Financial Engineering and Risk Management (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2024.070516
ISSN 2523-2576 Vol. 7 Num. 5

Research on the sustainability of property insurance based on risk assessment and decision protection models

Yingqi Liu[#,*], Ruohan Wang[#], Linghe Niu[#]

Houston International Institute, Dalian Maritime University, Dalian, 116026, China
[*]Corresponding author: yliu252@cougarnet.uh.edu
[#]These authors contributed equally.

Keywords: Property insurance, ARIMA, TOPSIS, PCA

Abstract: The insurance industry faces significant challenges as the frequency of extreme weather events increases. In this paper, two key issues are investigated, and solutions are proposed. First, an ARIMA model is used to predict the frequency of extreme weather events in the future, and the TOPSIS model evaluates insurance underwriting decisions. The results show that the United States has the highest similarity (0.438) in approaching the ideal solution's positive aspects. Next, principal component analysis (PCA) optimized the insurance decision-making model, which considered the number of extreme weather events, GDP per capita, population density, and resilience to determine the per capita premiums for the five regions. This paper provides a comprehensive set of risk assessment and decision-making protection programs, which improves the decision-making efficiency of insurance companies and provides a scientific basis for community leaders to develop effective building protection measures.

1. Introduction

In recent years, the frequency and intensity of extreme weather events have increased significantly with the intensification of global climate change. These changes caused severe damage to the natural environment and brought significant challenges to the social economy, especially the risk management problems faced by the insurance industry in the face of high-frequency and high-intensity natural disasters. According to a report by the United Nations, the economic losses caused by climate-related disasters have increased over the past two decades. In addition, according to the National Centers for Environmental Information (NCEI), in 2020 alone, the United States experienced 22 extreme weather events with a single loss of more than $1 billion. How to effectively evaluate and manage risks in the insurance industry under this background has become an urgent problem that needs to be solved.

Many scholars have conducted extensive research into this problem. Nguyen Van Anh et al. (2024) used machine learning models to predict and evaluate the risk of extreme weather events and proposed a data-driven insurance pricing strategy [1]. Chen S, Zou Q, Wang B, et al. (2023) adopted the time series analysis method to paper the impact of historical disaster data on future risk assessment and proposed an improved prediction model [2]. In addition, Esfandabadi Z S et al. (2023) built a multi-factor comprehensive assessment model through comprehensive risk analysis to improve the insurance industry's ability to cope with natural disasters [3]. In addition, Javadinejad S (2020) proposed a regionalized risk management strategy for extreme weather events in specific regions [4]. Pagano A J (2019) focused on enhancing insurance companies' anti-risk ability through policy and technical means [5].

Although the above studies have made significant progress in extreme weather event risk assessment and insurance strategy development, there are still some shortcomings. For example, many studies only focus on a single risk factor, ignoring the complexity of the combined effects of multiple factors. Some forecasting models show significant limitations in the face of different regions, and other types of extreme weather events lack universality and flexibility. This paper proposes a new risk assessment and decision protection framework by integrating the ARIMA and TOPSIS models, and principal component analysis (PCA) is used to optimize the insurance decision model. Our work improves the prediction accuracy and applicability of the model and provides the scientific basis and practical tools for the insurance industry to manage risk in the context of frequent extreme weather events. The data in this paper comes from: https://public.emdat.be/data, https://www. swissre. com/ china/, https://www.ncei.noaa.gov/access/monitorg/scec, https://www. imf.org/en/Home, https:// www. un.org/, https://www.usgs.gov/, https:// www.bmuv.de/, https://www.ga.gov.au/.

2. The basic fundamental Model

This paper used the ARIMA, TOPSIS, and Principal component analysis (PCA) models to assess and manage the risk of extreme weather events. Each model has its specific application field and working mechanism, which can be effectively combined to provide comprehensive solutions.

2.1 ARIMA Model

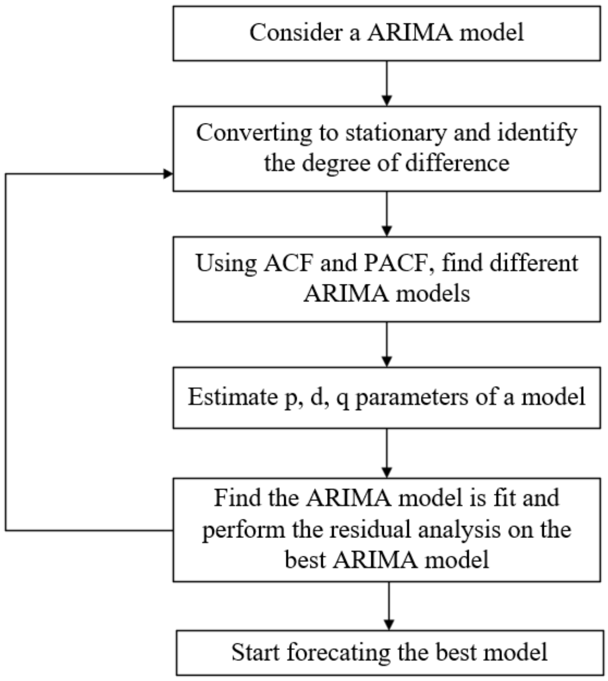

Figure 1: ARIMA model diagram

The model of ARIMA (autoregressive integrated moving average, self-regression integral sliding average) is a statistical model used in time series analysis. It can effectively predict the self-related characteristics of the time series data by combining the sum of the (arand) elements. In this paper, we used the ARIMA model to predict the frequency of extreme weather events in the future. The flow diagram of the ARIMA model is shown in Figure 1.

The ARIMA model is determined by three main parameters (p, d, q), where p is the order of the autoregressive part (AR), d is the difference number used to make a non-stationary time series stationary (I), and q is the order of the moving average part (MA).

The ARIMA prediction model can be written as follows:

$$\phi(B)\nabla^{d}Y_t = c + \theta(B)_{\epsilon_t}, \tag{1}$$

where Y_t is the actual value of t, c is a constant term, $\phi(B)$ is an autoregressive parameter, ϵ_t is the white noise term, $\theta(\mathrm{B})$ is the moving average parameter.

2.1.1 The Function of ACF and PACF—p, q Selection

ACF (autocorrelation function) and PACF (partial autocorrelation function) are both functions used to evaluate the linear relationship between historical data and current values. PACF can help determine the order 'p' of the AR part, while the ACF plot is used to identify the order 'q' of the MA part. The formula for ACF is

$$ACF(q) = \frac{\frac{1}{n-q}\sum_{j=q+1}^{n}(x_j - \bar{x})(x_{j-q} - \bar{x})}{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{2}$$

2.2 TOPSIS Model

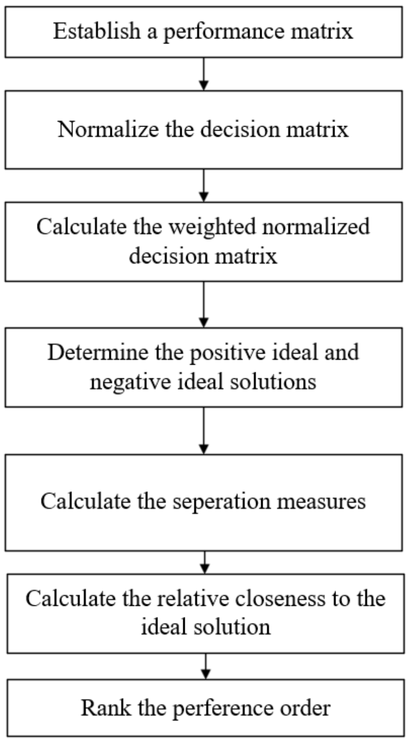

Figure 2: TOPSIS model diagram

The technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) model is a multi-criterion decision analysis method. It prioritizes alternatives by evaluating their distance from ideal and harmful ideal solutions. This paper used the TOPSIS model to evaluate insurance underwriting decisions in different regions.

The basic idea of the TOPSIS model is to determine each scheme's relative advantages and disadvantages by calculating the Euclidean distance between each scheme and the ideal solution (the best solution) and the negative ideal solution (the worst solution). The flow diagram of the TOPSIS model is shown in Figure 2.

We standardize the data to eliminate the dimensional effect of variables so that each variable has the same expressive force.

$$b_{ij} = \frac{a_{ij} - \overline{a}_j}{s_j}, i = 1,2, \ldots, m, j = 1,2, \ldots, n \tag{3}$$

Where

$$\overline{a}_j = \frac{1}{m}\sum_{i=1}^{m} a_{ij}, s_j = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(a_{ij} - \overline{a}_j)^2}, j = 1,2, \ldots, n \tag{4}$$

Then, we normalize the decesion matrix, the original data matrix is normalized to a dimensionless standardized matrix:

$$b_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^{m} {a_{ij}}^2}}, i = 1, \ldots, m, j = 1, \ldots, n \tag{5}$$

According to the importance of each index, ω_j is given to form a weighted standardized matrix:

$$\mathbf{V} = [v_{ij}] \tag{6}$$

Where

$$v_{ij} = \omega_j \cdot b_{ij} \tag{7}$$

The positive ideal solution c_j^* and negative ideal solution c_j^0 are respectively:

$$c_j^* = \begin{cases} max c_{ij}, j \text{ is a benefit attribute} \\ min c_{ij}, j \text{ is a benefit attribute} \end{cases}, j = 1,2, \ldots, n \tag{8}$$

$$c_j^0 = \begin{cases} min c_{ij}, j \text{ is a benefit attribute} \\ max c_{ij}, j \text{ is a benefit attribute} \end{cases}, j = 1,2, \ldots, n \tag{9}$$

Calculate the distance s_i^* from the ideal solution c_j^* and the distance s_i^0 from the negative ideal solution c_j^0.

$$s_i^* = \sqrt{\sum_{j=1}^{n}(c_{ij} - c_j^*)^2}, i = 1,2, \ldots, m \tag{10}$$

$$s_i^{\ 0} = \sqrt{\sum_{j=1}^{n}(c_{ij} - c_j^{\ 0})^2}, i = 1,2, \dots, m \tag{11}$$

where c_{ij} is the j attribute value of place i and$c_j^{\ *}$ is the j attribute value of the positive ideal solution.

The relative closeness of each scheme is:

$$f_i^{\ *} = \frac{s_i^{\ 0}}{(s_i^{\ 0} + s_i^{\ *})}, i = 1,2, \dots, m \tag{12}$$

According to the relative closeness $f_i^{\ *}$ from large to small, the larger the value, the better the scheme.

2.3 Optimize Decision Model

Decision optimization is a branch of mathematics that maximizes the output from many input variables that exert relative influence on the production [6]. Define risk factor R, which combines four independent variables: average annual occurrence of extreme weather (EW), gross domestic product (GDP) per capita, population density (PD), and resistant percentage (RP). Specific weight coefficients d, e, f, and g reflect their contribution to the overall risk. The definition formula is

$$R = d \cdot EW + e \cdot GDP + f \cdot PD + g \cdot RP \tag{13}$$

The weight coefficients d, e, f, and g measure each factor's contribution.

We use principal component analysis (PCA) for weight allocation to retain as much of the original data's variability as possible and reveal the data set's internal structure. PCA is a dimensionality reduction method often used to reduce the dimensionality of large data sets by transforming a large set of variables into a smaller one that still contains most of the information in the large set [7].

The detailed steps of PCA are as follows:

2.3.1 Data Standardization

To ensure that each variable contributes equally to the results, we standardized the raw data, and the standardized formula is

$$\overline{a_{ij}} = \frac{a_{ij} - \mu_j}{s_j}, i = 1,2, \cdots, 10, j = 1,2, \cdots, 4 \tag{14}$$

In the formula, $\mu_j = \frac{1}{10}\sum_{i=1}^{10} a_{ij}$;$s_j$=$\sqrt{\frac{1}{10-1}\sum_{i=1}^{10}(a_{ij} - \mu_j)^2}, j = 1,2,..,4$, μ_j, $s_j$are the sample mean and the sample difference of the JTH indicator.

Let

$$\bar{x}_j = \frac{x_j - \mu_j}{s_j}, j = 1,2, \cdots, 4 \tag{15}$$

as standardized indicator variable **R**

2.3.2 Calculate the Correlation Coefficient Matrix R=$(r_{ij})_{4\times4}$

$$r_{ij} = \frac{\sum_{k=1}^{17} \tilde{a}_{ki} \cdot \tilde{a}_{kj}}{17-1}, i,j = 1,2 \cdots, 4, \tag{16}$$

In this formula： $r_{ij} = 1$; $r_{ij} = r_{ji}$, r_{ij} is the correlation coefficient between index i and index j.

2.3.3 Calculate Eigenvalues and Eigenvectors

Calculate the eigenvalues of the correlation coefficient matrix R: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_4 \geq 0$, and the corresponding standardized feature vectoru1, u2…. u_j=$\left[u_{1j}, u_{2j}, \cdots u_{4j}\right]^T$， Four new index vectors are formed by the feature vectors

$$\begin{cases} y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \cdots + u_{41}\tilde{x}_4 \\ y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \cdots + u_{42}\tilde{x}_4 \\ \cdots\cdots \\ y_4 = u_{14}\tilde{x}_1 + u_{24}\tilde{x}_2 + \cdots + u_{44}\tilde{x}_4 \end{cases} \tag{17}$$

where y_1 is the first principal component, y_2 is the second principal component, y_3 is the third principal component, and y_4 is the fourth component.

2.3.4 Comprehensive Evaluation Value

The information contribution rate and cumulative contribution rate of the eigenvalue λ_j (j= 1, 2, 3, 4) are calculated. Let

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{4} \lambda_k}, j = 1,2, \cdots, 4 \tag{18}$$

is the information contribution rate of the principal component y_i;Let

$$\alpha_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{4} \lambda_k} \tag{19}$$

The cumulative contribution rate of principal components y_1, y_2, $y_3$and y_4.

3. Results

3.1 ARIMA Model Analysis

Calculating the autocorrelation function and partial correlation function determines d=2. AIC and BIC criteria are used to determine the order, and the ARIMA(0,2,1) model is adopted. Parameter estimates for ϕ = 0.2425 and σ = 0.9942 are obtained 0.9942.

$$(1 - 0.24258)(1 - B)X_t = (1 - 0.9942B)\varepsilon \tag{20}$$

Table 1 shows the calculated extreme weather forecasts for the United States and India over the next three years. The number of extreme weather events in the U.S. over the next three years shows an increasing trend each year. Specifically, the number is projected to be 26.10 in 2024, increase to 26.23 in 2025, and rise significantly to 27.84 in 2026. This trend suggests extreme weather events will likely become more frequent in the United States. The situation in India is slightly different from that in the United States. India is projected to experience 11.59 extreme weather events in 2024, which rises significantly to 14.91 in 2025 but falls back to 13.19 in 2026.

Table 1: The number of extreme weather events in the US and India

Country	2024	2025	2026
the US	26.10	26.23	27.84
India	11.59	14.91	13.19

Similarly, a model of the total annual economic losses in the United States and India can be obtained. The U.S. economic losses over the next three years show an upward trend from year to year. India's financial losses are more stable than the United States. The results are shown in Table 2.

Table 2: The total economic losses in the US and India

Country	2024	2025	2026
the US	72082889.18	76784851.25	78083599,48
India	26299819.14	26792747.16	26725362.71

As Table 3 shows, annual insurance payouts in the United States increased each year for the next three years, consistent with the increasing trend in economic losses. However, there was no significant change in insurance claims in India.

Table 3: The annual insurance payout in the US

Data	2024	2025	2026
the US	3800096	3856725	3869319
India	329287.1	329325.5	329326.7

3.2 Principal Component Analysis (PCA) Analysis

After calculation, d = 0.7517, the number of extreme weather events positively correlated with the risk level. e equals -0.1290, denoting that higher GDP generally means greater resilience to disasters, so we assign a negative weighting coefficient to GDP per capita. f=0.0987, meaning that higher-density areas are likely to result in more significant losses in the event of a disaster, thus increasing the level of risk. g=0.0207, the disaster resilience percentage, represents the stability of the building during extreme weather; the higher the stability, the lower the risk level.

According to the above calculation process, we can obtain

$$R = 0.7517 \cdot EW - O,1290 \cdot GDP + 0.0987 \cdot PD + 0.0207 \cdot RP \tag{21}$$

The following table is obtained by plugging the data for each region into the formula. Table 4 shows that the risk factor's R-value varies significantly from region to region. The United States, Australia, and Germany have negative R-values, indicating that these countries have a lower overall risk when faced with extreme weather and other economic impacts.

On the contrary, India and Ethiopia have positive R-values, implying that risk is higher in these regions. Therefore, these regions may need to strengthen disaster prevention measures and enhance emergency response capacity to reduce potential risks.

Table 4: Risk factor R for each region

Region	R
the US	-52.937
India	18.276
Australia	-20.0668
Germany	-24.339
Ethiopia	18.0164

3.3 TOPSIS Results and Analysis

Through calculation, we get the results shown in Table 5.

Table 5: Ranking index values for USA and India

Area	f_i^*
The US	0.438
India	0.331

The United States is closest to the positive ideal solution, indicating that its overall conditions are relatively optimal when considering these indicators. After iterating the data, we set the metric threshold to 0.35 so insurers can cover the United States for the next three years.

4. Conclusions

This paper addresses the risk assessment and decision-making issues facing the insurance industry in the context of frequent extreme weather events. By applying the ARIMA model, this paper successfully predicted the frequency of extreme weather events in the United States and India in the coming years. Then, using the TOPSIS model, this paper comprehensively evaluated insurance underwriting decisions in different regions and identified the optimal underwriting regions. The results show that extreme weather events are rising in both the United States and India, and the TOPSIS model effectively helps us identify priority areas for insurance coverage.

The practical significance of this paper is to provide insurance companies with a set of scientific risk assessment and decision-making methods and improve their coping ability under frequent extreme weather events. These methods can be applied to the insurance industry and other areas that require risk assessment and decision support, such as government disaster management and enterprise risk control. This paper provides a new perspective and tool for risk management in complex environments through the comprehensive use of time series forecasting and multi-criteria decision analysis.

Looking forward to the future, the model established in this paper still has some shortcomings in practical application. For example, ARIMA models rely heavily on historical data and may be less accurate in the face of sudden extreme weather events. However, the TOPSIS model has subjectivity in weight allocation, which may affect the objectivity of decision results. Therefore, future studies may consider introducing more external variables and data sources to improve the accuracy of the predictive model. At the same time, machine learning methods can be combined to optimize weight allocation and improve the robustness of the decision model. In addition, future improvements include building more dynamic and adaptive models to respond to changing climate and environmental conditions. These improvements will further enhance the model's application value and practical effect.

References

[1] Van Anh N, Duc T M. Big Data-Driven Predictive Modeling for Pricing, Claims Processing and Fraud Reduction in the Insurance Industry Globally[J]. International Journal of Responsible Artificial Intelligence, 2024, 14(2): 12-23.

[2] Chen S, Zou Q, Wang B, et al. Disaster risk management of debris flow based on time-series contribution mechanism (CRMCD): Nonnegligible ecological vulnerable multi-ethnic communities[J]. Ecological Indicators, 2023, 157: 111266.

[3] Esfandabadi Z S, Ranjbari M, Scagnelli S D. Prioritizing risk-level factors in comprehensive automobile insurance management: A hybrid multi-criteria decision-making Model[J]. Global Business Review, 2023, 24(5): 972-989.

[4] Javadinejad S, Dara R, Jafary F, et al. Climate change management strategies to handle and cope with extreme weather and climate events[J]. Journal of Geographical Research, 2020, 3(4): 22-28.

[5] Pagano A J, Romagnoli F, Vannucci E. Implementation of blockchain technology in insurance contracts against

natural hazards: a methodological multi-disciplinary approach[J]. Environmental and Climate Technologies, 2019, 23(3): 211-229.
[6] Sun Y, Zhou X, Huang Z. Optimal PID Controller Design for AVR System Based on Multi-objective Optimization and Multi-attribute Decision Making[C]//2021 33rd Chinese Control and Decision Conference (CCDC). IEEE, 2021: 592-597.
[7] Cui J, Li G, Yu M, et al. Aero-engine fault diagnosis based on kernel principal component analysis and wavelet neural network[C]//2019 Chinese Control And Decision Conference (CCDC). IEEE, 2019: 451-456.