Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

A Study of Property Insurance Sustainability Based on ARIMA and CGDAM-WRIR Models

Jiaxuan Zhang [1, *], Jiuji Rong [2]

[1] School of Management, Shenyang University of Architecture, Shenyang, China

[2] School of Information Engineering, Shenyang Chemical University, Shenyang, China

* Corresponding author: 1204711479@qq.com

Abstract. The property insurance industry faces significant challenges against the backdrop of the high incidence of extreme weather events around the world. This study focuses on insurers writing policies in high-risk areas and explores how communities and real estate developers are adapting their insurance models to enhance property resilience. We provide an in-depth analysis of the impact of climate change on the sustainability of the insurance industry and use ARIMA and CGDAM-WRIR models to predict the potential losses and impacts of future extreme weather events. Finally, polynomial fitting is used to analyze the effects of earthquakes on buildings, the economy, and human life. Findings suggest that increased demand for insurance is causing insurers to re-evaluate underwriting strategies, especially in high-risk areas. Community engagement and real estate development strategies are critical in reducing costs and expanding coverage, and governments, insurers and communities are called upon to work together to effectively address climate change challenges.

Keywords: ARIMA; CGDAM-WRIR; polynomial fit.

1. Introduction

With the frequency of extreme weather events around the world, the property insurance industry is facing unprecedented challenges [1]. Not only do these events have a huge impact on individuals and communities, but they also pose a serious test for the sustainability of insurance companies. With over 1 trillion $ in losses and a dramatic increase in insurance claims, the impact of climate change on the insurance industry is becoming more pronounced [2]. This paper looks at the issue of insurers writing policies in high-risk areas and how communities and real estate developers are adapting their insurance models to enhance property resilience. By ARIMA and CGDAM-WRIR models, we provide an in-depth analysis of the impact of climate change on the insurance industry and propose strategies to deal with it. Polynomial fitting is also used to analyze the impact of earthquakes on buildings, economy and human life, as well as to predict the losses caused by earthquakes. This study aims to provide new perspectives and recommendations for the sustainability of the insurance industry and calls for concerted efforts to address the challenges of climate change.

2. ARIMA Forecasting Model

2.1. Forecasting Future Losses

Considering that the time series data of extreme weather events have obvious trends and patterns, we chose the ARIMA model for in-depth analysis. The ARIMA model performs well in dealing with time series data and is particularly suitable for non-seasonal data that exhibit obvious trends or patterns [3]. The model works by combining past values and error values of time series data to predict future values.

The ARIMA model, known as the Autoregressive Integrated Moving Average Model, is designed to characterize the autocorrelation of one-dimensional time-series data and predict future points. It combines the strengths of the following three main components [4].

Autoregression (AR): - The central idea of this component is that the observations at the current point in time can be linearly predicted by the past values of the series. The order p of the autoregressive component indicates the number of previous p historical values used to predict the current value.

Differentiation (I): - The purpose of differentiation is to smooth a time series, i.e., to convert a non-stationary time series into a stationary one. The difference operation is usually expressed as $(1-L)^d$, where L is the lag operator and d is the number of times the difference is performed to ensure the smoothness of the data.

Moving Average (MA): The moving average component involves a linear combination of the error terms of the time series. The order q indicates the number of the first q prediction errors used to predict the current value.

The basic equation of the ARIMA model can be expressed as follows:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d y_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\epsilon_t \tag{1}$$

Here, $L^k y_t = y_{t-k}$,which means that the L lag operator acting k times on y_t will give values prior to period k. Each component in the model has a specific role: ϕ_i are autoregressive coefficients, which are multiplied by p prior values of the time series. θ_i are the moving average coefficients, which are multiplied by the q prior error terms of the model. d is the number of differencing that the model needs to perform to ensure the smoothness of the time series.

To determine the parameters p, d and q of the model, we need to rely on the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series. These functions help us visualize the correlation between data points and their independent correlation with other data points.

Autoregressive Order p: can be determined from the PACF plot and is usually the last significantly non-zero lag in the PACF plot.

Difference-in-differences order d: can be determined from the ADF test, as well as observing significant changes in the smoothness and seasonality of the data.

Moving Average Order q: Determined from the ACF plot, usually the last significantly non-zero lag in the ACF plot.

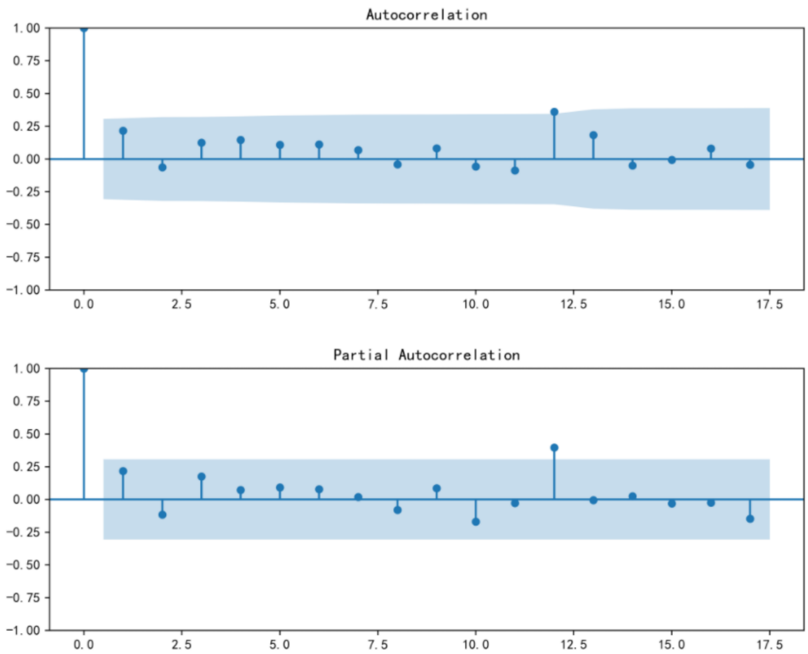

Figure 1. ACF and PACF

According to Fig. 1, we choose an ARIMA model to predict future disaster impacts. the PACF truncates after the first order, while the ACF exhibits decreasing fluctuations. we choose an ARIMA (1, d, 0) or ARIMA (1, d, 1) model, where d is determined by the smoothness of the data.

After fitting the ARIMA model, we obtain a model that predicts future values. The model is utilized to make predictions of total annual losses for the next seven years, and these predictions will help insurers understand the magnitude of potential future losses and adjust their strategies accordingly. The validity of the model needs to be assessed by residual analysis, confidence intervals for the predicted values and the fit to the historical data.

2.2. Forecasting Future Deaths

Impact of Disaster Events on Lives: This dataset demonstrates the significant impact of disaster events on lives, which is critical for emergency planning, resource allocation and risk management in the insurance industry.

Predicting Future Risks: By analyzing and predicting the time series of deaths, relevant agencies can better prepare for possible future extreme weather events and plan for disaster relief and rescue.

After fitting the ARIMA model, the results shown in Fig. 2 were obtained. The fluctuations and spikes in the data in Fig. 2 may indicate that the standard ARIMA model may not be sufficient to capture all the data characteristics, and that outlier handling or the use of more sophisticated models, such as outlier detection or decomposition methods for time series, may also need to be considered. In addition, residual analysis of the model will help us assess whether the model is capturing all the relevant information in the data or if there is still information that is not explained by the model. If the residuals show a clear pattern or autocorrelation, further model adjustments may be needed.

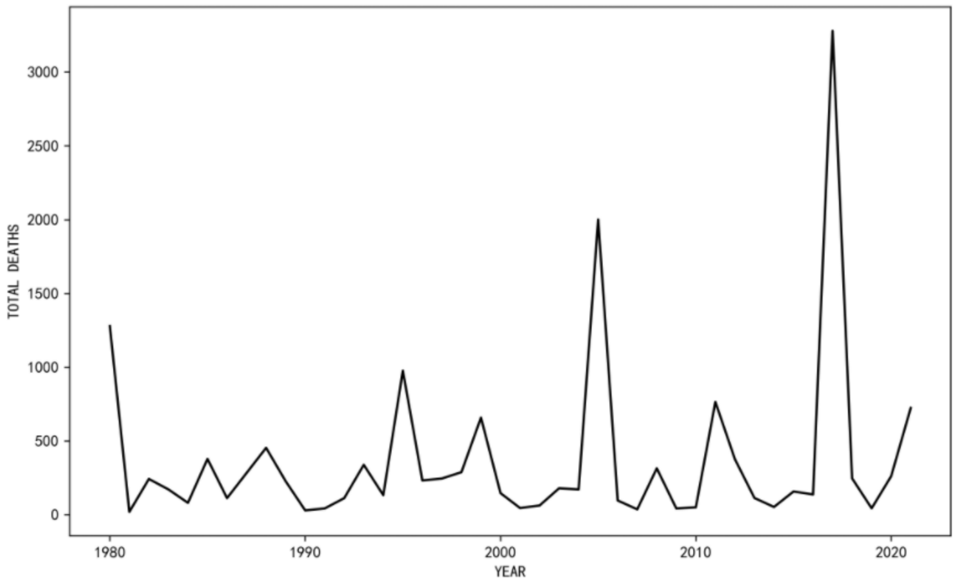

Figure 2. Forecasted results

The model's prediction results will help us understand the magnitude of potential future risks and provide decision support for governments and insurance companies. The confidence intervals of the prediction results will show the range of uncertainty in future predictions.

3. CGDAM-WRIR Model

3.1. Basic Framework of the Model

Given the global increase in extreme weather events and the potential impact of these events on human society and the economy, it is critical to choose an analytical approach that reveals regional risk levels. This simple descriptive statistical model provides policymakers and insurers with a visual map of risk by counting and visualizing the number of tornado events in each state. This approach directly reflects the frequency of past events and is the basis for predicting future risk and developing response strategies.

In our "Comprehensive Geospatial and Damage Assessment Model for Weather-Related Insurance Risks" (CGDAM-WRIR), we focus on the STATE and EVENT_TYPE fields in the statistics set [5]. The model generates a bar graph by grouping and counting these fields, as shown in Fig. 3, where each bar represents a state, and the height of the bar indicates the number of tornado events recorded in that state. This visualization allowed us to quickly identify the states that experienced the most tornado events. For example, in the graph generated by the CGDAM-WRIR model, if Texas has the tallest bar, this clearly indicates that Texas suffered the most frequent tornado events during the observation period.

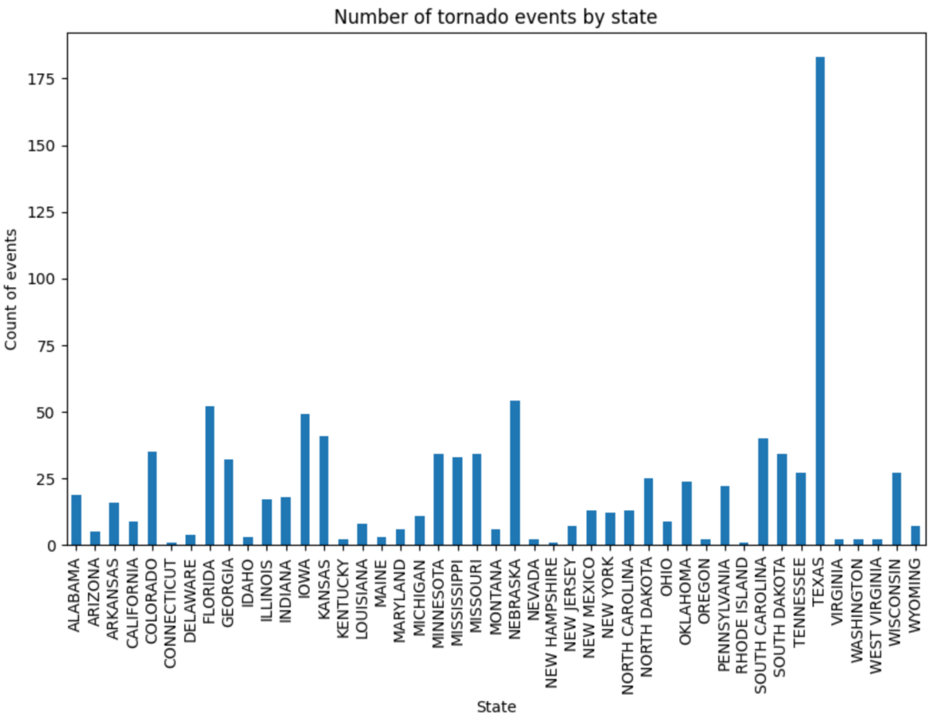

Figure 3. Number of tornado events

3.2. Parameter Estimation and Calibration

In the CGDAM-WRIR model, we are concerned with the accuracy and real-time availability of property loss data. The data preprocessing step ensures that loss values can be correctly parsed and converted into a numerical format, which is the basis for any further analysis. By combining the damage data with location (latitude and longitude), we can identify the hardest hit areas and understand the specific risks these areas may face.

Each dot in Fig. 4 represents a recorded weather event, and the color and size of the dot is proportional to the value of property damage. This visualization immediately highlights those areas with high economic losses, possibly due to frequent weather events or the high impact of a single event.

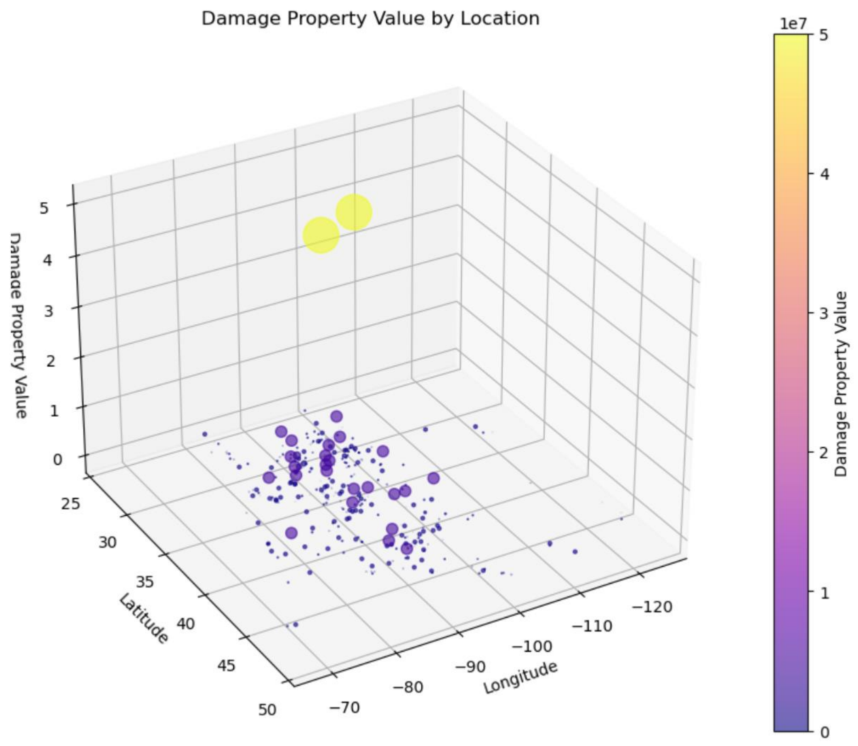

Figure 4. Economic loss value

As we can see in Fig. 4, there are a few locations with much higher-than-average loss values, and these points may represent particularly catastrophic events or areas with high property values. Such information is critical for insurers when deciding how to price insurance and which areas may require higher premiums or more risk mitigation measures.

3.3. Modeling to Assess Insurance Risk

To develop effective community education and response strategies, we need to understand which states are more frequently exposed to specific types of weather events and the average damage these events cause to properties and crops. This graphic provides that understanding by summarizing risk scores from historical data. Risk scores may be based on a variety of factors, including historical event frequency, intensity (e.g., Fujita scale for tornadoes), and event damage to properties and crops.

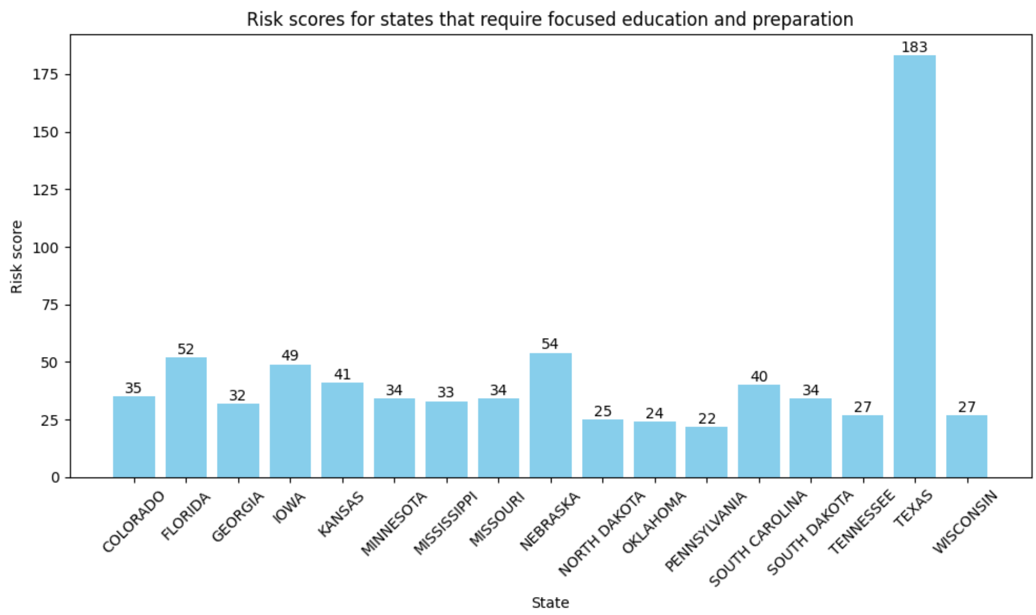

Figure 5. Risk scoring

By visualizing the risk scores, the model reveals the relative risk levels in different states, indicating which states need to increase education and preparedness. For example, Fig. 5 shows that Texas has a much higher risk score than other states, suggesting that the state may need more resources to increase public awareness of tornado risk and to strengthen preparedness measures for tornadoes.

3.4. Extent of Regional Impact of Tornadoes of Different Intensities

In weather disaster management, it is critical to understand the extent of regional impacts from tornadoes of different intensities. This information has significant implications for risk assessment, resource allocation, emergency response planning, and long-term infrastructure planning. In addition, insurance companies need this type of data when designing insurance products and assessing risk.

In this model, we focus on the Fujita scale of tornadoes-a measure of tornado intensity. By calculating the number of events in each class and converting it to a percentage, we can get a clear picture of which class is most common. This helps us identify which types of tornado damage are most prevalent and has practical implications for preparedness and response strategies.

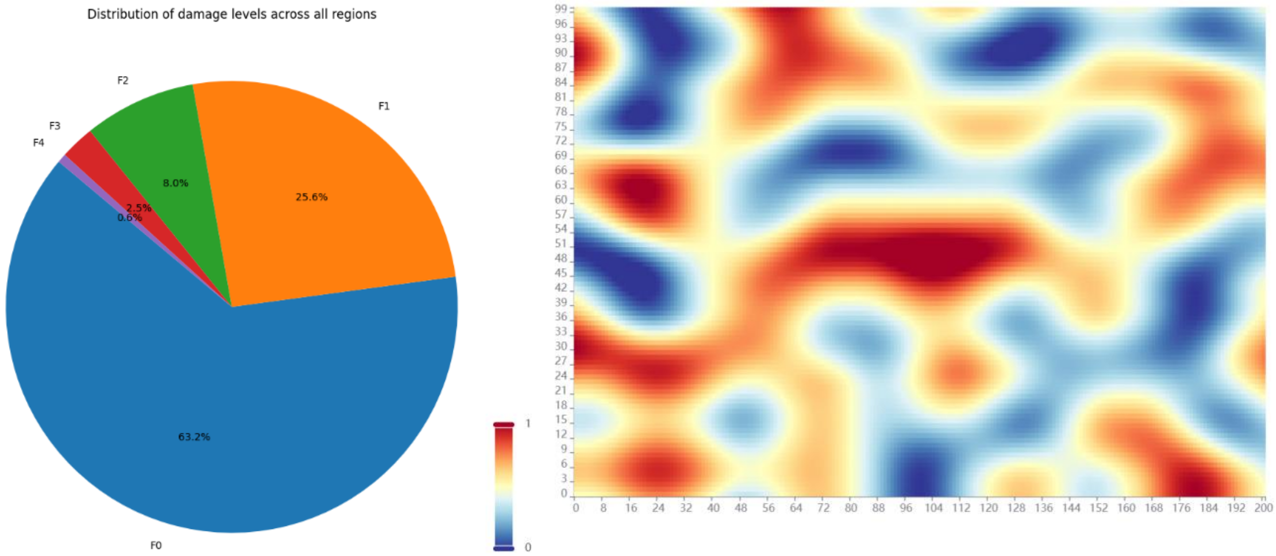

Figure 6. Damage level distribution

Fig. 6 shows the distribution of tornado damage classes in each region, where the damage classes are categorized according to the Fujita Scale (F-SCALE) from F0 (weakest) to F4 (strongest). The results show that most of the recorded events belong to lower class tornadoes (e.g., F0), while higher class tornadoes (e.g., F4 and F5) are rarer. This information can be used to optimize resource allocation, for example, there may be a need to increase building intensity or improve early warning systems in areas that are frequently struck by high magnitude tornadoes.

4. Underwriting Decision Model

4.1. Basic Framework of the Model

To safeguard the long-term stability and growth of a community, a risk assessment of the areas and buildings most affected by climate impacts is needed. By quantifying the level of risk for each building, community leaders and policy makers can make more informed decisions such as strengthening building structures, adjusting land use planning, optimizing emergency preparedness measures, setting appropriate insurance rates, and prioritizing investments.

In developing this model, the following key factors were considered:

Probability of Damage (P): this is the likelihood that a building will be exposed to an extreme weather event at a given time. It is usually calculated based on past event data and climate model projections.

Degree of Potential Damage (D): this is the degree of damage the building could suffer if the event occurs. It depends on the building's structural strength, design, location, and surroundings.

Building Importance (I): this reflects the cultural, historical, economic or social value of the building to the community. The higher the importance score, the more weight a building will have in the risk assessment model.

The mathematical model assumes that we want to assess the risk of building i. We can define the following parameters: P_i: the probability that building i will be affected by extreme weather. d_i: the expected level of damage if building i is affected. I_i: the importance score of building i, which may be based on cultural, historical, economic, and/or community importance. We can then define the risk score R_i for building i as follows:

$$R_i = P_i \times D_i \times I_i \tag{2}$$

This score can be used to rank buildings and determine which buildings need more urgent conservation measures.

Data preparation phase: first, we create a hypothetical dataset containing information about three buildings. This information includes the building's probability of damage (P_i), degree of potential damage (D_i), and importance score (I_i). These data are usually obtained based on historical data, building assessments, and community research.

Risk score calculation: next, we calculate the risk score for each building according to the previously defined risk score formula ($R_i = P_i v \times D_i \times I_i$). This score will be used to quantify the building's risk level under extreme weather events.

Sorting and analyzing the results: finally, we sort the buildings based on the calculated risk score to determine which buildings require more urgent protection measures. This helps decision makers prioritize the highest risk buildings and rationalize the allocation of resources.

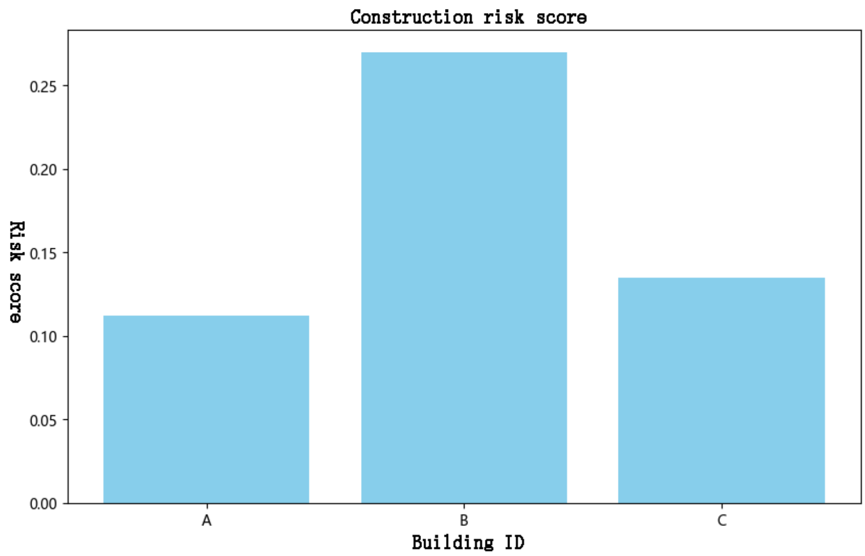

Figure 7. Construction risk score

As shown in Fig. 7, building B has the highest risk: according to the calculations, building B has the highest risk score (0.27), which means that it faces the highest risk under the current assessment model. This could be due to its high probability of damage, potential level of damage or significance score. Building B may be a historic building or a facility of significance to the community and therefore needs to be prioritized for preservation measures.

Lower risk for buildings A and C: buildings A and C have lower risk scores (0.112 and 0.135, respectively) compared to building B, suggesting that they are at less risk under the current assessment model. Nevertheless, they still require appropriate attention and protective measures to prevent potential risk events from occurring.

Importance of risk management: These risk scores emphasize the importance of conducting regular assessments and implementing effective risk management strategies. For buildings with high-risk scores, measures such as strengthening the building structure, improving emergency preparedness, or purchasing insurance may be required to reduce risk.

4.2. Polynomial Fitting of Damage Caused by Earthquakes

To better understand the impact of earthquakes on buildings, the economy, and human life, it is necessary to develop a model to predict the damage caused by earthquakes. Losses due to earthquakes are related to the intensity of the earthquake (e.g., Richter scale), population density, building construction, and other factors, and these relationships are often nonlinear. Therefore, using a polynomial fitting model to describe the relationships between these variables can help us to more accurately predict and assess the damage caused by earthquakes.

Polynomial fitting is a regression analysis technique used to model the nonlinear relationship between the dependent variable (y) and the independent variable (x) [6]. For the scenario of losses caused by an earthquake, we can consider the intensity of the earthquake as the independent variable (x) and the losses caused by the earthquake (e.g., economic losses, building damage, etc.) as the dependent variable (y). The general form of the polynomial fitting model can be expressed as:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \epsilon \tag{3}$$

y represents the value we want to predict, which is the damage caused by the earthquake. x is our independent variable, which can be the intensity of the earthquake. $\beta_0, \beta_1, \beta_2, \beta_3 \cdots \beta_n$ are the parameters of the model, these are the values we determine by fitting the data. The n is the order of the polynomial, which tells us the complexity of the model. ϵ represents the error term, which indicates the portion of the variation in the data that the model cannot account for.

To measure our model, a loss function is usually used. In the case of polynomial fitting, a commonly used loss function is the mean square error (MSE), which calculates the average of the squares of the differences between the model's predicted and actual values [7]. The mean square error is expressed as:

$$MSE = (1/m)^{*} \Sigma (yi - \hat{y}i) \wedge 2 \tag{4}$$

By applying the model to a test dataset, we can assess the predictive power of the model. If the model performs well on unseen data, it means that it not only captures patterns in the training data, but also has good generalization ability. By comparing the difference between the actual values of the test set and the predicted values of the model, we can further validate the effectiveness of the model.

5. Summary

The property insurance industry faces significant challenges in the face of frequent and intensifying extreme weather events due to climate change. Using ARIMA and CGDAM-WRIR models, we provide insights into how insurers are underwriting policies in high-risk areas and how communities and real estate developers are adapting their insurance models to enhance the resilience of their properties. Our findings show that the impact of climate change on the insurance industry is already significant, with more than 1 trillion $ in losses and a dramatic increase in insurance claims highlighting the severity of the challenge.

To ensure the sustainability of the insurance industry, insurers need to revisit their underwriting strategies, especially in high-risk areas. At the same time, adjustments in community engagement and real estate development strategies can help reduce insurance costs, expand insurance coverage, and enhance overall risk management capabilities. Addressing the challenges of climate change will require governments, insurers and communities to work together and synergize to develop and implement more flexible and sustainable insurance strategies to ensure that the property insurance industry is able to cope with the more complex and severe impacts of climate change in the future.

References

[1] Raymond C, Horton R M, Zscheischler J, et al. Understanding and managing connected extreme events [J]. Nature climate change, 2020, 10 (7): 611 - 621.

[2] Golnaraghi M. Climate change and the insurance industry-risks and opportunities for transitioning to a resilient low carbon economy[M]//Handbook of Business and Climate Change. Edward Elgar Publishing, 2023: 145 - 186.

[3] Adineh A H, Narimani Z, Satapathy S C. Importance of data preprocessing in time series prediction using SARIMA: A case study [J]. International Journal of Knowledge-Based and Intelligent Engineering Systems, 2020, 24 (4): 331 - 342.

[4] Yu C S, Chang S, Chang T H, et al. A COVID-19 pandemic artificial intelligence–based system with deep learning forecasting and automatic statistical data acquisition: development and implementation study [J]. Journal of medical Internet research, 2021, 23 (5): e27806.

[5] Ding Y, Fan Y, Du Z, et al. An integrated geospatial information service system for disaster management in China [J]. International Journal of Digital Earth, 2015, 8 (11): 918 - 945.

[6] Gómez D, Escalante D, Hajari E, et al. Assessing the effects of in vitro imposed water stress on pineapple growth in relation to biochemical stress indicators using polynomial regression analysis [J]. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 2020, 48 (1): 162 - 170.

[7] Kim Y, Oh H. Comparison between multiple regression analysis, polynomial regression analysis, and an artificial neural network for tensile strength prediction of BFRP and GFRP [J]. Materials, 2021, 14 (17): 4861.