



OPEN ACCESS

EDITED BY

Sebastian Schnaubelt,
Medical University of Vienna, Austria

REVIEWED BY

Francesco Amenta,
University of Camerino, Italy
Yen-Chiang Chang,
Dalian Maritime University, China

*CORRESPONDENCE

Hua Zhang
✉ zhanghuashelley@hotmail.com

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 14 March 2024

ACCEPTED 13 June 2024

PUBLISHED 27 June 2024

CITATION

Yang P, Cheng P, Zhang N, Luo D, Xu B and Zhang H (2024) Statistical machine learning models for prediction of China's maritime emergency patients in dynamic: ARIMA model, SARIMA model, and dynamic Bayesian network model.

Front. Public Health 12:1401161.

doi: 10.3389/fpubh.2024.1401161

COPYRIGHT

© 2024 Yang, Cheng, Zhang, Luo, Xu and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Statistical machine learning models for prediction of China's maritime emergency patients in dynamic: ARIMA model, SARIMA model, and dynamic Bayesian network model

Pengyu Yang^{1†}, Pengfei Cheng^{2†}, Na Zhang³, Ding Luo³,
Baichao Xu^{4,5} and Hua Zhang^{3*}

¹Department of Nursing, West China Hospital, Sichuan University, Chengdu, China, ²Department of Nursing, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China, ³International Nursing School, Hainan Medical University, Haikou, China, ⁴Department of Physical Education, Hainan Medical University, Haikou, China, ⁵Hainan Provincial Key Laboratory of Sports and Health Promotion, Hainan Medical University, Haikou, China

Introduction: Rescuing individuals at sea is a pressing global public health issue, garnering substantial attention from emergency medicine researchers with a focus on improving prevention and control strategies. This study aims to develop a Dynamic Bayesian Networks (DBN) model utilizing maritime emergency incident data and compare its forecasting accuracy to Auto-regressive Integrated Moving Average (ARIMA) and Seasonal Auto-regressive Integrated Moving Average (SARIMA) models.

Methods: In this research, we analyzed the count of cases managed by five hospitals in Hainan Province from January 2016 to December 2020 in the context of maritime emergency care. We employed diverse approaches to construct and calibrate ARIMA, SARIMA, and DBN models. These models were subsequently utilized to forecast the number of emergency responders from January 2021 to December 2021. The study indicated that the ARIMA, SARIMA, and DBN models effectively modeled and forecasted Maritime Emergency Medical Service (EMS) patient data, accounting for seasonal variations. The predictive accuracy was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) as performance metrics.

Results: In this study, the ARIMA, SARIMA, and DBN models reported RMSE of 5.75, 4.43, and 5.45; MAE of 4.13, 2.81, and 3.85; and R^2 values of 0.21, 0.54, and 0.44, respectively. MAE and RMSE assess the level of difference between the actual and predicted values. A smaller value indicates a more accurate model prediction. R^2 can compare the performance of models across different aspects, with a range of values from 0 to 1. A value closer to 1 signifies better model quality. As errors increase, R^2 moves further from the maximum value. The SARIMA model outperformed the others, demonstrating the lowest RMSE and MAE, alongside the highest R^2 , during both modeling and forecasting. Analysis of predicted values and fitting plots reveals that, in most instances, SARIMA's predictions closely align with the actual number of rescues. Thus, SARIMA is superior in both fitting and forecasting, followed by the DBN model, with ARIMA showing the least accurate predictions.

Discussion: While the DBN model adeptly captures variable correlations, the SARIMA model excels in forecasting maritime emergency cases. By comparing these models, we glean valuable insights into maritime emergency trends, facilitating the development of effective prevention and control strategies.

KEYWORDS

medical assistance at sea, prediction, ARIMA, SARIMA, emergency medical service, dynamic Bayesian network

1 Introduction

Working at sea is one of the world's most perilous occupations, characterized by a markedly high accident rate (1). Surveys reveal that there are 103 injuries per 1,000 full-time fishermen (2). In Australia, the annual fatality rate at sea is three times that of agriculture and four times that of the road freight sector (1). Similarly, in the United Kingdom, the fatal accident rate at sea from 2003 to 2012 was nearly five times higher than in the construction sector (3), highlighting the significant risk and mortality associated with maritime accidents. Delivering emergency care at sea presents considerable challenges exacerbated by factors such as geographic isolation, limited medical personnel, remoteness, and the scarcity of medical facilities aboard ships (4). These conditions render medical emergencies at sea among the most arduous prehospital situations for healthcare professionals and researchers (5). Furthermore, low- and middle-income countries encounter difficulties in providing maritime and aquatic first aid, contributing to a major global public health concern. According to Dykes et al. (6), approximately 19% of all marine search and rescue missions require medical evacuation at sea, significantly impacting patient outcomes due to delays in rescue operations and the absence of immediate responders.

To tackle the aforementioned challenges, the United States has initiated several telemedical maritime assistance service centers and extended telemedicine globally since 2003 to alleviate the shortage of medical personnel at sea (7). Scholars like Adrian P (8) have scrutinized sea and land ambulance response times in the Philippines, devising care strategies and spatial pathways to expedite detection, optimize human resource allocation, and reduce rescue durations. However, dynamically adjusting medical rescue personnel at sea remains an unresolved issue. Zhang et al. (9) have proposed dynamically allocating emergency resources; yet existing models struggle with uncertain data and lack the capability to adapt human resources promptly. The demand for Emergency Medical Services (EMS) at sea fluctuates on a monthly basis, particularly during periods of heightened migrant, refugee, and asylum seeker rescues, or mass casualties from disasters (10–13). Leveraging historical data to forecast future requirements aids in improved scheduling and staffing, thereby enhancing emergency supply reserves for high-risk days. Historical data has proven instrumental in predicting demand and understanding variability, facilitating more effective pre-hospital emergency care planning (14). Accurate prediction of human resource needs enables hospitals to circumvent the expenses associated with hiring temporary staff or implementing flexible schedules, thus reducing response times and resource wastage. The development of a dynamic scheduling system that aligns with patient needs, catering to both full-time and part-time staff, holds immense promise in addressing these challenges.

Hence, the creation of a precise model for predicting emergency patient numbers at sea is paramount for optimizing medical manpower allocation and shaping preventive measures and policies.

Recent research has increasingly embraced advanced technologies such as deep learning, neural networks, and big data for forecasting patient volumes. Key models in this domain include the Auto-regressive Integrated Moving Average (ARIMA), Seasonal Auto-regressive Integrated Moving Average (SARIMA), and Dynamic Bayesian Networks (DBN). ARIMA, renowned for its efficiency in capturing linear trends in time series data with minimal computational burden, is often employed to explore variable relationships or serve as a benchmark in testing hybrid models, albeit with mixed outcomes (15). For instance, Li et al. (16) in China utilized ARIMA to assess the impact of the Corona Virus Disease 2019 (COVID-19) on gonorrhoea trends, while Eyles et al. (17) in the United Kingdom (UK) employed it to generate precise short-term forecasts for patient admissions and bed occupancy, thereby improving predictions for medical specialties and lengths of stay. And ARIMA has proven capable of processing and predicting complex but stable time series data. It can also effectively manage short-term mutations and trends, making it suitable for rapid prediction of emergency events (18). Despite demonstrating high accuracy, ARIMA may struggle with seasonal fluctuations, potentially leading to prediction inaccuracies. SARIMA, on the other hand, excels in analyzing time series periodicity, trends, and disturbances, making it a staple choice for infectious disease forecasting (19–21). Almeida et al. (22) utilized SARIMA to study pediatric emergency department visits, while Zhang et al. (23) applied it to predict hospital blood demand, facilitating resource allocation. Moreover, the maritime environment has significant seasonal characteristics, such as monsoons and tides, so the SARIMA model can better capture and predict these cyclical changes. However, SARIMA's limitations in capturing dynamic inter-variable changes can impede prediction accuracy. In contrast, the DBN model, a graphical representation of variable correlations and temporal changes (24), finds extensive application in infectious disease (25–27) and diabetes research (28) for pathway identification and risk assessment. Emergencies at sea often involve multiple uncertain factors, such as weather changes, wave conditions, ship status, etc. The DBN model can effectively process and integrate this information for real-time dynamic prediction. Despite its utility, DBN typically requires at least 3 years of historical data (25), compared to ARIMA and SARIMA models, which necessitate a minimum of 50 time points (15). Consequently, while DBN models demand less data, they may be susceptible to specific errors. Furthermore, the comparative effectiveness of these models in forecasting emergency patient volumes and their sensitivity to dynamic predictions remain relatively unexplored.

In this study, we delved into the stability and predictive efficacy of three models to analyze the number of maritime emergency

patients. By comparing the fitting and predictive abilities of these models, our goal is to provide an early warning system, facilitating effective prevention and control strategies for maritime emergencies. This includes ensuring timely allocation of manpower and medical resources in anticipation of significant maritime incidents along China's coast. First, this study utilized medical records from the South China Sea region of Hainan Province spanning from 2016 to 2020 to develop ARIMA, SARIMA, and DBN models. These models were evaluated using 2021 data to improve the analysis and prediction of illness trends among offshore emergency patients. The rest of this document is organized as follows: Section 2 outlines the data collection locations, time frames, ethical considerations, inclusion and exclusion criteria, data entry methods, statistical models, their fundamental principles, analysis tools, evaluation criteria, and the overall modeling process. Section 3 elaborates on the modeling procedures of the three models, the selection of parameters, the comparison of their predictive performance, and the identification of the best model. In addition, Section 4 elaborates on why SARIMA excels in predicting the number of first aid workers at sea, the importance of forecasting marine emergencies, the strengths and limitations of this study, and recommendations for future policy. Finally, Section 5 presents the conclusions derived from this research.

2 Subjects and methods

2.1 Subjects

The researcher discovered through initial investigations that the Hainan Maritime Safety Bureau primarily directs sea medical rescues in the South China Sea to five key medical facilities: Haikou City's first aid center handles the Haikou region, while Sanya City, Dongfang City, Wenchang City, and the Yangpu Economic Development Zone each have a dedicated hospital for other areas. Consequently, maritime emergency patients treated at these five Hainan Province hospitals between January 2016 and December 2021 were chosen for this study. Previous research (15, 25) indicates that constructing ARIMA, SARIMA, and DBN models requires a minimum of 50 time points and 3 years of historical data, respectively. Accordingly, this study utilized EMS patient data from January 2016 to December 2020 for model development, and data from January to December 2021 served as an internal validation set to assess the models' predictive accuracy. The Hainan Medical University Ethics Committee approved this study (NO.: HYLL-2022-018), adhering to the Declaration of Helsinki's guidelines. Informed consent was obtained from all participants.

2.2 Data collection

2.2.1 Inclusion criteria

In this study, we examine various cases including: ① 120 instances of emergency vehicles arriving at the dock, port, harbor first aid station, etc.; ② communication with units such as the Maritime Bureau, Coastal Radio Division, port medical aid station, etc.; ③ the nature of calls involving ships affected by natural disasters like typhoons, as well

as other onboard emergencies such as poisoning, requiring immediate medical attention for patients, sudden illnesses or injuries, and the medical history of patients describing injuries or the onset of illnesses onboard, such as cable strangulation injuries, acute gastric perforation, and drowning due to ship sinking or jumping into the sea.

2.2.2 Exclusion criteria

This study excludes: ① 120 instances where emergency vehicles arrived at the scene but did not encounter the patient; ② medical records containing missing information, incomplete data (missing more than 3 items), errors, and duplicates; ③ patients who independently visited the hospital for prescriptions, examinations, or consultations without recorded statistics are also excluded.

2.2.3 Data entry

From the pre-hospital case management system of the five hospitals mentioned above, data of emergency patients at sea in the South China Sea from 2016 to 2021 were exported. If the hospital did not enter into the pre-hospital case management system in that year, the data was collected by manually flipping through records and using image records (taken by the camera) to gather relevant information. Afterward, a database was set up using Excel software to conduct the study, as shown in Table 1.

2.3 Statistical models and description

First, the data were loaded and preprocessed to create trend charts and related factor decomposition charts, and then a time series model was developed to forecast sea emergency patients using three predictive models: ARIMA, SARIMA, and DBN. The prediction framework for the time series model is depicted in Supplementary Figure 1.

2.3.1 ARIMA (p, d, q)

The ARIMA model is commonly represented as ARIMA (p, d, q), where Auto-Regressive (AR) signifies the auto-regressive function, I represents the differencing term, and Moving Average (MA) stands for the moving average function. Here, p indicates the count of autoregressive terms, q represents the number of moving average terms, and d signifies the levels of differencing applied to transform the original dataset into a smoother series. Below are the generalized formulas for the p -order AR model [Equation (1)] and the q -order MA model [Equation (2)].

$$AR(p): Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \varepsilon_t \quad (1)$$

In this sequence, every value can be depicted as a linear combination of its preceding p values. Here, Y_t represents any given observation within the sequence, μ stands for the sequence's average, β denotes the weight, and ε_t signifies the random disturbance.

$$MA(q): Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

where each value of the sequence can be represented as a linear combination of the previous q residuals. ε denotes the predicted residuals and θ is the weight.

TABLE 1 Emergency medical treatment at sea in Hainan region from January 2016 to December 2021.

Month	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
Year												
2016	18	14	20	19	16	13	10	17	17	17	29	20
2017	20	15	15	17	21	11	6	39	18	21	25	18
2018	16	18	16	18	15	7	6	13	21	21	19	19
2019	20	14	20	19	12	3	6	13	17	24	28	27
2020	31	27	29	19	15	9	13	16	25	23	27	22
2021	22	19	22	21	17	8	8	16	20	24	26	25

The ARMA (p, q) model merges the AR model [Equation (1)] with the MA model [Equation (2)], resulting in a unified model expressed mathematically as Equation (3).

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3}$$

where each sequence value is depicted as a linear blend of p previous observations and q residuals.

The ARIMA (p, d, q) model assumes that the time series data is non-stationary. The integration (I) component involves differencing the data to meet the model's requirements for smoothness, allowing further steps in the modeling process. After differencing the data d times, each point in the series is modeled as a linear combination of p past observations and q residuals. This process, which ensures the data adheres to the smoothness criterion, precedes the ARMA (p, q) modeling, represented by the subsequent Equation 4:

$$Y_t' = \mu + \beta_1 Y_{t-1}' + \beta_2 Y_{t-2}' + \dots + \beta_p Y_{t-p}' - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{4}$$

where Y_t' represents the differenced sequence.

2.3.2 SARIMA (p, d, q) (P, D, Q)

Data often exhibit seasonal trends, making a simple ARIMA model inadequate for capturing their correlations. This necessitates employing SARIMA, which further differentiates the data based on the time series' seasonal cycle, typically denoted as SARIMA (p, d, q) (P, D, Q)_[s]. In this notation, P represents the number of seasonal autoregressive terms, D the number of seasonal differencing orders, Q the number of seasonal moving average terms, and s the length of the seasonal cycle. The general SARIMA model is mathematically expressed as follows (Equation 5):

$$\Phi_P(B^m)\phi_P(B)(1-B^m)^D(1-B)^d Y_t = \Theta_Q(B^m)\theta_Q(B)w_t \tag{5}$$

Where Y_t represents the non-stationary time-series, w_t stands for the Gaussian white noise process, $\phi(B)$ denotes the non-seasonal auto-regressive polynomial, and $\theta(B)$ signifies the non-seasonal moving average polynomial. Additionally, D represents the seasonal

differencing term, which can adopt the values 1 or 2, among others. However, the value of $D=1$ effectively guarantees the data's stationarity. Furthermore, $\Phi_P(B^m)$ symbolizes the seasonal auto-regressive polynomial, while $\Theta_Q(B^m)$ represents the seasonal moving average polynomial. In this context, B is defined as the backshift operator, described as follows (Equation 6):

$$B^k Y_t = Y_{t-k} \tag{6}$$

2.3.3 DBN model

The Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm stands out as a method for variable selection, known for its high model stability. It systematically reduces coefficients by incorporating penalty terms during model estimation, streamlining the model and addressing overfitting and multicollinearity effectively (29). Tibshirani (30) is renowned for pioneering the LASSO approach, enabling simultaneous variable selection and coefficient estimation. The LASSO estimate is identified as the solution to Equation 7.

$$\min_{\beta} (\mathcal{Y} - \mathcal{X}\beta)'(\mathcal{Y} - \mathcal{X}\beta) + \lambda \sum_{j=1}^k |\beta_j|, \lambda \geq 0 \tag{7}$$

Least Absolute Shrinkage and Selection Operator regression becomes equivalent to ordinary least square (OLS) regression when the tuning parameter λ is set to 0. As the tuning parameter λ grows, it progressively reduces the magnitude of the unknown regression parameter vector β toward 0, leading to some regression parameters being precisely reduced to 0 for sufficiently high values of λ . The predictors linked to these zero-valued regression parameters are deemed inactive and are consequently excluded from the model.

The configuration and parameters of a DBN are typically derived by employing the LASSO algorithm (25). This method operates on the principle that, by introducing an L1 penalty to the least squares minimization function, it is possible to reduce the coefficients of less correlated variables to zero. This process refines the model by eliminating irrelevant variables, leaving only significant variables represented by directed arcs to the response variables. On the other hand, Bayesian Networks (BNs), a type of directed acyclic graph (DAG), depict the relationships between variables using nodes and arcs, with nodes symbolizing random variables and arcs illustrating the interactions among them. Given that complex systems evolve over time, static BN models fall short

in capturing these dynamics, leading to the development of DBNs. These models extend BNs by incorporating time, allowing for the analysis of time series data through arcs that connect variables at successive time points. This feature enables DBNs to capture both the interactions among variables and how these relationships evolve over time, offering a comprehensive framework for understanding dynamic systems. The calculation method is shown in Equation 8.

$$X_t = u_t + A_1 X_{t-1} + \dots + A_p X_{t-p} + a_t \tag{8}$$

Let $(t = 1, 2, \dots, T)$ where $X_t = (X_{i(t)})$, $(i = 1, \dots, k)$, represents (k) as the vector of observations of the number of first aiders at sea at time (t) . (A) is the matrix of coefficients to be estimated, and X_{t-p} represents the vector of observations of the number of first aiders at sea at lag order (p) . u_t denotes the constant term, and a_t represents the residuals. All arcs in the network are defined between two consecutive time points, with the set of arcs represented by the matrix $A_c (1 \leq c \leq t)$. If an element $a_{ij} \neq 0 (i \neq j)$ exists in A_c , then the network includes an arc from $X_{i(t-c)} X_{j(t)}$.

2.4 Analytical tools and model evaluation

The following tools are employed to evaluate the reliability of time series analysis: Auto-Correlation Function (ACF), Partial Auto-Correlation Function (PACF), Augmented Dickey-Fuller Test (ADF test), Akaike's Information Criterion (AIC), and Ljung-Box Test (LB test). These tools help identify the relationships between observations in a time series. ACF measures the correlation between the time series data and its previous values, whereas PACF identifies the correlation of the time series with its lagged values by specific time intervals. The ADF test assesses the smoothness of the time series, with the trend term parameter indicating the significance of the trend; a non-smooth series requires differencing before applying the ADF test. AIC serves as a penalized likelihood criterion, with a lower AIC value suggesting a more plausible model. The LB test checks if the series is purely random, using the p value to determine the presence of white noise.

The metrics employed in this study included Relative Error (RE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), along with the coefficient of determination (R^2).

2.4.1 ACF and PACF

Auto-correlation refers to the relationship between a current observation and one from a previous time step (lag) within time series data. An auto-correlation plot, which graphs auto-correlation against lag, visualizes this relationship. The ACF quantifies the linear connection between an observation at time t and another at time $t + k$. This concept is encapsulated in the time series Equation 9 as follows:

$$ACF(\mathcal{Y}_t, \mathcal{Y}_{t+k}) = \rho(\mathcal{Y}_t, \mathcal{Y}_{t+k}) = \frac{Cov(\mathcal{Y}_t, \mathcal{Y}_{t+k})}{Var(\mathcal{Y}_t)} \tag{9}$$

Where k represents the lag, it is defined by \mathcal{Y}_t and \mathcal{Y}_{t+k} . Lag k auto-correlation describes the correlation between observations that are k time periods apart. Conversely, partial auto-correlation measures the correlation between the time series and its lags, but

only after removing the influence of observations in between. Essentially, PACF "subtracts" the correlation that earlier lags have already accounted for. This feature is crucial for identifying the appropriate order for the AR model, as illustrated in the Equation 10.

$$PACF(\mathcal{Y}_t, \mathcal{Y}_{t+k}) = Corr(\mathcal{Y}_t - \mathcal{Y}_t'', \mathcal{Y}_{t+k} - \mathcal{Y}_{t+k}'') \tag{10}$$

\mathcal{Y}_t'' represents the estimated value of \mathcal{Y}_t , derived through linear regression analysis of variables $\mathcal{Y}_{t-1}, \mathcal{Y}_{t-2}, \dots, \mathcal{Y}_{t-k+1}$. Similarly, \mathcal{Y}_{t+k}'' signifies the estimated value of \mathcal{Y}_{t+k} , calculated using linear regression based on $\mathcal{Y}_{t+k-1}, \mathcal{Y}_{t+k-2}, \dots, \mathcal{Y}_t$. $Corr(\mathcal{Y}_t - \mathcal{Y}_t'', \mathcal{Y}_{t+k} - \mathcal{Y}_{t+k}'')$ denotes the correlation coefficient that measures the relationship strength between $\mathcal{Y}_t - \mathcal{Y}_t''$ and $\mathcal{Y}_{t+k} - \mathcal{Y}_{t+k}''$. $PACF(\mathcal{Y}_t, \mathcal{Y}_{t+k})$ is defined as the partial auto-correlation coefficient for a lag of k periods.

2.4.2 ADF test

The unit root test stands as the principal technique for assessing the smoothness of time series data. Within the framework of the ADF test, it is essential to conduct a regression analysis to derive the coefficients α , β , and γ , along with the residual error's variance, followed by a unit root test on the residual term. Absence of a unit root in this term indicates that the time series is smooth, leading to the rejection of the null hypothesis. Conversely, presence of a unit root signifies a non-smooth time series, necessitating the acceptance of the null hypothesis. In practical applications, the ADF test's p value is frequently employed to gauge the time series' smoothness. A p value below the significance threshold (commonly set at 0.05 or 0.01) leads to the rejection of the null hypothesis, indicating that the time series is indeed smooth. The Formula 11 is described as follows:

$$\mathcal{Y}_t - \mathcal{Y}(t+k) = \alpha \mathcal{Y}(t+k) + \beta_t + \gamma + \epsilon t \tag{11}$$

Where \mathcal{Y}_t represents the initial observation in the time series, $\mathcal{Y}(t+k)$ denotes a subsequent observation, α is the regression coefficient, β refers to the coefficient of the time trend, γ stands as the intercept (the point at which a line crosses the x- or y-axis), and γ signifies the error term (in statistics).

2.4.3 Akaike's information criterion

To elucidate the connections among variables, it is crucial to assess the generated models for their performance efficacy. This evaluation employs insightful criteria to gauge a model's ability to delineate a relationship accurately. The primary metric used is the AIC, which evaluates model quality by rewarding models that minimize errors and penalizing those with excessive parameters. AIC is mathematically formulated as follows (Equation 12).

$$AIC = -2LL + 2K \tag{12}$$

LL represents the log-likelihood function, and K stands for the number of parameters. Within the AIC framework, a lower score signifies the optimal model, characterized by a greater likelihood value. This aids the time series analyst in choosing the most suitable model from a restricted pool of possible models.

2.4.4 LB test

The LB test evaluates the presence of auto-correlation in a time series, examining not only white noise in the residuals but also testing for complete randomness, as implied by the nature of white noise, where residuals are uncorrelated. The test uses the p value to assess if the series is independent, random, and exhibits white noise characteristics. A p value below 0.05 leads to the rejection of the null hypothesis, suggesting the series is correlated; whereas a p value of 0.05 or higher indicates acceptance of the null hypothesis, portraying the series as an uncorrelated white noise sequence. This is represented in Equation (13).

$$Q = n(n + 2) * \sum \frac{(r^k)^2}{(n - k)} \tag{13}$$

Where n represents the sample size, r stands for the auto-correlation coefficient of the residual series, and k denotes the order of the auto-correlation coefficient. By referring to the critical values of Q and the degrees of freedom, one can conduct hypothesis testing to ascertain the presence of auto-correlation in the residual series.

2.4.5 Evaluation metrics

RE, MAE, RMSE, and R^2 are frequently utilized to assess the accuracy of the proposed model, as described by Equations (14–17). The R^2 value varies between 0 and 1, with values closer to 1 indicating a better model fit. Conversely, for RE, a smaller RMSE and MAE suggest a more accurately fitted model.

$$R^2 = 1 - \frac{\sum (\mathcal{Y}_{,i} - \widehat{\mathcal{Y}}_i)^2}{\sum (\mathcal{Y}_{,i} - \bar{\mathcal{Y}})^2} \tag{14}$$

$$RE = \left| \frac{(\mathcal{Y}_{,i} - \widehat{\mathcal{Y}}_i)}{\widehat{\mathcal{Y}}_i} \right| * 100\% \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\widehat{\mathcal{Y}}_i - \mathcal{Y}_{,i})^2} \tag{16}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\widehat{\mathcal{Y}}_i - \mathcal{Y}_{,i}| \tag{17}$$

Where $\mathcal{Y}_{,i}$ denotes the actual value, $\widehat{\mathcal{Y}}_i$ signifies the forecasted value, $\bar{\mathcal{Y}}$ stands for the average value, and n indicates the sample size.

2.5 Statistical methods

Data analyses were conducted using R version 4.3.1, utilizing the “forecast,” “tseries,” “lars,” and “bnlearn” packages. Hypothesis testing

was performed with a significance level of 0.05. For the analysis steps, refer to Figure 1.

2.5.1 ARIMA and SARIMA modeling

1. Employ the ADF unit root test to determine if the maritime emergency time series data is smooth or stationary, differencing any non-smooth information.
2. Generate auto-correlation and partial correlation plots of the maritime emergency time series to initially evaluate appropriate model parameters.
3. Fit possible ARIMA/SARIMA models according to step 2 and select the optimal model. Use the LB test to determine whether the residuals of the optimal model are white noise, and only the model whose residuals are white noise is a valid model.
4. Predictions utilized the optimal model, calculating RE, RMSE, MAE, and R^2 to compare the models’ predictive efficiency.

2.5.2 DBN modeling

1. Assess the smoothness of the maritime emergency time series data. If non-smooth, apply differencing before model fitting.
2. Determine the lag order of the DBN based on data characteristics. Utilize the LASSO algorithm to learn DBN parameters and structure.
3. The potential DBNs were adjusted as per step 2, and the best model was chosen. Subsequently, the LB test was applied to check if the residuals of the chosen model exhibit white noise characteristics, thereby confirming the model’s validity.
4. Utilize the optimal model for predictions. Calculate the RE, RMSE, MAE, and R^2 to evaluate model performance. Compare the predictive effectiveness of the models.

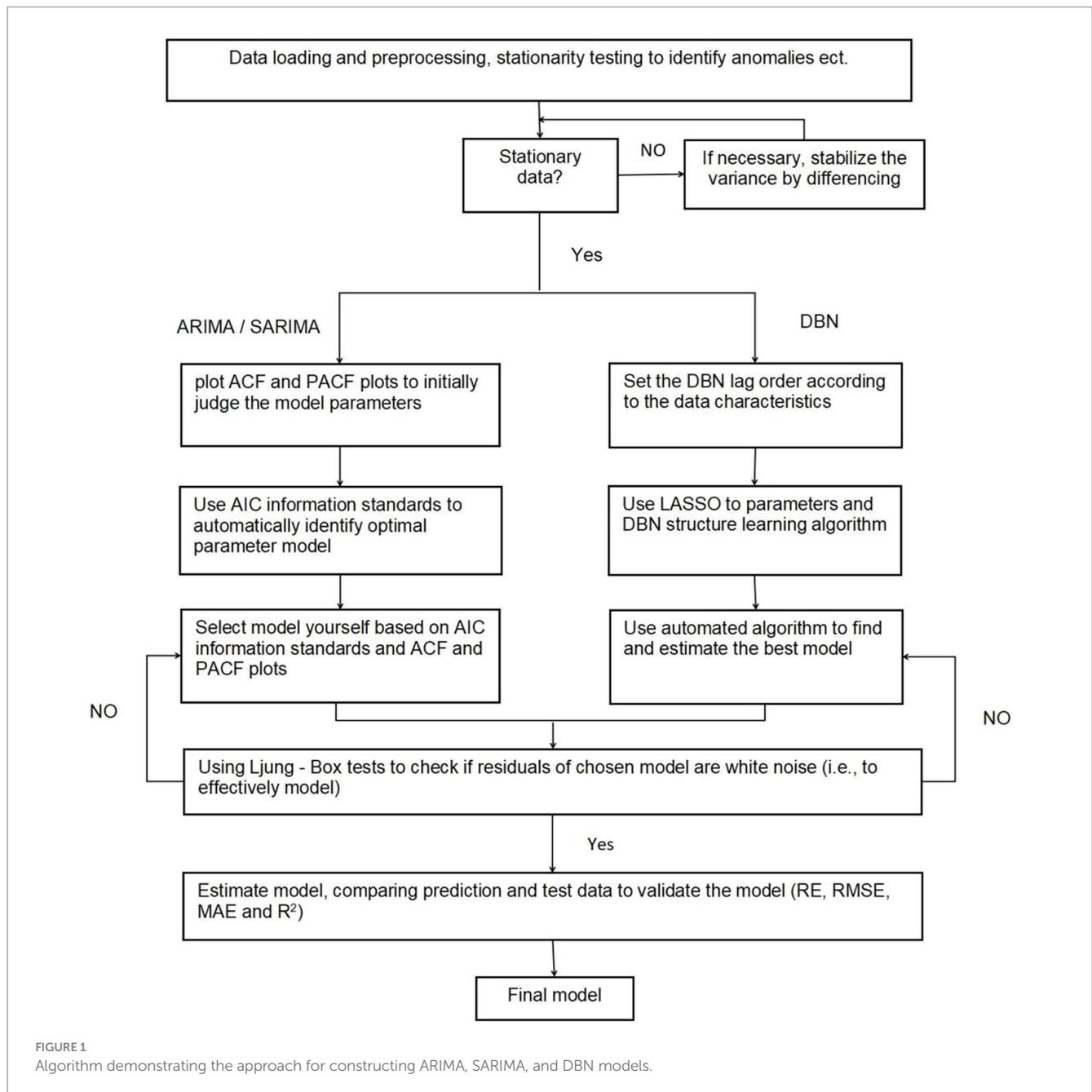
3 Results

3.1 Descriptive analysis

From January 2016 to December 2021, five hospitals in Hainan Province documented 1,312 maritime emergencies, with an average of 18.22 incidents per month. The process of case selection is outlined in Figure 2. The 60 time points of maritime first aid from January 2016 to December 2020, as detailed in Table 1, were inputted into R and depicted in a trend chart. Figure 4 displays the fluctuations in the number of maritime first aid incidents in Hainan Province during this timeframe. The data were preprocessed to decompose a series of factors influencing changes in maritime first aid in the South China Sea, as depicted in Supplementary Figure 2. Figure 4 and Supplementary Figure 2 present the time series analysis of maritime first responders in Hainan, unveiling a noticeable seasonal pattern. October, November, and December emerge as peak months for rescue and treatment services, while June and July experience the lowest activity, likely attributed to the summer season.

3.1.1 ARIMA modeling

The unit root test for the count of sea-based first responders from January 2016 to December 2020 yields $ADF = 0.01$, $p \leq 0.01$, confirming the series’ stationarity. Figures 3, 5 illustrate the ACF’s



sinusoidal trend and the PACF’s initial spike, followed by diminishing lags, suggesting an ARIMA (1, 0, 0) model. However, model selection can be subjective. Objectively, the AIC criterion prefers the ARIMA (1, 0, 0) model, with the lowest AIC at 462.8274 (see Table 2) and an R^2 value of 0.21. The LB test result, $P_{LB\text{ test}} = 0.8695$, exceeding 0.05, indicates the residuals are white noise, validating the model’s predictive accuracy for the number of sea-based first responders.

3.1.2 SARIMA modeling

The data remains consistent, yet the time series graph reveals a distinct seasonal trend. As a result, we implemented first-order seasonal differencing. The SARIMA model is expressed as: ARIMA (p, d, q) (P, D, Q)_[s], where s denotes the seasonal cycle, (p, d, q) represents the non-seasonal component of the model, and (P, D, Q)_[s] denotes the seasonal part of the model. The seasonal cycle is 12 months

per year (evident from the spikes at lags 12 in the ACF plot after differentiating the SARIMA model in Figure 6), hence $s = 12, D = 1$. By examining the auto-correlation and partial correlation of the maritime EMS trips time series (Figures 6, 7) and using the AIC for model selection, we identified SARIMA(1,0,0)(0,1,1)₁₂ as the optimal model. The AIC’s parameter estimation results are depicted in Figure 8. This model boasts an R^2 value of 0.54, and the LB test, with a p value of 0.6729—above the 0.050 threshold—verifies that the residuals are white noise, underscoring the model’s precision and reliability in predicting the number of maritime first-aiders.

3.1.3 DBN modeling

The original series is renowned for its fluidity. Consequently, the DBN model was precisely tailored with a maximum lag of 12 to accommodate seasonality. We utilized 12 sequences of lagged morbidity numbers,

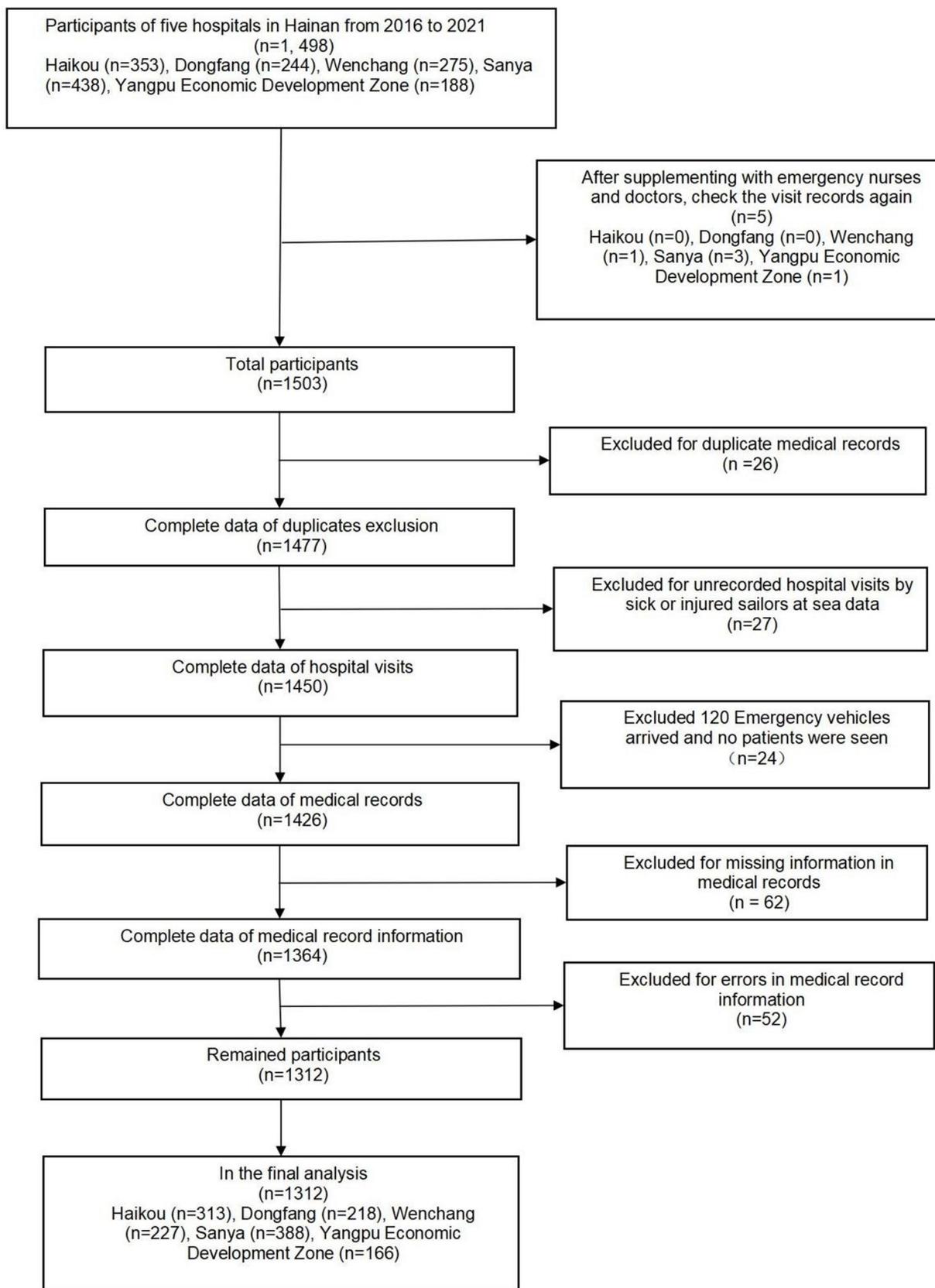


FIGURE 2
Flowchart of case screening.

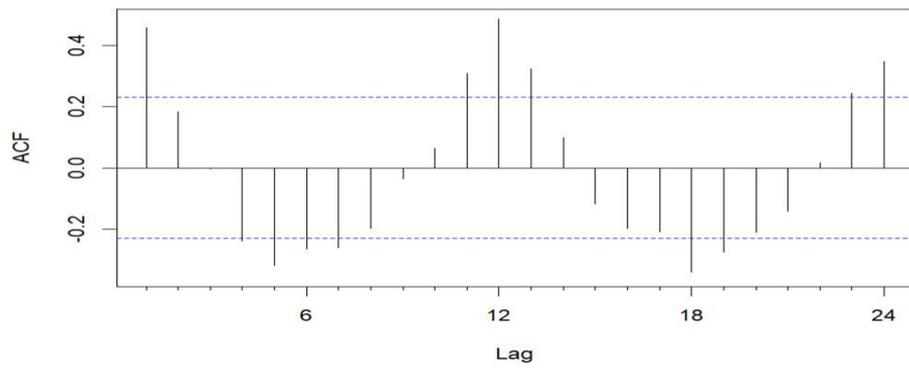


FIGURE 3
Sequential ACF chart of the number of emergency cases at sea from January 2016 to December 2020.

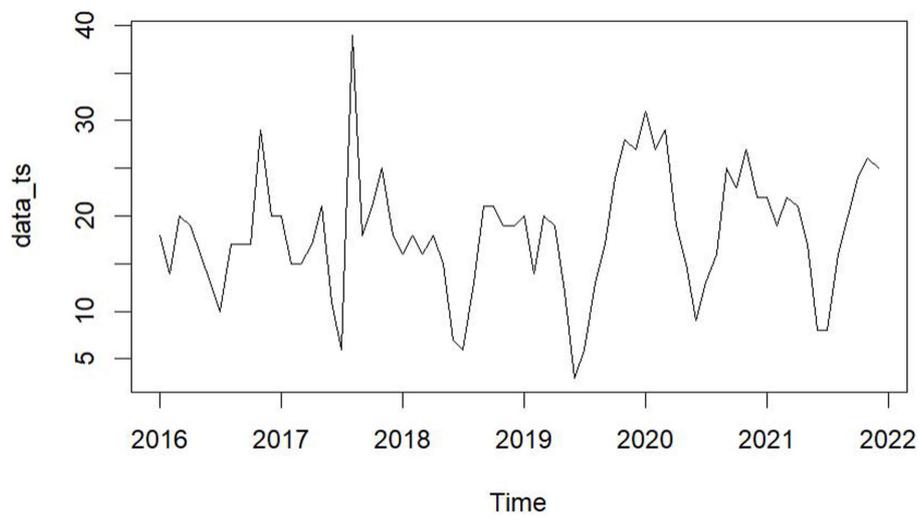


FIGURE 4
Sequence of maritime EMS attendances from January 2016 to December 2021.

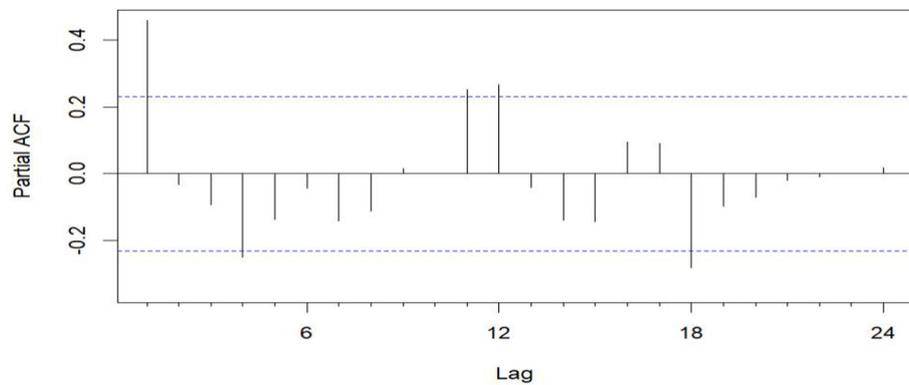


FIGURE 5
Sequential PACF chart of the number of EMS cases at sea from January 2016 to December 2020.

ranging from 1 to 12, as input. The LASSO algorithm facilitated parameter estimation, with outcomes detailed in Table 3. The nonzero coefficients of the 12 lagged morbidity numbers suggest a significant correlation with the current period's maritime first-aiders count. This relationship is mirrored in the DBN's structure, illustrated in Figure 9. The DBN's R^2 stood at 0.44. Moreover, the LB test yielded a p value of 0.708, surpassing

the 0.050 threshold, with the residuals classified as white noise, affirming the model's predictive accuracy for the number of sea first responders.

3.1.4 Comparison of three model fitting and prediction effects

The ARIMA, SARIMA, and DBN models were utilized for forecasting. Table 4 showcases the forecasted outcomes from each model, with SARIMA's forecasts more accurately mirroring actual rescues than those of ARIMA and DBN. Following this, the RMSE and MAE for each model were computed and are detailed in Table 5. The R^2 scores for ARIMA, SARIMA, and DBN range from 0 to 1, with SARIMA's score surpassing 0.5, signifying a robust model fit. Hence, SARIMA demonstrates superior performance over the ARIMA and DBN models. Specifically, SARIMA's RMSE and MAE saw reductions of 22.96 and 31.96%, respectively, in comparison to ARIMA. Additionally, compared to ARIMA, SARIMA's RMSE and MAE fell by 18.72 and 27.01%, respectively, whereas DBN's RMSE and MAE saw smaller declines of 5.22 and 6.78%, respectively. Figure 10 illustrates the fitting and prediction plots, indicating that SARIMA's fitting and forecast trajectories more closely match the actual data.

TABLE 2 ARIMA model least squares parameter estimation.

Model	AIC
ARIMA (2,0,2)	469.6173
ARIMA (0,0,0)	477.6239
ARIMA (1,0,0)	462.8274
ARIMA (0,0,1)	465.1882
ARIMA (0,0,0)	632.9414
ARIMA (2,0,0)	465.0171
ARIMA (1,0,1)	465.0313
ARIMA (2,0,1)	Inf
ARIMA (1,0,0)	483.6405

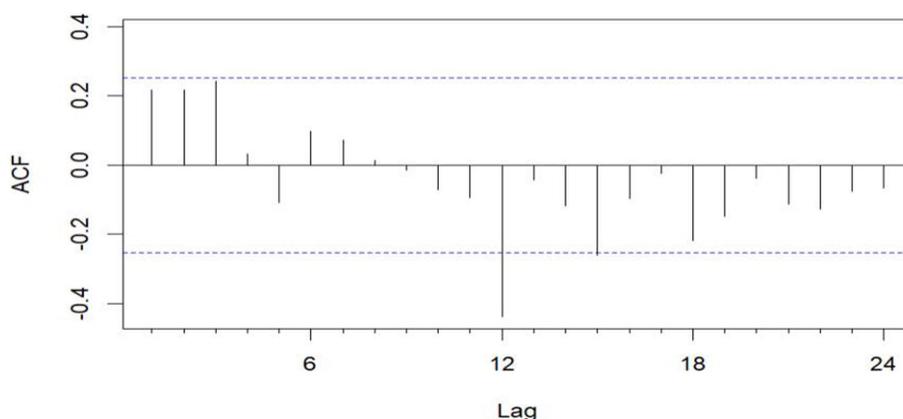


FIGURE 6 ACF plot after differential.

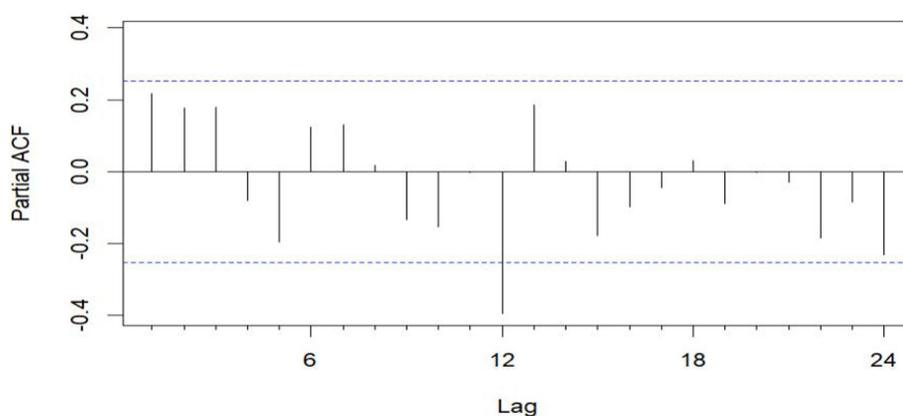


FIGURE 7 PACF plot after differential.

```

> auto.arima(data_ts, seasonal=T, trace=T)

ARIMA(2,0,2)(1,1,1)[12] with drift           : Inf
ARIMA(0,0,0)(0,1,0)[12] with drift         : 397.5069
ARIMA(1,0,0)(1,1,0)[12] with drift         : 381.0792
ARIMA(0,0,1)(0,1,1)[12] with drift         : 377.2092
ARIMA(0,0,0)(0,1,0)[12]                   : 395.4964
ARIMA(0,0,1)(0,1,0)[12] with drift         : 397.569
ARIMA(0,0,1)(1,1,1)[12] with drift         : Inf
ARIMA(0,0,1)(0,1,2)[12] with drift         : Inf
ARIMA(0,0,1)(1,1,0)[12] with drift         : 382.2681
ARIMA(0,0,1)(1,1,2)[12] with drift         : Inf
ARIMA(0,0,0)(0,1,1)[12] with drift         : 377.5367
ARIMA(1,0,1)(0,1,1)[12] with drift         : 377.3095
ARIMA(0,0,2)(0,1,1)[12] with drift         : 378.2873
ARIMA(1,0,0)(0,1,1)[12] with drift         : 376.392
ARIMA(1,0,0)(0,1,0)[12] with drift         : 396.8734
ARIMA(1,0,0)(1,1,1)[12] with drift         : Inf
ARIMA(1,0,0)(0,1,2)[12] with drift         : Inf
ARIMA(1,0,0)(1,1,2)[12] with drift         : Inf
ARIMA(2,0,0)(0,1,1)[12] with drift         : 377.131
ARIMA(2,0,1)(0,1,1)[12] with drift         : 379.2877
ARIMA(1,0,0)(0,1,1)[12]                   : 375.3083
ARIMA(1,0,0)(0,1,0)[12]                   : 394.7528
ARIMA(1,0,0)(1,1,1)[12]                   : Inf

ARIMA(1,0,0)(0,1,2)[12]                   : 377.5687
ARIMA(1,0,0)(1,1,0)[12]                   : 379.2181
ARIMA(1,0,0)(1,1,2)[12]                   : 379.9642
ARIMA(0,0,0)(0,1,1)[12]                   : 377.0919
ARIMA(2,0,0)(0,1,1)[12]                   : 375.7222
ARIMA(1,0,1)(0,1,1)[12]                   : 375.8081
ARIMA(0,0,1)(0,1,1)[12]                   : 376.3112
ARIMA(2,0,1)(0,1,1)[12]                   : 377.6999

Best model: ARIMA(1,0,0)(0,1,1)[12]

Series: data_ts
ARIMA(1,0,0)(0,1,1)[12]

Coefficients:
      ar1      sma1
    0.2539  -0.7293
s.e. 0.1247  0.1938

sigma^2 = 24.43: log likelihood = -184.44
AIC=374.88  AICc=375.31  BIC=381.16

```

FIGURE 8
Results of AIC parameter estimation.

4 Discussion

In our study, we constructed several forecasting models, including ARIMA, SARIMA, and DBN, to predict the monthly influx of maritime patients treated by hospitals across Hainan Province. By analyzing the treatment data of patients from the South China Sea region, we evaluated and compared the predictive accuracy of these models. Our results reveal that the SARIMA (1, 0, 0) (0, 1, 1)₁₂ model outperforms others in forecasting accuracy.

In our study, we analyzed data on maritime emergency patients rescued in the South China Sea, specifically within the Hainan region, from 2016 to 2021. This data, totaling 1,312 rescues, aligns closely with figures reported by the Hainan Maritime Bureau, encompassing police responses and search and rescue operations (31). Our dataset comprises notable events that influenced these figures, including the onset of the COVID-19 epidemic in 2019, Typhoon Pigeon in August 2017, and Super Typhoon Rey in late December 2021. These occurrences resulted in a surge in rescue operations, consistent with our collected data. This correlation, particularly in light of the unpredictable nature of such disasters,

emphasizes the accuracy of our predictive model and the objective, realistic selection of our data.

Aside from occasional spikes in Maritime EMS volume resulting from major disasters, there are noticeable seasonal fluctuations and trends. Peak periods for rescue and treatment operations typically occur during the winter months of October, November, and December, while quieter months are observed in June and July. Similar seasonal patterns have been identified in other studies (32, 33). However, both

the DBN model and the time series model effectively capture these seasonal fluctuations and trends in rescue numbers, serving as dynamic tools for forecasting data with periodic characteristics.

Prior research suggests that at least 3 years of historical data are necessary to meet sample size requirements (25), as insufficient data may diminish auto-correlation and hinder the extraction of periodic features, ultimately impacting prediction accuracy (34). This study relies on authentic, objective, and reliable medical records from Hainan hospitals specializing in maritime emergency care. The dataset spans 5 years and 60 months from 2016 to 2020, constituting a complete cycle and meeting the prerequisites for modeling. To the best of our knowledge, this is the first study to apply a DBN model and time series analysis to predict maritime EMS dispatches in China.

To enhance the management of first aid responders at sea, it is crucial to improve sea rescue prevention and control strategies, necessitating cooperation from all stakeholders. Accurately predicting emergency incidents at sea is vital for effective sea rescue operations. This study aims to provide predictive analytics for managing maritime first aid in the South China Sea. After extensive analysis, we identified the ARIMA (1, 0, 0) and SARIMA (1, 0, 0) (0, 1, 1)₁₂ models as optimal. By developing the DBN model and refining it through iterative debugging based on data characteristics, we established a DBN model with a maximum lag of 12 orders, starting with a non-zero lag order coefficient. Comparing the ARIMA (1, 0, 0) and SARIMA (1, 0, 0) (0, 1, 1)₁₂ models' fitting capabilities with the DBN model, using predicted values, RE, RMSE, MAE, and R², revealed that the SARIMA (1, 0, 0) (0, 1, 1)₁₂ model offers more precise predictions, lower RE, and better

TABLE 3 Results of DBN modeling parameter estimation.

Lag	Coefficient
X_t1	0.31678192
X_t2	0.19788793
X_t3	-0.05581918
X_t4	0.06195344
X_t5	-0.10013309
X_t6	-0.04453516
X_t7	0.02913787
X_t8	-0.08300831
X_t9	-0.07589329
X_t10	0.04222040
X_t11	0.04748903
X_t12	0.28105241

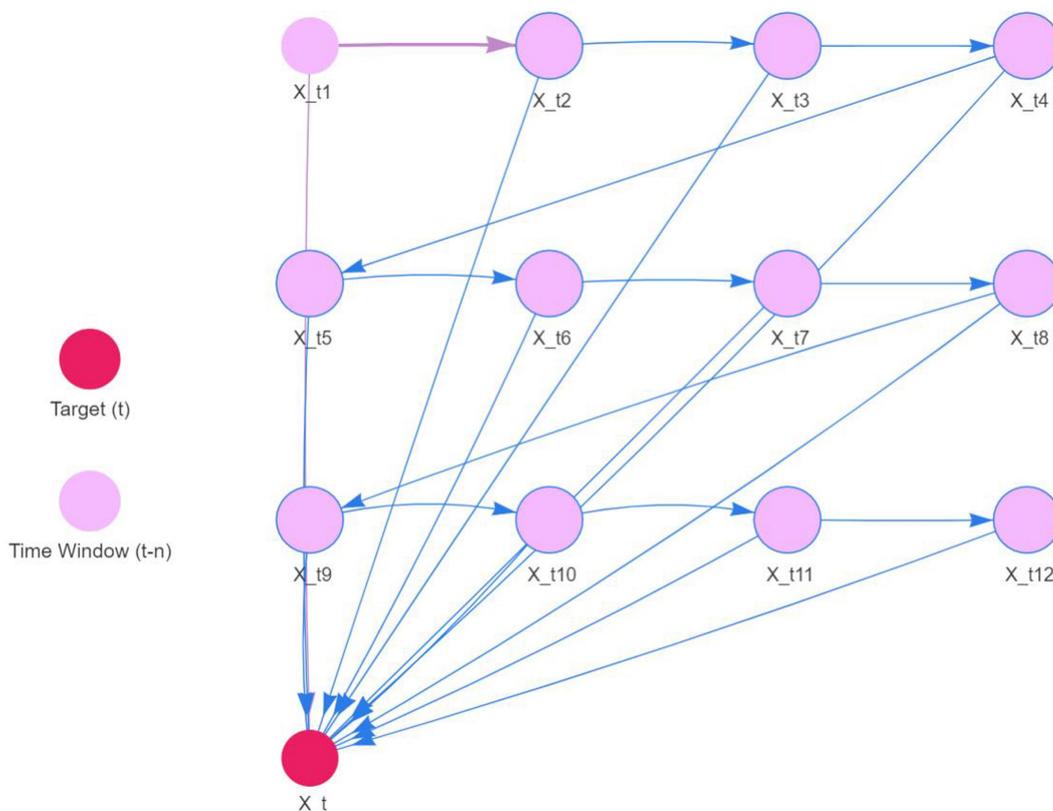


FIGURE 9 DBN model of first aid trips at sea.

TABLE 4 ARIMA, SARIMA, and DBN projections of the number of patients in maritime emergencies in Hainan from January 2021 to December 2021.

Month	Actual	ARIMA	RE	SARIMA	RE	DBN	RE
Jan.	22	20	9.09%	22	0.00%	26	18.18%
Feb.	19	20	5.26%	19	0.00%	25	31.58%
Mar.	22	19	13.64%	21	4.55%	22	0.00%
Apr.	21	20	4.76%	19	9.52%	19	9.52%
May	17	20	17.65%	16	5.88%	16	5.88%
Jun.	8	28	250.00%	9	12.50%	14	75.00%
Jul.	8	14	75.00%	9	12.50%	12	50.00%
Aug.	16	14	12.50%	18	12.50%	15	6.25%
Sep.	20	17	15.00%	20	0.00%	19	5.00%
Oct.	24	19	20.83%	22	8.33%	23	4.17%
Nov.	26	21	19.23%	26	0.00%	25	3.85%
Dec.	25	22	12.00%	22	12.00%	24	4.00%

TABLE 5 Evaluation metrics for ARIMA, SARIMA, and DBN model fitting and prediction accuracy.

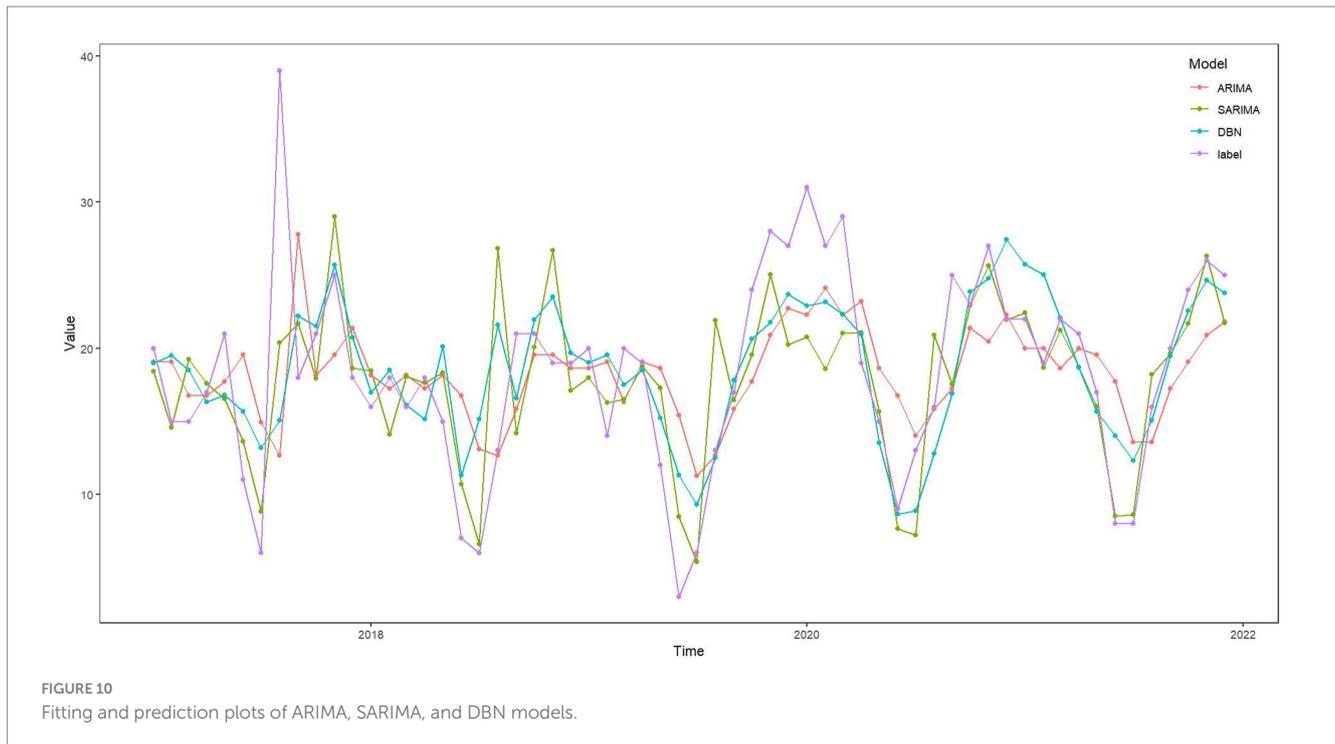
Model	RMSE	MAE	R ²
ARIMA	5.75	4.13	0.21
SARIMA	4.43	2.81	0.54
DBN	5.45	3.85	0.44
Percentage decrease in SARIMA over ARIMA	22.96%	31.96%	-157.14%
Percentage decrease in SARIMA over DBN	18.72%	27.01%	-22.72%
Percentage decrease in DBN over ARIMA	5.22%	6.78%	-109.52%

RMSE, MAE, and R² scores, outperforming the DBN model. The ARIMA (1,0,0) model showed lesser fitting capabilities. Therefore, the SARIMA (1,0,0)(0,1,1)₁₂ model is more effective for forecasting future maritime EMS operations, leading to more accurate EMS trend forecasts.

Numerous studies indicate that DBN methods deliver strong performance in predictive analytics (25, 35), yet their application in the medical sector, including areas like infectious diseases, outpatient volume, and emergency care, remains limited. DBN, a graphical model, effectively represents fitting results through network graphs. DBN estimation employs various machine learning algorithms, notably the LASSO algorithm in this research, along with James-Stein shrinkage estimation and first-order conditional dependence approximation (30, 36, 37). Despite this, the SARIMA method is also recognized for its robust fitting and predictive capabilities (34, 38, 39). Given our data’s specific characteristics, we found SARIMA to surpass both the DBN and ARIMA methods in performance. The SARIMA model excels at capturing and predicting seasonal components, making it effective for handling abnormal and fluctuating data with periodicity (40). This capability is particularly valuable for maritime emergency rescue data, which inherently follows a cyclical pattern. In contrast to the high computational demands and complexity of DBN (41), SARIMA offers high prediction accuracy at lower computational costs, making it well-suited for various real-time and near-real-time applications. While ARIMA struggles with seasonal data (15), SARIMA stands out for its ability to make precise predictions in complex datasets without the need for intricate assumptions or extensive prior knowledge. As illustrated in Figure 10, the comparison of the three models

demonstrates how SARIMA (1,0,0)(0,1,1)₁₂ adeptly fits historical data, offering a reliable forecast for the number of EMS sea trips.

Forecasting maritime emergency visits is crucial for managing healthcare in coastal areas. Emergency department visits serve as a vital measure of workload and the quality of care provided. Overcrowding occurs when the demand from patients surpasses the available resources during peak times (33). Thus, it is imperative to efficiently allocate medical staff. Precise predictions of maritime emergencies are essential for distributing hospital emergency resources effectively, ensuring the quality and safety of medical services, and optimizing the use of human, financial, and material resources for better economic and social outcomes. Failure to do so may result in the squandering of resources. Based on the findings, actionable steps include: (1) Boosting medical staff reserves in coastal hospitals from September to November, organizing medical staff more logically, and dynamically managing them according to the off-peak season characteristics to leverage their flexibility, adaptability, and synergy. (2) Adopting a flexible scheduling system to reduce work pressure and mental stress on team members, thus safeguarding their well-being and mitigating adverse effects. (3) Enhancing medical resource allocation by opening a fast-track process for resource approval during high-demand periods and allowing resource managers to review and augment resources in slower periods. This not only aids in the training and development of medical staff but also ensures that rescue organizations are adequately prepared for peak season challenges, thereby advancing the precise and sophisticated management of medical resources.



This study introduced the time series model and DBN model into maritime emergency medical rescue research, expanding the scope of the time series model. Through a systematic evaluation of ARIMA, SARIMA, and DBN models, it compares the effectiveness of different time series forecasting models for the number of first responders at sea, demonstrating their application in real-world scenarios. This provides practical cases for maritime administration officials and medical rescue personnel, enhancing the research's practical application value and laying a foundation for future development in emergency medical management for maritime rescue work. In the future, the prediction model constructed in this study can be extended to other coastal cities. Managers can combine SARIMA with machine learning models (such as Long Short-Term Memory) based on local data, leveraging the strengths of different models to improve prediction accuracy. Moreover, with improved computing power, future research can develop real-time prediction systems, using SARIMA and other efficient models for real-time data analysis and prediction. This will enhance marine monitoring and emergency response efficiency, providing new perspectives and methods for research combining emergency medicine and public health.

The study's policy recommendations are as follows: (1) Establish a prediction and early warning system. Medical and maritime managers should create a system based on the study's model. Each unit and institution can regularly upload emergency and environmental data, build a data-sharing platform, and improve data comprehensiveness and timeliness. (2) Enhance the dispatch of emergency rescue resources at sea. Based on predicted personnel needs, dynamically and scientifically allocate rescue vessels, medical supplies, and personnel to ensure sufficient and equitable distribution of resources and reduce response time. (3) Advance intelligent emergency management. Develop a dispatching system that integrates ARIMA, SARIMA, and DBN models into the emergency platform to achieve intelligent dispatching and optimized resource management. Conduct regular

emergency drills to validate the prediction model's accuracy and the dispatching strategy's feasibility, continuously refine the emergency plan, and enhance overall emergency preparedness.

However, this research encounters two primary constraints. Firstly, it did not encompass the entire Hainan region, resulting in unavoidable data omissions. Secondly, the study compiled only 1,312 medical records, a quantity considerably lower than that in similar studies. This reduced data pool likely impacted the predictive precision of the DBN model compared to the SARIMA model. To achieve more reliable long-term predictions, expanding data collection, collaborating with various centers, and closely examining the time distribution patterns of maritime emergency medical services are essential. Such initiatives will foster more efficient and scientifically accurate strategies for allocating medical staff. Another limitation is the study's lack of access to detailed environmental and meteorological data, which could have illuminated fluctuations in the need for first responders at sea. Enhancing the proposed model with this additional data could lead to improvements.

5 Conclusion

This study marks the first predictive analysis of maritime emergency medical personnel incidents in the South China Sea, specifically within China's Hainan region, covering the period from 2016 to 2021. It was found that the SARIMA, DBN, and ARIMA models are all effective in forecasting the need for emergency medical treatment at sea. Among these, the SARIMA method stood out for its superior accuracy over the DBN and ARIMA approaches. By applying the SARIMA (1, 0, 0) (0, 1, 1)₁₂ model to forecast the number of first aid responders at sea, this research provides valuable scientific insights for policymakers in management. It supports dynamic training of personnel, planning for human resource allocation, and optimizing resource use.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Hainan Medical University Ethics Committee (NO.: HYLL-2022-018). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

PY: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft. PC: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft. NZ: Investigation, Writing – review & editing. DL: Investigation, Writing – original draft. BX: Validation, Writing – review & editing. HZ: Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by Nature Science Foundation of Hainan (822MS071, China), Hainan Provincial Key Research and Development Project

References

- Penney G, Byrne W, Cattani M. Death at sea—the true rate of occupational fatality within the Australian commercial fishing industry. *Front Public Health*. (2022) 10:1013391. doi: 10.3389/fpubh.2022.1013391
- Chauvin C, Le Bouar G, Lardjane S. Analysis of occupational injuries in the sea fishing industry according to the type of fishery and the fishing activity. *Int Marit Health*. (2017) 68:31–8. doi: 10.5603/IMH.2017.0006
- Roberts SE, Nielsen D, Kotlowski A, Jaremin B. Fatal accidents and injuries among merchant seafarers worldwide. *Occup Med*. (2014) 64:259–66. doi: 10.1093/occmed/kqu017
- Bilir NA, Scheit L, Dirksen-Fischer M, Terschuren C, Herold R, Harth V, et al. Accidents, diseases and health complaints among seafarers on German-flagged container ships. *BMC Public Health*. (2023) 23:963. doi: 10.1186/s12889-023-15943-x
- Glassberg E, Lipsky AM, Abramovich A, Sergeev I, Hochman O, Ash N. A dynamic mass casualty incident at sea: lessons learned from the mavi marmara. *J Trauma Acute Care Surg*. (2013) 75:292–7. doi: 10.1097/TA.0b013e318294662d
- Dykes L, Thomas M, Hird J. A3 those in peril on the sea: the maritime work of the UK military. *Emerg Med J*. (2011) 28:979.3–979.980. doi: 10.1136/emmermed-2011-200644.3
- Battinelli G, Chintalapudi N, Gagliardi G, Amenta F. The use of radio and telemedicine by tmas centers in provision of medical care to seafarers: a systematic review. *J Pers Med*. (2023) 13:1171. doi: 10.3390/jpm13071171
- Kost GJ, Fuzery AK, Caratao L, Tinsay S, Zadran A, Ybanez AP. Using geographic rescue time contours, point-of-care strategies, and spatial care paths to prepare island communities for global warming, rising oceans, and weather disasters. *Int J Health Geogr*. (2023) 22:38. doi: 10.1186/s12942-023-00359-y
- Zhang W, Yan X, Yang J. Optimized maritime emergency resource allocation under dynamic demand. *PLoS One*. (2017) 12:e0189411. doi: 10.1371/journal.pone.0189411
- Huang J, Han N, Jiang BQ, Cao Y, Han XT. Water search and rescue (Sar) for ship accidents in China: analysis of 12 years' data. *Ann Transl Med*. (2022) 10:1207. doi: 10.21037/atm-22-4530
- Ryu JH, Yeom SR, Jeong JW, Kim YI, Cho SJ. Characteristics and triage of a maritime disaster: an accidental passenger ship collision in Korea. *Eur J Emerg Med Off J Eur Soc Emerg Med*. (2010) 17:177–80. doi: 10.1097/MEJ.0b013e328330f452
- van Boetzelae E, Fotso A, Angelova I, Huisman G, Thorson T, Hadj-Sahraoui H, et al. Health conditions of migrants, refugees and asylum seekers on search and rescue vessels on the central mediterranean sea, 2016–2019: a retrospective analysis. *BMJ Open*. (2022) 12:e53661. doi: 10.1136/bmjopen-2021-053661
- Yang C, Gao J, Du J, Wang H, Jiang J, Wang Z. Understanding the outcome in the Chinese Changjiang disaster in 2015: a retrospective study. *J Emerg Med*. (2017) 52:197–204. doi: 10.1016/j.jemermed.2016.08.013
- Yang W, Su A, Ding L. Application of exponential smoothing method and sarima model in predicting the number of admissions in a third-class hospital in Zhejiang province. *BMC Public Health*. (2023) 23:2309. doi: 10.1186/s12889-023-17218-x
- Wagner B, Cleland K. Using autoregressive integrated moving average models for time series analysis of observational data. *BMJ*. (2023) 383:2739. doi: 10.1136/bmj.p2739
- Li Y, Liu X, Li X, Xue C, Zhang B, Wang Y. Interruption time series analysis using autoregressive integrated moving average model: evaluating the impact of covid-19 on the epidemic trend of gonorrhoea in China. *BMC Public Health*. (2023) 23:2073. doi: 10.1186/s12889-023-16953-5

(ZDYF2022SHFZ306), High-level Talents Program of Hainan Provincial Natural Science Foundation of China (2019RC212), and Hainan Medical University, Scientific Research and Cultivation Fund project (HYPY201906).

Acknowledgments

The completion of this paper cannot be separated from the selfless guidance and unremitting efforts of many people. Here, I would like to express my heartfelt thanks to them, because the completion of this paper cannot be separated from their joint efforts.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2024.1401161/full#supplementary-material>

17. Eyles E, Redaniel MT, Jones T, Prat M, Keen T. Can we accurately forecast non-elective bed occupancy and admissions in the nhs? A time-series msarima analysis of longitudinal data from an nhs trust. *BMJ Open*. (2022) 12:e56523. doi: 10.1136/bmjopen-2021-056523
18. Jilani T, Housley G, Figueredo G, Tang PS, Hatton J, Shaw D. Short and long term predictions of hospital emergency department attendances. *Int J Med Inform*. (2019) 129:167–74. doi: 10.1016/j.ijmedinf.2019.05.011
19. Tian CW, Wang H, Luo XM. Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018. *Epidemiol Infect*. (2019) 147:e82. doi: 10.1017/S095026881800362X
20. Wang YB, Qing SY, Liang ZY, Ma C, Bai YC, Xu CJ, et al. Time series analysis-based seasonal autoregressive fractionally integrated moving average to estimate hepatitis b and c epidemics in china. *World J Gastroenterol*. (2023) 29:5716–17. doi: 10.3748/wjg.v29.i42.5716
21. Zhao D, Zhang H, Zhang R, He S. Research on hand, foot and mouth disease incidence forecasting using hybrid model in mainland China. *BMC Public Health*. (2023) 23:619. doi: 10.1186/s12889-023-15543-9
22. Almeida HS, Sousa M, Mascarenhas I, Russo A, Barrento M, Mendes M, et al. The dynamics of patient visits to a public hospital pediatric emergency department: a time-series model. *Pediatr Emerg Care*. (2022) 38:e240–5. doi: 10.1097/PEC.0000000000002235
23. Zhang X, Zhao X, Mou X, Tan M. Mixed time series approaches for forecasting the daily number of hospital blood collections. *Int J Health Plann Manag*. (2021) 36:1714–26. doi: 10.1002/hpm.3246
24. Suter P, Kuipers J, Beerenwinkel N. Discovering gene regulatory networks of multiple phenotypic groups using dynamic bayesian networks. *Brief Bioinform*. (2022) 23:bbac219. doi: 10.1093/bib/bbac219
25. Zhang T, Ma Y, Xiao X, Lin Y, Zhang X, Xin F. Dynamic bayesian network in infectious diseases surveillance: a simulation study. *Sci Rep*. (2019) 9:10376. doi: 10.1038/s41598-019-46737-0
26. Qiu J, Wang H, Hu L, Yang C, Zhang T, et al. Spatial transmission network construction of influenza-like illness using dynamic bayesian network and vector-autoregressive moving average model. *BMC Infect Dis*. (2021) 21:164. doi: 10.1186/s12879-021-05769-6
27. Johnson DP, Lulla V. Predicting covid-19 community infection relative risk with a dynamic bayesian network. *Front Public Health*. (2022) 10:876691. doi: 10.3389/fpubh.2022.876691
28. Marini S, Trifoglio E, Barbarini N, Sambo F, Di Camillo B, Malovini A, et al. A dynamic bayesian network model for long-term simulation of clinical complications in type 1 diabetes. *J Biomed Inform*. (2015) 57:369–76.
29. Alhamzawi R, Ali H. The bayesian adaptive lasso regression. *Math Biosci*. (2018) 303:75–82. doi: 10.1016/j.mbs.2018.06.004
30. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B*. (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
31. HaiNan Maritime Safety Administration of People's Republic of China (2023). Data center—Hainan maritime Administration of the People's Republic of China. Available at: https://www.hn.msa.gov.cn/cxgk_4_8/index.jhtml (Accessed November 16, 2023).
32. Bai L, Lu K, Dong Y, Wang X, Gong Y, Xia Y, et al. Predicting monthly hospital outpatient visits based on meteorological environmental factors using the Arima model. *Sci Rep*. (2023) 13:2691. doi: 10.1038/s41598-023-29897-y
33. Juang WC, Huang SJ, Huang FD, Cheng PW, Wann SR. Application of time series analysis in modelling and forecasting emergency department visits in a medical Centre in southern Taiwan. *BMJ Open*. (2017) 7:e18628. doi: 10.1136/bmjopen-2017-018628
34. Lin Y, Wan C, Li S, Xie S, Gan Y, Lu Y. Prediction of women and children's hospital outpatient numbers based on the autoregressive integrated moving average model. *Heliyon*. (2023) 9:e14845. doi: 10.1016/j.heliyon.2023.e14845
35. Orphanou K, Stassopoulou A, Keravnou E. Dbn-extended: a dynamic bayesian network model extended with temporal abstractions for coronary heart disease prognosis. *IEEE J Biomed Health Inform*. (2016) 20:944–52. doi: 10.1109/JBHI.2015.2420534
36. Lebre S. Inferring dynamic genetic networks with low order independencies. *Stat Appl Genet Mol Biol*. (2009) 8:9. doi: 10.2202/1544-6115.1294
37. Opgen-Rhein R, Strimmer K. Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics*. (2007) 8:S3. doi: 10.1186/1471-2105-8-S2-S3
38. Lin Y, Chen M, Chen G, Wu X, Lin T. Application of an autoregressive integrated moving average model for predicting injury mortality in Xiamen, China. *BMJ Open*. (2015) 5:e8491. doi: 10.1136/bmjopen-2015-008491
39. Wang M, Jiang Z, You M, Wang T, Ma I, Li X, et al. An autoregressive integrated moving average model for predicting varicella outbreaks - china, 2019. *China CDC Wkly*. (2023) 5:698–702. doi: 10.46234/ccdcw2023.134
40. Zheng Y, Zhang L, Wang C, Wang K, Guo G, Zhang X, et al. Predictive analysis of the number of human brucellosis cases in xinjiang, china. *Sci Rep*. (2021) 11:11513. doi: 10.1038/s41598-021-91176-5
41. Longato E, Morieri ML, Sparacino G, Di Camillo B, Cattelan A, Lo MS, et al. Time-series analysis of multidimensional clinical-laboratory data by dynamic bayesian networks reveals trajectories of covid-19 outcomes. *Comput Methods Programs Biomed*. (2022) 221:106873. doi: 10.1016/j.cmpb.2022.106873

Glossary

COVID-19	Corona Virus Disease 2019
U.K.	United Kingdom
EMS	Emergency medical services
ARIMA	Auto-regressive Integrated Moving Average
SARIMA	Seasonal Auto-regressive Integrated Moving Average
DBNs	Dynamic Bayesian networks
LASSO	Least Absolute Shrinkage and Selection Operator
OLS	Ordinary least square
BNs	Bayesian networks
DAG	Directed acyclic graph
ACF	Auto-correlation function
PACF	Partial auto-correlation function
ADF test	Augmented Dickey-Fuller Test
AIC	Akaike's information criterion
RE	Relative error
MAE	Mean absolute error
RMSE	Root mean square error
R^2	The coefficient of determination
MA	Moving Average
AR	Auto-Regressive
LB test	Ljung-Box Test