Actuarial study and statistical analysis of extreme wildfire insurance claims

by

Jiali Wang

A thesis submitted to
the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements
of the degree of

Master of Science

I.H. Asper School of Business
University of Manitoba
Winnipeg, Manitoba, Canada

Abstract

The wildfire season of 2023 was one of the most devastating in Canadian history and caused significant losses to the insurance industry. This study focuses on the large wildfire losses exceeding CAD 25 million in the last 16 years (2008-2023).

For severity analysis, mainly due to the largest amount of wildfire losses reaching CAD 4 billion, heavy-tail distributions have been applied to find the best fit for the loss claims history. We also estimate loss amounts in different return periods.

Traditional actuarial ratemaking methods in non-life insurance are based on the assumption of the absence of correlation between the frequency and severity of claims. In this thesis, for frequency analysis, logistic regression is used to investigate the relationship between the maximum temperature and the probability of extreme wildfire losses. Consistent with common sense, the probability of a large wildfire occurring is found to increase with increasing temperature. According to the requirements of Solvency II, we derive the 99.5% confidence interval for the claim frequency.

Overall, there are not many published papers and available data on Canadian wildfire insurance losses. Therefore, this thesis lists some possible research directions in property and casualty insurance, reinsurance, and life insurance in the future research directions section.

Acknowledgements

I would like to first express my gratitude to my supervisor, Dr. Xikui Wang, for the countless time and energy he devoted to me and my thesis. I firmly believe that this work could not have been done without encouragement and patient guidance from him. Dr. Xikui Wang's solid academic foundation and rich research experience gave me the right direction when I encountered difficulties. Thank you for providing me with the opportunity to show my research, at the 2024 Statistical Society of Canada (SSC) Annual Meeting and Winnipeg Actuaries' Club (WAC) Semi-Annual Meeting.

I also want to say thank you to Dr. Xuemiao Hao and Dr. Wenlong Yuan as committee members, for their invaluable feedback and comments that indeed helped me a lot to improve and complete my thesis work.

Finally, I would like to say a big thank you to my parents, for all of their understanding, support and love over the years. Thank you to them for their support in my decision to pursue my studies, especially during the COVID-19 period when they fully understood the various challenges in my study and work.

Contents

1

Introduction

1.1 Introduction

In this chapter, we introduce the issue of global warming. Then, some facts about wildfires, especially wildfires in Canada, are presented. In actuarial practice, the catastrophe model is a useful tool to price extreme risk and the two most commonly used models are introduced. Last, we propose research objectives, hypotheses and a summary of this thesis.

1.2 Some facts and effects of climate change

According to the National Oceanic and Atmospheric Administration (NOAA), global temperatures rose about 1.98° F (1.1°C) from 1901 to 2020. As shown in Figure 1, since 1900, the global average surface temperature has significantly increased. Before 1935, the annual global surface temperature was below the average between 1901 and 2020, while after 1980, the annual global surface temperature was above the average. Moreover, after 2010, the magnitude of annual global surface temperature increases significantly. Particularly, the 10 warmest years from 1850 to 2023 have all occurred in the past decade (2014-2023). In 2023, the global temperature has risen

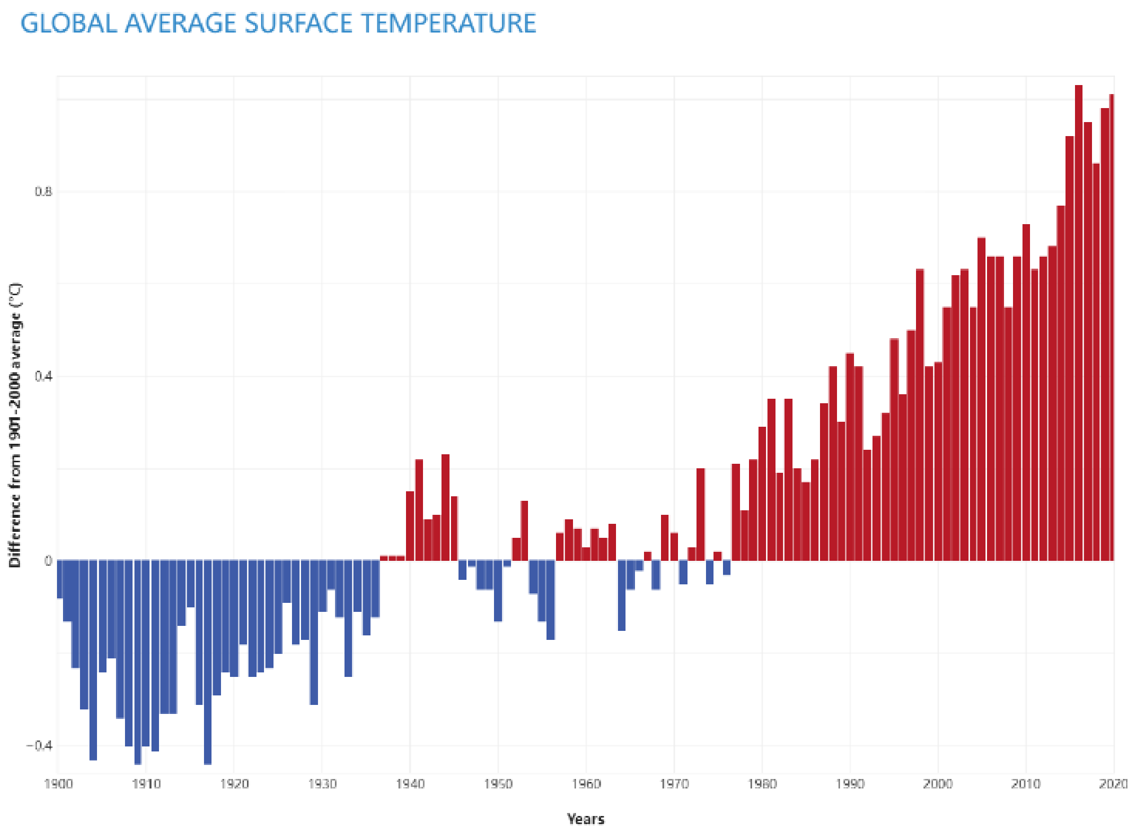

Figure 1.1: Global Average Surface Temerature [32]

2.12 °F (1.18 °C) above the 20th-century average and 2.43 °F (1.35 °C) above the pre-industrial average (1850-1900) [32].

Numerous other changes have resulted from the global temperature increase. Over the majority of the 20th century, sea level rise was 1.7 mm/year. Starting in 1993, it has accelerated to 3.2 mm/year [16]. Since 1980, the average thickness of thirty carefully examined glaciers has dropped by more than sixty feet, indicating that glaciers are receding. In 2019, the area covered by sea ice in the Arctic at the end of summer has decreased by roughly 40% since 1979 [15].

According to the Global Annual Temperature Outlook by the National Centers for Environmental Information (NCEI), there is a 22% probability that 2024 will be the warmest year on record and a 99% likelihood that it will rank among the top five warmest years ever [18].

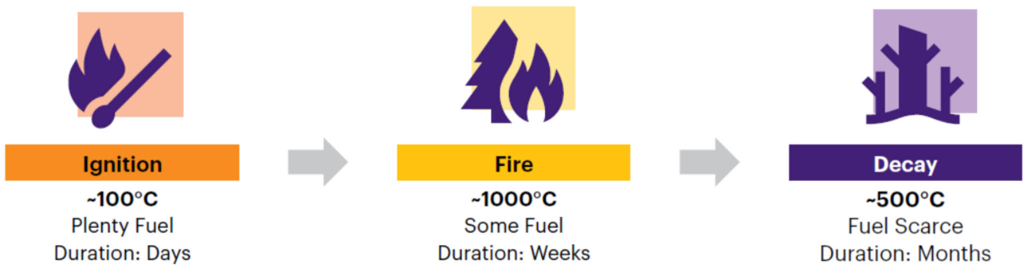

Figure 1.2: Life Cycle of a Wildfire [30]

1.3 Wildfire and its catastrophic economic impact

A wildfire means an uncontrolled wildfire that occurs in a wilderness area. Figure 1.2 shows the life cycle of a wildfire. Both man-made and natural events have the potential to start a wildfire. Lightning strikes, burning trash (or other wildfires), improperly extinguished campfires, cigarette butts, and even arson are common sources of wildfire ignition. A fire spreads quickly after it is started. Numerous variables, such as wind, drought, fuel availability, and exceptionally hot weather, can affect how quickly and far a wildfire spreads [34].

Wildfires can spread quickly and far if the right conditions are met, which is why quick suppression measures are crucial. In managing and suppressing wildfires, firefighting teams assess the fire based on its access points, weather, hazards, natural barriers, and firebreaks. Based on this information, they devise a plan for utilizing the available fire suppression resources. A wildfire will be allowed to burn under close observation if it has been contained and deemed non-threatening [34].

Instead of referring to the time between a wildfire's ignition and total extinguishment, duration in this thesis refers to the number of days from when the fire has already started to spread out of control to when it is under control.

Climate change affects a wildfire's ability to spread and intensify, though it can be started by humans or by natural events like lightning strikes [30].

From the United Nations Environment Programme, climate change and wildfires exacerbate each other. Climate change makes wildfires worse by causing longer, hotter, and drier fire seasons due to increased drought, high air temperatures, low relative humidity, lightning, and strong winds. In addition, wildfires exacerbate climate change by destroying delicate, carbon-rich ecosystems such as peatlands and rainforests. It becomes more difficult to stop rising temperatures as a result, turning landscapes into tinderboxes [50].

According to the Natural Resources Canada, approximately 7,300 forest fires have happened annually during the past 25 years in Canada. Although it varies greatly from year to year, the yearly total area burned averages roughly 2.5 million hectares. The cost of fighting fires has fluctuated between $800 million and $1.05 billion annually over the past ten years. These fires happen in grasslands, shrublands, and forests. Some wildfires go out of control due to human errors or lightning strikes [11].

Marc and Gerry (2019) proposed that wildfires tend to be rural and exhibit a notable proportion of overall losses to structures as a result of inadequate fire protection resources. In contrast, other natural calamities like hurricanes and earthquakes typically entail partial loss scenarios occurring across numerous locations and a significantly larger geographical area. The distinct characteristic of conflagration losses, where individual losses are substantial, contributes to the comparatively significant monetary losses stemming from wildfire incidents [13].

Canada experienced an unprecedented wildfire season in 2023. Approximately 18.5 million hectares (45.7 million acres) of land was burned, while the previous ten-year (2013-2022) average was only 2.97 million hectares, smashing the previous annual high in 1995 nearly three times over. Wildfires in 2023 caused notable structural and vehicular damage and resulted in economic losses of $1.5 billion and insured losses of at least $760 million. The costliest wildfires in 2023, the West Kelowna and Bush Creek fires occurred in British Columbia. Other severe wildfires

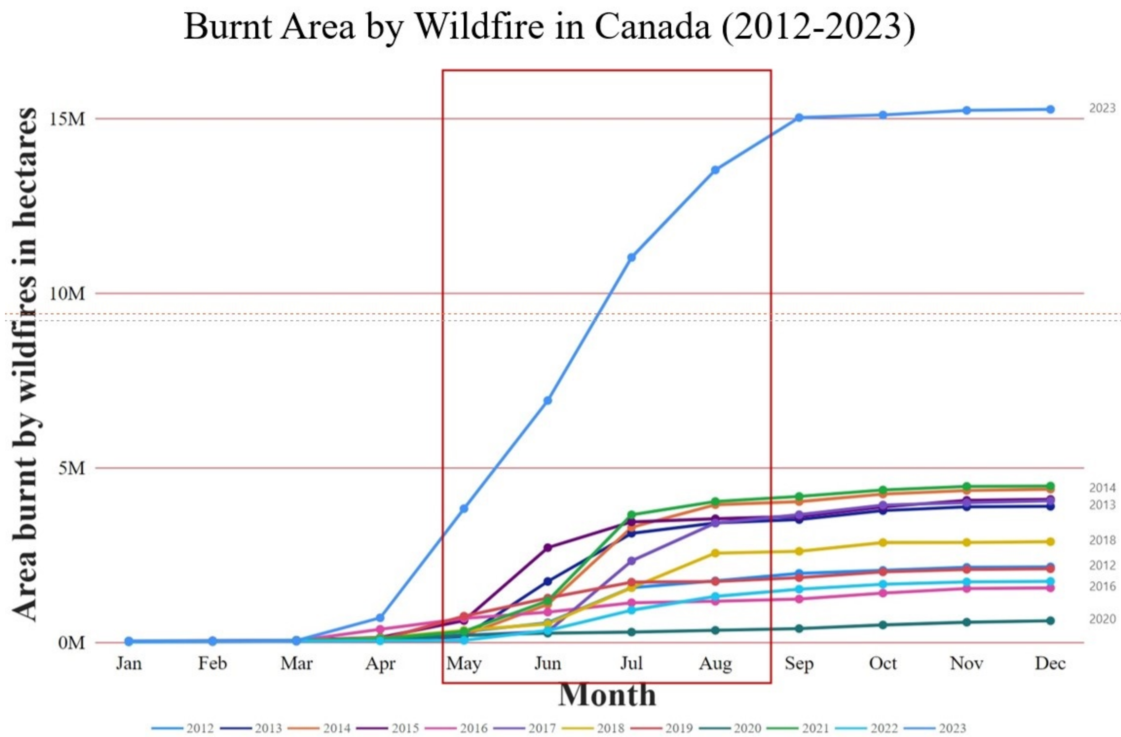

Figure 1.3: Burnt Area by Wildfire in Canada [52]

appeared across Northwest Territories, Québec, Nova Scotia, and Alberta. Figure 1.3 shows the burnt area by wildfire in Canada from 2012 to 2023 [52].

There are many related studies on the factors and predictions of economic losses caused by wildfire disasters. Jeffrey, Karen, and Krista (2008) used the Niño 3 Sea Surface Temperature Anomaly (SSTA), the Southern Indicators such as Observation Index (SOI) and North Atlantic Observation (NAO) to measure the degree of El Niño event to predict compression costs of wildfire seasons [49]. Meng and Daoyi (2022) concluded that there is a strong correlation between winter North Atlantic Sea Surface Temperature and spring wildfires over Europe [36]. Brian and William (2003) applied the Southern Oscillation Index (SOI) of ENSO to predict the area of wildfire [7]. Skinner et al. (2006) found that long-range forecasting for fire severity in Canada may be based on a 6-month lag link between large-scale SSTs and the unpredictability of forest fires in Canada [55].

Niño 3 SSTA serves as an indicator of sea surface temperatures in the central

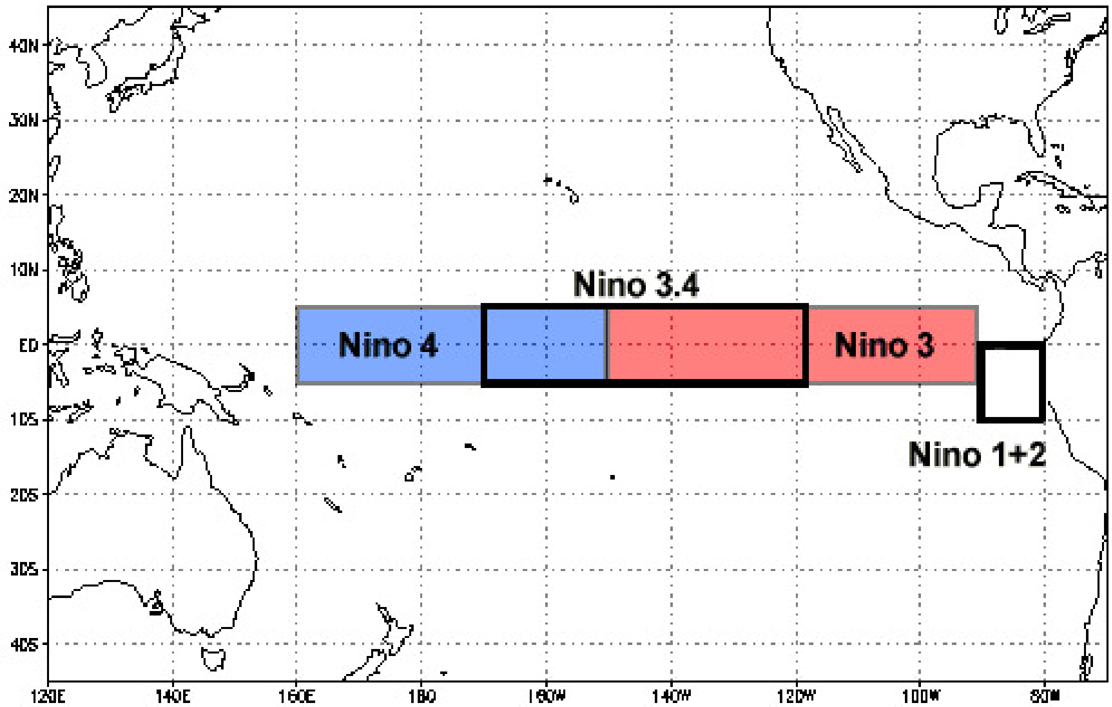

Figure 1.4: Niño Regions [17]

Pacific equatorial region [17]. The Southern Oscillation Index (SOI) is determined by the air pressures at sea level in Tahiti and Darwin, Australia. The SOI is connected to the same sea surface temperatures in the central Pacific that characterize El Niño and La Nina events [19]. The North Atlantic Oscillation (NAO) assesses the disparity between low pressure systems in the polar region and high pressure systems in the subtropical zone of the northern Atlantic Ocean [33].

In this thesis, in addition to the meteorological variables mentioned above, we also attempt to discover the relationship between other important variables and Canadian wildfire insurance losses, such as the three and six months before the wildfire accident of Niño1, Niño 3, Niño 4 and Niño 3.4 data. Figure 1.4 shows the location of Niño 1, Niño 3, Niño 4 and Niño 3.4. However, the results were not significant enough (P values are all larger than 0.1), so no further statistical analysis was conducted on these variables. Therefore, in this thesis, the temperature mainly refers to the temperature of the province where the catastrophe occurred.

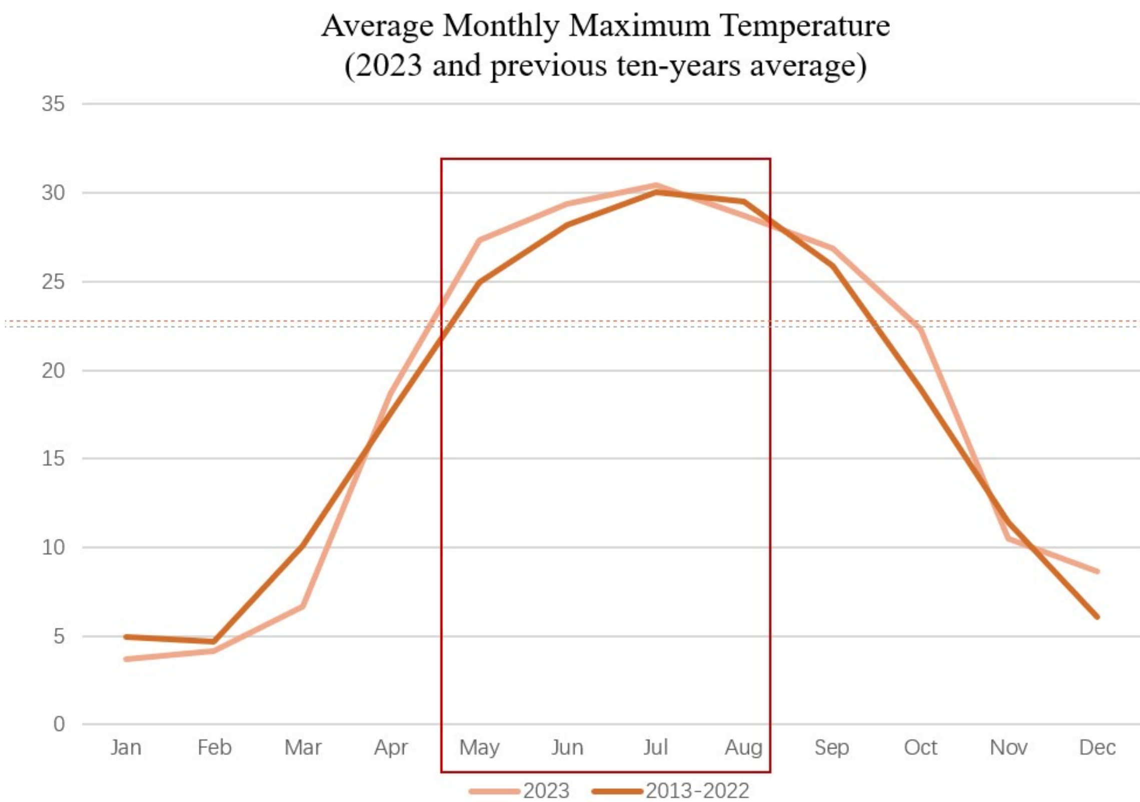

Figure 1.5: Average Monthly Maximum Temperature [43]

In the summer of 2023, as shown in figure 1.5, especially in May and June, the highest temperature was significantly higher than the average in the previous years. This situation is strongly correlated with the severity of wildfires. In the subsequent analysis, this relationship will be analyzed in detail.

Moreover, Canadian wildfires generated considerable secondary impacts throughout the year. In June, thick smoke plumes from the fires extended to south, and affected parts of north-eastern United States with unhealthy air quality conditions for tens of millions of people. Several cities, including New York, saw their worst air quality on record. Smoke from fires dropped solar farms production and delayed hundreds of flights in the region. Canadian wildfires produced 23 percent of the global wildfire carbon emissions for 2023: 480 megatonnes of carbon (MtC) out of the global total of 2,100 MtC. This amount was three times higher compared to the Canadian fossil carbon emissions of 150 MtC in 2022 [5].

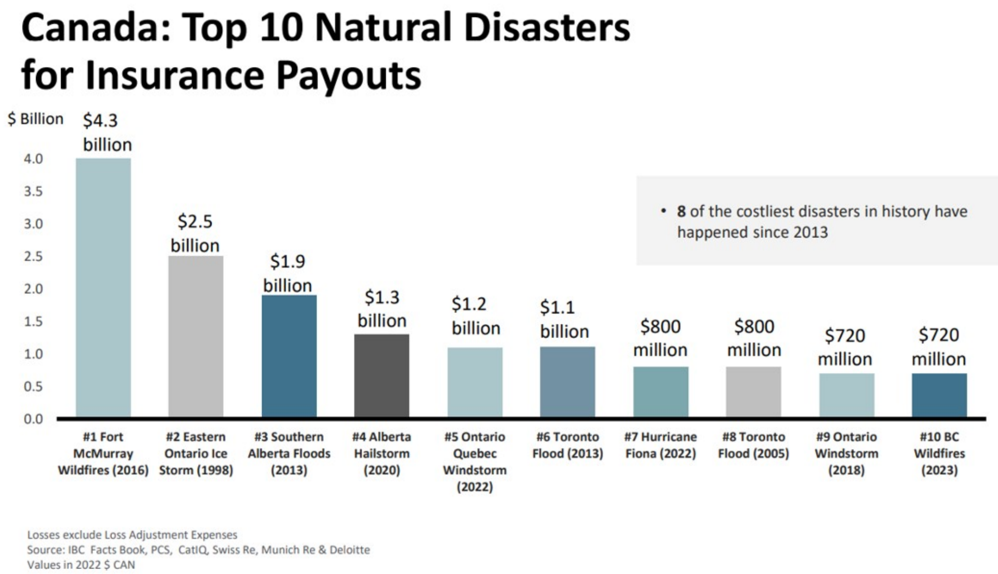

Figure 1.6: Top 10 Natural Disasters for Insurance Payouts in Canada [42]

The wildfire disaster that occurred in British Columbia in 2023 resulted in an insurance loss of CAD 720 million, making it the tenth largest natural disaster loss in Canada. Figure 1.5 lists top 10 natural disasters for insurance payouts in Canada. The most serious natural disaster in Canada is caused by wildfire, which occurred in Alberta in 2016 and caused insurance losses exceeding CAD 4 billion [42].

1.4 Catastrophe models in insurance

In the property and casualty insurance industry, the catastrophe model is a commonly method for estimating natural disaster losses. Bushfire and Wildfire Models by Verisk and Wildfire High-Definition (HD) Model by Moody's RMS are the two main models. Based on the Verisk (formerly AIR Worldwide) model, the disaster model is divided into three modules: The hazard component, the engineering component, and the financial component. The hazard component calculates the intensity of disasters at various locations based on simulating various types of catastrophic events. The engineering component mainly considers the risk exposure situation affected by catastrophic disasters. The financial component uses various probability distributions to estimate the amount of loss [3].

The model by Verisk meticulously examines six factors (Table 1.1) to effectively address the risk of wildfires at the individual property level. Subsequently, these factors can be used to establish a correlation with the level of risk associated with the property.

Factor	Description
Fuel	Fires are fueled by trees, bushes, and other plants. Close to a structure or densely packed vegetation can pose a serious risk.
Terrain	Wildfire spread rate is influenced by topography. A hillside's slope, or incline, can influence wind patterns, which affect the direction and speed of fires. The spread, however, can be slowed by natural obstacles like rocky terrain or bodies of water.
Access	Firefighting apparatus may be hindered by narrow roads and dead ends.
Wind patterns	Strong, dry wind patterns, like the Santa Ana and Sundowner winds, can carry embers, widen the area of a fire, or start new ones. The state of fighting fires may worsen due to these winds.
Special Hazard zones	Even though there may be less of a fire risk in these areas, businesses are nevertheless affected by smoke, ash, and business closures and evacuations.
Mitigation	Attempts to lower risk at the individual and community levels can lessen a particular property's vulnerability to loss or damage from wildfires by managing vegetation and maintaining the property.

Table 1.1: Wildfire Risk Factors Considered by Verisk [2]

In addition, Moody's has introduced the Canada Wildfire HD Model, an advanced tool that offers comprehensive probabilistic wildfire modeling for all 10 provinces in the country. This model utilizes a simulation-based framework with numerous realizations over a span of thousands of simulated years, taking into account crucial factors such as topography, fuel type, and weather parameters. Moreover, the model encompasses ignition and spread, along with the inclusion of ember footprints, smoke footprints, and urban conflagration [1].

For decades, researchers have attempted to predict wildfires and proposed a number of different approaches. The majority of research can be categorized into two groups: empirical studies that evaluate wildfire characteristics using historical data on wildfires, and physical models that replicate the physical processes of wildfires. The use of sophisticated Machine Learning (ML) models is likely the most notable development in recent years among empirical studies. The comparatively high performance of machine learning models in predicting non-linear phenomena is one of their main advantages. Due to the non-linear interactions between different factors that influence wildfire behavior, researchers have shown how well ML models perform both locally and globally in predicting wildfires, highlighting their potential contribution to field predictions [54].

1.5 Research objectives and hypotheses

This thesis first investigates the connection between temperature, precipitation and wind speed and the insurance loss amount of wildfires. Then, we choose the most important variables to forecast the eventual loss amount, serving as a guide for insurance pricing and other matters. This thesis mainly studies the insurance losses of wildfires in two aspects: the probability and severity of extreme wildfires.

This thesis hypothesizes that some variables, such as temperature, can significantly affect wildfire insurance losses.

1.6 Summary of the thesis

The insurance industry is facing more uncertainty as a result of the losses from catastrophic insurance brought on by climate change. This thesis focuses on the notable rise in wildfire disasters in Canada and looks for patterns in the relationship between insurance losses and meteorological indicators to provide references for future insurance loss predictions. The primary tool utilized in the research process is R software for statistical regression analysis.

The thesis contains five chapters in addition to the Introduction: Chapter 2 describes the available data, including response variables and predictor variables. Chapter 3 conducts severity analysis on wildfire aggregate claims. Chapter 4 applies logistic regression to investigate the possibility of extreme wildfires occurring. Conclusions and directions for future work are presented in chapter 5.

2

Data description and exploratory data analysis

2.1 Introduction

In this chapter, we describe the basic information of the CatIQ database, including data sources, the data collection process, and insurance loss analysis. Then, we investigate the relationship between factors such as temperature, precipitation, wind power and burned area and extreme wildfire losses.

2.2 Description of the database

Data was obtained from Catastrophe Indices and Quantification Inc. (CatIQ), launched in 2014. This thesis analyzes catastrophe losses from the Canadian insurance industry and includes 183 losses (containing 11 wildfire losses) between 2008 and 2023. CatIQ collects catastrophe loss data from the Canadian insurance industry with losses exceeding CAD 25 million. Insurance loss data is divided into three categories based on lines of business, including personal property, commercial property, and auto [24].

CatIQ releases preliminary estimated loss data 45 days after the catastrophic event and updates 90 days, 180 days, and 1 year later. If the catastrophe loss exceeds CAD 500 million, the final update will also be conducted in 2 years [26]. Considering that only two losses exceeded CAD 500 million in 2023 and the last update had a minor change in the data, this thesis selects catastrophic loss data from 2008 to 2023. In addition to the data on catastrophe losses, this thesis also uses the risk exposure data provided by CatIQ as of February 9, 2024. The loss data of CatIQ is before reinsurance and subrogation.

CatIQ provides two types of data for each loss: claims incurred and closed claims incurred. Claims incurred refers to the funds reserved by insurance companies for claims, including reserves, while closed claims incurred refers to the actual amount of claims paid by the insurance company. CatIQ no longer updates data after two years of claims, and some complex claims may not be closed within two years [24]. Therefore, this thesis uses claims incurred data.

All ten provinces and three territories across Canada are covered in the CatIQ database. The provinces are, in alphabetical order: Alberta (AB), British Columbia (BC), Manitoba (MB), New Brunswick (NB), Newfoundland and Labrador (NL), Nova Scotia (NS), Ontario (ON), Prince Edward Island (PE), Quebec (QC), and Saskatchewan (SK). The three territories are the Northwest Territories (NT), Nunavut (NU), and Yukon (YT). The types of catastrophic disasters include hail, flood, wind storm, water, winter storms, riot, fire, terrorism, earthquake and explosion [26].

The loss data for Canadian property and casualty insurance companies caused by catastrophe was downloaded on February 9, 2024. Figure 2.1 shows the changes in the catastrophe loss amount and catastrophe loss ratios of Canadian large property and accident insurance companies since 2008, where the catastrophe loss ratio is equal to the catastrophe loss divided by premium income. The premium income of Canadian property and casualty insurance companies was downloaded from the Office of the Superior of Financial Institutions (OSFI) [45].

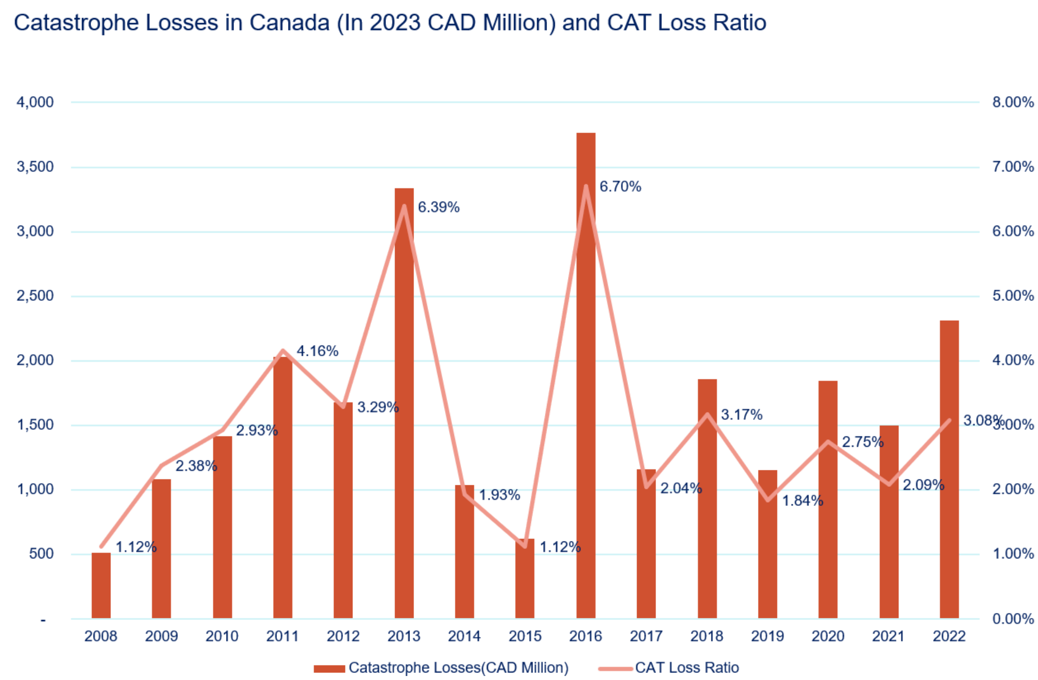

Figure 2.1: Catastrophe losses in Canada and Catastrophe loss ratio [25]

Since 2008, the losses caused by natural disasters have fluctuated greatly, and it was due to a wildfire that the amount of losses in 2016 was the highest in recent years. From the chart, the loss amount for catastrophic insurance in most years is below 2 billion CAD, with three years (2013, 2016, and 2022) having a loss amount exceeding 2 billion CAD. The main reason is that in 2013, the Southern Alberta Flood resulted in a loss exceeding 1.6 billion CAD and the Greater Toronto Area Flooding had a loss exceeding 1 billion CAD. In 2016, the Fort McMurray Wildfire, the largest catastrophe event in Canadian history, occurred with a loss amount exceeding 3.6 billion CAD. In 2022, the Ontario and Quebec Storms occurred, resulting in a loss amount exceeding 800 million CAD.

Figure 2.2 shows the distribution of cumulative loss amount for Canadian catastrophes from 2008 to 2022 by province and the exposure map of insurance coverage is as of February 9, 2024 (Others include Prince Edward Island, Newfoundland and

Labrador, New Brunswick, Northwest Territories, Nunavut, and Yukon). The darker the color of the province is, the greater the relevant data is. It can be observed that although Alberta only accounts for 13% of the insurance exposure, the loss accounts for nearly 50%. Ontario, which has the largest coverage (40%), only accounts for 27% of the national loss amount.

Province	Loss Amount (in CAD **Millions**)	Shares (%)
Alberta	12,674.56	49.49%
Ontario	6,970.56	27.22%
Quebec	2,424.44	9.47%
Saskatchewan	1,363.70	5.32%
British Columbia	638.48	2.49%
Nova Scotia	615.38	2.40%
Manitoba	390.03	1.52%
Others	534.54	2.09%

Province	Loss Amount (in CAD **Billions**)	Shares (%)
Ontario	6,351.83	40.34%
Quebec	3,157.13	20.05%
British Columbia	2,239.50	14.22%
Alberta	1,978.73	12.57%
Manitoba	590.25	3.75%
Saskatchewan	587.63	3.73%
Nova Scotia	348.69	2.21%
Others	493.10	3.13%

Figure 2.2: CAT loss and loss limit of Canada by province and territory [25]

2.3 Description of important variables

2.3.1 Description of response variables

To improve the comparability of the data, this thesis uses the inflation rate data of the World Bank (Table 2.1) [22] and treats the amounts for all years as Canadian dollars in 2023. According to Statistics Canada, the top three basket shares in CPI data from 2017 to 2022 include shelter, food, and transportation. This thesis assumes that CPI data can indicate changes in insurance companies' loss for catastrophic losses.

The Canadian Forest Fire Weather Index (FWI) System mainly considers the following four factors: temperature, relative humidity, wind speed, and 24-hour pre-

Year	Inflation, consumer prices (annual %)	Conversion Factors to 2023 CAD
2011	2.91	1.30
2012	1.52	1.26
2013	0.94	1.24
2014	1.91	1.23
2015	1.13	1.21
2016	1.43	1.19
2017	1.60	1.18
2018	2.27	1.16
2019	1.95	1.13
2020	0.72	1.11
2021	3.40	1.10
2022	6.80	1.07

Table 2.1: Inflation ratios and conversion factors in 2023 dolllar [22]

cipitation [10]. To study the factors that cause losses, this thesis mainly considers temperature, precipitation, and maximum wind speed. The main reason for not considering relative humidity is that reliable data sources have not been found. The following list provides details for these parameters. The first six meteorological data are from the Meteorological Service of Canada [43], and the following two factors are from CatIQ's description of accidents.

1. *Max_Temp* : The highest local temperature on the first day of a wildfire accident.

2. *Differences_in_Max_Temp* : The difference between *Max_Temp* and the five-year average of the highest temperature on the first day of a wildfire accident.

3. *Precip* : The cumulative precipitation in the area in the 90 days before the day of wildfire occurrence where the wildfire occurred.

4. *Differences_in_Precip* : The difference between *Precip* and the average precipitation of the previous five years in the area where the wildfire occurred.

5. *Max_Gust* : The strongest gust on the first day of a wildfire accident.

6. *Differences_in_Max_Gust* : The difference between *Max_Gust* and the average strongest gust in the five years before the same date.

7. *Burned_area* : Total area that was burned.

8. *Duration_Days* : The total number of days from the start of the fire losing control to being brought under control.

Figure 2.3 lists the summay of the factors, including the mean, standard deviation, minimum, 25th percentile, median, 75th percentile and maximum.

2.4 Some important summary statistics and some illustrative graphic and tabular summaries

Summary on Factors

Variable	Mean	Sd	Min	Pctile[25]	Median	Pctile[75]	Max
Max_Temp	31	8.4	20	24	31	35	49
Differences_in_Max_Temp	7.4	5.8	-4.2	6.4	8	8.4	21
Precip	45	28	9.4	33	38	44	115
Differences_in_Precip	-52	36	-119	-78	-44	-24	-3.3
Max_Gust	50	19	30	36	44	58	89
Differences_in_Max_Gust	13	18	-3.6	-2.1	7.4	19	59
Burned_area	156438	201684	950	24092	81139	184581	589552
Duration_Days	32	25	2	15	30	38	83

Figure 2.3: Summary on Factors

2.4.1 Correlation matrix

The covariance matrix shows the correlation between these three types of losses: personal property loss, commercial property loss and auto loss. The upper graph Figure 2.4 includes data from all 11 wildfire accidents, while the lower graph Figure 2.5 is based on data after removing the Forest McMurray wildfire. When taking into account all wildfire losses, the three categories of insurance losses exhibit a strong correlation. However, when excluding Fort McMurray wildfire, the correlation is not as significant as previously, the correlation is approximately 80%.

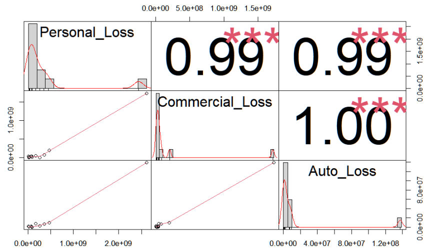

Figure 2.4: Correlation Matrix for all losses

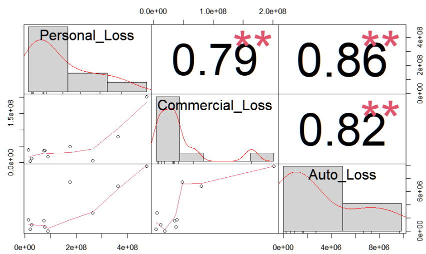

Figure 2.5: Correlation Matrix for Losses excluding Fort McMurray

2.4.2 Matrix of (partial) correlation coefficients

In the study of the relationship between the amount of loss and the eight variables mentioned above, only wind power has a significant correlation. In Figure 2.6, the number represents the Pearson correlation coefficient (PCC), representing the strength of the correlation between the variable and the amount of loss. In statistics, the Pearson correlation coefficient (PCC) is a metric used to assess the linear relationship between two datasets. It quantifies the degree of linear correlation by comparing the covariance of the variables to the product of their standard deviations. Consequently, the PCC offers a standardized measure of covariance, ensuring that the resulting value falls within the range of -1 to 1. Similar to covariance, this metric is limited to capturing linear associations between variables, disregarding other forms of relationships or correlations [20].

$$Pearson\ correlation\ coefficient\ (PCC) = \frac{1}{(n-1)s_x s_y}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) \quad (2.4.1)$$

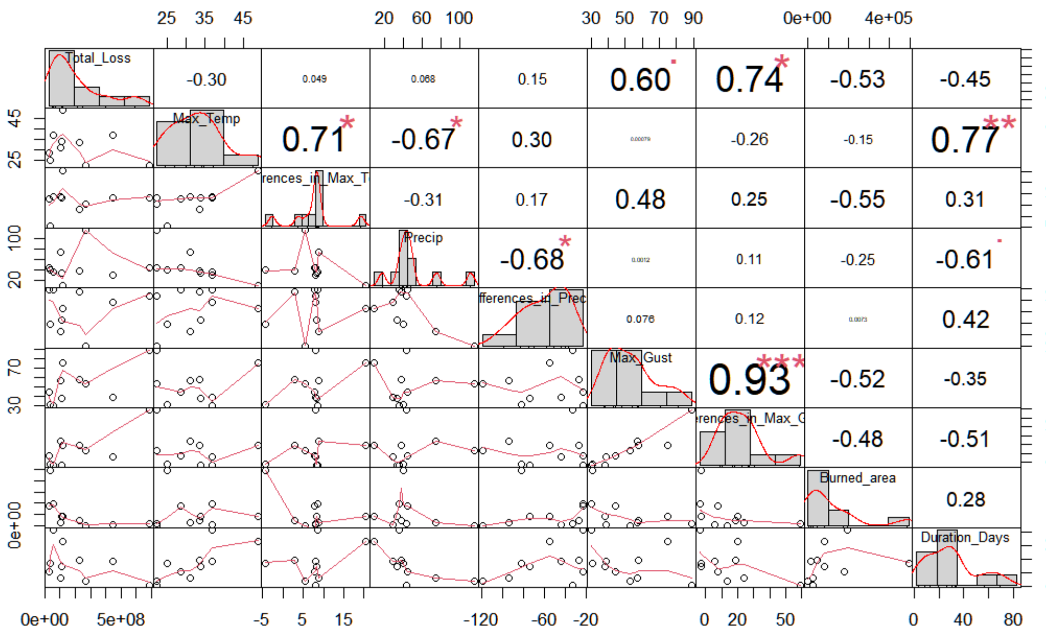

Figure 2.6: Correlation Analysis

From the results, there is a statistically significant correlation between the amount of loss and wind power, but there is an opposite relationship between the amount of loss and burned area and duration. The relationship between temperature and precipitation is not significant. From this analysis, it appears that the results do not align with common sense. Considering that the Pearson correlation coefficient does not consider relationships beyond linear relationships, the magnitude of the change in loss amount is much larger than the related variables. After this thesis, we will try to analyze such data from other perspectives.

Name	Province	Year	Month
Slave Lake Wildfire	AB	2011	May
Fort McMurray Wildfire	AB	2016	May
Williams Lake Wildfire	BC	2017	July
Elephant Hill Wildfire	BC	2017	July
Lytton Creek Wildfire	BC	2021	June
White Rock Lake Wildfire	BC	2021	August
Tantallon Wildfire	NS	2023	May
Kelowna Wildfire	BC	2023	August
Bush Creek East Wildfire	BC	2023	August
Hay River Wildfire	NT	2023	August
Behchoko' Yellowknife Wildfire	NT	2023	August

Table 2.2: Wildfire losses analyzed in this thesis

2.4.3 Other graphs and tables

Table 2.2 lists the wildfire disasters that this thesis analyzes. This thesis only analyzes insurance losses exceeding CAD 25 million since 2008. Most wildfires occurred in British Columbia and Alberta. There were 5 extreme wildfires that occurred in 2023. All wildfire losses listed occurred from May to August.

2.5 Some fundamental statistical analysis

Figure 2.7 shows the loss distribution of wildfire catastrophe events in Canada that have insurance losses exceeding 25 million CAD insurance claims, since 2008. From the histogram, the wildfire loss amount shows a heavy tail distribution and the wildfire in Alberta in May 2016 is the most expensive loss.

2.6 Summary

In this chapter, this thesis first describes the database, CatIQ, which mainly includes loss data and risk exposures. The loss data has been adjusted for inflation. Given that this thesis mainly focuses on wildfire disasters, detailed descriptions of the

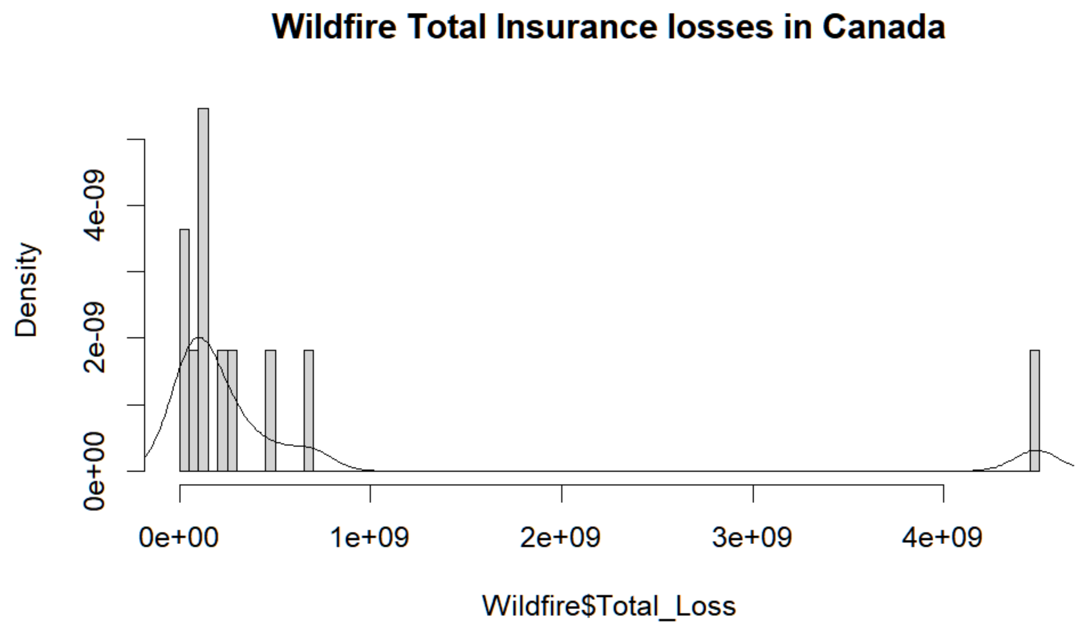

Figure 2.7: Wildfire Total Insurance Losses In Canada

11 related wildfire disasters are presented. Later, the eight factors and correlation matrix were described. The last section presented the loss distribution.

3

Regression analysis of wildfire aggregate claims

3.1 Introduction

In this chapter, we apply four normality tests to verify that the data on wildfire losses shows heavy tail distribution characteristics. Then, we select the best distribution among the five heavy tail distributions to predict the loss amount for return years.

3.2 Normality Tests

The assessments for normality can be summarized into two categories: descriptive statistics and theory-driven methods [46]. In Section 3.1, four types of theory-driven methods will be presented: Shapiro–Francia normality test, Anderson-Darling normality test, Lilliefors (Kolmogorov-Smirnov) normality test and Pearson chi-square normality test. In Section 3.2, Descriptive statistics measures, skewness and kurtosis, will be presented.

3.2.1 Shapiro–Francia normality test

Shapiro–Francia normality test statistic W' is a modification of the Shapiro-Wilk W statistic. Both are used to test whether a random sample comes from a normal distribution. Shapiro-Wilk W statistic is defined as follows.

Let $x_1 \leqslant x_2 \leqslant ... \leqslant x_n$ be the n ordered observations from a standard normal distribution. Then,

$$E(x_i) = m_i, i = 1, 2, ..., n \tag{3.2.1}$$

Let $y_1 \leqslant y_2 \leqslant ... \leqslant y_n$ be the n ordered random observations. If y_i is a normal sample, y_i can be expressed as $y_i = \mu + \sigma x_i$. The W test statistic is calculated by:

$$W = \frac{(\sum_{i=1}^{n} a_i y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{3.2.2}$$

where

$$a' = (a_1, a_2, ..., a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{\frac{1}{2}}} \tag{3.2.3}$$

and m' is the row vector of expected values of $x_1 \leqslant x_2 \leqslant ... \leqslant x_n$. m' is the transpose of m. V is the corresponding $n * n$ covariance matrix of the order statistics. V^{-1} is the inverse matrix of V.

The W' test statistic is calculated by [53]:

$$W' = \frac{(\sum_{i=1}^{n} b_i y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{3.2.4}$$

where

$$b' = \frac{m'}{(m'm)^{\frac{1}{2}}} \tag{3.2.5}$$

For the loss data, $W' = 0.44121$ is below the 1% significant level with a P-Value 0.9. Because a small value of W' indicates non-normality [51], there is statistically significant evidence of non-normality.

3.2.2 Anderson-Darling normality test

Let $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ be the n ordered observations, and let $u_i = F(x_i)$. The test statistic [4] is given by

$$W_n^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2i-1)[log u_j + log(1-u_{j+1})] \tag{3.2.6}$$

Our calculated result is 2.5851 which is large enough to reject the hypothesis of a normal distribution at the 1% significant level.

3.2.3 Lilliefors (Kolmogorov-Smirnov) normality test

When the mean and variance are unknown, Lilliefors (Kolmogorov-Smirnov) normality test can be used to test a set of observations to see if they are from a normal population. Given a set of n observations, the test statistic can be calculated by:

$$D = sup_x|F^*(x) - F_n(x)| \tag{3.2.7}$$

where $F^*(x)$ is the cumulative normal distribution function with $\mu = \overline{X}$, the sample mean, and $\sigma^2 = s^2$, the sample variance and $F_n(x)$ is the empirical cumulative distribution. [31].

Our calculated D is 0.38217 for the loss data, exceeding the critical value for a sample of size 11 at 0.1 significant level of 0.1 (the critical value is 0.36866). We reject the hypothesis that the data sets is from a normal population.

3.2.4 Pearson chi-square normality test

Let X_1, ..., X_n be independent and identically distributed random variables. The null hypothesis H_0 is defined as the vector being asymptotically normally distributed.

Define

$$P(X_i \leqslant x \mid H_0) = F(x;\theta), \theta = (\theta_1, ..., \theta_s)^T, x \in R^1 \tag{3.2.8}$$

Let

$$N_j^{(n)} = Card\{i : X_i \in \Delta_i, i = 1, ..., n\} \tag{3.2.9}$$

$$p_i(\theta) = \int_{\Delta_i} dF(x;\theta), j = 1, ..., r \tag{3.2.10}$$

where $Card$ stands for the cardinality (the number of elements) of a set. Δ_i s are non-intersecting grouping intervals. The standard Pearson chi-squared statistic $X_n^2(\theta)$ [6] is calculated by:

$$X_n^2(\theta) = V^{(n)T} V^{(n)} = \sum_{j=1}^{r} \frac{[N_j^{(n)} - np_j(\theta)]^2}{np_j(\theta)} \tag{3.2.11}$$

$$V_j^{(n)} = \frac{N_j^{(n)} - np_j(\theta)}{\sqrt{np_j(\theta)}} \tag{3.2.12}$$

where N_j is the observed frequency in cell j and np_j is the expected frequency in cell j under the null hypothesis.

Our calculated value of the statistic is 25.545, exceeding the critical value with a degree of freedom of 10 and a significant level of 0.01 (with a critical value of 23.209).

Table 3.1 presents the p-value of the five normality tests. The p-values are all extremely small, indicating that the null hypothesis of a normal distribution is rejected. Therefore, subsequent analysis will be based on heavy tail distributions.

Normality Tests	p-value
Shapiro-Francia normality test	3.195×10^5
Anderson-Darling normality test	4.57×10^7
Lilliefors (Kolmogorov-Smirnov) normality test	7.196×10^5
Pearson chi-square normality test	1.187×10^5

Table 3.1: Normality Tests

3.3 Skewness and Kurtosis

Skewness and Kurtosis are two key indicators of tail heaviness. Skewness, defined by (3.3.1), is a measure of a distribution's lopsidedness or lack of symmetry. The third standardized moment is commonly used to quantify it.

$$Skewness = \frac{E(y - Ey)^3}{(Var\ y)^{3/2}} \tag{3.3.1}$$

Kurtosis is a metric used to quantify peakedness or tail heaviness defined by

$$Kurtosis = \frac{E(y - Ey)^4}{(Var\ y)^2} - 3 \tag{3.3.2}$$

The fourth standard moment minus three is usually used to quantify it. A normal distribution can be verified by checking (3.3.2), which is why the "minus 3" is used. Leptokurtic distributions have positive kurtosis, while platykurtic distributions have negative kurtosis [20].

Skewness and Kutosis for the loss data are calculated to be 2.7276 and 5.6815 respectively. Both results indicate significant heavy tail properties and leptokurtic distribution.

3.4 Heavy tail distributions

Actuaries frequently come across situations where the data display heavy tails, indicating that unusual values are more probable than in a normally distributed dataset.

These distributions are often referred to as "fat", "heavy", "thick" or "long" compared to a normal distribution. Due to the fat-tailed distribution, the typical guidelines for estimating parameters no longer hold for large sample sizes. Traditional regression methods typically aim to reduce squared error losses, but in some cases, asymmetrical errors (either low or high) are of greater concern than symmetric errors [8].

The cumulative distribution functions (CDF) of the relevant distributions used in this thesis are:

(1) Log-Logistic distribution[58]:

$$F(x;\alpha,\beta) = \frac{1}{1+(\frac{x}{\alpha})^{(-\beta)}}, x > 0, \alpha, \beta > 0 \tag{3.4.1}$$

(2) Log-normal distribution[59]:

$$F(x;\mu,\sigma) = \Phi\left(\frac{ln(x)-\mu}{\sigma}\right), \sigma > 0 \tag{3.4.2}$$

where Φ is the cumulative distribution function of the standard normal distribution.

(3) Pareto distribution[60]:

$$F(x;\alpha,k) = 1-\left(\frac{k}{x}\right)^{\alpha}, x \geqslant k \tag{3.4.3}$$

(4) Weibull distribution[61]:

$$F(x;\lambda,k) = 1-e^{-(\frac{x}{\lambda})^k}, x \geqslant 0 \tag{3.4.4}$$

(5) Inverse Weibull distribution[57]:

$$F(x;\lambda,k,m) = e^{-(\frac{x-m}{\lambda})^{-k}}, x > m \tag{3.4.5}$$

Distributions	Shape Parameter	Scale Parameter	AIC	BIC
Log-logistic distribution	1.2781	1.6003×10^8	460.7075	461.5033
Log-normal distribution	19.0074 (meanlog)	1.3991(sdlog)	460.7667	461.5625
Pareto distribution	1.2542	2.2266×10^8	461.4329	462.2286
Weibull distribution	0.6405	3.9807×10^8	464.5707	465.3665
Inverse Weibull distribution	2.5399	1.6363×10^7	478.3827	479.1785

Table 3.2: Statistical results of heavy tail distributions

3.5 Summary of statistical results

James (2023) points out that Akaike information criterion (AIC) and Bayesian information criterion (BIC) can be used to evaluate the quality of a model.

The AIC criterion is defined for a large class of model fitting by the maximum likelihood

$$AIC = \frac{1}{n}(RSS + 2d\hat{\sigma}^2) \tag{3.5.1}$$

The BIC is derived from a Bayesian point of view and is deined as

$$BIC = \frac{1}{n}(RSS + log(n)d\hat{\sigma}^2) \tag{3.5.2}$$

Where n is the number of observations, RSS is the Residual Sum of Squares, d is the number of predictors and $\hat{\sigma}^2$ is an estimate of the variance of the error [21].

The definition of BIC changes the coefficient of AIC before d from 2 to $log(n)$. Considering that the sample size is usually larger than 8, the value of BIC is usually greater than that of AIC. The smaller the values of AIC and BIC, the lower the test error of the model which means a better model fit. The number of losses used in this thesis is only 11, thus the difference between using a natural logarithm of n or 2 is not significant.

Table 3.2 shows the shape parameter, scale parameter, AIC and BIC of the five distributions. The other four distributions, except for the log-normal distribution, are represented by the shape parameter and scale parameter, with the parameters for

Distributions	1-in-100-year wildfire loss
Log-logistic distribution	5,829,098,275
Log-normal distribution	4,659,231,966
Pareto distribution	8,533,277,697
Weibull distribution	4,320,317,835

Table 3.3: Estimated 1-in-100-year wildfire loss

the log-normal distribution being meanlog and sdlog. The results of AIC and BIC are the lowest for Log-logistic distribution. Future analysis in the next section will only use the log-logistic distribution. Considering the largest insurance loss so far, the Fort McMurray wildfire has a loss amount of approximately CAD 4 billion. According to the quantile function of the distributions, this amount of loss is approximately 1 in 100 years.

3.6 Actuarial applications and implications using some measures of risk

The value at risk (VaR) and the conditional tail expectation (CTE) are two commonly used risk measures. Value at risk (VaR) is initially applied in the financial field, where the downside risk of assets, especially the scale of losses in extreme situations, is considered. In actuarial science, it refers to the maximum possible loss of a disaster in a specific future period at a certain probability level (confidence level) given by

$$VaR_p(X) = \pi_p \tag{3.6.1}$$

$$P(X \geqslant \pi_p) = p \tag{3.6.2}$$

When analyzing catastrophe risk in insurance companies, excess probability (EP) and return period are commonly used expressions to indicate the amount of loss occurred. EP curves provide information on the probability of experiencing a loss equal to or exceeding a certain size within a specific years. A return period offers an alternative means of conveying the yearly EP probability and indicates the estimated chance of a loss of a particular amount happening over a specified years. For instance, a return period of 50 years suggests that a specific event or scenario will occur, on average, once in a series of 50-year intervals. The method to transition between these two measures is as follows [29]:

$$\text{Loss Return Period} = \frac{1}{\text{Exceedance Probability}} \tag{3.6.3}$$

Figure 3.1 shows the insurance loss amount of Canadian wildfires based on log-logistic distribution. The first graph shows insurance losses under different exceedance probabilities, and the second graph shows the return periods corresponding to different loss amounts.

3.7 limitations of AIC and BIC

Both AIC and BIC consider the error sum of squares of all data, while the tail risks may impact more for insurance companies. Due to the lack of extreme wildfire claims data, especially losses exceeding CAD 1 billion, AIC and BIC may not be the best benchmarks for the goodness-of-fit test in this situation. As more extreme wildfires occurred during the global warming trend, more precise distributions may be acquired in the future.

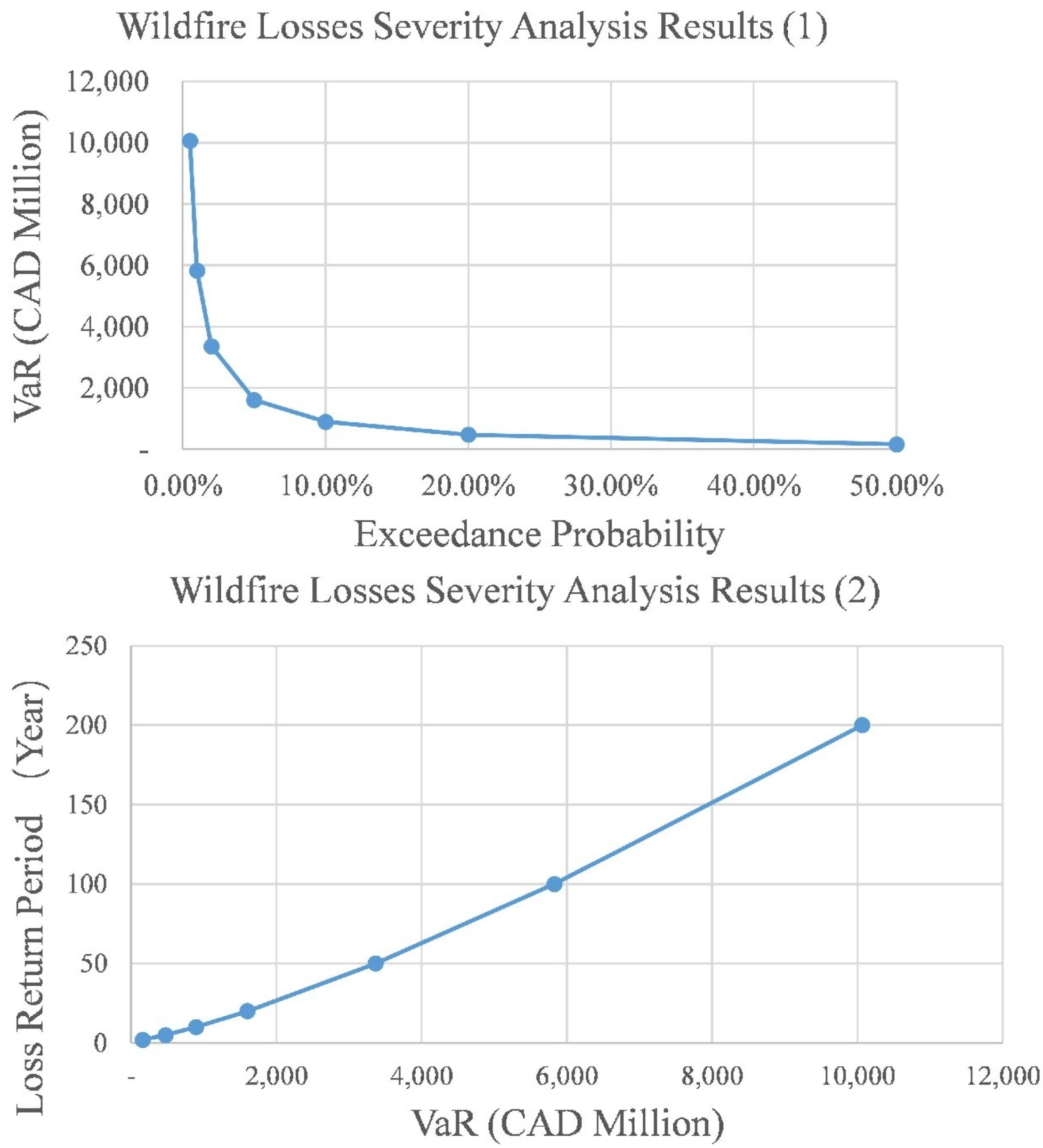

Figure 3.1: Wildfire Losses Severity Analysis Results

3.8 Mean excess loss and Generalized Pareto Distribution (GPD)

Mean excess loss (MEL) of x, denoted $e(d)$, is a function of the excess threshold d.

$$e(d) = E[X - d|X > d] \tag{3.8.1}$$

The MEL for the extreme wildfire loss data is shown in the following figure:

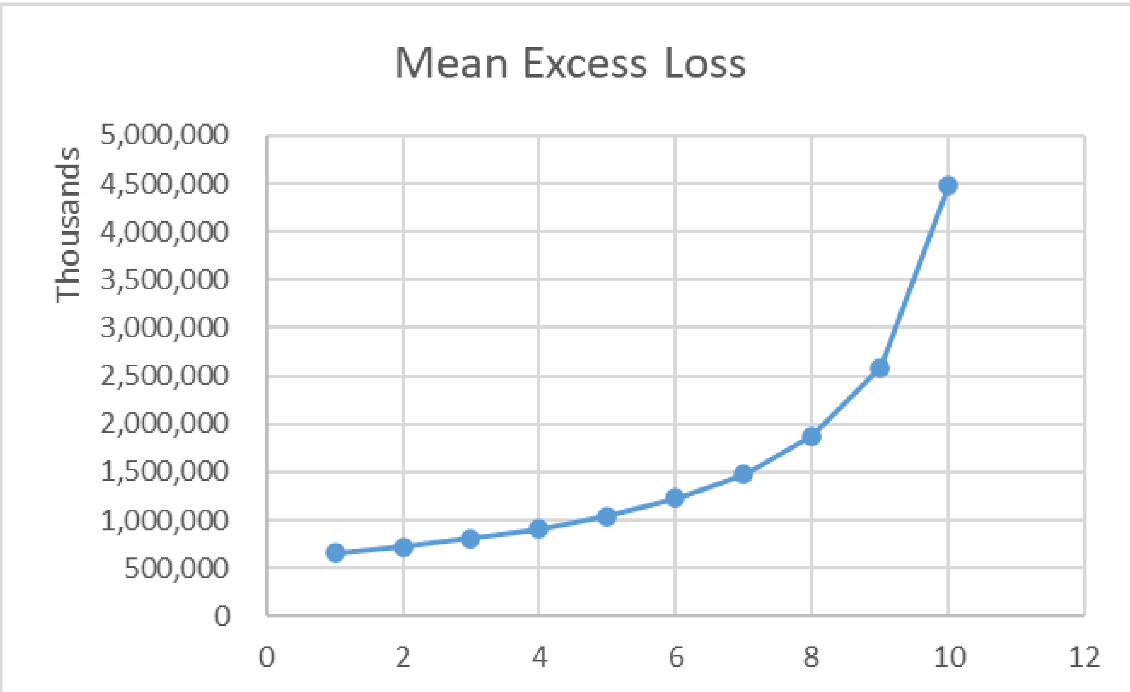

Figure 3.2: Mean excess loss for extreme wildfire

Generalized Pareto distribution (GPD) is a general asymptotic distribution for excess losses. For the GPD distribution, the MEL function is a straight line. If $\beta > 0$, the distribution function of the generalized Pareto distribution (GPD) is

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \xi \neq 0, \\ 1 - e^{-x/\beta} & \xi = 0, \end{cases} \tag{3.8.2}$$

Where $x \geqslant 0$ for $\xi \geqslant 0$ and $0 \leqslant x \leqslant -\beta/\xi$ for $\xi < 0$. β is a scale parameter and ξ is the shape parameter. [23].

Due to the lack of relevant loss, the threshold is set at 25,000,000, the minimum amount of losses included in CatIQ's database. The results of AIC and BIC for GPD are 462.4189 and 466.4189 respectively. The result of the 99th percentile is 8,653,848,075, much larger than the four distributions previously analyzed.

3.9 Summary

In this chapter, we have analyzed the loss data of wildfire accidents. First, we perform a normality analysis on the loss data and calculate skewness and kurtosis. Then, based on the significant heavy tail properties of the loss data, five heavy tail distributions are analyzed and the one with the best fit to the data is selected for simulation analysis.

4

Logistic regression analysis of wildfire claims

4.1 Introduction

In this chapter, we introduce basic concepts and assumptions of logistic regression. Then, we predict the probability of extreme wildfire losses under different temperatures using logistic regression.

4.2 Discussion on factors

Maximum temperature is the main factor included in the previous analysis due to accessibility and correlation with the extreme loss occurrence. According to the Canadian Forest Fire Weather Index (FWI) System, relative humidity, wind speed, and 24-hour precipitation may also be considered. However, we have not found any reliable data source of relative humidity. When searching data from the Government of Canada, most 24-hour precipitation data is 0. If extend the period to 30 days before the occurrence of wildfire, the null hypothesis that the true coefficient of precipitation is zero should not be rejected because of the high P-value (0.848).

For wind speed data, there is almost no relationship between the occurrence of extreme wildfire and the monthly maximum wind speed. The P-value is as high as 0.957, much higher than that of monthly maximum temperature (less than 0.1) and similar to the results of precipitation.

4.3 Basic concepts related to logistic regression

By fitting data to a logistic curve, logistic regression, also known as the logistic model or logit model, estimates the likelihood that an event will occur by analyzing the relationship between several independent variables and a categorical dependent variable. Binary logistic regression and multinomial logistic regression are the two main types of logistic regression models. When the independent variables are either continuous or categorical and the dependent variable is dichotomous, binary logistic regression is commonly employed. A multinomial logistic regression can be used when the dependent variable is not dichotomous and consists of more than two categories.

The probability that an event will occur divided by the probability that it won't occur is known as the odds of an event. The probability that an event won't occur is $(1-p)$ if the event has a probability p of happening. The value represents the corresponding odds is given by

$$Odds\ of\ Event = \frac{p}{1-p} \tag{4.3.1}$$

Logistic regression computes the likelihood of an event happening as opposed to the likelihood that it won't, so odds are typically used to explain the influence of independent variables. If $p = \alpha + \beta x$ is used to model the mean of the response variable p in terms of an explanatory variable x, then , extreme values of x will yield values of $\alpha+\beta x$ that do not lie between 0 and 1, so this model is not good. Peng, Lee, and Ingersoll (2002) state that the odds are transformed using the natural logarithm as the logistic regression solution to this problem. The natural log odds are modeled as a linear function of the explanatory variable using logistic regression as

$$logit(y) = ln(odds) = ln\left(\frac{p}{1-p}\right) = \alpha + \beta x \tag{4.3.2}$$

Where x is the explanatory variable and p is the probability of the desired outcome and α and β are the logistic regression parameters. This is the simple logistic model. An equation for predicting the likelihood of an interested outcome occurring can be derived by taking the antilog of the equation on both sides [48], as

$$p = P\left(Y = \frac{\text{interested outcome}}{X = x}, \text{a specific value}\right) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{1}{1+e^{-(\alpha+\beta x)}} \tag{4.3.3}$$

For data that are binary coded (0 for failure and 1 for success), logistic regression is a technique to fit a regression curve, $y = f(x)$. Logistic regression fits a logistic curve to the relationship between x and y when the response is a binary (dichotomous) variable and x is a numerical value [48]. Logistic is the simplest mathematical function that produces an S-curve, a visual symbol of cumulative growth [28], by

$$y = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{1}{1+e^{-(\alpha+\beta x)}} \tag{4.3.4}$$

where α and β are the logistic intercept and slope. The function above can be extended to the function as [48]

$$y = \frac{e^{x}}{1+e^{x}} = \frac{1}{1+e^{-x}} \tag{4.3.5}$$

4.4 Assumptions of logistic regression

Logistic regression differs from linear regression models in that it does not rely on several key assumptions such as linearity between variables, normal distribution of errors, homoscedasticity, and measurement level of variables. It can handle non-linear relationships between dependent and independent variables by using a non-linear log transformation. The distribution of errors does not need to be multivariate normal, although this may lead to a more stable outcome. Errors can have varying variances based on the level of independent variables. Logistic regression can accommodate both continuous and discrete data for independent variables [47].

Nonetheless, some assumptions remain relevant. First, logistic regression assumes that the dependent variable is discrete and predominantly dichotomous. This means that the observations under consideration must fall into distinct categories.

Second, given that logistic regression is concerned with estimating the probability of a specific event occurring ($P(Y = 1)$), it is crucial to appropriately encode the dependent variable to reflect this. Thus, the desired outcome should be represented by the numerical value 1 in the dataset.

Third, the logistic regression model must be fitted accurately to the data at hand. Overfitting, where extraneous variables are included leading to a model that performs well on the training data but poorly on the test data, must be avoided. Conversely, underfitting should also be prevented, which occurs when meaningful variables are not incorporated into the model, resulting in poor predictive performance.

Fourth, logistic regression necessitates that each observation in the dataset is independent of the others. Additionally, the model should exhibit minimal or no multicollinearity, indicating that the independent variables are not closely related or linear functions of each other. In essence, the variables should provide unique and independent information to the model without redundancy or overlap.

Fifth, logistic regression does not necessitate a linear relationship between the dependent and independent variables. Instead, it requires that the independent variables are linearly associated with the log odds of an event. Moreover, logistic regression mandates large sample sizes due to the reduced power of maximum likelihood estimates compared to ordinary least squares used in linear regression models for estimating unknown parameters [47].

4.5 Application to insurance losses in wildfires

Consecutive daily observations of temperature, relative humidity, wind speed, and 24-hour precipitation are the primary sources of data for the Canadian Forest Fire Weather Index (FWI) System [10]. The variable that most significantly affects the likelihood of wildfires is the maximum temperature. The relationship between the maximum temperature and the insurance losses from large-scale wildfires is analyzed using logistic regression in this study.

In this section, 1 means that there was a wildfire accident in a given month that resulted in an insurance loss greater than CAD 25 million, and 0 means that there was no such accident in that particular month. The average value of the highest temperature recorded by each meteorological station for the current month is downloaded from Canadian government data that is made publicly available [43]. The data only covers the temperatures in British Columbia and Alberta from 2011 to 2023 since these two provinces have the majority of related wildfires and the earliest wildfires in CatIQ data happened in 2011.

In the analysis encompassing the six data points detailing wildfire disasters, the mean highest temperature spans a spectrum from 23.3°C to 37.1°C, whereas in scenarios no such wildfire accidents, the average highest temperatures range between 23.0°C and 32.4°C. In June 2021, British Columbia province experienced an unprecedented heatwave of high temperatures. The provincial coroners' service confirmed

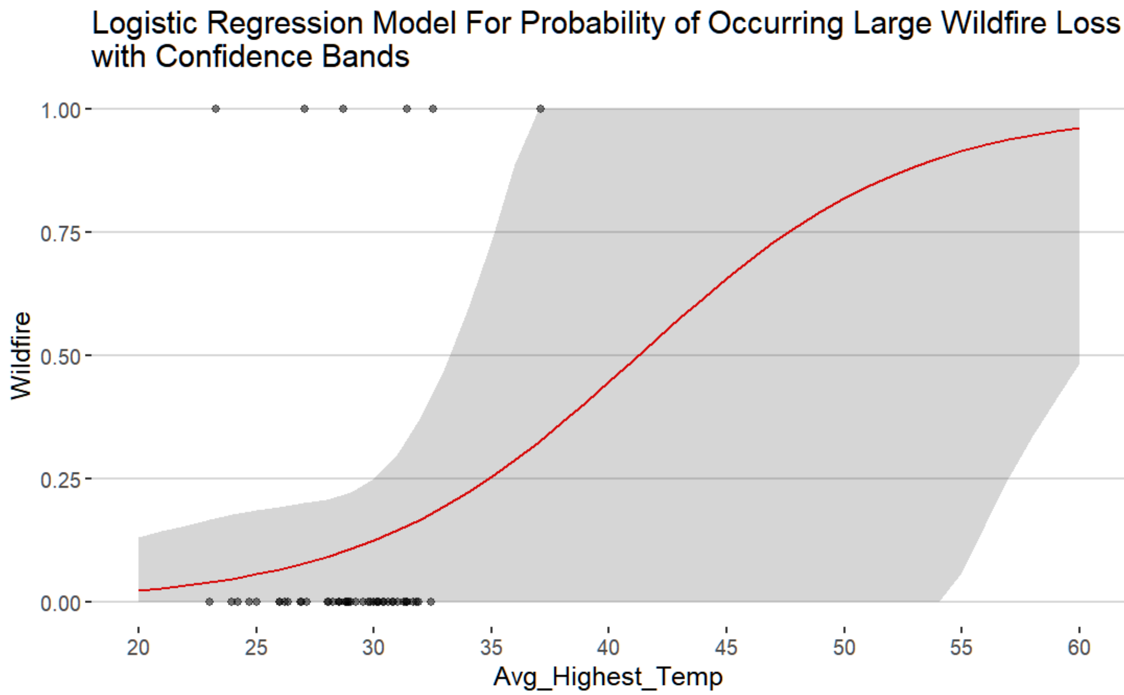

Figure 4.1: Logistic Regression Model For Probability of Occurring Large Wildfire Loss with Confidence Bands

over 600 fatalities due to the severe heat, with 93 percent happening between June 25 and July 1. Lytton experienced record-breaking Canadian heat, registering a temperature of 49.6°C on Tuesday, June 29, 2021 [41].

The rationale behind choosing a confidence interval of 99.5% primarily stems from the stipulations set forth by Solvency II concerning the Solvency Capital Requirement (SCR). With such a confidence level, the insurance entity would be prepared to endure a catastrophe loss that occurs once every 200 years within a single year [14].

From the upper limit of the confidence interval, the likelihood of a wildfire accident with an insurance loss amount exceeding CAD 25 million increases significantly after the temperature exceeds 30°C. In May 2024, a major wildfire disaster occurred in the Fort McMurray forest area in Alberta province. The Government of Alberta listed this wildfire disaster as the highest extreme danger from May 10th to May 16th [39]. The area of fire has exceeded 20,000 hectares. Figure 4.2 shows the burnt area from May 10 to May 16. At the start of this wildfire, the highest temperature

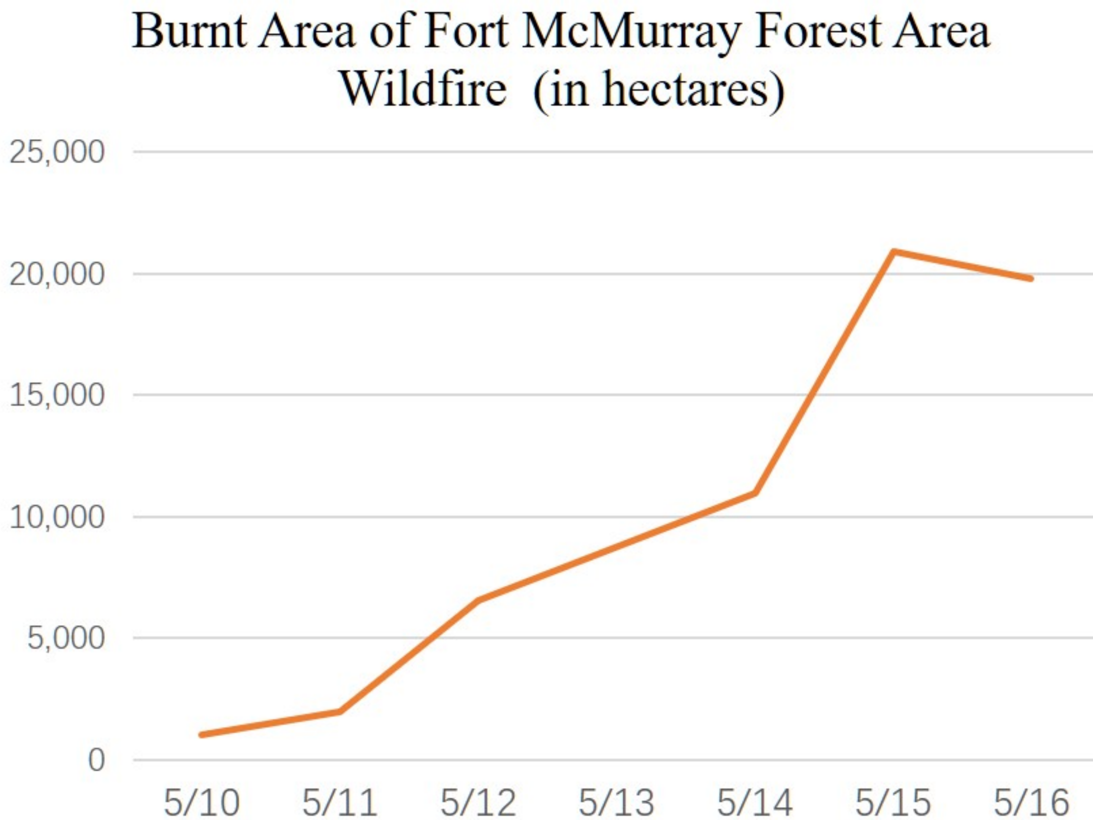

Figure 4.2: Burnt Area of Fort McMurray Forest Area Wildfire

was 28°C on May 10th. According to the previous model results, there is a 99.5% confidence that the probability of occurrence of wildfire with losses exceeding CAD 25M is less than 22%. The final result is consistent with the model results. On May 17th, the wildfire was brought under control and the wildfire danger was listed as low [39].

4.6 Summary

We first introduce the basic concept and five assumptions of logistic regression. Then, logistic regression is used to study the relationship between wildfire loss data and maximum temperature. Given that most wildfire losses occurred in British Columbia and Alberta, data from these two provinces is used.

5

Conclusion and future research

5.1 Conclusion

The results of this thesis can serve as a reference for insurance pricing. First, the insurance losses from wildfires exhibit clear heavy-tail distribution characteristics. However, given the current lack of relevant loss data, further adjustments are needed for the fitting and simulation analysis of the distribution. Then, according to the results of logistic regression, the likelihood of occurring wildfires exceeding CAD 25 million rapidly increases when the maximum temperature exceeds 30°C.

5.2 Future research in property and casualty insurance

Given the ongoing global warming trend, the likelihood of large-scale wildfires occurring is likely to increase. Related research will lead to more accurate predictions when more data are becoming available. If there is sufficient data available, small regional studies could be conducted, rather than relying on average maximum temperatures for BC and AB, as was used in this research. Moreover, large-scale wildfires also occurred in Nova Scotia and Northwest Territories in 2023. If accidents occur more frequently in areas other than British Columbia and Alberta in the future, other regions could also be included in the analysis of the probability of extreme wildfires.

Wildfires may be man-made, such as not extinguishing fires promptly during camping. It is also possible that it is caused by nature, such as lightning strikes. If the losses of wildfires are analyzed separately based on their causes, more accurate loss estimation may be predicted. Government of Alberta proposed that 67% of wildfires are human-caused [40]. However, it is difficult to classify the causes of a wildfire disaster. For example, the Fort McMurray wildfire, which can only be said to be suspected of human arson, was initially discovered by a helicopter forest crew 15 kilometers away from Fort McMurray [37].

When pricing, it is possible to consider combining wildfire disaster data with insurance coverage information to estimate the additional premium charged due to wildfire disasters. According to the damage ratio ($\frac{Loss\ amount}{Exposure\ data}$) calculated based on the exposure data in the CatIQ database, except for the Fort McMurray wildfire, the approximate range is between 0.2% and 1.3%. However, given the small number of such disasters in our study, it is difficult to find a stable trend. Further analysis can be conducted after collecting more data in the future, and provide a reference for insurance pricing in areas prone to wildfire disasters.

Using CPI data may not accurately reflect the trend of insurance claims. For instance, food makes up more than 10% of the CPI, while reading, recreation, and schooling make up roughly 10%. Insurance claims for losses resulting from wildfires are not particularly related to the aforementioned categories [9]. A report by Verisk states that many insurance companies have given careful consideration to reconstruction costs, including material costs and labor costs [38]. Only 35 to 40 percent of CPI is accounted for by shelter and household operations, furnishings, and equipment. To more precisely adjust the loss amount for previous years, these may be taken into

Layer	Limit	Deductible	Burning Cost Price		Ratio
			All losses	Wildfire Only	
1	50,000,000	50,000,000	23,769,595	8,321,002	35%
2	100,000,000	100,000,000	20,301,515	16,642,004	82%
3	100,000,000	200,000,000	16,244,463	16,244,463	100%
4	500,000,000	300,000,000			
5	1,000,000,000	800,000,000			

Table 5.1: Burning cost results of a Catastrophe excess of loss treaty

consideration in future research to modify the percentage of different costs [9].

5.3 Future research in reinsurance

In the reinsurance industry, purchasing catastrophe excess of loss treaty is the main way for insurance companies to protect aginst catastrophic losses.

Burning cost is one of the basic pricing methods for catastrophe excess of loss treaty. It is to calculate the loss amount of each layer based on the loss data, which is pure premium. In practice, loss data is first converted to the current year based on the inflation rate and premium growth rate. If there is only industry data, it is also necessary to consider the company's market share. Afterward, the losses covered by proportional treaty and facultative reinsurance will be reduced. Finally, we can add more factors, such as brokerage fees [56].

This thesis takes a Canadian insurance company as an example. The company has purchased a five-layer catastrophe excess of loss treaty with an deductible of CAD 50 million and a limit of CAD 1.8 billion. From Table 5.1, a large part of the price for the catastrophe excess of loss treaty is due to wildfire accidents.

The design of the excess of loss treaty layer requires consideration of multiple disasters, not just wildfire disasters. Therefore, more research can consider multiple causes of disasters to design and price the excess of loss treaty layer. Shree and Keven (2021) use occurred losses and order statistics to price catastrophe excess of loss treaties. Poisson distribution and gamma distribution are used to analyze frequency and severity respectively, and different parameters are applied to simulate losses over 500,000 years for different disasters such as hurricanes, earthquakes, wildfires, etc. And, different return periods are used to set the catastrophe excess of loss treaty layer. Some catastrophe data from the United States have been studies, which can also be applied to analyze catastrophes in Canada. They assume that frequency and severity are independent. However, in the context of global temperature changes, the frequency and severity of accidents may increase synchronously [27].

5.4 Future research in life insurance

The wildfire smoke includes various gases, particles and water vapour, containing ozone, methane, sulfur dioxide, nitrogen dioxide, and PM2.5 [44]. In terms of sales, temporarily elevated air pollution levels result in an increased need for health insurance [12].

In terms of losses, research released by Health Canada and other institutions in 2020 estimated the impact of wildfire smoke on Canada's entire country between 2013 and 2015, as well as for 2017-2018. From their research, British Columbia and Alberta are the provinces most affected by the smoke generated by wildfires in Canada in most years [35]. This result is consistent with the scenario studied in this thesis.

The analysis estimated annual mortality and non-fatal cardiorespiratory health outcomes of 54 to 240 short-term deaths and 570 to 2,500 long-term premature deaths. The economic loss due to health consideration during the study period was estimated to be between $410 and $18 million per year for acute health impacts and between $43 and $19 million per year for chronic health impacts [35]. Considering that such risks are similar to the hazards of smoking, insurance companies use differ-

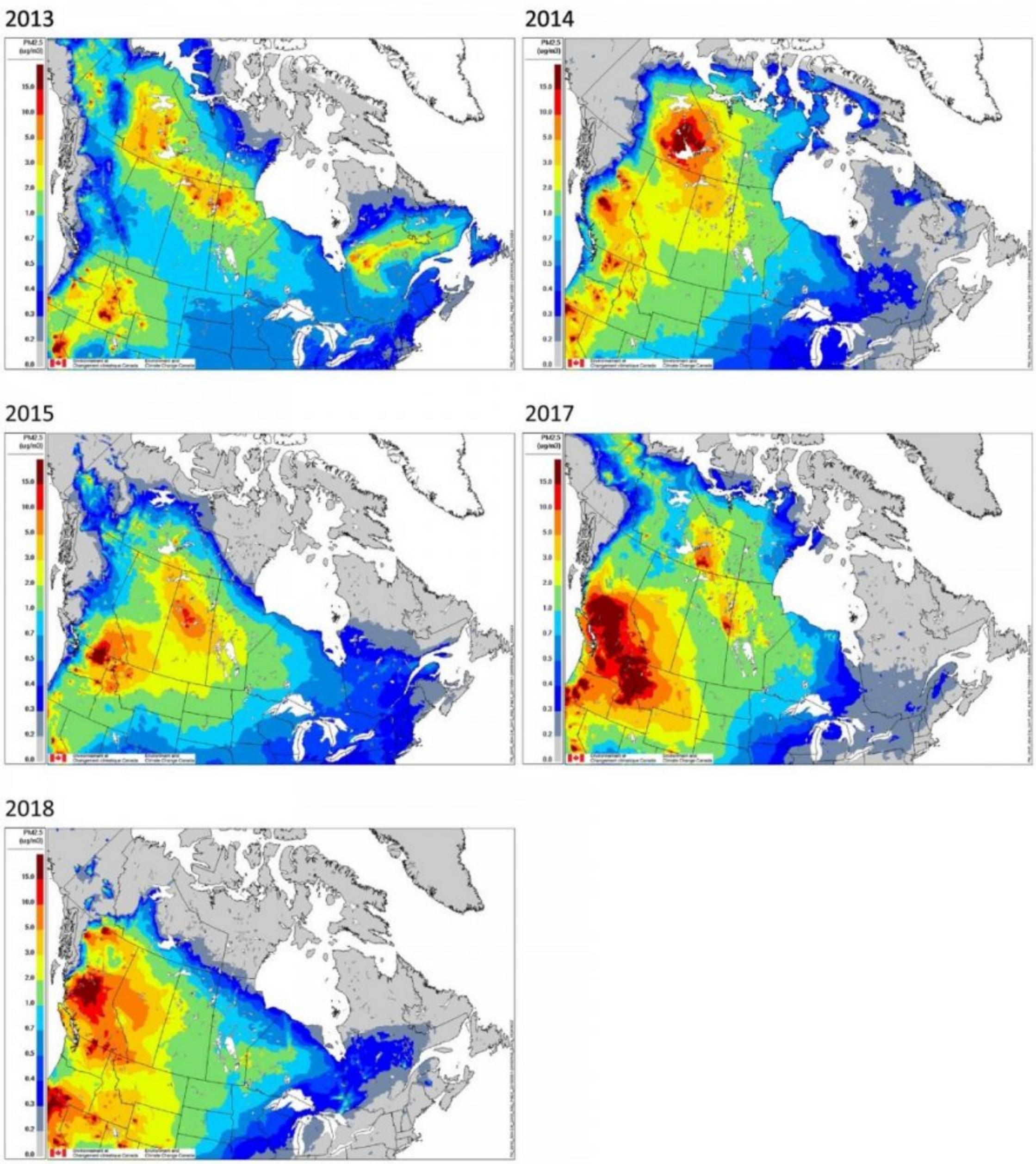

Figure 5.1: PM2.5 concentrations (May to September) attributable to wildfires for 2013–2015 and 2017–2018 based on FireWork and GEM-MACH [35].

ent mortality tables for smokers and non-smokers to set premiums. Future research can consider attempting to price life insurance by weighing the mortality rates of smokers and non-smokers in areas prone to wildfires.

Appendix A

R code

A.1 Wildfire Insurance losses in Canada

```
hist(x=Wildfire$Total_Loss,breaks = 20,freq = F,main="Wildfire Insurance losses
in Canada (CAD)")
lines(density(Wildfire$Total_Loss))
```

A.2 Correlation Matrix

```
library("PerformanceAnalytics")
Wildfire_Total <- Wildfire[, c(17,6:13)]
chart.Correlation(Wildfire_Total)
corrgram (Wildfire_Total, order = TRUE , lower.panel=panel.conf)
```

A.3 Q-Q Plot for wildfire loss and log of wildfire loss

```
qqnorm(Wildfire$Total_Loss, pch = 1, frame = FALSE)
```

qqline(Wildfire$Total_Loss, col = "steelblue", lwd = 2)

qqnorm(log(Wildfire$Total_Loss), pch = 1, frame = FALSE)

qqline(log(Wildfire$Total_Loss), col = "steelblue", lwd = 2)

A.4 Normality test on wildfire losses

library(nortest)

sf.test(Wildfire$Total_Loss)

ad.test(Wildfire$Total_Loss)

cvm.test(Wildfire$Total_Loss)

lillie.test(Wildfire$Total_Loss)

pearson.test(Wildfire$Total_Loss)

A.5 Skewness and kurosis

library(moments)

print(skewness(Total_Loss))

print(kurtosis(Total_Loss))

library(fitdistrplus) library(actuar)

A.6 Comparison of goodness-of-fit of different heavy-tail distributions

Fit_llogis_Total <- fitdist(Wildfire$Total_Loss, "llogis", start = list(shape = 1, scale = 500))

summary(Fit_llogis_Total)

gofstat(Fit_llogis_Total)

Fit_lnorm_Total <- fitdist(Wildfire$Total_Loss,"lnorm")

summary(Fit_lnorm_Total)

gofstat(Fit_lnorm_Total)

Fit_pareto_Total <- fitdist(Wildfire$Total_Loss, "pareto", start = list(shape = 1, scale = 1))

summary(Fit_pareto_Total)

gofstat(Fit_pareto_Total)

Fit_weibull_Total <- fitdist(Wildfire$Total_Loss, "weibull")

summary(Fit_weibull_Total)

gofstat(Fit_weibull_Total)

Fit_invweibull_Total <- fitdist(Wildfire$Total_Loss,'invweibull')

summary(Fit_invweibull_Total)

gofstat(Fit_invweibull_Total)

gofstat(list(Fit_llogis_Total,Fit_lnorm_Total, Fit_pareto_Total, Fit_weibull_Total, Fit_invweibull_Total))

A.7 Value at risk (VaR) of loglogistics distributions

library(fitdistrplus)

library(actuar)

qllogis(0.5, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

qllogis(0.2, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

qllogis(0.1, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE) qllogis(0.05, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

qllogis(0.02, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

qllogis(0.01, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

qllogis(0.005, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

qllogis(0.002, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

qllogis(0.001, shape = 1.278109e+00, scale = 1.600312e+08, lower.tail = FALSE, log.p = FALSE)

A.8 Logistic Regression Model for Probability of Occurring Large Wildfire Loss with Confidence Bands

library(tidyverse)

library(ggthemes)

set.seed(123)

my_data1 = data.frame(Avg_Highest_Temp = rnorm(50,30.14,21.87623), Wildfire = sample(c(0,1), prob = c(0.7,0.3), replace=TRUE, size= 50))

my_data2 = data.frame(Avg_Highest_Temp = rnorm(50,30.14, 21.87623), Wildfire = sample(c(0,1), prob = c(0.3,0.7), replace=TRUE, size= 50))

my_data = rbind(my_data1, my_data2)

model <- glm(Wildfire   Avg_Highest_Temp, data = ABBCWeather, family =

```
binomial())
preds <- data.frame(Avg_Highest_Temp = seq(20,60), length.out = 100)
preds$pred <- predict(model, preds, type = "response")
preds$upper <- predict(model, preds, type = "response", se.fit = TRUE)$fit +
2.57 * predict(model, preds, type = "response", se.fit = TRUE)$se.fit
preds$upper <- ifelse(preds$upper>1,1,preds$upper)
preds$lower <- predict(model, preds, type = "response", se.fit = TRUE)$fit -
2.57 * predict(model, preds, type = "response", se.fit = TRUE)$se.fit
preds$lower <- ifelse(preds$lower<0,0,preds$lower)
ggplot(ABBCWeather, aes(x = Avg_Highest_Temp, y = Wildfire)) + theme_hc()+
scale_colour_hc()+
geom_point(alpha = 0.5) +
geom_line(data = preds, aes(x = Avg_Highest_Temp, y = pred), color = "red",
inherit.aes = FALSE) +
geom_ribbon(data = preds, aes(x = Avg_Highest_Temp, ymin = lower , ymax
= upper), alpha = 0.2, inherit.aes = FALSE) +
ggtitle("Logistic Regression Model For Probability of Occurring Large Wildfire
Loss Confidence Bands")+
scale_x_continuous(breaks = seq(20, 60, 5))
```

References

[1] Moody's Analytics. North America wildfire HD model suite. Online; url="https://www.rms.com/models/wildfire"; accessed on July 8, 2024.

[2] Verisk Analytics. Ice, fire, wind, and water. Online; url="https://www.verisk.com/resources/campaigns/ice-fire-wind-and-water/"; accessed on July 8, 2024.

[3] Verisk Analytics. Modeling extreme event risk. Online; url="https://www.verisk.com/siteassets/extreme-event-solutions/modeling-extreme-event-risk.pdf"; accessed on July 8, 2024.

[4] Theodore W Anderson and Donald A Darling. A test of goodness of fit. *Journal of the American Statistical Association*, 49(268):765–769, 1954.

[5] AON. Climate and Catastrophe Insight 2024. Online; url="https://assets.aon.com/-/media/files/aon/reports/2024/climate-and-catastrophe-insights-report.pdf"; accessed on July 8, 2024.

[6] Narayanaswamy Balakrishnan, Vassilly Voinov, and Mikhail Stepanovich Nikulin. *Chi-squared goodness of fit tests with applications*. Academic Press, 2013.

[7] Brian Beckage and William J Platt. Predicting severe wildfire years in the florida everglades. *Frontiers in Ecology and the Environment*, 1(5):235–239, 2003.

[8] Jan Beirlant, Gunther Matthys, and Goedele Dierckx. Heavy-tailed distributions and rating. *ASTIN Bulletin: The Journal of the IAA*, 31(1):37–58, 2001.

[9] Inflation Calculator. Canadian CPI Basket of Goods and Services. Online; url="https://inflationcalculator.ca/cpi-basket/"; accessed on July 8, 2024.

[10] Natural Resources Canada. Canadian Forest Fire Weather Index (FWI) System. Online; url="https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi"; accessed on July 8, 2024.

[11] Natural Resources Canada. Facts about wildland fires in Canada. Online; url="https://natural-resources.canada.ca/our-natural-resources/forests/wildland-fires-insects-disturbances/forest-fires/13143"; accessed on July 8, 2024.

[12] Tom Y Chang, Wei Huang, and Yongxiang Wang. Something in the air: Pollution and the demand for health insurance. *The Review of Economic Studies*, 85(3):1609–1634, 2018.

[13] Marc Dahling and Gerry Hopkins. Underwriting Wildfire – Creating Your Hazard Map. Online; url="https://www.genre.com/content/dam/generalreinsuranceprogram/documents/iinapc1909-en.pdf.coredownload.inline.pdf"; accessed on July 8, 2024.

[14] RBC Dependencies and Calibration Working Party. Solvency ii standard formula and naic risk-based capital (rbc). In *CAS E-Forum*, volume 2, pages 1–38, 2012.

[15] Gloria Dickie. The Arctic and climate change (1979 – 2019): What the ice record tells us. Online; url="https://news.mongabay.com/2019/09/the-arctic-

and-climate-change-1979-2019-what-the-ice-record-tells-us/"; accessed on July 8, 2024.

[16] Amy Dusto. Reading between the tides: 200 years of measuring global sea level. Online; url="https://www.climate.gov/news-features/climate-tech/reading-between-tides-200-years-measuring-global-sea-level"; accessed on July 8, 2024.

[17] NOAA National Centers for Environmental Information. Equatorial Pacific Sea Surface Temperatures (SST). Online; url="https://www.ncei.noaa.gov/access/monitoring/enso/sst"; accessed on July 8, 2024.

[18] NOAA National Centers for Environmental Information. Global climate summary for January 2024. Online; url="https://www.climate.gov/news-features/understanding-climate/global-climate-summary-january-2024"; accessed on July 8, 2024.

[19] NOAA National Centers for Environmental Information. Southern Oscillation Index (SOI). Online; url="https://www.ncei.noaa.gov/access/monitoring/enso/soi"; accessed on July 8, 2024.

[20] Edward W Frees. *Regression modeling with actuarial and financial applications.* Cambridge University Press, 2009.

[21] James Gareth, Witten Daniela, Hastie Trevor, and Tibshirani Robert. *An introduction to statistical learning: with applications in R.* Spinger, 2023.

[22] The World Bank Group. Inflation, GDP deflator (annual %) - Canada. Online; url="https://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG?locations=CA"; accessed on July 8, 2024.

[23] Mary R Hardy and David Saunders. *Quantitative enterprise risk management*. Cambridge University Press, 2022.

[24] Castastrophe Indices and Quantification Inc. CatIQ about us. Online; url="https://public.catiq.com/catiq-about/"; accessed on July 8, 2024.

[25] Castastrophe Indices and Quantification Inc. CatIQ Data. (2024).

[26] Castastrophe Indices and Quantification Inc. CatIQ products. Online; url="https://public.catiq.com/products/"; accessed on July 8, 2024.

[27] Shree Khare and Keven Roy. Quantifying the role of occurrence losses in catastrophe excess of loss reinsurance pricing. *Risks*, 9(3):52, 2021.

[28] Dmitry Kucharavy and Roland De Guio. Application of sshaped curves. *Procedia Engineering*, 9:559--572, 2011.

[29] David Lalonde and Alissa Legenza. Modeling Fundamentals: Evaluating Risk Measures. Online; url="https://www.verisk.com/blog/Modeling-Fundamentals–Evaluating-Risk-Measures/"; accessed on July 8, 2024.

[30] Katherine Latham, Aimee Colgate, Elizabeth Doocey, and Scott St George. Quantifying and managing wildfire risks. Online; url="https://www.wtwco.com/-/media/wtw/insights/2023/12/get-ahead-of-wildfire-risks-in-2024-strategies-for-mitigation-management-and-measurement.pdf?modified=20231206155604"; accessed on July 8, 2024.

[31] Hubert W Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318):399–402, 1967.

[32] Rebecca Lindsey and Luann Dahlman. Climate Change: Global Temperature. Online; url="https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature"; accessed on July 8, 2024.

[33] Rebecca Lindsey and Luann Dahlman. Climate Variability: North Atlantic Oscillation. Online; url="https://www.climate.gov/news-features/understanding-climate/climate-variability-north-atlantic-oscillation"; accessed on July 8, 2024.

[34] Spiedr Ltd. The Life Cycle Of A Wildfire. Online; url="https://www.spiedr.com/blog/wildfire-life-cycle/"; accessed on July 8, 2024.

[35] Carlyn J Matz, Marika Egyed, Guoliang Xi, Jacinthe Racine, Radenko Pavlovic, Robyn Rittmaster, Sarah B Henderson, and David M Stieb. Health impact analysis of pm2. 5 from wildfire smoke in canada (2013–2015, 2017–2018). *Science of The Total Environment*, 725:138506, 2020.

[36] Meng Meng and Daoyi Gong. Winter north atlantic sst as a precursor of spring eurasian wildfire. *Geophysical Research Letters*, 49(18):e2022GL099920, 2022.

[37] CBC News. Someone likely sparked the Fort McMurray wildfire, but was it a crime? RCMP ask. Online; url="https://www.cbc.ca/news/canada/edmonton/fort-mcmurray-wildfire-cause-investigation-rcmp-1.3635241"; accessed on July 8, 2024.

[38] CBC News. Verisk 2021 wildfire risk analysis provides updated data on rising threat. Online; url="https://www.verisk.com/blog/verisk-2021-wildfire-risk-analysis-provides-updated-data-on-rising-threat/"; accessed on July 8, 2024.

[39] Government of Alberta. Fort McMurray Area Update. Online; url="https://srd.web.alberta.ca/fort-mcmurray-area-update/all"; accessed on July 8, 2024.

[40] Government of Alberta. Wildfire prevention and mitigation. Online; url="https://www.alberta.ca/wildfire-prevention-and-mitigation"; accessed on July 8, 2024.

[41] Government of Canada. Surviving the heat: The impacts of the 2021 western heat dome in Canada. Online; url="https://science.gc.ca/site/science/en/blogs/science-health/surviving-heat-impacts-2021-western-heat-dome-canada"; accessed on July 8, 2024.

[42] Insurance Bureau of Canada. Okanagan and Shuswap area wildfires cause over $720 million in insured damage. Online; url="https://www.ibc.ca/news-insights/news/okanagan-and-shuswap-area-wildfires-cause-over-720-million-in-insured-damage"; accessed on July 8, 2024.

[43] Meteorological Service of Canada. Historical Data. Online; url="https://climate.weather.gc.ca/historical_data/search_historic_data_e.html"; accessed on July 8, 2024.

[44] Public Health Agency of Canada. Wildfires in Canada: Toolkit for Public Health Authorities. Online; url="https://www.canada.ca/en/public-health/services/publications/healthy-living/wildfires-canada-toolkit-public-health-authorities.html#a5.4.1"; accessed on July 8, 2024.

[45] Superintendent of Financial Institutions. Financial data for property and casualty companies. Online; url="https://www.osfi-bsif.gc.ca/en/data-forms/financial-data/financial-data-property-casualty-companies"; accessed on July 8, 2024.

[46] Hun Myoung Park. Univariate analysis and normality test using sas, stata, and spss. 2015.

[47] Hyeoun-Ae Park. An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2):154–164, 2013.

[48] ChaoYing Joanne Peng, Kuk Lida Lee, and Gary M Ingersoll. An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1):3--14, 2002.

[49] Jeffrey P Prestemon, Karen Abt, and Krista Gebert. Suppression cost forecasts in advance of wildfire seasons. *Forest Science*, 54(4):381–396, 2008.

[50] United Nations Environment Programme. Spreading like Wildfire: The Rising Threat of Extraordinary Landscape Fires. Online; url="https://www.unep.org/resources/report/spreading-wildfire-rising-threat-extraordinary-landscape-fires"; accessed on July 8, 2024.

[51] JP Royston. A simple method for evaluating the shapiro–francia w' test of non-normality. *Journal of the Royal Statistical Society Series D: The Statistician*, 32(3):297–300, 1983.

[52] Veronika Samborska and Hannah Ritchie. Wildfires - Explore global and country-level data on the extent of wildfires and how they've changed over time. Online; url="https://ourworldindata.org/wildfires"; accessed on July 8, 2024.

[53] Samuel S Shapiro and RS Francia. An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337):215--216, 1972.

[54] Assaf Shmuel and Eyal Heifetz. A machine-learning approach to predicting daily wildfire expansion rate. *Fire*, 6(8):319, 2023.

[55] WR Skinner, A Shabbar, MD Flannigan, and K Logan. Large forest fires in canada and the relationship to global sea surface temperatures. *Journal of Geophysical Research: Atmospheres*, 111(D14), 2006.

[56] Geoff Werner and Claudine Modlin. Basic Ratemaking. Online; url="https://www.casact.org/sites/default/files/old/studynotes_werner_modlin_ratemaking.pdf"; accessed on July 8, 2024.

[57] Wikipedia. Fréchet distribution. Online; url="https://en.wikipedia.org/wiki/Fr%C3%A9chet_distribution"; accessed on July 8, 2024.

[58] Wikipedia. Log-logistic distribution. Online; url="https://en.wikipedia.org/wiki/Log-logistic_distribution"; accessed on July 8, 2024.

[59] Wikipedia. Log-normal distribution. Online; url="https://en.wikipedia.org/wiki/Log-normal_distribution"; accessed on July 8, 2024.

[60] Wikipedia. Pareto distribution. Online; url="https://en.wikipedia.org/wiki/Pareto_distribution"; accessed on July 8, 2024.

[61] Wikipedia. Weibull distribution. Online; url="https://en.wikipedia.org/wiki/Weibull_distribution"; accessed on July 8, 2024.