MT'24. 10th International Conference on Maritime Transport
Barcelona, June 5-7, 2024

THE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL AS A TOOL FOR PREDICTING ACCIDENTS IN THE MARITIME DOMAIN: THE CASE OF THE GALICIAN FISHING FLEET.

JAVIER RAMÓN SÁNCHEZ-GIRÓN
Department of Nautical Sciences and Marine Engineering, Higher Technical College of Nautical Sciences and Marine Engineering, University of A Coruña, Spain.
javier.sanchez5@udc.es
Orcid: 0000-0001-7302-313X

ÁLVARO FERNÁNDEZ THEOTONIO
Department of Social Psychology, Basic and Methodological Psychology, Faculty of Psychology, University of Santiago de Compostela, Spain.
alvarofernandez.theotonio@usc.es
Orcid: 0009-0003-5524-8142

ROSA MARY DE LA CAMPA PORTELA
Department of Nautical Sciences and Marine Engineering, Higher Technical College of Nautical Sciences and Marine Engineering, University of A Coruña, Spain.
rosa.mary.campa@udc.es
Orcid: 0000-0002-2770-4302

MARÍA-NATIVIDAD LÓPEZ-LÓPEZ
Department of Nautical Sciences and Marine Engineering, Higher Technical College of Nautical Sciences and Marine Engineering, University of A Coruña, Spain
natividad.lopezl@udc.es
Orcid: 0000-0002-2313-0516

Keywords

Forecast, fishing, accident, safety.

Abstract

The Autoregressive Integrated Moving Average (ARIMA) model has proven to be a powerful statistical prediction technique due to its simplicity and wide acceptance. Its main application lies within the realms of economics and social sciences, although not limited to these. This study explores its utilization in the context of maritime safety within the Galician fishing fleet. Specifically, its application is proposed for predicting the vessels involved in accidents quarterly index.

The Galician fishing environment, of significant socioeconomic relevance to the region, is affected by serious accidents compromising its sustainability. The vessels involved in accidents index, derived from investigation reports by the Maritime Accident Investigation Commission (CIAIM) and fleet data collected by the European Fishing Vessel Register, depicts the annual evolution of these accidents in relation to the decreasing fishing capacity of the fleet. Analysis of quarterly values for the period between 2011 and 2021 indicates that accidents are recurrent and persistent in nature.

Using the proposed ARIMA model, index values for the quarters of 2022 were forecasted and the model's effectiveness was validated by comparing the obtained results with actual observed data.

The outcomes support the utility of ARIMA model as a tool for predicting maritime accidents and pave the way for future research in the field of safety. In the studied case, this predictive capacity supports proactive safety management on board while providing a solid foundation for the adoption of preventive initiatives in both public and private sectors within the industry.

1 INTRODUCTION

ARIMA models are a class of statistical models used for analysing and forecasting time series data. They are particularly valuable in academic research due to their ability to capture and model complex temporal patterns in data. Therefore, ARIMA models find applications in various fields such as economics, finance, environmental science, epidemiology, and more (Yenidogan et al., 2018; Benvenuto et al., 2020; Lai and Dzombak, 2020). Researchers use ARIMA models to analyse historical data, understand underlying trends and patterns, and make forecasts for future observations. Additionally, ARIMA models serve as a foundation for more complex time series modelling techniques, making them a fundamental tool in time series analysis within academia.

In maritime research, ARIMA models had been utilized in different applications such as short-term forecasting of container throughput, ship trajectory planning for collision avoidance and prediction of freight and passenger traffic (Rashed et al., 2017; Aveta and Romano, 2020; Aivazidou and Politis, 2021; Abebe et al., 2022). These models provide valuable insights that aid stakeholders in decision-making and management practices.

In the field of maritime accident prevention, ARIMA models have served as a primary component in multifactorial predictive instruments and as a tool for assessing the outcomes of more complex prediction systems (Sui et al., 2023; Wang et al., 2023).

Given the widespread use and acceptance of ARIMA models, the purpose of this study is to determine a model that allows predicting the annual number of Galician vessels involved in accidents, a persistent phenomenon with serious consequences. Recently, on February 15, 2022, the trawler "Villa de Pitanxo" sank off the coasts of Newfoundland and Labrador, in Canadian waters, claiming the lives of 21 of its 24 crew members, the highest number of fatalities in an accident aboard a Galician fishing vessel since the sinking of the fishing boat "Marbel" in 1978. In the same year, four other ships based in Galicia suffered accidents, resulting in two fatalities and two of the vessels becoming irrecoverable.

Far from being an infrequent phenomenon, accidents in the Galician fishing fleet have characterized the profession's activity for at least forty years despite the progressive reduction in the number of vessels and their fishing capacity (Allegue, 1993; Perez-Labajos et al., 2006; Moreno and Gómez-Cano, 2014).

The consequences of this accident rate are usually very serious both from a human and material point of view. In fact, the fishing sector in Galicia presents the highest rate of severe occupational accidents and fatalities, with its figures consistently higher than other highly accident-prone sectors such as construction and industry (Sánchez and De la Campa, 2022).

The socioeconomic relevance of extractive fishing activity for Galicia is significant for the region in terms of Gross Domestic Product (GDP) and employment, as it acts as a backbone for the services and industry that have emerged around it, and in some small coastal communities, it constitutes the main source of employment (Surís-Regueiro and Santiago, 2014; Garza-Gil et al., 2017; García-Negro et al., 2018; Fernández-González et al., 2022).

It is for this reason, and in pursuit of the objectives outlined in Goal 8 (Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all) of the 2030 Agenda for Sustainable Development (Asamblea General de Naciones Unidas, 2015), that reducing accidents in the Galician fishing fleet deserves special attention, as its improvement would contribute positively to both onboard safety and the creation of a more attractive image of the profession (Johnsen and Vik, 2013; Penas, 2019), a matter of particular interest given the generational turnover problems faced by the sector (Consultora Sinerxia, 2018; Confederación Española de Pesca, 2019).

In summary, the present study aims to apply the predictive capabilities of ARIMA models to the specific case of maritime accidents in the Galician fishing fleet, considering its temporal nature and the negative effects it has on the extractive fishing sector, which is of great socio-economic importance to the region. To achieve this, five objectives are proposed:

- Describe maritime accidents in the Galician fishing fleet in relation to the evolution of the fishing fleet during the period 2011–2021 using the annual vessels involved in accidents index.
- Analyse the trend and seasonality of the vessels involved in accidents index during the period 2011–2021.
- Determine a valid ARIMA model for predicting the vessels involved in accidents index using the Akaike Information Criterion (AIC).
- Predict the vessels involved in accidents index for the year 2022 using the selected model.
- Assess the effectiveness of the method by comparing predicted results with observed ones.

The interest of this study lies precisely in finding evidence to support both public and private actions in the Galician fishing sector regarding maritime accident prevention. It is believed that having well-founded evidence of maritime accidents makes it imperative and unavoidable to take action to prevent them.

R and R Studio (R Core Team, 2023) will be used for the selection and determination of the model due to the wide range of packages available and specifically designed for time series analysis. These packages offer comprehensive functionalities for model fitting, diagnostics, and forecasting, providing a powerful and convenient tool for conducting sophisticated time series analysis with ARIMA models (Cryer and Chan, 2008; Cowpertwait, 2009).

2 THEMATIC DEVELOPMENT

Between 2011 and 2021, the CIAIM investigated 114 accidents involving fishing vessels[1] based in Galicia. This body of the Spanish administration, attached to the Ministry of Transport, Mobility, and Urban Agenda, is responsible for investigating "serious or very serious[2]" maritime accidents suffered by Spanish-flagged civilian vessels and periodically publishing the results of its inquiries to draw lessons on safety and recommendations for the prevention of similar events aimed at the parties involved (Ministerio de Fomento, 2011). Its reports thus constitute the main source of information on the number, time and nature of accidents occurring annually, as well as on the characteristics of the vessel or vessels involved. Among these are the the name and registration of the vessel or vessels affected, which, through data from the online European fishing registry, allows for determining their home port[3] and consequently the autonomous community to which they belong. From this information, a series can be constructed that relates the number of vessels involved in accidents to the established time frame, which in this case corresponds to the quarters of each year between the period 2011–2021. This division approximately corresponds to the duration of the natural seasons, which, if seasonal trends are perceived in the series, could be related to the effects that the climatological conditions of each season have on the probability of maritime accidents occurring, as suggested by the works of Jin et al. (2002) and Jin and Thunberg (2005).

However, the Galician fishing fleet progressively decreases in number and capacity each year due mainly to the common European policy of reducing fishing effort (Cordón et al., 2016; Carvalho et al., 2019; Sobrino and Oanta, 2021) (Figure 1). This is the reason why an index to assess the number of vessels involved in accidents in relation to the operational fleet in each quarter has been calculated. This approach thus provides a broader and contextualized description of the phenomenon of accidents from a temporal perspective.

---

[1] Fishing vessels are those intended for the commercial capture and extraction of fish and other living marine resources, but not fishing auxiliary vessels, auxiliary vessels for aquaculture operations, or devices dedicated to the cultivation or rearing of marine species.

[2] The circumstances leading to the investigation of an accident are established in Royal Decree 800/2011, of June 10, which regulates the investigation of maritime accidents and incidents and the CIAIM, so those that do not comply with them are dismissed by the commission's plenary. Article 3 of the aforementioned decree defines a "maritime accident" as "an occurrence, or series of occurrences, directly related to the operation of a ship that has resulted in any of the situations listed below:

i. The death or serious injury of a person;
ii. The loss of a person on board;
iii. The loss, presumed loss, or abandonment of a ship;
iv. Serious damage to a ship;
v. The grounding or major damage of a ship, or its involvement in a collision;
vi. Serious damage to maritime infrastructure other than the ship, posing a serious threat to the safety of the ship, another ship, or a person;
vii. Serious damage to the environment, or the possibility of serious environmental damage, as a result of damage suffered by one or more ships."

[3] The concept of home port is defined in Article 65 of Act 3/2001, of March 26, on State Maritime Fishing, stating that it is "the one from which the vessel carries out the majority of its activities related to the beginning of fishing trips, dispatch, and commercialization of catches" for vessels fishing in national waters, and as "the one with which a significant socioeconomic link is maintained, in accordance with what is established by regulations" for vessels fishing outside national waters.

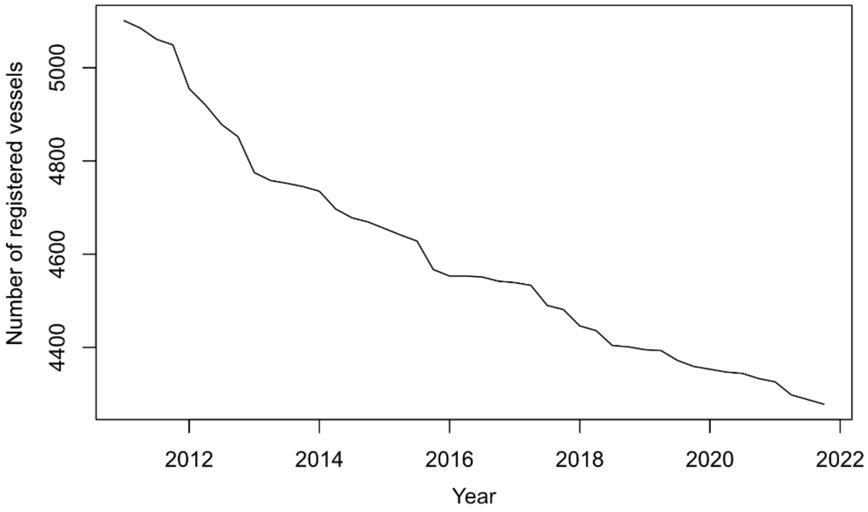

Fig. 1 - Number of registered Galician fishing vessels in each quarter from 2011-2021 (Source of the data: (Dirección General de asuntos marítimos y pesca, 2023a)

In addition to the above, two other indexes have been calculated, with the purpose of describing the consequences of accidents suffered by Galician vessels and assessing their evolution. The first, shown in the left side of Figure 2, relates the number of vessels involved in accidents annually in which one or more crew members or a person unrelated to the vessel perished[4], to the number of vessels comprising the operational fleet. The second, on the other hand, maintains the same denominator but uses the number of vessels involved in accidents which were not recoverable due to sinking or structural damage suffered as the numerator (Figure 2, right side).

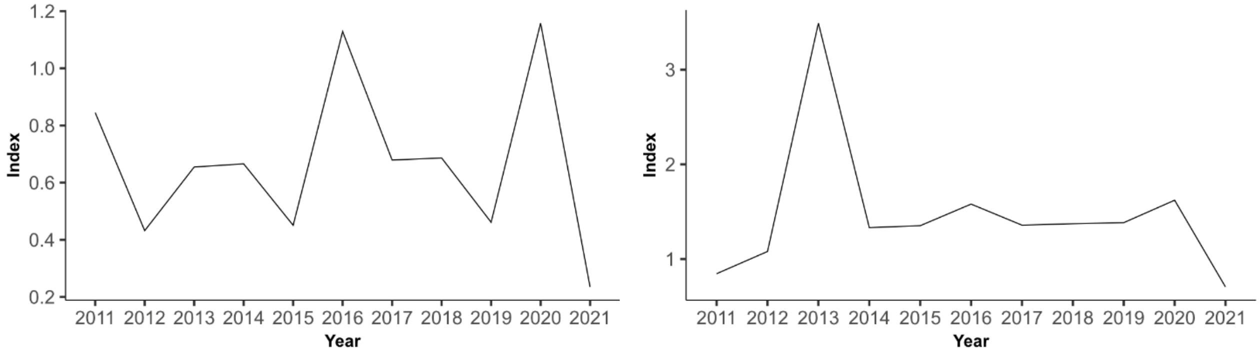

Fig. 2 – Left: Annually index of vessels involved in accidents in which one or more crew members or a person unrelated to the vessel would perish. Right: Index of vessels involved in accidents which are not recoverable due to sinking or structural damage suffered. (Source of the data: (Comisión de Investigación de Accidentes e Incidentes Marítimos, 2023; Dirección General de asuntos marítimos y pesca, 2023a)

The evolution of both indexes over the period does not show a clear trend, suggesting that the severity of accidents does not vary appreciably.

---

[4] Only one accident resulted in the loss of life of someone not related to the vessel's crew. This occurred when the fishing boat named "No Se" collided with a swimmer near the Punta Lagoa marina in Vigo in July 2014.

Regarding the vessels involved in accidents index, which constitute the main time series analysed in this study, the upper plot in figure 3, suggests the absence of both trend and seasonality in the series. This conclusion is reinforced by the results of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), which show values close to zero for most lags. In other words, there is no clear linear relationship between past and future observations, indicating that the maritime accident index does not exhibit a discernible increasing or decreasing trend over time, nor do they display predictable seasonal patterns. Thus, the sector finds itself in a situation that, due to the severe consequences of accidents, deviates from desirable onboard safety conditions according to national and international standards and the sustainability of the fishing activity.

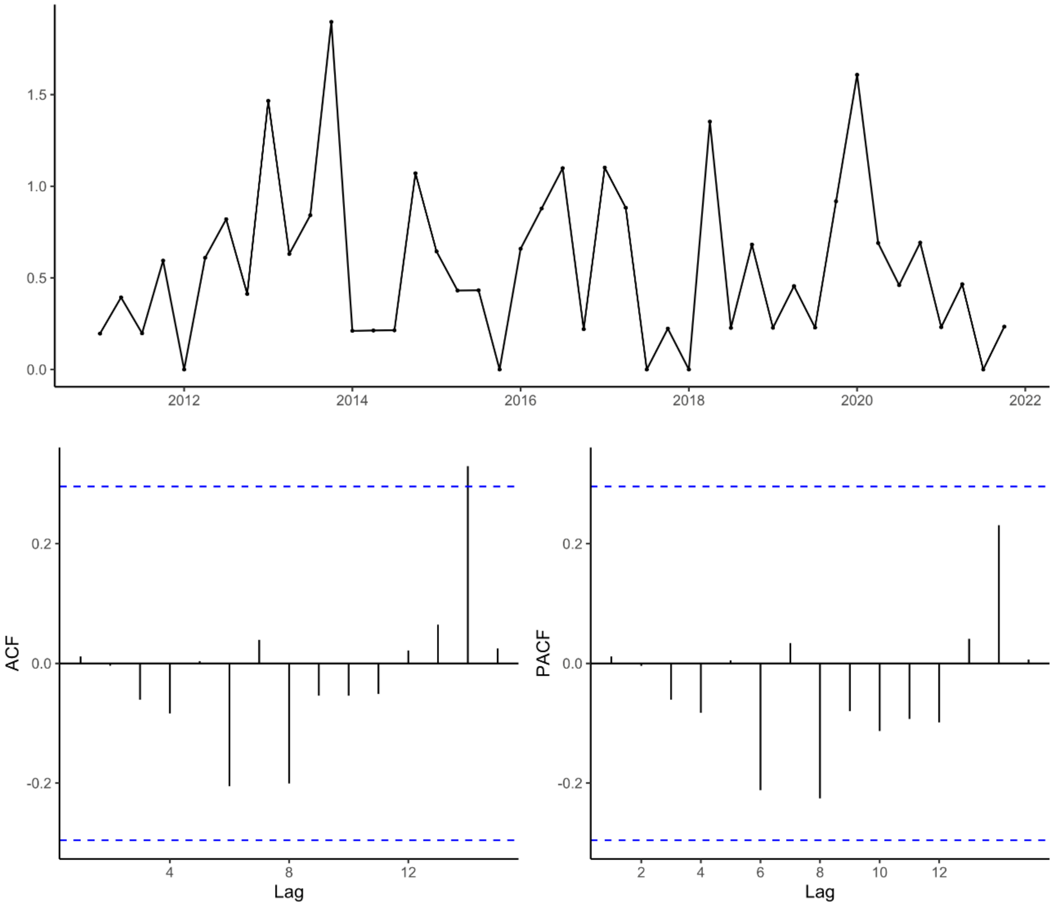

Figure 3 - Top: Time series of the vessels involved in accidents index during the period 2011 - 2021. Bottom: Results of the ACF and PACF. (Source of the data: (Comisión de Investigación de Accidentes e Incidentes Marítimos, 2023; Dirección General de asuntos marítimos y pesca, 2023a)

3 METHODS

3.1 Calculation of the vessels involved in accidents index: obtaining the number of maritime accidents and registered vessels in Galician ports.

The vessels involved in accidents index has been obtained using the following relationship (Perez-Labajos et al., 2006):

$$\frac{\text{Number of vessels involved in accidents} \cdot 1000}{\text{Number of Galician fishing vessels registered in the european registry}}$$

The number of vessels involved in accidents has been tallied for each quarter of each year in the studied period, with the first quarter spanning from January 1st to March 31st, the second quarter from April 1st to June 30th, the third quarter from July 1st to September 30th, and the fourth quarter from October 1st to December 31st. Meanwhile, the number of vessels registered has been obtained from instances of the online European registry, as it is the only one of the three existing registries[5] that allows historical information retrieval for any day of any past year.

Thus, the number of registered vessels was determined for the days February 14, May 16, August 15, and November 15, ensuring that the figure is representative of each period. To achieve this, the annotations of all Spanish-flagged vessels registered each year on the aforementioned dates were downloaded from the online page of the European registry of fishing vessels (Dirección General de asuntos marítimos y pesca, 2023a). From these, Galician vessels were identified and counted based on the "Place of registration" variable, which contains an alphanumeric code used to identify the vessel's home port, the relationship and meaning of which are documented in the Master Data Registry (MDR) code list (Dirección General de asuntos marítimos y pesca, 2023b).

The indexes related to vessels involved in fatal accidents and those deemed irrecoverable were obtained in the same way as described above, with the exception that the number of Galician fishing vessels recorded in the registry was obtained for December 31 of each year in the period.

Information regarding the number of vessels involved in accidents and the circumstances of the incident, detailing, among other factors, the number of fatalities and the damage sustained by the vessel involved, was sourced from investigation reports published by the CIAIM (Comisión de Investigación de Accidentes e Incidentes Marítimos, 2023). Each accounted-for vessel corresponds to one vessel involved in an accident; that is, in accidents involving two or more vessels, each vessel is counted as an independent vessel. All accident reports of vessels whose home port is in Galicia published between 2011 and 2021 were selected in accordance with the scope of the study. The home port of a vessel or craft involved in an accident is usually documented in the investigation report. In cases where such information was not provided, the Spanish Fishing Registry was consulted to determine the home port.

3.2 Determination of the model orders.

An ARIMA model combines autoregressive (AR), differencing (I), and moving average (MA) components to capture the temporal structure of the data. It is specified by three parameters: "p" (autoregressive order), "d" (differencing order), and "q" (moving average order). These parameters determine the structure and complexity of the model, allowing it to capture different patterns and trends present in the time series data.

An ARIMA model is an extension of an ARMA (Autoregressive Moving Average) model. It includes a differencing step to make the time series stationary, ensuring that it does not exhibit trend.

Thus, the first step in constructing an ARIMA model is determining its orders "p", "d", and "q", but as a preliminary step, it must be observed that the time series meets the stationary condition typical of ARIMA processes, that is, that the mean and variance do not change over time.

Compliance with this condition is studied through the graphical representation of the time series and the values of the ACF and PACF (Figure 3), obtained using the "forecast" package developed by Hyndman, Khandakar et al. (2008, 2023).

As mentioned in the previous section, the studied series does not appear to exhibit trend or seasonality

---

[5] The fishing registry encompass the annotations of all vessels that, at a given moment, are active; meaning, they have authorization to engage in commercial fishing activities according to the census in which they are registered, which in turn determines the fishing gear they can use and the fishing grounds where they are authorized to operate. The vessels comprising the active fishing fleet of Galicia will therefore be listed in the three fishing registries that exist: the Register of Fishing Vessels of the Autonomous Community of Galicia, the Spanish General Register of the Fishing Fleet, and the Register of the European Union fishing fleet. Both the European and Spanish registers only include fishing vessels registered in the third list.

(Fig. 3), leading to the belief that it could have been generated by a stationary process modellable with an ARMA. However, the data variability appears compromised between late 2013 and 2020, thus failing to meet the homoscedasticity condition characteristic of such processes.

To achieve stabilization of the data variability, transformation was performed using the family of potential transformations proposed by Box and Cox (1964), employing both the Guerrero method (1993) and the maximum likelihood (ML) method. However, more robust and direct results were obtained by applying a single differencing to the untransformed time series (Figure 4). Consequently, this procedure establishes the order "d" of the model, which will be equivalent to 1.

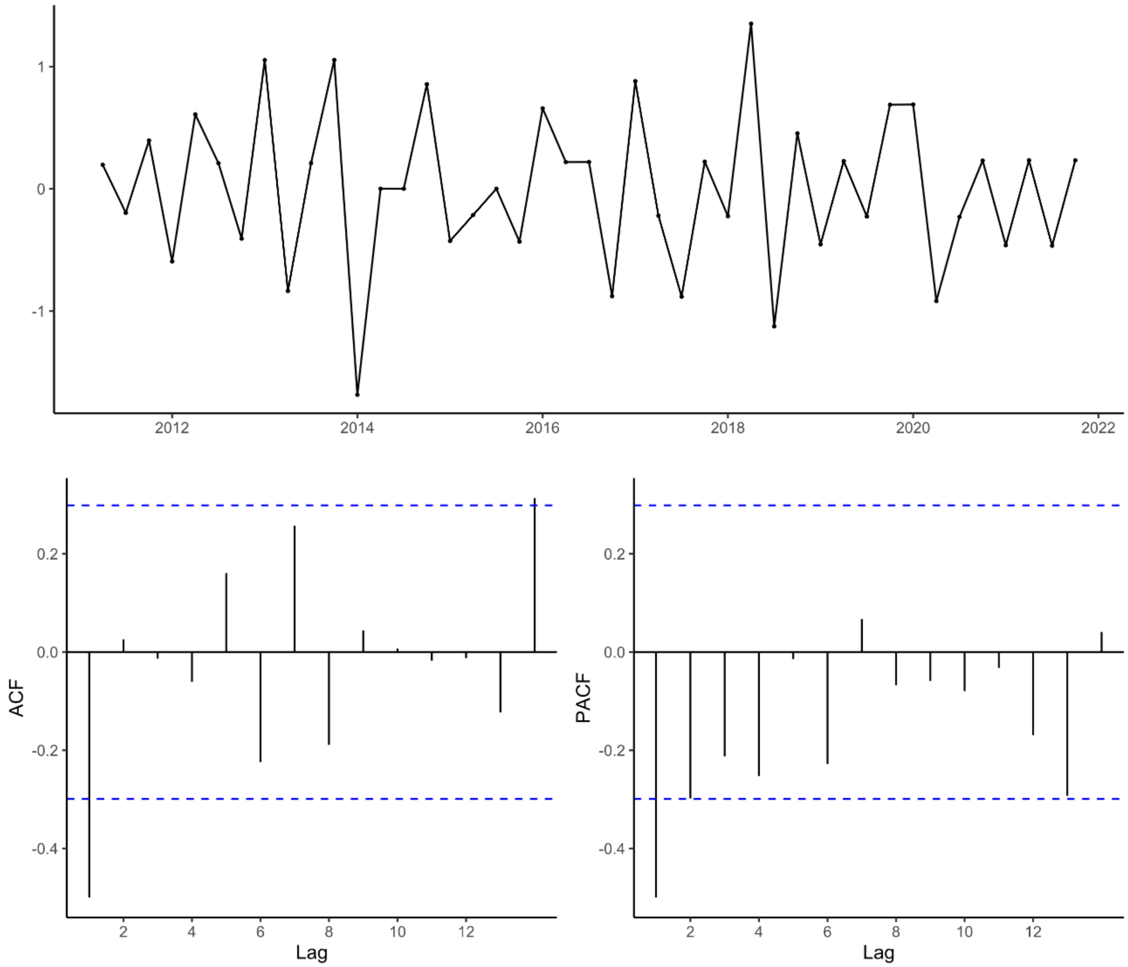

Figure 4 - Top: Time series of the vessels involved in accidents index during the period 2011 – 2021 differentiated. Bottom: Results of the ACF and PACF. (Source of the data: (Comisión de Investigación de Accidentes e Incidentes Marítimos, 2023; Dirección General de asuntos marítimos y pesca, 2023a)

The identification of the "p" and "q" orders of the model is carried out based on the interpretation of the results from the ACF and PACF plotted in their respective correlograms. Specifically:

- In the ACF correlogram, the lag at which the autocorrelation coefficient is significantly different from 0 is the first lag. Therefore, it is suggested that the studied time series may have been generated by a differenced (d =1) MA (1) process.
- In the PACF correlogram, the lag at which the autocorrelation coefficient is significantly different from 0 is also the first lag. Hence, it is suggested that the studied time series may have been generated by a differenced (d = 1) AR (1) process.

3.3 Model selection and estimation

To determine which of the proposed models best fits the studied series and whether there exists a more complex yet more suitable one, an automatic method will be used. Precisely, the *auto.arima* function from the "forecast" package determines an ARIMA model based on the ML method and the AIC. Maximum Likelihood estimation finds the parameter values of a stochastic model that maximize the likelihood of obtaining the observed data, while the AIC provides a score based on this likelihood and the number of parameters used by the model. Thus, the lower the AIC value of a model, the fewer parameters are used in it, a desirable condition for subsequent estimation and prediction phases.

In addition to AIC, the *auto.arima* function provides the value of the corrected AIC (AICc) and the Bayesian Information Criterion (BIC), both based on AIC but incorporating corrections to account for sample size. Of the three criteria, Hyndman and Khandakar (2008) suggest using AICc, which is also recommended for small samples (Hurvich and Tsai, 1989; Burnham and Anderson, 2002). The function yields an ARIMA (2,1,0) model with the lowest AICc value (73.28), the parameters, and standard error (SE) of which are shown in the following table:

Coefficients	**AR1**	**AR2**
Value	-0,6352	-0,2858
SE	0,1441	0,1430
sigma^2 = 0.2864: log likelihood = -33.33 AICc=73.28		

Table 1 – Parameters estimated by the *auto.arima* function for the ARIMA (2,1,0) model.

The suggestions made in the previous section, an AR(1) model (ARIMA 1,1,0) or MA(1) model (ARIMA 0,1,1), were also considered. However, given that the difference between the AICc of the model calculated by auto.arima and the previously proposed through the study of autocorrelations were more than 2 units for both of them, those were discarded (Sakamoto et al., 1986) and the ARIMA (2,1,0) accepted.

Regarding the model parameters, it was assessed whether the obtained values were significantly different from 0. To do this, assuming a normal distribution of the parameters, a confidence interval can be conceived with a significance level of 5% as ± 1.96 SE, so that if the value of each parameter is not contained within the confidence interval, the alternative hypothesis that it is different from 0 can be accepted. Thus, it is demonstrated that the values of the AR1 and AR2 parameters of the model are significantly different from 0, justifying their inclusion in the model.

3.4 Model diagnosis

The diagnostic phase of the model involves assessing, through various tests, whether the properties of the residuals resemble those of a Gaussian white noise process. This means they should exhibit no correlation, have a normal distribution, zero mean, and constant variance. Failure to meet these conditions would invalidate the estimated model as a potential generator of the time series. Residuals are defined as the differences between the observed value of the variable at a given time and the value estimated by the model at the same time.

The absence of correlation in the model can be initially assessed by examining the graphical representation of the standardized residuals. If the model is appropriate, these plots will suggest a rectangular dispersion around a horizontal zero level without trend. In the case of the model under study, this representation is confirmed in the first graph of Figure 5. The *tsdiag* function from the R "stats" package was used to obtain it.

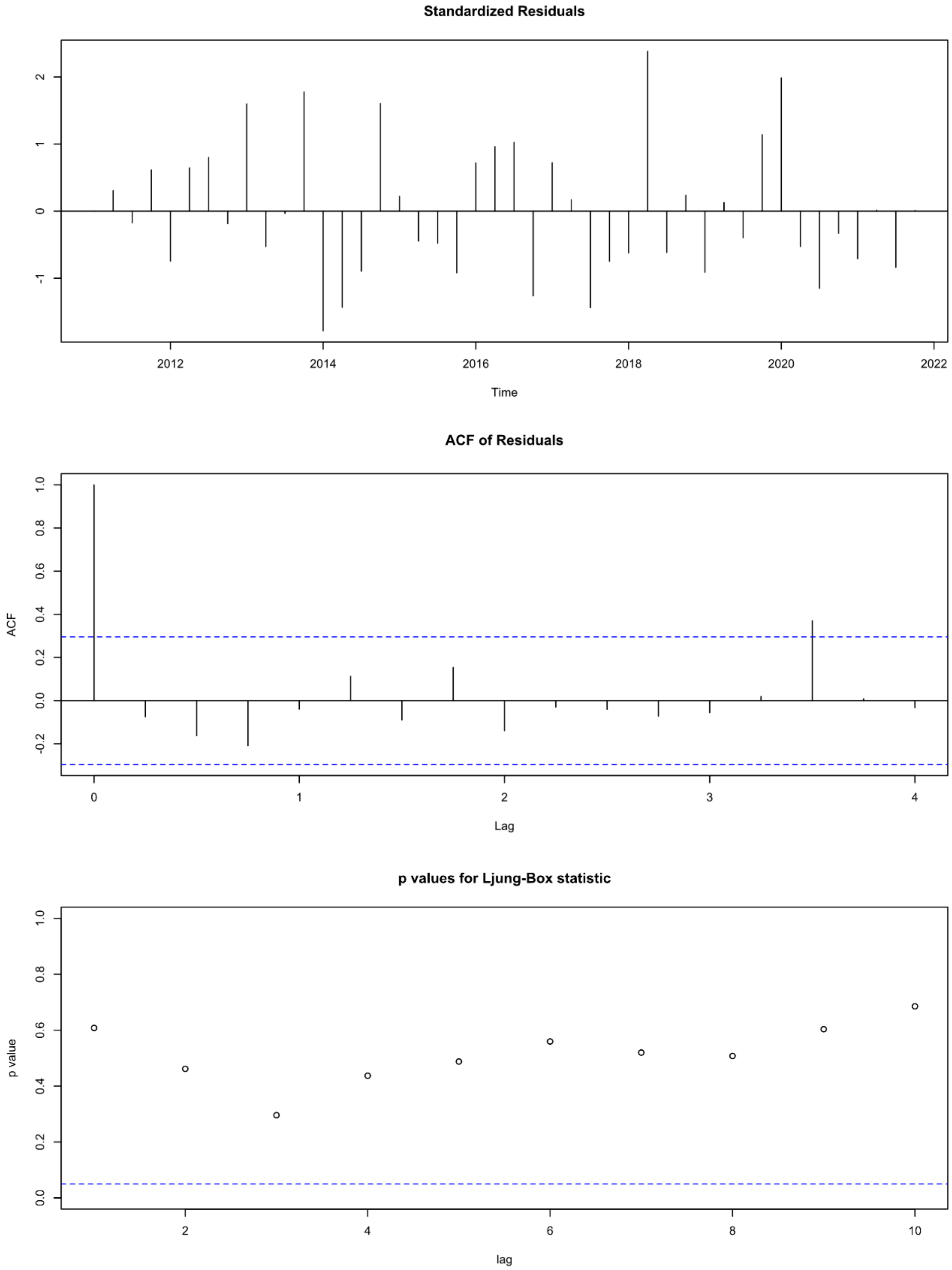

Figure 5 - Top: graphical representation of the residuals. Middle: Sample coefficients of the ACF function. Bottom: p-values for the Ljung-Box statistic for individual residuals.

Nonetheless, Peña (2005), Sumway and Stoffer (2006), as well as Cryer and Chan (2008) propose a more precise assessment of residual correlation through graphical analysis of the results of the ACF. If the independence condition is met, the coefficients of the function will be approximately random variables with a mean of zero, so their value should graphically fall within the space bounded by the 5% confidence intervals. In the second graph of Figure 5, it is observed that, except for the value of the lag 0 coefficient, which the *tsdiag* function defaults to a value of 1, only one significantly different value from zero is shown at a high lag. However, given the significance level, it is expected that this will occur in one out of every 20 coefficients, so the independence condition is considered proven through graphical analysis of the residuals' ACF.

Finally, the independence analysis is completed by the Ljung-Box test (Ljung and Box, 1978). Through this test, a Q statistic of a χ^2 distribution is obtained, so if its value exceeds the 0.95 percentile of this distribution, the null-hypothesis of residual independence will be rejected. The Q statistic calculated for the residuals with the *checkresiduals* function of the "forecast" package yields a result of 2.696 with a p-value of 0.2965 (Table 2), so there is enough evidence to accept the null-hypothesis of residual independence.

Variable	**Q^2**	***df***	***p***	**model df**	**Total lags used**
ARIMA (2,1,0) residuals	7,271	6	0,2965	2	8

Table 2 - Results of the Ljung-Box test on the residuals

To check if the mean of the residuals is equal to zero, the Student's t-test was proposed. Thus, the test result (Table 3), with a p-value of 0.9848, led us to accept the null-hypothesis that the means of the residuals are equal to 0.

Variable	**t**	***df***	***p***	**95% CI**	**Estimaded mean**
ARIMA (2,1,0) residuals	-0,019135	43	0,9848	[-0,1604, 0,1573]	-0,001507

Table 3 - Results of the t-test on the residuals

The evaluation of the normality of the residuals will be conducted through the analysis of a quantile-quantile (Q-Q) plot and the Jarque-Bera (1980) and Shapiro-Wilk (1965) normality tests.

The Q-Q plot (Figure 6) represents the sample quantiles of the residuals against the sample quantiles of a normally distributed sample, which appears as a straight line. Thus, the closer the distribution of the quantiles to the latter, the closer it will approach the condition of normality. Despite its deviation from normality on extreme quantiles, the displayed output suggests that these are not significant enough to invalidate the assumption of normality.

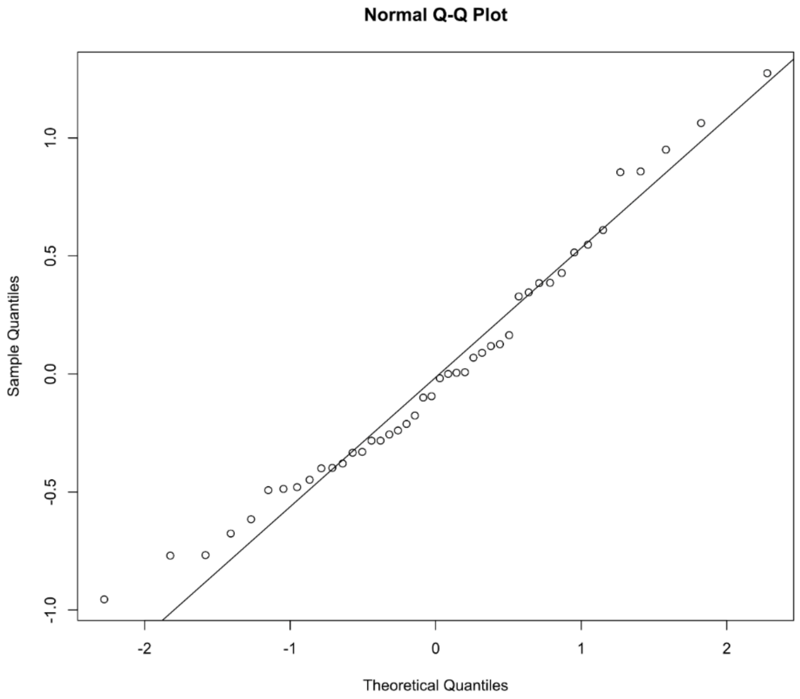

Figure 6 - Normal Q-Q plot of the residuals

Regarding the Jarque-Bera test, it compares the skewness and kurtosis of a given distribution with those of a normal distribution. The obtained test statistic is, under the null hypothesis of normality, distributed approximately as a χ^2 distribution with two degrees of freedom. The test result (Table 4), performed with the *jarque.bera.test* function of the "tseries" package (Hornik and Trapletti, 2023), supports the acceptance of the null hypothesis of normality.

Variable	X^2	*df*	*p*
ARIMA (2,1,0) residuals	2,1413	2	0,3428

Table 4 - Jarque-Bera test results on the residuals

Lastly, the Shapiro-Wilk test calculates a statistic W that evaluates whether a sample comes from a normally distributed population, with small values of the statistic indicating a worse fit to normality. The statistic values for different confidence levels and sample sizes were obtained by Pearson and Hartley (1972). Thus, for n = 44 and a 95% confidence level, W will have a value of 0.987, higher than the calculated one in R using the *shapiro.test* function of the "stats" package, which with a p-value of 0.2936, supports the acceptance of the null hypothesis of normal distribution.

The contrasts and evaluations performed in the preceding sections demonstrate the compliance with the properties of zero mean, normal distribution, and independence of the residuals. Therefore, these results prove that the residuals would have been generated by a Gaussian white noise process, a condition that, having been satisfied, validates the choice and fitting of the model.

As a summary, a graph (Figure 7) is attached with the representation of the standardized residuals, the coefficients of the ACF, and their distribution compared to a normal one generated in R using the *checkresiduals* function of the "forecast" package.

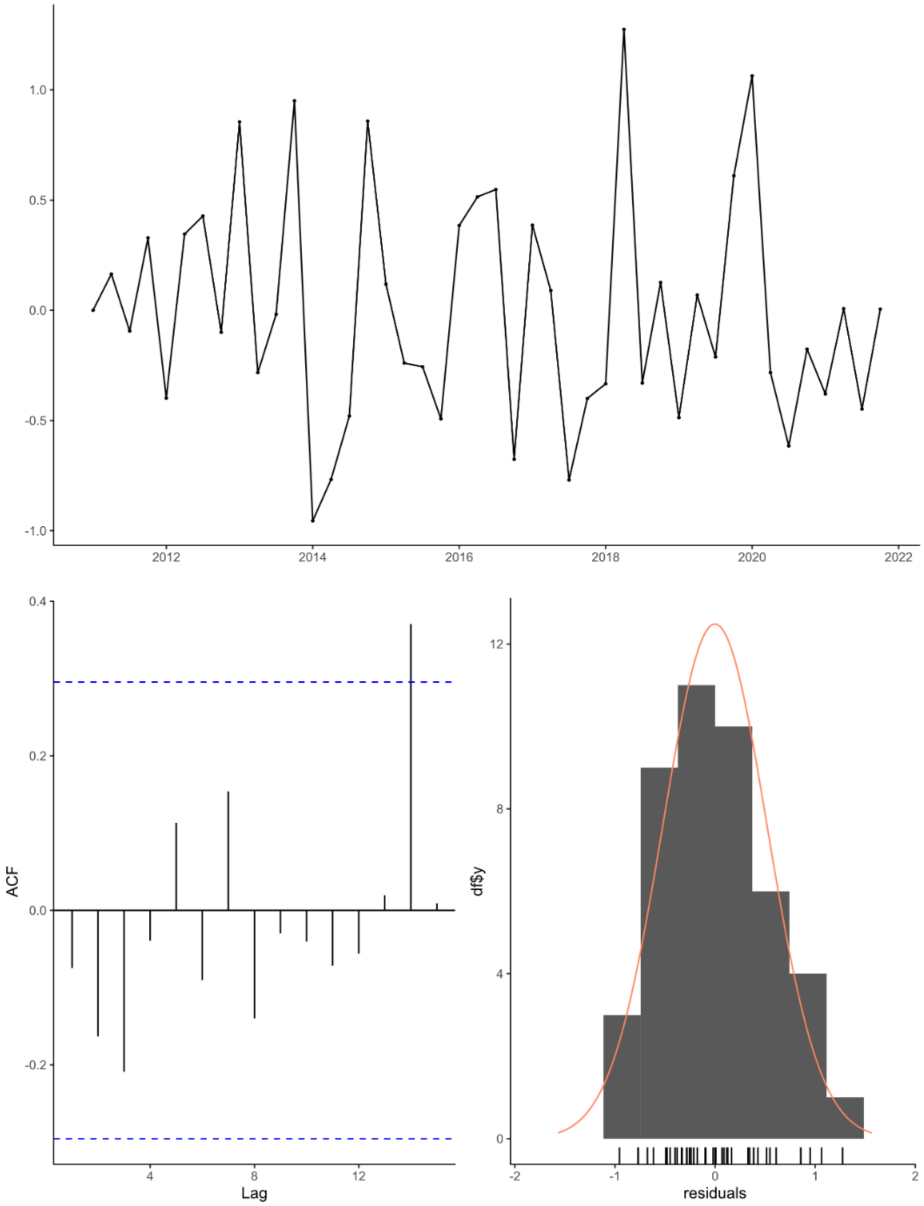

Figure 7 – Top: Residuals from ARIMA (2,1,0). Bottom left: ACF of the residuals. Bottom right: Distribution of the residual compared to a normal distribution.

3.5 Forecast.

The prediction of future values of the time series was obtained using the *forecast* function, and to evaluate the predicted values, confidence intervals were calculated. These confidence intervals are constructed based on the assumption that the residuals meet the conditions of normality and independence. However, due to the small sample size of 44 observations in the studied time series, the bootstrap technique was employed. Bootstrap allows for the generation of multiple random samples from a single original sample, ensuring robust prediction intervals with small sample sizes. In practice, while using the *forecast* function from the "forecast" package, the number of resamples was set to 1000. This approach ensures that the confidence intervals provide a reliable measure of uncertainty around the predicted values, contributing to the assessment of forecast accuracy. On the other hand, given that the forecasts for future values depend on immediate future values, the number of differencing in the model and the value of the constant will significantly influence them. For this reason, the prediction made is limited to only four future values.

4 RESULTS

The forecasted and observed values of the vessels involved in accidents index can be seen in Table 5, while their graphical representation is displayed in Figure 8. It is important to note that the mathematical calculation of confidence intervals yields both positive and negative bounds. However, in the context of this analysis, the negative bound holds no practical significance beyond the zero value, as the actual vessels involved in accidents index cannot be negative. Therefore, for clarity and practical interpretation, only the positive bounds of the confidence intervals are considered in the following discussion.

Time	Forecasted value	Observed value	Low 80%	High 80%	Low 95%	High 95%
2022 Q1	0,2149684	0,4683841	-0,3989804	1,0747750	-0,5531763	1,491177
2022 Q2	0,1568588	0,2344116	-0,5296340	0,9395977	-0,7595826	1,347762
2022 Q3	0,1958907	0,2350176	-0,5606331	1,0543004	-0,8502344	1,465096
2022 Q4	0,1844477	0,2354049	-0,6882632	1,0654127	-1,0552977	1,544494

Table 5 - Real and forecasted values of the vessels involved in accidents index for the year 2022 (Source of the data: (Comisión de Investigación de Accidentes e Incidentes Marítimos, 2023; Dirección General de asuntos marítimos y pesca, 2023a)

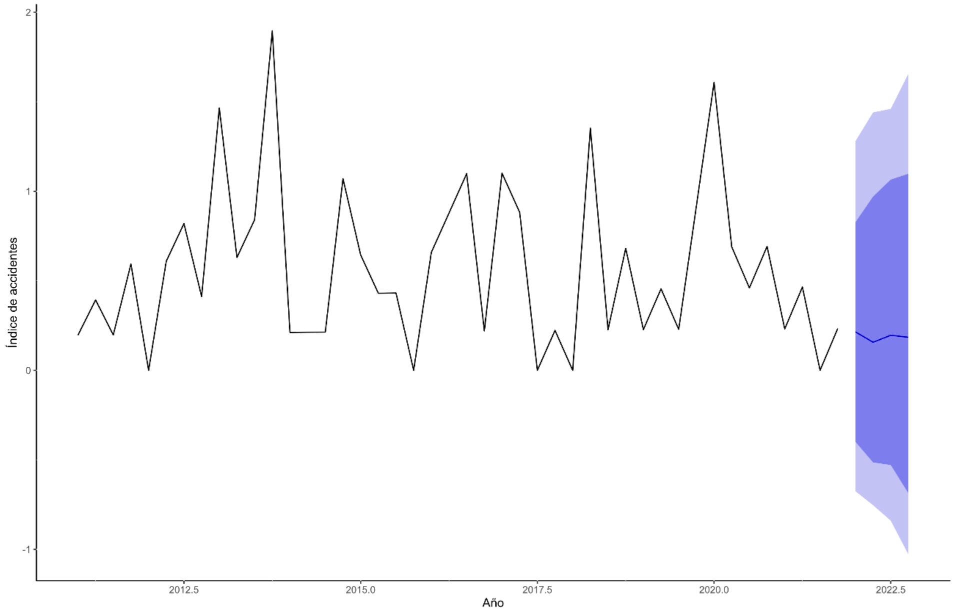

Figure 8 – Vessels involved in accidents index times series for the period 2011 – 2022. Lighter purple tones represent 95% confidence intervals while darker tones 80% confidence intervals. (Source of the data: (Comisión de Investigación de Accidentes e Incidentes Marítimos, 2023; Dirección General de asuntos marítimos y pesca, 2023a)

Based on the comparison between the predicted and actual values of the vessels involved in accidents index for the year 2022, it is notable that the observed values fall within the confidence intervals provided by the model for both the 80% and 95% levels.

For instance, in the first quarter of 2022, the actual index value was approximately 0.468, which falls within the 80% confidence interval of [0, 1.075] and the 95% confidence interval of [0, 1.491]. Similarly, in the second, third, and fourth quarters, the actual index values of 0.234, 0.235, and 0.235 respectively, also fall within the confidence intervals provided by the model.

Nonetheless, it can also be observed that the model generally tends to slightly underestimate the index. For example, in the first quarter of 2022, the actual index value was approximately 0.468, while the predicted value was approximately 0.215. Similarly, in the second, third, and fourth quarters, the actual index values were 0.234, 0.235, and 0.235 respectively, whereas the predicted values were 0.157, 0.196, and 0.184 respectively.

Despite the underestimation, the predicted values generally follow a similar trend to the actual values, indicating that the model captures some of the variability in the data. On the other hand, the model provides predictions that are consistent with the observed data, as the actual values lie within the predicted ranges. However, further analysis and refinement of the model may be necessary to improve its accuracy in forecasting vessels involved in accidents index.

5 CONCLUSIONS

The accidents suffered by Galician fishing vessels constitute a recurring phenomenon with serious human and material consequences. Their evolution, studied through the vessels involved in accidents index, has not shown any trend of improvement in recent years, nor have the indexes used to assess their consequences. However, the results of this study support the effective prediction of the phenomenon. Therefore, the developed ARIMA model emerges as a valuable asset in accident prevention within this sector. Despite its limitations, such as the tendency to slightly underestimate the index, the model's predictions consistently align with observed data, offering a reliable means of identifying periods of heightened accident risk. Notably, the actual values consistently fall within the confidence intervals provided by the model at both the 80% and 95% confidence intervals, underscoring its utility as a proactive tool in accident prevention efforts.

However, it's essential to acknowledge the study's limitations. With only 44 observations available for analysis, there may be inherent constraints in the model's predictive accuracy, particularly when extrapolating findings to broader contexts or making long-term forecasts. On the other hand, the number of vessels registered at a specific moment provides an incomplete measure of the exposure of the fleet to the risk of experiencing an accident, since it does not allow us to know the exact time a vessel spends fishing or sailing, or even if it performs any commercial o navigation activity at all.

Additionally, while the model provides valuable insights into accident likelihood, it does not account for all potential variables influencing accident occurrence, highlighting the need for further research to refine and expand upon its predictive capabilities. In fact, it's worth noting that both the data on the number of accidents and registered vessels are being treated in an aggregated manner, without considering the technical and operational diversity of the Galician fishing fleet. This limitation could influence the precision of predictions and the interpretation of results. Therefore, for a more effective prevention of maritime accidents, it would be crucial to complement this approach with statistical techniques that allow for the integration of these specific fleet characteristics into the analysis. Such an approach would not only enhance the accuracy of accident predictions but also provide a more nuanced understanding of the factors influencing maritime safety.

Precisely, the model's predictive capabilities hold promise for informing broader statistical studies on fisheries' accident rates. By providing insights into accident occurrence patterns and their potential drivers, the model can serve as a valuable resource for researchers seeking to delve deeper into the underlying factors contributing to maritime accidents.

At last, the methodology employed in this study holds promise for estimating the consequences of accidents, particularly in terms of fatalities and lost vessels. However, accurate estimation of fatalities would require, at least, access to specific and up-to-date data on the number of crew members regularly aboard vessels. While challenging, such data could enhance the model's predictive capabilities and contribute to a more comprehensive understanding of the human toll of maritime accidents. Additionally, extending the analysis to include estimates of lost vessels would provide valuable insights into the economic and environmental consequences of accidents.

In conclusion, while the model represents a significant advancement in accident prevention efforts within the Galician fishing industry, its broader applicability lies in its potential to inform statistical research on maritime accident rates. By leveraging historical data and predictive analytics, the model offers valuable insights into accident occurrence patterns, contributing to a deeper understanding of factors influencing maritime safety and informing proactive measures to reduce accidents in fisheries operations worldwide.

6 REFERENCES

Abebe, M.; Noh, Y.; Kang, Y.; Seo, C.; Kim, D.; Seo, J. Ship trajectory planning for collision avoidance using hybrid ARIMA-LSTM models. *Ocean Engineering* [online]. 15 July 2022, vol. 256, p. 111527. [Accessed: 14 May 2024]. eISSN: 1873-5258. Available at: https://dx.doi.org/10.1016/j.oceaneng.2022.111527

Aivazidou, E.; Politis, I. Transfer function models for forecasting maritime passenger traffic in greece under an economic crisis environment. *Transportation Letters* [online]. 2021, vol. 13, no. 8, p. 591-607. [Accessed: 14 May 2024]. eISSN: 1942-7867. Available at: https://www.sciencedirect.com/science/article/pii/S1942786722002703

Allegue, E. *Los pescadores gallegos : Análisis de la accidentabilidad en el sector pesquero gallego* [online]. Doctoral thesis. Universidade da Coruña. Departamento de Electrónica e Sistemas, 1993. [Accessed: 14 May 2024]. ISBN: 9788469280676. Available at: http://hdl.handle.net/2183/5625

Nacions Unidas, Asamblea General. T*ransformar nuestro mundo: la agenda 2030 para el desarrollo sostenible* [online]. Naciones Unidas, 21 de octubre de 2015. A/RES/70/1. 40 p. [Accessed: 14 May 2024]. Available at: https://unctad.org/system/files/official-document/ares70d1_es.pdf

Aveta, C.; Romano, C. A port planning study case: The freight strategy of the new central tyrrhenian sea port authority 2017-2020. *Transportation Research Procedia* [online]. 2020, vol. 45, p. 127-134. [Accessed: 14 May 2024]. eISSN 2352-1465. Available at: https://dx.doi.org/10.1016/j.trpro.2020.02.098

Benvenuto, D.; Giovanetti, M.; Vassallo, L.; Angeletti, S.; Ciccozzi, M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief* [online]. Apr 2020, vol. 29, p. 105340. [Accessed: 14 May 2024]. eISSN 2352-3409. Available at: https://dx.doi.org/10.1016/j.dib.2020.105340

Burnham, K.P.; Anderson, D.R. (eds.). *Model Selection and multimodel inference: a practical information-theoretic approach* [online]. 2nd ed. New York: Springer, 2002. ISBN 9780387953649. Also available at: https://doi.org/10.1007/b97636

Carvalho, N.; Keatinge, M.; Guillen, J. (eds.). *The 2019 annual economic report on the EU fishing fleet (STECF 19-06)* [online]. Luxenbourg: Publicatins Office of the European Union, 2019. [Accessed: 14 May 2024]. eISSN 2467-0715. Available at: https://data.europa.eu/doi/10.2760/911768

Comisión de Investigación de Accidentes e Incidentes Marítimos. Informes técnicos de accidentes e incidentes marítimos. In: Ministerio de Transportes y Movilidad Sostenible. *Comisión de Investigación de Accidentes e Incidentes Marítimos: publicaciones* [online]. Madrid: Gobierno de España. Ministerio de Transportes y Movilidad Sostenible, [2023?]. [Accesed: 14 september]. Available at: https://www.mitma.gob.es/organos-colegiados/ciaim/publicaciones

Confederación Española de Pesca. *Informe del sector pesquero español 2019* [online]. Madrid: Cepesca, [2020?]. [Accessed: 14 May 2024]. Available at: https://cepesca.es/wp-content/uploads/2019/06/INFORME-SECTOR-PESQUERO-ESPAN%CC%83OL-2019.pdf

Consultora Sinerxia. *Primera encuesta sobre situación sociolaboral y de contratación en el sector de la pesca en Galicia* [online]. A Coruña: Sinerxia, June 2018, 28 p. [Accessed: 14 May 2024]. Available at: https://sinerxia.es/wp-content/uploads/2018/11/ESTUDIO-SOBRE-LA-SITUACI%C3%93N-SOCIO-LABORAL-DEL-SECTOR_PESQUERO_GALICIA-2018.pdf

Cordón, E.; García, F.; García, J.J.. The determinants that cause small-scale vessels to exit fishing: The case of the spanish small-scale purse seine fishery. *Fisheries Research* [online]. September 2016, vol. 181, p. 155-162. [Accessed: 14 May 2024]. eISSN: 1872-6763. Available at: https://dx.doi.org/10.1016/j.fishres.2016.04.015

Cowpertwait, Paul S.P.; Metcalfe, Andrew V. *Introductory Time Series with R.* Dordrecht: Springer, 2009, 256 p. ISBN: 9780387886978. eISBN: 978-0-387-88698-5. Also Available at: https://doi.org/10.1007/978-0-387-88698-5

Cryer, J.D.; Chan, K.2008. *Time series analysis: with applications in R* [online]. New York: Springer, 2008, 491 p. ISBN: 9780387759586. eISBN: 9780387759593. [Accessed: 14 May 2024]. Available at: https://doi.org/10.1007/978-0-387-75959-3
Also available at: https://library.biblioboard.com/viewer/f2032f90-c05c-11ea-b48a-0a28bb48d135

Dirección General de Asuntos Marítimos y Pesca. *Fleet register* [database]. Fleet Register 1.1.0.0. Directorate General for Maritime Affairs and Fisheries, [2023?]. [Accesed: 14 september]. Available at: https://webgate.ec.europa.eu/fleet-europa/search_es

Dirección General de Asuntos Marítimos y Pesca. MDR vessel port codes. *In: Circabc* [online]. Version 1.26. DG Digit, 2023. [Accesed: 14 september]. Available at: https://circabc.europa.eu/ui/group/3cc8c417-0f2a-4eb4-8ff7-10d60638446a/library/9ad4e080-0b19-47bb-b77a-24a63cd2d7f7/details

Fernández-González, R.; Pérez-Pérez, M.; Hervés-Estévez, J.; Garza-Gil, M.D. 2022. Socio-economic impact of covid-19 on the fishing sector: A case study of a region highly dependent on fishing in spain. *Ocean & Coastal Management* [online]. 15 April 2022, vol. 221, p. 106131. [Accessed: 14 May 2024]. eISSN: 1873-524X. ISSN: 0964-5691. Available at: https://doi.org/10.1016/j.ocecoaman.2022.106131

García-Negro, M.D.C.; Rodríguez-Rodríguez, G.; González-Laxe, F. A importancia económica da pesca en galicia. *Revista Galega de Economía* [online]. 10 September 2018, vol. 27, no. 3, p. 35-48. [Accessed: 14 May 2024]. eISSN: 2255-5951. ISSN: 1132-2799. Available at: https://dialnet.unirioja.es/servlet/oaiart?codigo=6549962

Garza-Gil, M.; Surís-Regueiro, J.C.; Varela-Lafuente, M. Using input–output methods to assess the effects of fishing and aquaculture on a regional economy: the case of Galicia, Spain. *Marine Policy* [online]. November 2017, vol. 85, p. 48-53. [Accessed: 14 May 2024]. eISSN: ISSN: 1872-9460. ISSN: 0308-597X. Available at: https://dx.doi.org/10.1016/j.marpol.2017.08.003

Hornik, K.; Trapletti, A. T*series: time series analysis and computational finance.* [program]. Version 0.10-55. 6 December 2023. [Accessed: 14 May 2024]. Available at: https://CRAN.R-project.org/package=tseries

Hurvich, C.M.; Tsai, C.L. Regression and time series model selection in small samples. *Biometrika* [online]. June 1989, vol. 76, no. 2, p. 297-307. [Accessed: 14 May 2024]. Available at: https://doi.org/10.1093/biomet/76.2.297

Johnsen, J.P.; Vik, J. Pushed or pulled? understanding fishery exit in a welfare society context. *Maritime Studies* [online]. 26 April 2013, vol. 12, no. 4, p. 1-20. [Accessed: 14 May 2024]. ISSN 2212-9790. Available at: https://doi.org/10.1186/2212-9790-12-4

Lai, Y.; Dzombak, D.A. Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation. *Weather and Forecasting* [online]. June 2020, vol. 35, no. 3, p. 959-976. [Accessed: 14 May 2024]. eISSN: 1520-0434. ISSN: 0882-8156. Available at: https://doi.org/10.1175/WAF-D-19-0158.1

Ljung, G.M.; Box, G.E.P. On a measure of lack of fit in time series models. *Biometrika* [online]. August 1978, vol. 65, no. 2, p. 297-303. [Accessed: 14 May 2024]. Available at: https://doi.org/10.1093/biomet/65.2.297

Ministerio de Fomento. Real decreto 800/2011, de 10 de junio, por el que se regula la investigación de los accidentes e incidentes marítimos y la comisión permanente de investigación de accidentes e incidentes marítimos. In: *Boletín Oficial del Estado* [online]. Madrid: Agencia Estatal Boletín Oficial del Estado, 11 June 2011, no. 139, p. 60091-601082011. BOE-A-2011-10133. [Accessed: 14 May 2024]. eISSN: 2951-7915. Available at: https://www.boe.es/eli/es/rd/2011/06/10/800

Moreno, F.J.; Gómez-Cano, M. *Causas de los accidentes marítimos muy graves en la pesca 2008-2013* [online]. Madrid: Instituto Nacional de Seguridad e Higiene en el Trabajo (INSHT), November 2014, 67 p. [Accessed: 14 May 2024]. Available at: https://www.insst.es/documentacion/material-tecnico/documentos-tecnicos/causas-accidentes-maritimos-muy-graves-en-pesca-2008-2013-ano-2014

Penas Lado, E. *Quo Vadis Common Fisheries Policy?*. Hoboken: John Wiley & Sons, 2019. ISBN 9781119576860. eISBN:9781119576907. Also available at: http://doi.10.1002/9781119576907

Perez-Labajos, C.; Azofra, M.; Blanco, B.; Achutegui, J.; González, J. Analysis of accident inequality of the spanish fishing fleet. *Accident Analysis and Prevention* [online]. November 2006, vol. 38, no. 6, p. 1168-1175. [Accessed: 14 May 2024]. eISSN: 1879-2057. Available at: https://doi.org/10.1016/j.aap.2006.05.007

R Core Team. *R: A language and environment for statistical computin* [program]. Version 4.3.0. Austria: R Foundation for Statistical Computing, 2023. [Accessed: 14 May 2024]. Available at: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing

Rashed, Y.; Meersman, H.; Van de Voorde, E.; Vanelslander, T. Short-term forecast of container throughout: an ARIMA-intervention model for the Port of Antwerp. *Maritime Economics & Logistics* [online]. December 2017, vol. 19, no. 4, p. 749-764. [Accessed: 14 May 2024]. eISSN: 1479-294X. ISSN 1479-2931. Available at: https://doi.org/10.1057/mel.2016.8

Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. *Akaike Information Criterion Statistics.* Dordrecht: Springer, 1986. Mathematics and its Applications Series. ISBN 9027722536.

Sánchez, J.R.; De La Campa, R.M. Seguridad y salud laboral de los trabajadores de la flota pesquera gallega: retos del escenario post-pandémico. In: González Laxe, F.; Rodríguez Docampo, M.J. (eds.). *Congreso internacional marítimo sobre los factores de resiliencia de las regiones marítimas en el mundo post-covid : A Coruña, 5 y 6 de octubre, 2022: libro de actas*. A Coruña: Universidade Da Coruña. Servico de Publicacións, 2022, p. 145-165. [Accessed: 14 May 2024]. Available at: http://hdl.handle.net/2183/31857

Sobrino, J.M.; Oanta, G.A. The legal impact of the common fisheries policy on the galician fisheries sector. *Ocean & Coastal Management* [online]. January 2019, vol. 167, p. 87-99. [Accessed: 14 May 2024]. eISSN: 1873-524X. ISSN 0964-5691. Available at: https://doi.org/10.1016/j.ocecoaman.2018.10.011

Sui, Z.; Wen, Y.; Huang, Y.; Song, R.; Piera, M.A. Maritime accidents in the Yangtze River: a time series analysis for 2011-2020. *Accident Analysis and Prevention* [online]. February 2023, vol. 180, p. 106901. [Accessed: 14 May 2024]. eISSN: 1879-2057. Available at: https://dx.doi.org/10.1016/j.aap.2022.106901

Surís-Regueiro, J.C.; Santiago, J.L. Characterization of fisheries dependence in Galicia (Spain). *Marine Policy* [online]. July 2014, vol. 47, p. 99-109. [Accessed: 14 May 2024]. eISSN: 1872-9460. ISSN 0308-597X. Available at: https://dx.doi.org/10.1016/j.marpol.2014.02.006 DOI 10.1016/j.marpol.2014.02.006.

Wang, J.; Zhou, Y.; Zhuang, L.; Shi, L.; Zhang, S. A model of maritime accidents prediction based on multi-factor time series analysis. *Journal of Marine Engineering and Technology* [online]. May 2023, vol. 22, no. 3, p. 153-165. [Accessed: 14 May 2024]. ISSN 2046-4177. Available at: https://www.tandfonline.com/doi/abs/10.1080/20464177.2023.2167269

Yenidogan, I.; Cayir, A.; Kozan, O.; Dag, T.; Arslan, C. Bitcoin forecasting using ARIMA and PROPHET. In: *2018 3rd International Conference on Computer Science and Engineering : 20-23 September 2018, Sarajevo, Bosnia-Herzegovina* [online]. Piscataway : Institute of Electrical and Electronics Engineers, 2018, p. 621-624, [Accessed: 14 May 2024]. ISBN: 9781538678930. Available at: https://doi.org/10.1109/UBMK.2018.8566476