Research on Insurance Cost Prediction Model Based on Linear Regression

Xinyang Liu *, Siyu Wang, Xu Fang, Senyao Zhang, Bohan Lin

College of Aerospace and Civil Engineering, Harbin Engineering University, Harbin, China,150001

* Corresponding Author Email: l403x_y7@163.com

Abstract. The insurance industry is experiencing a crisis due to the frequent occurrence of extreme weather worldwide, resulting in significant losses across all sectors. This paper aims to mitigate the impact of extreme weather on the insurance industry by using Harbin and Haikou as case studies. The study employs linear regression (least square method) and Pearson correlation coefficient analysis to simulate the insurance cost of the two cities over the next ten years. The results offer insights into the potential profits and losses of insurance costs during different periods. Additionally, this section presents the management mode of urban insurance companies, taking into account local conditions. The model method responds to national policies, but further data is required to support in-depth research and improvement.

Keywords: Extreme Weather, Least Square Method, Prediction Model, Correlation Coefficient.

1. Introduction

In recent years, the frequency of extreme weather has been increasing, and the economic losses caused by it have been increasing. Of the thousands of alerts issued by International SOS in 2023, the number of alerts related to extreme weather events ranked second in the overall category. [1] According to statistics, the world has endured "more than $1 trillion in damage from more than 1,000 extreme weather events in recent years." [2] The insurance industry finds that natural disaster claims in the coming years will be "115 percent higher than the 30-year average." [2] China is one of the countries in the world that suffers the most serious losses from natural disasters. In the past 20 years, the average annual direct economic losses caused by various kinds of natural disasters have reached 360 billion yuan, and the affected population is about 220 million people each year. [3] At the same time, losses caused by severe weather such as floods, typhoons, drought and freezing damage may increase. In the face of the potential for more extreme weather events in the future, climate change has been closely related to security and social stability. [4] With the increasing vulnerability of natural disasters and the rising cost of insurance compensation, the future prospects of the insurance industry will become worse. [5] How to make insurance companies have the ability to withstand future underwriting risks and thus have better profitability has become the key issue at this stage.

In the past, most of the insurance cost predictions were based on the impact of an extreme weather on the local premium to establish the connection between the extreme weather and the premium, which is often inaccurate. With global warming, the frequency of extreme weather will be greater than before, so it is necessary to predict a total model that covers the insurance cost and all the frequent extreme weather. In this way, the insurance cost prediction model can be more accurate.

This paper takes two representative cities, Harbin and Haikou, as examples. Firstly, the frequency of four common extreme weather in ten years is calculated, and the insurance premiums of these two places in the past ten years are known by consulting the data. Using the frequency as the weight, Pearson correlation analysis is carried out on the extreme weather conditions and insurance costs in these two places respectively. It is found that there is indeed a strong correlation between the insurance cost and these four extreme weather conditions, indicating that the cost of the four extreme weather conditions and the known insurance premiums in the next ten years can be predicted by using the prediction model (linear regression). Finally, the predicted value formula is compared with that of ten years ago, and the profit and loss of the insurance cost in the two cities in different periods can be obtained. In this way, it is better for the two cities to offer suitable premiums at different times.

2. The insurance model of Harbin

2.1. Probability of extreme weather in Harbin

First, we conducted data analysis and statistics on the extreme weather in Harbin City. [Meteorological data source: China Meteorological Data Network - National Data Center for Meteorological Sciences (https://data.cma.cn/analysis/yearbooks.html)] In previous literature, there are generally four types of common extreme weather, including high temperature weather, low temperature weather, typhoon weather and heavy rain weather. According to the collected data, we can get the following four pie charts in Figure 1~4, which illustrate the probability of various extreme weather occurrences in Harbin City.

This month is not mentioned because the probability of extreme weather occurring is zero.

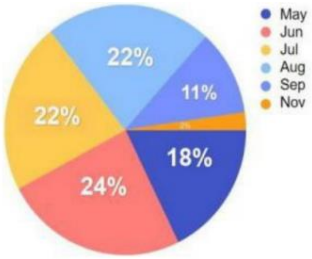

Figure 1. Frequency of rainstorm Weather in Harbin in Recent Ten Years

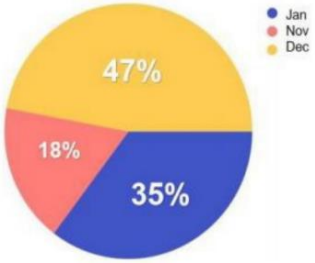

Figure 2. Frequency of low temperature Weather in Harbin in Recent Ten Years

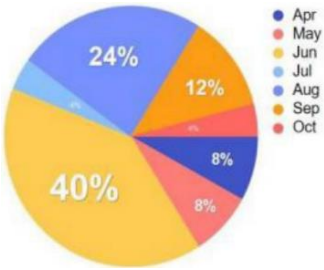

Figure 3. Frequency of tropical cyclone Weather in Harbin in Recent Ten Years

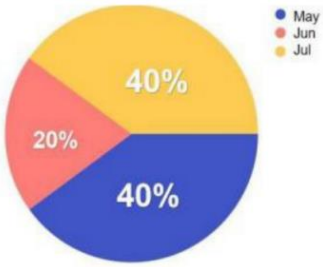

Figure 4. Frequency of high temperature Weather in Harbin in Recent Ten Years

2.2. The impact of extreme weather on insurance

Here, linear regression (i.e. least square method) is used to simulate the relationship between the intensity of four types of extreme weather and the time model, and the intensity of an extreme weather on a given date is calculated by calculating the formula.

After modeling with the least square method [6], it is also necessary to calculate the variance inflation factor (VIF) to detect the multicollinearity between the independent variables. A lower VIF value (generally less than 10) means that the independent variable may be less affected by other independent variables, improving the accuracy and stability of the model. The following four simulated VIF values are (shown in Figure 5~8) less than 10.

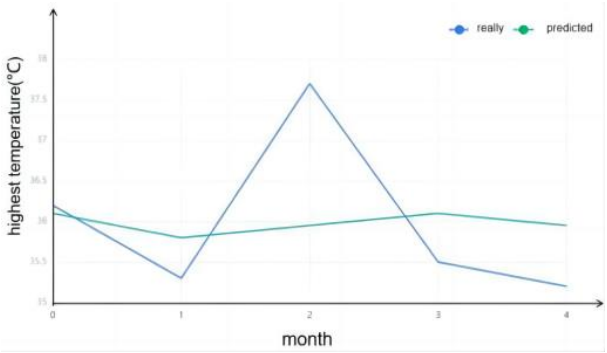

Figure 5. Future high temperature forecast map of Harbin

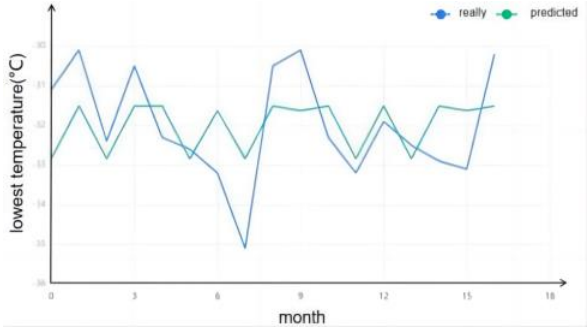

Figure 6. Future low temperature forecast map of Harbin

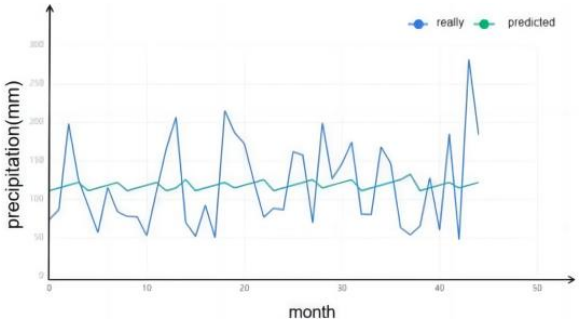

Figure 7. Forecast map of future precipitation in Harbin

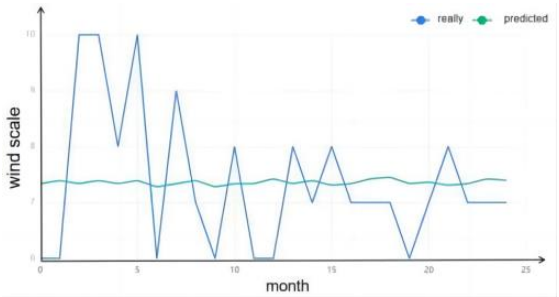

Figure 8. Forecast map of future tropical cyclone weather in Harbin

2.3. An insurance cost model for extreme weather forecasting

Firstly, we adopt Pearson correlation analysis method [7] for research. It can be seen from the literature that Pearson correlation analysis first assumes that the relationship between variables is linear, and measures the linear relationship between two variables when calculating the correlation coefficient. It is applicable to continuous variables and can detect the presence and strength of linear relationships. The conclusion shows that there is a correlation between the insurance cost and extreme weather in Harbin city (that is, the correlation coefficient is greater than 0.3), which can also explain the weight ratio of these four factors on the insurance impact in Figure 9. Finally, the quantitative relationship between insurance cost and four kinds of extreme weather events is obtained by linear regression method and least square method. [Insurance data Source: China Securities Regulatory Commission (http://www.csrc.gov.cn/pub/newsite/sjtj/)]

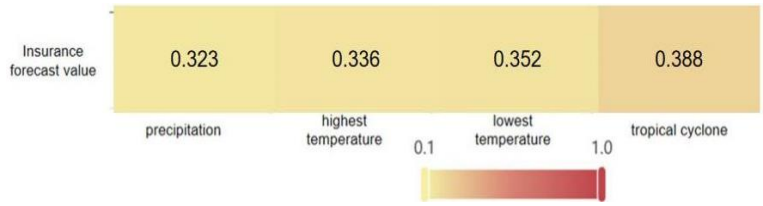

Figure 9. The relationship between insurance cost and four kinds of extreme weather events in Harbin

In order to test the construction effect of the model, we will also conduct F-test [7]. F-test is a method to test the significance of a statistical model based on the assumption that the coefficients of all the independent variables in the model are equal to zero, that is, the independent variables have no significant effect on the dependent variables. Another hypothesis is that the coefficient of at least one independent variable is not equal to zero, that is, the independent variable has a significant effect on the dependent variable. The F-test calculates the F-value by comparing the sum of squares of the model's regression squares and the sum of errors. If the F-value is greater than a certain critical value (which needs to be determined based on confidence levels and degrees of freedom), the null hypothesis is rejected and the model as a whole is significant because at least one independent variable in the model has a significant effect on the dependent variable.

According to the analysis of the F-test results, the significance P value is 0.000***, which shows significance at this level, rejecting the hypothesis that the regression coefficient is 0, so the model basically meets the requirements. For the collinearity of the variables, VIF is less than 10, so there is no multicollinearity problem in the model, and the model is well constructed. The calculation formula is as follows:

$$y = 2.421 + 0.165 \times T_{max} - 0.073 \times T_{min} + 1.835 \times C + 0.023 \times V \quad (1)$$

The comparison between the predicted result and the true value is shown in Figure 10 below.

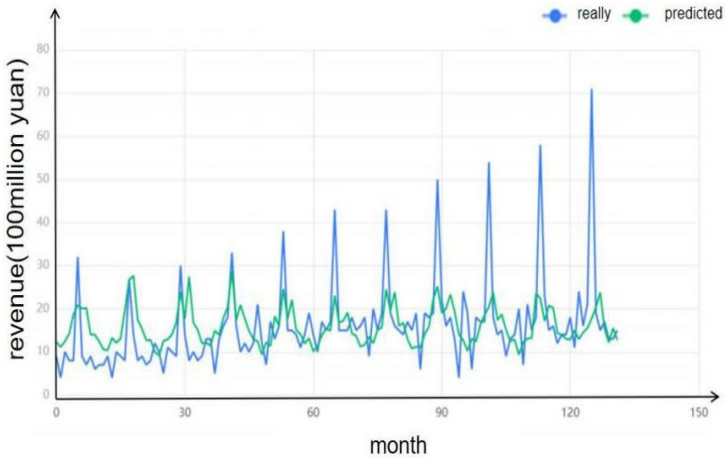

Figure 10. Comparison of actual and predicted insurance costs in Harbin

To sum up, in order to establish the model of the impact of extreme weather on insurance in Harbin, the probability of extreme weather is calculated first, and the intensity of extreme weather within a certain period is calculated through the calculation formula. Finally, by verifying the weak correlation between insurance and four kinds of extreme weather, the quantitative relationship between insurance cost and four kinds of extreme weather is derived and the formula is given. However, it should be noted that the forecast value of the insurance in June each year is always lower than the actual value, that is, the possibility of a loss in that month is high. As can be seen from the figure, the company is prone to losses in June, so it should reduce the insurance amount or appropriately reduce the compensation for losses caused by natural disasters, and it can choose to take risks to increase the insurance coverage in June (in line with the green peak). Other months can be predicted according to the green curve in the figure, that is, keep the forecast sales trend.

3. The insurance model of Haikou

Haikou is also one of the areas where extreme weather [8, 9] is on the rise. Unlike inland Harbin, Haikou is surrounded by the sea and located at a low latitude. Below, we will use a similar method to study Haikou's insurance model.

3.1. Probability of extreme weather in Haikou

Similarly, we conducted data analysis and statistics on the extreme weather [10] in Haikou, including four kinds of high temperature weather, low temperature weather, typhoon weather and heavy rain weather. The following four pie charts can be obtained in Figure 11~14, illustrating the occurrence probability of various extreme weather in Haikou.

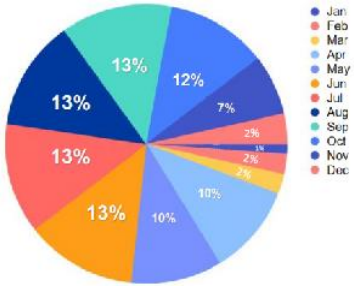

Figure 11. Frequency of rainstorm Weather in Haikou in Recent Ten Years

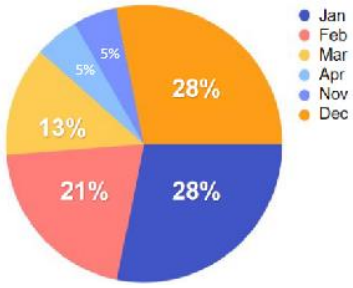

Figure 12. Frequency of low temperature Weather in Haikou in Recent Ten Years

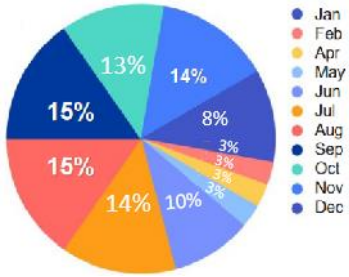

Figure 13. Frequency of tropical cyclone Weather in Haikou in Recent Ten Years

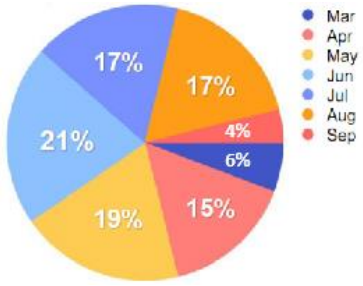

Figure 14. Frequency of high temperature Weather in Haikou in Recent Ten Years

3.2. The impact of extreme weather on insurance

Here, we also model the relationship between the intensity and timing of four extreme weather events by linear regression (least square method), we can calculate the intensity of some extreme weather events on a given date by calculating the formula, and obtain a quantitative prediction of the relationship between insurance costs and four extreme weather events.

After least square modeling, variance inflation factor (VIF) is also calculated to detect multicollinearity between the independent variables. A lower VIF value (generally less than 10) means that the independent variable may be less affected by other independent variables, improving the accuracy and stability of the model. The following four figures' simulated VIF values (shown in Figure 15~18) are all less than 10.

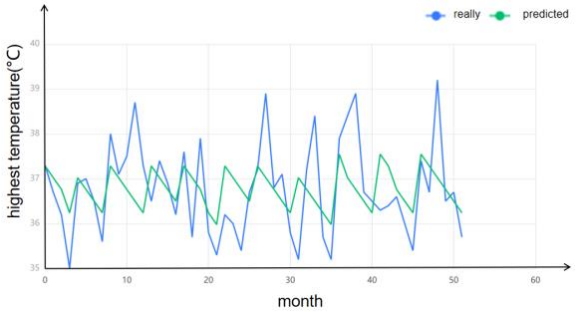

Figure 15. Future high temperature forecast map of Haikou

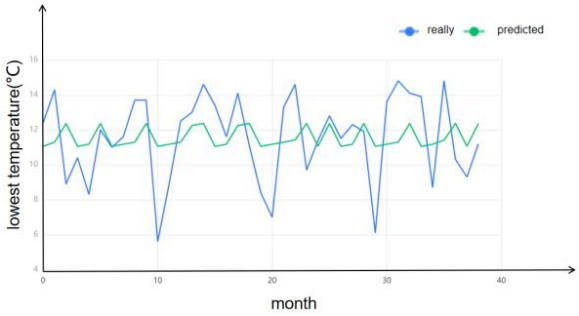

Figure 16. Future low temperature forecast map of Haikou

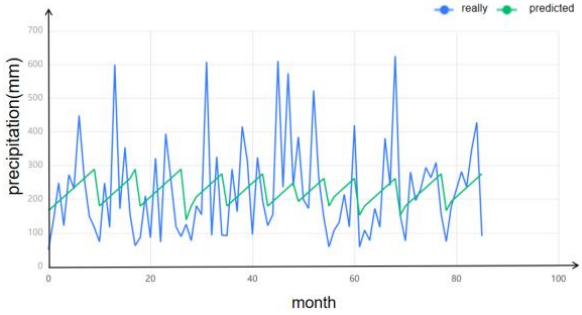

Figure 17. Forecast map of future precipitation in Haikou

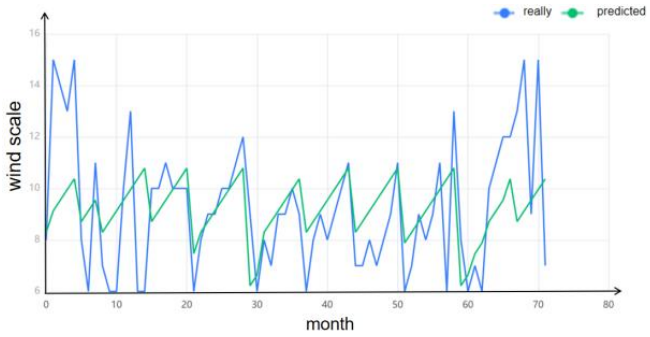

Figure 18. Forecast map of future tropical cyclone weather in Haikou

3.3. An insurance cost model for extreme weather forecasting

Unlike Harbin, the relationship between extreme weather and insurance costs in Haikou is in Figure 19.

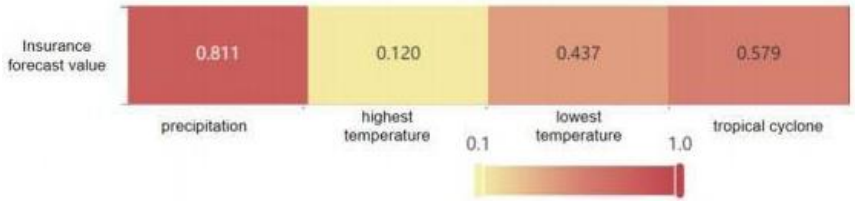

Figure 19. The relationship between insurance cost and four kinds of extreme weather events in Haikou

The linear regression prediction model (least square method) is used to obtain the model between the insurance and the predicted value: according to the analysis of the F-test results, the significance P value is 0.000***, which shows the significance at this level, rejecting the hypothesis that the regression coefficient is 0, so the model basically meets the requirements. For the collinearity of the variables, VIF is all less than 10, so there is no multicollinearity problem in the model, and the model is well constructed. The formula is as follows: (The green line in Figure 20 is the formula)

$$y = 7.513 - 0.096 \times T_{max} + 0.032 \times T_{min} - 0.001 \times C - 0.011 \times V \quad (2)$$

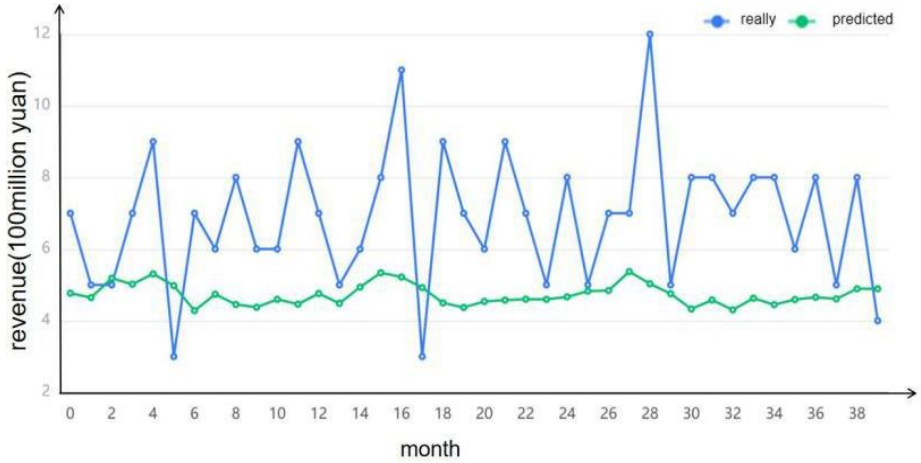

Figure 20. Comparison of actual and predicted insurance costs in Haikou

To sum up, the model of the impact of extreme weather on insurance in Haikou is basically consistent with the analysis of Harbin City, but the prediction model shows that insurance companies have been losing money, indicating that Haikou is not suitable for insurance companies to carry out long-term insurance.

4. Conclusion

In this paper, Pearson correlation analysis is firstly carried out on four kinds of extreme weather and insurance costs in Harbin and Haikou in the next ten years. Secondly, linear regression (least square method) is used to predict the insurance costs in the next ten years, and the image of insurance costs in the next ten years is successfully simulated and compared with the existing ten years' insurance costs. Losses and profits over the next ten years, so that the two cities can better offer suitable premiums at different times.

With strong data, logic, accuracy and comprehensiveness, this method can systematically simulate the loss and profit of insurance costs with data, which is consistent with the plan of gradually establishing climate change risk assessment in China, and can better play the role of economic compensation after disasters and realize data-based management. However, the model considers a limited variety of extreme weather and cannot include all extreme weather in the model, which leads to the fact that if unknown extreme weather occurs, we cannot use the model to predict it. Therefore, in the future model building work, we can specifically explore the factors of extreme weather occurrence, and explore the frequency of extreme weather occurrence through various factors to make the model more accurate.

References

[1] Wang Fangqi. International SOS call for enterprises to strengthen risk management [N]. The bank of China insurance quote, 2024-01-04 (007). The DOI: 10.28049/n.c. Nki NCBXB. 2024.000042.

[2] Boston Consulting Group. (2023, December 4). An Insurance Risk Framework for Climate Adaptation. Retrieved at: https://www.bcg.com/publications/2023/an-insurance-risk-framework-for-climate-adaptation.

[3] Li Lixiao, Zhang Lun, Sun Mengsi et al. Analysis of extreme natural disaster risk and disaster-causing model in typical urban areas under the background of urban size and population growth [J/OL]. Industrial building: 1-11 [2024-03-18]. http://kns.cnki.net/kcms/detail/11.2068.TU.20240307.1425.004.html.

[4] Wu Jiannan, Gao Xiaoping, Zhong Kaibin, etc. To enhance the resilience of urban governance under extreme weather [J/OL]. Exploration and contention: 1-23 [2024-02-28]. http://kns.cnki.net/kcms/detail/31.1208.c.20230103.1130.001.html.

[5] Guo Xiaojie, Liu Shu. Frequent extreme weather who will "insure" the insurance industry? [N]. The 21st century economic report, 2023-12-27 (008) DOI: 10.28723/n.c. Nki NSJBD. 2023.005140.

[6] Wang Rongzhong, Gao Qiuying, Wang Dejun. The least squares-based curve fitting and its simplified algorithm [J]. The Sensor World, and the 2021, 27 (10): 8-10+25. DOI: 10.16204/j.cnki.sw. 2021.10.002.

[7] SHI Biao, LI Yu Xia, YU Xhua, YAN Wang. Short-term load forecasting based on modified particle swarm optimizer and fuzzy neural network model [J]. Systems Engineering-Theory and Practice, 2010, 30 (1): 158-160.

[8] Society Disasters and accidents. Mao Zhihua, editor-in-chief, Hainan Yearbook, Hainan Yearbook Society, 2016, 484-486, Yearbook.

[9] Society Meteorology-Earthquakes. Wu Zeming Editor-in-chief, Hainan Yearbook, Hainan Yearbook Society, 2022, 511-517, Yearbook.

[10] Weather, earthquake and meteorology. Mao Zhihua, Hainan Yearbook, Hainan Yearbook Society, 2021, 479-483, Yearbook.