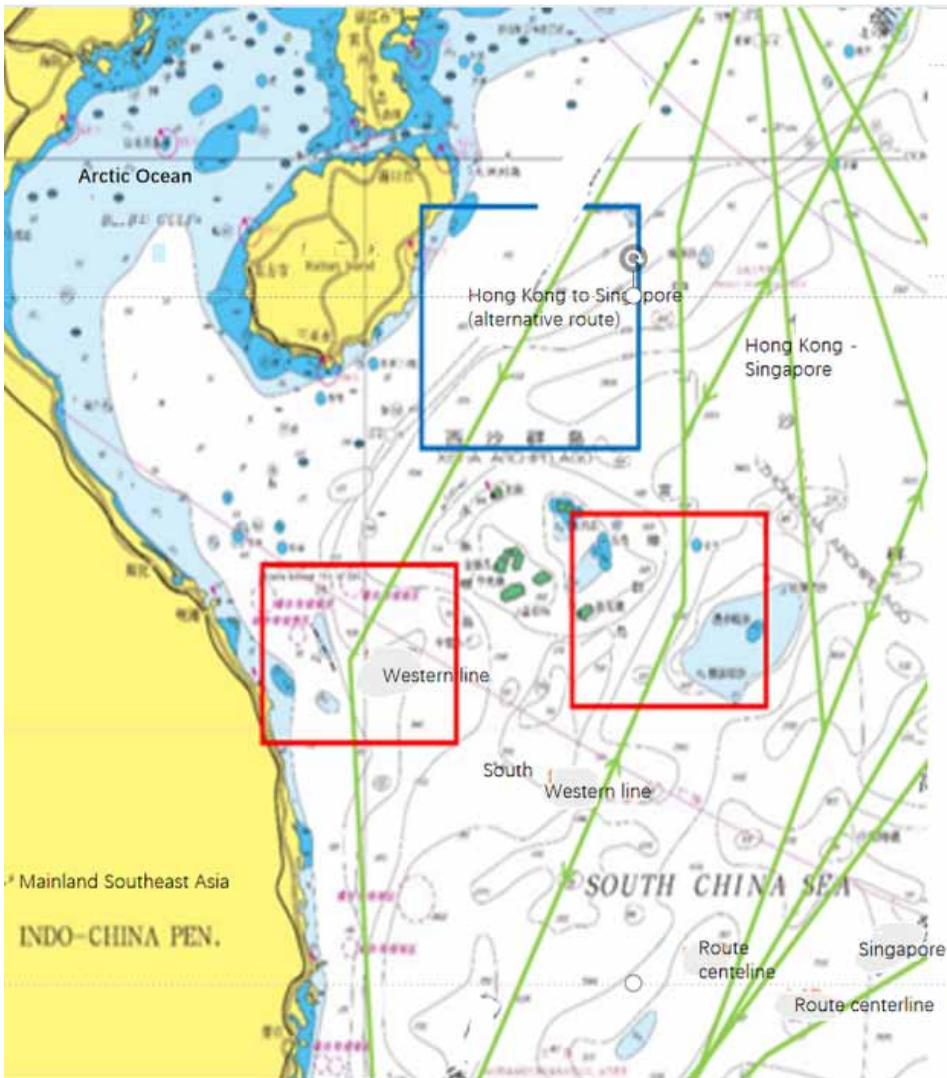
Data Restoration Fusion

Data repair fusion involves using a combination of database techniques and algorithms to repair and restore data within a database. It includes analyzing the data, detecting errors, localizing the issues, and applying appropriate repair strategies such as data transformation, imputation, integration, and restoration. The repaired data is then validated, and the database is optimized to prevent future errors. The process is documented and reported for reference and auditing purposes. Data repair fusion (Tang et al., 2021) refers to the repair and restoration of data through database and algorithmic techniques and the fusion of valid information in the database to form data with complete information. The objective is to generate a complete dataset by combining valid information from various sources in a database. This includes analyzing the data, repairing any problems such as inconsistencies or missing values, integrating the repaired data, fusing information from various sources, and validating the final dataset for quality. Data repair fusion increases data reliability, enables better decision-making, and improves data analysis capabilities.

For the main routes in the South China Sea, they were divided into straight section routes and curved section routes (near the route inflection point) for data restoration. In Figure 3 the red boxed line is the curved section, and the blue boxed line is the straight section.

The WLS-Hermite restoration algorithm is highly prominent in image processing and computer vision. It effectively enhances degraded images by combining WLS and Hermite interpolation methods. The algorithm's key features include adaptability to different types of degradation, preservation of edges and textures, parameter control, and computational efficiency. Its prominence is determined by its ability to improve image quality and its versatility in various applications. In the joint WLS-Hermite restoration algorithm, the weighted least squares filtering algorithm (WLS) provides a better fit to the straight line (Wang, 2019). The weighted least squares (WLS) filtering algorithm improves the fit to a straight line by assigning weights to data points based on their reliability. It minimizes the weighted residuals, prioritizing more reliable points and reducing the impact of outliers. The algorithm iteratively adjusts the line parameters, optimizing the fit to the weighted data points. Overall, WLS considers uncertainties, balances influences, and produces a line fit that better captures the underlying trend. The principle of the algorithm is to assume that a set of data is sampled from a particular row and that there is a certain error. In order to estimate a straight line that passes through the data points and satisfies the weighted least squares criterion, it is required that this line passes through the center of the data point set N to find the global optimal solution. The process of determining the weights for each square in the weighted least squares criterion involves assigning weights based on the inverse of the estimated variances or uncertainties associated with each data point.

Figure 3. South China Sea waters route division



The segmented cubic Hermite interpolation algorithm is used for data repair to achieve a smooth curve for the navigation track line when the ship is sailing at the turn point of the route (Li et al., 2017). The algorithm handles turn points in the ship's navigation track line by detecting significant changes in direction or curvature. It applies appropriate interpolation techniques, such as segmented cubic Hermite interpolation, to smooth out the track line and ensure continuity. Segmented cubic Hermite interpolation, also referred to as piecewise cubic Hermite interpolation, is an interpolation method that approximates a smooth curve or function between given data points. It achieves this by constructing segments of cubic polynomials. This approach offers flexibility and continuity in interpolation, allowing for the incorporation of both data points and derivative information. By dividing the data range into segments and constructing cubic polynomials for each segment, segmented cubic Hermite interpolation provides a versatile and seamless interpolation technique. In the segmented cubic Hermite interpolation algorithm, the key steps are as follows. Firstly, consecutive points before and after each turn point are identified. Then, tangent vectors are calculated at each point to capture

the track direction. Control points for each segment are determined using neighboring tangent vectors. Next, the cubic Hermite interpolation formula is applied to interpolate points between control points, ensuring smoothness. The segmented cubic Hermite interpolation algorithm has the feature that the values of the functions of the corresponding nodes are equal and the values of the corresponding derivatives and their higher order inverse are also equal, so that there are tangents at the corresponding turning points to ensure the smoothness of the curve and achieve the feature that the interpolated function and the interpolated function are in good agreement (Zhou et al., 2020). By utilizing tangent vectors at each point. These tangent vectors capture the direction and rate of change of the function, enabling the algorithm to interpolate points in a way that preserves both the function values and their derivatives at the corresponding nodes. WLS filtering reduces noise, while Hermite interpolation establishes smooth curves from data points. Hermite interpolation is accurate and requires derivative information. WLS filtering assigns weights to reduce noise. WLS filtering is used in signal processing and time series analysis, whereas Hermite interpolation is used in graphics and function approximation. A decision is made based on the desired level of smoothness and noise. Hermite interpolation utilizes derivative information to improve accuracy by allowing for a more precise approximation of the underlying function. By incorporating derivative values at interpolation points, Hermite interpolation captures the local behavior of the function more effectively, resulting in a more accurate representation.

Water Grid Methodology

In satellite AIS data, the position information of ships is discrete points in space, and in fixed waters, to obtain the distribution of ships' navigation trajectories and to analyze the ships' navigation characteristics, a gridding method (Yan et al., 2020) can be adopted for fixed waters, so that the traffic characteristics of ships sailing in those waters can be extracted precisely. The gridding method is crucial for dividing continuous data into a grid structure, enabling efficient analysis and visualization in fields like computer graphics and data analysis. It facilitates tasks such as interpolation, reconstruction, and numerical computations while aiding in data visualization, compression, and storage. Advantages of using Hermite interpolation over other methods include higher accuracy due to the incorporation of derivative information, flexibility in handling irregularly spaced data points, smoothness in maintaining continuity, capturing the local behavior of the function accurately, and enabling both interpolation and extrapolation beyond the given data range.

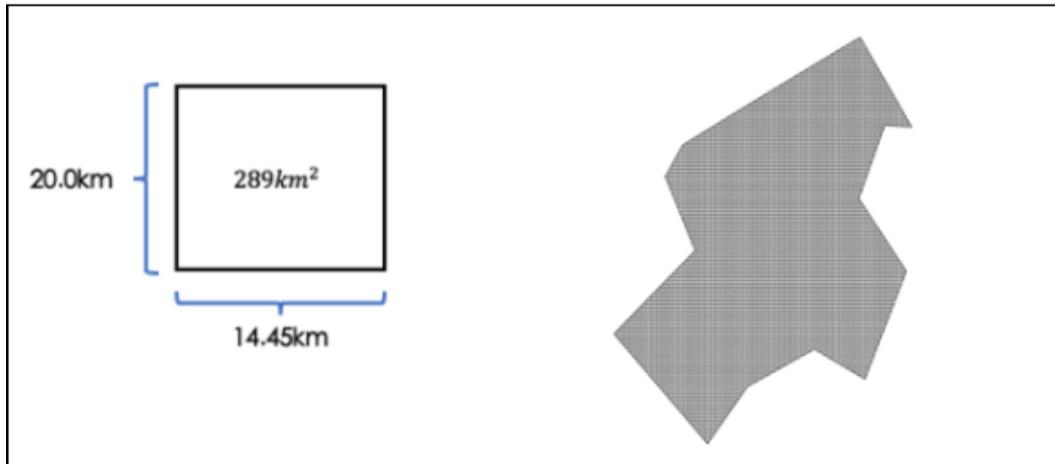
The Hermite interpolation algorithm is chosen for data repair in navigation track lines because it can produce a smooth curve. It has a high level of accuracy, ensuring that the original data points are precisely approximated. It effectively captures local variations by considering derivative information. As a result, the curves are less distracting and continuous, with fewer abrupt changes. Furthermore, Hermite interpolation allows for curvature control and is noise resistant, which improves the accuracy and reliability of the repaired track lines. Assume that scatters are used to represent the track point location information broadcast at intervals from the ship's AIS data. Then, according to the demand for three-dimensional management and service capability guarantee of maritime safety in the middle and long-range waters of the South China Sea, the entire study waters are gridded by using the gridding method based on the spatial location distribution of ships in longitude and latitude, with the grid granularity and area of 289km², as seen in Figure 4.

Data Analysis and Applications

Based on the well-processed static AIS data source and ship track data source in the South China Sea, the static and dynamic characteristics of ships in the South China Sea are analyzed from the spatial and temporal dimensions through data statistics, data extraction thinning, QGIS visualization, and water gridding methods.

Vessel static characteristics analysis and vessel dynamic characteristics analysis in the South China Sea aim to assess and understand vessel behavior. Static analysis examines physical attributes,

Figure 4. Grid-based data processing in South China Sea waters



while dynamic analysis focuses on performance in motion. These analyses inform decision-making for maritime safety, operations, and environmental management in the region. Due to reasons such as anchoring, berthing at a port, waiting for clearance, conducting maintenance activities, adverse weather conditions, or operational requirements.

Vessel static characteristics analysis. Seven ship observation lines are set up in the South China Sea waters, and the statistical analysis of ship length, ship speed, and ship draught are carried out through the static characteristics of ships in the observation line waters and the environmental characteristics of the South China Sea waters in the data source.

Vessel dynamic characteristics analysis. Using QGIS visualization technology and traffic flow analysis, the influence of navigable waters is comprehensively considered, and in-depth analysis of traffic conditions such as the distribution of navigation zones in the waters of the South China Sea is carried out using the setting of observation gate lines.

SATELLITE AIS VESSEL TRAFFIC FLOW OBSERVATION AND ANALYSIS IN THE SOUTH CHINA SEA

Traffic Information Condition Observation Gate Line Setting and Data Source Selection

Observation Gate Line Setting

The South China Sea are bounded by the Taiwan Strait and the Bashi Strait in the north and the Mindoro Strait and the Barabak Strait in the middle, and the southern part goes straight to Singapore, Indonesia, and Thailand. Therefore, based on historical routes and management and service needs, seven gate lines are set up clockwise from south to north, as shown in Figure 5 and Table 1.

Data Source

Due to the obvious monsoon climate characteristics in the South China Sea, the monsoon is highly correlated with the ship route traffic, and the ship activities in the navigation area show a regularity from year to year. Possible solutions for addressing ship route traffic in monsoon-affected regions include enhanced weather monitoring and forecasting, optimized route planning, improved vessel design and technology, diversification of ports and transshipment hubs, collaborative efforts and information sharing, seasonal planning and adjustments, and crew training. During the monsoon

Figure 5. South China Sea ship traffic condition observation gate line

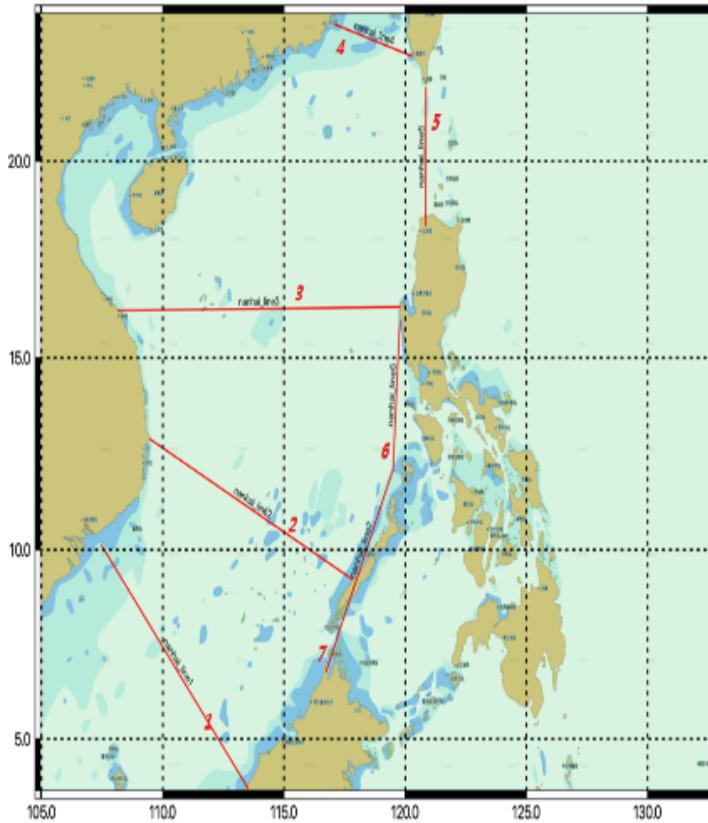


Table 1. Flow observation gate line basic information

Observation Gate Line Number	Observe the latitude and longitude of the goal line	Note
Nanhai_line1	03° 41' 57.8"N,113° 30' 17.0"E 10° 36' 29.7"N,107° 28' 56.8"E	By way of Nansha near Zengmuansha
Nanhai_line2	09° 13' 16.7"N,117° 49' 31.5"E 12° 53' 50.3"N,109° 28' 01.8"E	By way of Nansha near Taiping Island
Nanhai_line3	16° 18' 17.4"N,119° 46' 51.7"E 16° 12' 52.7"N,108° 09' 55.1"E	Passage near Xisha and Zhongsha Islands
Nanhai_line4	23° 23' 18.5"N,117° 06' 27.7"E 22° 37' 05.6"N,120° 15' 23.5"E	Near the southern waters of the Taiwan Strait
Nanhai_line5	18° 39' 58.9"N,120° 50' 55.2"E 21° 49' 44.9"N,120° 51' 44.2"E	By way of Bashi Strait
Nanhai_line6	15° 56' 19.1"N,119° 46' 42.2"E 12° 06' 03.7"N,119° 29' 54.9"E	Near Huangyan Reef and Mindoro Strait
Nanhai_line7	12° 06' 03.7"N,119° 29' 54.9"E 06° 46' 46.6"N,116° 43' 33.7"E	Near the Barabakh Strait

season, these solutions aim to minimize disruptions and ensure safer and more efficient maritime transportation. In view of the large volume and repetition of data sources throughout the year, the representative southwest wind prevailing season in July and the northeast monsoon prevailing period in December are selected to procure satellite AIS ship data sources for data mining.

Disposition of Vessel Density and Vessel Traffic

Vessel density (Liu et al., 2020) refers to the number of vessels in a certain area of water at a certain moment, which directly reflects the spatial distribution of vessels in a certain area of water and can be used to measure the degree of vessel traffic congestion and danger in the water to some extent. Vessel traffic density is measured by the number of vessels passing through each grid or unit area. AIS, radar, and satellite imagery are used to calculate vessel density. Each grid counts vessels passing through the area of interest. Vessel traffic density (vessels per unit area) is calculated by dividing the vessel count by the grid area. This data helps to identify congested areas and collision risks and optimize maritime operations for navigation, safety, and planning. To compute the density of vessel traffic (vessels per unit area), the total count of observed or recorded vessels within a specific area is divided by the corresponding area size of the grid or spatial unit used for analysis. This calculation results in a density value that represents the average number of vessels present within each unit of area. The calculation of vessel traffic density value is measured by the number of vessels passing through each grid, as seen in Figure 4, in 1h, and the calculation formula is:

$$\rho = Q/h \tag{1}$$

In Eq. (1) ρ is the ship density (ships/h), Q is the total number of ships passing through a certain area, and h is the time. A larger ρ indicates more ships passing through an area per unit time and more traffic congestion.

Vessel traffic volume (Zhang & Crabbe, 2021) is the number of all vessels passing through a specific location in the water at a certain time interval, and its size can characterize the busy traffic of the corresponding waterway to some extent; the larger the vessel traffic volume index is, the larger and busier the traffic flow of the corresponding waterway in the water is. The statistical model of traffic volume is introduced here as:

$$\bar{F} = \sum_{i=1}^n F_i \tag{2}$$

In Eq. (2) \bar{F} indicates the volume of ship traffic in a certain time period, F_i indicates the volume of traffic at a certain moment i , and n indicates the time. For the convenience of statistical analysis, the volume of marine vessel traffic through the study area is calculated based on the sampling data in terms of time (moment, month, season) to measure the marine traffic in the South China Sea. To assess marine traffic in the South China Sea, it involves analyzing the frequency of vessel sightings or AIS data over specific time intervals, such as moments, months, or seasons. By aggregating and analyzing the sampled data, the assessment of marine traffic in terms of time can be determined.

Statistics and Analysis of Major Traffic Flows in the Waters of Each Gate Line

The raw data of July and December 2020 were processed and mined for analysis such as combing, cleaning and screening, restoration fusion, and data extraction thinning, and the ship traffic was exported according to the seven gate-line distribution waters in Figure 5, and the statistical information is shown in Tables 2 and 3.

Table 2. Summary table of the statistical information breakdown of the observed goal line in July

Observation Gate Line Number	Number of Vessels /vessel	Average length/m	Average ship width/m	Average draft/m	Average over-the-line speed/kn.
Nanhai_line1	8050	179.638	29.306	7.611	10.553
Nanhai_line2	7868	190.787	30.340	8.782	12.011
Nanhai_line3	9481	201.699	31.972	9.049	12.176
Nanhai_line4	20503	147.383	22.899	6.058	9.929
Nanhai_line5	3177	235.879	38.346	11.073	12.759
Nanhai_line6	3084	218.435	34.439	9.860	12.243
Nanhai_line7	389	157.431	25.469	7.918	11.849

Table 3. Summary table of the statistical information breakdown of the observed goal line in December

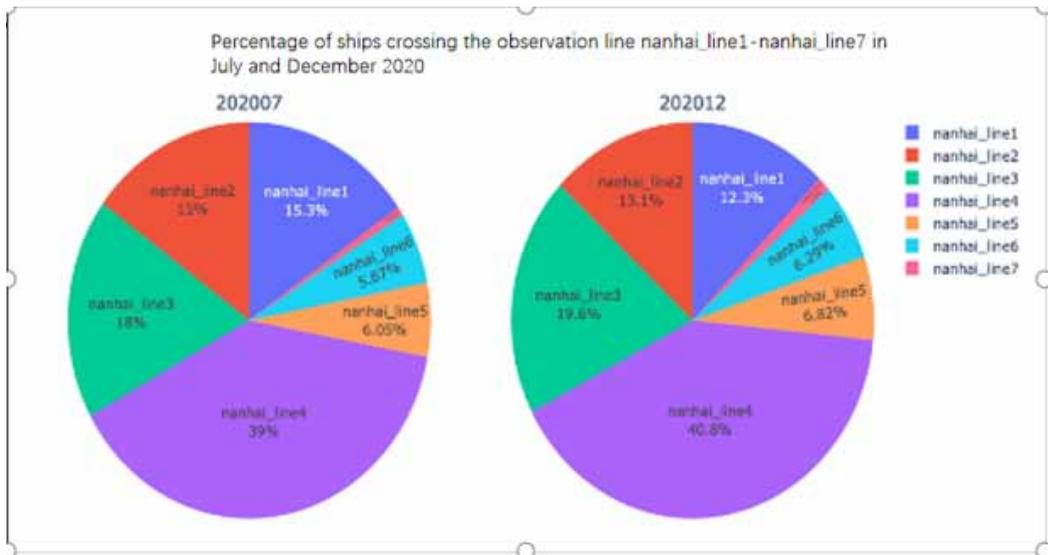
Observation Gate Line Number	Number of Vessels /vessel	Average length/m	Average ship width/m	Average draft/m	Average over-the-line speed/kn.
Nanhai_line1	5614	200.796	31.937	9.248	12.337
Nanhai_line2	5949	199.843	31.553	9.102	12.421
Nanhai_line3	8928	193.222	30.656	8.846	11.163
Nanhai_line4	18561	144.497	22.440	6.520	9.301
Nanhai_line5	3106	236.224	38.210	11.188	11.362
Nanhai_line6	2866	206.829	32.819	9.609	11.883
Nanhai_line7	524	169.461	27.058	8.125	11.702

The number of ships passing through each gate line in different months in the table is represented by pie charts, and the data are shown in Figure 6. It can be seen from Figure 6 and Tables 2 and 3 that the number of ships passing through gate line 1 and gate line 2 are basically similar, and the average length, width, draught, and speed are relatively large, so it is judged that the ships passing through these two places are mainly originally ocean-going ships sailing along the main channel, and some fishing vessels are active in the waters, which can pull down the average scale of vessels.

There is a significant increase in vessel traffic at gate line 3 compared to gate line 2. According to the scatters traffic flow maps in Figure 6, Figure 11, and Figure 12, it can be seen that there is a significant increase in vessel traffic along the coast of Vietnam at gate line 3, and thus the frequency of vessel activity along the coast of Vietnam can be characterized.

The number of ships passing through gate line 4 near the southern waters of the Taiwan Strait is significantly higher than other gate lines, while the scatters traffic flow map shows that there are more ships off the coast of Guangdong and Hainan, China. The average ship size is relatively small because the ship traffic at route gate line 4 is not only ocean-going ships passing through gate line 3 but also a large number of coastal transport ships and fishing vessels sailing through Guangdong and Hainan. The average ship length, width, and draft of the ships near the Bus Strait in gate line 5 are the largest among the seven gate lines, and it can be judged that the ships sailing through here are mostly large transiting ships crossing the South China Sea into the Pacific Ocean. The data show that the gate line 4 and 5 waters are the key waters for the safety control and service of ship navigation in the South China Sea.

Figure 6. The percentage of the number of vessels crossing the observation line 1 to 7 in July and December 2020

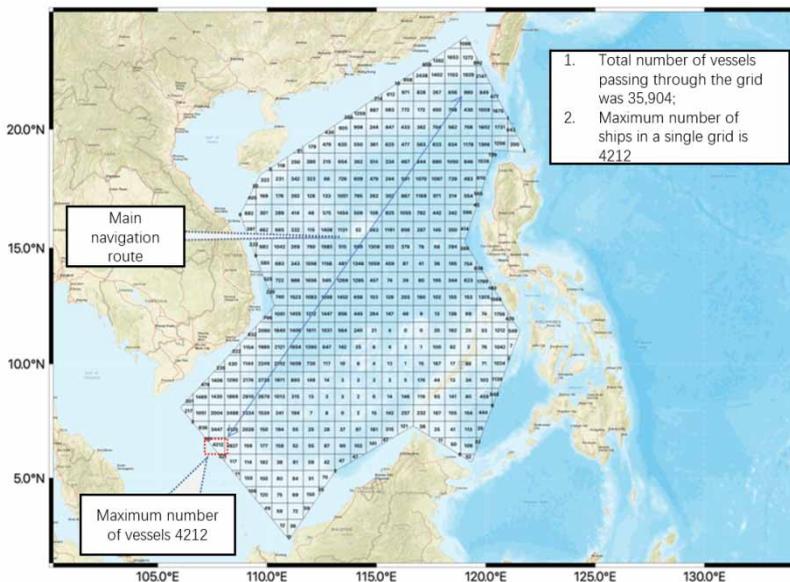


The average scale of ships near Huangyan Island Reef and Minduluo Strait by gate line 6 is relatively large, second only to the average scale of ships by gate line 5. At gate line 7 is the water via Barak Strait, whose number of ships is obviously reduced, and the average draft is smaller, but the speed of the ships passing through is larger.

Analysis of Vessel Traffic Flow Density Data in South China Sea

The distribution of ship density during the southwest monsoon season in July 2020 in the South China Sea is shown in Figure 7 after grid-based analysis. Statistics show that the number of ships in the main

Figure 7. Distribution of the number of vessels in the grid in July (Vessels/Month)

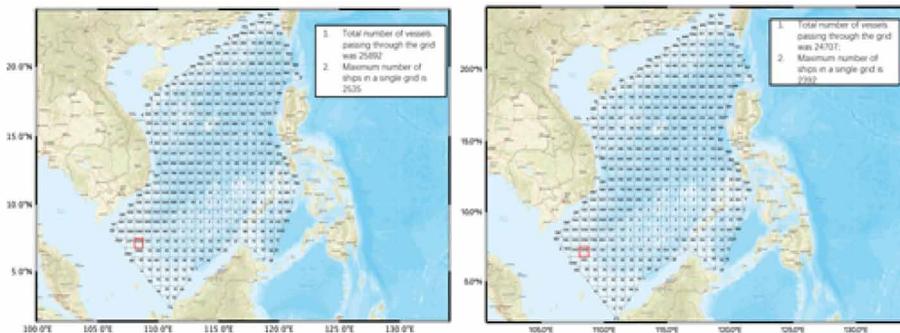


waters of the South China Sea in July (excluding waters within 30 nautical miles off the coast) reached 35,904, with a higher ship density in the main channel belt, and the southwest piece of waters at the southern end of the Spratly Islands in the south is more dense than the northern waters, where the largest density of the grid ship flow reached 4,212, located at the southwest end of the recommended route in the southern part of the South China Sea. The main channel midline waters in general have high ship density and wide voyage paths, which are the key waters for maritime safety control and service. These include the convergence of major shipping routes, substantial trade and economic activities in the area, navigational safety considerations prompting wider voyage paths, access to multiple ports, and the regional significance of the main channel as a vital transportation route.

According to the statistics of vessel heading (northward and southward), the ship flow in the grid with the highest density of vessels sailing northward reached 2,535, as seen in Figure 8a, and the highest density of vessels sailing southward reached 2,392, as seen in Figure 8b, whose location is still the southwest waters of the main channel in the southern part of the South China Sea, and the overall are consistent with the overall density distribution in July.

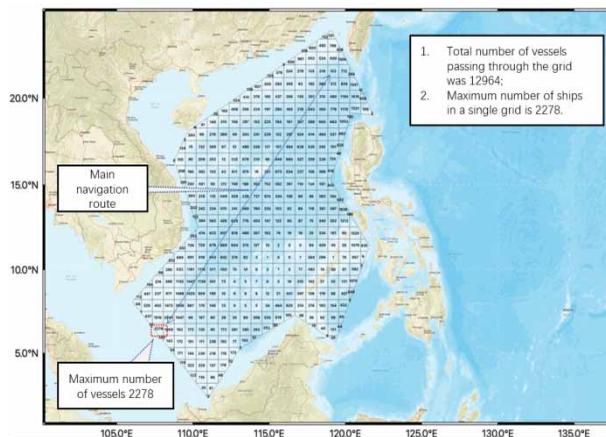
In December, under the influence of the northeast monsoon, there was a significant decrease in vessel density compared to July, with the highest density grid vessel traffic of 2,278, as seen in Figure 9, still located at the southwest end of the recommended route in the southern South China Sea.

Figure 8. Distribution of the number of vessels in the grid of northbound/southbound traffic flow in July (Vessels/Month)



Note: a) Northbound and b) Southbound.

Figure 9. Distribution of the number of vessels in the grid in December (Vessels/Month)



According to the statistics of vessel heading (northbound and southbound), the vessel traffic in the grid with the highest density of vessels sailing northbound is 1115, as seen in Figure 10a, and the vessel traffic in the grid with the highest density of vessels sailing southbound is 1207, as seen in Figure 10b.

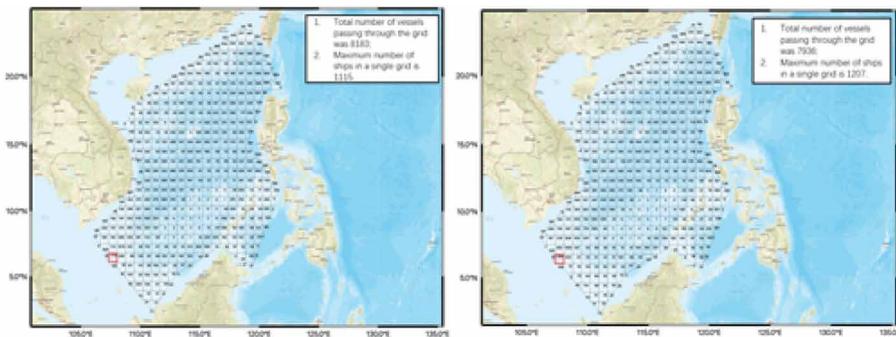
CHARACTERISTICS OF SHIP TRAFFIC FLOW IN SOUTH CHINA SEA

Ship Dynamic Characteristics in South China Sea

Vessel Track Belt Distribution

As shown in Figure 11 and 12, the track points in the grid are represented in the form of flow feature maps to obtain the ship satellite AIS data track distribution maps for July and December 2020. Since

Figure 10. Distribution of the number of vessels in the grid of northbound/southbound traffic flow in December (Vessels/Month)



Note: a) Northbound and b) Southbound.

Figure 11. July South China Sea ship navigation traffic flow track distribution

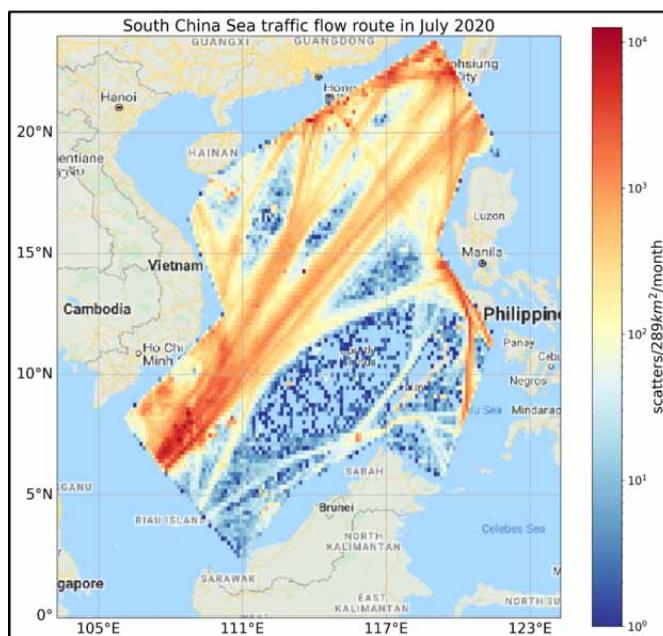
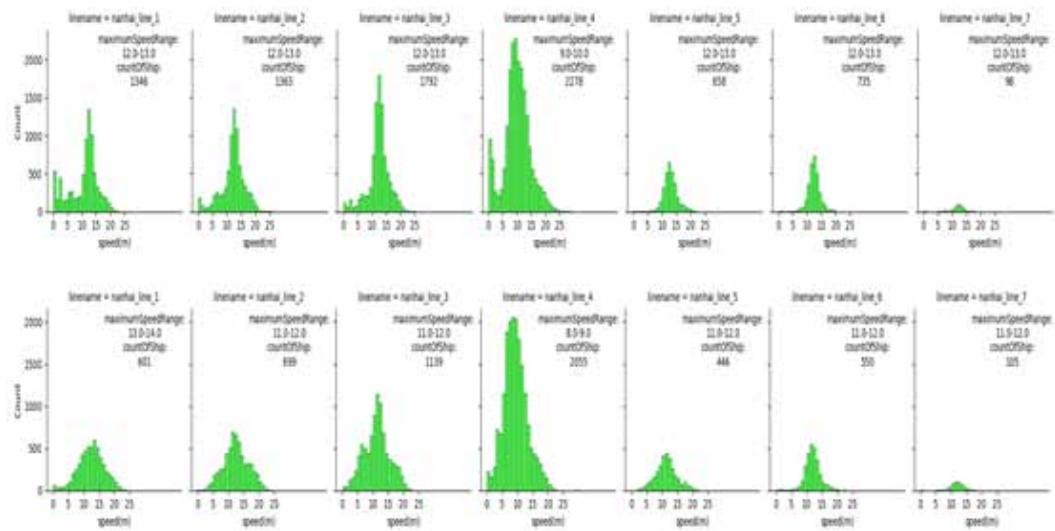


Figure 13. Speed distribution of boats at each gate line in July 2020 and speed distribution of boats at each gate line in December 2020



vessel speed range with the highest number of vessels at gate line 4 is 9kn to 10kn, with the number of vessels at 2278. And the ship speed range with the largest number of ships at gate line 5, 6, and 7 is 12kn to 13kn.

As shown in the December statistics shown in Figure 13b, the ship speed range with the highest number of ships at gate line 1 is all 13kn to 14kn. The distribution of ship speed at gate lines 2 and 3 is more similar, and the ship speed range with the highest number of ships is all 11kn to 12kn. The ship speed range with the highest number of ships at gate line 4 is 8kn to 9kn, with the number of 2,055 ship trips. The speed range with the highest number of ships at gate line 5, 6, and 7 is also 11kn to 12kn.

After comparison, it can be seen that the overall sailing speed of ships in the South China Sea in December is lower than the ship speed in July, mainly because the ships are affected by frequent windy weather in winter in December; the overall speed distribution of ships passing through the main track with each gate line water is stable.

Ship Static Characteristics in South China Sea

The static characteristics of the gate-line vessels in the South China Sea waters in July as well as December 2020 are analyzed from two aspects of vessel draught and vessel length, which are shown in Figure 14 and Figure 15.

Comparing the distribution of ship draught in July and December, the main body of the draught distribution of gate line 1 in the South China Sea waters in July is 0m~1.0m, which is dominated by fishing vessels operating nearby, and the main body of ship draught in December is 9m~10.0m. The draught of ships in gate line 2 and 3 is mainly 8.0m~9.0m. The distribution of 0m~1m draught of ships in gate line 4 is particularly prominent, and it can be seen that there are a large number of fishing vessels operating at the southern exit of the Taiwan Strait in addition to merchant vessels. The draught of ships in gate line 5 and 6 is concentrated in 8.0m~10.0m, and the flow rate in July and December is basically close to each other, which indicates that the ships active in this water are basically mainly medium and large ships.

Figure 14. Draught distribution at each goal line in July 2020 and draught distribution at each goal line in December 2020

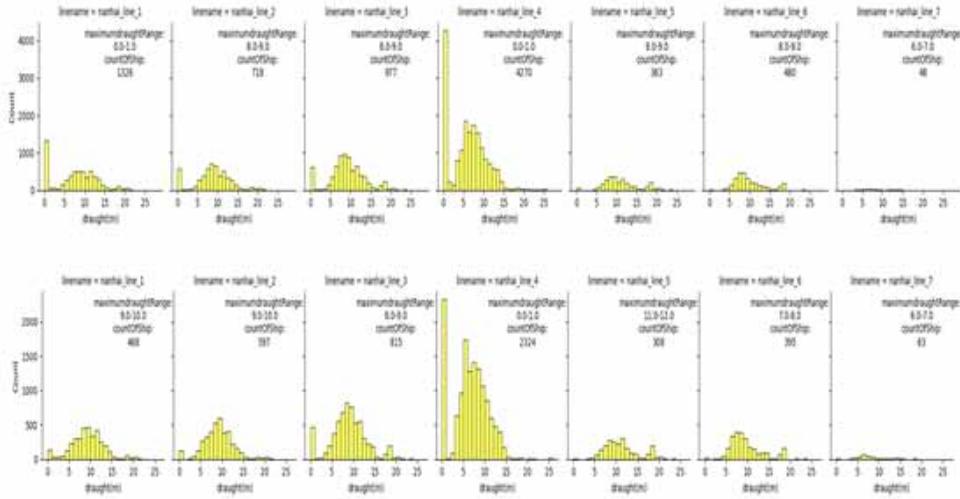
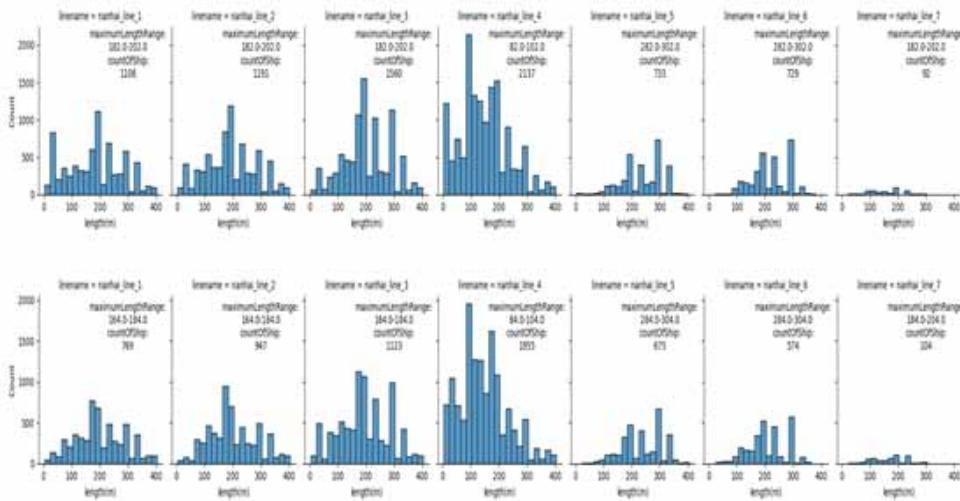


Figure 15. Length distribution of captains at each goal line in July 2020 and captain length distribution at each goal line in December 2020



A large number of fishing vessels were found in the statistics of crossing vessels in gate line 1 as well as gate line 4. According to the satellite AIS data, the main distribution of fishing vessels is located in the above-mentioned fishing grounds in the South China Sea as well as in the fishing grounds in the Nansha Islands, which are basically based on regional activities (Li et al., 2021). The intensity of fishing vessel activities shows block or strip distribution within 100 km of the coastline, and the number of fishing vessel points reflects more concentrated offshore fishing activities with a certain periodicity (Li et al., 2021). Therefore, the activities of fishing vessels have less influence

on the ship traffic in the main waters of the South China Sea and do not affect the analysis of the above-mentioned ship traffic situation (Yan et al., 2022).

According to Figure 15, the distribution of ship lengths at gate line 1, 2, and 3 is more similar in July and December, with ship lengths ranging from 182.0m to 202.0m, with large ships passing through the waters at gate line 3, with ship lengths ranging from 282.0m to 302.0m; the ship lengths at gate line 4 for the two months are mainly 82.0m to 104.0m, which coincides with the distribution of ship drafts. The ship length at gate line 5 and 6 is mainly 282.0m~304.0m, mainly through medium and large ships.

In summary, analyzed from the perspective of seasonal influence, the whole number of ships in December is less than that in July, with a large difference of 1,942 ships in gate line 4. The difference in the number of ships is highly concentrated in small and medium-sized ships, mainly fishing boats and small surface boats near the coast and islands affected by the stronger northeast monsoon in winter (Gao et al., 2021). However, the number of ships in gate line 5 (Bashi Strait waters) is basically not affected by the season (mainly large ships voyage). The general trend of the state distribution of each gate line in 2 months is basically the same, indicating that the flow of ships sailing the South China Sea is relatively stable, and there is no obvious change of ship type due to the change of season (Zhang et al., 2022).

From the comparison of the ship's length and draft scale, the number of ships corresponding to the length and draft at the seven gate lines are different but basically the same on the midline through the main channel, and the difference between inshore and island waters is obvious, which further verifies the validity of the data (Wang et al., 2021). Therefore, ships in the South China Sea are mainly free to choose their routes according to their tonnage/power, monsoon sea conditions, and other weather elements.

SPATIOTEMPORAL DATA MINING AND VISUALIZATION OF SHIP TRAJECTORIES

Recognition of Key Areas Based on Hierarchical Clustering

This section first extracts the ship berthing points in the water area and then uses the hierarchical clustering algorithm to cluster the ship berthing points, eliminate the clusters with less clustering sample point data, and analyze the spatial distribution of the remaining clusters, so as to mine the key areas of the distribution of the ship berthing points in the water area.

The paper extracted data from 33,819 mooring points in the South China Sea waters (August 1st, 2020-2020 August 23rd, 2020) for a total of 23 days. The extraction process is shown in Figure 16.

First, cluster analysis is carried out on the berthing points based on hierarchical clustering algorithm, and the minimum threshold of the berthing points is set to 200, so as to eliminate the cluster with a small number of berthing points. Then, the clustering results are evaluated using the contour coefficient SC. The clustering results retained 33,055 anchor points, removed 764 discrete anchor points, had a contour coefficient of 0.652, and had 18 categories. Finally, visualize the clustering results in Architecture Geographic Information System (ARCGIS), and count the number of mooring points included in different categories to explore key areas in the South China Sea waters.

The number statistics of hierarchical clustering results based on mooring points in South China Sea waters are shown in Figure 17, from which the following characteristics can be seen:

1. From the perspective of the number of berthing points, there are significant differences in the distribution of berthing points in the South China Sea waters. The number of berthing points in cluster 1 is the highest, reaching 6,117. Cluster 14 has the lowest number of mooring points, only 219. The number of berthing points in clusters 1, 4, and 5 ranges from 4,000 to 6,000, making them the most densely populated area for ship berthing. The number of berths in clusters 3, 7, and

Figure 16. Extraction process of mooring points in the South China Sea waters

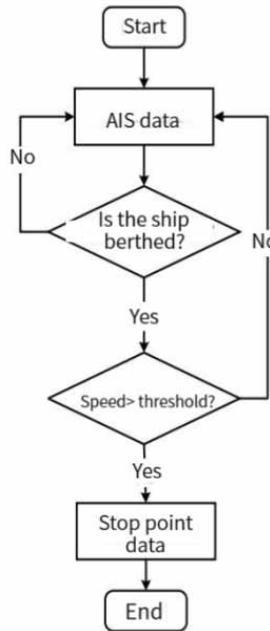
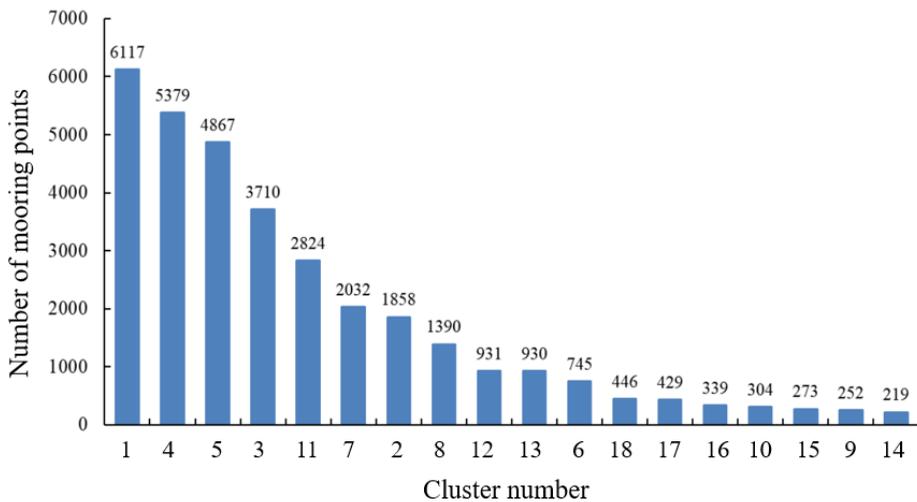


Figure 17. Number of hierarchical clustering berthing points of different categories



11 ranges from 2,000 to 4,000, making them densely populated areas. The number of berthing points for clusters 2 and 8 ranges from 1,000 to 2,000, while the number of berthing points for clusters 6, 12, and 13 is around 1,000. The number of berthing points in other clusters ranges from 200 to 500, which is a low-density area for ships to berth.

- From the spatial distribution of berthing points, it can be seen that the number of berthing points is unevenly distributed in the South China Sea waters. Clusters 1, 4, and 5 are densely populated

clusters of mooring points, mainly distributed around the main and secondary waterways. Clusters 3 and 11 are dense clusters of berthing points, mainly distributed around major ports. Clusters 2 and 7 are mainly distributed in the surrounding waters of medium-sized ports. The cluster with a relatively small number of berthing points is mainly distributed around small ports and islands.

Research has found that low-density clusters of berthing points are mainly distributed in small ports and waters near non main sea/inlet areas. High density clusters of berthing points are mainly distributed around large ports and major sea entrances/exits. The cluster with a large number of berthing points is mainly distributed in medium-sized ports, with secondary water areas near sea entrances/exits. The research shows that the major/minor inlets/outlets and medium and large ports are the key areas in the South China Sea waters.

Port Spatial Information Mining Based on Ship Trajectory

As an important relay station for the resource flow of water and land transportation, ports have transitioned from chaotic competition to hierarchical layout. Based on port shipping network analysis, it has become a research focus at this stage. Therefore, this section analyzes the shipping network of ports in the South China Sea based on complex network analysis. Firstly, determine the ports corresponding to the starting and ending points of the route. Then, the port network is constructed according to the complex network analysis method, and its relevant network parameters are analyzed. Finally, the port hierarchy is divided based on clustering analysis method.

The construction of port shipping network is based on each port as a node and the routes between two ports as edges. Ignoring the specific location, spatial scope, and edge shape of each port node, this section selects 22 ports in the South China Sea waters to build a port network for analysis. The degree of a port is the number of ports directly connected to that port. Port point strength refers to the number of all routes between the port and its directly connected ports. The port degree and point strength are direct manifestations of the port's distribution capacity throughout the entire network. The larger the degree of the port, the more direct connections it has with other ports. The greater the point strength, the more routes it has to travel to and from other ports. Network centrality analysis starts from two aspects: centrality and centrality. In network centrality analysis, there are various aspects or measures used to assess the importance or centrality of nodes within a network. Two commonly used aspects are degree centrality and betweenness centrality. Centrality represents the importance of a node in the entire network. The central potential represents the tightness of the entire network. This article determines the belonging port by determining the starting/ending points of each route and constructs the degree and point strength of the port, as shown in Figure 18.

Clustering coefficient represents the aggregation of nodes in the network. The greater the clustering coefficient, the closer the connection between the nodes is, and the closer the connection between the ports is. The clustering coefficient of each port is shown in Figure 19. A higher clustering coefficient indicates stronger connections between nodes, including the ports themselves. Figure 19 displays the clustering coefficient for each port, illustrating the extent of interconnectivity among the ports.

It can be seen from Figure 19 that the clustering coefficient of the port network in the South China Sea waters is 0.53, and the minimum value is 0.31, which is far higher than the theoretical value of the random network (0.15), indicating that the ports are closely connected.

Overall, the average path length of the port network in the South China Sea is 1.67, slightly lower than the theoretical average path length of a random network of the same size (approximately 1.93). It indicates that the interconnection between any two ports in the South China Sea requires 1-2 routes, indicating that the degree of mutual navigation between the two ports is relatively easy and the connectivity is strong.

Further hierarchical analysis is needed for each port to provide theoretical support for the development of port management systems by water related departments. In this paper, the

Figure 18. Scatter plot of port degree and point strength

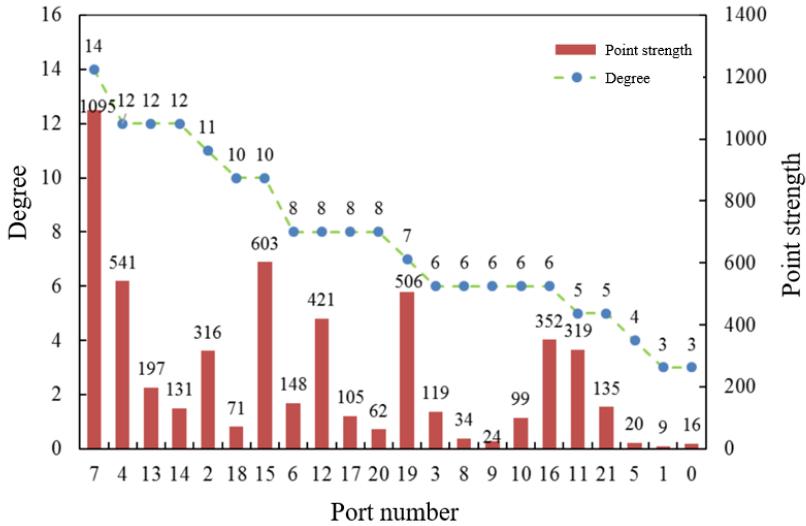
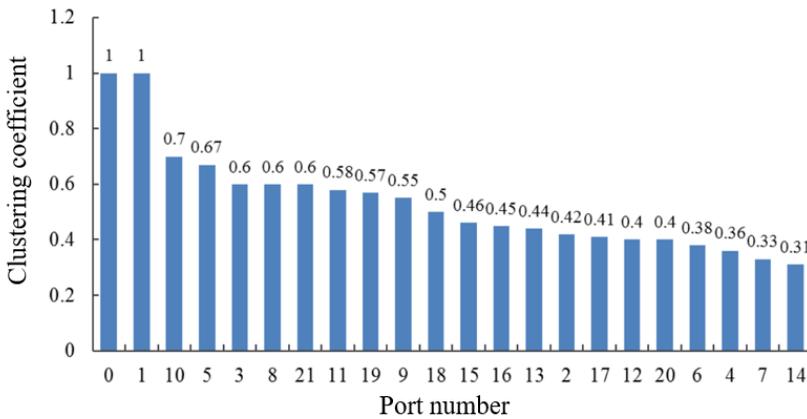


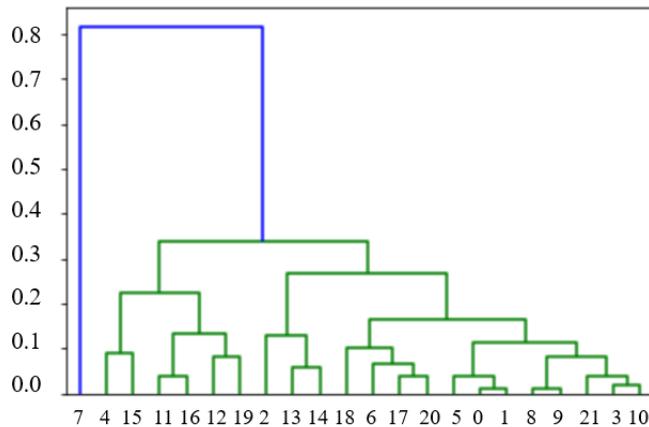
Figure 19. Distribution of port clustering coefficient



hierarchical clustering method is used to comprehensively distinguish the hierarchical structure of ports by considering the degree of direct and indirect connection between each port and other ports in combination with such indicators as port degree, point strength, and centrality. The hierarchical clustering results of ports are shown in Figure 20.

The study divides the ports in the South China Sea into five levels. The first level port degree, point strength, and centrality values are significantly higher than other levels of ports. This is attributed to the fact that these ports are the main ports of the Maritime Silk Road and important hubs of the national comprehensive transportation system. The first level port is an important distribution center and an important port open to the outside world in South China. The centrality values of the second level ports are similar to those of the first level ports, but the port strength is lower than that of the third level ports. This indicates that these ports have established connections with numerous ports, but the connections are not close enough, shipping and transportation are not yet developed, and port

Figure 20. Port hierarchical clustering results



investment and construction still need to be strengthened. The third level ports are mostly located around the main or secondary sea ports, with higher port strength and slightly lower centrality values than the fourth level ports. This indicates that the transit and distribution capabilities of these ports are not strong, and there are fewer ports directly connected to them, but the connections are relatively close, which can easily lead to regional aggregation. The fourth level port is located on a secondary waterway, with a centrality similar to that of the third level port, and a port strength below 100. This indicates that the routes between these ports are relatively scattered and their direct contact with the ports is not close. The centrality and strength of various types of ports at the fifth level are relatively low, with poor transit and distribution capabilities.

CONCLUSION

Taking the South China Sea waters as the research object, based on the main data such as AIS data, port data, and basic geographic information data, this paper uses a variety of methods such as spatio-temporal statistical analysis, linear density analysis cluster analysis, and complex network analysis to analyze the spatio-temporal characteristics of ship trajectories in the South China Sea waters, identify key areas, mine the characteristics of port information, and provide a reference for the maritime regulatory authorities to supervise ships. The main research conclusions are as follows.

The selection of sea routes is always based on the basic principle of “safety and economy,” and the results of satellite AIS data analysis show that the recommended routes for merchant ships in the South China Sea are consistent with those recommended by traditional navigation books. The satellite AIS data fully supports that the recommended routes in the South China Sea have been repeatedly chosen by the ships and the ships have stable speed and smooth navigation during the round trip.

AIS vessel traffic flow data in the South China Sea shows that the number of vessels passing through gate line 4 via the southern waters of the Taiwan Strait has increased significantly compared to other gate lines.

AIS ship traffic flow data mining shows that the average length, width and draft of ships near the Bashi Strait of route gate line 5 are the largest, and it can be judged that the ships sailing through here are mostly large transiting ships passing through the South China Sea into the Pacific waters.

The ship traffic flow in the South China Sea waters is relatively large, and the density of ships in non-coastal waters is mainly concentrated in the waters of the main channel belt. Through the route distribution mapping, it is found that the western route of Nansha Islands in the southern part of the

South China Sea waters is concentrated, and the northern waters of the South China Sea (Xisha, Zhongsha, and Dongsha waters) are distributed in multiple directions; the main width of the middle route belt reaches about 80 nautical miles, the track belt width of the western route is about 30 nautical miles, and the overall concentration varies from 30 to 80 nautical miles.

Based on the hierarchical clustering method, combining the degree, point strength, and centrality parameters of the port, the port is divided into five levels. There is a significant difference in the centrality of ports at different levels in the South China Sea waters, and the level of port development is uneven.

DATA AVAILABILITY

The figures and tables used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Funding for this research was covered by the author of the article.

ACKNOWLEDGMENT

The authors would like to show sincere thanks to those whose techniques have contributed to this research.

REFERENCES

- Chen, C., Sasa, K., Prpić-Oršić, J., & Mizojiri, T. (2021). Statistical analysis of waves' effects on ship navigation using high-resolution numerical wave simulation and shipboard measurements. *Ocean Engineering*, 229, 108757. doi:10.1016/j.oceaneng.2021.108757
- Du, L., Wen, Y. Q., Li, Z. Q., Sun, T. D., Xiao, C. S., & Zhou, C. H. (2016). Marine traffic macroscopic situation assessment model. *J. Dalian Maritime Univ.*, 42(1), 27–33.
- Du, Z., Negenborn, R. R., & Reppa, V. (2021). Multi-vessel cooperative speed regulation for ship manipulation in towing scenarios. *IFAC-PapersOnLine*, 54(16), 384–389. doi:10.1016/j.ifacol.2021.10.120
- Gao, H., Zhao, H., Han, G., & Dong, C. (2021). Spatio-temporal variations of winter phytoplankton blooms northwest of the Luzon Island in the South China Sea. *Frontiers in Marine Science*, 8, 637499. doi:10.3389/fmars.2021.637499
- Greidanus, H., Alvarez, M., Eriksen, T., & Gammieri, V. (2016). Completeness and accuracy of a wide-area maritime situational picture based on automatic ship reporting systems. *Journal of Navigation*, 69(1), 156–168. doi:10.1017/S0373463315000582
- Gunnarsson, B. (2021). Recent ship traffic and developing shipping trends on the Northern Sea Route—Policy implications for future arctic shipping. *Marine Policy*, 124, 104369. doi:10.1016/j.marpol.2020.104369
- Jenkins, H. (1973). *Ocean passages for the world*. Hydrographer of the Navy.
- Karagiannidis, P., & Themelis, N. (2021). Data-driven modelling of ship propulsion and the effect of data pre-processing on the prediction of ship fuel consumption and speed loss. *Ocean Engineering*, 222, 108616. doi:10.1016/j.oceaneng.2021.108616
- Li, H., Liu, J., Liu, R. W., Xiong, N., Wu, K., & Kim, T. H. (2017). A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis. *Sensors (Basel)*, 17(8), 1792. doi:10.3390/s17081792 PMID:28777353
- Li, H., Liu, Y., Sun, C., Dong, Y., & Zhang, S. (2021). Satellite observation of the marine light-fishing and its dynamics in the South China Sea. *Journal of Marine Science and Engineering*, 9(12), 1394. doi:10.3390/jmse9121394
- Li, X., Xiao, Y., Su, F., Wu, W., & Zhou, L. (2021). AIS and VBD data fusion for marine fishing intensity mapping and analysis in the northern part of the south China sea. *ISPRS International Journal of Geo-Information*, 10(5), 277. doi:10.3390/ijgi10050277
- Liu, Z., Wu, Z., & Zheng, Z. (2020). Modelling ship density using a molecular dynamics approach. *Journal of Navigation*, 73(3), 628–645. doi:10.1017/S0373463319000857
- Rosenberg, D., & Chung, C. (2008). Maritime security in the South China Sea: Coordinating coastal and user state priorities. *Ocean Development and International Law*, 39(1), 51–68. doi:10.1080/00908320701641602
- Tang, J., Zhang, X., Yu, T., & Liu, F. (2021). Missing traffic data imputation considering approximate intervals: A hybrid structure integrating adaptive network-based inference and fuzzy rough set. *Physica A*, 573, 125776. doi:10.1016/j.physa.2021.125776
- Vespe, M., Greidanus, H., & Alvarez, M. A. (2015). The declining impact of piracy on maritime transport in the Indian Ocean: Statistical analysis of 5-year vessel tracking data. *Marine Policy*, 59, 9–15. doi:10.1016/j.marpol.2015.04.018
- Wang, Q., Wu, W., Su, F., Xiao, H., Wu, Y., & Yao, G. (2021). Offshore hydrocarbon exploitation observations from VIIRS NTL images: Analyzing the intensity changes and development trends in the South China Sea from 2012 to 2019. *Remote Sensing (Basel)*, 13(5), 946. doi:10.3390/rs13050946
- Wang, Y. (2019). Development of AtoN real-time video surveillance system based on the AIS collision warning. In *2019 5th international conference on transportation information and safety (ICTIS)* (pp. 393–398). IEEE. doi:10.1109/ICTIS.2019.8883727

- Weng, J., Meng, Q., & Qu, X. (2012). Vessel collision frequency estimation in the Singapore Strait. *Journal of Navigation*, 65(2), 207–221. doi:10.1017/S0373463311000683
- Yan, Z., He, R., Ruan, X., & Yang, H. (2022). Footprints of fishing vessels in Chinese waters based on automatic identification system data. *Journal of Sea Research*, 187, 102255. doi:10.1016/j.seares.2022.102255
- Yan, Z., Xiao, Y., Cheng, L., He, R., Ruan, X., Zhou, X., Li, M., & Bin, R. (2020). Exploring AIS data for intelligent maritime routes extraction. *Applied Ocean Research*, 101, 102271. doi:10.1016/j.apor.2020.102271
- Yliskylä-Peuralahti, J., & Gritsenko, D. (2014). Binding rules or voluntary actions? A conceptual framework for CSR in shipping. *WMU Journal of Maritime Affairs*, 13(2), 251–268. doi:10.1007/s13437-014-0059-8
- Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2022). The dynamics of the fishing fleet in China Seas: A glimpse through AIS monitoring. *The Science of the Total Environment*, 819, 153150. doi:10.1016/j.scitotenv.2022.153150 PMID:35041965
- Zhang, Z., & Crabbe, M. J. C. (2021). Management of environmental streaming data to optimize Arctic shipping routes. *Arabian Journal of Geosciences*, 14(15), 1–8. doi:10.1007/s12517-021-07782-0
- Zhong, H., & White, M. (2017). South China Sea: Its importance for shipping, trade, energy and fisheries. *Asia-Pacific Journal of Ocean Law and Policy*, 2(1), 9–24. doi:10.1163/24519391-00201003
- Zhou, J., Sun, H., Wang, Z., Cong, W., Wang, J., Zeng, M., Zhou, W., Bie, P., Liu, L., Wen, T., Han, G., Wang, M., Liu, R., Lu, L., Ren, Z., Chen, M., Zeng, Z., Liang, P., Liang, C., & Fan, J. (2020). Guidelines for the diagnosis and treatment of hepatocellular carcinoma (2019 edition). *Liver Cancer*, 9(6), 682–720. doi:10.1159/000509424 PMID:33442540