Property risk assessment and dynamic premium pricing based on ARIMA

Xin Wu [1], Yimeng Zhao [1], Wenjing Song [1], Hao Luo [1, 2, 3, *]

[1] School of Medical Information Engineering, Gansu University of Chinese Medicine, Gansu, China, 730030

[2] School of Electronic and Information Engineering, Lanzhou Jiaotong University, Gansu, China, 730070

[3] Key Laboratory of Media convergence Technology and Communication,Gansu Province, Gansu, China, 730030

* Corresponding Author Email: luoh804@gszy.edu.cn

Abstract. This study aims to improve the efficiency of insurance underwriting decisions and assess community risks, and constructs a dynamic premium pricing model through the ARIMA model. In-depth analysis of historical data provides quantitative and optimal support for insurers to address premium setting and underwriting decision challenges. The results show that the model can predict future risks and losses, facilitate risk classification and premium determination. The classified statistical method analyzes the frequency and time correlation of extreme meteorological events, and provides basis for insurance policy adjustment. Research provides more accurate and dynamic premium pricing and underwriting decision-making tools, simplifies the risk assessment process, and is conducive to insurance market stability and resource optimization, service optimization and long-term market stability.

Keywords: Insurance underwriting, Risk assessment, ARIMA.

1. Introduction

1.1. Background

In recent years, meteorological disasters and climate change have caused immeasurable losses to the world, and with the increase of public risk awareness, the demand for property insurance is also increasing [1]. However, the real estate insurance industry is facing unprecedented challenges. Frequent extreme weather events have led to a significant increase in insurance claims and an increasing number of insurance applicants [2]. The number of applicants in 2022 is 115% higher than the 30-year average, leading to a widening of the global insurance protection gap. As a result, homeowners and insurance companies are also under financial pressure, facing the dilemma of how to maintain sustainable business development and how to solve the problem of profitability in an increasingly risky environment. Communities and real estate developers need to work closely with the insurance industry to develop new insurance models to make them a reality, adapt insurance decisions to this changing insurance environment, ensure they are both flexible and able to meet growing demand, and ensure long-term stable development [3].

1.2. Research purpose

The main purpose of this study is to address the following two core issues:

(1)In the context of climate risk, how insurers assess and adjust the amount of risk they cover to prevent taking too little or too much risk that could affect the company's financial health and long-term development. The aim of this study is to develop a quantitative model to guide insurance companies in deciding which policies are suitable for coverage, how to choose insurance policies with different levels of risk, and to determine the optimal role of insurance companies in the relationship with policyholders. The model is also used to evaluate the effectiveness of insurance policies in the face of frequent extreme weather events in different regions, such as two regions spanning different continents.

(2)In the field of real estate construction, taking into account the development needs to support expanding communities and population growth, the research aims to promote more rational construction planning by adjusting insurance models. This includes conducting an assessment of the location, method and feasibility of construction in a particular area, and the insurance model needs to adapt to these assessments to ensure that appropriate insurance coverage is provided to support sustainable real estate development in the face of different construction challenges.

2. Materials and methods

2.1. Data acquisition and preprocessing

The data description and preprocessing section can be rewritten as follows to improve clarity and fluency:

To analyze and compare the differences in risk of climate disasters between two countries on the European and American continents, we collected comprehensive weather data from the United Kingdom and the United States. By accessing https://ourworldindata.org/ and https://www.statista.com/ and other reliable sources, we get the two countries in different years record details of weather disasters.

For the UK, the dataset covers 323 meteorological disaster records between 1980 and 2021. These records contain six key characteristics: the name of the event, the type of disaster, the start date, the end date, the total cost adjusted for inflation (CPI) in millions of dollars, and the corresponding number of deaths. The US dataset includes 15,624 meteorological disaster records since 1994, which contain 29 different characteristics, including BEGIN_YEARMONTH, BEGIN_DAY, BEGIN_TIME and other information.

In the process of careful cleaning of these data, we did not find any missing or duplicate values, which indicates that the data we collected has a high degree of integrity and reliability, which provides a solid foundation for our analysis. Next, after confirming the quality of the data, we will conduct a deeper analysis of the data to reveal the impact of climate risk on insurance underwriting strategies and the related challenges of real estate construction.

2.2. Method introduction

2.2.1. Risk assessment model

To determine whether an insurance company is able to cover a certain area, and to assume that the insurance content of the insurance company is only financial claims, we only need to know the relative size of the relationship between premiums and claims. Consult relevant literature [4], the premium calculation formula is:

$$N_t = h_t \times \alpha_t \tag{1}$$

Where, N is premium, h is basic premium and α is risk factor. In addition, the basic premium h is calculated using the following formula:

$$h_t = r_t \times f_t \times s_t \tag{2}$$

Where, r is the probability of extreme weather, f is the claim cost, and s is the insurance rate. Since we consider the impact of risk on premiums by the proportion of risk to total assets, the risk coefficients are as follows:

$$\alpha_t = \frac{1}{(1+m_t)} \tag{3}$$

m is the probability of risk to total asset, and the total asset is M . Its calculation formula is as follows:

$$m_t = \frac{r_t \times f_t}{M} \tag{4}$$

If you want to get compensation, you must first know the loss, but in the actual situation, considering that the calculation of the loss is more complicated, the insurance company only makes financial compensation, so the compensation is directly defined as:

$$l_t = \frac{c_t}{n} \tag{5}$$

Where, c is the amount of property damage and n is the number of insured. Finally, the probability p of an extreme weather event is:

$$P = \theta\left(\beta_0 + \beta_1 \times \text{ year } + \beta_2 \times \text{ Disaster } + \beta_3 \times \text{ cost } + \beta_4 \times \text{ death }\right) \tag{6}$$

2.2.2. ARIMA

The ARIMA model is a widely used technique for analyzing and predicting time series data, especially for non-stationary time series with trends and seasonality. The general form of the ARIMA model is denoted as ARIMA (p, d, q), where p, d, q are the key parameters of the model, representing the autoregressive term, the difference number and the moving average term respectively [4].

The steps of establishing ARIMA (p, d, and q) model are roughly as follows. For a given time series Y_t, the autoregressive (AR) part can be expressed as [5, 6]:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \tag{7}$$

Where c is the constant, $\phi_1, \ldots, \phi_p$ is the coefficient of the AR term, $Y_{t-1}, \ldots, Y_{t-p}$ is the value of the first p time points, and ε_t is the error term.

The moving average (MA) part can be expressed as:

$$\varepsilon_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \omega_t \tag{8}$$

Where, $\theta_0, \ldots, \theta_q$ is the coefficient of the MA term, $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ is the error term of the first q time points, and ω_t is a purely random shock.

If the time series Y_t requires d-difference to be stationary, then we define $Y_t^{'}$ as:

$$Y_t^{'} = (1 - B)^d Y_t \tag{9}$$

Here, B is the shift operator such that:

$$B^k Y_t = Y_t - k \tag{10}$$

The ARIMA (p, d, and q) model is obtained by synthesizing AR and MA parts and including differences:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1 - B)^d Y_t = c + (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)\varepsilon_t \tag{11}$$

3. Model building and solving

3.1. Research on insurance model to cope with extreme weather

For the United Kingdom and the United States, we chose different time grainings to show the effect of the model: for the United Kingdom, we calculated with year grainings; In the United States, the monthly granularity is used. We will take the British region as an example. First, we need to assess the risks. Here, we assume that extreme weather occurs when property damage and deaths exceed the average. Using the logistic regression model established above, extreme weather is judged according to disaster data, and its prediction probability is obtained. At the same time, under this setting, the predicted probability value greater than or equal to 0.8 is the high risk level, less than 0.5 is the low risk level, and the rest is the medium risk level. The risk assessment results are as follows, only some of which are shown in Table 1:

Table 1. Predicted probability value and Risk assessment.

Britain		America	
Predicted probability value	Risk assessment	Predicted probability value	Risk assessment
0.00	low risk	1.00	high risk
0.00	low risk	0.00	low risk
1.00	high risk	1.00	high risk
. . .	. . .	. . .	. . .
1.00	high risk	0.00	low risk
0.00	low risk	0.00	low risk
0.99	high risk	0.00	low risk

Based on UK and US loss statistics based on different time granularity, the ARIMA model is used to predict losses for the next seven cycles. First, the adf is used to check whether the time series is stable. Taking the UK as an example, the following results are obtained: the test statistic is -4.759697, the p value is 0.000065, and the min aci is (2, 1, 3), indicating that the time series is stable. Autocorrelation and partial autocorrelation are shown in Figure 1:

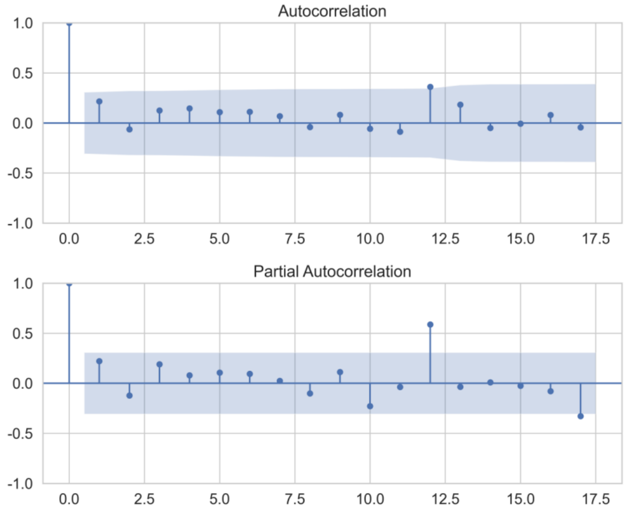

Figure 1. Autocorrelation partial autocorrelation graph (UK cost).

Similarly, we have also made risk projections for the next seven cycles. According to the risk assessment threshold set before, it is divided into different risk levels. As shown in Table 2:

Table 2. Risk assessment for the next seven cycles.

	Britain			America		
	Predict risk	Predict cost	Risk level	Predict risk	Predict cost	Risk level
1	0.05	4535521.00	low risk	0.34	113797.16	low risk
2	0.05	4615112.00	low risk	0.17	91240.83	low risk
3	0.05	5323129.00	low risk	0.11	118847.18	low risk
4	0.49	4867994.00	low risk	0.25	119267.55	low risk
5	0.05	5130364.00	low risk	0.24	129228.08	low risk
6	0.05	5043440.00	low risk	0.14	132427.91	low risk
7	0.05	5023884.00	low risk	0.19	136955.13	low risk

Assume that the total assets of the insurance company is M million, and the premium rate s of the company is 20% for high risk, 10% for medium risk, and 5% for low risk areas. The number of participants is 120. Calculate the premiums and claims for the next seven cycles according to formula (1.1) - (1.5), as shown in the table 3:

Table 3. Premiums and claims for the next seven cycles.

	America		Britain	
	Premium	Compensate	Premium	Compensate
1	3532.852	37796.01	1380.67	948.39
2	3577.542	38459.27	676.09	760.34
3	3636.527	44359.41	564.06	990.39
4	3526.315	40566.62	1151.63	993.89
5	3597.416	42753.03	1189.19	1076.90
6	3576.066	42028.67	785.76	1103.56
7	3575.246	41865.70	1018.97	1141.29

When the insurance premium is greater than or equal to the amount paid, the insurance company can carry out insurance, but when the insurance premium is lower than the amount paid, the insurance company's underwriting risk is greater.

3.2. The adaptability adjustment of insurance model to real estate development

Depending on where, how and how feasible it is to build a home in a particular location, we need to use weather event data to adjust the insurance model. It is mainly to analyze the risks in various areas and avoid construction in areas with higher risks [7].

3.2.1. Adaptive adjustment analysis

Before the adaptive adjustment of the model, it is necessary to analyze the meteorological conditions and get the idea and direction of the model adjustment. We will analyze the region of the disaster, the damage caused, the intensity and frequency of the disaster [8], so as to achieve the purpose of adaptive adjustment of the model.

(1)Analysis of disaster areas

First of all, from the statistical analysis of historical disaster events [9], the United States recorded 7,945 thunderstorms, 6,733 hail events and 947 tornadoes in 1994. It is clear from these data that thunderstorms occur with the highest frequency, while tornadoes are relatively the least frequent natural disasters. In addition, by integrating and analyzing the number of natural disasters occurring by state, we can identify which areas are at higher risk. Figure 2 clearly shows that Texas and Oklahoma are states that have been particularly affected by disasters, designating them as high-risk areas. Based on this information, it is recommended to avoid selecting these high-risk areas during the construction planning process. Conversely, areas with a low frequency of disasters, such as Washington state and Delaware, may be ideal sites for construction. Choosing these low-risk areas for construction can reduce the potential risks caused by natural disasters.

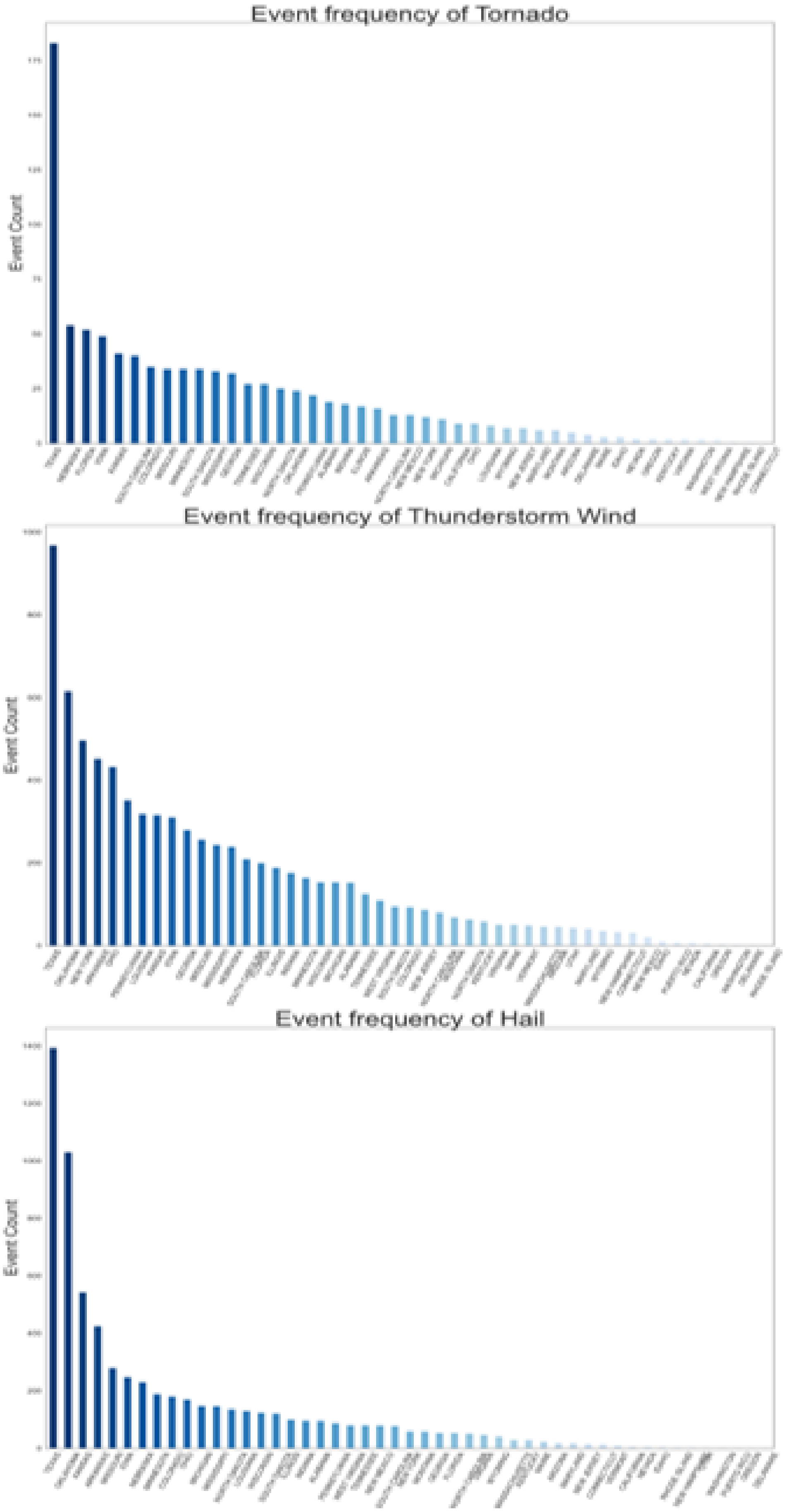

Figure 2. Event count.

(2)Damage assessment

Then, on the basis of understanding the frequency of different disasters and areas with high incidence of disasters, the loss caused by disasters is analyzed to provide a better reference for construction site selection and insurance rate adjustment. Through the analysis of two oil fields, DAMAGE_PROPERTY and DAMAGE_CROPS, the average loss of property and crops caused by different disasters was obtained [10], and the threat of different types of events to assets was understood, as shown in Figure 3. In the property damage analysis, it is found that the damage caused by tornadoes far exceeds the damage caused by thunderstorms and hail; in the crop loss analysis, tornadoes and hail caused more damage than thunderstorms. But overall, tornadoes cause the most damage and pose the greatest threat to assets.

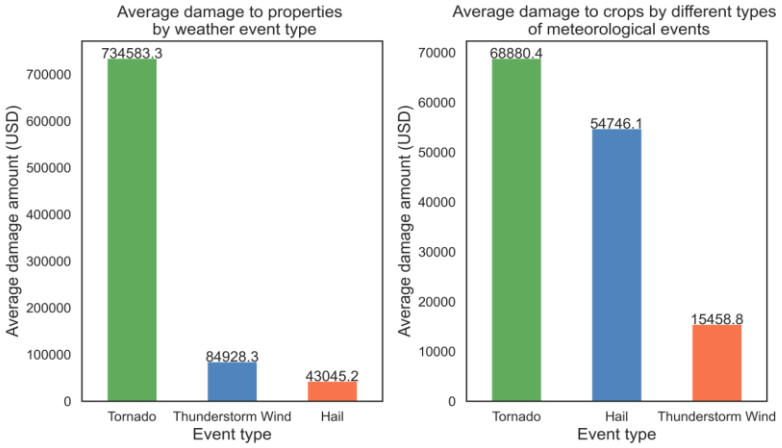

Figure 3. Average damage to properties or crops.

The longitude and latitude of property losses are shown in Figure 4:

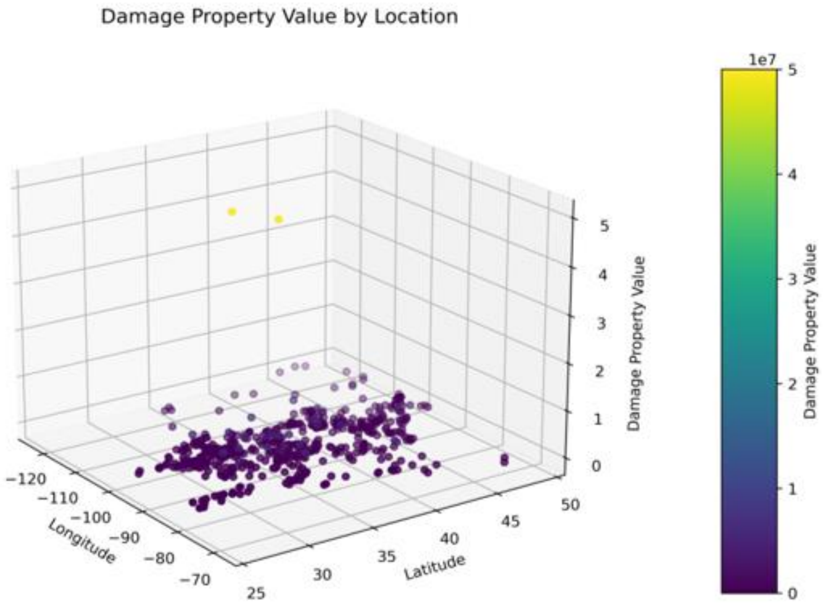

Figure 4. Property damage latitude and longitude.

After understanding the high incidence of disasters and the distribution of losses, the intensity of the disaster is analyzed in order to make a comprehensive assessment of the risk posed by the disaster. Through the analysis of TOR_F_SCALE, earthquake magnitude and other fields, it is found that the frequency of mild tornadoes is much higher than other tornadoes. Other disasters typically occur between 0 and 10 degrees in intensity. As can be seen from Figure 6, although the damage caused by tornadoes is the largest, it is likely that the damage caused by a disaster is not large, because the main influencing factor is mild tornadoes, and the damage caused by tornadoes is mainly a high number of occurrences, rather than a severe damage.

(3)Analysis of disaster high occurrence period

Finally, we do statistics on when different disasters occur to see if disasters have seasonal effects in order to better adjust and build. As can be seen from Figure 5 and Figure 6, thunderstorms in June and July, hail in April, May and June, and tornadoes in April and July. To sum up, the peak of incidents occurred mainly in June.

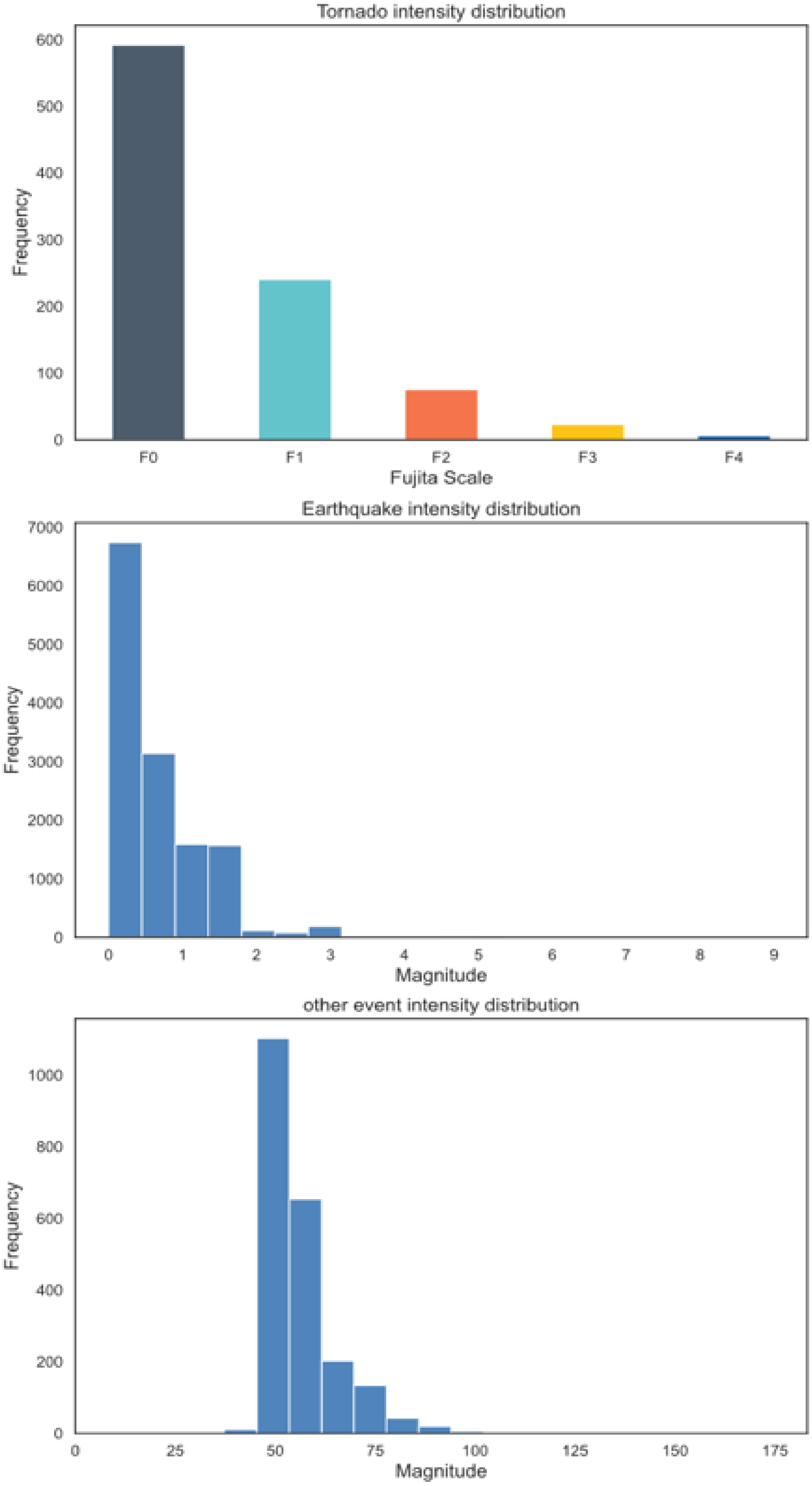

Figure 5. Intensity distribution.

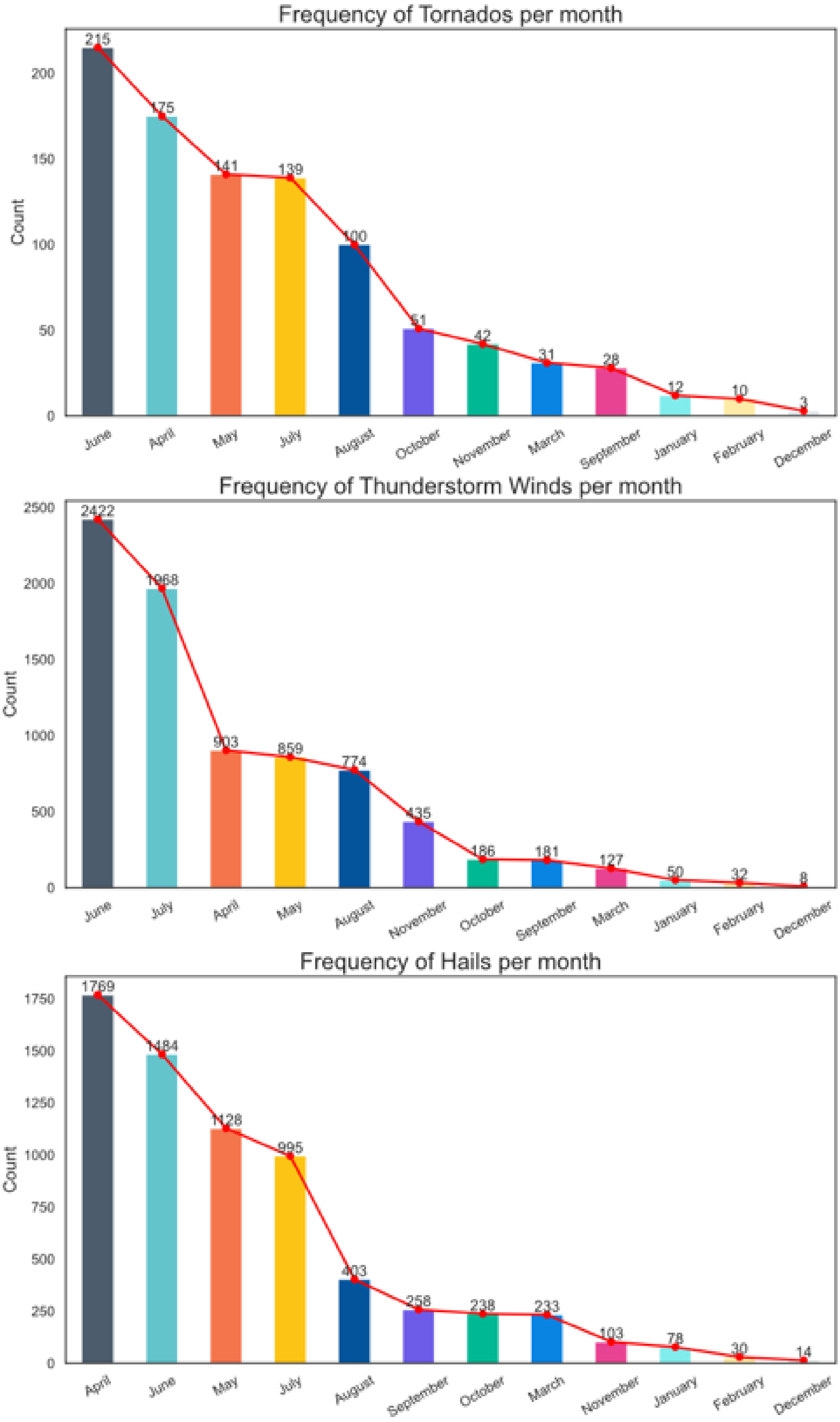

Figure 6. Frequency of per month.

4. Conclusions

Combining the findings of the two studies, we can conclude that to limit exposure, insurers should set premiums at least equal to expected payouts. Disaster analysis shows that although tornadoes do not occur frequently, they cause more damage to property and crops than any other natural disaster, especially in disaster-prone states such as Texas and Oklahoma, which should be avoided as construction sites. In contrast, less frequent areas such as Washington state and Delaware are more suitable for construction. The study further reveals the seasonal trend of these events, particularly peaking in June, with key implications for the seasonal adjustment of construction scheduling and premium policies. Since minor tornadoes are rare but have the highest cumulative costs, this suggests that we should consider both severity and frequency of events when developing risk management strategies. Therefore, insurance companies and the construction industry should conduct detailed risk assessment and careful resource allocation based on these conclusions when evaluating future project sites and adjusting insurance rates.

Acknowledge

This work is supported by the Natural Science Foundation for Young Scientists of Gansu Province(Grant No. 22JR5RA595), the Youth Project of Social Science Planning of Gansu Province(Grant No. 2021QN027), and the Scientific Research and Innovation Fund Project of Gansu University of Chinese Medicine(Grant No. 2021KCYB-10).

References

[1] Türkeş M, Deniz A Z. Climatological/meteorological and hydrological disasters and the insurance sector [J]. International Journal of Human Sciences, 2010, 7 (2): 996.

[2] Samuel R, J P R, W J W B. Insights into the complementarity of natural disaster insurance purchases and risk reduction behavior. [J]. Risk analysis: an official publication of the Society for Risk Analysis, 2023, 44 (1): 141-154.

[3] Zhang Shuai, Abdulwali Aibai The impact of extreme weather on regional financial risks [J]. Finance and Economics, 2024, (01): 30-40.

[4] HUDSON, P., DE RUIG, L. T., DE RUITER, M. C., et al. An assessment of best practices of extreme weather insurance and directions for a more resilient society [J]. Environmental Hazards. 2020, 19(3):301-321.

[5] Bai Tong, Zhao Yu, Cheng Boyu. A Brief Discussion on Time Series Analysis - Taking ARIMA as an Example [J] Science and Information Technology, 2021, (16):173/175.

[6] Siyu C, Qiang Z, Bin W, 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(15):111266.

[7] Liu Weili, Wang Chao, Wang Qiongqiong, etc Green evaluation system for the entire life cycle of building structures in extreme climate zones [J] Southern Architecture, 2023 (1): 19-26.

[8] Takamune K. Archaeology for Disaster Management [J]. Inter Faculty, 2022, 11:269-276.

[9] T. K S, Junia H, R. J E. Disasters, local organizations, and poverty in the USA, 1998 to 2015 [J]. Population and Environment, 2018, 40 (2): 115-135.

[10] [10] Seamon E ,Gessler E P ,Abatzoglou T J , et al. Climatic Damage Cause Variations of Agricultural Insurance Loss for the Pacific Northwest Region of the United States [J]. Agriculture, 2023, 13 (12):2214.