Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 9
GBMS 2024

A study based on the impact of extreme weather on the insurance industry

Hengjin Kang *, Duo Zhang, Shengping Kang

Higher Vocational College, ChengDu Neusoft University, Chengdu, China, 611844

* Corresponding Author Email: 18583719930@163.com

Abstract. Extreme weather has become a significant crisis in property insurance, necessitating the avoidance of high-risk areas for investment and asset protection. This paper establishes the Underwriting Model (UM) and Underwriting Decision Model (UDM) to develop strategies that mitigate extreme weather impacts. For UM, Munich Re Worldwide statistics from 2016 to 2023 were combined with Bayes-LSTM modeling to predict the likelihood of weather and disasters on each continent over the next decade. Using the Bayes-TOPSIS model, the risk of each continent was rated, revealing Europe (0.33), North America (0.21), Asia (0.57), Oceania (0.55), Africa (0.51), and South America (0.73) as varying risk levels. The combined Bayes-TOPSIS scores inform insurers' decisions on underwriting. The XgBoost Algorithm was then applied to formulate insurance strategies for the United States and Chile. For UDM, the ARIMA algorithm projected the global population to reach 10.124 billion in 50 years, with an average annual growth rate of 0.49%. Disaster frequency predictions, using the Random Forest Algorithm, and cost-benefit analysis informed the developer's decision-making model, emphasizing earthquake and lightning strike data.

Keywords: LSTM, TOPSIS, ARIMA, XgBoost Algorithm, Random Forest Algorithm.

1. Introduction

With the increasing severity of global climate change, the frequency and intensity of extreme weather events are rising, creating unprecedented challenges for the property insurance industry. Insurers must assess and price risks accurately, especially in regions where such events are increasing, to ensure sustainable operations and long-term stability.[1] Community planners and real estate developers also need to adapt, focusing on improving the resilience of buildings in new construction and retrofitting projects. This includes selecting appropriate materials and strengthening existing structures. Recent studies highlight that insurers must enhance their risk assessment and pricing mechanisms, particularly in high-risk areas, to maintain sustainable operations and long-term stability. This necessitates advanced modeling techniques and robust data analytics to accurately predict weather-related risks.

Community planners and real estate developers also play a crucial role in this adaptive landscape. They need to prioritize the resilience of new constructions and retrofitting projects by selecting appropriate materials and reinforcing existing structures to withstand extreme weather conditions. Moreover, there is a pressing need for effective strategies to protect buildings of cultural, historical, economic, or community importance. Community leaders must ensure these structures are resilient against future extreme weather events, preserving their value and significance. Recent research underscores the importance of integrating climate resilience into building codes and urban planning policies. Insurers are increasingly relying on sophisticated climate models and real-time data to refine their underwriting processes and pricing strategies, ensuring they can cover the heightened risks associated with extreme weather events while remaining financially viable. Insurers require effective methods to assess underwriting risks in high-risk areas and determine pricing strategies. [2]Additionally, community leaders need strategies to protect buildings of cultural, historical, economic, or community value, ensuring their resilience against future extreme weather events.

2. Underwriting decision model

2.1. Disaster data display analysis and forecasting

This text collected data for six continents in recent years, the total losses incurred at various times, the number of disasters, and financial losses on different continents. The data collected consisted of three main areas: physical events, meteorological events, and climatic events. As shown in fig.1. The vertical axis represents the financial losses in 1000 USD, the horizontal axis represents the different continents, including Europe, North America, Asia, Oceania, Africa, and South America.

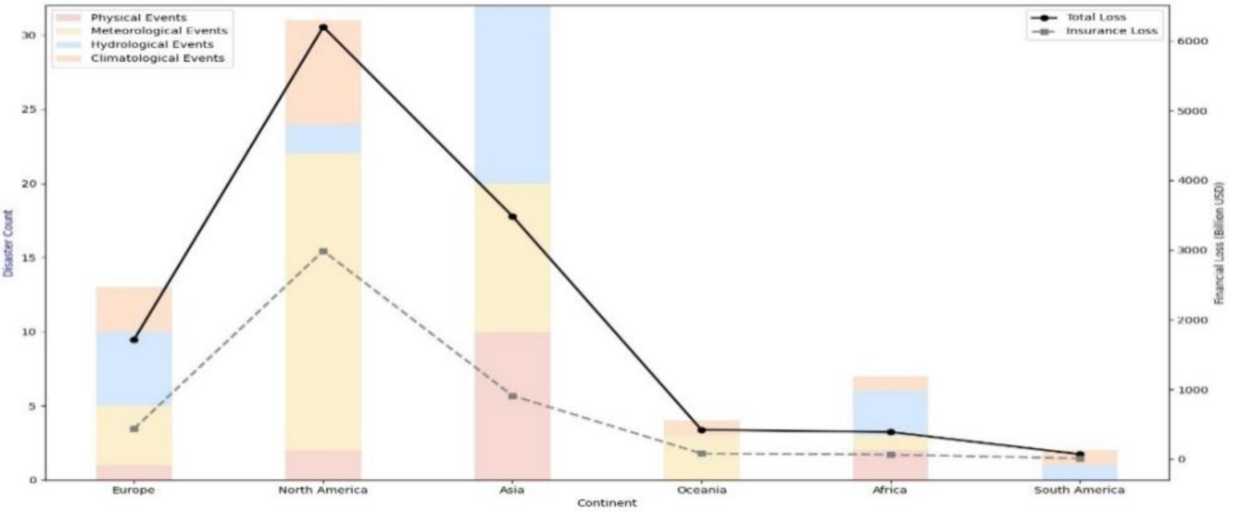

Fig 1. Number of Disasters and Number of Losses by Continent.

From the fig.1, this text can observe that North America has the highest financial losses of all continents, which may be related to the higher level of economic development in the region. Meanwhile, Asia also has higher financial losses, which may be related to its higher population density and the frequency of natural disasters. This may indicate that meteorological and physical events have a greater social and economic impact and that more effective measures need to be taken to minimize the impact of these hazards, and a table of hazard probability statistics has been developed based on hazard data for each continent, as shown in Table 1:

Table 1. Statistics on disaster probability by continent

	South America	Meteorological event	Hydrographic event	Climatic event	Total loss (billions of dollars)	Insurance loss (billions of dollars)
European	0.0095	0.0383	0.048	0.02875	214.5	54.5
North America	0.008	0.080625	0.008	0.028125	775.125	373.45
Asain	0.039	0.039	0.45875	0.005	435.75	113.725
Oceania	0.005	0.09375	0.005	0.03125	52.375	9.96
African	0.035625	0.01775	0.0535	0.1775	48	7.912
South America	0.005	0.005	0.0625	0.0625	8.75	1.275

To visualize the relationship between insurance losses and the occurrence of natural disasters [3-5] ,the model plotted the relationship between insurance losses by continent and year, as shown in fig.2:

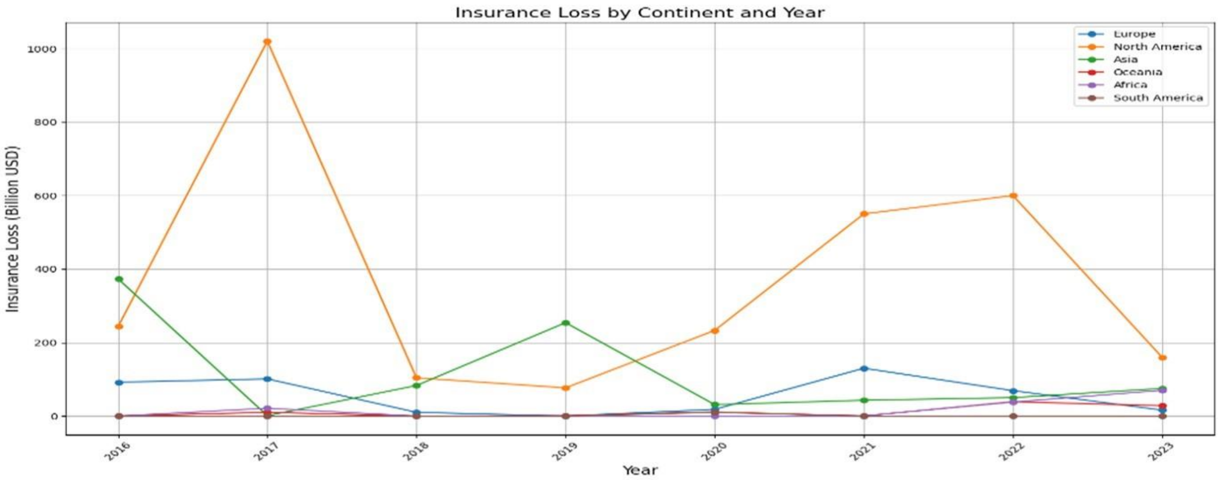

Fig 2. Insurance Loss by Continent and Year.

From the fig.2, shows the insurance losses for different continents between 2016 and 2023. The horizontal axis represents the year and the vertical axis represents the insurance losses, from this graph, can observe that the insurance losses in North America are relatively high between 2016 and 2023, which may be related to the frequency of natural disasters or high economic losses in this region. In contrast, Oceania has relatively low insurance losses.

2.2. Bayes-LSTM enhancement in the case of the TOPSIS model

The text outlines the need to define the factors influencing the insurance program and evaluate the probability and severity of disasters using historical weather data and climate model forecasts. Additionally, the anticipated cost of claims should be assessed by considering disaster risk levels and historical claims data. The stability of the regional environment should be examined using GIS data. Initially, global geographic data was gathered.

To integrate the Bayes-LSTM model, first utilize the LSTM to forecast the probability of occurrence for extreme weather events, which is denoted as P(A|LSTM). This probabilistic output is then treated as a dynamic prior within the Bayesian framework, allowing for a realtime update of our belief system based on the latest sequence data. The modified Bayesian formula incorporating the LSTM predictions is as follows:

$$\mathrm{P(A|B,C,LSTM)} = \frac{\mathrm{P(B|A,C)\cdot P(A|LSTM)}}{\mathrm{P(B|C)}} \tag{1}$$

This model a Bayesian network model. Event A denotes the occurrence of an extreme weather event, event B denotes the insurance company underwriting the policy, and event C denotes the homeowner taking steps to reduce the risk. We can build the following conditional probability table: P(A): the probability of the extreme weather event occurring, P(B): the probability of the insurance company underwriting the policy, P(C): the probability of the homeowner taking [6]measures to reduce the risk, P(B|A, C): the probability that the insurance company underwrites the policy in the event of the extreme weather event and the homeowner taking measures to reduce the risk.

With the training data, this model estimate these probabilities and calculate the conditional probabilities using Bayesian formulas:

$$P(A|B,C) = \frac{(\mathrm{P(B|A,C)}*\mathrm{P(A)})}{P(B,C)} \tag{2}$$

Where P (A|B, C) denotes the probability of an extreme weather event occurring if the insurance company underwrites the policy and the homeowner takes steps to reduce the risk. The specific method of solving the model can be done by using an optimization algorithm to model the following machine learning XgBoost algorithm:

$$\text{maximize } \Sigma(P(B|A, C) * P(A) * P(C) * C) - \Sigma(P(B|A', C') * P(A_0) * P(C_0) * C) \tag{3}$$

Where P (B|A, C) denotes the probability that the insurance company will underwrite the policy if an extreme weather event occurs and the owner takes steps to reduce the risk, P(A) denotes the probability that an extreme weather event occurs, P(C) denotes the probability that the owner takes steps to reduce the risk, C denotes the profit of the insurance company, and Σ denotes the summation of all possible scenarios.

Then TOPSIS score: A TOPSIS score is assigned to each area and the risk level is determined based on the resulting score. The formula for calculating the TOPSIS score is as follows:

$$\text{TOPSIS}_{\text{scorei}} = \frac{C_i - \min(C_1, C_2, \ldots, C_n)}{\max(C_1, C_2, \ldots, C_n) - \min(C_1, C_2, \ldots, C_n)} \tag{4}$$

Districts are categorized into different risk classes based on TOPSIS scores. For example, risk levels can be categorized as high, medium, or low.

Utilizing the predictive strength of the Bayes-LSTM model, this model forecast the likelihood of extreme weather events with higher accuracy. These forecasts serve as input for our Bayes-TOPSIS model, which then evaluates the risk levels of different geographic areas[7]. The Bayes-TOPSIS model assigns a score to each area, which is directly influenced by the LSTM predictions of extreme weather probabilities.

2.3. Presentation and Analysis

The model can bring the required data into the TOPSIS scores for calculation, respectively, to get the TOPSIS scores for the six continents, respectively, 0.33 for Europe, 0.21 for North America, 0.57 for Asia, 0.55 for Oceania, 0.51 for Africa, 0.73 for South America, and it is known that a TOPSIS score of less than 0.3 is a high-risk region, and a score of more than 0.7 is a lowrisk region. Next, for the insurance losses on appeal, analyzed each continent and the results are shown in Table 2:

Table 2. Summary of insurance by continent.

Physical event	Meteorological event	Hydrographic event	Climatic event	Total loss (billions of dollars)	Insurance loss (billions of dollars)
1	4	5	3	1716	436
2	20	2	7	6201	2987.6
10	10	12	0	3486	909.8
0	3	0	1	419	79.7
2	1	3	1	388	63.3
0	0	1	1	70	10.2

Based on the TOPSIS scoring rules and the factors that may cause an insurer to lose money, the text know that North America has the highest level of risk and South America has the lowest, so model demonstrate the model on these 2 continents.

In North America, insurers need to balance higher coverage to offset property risks with adjustments for policyholders’ risk-reduction measures like safety enhancements and additional insurances. Due to frequent natural disasters, coverage may be restricted for those not adopting these measures.

In Chile, lower risk allows for more flexible insurance offerings, from property to health insurance, with reduced premiums to expand the customer base and promote insurance awareness. Despite this, insurers must focus on robust risk management, efficient claims handling, and enhanced customer service to ensure stability and customer loyalty.

3. Developer Decision Modeling for Housing Construction

3.1. Construction of the overall decision model

First in perform the integration of demographic data and use demographic projection models to forecast future trends in world population growth. The modeluse a time series model for predicting population size and trends.

Model use the ARIMA model for population projections:

$$y_s = \varphi_1 y_{s-1} + \varphi_2 y_{s-2} + \cdots + \varphi_p y_{s-p} + \epsilon_s - \Theta_2 \epsilon_{s-2} - \cdots \theta_q \epsilon_{s-q} \tag{5}$$

Where ys is the population size of the time series at moment t. φi (i=1, 2,...,p) and θj(j=1, 2,...,q) is a parameter of the model, estimated by least squares or other optimization algorithms. p and q are the orders of autoregression and differencing, respectively, [8]chosen according to the characteristics of the time series.

In the ARIMA model, model assume that the population size is a smooth time series, i.e., its statistical properties (e.g., mean and variance) do not vary over time. The text use differencing to make the time series smooth and autoregressive and moving average terms to capture the dynamics of the population size. [9] By estimating these parameters, model can use the ARIMA model to predict future population growth trends.

Then when use the ARIMA model for disaster prediction, the specific formula is as follows:

$$H_t = \alpha_1 H_{t-1} + \alpha_2 H_{t-2} + \cdots + \alpha_p H_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1-} \beta_2 \epsilon_{t-2} - \cdots \beta_q \epsilon_{t-q} \tag{6}$$

Where Ht is the disaster risk or impact of the time series at moment t. αi (i=1, 2,...,p)and βj (j=1,2,...,q) are the parameters of the model. Estimation was performed by least squares or other optimization algorithms. p and q are the orders of autoregression and difference, respectively, chosen based on the characteristics of the disaster data. Should have compiled and visualized data on global earthquake magnitudes in 2017, earthquake damages from 1970 to 2021, and earthquake frequency data from 1970 to 2015, to assist insurance companies in disaster prediction, as follows fig.3:

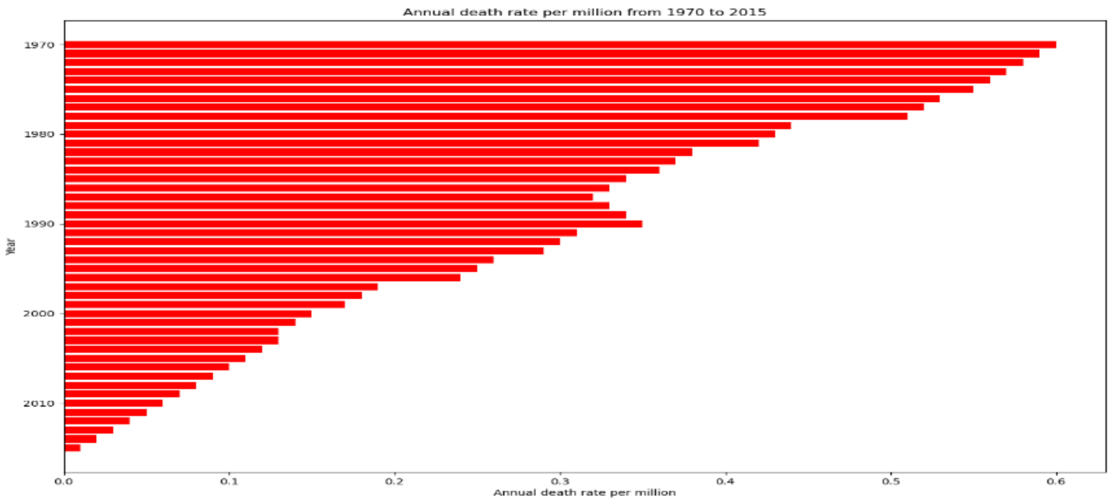

Fig 3. Annual death rate per million from 1970 to 2015.

In the ARIMA model for disaster forecasting, the text assume that disaster risk or impact is a smooth time series, i.e., its statistical properties (e.g., mean and variance) do not vary over time. Model use differencing to make the time series smooth and autoregressive and moving average terms to capture the dynamics of disaster risk. By estimating these parameters, text can use the ARIMA model to predict future disaster risks or impacts.

3.2. Disaster prediction optimization model based on population prediction

As climate change worsens, the rise in extreme weather challenges communities and developers. Adapting insurance models for future risks is essential. Insurers and developers must understand climate trends and enhance building resilience against disasters, incorporating these insights into decision-making to mitigate risks.

For the model, inputs are historical disaster details (type, time, location, impact, and damage), and outputs are predictions of future disasters, typically as probabilities. Due to the complexity of forecasting, should be may use methods like random forests, which consider various features and their interactions for accurate predictions.

$$O = \frac{1}{N}\sum_{i=1}^{N} f_i(X) \tag{7}$$

The process of model training typically uses historical data. First, divide the data into a training set and a validation set. The training set is used to train the model, and the validation set is used to evaluate the model's performance. For the evaluation of model performance use the Mean Squared Error (MSE) metric. The mathematical formula for Mean Squared Error is as follows:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{8}$$

Where y_i is the actual label value, $\hat{y}_i$is the model's predicted value, and N is the number of samples. To better evaluate the model's generalization ability can use cross-validation techniques. The basic idea of crossvalidation is to divide the dataset into K parts, using K−1 parts for training and the remaining part for validation in each round.

Given the complexity of population changes and the variety of influencing factors will choose to use methods of time series analysis and regression analysis. Time series analysis can reveal the time trends and cyclical patterns of population changes, while regression analysis can help us understand the contribution of different factors to population changes. Specifically, time series analysis can use methods such as the ARIMA model or exponential smoothing. The mathematical formula for the ARIMA model is as follows:

$$y_t = c + \phi_2 y_{t-1} + \cdots + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \tag{9}$$

Regression analysis can be conducted using methods such as linear regression, polynomial regression, or logistic regression. The mathematical formula for linear regression is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_n \tag{10}$$

Where y is the population size, xn are the various influencing factors, and βi are the model parameters.

Text divide the data into a training set and a validation set. The training set is used for training the model, and the validation set is used for evaluating the model's performance.

For the evaluation of model performance use metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Rsquared value. These metrics can help us understand the accuracy of the model's predictions. The mathematical formula for Mean Squared Error is as follows:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (11)$$

Where N is the number of samples, representing the number of historical population data points used for training and validating the model, $\hat{y}_i$is the predicted population value of thei_{th}historical data point, that is, the future population estimate given by the model based on input features and algorithm. $y_i$is the actual population value of the i_{th} historical data point.

To better evaluate the model's generalization ability can use time series crossvalidation techniques. The basic idea of cross-validation is to divide the dataset into K parts, training with $K - 1$ parts each time and validating with the remaining part. This is repeated K times, with each data point used for validation once. Finally, take the average of the K validation results as the final evaluation outcome.

Train the SVM model using the chosen kernel function and training data. Find the optimal hyperplane by optimizing the following objective function:

$$\min_{w,b}\frac{1}{2}|w|^2 + C\sum_{i=1}\xi_i \quad (12)$$

Where w represents the normal vector of the hyperplane, b is the bias term, ξ_i is the slack variable, and C is the regularization parameter.

Decision trees use a treelike model to make decisions, with nodes representing decisions based on attributes, and leaves indicating outcomes. They involve selecting key attributes such as disaster risk and population changes. The process involves splitting nodes on the best attribute from the root down, stopping when reaching a set depth or sample size.

4. Conclusions

This study explores how the insurance industry and the real estate development sector respond to extreme weather events by constructing a model of insurance company underwriting and risk and a model of developers' decision-making for housing construction. Model use Bayesian networks and TOPSIS methods to optimize insurance company underwriting strategies and ARIMA models to predict population growth trends and disaster risks to assist developers in making informed decisions.[10]The results show that North America is the region with the highest insurance losses, while South America has relatively low risk. Text recommend that insurers adjust their underwriting strategies based on risk levels while using demographic data and ARIMA models to predict developers' decisions in new construction projects. These models provide a powerful tool for responding to extreme weather events, helping industry decision-makers mitigate risk and achieve sustainable development.

References

[1] Qimeng P, Lysa P, Hong L. Assessing the Effectiveness of the Actuaries Climate Index for Estimating the Impact of Extreme Weather on Crop Yield and Insurance Applications[J]. Sustainability,2022,14(11): 165-196.

[2] University of Waterloo; Waterloo study says insurance industry dangerously unprepared for extreme weather[J]. Global Warming Focus,2018:66-79.

[3] Xu H, Ge Z, Ao W. Research on Climate Change Prediction based on ARIMA Model and its Impact on Insurance Industry Decision-Making[J]. Frontiers in Computing and Intelligent Systems,2024,8(1):1-5.

[4] Sitong Y, Shouwei L, Xue R, et al.The impact of climate risk on the asset side and liability side of the insurance industry: evidence from China[J].Economic Change and Restructuring,2024,57(3):44-53.

[5] Bing L, Weijun Y, Gang C, et al.The threshold effect of climate risk and the non-linear role of climate policy uncertainty on insurance demand: Evidence from OECD countries[J].Finance Research Letters,2023,55(PA):53-77.

[6] Jakob T, Clare M, Vincent H J, et al. Do you manage what you measure? Investor views on the question of climate actions with empirical results from the Swiss pension fund and insurance sector[J]. Journal of Sustainable Finance & Investment,2021,11(1):47-61.

[7] Gatzert N, Reichel P. Awareness of climate risks and opportunities: empirical evidence on determinants and value from the U.S. and European insurance industry[J]. The Geneva Papers on Risk and Insurance - Issues and Practice,2021,47(1):1-22.

[8] New Report Highlights U.S. Insurance Sector's Efforts in Addressing Climate Change Risk[J].M2 Presswire,2023: 21-43.

[9] Brown E. Climate Change Connects the Insurance Industry, Bond Rating Agencies, and Water Utilities[J]. Journal - American Water Works Association,2020,112(4):74-76.

[10] Tatjana M, Dragan M, Karsten M. Examining the impact on mortality arising from climate change: important findings for the insurance industry[J]. European Actuarial Journal,2018,8(2):363-381.