JANUARY 2020

ACTUARIES CLIMATE RISK INDEX

Preliminary Findings
January 2020

American Academy of Actuaries
Objective. Independent. Effective.™
ACTUARY.ORG

Acknowledgments

Primary author: Steve Jackson, Ph.D.
We would also like to acknowledge the contributions of the following individuals who assisted, primarily through their participation in the Climate Index Working Group, in the development of this research.

Doug Collins, MAAA, FCAS (Chair)
Rich Gibson, MAAA, FCAS
Steve Kolk, MAAA, ACAS
Caterina Lindman, FCIA, FSA
Stuart Mathewson, MAAA, FCAS
Dale Hall, MAAA, FSA
Yves Guérard, FCIA, FSA

In addition, we would like to acknowledge the contributions of several peer reviewers from both the actuarial and climate modeling communities who provided invaluable critiques and comments on earlier drafts of this paper.

Introduction

The Academy has long been the most reliable and credible source of objective, independent, and nonpartisan information about actuarial matters that can and do affect public policy decisions in the U.S. We have long sought to provide an objective voice about matters related to risks from climate, which is an area that can only benefit from objective and independent actuarial analysis. We are now releasing the Actuaries Climate Risk Index (ACRI) to provide that objective and independent analysis to assist in answering the question: Are the extreme weather conditions that result from a changing climate producing increased property losses?

The findings contained in version 1.0 of the ACRI are the culmination of years of research. We are presenting them now in the spirit of objective, transparent scientific inquiry and statistical rigor. This release is not a political statement. We fully understand and have heard from some who would prefer that actuaries make a political statement. This is not the Academy's mission or undertaking.

This project could not have happened without the tireless and dedicated work of Steve Jackson, Ph.D., the Academy's assistant director for research (public policy), as well as the members of the Climate Index Working Group. Many thanks for all their efforts to bring this project to fruition.

Release of version 1.0 of the ACRI is one that we and we anticipate other stakeholders will continue to build upon with the same objective, transparent inquiry and rigor; such is the nature of scientific investigation.

We welcome feedback and suggestions for enhancements. Please email acri@actuary.org with your thoughts.

Yours sincerely,

D. Joeff Williams	Shawna Ackerman
Academy President	Academy Past President

Executive Summary

In November 2016, the American Academy of Actuaries, the Canadian Institute of Actuaries, the Casualty Actuarial Society, and the Society of Actuaries launched the Actuaries Climate Index (ACI). The ACI provides an objective indicator of the frequency of extreme weather events and the extent of sea level change for 12 regions in the United States and Canada. This index is updated four times a year. Reflecting on the results of the ACI invites us to ask the question: Is there a statistical relationship between the weather components of the Actuaries Climate Index and damages to life and property caused by severe weather? This paper summarizes research to model this relationship. The Actuaries Climate Risk Index (ACRI) was developed from this model. This new index, the ACRI, is intended to measure the change in damages resulting from environmental conditions in excess of those observed in the reference period, as measured by the ACI.

In undertaking this effort, the American Academy of Actuaries is mindful of the results and the messages offered by prior research. First, that losses due to extreme weather events are large and increasing, yet most of the losses are due to increasing wealth and population yielding increased exposure to risk. Second, that estimates of loss due to extreme weather have been, are, and are likely to be very imprecise, yet imprecise results may be useful.

The examination began by looking at the relationship between environmental variables, as captured in the ACI components, and losses captured in publicly available databases that matched the ACI's geographic and time reference periods. For the United States, the SHELDUS[1] database (built largely with data from the National Oceanographic and Atmospheric Administration [NOAA] Storm Database) was identified as the most appropriate set of data. For Canada, the Major Storms database was identified as the best available dataset to use for the analysis. However, due to the limited number of events covered by the Canadian database, it was decided for version 1.0 of the ACRI to restrict attention to the United States and its seven regions. Moreover, the model thus far has only been developed to quantify impacts on property losses, although the same framework is believed likely to perform similarly for deaths and injuries.

1 Spatial Hazard Events and Losses Database for the United States.

To find the best correlation between weather variables and property losses, the impact of inflation, exposure, region, and seasonality have been controlled for. The Academy has have analyzed a dependent variable expressing losses in dollars and have treated each month of each year for each region as a separate observation. To allow for non-linearity in relationships between weather conditions and losses, to allow for interaction among weather conditions, and to mitigate the impact of the highly skewed distribution of losses, a model has been estimated in which both independent and dependent variables are log-transformed. To identify statistically significant parameters, the Academy used backward regression on the dependent variable and the ACI to select the best estimated model in which all parameters were statistically significant at the 90 percent confidence level.

Based on this estimated relationship between the ACI and losses, the ACRI is calculated as the difference in modeled losses due to ACI components being above (or below) their reference period mean values. In order to exclude the impact of changes in exposure on the ACRI, the reference period mean modeled losses are exposure-adjusted. The resulting ACRI totals $24 billion during the post-reference period, 1991–2016, equal to approximately 3.3 percent of the exposure-adjusted losses during that period.

The model has a large amount of uncertainty, because each region-month currently only has 56 data points on which to base the parameters, 30 points during the reference period and 26 points subsequent to the reference period. The Academy has estimated uncertainty in two ways. Based on the intrinsic uncertainty associated with the regression estimates from which the ACRI is built, a 90 percent confidence interval is estimated around the best estimate for total ACRI losses of $16 billion to $36 billion. However, the broader, extrinsic uncertainty associated with only having one "draw" of the weather distributions, both for the reference and the post-reference periods, has been estimated using a stochastic model of synthetic datasets based on randomly selected observations from the original data. With this broader definition of uncertainty, it is estimated that a 90 percent confidence interval for total ACRI losses stands at $2 billion to $45 billion. Of course, even these two measures of uncertainty are somewhat uncertain. There are several ways in which these confidence intervals for both intrinsic and extrinsic uncertainty could have been created, and different methods might produce materially different estimates.

Weaknesses and limitations are outlined throughout this documentation that serve as cautionary notes, pointing to the need to interpret these current results in light of their inherent uncertainty. Chief among these limitations are:

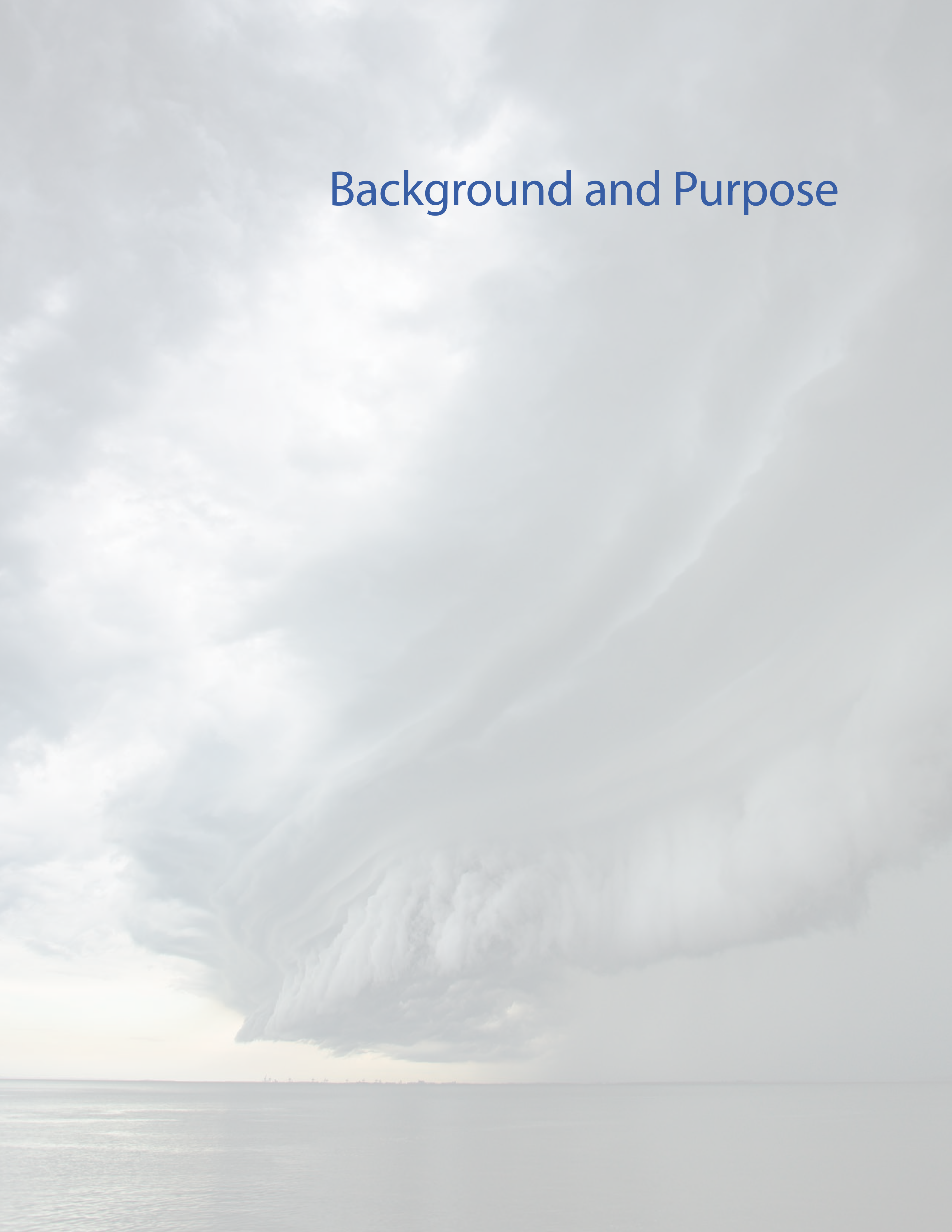
- As noted, while the model has an r-squared of 0.62 on log-transformed values, the r-squared on dollars of modeled and actual losses is only 0.03.
- The model performs most dependably at the national level, less so at the regional level (mean r-squared equals 0.36), and even less well at the region-month level (r-squared equals 0.24).
- The ACI metrics used in the model are averaged over large geographic areas, while the most damaging events are concentrated in much smaller areas.
- The ACI metric for wind, based on average monthly wind speeds in these large geographic areas, is not shown by the model to be a close estimate of large losses, which are driven primarily by windstorms.
- Equation coefficients are quite inconsistent from one month to the next in a given region, which does not provide a logical explanation for the ACRI values.

These weakness and limitations also spur the Academy to proceed to version 2.0 of both the ACI and the ACRI to seek better data and develop more effective metrics and more robust analysis. Others are encouraged to build on this work by conducting research using weather metrics and proprietary insurance company loss data, which would be available in precise geographic detail.

Actuaries Climate Risk Index: Preliminary Findings

Table of Contents

Background and Purpose



Background

In November 2016, the American Academy of Actuaries, the Canadian Institute of Actuaries, the Casualty Actuarial Society, and the Society of Actuaries ("the ACI Actuarial Associations") introduced the Actuaries Climate Index (ACI). The ACI provides an objective indicator of the frequency of extreme weather events and the extent of sea level change. The six components of the ACI are drought, precipitation, high temperature, low temperature, sea level, and high winds. The data for each category are standardized in respect to a reference period, 1961–1990, and those standardized values are combined to produce the ACI. On the ACI website (actuariesclimateindex.org), both the components' index values and the composite index value (an amalgam of the components) are provided for seven regions within the U.S. and five regions within Canada, as well as for both countries as a whole. The website also documents the methodology and data used to develop the indexes. The data is updated quarterly and presented both monthly and by meteorological season.

The ACI documents the variation in a set of extreme weather and hydrologic measures across time and place. Having the ACI invites us to ask the question: How much damage is done to life and property when the distribution of environmental events differs from those observed during a reference period, 1961–1990? The Actuaries Climate Risk Index (ACRI) seeks to answer that question. Based on identified relationships between the atmospheric and hydrologic conditions assessed in the ACI and data on harm to people and damages to properties due to climate-related events, the ACRI is intended to estimate the impacts resulting from environmental conditions in excess of (or below) the average in the reference period. Just as the ACI reports values monthly and by season for each of seven regions in the U.S. and five regions in Canada, the ACRI aims to also reflect the damage done monthly and by seasons in the same regions.

In November 2015, the ACI Actuarial Associations received a report commissioned from Solterra Solutions describing a procedure for creating the ACRI. Data from the ACI was combined with data from the SHELDUS database on losses from storm events in the U.S., and from the Major Storms database in Canada, to assess relationships between environmental conditions and losses. For each region, Solterra looked for the best fit between an element of the ACI (e.g., wind) and losses from associated events (e.g., floods). Based on the set of best-fitting regressions, Solterra created an index for the ACRI that allowed information from the different regions, based on different environmental conditions, to be combined. While never endorsed by the ACI Actuarial Associations, nor launched on the ACI website, this method was the basis for numerous presentations in recent years. This index will be referred to as ACRI version 0.1.

After identifying weaknesses in version 0.1 and receiving peer review inputs from the Institute and Faculty of Actuaries (U.K.), the Academy decided to create version 1.0, responding to the five main limitations in the regressions that serve as the ACRI's foundation version 0.1:

1. Introduce control for the risk exposure and other intervening factors in the modeling of the relationship between weather and hydrologic conditions and losses;
2. Find statistically significant relationships that accounted for a reasonable share of the variation observed;
3. Move from a 1-10 scale to a dollar scale;
4. Look at regressions that included multiple environmental variables simultaneously rather than one variable at a time; and
5. Improve the analysis of heat-related losses.

ACRI version 1.0 uses the same data sources that Solterra used for ACRI version 0.1, which are still considered the best readily available databases.

As the Intergovernmental Panel on Climate Change (IPCC) noted in its 5th Assessment in 2014, "Studies analyzing changes in climate variables and insured losses in parallel are still rare."[2] But, as the Government Accountability Office (GAO) concluded in its 2017 analysis of climate change and economic losses, "Methods used to estimate the potential economic effects of climate change in the United States—using linked climate science and economics models—are based on developing research. The methods and the studies that use them produce imprecise results because of modeling and other limitations but can convey insight into potential climate damages across sectors in the United States."[3]

This paper will discuss the most closely related comparable efforts, but, attempting to create a sustainable index that can be easily updated quarterly and that is reflecting damages from extreme or moderately extreme environmental conditions at the regional level within the U.S. and Canada has not been done before. The Academy intended to build ACRI 1.0 using the best practices developed for similar efforts. This version of the ACRI is expected to be updated to version 2.0 as soon as the Academy can explore additional data sources, environmental metrics, and methods of analysis.

The following sections describe in detail the methodology, choices made, and data used to create ACRI version 1.0.

2 IPCC, 2014: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
3 GAO 17-720, "GAO Highlights," 2017.

Normally, building indexes involves four tasks:

1. Deciding what modeling approach to use;
2. Deciding what data to use;
3. Modeling the relationship between climate events and damage in a robust manner; and
4. Constructing the index (or versions of the index).

However, given the novelty of this effort and the limited relationships found, two more tasks are added as follows:

5. Subjecting results on the relationship between environmental conditions and damages to deeper scrutiny;
6. Examining other models of the relationship between weather conditions and damages.

Purpose

The Actuaries Climate Risk Index (ACRI) is intended to provide an actuarial perspective on impact on property and human death/injury of extreme or moderately extreme environmental conditions.

Using measures of four categories of weather conditions, the ACRI is built upon a modeling of the relationship between those changing climate conditions and the damage done by climatological events. Controlling for risk exposure and regional and seasonal factors, the ACRI produces a measure of the effect on economic losses of deviations from benchmark climate conditions in each region, in each season.

A robust version of the ACRI might be useful to different audiences in different ways:

- For the **general public**, the ACRI can provide a means to understand to what extent extreme climate events and their increasing frequency have been correlated with economic losses, allowing for a greater understanding of the impact of climate on costs involved.
- For **public policymakers**, the ACRI can provide a measure that may be useful in leveraging the costs of prevention and mitigation policies.
- For **public and private decision-makers**, it can provide a base for planning the capacity to assume larger risks associated with changes in environmental conditions.
- For **actuaries**, the ACRI can provide insight into the risks potentially associated with extreme or moderately extreme climate events. Information on potential losses due to the increasing frequency or severity of extreme events helps with setting parameters for those losses in stochastic models used to project possible losses in the future. New considerations for increasing contingency margins could also result. However, by nature an index such as the ACRI provides information at a macro level but not at the level of granularity that is required for reserving or pricing. For these purposes, actuaries would need to adapt the ACRI methodology by incorporating more specific information about environmental conditions and insurance claims related to a specific purpose.

While ACRI 1.0 provides estimates of the impact of extreme environmental conditions, and assessments of the uncertainty there of, it is not the robust version that is sought.

The conclusions of prior research have been considered in presenting the ACRI 1.0 and the results of this analysis:

- Losses due to extreme weather events are large and increasing; a recent study by the Universal Ecological Fund (FEU) estimated that between 2007 and 2017, annual losses from extreme weather events in the United States averaged $42 billion.[4]
- Most of the losses are due to increasing wealth and population yielding increased exposure to risk; as the IPCC 2014 Report concludes: "Economic costs of extreme weather events have increased over the period 1960–2000. … However, the greatest contributor to increased cost is rising exposure associated with population growth and growing value of assets."[5] One of the few studies that has sought to quantify the relative impacts of climate events and exposure on observed, as opposed to future, losses for the United States as a whole over a long period of time concluded: "[T]he increase in losses due to socio-economic changes was approximately three times higher than that due to climate-induced changes."[6]
- Estimates of loss due to extreme weather have been, are, and are likely to be very imprecise; as the GAO concluded from its review of prior research seeking to assess the economic impacts of climate change and discussions with experts, "the methods produce imprecise results."[7] As elaborated by Schmidt et al., "[I]t is generally difficult to obtain valid quantitative findings about the role of socio-economics and climate change in loss increases. This is because of criteria such as the stochastic nature of weather extremes, a shortage of quality data, and the role of various other potential factors that act in parallel and interact."[8]
- Imprecise results may be useful. As the GAO noted, imprecise results "can convey useful insight into broad themes about potential climate damages across sectors in the United States. For example, according to several experts interviewed, these methods can provide valuable research information about the potential magnitude of economic effects and potential areas of greatest concern, including where assets may be at greatest risk. Some other experts told us that using the methods can help identify areas where additional research would be most useful."[9]

4 Sir Robert Watson, Dr. James J. McCarthy and Liliana Hisas, "The Economic Case for Climate Action in the United States," Universal Ecological Fund (FEU), September 2017.
5 IPCC, 2014: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
6 Schmidt, Silvio; Kemfert, Claudia; Höppe, Peter (2008): "The impact of socio-economics and climate change on tropical cyclone losses in the USA, DIW Discussion Papers, No. 824, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.
7 U.S. GAO, "Climate Change: Information on Potential Economic Effects Could Help Guide Federal Efforts to Reduce Fiscal Exposure," September 2017.
8 Schmidt, op. cit.
9 U.S. GAO, "Climate Change: Information on Potential Economic Effects Could Help Guide Federal Efforts to Reduce Fiscal Exposure," September 2017.

Overview of Models of Weather-Related Risks to Property and Lives

There are at least four types of models that attempt to relate extreme climate events to socioeconomic harm. First are the catastrophe models now routinely used by insurers to estimate the property insured losses that are likely to occur as a result of natural disasters such as earthquake, flood, convection and snowstorms, etc. Second are the integrated models that underline IPCC's periodic assessments. Third are the social-cost-of-carbon models that were used by the Interagency Working Group on Social Cost of Greenhouse Gases (United States Government) to try, among other things, to provide metrics for evaluating environmental regulations, based on the economic damage done by increasing levels of greenhouse gases. Fourth are the models used to generate the Disaster Risk Index for the United Nations Development Programme (UNDP), which aims to assess the number of deaths resulting from natural catastrophes, taking into account the varying levels of socioeconomic development in different countries.

Catastrophe models proceed in three stages to estimate the likely costs of natural disasters.[10] In the first stage, the probabilities of certain events occurring at certain magnitudes and/or frequencies are calculated. As a second step, the physical damage that would occur in a particular region if it were subjected to a certain magnitude of events is estimated, given the characteristics of both the built and the natural environment. Finally, in the third stage, the insured losses that would occur given specifics insurance protection and conditions are estimated. This type of model requires detailed data not available for the ACRI intended purposes.

The integrated models used by the IPCC rely on underlying research that establishes likely consequences of particular climate events on particular outcomes.[11] The models then aim to integrate the consequences of these various events, taking into account the interactions among the climatological events and, in principle, among the effects—interactions which might either increase or decrease the magnitude of impacts. In the context of ACRI, the limitation of these models is that they depend on establishing a very large number of relationships between discrete past and future environmental events and discrete past and future harms in a large number of locations. These integrated models also depend on establishing or assuming the interactions among causes and effects and again, in ways that might differ in different locations. As the ACRI is intended to be an objective, retrospective measure of the relationship between environmental effects and economic losses, this type of model was put aside.

10 See AIR description or RMS description of Catastrophe Modeling.
11 See Intergovernmental Panel on Climate Change, *Fifth Assessment Report*, 2014, Chapter 10, page 681.

Social-cost-of-carbon models estimate the harm done by the increases in surface temperature driven by the increases in the level of greenhouse gases over a reference period point in time.[12] These models are built on substantial underlying research on the impact of temperature on various elements of the ecosystem and the socioeconomic environment. They also include varying degrees of interaction effects among the varying ecological and socioeconomic elements. Given that these models are forward-focused and depend on strong underlying research, they do not fit the purposes for which the ACRI is being developed.

Finally, models aiming to broadly assess the impact of extreme events for large numbers of countries are exemplified by the Disaster Risk Index developed for the UNDP. The general objective here was to develop a method of estimating how many deaths would occur in each country based on climate-related natural disasters given the socioeconomic status of the country. Compared to the previous three classes of models, this is a simpler modeling effort aiming to broadly illustrate the impact of varying socioeconomic conditions on the relationship between natural disasters and lives lost. While the ACRI aims to assess damage to property as well as number of deaths for two countries, the Academy concluded that the general objective of the Disaster Risk Index is similar to that of the ACRI.

Leveraging these developed models while keeping in mind the specific objectives underlying the ACRI model, the relationship, if any, between environmental conditions and economic losses or deaths from environmental events is defined in the following general functional form:

Equation (A) Loss = f(Risk Exposure, Environmental conditions, Geography, Season)

12 See the 2016 update to the Technical Documentation of the SOC models.

Data Used in Constructing the ACRI

The model requires data on damages due to environmental events given risk exposure, geographic variability, and seasonality. Those factors—risk exposure, geography, and season—set the value and resilience of the built environment, and therefore might shape the relationship between environmental conditions and losses observed.

Economic Damages: While insurers have accurate data on covered losses (both in property and life insurance) generated by insured climate events, their data do not include losses that were not covered, and, as a result, provide an incomplete picture of the damage done. Further, the data is generally proprietary.

For the United States, the NOAA Storm Events Database documents:

> "The occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce; Rare, unusual, weather phenomena that generate media attention, such as snow flurries in South Florida or the San Diego coastal area; and Other significant meteorological events, such as record maximum or minimum temperatures or precipitation that occur in connection with another event."[13]

The SHELDUS database[14] builds upon the foundations of the NOAA Storm Events data and, especially for the years prior to 1996, supplements that data with occasional additional reports.

While the NOAA database contains information on more than 50 types of incidents, the SHELDUS database reduces that to 18 categories, of which 17 are relied on for the ACRI: Avalanche, Coastal, Drought, Flooding, Fog, Hail, Heat, Hurricane/Tropical Storm, Landslide/avalanche, Lightning, Severe Storm/Thunder Storm, Tornado, Tsunami, Volcano,[15] Wildfire, Wind, and Winter Weather. For each event, in addition to the date and location of the event, the database reports estimates for property damage, crop damage, lives lost, and injuries. Because these categories do not match exactly the ACI weather event types, the Academy decided to examine the relationship between all ACI components and all losses (from these specified sources) in modeling. As a result, the measure of losses for a particular region in a particular month is the sum of all damage done from all reported events, in constant 2016 dollars.[16]

13 NOAA Storm Events Database.
14 SHELDUS Database.
15 Losses from volcano eruptions will be excluded in loss totals in the next iteration of this analysis. Overall, volcanos accounted for less than 1 percent of the reported losses over the time period, 1961–2016. Geophysical losses were excluded.
16 Controlling for inflation was approximated by using constant 2016 U.S. dollars. This year was selected because it was the most recent year for which ACI and loss data existed when modeling began.

The SHELDUS database has advantages as well as disadvantages. The advantages include: data coverage from the present back to before 1961, the starting point for the ACI reference period;[17] coverage of losses from a wide range of weather event types; losses are designated at the county level; losses include property losses, crop losses, lives lost, and injuries sustained. Disadvantages include: concerns about the completeness of reporting of events, especially prior to 1996; and concerns about the reported losses. Property losses for hurricanes and wildfires are less than those reported by the Insurance Industry Institute.[18] Nonetheless, of the publicly available databases, the SHELDUS database (or the NOAA Major Storms Database on which SHELDUS is largely based) is the best for the ACRI's purposes.

For Canada, the best source of comparable data is the Canadian Disaster Database (CDD), which

> "contains detailed disaster information on more than 1000 natural, technological and conflict events (excluding war) that have happened since 1900 at home or abroad and that have directly affected Canadians. The CDD tracks 'significant disaster events' which … meet one or more of the following criteria: 10 or more people killed; 100 or more people affected/injured/infected/evacuated or homeless; an appeal for national/international assistance; historical significance; or significant damage/interruption of normal processes such that the community affected cannot recover on its own."[19]

While this database contains information on almost 60 categories of disasters, 13 of them are "Meteorological/Hydrological." Of those, 10 of which were selected that are most likely to result from the categories of climate events incorporated into the ACI: Cold Event, Drought, Flood, Heat Event, Hurricane/Typhoon/Tropical Storm, Storm Surge, Storms and Severe Thunderstorms, Tornado, Wildfire, and Winter Storm. As with the data for the U.S., the damage done by all of these events have been added together into a single measure for a given region in a given month. Unfortunately, during the entire time period covered by the analysis, 1961–2016, there were only 275 region-months in Canada (of 3,360 region-months) with nonzero losses.[20] This small quantity of data made it impossible to estimate credible relationships for the Canadian regions using the same model as in the U.S. Because of this data constraint, ACRI 1.0 is focused solely on the United States.

17 Availability of data was a major factor in setting 1961 as the beginning of the reference period for the ACI.
18 Calculations by author from SHELDUS data compared, for example, to losses cited in "Facts + Statistics: Wildfires," Insurance Industry Institute (accessed Dec. 19, 2019), and "Facts + Statistics: Hurricanes," Insurance Industry Institute (accessed Dec. 19, 2019). More generally, issues with the reliability of SHELDUS data (compared to that of other databases) is well covered in: "When Do Losses Count? Six Fallacies of Natural Hazards Loss Data," Melanie Gall, Kevin A. Borden, and Susan L. Cutter, *Bulletin of the American Meteorological Society*, June 2009.
19 Canadian Disaster Database.
20 The limited number of reported losses in Canada is due to the definition of "disasters," compared to the tracking of storms and weather events in the US data. In contrast to Canada, fewer than 8 percent of region-months in the U.S. had zero losses; more than 92 percent had nonzero losses.

Risk Exposure: Adapting a well-established method proposed by Collins and Lowe,[21] the value of property at risk of damage in each region in the U.S. in each month has been approximated. On the assumption that residential property value is correlated with total property value, the ACRI uses median house prices multiplied by the number of housing units, which is available from the Census Bureau on an annual basis at the state level. That data has been aggregated to the region level, as defined for the ACI, and interpolated to obtain values for each month. Exposures are then expressed in constant 2016 U.S. dollars.[22]

Environmental Conditions: While constructing the ACI, the ACI Actuarial Associations evaluated many sources of environmental data and many ways to construct indicators of extreme weather events.[23] In that process, they settled on indicators for six categories of weather and hydrologic events: Drought, Precipitation, High Temperature, Low Temperature, Sea Level, and Wind. To combine information about these disparate categories into a single index (the ACI), the ACI Actuarial Associations decided to standardize the individual measures by relating the raw measure to the mean and standard deviation of that measure for a particular region and a particular month during a reference period, from 1961–1990.

Because the manner in which ACI components capture weather and hydrologic "extremes" is highly relevant to the way in which the meaning of the ACRI is interpreted, it is important to detailed here how they were constructed. Consider T90, the indication of temperatures above the 90[th] percentile, keeping in mind that all components were constructed the same way.

T90 is a temperature metric included in the GHNCNDEX dataset24 and is equal to the percentage of days in a month for which the temperature of the day falls above the 90th percentile of the distribution of daily high temperatures during the reference period. It is tabulated for each month and for each region over the 1961–1990 time period. A separate temperature distribution is used for each of the 12 calendar months and for each geographic location.

21 Douglas Collins and Stephen Lowe, Collins (2001). "A macro validation dataset for U.S. hurricane models." Casualty Actuarial Society, *Winter Forum*, pp. 217–52.
22 This was one of eight measures of exposure that was tested, and which improved the correlation of estimates most. The other seven were Population; Housing Units; Total Income, computed by multiplying each county's per capita income by its population, and then summing across all counties in an ACI region; Net Worth, computed from national net worth by assuming that wealth is proportional to the population in each region; Net Worth, computed from national net worth by assuming that wealth is proportional to the number of housing units in a region; Fixed Assets and Consumer Durable Goods, computed from national data by assuming proportionality with number of housing units in a region; Value of Housing Stock, equal to Median house price * number of housing units, using national-level house price data; Value of Housing Stock, equal to Median house price * number of housing units, using state-level house price data.
23 See the *Actuaries Climate Index Development and Design* for details on data sources and measurement techniques.
24 Global Historical Climatology Network (GHCN) Daily, from the NOAA Satellite and Information Service, is an integrated database of daily climate summaries from land surface stations across the globe. The grids each cover a surface area of 2.5 degrees longitude by 2.5 degrees latitude. GHCNDEX is a dataset based on GHCN Daily. It provides gridded, station-based indices of temperature- and precipitation-related climate extremes and was developed by the Climate Change Research Centre and the Australian Research Council's Centre of Excellence for Climate System Science. (Donat, M.G., L.V. Alexander, H. Yang, I. Durre, R. Vose, J. Caesar, "Global Land-Based Datasets for Monitoring Climatic Extremes," *Bulletin of the American Meteorological Society*, July 2013.)

The GHCNDEX data is geographically "gridded." The grid consists of pairs of latitude and longitude points with 2.5 degrees of spacing between them. Each of the ACI's geographic regions contains numerous GHCNDEX's grid points. The T90 data for these grid points is averaged across each ACI's geographic region separately for each of the 12 calendar months, thus producing 12 time series for each region. Each of these monthly time series is then standardized by region by (i) subtracting the region's mean value of T90 computed across the 1961–1990 reference period, and (ii) by dividing the result by the region standard deviation of T90 computed across the reference period. A positive standardized value indicates that a particular T90 value is above the reference period mean, while a negative value indicates that it is below the reference period mean.

If the reference period means of the ACI components represent the cutoff between ordinary and extreme in weather conditions, that would allow us to interpret without qualification positive values of the ACI as extreme and the impact of those positive values as the impact of extreme weather conditions. Unfortunately, it is not quite that simple. If a month has 30 days, then, on average, 3 of those days should exceed the 90th percentile temperature during the reference period. But, that means that some months within a time series will have 0, 1, or 2 days and others will have 4, 5, 6 or more days exceeding the 90th percentile temperature. If the reference period percentage of days above the 90th percentile is defined as the average of all months including those with fewer and those with more than the average, the reference period will capture extreme conditions in certain ways, but not in others. In other words, T90 captures the number of days which exceed the 90th percentile of the temperature distribution, thus providing an initial basis for measuring extreme conditions. But, when results are averaged across all months, both extreme months and non-extreme or ordinary months are included—thus moderately extreme weather conditions are being measured rather than extreme weather conditions. In this paper, references to "extreme" or "unusual" weather conditions indicate these "moderately extreme" conditions which the ACI currently captures.

While different data or different measures of climate events for the ACRI could have been sought, it seemed closest to the intent of the ACRI to rely as much as possible on the same data and the same measures as were (and are) created for the ACI. While models were tested on standardized data and other data sources were explored, the results obtained were not as strong as those obtained using the unstandardized measures of extreme or moderately extreme climate events underlying ACI and were sufficient for ACRI version 1.0. It should be noted that this version of the ACRI does exclude two of the six ACI-measured conditions: Sea Level and Continuous Dry Days (Drought). Sea Level was excluded because it was not measured for one of the U.S. regions; only those measures which existed for all regions were included. Moreover, preliminary analyses did not indicate very much explanatory power was lost by the exclusion of Sea Level.[25] Continuous Dry Days was excluded because it is the only element in the ACI that comes from annual data and is then interpolated to obtain monthly values. Perhaps due to the relative infrequency of the data, this variable also proved, in preliminary analyses, to be without significant explanatory power.

Geography: Because the relationship between extreme weather events and property losses varies by geographic region, the ACRI controls for geographic variability by estimating the parameters for each region separately.

Seasonality: Explicitly represented in the ACRI model Equation (A), seasonal effects—where, for example, high temperatures are more likely to cause damage during summer months than during winter months—is implicit in the definition of hazards introduced by Peduzzi in the Disaster Risk Index modeling.[26] Rather than combining the measures for losses, exposure, geography, and seasonality into a single hazard measure, months of the year are used as a way to control for the seasonal variation in damage resulting from environmental conditions. The ACRI controls for exposure, geography, and seasonality by estimating parameters separately for each region-month combination, while including exposure as an independent variable in the model.[27]

25 Sea level was primarily removed because of its absence from one of the ACI regions. If it can be incorporated into future versions of the ACRI, sea level may well have significant impacts.
26 P. Peduzzi, H. Dao, C. Herold, and F. Mouton, "Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index," *Natural Hazards and Earth System Sciences*, (2009) 9, 1149–1159.
27 For some univariate and bivariate views of the data elements, see Appendix 1.

Modeling the Data

To assess the relationship between weather and losses described in Equation (A) above, the ACRI relies on the formulation of Peduzzi,[28] shown in Equation B:

(B) $Loss = I * Exposure^{e} * Precipitation^{p} * Low\ Temperature^{l} * High\ Temperature^{h} * Wind^{w}$

Where:

Loss: Property losses in dollars for a particular region in a particular month;

I: Intercept which scales losses to account for factors other than those included in the model;

Exposure: an estimate of the property value at risk in a given region in a given month;

Precipitation (Rx5Day): the maximum 5-day precipitation in the month;

Low Temperatures (T10): the change in frequency of colder temperatures below the 10th percentile, relative to the reference period of 1961 to 1990;

High Temperatures (T90): the change in frequency of warmer temperatures above the 90th percentile, relative to the reference period of 1961 to 1990;

Wind: Daily average wind speed measurements are converted to wind power, which is proportional to the cube of the wind speed. Wind power is used because impacts from high winds (i.e., damages) have been shown to be more closely related to the cube of wind speed. The procedure used for temperatures is followed, by finding the 90th percentile of wind power for each month or season and subtracting the 90th percentile of wind power for that month over the reference period.

$^{e,\ p,\ l,\ h,\ w}$: If statistically significant, these are the exponents corresponding to the independent variables and reflect the sensitivity of loss to changes in these variables.

Taking the natural log of both sides of equation (B) does not change the equality and produces an equation estimable by linear regression, as shown in equation (C):

28 P. Peduzzi, op cit.

(C) $Ln(Loss) = ln(I) + e*ln(Exposure) + p*ln(Precipitation) + l*ln(Low Temperatures) + h*ln(High Temperatures) + w*ln(Wind)$.

The log transformation of the dependent variable, Loss, is useful, given the distribution of Losses. Examination of that distribution revealed a significantly positively skewed distribution, as seen in Table 1 below.[29] To reduce the skew while preserving as much as possible of the remaining characteristics of the distribution, the log transformation is a common procedure. The resulting distribution, also shown in Table 1, is materially less skewed.

Table 1: Distributions of Monthly Losses, Original and Log-Transformed

	Loss ($)	Ln(Loss) ($)
Min	0	0
Median	5,479,140	16
Mean	139,947,401	19
95^{th}	355,898,211	20
99^{th}	2,173,762,739	21
Max	92,905,914,368	25
Skewness	42.80	-1.66
Coefficient of Variation	11.86	0.35

The parameters of equation (C) for each region-month have been estimated over the time period 1961–2016, using a pooled cross-sectional time series analysis. In this form of estimation, an assumption was made that excluded factors have a common distribution of impacts in all region-months. This effect is captured by a shared error term and by a general intercept for the equation as a whole. But, the Academy further assumes that all of the included variables—both weather-related and exposure—have impacts specific to the particular region-month. Dummy variables were used for both intercepts and slopes to pool the region-months into a single equation and have used backwards regression (with a 90 percent confidence level) to identify statistically significant parameters. The model estimated, with its dummy variables for each region, is as follows:

29 For more detailed analysis of the Loss variable, see Appendix 1.

(D) $Ln(Loss) = Dum_{11}*ln(I_{11}) + \ldots + Dum_{712}*ln_{(I712) +}$

$Dum_{11}*e_{11}*ln(Exposure) + \ldots + Dum *e_{712}*ln(Exposure) +$

$Dum_{11}*p_{11}*ln(Precipitation) + \ldots + Dum_{712}*p_{712}*ln(Precipitation) +$

$Dum_{11}*l_{11}*ln(Low\ Temperatures) + \ldots + Dum_{712}*l_{712}*ln(Low\ Temperatures) +$

$Dum_{11}*h_{11}*ln(High\ Temperatures) + \ldots + Dum_{712}*h^{712*}ln(High\ Temperatures) +$

$Dum_{11}*w_{11}*ln(Wind) + \ldots + Dum_{712}*_{w712}*ln(Wind).$

Where:

$I_{11} \ldots I_{712}$: Intercept for region 1, month 1 … region 7, month 12 which scales losses to account for factors other than those included in the model;

e_{11} **…** e_{712}**:** If statistically significant, these are the estimated exponents for region 1, month 1 … region 7, month 12 for exposure, and similarly for the four weather components.

The pooled, cross-sectional model produced an r-squared of 0.63, an adjusted r-squared of 0.62, and a Durbin-Watson statistic of 1.76.[30] However, a measure of heteroskedasticity proposed by MacKinnon and White indicated rejection of the null hypothesis of homoskedasticity at the 99.99 percent confidence level.[31] As a result, the equation was re-estimated with a consistent, adjusted covariance matrix, as suggested by MacKinnon and White. The results reported are those derived from these corrected estimates.

The corrected, pooled, cross-sectional model produced an r-squared of 0.62 and an adjusted r-squared of 0.61. When the parameters estimated from the pooled data are applied to the data by region, the mean r-squared for the seven regions is 0.36. In other words, about 60 percent of the explained variation in losses occurs within regions while the other 40 percent of the variation is explained by differences across regions. When those same parameters are applied to the data by region-month, with 56 observations per sub-sample (one for each year, 1961–2016), the mean r-squared is 0.24. These results, at different levels of observation, suggest the model is strong at the national level, a little weaker at the regional level, and weaker still at the region-month level.

30 The Durbin-Watson statistic indicates a likelihood of positive serial correlation (probability of rejecting the null hypothesis = 0.0001). With a first order autocorrelation of 0.12, the Academy elected to use the Cochrane-Orcutt procedure to correct for this autocorrelation. The modeled values from the corrected regression produce an r-squared of 0.99 when regressed against the modeled values from the original estimation, a result that suggests that serial correlation is not a material issue. In addition to the Durbin-Watson test and evidence of heteroskedasticity discussed in the text immediately following this note, results were examined as closely as possible for other violations of assumptions as suggested by Nau's Regression Diagnostics. Not surprisingly, there are some observations exercising undue influence. Otherwise, the results are well-behaved according to Nau's recommended tests.

31 James G MacKinnon and Halbert White, "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties" *Journal of Econometrics*, Volume 29, Issue 3, September 1985, Pages 305-325 (https://doi.org/10.1016/0304-4076(85)90158-7).

A detailed summary of the methods used to estimate Equation D, as well as the methods used to calculate the ACRI and to assess its uncertainty, are presented in Appendix 2.

The parameter results are presented in Appendix 3 and summarized in Table 2. There are several noteworthy features of these parameter results. For the four climate conditions, there are 336 possible slope estimates (four climate conditions* seven regions * 12 months); of those, 25 percent are statistically significant. For the four climate conditions, the respective percentage of region-months which are statistically significant are: 54% (Rx5Day), 12% (T10), 19% (T90), and 15% (Wind).[32] For those region-months with statistically significant slopes, the average parameters, respectively, for the four weather conditions are 4.13 (Rx5Day), 1.12 (T10), 1.11 (T90), and 2.80 (Wind). Finally, across all region-months, whether estimated parameters are statistically significant or equal to zero, the average parameters for the four weather conditions are 2.21 (Rx5Day), 0.13 (T10), 0.21 (T90), and 0.43 (wind). In three different ways, among the weather elements. Precipitation (Rx5Day) is the most important factor driving results, with Wind also important.

Table 2: Summary of Parameter Estimates Significant at the 90% Confidence Level
(based on estimates for 84 region-months)

	Statistically Significant	Average Value for Region-Months With Statistically Significant Values	Average Value for All
Exposure	70%	1.84	1.29
Rx5Day	54%	4.13	2.21
T10	12%	1.12	0.13
T90	19%	1.11	0.21
Wind	15%	2.80	0.43

It is worth noting that with an r-squared of 0.62, there is still significant unexplained variation. It is also worth noting that the included variables might also be capturing effects of excluded variables that are correlated with included variables. In particular, note that the parameter estimates for exposure (displayed in Appendix 3) could reflect non-exposure-related issues that, like exposure, change across time. For example, consider the fact that the completeness of the data has increased across time. This measurement problem could boost the exposure exponent, even if the problem is unrelated to exposure.

32 For comparison, exposure is significant in 70 percent of the region months.

Constructing the ACRI

The ACRI intends to describe the losses that resulted from unusual levels of precipitation, temperature, and wind compared to the reference period, 1961–1990. In other words, if one asks how much loss (in 2016 dollars) occurred in a particular region in a particular month in a particular year between 1961 and 2016, the ACRI can answer that by looking directly at the data on losses from NOAA. If an observer then wants to ask how much of that loss (again in 2016 dollars) occurred as a result of unusual environmental conditions, estimates for Equation (D) provide an answer to that question.

Ideally, the ACRI would satisfy the following criteria in its construction:

- the sum of ACRI for each region-month during reference period equals zero. In the same way that the reference period creates a baseline for the evaluation of changes in weather in the ACI, the reference period ought to create a baseline for the losses associated with weather conditions;
- the ACRI for each region-month for which ACI elements are statistically insignificant equals zero. If weather is not statistically significant, then no losses should be attributable to weather in the ACRI;
- the partial correlation of ACRI and Exposure for each region-month equals zero (i.e., controlling for ACI). ACRI should not reflect changes in exposure, only changes in weather;
- the ACRI should be expressed in dollars;
- finally, there should be no artefactual bias either upward or downward in the estimated ACRI.

For each region-month-year, the parameter estimates produce a modeled value of losses. To produce a value for the ACRI, from the modeled loss is subtracted that loss which would have occurred had environmental conditions not been unusual. There are two ways in which the losses associated with "usual" conditions might be estimated. First, using the parameter estimates from Equation (C), for each region-month in each year the modeled losses can be calculated if the value of each weather condition equaled its reference period average. To calculate the ACRI for a particular region-month-year, simply subtract the modeled losses under reference period mean weather conditions from the modeled losses with observed weather conditions. These values are then aggregated by month, season, or year, to produce ACRI estimates of losses due to unusual environmental conditions.

Unfortunately, the non-linearity of the estimating equation builds an upward artefactual bias into this estimating method. With exponents greater than 1 (which is generally the case in the estimates), a weather condition 10 percent above average will produce more than a 10 percent increase in losses, while a weather condition 10 percent below average will produce less than a 10 percent decrease in losses. In this way, average weather conditions in two months (one above by 10 percent and one below by 10 percent) will produce more losses than average weather conditions in each of the two months would.

A second method avoids this bias due to non-linearity. In this method, the average losses modeled during the reference period, 1961–1990, are calculated for a particular region-month.[33] These losses represent the losses that are estimated by the ACRI model with the distribution of weather conditions which were observed in a region-month during the reference period. That average reference period loss is then taken as the estimate of the losses expected in a region-month experiencing the "usual" weather of the reference period. The ACRI is then calculated by subtracting from the modeled losses for a region-month in a given year the reference period average modeled loss for that region-month.

However, the ACRI measured in this way captures some of the impact of exposure changes in addition to the impact of changes in weather patterns. It also reflects, in undiscernible ways, differences in resilience. If, over time, some regions are adopting measures (e.g., enhanced building codes) to reduce losses due to extreme weather events, this estimating method, which assumes constant parameters over time, will be underestimating the impact of weather early in the time period and overestimating the impact in later years. As Bouwer notes: "[T]he potential effects of past risk-reduction efforts on the loss increase are often ignored, because data that can be used to correct for these effects are not available."[34] With currently available data, no method presents itself to adjust for changes in resilience; however, the ACRI has incorporated an adjustment for changes in exposure.

33 Before averaging, the ln(Loss) is converted into dollars through exponentiation.
34 Laurens M Bouwer, "Have Disaster Losses Increased due to Anthropogenic Climate Change," *Bulletin of the American Meteorological Society*, January 2011.

In order to control for the impact of increases in exposure, the reference period average modeled losses have been exposure-adjusted before subtracting them from the modeled losses for a particular region-month-year. When exposure-adjusting the reference period averages, the logic of the model might suggest adjusting only those region-months that produced statistically significant estimates for exposure. However, accepting that some (or all) of the 30 percent of region-months without statistically significant parameters might have been misestimated, the Academy has chosen to exposure-adjust all region months. Compared to the unadjusted total ACRI, the adjustment of all region-months reduces the total by twice as much as would the adjustment only of those region-months with statistically significant parameter values.[35]

Table 3 calculates for the post-reference period, 1991–2016, for the USA as a whole and for its seven ACI regions the sum of all ACRI losses and displays them alongside the observed losses as well as the exposure-adjusted losses. Several noteworthy points arise. First, in the post-reference period, 1991–2016, an estimated total of $24B of losses are attributable to unusually high precipitation, extreme temperatures, and high winds.[36] Further, those losses attributable to unusual environmental conditions amount to approximately 5 percent of the $493B in observed losses during that same period, and 3.3 percent of the exposure-adjusted losses. The vast majority of the ACRI total losses originate in the Southeast Atlantic region, which experienced ACRI losses of $22B in the post-reference period, out of a total of $278B observed losses (8 percent) and $421B in exposure-adjusted losses (5 percent). Of the other six regions, three reveal no material impact from extreme weather (ALA, CWP, and MID), one had somewhat less loss as a result of changes in weather conditions (CEA), and two had modest losses (SPL and SWP).

35 Without adjustment for changes in exposure, the ACRI total estimate is approximately $74 billion. With the adjustment for all region-months implemented, that total decreases to $24 billion. If the adjustment were only applied to region-months with statistically significant parameters for exposure, the ACRI total would be approximately $50 billion.

36 Precipitation and wind increase, on average, relatively small amounts after the reference period, although volatility increases somewhat more. The losses result from the high sensitivity to those changes.

Table 3: ACRI Losses, Observed Losses, Exposure-Adjusted Losses, by Region, 1991–2016, (in billions)

	ACRI	Observed Losses	Exposure-Adjusted Losses
USA	$23.78	$493.61	$711.95
ALA	$0.01	$0.51	$0.72
CEA	-$3.00	$51.53	$60.71
CWP	-$0.16	$5.39	$8.19
MID	$0.28	$57.07	$79.54
SEA	$22.42	$277.65	$420.84
SPL	$2.69	$65.81	$92.18
SWP	$1.55	$35.65	$49.77

Figure 1 displays annual totals for the USA for ACRI, observed losses, and modeled losses. A close relationship exists between modeled losses and the ACRI; this makes sense given the method used to generate the ACRI. The modeled losses match imperfectly the observed losses in two ways: 1) the modeled losses are substantially less than the observed losses; and 2) in most years, the peaks in the ACRI do not match peaks in observed losses. In most of those cases, the observed losses are elevated when the ACRI peaks, but not necessarily at their peaks. While there is no guarantee that ACRI peaks and those of observed losses should correspond exactly, that certainly would be a prior expectation.

Figure 1: ACRI, Modeled Losses, Observed Losses: Annual Totals, 1961–2016; USA

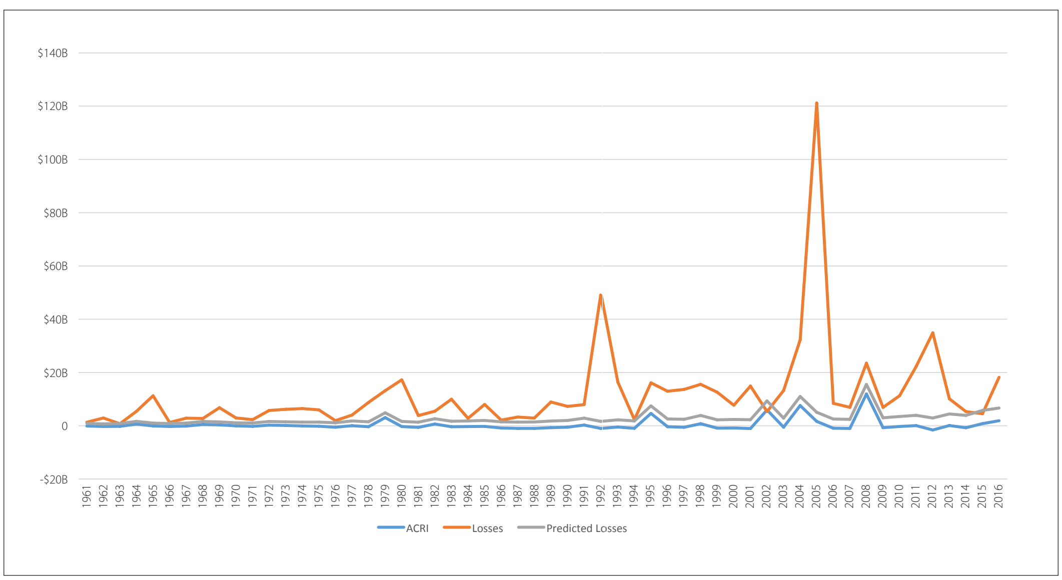
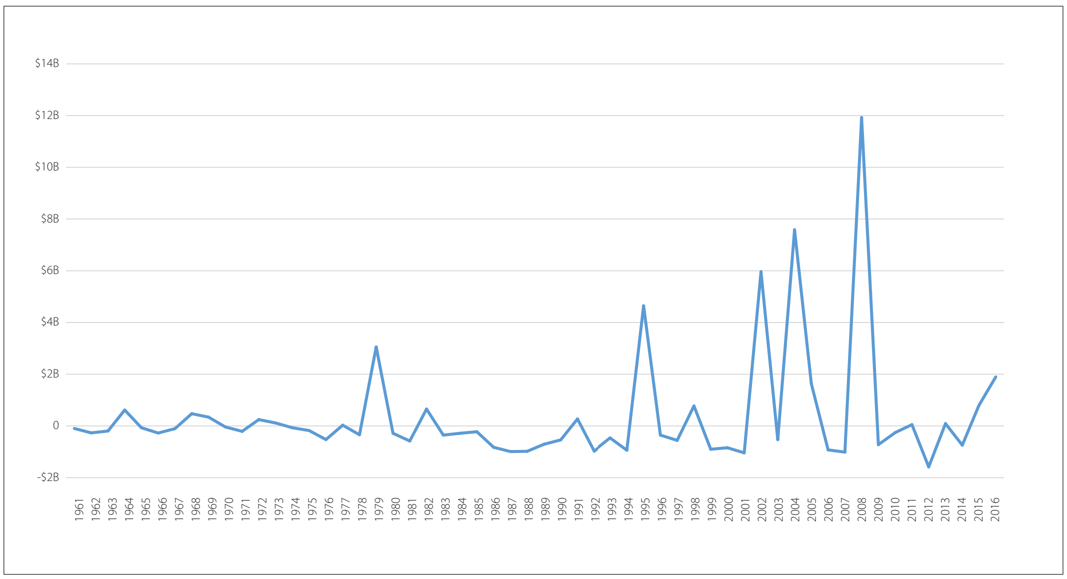


Figure 2 isolates the graph of the ACRI (the same one shown in Figure 1); separated in this way, the movement of the ACRI is more plainly visible. If we look first at the ACRI during the 30-year reference period (1961–1990), we see that on one occasion the ACRI exceeded $3B, reaching a high of $3.1B in 1979. On 22 occasions, the ACRI was negative, and on the seven occasions (other than 1979) when the ACRI was positive, it was less than $620 million. In the 26-year period since 1990, we see four occasions when the ACRI exceeded $3.1B, the highest of those being greater than $11B. The ACRI is still negative 15 times since 1990, indicating that weather conditions less extreme than during the reference period reduced losses. In sum, the ACRI in the period since 1990 reaches higher heights, reaches those heights more frequently, and dips down to negative levels frequently, if not quite so frequently as during the reference period. This describes a pattern of both increasing values of the ACRI and increasing volatility.[37]

Figure 2: ACRI: Annual Totals, 1961–2016; USA

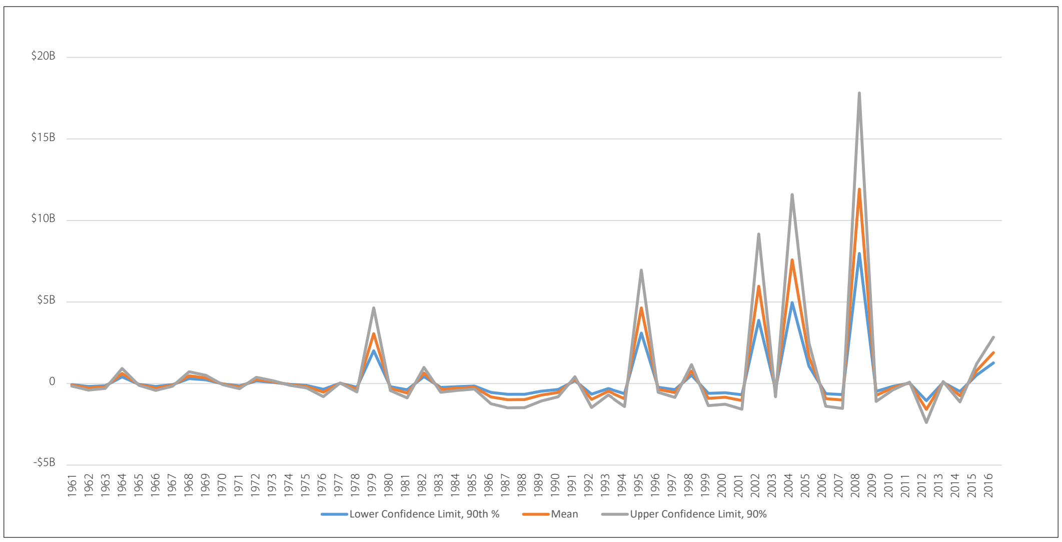


37 It is important to recall that at least some of the increase in the ACRI is due to increasing exposure.

This line showing ACRI losses may suggest more certainty than the Academy believes appropriate. There are many sources of uncertainty surrounding these estimates; some of those sources are internal to the calculations, and some are external (including data sources, singularity of very large losses, and the possibility that the results hinge on unusual characteristics of the observed weather distributions). To gauge the uncertainty with which the estimates ought to be viewed, that uncertainty has been assessed in two ways. The first way sought to gauge the intrinsic uncertainty—the uncertainty that attaches to any regression results based on the uncertainty of the fitted equation. The second sought to gauge the extrinsic uncertainty, particularly that which follows from the particularities of the distribution of weather events presented historically.

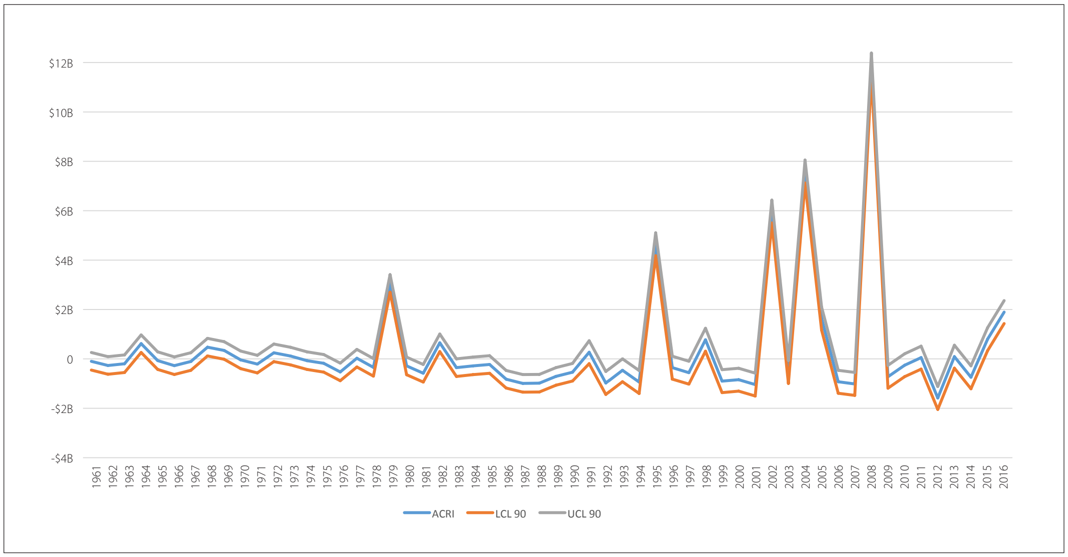
Figure 3 shows the ACRI graph bounded by the upper and lower limits of the 90th percent confidence interval.[38] This interval only reflects the uncertainty due to the probabilistic errors associated with the regression method employed here. Figure 4 shows the ACRI graph bounded by the 90th percent confidence intervals generated by a stochastic model based on alternative distributions of observations drawn from the historical record. For both of these techniques, see Appendix 2 for more details on the procedures followed.

Figure 3: ACRI, 1961–2016, Across All Regions and Months With 90% Confidence Level, Based on Intrinsic Uncertainty



38 The 90th percentile confidence interval for the ACRI was calculated by using the 90th percentile limits for the modeled values from the underlying estimates of the relationship between the weather metrics and losses. These confidence intervals are themselves subject to uncertainty, as they could be constructed in several different ways.

Figure 4: ACRI, 1961–2016 Across All Regions and Months With 90% Confidence Level, Based on Extrinsic Uncertainty



These results show good reason to be cautious in attaching too much significance to the precise values estimated. With intrinsic uncertainty accounted for, only 11 of the 26 post-reference period years are likely, at the 90 percent confidence level, to exhibit a positive ACRI. When accounting for the extrinsic uncertainty, only eight of the 26 post-reference years are likely to have positive values for the ACRI. Yet, the years that *are* positive are relatively large. Hence, it is useful to return to the breakdown for the regions and the country as a whole of ACRI totals for the entire post-reference period, as shown in Table 3. Table 4 shows those same values, but now with Lower and Upper confidence limits, from both internal and external assessments.

Table 4: ACRI Losses by Region, With Confidence Intervals (in billions)

	1991–2016	Intrinsic, Lower Limit	Extrinsic, Lower Limit	Intrinsic, Upper Limit	Extrinsic, Upper Limit
USA	$23.78	$15.72	$2.42	$35.98	$45.15
ALA	$0.01	$0.00	-$0.06	$0.01	$0.08
CEA	-$3.00	-$4.47	-$3.65	-$2.01	-$2.35
CWP	-$0.16	-$0.23	-$3.82	-$0.10	$3.51
MID	$0.28	$0.20	-$1.26	$0.39	$1.82
SEA	$22.42	$14.82	$10.90	$33.91	$33.94
SPL	$2.69	$1.79	-$15.66	$4.05	$21.03
SWP	$1.55	$1.02	-$0.19	$2.33	$3.29

This table illuminates the caution with which these estimates might best be treated. Taking only intrinsic uncertainty into account, one might conclude with 90 percent confidence that the U.S. as a whole and five of the seven regions had positive losses in the period 1991–2016 due to increases in the extremity of weather events compared to the reference period. However, taking extrinsic uncertainty into account, one might only conclude with 90 percent confidence that the U.S. and one of the seven regions had positive losses.

Throughout the process of estimating the relationship between weather and losses, and extending to the analysis of ACRI, the results are more certain at the national level than at the region level, and more certain at the region level than at the region-month level. Estimations at the region-month level with log-log transformed variables that exhibit a mean RSQ of 0.26, and an RSQ approaching 0 when transformed into dollars, produce estimates with lots of room for error. While each estimate might be far off in dollars from observed losses, the results in which higher levels of aggregation are more certain than lower levels suggests that at least some of the errors are the result of predictable over- and underestimation of particular observations.

Assessing the Modeled Relationship

The relationship modeled between the ACI components and property losses is the basis for the calculation of the ACRI. The credibility of the ACRI depends, in large part, upon the credibility of the modeled relationship. Going beyond the standard tools used to assess the validity of least squares regression models (primarily the significance tests for the parameters), it is important to illuminate as much as possible the extent and the limits of the robustness of this model. Throughout this estimation effort, the Academy has been trying to balance a set of conflicting objectives: a model robust with respect to both time and place, sensitive to both low-medium values of losses as well as to the extreme values in the loss distribution. The results in this section of the paper shed light on how successful we have been.

While r-squared is a reasonable, if statistically problematic,[39] yardstick for the ACRI model, the conventional r-squared reported examines all cases. As a result, it combines the effects which the model estimates within regions with some part of the differences across regions. To see whether the model is robust enough so that its explanatory power remains when the r-squared at the level of each region is examined, Table 5 shows the r-squared from the estimated model for each of the seven regions in the U.S., as well as for the U.S. as a whole. It has already been noted that the mean r-squared for the regions is a little more than half of the r-squared for the nation as a whole. What we see beyond that is that the range of r-squareds runs from a low of 0.24 to a high of 0.49, with only one region, ALA, exhibiting r-squared below 0.30. These regional results suggest that the model is similarly effective in most regions, with some small variation.

Looking at the results in another way, the r-squareds have been recalculated not in terms of Ln(Loss) (observed vs. modeled) but rather, in terms of the original Losses in dollars. When this is done, the r-squareds decline precipitously (as seen in the final row of Table 5). This difference is a measure of the impact of a log transformation on a highly skewed dependent variable. It indicates that the current state of the ACRI modeling captures a good deal of the variation in losses as long as those losses are log-transformed.

39 R-squared is potentially problematic in two senses. First, it is problematic because r-squared has no defined statistical distribution; as a result, it is impossible to identify a confidence level for r-square. Another problem with r-squared, in this application, is that r-squared increases with the number of variables. Adjusted r-squared would mitigate this problem. However, given that either when the sample is taken as a whole, or when region-months are considered as units within the sample, the number of estimators is less than 10 percent of the number of cases, the adjusted r-squared is never more than 0.02 less than r-square.

Table 5: R-Squared by Region, Ln(Loss) and Loss in $

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	USA
R-Squared, Ln(Loss)	0.22	0.36	0.26	0.50	0.39	0.47	0.32	0.36	0.62
R-Squared, Loss in $	0.00	0.02	0.00	0.07	0.02	0.07	0.14	0.05	0.03

Are these results being dominated by extreme values in the loss distribution? While the values of loss have been log transformed, with the effect of trimming in the most extreme values, those highest values are still significantly higher than the median values. As one can see in Table 1, the median Ln(loss) is 16, while the value of the largest Ln(loss) is 25. Table 6 repeats the r-squared by region from Table 5 and adds a new row reporting the r-squareds calculated post-regression without the inclusion of the top 1 percent of observations. Clearly, the results are not being heavily influenced by the extreme values in the loss distribution. On average, the regions decline by 1 percent, while the U.S. as a whole does not decline. It is clear that estimating the relationship with a log-transformed dependent variable has made it so that those cases have the same impact on the regression as all others.

Table 6: R-Squared by Region, Ln(Loss): Whole Sample and Bottom 99 Percent

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	USA
R-Squared, All	0.22	0.36	0.26	0.50	0.39	0.47	0.32	0.36	0.62
R-Squared, Bottom 99%	0.24	0.35	0.26	0.50	0.34	0.47	0.31	0.35	0.62

Given that the pooled cross-section includes many effects—both explicitly and implicitly—the Academy also wanted to ensure that the multivariate correlations observed are due, at least in part, to the environmental conditions measured by the ACI components. It would be possible for the r-squareds to derive solely or primarily from cross-regional or cross-monthly differences in losses. Results in Table 7, in which both the r-squareds by region of the estimated equation and the equivalent r-squareds for an equation estimated without any ACI components are presented, suggest that differences in average losses across region-months is playing a major role. The r-squared for the U.S. as a whole drops from 0.62 to 0.54 without the ACI components, indicating that only 8 percent of the impact observed originates with the ACI components. In the regions, on average, a little more than half (56 percent) of the impact appears to be due to the climate index components, with the average regional r-squared declining from 0.36 to 0.16. While a larger impact of the ACI components would be desirable (especially on the national level) to assure us that the results do reflect, in part, the impact of the weather metrics, the substantial overall r-squared combined with statistically significant estimates (at the 90 percent confidence level) for one-quarter of the ACI component parameters indicates a relatively robust result.

Table 7: R-Squared by Region: With and Without ACI Components

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	USA
R-Squared, With ACI	0.22	0.36	0.26	0.50	0.39	0.47	0.32	0.36	0.62
R-Squared, Without ACI	0.16	0.10	0.11	0.26	0.09	0.29	0.12	0.16	0.54

Finally, two tests of robustness of the results and a brief look at results by region-month. In Table 8, the Academy randomly split the sample of region-months in half, and estimated the relationship between losses and the ACI components. In almost every instance, the r-squareds for the two samples are similar to each other, and similar to the results for the whole sample. The r-squared for the U.S. is 0.63 for the whole sample, and 0.64 and 0.69 for the first and second random samples, respectively. The means of the regional r-squareds are also comparable, 0.36 vs. 0.43, with an r-squared of 0.37 for the sample as a whole.

Differences are observed in the parameter estimates. Only 54 percent of the parameters estimated as statistically significant in the first half of the sample are also significant in the second half, while 55 percent of the parameters estimated as significant in the second half are statistically significant in the first half as well. This lack of robustness across geography and seasonality is not surprising. Given the disproportionate impact of extreme values on the estimates, losses from a single major storm that appears in one-half of the sample and not the other can easily change results. This might serve to remind observers that these descriptive results do depend on losses as they occurred, with some very large losses having a large impact on the particular parameters estimated.

Table 8: Corrected for Heteroskedasticity
R-Squared by Region: Whole Sample, and Randomly Split Into Sample A and Sample B

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	USA
R-Squared, Whole Sample	0.22	0.36	0.26	0.50	0.39	0.47	0.32	0.36	0.62
R-Squared, Random Sample A	0.32	0.32	0.37	0.60	0.41	0.48	0.40	0.41	0.67
R-Squared, Random Sample B	0.24	0.38	0.32	0.48	0.39	0.50	0.33	0.38	0.63

The second test of robustness looked to intertemporal robustness; namely, do estimates of the relationship based on one time period produce reasonable correlations between observed and modeled values in a different time period. Most of the work here uses in-sample estimates; estimates for the years 1961–2016 are based on parameters estimated on those same years. This intertemporal robustness test is looking at out-of-sample estimates. Specifically, the Academy has estimated parameters based on the time period, 1961–2015, and then examined the model estimates for 2016. The results reported in Table 9 show an r-squared for the USA as a whole that is a little less than one-third for the modeled 2016 relationship compared to that estimated for the whole sample reported above. The average for regional r-squareds (based on only 12 observations for each region in 2016) is similarly reduced, declining from 0.31 to 0.09. However, this average masks differences. For one regions (MID), the r-squared declines by roughly 50 percent when moving from in-sample to out-of-sample testing; for two regions (CWP and SEA), the r-squareds decline by roughly two-thirds for the out-of-sample test compared to the in-sample results; and for the four remaining regions, the out-of-sample results are more than 80 percent lower than the in-sample ones. This test indicates questions about the intertemporal stability of the estimates. While the Academy does not now nor intends in the future to use these estimates as the basis for predicting future outcomes, intertemporal stability would still be a desirable characteristic of the process used to generate ACRI values.

Table 9: R-Square by Region: Model Estimates for 2016 Based on 1961–2015 Estimates

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	USA
R-Squared, 1961–2016	0.14	0.34	0.25	0.45	0.35	0.42	0.24	0.31	0.59
R-Squared, 2016, based on 1961–2015 estimates	0.02	0.06	0.08	0.22	0.12	0.07	0.03	0.09	0.20

Conclusion

This paper reports on progress made on the path to meeting the following goals:

- **To assist** policymakers, the public, and actuaries with an indication of the relationship between environmental conditions and damages;
- **To use the data** and defined components of the Actuaries Climate Index in the construction of the Actuaries Climate Risk Index, as much as possible; and
- **To create an index** which can be updated regularly and made accessible to all users via the website maintained by the Academy.

This paper also reflects the challenges that remain with the data used both for the ACI and the ACRI, the metrics underlying the ACI components, the modeling of the relationship between the ACI components and losses, and the construction of the ACRI. Given the identified sources of uncertainty in the current estimates, the work has not yet extended to produce ACRI estimates for losses of life and for injuries, nor develop an ACRI measure for Canada, although the plan in the future is to include those.

Much remains to be done to improve upon the ACI and the ACRI in versions 2.0 and beyond. While others have undertaken efforts similar in some respects to the ACI and ACRI, the differences in the sponsoring organizations' objectives for both indexes create novel challenges. While the IPCC and the U.S. National Climate Assessments aim at a goal related to that of the ACI, the focus of their efforts is both longer term and predictive. Unlike those efforts, the ACI aims to describe what has already happened and to update those descriptions quarterly. While the UNDP Disaster Index sought to establish a relationship between environmental conditions and economic losses, it did so with an annual national database. Moreover, while catastrophe models have become quite good at identifying relationships between certain extreme events and associated economic losses, their focus is on estimation and prediction, which is usually narrowly circumscribed both by types of event and geographic domain.

The ACRI, like the ACI, aims not to predict but to describe what has happened and to update those descriptions regularly. The ACRI also aims to apply to large regions within two countries (U.S. as outlined in this paper and Canada) with a common model. These projects—both the ACI and the ACRI—are inherently difficult.

To find the best correlation between weather variables and property losses, the impact of inflation, exposure, region, and seasonality has been controlled for. A dependent variable expressing losses in dollars was analyzed, and each month of each year for each region has been treated as a separate observation. To allow for non-linearity in relationships between weather conditions and losses, to allow for interaction among weather conditions, and to mitigate the impact of the highly skewed distribution of losses, a model has been estimated in which both independent and dependent variables are log transformed. To identify statistically significant parameters, backwards regression was used on the dependent variable and the ACI to select the best estimated model in which all parameters were statistically significant at the 90 percent confidence level. The pooled, cross-sectional model produced an r-squared of 0.62 with the log-transformed modeled values, although the r-squared for the dollar equivalents was 0.03. Once corrected for heteroskedasticity, the r-squareds remained largely unchanged.

Based on this estimated relationship between the ACI and losses, the ACRI was calculated as the difference in modeled losses due to ACI components being above (or below) their reference period mean values. In order to exclude the impact of changes in exposure on the ACRI, the reference period mean modeled losses have been exposure-adjusted. The resulting ACRI totals $24 billion during the post-reference period, 1991–2016.

The model has a large amount of uncertainty, because each region-month currently only has 56 data points on which to base the parameters: 30 points during the reference period and 26 points subsequent to the reference period. This uncertainty has been estimated in two ways. Based on the intrinsic uncertainty associated with the regression estimates from which the ACRI is built, a 90 percent confidence interval is estimated around the best estimate for total ACRI losses of $16 billion to $36 billion. However, the broader extrinsic uncertainty associated with only having one “draw” of the weather distributions, both for the reference and the post-reference periods, has been estimated using a stochastic model of synthetic datasets based on randomly selected observations from the original data. With this broader definition of uncertainty, a 90 percent confidence interval has been estimated for total ACRI losses of $2 billion to $45 billion.

Throughout this documentation, weaknesses and limitations are outlined that serve as cautionary notes, pointing to the need to interpret these current results in light of their inherent uncertainty. Chief among these limitations are:

- As noted, while the model has an r-squared of 0.62 on log-transformed values, the r-squared on dollars of modeled and actual losses is only 0.03.
- The model performs most dependably at the national level, less so at the regional level (mean r-squared equals 0.36), and even less well at the region-month level (r-squared equals 0.24).
- The ACI metrics used in the model are averaged over large geographic areas, while the most damaging events are concentrated in much smaller areas.
- The ACI metric for Wind, based on average monthly wind speeds in these large geographic areas, is not shown by the model to be very good estimates of large losses that are driven primarily by windstorms.
- Equation coefficients are quite inconsistent from one month to the next, in a given region, which does not provide a logical explanation for the ACRI values.

These weakness and limitations also suggest a direct proceeding to version 2.0 of both the ACI and the ACRI to seek better data and develop more effective metrics and more robust analysis. Others are encouraged to build on this work by conducting research using weather metrics and proprietary insurance company loss data, which would be available in precise geographic detail.

Appendix 1: Statistical Appendix

The tables below present data tabulations for each of the seven geographic regions. For each region, the results are tabulated across all 56 years and 12 months of data.

Table 10: Univariate Analysis: 99th Percentile / 50th Percentile

Region	Rx5day	Wind	T90	T10	Property Loss
ALA	2.2	2.1	4.1	6.3	3,133.9
CEA	2.1	2.4	3.1	3.3	175.1
CWP	2.7	2.3	3.0	4.0	1205.3
MID	1.8	2.1	3.5	3.6	130.0
SEA	1.6	2.2	3.4	3.2	281.6
SPL	2.0	2.0	2.7	3.1	104.0
SWP	2.7	2.9	2.7	3.2	443.9

Table 11: Univariate Analysis: Average / 50th Percentile

Region	Rx5day	Wind	T90	T10	Property Loss
ALA	1.07	1.05	1.21	1.39	434.0
CEA	1.04	1.03	1.12	1.14	14.7
CWP	1.07	1.04	1.14	1.23	82.6
MID	1.00	1.06	1.16	1.12	8.6
SEA	1.00	1.05	1.14	1.12	20.6
SPL	1.05	1.04	1.09	1.09	6.4
SWP	1.08	1.08	1.09	1.13	19.4

Table 12: Average Weather Metric 1991–2016 Divided by Average Weather Metric 1961–1990

Region	Rx5day	Wind	T90	T10
ALA	1.01	0.91	1.33	0.65
CEA	1.06	0.75	1.20	0.73
CWP	1.02	0.87	1.19	0.77
MID	1.05	1.16	1.06	0.87
SEA	1.03	0.94	1.20	0.80
SPL	1.04	1.09	1.15	0.86
SWP	0.98	0.81	1.32	0.76

Table 13: 90th Percentile of Weather Metric 1991–2016 Divided by 90th Percentile of Weather Metric 1961–1990

Region	Rx5day	Wind	T90	T10
ALA	1.02	0.94	1.31	0.70
CEA	1.05	0.94	1.17	0.81
CWP	1.00	0.87	1.16	0.80
MID	1.04	1.10	1.00	0.93
SEA	1.01	1.00	1.16	0.86
SPL	1.00	1.08	1.16	0.91
SWP	0.94	0.73	1.28	0.76

Table 14: Correlation of Weather Metric With Logged Loss

Region	Rx5day	Wind	T90	T10
ALA	6.6%	5.7%	13.6%	-7.6%
CEA	48.8%	15.9%	2.3%	-2.6%
CWP	26.8%	20.3%	15.5%	-0.4%
MID	54.1%	22.1%	4.3%	-4.2%
SEA	43.0%	32.3%	16.0%	-4.9%
SPL	49.5%	16.9%	1.4%	4.6%
SWP	24.6%	8.3%	12.0%	-1.1%

Table 15: Correlation of Weather Metric With Loss in Dollars

Region	Rx5day	Wind	T90	T10
ALA	-4.4%	-6.8%	-5.4%	1.4%
CEA	15.1%	1.4%	1.3%	-2.2%
CWP	7.4%	0.6%	0.4%	0.9%
MID	19.9%	8.2%	-3.0%	-3.1%
SEA	12.1%	7.5%	5.2%	-0.8%
SPL	22.0%	4.8%	2.0%	2.0%
SWP	14.5%	3.5%	9.0%	4.9%

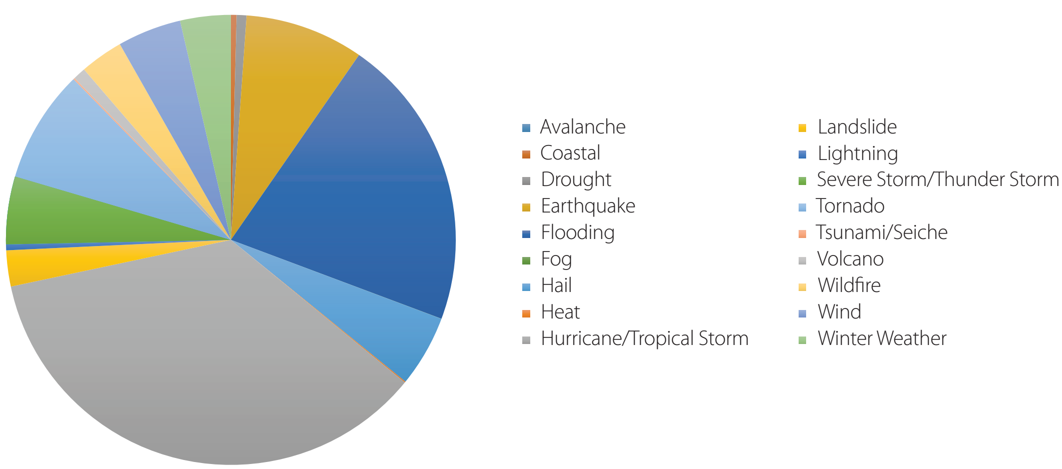
Table 16: Exposure in Year "Y" / Exposure in Year "X"

Region	2016 / 1961	2002 / 1975
ALA	9.25	2.42
CEA	3.61	1.91
CWP	8.09	2.83
MID	3.25	1.71
SEA	6.78	2.56
SPL	4.88	2.03
SWP	8.02	2.62
Note: The exposure data is in 2016 USD, so the ratios above capture real as opposed to nominal exposure growth.		

Table 17: Average Modeled Loss 1991–2016 Divided by Average Modeled Loss 1961-1990

Region	Loss	Loss / Exposure
ALA	5.64	1.93
CEA	0.34	-0.29
CWP	1.44	-0.28
MID	0.75	-0.01
SEA	2.86	0.48
SPL	1.32	0.07
SWP	2.25	0.29
Note: The estimates are in 2016 USD, so the ratios above reflect real as opposed to nominal changes.		

Figure 5: Distribution of Property Losses by Peril



Note that losses arising from volcanoes—which represent less than 1 percent of total property losses—were included in this analysis by mistake. This error, however, has no material effect on results.

Appendix 2: Procedures Used in Estimation of OLSQ Equation and ACRI Calculation

Estimation

The relationship between weather and losses is estimated in three steps, using data from seven U.S. regions, over each of the 12 months, over a 56-year period, from 1961 to 2016. The 30-year period from 1961 to 1990 is treated as the reference period; the 26 subsequent years are discussed as the post-reference period. Estimation was performed using SAS statistical software.

The exponential equation in Equ B as estimated by taking natural logarithms of all independent and dependent variables. For Losses, which sometimes have a value of 0, 1 was added to all dollar values of losses prior to transformation. When losses are restored to dollars by exponentiation, 1 dollar was subtracted.

In the first phase, four weather elements (Precipitation, High Temperatures, Low Temperatures and Wind), along with Exposure, were entered into an ordinary least squares (OLSQ) regression estimating equation with dummy variables for each region-month combination. A total of 420 variables were entered in this fashion (seven regions, 12 months, five variables). In addition to a global intercept, there was an intercept for each region-month. Backward, stepwise regression was then employed to eliminate variables not significant at the 90 percent confidence level.

In the second phase, the variables found to be significant in the first phase, were entered into an OLSQ regression. However, in this phase the covariance matrix was adjusted to correct for heteroskedasticity using the method describe by White (1980).[40] With the corrected covariances, tests of significance were repeated and some variables were identified as insignificant.

In the third phase, the variables continuing to be significant after correction for heteroskedasticity were entered into an OLSQ regression. This equation was used to produce the modeled values (and the confidence interval around those values), which were used as the basis for calculating the ACRI and its intrinsic confidence interval.

40 Halbert White, "A Heteroskedastic Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* (48,4), May 1980.

Calculating the ACRI

First, all modeled losses were converted to dollars. Second, the average modeled loss for a particular region month over the 30-year reference period was calculated. This provides a baseline against which to measure subsequent losses; it is a baseline which controls, in a sense, for the distribution of weather and exposures in that region-month during the reference period.

The intention in the ACRI is to assess the losses that are due to differences between weather in the post-reference period and that which prevailed during the reference period. However, if the modeled losses in a particular post-reference period are compared to the reference period average, it will be the product of differences in both weather and exposure. In order to control for the effect of changes in exposure, and isolate the effect of changes in weather, the average modeled losses from the reference period are adjusted for changes in exposure. In particular, the average for a region-month are multiplied by the ratio of Exposure in a particular year to the average Exposure in that region-month during the reference period. For example, if the average modeled loss in a particular region-month was $1B during the reference period, and the average exposure for that region-month during the reference period was $1T, then if the ACRI is being calculated for a later period in which the exposure was $10T, one would treat $10B as the exposure-adjusted average modeled loss during the reference period for that region-month.

To calculate the ACRI for each region-month each year, the exposure-adjusted average modeled loss for the region-month was subtracted from the modeled loss for that region-month-year. Summing these values across all months and all years from 1991 to 2016 produced estimates of ACRI for each region in the post-reference period. Summing the individual ACRI values across all region-months in a given year produced estimates of the ACRI for each year.

Estimating the Uncertainty of the ACRI Estimates

Uncertainty of the ACRI has been estimated in two ways:

First, the intrinsic uncertainty of the ACRI estimates has been estimated. This is the uncertainty arising from the OLSQ method with the observed data. Using the standard error of the regression, the lower and upper limits for the 90 percent confidence level of each modeled loss have been derived. The lower limits of the modeled losses were then used to create a lower bound for the ACRI, following exactly the same steps to calculate the ACRI as described above, substituting the lower bound for the best estimate. Similarly, the upper bound of the ACRI confidence level has been calculated using the upper bound limits of the modeled losses. This produces the picture shown in Figure 3, with a fairly tight confidence band around the ACRI annual estimates.

Second, the extrinsic uncertainty of the ACRI estimates has been estimated. This is the uncertainty arising from reliance on a single observed distribution of weather events and losses. While this distribution is the only historical set of observations available, one could think of this set of observations as drawn from a larger pool of potential observations. This second estimate of uncertainty tries to capture this extrinsic source of uncertainty, while preserving as much as possible the structure of the observed data.

In order to accomplish this objective, a synthetic data set has been created with values for each region-month-year. To create the values for the 30 reference-period years for a particular region-month, one of the observations has been randomly selected from the reference period for that region-month and its values for weather and losses have been assigned to the first year. This process was then repeated for each of the 30 years, treating as the pool of possibilities the original 30 observations with replacement. In creating this synthetic set of values for the reference period, some observations may be excluded, and some included more than once. The same procedure was then done for the 26 post-reference period values for each region-month, drawing the synthetic values from the observed values in the post-reference period. Having created the synthetic data base for all regions, months, and years, the equations were then estimated in the three-phase process described above. Based on those results, the ACRI was then calculated for each region-month-year, also as described above.

Early attempts revealed that a single region-month—Alaska in April—often produced outlandish results, sometimes exceeding the total losses for all regions for all years by several orders of magnitude. This was the result of the region-month experiencing one large storm early in the reference period (and thus it became larger when exposure-adjusted). The doubling or the exclusion of this event caused enough differences in estimates for this region that the Academy excluded it from all calculations of uncertainty. No other region-month demonstrated any comparable volatility.

This process was repeated 30 times. This process generated a mean and a standard error for the annual total ACRI values, and for regional, post-reference period ACRI values, as shown in Figure 4 and Table 4. While more repetitions would have been desirable, the time required for this exercise was a limiting factor. However, plotting the mean and standard errors against the number of simulation runs indicates that both have begun to converge on their limiting values, as seen in figures 6 and 7 below. The values for the ACRI derived from the historical observations were combined with the standard errors derived from this method to produce a 90 percent confidence interval for the ACRI estimates based on extrinsic uncertainty.

Figure 6: ACRI USA Total, 1991–2016 Simulation Mean, v. # of Simulation Trials

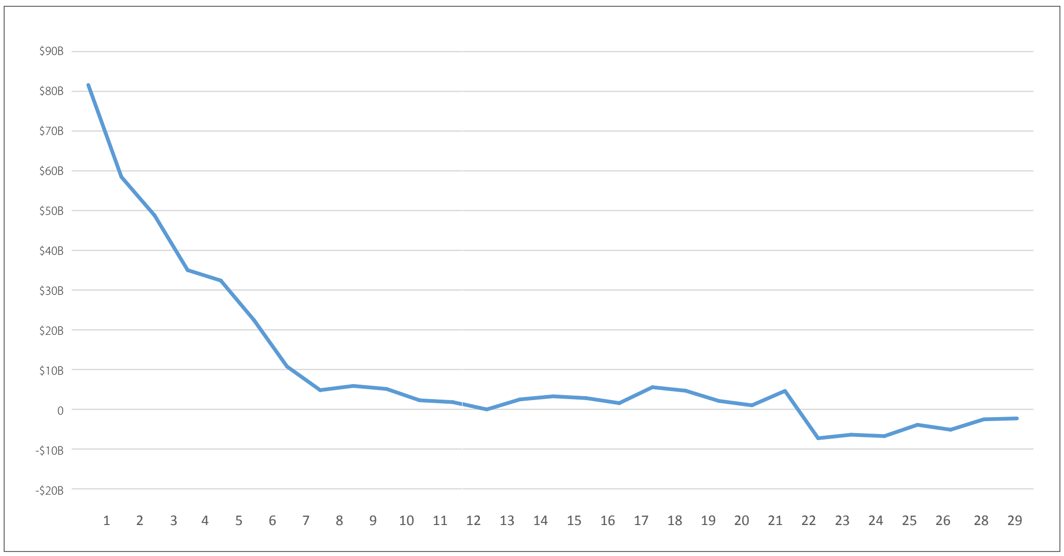
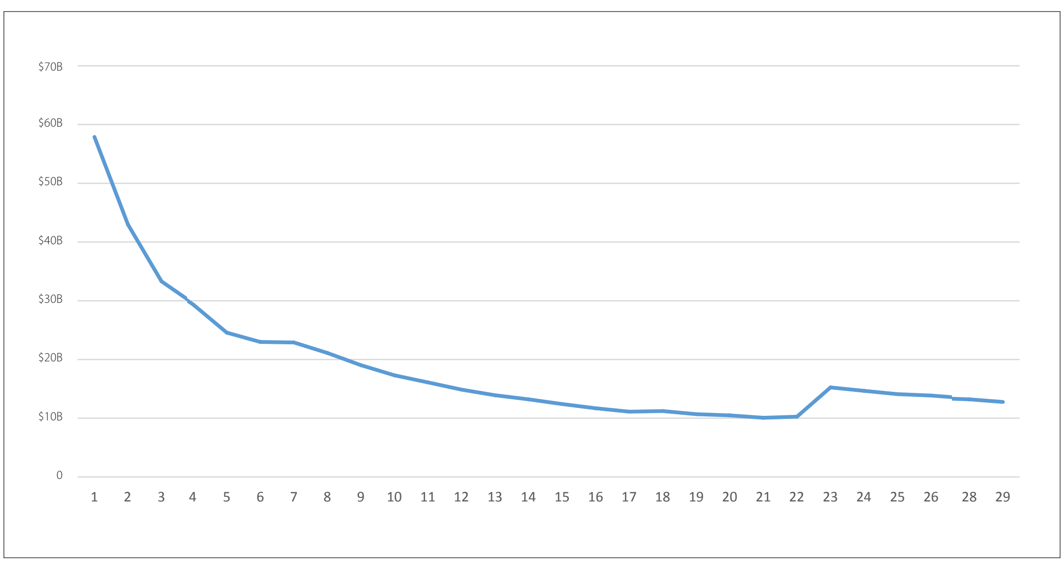


Figure 7: ACRI USA Total, 1991–2016 Simulation Standard Error, v. # of Simulation Trials



Appendix 3: Statistically Significant Parameters (90 Percent Confidence Level) For Equation D

	Region	ALA	CEA	CWP	MID	SEA	SPL	SWP			Region	ALA	CEA	CWP	MID	SEA	SPL	SWP
Variable	Month									Variable	Month							
Intercept	0	-32.64	-32.64	-32.64	-32.64	-32.64	-32.64	-32.64		T10	1							
Intercept	1		37.17				-56.76			T10	2			0.84			0.91	0.89
Intercept	2	-72.80	38.53	-47.79	-93.33					T10	3			1.04				
Intercept	3	38.40	36.04		36.36	50.66				T10	4							
Intercept	4	-37.76	33.96		50.27					T10	5							1.70
Intercept	5	-73.68	37.33							T10	6							
Intercept	6	-40.32	33.09			49.85	51.35			T10	7							
Intercept	7	-104.16	49.72		50.56	49.69				T10	8	4.80		2.01				
Intercept	8		34.99	45.93						T10	9							
Intercept	9	-62.81	29.15			29.37	49.60	42.11		T10	10					-1.47		
Intercept	10							-46.34		T10	11							1.68
Intercept	11			-50.81			-43.57			T10	12						-1.21	
Intercept	12	-54.02	37.53	46.46			28.75			T90	1						-1.78	
Exposure	1	1.70		0.69	0.92	1.10	3.18	1.16		T90	2				-0.86			
Exposure	2	4.40		2.52	4.23	1.61	1.40	1.33		T90	3	-1.67						-1.23
Exposure	3			1.00			1.79	1.37		T90	4							
Exposure	4	2.99		1.06		1.24	1.84	1.51		T90	5	1.17		1.12				
Exposure	5	4.69		1.65	1.76	1.28	1.85	1.67		T90	6							
Exposure	6	3.23		1.21	1.32			1.70		T90	7			1.19				
Exposure	7	5.12		1.29			1.80	1.70		T90	8	3.79		2.96				
Exposure	8	1.58			1.74		1.79	1.71		T90	9					1.69		
Exposure	9	3.32		1.38	1.28					T90	10			2.12				
Exposure	10	1.53	0.96	0.77	0.96	1.30	1.35	2.55		T90	11	2.32		3.07		1.19		3.73
Exposure	11	0.96	0.95	3.53	1.02	1.57	2.56	1.45		T90	12	-1.04						
Exposure	12	3.91			1.08	1.07		1.11		Wind	1						5.29	2.21
Rx5Day	1		3.17	6.88	6.19	4.48	2.23	4.17		Wind	2				1.62		3.60	
Rx5Day	2		2.82	6.24	6.00		3.03	3.50		Wind	3	3.98				1.67		
Rx5Day	3		3.39	5.07	3.66			1.63		Wind	4							1.32
Rx5Day	4		3.81	3.81		3.72				Wind	5							
Rx5Day	5		2.93			3.35				Wind	6							
Rx5Day	6		4.21	3.47	3.22					Wind	7							
Rx5Day	7									Wind	8							
Rx5Day	8		3.72			11.07				Wind	9							
Rx5Day	9	5.93	4.90		3.18	4.10		2.14		Wind	10					2.92		
Rx5Day	10		4.02	4.66	5.35		3.28	2.83		Wind	11		1.85			4.19		
Rx5Day	11	6.62	5.21		3.32		4.46	1.33		Wind	12				2.72	1.72	3.36	
Rx5Day	12		2.85		4.83	4.43	3.47	3.25										

Appendix 4: Alternative Forms of the Model

In developing the model presented here, several functional forms have been explored for the estimating model, and others have been contemplated but not yet fully explored. In each case, other forms had weaknesses that seemed greater than those of the model presented. Without revisiting all of the analysis, consider next three illustrative forms accompanied by the reasoning that led the Academy to reject these forms in favor of the proposed model (Equation D).[41]

Loss per dollars of exposure was treated as the dependent variable in many efforts. When the same model estimated as Equation (C) but dividing Losses by Exposure before log-transforming was run, results similar to those presented in the body of the paper were obtained.

These results (see Table 19, R-Squared, Log-Log Form) reveal an r-squared for the U.S. as a whole of 0.46, and an r-squared dropping to 0.04 for the U.S. and to 0.06 for the mean of the regions when the exponentiated values of losses in dollars are correlated with modeled losses in dollars. These results are similar but inferior to the results presented in the paper.

Table 19: R-Squared by Region: Log-Log Transformed Model in Logged Units and in $

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	U.S.
R-Squared, Log-Log Form	0.13	0.33	0.33	0.48	0.39	0.48	0.26	0.34	0.46
R-Squared in $	0.00	0.02	0.00	0.16	0.04	0.10	0.08	0.06	0.04

In an effort to find a model where the r-squared would be higher in dollar-dollar correlations, a simple linear model was tried, without log-transformations. The dependent variable was dollars per dollar of exposure. But, because of the skewness of the distribution of losses, this variable was truncated at the 99th percentile, with values at the 99.5th percentile substituted. Results, shown in Table 20, show both strength and weakness relative to the model presented in the text.

Table 20: R-Squared by Region, Loss/Exposure and Loss in $

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	USA
R-Squared, Loss/Exposure	0.08	0.16	0.13	0.09	0.32	0.18	0.12	0.15	0.21
R-Squared, Loss in $	0.00	0.01	0.00	0.14	0.09	0.15	0.10	0.07	0.09

41 The results reported in this appendix do not include corrections for heteroskedasticity.

GLM models, commonly used in catastrophe modeling, aim primarily to estimate distributions of losses rather than to generate modeled values for particular observations. While in Version 2.0 of the ACRI this distributional approach may be pursued, and GLM and other models may be more useful, the current effort aimed to generate a fit to the data such that each region-month in each year had a reasonable model estimate of the observed loss (i.e., a classical regression fit). In exploring GLM models, given limitations from available technology,42 the Poisson distribution provides the best fit of the data. In the GLM model presented in Table 21, results assessed for the correlation of observed and modeled Loss/Exposure is similar to, and stronger in some respects, than the current Loss/Exposure proposed model. The same conclusion holds true when considering correlations in terms of dollar losses.

However, when looking at the correlations within region-months (for example, calculating the r-squared for January in Region 1, etc.), overfitting is immediately evident. Table 22 shows certain characteristics of the distribution of these 84 region-month r-squareds. Note that the 90th percentile cut-off for r-squareds among the region-months is 0.79 means that 10 percent of the region-months have r-squareds greater than 0.79. This is an implausible result without overfitting. The data possesses too much error and the relationships are sufficiently obscure that r-squareds above 0.4 or 0.5 ought to raise red flags. As a result of this overfitting, the Academy has put aside the GLM models for the current version of the ACRI.

Table 21: R-Squared by Region: GLM Model, in Loss/Exposure, and in $

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	U.S.
R-Squared, Loss/Exposure	0.33	0.30	0.38	0.33	0.34	0.27	0.44	0.34	0.33
R-Squared, Loss in $	0.01	0.05	0.01	0.49	0.08	0.20	0.32	0.16	0.09

Table 22: Distribution of R-Squareds, Loss/Exposure, GLM model, 84 Region-months

	R-Squared
Minimum	0.00
Median	0.19
Mean	0.30
80th Percentile	0.59
90th Percentile	0.79
Max	0.96
SWP	2.25

42 Work was done in SAS and R. Without programming proprietary density functions, work was limited to the distributions available.

Finally, in an effort to separate estimation of the frequency of losses from the severity of those losses, the Academy follows the example of Wanik (2012), who applied separate models to these two tasks in an effort to predict damages to an electric grid. Combining the results for frequency and severity would, in principle, produce a more refined picture of the relationship between ACI components and losses as a basis for creating the ACRI.

To estimate frequency, any region-month with any non-zero losses was coded a 1; all others were coded 0. A logistic regression with a probit transform was executed and the results were evaluated by comparing region-months that experienced (or did not) any losses with those region-months in which the estimated probability of experiencing any loss was greater (or less than) 50 percent. The results are reported in Table 23. On the face of it, these results are quite positive, with more than 92 percent of region-months correctly modeled (Yes-Yes and No-No in the table). The problem is that almost all cases where there were no losses are mismodeled as likely to have losses. In a sample where 92 percent of cases are positive, estimating that virtually all cases (99.70 percent) will be positive is an effective strategy for getting more than 90 percent correct estimates. This model is therefore not effectively discriminating between cases that will and will not have losses. This problem was exacerbated when also looking at the Canadian data, where 90 percent of the region-months had zero losses and virtually all region-months were modeled as zero losses. Again, the model is showing no ability to discriminate non-zero from zero-loss observations.

Table 23: Probit Regression Estimating Frequency of Observing Losses Percentage Correct and Incorrect Predictions

Predicted Probably of Losses	Observed Losses		
	Yes	No	Total
Yes (or >50%)	92.01%	7.70%	99.70%
No (or < 50%)	0.17%	0.13%	0.30%
Total	92.18%	7.82%	

For the U.S. as a whole, this problem is not very material given the limited number of observations with zero losses. Hence, estimates of conditional severity would be considered—that is, the relationship between ACI components and Loss/Exposure restricted to region-months with non-zero losses. However, the results of this model reported in Table 24 are not significantly different from those of the proposed model where the same model is applied to both non-zero and zero-loss observations. Simplicity, combined with the weakness of the frequency modeling, suggests using the unified model rather than a two-stage model attempted here.

Table 24: R-Squared by Region: Severity Model for Non-Zero Loss Region-Months

Region	ALA	CEA	CWP	MID	SEA	SPL	SWP	Mean	U.S.
R-Squared, non-zero losses, Loss/Exposure	0.15	0.17	0.14	0.09	0.33	0.18	0.09	0.17	0.23

AMERICAN ACADEMY of ACTUARIES
Objective. Independent. Effective.™

AMERICAN ACADEMY OF ACTUARIES
1850 M STREET NW, SUITE 300, WASHINGTON, D.C. 20036
202-223-8196 | **ACTUARY.ORG**

© 2020 American Academy of Actuaries. All rights reserved.