Where the Home Insurance Meets the Climate Change: Making Sense of Climate Risk, Data Uncertainty, and Projections

Authors: Vyacheslav Lyubchich[1], Kelly H. Kilbourne[1], Yulia R. Gel[2*]

Affiliations:

[1]Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, 146 Williams St., Solomons, MD 20688, United States

[2]Department of Mathematical Sciences, University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, United States

*Correspondence to: Yulia R. Gel, ygl@utdallas.edu

Abstract:

We present an attribution analysis of residential insurance losses due to non-catastrophic weather events and propose a comprehensive statistical methodology for assessment of future claim dynamics in the face of climate change. We also provide valuable insight into uncertainties of the developed forecasts for claim severities with respect to various climate model projections and greenhouse emission scenarios. The results of our study pave the way for more accurate short- and long-term cost-benefit assessment of climate adaptation in the insurance sector.

Keywords: climate-related risk management, property damage, forecasting, climate adaptation

1. Introduction

There is a growing body of scientific evidence that extreme weather events are increasing in frequency and intensity (IPCC, 2014a). This phenomenon has already led to significant property damages and mounting insurance losses due to floods, storms, hurricanes, and other natural disasters (e.g., see Bouvet & Kirjanas, 2016; LSE, 2015; NAIC, 2016; Smith & Matthews, 2015 and references therein). To mitigate the adverse effects of this upward trend, insurers need a comprehensive attribution analysis of climate-related claim dynamics at a local level and a reliable quantification of future climate-related risks.

Whereas a vast literature exists on modeling insurance risks due to weather disasters and catastrophic events (see, for instance, overviews by Grossi et al., 2005; Kunreuther & Michel-Kerjan, 2012; Toumi & Restell, 2014), little is known about how to address risks due to less extreme events, the so-called "*low individual but high cumulative impact*" events, which have exhibited rising frequencies over time. The yet scarce literature on this topic primarily focuses on modeling a number of weather-related insurance claims (Scheel et al., 2013; Soliman et al., 2015) rather than on the dynamics of related losses. A few recent studies on predicting insurance losses due to extreme but non-catastrophic weather events include Haug et al. (2011) who consider Norwegian home insurance data aggregated at a provincial level, Cheng et al. (2012) who analyze rainfall-related damage in four cities in Ontario, Canada, and Held et al. (2013) who use various downscaling methods to evaluate winter-storm induced residential losses in Germany.

In this paper, we develop an attribution analysis of weather-related insurance losses on a daily scale and present a comprehensive statistical methodology for assessing future loss dynamics. Our study also provides a valuable insight into uncertainties of the developed

forecasts for claim severities in respect to various climate model projections and greenhouse emission scenarios. While the current paper is only one of the very first steps on the path of a systematic and sound analysis of future insurance risk due to non-catastrophic weather events, the results of our study pave the way for more accurate short- and long-term cost-benefit assessment of climate adaptation in the insurance sector.

The strategy we employ includes an analysis of nonlinear dependencies between weather and claims data; detection of critical thresholds, or tipping points leading to an increased number of claims; and joint frequency-severity predictive modeling of weather-related daily losses. A summary of the proposed analysis strategy is presented in Figure 1; all the details are explained further in the paper.

The remainder of this paper is organized as follows. Section 2 describes the data: recorded insurance claims and weather conditions as well as climate projections (Step 1 in Figure 1). Section 3 showcases the methods of attribution analysis and tail comparison (Steps 2 and 3 in Figure 1). Section 4 presents joint frequency-severity predictive modeling and forecasts (Step 4 in Figure 1). The paper is concluded with a discussion in Section 5.
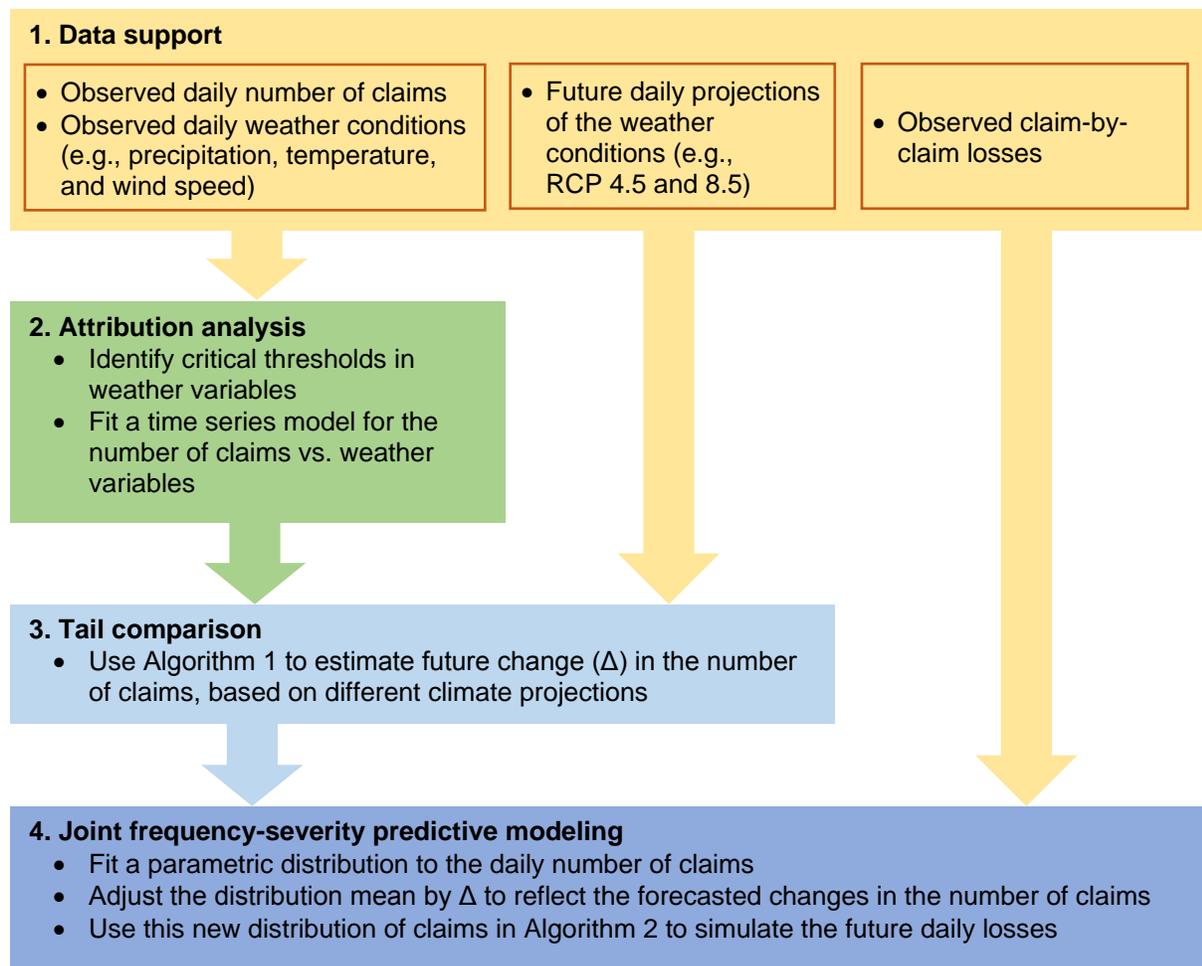
1. Data support
- Observed daily number of claims
- Observed daily weather conditions (e.g., precipitation, temperature, and wind speed)

- Future daily projections of the weather conditions (e.g., RCP 4.5 and 8.5)

- Observed claim-by-claim losses

2. Attribution analysis
- Identify critical thresholds in weather variables
- Fit a time series model for the number of claims vs. weather variables

3. Tail comparison
- Use Algorithm 1 to estimate future change (Δ) in the number of claims, based on different climate projections

4. Joint frequency-severity predictive modeling
- Fit a parametric distribution to the daily number of claims
- Adjust the distribution mean by Δ to reflect the forecasted changes in the number of claims
- Use this new distribution of claims in Algorithm 2 to simulate the future daily losses

Figure 1. Flow chart of the attribution analysis and joint forecast of daily number and severity of weather related home insurance claims

2. Data

In this study, we use Canadian residential insurance claim data, local weather station data, gridded instrumental data products, and regional climate model data that are described below.

Insurance Data The insurance data consists of redacted, anonymized and aggregated home insurance claims (counts and dollar losses) incurred due to weather damage reported by postal code over ten years spanning 2002–2011. The data have been internally validated for quality

control, and the selected geographic locations represent four urban areas in Canada, which we call City A, City B, City C, and City D. City A is situated in the continental climate zone in the high prairie, City B directly on a lake in the Great Lakes area, City C in eastern Canada with cold and temperate continental climate characterized by four distinct seasons, and City D in eastern Canada with a similar hemiboreal climate with no dry season but at double the elevation of City C. All four selected cities are medium-sized and have similar population densities. At the pre-processing stage, we normalize the daily number of claims by the number of policyholders on each day and adjust the dollar amounts by the metropolitan area composite quarterly index of apartment building construction (Table 327-0044, http://www5.statcan.gc.ca). Since we have no data on collected premiums and thus are unable to evaluate the loss ratios, similarly to Frees et al. (2012), we focus on the assessment of claim frequencies and severities (see Frees et al., 2014 for a detailed discussion and literature overview).

Weather Observations The weather observations we use in our analysis are provided by Environment Canada in their Digital Archive of Canadian Climatological Data and are available from http://climate.weather.gc.ca/. Daily climatological data are obtained for each city, from which we use mean temperature, total precipitation (mm, snow is converted to water equivalents), and speed of maximum gust (km/h) for our analyses. Gust speeds below 31 km/h (8.6 m/s) are not reported, but this is well below the wind level that would be expected to cause damage claims (Stewart, 2003).

Climate Model Output Data from three different climate model experiments used in this study are derived from the CanRCM4 version of the Canadian Regional Climate Model run by the Canadian Center for Climate Modeling and Analysis. The model is described in Scinocca et al. (2016) and the model experiments used here were part of the North American Coordinated

Regional Climate Downscaling Experiment (NA-CORDEX) (Giorgi & Gutowski, 2015). Regional climate models require lower resolution input data from either processed observational datasets or a global-scale climate model.

To evaluate uncertainty due to various scenarios of global warming, we use available data at 25 km (0.22 degree) spatial resolution from two CanRCM4 experiments run with the second generation Canadian Earth System Model, CanESM2, as input data. In particular, we consider two runs from 2006–2100, forced with greenhouse gas scenarios known as RCP 4.5 and RCP 8.5. These Representative Concentration Pathways (RCPs) are commissioned by the Intergovernmental Panel on Climate Change (IPCC) and have been used widely in different efforts to study potential impacts of climate change. The RCP 4.5 scenario stabilizes greenhouse forcing to 4.5 W/m^2 before 2100 and represents a world where greenhouse gas emissions are brought under control by a variety of human actions. RCP 8.5 is a business as usual scenario where greenhouse gas forcing from human activities is not curtailed and rises to 8.5 W/m^2 by 2100. The climate in each of the model runs develops independently, so small differences between model runs are expected due to normal stochastic noise in the climate system.

In addition, to assess uncertainty due to the potential discrepancy between climate model output and weather observations, we consider a third experiment of the CanRCM4 regional model that was run for 1989–2009 with ERA-Interim reanalysis data (Dee et al., 2011) as input data. A climate reanalysis provides an estimate of the state of the climate system over the whole globe based on observations, by coupling models with the observations. Like global climate model output, reanalysis data are continuous (unlike true observations) and can be used as the input data for the regional climate model. The ERA-Interim run of CanRCP4 represents a

validation experiment to test how well the model represents the real world when the boundary condition inputs are from the real world rather than a model simulation.

3. Methods

To develop an attribution analysis of current weather-related losses and to assess dynamics of future claim severities, we combine the approach of Soliman et al. (2015) to modeling a number of weather-related claims, with a collective risk model simulation algorithm of Meyers et al. (2003). As a result, we propose a new method for joint frequency-severity predictive modeling of weather-related claim severities.

3.1. Triggering thresholds and attribution analysis

The first step of our methodology is to find critical thresholds for weather variables after which the severity of daily claims starts to increase monotonically (Soliman et al., 2015). This knowledge of the weather effects is useful and important not only for accommodating nonlinearities in statistical models, but also for the insurance industry to: 1) dispatch insurance inspectors more efficiently (short-term); 2) establish appropriate building guidelines (long-term), and 3) formulate mitigation and prevention policies (long-term).

We use two methods to find the thresholds for observed insurance claim counts vs. precipitation and wind speed: Classification and Regression Trees (CART; Breiman et al., 1984; Derrig & Francis, 2006; Hadidi, 2003) and Alternating Conditional Expectations (ACE) transformations (Breiman & Friedman, 1985). In addition to precipitation and wind speed, we considered a set of other atmospheric variables as potential predictors, including surface air temperature and pressure. However, our model selection analysis based on the Akaike Information Criterion (AIC) indicated that surface air temperature and pressure are not

significant predictors and do not provide an additional predictive gain. (AIC is a model selection criterion that is widely used in numerous actuarial applications, see Brockett, 1991; Frees et al., 2014, and references therein.)

The key ideas of CART (Breiman et al., 1984) are to:

- use piecewise constant functions to approximate non-linear relationship between covariates (weather events) and a response variable (the number of house insurance claims due to adverse weather);

- partition the predictor space with binary splits formulated in terms of the covariates, e.g., identify thresholds in precipitation amounts that iteratively split the range for this variable into intervals;

- divide the whole domain into relatively homogenous subdomains, so the thresholds are chosen to provide the most homogeneous responses (claim counts) within each subdomain.

The tree-based models do not aim to construct an explicit global linear model for prediction or interpretation. Instead, the nonparametric tree-based approaches bifurcate the data recursively at critical points of covariates in order to split the data into groups that are as homogeneous as possible within and as heterogeneous as possible in-between.

ACE is a nonparametric procedure that smoothly transforms predictors X_j to maximize the correlation between the response Y and transformed predictors $f_j(X_j)$ in a linear additive model (Breiman & Friedman, 1985):

$$Y = \alpha + \sum_{j=1}^{k} f_j(X_j) + \epsilon,$$

where $f_j(\cdot)$ are some smooth nonlinear functions, k is the number of predictors, and ϵ are independent and identically distributed random variables. ACE can also employ a smooth transformation of response, i.e., $\theta(Y)$, with the goal to minimize

$$E\{\theta(Y) - \sum_{j=1}^{k} f_j(X_j)\}^2.$$

ACE transformations are splines with a pre-defined degree of smoothness, which is a special case of the Generalized Additive Models (GAMs) (see Frees et al., 2014, for an overview on GAMs in actuarial sciences). In addition, a user can define such conditions as monotonicity and periodicity. In our analysis, we apply monotonic transformations to predictors without transforming the response.

Similar to Soliman et al. (2015), we identify the thresholds from ACE transformation plots as the location of a breakpoint, which separates values of a weather predictor with trivial or no impact on the claim counts from the values having a significant impact. We finally truncate the weather predictors by setting the observations below selected thresholds to zero.

For the attribution analysis, we use a Generalized Autoregressive Moving Average (GARMA) approach by Benjamin et al. (2003), which accommodates integer-valued time series as well as non-Gaussianity of the variables. Let $Y_1,\ldots, Y_t$ be observed daily number of claims and X_t be a matrix of some exogenous regressors (precipitation, wind speed, etc.). Then, we can model conditional distribution of Y_t, given $Y_1,\ldots, Y_{t-1}, X_1,\ldots, X_t$, as

$$g(\mu_t) = \alpha + X_t'\beta + \sum_{j=1}^{p} \phi_j\{g(Y_{t-j}) - X_{t-j}'\beta\} + \sum_{j=1}^{q} \theta_j\{g(Y_{t-j}) - g(\mu_{t-j})\}, \qquad (3.1)$$

where $g(\cdot)$ is an appropriate link function. We use a zero-adjusted Poisson distribution to model daily claims and link function $g(\cdot)$ is $\ln(\cdot)$. See Cameron and Trivedi (2013), Creal et al. (2013), Fokianos (2015), and references therein for discussion of GARMA and other models for time series of counts.

The ACE-selected thresholds for precipitation and wind speed are reported in Table 1, along with the modeling results of the number of claims vs. the weather predictors in GARMA framework (3.1), where model order q was selected using AIC. The variability of estimated parameters is not high and City B has generally higher β's implying higher vulnerability to severe weather, which may be related to the city's location directly in the Great Lakes area.

Table 1. Results of attribution analysis and modeling of daily number of weather related claims using weather predictors (precipitation and wind speed) in 2002–2011. Standard errors are in parentheses

	Thresholds		GARMA(0, q) model parameters		
City	*Precipitation, mm/day*	*Wind speed, km/h*	$\hat{\beta}_{pcp}$	$\hat{\beta}_{wind\ speed}$	q
A	5	50	0.032 (0.004)	0.013 (0.001)	1
B	1	50	0.034 (0.004)	0.037 (0.002)	6
C	1	45	0.024 (0.002)	0.021 (0.001)	3
D	5	55	0.026 (0.002)	0.014 (0.001)	1

Remark. It is important to emphasize that atmospheric thresholds that are currently adopted in civil engineering for building codes and standards primarily target extreme catastrophic events (see, for instance, Crawford & Seidel, 2013; Gillespie et al., 2016; Obama, 2015) and do not account for the *low individual but high cumulative impact* events, which have exhibited rising frequencies over time but still remain a gray zone for risk management in the public sector. Hence, the burden of handling the outcomes of such low individual but high cumulative impact events is still on the insurers and their clients. The study proposed here allows to shed a light on the critical thresholds and methods to derive them that could be used by the insurance companies to mitigate the risks and minimize their losses – before the public sector entities recognize the challenges of the gray zone of low individual but high cumulative impact events.

3.2. Adaptive extrapolation of the claim frequencies

The selected thresholds and estimated model parameters further play a role in assessing the future trends in the number of insurance claims. We obtain future values of the atmospheric variables at daily resolution from the climate model projections. Similar to the modeling step (Section 3.1), we focus on the distribution tails above certain thresholds, which are both most important in insurance claims modeling and most dynamic as they exhibit greater changes than the rest of the distribution (Figure 2). Algorithm 1, which is a multivariate extension of the quantile-based method of Soliman et al. (2015), is based on the idea of numerical integration and allows us to naturally compare tails of two distributions (e.g., observed and forecasted precipitation) and use this information together with GARMA output to incorporate the climate change effects into claims forecasts.
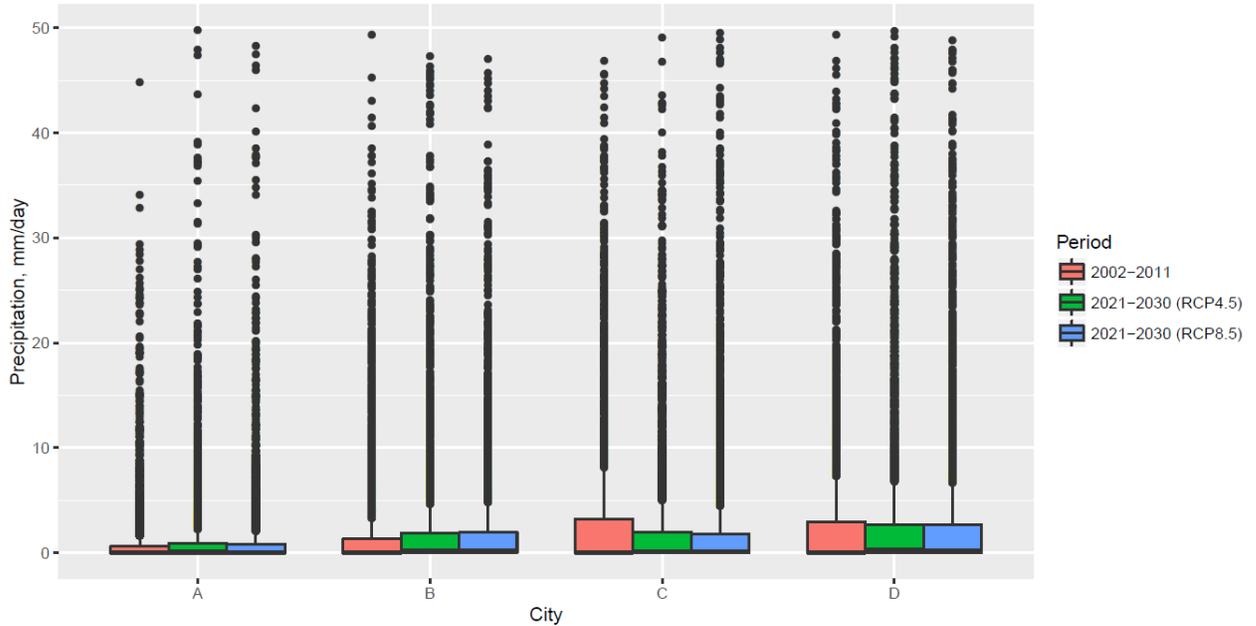

Figure 2. Box-plots of daily precipitation below 50 mm/day observed in 2002–2011 and projected for 2021–2030 under RCP 4.5 and RCP 8.5

11

Algorithm 1. Incorporating the effects of weather extremes in extrapolation of the claim frequencies (modified multivariate version of the quantile-based method by Soliman et al., 2015)

1. Select a percentile threshold l_i for the i-th weather predictor (i=1,…,k), based on the baseline (reference, observed) distribution. Here we set l_i as the 90th percentile of observed distribution for all i.

2. Let N_i be the number of observations above l_i, which are separated into w_i equally-filled intervals with $n_i = N_i/w_i$ observations in each. The number of intervals w_i can be automatically selected, e.g., using the Sturges' formula: $w_i = \lceil \log_2 N_i \rceil + 1$.

3. Let C_{ij} be a center of j-th interval, then find quantile distances between climate projections and observations: $P_{ij} = C_{ij}^{fcst} - C_{ij}^{obs}$, where j=1,…,w_i.

4. Repeat the steps above for each i=1,…,k and calculate the overall change in the number of claims suggested by the tail dynamics of the weather predictors:

$$\Delta = \sum_{i=1}^{k} \sum_{j=1}^{w_i} G(\widehat{\beta}_i P_{ij}) n_i,$$

where $G(\cdot)$ is an inverse function of $g(\cdot)$.

We now apply Algorithm 1 to the claim dynamics in four cities, and summarize our findings in Table 2. In particular, Table 2 suggests that claim counts will likely exhibit increasing trends in all four cities over the next decade of 2021–2030. The major increase is expected for City B that is directly on a lake, indicating that changes in lake-derived moisture and its representation in the model should be explored with detail to fully understand this result. In general, higher elevation leads to a lower increase in claim counts, but it is very uncertain whether this is due to a reduced change in climate with elevation, or simply an artifact of the model representation of elevation

and its dynamical impacts. There is not a large difference between the two scenarios for this time period because the climate projections do not diverge enough by 2021–2030 to make a substantial difference in the analysis. The small differences of a few percentage points between scenarios are likely simply due to climate system noise expected between different model runs.

Table 2. Summary of projected increase (Δ, %) in weather-related claim counts in 2021–2030, relative to the baseline level of 2002–2011

| City | Climate scenario for 2021–2030 | |
	RCP 4.5	RCP 8.5
A	20	21
B	89	99
C	35	33
D	30	28

3.3. Future projections in respect to different climate comparison scenarios

We now conduct an extensive study to examine the discrepancies arising from using different climate model outputs and observations. Specifically, our previously discussed analysis has been based on weather observations in 2002–2011, and future projections have been based on comparing future climate model output against the weather observations.

To evaluate the uncertainty due to climate model projections and the model discrepancies in estimating the "true" state of atmosphere, here we use an alternative to weather observations – output of the same climate model produced during an evaluation run for the recent past (1989–2009) which was forced with ERA-Interim reanalysis data. This analysis allows us to compare outputs of a climate model representation of the recent past and of the future (climate scenarios RCP 4.5 and RCP 8.5), at the same space-time resolution, to evaluate discrepancies in our analysis that arise due to differences between the climate model output and the real world

observations. A disadvantage of using ERA-Interim forced output is the lack of model output data for certain years and varying data availability for different climate models. In our case, evaluations continued up to the year 2009, whereas available insurance data cover 2002–2011, thus, we had to reduce the period of analysis from ten to eight years: 2002–2009.

The climate model adequately represents the variables of the climate system that are important for our analysis. The model data are characterized by a higher proportion of days with very low precipitation (Figure 3), which is compensated by also higher extremes, such that the centers of the distributions, i.e., the 1st–3rd quartiles, remain close to each other (Table 3).

Table 3. Summary comparison of daily observed precipitation and output of the ERA-Interim evaluation, for the 8-year period of 2002–2009 (mm/day)

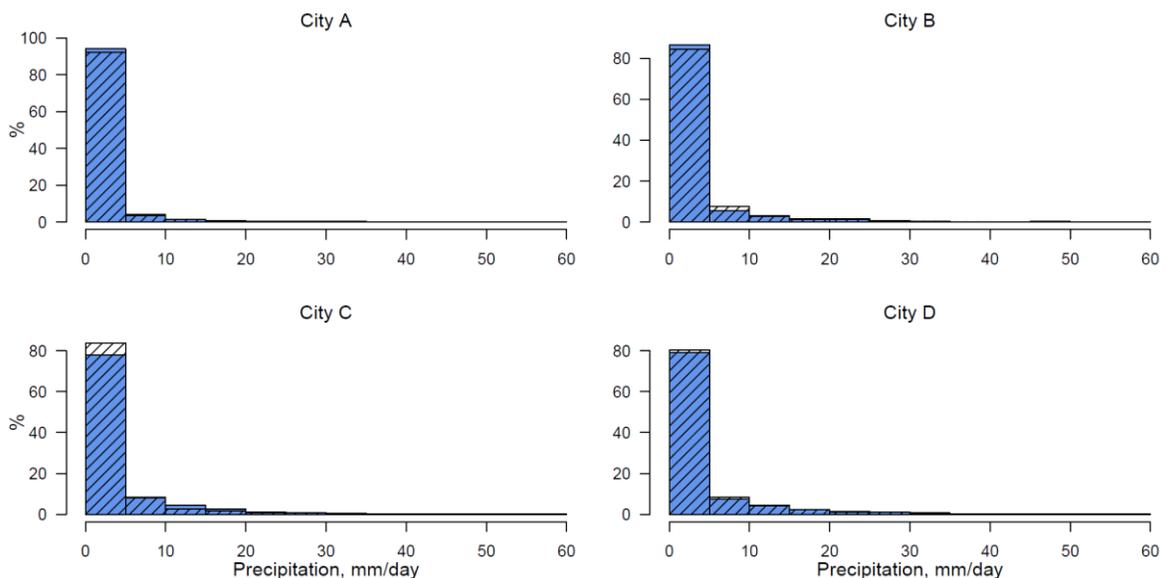| City	Data	Min.	1st quartile	Median	Mean	3rd quartile	Max.
A	Observations	0	0	0.0	1.0	0.6	34.1
	ERA-Interim evaluation	0	0	0.1	1.6	1.0	82.0
B	Observations	0	0	0.0	2.2	1.3	62.9
	ERA-Interim evaluation	0	0	0.2	2.9	2.4	107.3
C	Observations	0	0	0.0	3.2	3.2	75.5
	ERA-Interim evaluation	0	0	0.2	3.2	2.7	153.3
D	Observations	0	0	0.0	3.2	3.0	88.3
	ERA-Interim evaluation	0	0	0.3	3.7	3.2	114.6

Figure 3. Frequency distribution of original (blue) daily precipitation and output of the ERA-Interim reanalysis (shaded) for the 8-year period of 2002–2009

Table 4. Projected change in the number of claims (Δ, %), depending on the source of weather data and scenario of comparisons. The compared 8-year periods are current (2002–2009) and future (2021–2028)

Comparison scenario:		Scenario 1		Scenario 2		Scenario 3	
Current weather data (used in attribution analysis):		Observations		ERA-Interim evaluations		Observations	
Future is compared against current:		Observations		ERA-Interim evaluations		ERA-Interim evaluations	
Future climate scenario:		RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
City:	A	21	22	25	26	25	25
	B	94	100	118	116	119	116
	C	32	31	37	36	35	35
	D	32	29	37	36	38	36

The projected number of weather-related home insurance claims were compared between three possible analysis scenarios summarized in Table 4. The scenarios differ based on (i) which weather data source (observations or climate model output) we use for estimating critical thresholds and GARMA modeling; and (ii) whether we compare future climate model output with the model itself (ERA-Interim evaluations) or with the observed data.

Scenario 1. Model number of claims (*ncl*) as a function of observed data (*obs*), i.e., *ncl* ~ $b_1 \times obs$. Compare tails of future climate projections with observed data and estimate the change δ_1. The projected change in a future number of claims Δ is a function of $b_1 \times \delta_1$.

Scenario 2. Model number of claims (*ncl*) as a function of the reanalysis data (*era*), i.e., *ncl* ~ $b_2 \times era$. Compare tails of future climate projections with *era* and estimate the change δ_2. The projected change in a future number of claims Δ is a function of $b_2 \times \delta_2$.

Scenario 3. Model number of claims (*ncl*) as a function of observed data (*obs*), i.e., *ncl* ~ $b_1 \times obs$. Compare tails of future climate projections with *era* and estimate the change δ_3. The projected change in a future number of claims Δ is a function of $b_1 \times \delta_3$.

Remarkably, all three comparison scenarios agree not just in a common upward trend (number of claims is increasing) across the four considered cities, but also in similar forecasted changes for each city across the three comparison scenarios. This validates our original approach (Scenario 1), which has an advantage of using a longer data set (subject to insurance data availability, which is the period of 2002–2011 in our case) for reducing the uncertainty in statistical models. Hence, in our further analysis we proceed with Scenario 1.

4. Joint Frequency-Severity Predictive Modeling

To forecast the dollar amounts (severity) of incurred losses, we use the Collective Risk Model (CRM) approach. This approach entails integrating our GARMA-based projections for claim frequencies with the simulation algorithm of Meyers et al. (2003) to evaluate the distributions of future claim severities. We start from the CRM method of Meyers et al. (2003) adapted for one line of insurance (Algorithm 2).

Algorithm 2. Collective risk model (CRM) simulation (a modified version of the CRM method by Meyers et al., 2003)

1. Select a random count K from a distribution of daily number of claims with mean λ, where λ is the expected claim count.

2. Select random claim sizes Z_k, where $k = 1, \ldots, K$.

3. Calculate daily loss $X = \sum_{k=1}^{K} Z_k$.

4. Repeat the three steps above B times and use $\{X\}_1^B$ to construct a distribution for daily claim severities.

Algorithm 2 can be applied using the observed distributions of daily number of claims and claim-by-claim losses. However, to obtain forecasts using Algorithm 2, we need an approximation to the distribution of claim counts, so we can adjust it to the future scenarios of claim frequencies. We considered a range of distributions to approximate the daily number of claims (Delaporte, negative binomial of types I and II, Poisson, Poisson inverse Gaussian, Sichel, and zero-inflated/adjusted Poisson) (Frees et al., 2014; Trowbridge, 1989) and found that the Sichel distribution (Figure 4) delivers the most competitive results based on AIC. The Sichel distribution is a three-parameter compound Poisson distribution; the two-parameter form is known as the inverse Gaussian Poisson (Stein et al., 1987). The Sichel distribution provides a flexible alternative for modeling highly skewed distributions of observed counts, and its utility in actuarial applications has been introduced by Willmot (1986), Panjer and Willmot (1990), and Willmot (1993), and then extensively studied in a variety of insurance settings (see, for instance, Bermúdez & Karlis, 2012; Hera-Martines et al., 2008; Tzougas & Frangos, 2014, and references therein). Table 5 reports the estimated parameters of the Sichel distribution fitted to the number of claims in each city in the period 2002–2011. The Kolmogorov–Smirnov and Anderson–Darling goodness of fit tests at 5% significance level do not provide evidence against the fitted Sichel distributions for City A and City C. The null hypothesis is rejected by these tests in the case of City B and City D, due to outliers, however, the Sichel distribution still provides the best fit among all other considered distributions.

Table 5. Estimated parameters of the Sichel distributions for the normalized number of claims

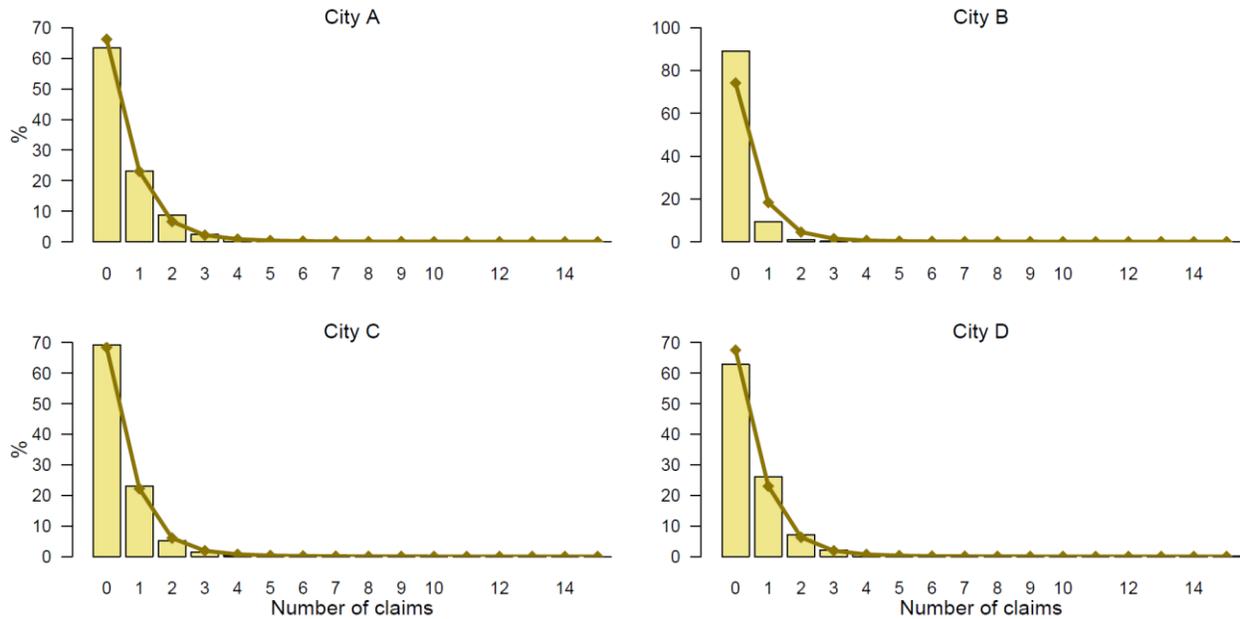| Parameter	City			
	A	B	C	D
μ	4.99	1.82	2.83	2.31
σ	4.85	3.77	3.00	2.10
ν	-1.89	-1.45	-1.88	-2.01

Figure 4. Results of fitting Sichel distribution curves to the series of normalized number of claims in four cities (bars); x-axes are truncated to the same value

To validate the approach of approximating the number of claims with a Sichel distribution, we first consider the distributions of the observed claim severities vs. their approximations using Algorithm 2 (Figure 5). The worst fit is observed for City B, which is not surprising given that it has very few (about 11%) days with non-zero losses, compared with more than 30% of days with losses in the other three cities. In the cases with more non-zero data (City A, City C, and City D), the simulated distributions mimic the distributions of observed severities well; therefore we

proceed with using the Sichel distribution in the CRM approach (Algorithm 2) for assessing the projected distributions of claim severities.
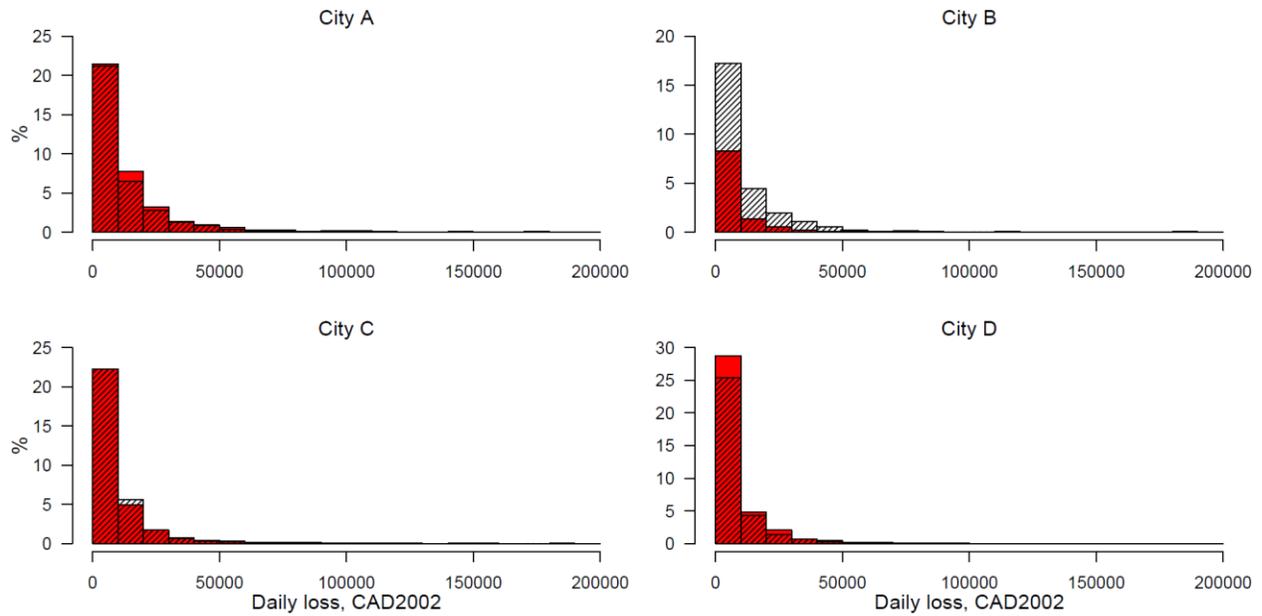

Figure 5. Original (red) and approximated with Algorithm 1 (shaded) distributions of daily losses in the four cities. Y-axis represents the percent of the total number of days in the period 2002–2011. (Days with zero losses are counted toward the total of 100% but not plotted)

Further, we employ results of implementing Algorithm 1 (Section 3.2) under comparison Scenario 1 (Section 3.3) for the longest period of data availability for the given scenario (10 years: 2002–2011). Table 2 shows the estimated increase in claim frequencies, which implies the respective changes in the parameter μ of a Sichel distribution. We now apply the CRM Algorithm 2, where simulations of the future number of claims in the Step 1 are based on the Sichel distribution with parameters $\mu+\Delta$.

Table 6. Summary of the observed and projected percentiles of daily weather-related losses in CAD2002

City	Percentile	Observed in 2002–2011	Climate scenario for 2021–2030	
			RCP 4.5	RCP 8.5
A	75th	4,258	4,124	4,237
	90th	15,451	15,868	16,250
	95th	25,814	28,161	29,559
B	75th	0	4,205	4,688
	90th	704	15,738	17,049
	95th	5,015	29,228	30,689
C	75th	1,722	3,347	3,370
	90th	8,829	12,466	12,144
	95th	15,847	20,603	20,647
D	75th	2,506	2,637	2,636
	90th	8,393	9,217	9,175
	95th	15,770	16,289	15,840

Table 6 presents the observed and projected percentiles for daily losses under the scenarios RCP 4.5 and RCP 8.5. All four cities exhibit a noticeable increase in losses, and the most profound changes are suggested for City B (both in lower and upper percentiles). Such findings can be explained by the greater expected increase in claim frequencies (Table 2), the location of City B (directly on the lake) as well as higher variability of fit for City B (the approximation of the number of claims by the Sichel distribution is less accurate for this city, which affects the approximation of the losses).

Remarkably, although the projected changes in mean frequency of claims in other cities are relatively synchronous (20–35% increase for City A, City C, and City D; Table 2), the changes in individual percentiles are more variable and range from a decline of 3% (in 75th percentile under RCP 4.5 for City A) to an increase of 94–96% (in 75th percentile for City C).

5. Conclusion

In this paper, we present a new method for modeling and forecasting house insurance losses due to non-catastrophic weather events. Our data-driven approach is based on integrating two algorithms, that is, a multivariate extension of a nonparametric method for analysis of weather-related claim counts by Soliman et al. (2015) and a simulation-based algorithm for collective risk modeling by Meyers et al. (2003). We evaluate uncertainty due to potential differences between climate model output and weather observations and evaluate the estimated losses from two different global warming scenarios. As a result, we produce a projected distribution of weather-related claim severities that intrinsically accounts for local geographical characteristics and, hence, is suitable for short- and long-term management of insurance risk due to adverse weather events. For instance, in the case of short-term risk management, when the projected precipitation is expected to exceed the critical threshold defined by our methodology, the insurance company may pro-actively hire more insurance assessors and loss adjusters, to ensure timely processing of claims. In a case of medium- and long-term risk projections, the results of this study can be used to re-evaluate insurance premiums and a range of insurance products offered in a particular area, as well as to offer incentives for homeowners who reduce vulnerability of their property to adverse weather effects, and more closely collaborate with public sector entities on introducing new weather-resistance building codes and standards.

Despite the well-documented effects of adverse weather on the insurance industry, our study is one of the first attempts to understand and evaluate the impact of non-catastrophic atmospheric events on the insurance sector, particularly, home insurance. Hence, we would like to conclude the paper with discussing a number of caveats, limitations, and open research directions.

In this proof-of-concept study, we use data from only a few locations over 10 years, but obviously, a broader spatial and temporal extent of data (to which we did not have access) would make the results more practically applicable. Climatologists traditionally use 30 years of weather data to describe the expected "normal" weather, however in a changing climate, that length of record may be too long since we expect the weather to change significantly over 30 years (Huang et al., 1996; Livezey et al., 2007). It has been suggested that temperature and precipitation over the continental United States can be well characterized by only 10 and 15 years of data, respectively, providing the optimal predictive skill for the following year (Huang et al., 1996). Thus, our analysis of 10 years is reasonable, though an analysis of the length of record needed for optimum predictive skill in each variable of interest could be employed by industry analysts with access to larger actuarial data sets and existing publically available weather and climate data. Expansion of this method to spatially more extensive data sets in the future would provide important information about regional patterns of expected future climate-claims relationships. With such information, we expect companies could adjust rates to more appropriately reflect the changing potential risk of "normal" weather hazards in a region.

From a practical standpoint, this type of analysis would need to be updated as model data become available (approximately every 5 years) to continue to use cutting edge information of future climate. Climate models, both regional and global, are constantly being improved. We used the latest available North America regional dynamically downscaled climate data at the time of our analysis, but the global model projection data are updated every five years for the Intergovernmental Panel on Climate Change reports and new regional model data are available accordingly, if funding for the science is continued.

The approach of our study side-steps the traditional notion that the past is a feasible predictor of the future. In social and natural sciences, information from the past behavior of a system is often used as a predictor of the future behavior of the system. However, this assumption is valid only as long as there is a reasonable expectation that the dynamics of the system does not vary (significantly) over time. Anthropogenic climate change, however, causes a relatively rapid change in the Earth's climate system, which puts that assumption in jeopardy for many social and environmental systems (IPCC, 2014b). Milly et al. (2008) put it succinctly for the field of water management in the title of their article, "Stationarity is dead: whither water management?" How can we assess the impact of such changes in the Earth's climate system and how can these changes be incorporated into insurance risk management? How are these changes distributed in space and time? This is of particular importance for regional risk management as weather-related events are long known to induce spatial dependencies among insurance policies, often leading to undercapitalization and excessive risk-taking (Cooke & Kousky, 2010).

Although insurance policies are adjusted regularly and therefore can theoretically be changed to reflect current conditions, the episodic nature of many weather-related perils and relatively rapid changes cause a lag between the observed event statistics and the current real probability of an event that is increasing in frequency. How can we get ahead of this lag?

These and many other related fundamental questions on insurance risk management at the face of climate change can be addressed only if the gaps between paradigms of actuarial theory and practice, statistics, and climate sciences are bridged together, forging joint interdisciplinary efforts.

References

Benjamin, M. A., Rigby, R. A., & Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American statistical Association, 98*(461), 214-223.

Bermúdez, L., & Karlis, D. (2012). A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics & Data Analysis, 56*(12), 3988-3999.

Bouvet, L., & Kirjanas, P. (2016). *Global climate 500 index 2016: Insurance sector analysis*. London: AODProject.

Breiman, L., & Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association, 80*(391), 580-598.

Breiman, L., Friedman, J. H., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC Press.

Brockett, P. L. (1991). Information theoretic approach to actuarial science: A unification and extension of relevant theory and applications. *Transactions of the Society of Actuaries, 43*, 73-135.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53): Cambridge University Press.

Cheng, C. S., Li, Q., Li, G., & Auld, H. (2012). Climate change and heavy rainfall-related water damage insurance claims and losses in Ontario, Canada. *Journal of Water Resource and Protection, 4*, 49-62.

Cooke, R., & Kousky, C. (2010) Climate dependencies and risk management: Microcorrelations and tail dependence. *Vol. 10-13*. Washington: Resources for the Future.

Crawford, M., & Seidel, S. (2013). *Weathering the Storm: Building Business Resilience to Climate Change*. Arlington: Center for Climate and Energy Solutions.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics, 28*(5), 777-795.

Derrig, R., & Francis, L. (2006). *Distinguishing the forest from the TREES: a comparison of tree based data mining methods.* Paper presented at the Casualty Actuarial Society Forum.

Fokianos, K. (2015). Statistical Analysis of Count Time Series Models: A GLM Perspective. In R. A. Davis, S. H. Holan, R. Lund, & N. Ravishanker (Eds.), *Handbook of Discrete-Valued Time Series* (pp. 3-28): CRC Press.

Frees, E. W., Derrig, R. A., & Meyers, G. (2014). *Predictive Modeling Applications in Actuarial Science*: Cambridge University Press.

Frees, E. W., Meyers, G., & Cummings, A. D. (2012). Predictive modeling of multi-peril homeowners insurance. *Variance, 6*(1), 11-31.

Gillespie, C. W., Antes, M., & Donnelly, P. (2016). *Climate Change and the Electricity Sector: Guide for Climate Change Resilience Planning*. Washington: U.S. Department of Energy.

Giorgi, F., & Gutowski, W. J. (2015). Regional Dynamical Downscaling and the CORDEX Initiative. In A. Gadgil & T. P. Tomich (Eds.), *Annual Review of Environment and Resources* (Vol. 40, pp. 467-490). Palo Alto: Annual Reviews.

Grossi, P., Kunreuther, H., & Windeler, D. (2005). An introduction to catastrophe models and insurance *Catastrophe modeling: A new approach to managing risk* (pp. 23-42): Springer.

Hadidi, N. (2003). *Classification ratemaking using decision trees.* Paper presented at the Casualty Actuarial Society Forum.

Haug, O., Dimakos, X. K., Vårdal, J. F., Aldrin, M., & Meze-Hausken, E. (2011). Future building water loss projections posed by climate change. *Scandinavian Actuarial Journal, 2011*(1), 1-20.

Held, H., Gerstengarbe, F.-W., Pardowitz, T., Pinto, J. G., Ulbrich, U., Born, K., Donat, M. G., Karremann, M. K., Leckebusch, G. C., Ludwig, P., Nissen, K. M., Osterle, H., Prahl, B. F., Werner, P. C., Befort, D. J., & Burghoff, O. (2013). Projections of global warming-induced impacts on winter storm losses in the German private household sector. *Climatic Change, 121*(2), 195-207.

Hera-Martines, A., Gil-Fana, J. A., & Vilar-Zanon, J. L. (2008). Claim counts modeling and stable distributions. *Astin Bulletin, 37*(7), 1-9.

Huang, J., Dool, H. M. v. d., & Barnston, A. G. (1996). Long-Lead Seasonal Temperature Prediction Using Optimal Climate Normals. *Journal of Climate, 9*(4), 809-817. doi:10.1175/1520-0442(1996)009<0809:llstpu>2.0.co;2

IPCC. (2014a). *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* (C. B. Field, 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, & L. L. White Eds.). Cambridge, United Kingdom and New York, USA.

IPCC. (2014b). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (R. K. Pachauri & L. A. Meyer Eds.). Geneva, Switzerland: IPCC.

Kunreuther, H., & Michel-Kerjan, E. (2012). *Policy options for reducing losses from natural disasters: Allocating $75 billion*

Livezey, R. E., Vinnikov, K. Y., Timofeyeva, M. M., Tinker, R., & Dool, H. M. v. d. (2007). Estimation and Extrapolation of Climate Normals and Climatic Trends. *Journal of Applied Meteorology and Climatology, 46*(11), 1759-1776. doi:10.1175/2007jamc1666.1

LSE. (2015). *Evaluating the economics of climate risks and opportunities in the insurance sector*. Retrieved from

Meyers, G. G., Klinker, F. L., & Lalonde, D. (2003). *The aggregation and correlation of insurance exposure.* Paper presented at the Casualty Actuarial Society Forum.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science, 319*(5863), 573-574. doi:10.1126/science.1151915

NAIC. (2016, 06/13/2016). Climate change and risk disclosure. Retrieved from http://www.naic.org/cipr_topics/topic_climate_risk_disclosure.htm

Executive Order 13690 of Jan 30, 2015 -- Establishing a Federal Flood Risk Management Standard and a Process for Further Soliciting and Considering Stakeholder Input, 80 C.F.R. (2015).

Panjer, H. H., & Willmot, G. E. (1990). *Insurance Risk Models*: JSTOR.

Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O., Hinnerichsen, M., & Meze-Hausken, E. (2013). A Bayesian hierarchical model with spatial variable selection: The effect of weather on insurance claims. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 62*(1), 85-100.

Scinocca, J. F., Kharin, V. V., Jiao, Y., Qian, M. W., Lazare, M., Solheim, L., Flato, G. M., Biner, S., Desgagne, M., & Dugas, B. (2016). Coordinated global and regional climate modeling. *Journal of Climate, 29*(1), 17-35. doi:10.1175/jcli-d-15-0161.1

Smith, A. B., & Matthews, J. L. (2015). Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. *Natural Hazards, 77*(3), 1829-1851. doi:10.1007/s11069-015-1678-x

Soliman, M., Lyubchich, V., Gel, Y. R., Naser, D., & Esterby, S. (2015). Evaluating the impact of climate change on dynamics of house insurance claims. In V. Lakshmanan, E. Gilleland, A. McGovern, & M. Tingley (Eds.), *Machine Learning and Data Mining Approaches to Climate Science* (pp. 175-183): Springer.

Stein, G. Z., Zucchini, W., & Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American statistical Association, 82*(399), 938-944.

Stewart, M. G. (2003). Cyclone damage and temporal changes to building vulnerability and economic risks for residential construction. *Journal of Wind Engineering and Industrial Aerodynamics, 91*(5), 671-691. doi:10.1016/S0167-6105(02)00462-2

Toumi, R., & Restell, L. (2014). *Catastrophe modelling and climate change*: Lloud's.

Trowbridge, C. L. (1989). *Fundamental concepts of actuarial science*: Actuarial Education & Research Fund.

Tzougas, G., & Frangos, N. (2014). The Design of an Optimal Bonus-Malus System Based on the Sichel Distribution *Modern Problems in Insurance Mathematics* (pp. 239-260): Springer.

Willmot, G. E. (1986). Mixed compound Poisson distributions. *Astin Bulletin, 16*(S1), S59-S79.

Willmot, G. E. (1993). On recursive evaluation of mixed Poisson probabilities and related quantities. *Scandinavian Actuarial Journal, 1993*(2), 114-133.