Extreme Weather Risk Assessment and Insurance Decision Making Based on ARIMA Model and Visual Analysis

Hongyu Liu [1], Jingyu Meng [2], Weizhou Li [2], Shuxian Liu [*, 2]

[1] School of Electrical Engineering, Xinjiang University, Urumqi, China, 830018

[2] School of Cyberspace Security, Xinjiang University, Urumqi, China, 830018

* Corresponding Author Email: 75862327@qq.comc

Abstract. Extreme weather events have occurred frequently in recent years, posing great challenges to the insurance industry. In this paper, ARIMA model is used to predict future losses caused by extreme weather, and data visualization analysis is combined to provide risk assessment and decision support for insurance companies. The results show that the ARIMA model can effectively predict future losses, and insurance companies can adjust premiums and invest in disaster mitigation projects according to the forecast results. In addition, through data visualization analysis, the frequency and loss degree of different regions, seasons and disaster types are found, which provides a basis for insurance companies to formulate differentiated premium strategies. This study provides important technical support for insurance companies to make decisions under extreme weather risks.

Keywords: ARIMA model; Data visualization; Extreme weather; Risk assessment; Insurance decision.

1. Introduction

In recent years, the world has suffered more than $1 trillion in losses due to extreme weather, and the property insurance industry has seen a 115% increase in claims for natural disasters in 2020 compared to the 30-year average. This phenomenon has led to rising insurance costs and a growing insurance protection gap, which currently stands at 57% globally. This challenge threatens the sustainability and resilience of insurance markets. If insurers are reluctant to underwrite policies in too many cases, this can lead to too few customers and hence insolvency. Conversely if underwriting policies is too risky, it will lead them to pay too many claims, which in turn will lead to low rates or insolvency. Therefore, in the face of the above crises, a model with the ability to assess extreme weather risk and deploy adjustments is imminent (Figure 1).

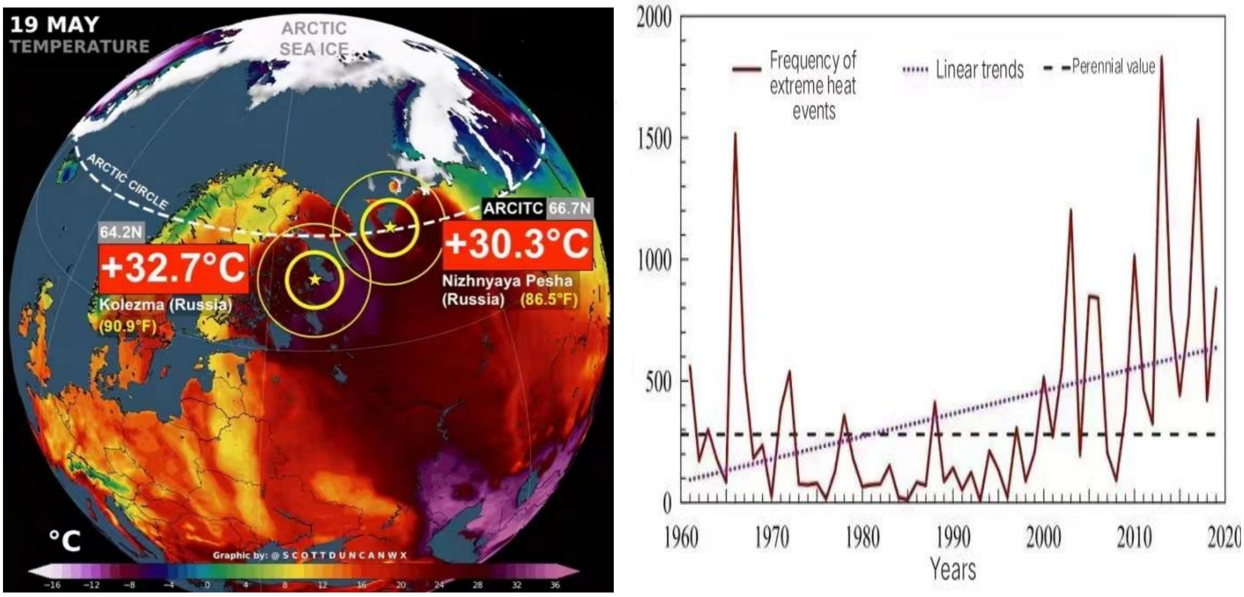

Figure 1. Extreme heat sweeps across North America and Frequency of extreme heat events/station day

With the sustainability of property insurance, the event becomes a crisis for homeowners and insurance companies. As climate change increases, the likelihood of more severe weather and natural disasters increases, with the main events involved being floods, hurricanes, earthquakes, droughts and wildfires along with severe weather, among others.

In recent years, the impacts of global climate change have been increasing and the frequency and intensity of extreme weather events have been rising. This phenomenon has led to a rapid rise in insurance premiums, with climate-driven premiums expected to rise by 30-60% by 2024.The above has led to changes in how and where insurers underwrite coverage. COMAP's mega-catastrophe Insurance Modeler (ICM) wants to deploy property insurance that is resilient period to period, paying out future compensation while ensuring the long-term health of the insurer. What happens when an insurer assumes a policy, when it assumes the risk, and what the property owner does can influence that decision. Adjusting the insurance model and evaluating where and how to build sites. And recommending the future of the community's precious ground, in terms of time and cost, based on four items: culture, history, economics, and risk of damage.

Considering the background information and restricted conditions identified, we need to take several measures to solve the following problems:

(1) Develop a model to predict future losses due to disasters by analyzing the damage in previous years. Inform insurance companies about their insurance efforts. (Indicators and measures to determine whether insurance companies will cover policies in the event of natural disasters and extreme weather)

Our work: The ARIMA model was fitted to determine the parameters of the model to predict the number of deaths in the future disaster through the number of deaths in the past disaster in an area, and the prediction results can provide a reference for the insurance company to make decisions. And the model is analyzed by ACF, PACF, and the parameters are adjusted by the derived data.

(2) Derive decision-making ideas about when insurers should take on risk

Our work: Statistical disaster area data are visualized and analyzed to get the connectivity between area and time disasters, and decision-making ideas are analyzed in this regard.

2. The ARIMA model predicts future extreme weather losses and provides decision support for insurance companies

2.1. Model Preparation

This model belongs to one kind of time series analysis model, which is merged by AR (Autoregressive) model and MA (Moving Average) model.AR model can capture the trend of data with historical trend; MA model can deal with sudden changes or noisy time series data; ARIMA model combines the advantages of the two and can deal with complex time series problems. For the requirement of the question, it is necessary to analyze the losses caused by natural disasters in the past years, and the ARIMA model can fulfill this requirement.The ARIMA model contains `p` autoregressive terms,` q` moving average terms, and the order of the difference` d`. The model can be expressed as:

$$\left(1-\sum_{i=1}^{p}\varphi_i K^i\right)(1-K)^d Y_t=\left(1+\sum_{i=1}^{q}\theta_i K^i\right)\epsilon_t \tag{1}$$

p, d, and q determine the model type, as shown in Figure 2:

Figure 2. ARIMA Model Criteria

The value of d is determined by ACF (autocorrelation function) and PACF (partial autocorrelation function). Where ACF and PACF take values in the range of [-1,1], the absolute value of the number indicates the magnitude of the correlation, and positive or negative indicates positive or negative correlation.

The following are the three cases of values and the corresponding representations:

(1) If the ACF rapidly decreases to 0 at a certain point, it indicates that the time series data is caused by a short-term series

(2) If the ACF slowly decays or fluctuates, it indicates that there may be a more complex correlation structure in the data

(3) If the data is smooth, it indicates the need for differential processing

For any time, series, when the ACF image presents a trailing and the PACF image presents a truncated state, the AR model is applied to the current time series, and the lag order of the PACF truncation is the ideal value of the hyperparameter p. For any time, series, when the PACF image presents a trailing and the ACF image presents a truncated state, the MA model is applied to the current time series, and the lag order of the ACF truncation is the ideal value of the hyperparameter q.

After fitting the ARMA model, we obtained a model for predicting future values.

2.2. Model Establishment

We used 'Pandas' library to extract the CSV file (US annual deaths due to disasters and annual losses due to disasters) and then cleaned the extracted data.

Immediately after that we used 'matplotlib', 'seaborn' library and 'WPS' to create graphs showing the number of events, total cost and number of deaths for different disaster types and showing the total cost and deaths by year. and fatalities for different types of disasters and also show the trend of total cost and fatalities by year to observe the changes over time.

Finally, we get Figures 3 and 4:

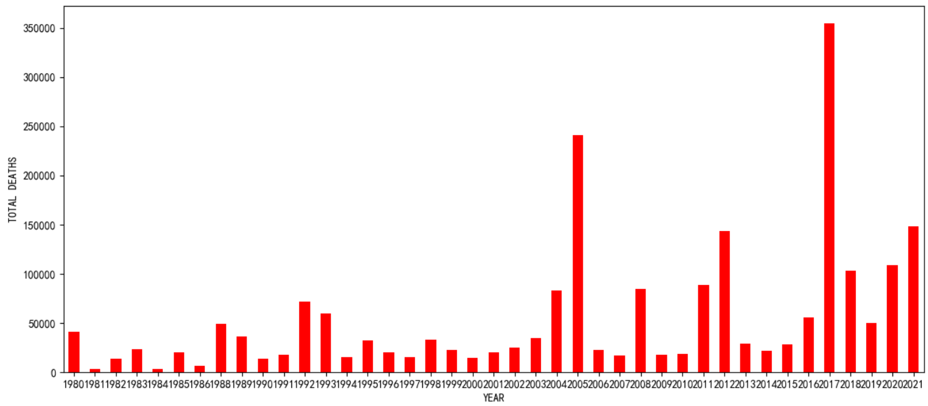

Figure 3. Annual losses from disasters in the United States

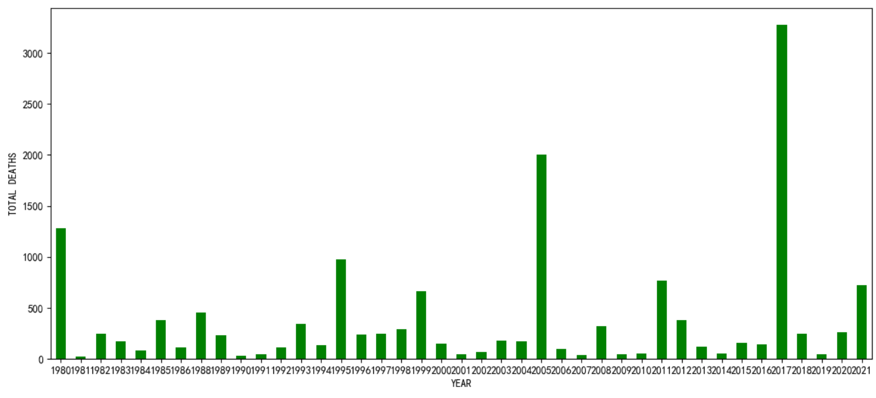

Figure 4. Annual deaths due to disasters in the United States

We then utilized the software 'vscode' to come up with Figure 5:

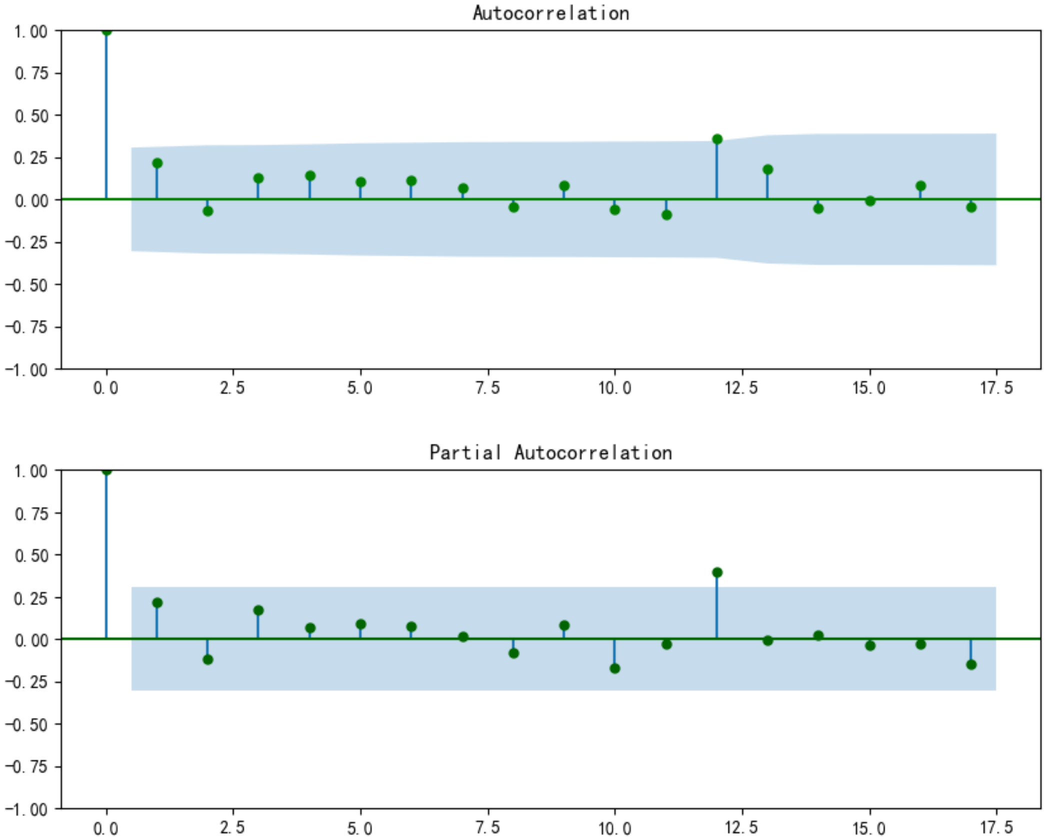

Figure 5. PACF and ACF charts for U.S. disaster year losses

By analyzing the truncated tail of the PACF and the fluctuation of the ACF choose an appropriate ARIMA model to predict the losses caused by future disasters.

For the U.S. data we used the ARIMA (1, d, 1) model for calculation, through which we can predict the value of future losses in the U.S. The prediction results are shown below(Figure 6):

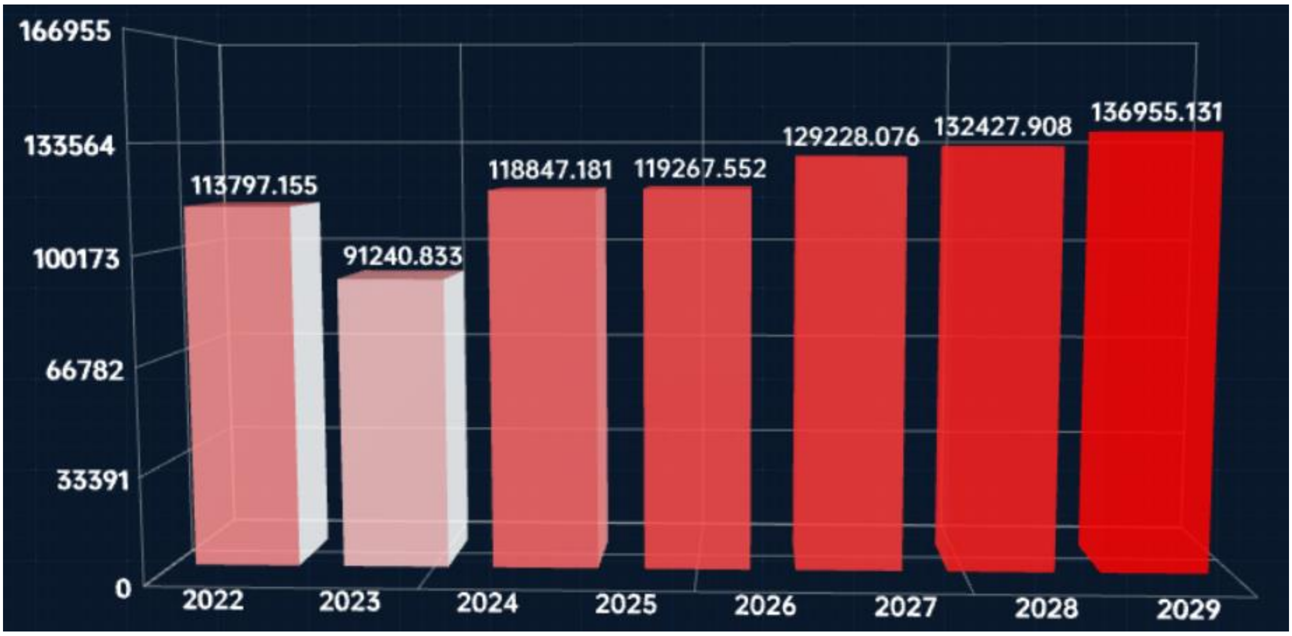

Figure 6. Forecasting the results of projections of future loss values in the United States (Million dollars)

In the same way we can get a forecast of the number of deaths in the United States, and the forecasts go into the Figure 7 below:

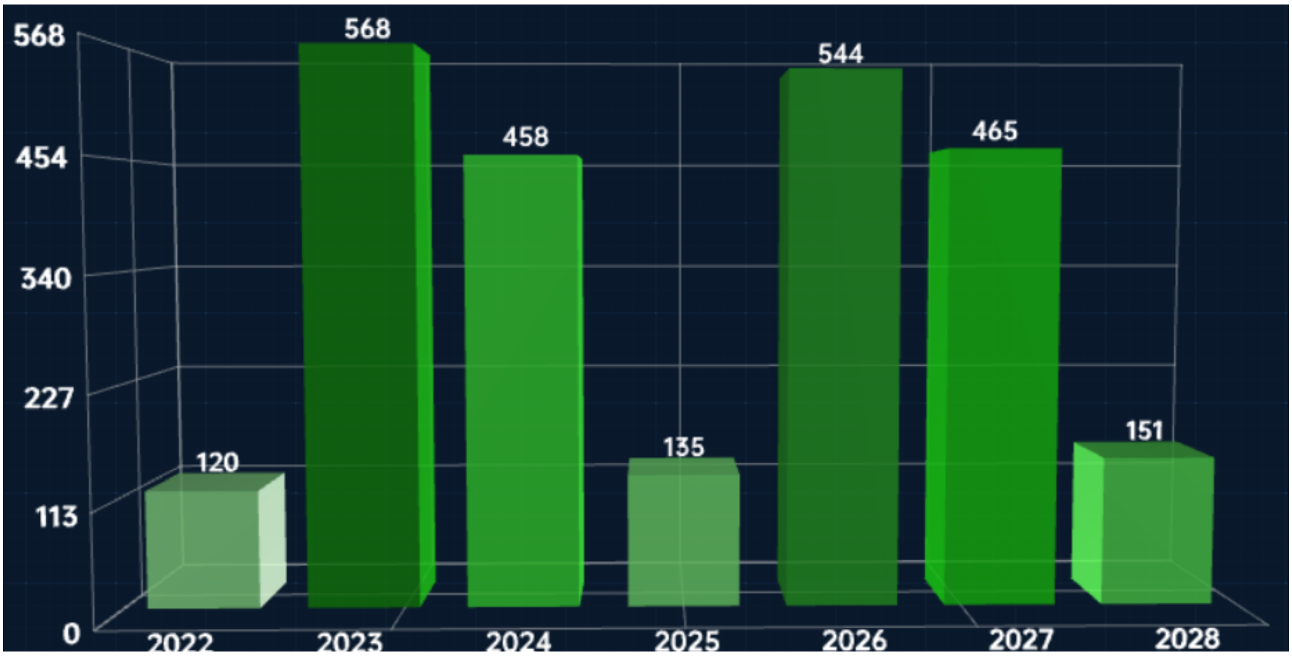

Figure 7. Forecast of the number of deaths in the United States

After fitting, the PACF and ACF of the residuals are calculated to ensure that the model is not missing important information. If the residuals show randomness, the model is a good fit. Conversely, the fit is poor, and another model should be tried.

After calculating and analyzing that data, we chose a model that has the ability to predict the data.

2.3. Model Prediction

We used the model to predict the economic losses in Guangdong and Africa for the next few years. The results are shown below(Figure 8):

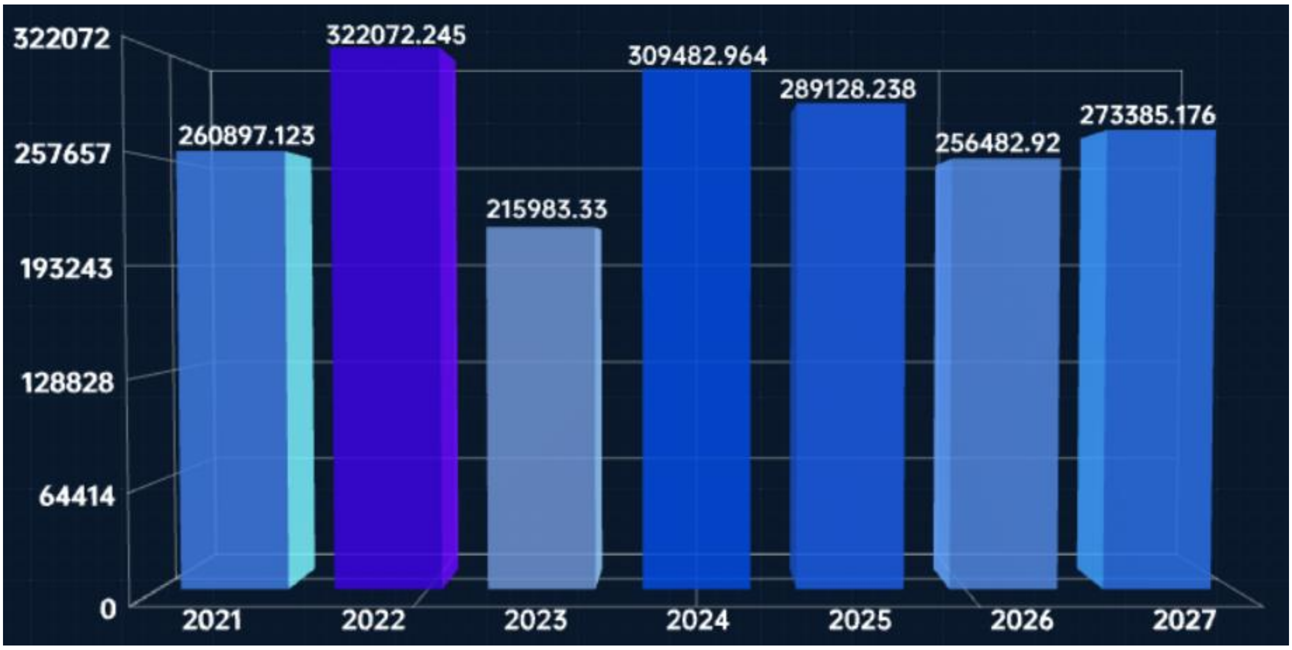

Figure 8. The expected annual loss in Guangdong Province,China(Million dollars)

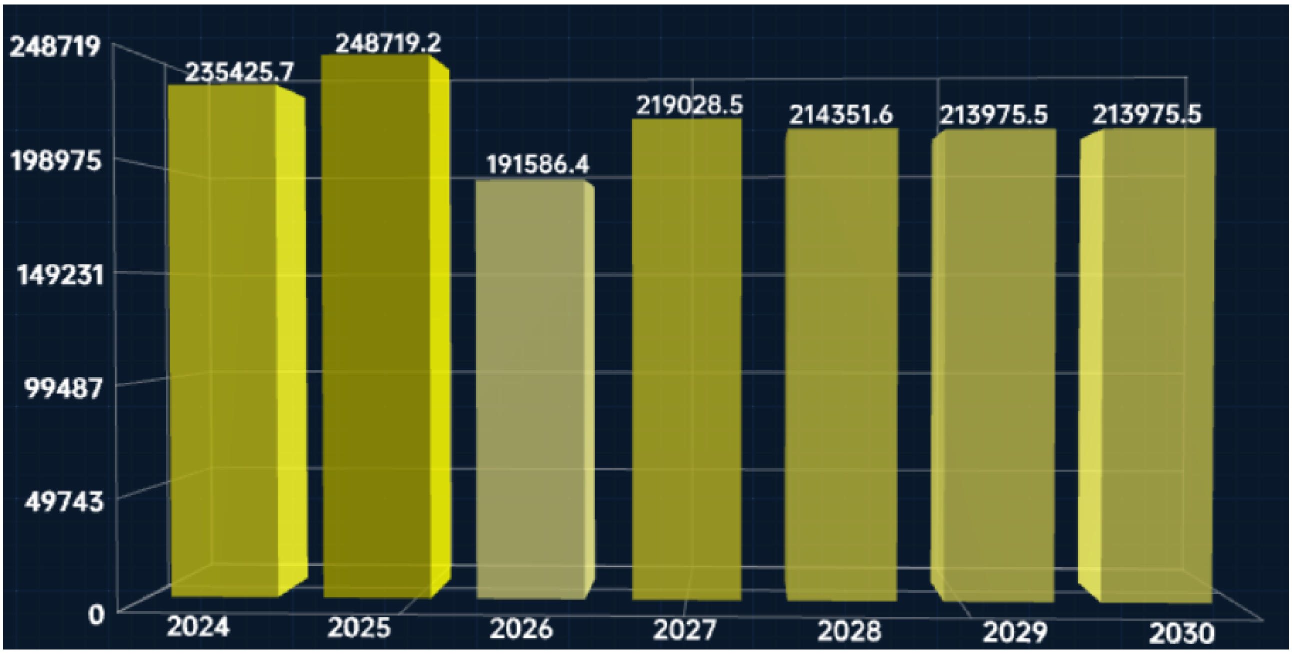

Figure 9. Estimated annual losses for Africa(Million dollars)

2.4. Analysis of the Result And Promotion

Taking into account the expected profit (P), the amount of compensation to be paid per person (N), the cost of each person's specific participation is adjusted to ensure that the company realizes a profit.

The calculation is as follows:

Assuming a specific enrollment cost per person (C), the revenue from total premiums is (M×C)

According to Eq:

$$C = k + \frac{P}{M} + \frac{N}{M} \tag{2}$$

Note: The formula consists of three parts

(1) Base cost component (k): base premium rate paid per enrollee

(2) Expected profit sharing portion ($\frac{P}{M}$): expected profit is divided equally according to the number of participants (M)

(3) Amount to be compensated per person apportionment part ($\frac{N}{M}$): the amount to be compensated per participant (N) is apportioned to each participant

According to the model, it is possible to forecast the insurance company's costs for the next few years, and through the above formula, the insurance company can adjust the premiums accordingly for the corresponding areas. Premiums can be increased when the prediction results in higher loss costs in the current year, or when local losses are higher.

N can be roughly the number of expected losses per year divided by the number of deaths per year (assumption: the number of people affected by the disaster is the same as the number of deaths). When the number of homeowners changes, a change in M causes the insurance company's premium to change based on N and M each year for the same profit (constant P-value). Insurance companies can adjust premiums based on our model predictions.

3. Data visualization analysis of extreme weather events frequency and extent of damage

3.1. Model Preparation

We chose to use data visualization as a first step in processing the data to visualize the values of natural disasters in terms of seasons, regions, and damages.

Data visualization is primarily aimed at conveying and communicating information clearly and effectively through graphical means. Ideas can be communicated effectively, and aesthetic form and function need to go hand in hand to achieve deeper insights into rather sparse and complex data sets by visually communicating key aspects and features.

3.2. Model Establishment

By finding historical weather data on the official website and organizing it, a set of datasets is obtained that can be analyzed. The dataset includes the season, frequency, and economic losses caused by each disaster in a region.

Then we divided the problem into three points:

(1) The relative frequency of a disaster in those areas.

(2) The damage caused by that kind of disaster is relatively greater.

(3) The relationship between the frequency of disasters and the seasons.

We find the number of occurrences of each type of disaster in each state of the United States, using (code, tool), and draw a bar graph, from which we can get the association between regions and disasters.

We analyzed the economic losses caused by each type of disaster in the U.S., (tool, code) by processing the fields in the data about DS-Type(the type of disaster) and LCD(the loss caused by disaster), and drew a line graph, where each time the point is the loss caused by this type of disaster, and accordingly we can analyze which type of disaster caused more losses.

To solve the third problem, we collected the frequency of different disasters in the United States during different seasons and made a line graph, where each line represents a type of disaster, and each point represents the frequency of that disaster during that season.

3.3. Results

(1) Taking typhoons as an example, we get the number of typhoons that have occurred in each state of the United States, as shown in the figure 9 below:

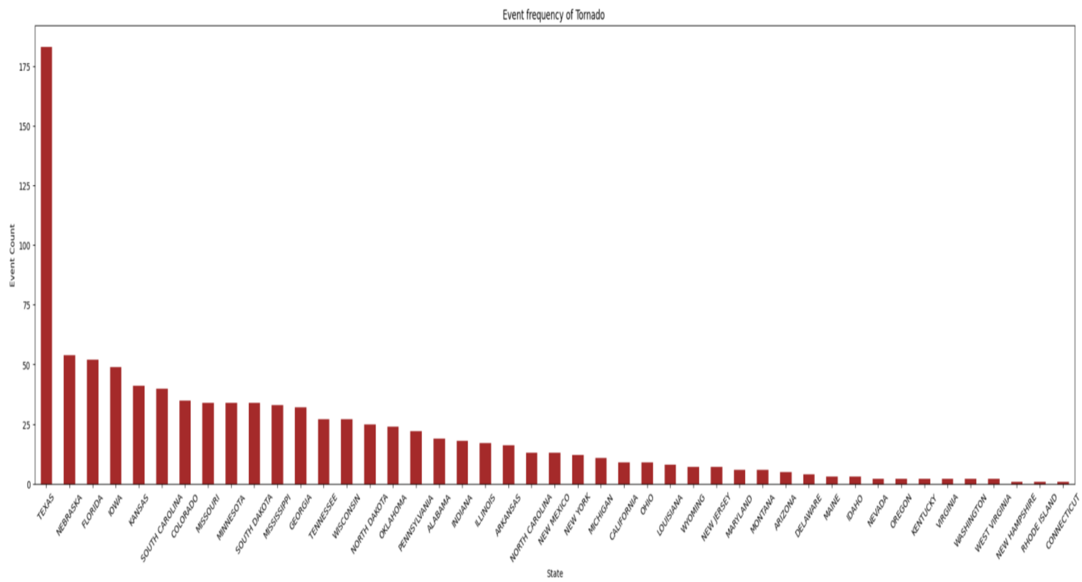

Figure 9. Event frequency of Tornado

Draw a line graph of the damage caused to the United States by different disasters. As shown below (Figure 10):

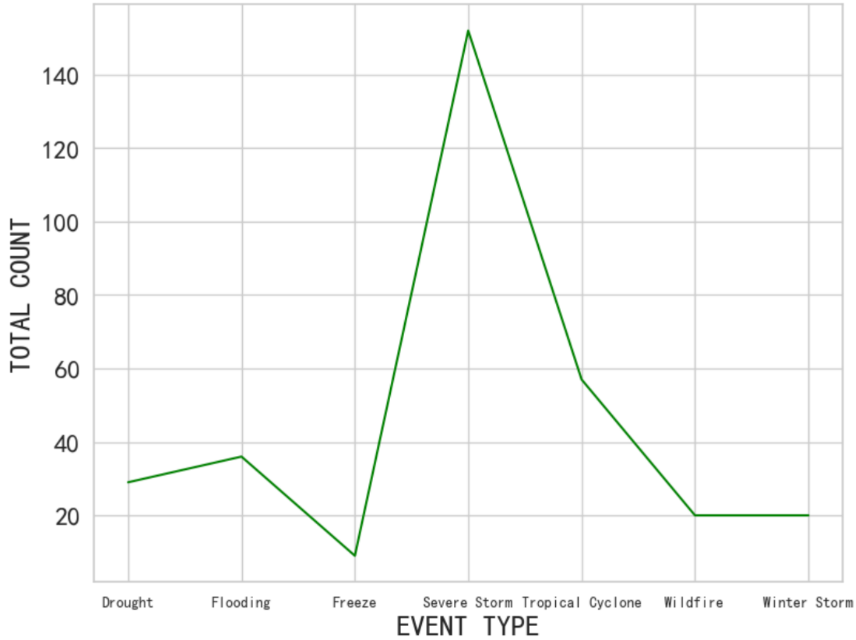

Figure 10. The Damage caused to the United States by different disasters.

Analyze the frequency of different disasters in different seasons, using hurricanes, tsunamis, and hails. As examples in the Figure 11 below:

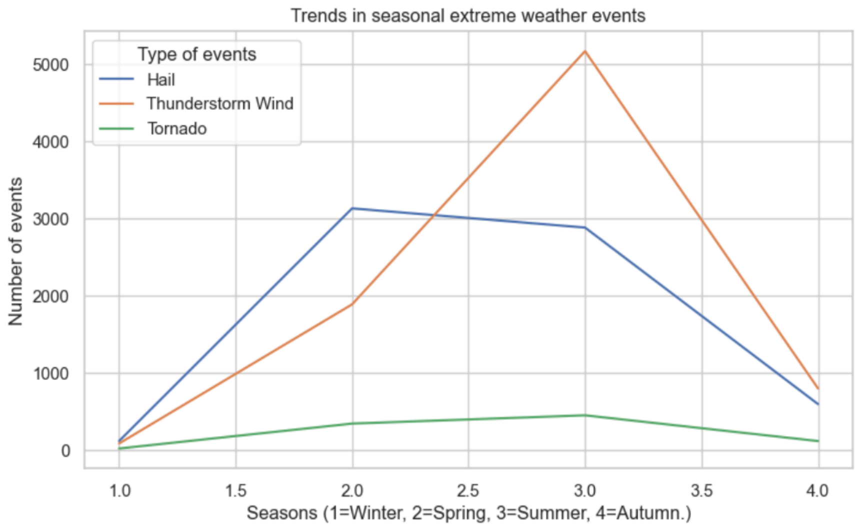

Figure 11. Trends in seasonal extreme weather events

3.4. Analysis of the Result

- According to Figure 9, we can see that texas, in terms of tsunami frequency is much higher than other states, \`connecticut\` has very few tsunamis.
- According to Figure 10, we can see that \`storm\` cause the most damage.
- According to figure 11, we can see that hurricanes are concentrated in summer and hail is concentrated in spring, and tsunamis are less frequent in summer and slightly more frequent in summer.
- Insurance companies can adjust local premiums according to the frequency of local disasters and the value of losses, for example, increase premiums in areas where disasters occur frequently and cause greater losses; insurance companies can also be based on the changes in the time of the

occurrence of the disaster law, according to the different seasons to set different premiums, for example, in the summer of the hurricanes increase the premiums for the type of insurance, the premiums in the spring and winter can be reduced accordingly. Insurance companies can also choose to invest in some disaster prevention and mitigation projects in certain disaster-prone areas, such as water conservancy projects, etc., in order to reduce losses and thus reduce the cost of compensation and thus increase profits.

- Real estate developers can choose to invest in construction in areas where disasters are less frequent according to the frequency of disasters in each area.

4. Conclusions

In this paper, ARIMA model is used to predict future extreme weather losses, and data visualization analysis is combined to provide risk assessment and decision support for insurance companies. The results show that the ARIMA model is effective in predicting future losses, which provides a basis for insurance companies to formulate premium strategies. In addition, the data visualization analysis reveals the frequency and extent of losses across regions, seasons and disaster types, providing insurers with the basis for differentiated premium strategies. This study provides important technical support for insurance companies to make decisions under extreme weather risks. Future studies may further consider combining other models with the ARIMA model to improve prediction accuracy and study the differences between applicable models in different regions.

References

[1] Jianing L, Minglei B, Ziyang L, et al.Spatio-temporal evaluation of electricity price risk considering multiple uncertainties under extreme cold weather[J].Applied Energy,2022,328.

[2] DING Yong, Kong Qingqing, Shu Haibin et al. Research and Development of Natural Disaster Risk Assessment System for Highway Property Insurance [C]// Maintenance and Management Branch of China Highway Society. Proceedings of the 12th Annual Conference of Maintenance and Management Branch of China Highway Society. Ningbo University; Highway and Transportation Management Center, Haishu District, Ningbo City; China Pacific Property Insurance Co., LTD. Ningbo Branch; 2022:7.

[3] P. R L, Muhammad R S, Laila K, et al.Perspectives from the field: Evaluation of a relational model of risk communication in the context of extreme weather[J].Climate Risk Management,2022,37.

[4] Yin L, Wang L, Huang L, et al.Real estate advertising campaigns in the context of natural hazards[J].Disaster Prevention and Management,2019,28(2):183-200.

[5] Sheng L, Yu-Ting K, Kevin L, et al.Applicability of different extreme weather datasets for assessing indoor overheating risks of residential buildings in a subtropical high-density city[J].Building and Environment,2021,194.

[6] Cai M, Xiu W, Wei G. Expected loss utility for natural hazards and its application in pricing property insurance products[J]. Environment Systems and Decisions,2021,41(3):1-15.

[7] Elaina M, Sanjay K, Anthea D, et al.Evidence synthesis - Evaluating risk communication during extreme weather and climate change: a scoping review.[J].Health promotion and chronic disease prevention in Canada : research, policy and practice,2019,39(4):142-156.

[8] Chiara B, Bruno A, Akli B, et al.Wildfire Hazard and Risk Assessment in Pulp Paper Industrial Properties under Extreme Weather Conditions: A Case Study in North-Western Portugal †[J].Environmental Sciences Proceedings,2022,17(1):126-126.

[9] C. Q N, Y. H T N, H. M T V, et al.Assessing the potential risks of extreme weather events causing flood hazards for rice cultivation regions in Quang Nam Province[J].Research on Crops,2022,23(3):481-487.