NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

IDENTIFYING PROBABLE MARITIME PIRACY EVENTS USING MARITIME INCIDENT DATA

by

Jun Yi Yeo

September 2023

Thesis Advisor:	Robert A. Koyak
Second Reader:	Ross J. Schuchard

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2023	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE IDENTIFYING PROBABLE MARITIME PIRACY EVENTS USING MARITIME INCIDENT DATA	**5. FUNDING NUMBERS**
6. AUTHOR(S) Jun Yi Yeo	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Maritime security and piracy incidents continue to pose significant risks to commercial navigation in spite of the fact that these incidents have decreased. In 2021, the strategically important Gulf of Guinea and Singapore Straits remained as the most prominent piracy hotspots. Using data from 2014, Desai and Shambaugh estimated the economic cost to the shipping industry from maritime piracy at more than U.S. $16 billion per year, which impacts international shipping industries and is often associated with violence in which hostages with injuries or deaths are involved. In addition, shipping companies have also privately succumbed to pirates' demands through failing to report incidents of attack, and these suppressions worsen maritime security.

As part of International Maritime Organization's (IMO) agenda to enhance maritime security, the organization publishes acts of piracy and armed robbery incidents as part of the Global Integrated Shipping Information System (GISIS). Our study aims to exploit the available dataset with geospatial and vessels' static details to provide descriptive statistical information about maritime piracy events, such as piracy hotspots. Time series analysis and clustering analysis is applied in the study to predict incident trends and identify piracy hotspots within the given time and space.

14. SUBJECT TERMS maritime security, piracy, data analysis, shipping, maritime incidents	**15. NUMBER OF PAGES** 69
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

IDENTIFYING PROBABLE MARITIME PIRACY EVENTS USING MARITIME INCIDENT DATA

Jun Yi Yeo
Civilian, DSTA
Bachelor of Science in Applied Physics, Nanyang Technological University, 2016
MBA, Quantic School of Business and Technology, 2022

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 2023

Approved by: Robert A. Koyak
Advisor

Ross J. Schuchard
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Maritime security and piracy incidents continue to pose significant risks to commercial navigation in spite of the fact that these incidents have decreased. In 2021, the strategically important Gulf of Guinea and Singapore Straits remained as the most prominent piracy hotspots. Using data from 2014, Desai and Shambaugh estimated the economic cost to the shipping industry from maritime piracy at more than U.S. $16 billion per year, which impacts international shipping industries and is often associated with violence in which hostages with injuries or deaths are involved. In addition, shipping companies have also privately succumbed to pirates' demands through failing to report incidents of attack, and these suppressions worsen maritime security.

As part of International Maritime Organization's (IMO) agenda to enhance maritime security, the organization publishes acts of piracy and armed robbery incidents as part of the Global Integrated Shipping Information System (GISIS). Our study aims to exploit the available dataset with geospatial and vessels' static details to provide descriptive statistical information about maritime piracy events, such as piracy hotspots. Time series analysis and clustering analysis is applied in the study to predict incident trends and identify piracy hotspots within the given time and space.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AHP	Analytical Hierarchical Process
ARIMA	Autoregressive Integrated Moving Average
BIMSTEC	Bay of Bengal Initiative for Multi-Sectoral Technical and Economic Cooperation
BMP	Best Management Practices
DBSCAN	Density-Based Spatial Clustering of Application with Noise
GISIS	Global Integrated Shipping Information System
HW	Holt-Winters
ICC	International Chamber of Commerce
IMB	International Maritime Bureau
IMO	International Maritime Organization
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MSC/Circ	Maritime Safety Committee/Circular
NPS	Naval Postgraduate School
ReCAAP	Regional Agreement on Combating Piracy and Armed Robbery against Ships in Asia

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Maritime piracy refers to the act of forceful boarding or hijacking of ship for various criminal purposes, such as theft, kidnapping, and ransom demands (United Nations Office on Drugs and Crime 2010). Piracy incidents have been reported globally, and these incidents post significant maritime security threats to the international shipping industries, which results in both economic losses and even loss of lives at times. Despite continuous efforts to combat piracy, piracy occurrences persist, which highlights the importance of sustained cooperation among stakeholders involved.

This study seeks to answer two main questions through utilizing data analysis techniques: First, it aims to measure the effectiveness of time series models to forecast incident occurrences. Second, it aims to identify prominent piracy hotspots clusters across the globe in the past decade.

With the historical maritime piracy and armed robbery incidents dataset from International Maritime Organization (IMO) during the period of 2010 to 2022, the data is extracted, transformed, and loaded for the purpose of time series analysis and clustering. The time series analysis models used in this study for forecasting monthly piracy incidents count include Naïve, Seasonal Decomposition, Holt Winters (HW) Exponential Smoothing, Auto Regressive Integrated Moving Average (ARIMA), and Ensemble. The performance metrics of Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE) are used to evaluate the models.

We find that the ARIMA model is optimal for forecasting monthly piracy occurrences for year 2023 in the study, with a MAPE of 31 percent. When comparing the forecasted results with the actual occurrences in the first half of 2023, a MAPE of 32.9 percent is determined.

Three clustering algorithms are also applied in this study for grouping of piracy incidents based on spatial similarity with the geographical coordinates in order to calculate the distance matrix. The algorithms that we considered are the following: Centroid-Based (K-means), Connectivity-Based (Hierarchical), and Density-Based

Spatial Clustering of Application with Noise (DBSCAN). We find that DBSCAN is effective for further analysis to study the changes in piracy concentration across the globe in the past decade.

Using data between the period of 2010 to 2022, this study identifies several prominent piracy hotspots, which include the following regions: Caribbean Sea, Gulf of Guinea, Gulf of Aden and Arabian Sea, Bay of Bengal, and South East Asia. Interestingly, piracy around South East Asia, particularly the Singapore Straits, have recently seen year-on-year increases of piracy occurrences (International Chamber of Commerce International Maritime Bureau 2023). Thus, there is a need for more attention and engagement from involved stakeholders to strengthen collaboration and enhance the maritime security within the region.

In summary, this thesis dives into the analysis of maritime piracy incidents using data analysis techniques of time series analysis and clustering algorithm. By examining piracy hotspots, predicting incident occurrences, and understanding influencing factors of maritime piracy events, the study enhances the understanding of maritime security challenges and potential strategies to combat piracy in different regions. The thesis concludes with the recommendations of enhancing maritime security through data-driven insights and future research possibilities.

References

International Chamber of Commerce International Maritime Bureau (2023) Piracy and Armed Robbery Against Ships. ICC IMB Piracy and Armed Robbery Against Ships Report-January-December 2022, United Kingdom, https://www.icc-ccs.org/reports/2022%20Annual%20IMB%20Piracy%20and%20Armed%20Robbery%20Report.pdf.

United Nations Office on Drugs and Crime (2010) 9. Maritime Piracy. Transnational Organized Crime Threat Assessment Report 2010 E.10.IV.6, New York, https://www.unodc.org/documents/data-and-analysis/tocta/9.Maritime_piracy.pdf.

ACKNOWLEDGMENTS

I would like to thank Professor Koyak for being my thesis advisor and dedicating his precious time to guide me throughout my one year in Naval Postgraduate School (NPS). I am also thankful for having LTC Schuchard for being my second reader.

I am grateful for the opportunity given by Defence Science and Technology Agency (DSTA), Singapore, to gain valuable experience in the field of Operations Research.

Lastly, I would like to thank my family for taking care of my precious wife, Jasmine, and my children, Jerynn and Jenelle, and for providing me with their fullest support while I was away during this period.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Maritime piracy encompasses the criminal act of boarding or hijacking a ship by force, with the intent to steal a maritime vessel or its cargo; to kidnap or harm crew members, or to rob or demand ransom from crews or passengers (United Nations Office on Drugs and Crime 2010). Over the years, piracy incidents have been reported in several parts of the world: Southeast Asia, which surrounds South China Sea; the Straits of Malacca and Singapore; as well as the Gulf of Guinea and Gulf of Aden, which surrounds Africa, are the most affected regions (International Maritime Organization [IMO] 2019).

Piracy poses a significant maritime security threat to the international shipping industry, estimated to inflict U.S. $16 billion a year in economic losses (Desai and Shambaugh 2021). Piracy incidents can also have serious non-monetary consequences, including physical harm and even death of crew members. In some cases, the owners of targeted vessels privately accede to pirates' demands or fail to report incidents of attack, and these suppressions may worsen maritime security.

Continuous efforts have been made internationally to help combat maritime piracy incidents, such as the deployment of naval forces, improvement on governance of maritime security, the use of private security personnel, and the implementation of Best Management Practices (BMP) by ship owners and operators (IMO 2019). Despite these efforts, maritime piracy incidents continue to persist, highlighting the need for continued vigilance and cooperation among stakeholders in the maritime industry.

In this thesis, we examine maritime piracy incidents data to identify piracy hotspots based on the twelve-year period of 1 January 2010 to 31 December 2022.

A. RESEARCH OBJECTIVE

The purpose of this study is to analyze maritime piracy events using data analysis techniques (such as spatio-temporal analysis, time series analysis) to identify piracy hotspots, estimate incident occurrences, and evaluate potential influencing factors resulting in piracy events.

This study seeks to answer the following questions:

1. What are the prominent piracy hotspots clusters and trends in the world over the past decade?

2. How well can time series analysis models forecast and predict the number of maritime piracy events based on historical trends?

B. THESIS ORGANIZATION

This thesis entails the following structures:

- Chapter II details the literature reviews of research papers with regards to data analysis techniques on maritime piracy incidents and relevant studies conducted in Naval Postgraduate School (NPS).

- Chapter III describes the data the necessary data transformation required for the clustering techniques and time series models applied.

- Chapter IV presents the results of the data analysis with the model comparison and selection.

- Chapter V summarizes the conclusions of the study and presents potential research areas for future developments.

II. LITERATURE REVIEW

In this chapter, we review the various research studies related to analysis of maritime piracy incidents, as well as several interesting studies relevant to maritime piracy that had been conducted at NPS.

A. SPATIO-TEMPORAL ANALYSIS TECHNIQUE

To examine the complex events of maritime incidents, Marchione and Johnson (2013) apply methods from spatial-temporal analysis of urban crime to provide empirical analyses of piracy events. In terms of spatial analysis, the authors applied spatial autocorrelation measures to determine statistical significance between clusters. In terms of temporal analysis, simple monthly time series are used to compare across certain sub regions, such as Gulf of Guinea, Somalia, etc. Mantel test statistics and the Knox test are also used to study the space-time dynamics of piracy event clusters.

B. MODELLING AND SIMULATION TECHNIQUE

Using modelling and simulation, Dabrowski and De Villiers (2015) apply dynamic Bayesian networks to generate piracy behavioral data with synthetic locations of pirate attacks. The authors show that this type of model is useful to generate realistic results to aid in training, testing and evaluating maritime piracy threats assessment.

Similarly, Pristrom et al. (2016) use a Bayesian network model to analyze the complexity of multiple influential factors such as weather (e.g., wind and rain conditions) and ship characteristics (e.g., speed, vessel type) to estimate the probability of successful piracy hijacking of ship in the Western Indian/East Africa region.

C. PROBABILITY PREDICTION-BASED TECHNIQUE

Nwokedi et al. (2022) use historical occurrences of events to estimate the probability of events related to piracy taking into account different ship types, as well as different types of piracy attack, such as kidnapping for ransom, hostage taking of crew members, etc.

Jin et al. (2019) employ the use of binary logistic regression to estimate the probability of piracy attack and probability of successful boarding. They utilize the labeling of a vessel being attacked or otherwise, and determine the probability of incident occurrence, which aids in the decision making of stakeholders in anti-piracy domain.

D. ANALYTICAL HIERARCHICAL PROCESS TECHNIQUE

Tsioufis et al. (2023) apply the Analytical Hierarchical Process (AHP) approach focusing on vessel type related to the transport of fossil fuels to predict piracy behavior in areas such as ports and territorial or international waters. In this approach, the authors rank the relative importance of various AHP criteria based on security, time, topography, and fuel consumption. They also determine that piracy events usually occur around territorial water.

E. PIRACY STUDIES CONDUCTED AT NPS

The maritime piracy issue is vast and dynamic, requiring a multidisciplinary domain to study the problem at different aspects. There have been continuous efforts to research various areas of maritime security domain to combat against piracy incidents.

For instance, several security studies have been conducted at NPS with regards to maritime piracy in certain piracy hotspots. Ajeagah (2022) studies the maritime security architecture in the region of Gulf of Guinea, and highlights the importance of regional collaboration, and challenges that need to be addressed to combat maritime piracy. In this study, he delves into the capabilities of individual nations around the region and highlights the need of resource enhancement.

Matthews (2015) focuses on the different maritime policies relating to Indonesia and explores on the impact of Indonesia’s sensitivity regarding its sovereignty around the Malacca Straits. The author also analyzes the actions taken by Indonesia to engage in multilateral cooperation or rejection of certain security initiatives.

Interestingly, in one of system engineering capstone projects at NPS, Cabungcal et al. (2014) adopt the system engineering process to develop cost-effective solutions to prevent pirates from boarding commercial vessels. They design a system to be placed on board vessels and generate simulations to support their recommendations.

III. METHODOLOGY

In this chapter, we introduce the dataset, how it is processed, and the techniques we use to analyze the data. All our analyses were conducted using the R statistical programming language (R Core Team 2023) with RStudio (Posit Team 2023).

A. DATA DESCRIPTION

The dataset consists of piracy and armed robbery incidents that were downloaded from the International Maritime Organization's (IMO) Global Integrated Shipping Information System (GISIS), and covers the twelve-year period of 1 January 2010 to 31 December 2022. The information extracted is shown in Table 1.

Table 1. Dataset Headers, Description and Data Type. Adapted from International Maritime Organization (2023).

Header	Description	Data Type
Date	Date of Incident	Character
Ship Name	Vessel Detail	Character
Ship Type	Vessel Detail	Character
IMO No.	Vessel Detail	Character
Area	Occurred in territorial waters, port area or international waters	Character
Latitude	Location Coordinates in Degree, Minutes, Seconds	Character
Longitude	Location Coordinates in Degree, Minutes, Seconds	Character
Incident details	Free text of Incident Report	Character
Consequences for crew	Free text of Incident Report	Character
Action taken by master/crew	Free text of Incident Report	Character
Reported	True/False	Boolean
Reported to	Free text of Incident Report	Character
Reporting State	Reporting Country/State	Character
Coastal State Action Taken	Free text of Incident Report	Character
Maritime Safety Committee/Circular (MSC/Circ)	Circular Report Number	Numeric

B. DATA PROCESSING

In this study, we focus on data preparation for time series analysis and clustering of piracy incidents. First, we transform incident date from text to year and month to tabulate the number of incidents in each year and month for the preparation of time series trend analysis of incident counts in the individual months. Second, we parse Latitude and Longitude into degrees, minutes, and direction for conversion to numerical latitude and longitude columns. This is done in preparation for identifying the individual incident points to generate clusters of piracy hotspots.

C. MODEL DESCRIPTION

1. Times Series Analysis

Through using the monthly maritime piracy incident counts from 2010 to 2022, we forecast the monthly forecast for up to one year ahead, from January 2023 to December 2023.

The time series patterns consist of trends, seasonal effects, cycles, and noise. Hyndman and Athanasopoulos (2018) and Yoshida (2023b) describe the time series components in the following definition:

1. Trend: The long-term data points which show increasing or decreasing trend.
2. Seasonal: The data is influenced by seasonal effects (e.g., incidents occur in certain month of year or time of day), and it has a fixed and known frequency.
3. Cycles: The data shows an upwards or downwards trend without fixed intervals. They include economic factors and other factors such as pandemics.
4. Noise: The variance that remains after removal of the previous three components.

Figure 1 depicts the general decreasing trend over the past 12 years with a peak in January 2011. The trend analysis is generated with the "stl()" function within R (Ripley

2023). Looking at seasonality, piracy incidents tend to peak around March to May and November to December each year.

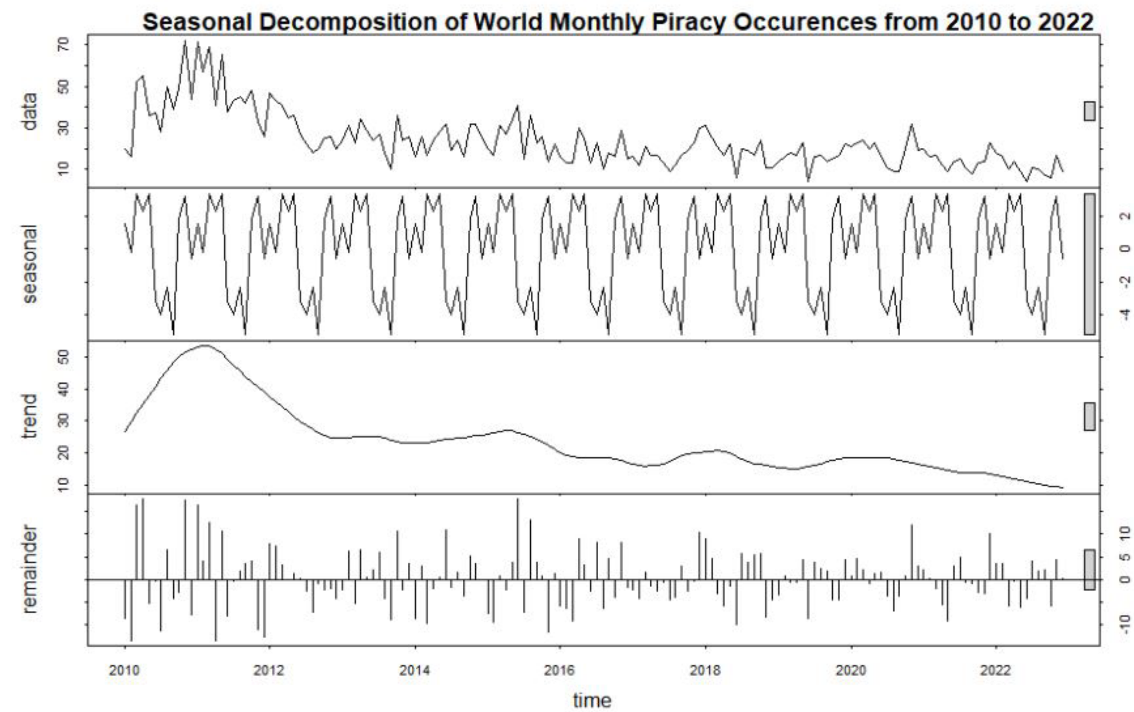

Figure 1. Seasonal Decomposition of World Monthly Piracy Incidents from 2010 to 2022

In this study, five forecasting techniques are applied and their respective performance are compared against using Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE). The models are fitted through a rolling horizon design to forecast every month from January 2013 to December 2022. This implies that a total of 600 models are fitted (five models over 120 periods). Below, we elaborate on further detail on the techniques (Hyndman and Athanasopoulos 2018; Yoshida 2023b) applied in R:

a. Naïve: This is a simple method of time series forecasting that sets the forecast to be the most recent past observed value. This usually works well when data mimic a random walk. The mathematical formulation is as such:

$$\hat{y}_{T+h|T} = y_T$$

where $\hat{y}_{T+h|T}$ is the estimate of y_{T+h} based on the previous observed data, y_T .

b. Seasonal Decomposition: This is the naïve forecast with seasonal adjustment.

c. Holt-Winters (HW): This is a type of exponential smoothing which utilizes weighted averages of historical data with weights that are reduced as observations become outdated. It is based on the consideration of three smoothing components (level, trend and seasonal). In this study, the HW additive method is adopted, and formulated as follows:

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)}$$

where l_t is level estimate at time t, b_t is trend estimate at time t, s_t is seasonal estimate for time t, h is the number of forecasts into future, and m is the seasonality frequency (monthly data implies $m = 12$). In R, the "HoltWinters()" function is applied to determine the parameters for forecasting (Meyer 2023).

d. Autoregressive Integrated Moving Average (ARIMA): This is a classical time series model that takes into account three terms (order of Auto Regressive (AR) component, degree of differencing (I) and order of Moving Average (MA) component). In R, the "auto.arima()" function is applied to determine the best parameters for the ARIMA model (Hyndman 2023).

e. Ensemble: We aggregate the multiple forecasting results from Seasonal Decomposition, HW, and ARIMA to obtain the average result and create a combined model.

After applying multiple techniques, the "forecast()" function is applied to generate the time series prediction (Hyndman et al. 2023). To evaluate the forecast models, the performance comparison measures, MAPE and MASE, are then computed for each model (Yoshida 2023b).

a. The mathematical formula for computing MAPE is the following:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|Y_t^* - F_t\right|}{\left|Y_t^*\right|}$$

where Y_t is time series observation at time t, and F_t is forecasted observation at time t. N is the last observation in the time series. As $Y_t = 0$ causes MAPE to be undefined, we set $Y_t^* = Y_t + 1$ to address the invalid MAPE issues. MAPE produces a range of values between 0 and 1, where a lower value implies a better predicted model.

b. The mathematical formula for computing MASE is as such:

$$MASE = \frac{\frac{1}{n}\sum_{t=1}^{N}\left|Y_t^* - F_t\right|}{\frac{1}{n}\sum_{t=1}^{N}\left|Y_t^* - Y_{t-12}^*\right|}$$

While the parameters are similar to MAPE, MASE is scale free, and the value do not depend on the scale of observation Y_t. This is due to MASE scale model based on performance of the Naïve model (as shown in denominator). Y_{t-12}^* implies that we are scaling against the Naïve model of the same month in the previous year. MASE producing a value of < 1 implies that the predictive power of the forecasted model is better than that of the Naïve model.

2. Clustering

Clustering is an unsupervised learning technique that we use to group maritime piracy incidents. This allows us to study the prominent piracy hotspots in recent years. As clustering algorithms generally use distance measures, the transformed latitude and longitude in the dataset are used to calculate a distance matrix. Note that Haversine distance is used for our distance measures. The main idea in the study is to bound the dataset within the geographical space to measure the similarity of items within a cluster and distance of data points between clusters.

Three main clustering algorithms are adopted for this study (; Yoshida 2023a):

a. Centroid-Based (K-means clustering): Using centroids to represent each cluster, observations are associated to clusters based on their distances to individual centroids. The number of clusters, parametrized by K, is determined using an elbow or silhouette plot to determine the optimal number of clusters:

i. Elbow plot: Identifying the point where more clusters do not help to minimize the sum of squared distance.
ii. Silhouette plot: Through computing of silhouette coefficient, the coefficients are compared across to determine the optimal clusters that are well separated.

b. Connectivity-Based (Hierarchical clustering): A hierarchical tree, also known as dendrogram, groups the observations based on similarity or dissimilarity measures. Divisive (top-down) or agglomerative (bottom-up) coefficients are computed as a measure of fit with a linkage method (Mechler et al. 2022). An optimal point is then chosen to cut the tree to group observations into its respective clusters:
i. Divisive clustering: Suitable for finding large clusters.
ii. Agglomerative clustering: Suitable for finding small clusters.

c. Density-Based (Density-Based Spatial Clustering of Application with Noise (DBSCAN)): DBSCAN clusters points by scanning a neighborhood around each point for a defined distance and minimum number of points to establish a link to a cluster (Hahsler et al. 2022). It allows high-density clusters that are non-linear to be determined. However, some points are not assigned a cluster as compared to the earlier algorithms.
i. The epsilon parameter defines the neighborhood distance between two points, and is determined via k-distance graph. For small epsilon chosen, the majority of the data may be considered outliers, while if large epsilon is chosen, clusters are merged easily. With the determined epsilon, the core data points are identified to form well defined clusters.
ii. The minPts parameter defines the minimum number of data points within the chosen epsilon distance. Minimum value of minPts must be at least 3. With a larger number of data points, a higher minPts value is preferred to form clusters that are better defined.

IV. RESULTS AND ANALYSIS

To address our research objective, we conduct analysis on the dataset based on the methods described in Chapter III. The results and analysis of our investigation into maritime piracy incidents are presented through the application of time series analysis and clustering techniques. Various models are evaluated and compared to select the best model.

For time series analysis, the temporal patterns and trends in piracy incidents over the period of January 2010 to December 2022 are examined to provide insights into the seasonality and long-term trend. Furthermore, the incident counts are aggregated into monthly counts instead of daily counts. This helps to uncover underlying patterns and to detect any shift in piracy activities.

For clustering analysis, we evaluate the strengths and weaknesses of the various clustering techniques described in Chapter III. Through the clustering algorithm, piracy incidents of similar geographic properties are grouped, from which we can identify distinct piracy hotspots and gain a deeper understanding of the spatial dynamics and risk zone. In this study, the visualizations of clusters are generated using the “mapview” package (Appelhans et al. 2022) in R.

Integrating temporal analysis into geospatial analysis also offers a comprehensive approach to unraveling the complexities of maritime piracy, which allows decision makers to make informed decisions and target high risk areas for intervention in maritime security.

A. TIME SERIES ANALYSIS

In the dataset, there are 3714 valid records between the period of 1 January 2010 to 31 December 2022 that contain valid dates. The dates are parsed and transformed for the purpose of time series analysis.

1. Model Comparisons for Time Series Analysis

Model comparisons for time series analysis with the use of a rolling horizon design allows us to evaluate the performance of the different forecasting models from January 2013 to December 2022. The training model uses three years of observations, from January 2010 to December 2012, and then the forecast starts from January 2013 onwards and iteratively generates the monthly forecast to update the model at each time step.

By comparing the forecasted values against the observed values for each month, the accuracy of each model can be evaluated against real-world data. We can also observe how well each model adapts to changing patterns, trends, and seasonality throughout the time series data.

Complementing the results of forecasted values in the rolling horizon design allows us to capture the dynamic nature of time series data and identify the most robust and reliable forecasting model. With the help of graphical plots, each model can be better assessed through the performance in different periods, and it enables analysts to make informed decisions in selecting appropriate model for prediction purposes. In addition, forecasting errors become rather obvious to detect, and we can validate the model visually to build confidence in the model chosen.

Furthermore, from Figure 1, it is observed that there might be unequal variances leading to heteroscedasticity problems such as biased estimates. Therefore, to stabilize the variances, the square root transformation is applied to the observed dataset for further model evaluation. Similarly, Figure 2 shows the trend analysis generated with square root transformation. It is observed that seasonality and trend effects remain similar to Figure 1, while the remainder shows that the variances are more proportional and equal across time in Figure 2 as compared to Figure 1. The count of incidents shown in Figure 2 is reduced due to square root transformation, which is then used for forecast comparison. There is a need to reverse the transformation of the forecasted results by squaring it back. This is to ensure that the measurement of incident counts is presented in the original scale.

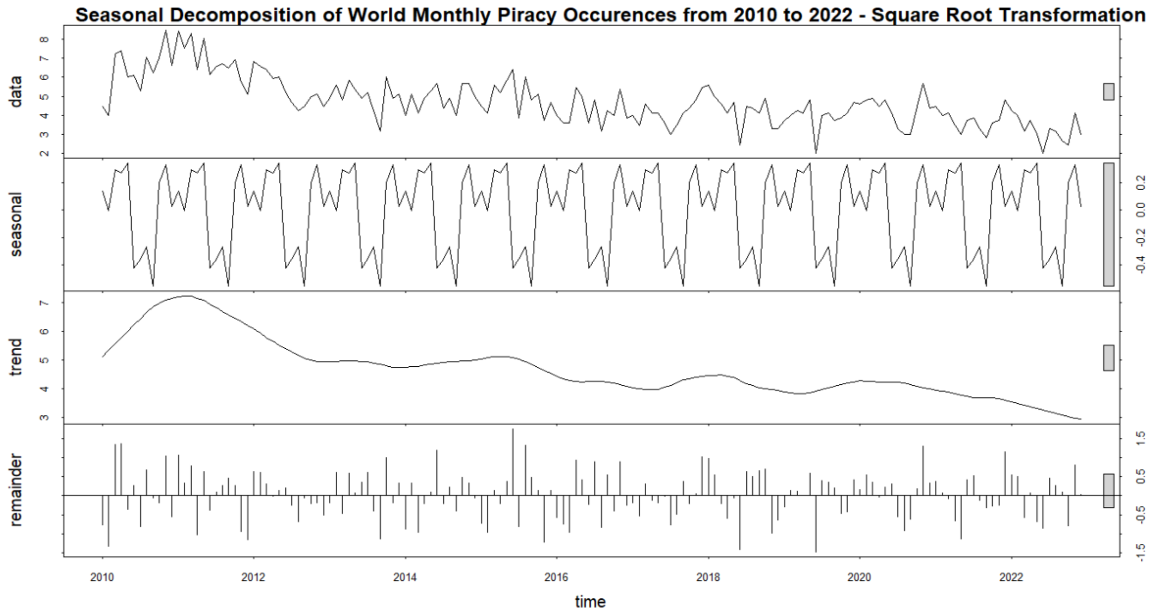

Figure 2. Seasonal Decomposition of World Monthly Piracy Incidents from 2010 to 2022 with Square Root Transformation

Looking at the results of maritime piracy incident monthly forecasts using the rolling horizon design, we break the time series analysis comparison into individual years for comparison purposes. For model comparison purposes, the square root transformed data is also used for forecasting, followed by squaring the forecasted result to ensure that the incident counts are on the proper scale.

In Figure 3, with the use of Naïve model, the discrepancies between the forecasted values and observed values are especially obvious during certain periods. In addition, it is observed that Naïve model with or without square root transformation results in the same forecasted results. Both forecasted results overlap with one another as shown in Figure 3.

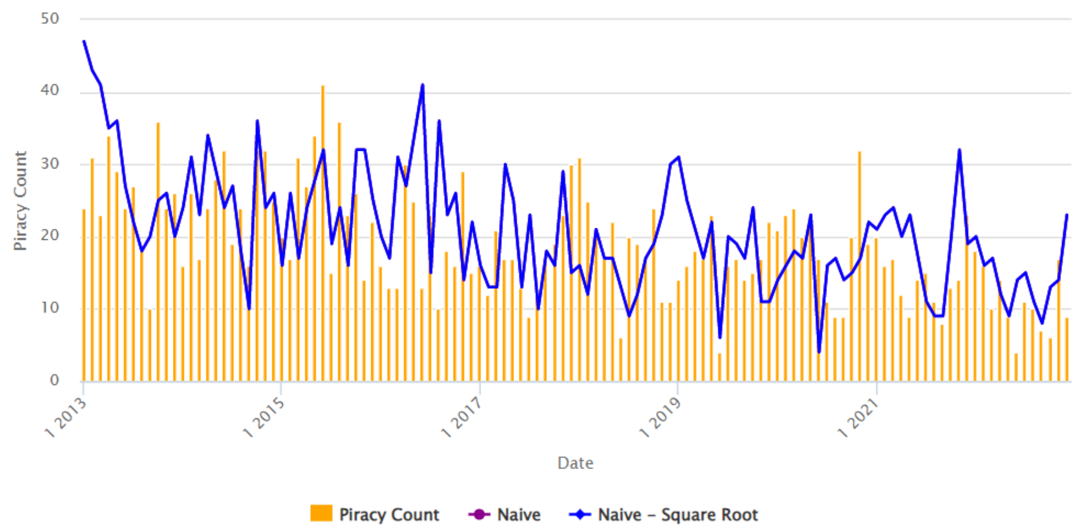

Figure 3. Rolling Horizon Design for Monthly Forecast of Maritime Piracy Incidents Using Naïve Model

From Figure 3, in the initial forecast period in 2013, the forecasted value is especially high. It is due to the high number of piracy incidents that occurred in 2012. which Naïve model use the previous observed value in 2012 for the current predicted value in 2013. Thus, if we shift the monthly forecast by a year to the left, it can be observed that the forecasted values replicate the observed values. Naïve model presents a straightforward forecasting approach and serves as the baseline for comparison and evaluation of the other more advanced forecasting model.

In Figure 4, Seasonal Decomposition model is applied to the same dataset. Seasonal Decomposition model aims to capture the trend and seasonality in time series data. In the forecasted results, both models with and without square root transformation seem to reflect the underlying decreasing trend of piracy incidents, as well as the recurring seasonal fluctuations. However, the movement of forecasted result appears to lag observed values by one to two months. Observing the differences between both models in Figure 4, the square root transformed results seem to bound within the range of the seasonal model without transformation.

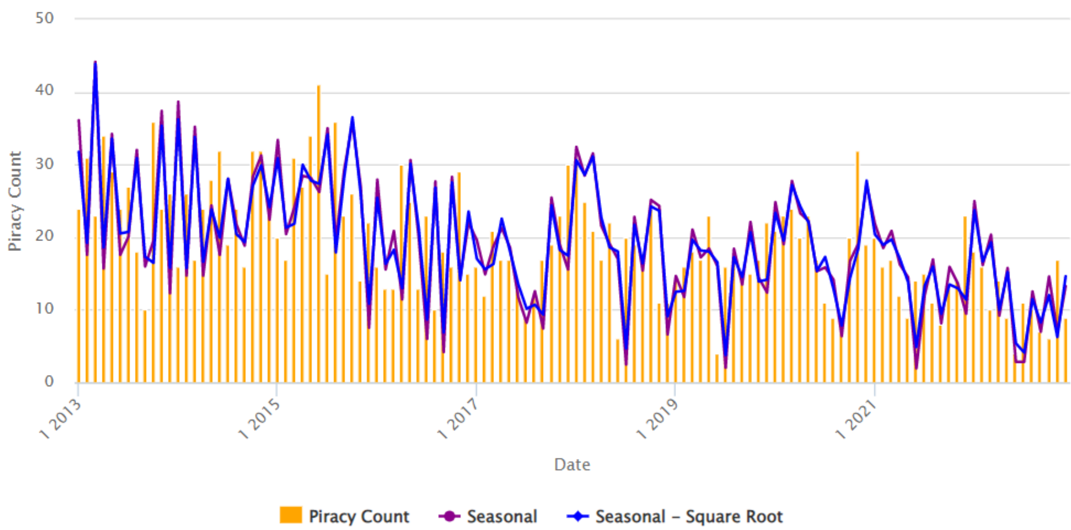

Figure 4. Rolling Horizon Design for Monthly Forecast of Maritime Piracy Incidents Using Seasonal Decomposition Model

In Figure 5, HW Exponential Smoothing model is applied. Similarly, this method considers the three main components (level, trend and seasonal) as described in Chapter III. Visually, the overall forecasted results do not capture the seasonality pattern well and the general trend between both models in Figure 5 is relatively similar.

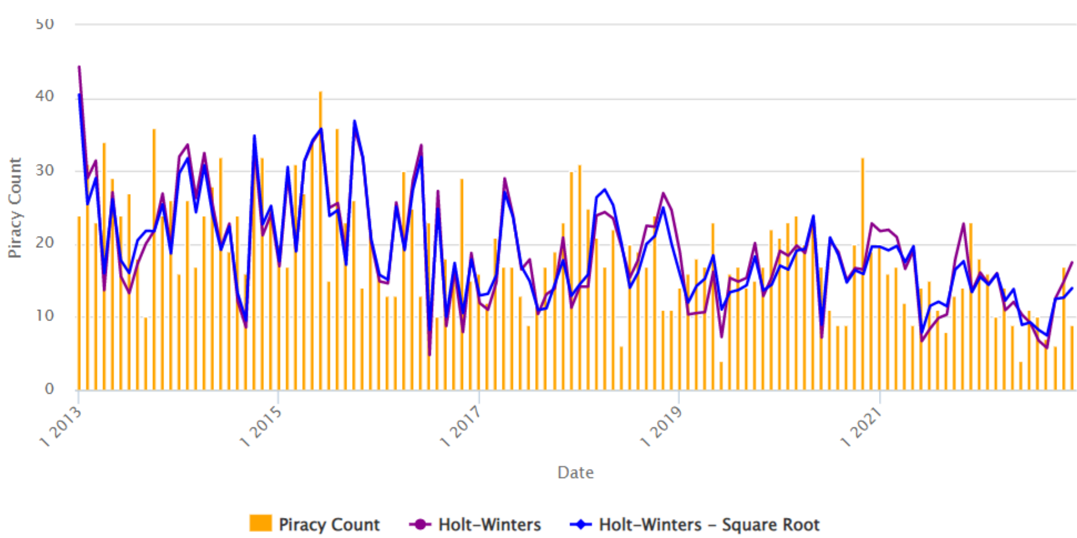

Figure 5. Rolling Horizon Design for Monthly Forecast of Maritime Piracy Incidents Using HW Exponential Smoothing Model

In Figure 6, the ARIMA model is applied. The forecasted values show hints of similar prediction following along the general decreasing trend. Although the results seem to be rather stable, which provide relatively constant forecast errors, it can be observed that the forecasted values seem to be over- or under-estimated throughout the time series. Both models with and without square root transformation in ARIMA do not have obvious discrepancies as compared to HW Exponential Smoothing.

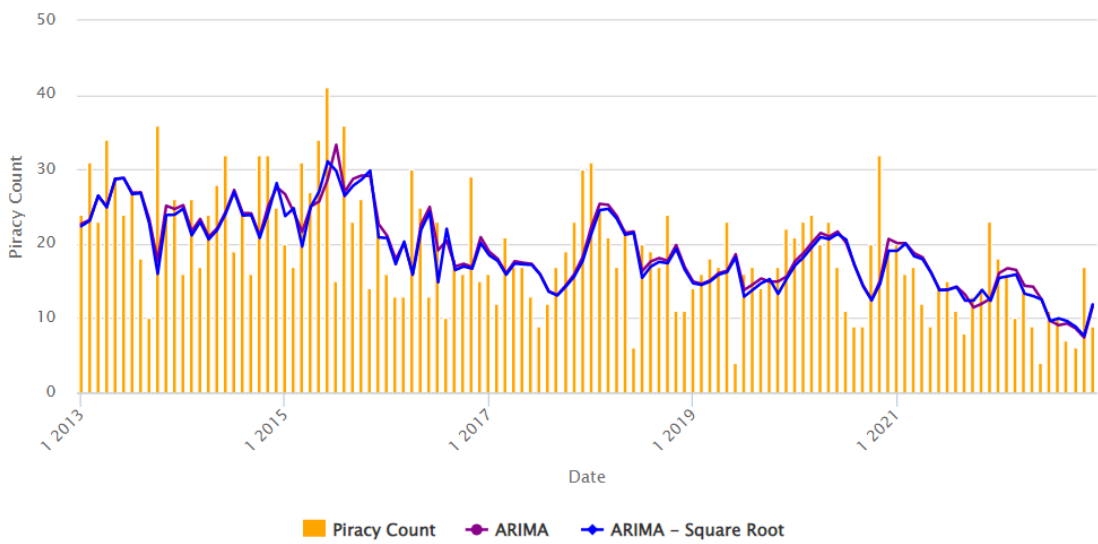

Figure 6. Rolling Horizon Design for Monthly Forecast of Maritime Piracy Incidents Using ARIMA Model

The Ensemble method of combining the earlier three forecasting model results (Seasonal Decomposition, HW Exponential Smoothing and ARIMA) to obtain the average forecasted value is applied and presented in Figure 7. By leveraging on the strength of each type of forecasting model, the Ensemble model aims to improve the accuracy and performance of the time series forecasted result. However, it also comes with the risk of being unreliable when models selected are not accurate in the first place.

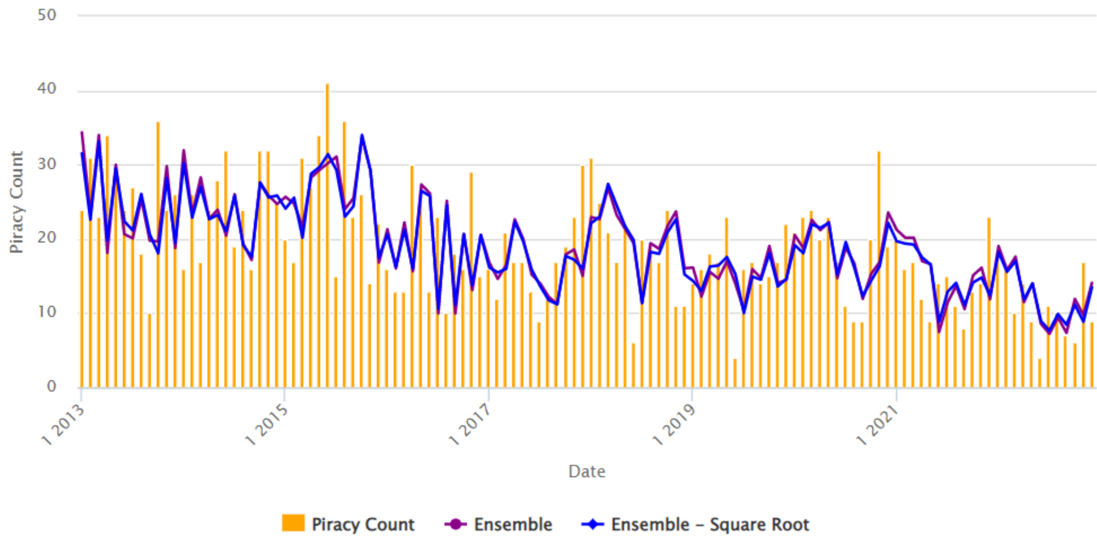

Figure 7. Rolling Horizon Design for Monthly Forecast of Maritime Piracy Incidents Using Ensemble Model

With the combination of the three models for both with and without square root transformation, it can be observed in Figure 7 that there is an opportunity for the Ensemble model to perform well. This is because the accuracy and performance of the three models are not high to begin with.

Other than visually comparing the various models, we also evaluate the performance metrics for model selection in this study.

2. Model Selection for Time Series Analysis

In our time series analysis, five forecasting techniques (Naïve, Seasonal Decomposition, HW Exponential Smoothing, ARIMA and Ensemble) are compared against each other through 1-step MAPE, 12-step MAPE, 1-step MASE and 12-step MASE. Noting that in this study, 1-step ahead is a month, while 12-step ahead is 12 months which is a year.

In 1-step ahead forecasting, the accuracy performance is compared through the forecasted value at specific time point to its actual observed value, which provides short-terms evaluation of model accuracy. While in 12-step ahead forecasting, the accuracy performance measures the forecasted value at twelve steps ahead into the future, which provides for long-terms evaluation of model accuracy.

The objective of model selection is to identify the most suitable forecasting model for predicting maritime piracy incidents accurately. In Table 2, it presents the model performance comparison results of our model selection process for time series analysis without transformation on incident counts.

Table 2. Time Series Model Performance Comparison – without Transformation

Model	1-step MAPE	12-step MAPE	1-step MASE	12-step MASE
Naïve	0.40	0.39	1.00	1.00
Seasonal Decomposition	0.39	0.43	1.10	1.10
HW Exponential Smoothing	0.39	0.41	1.00	1.10
ARIMA	0.31	0.35	0.78	0.87
Ensemble	0.32	0.36	0.84	0.92

Likewise, the model performance for the models forecasted through square root transformation are reflected in Table 3.

Table 3. Time Series Model Performance Comparison – with Square Root Transformation

Model	1-step MAPE	12-step MAPE	1-step MASE	12-step MASE
Naïve	0.40	0.39	1.00	1.00
Seasonal Decomposition	0.36	0.41	1.10	1.10
HW Exponential Smoothing	0.36	0.38	1.00	1.10
ARIMA	0.32	0.35	0.78	0.87
Ensemble	0.32	0.36	0.84	0.92

As described briefly in Chapter III, a lower MAPE indicates better forecasting accuracy, which suggests that forecasted values are closer to observed values. And MASE, MASE less than 1 results in a better predictive power as compared to the Naïve model. Among the techniques evaluated in this study, ARIMA stands out as the best model as shown in both Table 2 and Table 3, by achieving the lowest 1-step and 12-step ahead forecasting for both MAPE and MASE. This signifies that the accuracy in predicting short-term and long-term trend of piracy incident is better with ARIMA, outperforming the Naïve model benchmark.

The model performance result also shows that the Ensemble model performs reasonably well as compared to Naïve, Seasonal Decomposition and HW Exponential Smoothing model. However, the forecasting accuracy of the other models lags behind, which causes the combination of models to have a relatively lower performance and predictive power as compared to ARIMA model.

When comparing performance between the results between Table 2 and Table 3, both Seasonal Decomposition and HW Exponential Smoothing have slightly improved the MAPE score, where their 1-step MAPE has improved from 0.39 to 0.36. However, in ARIMA model, the 1-step MAPE score drops slightly from 0.31 to 0.32.

Based on recommendation by Swanson (2015), a MAPE greater than 0.25 is considered low accuracy prediction, and MAPE between 0.10 and 0.25 is considered low but acceptable accuracy. However, as a general guideline shown in Table 4, Allwright (2022) opined that the MAPE of 0.20 to 0.50 is considered acceptable and MAPE that is lower than 0.20 is considered good for forecasting since this implies that the predicted values for the dataset is less than 20 percent away from observed values.

Table 4. Interpretation of MAPE Results. Source: Allwright (2022).

MAPE	**Interpretation**
< 0.10	Very Good
0.10 to 0.20	Good
0.20 to 0.50	OK
> 0.50	Not Good

In this study, the ARIMA model returns a 1-step MAPE of 0.31 and 1-step MASE of 0.78, implying that the predicted value of dataset is approximately 31 percent away from observed value, and the predictive power is approximately 22 percent better than the Naïve model. Thus, the model selected for the study for monthly prediction of maritime piracy incidents is ARIMA.

3. Forecast of Maritime Piracy Incident Using ARIMA

With the selected model using the performance metrics, ARIMA is used for the one year forecast with a monthly prediction of maritime piracy incidents. Figure 8 depicts the fitted ARIMA trendline as shown in black, and the forecast as shown in blue. The shaded regions present the 80 percent and 95 percent prediction interval, which is the level of uncertainty in the forecasted result.

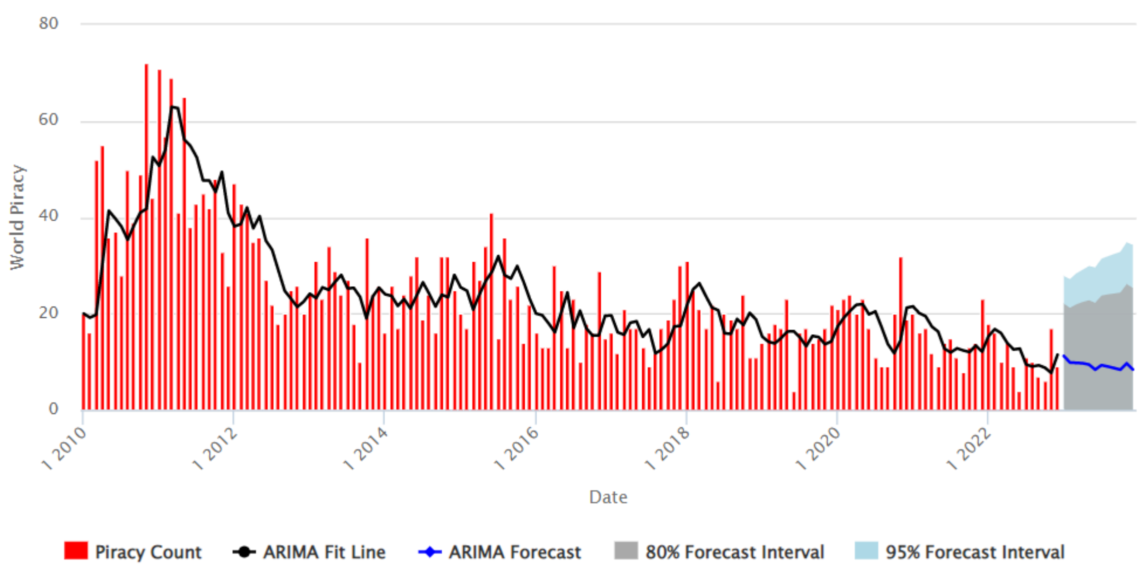

Figure 8. Forecast of Piracy Incidents in 2023 Using ARIMA Model

In the one year forecast of piracy incidents, it is observed that the level of uncertainty progresses as the prediction moves further. In addition, the trend of forecasted incidents counts show signs of decreasing pattern through 2023, with a slight increase in November 2023 possibly due to the seasonality effect.

B. CLUSTERING ANALYSIS

In clustering analysis, we evaluate the differences between the three main algorithms adopted in the study. Each of the clustering algorithms have its strengths and limitations, and we aim to choose one that is suitable for spatio-temporal analysis that we are conducting for maritime piracy events.

In this study, actual latitude and longitude points are used for geographical representation of piracy incidents. Therefore, the distance measure in this study utilizes Haversine distance, which is suitable for geospatial application. It measures the distance between the points with consideration of the great-circle distance based on curvature of Earth. The "geosphere" package (Hijmans 2022) in R is used to compute the Haversine distance between points.

Out of the 3714 records with valid date parsed, there are only 3349 records that contain latitude and longitude information between the period of 1 January 2010 to 31 December 2022. Thus, these 3349 records are used for the purpose of clustering analysis.

1. K-means Clustering

There is a need to determine the number of clusters that is required for the dataset with the use of K-means clustering algorithm (Hartigan and Wong 1979) in R. As the entire globe may consist of many different regions of piracy hotspot, the maximum number of clusters to be explored is set at 50. The sum of squared distance of clusters is computed for elbow plot, while the average silhouette width is computed for silhouette plot to determine the optimal K parameter (number of clusters).

Figure 9 shows the elbow plot with the sum of squared distances against the number of clusters, K. From Figure 9, the elbow point where the sum of squared distances decreases significantly and starts to slow down when K = 5. This indicates that the optimal K is 5, where it strikes a balance of having compact clusters and avoids having too many clusters.

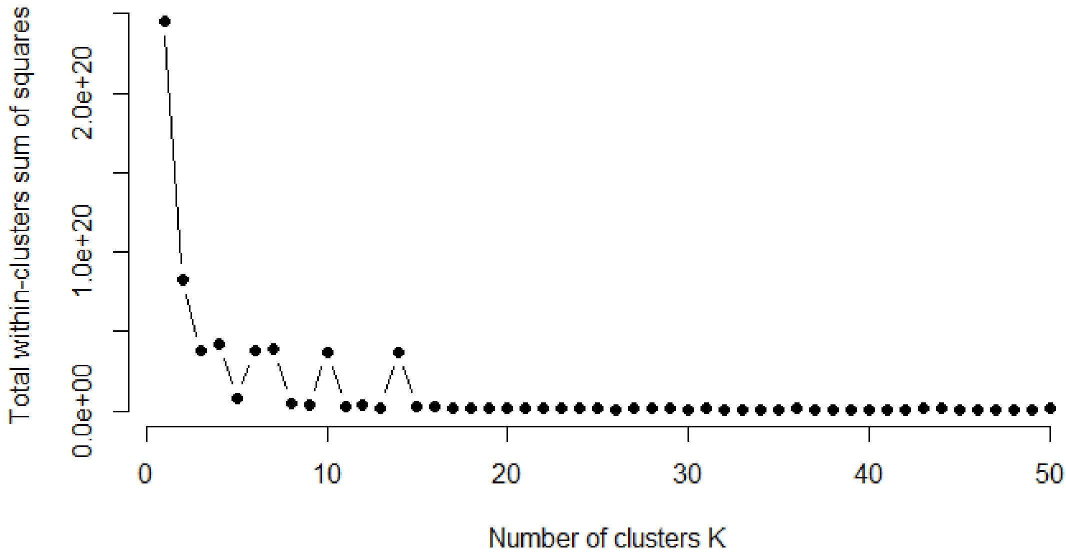

Figure 9. Elbow Plot

Figure 10 illustrates the silhouette plot, where silhouette width represents the clustering quality. Higher silhouette width implies that clusters are well separated while lower implies that clusters may be overlapping and are poorly separated. Similar to the determined optimal K in elbow plot, the silhouette plot also shows that the optimal number of clusters is approximately five.

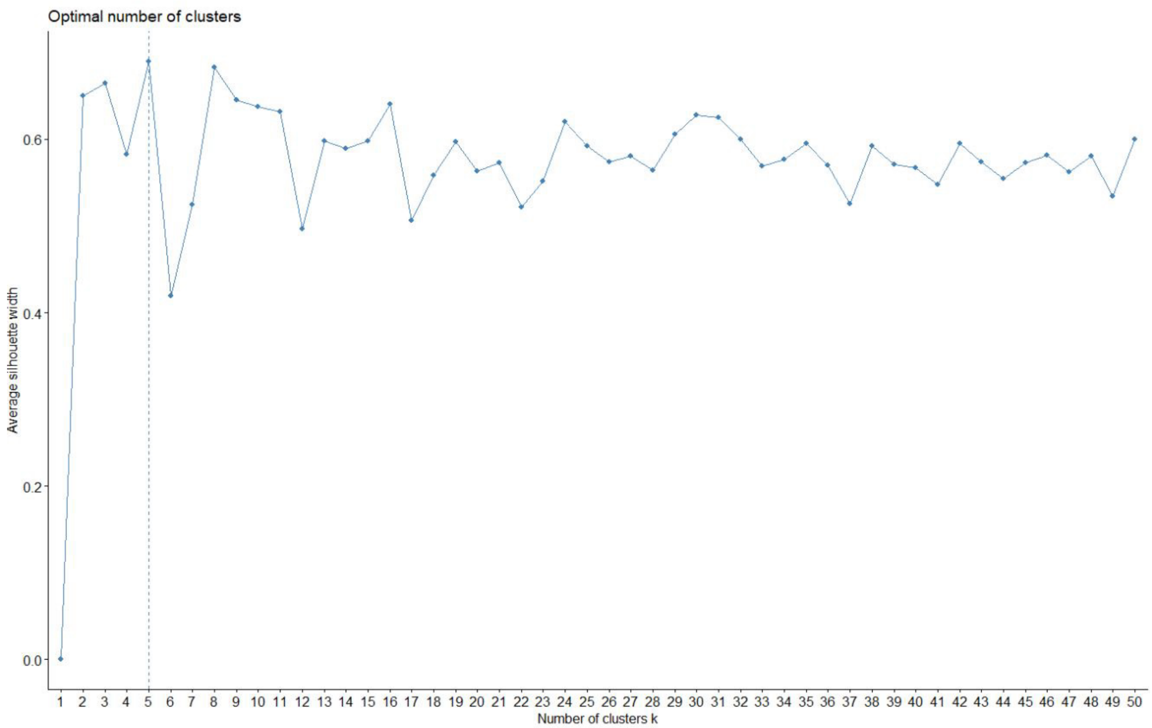

Figure 10. Silhouette Plot

Thereafter, the initial clustering with K-means algorithm proceeds with the visualization of clusters generated. The K-means clustering algorithm iteratively assigns each point to the nearest centroid and updates the centroids based on the mean of the assigned points. Figure 11 shows the results of K-means clusters, splitting the piracy event into five main regions.

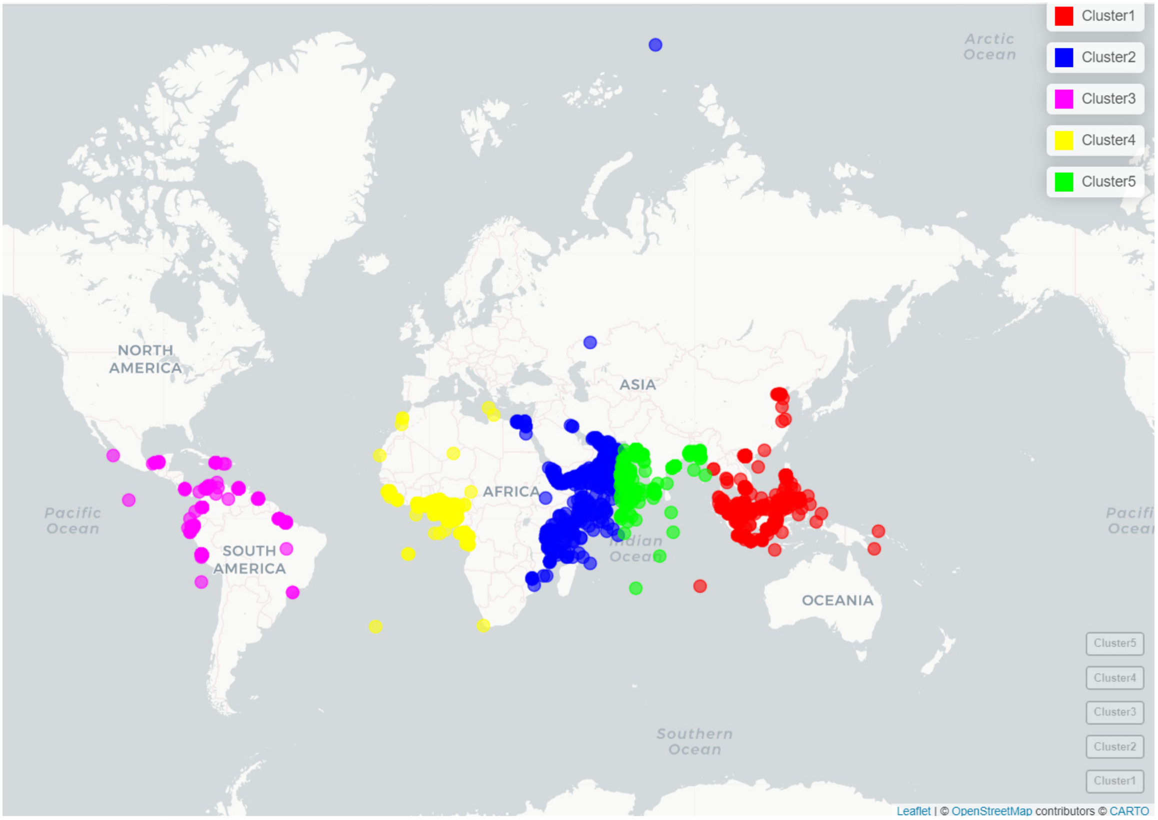

Figure 11. K-means clustering

From Figure 11, it is observed that maritime piracy incidents are split into the following five main regions:

- Cluster 1: South East Asia along Strait of Malacca and Strait of Singapore, Indonesian and Philippine's water, and East Asia along Yellow sea and South China Sea
- Cluster 2: East Africa and Middle East consisting of Gulf of Aden and part of Arabian Sea
- Cluster 3: Between North America and South America
- Cluster 4: West Africa around Gulf of Guinea
- Cluster 5: South Asia with part of Arabian Sea and Bay of Bengal

In addition, the clusters separated do not seem to fit very well with geospatial data that possibly are irregular shapes, since K-mean clustering may struggle to fit clusters with varying densities. The separation between Cluster 2 (in blue) and Cluster 5 (in green) is observed and may imply that the cluster is separated sub-optimally.

2. Hierarchical Clustering

In hierarchical clustering, both agglomerative coefficient and divisive coefficient are computed to evaluate the measure of fit. For both agglomerative and divisive coefficient, it ranges between 0 to 1, and a higher value is preferred for more balanced and distinct clusters (Boehmke and Greenwell 2020).

The tabulated measure of fit coefficient is shown in Table 5. The agglomerative hierarchical clustering with linkage type, Ward's minimum variance, has the best fit of 0.9999212, and is selected to plot the dendrogram.

Table 5. Hierarchical Clusters Distance Measures

Partition Approach	**Linkage Type**	**Coefficient Value**
Agglomerative	Complete	0.9984316
Agglomerative	Average	0.9979099
Agglomerative	Single	0.9930026
Agglomerative	Weighted	0.9975309
Agglomerative	Ward	0.9999212
Divisive	-	0.9986764

The dendrogram illustrates the individual data points being split or merged into the various branches based on the measure of similarity or dissimilarity in Figure 12, and it appears to lack legibility due to the high number of observations of maritime piracy incidents. With Ward's minimum variance method, it aims to minimize the total within-cluster variance during the clustering process and create clusters that are more compact.

To determine the optimal number of clusters for a dendrogram, it is required to determine the height to cut the dendrogram. In this study, we choose to partition the agglomerative hierarchical clustering model into four clusters.

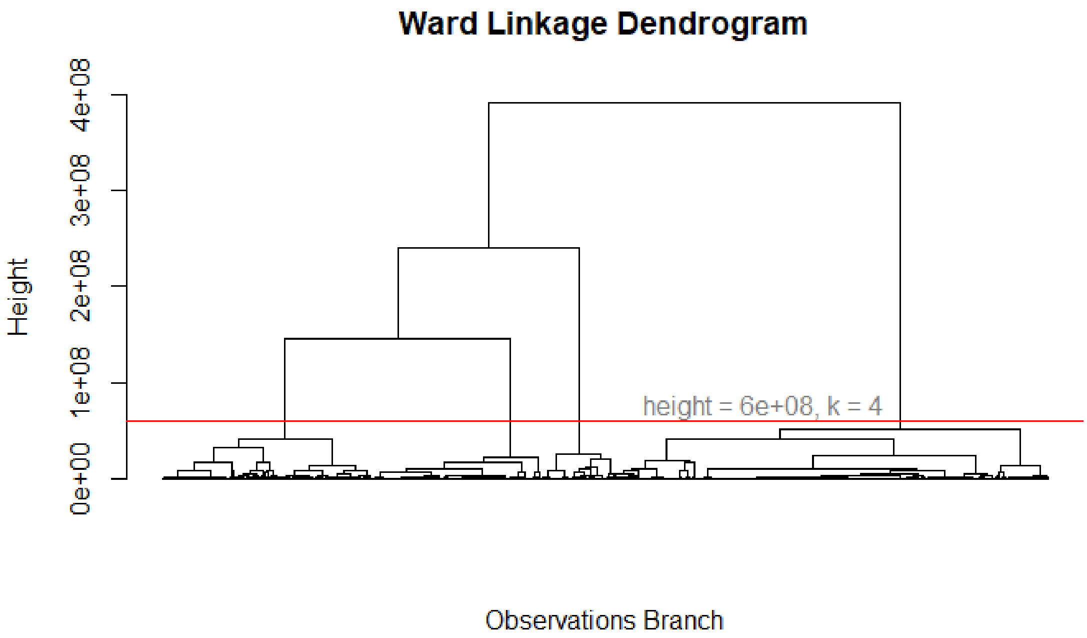

Figure 12. Dendrogram of Agglomerative Hierarchical Clustering with Ward Linkage

Figure 13 shows the visualization of hierarchical clustering model, which split the piracy incidents into 4 main regions:

- Cluster 1: East Africa and Middle East consisting of Gulf of Aden and Arabian Sea
- Cluster 2: South East Asia along Strait of Malacca and Strait of Singapore, Indonesian and Philippine's water, East Asia along Yellow sea and South China Sea, and South Asia around Bay of Bengal
- Cluster 3: West Africa around Gulf of Guinea
- Cluster 4: Between North America and South America

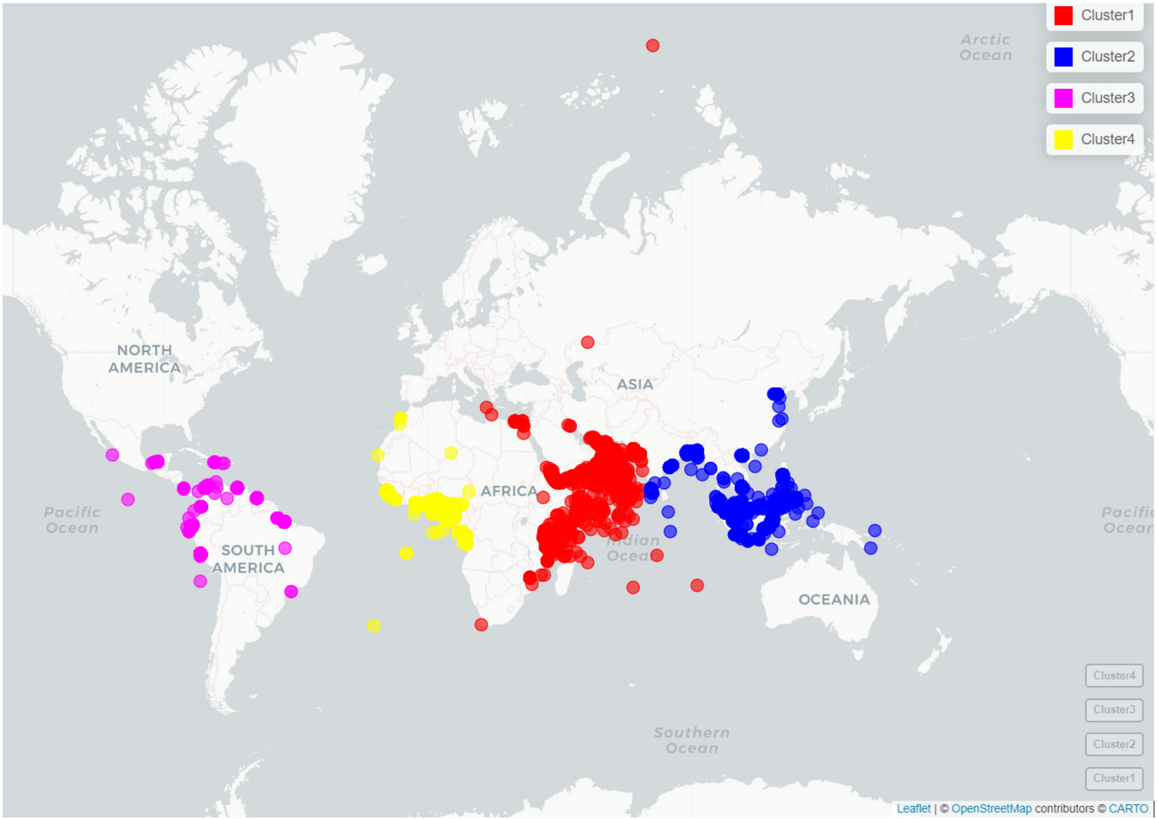

Figure 13. Hierarchical Clustering

When comparing between both K-means clustering and hierarchical clustering, it seems that hierarchical clustering is more appropriate. Figure 13 exhibits a clearer and more defined separation between Cluster 1 (in red) and Cluster 2 (in blue).

3. DBSCAN Clustering

DBSCAN is a popular approach for geospatial analysis due to its capability to effectively cluster high density region, as well as non-linear boundaries. However, this approach does not allocate all observations into a cluster unlike the previous two. It is also necessary to determine the parameters of epsilon and minPts. In this study, a knee plot is applied to determine epsilon with the use of "kNNdistplot()" function in the dbscan package (Hahsler et al. 2022) as shown in Figure 14. The minimum number of nearest neighbors within the epsilon radius used for the distance calculation is set to 15.

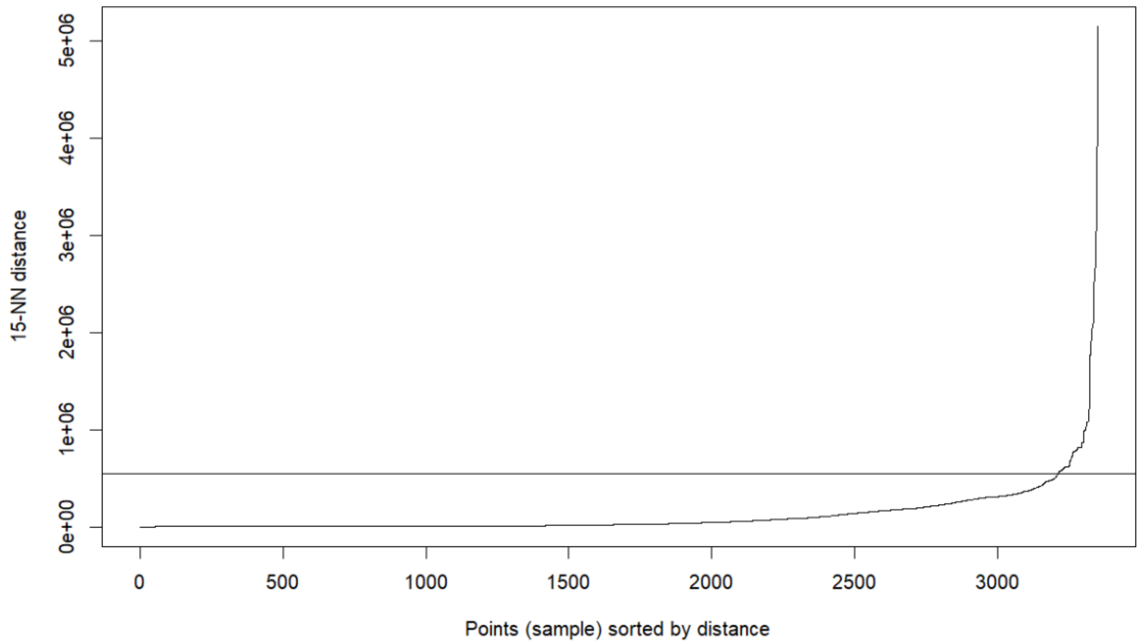

Figure 14. Knee Plot

From Figure 14, the knee plot depicts the k-nearest neighbor (kNN) distances of all points, arranged in ascending order from smallest to largest. Based on the knee plot, we visually identify the point of highest curvature where the kNN distance deviation starts to rise sharply. Thus, the optimal epsilon value is determined to be 550000.

With the defined parameter inputs of epsilon = 550000 and minPts = 15, the DBSCAN algorithm generates the outputs with 14 clusters as presented in Figure 15. A convex hull of 101 noise data points is also formed.

```
DBSCAN clustering for 3349 objects.
Parameters: eps = 550000, minPts = 15
Using unknown distances and borderpoints = FALSE
The clustering contains 14 cluster(s) and 101 noise points.

  0    1    2    3    4    5    6    7    8    9   10   11   12   13   14
101  746 1377   52   32  581  145   51   29   64   83   24   31   15   18
```

Figure 15. Parameters and Results of DBSCAN clustering

Likewise, the generated outputs are also displayed visually in Figure 16. Using DBSCAN algorithm, piracy incidents can be distinguished between clusters of high density and those with low density. There are three prominent piracy hotspots of higher density observed at Cluster 1 (in red), Cluster 2 (in blue) and Cluster 5 (in green). Outliers which have been classified as noise are also listed in the convex hull layer.

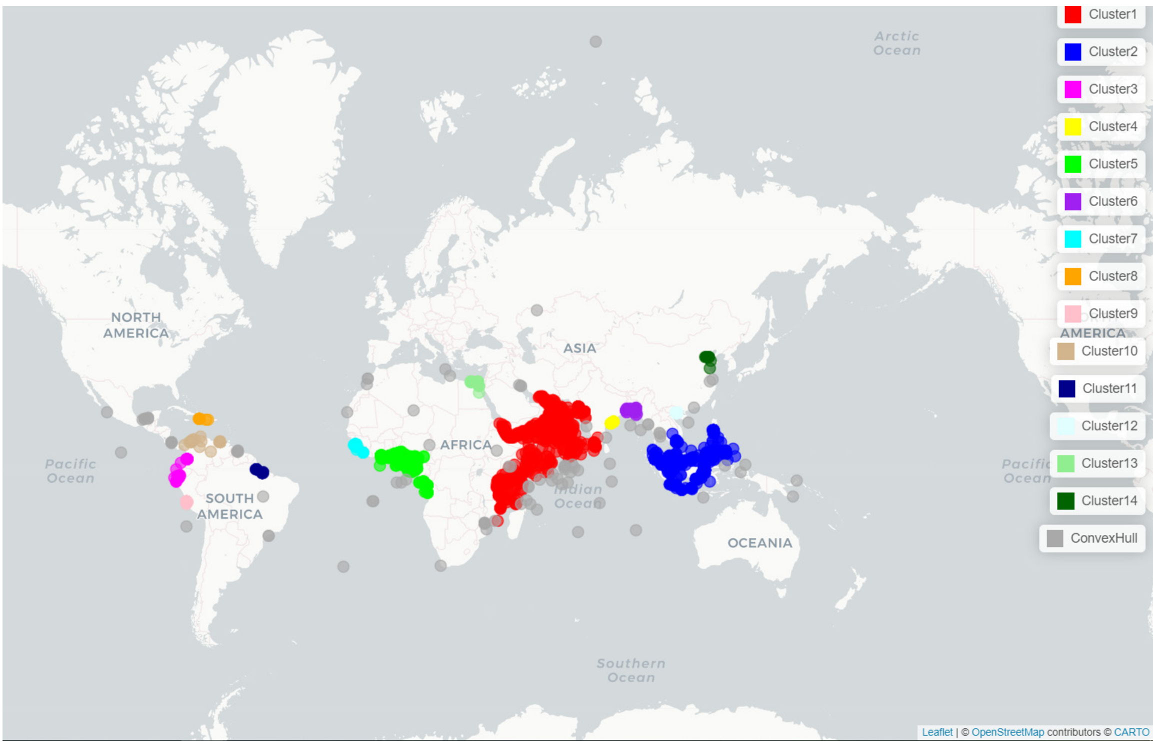

Figure 16. DBSCAN Clustering

The 14 clusters are divided into the following areas:

- Cluster 1: East Africa and Middle East consisting of Gulf of Aden and Arabian Sea
- Cluster 2: South East Asia along Strait of Malacca and Strait of Singapore, Indonesian, Southern Vietnam and Philippine’s water
- Cluster 3: Around West of Ecuador and West of Columbia

- Cluster 4: Around Visakhapatnam and Kakinada of India
- Cluster 5: Gulf of Guinea
- Cluster 6: Between southeastern coast of Bangladesh and Eastern region of India
- Cluster 7: Around Conakry (Capital of Guinea), Freetown (Capital of Sierra Leone) and Monrovia (Capital of Liberia)
- Cluster 8: Around Port-au-Prince (Capital of Haiti) and Santo Domingo (Capital of Dominican Republic)
- Cluster 9: Around Lima (Capital of Peru)
- Cluster 10: Northern Colombia and Northern Venezuela
- Cluster 11: Macapá and Belém of Brazil
- Cluster 12: Northeastern Vietnam
- Cluster 13: Port of Alexandria and Suez Canal around Egypt
- Cluster 14: East coast of China

4. Model Selection for Clustering Analysis

To select the appropriate clustering algorithm in geospatial analysis, we consider the nature of our geospatial dataset and how each algorithm differs in results generated. Among three algorithms investigated in this study, only DBSCAN can eliminate noise to generate clusters, enabling us to concentrate on higher density piracy hotspots to meet research objective.

In addition, as the number of piracy clusters is unknown, DBSCAN is a suitable option because it allows for ease of specification of the optimal number of clusters for conducting further analysis, unlike K-means and Hierarchical clustering. DBSCAN is also commonly adapted for geospatial analysis.

C. SPATIO-TEMPORAL ANALYSIS

With spatio-temporal analysis applied in maritime piracy, we analyze the interaction between time and space factors and detect clusters of incidents in specific regions and identify factors that affect piracy activities. This allows us to identify piracy hotspots across the globe and its changes in the concentration of piracy events across time.

To gain insight into the distribution and evolution of piracy events in specific regions, DBSCAN is applied for further analysis on the dataset. The dataset is similarly split into yearly piracy occurrences with the Haversine distance matrix computed, then the parameter of epsilon and minPts are determined for the yearly dataset. With these parameters' inputs, we then evaluate the cluster of the regions in a yearly manner.

The minimum number of nearest neighbors within the epsilon radius used for the distance calculation is standardized at 10 across the yearly data. The heuristic approach of using knee plot to determine epsilon is also adopted yearly as shown in Figure 17. The epsilon determined for the individual years are 2010 (500000), 2011 (750000), 2012 (1000000), 2013 (1000000), 2014 (1500000), 2015 (1000000), 2016 (750000), 2017 (1000000), 2018 (1000000), 2019 (1000000), 2020 (1000000), 2021 (1000000), 2022 (1000000).

With the defined parameter inputs of epsilon listed and minPts = 10, the DBSCAN algorithm is applied to generate the outputs for the yearly clusters as presented in the Appendix. The generated clusters for the individual years are then displayed visually on the world map in Figure 18 and Figure 19. Note that the cluster for convex hull is treated as noise and is not plotted in both Figures 18 and 19. The following packages are used in R for the illustration and evaluation in the study: ggplot2, ggmap, ggpubr, gganimate and gifski (Wickham et al. 2023; Kahle et al. 2023; Kassambara 2023; Pedersen and Robinson 2022; Ooms 2023).

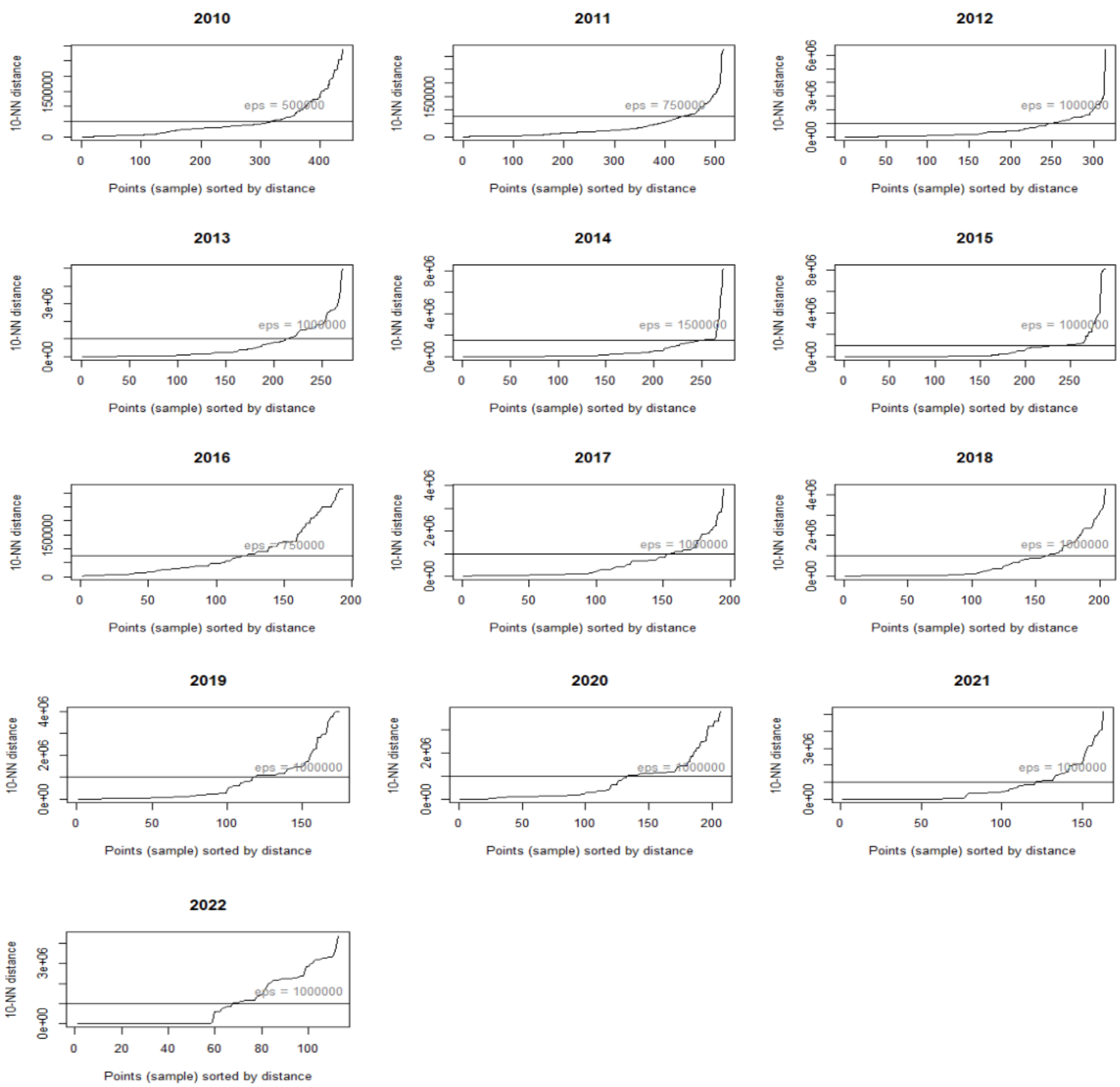

Figure 17. Knee Plot of Yearly Maritime Incidents from 2010 to 2022

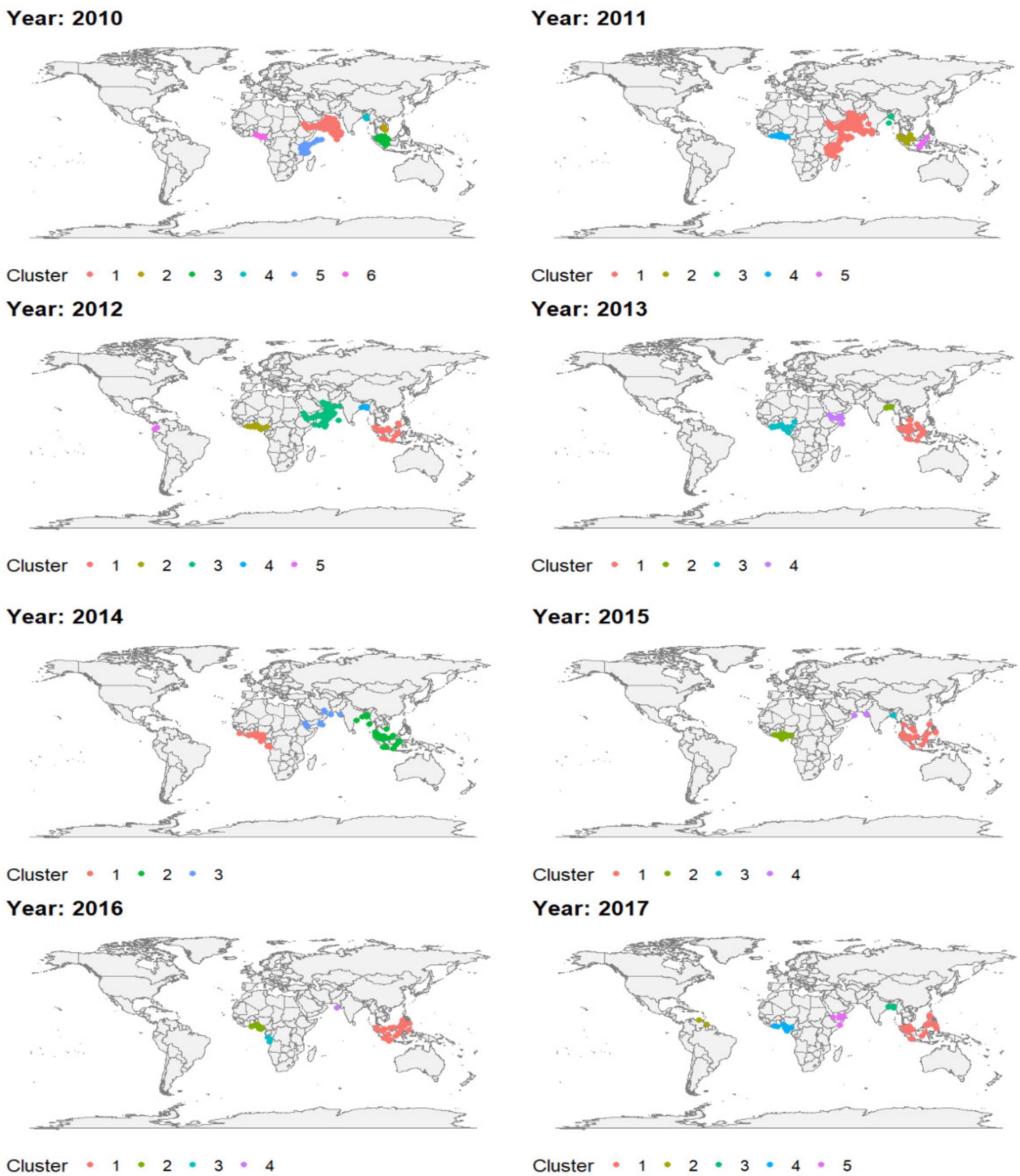

Figure 18. DBSCAN Clustering from 2010 to 2017

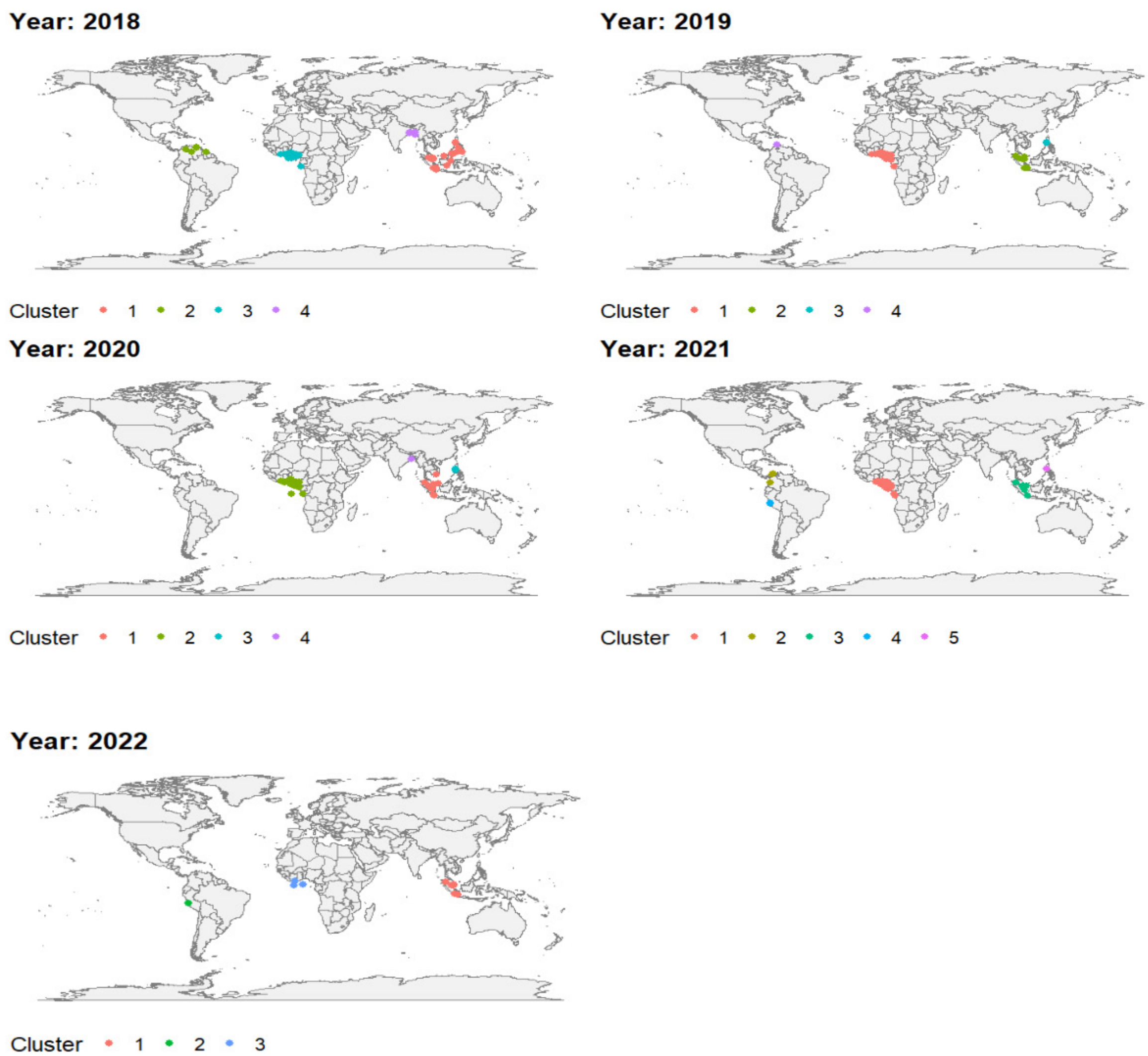

Figure 19. DBSCAN Clustering from 2018 to 2022

To evaluate the various clusters from 2010 to 2022, the plots in Figure 18 and Figure 19 are also used to form an animated image to view changes in piracy hotspot. We then observed the changes in concentration of piracy incidents across the individual years.

The piracy hotspots are mainly grouped into these areas:

- Caribbean Sea
- Gulf of Guinea
- Gulf of Aden and Arabian Sea

- Bay of Benegal
- South East Asia

1. Piracy around the Caribbean Sea

Caribbean Sea lies at the western Atlantic Ocean, and is well-known for its piracy activities with presence of drugs and weapons trafficking activities going on in the past decades (Dryad Global 2023). Compared against other piracy hotspots in the past decade, it appears that the Caribbean cluster is comparatively smaller. However, DBSCAN generated piracy clusters around Caribbean Sea and Peru in the past six years.

Gard (2023) reported the five-year statistics of piracy and armed and robbery incidents around South and Central America as illustrated in Figure 20, with half of the incidents contributed by Peru. Although the number of incidents declined in 2022 as compared to 2021, it is noteworthy that the incidents have been increasing in previous years from 2017 to 2022. There is a need for neighboring countries to stay vigilant and these incidents tend to happen at night around ports' anchorage area.

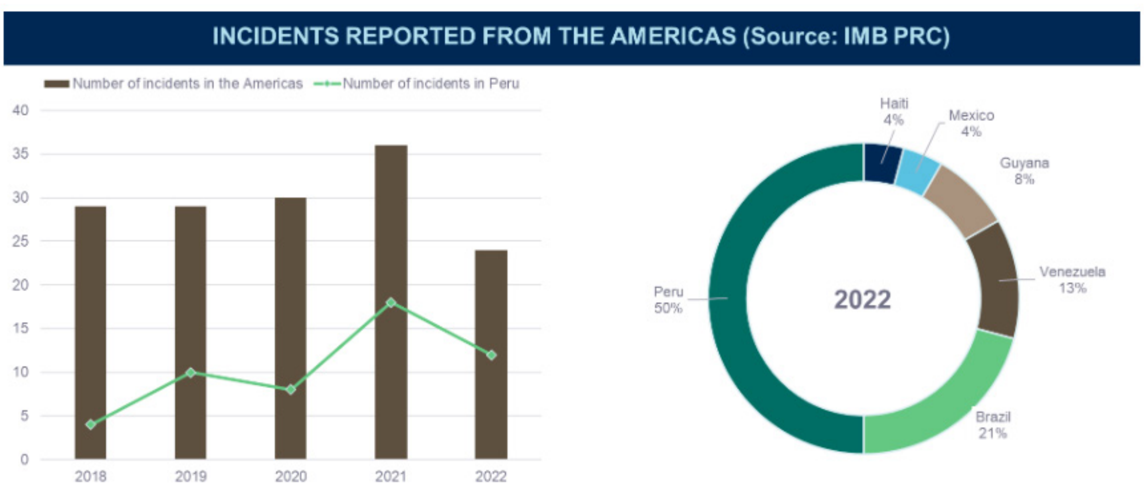

Figure 20. Five-Year Statistics of Piracy Incident Reported from South and Central America. Source: Gard (2023).

2. Piracy around the Gulf of Guinea

Gulf of Guinea is part of western Africa coast along 17 countries and is prominent for black-market involving oil and gas industries. There are around 20 commercial ports in this region which constitute a quarter of the total maritime activities in Africa, serving as the primary transportation route for the two Africa's main oil producers, namely Nigeria and Angola. In the past decade, Gulf of Guinea has been one of the piracy hotspots around the world, and accounts for 43 percent of reported piracy incidents and 95 percent of the maritime kidnapping incidents around the world in 2021 (Teixeira and Pinto 2022).

In 2022, piracy incidents around Gulf of Guinea have declined due to the increasing efforts from regional and international partners to combat against piracy. With the incorporation of the Yaoundé Architecture, it has enhanced the interregional maritime security around Gulf of Guinea states through maximizing the limited naval resources and enhancing information exchange (United Nations Press 2023).

The emphasis of regional and international partners' support is crucial to counter against piracy and armed robbery incidents, which led to the steady decline of incidents in Gulf of Guinea. Continuous cooperation between the various regional partners is required, complementing with strengthened legal framework to tackle piracy acts.

3. Piracy around the Gulf of Aden and Arabian Sea

The Gulf of Aden and Arabian Sea, situated around the northern region of Indian Ocean, represent one of the globe's most heavily pirated zones, particularly around the Somalia coastline. Due to the high numbers of piracy incidents around Somalia since 2008, Operation Ocean Shield is initiated during the period of 2009 to 2016 to aid in deterring and obstructing pirate assaults. Through diverse military operations and the presence of several international partners, it introduced stability and elevated maritime security measures in the region (North Atlantic Treaty Organization 2022).

Soy (2018) had reported that the piracy incidents peaks in 2011 around Somalia and decrease significantly in Figure 21. This is attributed to a change in policy to allow the deployment of naval and air forces in the area to suppress piracy. However, the dependence on international partners' efforts represents a short-term solution to sustain maritime security. It is imperative for regional nations around Gulf of Aden to strengthen the necessary infrastructure to enhance maritime capabilities and allocate sufficient resources to address the challenges posed by piracy related activities.

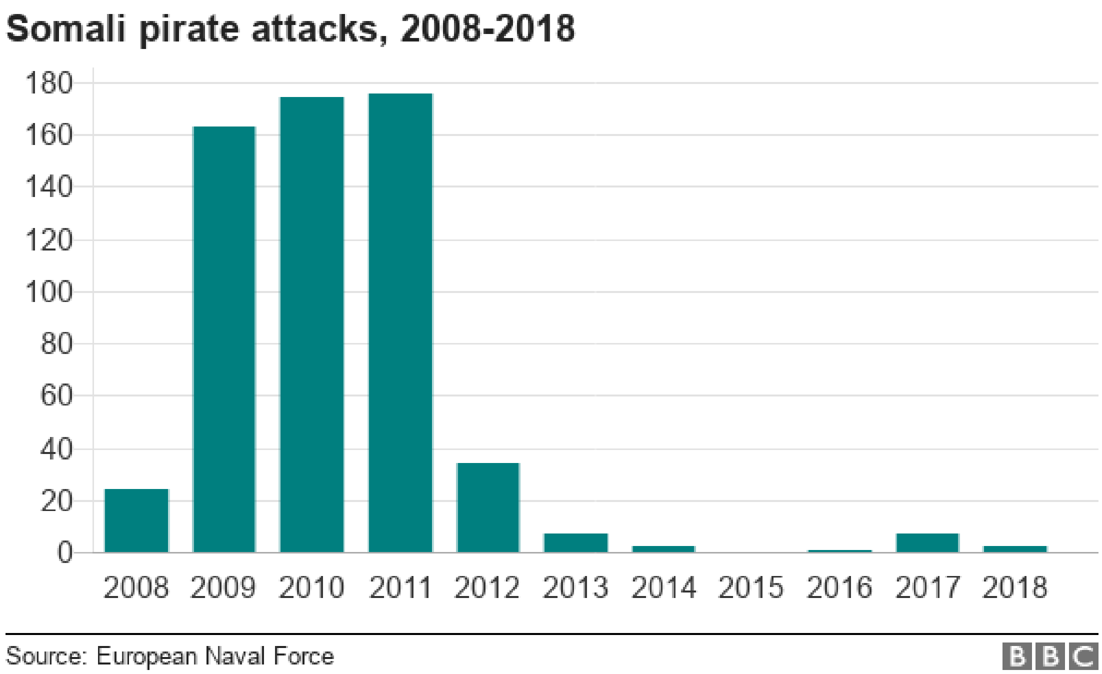

Figure 21. Piracy Incidents around Somalia from 2008–2018. Source: Soy (2018).

4. Piracy around the Bay of Benegal

Bay of Benegal is located between India and Myanmar, and south of Bangladesh at northeastern region of Indian Ocean. Chowdhury (2022) reported that human trafficking is rampant, and pirates often target the fishermen and cargo trawlers in the offshores area to loot, rob and kidnap fishermen. In addition, the increased occurrences of piracy incidents resulting from arrival of Rohingya refugees in Bangladesh prompt United States to introduce an "Indo-Pacific strategy" to help address both conventional and unconventional challenges to counteract piracy and human trafficking in 2017. In recent years, with the strengthened defense relationship from international and regional partners to combat maritime threats, piracy and trafficking activities are disrupted.

According to Koh (2023), multiple initiatives, such as Bali Process and Bay of Bengal Initiative for Multi-Sectoral Technical and Economic Cooperation (BIMSTEC), have been launched to ensure the well-being and safety in the region. These initiatives underscore the significance of multilateral maritime security collaborations, advancement of capabilities and broader engagement from neighboring nations to facilitate the establishment of a more fortified maritime setting in Bay of Bengal and Andaman Sea.

5. Piracy around South East Asia

The waters of South East Asia span across Indonesia, Malacca and Singapore Straits venturing into South China Sea, which function as one of the busiest global maritime routes. Between 1995 and 2013, Mccauley (2014) noted that South East Asia accounted for 41 percent of world's piracy occurrence, and the actual number of hijacking incidents could be more than what is reported.

With the evolving threats in Asia waters, Lee (2022) informed that Regional Agreement on Combating Piracy and Armed Robbery against Ships in Asia (ReCAAP) is established on 4 September 2006 to promote international and regional collaboration with 21 states. Furthermore, prevalent activities, such as vessel hijacking, oil cargo theft and crew kidnapping, are prevalent during different timeframe from 2007 to 2021.

According to International Chamber of Commerce International Maritime Bureau (ICC IMB) (2023a) Piracy and Armed Robbery Against Ship Report for 2022, there has been a continuous rise in incidents from 2018 to 2022 as illustrated in Figure 22. In 2022, 38 incidents are recorded, with incidents primarily taking place during night hours while vessels are underway.

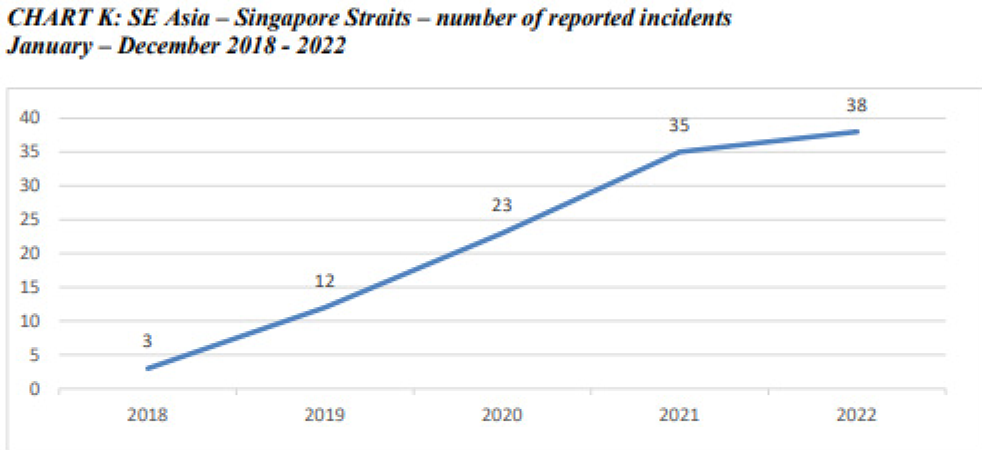

Figure 22. Number of Reported Piracy Incidents in Singapore from 2018 to 2022. Source: ICC IMB (2023a).

6. Piracy Updates in First Half of 2023

ICC IMB (2023b) reports a total of 65 piracy incidents in first half of 2023 in the mid-year report as compared to 58 piracy incidents during the same period in 2022. The primary locations for these incidents are Singapore Straits, Indonesia, and Peru, which accounts for 54 percent of incident reported in 2023. Notably, piracy incidents in the Singapore Straits are persistently increasing, solidifying its position as the global piracy hotspot.

Comparing the actual reported numbers depicted in Figure 23 with the forecasted results in Figure 8, we obtained the MAPE for individual months and an average MAPE of 0.329 as shown in Table 6. Therefore, the forecasted value of dataset is approximately 32.9 percent away from actual reported number of incidents in this study. Although the number of actual reported incidents is within the 80 percent (range from 0 to 22.77) and 95 percent forecast interval (range from 0 to 29.84) during the first half of 2023, the MAPE of ARIMA model in the study is considered relatively high at 32.9 percent.

CHART B: Monthly comparison of incidents during January – June 2023

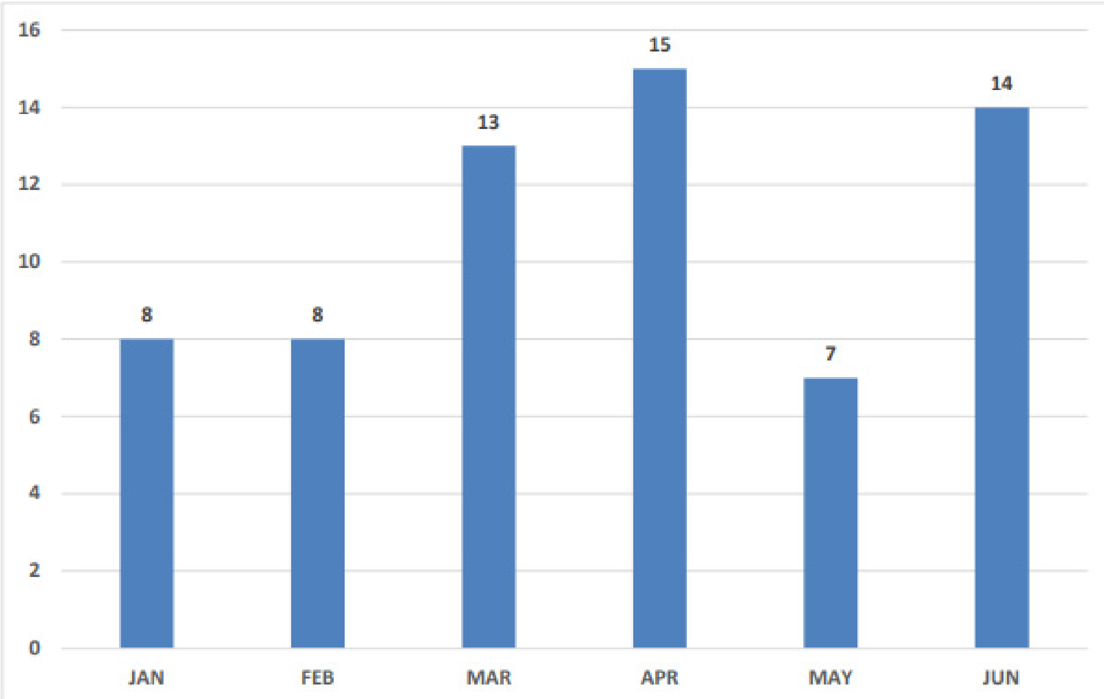

Figure 23. Monthly Reported Piracy Incidents from January 2023 to June 2023. Source: ICC IMB (2023b).

Table 6. MAPE of Forecasted Results against Reported Piracy. Adapted from ICC IMB (2023b).

Month	Actual Number of Reported Piracy	Forecasted Number of Piracy	MAPE
January	8	11.18	0.397
February	8	9.81	0.226
March	13	9.75	0.250
April	15	9.67	0.355
May	7	9.41	0.344
June	14	8.35	0.404
		Average MAPE	**0.329**

V. CONCLUSION

This chapter entails the summary of the study conducted and potential research areas in the topic of maritime piracy. In this study, we explore the use of time series analysis and clustering analysis to examine piracy events.

A. SUMMARY

The primary objectives in this study are to pinpoint areas of high piracy activities, examine potential factors contributing to piracy events and determining the effectiveness of time series analysis in forecasting the frequency of maritime piracy incidents.

In time series analysis, different models are evaluated, including, Naïve, Seasonal Decomposition, HW Exponential Smoothing, ARIMA and Ensemble models. Although ARIMA emerges as the optimal model in our study, the MAPE of 31 percent from modeling and MAPE of 32.9 percent when compared with actual reported numbers in first half of 2023, the result is relatively moderate and falls shorts of being considered as a highly accurate prediction model.

With the use of Haversine distance for the calculation of distance matrix in clustering analysis, DBSCAN clustering algorithm is selected for being able to remove noise and serve to identify piracy hotspots of varying densities across the globe. Furthermore, DBSCAN can handle the clustering of irregular shape of data points which can conform to the natural coastal line, making it an excellent choice for geospatial analysis.

Spatio-temporal analysis in maritime piracy involves studying the interplay between time and space factors to identify incidents cluster across time and allow us to understand factors influencing piracy. This aids in the identification of global piracy hotspots and the changes in concentration over time and may allow us to evaluate the effects of certain policy changes in the region. Decision makers can focus their attention on identified prominent piracy hotspots and discussed the necessary intervention and initiatives to counteract piracy events.

Prominent hotspots, such as Caribbean Sea, Gulf of Guinea, Gulf of Aden and Arabian Sea, Bay of Bengal, and South East Asia, require more attention over the period of 2010 to 2022. With the rising trend in Singapore Straits, there is a need for enhancing of maritime security measures and effective strategies, such as international and regional collaboration, to be put in place to help raise awareness and combat against piracy incidents.

B. FUTURE WORK

Some considerations to improve the forecasting accuracy models can include:

- The use of forecasting models beyond this study, such as machine learning model (e.g., recurrent neural network), vector autoregressive model etc.
- Incorporating external factors such as weather data, or socio-economic indicators to better understand factors influencing piracy events and improve forecasting accuracy of time series models.

Follow-on research opportunities can consist of:

- Merging vessel static details, such as vessel type, size, tonnage etc., to further analyze target of interests across the prominent piracy hotspots.
- Looking into specific area of occurrence (territorial, port, international water) to identify focus areas and improve policy within specific regions (e.g., Singapore and Malacca Straits).
- Analyzing and classifying of the type of piracy incidents (e.g., kidnap, smuggle, robbery, etc.) to discern specific patterns associated with different types of piracy and gain deeper understanding of the motive and modus operandi behind piracy activities.

APPENDIX. DBSCAN CLUSTERING OF YEARLY DATA

```
[1] 2010
DBSCAN clustering for 437 objects.
Parameters: eps = 5e+05, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 6 cluster(s) and 116 noise points.

  0   1   2   3   4   5   6 
116 153  11  73  24  34  26 

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2011
DBSCAN clustering for 517 objects.
Parameters: eps = 750000, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 5 cluster(s) and 80 noise points.

  0   1   2   3   4   5 
 80 288  82  13  42  12 

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2012
DBSCAN clustering for 314 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 5 cluster(s) and 64 noise points.

 0  1  2  3  4  5 
64 96 51 85 15  3 

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2013
DBSCAN clustering for 271 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 4 cluster(s) and 57 noise points.

  0   1   2   3   4 
 57 142  17  43  12 

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2014
DBSCAN clustering for 273 objects.
Parameters: eps = 1500000, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 3 cluster(s) and 20 noise points.

  0   1   2   3 
 20  37 196  20 

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2015
DBSCAN clustering for 288 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 4 cluster(s) and 46 noise points.

  0   1   2   3   4 
 46 202  18  11  11 
```

Figure 24. Parameters and Results of DBSCAN clustering from 2010 to 2015

```
[1] 2016
DBSCAN clustering for 193 objects.
Parameters: eps = 750000, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 4 cluster(s) and 66 noise points.

 0  1  2  3  4
66 76 39 11  1

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2017
DBSCAN clustering for 195 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 5 cluster(s) and 38 noise points.

 0  1  2  3  4  5
38 83 13 14 39  8

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2018
DBSCAN clustering for 204 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 4 cluster(s) and 45 noise points.

 0  1  2  3  4
45 62 14 67 16

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2019
DBSCAN clustering for 174 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 4 cluster(s) and 56 noise points.

 0  1  2  3  4
56 53 53 11  1

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2020
DBSCAN clustering for 207 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 4 cluster(s) and 73 noise points.

 0  1  2  3  4
73 54 66 13  1

Available fields: cluster, eps, minPts, dist, borderPoints
[1] 2021
DBSCAN clustering for 163 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 5 cluster(s) and 39 noise points.

 0  1  2  3  4  5
39 30  5 64 14 11

Available fields: cluster, eps, minPts, dist, borderPoints

[1] 2022
DBSCAN clustering for 113 objects.
Parameters: eps = 1e+06, minPts = 10
Using unknown distances and borderpoints = FALSE
The clustering contains 3 cluster(s) and 44 noise points.

 0  1  2  3
44 55 11  3

Available fields: cluster, eps, minPts, dist, borderPoints
```

Figure 25. Parameters and Results of DBSCAN clustering from 2016 to 2022

LIST OF REFERENCES

Ajeagah FN (2022) Combating Piracy in the Gulf of Guinea: Understanding The Challenges of The Yaoundé Architecture for Maritime Security. Master's thesis, Department of National Security Affairs. Naval Postgraduate School, California, https://hdl.handle.net/10945/70997.

Allwright S (2022) What is a good MAPE score? Stephen Allwright. Accessed July 19, 2023, https://stephenallwright.com/good-mape-score/.

Appelhans T, Detsch F, Reudenbach C, Woellauer S, Forteva S, Nauss T, Pebesma E et al. (2022) mapview: Interactive Viewing of Spatial Data in R, version 2.11.0. Accessed July 23, 2023, https://cran.r-project.org/web/packages/mapview/index.html.

Boehmke B, Greenwell B (2020) *Hands-On Machine Learning with R*, https://bradleyboehmke.github.io/HOML/

Cabungcal J, Kaniss D, Laing C, Mastran K, Powell J, Quijano N, Rosenberg E, Walsh G (2014) Preventing pirates from boarding commercial vessels – a systems approach. Capstone Project Report, System Engineering Department. Naval Postgraduate School, California, https://hdl.handle.net/10945/43991.

Chowdhury J (2022) Sri Lanka And Bangladesh's Role In Combatting Maritime Threat In The Bay Of Bengal – OpEd. Eurasia Review. Accessed August 7, 2023, https://www.eurasiareview.com/17012022-sri-lanka-and-bangladeshs-role-in-combatting-maritime-threat-in-the-bay-of-bengal-oped/.

Dabrowski JJ, De Villiers JP (2015) Maritime piracy situation modelling with dynamic Bayesian networks. *Information Fusion* 23:116–130, https://linkinghub.elsevier.com/retrieve/pii/S1566253514000840.

Desai RM, Shambaugh GE (2021) Measuring the global impact of destructive and illegal fishing on maritime piracy: A spatial analysis. *PLoS One* 16(2), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7904167/.

Dryad Global (2023) Caribbean and Gulf of Mexico Regional Hotspot. Accessed August 5, 2023, https://dg.dryadglobal.com/caribbean-gulfofmexico.

Gard (2023) Piracy trends and high risk areas. Accessed August 5, 2023, https://www.gard.no/web/articles?documentId=34977995.

Hahsler M, Piekenbrock M, Arya S (2022) dbscan: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Related Algorithms, version 1.1-11. Accessed July 9, 2023, https://cran.r-project.org/web/packages/dbscan.

Hartigan JA, Wong MA (1979) kmeans: K-Means Clustering. Accessed July 9, 2023, https://rdrr.io/r/stats/kmeans.html.

Hijmans RJ (2022) geosphere: Spherical Trigonometry, version 1.5-18. Accessed July 24, 2023, https://cran.r-project.org/web/packages/geosphere/index.html.

Hyndman RJ (2023) auto.arima: Fit best ARIMA model to univariate time series. Accessed July 9, 2023, https://rdrr.io/cran/forecast/man/auto.arima.html.

Hyndman RJ, Athanasopoulos G (2018) *Forecasting: Principles and Practice (2nd ed)*, https://otexts.com/fpp2/

Hyndman RJ, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, Kuroptev K, O'Hara-Wild M et al. (2023) forecast: Forecasting Functions for Time Series and Linear Models, R package version 8.21. Accessed July 9, 2023, https://cloud.r-project.org/web/packages/forecast/.

International Chamber of Commerce International Maritime Bureau (2023a) Piracy and Armed Robbery Against Ships. ICC IMB Piracy and Armed Robbery Against Ships Report-January-December 2022, United Kingdom, https://www.icc-ccs.org/reports/2022%20Annual%20IMB%20Piracy%20and%20Armed%20Robbery%20Report.pdf.

International Chamber of Commerce International Maritime Bureau (2023b) Piracy and Armed Robbery Against Ships. ICC IMB Piracy and Armed Robbery Against Ships Report-January-June 2023, United Kingdom, https://www.icc-ccs.org/reports/2023%20Jan%20-%20Jun%20IMB%20Piracy%20and%20Armed%20Robbery%20Report.pdf.

International Maritime Organization (2019) Maritime Security. IMO. Accessed April 16, 2023, https://www.imo.org/en/OurWork/Security/Pages/Default.aspx.

International Maritime Organization (2023) Global Integrated Shipping Information System, Piracy and Armed Robbery Module. International Maritime Organization. Accessed January 25, 2023, https://gisis.imo.org/Public/Default.aspx.

James G, Witten D, Hastie T, Tibshirani R (2021) *An Introduction to Statistical Learning with Applications in R, Second Edition*, https://hastie.su.domains/ISLR2/ISLRv2_website.pdf

Jin M, Shi W, Lin KC, Li KX (2019) Marine Piracy Prediction and Prevention: Policy Implications. *Marine Policy* 108(103528), https://torpedo.nrl.navy.mil/tu/ps/doc.html?vol=108&dsn=18877769&ssn=995&iss=C&st=JRNAL.

Kahle D, Wickham H, Jackson S (2023) ggmap: Spatial Visualization with ggplot2, version 3.0.2. Accessed July 9, 2023, https://rdrr.io/cran/ggmap/.

Kassambara A (2023) ggpubr: "ggplot2" Based Publication Ready Plots, version 0.6.0. Accessed July 9, 2023, https://cran.r-project.org/web/packages/ggpubr/.

Koh CSL (2023) Promoting Maritime Security in the Bay of Bengal and Andaman Sea. Centre for Social and Economic Progress. Accessed August 7, 2023, https://csep.org/reports/promoting-maritime-security-in-the-bay-of-bengal-and-andaman-sea/.

Lee YM (2022) Piracy and Armed Robbery as an Evolving Threat to Southeast Asia's Maritime Security. Asia Maritime Transparency Initiative. Accessed August 7, 2023, https://amti.csis.org/piracy-as-an-evolving-threat-to-southeast-asias-maritime-security/.

Marchione E, Johnson SD (2013) Spatial, Temporal and Spatio-Temporal Patterns of Maritime Piracy. *Journal of Research in Crime and Delinquency* 50(4):504–524, https://journals.sagepub.com/doi/epub/10.1177/0022427812469113.

Matthews AD (2015) Indonesian maritime security cooperation in the Malacca Straits. Master's thesis, Department of National Security Affairs. Naval Postgraduate School, California, https://hdl.handle.net/10945/45897.

Mccauley A (2014) The Most Dangerous Waters in The World. *TIME* (August), https://time.com/piracy-southeast-asia-malacca-strait/.

Mechler M, Rousseeuw P, Struyf A, Hubert M, Kurt Hornik (2022) cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al., version 2.1.4. Accessed July 9, 2023, https://cran.r-project.org/web/packages/cluster/.

Meyer D (2023) HoltWinters: Holt-Winters Filtering. Accessed July 9, 2023, https://rdrr.io/r/stats/HoltWinters.html.

North Atlantic Treaty Organization (2022) Counter-piracy operations (2008-2016). Accessed August 6, 2023, https://www.nato.int/cps/en/natohq/topics_48815.htm.

Nwokedi TC, Anyanwu J, Eko-Rapheaels M, Obasi C, Akpufu ID, Ogola DB (2022) Probability Theory Analysis of Maritime Piracy and the Implications for Maritime Security Governance. *Journal of ETA Maritime Science* 10(2):133–143, https://doi.org/10.4274/jems.2022.37132.

Ooms J (2023) gifski: Highest Quality GIF Encoder, version 1.12.0-1. Accessed July 9, 2023, https://cran.r-project.org/web/packages/gifski/.

Pedersen TL, Robinson D (2022) gganimate: A Grammar of Animated Graphics, version 1.0.8. Accessed July 9, 2023, https://cran.r-project.org/web/packages/gganimate/.

Posit Team (2023) RStudio: Integrated Development Environment for R, version 2023.06.0+421. Accessed February 16, 2023, http://www.posit.co/.

Pristrom S, Yang Z, Wang J, Yan X (2016) A novel flexible model for piracy and robbery assessment of merchant ship operations. *Reliability Engineering & System Safety* 155:196–211, https://doi.org/10.1016/j.ress.2016.07.001.

R Core Team (2023) R: A language and environment for statistical computing, version 4.3.1. Accessed February 16, 2023, https://www.R-project.org/.

Ripley BD (2023) stl: Seasonal Decomposition of Time Series by Loess. Accessed July 8, 2023, https://rdrr.io/r/stats/stl.html.

Soy A (2018) Somalia piracy: How foreign powers are tackling it. *BBC News* (December 11), https://www.bbc.com/news/world-africa-46454055.

Swanson DA (2015) On the Relationship among Values of the Same Summary Measure of Error when it is used across Multiple Characteristics at the Same Point in Time: An Examination of MALPE and MAPE. *Review of Economics and Finance* 5:1–14, https://escholarship.org/content/qt1f71t3x9/qt1f71t3x9.pdf?t=o5wul1.

Teixeira CA, Pinto JN (2022) Maritime piracy in the Gulf of Guinea. GIS Reports. Accessed August 5, 2023, https://www.gisreportsonline.com/r/piracy-gulf-guinea/.

Tsioufis M, Fytopoulos A, Kalaitzi D, Alexopoulos TA (2023) Discovering maritime-piracy hotspots: a study based on AHP and spatio-temporal analysis. *Annals of Operations Research*:1–23, https://www.researchgate.net/publication/370366835_Discovering_maritime-piracy_hotspots_a_study_based_on_AHP_and_spatio-temporal_analysis.

United Nations Office on Drugs and Crime (2010) 9. Maritime Piracy. Transnational Organized Crime Threat Assessment Report 2010 E.10.IV.6, New York, https://www.unodc.org/documents/data-and-analysis/tocta/9.Maritime_piracy.pdf.

United Nations Press (2023) Ongoing Decline in Gulf of Guinea's Piracy, Armed Robbery Encouraging, But Support Needed to Fully Implement Yaoundé Architecture, Briefers Tell Security Council. Meetings Coverage and Press Releases. Accessed August 5, 2023, https://press.un.org/en/2023/sc15331.doc.htm.

Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D (2023) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics, version 3.4.2. Accessed July 9, 2023, https://cran.r-project.org/web/packages/ggplot2/.

Yoshida R (2023a) OA4106 Advanced Data Analysis. Lecture, Data Science Clustering, Feb 16, Department of Operations Research, Naval Postgraduate School, Monterey, CA.

Yoshida R (2023b) OA4106 Advanced Data Analysis. Lecture, (Univariate) Time Series Overview, Jan 18, Department of Operations Research, Naval Postgraduate School, Monterey, CA.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California