



# Integrating structure time series forecasting and multicriteria decision analysis for adaptive operational risk assessment: an empirical study using real-time data

Guicang Peng<sup>1</sup> · Jon Tømmerås Selvik<sup>1</sup> · Eirik BJORHEIM Abrahamsen<sup>1</sup> · Knut Erik Bang<sup>1</sup> · Tore Markeset<sup>1</sup>

Received: 6 September 2023 / Revised: 1 December 2023 / Accepted: 21 March 2024  
© The Author(s) 2024

**Abstract** This study propose a framework for integrating Multi-Criteria Decision Analysis (MCDA) and Structure Time Series (STS) prediction for multivariate operational risk assessment often with highly dynamic risk determinants. In particular, by utilizing the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) as an MCDA method, the framework is able to prioritize risk determinants according to their inherent uncertainties' impacts on their respective operational objectives, and by employing Seasonal Autoregressive Integrated Moving Average (SARIMA) as an STS technique, the framework emphasizes real-time knowledge utilization for iteratively reducing uncertainty. By integrating SARIMA and TOPSIS, the framework aims to construct a multivariate operational risk assessment profile that is prioritized and continuously updated by the latest data and knowledge. Based on the proposed framework, the study constructs a mathematical model coded in Python to perform an empirical assessment of 161 countries' operational risk using real-time data from the Armed Conflict Location & Event Data Project's. A comprehensive analysis of the model's functionality, quality, and sensitivity based on the assessment result is provided. Conclusions and limitations are also discussed, highlighting the model's theoretical novelty and practical implications.

**Keywords** Operational risk assessment · MultiCriteria decision analysis · Structure time series prediction

## 1 Introduction

The term “operational risk” was initially defined by the Basel Committee on Banking Supervision (BCBS) as the potential for direct or indirect losses due to insufficient or failed internal procedures, personnel issues, system failures, or external events (Basel Committee on Banking Supervision 2011). This concept, initially exclusive to the banking sector, has since expanded and been tailored to various other industries (Pinto 2015; Kenett and Raanan 2011) and become fundamentally interdisciplinary (McNeil et al. 2015). Operational risk assessment can then be defined as a systematic process that evaluates the potential for direct or indirect losses arising from a wide array of interdisciplinary risk determinants (Aven 2015). The assessment process aims to identify, quantify, and prioritize these risk determinants, enabling organizations to understand their inherent uncertainties and their impacts on operational objectives (Haimes 2005).

Furthermore, according to Aven et al., risk can be described as the subjective probability expressing the uncertainty of the occurrences of the initiating events or scenarios and their consequences, as well as the uncertainty about the underlying factors influencing the initiating events and consequences. The subjective probability expressing the uncertainty is often given based on the background knowledge (Aven 2012). Consequently, for operational risk assessment, it becomes imperative to integrally consider two aspects:

Firstly, given that operational risks often intersect various disciplines, consequently, the operational risks are evaluated using diverse criteria. In that sense, it is appropriate to apply the MultiCriteria Decision Analysis (MCDA) framework to operational risk management (Chen and Tzeng 2004). MCDA is a systematic method used in complex decision-making scenarios with often conflicting criteria. It involves

✉ Guicang Peng  
guicang.peng@uis.no

<sup>1</sup> Department of Safety, Economics and Planning, University of Stavanger, 4036 Stavanger, Norway

identifying the problem, generating alternatives, setting evaluation criteria, and then weighing these alternatives to find the most suitable solutions (Wallenius et al. 2008; Belton and Stewart 2002). By integrating the MCDA framework into multi-objective operational risk assessment, one can obtain a prioritized risk profile that distinguishes risk determinants based on their overall importance as well as their relevance to specific objectives.

Secondly, since knowledge plays an important role in uncertainty reduction, for operations with uncertainty induced by inadequate knowledge, the continuous accumulation and utilization of knowledge is a pivotal step in operational risk assessment. In this case, STSP is a potential fit for this purpose. STSP is a key method in time series analysis, focusing on the prediction of future data by understanding available time series data, with underlying components like trend, seasonality, cyclical patterns, and noise (Box et al. 2015). STSP enables a dynamic and systematic approach to managing uncertainty, allowing decision-makers to adapt their understanding as new knowledge emerges.

While there is an extensive body of literature and many industrial practices addressing each of the two aspects discussed above independently, integrating these aspects holds a certain level of theoretical novelty and practical necessity. The aim of this paper is to propose a framework to iteratively reduce uncertainty within the MCDA process by applying continuously updated interdisciplinary knowledge based on the STSP techniques. In particular, the framework aims to achieve the following targets:

- By employing STSP techniques, to enable the utilization of updated knowledge in the operational risk assessment process. This is important for operations where decisions must be made based on real-time knowledge and data.
- By employing MCDA methods, to obtain an operational risk assessment consisting of a spectrum of risks that are prioritized based on their effects on the overall, as well as subordinated, operational objectives. This is to avoid silo risk assessment which often ignores the synergy or competitiveness of the interdisciplinary risk determinants.
- By integrating the MCDA and STSP techniques, to form a framework that is particularly suitable for operational risk assessment which is interdisciplinary and multi-objective, as well as critically depending on real-time knowledge and data updates.

To achieve the above targets, this study proposes the integration of the SARIMA (Seasonal Autoregressive Integrated Moving Average) as an STSP techniques (Box et al. 2015), with the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) as an MCDA method, to construct an adaptive operational risk assessment model. The model is coded with Python and applied for 161 countries'

operational risk assessment based on ACLED's (Armed Conflict Location & Event Data Project) real-time data (Raleigh et al. 2010). The model's assessment and forecast results were thoroughly analyzed, revealing insights into the model's functionality, assessment and forecast quality, as well as the sensitivity to model setting variations and data input quality.

The remainder of this paper is organized as follows: Section 2 presents a literature review on MCDA and STSP integration and applications. Section 3 introduces a mathematical model to integrate the SARIMA and TOPSIS methodologies, and the data ingesting process for the modeling process. Section 4 details the modeling results of the empirical operational risk assessment for 161 countries based on the proposed model with ACLED's real-time data. Section 5 discuss the assessment of result thoroughly including model sensitivity analysis. Section 6 concludes the study by summarizing the findings and discussing the limitations of the model, and possible future Improvements. Section 7 discuss the compliance of this paper with ethical standards.

## 2 Related research and industrial applications

MCDA was initially proposed to solve problems with fixed terms and conditions. This did not include treating elements of uncertainty. For MCDA to be employed in the operational risk management process, the method should be able to handle uncertainty. Roy proposed the ELECTRE (Elimination and Choice Translating Reality) method to consider theoretical interaction between MCDA and uncertainty, introducing the concept of robustness analysis to tackle inherent decision-making imprecision (Roy 1991). Figueira et al. conducted a comprehensive survey on MCDA framework and methodology groups, which comprised techniques that are able to deal with uncertainties, such as Stochastic Multicriteria Acceptability Analysis (SMAA), Evidential Reasoning (ER), and the ELECTRE method (Figueira et al. 2005). These methods offered distinct advantages in the face of uncertainty, with SMAA, for instance, integrating a probability distribution function into the decision-making process to accommodate uncertain, stochastic, or imprecise data. Similarly, ER provides a systematic framework for integrating and analyzing diverse forms of information, including uncertain or subjective judgments. In addition to MCDA methods that inherently address uncertainty like SMAA, ER, and ELECTRE, MCDA approaches that address uncertainty by integrating external frameworks and methodologies are also commonly seen in academic research and industrial applications. Examples include Bayesian network-based MCDA for operational risk management (Dalla Valle and Giudici 2008; Wathayu and Peng 2004), interval-valued

fuzzy MCDA (Wang et al. 2022; Wang 2019), and Dempster-Shafer theory-based MCDA (Hamid et al. 2022; Wu and Liao 2022). These methodologies implement MCDA with uncertainty in a wide range of contexts, from disaster risk reduction to supply chain management.

In particular, Bayesian Inference (BI), as a probabilistic method, integrating with MCDA methods allows for continuous uncertainty reduction. The iterative approach in BI lies in its fundamental premise of continuously updating prior beliefs based on newly observed data (Jaynes 2003). This process, often referred to as sequential analysis, allows for dynamic uncertainty reduction, where the degree of uncertainty is reduced with each additional data input (Sinha 1993). Recent theoretical advancements of BI focus on broadening its applicability and efficiency. Non-parametric methods have been proposed by Ghosal and Van Der Vaart for broadening the scope and adaptability of BI application, increasing its versatility beyond standard parametric models (Ghosal and Van der Vaart 2017). For industrial applications, Shevchenko and Wüthrich delve into the structural modeling of operational risk by fusing empirical loss data with expert opinions, through Bayesian inference (Shevchenko and Wüthrich 2009). Their study showcases the practical application of Bayesian techniques in operational risk assessment, leveraging both quantitative and qualitative information to formulate more robust risk management strategies. TOPSIS, as an MCDA technic, is widely used in performance ranking with multicriteria. The BI-integrated TOPSIS is designed to iteratively reduce the model uncertainties inherent in the performance determinants used in the ranking process. Gul and Yucesan present a Bayesian Best-Worst Method (Bayesian BWM) integrated TOPSIS model to rank 189 public and private Turkish universities (Gul and Yucesan 2022). Bayesian BWM is utilized to achieve the first ranking goal, followed by adopting the TOPSIS method, resulting in a comprehensive performance evaluation of universities. Lo and Liou propose an integrated Bayesian BWM and classifiable TOPSIS model to rank critical failure modes for risk assessment (Lo and Liou 2021). The study leverages Bayesian BWM for the initial ranking and integrates it with the classifiable TOPSIS technique, contributing to robust risk assessment methodologies.

A more recent study by Wang et al. (2023) targets the complexity in multivariate long sequence time-series forecasting. It specifically addresses the oversight of interdependencies among variables (Wang et al. 2023). Their solution involves integrating Graph Convolutional Networks (GCNs) with the Transformer model, further augmented by Temporal Convolutional Network (TCN) within the self-attention layer. This approach effectively improves prediction accuracy over several benchmarks, demonstrating the model's capability in managing variable interdependencies in long sequences.

Conversely, Kim and Moon explores multivariate time series data across various fields using a Bi-directional Long Short-Term Memory (BiLSTM) model (Kim and Moon 2019). Their innovation lies in incorporating field-specific features into the forecasting model, distinguishing their approach from traditional methods that often overlook such nuances. This adaptation results in enhanced predictive accuracy across multiple fields, underscoring the model's adaptability and efficiency.

Both studies contribute to the time-series forecasting field by introducing more nuanced, context-aware models capable of handling complex real-world data. Wang et al.'s research advances our understanding of variable interdependencies in extended time series, while Kim and Moon's work highlights the significance of field-specific adaptations in forecasting models. These studies collectively inform the trajectory of time-series forecasting, emphasizing the need for both technical sophistication and domain-specific tailoring, especially for multivariate scenarios.

## 2.1 The need for integrating MCDA and STSP

The literature survey presented above highlights a critical gap in the capabilities of existing MCDA methodologies, particularly in their ability to process and adapt to continuously evolving information. These MCDA methods, according to the literature survey, although robust in various aspects, are notably deficient in their iterative execution capability, which is essential for effectively incorporating the latest developments and data. In contrast, STSP methodologies excel in managing sequential data, but fall short in integrating multivariate or multiobjective considerations. This shortfall is largely attributable to the lack of a comprehensive MCDA framework within TSA approaches. The evident dichotomy between these two methodologies underscores the urgent need for their integration, aiming to leverage the strengths of both approaches for enhanced decision-making processes.

Expanding on this understanding, it becomes apparent that integrating MCDA with STSP could pave the way for a more objective and comprehensive approach to uncertainty management. Such an integration would not only facilitate a dynamic adaptation to changing data over time, but also allow for the incorporation of multiple criteria and objectives; a feature notably absent in standalone TSA methods. This synergy could lead to a more holistic decision-making process, where time-dependent data is analyzed not just in isolation, but in conjunction with a range of other relevant factors and criteria.

Moreover, the integration of MCDA with STSP could significantly enhance the predictive accuracy and reliability of the analyses. By combining the methodological rigor of MCDA in handling multiple criteria with the temporal

precision of STSP, decision-makers could gain deeper insights into complex scenarios, where variables and outcomes evolve over time. This integrated approach would enable a more nuanced understanding of trends, patterns, and potential future outcomes, thereby facilitating more informed and strategic decisions.

In addition to improving decision-making quality, this integration could also introduce a new dimension of flexibility in MCDA methods. By incorporating the dynamic aspects of STSP, MCDA methodologies could become more adaptable to changing circumstances, allowing for real-time adjustments and updates in the decision-making process. This would be particularly beneficial in environments characterized by rapid changes and uncertainty, where the ability to quickly adapt to new information is crucial.

Therefore, in the next chapter, we develop a mathematical model to integrate SARIMA as a STS method and TOPSIS as a MCDA method to form an operational risk assessment framework to test our concept.

### 3 Modeling method and data ingesting

This section introduces a mathematical model to integrate TOPSIS and SARIMA to achieve two functionalities. Firstly, it employs TOPSIS approach to create a comprehensive operational risk profile for countries. This profile integrates various weighted risk determinants, enabling a multi-objective risk assessment. Secondly, the SARIMA method is integrated into the TOPSIS. This addition is aimed at refining the model's time series analysis capabilities, allowing for the incorporation of historical trends and continuous updating of the risk profile with the most recent data and information.

#### 3.1 Country operational risk assessment based on TOPSIS

The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a well-known MCDA method often used to assess and rank among different alternatives based on multiple and often conflicting criteria (Hwang and Yoon 1981). In our paper, we use TOPSIS to score and rank different countries on their public security risk level. The basic idea behind TOPSIS is to find the alternative that has the shortest distance to the positive ideal solution (PIS) and the greatest distance from the negative ideal solution (NIS). The PIS is a hypothetical solution that has the best possible values for all of the criteria, while the NIS is a hypothetical solution that has the worst possible values for all of the criteria. In our case, all the risk determinants' value for public security has a negative preference direction—the lower their values, the better. Therefore, to use the TOPSIS to assess a

country's operational risk among a group of countries is to find a country's risk determinant value's difference compared to the hypothetical country with the worst value of all three public security risk determinants of that country group. The value of the risk determinants is often named as utility in the MCDA process, and we define the utility function of the operational risk determinants through the following steps.

Given that the public security risk category has  $M$  distinct risk determinants, for example, as numbers of wars, crimes, social unrest, and terrorism activities in a country within a given period (referred to as  $RD_m$ , where  $m$  belongs to set  $M$ ,  $M = \{\text{war, crime, unrest, terrorism} \dots\}$ ), we are assessing  $N$  countries (notated as  $CO_n$ , where  $n$  belongs to set  $N$ ).

The utility function  $f_n(m)$  quantifies the preference level of an individual risk determinant  $RD_m$  for a given country  $CO_n$ . Broadly speaking, a lower probability of encountering adverse events such as war, crime, unrest, and terrorism activities is more desirable. As such, the utility function for individual operational risk determinants can be computed directly based on the likelihood that the total number of adverse events in the evaluated country for a certain period is below a certain threshold.

We introduce  $AV_w(m)$  to represent the worldwide monthly average number of adverse events per million people. For each country  $n$ , we also define  $AV_n(m)$  as the monthly average number of adverse events per million of the country  $n$ 's population. The utility function for  $RD_m$  in country  $n$  is given by:

$$f_n(m) = \frac{AV_n(m)}{AV_w(m)} \quad (1)$$

This equation essentially measures the country's adverse event rate relative to the worldwide average, providing a standardized comparison of the operational risk determinants across countries.

To utilize the TOPSIS framework, we need to generate a weighted decision matrix. We firstly assign weights  $w_m$  to each risk determinant  $RD_m$ . The weights can be assigned based on the importance of each risk determinant. Then, the weighted utility is generated by multiplying each  $f_n(m)$  by its corresponding weight  $w_m$ . The weighted utility matrix is then:

$$F_n(m) = f_n(m) \times w_m \quad (2)$$

As all the risk determinants are harmful indicators, with lower values being better, the ideal solution (IS) will be the one with the minimum  $F_n(m)$  value for each  $m$  across all countries  $n$ , and the negative ideal solution (NIS) will be the one with the maximum  $F_n(m)$  value for each  $m$  across all countries. Therefore, for each  $RD_m$ ,



$$IS_m = \min(F_n(m)), \quad NIS_m = \max(F_n(m)) \quad (3)$$

Next, calculate the separation of each country from the ideal solution and the negative ideal solution, using the Euclidean distance. The separation  $S_n^+$  from the ideal solution  $IS_m$  is:

$$S_n^+ = \sqrt{\sum_{m=1}^M (F_n(m) - IS_m)^2} \quad (4)$$

The separation  $S_n^-$  from the negative ideal solution  $NIS_m$  is calculated as:

$$S_n^- = \sqrt{\sum_{m=1}^M (F_n(m) - NIS_m)^2} \quad (5)$$

$CO_n^{ci}$ , as the relative closeness to the IS, is the country's public security utility score and can be calculated as

$$CO_n^{ci} = \frac{S_n^-}{S_n^+ + S_n^-} \quad (6)$$

The country with the highest  $CO_n^{ci}$  is considered to have the lowest public security risk.

### 3.2 Integrating SARIMA into TOPSIS

As introduced previously, STSP is a pivotal method in time series analysis. It focuses on predicting future data points by understanding underlying components such as trend, seasonality, and residual randomness. Various models are available for this purpose, including but not limited to State Space Models and Exponential Smoothing Models. These models employ techniques like the Kalman filter to allow for the simultaneous estimation of these elements.

Since operational risk assessment often relies on time series data, it is inevitable to consider the following elements:

- **Cross-year trend:** The public security level of a country might degrade, improve, or fluctuate over the surveyed years. This can be discerned from the historical count of adverse events.
- **Seasonal trend:** Public security levels also exhibit seasonal variations. For example, public security generally improves during winter compared to summer, primarily because people are more inclined to stay indoors during colder months.
- **Residual randomness:** Apart from cross-year and seasonal trends, countries' security levels also exhibit randomness. This could be attributable to yet-to-be-identified underlying risk determinants, stochastic

errors in data collection, or the influence of external covariate factors.

To capture the three aspects of the time series dataset, we employ the Seasonal Autoregressive Integrated Moving Average (SARIMA) as a representative STSP method. SARIMA decomposes the observed time series into various components, namely cross-year trend ( $T_t$ ), seasonal trend ( $S_t$ ), and residual randomness ( $E_t$ ). Let  $X_t$  represent the dataset of the historical count of adverse events at time  $t$ . The SARIMA model can then be represented as:

$$X_t = T_t + S_t + E_t \quad (7)$$

The components  $T_t$  and  $S_t$  can be modeled as follows:

$$\begin{aligned} T_t &= \mu + \beta \cdot t, \\ S_t &= \gamma \cdot \sin(2\pi ft + \phi), \\ E_t &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \quad (8)$$

where  $\mu$  is the mean level of  $T_t$ ,  $\beta$  is the slope of  $T_t$ ,  $\gamma$  is the amplitude of  $S_t$ ,  $f$  is the frequency of  $S_t$ ,  $\phi$  is the phase shift, and  $\sigma^2$  is the variance of the normally-distributed random variable  $E_t$ .

The parameters  $(\mu, \beta, \gamma, f, \phi, \sigma^2)$  can be continuously updated as new data becomes available. This, in turn, produces new  $X_t$  values which update  $AV_w(m)$  and  $AV_n(m)$  accordingly. These continuously updated  $AV_w(m)$  and  $AV_n(m)$  can then be employed to calculate the country's public security utility score  $CO_n^{ci}$  through the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method at each data-updating point. The integration process of TOPSIS and SARIMA is illustrated in Fig. 1.

### 3.3 Operational risk assessment data ingesting

The empirical study presented in this section is designed to evaluate the functionality of the integrated SARIMA and TOPSIS model proposed in Sect. 3. Specifically, it tests the model's capacity for comprehensive operational risk assessment based on multifaceted risk determinants. It further explores the model's ability to integrate new data and information over time to improve its forecasting performance. The empirical study includes three steps, as illustrated in Fig. 1.

The empirical study presented in this section is designed to evaluate the functionality of the integrated SARIMA and TOPSIS model proposed in Sect. 3. Specifically, it tests the model's capacity for comprehensive operational risk assessment based on multifaceted risk determinants. It further explores the model's ability to integrate new data and information over time to improve its forecasting performance. The empirical study includes three steps, as illustrated in Fig. 2.

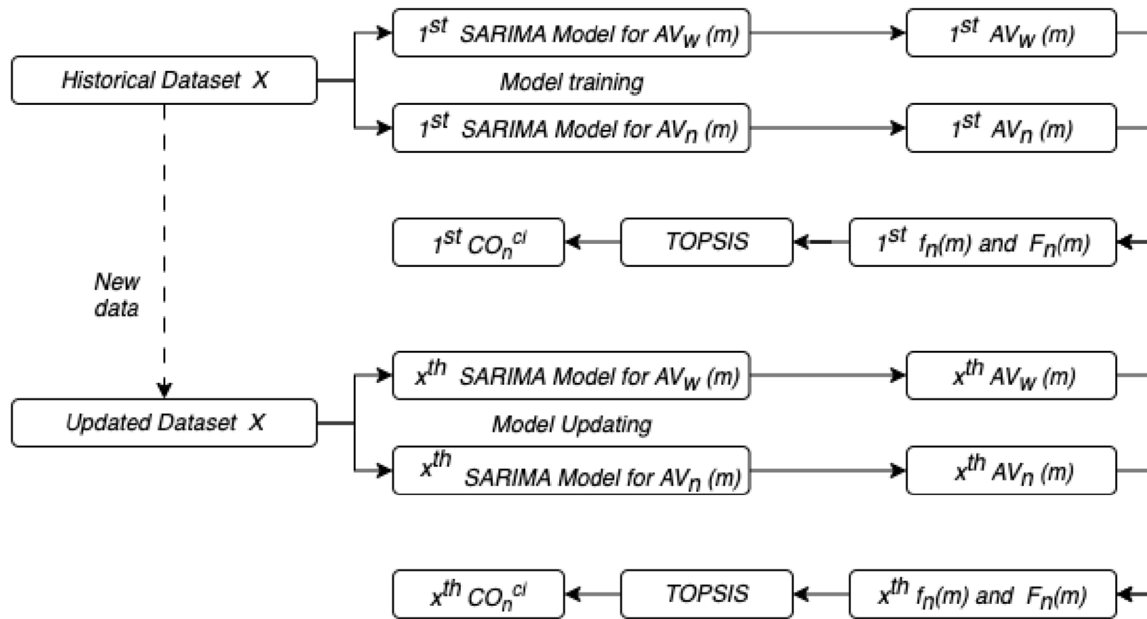


Fig. 1 Integration of TOPSIS and SARIMA

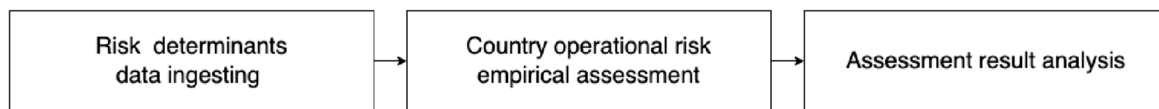


Fig. 2 Empirical study workflow

### 3.3.1 Operational risk determinant

As previously discussed, country operational risk often includes a complex interplay of various interdisciplinary risk determinants, which are notably dynamic and intricately interconnected. For modeling purposes, and to maintain the tractability of our analysis, we narrow our focus to four primary risk determinants for country operational risk assessment, as explained in detail below:

- **War:** This risk determinant measures a country's overall exposure to both civil and interstate warfare and conflicts that are conducted by political, religious, and military entities. Operational risks can arise in various ways from war and conflict of significant magnitude, including infrastructure damage, supply chain disruption, loss of market access, or potential harm to employees.
- **Crime:** This determinant evaluates the overall prevalence of crime within a country, encompassing both organized and individual criminal activities. These activities can vary from minor offenses to serious violent incidents perpetrated by non-governmental, non-military, and non-political entities. This determinant can be quantitatively evaluated using real-time national crime statistics, such

as crime report cases, homicide rates, incarceration rates, and the circulation of illicit small firearms.

- **Terrorism:** This risk determinant captures the level of terrorist activities within a country, which can be fueled by extreme religious, political, and ethnic ideologies. This determinant can be quantitatively measured using data such as the number of terrorist attacks and resulting fatalities.
- **Unrest:** This risk determinant reflects the overall intensity of social discontent within a country. The causes of social unrest can be diverse, encompassing income inequality, public resentment due to poverty and unemployment, exploitation of public resources by elites, or divisions based on political, religious, ethnic, or community lines. This determinant can be quantitatively assessed by monitoring the frequency of demonstrations and riots in a country.

We use the Armed Conflict Location & Event Data Project (ACLED) as the real-time data source for war, crime, social unrest, and terrorism data collection. ACLED is a leading provider of real-time data on political violence and protest events around the world. The project was originally conceived by Prof. Clionadh Raleigh and launched in 1997 at

the University of Sussex (Raleigh et al. 2010). ACLED has since grown and expanded in scope, providing data from 180 countries from 1979 onward. ACLED's data is used widely across academia and industry. In academia, researchers have free access to use it to study conflict patterns, violence against civilians, the impact of climate change on violence, and many other topics. In industry, particularly in risk management, humanitarian, and development sectors, ACLED data supports threat analysis, forecasting, and strategic planning. The information ACLED provides is especially valuable for organizations operating in conflict zones or areas with high levels of political instability. Please also note that ACLED is not exhaustive. While it is widely used and freely accessible for academic research, other sources exist that might provide more extensive coverage and be updated more frequently but are not publicly available. Meanwhile, although ACLED is a widely utilized resource for studying conflict patterns, it is not exempt from potential biases (Miller et al. 2022). For instance, ACLED relies heavily on media reports, which are susceptible to biases in terms of over-reporting violent incidents and under-reporting peaceful ones, especially in conflict-prone regions. Additionally, there can be an urban bias due to the concentration of media in cities, which could lead to over-reporting of urban incidents as compared to rural ones. ACLED's primary use of English-language sources can introduce bias in non-English-speaking regions. The coverage and accuracy of ACLED's data have also improved over time, which might introduce a temporal bias in longitudinal studies. Lastly, ACLED's event classification might differ from other sources, possibly introducing biases in this area.

In our data experiment, we establish event retrieval criteria for each risk determinant, following the guidelines set out in the ACLED Codebook (Raleigh et al. 2010), as

detailed in Table 1. However, it is crucial to recognize that, due to the inherent complexities and potential inaccuracies in event recording and categorization, the retrieved events may not perfectly align with our criteria. There could be overlaps, where events returned for a specific determinant may actually pertain to another category. Similarly, there may be omissions where certain events that should have been classified under a specific determinant are not retrieved due to imprecise categorization. Further refinement of the retrieval criteria, such as including more detailed specifications regarding actors involved, could enhance data precision and consequently the accuracy of the results.

We utilize Python to implement the above event retrieving criteria, by using ACLED's Data Application Programming Interface (API). Based on the events returned, we generate four separate tables as the training dataset for each of the four types of adverse event as war, crime, terrorism, or unrest. The structure of these tables is exemplified in Table 2.

In Table 2, the column *CountryISO* designates the country, while *EventNum* refers to the aggregated adverse events for a particular month, normalized by the country's population in million units. *Event\_month* identifies the specific month within the modeling period when the adverse event is tabulated for that country, and *EventNum\_average* represents the mean of *EventNum* for all countries during that month. If no adverse event of a specific type is reported by ACLED for a given country and month, a zero is recorded

**Table 2** ACLED adverse event monthly rate table structure

CountryISO	Event_month	EventNum	EventNum_average
...	...	...	...

**Table 1** Criteria for retrieving events by determinant

Determinants	Events retrieving criteria
War	For war, we intend to return the events that often involve state and non-state actors and the gaining of territory through battles and armed clashes Battles: a) Government regains territory; b) Non-state actor overtakes territory; c) Armed clash
Crime	For crime, we intend to return the crime events that are conducted mostly person to person, physically. This is to separate crime events from terrorism, as well as from war and conflict event categories Violence against civilians, including a) Sexual violence; b) Attack; c) Abduction/forced disappearance; d) Disrupted weapons use; e) Mob violence; f) Looting/property destruction
Terrorism	For terrorism activity, we primarily concentrate on incidents carried out remotely or suicidal, by explosive or chemical weapons and devices Explosions/remote violence: a) Chemical weapon; b) Air/drone strike; c) Suicide bomb; d) Remote explosive/landmine/IED; c) Grenade
Unrest	For social unrest, we intend to return demonstration and riot events of any kind Demonstration events: a) Protests: peaceful protest / protest with intervention / excessive force against protesters; b) Riots: violent demonstration / mob violence
Data period	2000 days before 2023.07.26, which is the date of the modelling process

for the *EventNum* value, ensuring data completeness across all four tables. This method enables seamless merging of the tables using *CountryISO* and *Event\_month* as keys, eliminating potential issues with null values. The data is organized into four distinct tables, each containing the monthly rates for one of the four adverse events, spanning 161 countries and a period of 65 months, from March 2019 to July 2023. These tables subsequently serve as the historical training dataset for the modeling process.

### 3.3.2 Risk determinants correlation assessment

Understanding the correlations between different factors or variables is pivotal in constructing accurate, reliable, and interpretable models. In this section, we first investigate the autocorrelations of adverse event rates across the entire time frame, aiming to discern how significantly past data influences future occurrences for each of the four types of adverse events. This analysis is vital for time-series prediction models such as SARIMA, where understanding temporal dependencies can enhance predictive accuracy. Subsequently, we assess the cross-correlations between the four adverse event rates as time-series datasets. This examination aims to determine the degree of interdependence among these events, a critical consideration for TOPSIS ranking.

*Adverse event rate autocorrelations across the time Frame*  
We calculate the autocorrelation of the four adverse event rates (crime, unrest, terrorism, and war) throughout the time frame across all countries. This analysis aims to understand how historical data correlates with the present occurrence of each adverse event, an essential aspect for time-series prediction models such as SARIMA. The findings reveal diverse autocorrelations for different adverse event rates across various countries, with both positive and negative autocorrelation values present. Interestingly, war exhibits the highest average autocorrelation coefficient across all 161 countries at 0.19, followed by unrest events at 0.17, terrorism events at 0.16, and crime events at 0.13. This ordering is intuitive: war, often marked by persistence and lagging effects, has the most substantial correlation, while unrest and terrorism show repetition with varying magnitudes and patterns for a country or territory. Crime, on the other hand, is often more individualized and exhibits more random

patterns, thus resulting in the smallest autocorrelation coefficient.

While the average autocorrelation coefficient appears relatively small and may not seem ideal for SARIMA modeling, it is important to recognize that the autocorrelation coefficient for individual countries can vary significantly and, in some cases, be quite high in both positive and negative directions. Countries with substantial adverse event data often exhibit stronger positive or negative autocorrelation values. Such autocorrelations facilitate more accurate forecasting through the SARIMA model. As illustrated in Table 3 below, countries like Afghanistan, Iran, Libya, and Poland demonstrate notable autocorrelation coefficients across various adverse event types, such as war, unrest, terrorism, and crime, from 2020 to 2023. These higher coefficients underscore the potential of SARIMA to effectively model and predict adverse events.

*Cross-correlations for adverse event rates as time series dataset*  
Cross-correlation is a statistical measure used to describe the degree of similarity or dependence between two time series at varying time lags. The level of dependence between different criteria is an important consideration in the TOPSIS process. In general, TOPSIS prefers lower dependence between criteria, as it ensures that each criterion is evaluated on its unique merits and contributes distinct information to the decision-making process. High interdependence between criteria can lead to redundancy and potentially bias the ranking and weighting process, undermining the accuracy and integrity of the overall evaluation.

We calculate the cross-correlations for the four adverse event rates as four time series data pair-wisely. To streamline the analytical process, we made a simplifying assumption that the cross-correlation remains invariant when reversing the order of the time series in each pair. Similar to adverse event monthly rate autocorrelation, the results reveal diverse cross-correlations for different adverse event pairs and for different countries, with both positive and negative values existing. The average cross-correlation coefficients between various pairs of adverse event rates across 161 countries are listed in Table 4

In Table 4, the moderate correlation between terrorism and war (0.209) may indicate underlying factors influencing both. Meanwhile, the weaker correlations, such as crime

**Table 3** Rank and adverse event monthly rate autocorrelation coefficient from 2020 to 2023

Rank	actual_value_war		actual_value_unrest		actual_value_terrorism		actual_value_crime	
	Country	Value	Country	Value	Country	Value	Country	Value
1	Afghanistan	0.929	Iran	0.734	Afghanistan	0.937	Burkina Faso	0.797
2	Libya	0.909	Poland	0.693	Ukraine	0.919	Poland	0.780
3	Azerbaijan	0.902	Ukraine	0.664	Myanmar	0.834	Angola	0.778
4	Yemen	0.854	Bangladesh	0.651	Libya	0.801	Slovakia	0.776
5	Myanmar	0.841	USA	0.612	Yemen	0.792	Sudan	0.746



**Table 4** Cross-correlation coefficients of adverse event pairs

Adverse event pair	Terrorist versus war	Crime versus war	Crime versus terrorist	War versus unrest	Crime versus unrest	Terrorist versus unrest
Cross-correlation coefficient	0.209	0.091	0.074	0.074	0.064	0.033

vs. war (0.091), crime vs. terrorist (0.074), war vs. unrest (0.074), crime vs. unrest (0.064), and terrorist vs. unrest (0.033), suggest a mild relationship between these risk factors. In the context of TOPSIS analysis, which requires a certain level of independence between criteria, the relatively low values of these coefficients may be seen as beneficial. It implies that the adverse events are not highly interdependent, allowing for a more accurate ranking of countries based on diverse and distinct aspects of risk.

However, the average cross-correlation coefficient for all 161 countries does not uniformly represent each individual country, as there are significant variations in the coefficients among different countries. Some countries exhibit much larger positive or negative cross-correlated coefficients. These highly correlated criteria, whether positive or negative, may introduce redundancy and bias into the weighting process of the TOPSIS analysis for specific countries, thus potentially affecting the overall accuracy and representativeness of the model.

The analysis in this section identifies specific autocorrelations and cross-correlations within the adverse event rates. The observed autocorrelations, while varying significantly among individual countries, reveal a measurable relationship between historical data and current occurrences, thus meeting the underlying assumptions for SARIMA modeling. Simultaneously, the cross-correlation coefficients, which indicate the degree of dependence between criteria, are mostly low. This low level of dependence is consistent with the requirements of TOPSIS, where independence between criteria is desired, to avoid redundancy and bias. Collectively, these results provide evidence that the data meets the necessary criteria for further analysis using both SARIMA and TOPSIS modeling.

### 3.4 Model contributions

The proposed model integrates MCDA and STSP in a novel manner. While both methodologies have been extensively utilized independently, their integration presents a theoretical innovation. This integration allows for operational risk assessment to adaptively work with multivariate risk factors using time series data. This approach is relatively new and not commonly seen in existing research. The application of MCDA provides a systematic method for handling complex decisions with conflicting criteria. This is crucial

in operational risk management where risks often span multiple disciplines and objectives. The incorporation of STSP, specifically SARIMA, enables the continuous update and utilization of knowledge in risk assessment. This is particularly important in environments where real-time data and knowledge significantly influence decisions.

## 4 Modeling results

In our Python-based model, constructed based on the steps illustrated in Fig. 1, we leverage libraries like NumPy, Pandas, and SciPy for data handling and analysis. For specialized tasks, we use 'pymcdm' library for MCDA and 'statsmodels' library for SARIMA forecasting.

### 4.1 SARIMA training and forecasting results

We partition each of the four ACLED event datasets as exemplified in Table 2 into a 24-month training subset, followed by a 41-month testing subset, and extend for another three months beyond the latest data point, for forecasting.

The SARIMA training on the four partitioned ACLED event datasets generates four new datasets, each one carrying 41 months' past data, with a corresponding forecast value for each data point, and the Mean Squared Error (MSE) of the forecast. The model also decomposes the overall trends of the 41 months' past data into yearly trends, seasonal trends, and residual randomness, for each of the countries. Besides the 41 months' past data, the datasets also include forecasts for the upcoming three months. The forecast value does not have actual values for comparison, as they represent future predictions. The Utility score within each of the four datasets is generated as dividing  $Forecast - EventNum$  by  $Forecast - EventNum - average$  for all historical data points and forecasting values. The SARIMA training result table is exemplified in Table 5.

### 4.2 TOPSIS overall utility score

The calculation of the overall utility using TOPSIS begins by creating a risk matrix as per function (2) defined in Section 3.1. This is done by joining the utility scores for each country-month combination across the four SARIMA training and forecasting result tables, as exemplified in

**Table 5** SARIMA training and forecasting results

Country/ISO	Event month	Event num	Trend	Seasonal	Residual	Event num average	Forecast event num	MSE event num	Forecast average	Utility score
:	:	:	:	:	:	:	:	:	:	:

**Table 6** Risk determinant weight distribution

Risk determinant	War	Terrorism	Crime	Unrest
Weight distribution	0.3	0.3	0.2	0.2

**Table 7** Operational risk matrix

Coun-tryISO	Event month	War util-ity	Crime utility	Unrest utility	Terrorism utility
...	...	...	...	...	...

**Table 8** Country operational risk assessment result

CountryISO	Event month	Overall utility score	Overall rank
...	...	...	...

Table 6, in conjunction with the predetermined weight distribution of the four risk determinants outlined in Table 6.

This operational risk matrix incorporates weighted utility scores for 161 countries and regions, spanning the past 41 months, with the upcoming three months forecasting values, as exemplified in Table 7.

Subsequently, we dissect the operational risk matrix to a subset corresponding to a unique month in the Event month column. To calculate a country’s overall operational risk utility score and the score ranking among all countries for each month separately, we apply TOPSIS calculation functions (3), (4), (5), and (6), as defined in Sect. 3.1, on the subsets of the operational risk matrix, one at a time, until all months are covered. As a result, a country’s overall operational risk utility score and its rank is specific to its monthly performance not across the entire time frame. This method provides a more direct comparison of a country’s current operational risk level relative to others for each specific month. The final result of the modeling is presented in Table 8.

### 5 Result analysis

The analysis of the results from integrating SARIMA and TOPSIS modeling methods in this section proceeds through three consecutive steps:

1. Trends Analyzing and Forecasting by SARIMA: The initial step assesses the model’s ability to assess cross-year trend and seasonal trends within the time-series adverse event rates. This includes an investigation into how effectively the model can incorporate new data to

refresh these yearly and seasonal trends, leading to more accurate forecasting of adverse event rates.

2. **Examining Operational Risk Assessment by integrating TOPSIS and SARIMA:** This step is concerned with the model's capacity to perform a comprehensive operational risk assessment at specific data points, based on forecasted adverse event rates for multiple risk determinants.
3. **Conducting Model Sensitivity Analysis:** The final step involves an examination of model sensitivity. The primary focus here is on testing the sensitivity of weight assignments to various risk determinants, as well as understanding how fluctuations in the data influence the modeling results. This analysis helps to gauge the robustness of the model and its responsiveness to changes in the underlying risk determinants.

Together, these three steps form a structured approach to exploring the capabilities and limitations of the integrated modeling technique, offering insights into its applicability and reliability in assessing time-related trends and risks.

### 5.1 SARIMA forecasting performance

In this section, we assess the SARIMA model's capabilities in utilizing the most recent data to augment forecasting precision at different estimation points. Our evaluation process is twofold: initially, we analyze the model's proficiency in discerning the trends of adverse event occurrences; subsequently, we estimate the degree of accuracy with which the model can predict future values grounded in these identified trends.

*Adverse event trend analysis* Our goal is to evaluate the ability of the SARIMA model to discern the cross-year trend, seasonal fluctuations, and random variations in the occurrences of specific types of adverse events, such as wars, terrorism, and social unrest, within particular geographical and climatic contexts. To facilitate this analysis, we have categorized the 161 countries in our study into four groups, based on their average yearly surface temperature: (1) Cooler countries, with temperatures generally less than 15 °C; (2) Moderate-temperature countries, ranging between 15 °C and 25 °C; (3) Warmer countries, between 25 °C and 35 °C; and (4) Hot countries, where temperatures usually exceed 35 °C. We decomposed the average trends of adverse event rates within these four categories. Cooler countries display the most substantial seasonal differences, with adverse event rate (monthly event number per million of the population) differences of  $-0.1$  to  $0.1$ , greater than those in moderate ( $-0.02$  to  $0.02$ ), warm ( $-0.01$  to  $0.01$ ), and hot ( $-0.01$  to  $0.02$ ) countries. This is probably because moderate, warmer and hot countries tend to have the smallest seasonal event rate gaps, due to lesser temperature variations. The analysis

of terrorism events reveals a weaker seasonal trend and a greater randomness across the four categories, aligning with the understanding that terrorism, often small-scale and individualized, is least affected by environmental and climatic conditions. On the other hand, social unrest clearly exhibits seasonal patterns, with cooler countries having the highest seasonal terrorism event rate difference ( $-0.1$  to  $0.2$ ) and hot countries the lowest ( $-0.01$  to  $0.02$ ). These variations can be attributed to the consistent year-round hot temperatures in hot countries and the significant temperature fluctuations across seasons in cooler countries, both of which profoundly influence the frequency and intensity of group outdoor events, such as demonstrations and riots.

In addition to testing the SARIMA and TOPSIS integrated model's ability to analyze adverse event trends across country groups, we also assess the model's capacity to dissect adverse event trends for specific countries, ensuring the alignment of both levels of analysis. As an example, we examine the analysis and forecast of terrorism event trends in Iraq, as shown in Fig. 3. The SARIMA model generates a cross-year trend, illustrating a consistent yet gradual increase from March 2020 to October 2023. This trend, depicted by the black line in the second panel from the top in Fig. 4, is in line with the actual terrorism event rates, represented by the red trend line at the top of the figure. Additionally, the seasonal trend line in the third panel exhibits a distinct seasonal pattern that echoes the repetitive pattern in the actual event rate (the red curved line in the top panel), particularly in data periods 1, 2, and 3. Interestingly, the range of the seasonal trend ( $-1$  to  $1$ ) is much less than the range of the cross-year trend, which spans from 3 to 6. This indicates that, for terrorist events in Iraq, the cross-year trend is more significant than the seasonal trend. This observation aligns with the previously drawn conclusion that terrorism, in general, is less influenced by environmental and climatic conditions, compared to other factors.

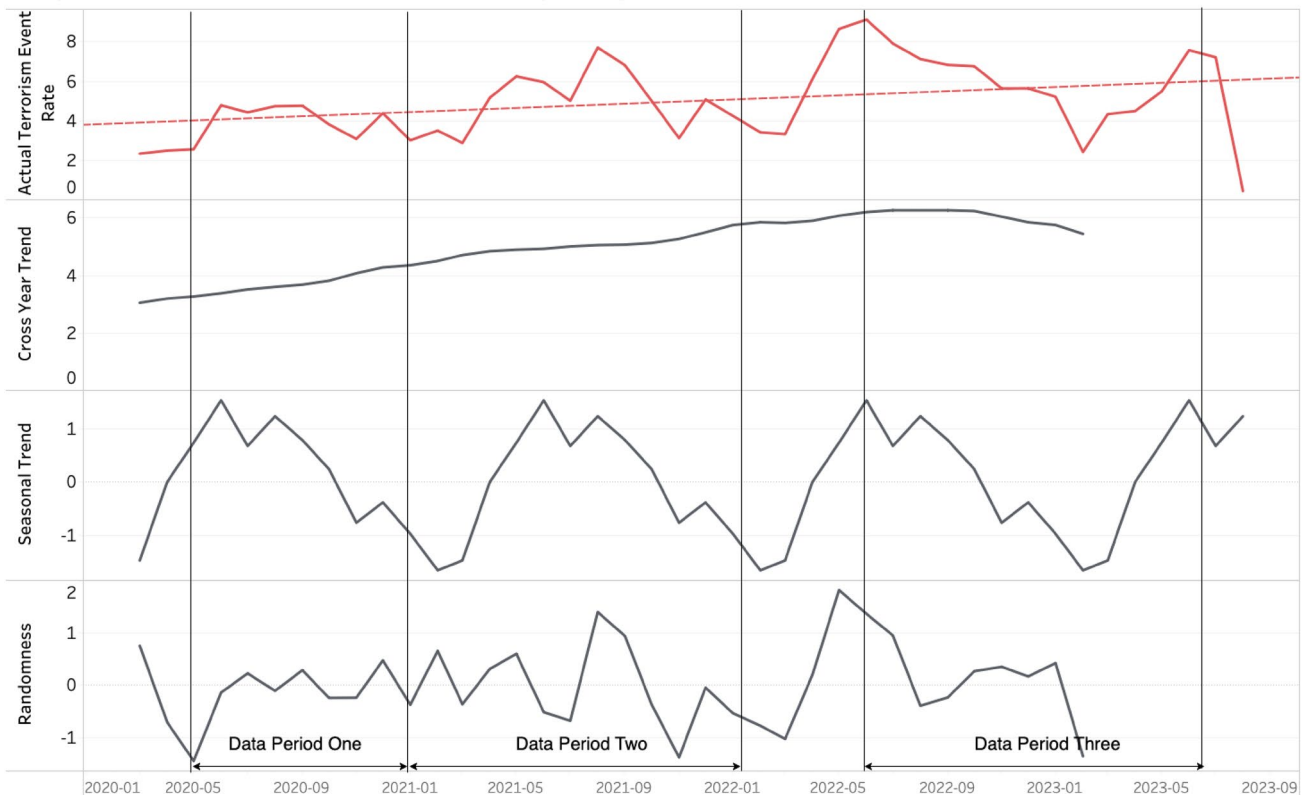
The trends derived from the SARIMA modeling not only align with common understanding but also exhibit consistency between the global adverse event trend analysis and the analysis for specific countries. This consistency provides a reliable foundation for future forecasting.

*Adverse event rate forecasting quality* We are interested in exploring whether the SARIMA model can effectively utilize the latest data to forecast future values. The SARIMA model's forecasts are grounded in the identified cross-year trends, seasonal trends, and randomness found in the historical data. These trends are constantly recalibrated and updated as new data becomes available, which in turn refines the subsequent forecasts. To validate this ability, we refer to our previous example analyzing terrorism event trends in Iraq. We plot the running forecast of terrorism event rates for each month with the actual terrorism event rates for the same period. By comparing

**Table 9** SARIMA model predication quality distribution

Predication quality category (cross correlation coefficient between actual and forecasted values)	Number of countries for the four adverse event types			
	War	Crime	Terrorism	Unrest
0.5 to 1.0: Good Forecasted Quality	17	15	12	8
0.3 to 0.5: Moderate Forecasted Quality	13	18	9	22
0.0 to 0.3: Weak Forecasted Quality	60	69	49	78
Negative: Wrong Forecasts	33	53	33	48
Not Meet Minimum Data Requirement	38	17	58	5

**Iraq Terrorism Event Rate Trend Decomposing** (Unit: Monthly Number of Events per Million Population)



**Fig. 3** Iraq terrorism trend decomposing

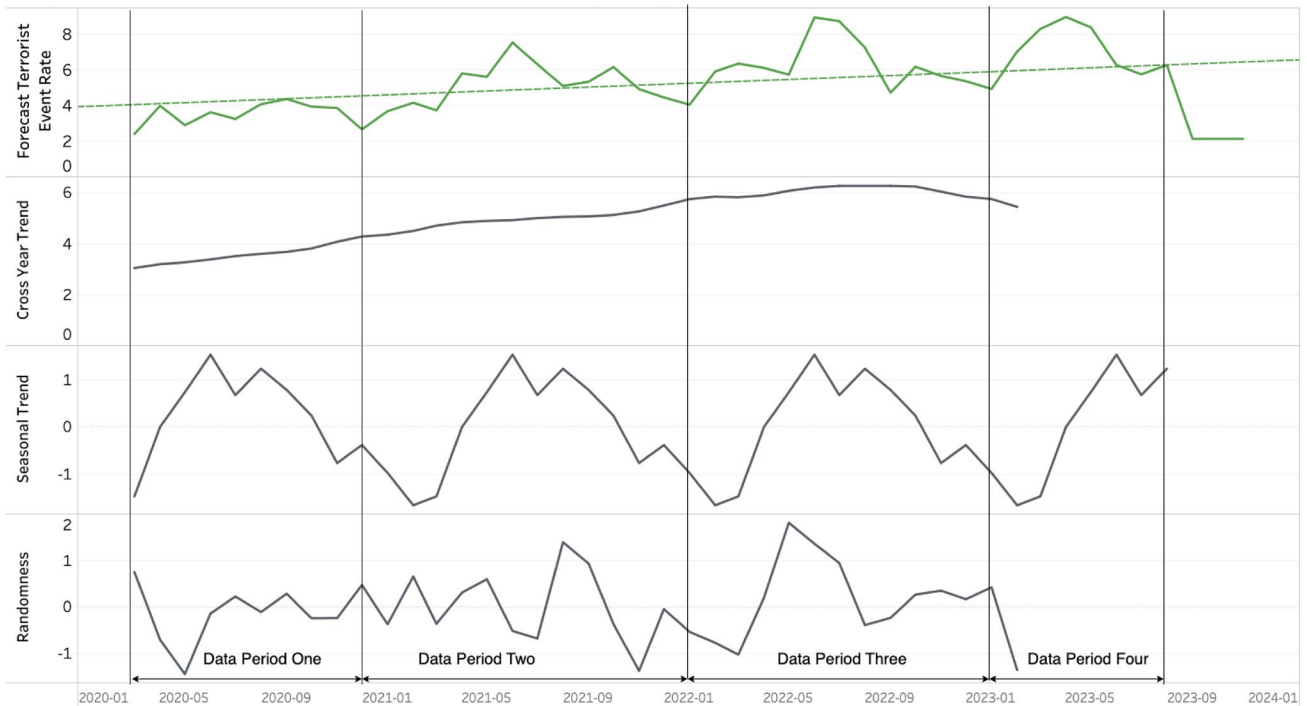
both the forecasted and actual event rates, along with their trends for upcoming values, we are able to assess the model’s capacity for accurate and dynamic forecasting.

First, as illustrated in Fig. 4, we observe that the forecast value curve in the top panel exhibits fluctuations that closely mirror those found in the seasonal trend line, identified based on historical data, and depicted in the third panel from the top for data periods 1, 2, 3, 4. Moreover, the general trend line of the forecast value aligns with the cross-year trend derived from historical data, which is demonstrated in the second panel from the top. This alignment indicates that the forecast has appropriately

incorporated both the cross-year and seasonal trends, reflecting a coherent understanding of the underlying patterns.

The actual rate of terrorism events can experience substantial fluctuations, owing to a multitude of unforeseeable factors, while the forecasted values are often derived from fixed modeling parameters at discrete forecasting intervals. As a result, discrepancies between actual and forecasted values are not only possible but expected. To evaluate the extent of these discrepancies, the cross-correlation coefficient between the forecasted adverse event rate and the actual adverse event rate has

## Iraq Terrorism Event Rate Forecast Trend Decomposing (Unit: Monthly Number of Events per Million Population)



**Fig. 4** Iraq terrorism trend forecasting

been calculated as time-series data. This coefficient provides a quantitative measure of the similarities between the forecasted and actual values in terms of both absolute differences and directional changes over time. The cross-correlation coefficients are categorized into five distinct levels: (1) 0.5–1.0, indicative of strong positive correlation and good prediction accuracy; (2) 0.3–0.5, denoting moderate positive correlation and moderate prediction accuracy; (3) 0.0–0.3, reflecting weak positive correlation and limited prediction success; (4) negative coefficients, signaling incorrect predictions; (5) absence of value, suggesting insufficient data to make predictions. The distribution of countries across these five categories has been analyzed for each adverse event type and is detailed in Table 9.

The SARIMA modeling returns mixed forecast quality. There are strong positive correlations for war, suggesting good prediction accuracy. Unrest has fewer countries with strong positive correlations. Moderate and weak positive correlations are more prevalent for crime and unrest, potentially indicating inconsistencies or noise.

While the areas of crime and unrest have fewer countries that do not meet the minimum data requirements (with crime at 17 and unrest at 5), they exhibit the largest number of incorrect predictions (with crime at 53 and unrest at 48). This inconsistency may stem from two factors. Firstly, the model's methodology may lack the capacity to accurately

forecast events categorized as crime and unrest. Secondly, the inherent randomness of these events may require a more substantial quantity of data to improve forecast quality, compared to areas like war and terrorism, which have stronger repetitive patterns.

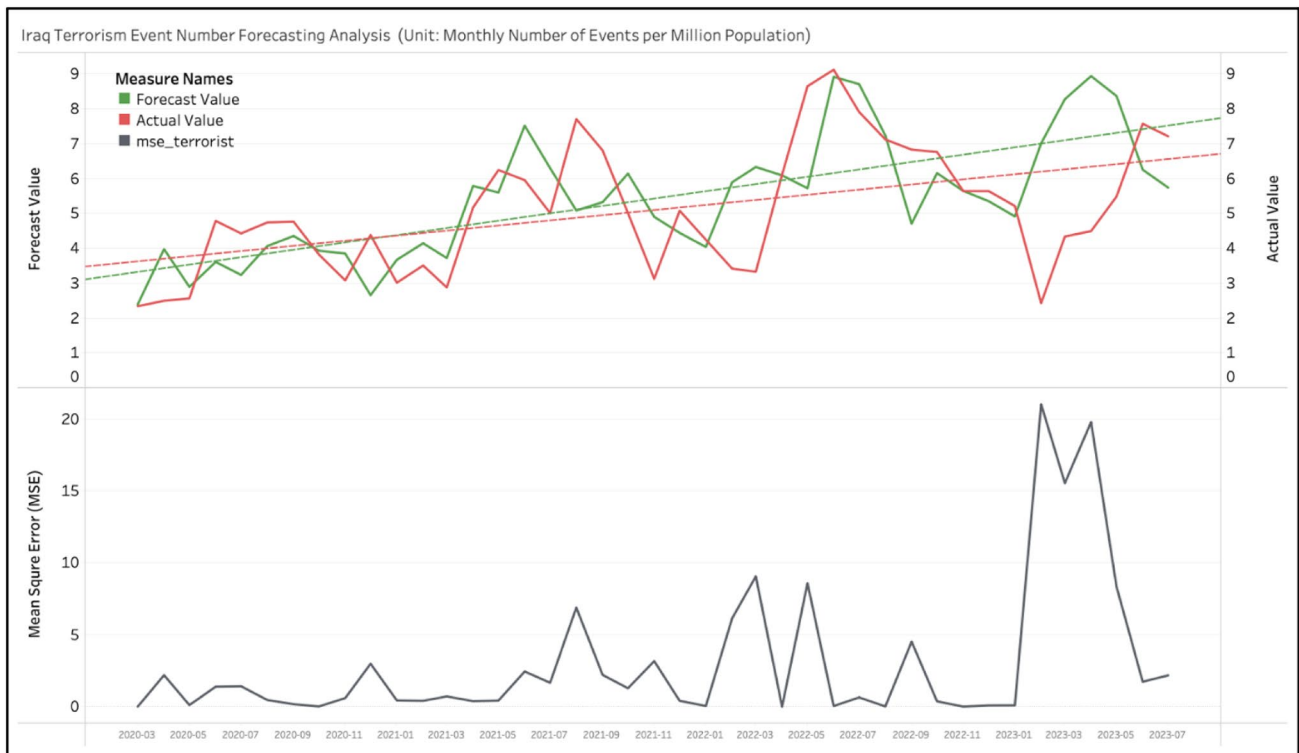
The presence of no data in several instances, especially for terrorism, might also affect the overall interpretation. Overall, the forecast quality is varied, with some success in predicting war but limitations and inconsistencies in other areas.

*Adverse event rate forecasting adaptability* A critical question that may arise is the ability of the forecast to adapt when there is a significant discrepancy between actual and forecasted values. To explore this issue, we examine the forecasted and actual terrorism event rates for Iraq as a representative example.

The forecasted and actual values are juxtaposed on the same time frame, as shown in Fig. 6, to verify both the absolute differences in value and the directional changes of the two measures. The green line, representing the forecast, generally aligns with the actual value's oscillations (depicted by the red line), maintaining a similar range. Notably, when divergence between the forecasted and actual values occurs, or there is a change in direction, the forecast line adjusts its course to realign with the actual value line.

This behavior is further examined through an analysis of the Mean Square Error (MSE) for the forecast, quantifying





**Fig. 5** Iraq terrorism event rate forecasting

the differences between the predicted and actual values. Plotting the MSE value on the same time frame as the forecast and actual values on the bottom of Fig. 5, it is observed that the MSE line rises when a significant disparity occurs between the forecast and actual values. These peaks are followed by declines, as the forecast readjusts to match the actual value's level and direction of change.

The patterns in the MSE line indicate a responsive characteristic in the model. The absence of sustained high MSE values suggests that the forecast continually adjusts along the timeline. This aspect could be interpreted as a sign of adaptability in the model, allowing for adjustments in response to actual event rates of terrorism. The adaptability of the forecast to the latest updated data indicates that the model can respond to new information efficiently. This makes the model not just a tool for prediction but an ongoing, dynamic system for tracking the actual adverse events rate.

## 5.2 TOPSIS operational risk assessment performance

In this section, our goal is to analyze the model's operational risk assessment quality. Since there is no universally recognized standard for operational risk assessment, a range of results may emerge from different models. Even if a universal standard did exist, our model, which utilizes unique risk determinants and assessment methods for experimental

purposes, may produce results that diverge from standard outcomes. This divergence does not necessarily indicate an error or inaccuracy in the modeling methodology we employ. However, recognizing this complexity, our analysis in this section turns to the Global Peace Index (GPI), to gauge the quality of the modeling results (Institute for Economics & Peace 2023). GPI is a pioneering measure that gauges the relative peacefulness of countries and regions worldwide. It includes specific ranking categories, such as the Societal Safety and Security rank, which provides a ranking of how each country performs concerning its population's safety and security. This category assesses diverse factors, including crime rates, political instability, interpersonal trust, terrorist activity, and homicide rates. In our analysis, we use the GPI Societal Safety and Security ranks as a reference point to evaluate our own ranking results. Given that there is no universally recognized standard for operational risk assessment, a difference of 10 in rankings can be considered reasonable and acceptable. This tolerance allows for variations that may arise from the unique risk determinants and assessment methods employed in our experimental model. It is also essential to recognize another aspect of our model: it ranks countries on a monthly basis, allowing for more frequent variations in response to updated information. In contrast, the GPI is updated annually. As a result, the

fluctuations in our ranking results are inherently higher than those in the GPI, leading to additional discrepancies.

Taking all the aforementioned factors into consideration, we classify countries with a rank difference of 10 or fewer as “similar” to our model’s results and the GPI’s rankings. This threshold acknowledges the inherent variability in different modeling approaches, including our model’s unique risk determinants and frequent updating. The percentage of cases that fall within this range serves as a quantitative metric to assess our model’s alignment with the recognized GPI standard. In our specific analysis, we compared the ranking results from July 2023 of our model with the GPI’s Societal Safety and Security rank for the year 2022, resulting in a similarity percentage of 26.14%, which is quantitatively weak.

However, it is essential to recognize the context of our model’s methodology, which is non-supervised and purely quantitative, relying solely on data-driven techniques based on four risk determinants from the same data source for mainly experimental purposes. In contrast, the GPI ranking is derived from both quantitative and qualitative methods, incorporating expert insights and a broader array of risk determinants. The fact that over a quarter of our ranks align closely with such a comprehensive and multifaceted standard indicates that our model possesses a certain degree of validity and quality.

The operational risk rankings generated by our model for July 2023 have been visually represented on the map, as illustrated in Fig. 6. As a more detailed example, Table 10 lists the top 20 countries that have the highest overall operational risk rankings, along with the countries with the highest rankings for subordinate risk determinants.

The rankings presented in Fig. 6 and Table 10 demonstrate alignment with the prevailing global security status. For instance, the top 20 countries with the highest overall operational risk all confront various safety and security challenges. These encompass recent civil and interstate conflicts, sustained social unrest, persistent terrorist threats, and significant civilian crimes. Specific examples include the Ukraine war that began in February 2022, border disputes between Armenia and Azerbaijan from April to July 2023, ongoing terrorist threats in Yemen, Iraq, and Somalia, and the extended social unrest in France, Guyana, and Kenya. The model’s ability to incorporate these recent events into its rankings underscores its effectiveness in considering the latest data and, subsequently, forecasting risk.

Significant discrepancies are also evident. For instance, Afghanistan emerged as the country with the lowest operational risk, ranking at 161, while Iceland’s unexpectedly high rank as the fourth in crime score deviates from common perceptions. An exploration of Table 3, containing the SARIMA training results, uncovers

**Table 10** Country operational risk assessment result

Rank	Overall operational risk		Risk_determinant_war		Risk_determinant_terrorism		Risk_determinant_crime		Risk_determinant_unrest	
	Country	Score	Country	Score	Country	Score	Country	Score	Country	Score
1	Ukraine	0.62	Ukraine	34.16	Ukraine	143.11	Armenia	38.03	Guyana	18.12
2	Armenia	0.44	Azerbaijan	20.75	Syria	9.42	Cyprus	13.72	Eswatini	6.57
3	Azerbaijan	0.34	Armenia	19.03	Myanmar	8.07	Greece	6.91	Cyprus	6.39
4	Guyana	0.33	Somalia	14.48	Yemen	7.92	Iceland	5.14	Haiti	4.63
5	Somalia	0.28	Syria	12.74	Iraq	7.86	Nicaragua	4.17	France	4.54
6	Syria	0.27	Myanmar	8.78	Somalia	4.10	Mali	3.84	Jamaica	3.17
7	Myanmar	0.23	Sudan	7.14	Armenia	3.32	Myanmar	2.71	Kenya	2.43
8	Cyprus	0.23	Jamaica	5.63	Burkina Faso	2.83	Burkina Faso	2.45	Nepal	2.08
9	Sudan	0.20	Yemen	4.64	Russia	2.80	North Korea	2.12	Greece	2.00
10	Jamaica	0.20	Burkina Faso	4.09	Mali	1.38	Eswatini	1.85	Bolivia	1.99
11	Eswatini	0.20	Brazil	3.21	Cyprus	1.07	Georgia	1.71	Chile	1.99
12	Haiti	0.19	Iraq	3.06	Azerbaijan	1.07	Kuwait	1.66	Paraguay	1.71
13	Yemen	0.18	Haiti	3.05	Kyrgyzstan	0.87	Lebanon	1.55	South Africa	1.63
14	Burkina Faso	0.17	Mali	2.89	Tajikistan	0.81	Afghanistan	1.50	Mauritania	1.61
15	Mali	0.16	Lebanon	2.76	Sudan	0.79	Mozambique	1.49	Peru	1.57
16	France	0.16	Kyrgyzstan	2.52	Libya	0.76	Benin	1.44	Colombia	1.48
17	Iraq	0.16	Puerto Rico	2.24	Afghanistan	0.68	Ukraine	1.38	Israel	1.46
18	Brazil	0.16	Central African	2.19	Benin	0.60	Haiti	1.35	Guinea	1.24
19	Greece	0.16	Cameroon	2.03	Lebanon	0.54	Sudan	1.29	Cape Verde	1.18
20	Kyrgyzstan	0.15	Venezuela	1.91	Colombia	0.38	Mauritania	1.27	Equatorial Guinea	1.15



**Fig. 6** Iraq terrorism event rate forecasting

substantial divergence between the forecasted and actual crime event rates for both Afghanistan and Iceland in July 2023. This discrepancy may be traced back to anomalous events that occurred prior to July 2023, leading to a skew in the SARIMA model's forecasting.

Nevertheless, the subsequent operational risk ranking, based on the forecast values for August 2023, exhibits a discernible shift towards more typical standings, with Afghanistan ranked at 128 and Iceland at 78. Although these rankings still defy normal expectations, the evident change in course signals that the model's forecast adjustment mechanism is actively working. This demonstrates the model's capacity to update the operational risk ranking responsively, based on the most recent data and historical trends, underscoring its adaptability in a changing environment.

### 5.3 Model sensitivity analysis

We conducted a sensitivity analysis to assess our model's robustness concerning weight assignments for various risk determinants. Moreover, an analysis was undertaken to gauge the influence of data quality and fluctuations on the ranking outcomes through different time frames.

*Weight assignment sensitivity* In weight assignment sensitivity analysis, we defined five distinct sets of weight configurations to be utilized in the TOPSIS ranking process,

**Table 11** Weight sets for four risk determinants

Weight sets	Crime	Unrest	Terrorist	War
Set 1	0.2	0.2	0.3	0.3
Set 2	0.3	0.2	0.3	0.2
Set 3	0.25	0.25	0.25	0.25
Set 4	0.1	0.3	0.3	0.3
Set 5	0.05	0.45	0.45	0.05

as outlined in Table 11. These selected weight sets were designed to introduce both distribution and magnitude variations. By incorporating these variations, we aimed to assess the model's stability and resilience against extreme weight assignment fluctuations.

We then computed the Spearman rank-order correlation coefficient (often referred to as Spearman's rho) for July 2023 against the five weight assignments. A Spearman's rho value of 0.8972 was derived, indicating a strong consistency in rankings across different weight assignments. This outcome can be interpreted in two ways: From the positive side, even when altering weight assignments, the model's country rankings exhibit marginal variance, suggesting the model's resilience to different hypothetical scenarios. On the other side, the marginal ranking variance suggests that variances in decision-maker preferences exert a minimal influence on country rankings.

Significant discrepancies are also evident. For instance, Afghanistan emerged as the country with the lowest operational risk, ranking at 161, while Iceland's unexpectedly high rank as the fourth in crime score deviates from common perceptions. An exploration of Table 3, containing the SARIMA training results, uncovers substantial divergence between the forecasted and actual crime event rates for both Afghanistan and Iceland in July 2023. This discrepancy may be traced back to anomalous events that occurred prior to July 2023, leading to a skew in the SARIMA model's forecasting.

Nevertheless, the subsequent operational risk ranking, based on the forecast values for August 2023, exhibits a discernible shift towards more typical standings, with Afghanistan ranked at 128 and Iceland at 78. Although these rankings still defy normal expectations, the evident change in course signals that the model's forecast adjustment mechanism is actively working. This demonstrates the model's capacity to update the operational risk ranking responsively, based on the most recent data and historical trends, underscoring its adaptability in a changing environment.

#### 5.4 Model sensitivity analysis

We conducted a sensitivity analysis to assess our model's robustness concerning weight assignments for various risk determinants. Moreover, an analysis was undertaken to gauge the influence of data quality and fluctuations on the ranking outcomes through different time frames.

*Weight assignment sensitivity* In weight assignment sensitivity analysis, we defined five distinct sets of weight configurations to be utilized in the TOPSIS ranking process, as outlined in Table 11. These selected weight sets were designed to introduce both distribution and magnitude variations. By incorporating these variations, we aimed to assess the model's stability and resilience against extreme weight assignment fluctuations.

We then computed the Spearman rank-order correlation coefficient (often referred to as Spearman's rho) for July 2023 against the five weight assignments. A Spearman's rho value of 0.8972 was derived, indicating a strong consistency in rankings across different weight assignments. This outcome can be interpreted in two ways: From the positive side, even when altering weight assignments, the model's country rankings exhibit marginal variance, suggesting the model's resilience to different hypothetical scenarios. On the other side, the marginal ranking variance suggests that variances in decision-maker preferences exert a minimal influence on country rankings.

*Data quality and fluctuations sensitivity* This might raise concerns in certain contexts, for instance where diverse stakeholder perspectives require a more reactive model. However, our specific choice of risk determinants for the

data experiment (war, crime, terrorism, and unrest) tends to have narrower preference disparities. Given this circumstance, the observed level of robustness is acceptable.

For data quality and fluctuation sensitivity analysis, since our model ranks overall operational risk on a monthly basis, we employed Spearman's rho to correlate the rankings across all months, aiming to understand how stable these rankings are over time with monthly data updates and variations. We found an average Spearman's rho of 0.72 across all months. Although this value does indicate some variation in rankings, it also suggests a reasonable level of stability. In practical terms, this means that, while the monthly data updates do have some impact, the operational risk levels for most countries remain fairly stable, without dramatic changes on a month-to-month basis. This reflects a consistent pattern in the operational risk landscape and reinforces the reliability of our model.

In conclusion, the SARIMA and TOPSIS integrated model is able to generate overall acceptable operational risk rankings with some notable exceptions at a given point of time, based on multifaced forecast risk determinant values. The model demonstrates commendable stability in relation to the distribution of risk determinant weights. However, while it exhibits a certain level of sensitivity to data variations, this sensitivity remains within an acceptable and practical range.

## 6 Conclusion

This paper proposes a framework that integrates SARIMA into TOPSIS for continuous assessment of operational risks with multifaced risk determinants. Specifically, the SARIMA method is utilized as part of the STS techniques, enabling the forecasting and updating of risk determinants, based on real-time data and knowledge. Thereafter, the TOPSIS, as an established MCDA technique, is employed to construct a prioritized operational risk matrix. This risk matrix is used to assess operational risks based on the risk determinants, which are continuously updated by the SARIMA model, thereby forming a continually updated operational risk assessment profile that can adjust to immediate changes in safety and security conditions. To test the model's functionality and performance, we use Python to code the model and ACLED as the real-time data source for the model, to assess 161 countries' operational risk level, based on the four risk determinants of war, terrorism, crime, and unrest. The model's performance is thoroughly assessed and concluded upon as follows:

- The integrated SARIMA and TOPSIS operational risk assessment model is functionable and possesses a moderate degree of validity and quality.

- The SARIMA model is able to analyze and identify the cross-year trend, seasonal trend, and randomness of adverse event rates for the risk determinant, and to apply the trends identified to forecast future values. However, the forecast is of mixed quality. Insufficient and sparse data is the major reason for the inaccurate or even wrong forecasting across the four adverse event types. However, the SARIMA is also able to adopt the latest data updates to swiftly adjust its forecast.
- The TOPSIS model, based on the SARIMA forecasted risk determinants as inputs, is able to generate an overall operational risk ranking result that is in general alignment with common understandings, but with notable discrepancies. However, the TOPSIS model result can reflect on concurrent events through SARIMA forecast adjustment, as well as adjusting the ranking result when significant discrepancies occur.
- The sensitivity analysis reveals that the SARIMA and TOPSIS integrated model is relatively insensitive to the risk determinant weight assignment variations, mainly because the four risk determinants we chose have larger cross-correlation coefficients for some countries, which brings redundancies to the model, to reduce insensitivity, regardless of weight assignment changes
- The sensitivity analysis also reveals that the SARIMA and TOPSIS integrated model is sensitive to monthly data updates and fluctuations, but with reasonable stability to reflect that a country and region's operational risk often does not change dramatically, at least monthly.

In conclusion, this study reveals that the integrated model combining SARIMA and TOPSIS exhibits the intended functionality, offering a continuous and comprehensive assessment, grounded in multiple interdisciplinary data sources, as well as the capacity to assess, adjust, and refine, utilizing the most recent data and information. This proof the novel idea that integrating STS and MCDA can provide a functionable framework for iterative uncertainty reduction for multifaced analysis and assessment. In particular, integrating the frequentist STS analysis method as an alternative to a probabilistic approach can handle the scenario in which prior knowledge is hard to obtain and objective uncertainty representation is needed.

### 6.1 Model limitations

Notwithstanding its strengths, the model sometimes displays significant inaccuracies for individual countries at specific points in time, as discussed as following:

- **Data Quality and Availability:** One of the most significant limitations is the dependency on data quality. Just like most of the statistic-driven models, the SARIMA

model's forecasting accuracy is directly impacted by the availability and reliability of data. In cases of insufficient or sparse data, especially concerning adverse event types, the model may produce inaccurate or even incorrect forecasts. This limitation is particularly evident in less-documented regions, or for less-reported types of risk events.

- **Model Sensitivity and Stability:** Although the model shows stability in reflecting operational risks, its sensitivity to monthly data updates can be a double-edged sword. On one hand, this allows for a dynamic and responsive risk assessment; on the other, it can lead to fluctuating assessments that may not accurately represent long-term trends or risks.
- **Cross-Correlation Challenges:** The sensitivity analysis reveals that the model struggles with redundancies due to larger cross-correlation coefficients among the four risk determinants. This reduces the model's sensitivity to weight assignment changes, potentially leading to an overemphasis, or underrepresentation, of certain risk factors.
- **Generalization and Adaptability:** The model's current configuration may not be universally applicable across different geographic regions or risk types. The methodologies and algorithms might need to be tailored to suit various contexts, which can be a complex and resource-intensive process.

### 6.2 Future improvements

Addressing these challenges necessitates concerted efforts in both academic research and industrial practices. The following strategies are proposed:

- **Analyzing Risk Determinants:** A detailed analysis of available risk determinants should be conducted, selecting those with strong autocorrelation and weak cross-correlation within their time series records. This selection process enhances the pattern and trend detection and forecasting capabilities of SARIMA, as well as the balanced analysis within TOPSIS.
- **Ensuring Data Quality:** Adequate and high-quality data is pivotal for a statistics-driven model such as SARIMA and TOPSIS. Employing Big Data technology is advisable for operational risk assessment, given the often requisite interdisciplinary and real-time data in massive quantities.
- **Exploring Alternative Methods:** There are other BI and MCDA methodologies with specific strengths and weaknesses tailored to certain risk assessment scenarios. Developing a model and its settings should involve a detailed theoretical study of these methods, aligned with practical reality checks.



- **Integrating Expert Inputs:** Recognizing that a purely statistics-driven model may not be error-free, even if all necessary requirements are met, the integration of expert inputs for evaluating and calibrating the modeling results is vital. Further academic research into when and how to merge modeling results and expert insights is necessary.

**Funding** Open access funding provided by University of Stavanger & Stavanger University Hospital. The research is self-funded.

**Data availability** The data set and the Python Code used in this study will be provided upon request.

#### Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethical approval** The research was carried out in strict compliance with ethical standards pertinent to academic research.

**Human and animal participants** No human participants or animals were involved in this study, thus negating the need for ethical approval for this particular aspect.

**Informed consent** As the study did not involve human participants, the requirement for obtaining informed consent was waived.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Aven T (2012) The risk concept—historical and recent development trends. *Reliab Eng Syst Saf* 99:33–44

Aven T (2015) *Risk analysis*. John Wiley & Sons, New Jersey

Basel Committee on Banking Supervision (2011) Principles for the sound management of operational risk. Bank for International Settlements

Belton V, Stewart T (2002) *Multiple criteria decision analysis: an integrated approach*. Springer, Berlin, Heidelberg

Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) *Time series analysis: forecasting and control*. John Wiley & Sons, New Jersey

Chen M-F, Tzeng G-H (2004) Combining grey relation and TOPSIS concepts for selecting an expatriate host country. *Math Comput Model* 40(13):1473–1490

Dalla Valle L, Giudici P (2008) A Bayesian approach to estimate the marginal loss distributions in operational risk management. *Comput Stat Data Anal* 52(6):3107–3127

Figueira J, Greco S, Ehrgott M (2005) *Multiple criteria decision analysis: state of the art surveys*

Ghosal S, Van der Vaart A (2017) *Fundamentals of nonparametric Bayesian inference*, vol 44. Cambridge University Press, Cambridge

Gul M, Yucesan M (2022) Performance evaluation of Turkish Universities by an integrated Bayesian BWM-TOPSIS model. *Socio-Econ Plan Sci* 80:101173

Haimes YY (2005) *Risk modeling, assessment, and management*. John Wiley & Sons, New Jersey

Hamid RA, Albahri AS, Albahri OS, Zaidan AA (2022) Dempster–Shafer theory for classification and hybridised models of multi-criteria decision analysis for prioritisation: a telemedicine framework for patients with heart diseases. *J Ambient Intell Humaniz Comput* 13(9):4333–4367

Hwang C-L, Yoon K (1981) *Multiple criteria decision making*. Lect Notes Econ Math Syst 186:58–191

Institute for Economics & Peace (2023) *Global peace index 2023: measuring peace in a complex world*. Technical report, Sydney

Jaynes ET (2003) *Probability theory: the logic of science*. Cambridge University Press, Cambridge

Kenett RS, Raanan Y (2011) *Operational risk management: a practical approach to intelligent data analysis*, vol 106. John Wiley & Sons, New Jersey

Kim J, Moon N (2019) BiLSTM model based on multivariate time series data in multiple field for forecasting trading area. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-019-01398-9>

Lo H-W, Liou JJ (2021) An integrated Bayesian BWM and classifiable TOPSIS model for risk assessment. In: Ren J (ed) *Multi-criteria analysis for risk assessment and management*. Springer, Cham, pp 21–51

McNeil AJ, Frey R, Embrechts P (2015) *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton University Press, Princeton

Miller E, Kishi R, Raleigh C, Dowd C (2022) An agenda for addressing bias in conflict data. *Sci Data* 9(1):593. <https://doi.org/10.1038/s41597-022-01705-8>

Pinto CA (2015) *Operational risk management*. Momentum Press, New York

Raleigh C, Linke A, Hegre H, Karlsen J (2010) Introducing ACLED: an armed conflict location and event dataset. *J Peace Research* 47(5):651–660

Roy B (1991) The outranking approach and the foundations of ELECTRE methods. *Theory Decis* 31:49–73

Shevchenko PV, Wüthrich MV (2009) The structural modelling of operational risk via Bayesian inference: combining loss data with expert opinions. arXiv preprint [arXiv:0904.1067](https://arxiv.org/abs/0904.1067)

Sinha D (1993) Semiparametric Bayesian analysis of multiple event time data. *J Am Stat Assoc* 88(423):979–983

Wallenius J, Dyer JS, Fishburn PC, Steuer RE, Zionts S, Deb K (2008) *Multiple criteria decision making, multiattribute utility theory: recent accomplishments and what lies ahead*. *Manag Sci* 54(7):1336–1349

Wang Y-J (2019) Interval-valued fuzzy multi-criteria decision-making based on simple additive weighting and relative preference relation. *Inf Sci* 503:319–335

Wang Y-J, Liu L-J, Han T-C (2022) Interval-valued fuzzy multi-criteria decision-making with dependent evaluation criteria for evaluating service performance of international container ports. *J Mar Sci Eng* 10(7):991

Wang X, Wang Y, Peng J, Zhang Z (2023) Multivariate long sequence time-series forecasting using dynamic graph learning. *J Ambient Intell Humaniz Comput* 14(6):7679–7693. <https://doi.org/10.1007/s12652-023-04579-9>

Wattayau W, Peng Y (2004) A Bayesian network based framework for multi-criteria decision making. In: Proceedings of the 17th international conference on multiple criteria decision analysis

Wu X, Liao H (2022) A Dempster–Shafer-theory-based entry screening mechanism for small and medium-sized enterprises under uncertainty. *Technol Forecast Soc Change* 180:121719

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.