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

by Vyacheslav Lyubchich, K. Halimeda Kilbourne, and Yulia R. Gel

ABSTRACT

We present an attribution analysis of residential insurance losses due to noncatastrophic 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

15080-09_Lyubchich-2ndPgs.indd 278 4/25/19 10:55 AM

1. Introduction

There is a growing body of scientific evidence that extreme weather events are increasing in frequency and intensity (Intergovernmental Panel on Climate Change [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 and Kirjanas 2016; LSE 2015; NAIC 2016; Smith and 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, e.g., overviews by Grossi, Kunreuther, and Windeler 2005; Kunreuther and Michel-Kerjan 2012; Toumi and 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; Lyubchich and Gel 2017) rather than on the dynamics of related losses. A few recent studies on predicting insurance losses due to extreme but noncatastrophic 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 with 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 noncatastrophic 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 1.1; 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 1.1). Section 3 showcases the methods of attribution analysis and tail comparison (Steps 2 and 3 in Figure 1.1). Section 4 presents joint frequency-severity predictive modeling and forecasts (Step 4 in Figure 1.1). The paper is concluded with a discussion in Section 5.

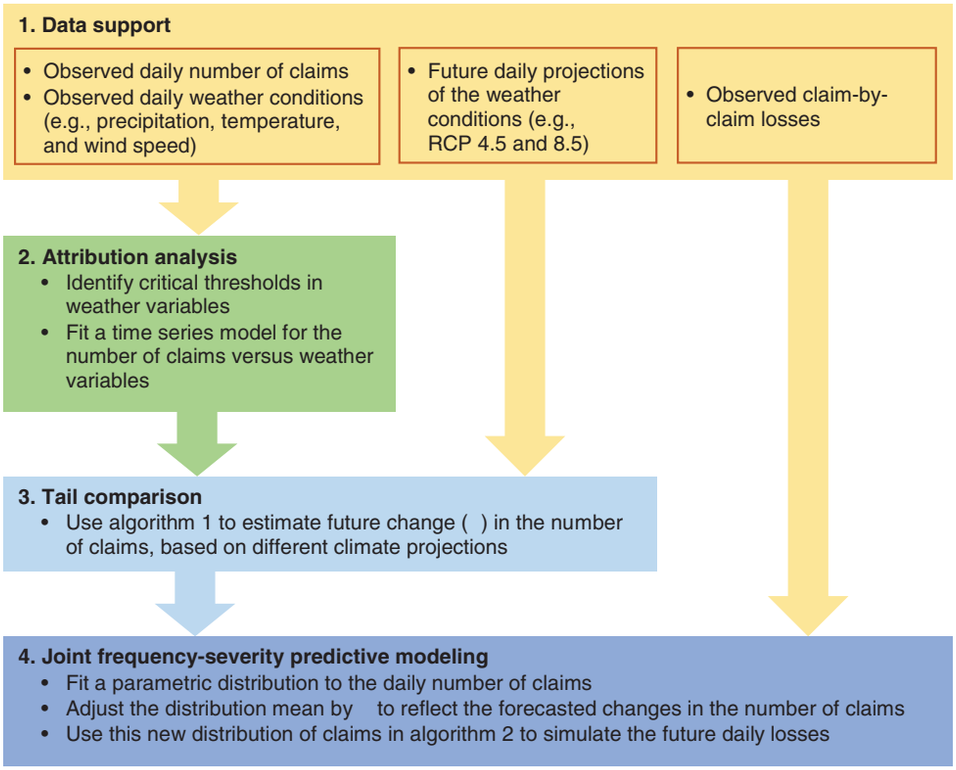
2. Data

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 consist of redacted, anonymized, and aggregated home insurance claims (counts and dollar losses) incurred due to weather damage reported by postal code during 10 years spanning 2002 to 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

15080-09_Lyubchich-2ndPgs.indd 279 4/25/19 10:55 AM

Figure 1.1. Flow chart of the attribution analysis and joint forecast of daily number and severity of weather-related home insurance claims



Note: RCP = Representative Concentration Pathway.

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 preprocessing 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, Meyers, and Cummings (2012), we focus on the assessment of claim frequencies and severities (see Frees, Derrig, and Meyers 2014 for a detailed discussion and literature overview).

Weather Observations The weather observations we use in our analysis are provided by Environment Canada in its 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 (millimeters, snow is converted to water equivalents), and speed of maximum gust (kilometers per hour) for our analyses. Gust speeds less than 31 kilometers per hour (8.6 meters per second) 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 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 (Giorgi and Gutowski 2015). Regional climate models require lower-resolution input data from either processed observational data sets or a global-scale climate model.

15080-09_Lyubchich-2ndPgs.indd 280 4/25/19 10:55 AM

To evaluate uncertainty due to various scenarios of global warming, we use available data at 25 kilometers (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 to 2100, forced with greenhouse gas scenarios known as RCP 4.5 and RCP 8.5. These Representative Concentration Pathways (RCPs) are commissioned by the 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 watts per square meter 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 watts per square meter 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 to 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 approaches of Soliman et al. (2015) and Lyubchich and Gel (2017) to modeling a number of weather-related claims, with a collective risk model (CRM) simulation algorithm of Meyers, Klinker, and Lalonde (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 (Lyubchich and Gel 2017). 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 versus precipitation and wind speed: Classification and Regression Trees (Breiman et al. 1984; Derrig and Francis 2006; Hadidi 2003) and Alternating Conditional Expectations (ACE) transformations (Breiman and 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 important predictors. (AIC is a model selection criterion that is widely used in numerous actuarial applications; see Brockett 1991; Frees, Derrig, and Meyers 2014; and references therein.)

The key ideas of Classification and Regression Trees (Breiman et al. 1984) are to

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

15080-09_Lyubchich-2ndPgs.indd 281 4/25/19 10:55 AM

- partition the predictor space with binary splits formulated in terms of the covariates, for example, identify thresholds in precipitation amounts that iteratively split the range for this variable into intervals; and
- divide the whole domain into relatively homogeneous 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 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 and Friedman 1985):

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

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

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

ACE transformations are splines with a predefined degree of smoothness, which is a special case of the generalized additive models (see Frees, Derrig, and Meyers 2014 for an overview of generalized additive models 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) and Lyubchich and Gel (2017), 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, Rigby, and Stasinopoulos (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 mean of Y_t, given $Y_1, \ldots, Y_{t-1}, X_1, \ldots, X_t$, as

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

where $g(\cdot)$ is an appropriate link function, β is the vector of regression coefficients, ϕ are coefficients of the autoregressive part of order p, and θ are coefficients of the moving average part of order q. 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, Koopman, and Lucas (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 3.1, along with the modeling results of the number of claims versus the weather predictors in GARMA(p, q) 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 βs, implying higher vulnerability to severe weather, which may be related to the city's location directly in the Great Lakes area.

Remark. Atmospheric thresholds that are currently adopted in civil engineering for building codes and standards primarily target extreme catastrophic events (see, e.g., Crawford and Seidel 2013; Gillespie,

15080-09_Lyubchich-2ndPgs.indd 282 4/25/19 10:55 AM

Table 3.1. Results of attribution analysis and modeling of daily number of weather-related claims using weather predictors (precipitation and wind speed) in 2002–2011

	Thresholds		GARMA(0, q) model parameters		
City	Precipitation, millimeters/day	Wind speed, kilometers/hour	$\hat{\beta}_{pcp}$	$\hat{\beta}_{wind\ speed}$	q
A	5	50	.032 (.004)	.013 (.001)	1
B	1	50	.034 (.004)	.037 (.002)	6
C	1	45	.024 (.002)	.021 (.001)	3
D	5	55	.026 (.002)	.014 (.001)	1

Note: Standard errors are in parentheses.

Antes, and Donnelly 2016; Obama 2015) and do not account for 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 insurers and their clients. The study proposed here allows us to shed light on the critical thresholds and methods to derive them that could be used by the insurance companies to mitigate 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 claim frequencies

The selected thresholds and estimated model parameters further play a role in assessing future trends in the number of insurance claims. We obtain future values of 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 3.1). 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.

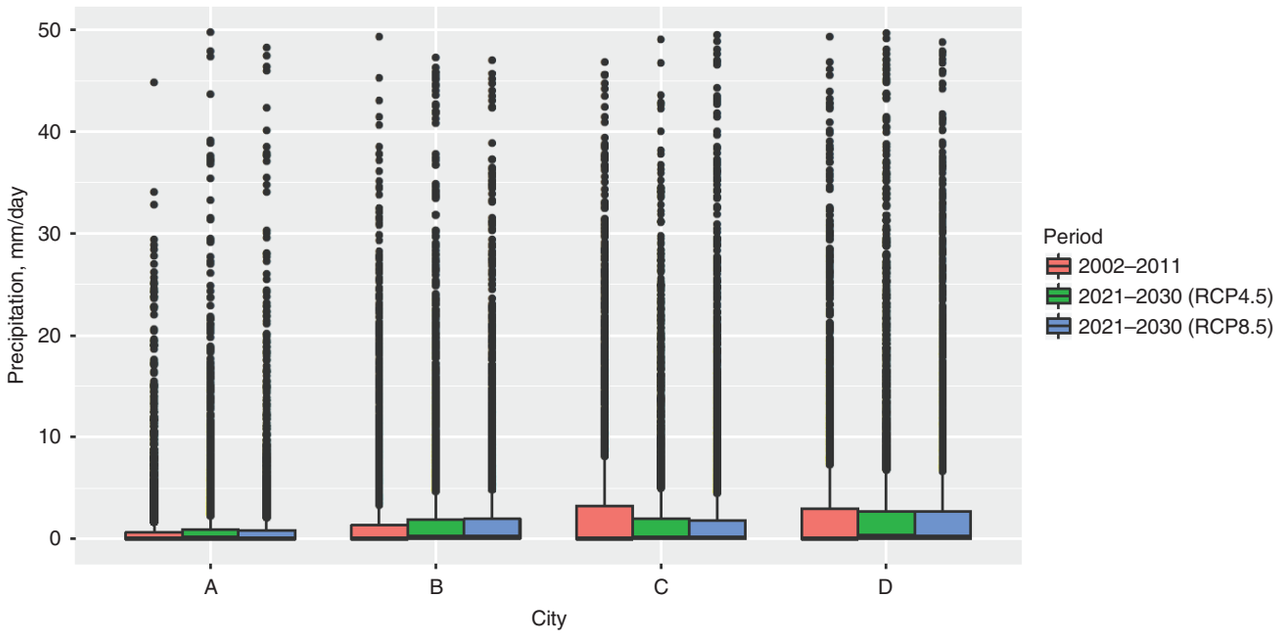
We now apply Algorithm 1 to the claim dynamics in four cities and summarize our findings in Table 3.2. In particular, Table 3.2 suggests that claim counts will likely increase in all four cities during 2021–2030. The major increase is expected for City B, which is directly on a lake, indicating that changes in lake-derived moisture and its representation in the model should be explored in detail to fully understand this result. In general, higher elevation corresponds to a smaller increase in claim counts, but it is uncertain whether this is due to a reduced change in climate with elevation or is 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 due to climate system noise expected between different model runs.

3.3. Future projections with 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 was based on weather observations in 2002–2011, and future projections

15080-09_Lyubchich-2ndPgs.indd 283 4/25/19 10:55 AM

Figure 3.1. Box-plots of daily precipitation less than 50 millimeters per day observed in 2002–2011 and projected for 2021–2030 under RCP 4.5 and RCP 8.5



have been based on comparing future climate model output against the weather observations.

To evaluate the uncertainty due to climate model projections and 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 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 2009, whereas available

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 ($i = 1, \ldots, k$), 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, for example, 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 $j = 1, \ldots, w_i$.
4. Repeat the steps above for each $i = 1, \ldots, k$, and calculate the overall change in the number of claims suggested by the tail dynamics of the weather predictors:

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

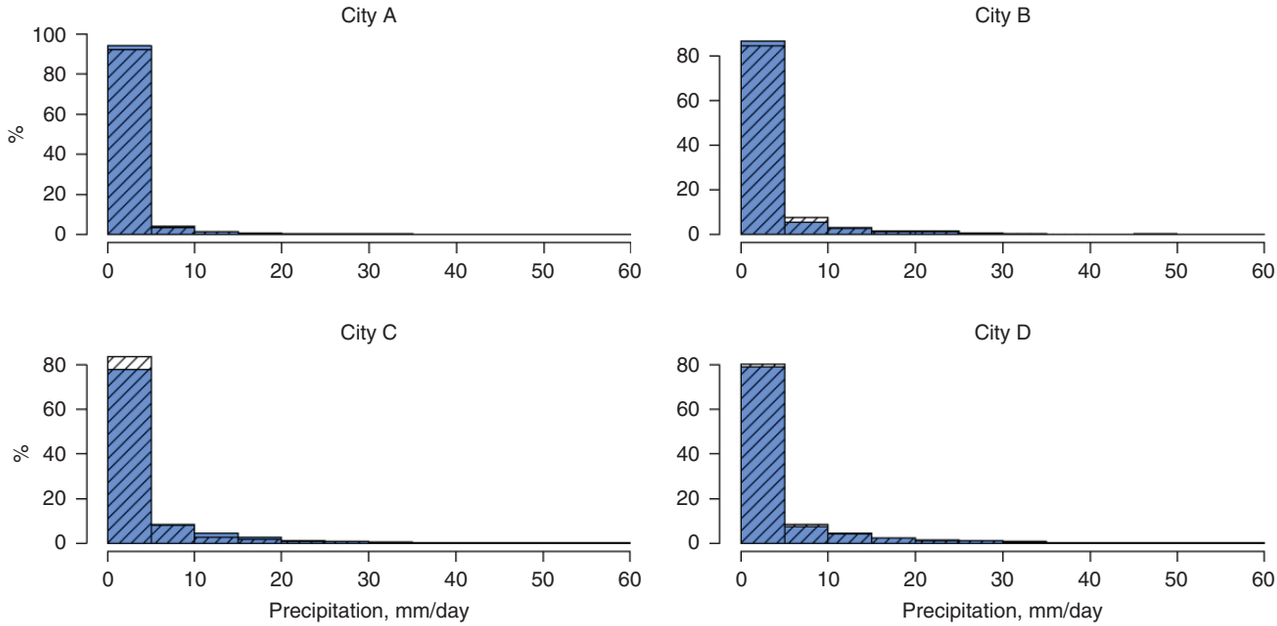
where $G(\cdot)$ is an inverse function of $g(\cdot)$.

Table 3.2. Summary of projected increase (Δ, percentage) in weather-related claim counts in 2021–2030, relative to the baseline level of 2002–2011

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

15080-09_Lyubchich-2ndPgs.indd 284 4/25/19 10:55 AM

Figure 3.2. Frequency distribution of original (blue) daily precipitation and output of the ERA-Interim reanalysis (shaded) for the eight-year period of 2002–2009



insurance data cover 2002–2011; thus, we had to reduce the period of analysis from 10 to 8 years: 2002 to 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 3.2), which is compensated for by higher extremes, such that the centers of the distributions, that is, the first to third quartiles, remain close to each other (Table 3.3).

The projected numbers of weather-related home insurance claims were compared between three possible analysis scenarios summarized in Table 3.4. The scenarios differ based on (1) which weather data source (observations or climate model output) we use for estimating critical thresholds and GARMA modeling and (2) 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), that is, $ncl \sim b_1 \times obs$.

Table 3.3. Summary comparison of daily observed precipitation and output of the ERA-Interim evaluation, for the eight-year period of 2002–2009 (millimeters per day)

City	Data	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
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

15080-09_Lyubchich-2ndPgs.indd 285 4/25/19 10:55 AM

Table 3.4. Projected change in the number of claims (Δ, percentage), depending on the source of weather data and scenario of comparisons. The compared eight-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

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*), that is, $ncl \sim b_2 \times era$. 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*), that is, $ncl \sim b_1 \times obs$. 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 $b_1 \times \delta_2$.

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 CRM approach. This approach entails integrating our GARMA-based projections for claim frequencies with the simulation algorithm of Meyers, Klinker, and Lalonde (2003) to evaluate the distributions of future claim severities. We start from the CRM method of Meyers, Klinker, and Lalonde (2003) adapted for one line of insurance (Algorithm 2).

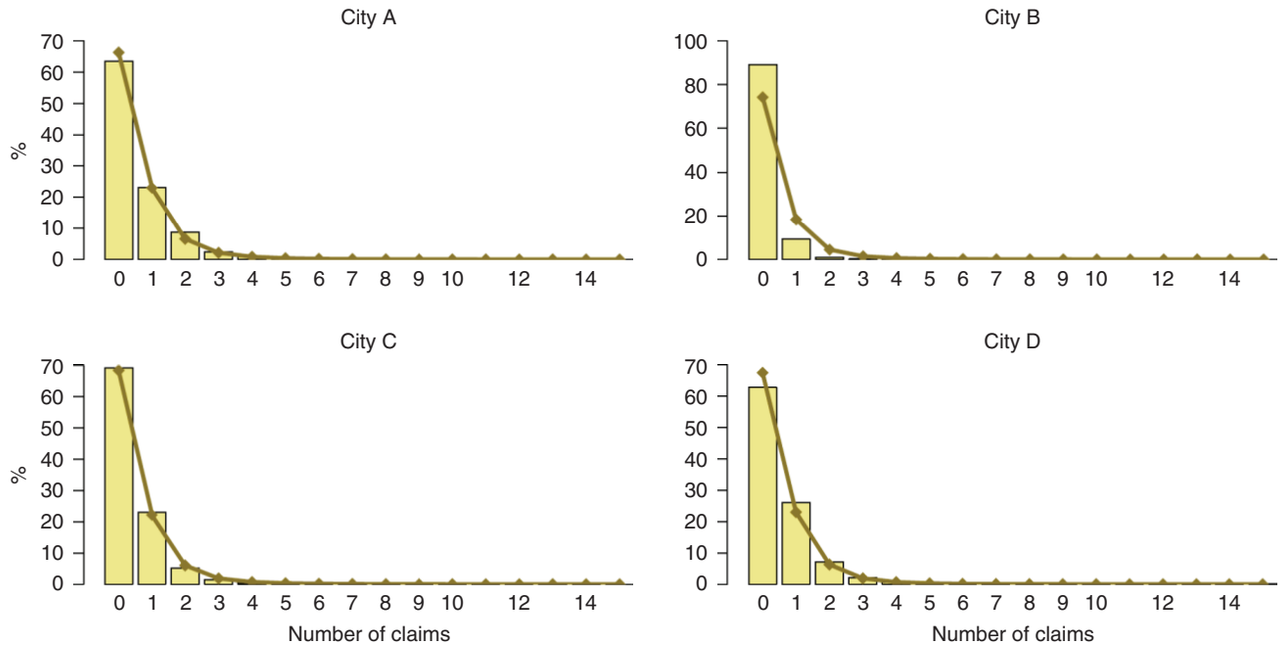
Algorithm 2 can be applied using as Z_k the observed distributions of daily number of claims and claim-by-claim losses. However, to obtain forecasts using Algorithm 2, we need an approximation of the distribution of claim counts so we can adjust it to the future scenarios of claim frequencies. We considered 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, Derrig, and Meyers 2014; Trowbridge 1989) and found that the Sichel distribution (Figure 4.1) 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, Zucchini, and Juritz 1987). The Sichel distribution provides a flexible alternative for modeling highly

Algorithm 2. CRM simulation (a modified version of the CRM method by Meyers, Klinker, and Lalonde 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.

15080-09_Lyubchich-2ndPgs.indd 286 4/25/19 10:55 AM

Figure 4.1. 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



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, e.g., Bermúdez and Karlis 2012; Hera-Martines, Gil-Fana, and Vilar-Zanon 2008; Tzougas and Frangos 2014, and references therein). Table 4.1 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 the 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 City B and City D due to outliers; however, the Sichel distribution still provides the best fit among all other considered distributions.

Table 4.1. Estimated parameters of the Sichel distributions for the normalized number of claims

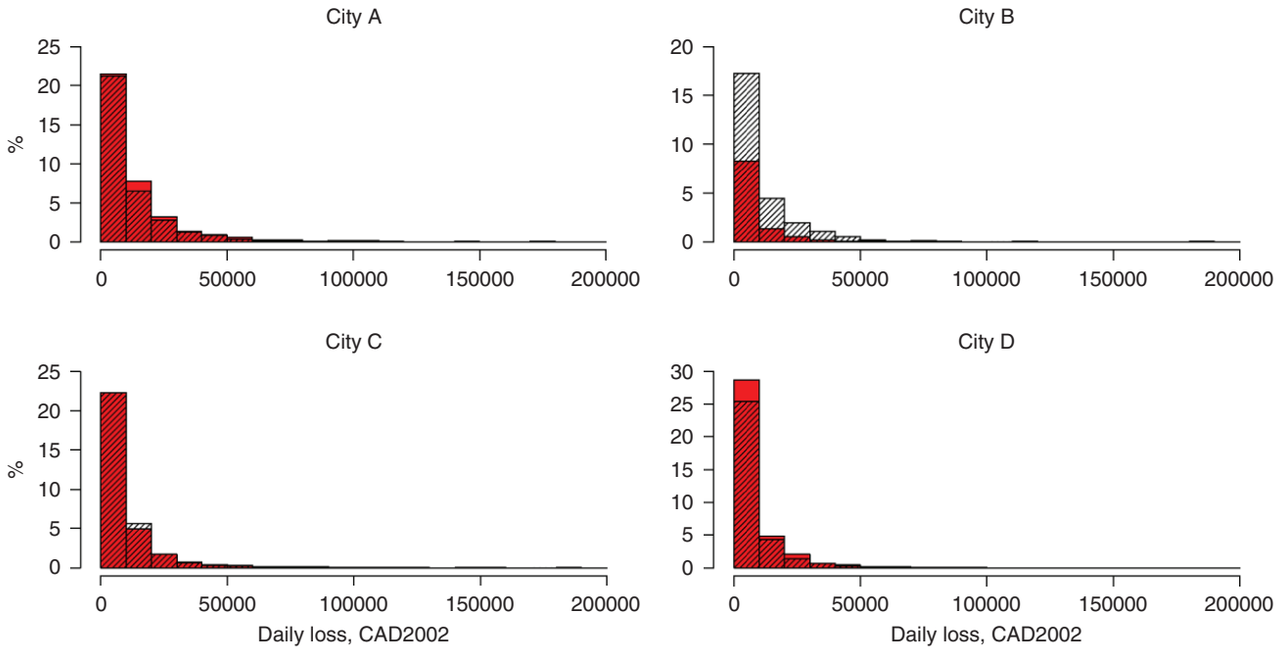
	City			
Parameter	A	B	C	D
μ	4.99	1.82	2.83	2.31
σ	4.85	3.77	3.00	2.10
ν	–1.89	–1.45	–1.88	–2.01

To validate the approach of approximating the number of claims with a Sichel distribution, we first consider the distributions of the observed claim severities versus their approximations using Algorithm 2 (Figure 4.2). The worst fit is observed for City B, which is not surprising given that it has very few (about 11%) days with nonzero losses, compared to more than 30% of days with losses in the other three cities. In the cases with more nonzero 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.

Further, we employ results of Algorithm 1 (Section 3.2) for estimating Δ under comparison Scenario 1 (Section 3.3) for the longest period of data availability for the given scenario (10 years: 2002–2011). Table 3.2 shows the estimated increase in claim frequencies, which implies the respective changes in parameter μ of a Sichel distribution. We now apply CRM Algorithm 2, where simulations

15080-09_Lyubchich-2ndPgs.indd 287 4/25/19 10:55 AM

Figure 4.2. Original (red) and approximated with Algorithm 1 (shaded) distributions of daily losses in the four cities. Y-axis represents the percentage of the total number of days in the period 2002–2011 (days with zero losses are counted toward the total of 100% but are not plotted).



Note: CAD2002 = Canadian dollars adjusted for inflation to the level of 2002.

of the future number of claims in Step 1 are based on the Sichel distribution with parameters $\mu + \Delta$.

Table 4.2 presents the observed and projected percentiles for daily losses under Scenarios RCP 4.5 and RCP 8.5. All four cities exhibit noticeable increases in losses, and the most profound changes are suggested for City B (in both lower and upper percentiles). Such findings can be explained by the greater expected increase in claim frequencies (Table 3.2), the location of City B (directly on the lake), and the 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).

Table 4.2. 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	Climate scenario for 2021–2030 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

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 3.2), 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 noncatastrophic weather events. Our data-driven

15080-09_Lyubchich-2ndPgs.indd 288 4/25/19 10:55 AM

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, Klinker, and Lalonde (2003). We evaluate uncertainty due to potential differences between climate model output and weather observations and evaluate the estimated losses from two 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 proactively hire more insurance assessors and loss adjusters to ensure timely processing of claims. In the case of medium- and long-term risk projections, the results of this study can be used to reevaluate 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 properties 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 noncatastrophic atmospheric events on the insurance sector, particularly home insurance (see the literature review by Lyubchich et al. 2019). Hence, we would like to conclude the paper with a discussion of 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 expected "normal" weather; however, in a changing climate, that length of record may be too long since we expect the weather to change significantly during 30 years (Huang, Dool, and Barnston 1996; Livezey et al. 2007). It has been suggested that temperature and precipitation across 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, Dool, and Barnston 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 five years) to continue to use cutting-edge information about the future climate. Climate models, both regional and global, are constantly being improved. We used the latest available North American regional dynamically downscaled climate data at the time of our analysis, but the global model projection data are updated every five years for the IPCC reports, and new regional model data are available accordingly if funding for the science is continued.

The approach of our study sidesteps 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 often is 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 do not vary (significantly) over time. Anthropogenic climate change, however, causes a relatively rapid

15080-09_Lyubchich-2ndPgs.indd 289 4/25/19 10:55 AM

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 have long been known to induce spatial dependencies among insurance policies, often leading to undercapitalization and excessive risk taking (Cooke and 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 about 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.

Acknowledgments

The authors thank Natalia A. Humphreys, University of Texas at Dallas; Douglas Collins, Casualty Actuarial Society (CAS); Stephen J. Ludwig, CAS; and other members of the Climate Change Committee of CAS for motivating discussions, helpful comments, and suggestions. This work was supported in part by CAS.

References

Benjamin, M. A., R. A. Rigby, and D. M. Stasinopoulos, "Generalized Autoregressive Moving Average Models," *Journal of the American Statistical Association* 98: 461, 2003, pp. 214–223.

Bermúdez, L., and D. Karlis, "A Finite Mixture of Bivariate Poisson Regression Models with an Application to Insurance Ratemaking," *Computational Statistics and Data Analysis* 56:12, 2012, pp. 3988–3999.

Bouvet, L., and P. Kirjanas, *Global Climate 500 Index 2016: Insurance Sector Analysis.* London: AODProject, 2016.

Breiman, L., and J. H. Friedman, "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association* 80:391, 1985, pp. 580–598.

Breiman, L., J. H. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees.* Boca Raton: CRC, 1984.

Brockett, P. L., "Information Theoretic Approach to Actuarial Science: A Unification and Extension of Relevant Theory and Applications," *Transactions of the Society of Actuaries* 43, 1991, pp. 73–135.

Cameron, A. C., and P. K. Trivedi, *Regression Analysis of Count Data,* vol. 53, Cambridge, UK: Cambridge University Press, 2013.

Cheng, C. S., Q. Li, G. Li, and H. Auld, "Climate Change and Heavy Rainfall-Related Water Damage Insurance Claims and Losses in Ontario, Canada," *Journal of Water Resource and Protection* 4, 2012, pp. 49–62.

Cooke, R., and C. Kousky, *Climate Dependencies and Risk Management: Microcorrelations and Tail Dependence,* vols. 10–13, Washington, DC: Resources for the Future, 2010.

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

Creal, D., S. J. Koopman, and A. Lucas, "Generalized Autoregressive Score Models with Applications," *Journal of Applied Econometrics* 28:5, 2013, pp. 777–795.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, "The ERA-interim Reanalysis: Configuration and Performance of the Data Assimilation System," *Quarterly Journal of the Royal Meterological Society* 137:656, 2011, pp. 553–597.

Derrig, R., and L. Francis, *Distinguishing the Forest from the TREES: A Comparison of Tree Based Data Mining Methods,* Casualty Actuarial Society E-Forum, 2006, https://www.casact.org/pubs/forum/06wforum/06w05.pdf.

Fokianos, K., "Statistical Analysis of Count Time Series Models: A GLM Perspective," in *Handbook of Discrete-Valued Time Series,* pp. 3–28, edited by R. A. Davis, S. H. Holan, R. Lund, and N. Ravishanker, Boca Raton, FL: CRC, 2015.

15080-09_Lyubchich-2ndPgs.indd 290 4/25/19 10:55 AM

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

Frees, E. W., G. Meyers, and A. D. Cummings, "Predictive Modeling of Multi-Peril Homeowners Insurance," *Variance* 6, 2012, pp. 11–31.

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

Giorgi, F., and W. J. Gutowski, "Regional Dynamical Downscaling and the CORDEX Initiative," *Annual Review of Environment and Resources* 40, edited by A. Gadgil and T. P. Tomich, 2015, pp. 467–490, Palo Alto, CA: Annual Reviews.

Grossi, P., H. Kunreuther, and D. Windeler, *An Introduction to Catastrophe Models and Insurance Catastrophe Modeling: A New Approach to Managing Risk,* pp. 23–42, Berlin: Springer, 2005.

Hadidi, N., *Classification Ratemaking Using Decision Trees,* Casualty Actuarial Society E-Forum, 2003, https://www.casact.org/pubs/forum/03wforum/03wf253.pdf.

Haug, O., X. K. Dimakos, J. F. Vårdal, M. Aldrin, and E. Meze-Hausken, "Future Building Water Loss Projections Posed by Climate Change," *Scandinavian Actuarial Journal* 2011:1, 2011, pp. 1–20.

Held, H., F.-W. Gerstengarbe, T. Pardowitz, J. G. Pinto, U. Ulbrich, K. Born, M. G. Donat, M. K. Karremann, G. C. Leckebusch, P. Ludwig, K. M. Nissen, H. Osterle, B. F. Prahl, P. C. Werner, D. J. Befort, and O. Burghoff, "Projections of Global Warming-Induced Impacts on Winter Storm Losses in the German Private Household Sector," *Climatic Change* 121:2, 2013, pp. 195–207.

Hera-Martines, A., J. A. Gil-Fana, and J. L. Vilar-Zanon, "Claim Counts Modeling and Stable Distributions," *ASTIN Bulletin* 37:7, 2008, pp. 1–9.

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

Intergovernmental Panel on Climate Change, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects,* Cambridge, UK, and New York: Intergovernmental Panel on Climate Change, 2014a.

Intergovernmental Panel on Climate Change, *Climate Change 2014: Synthesis Report.* Geneva, Switzerland: Intergovernmental Panel on Climate Change, 2014b.

Kunreuther, H., and E. Michel-Kerjan, *Policy Options for Reducing Losses from Natural Disasters: Allocating $75 Billion,* Wharton, Copenhagen Consensus 2012, https://www.copenhagenconsensus.com/sites/default/files/natural disasters.pdf.

Livezey, R. E., K. Y. Vinnikov, M. M. Timofeyeva, R. Tinker, and H. M. van den Dool, "Estimation and Extrapolation of Climate Normals and Climatic Trends," *Journal of Applied Meteorology and Climatology* 46:11, 2007, pp. 1759–1776, doi: 10.1175/2007jamc1666.1.

LSE (London School of Economics and Political Science), *Evaluating the Economics of Climate Risks and Opportunities in the Insurance Sector,* 2015, https://www.cccep.ac.uk/wp-content/uploads/2015/11/14-0964-Munich-Brochure-interactive-FINAL.pdf.

Lyubchich, V. and Y. R. Gel, "Can we Climate Proof our Insurance?" *Environmetrics* 28:2, 2017, pp. 2433.

Lyubchich, V., N. K. Newlands, A. Ghahari, T. Mahdi, and Y. R. Gel, "Insurance Risk Assessment in the Face of Climate Change: Integrating Data Science and Statistics," forthcoming in *WIREs Computational Statistics*, 2019.

Meyers, G. G., F. L. Klinker, and D. Lalonde, *The Aggregation and Correlation of Insurance Exposure,* Casualty Actuarial Society E-Forum, 2003, https://www.casact.org/pubs/forum/03spforum/.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, "Stationarity is Dead: Whither Water Management?" *Science* 319:5863, 2008, pp. 573–574, doi: 10.1126/science.1151915.

NAIC, *Climate Change and Risk Disclosure, 2016,* http://www.naic.org/cipr_topics/topic_climate_risk_disclosure.htm.

Obama, B., 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., and H. E. Willmot, *Insurance Risk Models,* Schaumburg, IL: Society of Acturaries, 1990.

Scheel, I., E. Ferkingstad, A. Frigessi, O. Haug, M. Hinnerichsen, and E. Meze-Hausken, "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, 2013, pp. 85–100.

Scinocca, J. F., V. V. Kharin, Y. Jiao, M. W. Qian, M. Lazare, L. Solheim, G. M. Flato, S. Biner, M. Desgagne, and B. Dugas, "Coordinated Global and Regional Climate Modeling," *Journal of Climate* 29:1, 2016, pp. 17–35, doi: 10.1175/jcli-d-15-0161.1.

Smith, A. B., and J. L. Matthews, "Quantifying Uncertainty and Variable Sensitivity within the US Billion-Dollar Weather and Climate Disaster Cost Estimates," *Natural Hazards* 77:3, 2015, pp. 1829–1851, doi: 10.1007/s11069-015-1678-x.

Soliman, M., V. Lyubchich, Y. R. Gel, D. Naser, and S. Esterby, "Evaluating the Impact of Climate Change on Dynamics of House Insurance Claims," pp. 175–183, *Machine Learning and Data Mining Approaches to Climate Science*, eds. V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, Berlin: Springer, 2015.

Stein, G. Z., W. Zucchini, and J. M. Juritz, "Parameter Estimation for the Sichel Distribution and Its Multivariate Extension," *Journal of the American Statistical Association* 82:399, 1987, pp. 938–944.

15080-09_Lyubchich-2ndPgs.indd 291 4/25/19 10:55 AM

Stewart, M. G., "Cyclone Damage and Temporal Changes to Building Vulnerability and Economic Risks for Residential Construction," *Journal of Wind Engineering and Industrial Aerodynamics* 91:5, 2003, pp. 671–691, doi: 10.1016/S0167-6105(02)00462-2.

Toumi, R., and L. Restell, *Catastrophe Modelling and Climate Change,* London: Lloyd's, 2014, https://www.lloyds.com/news-and-risk-insight/risk-reports/library/natural-environment/catastrophe-modelling-and-climate-change.

Trowbridge, C. L., *Fundamental Concepts of Actuarial Science,* Actuarial Education and Research Fund, 1989, https://www.actuariayfinanzas.net/images/sampledata/Conceptos-fundamentales-de-la-Ciencia-Actuarial.pdf.

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

Willmot, G. E., "Mixed Compound Poisson Distributions," *ASTIN Bulletin* 16:S1, 1986, pp. S59–S79.

Willmot, G. E., "On Recursive Evaluation of Mixed Poisson Probabilities and Related Quantities," *Scandinavian Actuarial Journal* 1993:2, pp. 114–133.

15080-09_Lyubchich-2ndPgs.indd 292 4/25/19 10:55 AM