Research Article

A GIS-Based Spatial-Temporal Autoregressive Model for Forecasting Marine Traffic Volume of a Shipping Network

Xianzhe Zhang,[1,2,3] **Gang Chen,**[1,2,3] **Jiechen Wang,**[1,2,3] **Manchun Li** ᵢᴅ **,**[1,2,3] **and Liang Cheng**[1,2,3]

¹Collaborative Innovation Center of South China Sea Studies, Nanjing University, Nanjing 210093, China
²Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing 210023, China
³School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China

Correspondence should be addressed to Manchun Li; limanchun_nju@outlook.com

Research on the forecasting of marine traffic flows can provide a basis for port planning, planning the water area layout, and ship navigation management and provides a practical background for sustainable development evaluation of shipping. Most of the traditional marine traffic volume forecasting studies focus on the variation of the traffic volume of a single port or section in time dimension and less research on traffic correlation of associated ports in shipping networks. To reveal the spatial-temporal autocorrelation characteristics of the shipping network and to establish a suitable space-time forecasting model for marine traffic volume, this paper uses the AIS data from 2011 to 2016 for the South China Sea to construct a regional shipping network. The adjacent discrimination rule based on network correlation is proposed, and the traffic demand between ports is estimated based on the gravity model. On this basis, STARMA (space-time autoregressive moving average) model was introduced for deducing the interaction between he traffic volumes of adjacent ports in shipping network. The experimental results show that (1) there is a significant positive correlation between time and space in the South China Sea shipping network, and this spatial-temporal correlation has the characteristics of time dynamics and spatial heterogeneity; (2) the forecasting accuracy of the marine traffic volume based on the spatial-temporal model is better than the traditional time-series-based forecasting model, and the spatial-temporal model can better portray the spatial-temporal autocorrelation of maritime traffic.

1. Introduction

With the deepening of economic globalization, marine transportation as an important pillar of international trade has developed greatly [1]. At the same time, with the advancement of science and technology and the market demand continuously being upgraded, the marine traffic is becoming increasingly busy, and the contradiction between busy marine traffic and navigation safety, navigation efficiency, and navigation resources is becoming increasingly prominent [2]. To alleviate this contradiction, better safeguard navigation, improve navigation efficiency in water areas, and scientifically and rationally plan the layout of water areas, it is necessary to determine the future development trends by forecasting and

researching marine traffic flows and to promote the sustainable development of shipping.

Marine traffic volume forecasting uses the historical data from society, economy, and transportation to predict traffic conditions in the sea for a period of time in the future, and whether the selection of the forecasting method is reasonable is the key to predicting success or failure. According to the nature of the forecasting method itself, the common forecasting methods can be summarized into three categories: linear-based forecasting methods, nonlinear-based forecasting methods, and combined forecasting methods. The forecasting method based on linear theory realizes prediction in a certain period of time in the future based on the periodic pattern of time series and is widely used in traffic

volume forecasting [3]. However, the disadvantage of the time-series model is that it is only applicable to the situation where the traffic volume trend is stable. As the nonstationary of traffic volume increasing, there is a large deviation between the forecast result and the observed value [4]. With the development of emerging intelligent methods, such as artificial intelligence, machine learnings, neural networks [5, 6], support vector machine [7, 8], and genetic algorithms, the forecasting model of knowledge discovery provides a new method for modelling spatial-temporal series. Researchers use the time series as an input variable to intelligent calculation methods to forecast future marine traffic to reflect seasonal changes in the marine traffic volume [9], and the results show that the nonlinear forecasting results are more accurate and reliable [10]. Although the marine traffic forecasting based on the intelligent method can be applied to the prediction of nonstationary marine traffic, there are problems such as a slow convergence rate and the ease of fall into a local optimal solution [11], and it is difficult to achieve higher accuracy relying on a single forecasting model. To overcome the deficiencies of a single forecasting model, scholars have studied the use of multiple forecasting models for joint forecasting and have achieved better results [12]. The forecasting model based on hybrid theory can effectively improve the forecasting accuracy of traffic volume to a certain extent, but due to the relatively complex model, the setting of its weight parameter is the key to the accuracy of the model [13].

In summary, it can be seen that each forecasting model has its applicable conditions and that the merits of a certain model cannot be determined simply. In addition, most of the above methods for predicting the maritime traffic volume focus on analyzing the variation characteristics of the traffic volume of a single port or section in the time dimension [14, 15] and are less involved in large-scale shipping networks, taking into account the link to traffic conditions between ports. Urban road network, which belongs to the transportation network as the shipping network, mainly defines the spatial adjacency [16] by means of the upstream and downstream relationship of the road network [17]. However, marine traffic is a complex system, and the spread of influence of traffic conditions between adjacent routes is not isotropic [18]. At the same time, the influence of different network topologies on the diffusion of traffic conditions also has a greater impact and constraints [19, 20]. It is biased to rely solely on simple spatial adjacency to determine the autocorrelation of traffic conditions. In addition, to more accurately depict the complex geographical phenomena existing in the shipping network, it is necessary to avoid isolated use of a temporal or spatial model. The spatial-temporal models such as space-time dynamics models and space-time Bayes, are widely used in climate change [21], environmental monitoring [22], urban transport [23], public health [24], and other fields. However, most of the spatial-temporal models are built for specific applications and do not adequately reflect the geographic characteristics of spatial-temporal data. As a very important spatial-temporal statistical model, the STARMA (space-time autoregressive moving average) model characterizes the spatial-temporal

dependence between different regions by deriving the interaction between adjacent regions within the system and is more suitable for modelling geographic spatial-temporal data [25]. Therefore, in order to reveal the spatial-temporal autocorrelation characteristics of shipping networks and to establish a suitable spatial-temporal prediction model for marine traffic volume, this paper proposes an adjacent rule based on network correlation based on the definition of road network adjacency and extends STARMA model to derive the relationship between traffic volumes between ports.

The basis for marine traffic volume forecasting is historical data on marine traffic. Not much data were available to use in previous marine traffic surveys, and usually only a few major ports were available. The lack of accurate marine traffic data has become a bottleneck in the previous study, which has brought great difficulties to the forecasting of marine traffic volume [26]. AIS (automatic identification system) data contain a large number of marine traffic characteristics, from which we can obtain the characteristics of ship behavior. In recent years, the number of shore-based base stations and low-orbit satellites in the AIS system has continuously increased, and the continuity and effectiveness of the ship's trajectory are continuously improving, making it possible to forecast the marine traffic volume.

In view of the above, this paper studies the STARMA model to establish the spatial-temporal model of traffic volume in the shipping network. First, the time series of traffic flow attached to the port are taken as the objects, and the spatial-temporal correlation characteristics of the marine traffic flow network are analyzed by using spatial statistics. Second, the port correlation in the network that meets the adjacency characteristics is determined based on the subdivision hierarchy of the marine traffic flow network, and the traffic demand between ports is estimated based on the gravity model. On this basis, the spatial weight matrix is constructed. Finally, we use the measured marine traffic volume of ports in the South China Sea and its surrounding sea to verify the applicability of the space-time autoregressive model.

2. Materials and Methods

2.1. Study Area. The study area of this paper is in the range of 96°E~126°E and −5°S~26°N, and the specific range is shown in Figure 1. The total study area is approximately $4.2 \times 10^6 \, \text{km}^2$ and includes the areas surrounding the South China Sea, the Gulf of Thailand, the Philippines, the Strait of Malacca, and the sea surrounding Kalimantan. The layout of the shipping network is regional, to more accurately reflect the regional shipping system and the study area in this paper covers the national ports around the South China Sea. Hereafter "South China Sea" in this article refers to the study area in this article.

The South China Sea is one of Asia's three largest marginal seas and is known as the "Mediterranean of Asia." The South China Sea shipping system is composed of numerous ports and routes, and the differences among the ports are relatively large. Due to the large number of people
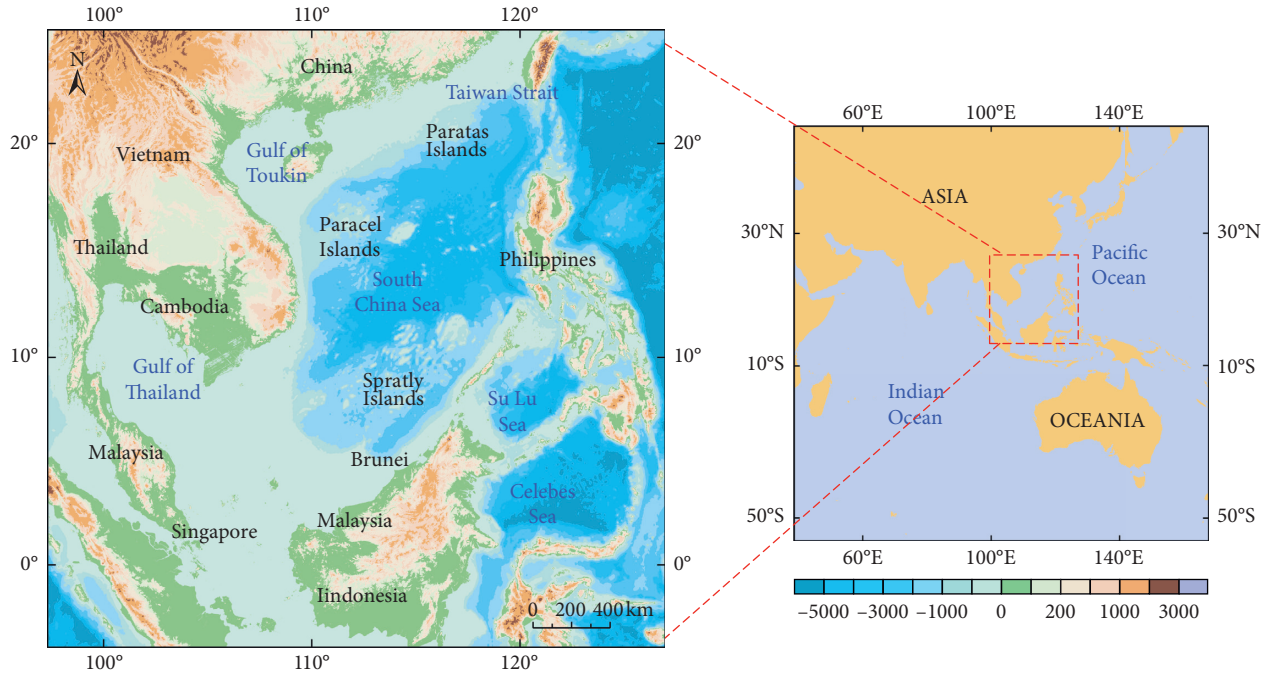
FIGURE 1: The study area of this paper.

involved and the large economic aggregates, its development potential is huge. As a very important location, the South China Sea is a must-have destination for East Asian and Southeast Asian westbound routes, and its special geographical location means that it has a large traffic flow within its sea area [27]. According to the statistics of the World Shipping Council, 25% of the world's sea transport takes place through the South China Sea, and an average of 300 ships travel through the South China Sea every day. In addition, the Malacca Strait, which is located between the Indian Ocean and the Pacific Ocean, is one of the most important channels for international trade transport, and it bears one-third of the global trade volume of goods and more than one-half of the oil transport volume. Based on the above reasons, this paper focuses on analyzing the spatial-temporal characteristics of traffic flow around the South China Sea and, based on this, builds a spatial-temporal model of the marine traffic volume.

2.2. Data and Data Processing. The experimental data used in this paper are the AIS data within the study area from January 2011 to December 2016. The ship types in the AIS data include cargo ships, passenger ships, tugboats, pilot ships, and oil tankers. The other types of data are excluded, and only the cargo ship data are retained as the data for the experiment; the amount of AIS data involved in each year is 1.27, 1.36, 1.52, 1.64, 1.58, and 175 million, respectively.

In this paper, the grid DBSCAN algorithm [28] is used to obtain the area of the ship's stay area in the study area, and the near-shore section is selected as the ship arrival port identification. On this basis, the ship's stop position at the port of departure (Lon_{od}, Lat_{od}), start time

t_o and end time t_d, and ship track point series Traj = $\{p_0, p_1, p_2, \ldots, p_k\}$ in the open sea area are used to extract the ship trip information.

Through the above steps, thousands of ship stay areas were obtained. However, after analyzing the data, it was found that a port city has multiple stopovers, and in reality, many of the ship's stay areas are only one of the many port areas in the port city, such as Singapore's main port areas of Jurong, Pulau Bukom, Sembawang, Tanjong, and Penjuru. In this paper, we obtained 155 port cities by integrating the obtained stay areas; that is, this article uses the port city as a shipping network node. Based on this, we obtained a total of 368 pairs of ports with a total of more than 50 voyages in the surrounding ports of the South China Sea, which constitutes the connections between the ports in the shipping network. In addition, to excavate the internal correlation characteristics of the ports in the shipping network and to reveal the spatial-temporal correlation of the marine traffic flow, the time series of each port's traffic flow were calculated. The skewness and kurtosis values of some ports were found to be relatively large (skewed > 20 and kurtosis > 5), and after further verification by the normal distribution test, the abnormal ports are deleted. Finally, the 60 ports, which mean monthly voyages are greater than 10, were selected as the modelling objects of the spatial-temporal model (Table 1).

Figure 2(a) is the spatial distribution map of 60 representative cargo ports in the study area. Figure 2(b) shows the monthly time series of marine traffic in the three ports of Thailand, Hong Kong, and Singapore in the South China Sea shipping network for 6 years. It has an upward trend with obvious periodic characteristics. During the experiment, to test the accuracy of the spatial-temporal model, data from 2011 to 2015 were used as training samples, data

TABLE 1: Monthly voyage time-series statistics for ports in the shipping network.

| Rank	Name	Numb	Mean	STDEV	Skewed	Kurtosis
1	Tainan	72	782.6250	91.4560	18.8483	−3.7311
2	Balikpapan	72	42.8472	19.2475	−0.0837	−1.0373
3	Bangkok	72	469.6388	131.894	2.7460	−1.8795
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	Fangchenggang	72	533.2916	119.1145	1.0687	−1.0374
29	Manila	72	254.7222	78.7479	0.6890	−0.1742
30	Beihai	72	142.7361	55.6224	0.3134	0.4706
⋮	⋮	⋮	⋮	⋮	⋮	⋮
58	Zhanjiang	72	301.4583	76.5235	0.9795	0.0841
59	Lungsod ng Cebu	72	112.2083	62.2414	−1.0059	0.4421
60	Samarinda	72	93.9583	57.6522	−1.3028	−0.6015
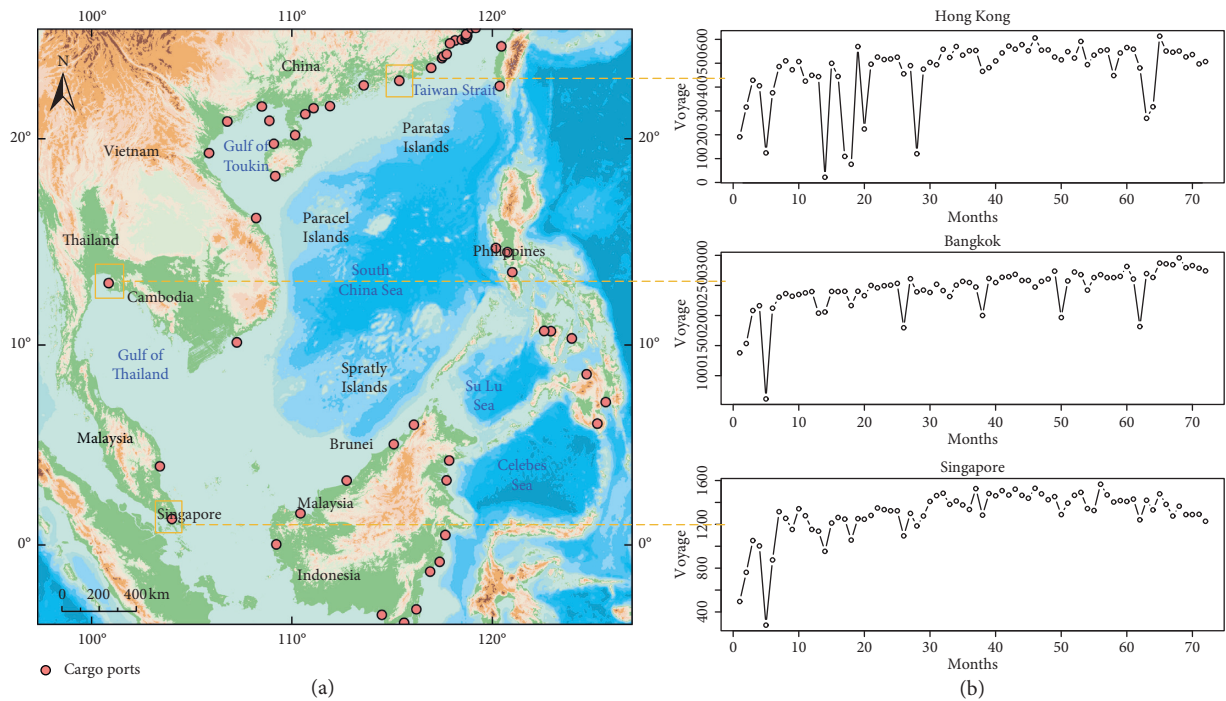
FIGURE 2: Spatial-temporal modelling of port spatial distribution and voyage time-series characteristics. (a) The spatial distribution of 60 ports. (b) Thailand, Hong Kong, and Singapore voyage time series.

from 2016 were used as test samples, and the forecast time was 12 months.

3. Method

The STARMA (space-time autoregressive moving average) model fully considers the autocorrelation existing in the geographic spatial-temporal data set and is suitable for processing spatial-temporal series. However, the STARMA has certain limitations in some aspects, such as spatial and temporal nonstationary time-series data modelling and effectively describing the correlation between spatial variable series by existing spatial weight matrix decision rules. In view of the above problems, this paper builds the spatial-temporal model of marine traffic volume in shipping network based on the STARMA model. First, Moran'I and temporal autocorrelation function are used to test the spatial and temporal autocorrelations of spatial-temporal data. Second, the shipping network hierarchical subdivision structure is used to determine the port correlation in the network that meets the adjacent rules, and the traffic demand between the supply end and the demand side is comprehensively considered to estimate the traffic demand between ports. On this basis, the spatial weight matrix of the spatial-temporal model of the shipping network is constructed. Finally, the applicability of the spatial-temporal model based on network correlation proposed in this paper is verified by using the measured ship traffic volume data of the ports in the South China Sea and its surrounding seas.

3.1. Spatial and Temporal Characteristics of Traffic Volume in Shipping Network. The traditional linear regression method cannot capture the autocorrelation characteristics contained in spatial-temporal data, which makes it impossible to find

the spatial-temporal patterns. Therefore, the premise of using above models to simulate traffic volume in the shipping network is that the data are statistically spatially independent and evenly distributed. In order to explore the spatial-temporal heterogeneity and nonstationary characteristics of the spatial-temporal series in the shipping network, we use Moran's I statistic to describe the spatial autocorrelation characteristics of traffic volume in shipping network and use time autocorrelation function and partial autocorrelation function to describe the temporal autocorrelation characteristics of traffic volume in the shipping networks.

Moran's I statistic is the most common type of spatial autocorrelation statistic in current applications and is divided into global Moran's I and local Moran's I according to the difference in the metric regions. The Global Moran's I statistic is used to characterize the degree of correlation between all spatial objects in the spatial range to show whether the spatial distribution pattern among the spatial objects is significant. The local Moran's I statistic is used to characterize the local spatial distribution pattern.

Time autocorrelation is one of the main features of time-series data, which can express the dependence of object attributes in time. The ACF illustrates the degree of correlation between time-series data at different times, ranging from −1 to 1, and the closer the value to 1 is, the higher the degree of autocorrelation of the time series is. The ACF map shows how the correlation of the time-series changes with the lag period k. The partial autocorrelation function is another method for describing the structural characteristics of a stochastic process in a stationary time-series and each regression coefficient φ_{kk} in the function indicates the autocorrelation coefficient between x_t and x_{t-k} which excludes the influence of the intermediate variables $x_{t-1}, x_{t-2}, x_t, \ldots, x_{t-k+1}$.

3.2. Spatial Weight Matrix considering Network Node Correlations.

The purpose of the spatial weight matrix is to accurately and comprehensively reflect the correlation between spatial geographical units. However, because of the economic development level of the port hinterland and the geographic location of the traffic flow, the spread of marine traffic flow is not isotropic. Even if the geographical distance is closest to the port, the traffic status may not be directly related. This nonuniformity of the proliferation of the marine traffic flow makes it difficult to find the spatial correlation of marine traffic by measuring the adjacent ports by geographical distance alone. To solve the above problems, first, based on the subdivision hierarchy structure of shipping network obtained in the prestudy [29], the port correlation in the network that meets the adjacent rules is determined. Second, the demand for traffic between ports based on the gravity model is estimated. Finally, a spatial weight matrix is constructed based on the above steps.

3.2.1. Definition of Spatial Adjacency.

In the spatial weight setting of the shipping network, it is first necessary to define spatial adjacency, which is also the basis of the spatial pattern measure. The spatial adjacency definition methods commonly used mainly include methods such as coboundary method, topological graph method, and distance method. However, the abovementioned spatial adjacency rules are mostly based on their own related perspectives, ignoring the network aggregation characteristics of the traffic flow in the shipping network. This paper defines the adjacency rules of any two ports in the shipping network as follows: (1) the two ports are directly adjacent; that is, there is navigation between the two ports; (2) the two ports are not directly adjacent, and at this time, the connectivity of the two ports needs to pass through one or more transit ports. Thus, based on the perspective of a complex network, the shipping network is identified as a hierarchical subdivision structure (Figure 3).

In the previous study of this paper [29], the shipping network in the South China Sea has been proven to be a typical scale-free network. There are obvious hierarchical and community characteristics in the network, and there is a direct connection with the ports at the adjacent level within the community. This kind of relationship between nodes in a complex network is just an expression of the adjacent relationship of the spatial adjacency matrix of the port nodes. Thus, the spatial adjacency matrix in the shipping network is defined as follows:

$$
\mathbf{W}_{ij} = \begin{cases} 1, & i,\ j \in M,\ i,\ j \text{ belongs to adjacent hierarchical,} \\ 0, & i \in M,\ j \in N, \end{cases}
$$

(1)

where i and j are the nodes in a shipping network; M and N are the different communities in a shipping network; $\mathbf{W}_{ij} = 1$ indicates that i and j have an adjacent relationship in the network, that is, belong to the same community; and two nodes are hierarchically adjacent. $\mathbf{W}_{ij} = 0$ indicates that nodes i and j are not adjacent in the network; that is, they do not belong to the same community, or two nodes are not hierarchically adjacent.

3.2.2. Definition of Spatial Weight.

Based on the spatial weight matrix created by the influence of geographical spatial relationship and attribute values, it is easier to reflect the spatial agglomeration and spatial differences in spatial units. The monthly forecast of marine traffic volume is mainly affected by many factors, such as weather conditions and navigable water level, and the marine traffic volume shows a certain degree of periodicity and seasonality. According to previous scholars' research, the gravity model can better explain the spatial adjacency of ports in shipping networks.

Spatial weight is a quantitative measure of adjacency. The gravity model is an important method to explain the spatial adjacency among ports in shipping networks [30]; that is, voyages between adjacent ports are more frequent than voyages between distant ports. If d_{ij} is the distance from port i to j, then the trend of the interaction between them is expressed by the lag function $f(d_{ij})$, and the voyage expression between them $V_{ij} = O_i D_j f(d_{ij})$, where O_i is the
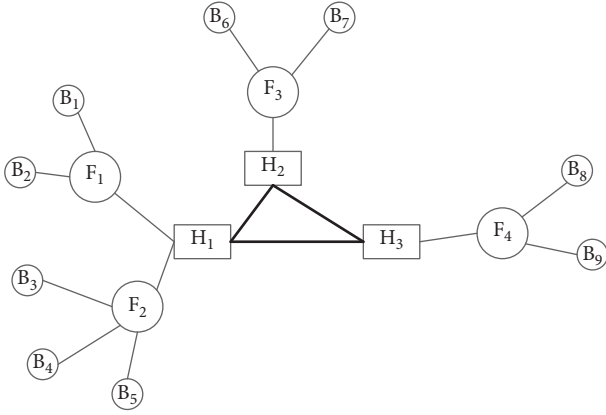
H: hub port
F: transit port
B: terminal port

FIGURE 3: Diagram for port correlation in ship traffic flow network.

total number of voyages leaving port i and D_j is the total number of voyages arriving at port j.

In estimating the weight of marine traffic between ports, the traffic attraction between the supply end and the demand end of the marine traffic volume is taken into consideration, and the gravity model is used to estimate the traffic volume between ports. The amount of traffic on the supply end and the demand end is proportional to the traffic and is inversely proportional to the traffic impedance between the two. The following equation is used to estimate the traffic volume between ports:

$$V_{ij} = k \frac{O_i^{\alpha} D_j^{\beta}}{e^{d_{ij}^{-\gamma}}}, \tag{2}$$

where V_{ij} is the amount of traffic generated between ports i and j, O_i^{α} and D_j^{β} are the total traffic volumes of ports i and j, and α and β are the traffic volume coefficients ranging from 1 to 2, which are set to 1. In this paper, $e^{d_{ij}^{-\gamma}}$ is the distance attenuation function and γ is set to 0.59 [31].

3.3. STARMA Model.
The STARMA model uses spatial-temporal delay operators to expresses the spatial-temporal variables that are influenced by both temporal lag and spatial lag [25]. The spatial lag operator L indicates that the traffic volume of a port in space is affected by the adjacent port, and the forecasting value of the port traffic volume can be expressed as a weighted average of the values of the adjacent port traffic variables. Its quantified expression is as follows:

$$z_i(t) = \sum_{k=1}^{p} \sum_{h=1}^{m_k} \varphi_{kh} L^{(h)} z_i(t-k) - \sum_{l=1}^{q} \sum_{h=0}^{n_l} \theta_{kh} L^{(h)} \varepsilon_i(t-l) + \varepsilon_i(t), \tag{3}$$

where p is the time autocorrelation coefficient, q is the temporal moving average order, $L^{(h)}$ is the spatial lag operator, m_k is the spatial order of the k temporal autocorrelation term, n_l is the spatial order of the l temporal moving average term, φ_{kh} and θ_{kh} are parameters, and $\varepsilon_i(t)$ represents the random error.

The basic idea of the STARMA model is to derive item by item according to the spatial-temporal series and calculate the spatial-temporal average of each item in turn to realize the simulation of long-term trend change. The periodic influence factors of traffic volume in the shipping network make the traffic volume to show a strong seasonality, while the trend influencing factors will make the traffic volume change to have long-term trend characteristics. Therefore, according to the time-series statistics, the ship traffic flow has the characteristics of short-term time autocorrelation and long-term trend change. In this paper, the traffic flow time-series data are decomposed into large-scale spatio-temporal nonlinear trend u and small-scale spatiotemporal variation trend e. The large-scale variability refers to the deterministic global trend, while the small-scale variability refers to the stationary sequence separated by spatial-temporal data. Such a nonstationary spatiotemporal sequence can be expressed as $z_i(t) = \mu_i(t) + e_i(t)$, where $z_i(t)$ is the observation, $\mu_i(t)$ is the large-scale space-time deterministic trend, and $e_i(t)$ is the periodic trend of the mean constant. The linear part $e_i(t)$ is modelled by STARMA, and the nonlinear part $\mu_i(t)$ is modelled by the neural network. On this basis, the linear forecasting results are combined with the nonlinear results to obtain the final results.

4. Results

4.1. Traffic Flow Time-Series Analysis

4.1.1. Spatial Autocorrelation Analysis.
Spatial autocorrelation is an important feature of the spatial geography unit, and this paper analyzes the distribution of marine traffic of each port from two aspects, global, and local as follows:

(1) Global Spatial Autocorrelation. This paper uses the global Moran's I to perform global spatial autocorrelation analysis of arrival voyages in the surrounding ports of the South China Sea during 2011~2016. Taking the arrival voyages of each port as variables, based on the distance weight matrix, we use GeoDa software to calculate the global Moran's I value. As Table 2 shows, the global Moran's I value of the arrival voyages to the ports of the South China Sea is positive, and the result passed a Z-value test with a significance level of 0.01. The fact that the global Moran's I index value is greater than 0 in each year indicates that the marine volume of the surrounding ports in the South China Sea has a positive spatial correlation and shows a certain degree of spatial aggregation. At the same time, Moran's I index significance level values are all below 0.001, indicating that the spatial correlation of shipping development in the ports of the South China Sea is relatively low, and the overall spatial agglomeration effect is not obvious. In general, marine traffic flow is spatially weakly stable, and in the modelling, the marine traffic flow in the shipping network can be considered as smooth data.

(2) Local Spatial Autocorrelation. The spatial correlation distribution of the marine traffic volume in the South China Sea was obtained through the local Moran's I analysis. As Figure 4 shows, the marine traffic flow mainly presents a

TABLE 2: Multiyear global autocorrelation statistics.

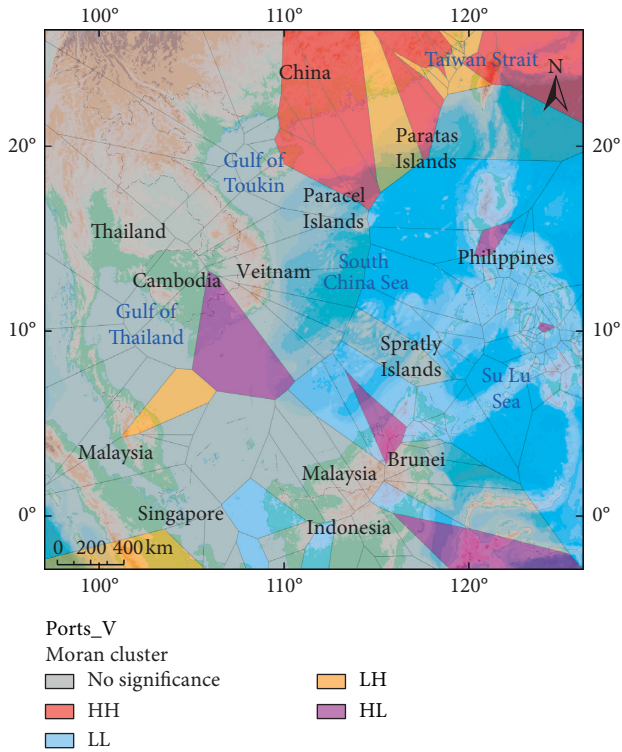| Year	2011	2012	2013	2014	2015	2016
Moran's I	0.069	0.079	0.080	0.083	0.072	0.073
Z-value	3.593	3.922	3.880	4.012	3.580	3.687
Significance level	0.0003	0.0001	0.0001	0.0001	0.0003	0.0002

FIGURE 4: Local spatial autocorrelation of ports in the South China Sea.

high-high aggregation model in China's coastal area, while in the Philippines, a low-low aggregation model is mainly presented. It is not difficult to understand as the route density in China's coastal areas is large and the links between ports are closer than those in other regions. The coastal areas in China have a high density of routes, and the connections between ports are closer than in other regions. It is not difficult to understand that China's foreign trade has continued to grow rapidly and voyages between ports have continued to hit new highs, while the Philippines has fewer voyages due to its smaller economic aggregates and low dependence on foreign trade. In addition, Ho Chi Minh, Manila, and surrounding areas are high-low correlation models. The abovementioned ports are either economic centers of the countries around the South China Sea or occupy important geographical locations, and all bear the pivotal role of connecting with other ports. Therefore, their marine traffic is relatively large and forms a high-low correlation model with the surrounding ports.

4.1.2. Temporal Autocorrelation Analysis. The ACF map can represent the change in the autocorrelation coefficient between any time periods in the time series. Figure 5 shows the ACF and PACF plot of the monthly time-series data of the total traffic volume of the surrounding ports in the South China Sea from January 2011 to December 2016. Among them, Figures 5(a) and 5(b) are the original time-series data of the total traffic volume and the time-series data after the difference; Figure 5(c) shows the ACF plot of the spatial-temporal series data for each port after the difference, where the abscissa is the lag period, the ordinate is the corresponding autocorrelation function value (that is, the correlation degree), and the dashed line is the confidence level with a 95% confidence interval approximation; Figure 5(d) is the PACF plot of the spatial-temporal series for each port after the difference, and the ordinate is the corresponding partial autocorrelation function value (that is, the degree of correlation).

In Figure 5(c), the step length of each lag period is one month, and within 12 months (lag period $k < 13$), the ACF is greater than 0.2; the ACF gradually decreases with the increase of the lag period k. The above content shows that the correlation degree ρ_k of the spatial-temporal series data for each port's traffic volume gradually weakens when the lag period k increase. In addition, the traffic volume for each port has a strong positive correlation in a certain time range and the traffic volume in the current time period has a certain degree of correlation with the traffic volume in the subsequent time period. As can be seen in Figure 5(d), the PACF rapidly approaches zero from 0.7 and exhibits an alternating exponential decay, which shows that the data after the difference are time-stable time series. Furthermore, the PACF is truncated after the first-order time lag, indicating that the time-shifted average degree is 1.

4.2. STARMA Model Building. In the previous study of this paper [29], the port communities in the shipping network are named China Coastal Community, Taiwan Community, Beibu Community, ASEAN Community, and Philippine Community (Figure 6). Combined with the centrality of ports, the network is presented as a subdivision hierarchy structure. From the above definition of the spatial adjacent relationship, the first-order and second-order spatial adjacencies of the ports around the South China Sea are obtained. The diagonal element of the spatial adjacency matrix is 0. If the off-diagonal element is 0, there is no adjacent relationship between the spatial units corresponding to the row and the column; if the nondiagonal elements is 1, the spatial unit corresponding to the row and the column corresponds to adjacency. Then, by weighting of equation (2) and normalization of the adjacent matrix, the first-order spatial weight matrix $\mathbf{W}^{(1)}$ and the second-order spatial weight matrix $\mathbf{W}^{(2)}$ are obtained.

In this paper, the ST-ACF and ST-PACF values are calculated based on spatial-temporal series data for each port in the shipping network. The calculation results are as follows:

It can be seen from Tables 3 and 4 that the values of the ST-ACF and ST-PACF show a progressively decreasing phenomenon in both time dimension and the spatial
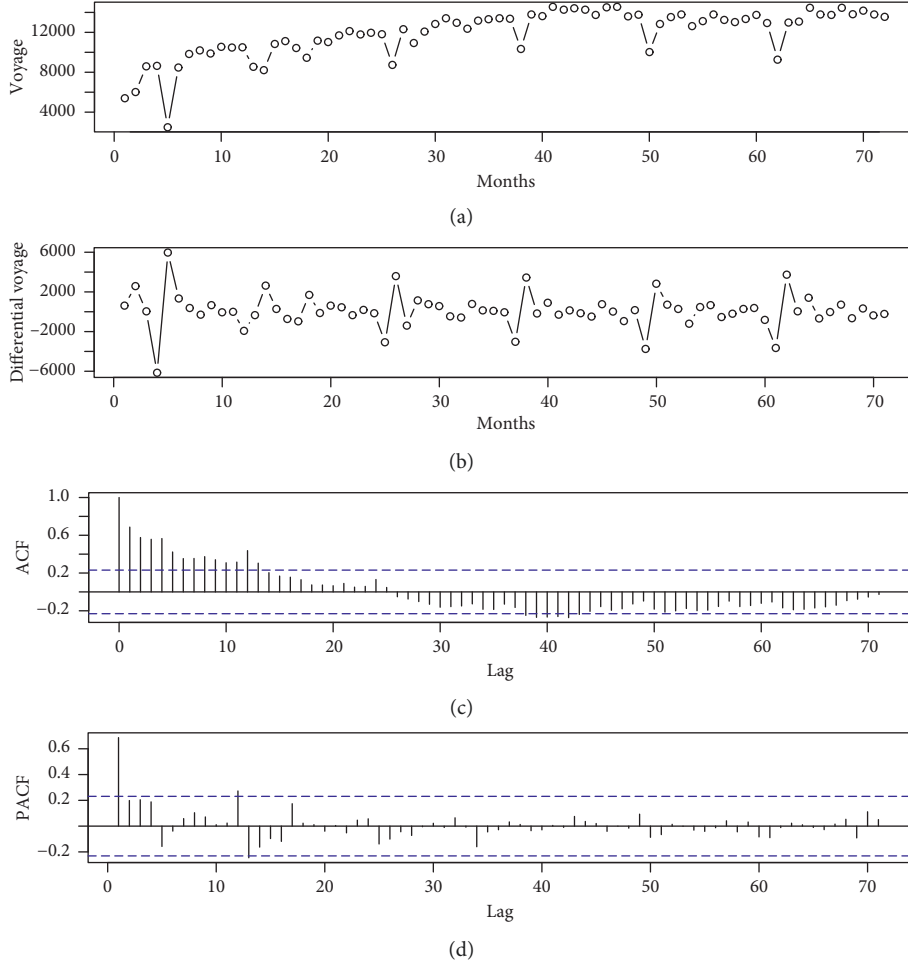
(a)

(b)

(c)

(d)

FIGURE 5: (a) The original time-series data, (b) the time-series data after the difference, (c) spatial-temporal autocorrelation function, and (d) partial autocorrelation function plot of the monthly traffic volume for ports in the shipping network.

dimension. The ST-PACF shows truncation after the first-order spatial and temporal lags. Therefore, it can be concluded that the sample is a spatial-temporal autocorrelation process of first-order time lag and second-order spatial lag. The initial judgment of the STARMA equation is STARMA (1, 0). The specific model is as follows:

$$z(t) = \varphi_{10}z(t-1) + \varphi_{11}\mathbf{W}^{(1)} \cdot z(t-1) + \varphi_{12}\mathbf{W}^{(2)} \cdot z(t-1) + \varepsilon(t). \tag{4}$$

After obtaining the above model, the parameters in the above equation are further estimated using the sample data from 2011 to 2015. The estimated value of each parameter in the equation and its test value are shown in Table 5.

Table 5 lists the parameter estimates of φ_{10}, φ_{11}, and φ_{12} in the equation and the corresponding significance test results. It can be seen that the p values of the three parameters are all less than 0.05; therefore, the assumption that the equation coefficient is not related to the dependent variable should be rejected; that is, the independent variable can explain the change in the dependent variable. Therefore, the spatial-temporal model can be expressed as follows:

$$z(t) = 0.4227z(t-1) + 0.2628\mathbf{W}^{(1)} \cdot z(t-1) + 0.3685\mathbf{W}^{(2)} \\ \cdot z(t-1) + \varepsilon(t). \tag{5}$$

4.3. Model Comparison. To test the forecasting model of shipping network traffic, we used the nonspatial time-series prediction method STL (seasonal-trend decomposition procedure based on Loess) to verify the results obtained. The STL method is a time-series decomposition method proposed by Cleveland in 1990 [32], and this method decomposes the time-series data into independent long-term trend terms, seasonal components, and residuals to achieve predictions. Although the STL method cannot achieve a spatial-temporal forecast, it can obtain the evolutionary characteristics and rules of things over time by processing the time-series data and then forecast the future development of things [33].

Based on the obtained shipping network model equation (5), spatial-temporal forecasts are made using the first-order and second-order spatial weight matrices constructed previously. By taking the traffic volume time-series data of the
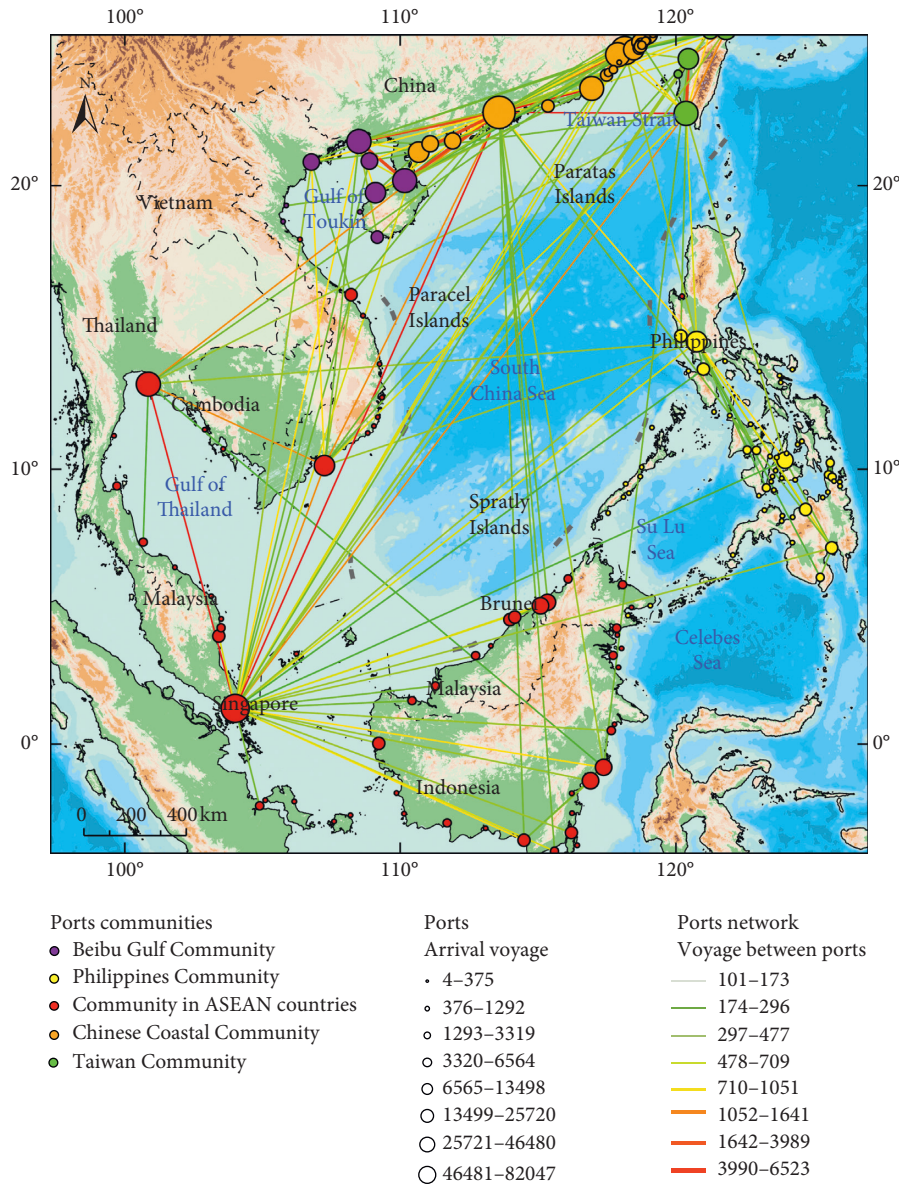
FIGURE 6: Community detection result of shipping network in the South China Sea.

TABLE 3: ST-ACF plot of spatial-temporal series data of arrival voyages for 60 ports

Temporal lag (h)	Spatial lag (h)		
	0	1	2
1	0.8828	0.8987	0.8142
2	0.8383	0.8684	0.7736
3	0.7736	0.8189	0.7130
4	0.7343	0.7742	0.6758
5	0.6003	0.6188	0.5490
6	0.5496	0.5387	0.5030

TABLE 4: ST-PACF plot of spatial-temporal series data of arrival voyages for 60 ports.

Temporal lag (h)	Spatial lag (h)		
	0	1	2
1	0.9676	0.2668	0.0166
2	0.0177	−0.0748	−0.1821
3	−0.0571	0.0912	−0.0235
4	0.0032	0.0442	−0.0132
5	−0.0102	−0.0164	−0.0384
6	−0.0389	0.0712	−0.0703

60 ports around the South China Sea from 2011 to 2015 as training data and taking the traffic volume time-series data in 2016 as the test sample of the model's forecasting ability, the spatial-temporal forecast results were obtained. In the spatial dimension, Figure 7 shows the spatial-temporal models and STL forecasting results of marine volume in the South China Sea shipping network in 2016. In the temporal dimension, Figure 8 shows the forecasting results of marine traffic volume in the South China Sea shipping network in 2016.

TABLE 5: STARMA model parameter estimates and their test values.

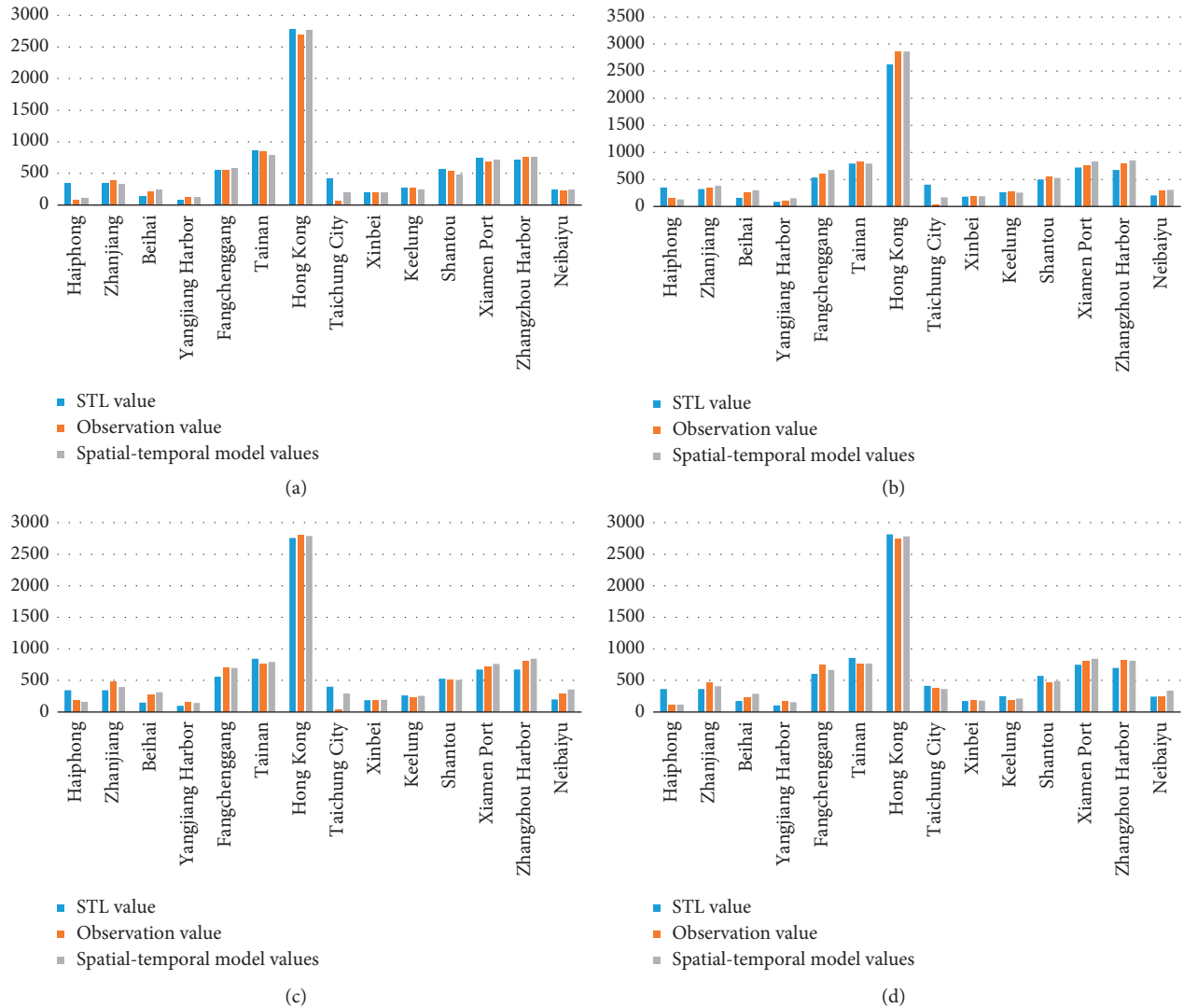|	φ_{10}	φ_{11}	φ_{12}
Value	0.4227	0.26280	0.3685
Standard deviation	0.0153	0.0109	0.0151
t value	26.7543	23.8937	24.2650
p value	0.0000	0.0000	0.0000

(a)

(b)

(c)

(d)

FIGURE 7: The marine volume prediction results of the STL and spatial-temporal model for some ports in the shipping network in (a) March, (b) June, (c) September, and (d) December of 2016.

From the forecast results in Figure 7, in the spatial domain, compared with the STL method, the forecast results of the spatial-temporal model of the shipping network are found to adapt better to the prediction results of different port sizes and are closer to the observations values. In Figure 8, in the temporal domain, it can be found that both models can obtain a good fit. Furthermore, it can also be found that when the time-stationary of the observations is high, the two models can approach the observations very well; when the observations vary greatly with time, the

spatial-temporal model can more easily capture such changes.

4.4. Model Accuracy.
In the previous section, we compared the time domain and space domain fitting results of marine traffic volume. In this section, we compare and analyze the two generalization accuracy evaluation results of traffic volume forecasting of the shipping network in the South China Sea. In order to compare the forecasting accuracy
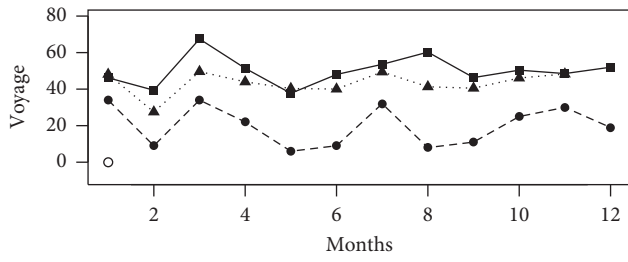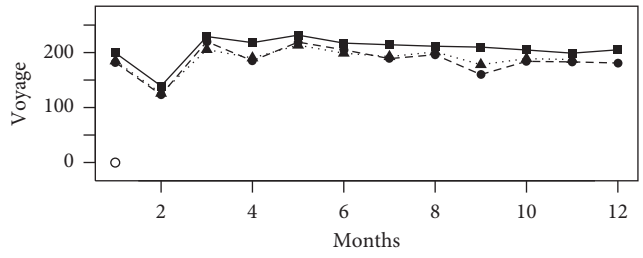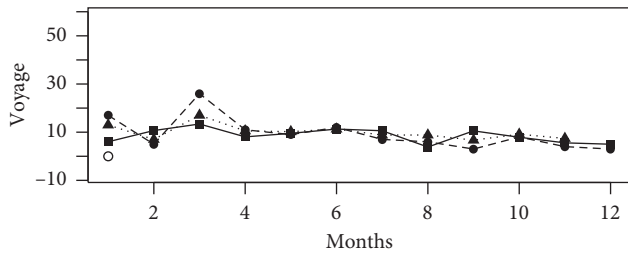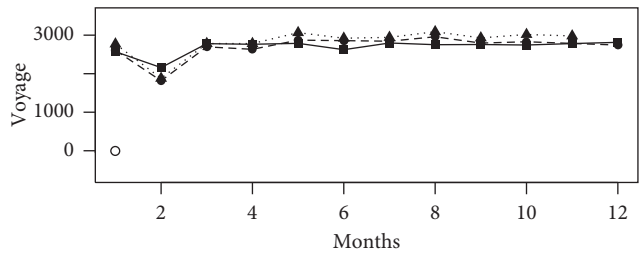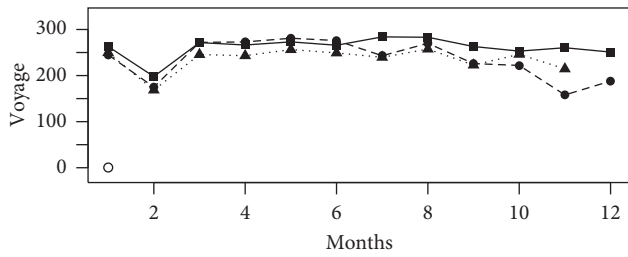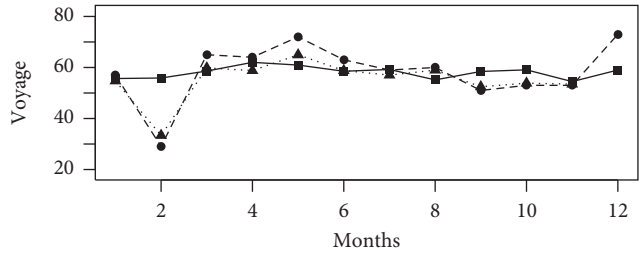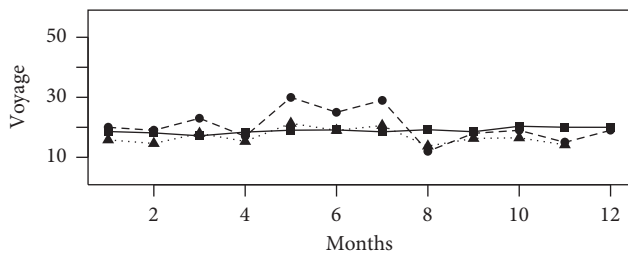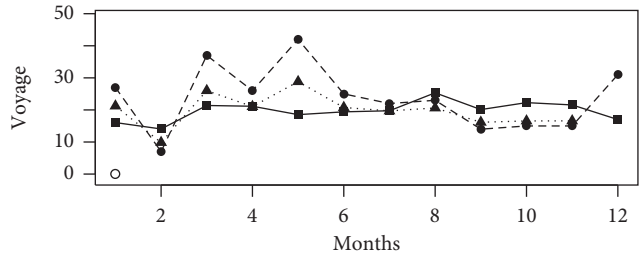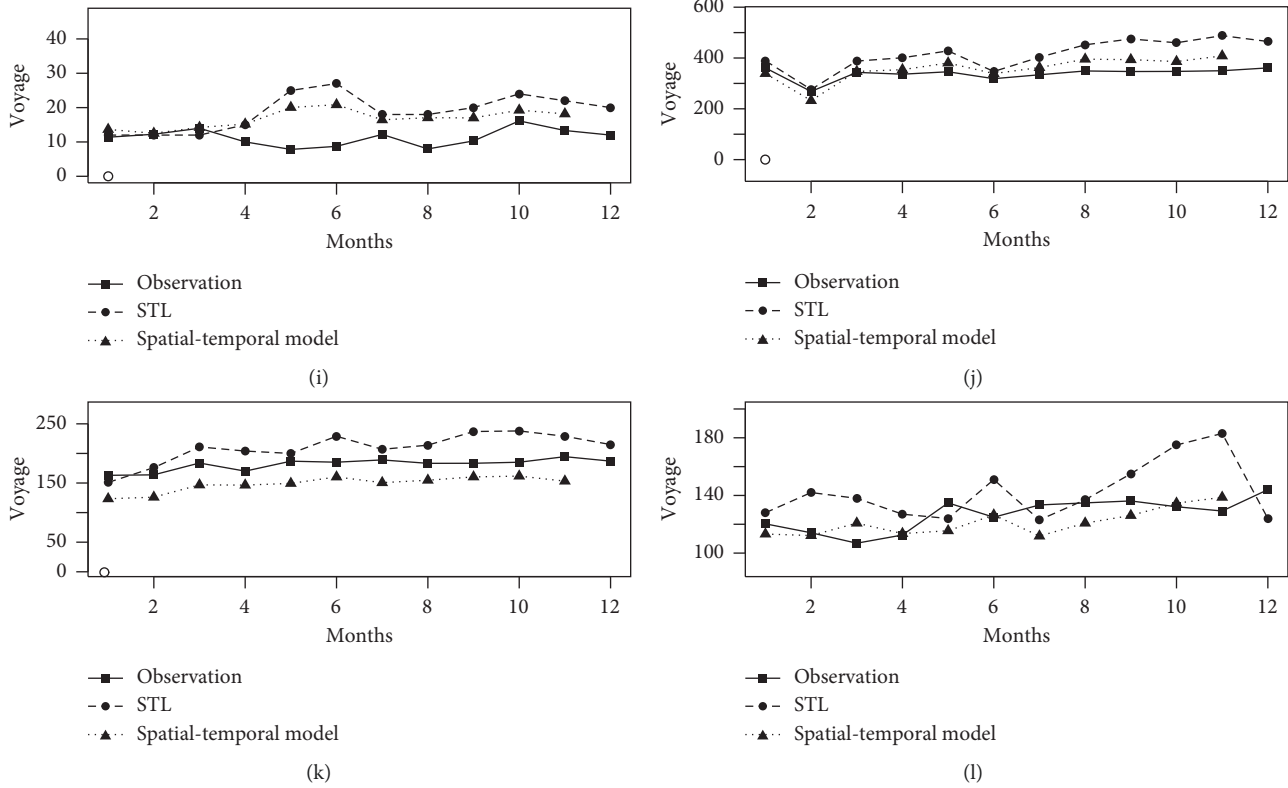
FIGURE 8: Continued.

FIGURE 8: Marine traffic volume forecasting plots of the spatial-temporal model and the STL model. (a) Jeishi, (b) Quanzhou, (c) Kinarut, (d) Hong Kong, (e) Keelung, (f) TigerJuyu, (g) Iloilo, (h) Pinghai, (i) Tarakan, (j) Zhanjiang, (k) Lungsod ng Cebu, and (l) Samarinda.

of the two models, three indicators, namely, Pearson's correlation coefficient (R), root mean square error (RMSE), and mean absolute percentage error (MAPE) are selected. Table 6 shows the evaluation results.

As shown in Table 6, the correlation coefficient R (0.8827) based on the STARMA model is significantly higher than the correlation coefficient (0.4469) of the STL model, indicating that compared with the time-series model, the spatial-temporal model is more likely to capture the periodicity and trend of the monthly traffic volume of the port in the shipping network. The RMSE of the STARMA model is 12.15% higher than the RMSE of the STL model, indicating that the spatial-temporal model is more sensitive to outliers of traffic observations; that is, it is easier to explain spatial-temporal variations. Similarly, the MAPE accuracy of the STARMA model is also better than the MAPE of the STL model, which has an increase of 19.66%, indicating that the forecasting accuracy of the spatial-temporal model is better than that based on the time series.

Through model comparison, it is found that the spatial-temporal model fits better than the time-series model in forecasting the regional shipping network traffic volume and can better explain the spatial-temporal variation in regional shipping networks. The analysis shows that the reason is that it not only extracts the large-scale nonlinear trends in the spatial-temporal series but also captures the small-scale random spatial-temporal variation. Therefore, the spatial-temporal model can improve the generalization accuracy of the fitting to some extent.

5. Conclusion

This study uses ship AIS data from 2011 to 2016 in the waters surrounding the South China Sea to build a regional shipping network based on data mining technology. Then, considering the influence of the correlation between the ports in the shipping network on the traffic volume, a set of closely connected ports is identified by utilizing the topological features of the shipping network and is used to construct the spatial weight matrix of the shipping network in the South China Sea. Finally, the STARMA model was used to derive the interaction between traffic volumes at adjacent ports in the shipping network and to obtain traffic forecasts for ports in the regional shipping network. The conclusions are as follows:

(1) By analyzing the spatial and temporal characteristics of the time-series data of ship traffic in various ports of the South China Sea shipping network, it can be concluded that the maritime traffic volume has certain regularity and difference in both time and space; that is, the port traffic volume in each region shows a certain degree of local autocorrelation in the traditional distance-based weight matrix, and it shows a clear spatial-temporal correlation in the network environment.

(2) Marine traffic has strong regularity. This regularity is not only reflected in the time-series self-correlation characteristics of the port's traffic volume but also in

TABLE 6: Forecasting accuracy of marine traffic volume for the spatial-temporal model and STL model

Port name	R		RMSE		MAPE	
	STARMA	STL	STARMA	STL	STARMA	STL
Tainan	0.7459	0.6444	71.7818	56.2860	7.7058	5.9089
Balikpapan	0.2879	−0.1428	31.5497	7.7862	60.5876	13.5260
Bangkok	0.9751	0.3573	172.3342	108.7648	31.2253	20.1060
Zhangzhou	0.9942	0.9629	20.3072	114.7340	2.5310	13.4685
Manila	0.9377	0.3327	148.2501	228.3151	62.0349	54.8789
Haiphong	0.9506	0.5281	21.2210	227.1680	16.9982	205.8779
. . .	. . .	. . .	. . .	. . .	. . .	. . .
Kuching	0.9723	0.4867	2.2821	4.2561	14.1942	36.2714
Zhanjiang	0.9559	0.8282	55.5359	86.1244	12.4189	17.2552
Lungsod ng Cebu	0.9572	0.7828	61.8685	33.1485	28.4573	13.8344
Samarinda	0.9441	0.5758	25.3620	26.4804	15.0513	14.8198
Mean	0.8827	0.4469	40.7082	46.3403	44.4811	55.3729

the strong correlation between spatially-related ports. The network adjacent rule based on the characteristics of the shipping network community proposed in this paper not only considers the geographical constraints but also takes into account the role of the port in the network and can better capture the spatial correlation of the marine traffic volume than the spatial adjacent rule that considers only the geographical distance.

(3) Combined with the measured marine traffic volume data of the surrounding ports in the South China Sea, the applicability of the ship traffic volume forecasting method based on the spatial-temporal correlation characteristics proposed in this paper is validated. Several evaluation indexes such as R, MAPE, and RMSE are used to evaluate the prediction model's simulation results. The results show that the fitting result of the spatial-temporal model is better than the time-series model, and it can better explain the spatial-temporal variation in the regional shipping network. The above results show that the prediction model proposed in this paper is reasonable and effective. In particular, the method proposed in this paper is to use the data in a period of time as training data to forecast the subsequent traffic volume. Therefore, the method proposed in this paper is also applicable to the forecasting of traffic volume in the future, that is, 2018 and 2019.

In addition, the results obtained from this paper also reflect that the model constructed in this paper reflects the system characteristics and spatial-temporal correlation characteristics of the regional shipping system and summarizes the integrity of the shipping network traffic and the interrelationship of various parts within the system. To a certain extent, this paper reveals the complexity, changing laws, and development trends of regional shipping systems, provides a scientific basis for the scientific planning of the future development scale of ports and the formulation of long-term strategic decisions, and provides a certain supplement for port system research. Since the shipping is a complex and volatile industry and will rise and fall with the global economy, in the medium and long term, it is recommended that each country should analyze the future development trend of the regional shipping system. On this basis, scientifically plan the layout adjustment of domestic ports, and maintain the moderate construction of the port so as to achieve the sustainable development of regional shipping.

This paper starts with the influence of the correlation of the shipping network nodes on the maritime traffic, analyzes the spatial-temporal autocorrelation of the ship traffic volume on the sea, and builds a corresponding forecasting model on this basis. However, deficiencies remain and need to be further explored. This paper mainly considers the impact of node correlation of a shipping network on maritime traffic. The influence of the spatial structure and topological characteristics on maritime traffic needs further study. In addition, different historical data volumes lead to different prediction results, and sensitivity analysis for historical data volume should be further studied.

Data Availability

The AIS data used to support the findings of this study were supplied by Elane Inc. (http://www.elane.com) under license and so cannot be made freely available.

Conflicts of Interest

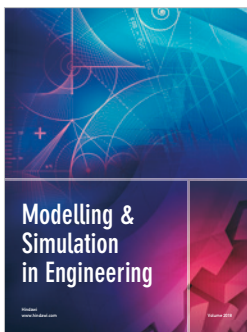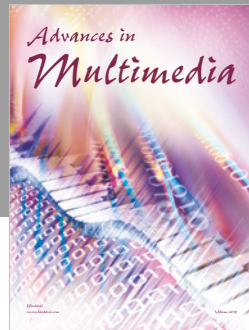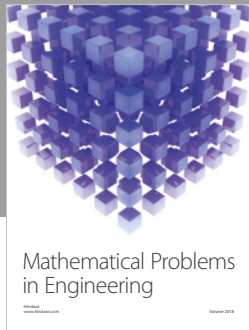The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. F. Wang, Q. L. Dong, Z. M. Peng, S. Khan, and A. Tarasov, "The Green Logistics Impact on International Trade: Evidence from Developed and Developing Countries," *Sustainability*, vol. 10, no. 7, p. 2235, 2018.

[2] R. J. Sánchez, J. Hoffmann, A. Micco, G. V. Pizzolitto, M. Sgut, and G. Wilmsmeier, "Port efficiency and international trade: port efficiency as a determinant of maritime transport costs," *Maritime Economics and Logistics*, vol. 5, no. 2, pp. 199–218, 2003.

[3] M. C. Tan, S. C. Wong, J. M. Xu, Z. R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 60–69, 2009.

[4] Y. Rashed, H. Meersman, E. Van de Voorde, and T. Vanelslander, "Short-term forecast of container throughout: an ARIMA-intervention model for the port of Antwerp," *Maritime Economics and Logistics*, vol. 19, no. 4, pp. 749–764, 2017.

[5] B. Sahin, S. Gürgen, B. Ünver, and Y. Altın, "Forecasting Baltic Dry Index by using artificial neural network approach," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 26, no. 3, 2017.

[6] L. Henesey, "Using systems dynamic modeling for forecasting global container demand," in *Proceedings of International Association of Maritime Economists (IAME)*, Norfolk, VA, USA, 2014.

[7] H. X. Feng, F. C. Kong, and Y. J. Xiao, "Vessel traffic flow forecasting model study based on support vector machine," in *Advanced Research on Electronic Commerce, Web Application, and Communication*, pp. 446–451, Springer, Berlin, Germany, 2011.

[8] X. Wang, N. Zhang, Y. Zhang, and Z. Shi, "Forecasting of short-term metro ridership with support vector machine online model," *Journal of Advanced Transportation*, vol. 2013, article 3189238, 13 pages, 2018.

[9] W. H. K. Lam, P. L. P. Ng, W. Seabrooke, and E. C. M. Hui, "Forecasts and reliability analysis of port cargo throughput in Hong Kong," *Journal of Urban Planning and Development*, vol. 130, no. 3, pp. 133–144, 2004.

[10] M. M. Mostafa, "Forecasting the Suez Canal traffic: a neural network analysis," *Maritime Policy and Management*, vol. 31, no. 2, pp. 139–156, 2004.

[11] B. Yang, "Prediction model for grid resources optimized BP neural network based on genetic algorithm," *Computer Simulation*, vol. 29, pp. 243–246, 2012.

[12] S. Chen and W. Wang, "Grey neural network forecasting for traffic flow," *Journal of Southeast University*, vol. 45, pp. 34–49, 2004.

[13] J. X. Liu, T. Zhang, and W. Liu, "Research on combination forecasting method for ship traffic flow," *Navigation of China*, vol. 32, pp. 80–84, 2009.

[14] V. Gosasang, T. L. Yip, and W. Chandraprakaikul, "Application of forecasting container throughputs for port planning: case of bangkok port," in *International Forum on Shipping, Ports and Airports (IFSPA) 2013: Trade, Supply Chain Activities and Transport: Contemporary Logistics and Maritime Issues*, , 2013.

[15] J. Farhan and G. P. Ong, "Forecasting seasonal container throughput at international ports using SARIMA models," *Maritime Economics and Logistics*, vol. 20, no. 1, pp. 131–148, 2018.

[16] M. Goodchild, *Spatial Autocorrelation*, GeoBooks, Norwich, UK, 1986.

[17] S. Lopes, N. Brondino, and A. Rodrigues da Silva, "GIS-based analytical tools for transport planning: spatial regression models for transportation demand forecast," *Isprs International Journal of Geo-Information*, vol. 3, no. 2, pp. 565–583, 2014.

[18] Y. Wang and K. Cullinane, "Traffic consolidation in East Asian container ports: a network flow analysis," *Transportation Research Part A: Policy and Practice*, vol. 61, pp. 152–163, 2014.

[19] C. Ducruet and I. Lugo, "Structure and dynamics of transportation networks: models," in *The SAGE Handbook of Transport Studies*, Vol. 347, SAGE Publications, Thousand Oaks, CA, USA, 2013.

[20] C. Wang and J. Wang, "Spatial pattern of the global shipping network and its hub-and-spoke system," *Research in Transportation Economics*, vol. 32, no. 1, pp. 54–63, 2011.

[21] C. Kooperberg and F. O'sullivan, "Predictive oscillation patterns: a synthesis of methods for spatial-temporal decomposition of random fields," *Journal of the American Statistical Association*, vol. 91, no. 436, pp. 1485–1496, 1996.

[22] A. S. Hering, "Space-time forecasting and evaluation of wind speed with statistical tests for comparing accuracy of spatial predictions," vol. 43, pp. 1577–1583, Texas A&M University, College Station, TX, USA, 2009, Dissertations and Theses-Gradworks.

[23] W. G. Han, J. F. Wang, Y. G. Gao, and H. U. Jian-Jun, "Forecasting and analysis of regional traffic flow in space and time," *Journal of Highway and Transportation Research and Development*, vol. 24, no. 6, pp. 92–96, 2007.

[24] D. L. Buckeridge, M. A. Musen, P. Switzer, and M. Crubézy, "An analytic framework for space–time aberrancy detection in public health surveillance data," in *Proceedings of AMIA Annual Symposium*, pp. 120–124, Washington, DC, USA, November 2003.

[25] P. E. Pfeifer and S. J. Deutsch, "A STARIMA model-building procedure with application to description and regional forecasting," *Transactions of the Institute of British Geographers*, vol. 5, no. 3, pp. 330–349, 1980.

[26] M. Zachcial and C. Heideloff, *ISL Shipping Statistics Yearbook*, Institute of Shipping Economics and Logistics, Bremen, Germany, 2001.

[27] UNCTAD, "Review of maritime transport 2016," *Review of Maritime Transport*, vol. 13, pp. 251–257, 2016.

[28] D. Huang and W. Peng, "Grid-based DBSCAN algorithm with referential parameters," *Physics Procedia*, vol. 24, pp. 1166–1170, 2012.

[29] X. Z. Zhang, Y. M. Chen, and M. C. Li, "Research on geospatial association of the urban agglomeration around the South China Sea based on marine traffic flow," *Sustainability*, vol. 10, no. 9, p. 3346, 2018.

[30] J. M. Drake and D. M. Lodge, "Global hot spots of biological invasions: evaluating options for ballast-water management," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 271, no. 1539, pp. 575–580, 2004.

[31] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, "The complex network of global cargo ship movements," *Journal of the Royal Society Interface*, vol. 7, no. 48, pp. 1093–1103, 2010.

[32] R. B. Cleveland, "STL: a seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, pp. 3–33, 1990.

[33] C. Bergmeir, R. J. Hyndman, and J. M. Benítez, "Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation," *International Journal of Forecasting*, vol. 32, no. 2, pp. 303–312, 2016.