Application of time series analysis techniques in proactive disaster management in the state of Rio de Janeiro

Crossref doi
https://doi.org/10.56238/interdiinovationscrese-006

Samir Batista Fernandes
Specialist in Civil Defense - UNISUL
MSc in Security and Civil Protection - UFF
Specialist in Data Science - USP
Special Advisor to SEDEC-RJ

ABSTRACT

In this study, an in-depth analysis of time series was conducted to enhance disaster management and contingency planning in the state of Rio de Janeiro. Employing time series analysis techniques, the study aimed not only to validate pre-existing hypotheses but also to discover new avenues for evidence-based decision-making. The analysis revealed distinct seasonal patterns, particularly between November and April, an identified critical period for effective response strategy implementation. The research underscored the significant contribution of specific occurrence categories to the observed trend and seasonality in the time series, providing a solid foundation for reducing uncertainties in contingency planning. Furthermore, the potential for result extrapolation paves the way for more accurate forecasts of future events, facilitating a swifter and more effective response to adverse events. In conclusion, the study serves as a robust model for data-driven disaster management, indicating a path towards a more scientific and effective approach in the development and implementation of contingency plans. It represents a valuable contribution to existing literature, displaying the potential of machine learning and time series analysis in promoting proactive, data-driven disaster management.

Keywords: Contingency Plan, SEDEC, ICTDEC, Data-Driven.

1 INTRODUCTION

In recent years, the frequency and magnitude of natural disasters and adverse events have intensified, raising concerns at global and local levels (CRED & UNISDR, 2018). In this context, the creation of effective contingency plans, based on robust scientific evidence, becomes a pressing need (Perry, 2007).

The analysis of historical time series, available in the databases of civil protection and defense agencies, presents an invaluable opportunity to unravel the patterns and trends of these adverse events (Zhang et al., 2015). The major advantage of employing advanced time series analysis techniques on vast datasets lies in the ability to overcome the limitations of basic statistical analyses, such as frequency counting, mode, and meaning, commonly employed in the historical records of occurrences provided by civil protection and defense agencies. These advanced techniques are able to identify complex patterns that may remain hidden for more superficial analysis (Hyndman & Athanasopoulos, 2018).

In this way, the use of advanced time series techniques not only goes beyond basic statistical analysis, but also facilitates more solid evidence-based decision-making, reducing uncertainties, allowing the validation of hypotheses and the projection of results with greater accuracy (Shmueli, 2010). The application of time series techniques, in this context, further expands the possibilities, allowing for more accurate forecasts and a more agile, data-driven response to disasters.

The integration of these analyses into contingency plans can promote a more efficient allocation of resources, thus minimizing the impact of these adverse events on affected communities (UNISDR, 2015).

The justification for this study lies in the hypothesis of the increase in occurrences in the summer months in the state of Rio de Janeiro and the growing need for evidence-based approaches for the management of disaster reduction or adverse events. Additionally, identifying seasonal patterns and trends can serve as a vital tool for decision-makers, aiding in the implementation of proactive measures that can save lives and conserve resources.

This study therefore sought to contribute to the existing literature by providing an in-depth analysis of adverse event occurrences, with a particular focus on identifying seasonal patterns and trends.

The main objective of this study was to analyze the historical time series of occurrences attended by municipalities in the state of Rio de Janeiro that use the platform created by the State Secretariat of Civil Defense of the state of Rio de Janeiro called the occurrence registration program – PRODEC, in order to identify patterns of seasonality and trends that can inform the creation of more effective contingency plans. In addition, the study aimed to highlight the importance of data analysis in the development of evidence-based disaster management strategies, thereby promoting a more effective and efficient response to adverse events.

This article is structured as follows: section 2 will describe the theoretical framework, and section 3 will adopt the methodology for data analysis, including a detailed description of the data used and the analysis techniques used. Section 4 will present the analysis of the data, highlighting the main results in relation to seasonality and trends observed in the time series. Section 5 will discuss the results obtained and highlight their practical implications. Finally, section 6 will offer a conclusion to the study, summarizing the key findings and suggesting possibilities for future research.

2 THEORETICAL FRAMEWORK

In this section, the theoretical foundation that underpins this study was outlined, covering the crucial concepts in disaster management, using data science as a tool in disaster risk reduction, and the advanced time series techniques that were employed.

2.1 FUNDAMENTAL CONCEPTS IN DISASTER MANAGEMENT

In the area of disaster management, it is essential to understand and differentiate key concepts such as threat, disaster, adverse events, vulnerability, and occurrences:

- **Threat**: Refers to a potentially damaging event that may occur in an uncertain future, usually as a result of natural phenomena or human activities (UNISDR, 2009).
- **Disaster**: It is the realization of a threat, where the harmful event occurs, causing significant damage and often human losses (CRED, 2015).
- **Adverse Events**: These are incidents that can lead to disasters, usually identified as early signs of an emerging threat (Perry, 2007).
- **Occurrences**: These refer to specific events that are recorded by civil protection and defense agencies, and can be both adverse events and disasters (Alexander, 2016). It should be noted that in the present study, the nature of the occurrences attended by civil protection and defense agencies are classified into distinct groups, such as: adverse event, vulnerability, threat, not applicable or incomplete record.
- **Vulnerability**: It represents the degree of exposure and susceptibility of a community or system to threats, being a crucial component in disaster management and mitigation (Cutter, 1996).

2.2 DATA SCIENCE AS A SUPPORTING TOOL

Data science has emerged as a powerful tool in disaster management, offering advanced techniques that go beyond traditional statistical analysis. Time series analysis, an underlying approach in data science, is complemented by other equally important techniques such as decomposition, exponential smoothing, ARIMA, regression, neural networks, etc. which together provide a more holistic and in-depth analysis of the data (Provost & Fawcett, 2013).

2.3 FUNDAMENTAL CONCEPTS IN TIME SERIES APPLIED IN THIS RESEARCH

- **Stationarity:** Stationarity refers to a property of time series where statistics, such as mean and variance, remain constant over time. This means that there are no significant trends or systematic patterns that change over time (Brockwell & Davis, 2016).
- **Time series:** A time series is a set of observations or data collected at regular time intervals. This data is organized chronologically and can be used to analyze trends and patterns over time (Chatfield, 2003).
- **Decomposition**: Decomposition is a process used to separate a time series into its main components, which are usually trend, seasonality, and waste. Trend represents the general pattern of increase or decrease over time, seasonality captures periodic variations, and

residuals are the variations not explained by trends and seasonality (Hyndman & Athanasopoulos, 2018).

- **Trend**: The trend in a time series represents the overall direction in which the data is moving over time. It can be an upward (increasing) or downward (decreasing) trend (Montgomery, Johnson & Gardiner, 2012).
- **Seasonality:** Seasonality in a time series refers to variations that occur at regular intervals of time, usually in a cyclical pattern. For example, we can observe a significant increase in the number of occurrences recorded by the Civil Defense during periods of higher rainfall, such as the heavy rainy season (Wei, 2006).

3 METHODOLOGY

The data used in this study were collected from the PRODEC platform, which aggregates information on adverse events from several municipalities in the state of Rio de Janeiro. During the study period, which covers the period from 2013 to 2023, a total of 70 municipalities were registered and actively contributing data to the platform.

For data analysis, Python was used, a programming language widely recognized for its applicability in data science and artificial intelligence. In time series analysis, techniques were employed to extract meaningful patterns and information from the data. Some of the techniques used in this study include:

- Augmented Dickey-Fuller (ADF) test: The Dickey-Fuller unit root test technique, often referred to as the ADF (augmented Dickey-Fuller) test, was initially developed by Robert F. Engle and Clive W.J., and later perfected by David A. Dickey and Wayne A. Fuller in 1979. A technique for verifying the stationarity of a time series, which is a prerequisite for many subsequent analyses (Said & Dickey, 1984).
- Holt-Winter Exponential Smoothing: The technique was developed by Charles C. Holt and Peter G. Winters in 1976 and used to identify seasonal trends and patterns in time series, allowing for accurate forecasts (Hyndman & Athanasopoulos, 2018).
- ETS decomposition: A technique for separating a time series into trend, seasonality, and residuals components, facilitating individual analysis of each component (Cleveland et al., 1990).

In addition, a second decomposition was performed only with the groups of occurrences. Namely: threats, vulnerabilities, adverse event, incomplete record, and not applicable.

Despite the results that can be derived from this analysis, it is important to recognize some limitations. First, the generalization capacity of the study may be restricted, given that the data are specific to the municipalities in the state of Rio de Janeiro that participate in the PRODEC program.

In addition, the analysis is subject to the limitations inherent to erroneous data entries by municipalities, which can be motivated by various reasons ranging from lack of knowledge, typing errors, conceptual errors, etc. Finally, the quality of the results depends on the data collected through the PRODEC platform and can influence the results, potentially introducing bias into the analysis.

4 DATA ANALYSIS

In the initial phase of the analysis, a descriptive analysis was conducted to understand the distribution and general characteristics of the data. In the time series, the Augmented Dickey-Fuller Test (ADF) was applied to verify the stationarity of the time series (Said & Dickey, 1984).

The enhanced Dickey-Fuller test returned the following results:

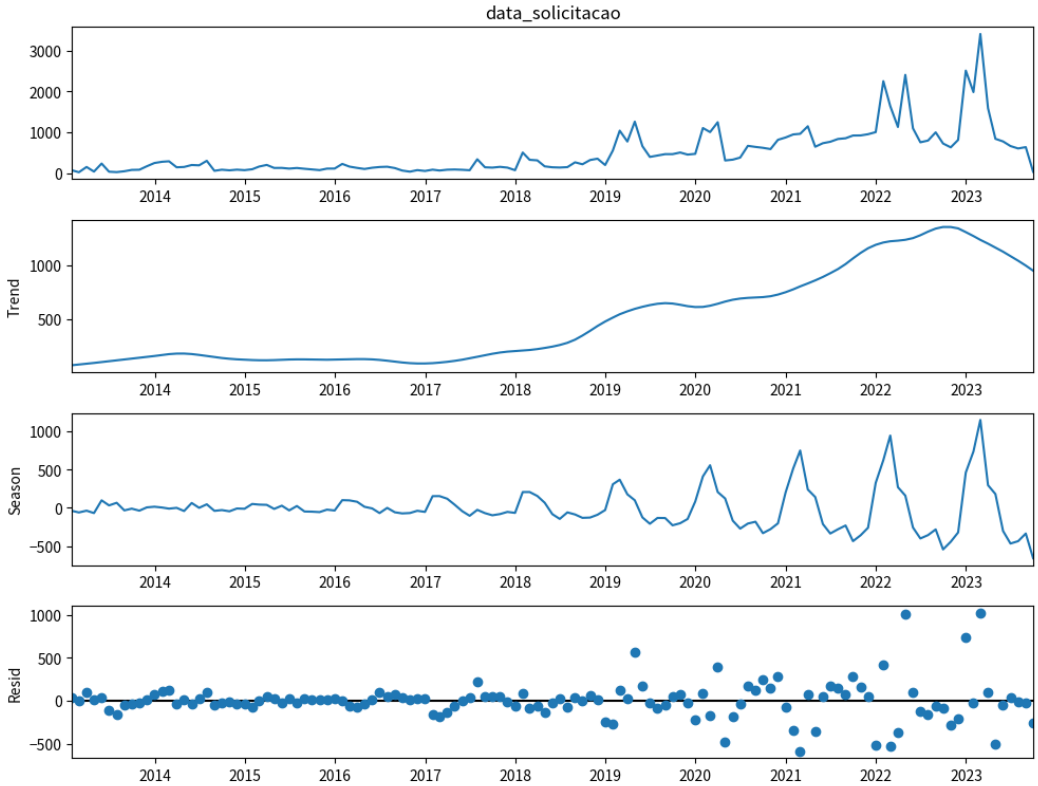
Test statistic value: -4.4114; p-value: 0.000283; Number of lags *used: 30 (*lags are the number of previous periods to be considered when modeling time dependency on data); Number of observations used: 3869; Critical values:**1%**: -3.4320; **5%**: -2.8623; **10%**: -2.5672

The p-value is very small (less than 0.01), which allows us to reject the null hypothesis that the time series has a unit root, indicating that the series is stationary and does not have a random time series structure.

This test was crucial to determine whether the time series has a dependent time structure, a constant mean and variance over time, thus facilitating the identification of seasonal patterns and trends (Box et al., 2015).

The subsequent analysis focused on the decomposition of time series to identify and understand the main components: trend, seasonality, and residuals. Using the ETS decomposition technique, it was possible to isolate each of these components and analyze their distinct characteristics.

Figure 1 - Monthly Time Series Decomposition



Source: The Author(2023)

The analysis revealed clear seasonal patterns, with periods of increased occurrences identified mainly between the months of November and April. As shown in figure 1 with four graphs: the monthly time series, the trend, the seasonality and the residuals.

After decomposition, a second Dickey-Fuller test was performed on the residuals to verify whether the time structure remains dependent over time, as well as whether the mean and variance remain constant, i.e., ultimately whether the data have stationarity characteristics (Dickey & Fuller, 1979).

The enhanced Dickey-Fuller test returned the following results:

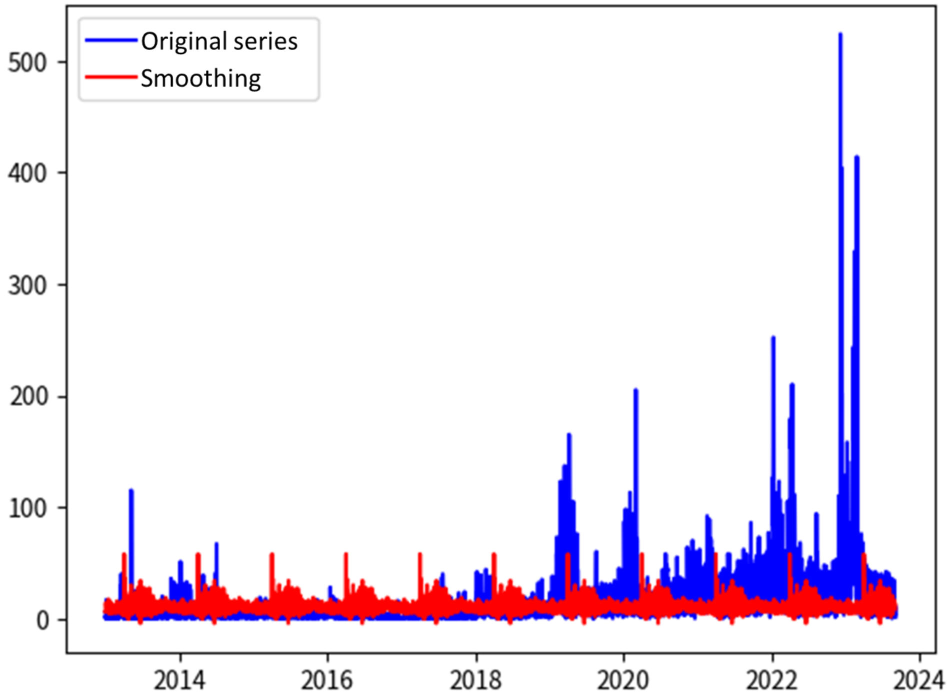
- Test statistic value: -25.1492; P-value: 0.0; Number of lags used: 28; Number of observations used: 3869; Critical values:**1%**: -3.4320; **5%**: -2.8622; **10%**: -2.5671

The p-value less than 0.05 and the test statistic below the critical values at all levels of significance (1%, 5% and 10%) indicate that the residues of the series, after the application of ETS decomposition, are stationary.

Therefore, based on the results of the Dickey-Fuller test, it can be stated that the data in the series are representative in their trend and seasonality, making them appropriate for possible use of traditional *machine learning* processes such as classification, regression, correlation, or association (Hyndman, 2018).

By exploring the exponential smoothing technique using Holt-Winters, it was possible to obtain the following results:

Figure 2 - Holt-Winters exponential smoothing technique



Source: The Author(2023)

After smoothing, the Dickey-Fuller test was performed on the residuals to verify whether the trend and seasonality are being captured in the time series. The results obtained were as follows:

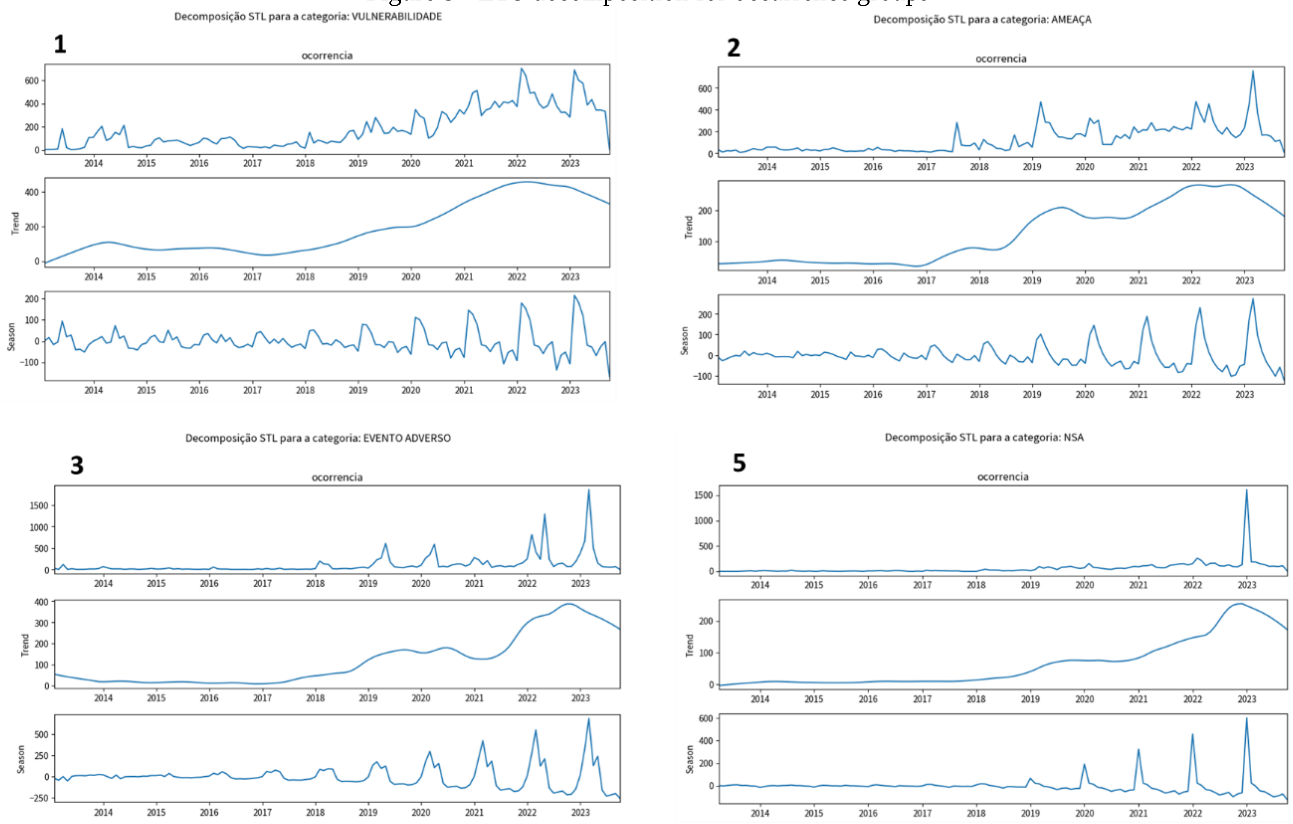
The enhanced Dickey-Fuller test returned the following results:

Test statistic value: -15.0756; P-value: 8.598...9e-28; Number of lags used: 30; Number of observations used: 3869 ; Critical values:**1%**: -3.4320; **5%**: -2.8622; **10%**: -2.5671

Therefore, one can reject the null hypothesis and admit that the residual data are stationary. Thus, both the decomposition technique and the smoothing were able to describe the trend, seasonality, and residuals with a level of statistical significance. In this way, we were able to choose between two distinct time series techniques: decomposition or exponential smoothing. We chose to use decomposition and mentioned smoothing, because in the future if you want to use smoothing techniques it is possible that this approach will offer a better performance in forecasting.

In figure 3 we have the next data treatment taking into account the classification of data by groups of occurrences as follows, decomposition of vulnerability occurrences (1), decomposition of threat occurrences (2), decomposition of adverse event occurrences (3) and decomposition of occurrences does not apply (5). Data from incomplete records were not used.

Figure 3 - ETS decomposition for occurrence groups



Source: The Authors(2023)

Categories 1, 3, and 5 showed more pronounced seasonal patterns and clear trends, indicating that these categories may be of particular interest for the development of effective contingency plans. The analysis also highlighted the presence of significant seasonal components, suggesting that certain periods of the year are associated with an increase in adverse event occurrences that will be discussed in the next section.

5 DISCUSSION

The results obtained through time series analysis provided important information that facilitates evidence-based decision-making, a crucial pillar for effective disaster management. The identification of a clear seasonality, especially between the months of November and April, validated the hypothesis of the increase in occurrences in the summer periods, previously established, but also helped to reduce uncertainties associated with contingency planning for the periods with the highest number of occurrences.

The exploration of the groups of occurrences that are classified as vulnerability, threat, incomplete records, adverse event or not applicable made it possible to better understand how each group of occurrence contributes to the time series by components (trend, seasonality or residuals).

This approach allowed us to understand the contribution of each group of occurrence throughout the year, so we have below a summary of the analysis of the components of each group.

5.1 TENDENCY

- **Vulnerability**: The trend shows an overall decrease over time, with some fluctuations.
- **Threat**: The trend is relatively stable, with small fluctuations over time.
- **Averse Event**: Shows an increasing trend, indicating an increase in occurrences over time.
- **Incomplete record**: The trend is fairly stable, with no significant changes over time.
- **Not Applicable Above**: The trend shows an initial decrease followed by an increase, indicating a change in the pattern of occurrences over time.

5.2 SEASONALITY

- **Vulnerability**: It has a clear seasonality, with regular peaks and valleys, indicating specific seasonal patterns.
- **Threat**: Seasonality is less pronounced, with smaller fluctuations throughout the year.
- **Averse Event**: Shows clear seasonality, with well-defined patterns of increase and decrease.
- **Incomplete record**: Seasonality is less evident, with minor fluctuations throughout the year.
- **Not applicable above**: It has a clear seasonality, with well-defined patterns of increase and decrease.

5.3 WASTE

- **Vulnerability**: Residuals are relatively small, indicating that most variations can be explained by trend and seasonality.
- **Threat**: Residues are larger, indicating variations that are not captured by trend and seasonality.
- **Adverse Event**: Residuals are small, suggesting that trend and seasonality explain most of the variations.
- **Incomplete record**: Residuals are larger, indicating significant variations not captured by trend and seasonality.
- **Not applicable above**: Residuals are small, suggesting that most of the variations are explained by trend and seasonality.

When examining the trend, it is observed that vulnerability exhibits an overall decrease over time, with some fluctuations, but it is important to consider that the agencies responsible for classifying occurrences may make errors in categorizing groups, which may influence the interpretation of the results. For example, cases of infiltration in low-income households may be erroneously classified as threat events, due to the complexity of the solution, taking into account the problem-solving capacity of that family group. However, these cases have a strong link to the socio-economic vulnerability of the affected communities.

When analyzing seasonality, it can be seen that vulnerability exhibits clear seasonal patterns, while the threat exhibits minor fluctuations throughout the year. However, it is important to remember that civil protection and defense agencies' knowledge of the terms and concepts of disaster risk management can vary, leading to different interpretations. For example, mass movement possibility events should be categorized as threat, due to the perception of imminent danger, even if there is an underlying condition of vulnerability in the affected areas.

Adverse events show an increasing trend, indicating an increase in occurrences over time. In this case, despite a more latent classification due to the consummation of the undesired effect, it is possible to have a combination of threat, vulnerability and new adverse events simultaneously, which complicates the accurate classification over time. Not infrequently, it is the civil protection and defense agencies that return to the same "address" of service due to the problem never being remedied and under different conditions, sometimes in a context of inspection to attend to an adverse event, sometimes in a threat or vulnerability.

In addition, "not applicable" or "incomplete registration" events may represent specific non-typified services or cases in progress, reflecting the complexity of the reality of civil protection and defense agencies.

6 CONCLUSION

The analysis of the historical time series of occurrences attended by municipalities in the state of Rio de Janeiro, using the PRODEC platform of the State Secretariat of Civil Defense, allowed the identification of seasonality patterns and trends, especially the increase in occurrence from November to April. These seasonal patterns and trends were identified in the components of the occurrences.

In addition, the study successfully highlighted the importance of data analysis in formulating evidence-based disaster management strategies. The research provided valuable information for the development of contingency plans by identifying the months that contribute the most to occurrences, through robust inferential statistics, allowing for better preparation and response to adverse events.

Despite the challenges in classifying occurrences, the information derived from this analysis has significant practical implications for disaster management in the state of Rio de Janeiro. Identifying seasonal patterns, even with potential variations in classification, allows for better preparedness and response to adverse events, potentially minimizing their impact on affected communities.

In addition, data-driven analysis facilitates the extrapolation of results, offering the possibility of more accurate predictions or techniques for investigating causality between dependent and independent variables through artificial intelligence.

The present study, based on a time series analysis, represented a significant advance in disaster management and contingency planning in the state of Rio de Janeiro.

For the future, we suggest that further research focus on standardizing the classification of occurrences, applying machine learning techniques for more detailed analysis, and adopting artificial intelligence for predictions. These initiatives have the potential to further strengthen disaster management, making it more data-driven and effective in responding to adverse events. With these research directions, we are confident that disaster management in the state of Rio de Janeiro will continue to evolve and enhance its ability to protect communities at risk.

REFERENCE

Alexander, D. Resilience and disaster risk reduction: an etymological journey. *Natural Hazards and Earth System Sciences*, 16(11), 2707-2716, 2016.

Bracewell, R. N. The Fourier Transform and Its Applications. 2ª ed. *McGraw-Hill*, 1986.

Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C.; Ljung, G. M. Time Series Analysis: Forecasting and Control. 5ª ed. *Wiley*, 2015.

Cleveland, R. B.; Cleveland, W. S.; Mcrae, J. E.; Terpenning, I. ETS: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1), 3-73, 1990.

CRED. The Human Cost of Natural Disasters. *Centre for Research on the Epidemiology of Disasters*, 2015.

Cutter, S. L. Vulnerability to environmental hazards. *Progress in Human Geography*, 20(4), 529-539, 1996.

Perry, R. W. What Is a Disaster? In: HANDELMER, H.; TORRES, A. (Eds.), *Emergency Management: Principles and Practice for Local Government*. 2ª ed. ICMA Press, p. 1-15, 2007.

Provost, F.; Fawcett, T. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. *O'Reilly Media*, Inc., 2013.

Said, S. E.; Dickey, D. A. Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, 71(3), 599-607, 1984.

UNISDR. UNISDR Terminology on Disaster Risk Reduction. *United Nations International Strategy for Disaster Reduction*, 2009.

CRED & UNISDR. Economic Losses, Poverty & Disasters 1998-2017. *Centre for Research on the Epidemiology of Disasters - CRED*, 2018.

Hyndman, R. J., & Athanasopoulos, G. Forecasting: principles and practice. *OTexts*, 2018. Disponível em: https://otexts.com/fpp3/

Shmueli, G. To explain or to predict? *Statistical Science*, 25(3), 289-310, 2010.

Montgomery, D. C., Johnson, L. A., & Gardiner, J. S. Forecasting and Time Series Analysis. *McGraw-Hill Education*, 2012.

Shumway, R. H., & Stoffer, D. S. Time Series Analysis and Its Applications: With R Examples. *Springer*, 2010.

Wei, W. W. S. Time Series Analysis: Univariate and Multivariate Methods. *Pearson*, 2006.

Chatfield, C. The Analysis of Time Series: An Introduction. *CRC Press*, 2003.

UNISDR. Sendai Framework for Disaster Risk Reduction 2015-2030. *United Nations Office for Disaster Risk Reduction*, 2015.

Zhang, Y., Wang, Y., & Grubesic, T. H. Exploring the impact of seasonality on spatial patterns of urban residential burglary. *Journal of Criminal Justice*, 43(6), 533-543, 2015.