PROJECTING THE ULTIMATE LOSS OF CATASTROPHIC EVENTS

A Major Qualifying Project Report submitted to the Faculty of WORCESTER POLYTECHNIC INSTITUTE in partial fulfillment of the requirements for the Degree of Bachelor of Science

By:
Siro Fritzmann
Elizabeth Hansen
Amber Munderville
Mike Sullivan
Tyler Turchiarelli

Advisors:
Professor Jon Abraham
Professor Barry Posterro

Project Sponsor:
The Hanover Insurance Group

Table of Contents

List of Figures

List of Tables

Abstract

In collaboration with the Hanover Insurance Group, we worked to develop a predictive model that projects the ultimate losses associated with catastrophes. The model we created is interactive and user-friendly. Users are able to select various characteristics that define the parameters for the model. The printable summary report presents graphical views of the expected development patterns over time and additional information regarding the projections. This model can be updated on a daily basis to ensure accuracy and will ultimately aid the catastrophe reserving process at Hanover.

Acknowledgements

Many parties assisted our Major Qualifying Project and we would like to recognize them. We would first like to express our thanks to Jaris Wicklund, Katelyn Scannell, Andrew Evans, Anthony Giangreco-Marotta, and Rick Burt of the Hanover Insurance Group for their guidance and support throughout the duration of this project. They were always enthusiastic and helpful. We have all learned so much through this project and are sincerely grateful to have had the opportunity to work with you.

Additionally, we would like to thank our project advisors, Professor Jon Abraham and Professor Barry Posterro. They were constantly teaching us new things and helping us to think critically about the project at hand. Their encouragement and wisdom helped us along the way to complete our goal. Their guidance was invaluable to our final product. We would also like to thank Worcester Polytechnic Institute and the Hanover Insurance Group for giving us this learning opportunity.

Executive Summary

Our project focused on catastrophe reserving for The Hanover Insurance Group, a medium-sized property and casualty insurance company based in Worcester, MA. The goal of the project was to help the company improve on its current methodology to predict how many insurance claims would be received following a catastrophic event and the corresponding amount of loss. Currently, the company has an internal process for predicting the fiscal impact of these catastrophes. Their current method is somewhat time-consuming and not entirely accurate as catastrophes, especially weather-related events, are very difficult to predict. Our project focused on creating a spreadsheet model to streamline this predictive process and increase accuracy in developing catastrophe reserves.

The first step we took in this project was to learn more about both The Hanover Insurance Group and about catastrophe modeling in general. The team read actuarial textbooks to understand how to predict claim counts and values when a catastrophe occurs, and used this knowledge when reviewing the data supplied by The Hanover Insurance Group.

The Hanover Insurance Group provided us with two data sets – one detailing claims per catastrophe and one containing estimated costs on a monthly basis per claim. The team sorted through this data and then combined the two sets to allow for one cohesive spreadsheet model. The team created a model which utilizes data corresponding to key features of peril type, line of business, and region. The model creates an average prediction of claim count and storm expense based on this historic data set and produces an output of graphical and numerical projections.

Throughout the duration of the project, we worked almost exclusively in Microsoft Excel and Microsoft Excel's Visual Basic for Applications. The spreadsheet model is contained in one Excel file and can be utilized by The Hanover Insurance Group moving forward.

We recommend the use of the spreadsheet model for each catastrophe. The model can be used on a daily basis for the first few days following a catastrophe, and then utilized monthly after the majority of claims have been received. The model will also need to be updated at least monthly to ensure accuracy in the utilization of most recent catastrophes. We feel the model will be used best by The Hanover Insurance Group actuarial team, and the final output pages can be utilized by anyone in the company as they are easily understandable and have graphical representations which are simple to follow. In the future, The Hanover Insurance Group can choose to expand upon this model by creating more frequent views of claim costs in the data set and by utilizing more specific features when selecting which historical data to utilize in the modeling.

1. Introduction

The insurance industry relies heavily on the predictability of insurable events. Many of these events are relatively easy to prepare for but certain events, known as catastrophes, are less predictable. Catastrophes are generally natural phenomena that can bring about devastating damage. The issue that catastrophes pose for insurance companies is their unpredictable nature. In general, it is difficult to determine when and where a catastrophe will occur, and even more difficult to predict the cost of damages that the event will bring.

In the early days of insurance, the ability to predict catastrophes was extremely limited. However, as technology advanced, insurance companies became better-equipped to make more accurate predictions about catastrophes. Now, in this age of data, the idea of catastrophe modeling has surfaced; this form of modeling leverages vast amounts of historical data in order to predict future events. By matching characteristics of past storms to current storms, insurance companies can determine an accurate measure of the potential losses associated with a given event. These methods help companies in the reserving process. By predicting the potential losses, a company will know how much money to set aside in a reserve in order to pay out potential claims reported by its customers.

Catastrophe (CAT) modeling has ultimately become an essential risk management strategy in the insurance industry. This technique lends itself particularly well to the Property & Casualty (P&C) side of insurance. Since catastrophes can afford major losses on property, P&C insurance companies rely on CAT modeling to mitigate its losses.

The Hanover Insurance Group located in Worcester, Massachusetts is one such P&C insurance company that uses these CAT modeling techniques. The company writes business in many areas of the country which exposes it to many different types of catastrophes. Its current catastrophe methodology delivers its projections at an ultimate basis (i.e. the fully developed point when all claims have been processed). Additionally, the current system requires a research component to match characteristics of the storm and geography, and uses separate Excel files for each evaluation. Since the current system is a bit cumbersome, there is room for improvement. As a result, this project aims to streamline the CAT modeling process at Hanover.

In this report, we will document the details and methods that helped accomplish this goal. Chapter 2 will deliver an in-depth discussion of claim reporting and CAT modeling as well as some insight into our sponsor, the Hanover Insurance Group. In Chapter 3, we will discuss the project objectives and describe the methods we used to complete these objectives. Chapter 4 will explain our results and findings, and Chapter 5 will display our deliverables: the dashboard to the CAT modeling tool complete with relevant documentation. Finally, we conclude our report in Chapter 6 with our final thoughts and recommendations.

2. Background

The following chapter details necessary background information. This section includes overviews of the claim reporting process and methods of predicting claims, as well as an introduction to CATs and CAT modeling. Additionally, this section delivers insight into The Hanover Insurance Group and its current CAT modeling methodology.

2.1 Claims Reporting

The ability to predict claims is what allows insurance companies to make money. The better they can predict claims for various products, the more profit they can bring in. Claims distributions can vary between different products and regions, and can be greatly affected by technological advancements or even political policies. The ability to compile and organize these claims so that they can be easily analyzed is essential to an insurance company.

While many insurance companies handle their own claims, many larger firms use third party administrators (Friedland, 2010). These third party administrators or TPAs are responsible for processing the claims for an insurance company and are compensated based on the entire book of business and not just a fixed rate. In addition to TPAs, sometimes independent adjusters are used for special cases, such as a catastrophe that brings in an abnormally high number of claims over a short period of time that the company may not be able to handle by itself (Friedland, 2010).

To determine an accurate report of claims in a period, the time at which various transactions related to the claim, as well as when the claim itself is paid out are important. While some claims can simply be paid out, others may have other transaction costs, especially when a claim is associated with a trial or court case. Cases can also be reopened at a later date- often due to the reopening of a court case in which that claim is associated. Examples of how claims can be paid out over time based on various circumstances are illustrated in **Error! Reference source not found.**.

Table 1: Examples of Changes in Reported Values of Claims (Friedland, 2010)

	At December 31, 2007			Transactions During 2008			At December 31, 2008		
Example Number	Cumulative Paid Claims	Case O/S	Reported Claims	Paid Claims	Change in Case O/S	Reported Claims	Cumulative Paid Claims	Case O/S	Reported Claims
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	-	-	-	100	-	100	100	-	100
2	200	-	200	50	-	50	250	-	250
(Making payments where there had been no previous case outstanding increases reported claim.)									
3	-	-	-	-	1,000	1,000	-	1,000	1,000
(Establishing a case outstanding increases reported claim by the amount of the case outstanding.)									
4	-	1,000	1,000	100	(100)	-	100	900	1,000
(Payment with offsetting case outstanding reduction has no effect on reported claim.)									
5	500	5,000	5,500	200	(1,000)	(800)	700	4,000	4,700
(If case outstanding is reduced by a larger amount than the claim payment, the impact is a reduction to reported claim.)									
6	5,000	10,000	15,000	12,000	(10,000)	2,000	17,000	-	17,000
(If payment on closing exceeds case outstanding, reported claim transaction is positive.)									
7	5,000	10,000	15,000	6,000	(10,000)	(4,000)	11,000	-	11,000
(If payment on closing is less than case outstanding estimate, reported claim transaction is negative.)									
8	5,000	15,000	20,000	4,500	-	4,500	9,500	15,000	24,500
(Claim payment with no change in case outstanding increases the reported claim.)									
9	3,000	10,000	13,000	-	(4,000)	(4,000)	3,000	6,000	9,000
(No payment and decrease in case outstanding decreases the reported claim.)									
10	2,000	10,000	12,000	1,000	5,000	6,000	3,000	15,000	18,000
(Payment and increase in case outstanding result in increase in reported claim.)									

While individual claims may seem to come in sporadically, they can generally be predicted in aggregate to varying levels of precision once organized correctly. To improve accuracy, one can generally subdivide claims for a particular product line into smaller groups that display similar characteristics such as the time to settle, average settlement value, time to report a claim after an incident, and the area of the incident (Friedland, 2010). However, deciding how to organize claims can be product dependent as claims for one incident can last years, while another lasts for only a few months. The data within these smaller groups must also be credible, which means that the subsets should have a sufficient number of claims in order to draw credible conclusions about each subset (Friedland, 2010). The method in which claims are aggregated should be taken into account before analysis. Claims can be aggregated by calendar year, incident year, policy year, report year, and numerous other metrics (Friedland, 2010). All of these aggregation methods have their own advantages and disadvantages, and the best one to use will likely depend on the product line and the type of claims.

The data utilized should be verified to be reliable and sufficient for the intended analysis. The data can be verified by checking that the data is consistent with financial statement data as well as with prior years. Additionally, the data should be reasonable, and any unquestionable data should be investigated. If done correctly, one should be able to more accurately predict characteristics of a claim or claims from an incident (Friedland, 2010).

2.2 Prediction of Unpaid Claims

In order to accurately predict the total value of claims for a particular period, an insurance company must be able to predict unpaid claims. This includes claims that are incurred but not reported, or IBNR claims. These claims include instances where damage has occurred, but it has not yet been reported to the insurance company (Friedland, 2010). Estimating unpaid claims begins by analyzing existing data to find patterns and identify possible anomalies. If not enough data is available, the existing data can be balanced using data from other sources, including the Insurance Services Office, Inc. and NAIC Annual Statement data (Friedland, 2010). Possible conflicts associated with various methods should be identified and evaluated to determine the best method to use. Using the chosen method, a projected ultimate value of claims should be calculated (Friedland, 2010). This value should then be altered and adapted as more data becomes available.

2.2.1 Development Triangles

Development triangles show changes in various elements over a period of time. They are useful in analyzing different values, such as reported claims, reported claim counts, and paid claims. The development triangle is commonly used to identify and analyze patterns in data and is helpful in estimating unpaid claims. Table 2 shows an example of a development triangle.

Table 2: Development Triangle

Year	**12 Months**	**24 Months**	**36 Months**
2010	90	120	125
2011	115	130	
2012	110		

Each row, column, and diagonal in the development triangle represents a different element. In the above example, rows represent one accident year. They can also be used to represent a specific event, such as a catastrophe. Columns represent age, or maturity as related to their corresponding row. Diagonals represent a valuation date. In the above example, the second diagonal is the December 31, 2011 valuation for accident years 2010 and 2011 (Friedland, 2010). The number of claims associated with a particular year reported will typically increase as time progresses until all claims are reported. However, it is possible for this to decrease. Circumstances where this occurs include claims settled for no payment or claims settled for a lower value than the original estimate (Friedland, 2010).

2.3 CATs

Many outside factors influence the processes of claim handling and predicting claims. One such factor that adds an element of complication to these processes is a catastrophe, otherwise known as a CAT. In the insurance sector, the term CAT describes a natural or man-made disaster that is unusually intense. If a loss amount is over $25 million (threshold may change over time) and affects more than a certain number of policyholders as well as insurance companies, it is labeled a CAT. Some of the most recent and costly CATs are Hurricane Sandy in 2012 and Hurricane Katrina in 2005 (Insurance Information Institute, Inc., 2016). A CAT for the Insurance sector, however, does not have to be a CAT for Hanover. If the CAT takes place in an area where Hanover has written few policies, it will not greatly affect its balance sheets.

Different CAT types include hurricanes, tornadoes, fire, wind, hail, floods, lightning, winter storms, earthquakes and volcanic eruptions. Hurricanes, though, were the costliest in recent years for the United States insurance industry (Kunreuther, Michel-Kerjan, & Doherty, 2011).

Table 3: Most costly insured catastrophes in the world, 1970 – 2008 (Kunreuther, Michel-Kerjan, & Doherty, 2011)

Cost	Event	Victims (dead or missing)	Year	Area of primary damage
$46.3	Hurricane Katrina	1,836	2005	United States, Gulf of Mexico
35.5	9/11 attacks	3,025	2001	United States
23.7	Hurricane Andrew	43	1992	United States, Bahamas
19.6	Northridge earthquake	61	1994	United States
16.0	Hurricane Ike	358	2008	United States, Caribbean
14.1	Hurricane Ivan	124	2004	United States, Caribbean
13.3	Hurricane Wilma	35	2005	United States, Gulf of Mexico
10.7	Hurricane Rita	34	2005	United States, Gulf of Mexico
8.8	Hurricane Charley	24	2004	United States, Caribbean
8.6	Typhoon Mireille	51	1991	Japan
7.6	Hurricane Hugo	71	1989	Puerto Rico, United States
7.4	Winterstorm Daria	95	1990	France, United Kingdom
7.2	Winterstorm Lothar	110	1999	France, Switzerland
6.1	Winterstorm Kyrill	54	2007	Germany, United Kingdom, Netherlands, France
5.7	Storms and floods	22	1987	France, United Kingdom
5.6	Hurricane Frances	38	2004	United States, Bahamas
5.0	Winterstorm Vivian	64	1990	Western/Central Europe
5.0	Typhoon Bart	26	1999	Japan
5.0	Hurricane Gustav	135	2008	United States, Caribbean
4.5	Hurricane Georges	600	1998	United States, Caribbean

Sources: Wharton Risk Center with data from Swiss Re and Insurance Information Institute.
Note: This table excludes payments for flood by the National Flood Insurance Program in the United States.
[a] In billions, indexed to 2007, except for 2008, which is in current dollars.

Since CATs may have a big impact on Hanover's balance sheet and, therefore, on investors' confidence, studying such disasters is very important. The sooner management presents precise data and expectations of a recent CAT, the less investors worry about company losses. Additionally, the reserve department needs up-to-date information to allocate the right amount of money to pay out to the customers. According to Gen Re, efficient claim handling includes a faster payout and results in better brand awareness, less bad-faith exposure, and a lower probability of claim escalation (Griffin & Kelley, 2014).

These events pose huge risks for insurers due to their unpredictable nature and the difficulty to react to them. Additionally, rating agencies typically consider losses from CAT events as the primary threat to an insurance company's solvency (Ellis, 2012). Therefore, being prepared for these disasters is a high priority for many insurers.

2.4 CAT Modeling

In order to properly prepare for unpredictable CAT events, insurance companies have increasingly relied on the utilization of CAT modeling in order to mitigate risk. As mentioned, CAT events include a wide array of natural and man-made disasters. Many models focus on the natural events, but some models even focus on man-made CATs to deal with emerging risks like terrorism (Wilkinson, 2008).

As a result, companies turn to CAT modeling techniques as a means of predicting the losses from such events. These techniques make use of vast amounts of historical data and advanced statistical methods to predict various aspects of risk associated with potential CAT events. In general, a CAT model evaluates the historic trends of previous CATs and generates probabilities associated with occurrences of these events and estimated losses (Jain, 2014). The figure below displays the general design of a CAT model.

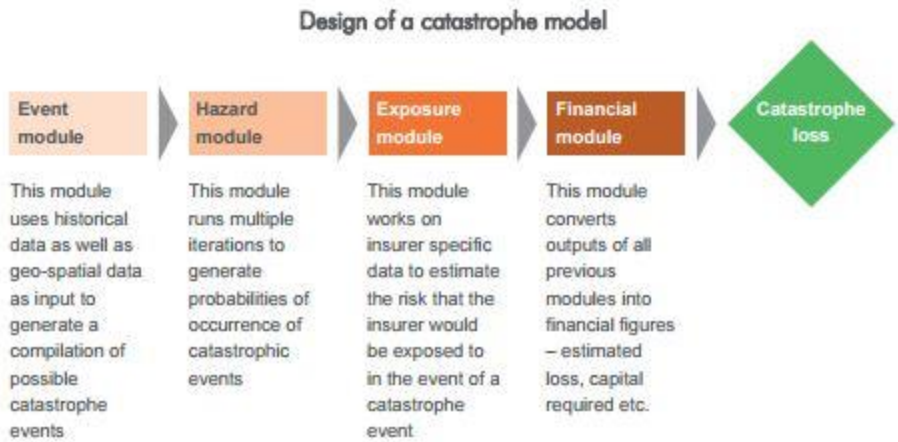


Figure 1: Design of a Catatrophe Model (Jain, 2014)

Insurers can draw numerous conclusions from these outputs. The companies can use figures from the models to map potential events and keep the overall risk associated with their books of business in check. Furthermore, an important function of CAT modeling is its use in a company's pricing and reserving strategy (Jain, 2014). By predicting aspects of potential catastrophes, companies can properly price the resultant risks. This information also allows companies to set aside an appropriate amount of money into a reserve in case the general pricing structure does not cover the full loss of a catastrophe. As a result, CAT modeling has become an important part of the operations at an insurance company.

2.5 The Hanover Insurance Group

The Hanover Insurance Group is a Property & Casualty insurance company headquartered in Worcester, MA. Founded in 1852, The Hanover is one of the longest operating businesses in the insurance industry. Hanover offers personal, commercial, and specialty insurance policies. The company offers these coverages in various locations across the United States and writes its business through third party agents. As a result, employees only interact with policyholders in regards to claims, while these agents deal with the purchasing efforts.

We quickly learned that this project focuses on Hanover's Personal and Commercial lines of business. Specifically we would be concentrating on the homeowners and personal automobile policies on the Personal Lines side the commercial property, commercial automobile, and inland marine policies on the Commercial Lines side.

2.6 Current Methods for CAT Modeling

The Hanover currently has a set method for reserving funds to pay out claims due to catastrophes, such as hailstorms, hurricanes, tornadoes, etc. When creating a reserve for a catastrophe, the company must estimate claims that are Incurred But Not Reported (IBNR). Essentially, the reserving department of Hanover must project the amount they expect to lose in claims without having actual filed claim reports. This is a process which requires the utilization of historical data from similar perils to project future losses. The data used in the current method is frequency data by day for each previous catastrophe as well as daily severity development for these catastrophes. With this data, Hanover projects catastrophic losses on an ultimate basis with regards to peril type, line of business, and geographic location. The projections create an estimated ultimate count of claims and an average severity or cost of each claim. The projection of the total loss amount is acquired from the product of these findings, which can then be used to find IBNR by subtracting the current reported losses.

To determine these values, Hanover uses a tool made in Excel. The tool draws on inputs of the type of peril and the states affected to output a spreadsheet of results. This output spreadsheet displays several tables populated with numerical projections by line of business. This view allows the user to see the ultimate projections but does not include the development

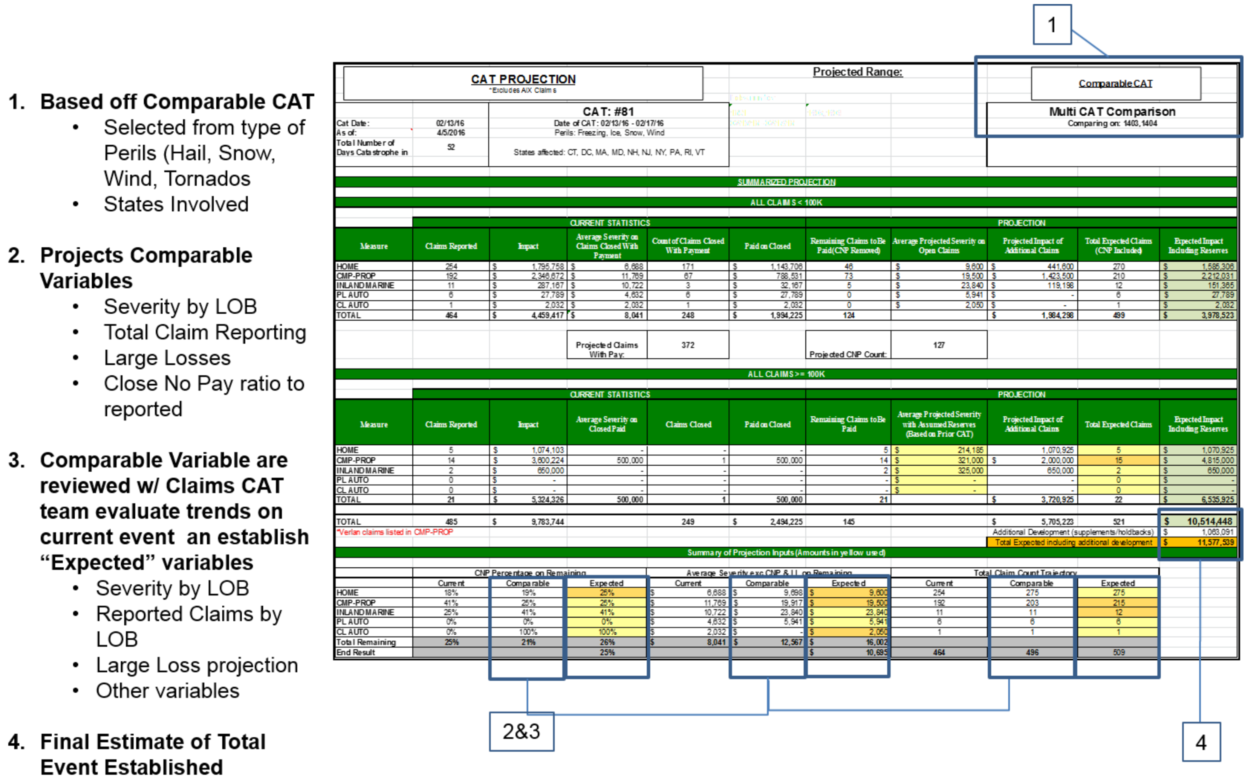


Figure 2: Hanover’s Current CAT Modeling Process

pattern that leads to this estimate. Additionally, this tool has the ability to be updated and monitored on a daily basis. Figure 2 is a quick view of the tool.

3. Methodology

This section details the methods we used to create the final spreadsheet model. These approaches included individual research and weekly meetings to understand the problem, followed by data analysis and selection of process to ultimately create the final product.

3.1 Understanding the Goal

While we knew the end goal was to create a catastrophe (CAT) model in Excel, we first had to familiarize ourselves with the idea of CAT modeling and then learn the specific elements that Hanover wanted in this CAT model. To begin understanding CATs and CAT modeling, we initially performed background research. This research included textbooks devoted to insurance and catastrophe predictions, as well as online sources, such as the EM-DAT International Disaster Database and articles from the Insurance Information Institute. The goal of this stage was to learn as much as possible about how the issue of catastrophe modeling has been tackled in the past, in order to align these ideas with Hanover's desires.

In particular, the textbook Estimating Unpaid Claims Using Basic Techniques by Jacqueline Friedland was key to our understanding of catastrophe modeling as it described general claim reporting and reserving processes. One team member also attended a webinar hosted by the Casualty Actuarial Society (CAS), which focused on catastrophe modeling. The event discussed the reasons for CAT modeling and introduced the three general components of modeling catastrophes: events, damage, and losses.

Such individual research was supplemented with frequent meetings and discussion so that each member of the team was familiar with the subject matter. Additionally, we held weekly meeting with the key Hanover personnel involved in the project. In the early stages of the project, these meetings allowed us to acquire further insight into the requirements for the CAT model. As the project progressed, these meetings allowed us to present our ideas of the model to the Hanover employees. Since we had total creative control over the model's layout, and we exhibited our model to the Hanover employees along the way to gain feedback and revisions. In doing so, we used our creativity to build a model that conformed to any requirements and additions requested by the Hanover personnel.

3.2 Understanding the Data

In addition to background research, we spent many of the early weeks of the project working to fully understand the data sets provided. Hanover provided two sets of CAT data, each dating back ten years. The first set was frequency data which allowed us to view the claim count progression over time for each CAT, and the second set was severity data which detailed the estimated cost per claim over time for each CAT. The section below will go into greater detail of our data analysis as well as how we decided to integrate the data into the CAT model.

3.2.1 Data Analysis

The current CAT methodology could surely be improved and streamlined, which is where we come in. In collaboration with Hanover we worked to streamline the CAT modeling process. In order to do so, Hanover provided us with a set of data off of which we would build a new model. This data file consists of 195,661 rows, with 19 columns of information. Each record, or row, is for a single claim. The following is some discussion by column.

3.2.1a Report Date

The first field describes the Report Date for each claim. The report date is the date on which the claimant called in the claim, making that the first date that Hanover knew about the claim. The earliest date of a claim in this data set is January 1, 2006. The most recent Report

Date is September 9, 2016. The table below shows how many claims were received in past years and how many have been received so far this year.

Table 4: Pivot Table of Report Date

Row Labels	Count of Claims
<5/18/2005	
2005	1
2006	11163
2007	9454
2008	29573
2009	11515
2010	23162
2011	40774
2012	32053
2013	10662
2014	13470
2015	6754
2016	7078
2020	1
Grand Total	195660

3.2.1b DOI

The second field is the DOI, which stands for Date of Incidence, for each claim. The Date of Incidence (DOI) describes the date on which the peril occurred. The earliest date for this column is January 1, 2006. The most recent DOI is May 12, 2016.

Table 5: Pivot Table of DOI

Row Labels	Count of Claims
<1/1/2006	
2006	11552
2007	9481
2008	30896
2009	10505
2010	22818
2011	41713
2012	32350
2013	9240

2014	**13729**
2015	**6957**
2016	**6419**
Grand Total	**195660**

3.2.1c Lag

The third field relates the Report Date and DOI by displaying the Lag. This value is the difference in days between the DOI and the Report Date and tells how many days after a peril it took for the insured to report a claim. The Lag ranges from -331 to 4745. There are a few negative values for Lag which are likely errors and the 4745 value may also be an error so these will not be taken into account. The figure below is a breakdown of the count of lag.

Table 6: Pivot Table of Lag

Row Labels	Count of Lag Buckets
<0	6
0-99	181819
100-199	7807
200-299	2855
300-399	2174
400-499	450
500-599	192
600-699	117
700-799	104
800-899	59
900-999	30
1000-1099	13
1100-1199	8
1200-1299	6
1300-1399	4
1400-1499	13
1500-3000	2
>3000	1
Grand Total	**195660**

The Lag will be an important value in modelling claim volume day-by-day. The graph below displays the number of claims separated by year for a given lag value.

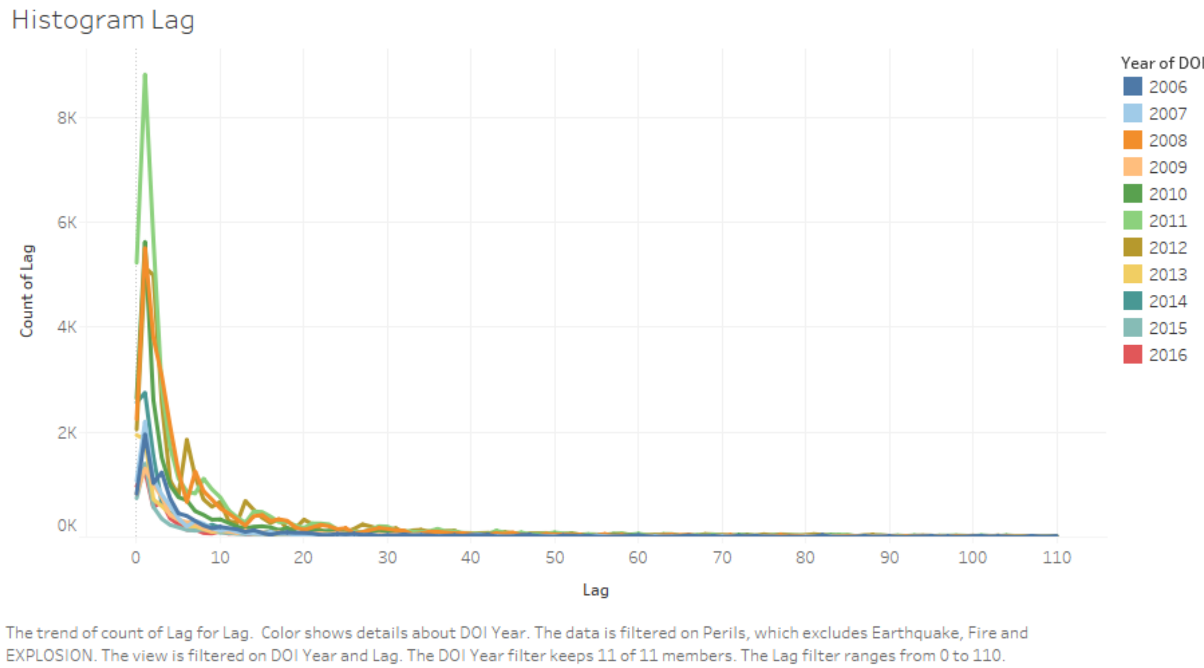


Figure 3: Histogram of Lag

3.2.1d AY

Field four is labeled AY, which stands for accident year. The information in this column is a specific year that matches the year of the DOI. Essentially, the accident year is the year in which the event that led to a claim occurred. In some cases, the accident year will not match the year in which the claim was reported. This is due to the previously mentioned lag that many claims experience. Therefore, if a loss event occurred at the end of a calendar year (in December for example), and the claimant did not report it until January, then the accident year of that claim remains as the prior year.

Table 7: Pivot Table of AY

Row Labels	Count of Claims
2006	11552
2007	9481
2008	30896
2009	10505
2010	22818
2011	41713
2012	32350
2013	9240
2014	13729
2015	6957
2016	6419
Grand Total	**195660**

3.2.1e Cat Numb

Field five is the CAT number. This number is assigned to a particular storm in a given calendar year. CAT numbers can be recycled after a few years. For example, CAT 06 shows up in four different calendar years (2008, 2011, 2014, 2016), and represents four different storms. There are 341 different CATs in the data set.

3.2.1f LOB

Field six is the line of business (LOB) related to each claim. The information here specifies which line of business writes the policy associated with each claim. These lines include: Commercial Line Automotive, Commercial Property, Home, Inland Marine, and Personal Line Automotive.

Table 8: Pivot Table of LOB

Row Labels	Count of LOB
CL_AUTO	2609
CMP-PROP	23311
HOME	144842
IM	2145
PL_AUTO	22753
Grand Total	**195660**

Lines that have the "Commercial" label are specifically designated as policies for businesses or companies. Alternatively, if the line of business has the "Personal Lines" designation, it is a policy that has been written to an individual.

3.2.1g Claim Numb

Field seven is the claim number. This number is distinctive for the claim and its format is 2 digits followed by a dash with six more digits. For example, 03-544393 is a claim number for a specific claim reported in January of 2006 for wind damage to a building.

3.2.1h Claims

Field eight is the number of claims that is associated with a given claim number. In the file, this number is always one, meaning that there is one claim associated with each claim number. Given a certain set of circumstances, this could be another number if there are multiple claims associated with a given claim number.

Table 9: Count of Claims

Count of Claims
195660

3.2.1i Feature Numb

Field nine is the feature number. This value is associated with the claim number and is used to further break down a claim by the loss type, peril, or coverage. Each of these features also have a transaction associated with it. A claim number can have multiple feature numbers associated with it.

3.2.1j State

Field ten is the state abbreviation. This represents the state in which the claim occurred. **Error! Reference source not found.** shows the total count of claims in each state. Michigan had by far the most claims, more than doubling the state with the second highest amount. The other most frequent states are (in decreasing order) Massachusetts, Indiana, Illinois, Louisiana, Arkansas, Connecticut, Georgia, and Maine. Various other states have less than 1,000 claims.

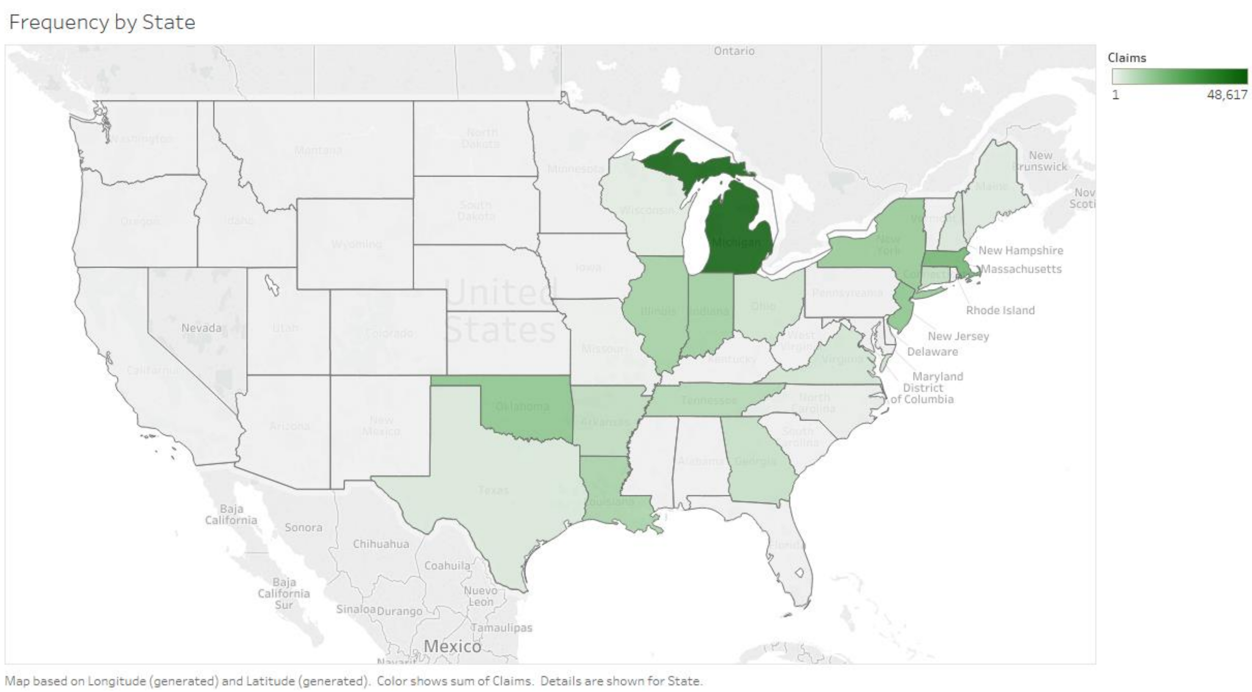


Figure 4: Frequency by State

3.2.1k Description

Field eleven states a short description of what each claim was for. These descriptions include things such as property loss or wind and hail.

3.2.1l OS_Loss

Field twelve indicates the outstanding loss associated with a claim and is labelled "OS_Loss". This column can contain a value from 0 to the full estimate of a claim. If the column contains a '0', then the claim has been paid in full. This would imply that there is no outstanding value and no funds need to be reserved to pay out the claim. If the column contains the full estimate of a claim, then the claim has not been paid at all yet. The column can also contain a value anywhere in between 0 and the total estimate of the claim, implying that some but not all of the claim has been paid.

Table 10: Pivot Table of OS_Loss

Row Labels	Count of OS_Loss
2006	11552
2007	9481
2010	22818
2011	41713
2012	32350
2013	9240
2014	13729
2015	6957
2016	6419
Grand Total	**195660**

3.2.1m PD_Expen

Field thirteen represents allocated loss adjustment expenses or ALAE. These expenses are those attributed to the processing and paying out of a particular claim. This value is not used to calculate the incurred loss value, as ALAE and loss are thought of separately.

Table 11: Pivot Table of PD_Expen

Row Labels	Count of PD_Expen
2006	11552
2007	9481
2010	22818
2011	41713
2012	32350
2013	9240
2014	13729
2015	6957
2016	6419
Grand Total	**195660**

3.2.1n PD_Loss

The fourteenth field is labeled "PD_Loss" which stands for Paid Loss. This column indicates how much loss has been paid so far. This value may or may not be equal to the total cost of the claim. If claim is still being settled, it may be only paid in part thus far. The claim might ultimately cost $1,000, but at this point in time Hanover has only paid $100. This would make the paid loss column reflect a value of 100, and this value can be changed over time to reflect how much more of the loss has been paid to the claimant, until it reaches the ultimate value.

Table 12: Pivot Table of PD_Loss

Row Labels	Count of PD_Loss
2006	11552
2007	9481
2010	22818
2011	41713
2012	32350
2013	9240
2014	13729

2015	6957
2016	6419
Grand Total	**195660**

3.2.1o Salv_Rcvr

Field fifteen represents the salvage and recovery value of damaged property. This means, for example, if a car worth $10,000 can be salvaged for a value of $500 in parts or scraps, the total loss to the insurer would be $9,500.

Table 13: Pivot Table of Salv_Rcvr

Row Labels	**Count of Salv_Rcvr**
2006	11552
2007	9481
2010	22818
2011	41713
2012	32350
2013	9240
2014	13729
2015	6957
2016	6419
Grand Total	**195660**

3.2.1p Subro_Rcvr

The sixteenth field represents the subrogation recovery amount. This is the amount that Hanover recovers from other insurance companies. This happens when Hanover makes its own claim against others who may have caused the claim. Since Hanover recovers this money, it is subtracted from the incurred loss of the claim.

Table 14: Pivot Table of Subro_Rcvr

Row Labels	**Count of Subro_Rcvr**
2006	11552
2007	9481

2010	22818
2011	41713
2012	32350
2013	9240
2014	13729
2015	6957
2016	6419
Grand Total	**195660**

3.2.1q Sum Incurred

Field seventeen contains the sum incurred for each claim. The sum incurred is the total amount that a claim is worth. This is the sum of two previous columns, OS_Loss and PD_Loss. Combining the unpaid claim amount, or outstanding loss, with the paid claim amount, or paid loss, gives the total amount of a claim. The sum incurred will be very crucial information in tracking a catastrophe's severity, or total loss. This number is relevant for Hanover to plan to have enough money on hand to cover these losses, and the data about the sum incurred can be used to create a model to predict an upcoming catastrophe's final cost.

Table 15: Pivot Table of Sum Incurred

Row Labels	**Count of Sum Incurred**
2006	11552
2007	9481
2010	22818
2011	41713
2012	32350
2013	9240
2014	13729
2015	6957
2016	6419
Grand Total	**195660**

3.2.1r Perils

Peril type is located in field eighteen. This column indicates the type of peril or perils associated with a particular claim. Examples of perils include flood, wind, hail, hurricane, and more. Many claims have multiple perils associated with them, such as “Wind, Rain, Flood, Freezing”. This column indicates that all four of those perils occurred during the catastrophe that caused the claim in question.

Table 16: Number of Claims and Storms by Peril

Peril	Number of Claims	Number of Storms
WIND	174583	324
RAIN	4194	2
FLOOD	142773	262
FREEZING	29510	30
HAIL	123537	259
TORNADO	122839	215
ICE	34363	31
SNOW	32465	27
EARTHQUAKE	4	1
FIRE	22	7
EXPLOSION	37	2
WEIGHT OF ICE & SNOW	1984	3
THUNDERSTORM	4321	3
COLLAPSE	340	1
HURRICANE	28868	6
HURRICANE DOLLY	35	1
HURRICANE GUSTAV	7491	1
HURRICANE IKE	4459	2
TROPICAL STORM	225	4
TROPICAL STORM FAY	100	1
TROPICAL STORM HANNA	82	1
THAWING	761	1
LIGHTNING	3307	1
RIOT	7	1
CIVIL DISORDER	7	1
POWER OUTAGE	776	1

This field is extremely important in creating a catastrophe model, as each catastrophe is different. Each catastrophe may have a different amount of time for claims associated with it, or a different volume of claims. Different peril types also tend to have different severities associated with them, as some cause more damage than others. This column will be crucial in determining exactly what type of a catastrophe caused a given claim and how that claim will then play into the model.

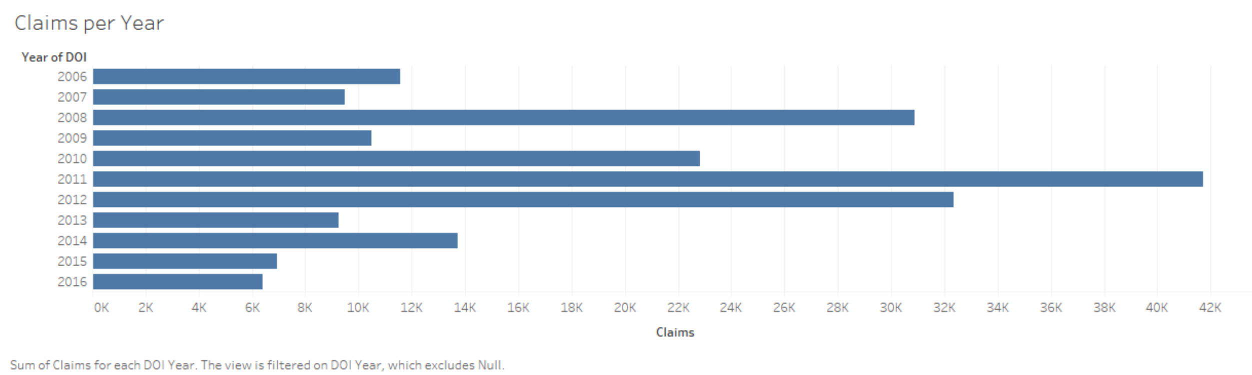


Figure 5: Claims per Year

3.2.1s Open/Closed

The nineteenth and final field in the data set is call "Open/Closed". This column indicates whether a claim is currently open or currently closed. An open claim has not been paid in full, and could still require more funds from Hanover to settle the claim. A closed claim has been settled and paid in full, and no longer needs the same attention to checking on funds available for the claim.

Table 17: Pivot Table of Open/Closed

Row Labels	Count of Claims
closed	195018
open	642
Grand Total	**195660**

3.2.1t Frequency and Severity Data Storms

We received two separate spreadsheets for the claim count and loss dollars data. There were many storms with representation in each data file; however, some were only represented in one spreadsheet. The figure below details the distribution in the format of a Venn diagram, with the storms shown in both spreadsheets represented by the intersection of the circles.

Figure 6: Storm Representation by Data File

3.2.2 Data Reconciliation

After analyzing the different fields included in the data sets, we worked to reconcile all of the data in order to ensure its consistency. To carry out this task, we focused on making sure numerical fields added up to the proper values. For example, in the frequency data we confirmed that the value contained in the Sum Incurred field matched the sum of the Paid Loss and Outstanding Loss fields. We also analyzed the perils associated with each CAT to ensure that, indeed, each individual CAT had a specific set of perils. Additionally, we created a field to distinguish between each CAT event. We built a "key" field which linked the year of incidence and CAT number, thereby easily breaking down the catastrophes into unique storms.

A large obstacle we faced while reconciling frequency data regarded the “lag” field. We found some claims that had extreme lags such as 1,400 days or more after the storm, or claims that had negative lags, as if the claim had been reported before the storm ever occurred. We sorted through these outlier lags and determined where the issues stemmed from incorrect reporting dates. We ultimately fixed these incorrect lags by correcting the report dates thereby allowing us to continue using this data.

We faced a similar issue in the severity data as there was no “lag” field. Instead we created a field called “Severity Lag” that operated under similar logic to the frequency data’s lag field. In order to line up with the frequency data, this field displays the lag day on which the estimated dollar amount for each claim was given, thereby allowing day-by-day progression of the severity.

These added fields, among others that were also necessary for the model, were then built into the model’s document. Logic also exists that will automatically calculate these fields for any new data that is loaded into the system.

3.2.3 Exploring Trends

In order to best interpret the claim data, we created numerous graphs and charts outlining the data through different filters to discover what distributions or trends may be present. We created such views in Tableau, a data visualization program. We initially created graphs focusing on the lag in order to visualize the claim count development of a storm. These graphs displayed claim counts at each lag day as a cumulative percentage for each peril in order to show the difference in reporting time by peril. Some of these can be seen below. A complete view of all graphs can be found in Appendix B.

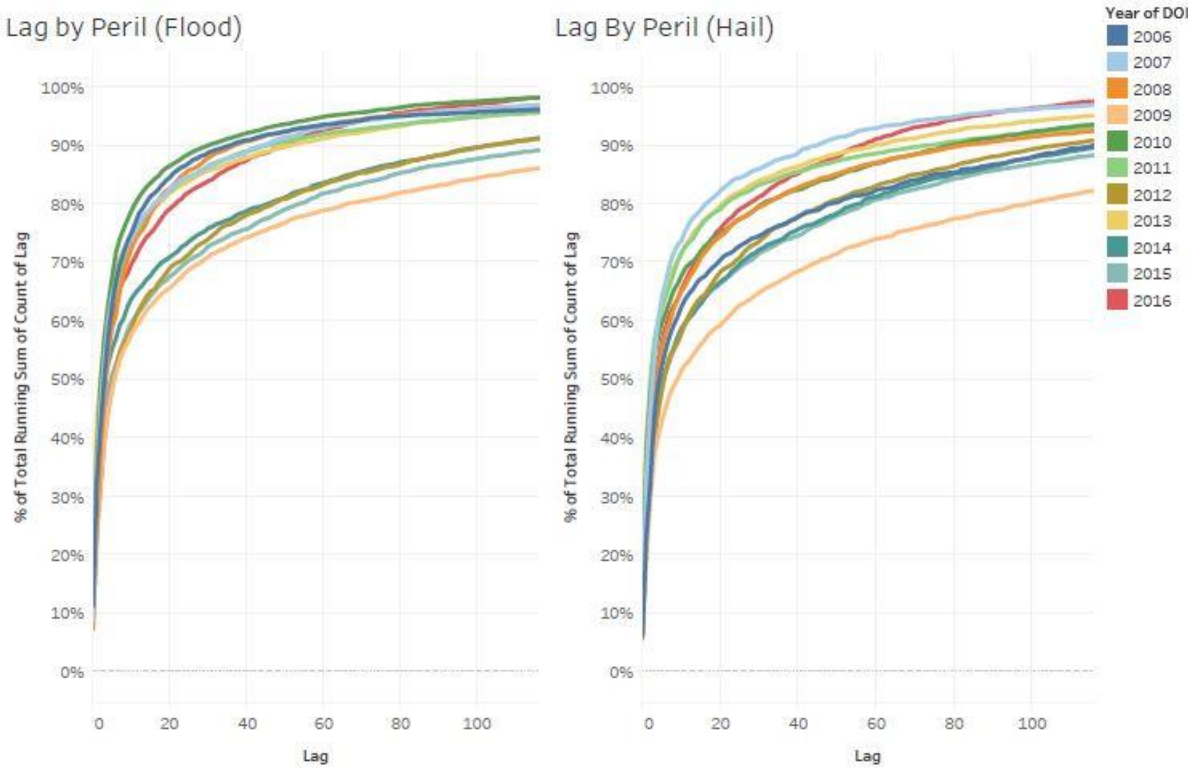


Figure 7: Example of Lag by Peril Graphs

Another helpful view was the graphical representation of lag in relation to claims as a percentage of total claims. This graph shown below helped to visualize what percentage of claims were received on each day following a catastrophe and was particularly useful in understanding the pattern of claim development for catastrophes.

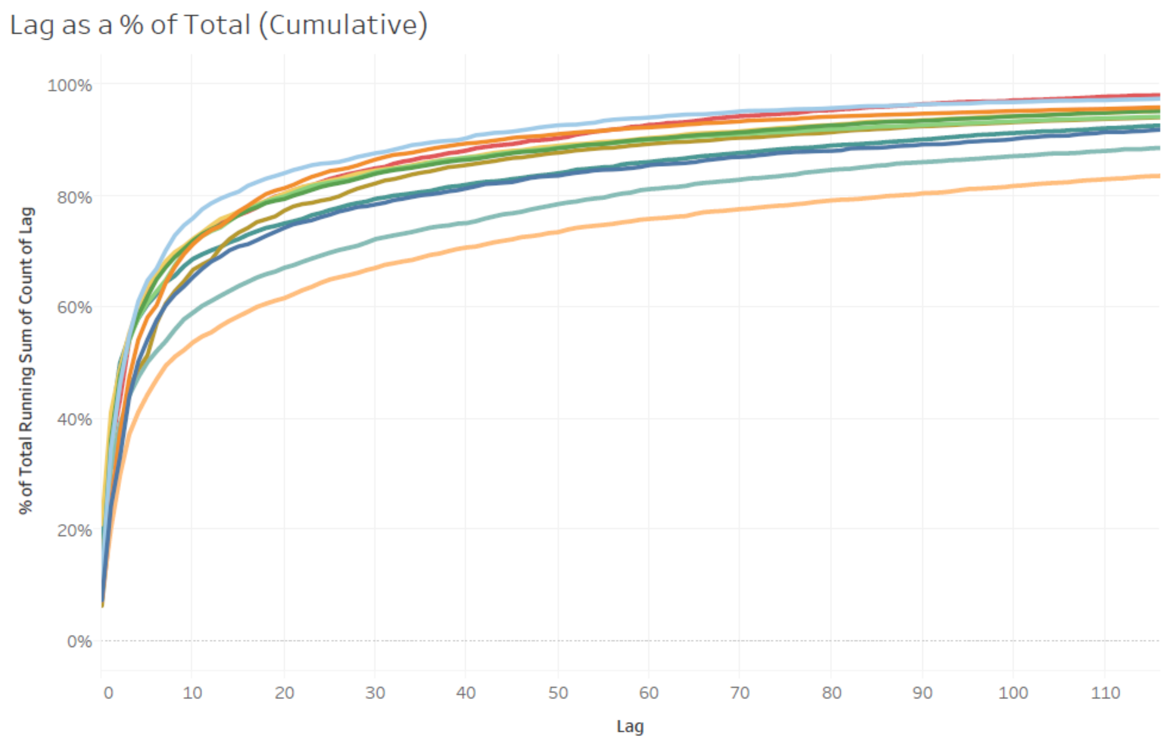


Figure 8: Lag as a Percent of Total (Cumulative)

Additionally, charts were created to assess total claims and sum of claim values incurred by line of business, as well as the correlation between perils. These charts can be found in Appendix C and D, respectively.

3.2.4 Creating Filters

Drawing on our research and weekly meetings, we decided an important aspect of the final spreadsheet model would be the user's ability to filter on different criteria to accurately match the CAT he or she wishes to model. After reviewing the graphs in Figure 2.1 and other similar graphs, it became evident that CATs perform differently across different perils, lines of business, and states. As a result, the criteria we settled on for the filters would be Peril Type, Line of Business (LOB), and Region. Therefore, by specifying these conditions, the output of the model will return accurate predictions for the exact type of storm the user is modeling.

The next step was to parse out these filters. The LOB filter was simple as we only dealt with five lines of business and each claim corresponded to exactly one line of business. Similarly, building the Region filter was a simple process. We assigned the states to regions according the U.S. Census Bureau region definitions. Again, each claim corresponded to exactly one state and, therefore, exactly one region. The tables below display the five lines of business given by their short and long names, and the regions with the list of states contained in each region.

Table 18: Line of Business and Region Filters

LOB Filter

Short Name	**Description**
HOME	Personal Lines Homeowners
CMP-PROP	Commercial Lines Property
IM	Commercial Lines Inland Marine
CL_AUTO	Commercial Lines Automobile
PL_AUTO	Personal Lines Automobile

Region Filter

Region	**States**
Northeast	CT, ME, MA, NH, NJ, NY, PA, RI, VT
Midwest	IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI
South	AL, AR, DE, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA
West	AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR,UT, WA, WY

Separating the types of perils into buckets, on the other hand, was not such a simple task. In order to determine these buckets, we brainstormed the categories that would include all possible catastrophes. We looked through the data set to determine most frequent perils, and discovered that wind was present in over 85% of the claims, and flood was also contained in a very high percentage of the claims. Because of this, we decided to move forward disregarding these two peril types to get the most accurate and specific buckets possible. After referring to many outside sources, including other insurance/reinsurance companies, as well as an online catastrophe database, the buckets were finalized. The buckets we created are Hurricane/Tropical

Storm, Winter Storm, Tornado/Hail, Other Wind Event, and Other Catastrophes. The table below contains the explanation of which claims are in each bucket.

Table 19: Bucket Contents

Bucket	Contents
Hurricane/Tropical Storm	All claims with hurricane or tropical storm listed as a peril
Winter Storm	All claims with perils of snow, ice, freezing, thawing, and/or weight of snow and ice
Tornado/Hail	All claims with tornado or hail listed as perils that do not list hurricane or any of the winter storm perils
Other Wind Event	All claims with wind listed that had not been categorized in any of the three previous buckets
Other Catastrophes	All claims that have not been previously categorized (for example, earthquakes and fires)

3.3 Building the Model

Once the underlying data had been thoroughly analyzed and manipulated, we began work on the model itself. We created the final spreadsheet model through Excel utilizing Excel's pivot chart and macro features through the Visual Basic for Applications (VBA) programming language. The spreadsheet model was drafted and changed multiple times to fully align its process with the goal that Hanover had laid out for the project.

We developed the tool for the two data sets (frequency and severity) separately and then combined them in the final tool for estimating the total loss dollar. Even though we constructed the frequency and severity part differently, the structure remained similar. Hence, the merge worked smoothly and we omitted the code associated with the separate tool.

To create the report, we used different features of Excel. In general, we used formulas for data processing, pivot tables for data aggregation, and VBA for execution order as well as user interaction.

The first step was to filter the data according to the aforementioned filter fields. Each CAT which matches all the chosen descriptions (Peril Type, LOB, and Region) would be arranged in pivot table. These tables display the claim count and severity, respectively, as a running percent of the total for each CAT (similar to Figure 8 above). To compute the standard deviations (SD) for each field we calculated the changes relative to the current lag day within the actual data. For example, if a user is modeling a CAT on day 15, the model will determine the difference in claim count and severity relative to lag day 15. This gave us the necessary values to compute the interval of certainty at one and two standard deviations. After those calculations, we created charts for the report. The figure below summarizes this coding methodology into a flowchart.

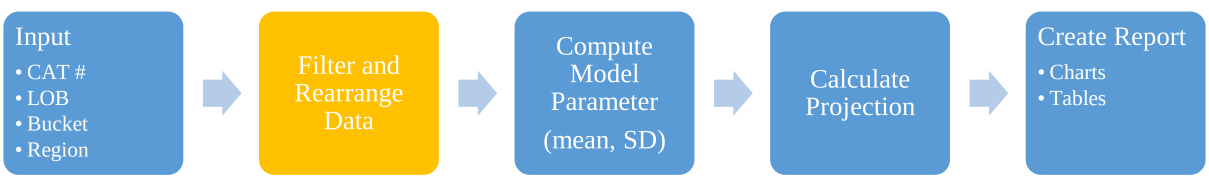


Figure 9: Coding Methodology Flowchart

The sequential arrangement was programmed with VBA. This was a crucial step for the calculation time. We had to consider which sheet or cell must be updated at which step. To reduce the calculation time, the goal was to update as little as possible. We also used VBA for the interaction with the user. Pop-up windows were easily created, and we implemented certain useful features (e.g. print PDF).

4. Results and Findings

The following chapter details the results and findings of the project team. This chapter aims to fully explain the spreadsheet model created through this project, as well as to describe how to use and update the model for future users.

4.1 Explanation of Spreadsheet Model

The final model essentially intakes user information to determine what data is relevant to a given catastrophe or catastrophes. It then selects the relevant data and utilizes only this data to create its future projections. The model generates predictions for the current catastrophe or catastrophes by utilizing trends found in the relevant historical data selection. Below is a table listing the inputs and outputs of the spreadsheet model.

Table 20: Model Inputs and Outputs

Inputs	**Outputs**
Catastrophe Number	Ultimate Count Projection with Confidence Intervals
Date of Event	Ultimate Loss Dollars Projection with Confidence Intervals
Reporting Day (this will allow the model to calculate lag day)	Cone Graph
Number of Claims Reported to Date *	Figures Depicting Prior Storms with Same Peril Set (Specific Types Selected in Inputs)
Reported Loss Dollars to Date *	Other Figures (Selected in Inputs)
Peril (s) *	
Region(s) *	
Line(s) of Business *	

Fields denoted with a * can be automatically populated, a feature to be discussed later.

Based on a user's customized inputs and on the specific CAT(s) he wishes to model, this tool outputs a series of graphs and tables displaying both the current trends and the projected future trends of claim development. The tool outputs a comprehensive report allowing the user to read and visualize the projected development of total loss dollars, claim count, and severity. The tool also delivers a view of the actual claim development compared to the average claim development associated with the specific type of CAT being modeled. The next chapter explains in detail exactly how to use and interpret this CAT model.

5. Deliverables

In this chapter we will display the input and output pages of the finalized tool. Additionally, this chapter will provide documentation that will describe the interface itself and how to use the model to output the user's desired results.

5.1 Input Page

The input page, displayed below, is the start page that the user will see upon opening the file. On this page, the user can specify his or her inputs based on the catastrophe he or she wishes to model.

Figure 10: Spreadsheet Model Input Page

Upon opening the document, the user will be greeted with a pop-up window with the options to either go directly to the input page or to view the documentation. Choosing the option for the documentation will bring the user to a sheet that display complete documentation for the tool that includes instructions on how to use and interpret the model. This documentation can be viewed in full in Appendix F. The other option brings the user to the input page of tool where he or she will find several input fields.

The first field that the user will see is the DOI, or Date of Incidence, field where the user must specify the date on which the CAT occurred. The next input is the CAT Number field where the user can specify the CAT event that he or she wants to model. This field can take up to five entries in the event that the user wishes to model multiple active CATs. In the event that multiple CAT numbers are entered, the tool will utilize the earliest date of incidence among the specified CATs.

The input page is equipped with some automation such that after the previously described fields have been entered, other fields will populate. Once the user is comfortable with the DOI and CAT Number fields, he or she may click the button labeled “Populate Suggestions and Claims to Date.” Doing so will populate the table labeled Claims to Date. The values that appear in this table will correspond to actual values of total claim count and estimated loss dollars to date. Again, in the event multiple CAT numbers are entered, the values in this table will represent the sum of claim count and loss dollars across those CATs. Additionally, the button will return suggested selections for the Peril, LOB, and Region fields that are described below.

The next items that the user will see are three checklists labeled Peril, LOB, and Region, respectively. In each checklist the user can check one or multiple criteria based on the characteristics of the storm which he or she wants to model. For example, if the user is concerned with an ice storm that affected both commercial and personal property in Michigan, he or she would check only Winter Storm for in the Peril checklist, both HOME and CMP-PROP in the LOB checklist, and only Midwest in the Region checklist. Essentially, the user tells the model to base its output only on historical data for CATs that match the specified criteria, thereby offering an accurate view of this specific storm’s claim development. If the user is unsure about the criteria to use, he or she can simply refer to the suggested selections which list the criteria that correspond to the CATs that have been entered.

Once the user has decided on the desired inputs, he or she may then click the button labeled Create Report. At this point, the model will run through its processes and deliver its output in the form of a report on a separate sheet in the same Excel workbook.

5.2 Output Page

Once the user has created the report, it is now up to him or her to interpret the results on the output page. Each time the model is run, the graphical depictions on the output page will change, as will with numerical representations of projections. The specifics for the report including peril, CAT number, line of business, and region, will change as well. Appendix E is an example of what the output page may look like once a particular event is modeled. This example is of CAT numbers 5 & 6, modeled against all peril types, regions, and lines of business.

The user will immediately notice that the output is separated into three distinct sections. The first section has a title that matches the CAT number(s) that the user input, followed by a one-sentence summary of the information that is presented on this output page. It also displays the projections for the total loss dollars and a table exhibiting the input criteria for reference. The second section displays the projections for the claim count and severity. And, the third section depicts a zoomed in view comparing the actual claim development to the predicted average claim development to date.

A series of graphs is present throughout the report, and these graphs fall into two distinct types. The user will see the first type of graph in the ultimate projections for loss dollars, claim count, and severity. These graphs exhibit cumulative development over time, and the user will notice that each graph has different colored portions that fan out from a certain point. The point at which the graphs fan out represents the lag day relative to the date of incidence on which the model has been run. Before reaching this point the user will notice two lines: a blue line that represents the actual development of this CAT over time and a dark green line that represents average development for a CAT matching the input criteria. We display both of these lines so that the user may view how the current storm has developed compared to the average. Once the lines reach the current lag day, the blue line stops and the dark green line continues to represent the average projected development.

However, the actual development of the storm being modeled may not adhere strictly to this average projected development. To present an intuitive view of how the storm could develop, the previously mentioned fanned out portions of the graph begin at the current lag day and represent intervals of certainty for the projected development. The outer interval shown in light green is two standard deviations away from the average projected development. As a result, the user's interpretation is that the actual development is highly likely to fall within this light green interval. The inner interval represented by a solid green portion is one standard deviation away from the average projected development. The user's interpretation here would be that the actual development is likely, but not highly likely, to fall within the solid green interval. This logic holds for all three ultimate projection graphs.

As mentioned previously, the first graph represents projected loss dollars allowing the user to formulate a prediction of the ultimate loss amount associated with the particular storm(s) he or she is modeling. Similarly, the second and third graphs display projected claim count and severity, respectively. These views offer the user a means of predicting the ultimate number of claims associated with the storm(s) as well as the average dollar loss per claim. Additionally, we provided numerical values for each of these projections in respective tables below the graphs. The tables exhibit the average ultimate projections in bold along with the one- and two-standard deviation intervals below. This aspect allows the user to visualize the developments graphically and relate actual numbers to the visual aids.

The second type of graph is present in the third section and is a simple line graph. The graphs in this section display the actual development versus the average historical development to the current lag day of loss dollars and claim count. This view allows the user to determine if the CAT being modeled is following a similar development pattern to what has normally been seen in the past for this type of CAT, or if it is drastically different from the norm. Finally, this output report is printable and will be overwritten each time the user runs the model.

5.3 Updating the Underlying Data

As new catastrophes occur, data associated with these needs to be updated into the data portion of the model. This will allow the model to most accurately model both these new catastrophes when they are open storms, and future catastrophes based on this relevant, new data. This new data will be updated into the spreadsheet data tab, and the data from an active storm that a user wishes to model can be pulled in automatically to the input page to show claim development to date.

In order to update this data, the user must copy the frequency and severity information from the respective files kept locally at Hanover. Though we have other calculated fields that may not be present in the original data, the formulas and code included in the model will calculate these fields. Ultimately, this process is simple and will keep the underlying data current and accurate.

5.4 Users

At the conclusion of this project the team handed over the tool to the Hanover Insurance Group. The file containing the model will remain in the Hanover system. Its primary users will be individuals on the Actuarial Reserving team as well as individuals in the Claims/Finance Group. Since the team is not available to train these potential users, we have included full documentation within the file. These comprehensive instructions will allow individuals to understand how to use and interpret this model independently. The documentation and training manual can be found in Appendix F.

6. Recommendations and Conclusions

This section presents our concluding thoughts on this project and offers opportunities for future updates to the spreadsheet model.

6.1 Conclusion

Throughout the duration of this project, we have found opportunity for improvement in The Hanover Insurance Group's catastrophic reserving process. We utilized the current methods of projecting losses due to catastrophes to build a new spreadsheet model which can help increase the accuracy of Hanover's reserving process.

This project gave us the opportunity to automate and quicken a tedious process. We were able to do so with the use of Microsoft Excel and Microsoft Excel's Visual Basic for Applications. The goal to create a quick and functional model guided the project team along the way. The final model is simple and user-friendly, making it useful and easy to understand, so it can be shared with personnel of varying degrees of understanding of catastrophe modeling. We hope this speedy reserving model will serve the Hanover well in the future.

6.2 Future Recommendations

The project team recommends updating the data utilized in the spreadsheet model at least monthly to ensure accuracy. The team also recommends some future updates for the model including filtering by smaller regions and projecting loss dollars on the claim level rather than the storm level.

Bibliography

Ellis, D. (2012). *CAT Modeling...Why All the Fuss?* Retrieved from Marsh Insights: Property: https://www.marsh.com/content/dam/marsh/Documents/PDF/US-en/Marsh-Insights-Property-Fall-2012.pdf

Friedland, J. (2010). *Estimating Unpaid Claims Using Basic Techniques.* Casualty Actuarial Society.

Griffin, M., & Kelley, K. (2014). *Good Cat Claim Handling—Planning.* Stamford, CT: General Reinsurance Corporation.

Insurance Information Institute, Inc. (2016, July). *Inurance Topics*. Retrieved from Catastrophes: Insurance Issues: http://www.iii.org/issue-update/catastrophes-insurance-issues

Jain, A. (2014). *Catastrophe Modeling in P&C Insurance.* Retrieved from Everest Group Research: http://www.genpact.com/docs/default-source/resource-/catastrophe-modeling-in-p-c-insurance

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss Models: From Data to Decisions.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Kunreuther, H. C., Michel-Kerjan, E. O., & Doherty, N. A. (2011). *At War with the Weather : Managing Large-Scale Risks in a New Era of Catastrophes.* Boston: The MIT Press.

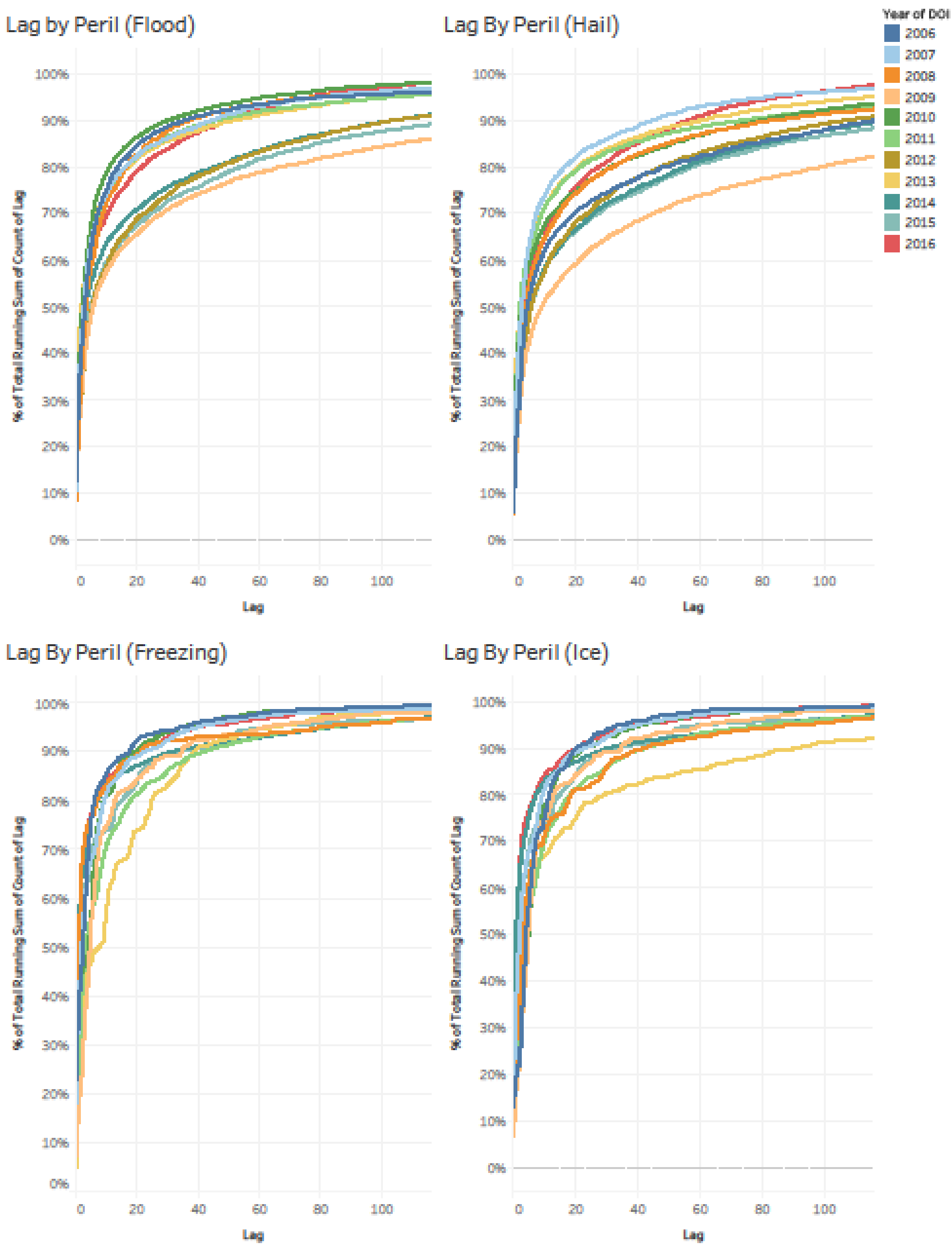
Wilkinson, C. (2008). *Catastrophe Modeling: A Vital Tool in the Risk Management Box*. Retrieved from Insurance Information Institute: http://www.iii.org/article/catastrophe-modeling-vital-tool-risk-management-box

Appendices

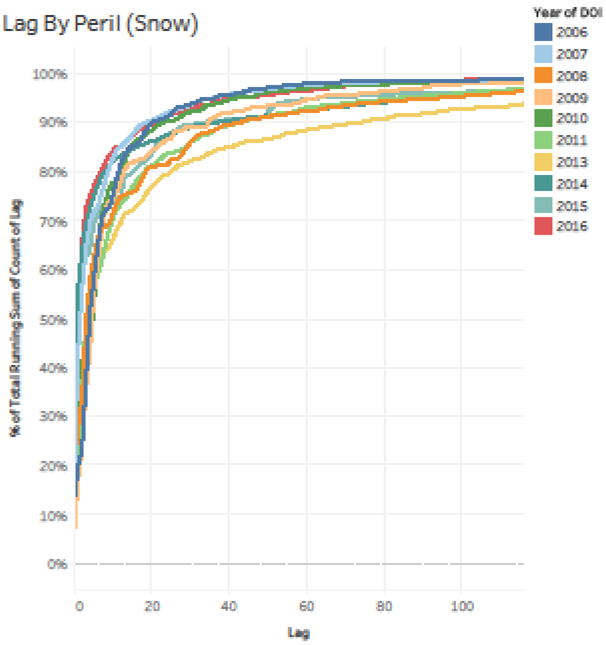
Appendix A: Data Analysis Summary

Field	Type	Description
Report Date	Date	Report Date is the date that the claim was reported to Hanover.
DOI	Date	DOI is the Date of Incidence and is the date that the event occurred.
Lag	Number (nominal)	Lag is the number of days between the DOI and the Report Date.
AY	Year	AY is the Accident Year and is the year that an event occurred.
CAT Numb	Number (descriptive)	CAT Number is number from 1-94 assigned to a certain catastrophe; values are recycled.
LOB	Text	LOB is the Line of Business which specifies the business area that writes the policy associated with each claim.
Claim Numb	Number (descriptive)	Claim Number is a unique number assigned to each claim.
Claims	Number (nominal)	The Claims field contains a value describing the number of claims associated with a given claim number.
Feature Numb	Number (nominal)	This is a nominal number used to further break down the claim by loss type, peril, or coverage.
State	Text	The State field indicates the state in which the claim occurred.
Description	Text	The Description is a brief summary of what was damaged during the catastrophe.
OS_Loss	Number (nominal)	OS_Loss is the Outstanding Loss and is a value indicating the estimated loss that has yet to be paid out on a particular claim.
PD_Expen	Number (nominal)	PD_Expen is Paid Expenses and is a value indicated the amount that has already been paid out in expenses for a claim.
PD_Loss	Number (nominal)	PD_Loss is Paid Loss and is a value indicating the amount that has already been paid out on a particular claim.
Salv_Rcvr	Number (nominal)	Salv_Rcvr represents the salvage recovery value of damaged property.
Subro_Rcvr	Number (nominal)	Subro_Rcvr represents the subrogation recovery amount.
Sum Incurred	Number (nominal)	Sum Incurred is a value indicating the estimate of the total claim value and is calculated as the sum of OS_Loss and PD_Loss.
Perils	Text	Perils indicates the type of events associated with each claim.
Open/Closed	Text	The Open/Closed field indicates if a given claim is open or closed.

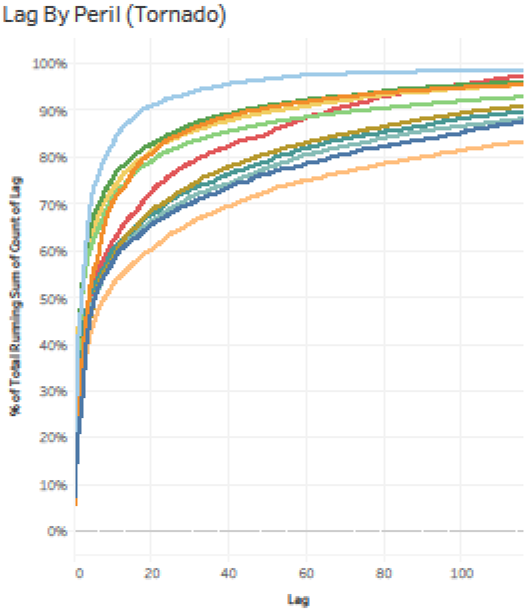
Appendix B: Lag by Peril



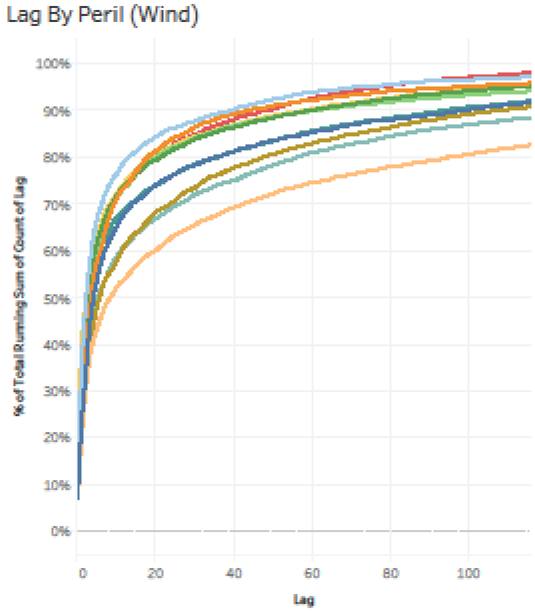
Lag By Peril (Snow)

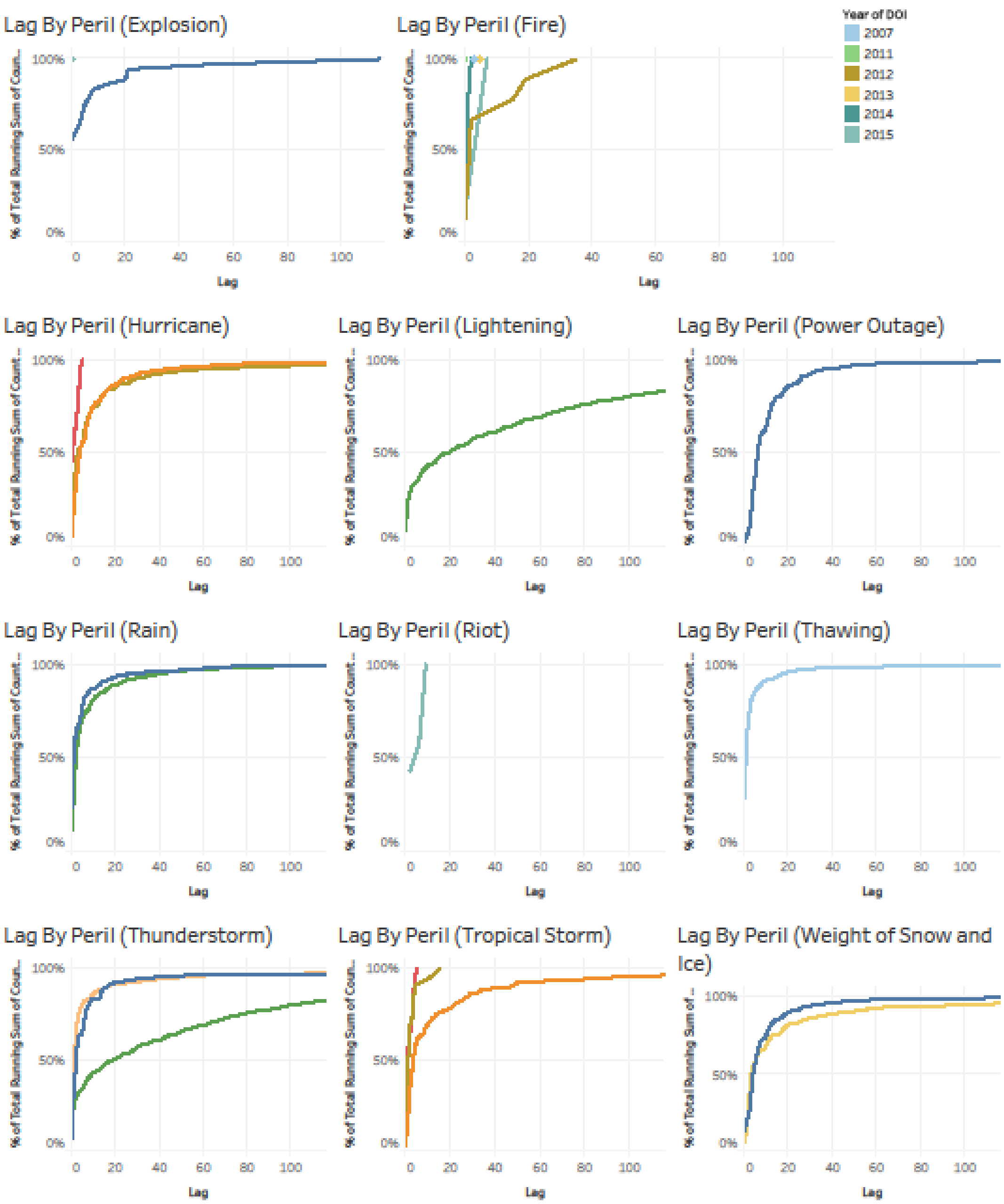


Lag By Peril (Tornado)



Lag By Peril (Wind)




Lag By Peril (Explosion)
Lag By Peril (Fire)
Year of DOI
2007
2011
2012
2013
2014
2015
Lag By Peril (Hurricane)
Lag By Peril (Lightening)
Lag By Peril (Power Outage)
Lag By Peril (Rain)
Lag By Peril (Riot)
Lag By Peril (Thawing)
Lag By Peril (Thunderstorm)
Lag By Peril (Tropical Storm)
Lag By Peril (Weight of Snow and Ice)
% of Total Running Sum of Count...
100%
50%
0%
0
20
40
60
80
100
Lag

Appendix C: Peril vs. Line of Business Pivot Table

LOB:	CL_AUTO		CMP-PROP		HOME		IM		PL_AUTO	
peril	Claim Count	Avg Sum Incurred	Claim Count	Avg Sum Incurred	Claim Count	Avg Sum Incurred	Claim Count	Avg Sum Incurred	Claim Count	Avg Sum Incurred
WIND	2120	5,518.20	18838	27,861.29	130848	5,088.06	1636	57,830.97	21141	2,969.03
RAIN	33	3,895.81	446	11,281.83	3467	3,252.12	36	3,895.81	212	3,731.36
FLOOD	1644	5,726.59	15101	28,014.87	108686	5,726.59	1374	58,197.43	15968	3,049.19
FREEZING	207	5,590.75	4529	24,202.30	22833	3,936.33	312	68,500.71	1629	2,639.69
HAIL	1822	4,387.67	11023	33,237.45	91465	5,781.83	1076	58,523.51	18151	2,942.26
TORNADO	1469	5,619.14	12625	29,172.72	91890	5,422.68	1192	61,579.77	15663	2,978.47
ICE	221	5,441.95	4788	22,798.37	26794	3,740.85	352	64,369.07	2208	2,513.84
SNOW	237	4,312.49	4276	24,030.64	25314	3,756.87	326	69,071.84	2312	2,563.03
EARTHQUAKE	0	0.00	4	0.00	0	0.00	0	0.00	0	0.00
FIRE	1	0.00	16	306,730.90	2	0.00	3	13,393.33	0	0.00
EXPLOSION	0	0.00	6	66,475.02	19	24,147.35	12	38,990.68	0	0.00
WEIGHT OF ICE & SNOW	18	4,266.56	363	10,182.86	1384	1,959.83	75	122,423.64	144	2,202.10
THUNDERSTORM	62	1,946.60	128	21,289.60	3316	5,909.68	15	91,425.65	800	2,227.86
COLLAPSE	0	0.00	61	7,830.35	269	3,477.46	1	69,500.00	9	296.79
HURRICANE	514	20,671.84	6301	17,542.44	19540	3,449.08	650	80,014.29	1863	7,663.02
HURRICANE DOLLY	0	0.00	22	4,824.33	10	1,684.40	3	12,551.84	0	0.00
HURRICANE GUSTAV	12	4,382.77	1136	14,895.84	6047	3,281.55	52	26,338.36	244	2,675.95
HURRICANE IKE	22	4,586.45	1237	28,541.74	2981	3,109.09	140	91,579.61	79	4,106.31
TROPICAL STORM	2	4,260.39	34	2,894.63	143	3,951.89	3	0.00	43	3,097.15
TROPICAL STORM FAY	0	0.00	17	1,468.49	65	4,499.98	0	0.00	18	1,670.50
TROPICAL STORM HANNA	0	0.00	11	3,495.72	61	3,649.31	0	0.00	10	5,925.21
THAWING	0	0.00	178	19,065.74	576	6,567.19	4	4,482.00	3	1,568.93
LIGHTNING	24	2,606.60	73	27,196.88	2481	6,807.14	5	55,487.75	724	2,183.70
RIOT	0	0.00	6	308,587.14	0	0.00	1	1,630.00	0	0.00
CIVIL DISORDER	0	0.00	6	308,587.14	0	0.00	1	1,630.00	0	0.00
POWER OUTAGE	0	0.00	180	3,952.89	561	1,203.28	9	10,146.43	26	1,202.38

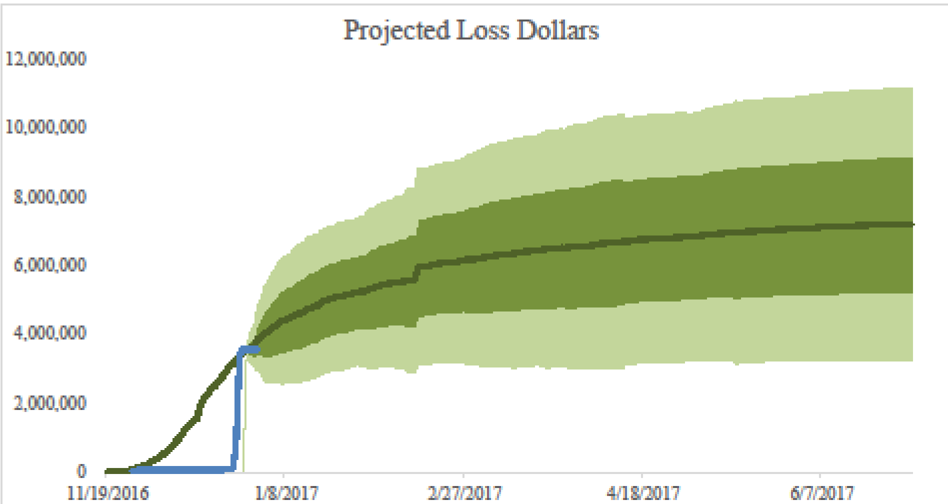
Appendix D: Lower-Triangular Peril Comparison Chart

	WIND	RAIN	FLOOD	FREEZING	HAIL	TORNADO	ICE	SNOW	EARTHQUA	FIRE	EXPLOSION	WEIGHT OI	THUNDERS	COLLAPSE	HURRICAN	HURRICAN	HURRICANI	HURRICAN	TROPICAL !	TROPICAL !	TROPICAL !	THAWING	LIGHTNING	RIOT	CIVIL DISOI	POWER OU
WIND																										
RAIN	2.40%																									
FLOOD	80.14%	100.00%																								
FREEZING	14.53%	14.85%	15.65%																							
HAIL	70.76%	0.00%	67.53%	11.73%																						
TORNADO	70.36%	0.00%	71.56%	13.47%	86.37%																					
ICE	17.36%	0.00%	19.24%	95.09%	1.45%	3.89%																				
SNOW	17.01%	0.00%	19.35%	87.61%	1.61%	4.05%	92.44%																			
EARTHQUAKE	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%																		
FIRE	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%																	
EXPLOSION	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%																
WEIGHT OF ICE & SNOW	1.14%	0.00%	1.39%	2.41%	0.57%	0.58%	4.32%	6.11%	0.00%	0.00%	0.00%															
THUNDERSTORM	2.48%	0.00%	0.57%	0.00%	1.50%	0.66%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%														
COLLAPSE	0.19%	0.00%	0.24%	0.00%	0.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%													
HURRICANE	6.88%	0.00%	8.41%	0.00%	0.00%	9.77%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%												
HURRICANE DOLLY	0.02%	0.00%	0.02%	0.00%	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.12%											
HURRICANE GUSTAV	4.29%	0.00%	5.25%	0.00%	0.00%	6.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.95%	0.00%										
HURRICANE IKE	2.55%	0.00%	3.12%	0.00%	0.00%	3.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	15.45%	0.00%	0.00%									
TROPICAL STORM	0.13%	0.00%	0.16%	0.00%	0.00%	0.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%								
TROPICAL STORM FAY	0.06%	0.00%	0.07%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	44.44%							
TROPICAL STORM HANNA	0.05%	0.00%	0.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	16.44%	0.00%						
THAWING	0.44%	0.00%	0.00%	2.58%	0.00%	0.00%	2.21%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%					
LIGHTNING	1.89%	0.00%	0.00%	0.00%	2.68%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	76.53%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%				
RIOT	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%			
CIVIL DISORDER	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%		
POWER OUTAGE	0.44%	0.00%	0.54%	0.00%	0.00%	0.00%	2.26%	2.39%	0.00%	0.00%	0.00%	39.11%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

Appendix E: Output Page

Projected Ultimates of CAT 79

This report details the projected loss amount and projected total claim count for CAT 79, given the inputs shown below. This report is based off of 269 storms with similar perils and geographies.



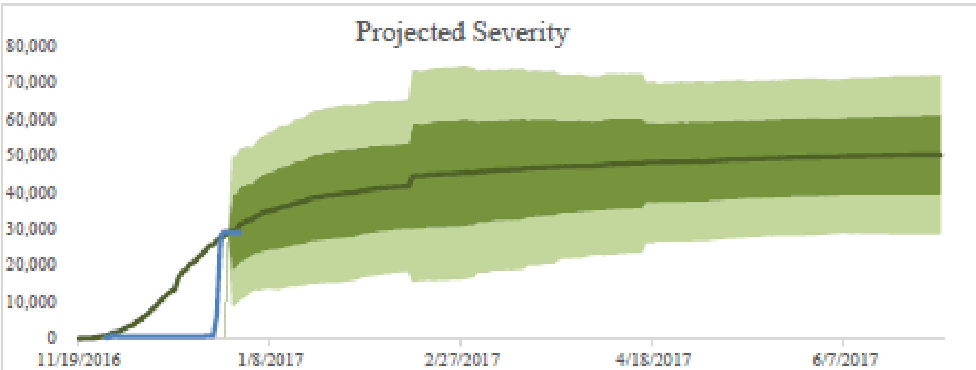
This graph shows the projected ultimate loss amount with a confidence interval of one and two standard deviations. The blue line shows the actual loss to date.

Projected Loss Dollars:	7,915,563
Confidence Interval (1 SD)	5,559,634 - 10,271,491
Confidence Interval (2 SD)	3,203,706 - 12,627,420

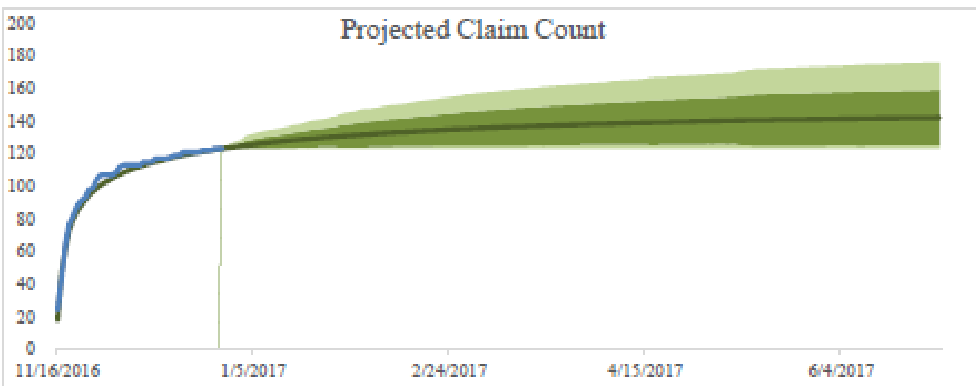
Claims above $100,000	
Count	0
Total Monetary Value	0

DOI	11/16/2016
Reporting Date	1/15/2020
Peril	Tornado/Hail
LOB	Home, Cmp-Prop, IM, Cl-Auto, Pl-Auto
Region	Northeast, Midwest, South, West
Claims Reported	4,587
Loss Reported	23,965,317
Current Severity	5,225
CATs Utilized	269

12/2/2016 1/3



This graph shows the projected severity per claim with a confidence interval of one and two standard deviations. The blue line shows the projected severity to date.

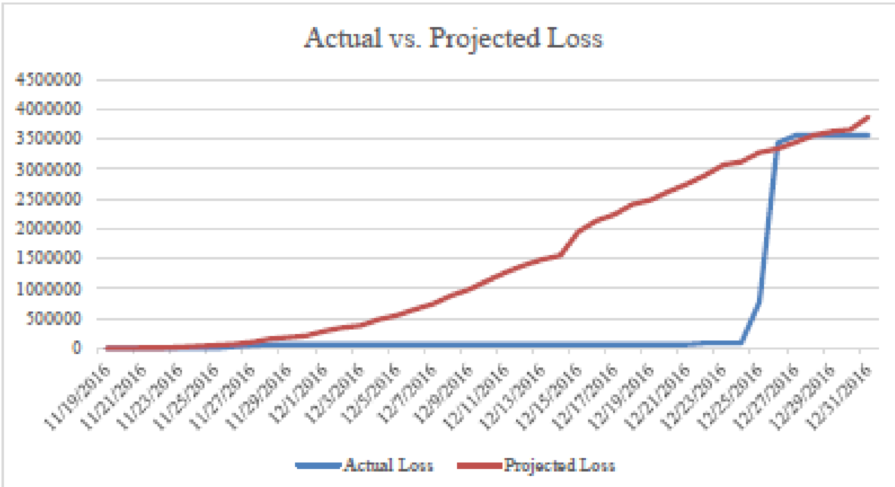


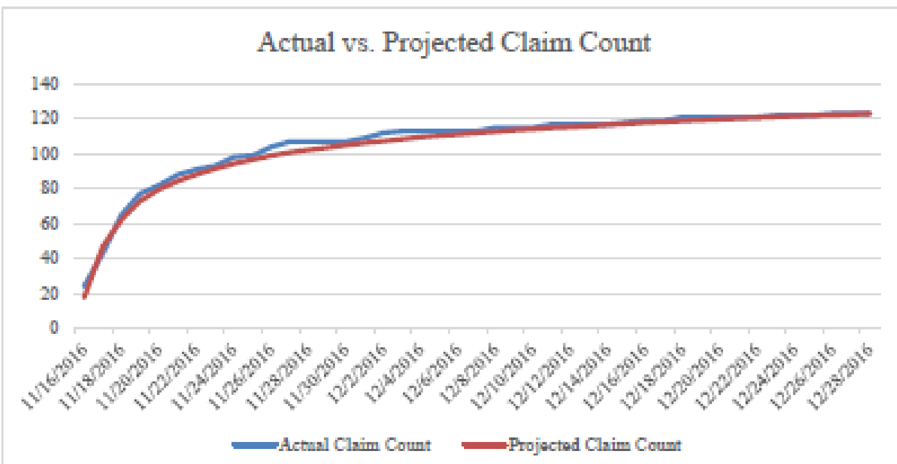
This graph shows the projected claim count with a confidence interval of one and two standard deviations. The blue line shows the actual claim count to date.

Projected Severity	53,519
Confidence Interval (1 SD)	46,345 - 60,693
Confidence Interval (2 SD)	39,171 - 67,867

Projected Claim Count:	147
Confidence Interval (1 SD)	103 - 190
Confidence Interval (2 SD)	103 - 190

Actuals vs. Projections





Appendix F: Documentation and Training Guide

The following documentation has been provided in a tab within the spreadsheet model so that a user may easily access the instructions while utilizing the model:

Instructions
The following sections detail how to use and interpret this CAT model.
Input Page:
The input page contains fields that require manually typed inputs, checkboxes, and automated entries.
The first manual field is the date of incidence (DOI) in which the user must type the earliest date on which the CAT or group of CATs occurred.
Next is another manual field where the user can specify up to five CAT numbers. The CAT number(s) entered will correspond to the particular storm(s) that the user wishes to model.
After entering this information, click the button labeled "Populate Suggestions and Claims to Date." This will populate a table of suggested selections for the Peril Type, LOB, and Region checklists described below. This button also populates the Claims to Date table which is described in detail below.
The user will then see a series of checklists labeled Peril Type, LOB, and Region:
• In the Peril Type list, the user can distinguish the type of storm that he or she wishes to model.
• In the LOB list, the user can specify the Line(s) of Business affected by the storm.
• In the Region list, the user can specify the region(s) of the country affected by the storm.
*Multiple boxes can be checked in each field based on the specifications of the storm being modeled
*Defining these criteria ensures that model bases its output only on past events that also exhibit that criteria, thereby offering an accurate model
The final portion of the input page is the 'Claims to Date' field. This table is auto-populated with current information about the CAT(s) specified in the CAT number field. The table will automatically display values of total claim count and total estimated loss dollars associated with the storm(s) on each lag day relative to the earliest date of incidence. If the user so chooses, they may choose to manually input the number of claims reported and loss dollars reported columns. Please note if you are manually inputting data, these columns are running sums.
Once all the information is entered the user simply clicks the button labeled "Create Report," at which point the user will be brought to a separate worksheet that displays the output report.
Output Page:
The output page will display a report that must be interpreted by the user.
The report is split into three sections:
1) Loss dollars projection to ultimate 2) Frequency and Severity projections to ultimate 3) Comparison of actual development to date and average historical development to date for loss dollars and frequency.
This report contains two types of graphs that display the results: (1) a "fan graph" used to show the ultimate projections for loss dollars, frequency, and severity and (2) a line graph to actual vs. average development comparison.

1) Each "fan graph" has a dark green line that represents average projected development. There is also a blue line that displays the actual development to date. Upon reaching the current lag day, the actual development line ends, but average projected development continues to ultimate. Also, at this point, each graph "fans out" into intervals of certainty. The solid green interval represents one standard deviation above and below the average development line, telling the user that the actual development pattern is likely to stay within this interval. The light green interval represents two standard deviations above and below the average development line, telling the user that the actual development pattern is highly likely to stay within this interval. Below these graphs are tables that correspond to each graph and display the ultimate projections numerically. The tables display the average ultimate projections in bold, with the one- and two-standard deviation intervals below. 2) The report contains three line graphs which correspond to the loss dollars development, frequency development, and severity development. Each graph contains two lines: a red line indicating average historical development to the current lag day and a blue line indicating actual development to date. By comparing these lines, the user can determine whether or not the current development pattern is drastically from what has been seen in the past.
An additional table in the first section of the report displays the input information for the the user to reference while interpreting the output.
Updating Data:
Data can be updated simply through the copy-paste function in order to keep the underlying data current and keep the model accurate.
The user must copy the frequency and severity data from their respective files kept locally at Hanover and paste the information into the data tabs of the CAT Model file.
Several calculated fields exist in the model that may not be present in the original data, but these formulas will be copied down properly upon running the model, thereby keeping the data consistent.