A Simulation Approach for Estimating Hurricane Risk over a 5-yr Horizon

A. BONAZZI
Risk Management Solutions, London, United Kingdom

A. L. DOBBIN
Risk Management Solutions, and University College of London, London, United Kingdom

J. K. TURNER, P. S. WILSON, C. MITAS, AND E. BELLONE
Risk Management Solutions, London, United Kingdom

(Manuscript received 11 April 2013, in final form 30 September 2013)

ABSTRACT

We develop a stochastic North Atlantic hurricane track model whose climate inputs are Atlantic main development region (MDR) and Indo-Pacific (IP) sea surface temperatures and produce extremely long model simulations for 58 different climates, each one conditioned on 5 yr of observed SSTs from 1950 to 2011—hereafter referred as medium-term (MT) views.

Stringent tests are then performed to prove that MT simulations are better predictors of hurricane landfalls than a long-term view conditioned on the entire SST time series from 1950 to 2011.

In this analysis, the authors extrapolate beyond the historical record, but not in terms of a forecast of future conditions. Rather it is attempted to define—within the limitation of the modeling approach—the magnitude of extreme events that could have materialized in the past at fixed probability thresholds and what is the likelihood of observed landfalls given such estimates.

Finally, a loss proxy is built and the value of the analysis results from a simplified property and casualty insurance perspective is shown. Medium-term simulations of hurricane activity are used to set the strategy of reinsurance coverage purchased by a hypothetical primary insurance, leading to improved solvency margins.
et al. 1998), and as such the risk of insurance losses increases. Owing to the lead time necessary for insurance pricing decisions, however, seasonal forecasts that incorporate the effect of ENSO or the North Atlantic Oscillation (NAO), very few of which show skill before spring or early summer of a given year (Vecchi et al. 2011), are not practical. Business planning (and certain insurance contracts such as catastrophe bonds) along with a market requirement for stable pricing require an estimate of hurricane activity on time scales longer than one season. Characterizations of Atlantic hurricane activity that are based on more slowly changing trends in SSTs are preferable to interannual predictions (Elsner et al. 2008) and will be considered in the work presented here. Interannual changes in tropical SST, if averaged over several years to eliminate high-frequency variability, can be skillfully forecasted using simple statistical models (Laepple and Jewson 2007; Laepple et al. 2007, 2008).

For the purposes of this paper we identify two climate indices: the seasonal area averages of SST in the Atlantic main development region (MDR) and tropical Indo-Pacific (IP), the two of which affect hurricane frequency and life cycle in different ways. Geographical definitions of MDR and IP are given in section 2. The time series of MDR–IP SST data used in this analysis is shown in Fig. 1. MDR SSTs experience multidecadal changes that have been described as the Atlantic multidecadal oscillation (AMO) (Delworth and Mann 2000; Goldenberg et al. 2001) and as the Atlantic meridional mode (AMM) (Vimont and Kossin 2007). In the positive phase of both modes, as has been the case since approximately 1995, there are anomalously warm SSTs. Warm MDR SSTs are associated with a pattern of lower wind shear in the same region. During periods of low shear in the MDR, the environment is favorable for the development and intensification of African easterly waves into hurricanes near the Cape Verde islands and hurricanes become more frequent. However, shifts in genesis during warm SST years (Kossin et al. 2010; Wu et al. 2010) and changes to atmospheric steering currents caused by an increased Atlantic warm pool (Wang et al. 2011) can mean that hurricanes are less likely to make landfall even though they become more frequent. Colbert and Soden (2012) used both historic and simulated tropical cyclone tracks to illustrate and investigate the relative importance of both genesis location within the MDR and large-scale steering flow on the tracks of tropical cyclones. A distinct difference in the average tropical cyclone track between positive and negative phases of the AMM was attributed to a shift in genesis location, associated with warmer SSTs during the positive phase.

Conversely, warm IP temperatures are associated with higher wind shear in the western MDR and Caribbean Sea (Wang and Lee 2008) and inhibited Atlantic cyclone formation through a warmed upper atmosphere (Vecchi and Soden 2007). IP SST-driven wind shear affects the regionalization of genesis (Aiyyer and Thornicroft 2006) by inhibiting hurricane formation (Gray 1968; McBride 1981; McBride and Zehr 1981; DeMaria 1996; DeMaria et al. 2001).

The aim of this work is to evaluate the extent to which hurricane multiannual variability can be modeled by MDR and IP SSTs and to test whether reducing the aleatory uncertainty of MDR–IP SSTs would translate

![Fig. 1. Time series of Atlantic main development region and Indo-Pacific SSTs. (a) Observed yearly and 5-yr running-mean values. (b) The same data are presented as a scatterplot. The Euclidean distance in the MDR–IP SSTs space of 5-yr average values between MT$_{1950-54}$ and MT$_{2007-11}$ is shown as a gray line.](image_url)
into a better representation of the hurricane activity in the Atlantic basin in a stochastic track model. We compare a long-term (LT) simulation of hurricane tracks conditioned by MDR–IP SSTs from 1950 to 2011 with a set of medium-term (MT) simulations conditioned by 5-yr intervals of MDR–IP SSTs. This approach is meant to mimic a real life situation whereby a user interested in the hurricane activity in a certain year, for instance 2011, is provided with two stochastic catalogs: (i) a LT catalog representing the overall climatological distribution of hurricanes over the period 1950–2011 and (ii) a medium-term catalog that describes the hurricane climatological distribution in a 5-yr interval, such as 2007–11. It is of interest to score one against the other.

A natural continuation of this study will consider SST forecasts, but here we will only assess the skill scores of the time-varying models given a perfect knowledge of SST conditions.

The paper is organized as follows: Section 2 briefly describes the various datasets used in the study, while in section 3 we describe the statistical track model used to perform MT and LT simulations. The ability of MT simulations to mimic observed landfalls is discussed in section 4. Finally in section 5, a simple insurance/reinsurance framework is presented to assess the value of MT simulations from a practical business perspective.

2. Data

Historical hurricane tracks in the Atlantic basin were used to build the stochastic model. These tracks were taken from the Atlantic hurricane database (HURDAT) historical catalog (Landsea and Franklin 2013) that is maintained by the National Oceanic and Atmospheric Administration National Hurricane Center (NHC). It can be downloaded online (http://www.nhc.noaa.gov). For U.S. landfalls, we use the additional information in the landfill summary to estimate the maximum 1-min wind speeds when the hurricane center crossed the coastline. HURDAT track data are only used after 1950.

Although the HURDAT catalog extends back to the mid-nineteenth century, it is known to contain missing storms in the open ocean of the eastern Atlantic (Landsea 2007) from the years before airplane observations. Some data issues remain after 1950, for instance, the increasing detection of short-lived storms in the postsatellite era (Landsea et al. 2010).

The extended best-track dataset (Demuth et al. 2006) has been used to supplement information about the radius of maximum wind. This catalog includes all storms from 1988 onward.

The SSTs used in this study are from the Met Office’s Hadley Centre dataset (Rayner et al. 2003), also since 1950. We use monthly averaged values interpolated to a 2.5 × 2.5 latitude–longitude grid. For use in fitting the stochastic model, a seasonal area average over the MDR (10°–20°N, 15°–71°W) and IP (0°–15°N, 40°E–80°W) is calculated, using inclusive data from July to September. This paper frequently refers to the east MDR and west MDR, meaning a simple central division in the MDR at −42.5°W.

Finally the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis (Kalnay et al. 1996) zonal wind is used at the 700- and 300-hPa level in order to calculate wind shear in the analysis of the results. We define wind shear as the climatology of the magnitude of the vertical shear between the between the upper (300 hPa) and lower (700 hPa) troposphere. The climatological means are calculated in 28-day windows from the 1950 to 2007 period.

3. Models

Stochastic models generate more hurricanes than exist in the historical observations or in numerical model outputs, while maintaining the statistics of the observational history. Some early published studies on stochastic models of hurricane wind speeds are by Darling (1991) and Chu and Wang (1998). Later, Vickery et al. (2000) created a model of Atlantic hurricane tracks with an autoregressive incrementation of speed and direction and a simulated error term, a similar technique to the model used in this study. Other notable hurricane models in the published literature are Casson and Coles (2000), James and Mason (2005), Emanuel et al. (2006), and Rumpf et al. (2007).

The models for TC genesis, path, and intensity closely resemble—both in formulation and fitting methods—those of Hall and Jewson (2007), Yonekura and Hall (2011), and Hall and Yonekura (2013). As in Hall and Yonekura (2013), we use environmental predictors to capture seasonal and interannual climate variability.

Genesis is assumed to be a spatial Poisson process with spatially, seasonally, and interannually varying intensity lambda. Maps of monthly averages of SSTs and wind shear are used in combination with yearly averages of MDR and IP SSTs as predictors in a Poisson regression model. The formulation of the genesis component has the same formulation and similar SST predictors to the model used in Hall and Yonekura (2013).

The 6-hourly track steps in the zonal direction are modeled as Gaussian random variables; predictors are MDR and IP SST yearly averages and monthly varying maps of the average zonal wind at 300 hPa from the reanalysis data. As steering winds, defined as the average zonal winds at 300 hPa, do not have as great an
influence on track direction in the meridional direction, the mean of the meridional displacements is calculated based on the historical values of the displacements themselves. Computation of the variance about the mean displacements is based on the assumption that the displacement anomalies in a local region are well described by a bivariate normal distribution, which is, in general, anisotropic and correlated. However, using a covariance ellipse formulation, if the coordinate axes are rotated to match the principal axes of the ellipse, the standardized anomalies in the rotated space (parallel to the major and minor axis of the ellipse) are uncorrelated and can therefore be modeled independently (Yonekura and Hall 2011). As in Hall and Jewson (2007), autocorrelation of displacement anomalies are modeled as weighted spatial averages of all historical values (still working in rotated space).

The overwater 6-hourly changes in central pressure (CP) are modeled by a linear regression on the previous CP value, the total change in CP since genesis, zonal and meridional track increments, maps of monthly averages of SSTs, and yearly averages of MDR–IP SSTs. The model coefficients vary in space as for the genesis and track path components. When a storm makes landfall, pressure is assumed to fill exponentially. The filling rate varies by region and depends on storm (and landfall) specific parameters (see Colette et al. 2010)—is used to sample the time at which transitioning initiated. Linear interpolation is then used to define the intermediate status of the transitioning amount.

A truncated lognormal model is used to simulate radii of maximum wind ($R_{max}$) with predictors CP and latitude. Truncation is applied to remove values of $R_{max}$ beyond the range of observed data. Parameters are estimated for tropical and extratropical conditions separately, and a combination of the two models is used for the transitioning phase. Finally, maximum wind velocity ($V_{max}$) is modeled as a function of central pressure and latitude. Separate models are formulated for tropical and extratropical conditions.

A summary of track model components is given in Table 1.

<table>
<thead>
<tr>
<th>Stochastic track parameter (function or variable)</th>
<th>Functional form of distribution</th>
<th>Data source</th>
<th>Year range used</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genesis</td>
<td>Poisson</td>
<td>HURDAT</td>
<td>1950–2011</td>
<td>Predictors are maps of monthly averages of SST and wind shear and annual averages of MDR and IP SSTs.</td>
</tr>
<tr>
<td>Translational speed and heading</td>
<td>Gaussian</td>
<td>HURDAT</td>
<td>1950–2007</td>
<td>Predictors are zonal and meridional track steps and IP and MDR SSTs; see Hall and Jewson (2007).</td>
</tr>
<tr>
<td>Central pressure over water</td>
<td>Gaussian</td>
<td>HURDAT</td>
<td>1900–2008</td>
<td>Predictors are central pressure at previous time step, total pressure drop from genesis, meridional and zonal track steps, and IP and MDR SSTs.</td>
</tr>
<tr>
<td>Inland filling rate</td>
<td>Gaussian</td>
<td>HURDAT NHC NWP Simulated</td>
<td>1988–2008</td>
<td>The methods used to estimate, select, and validate the model are described in Colette et al. (2010).</td>
</tr>
<tr>
<td>$V_{max}$</td>
<td>Lognormal</td>
<td>HURDAT</td>
<td>1900–2008</td>
<td>Predictors are central pressure and latitude (Knaff and Zehr 2007). Parameters are estimated separately for tropical and extratropical storms.</td>
</tr>
<tr>
<td>$R_{max}$</td>
<td>Lognormal (truncated)</td>
<td>Extended best track</td>
<td>1988–2008</td>
<td>Parameters are estimated separately for tropical and extratropical storms.</td>
</tr>
</tbody>
</table>

4. Results

We simulate tropical storm tracks using a two-step procedure. First, we obtain samples of MDR–IP SSTs from the histogram of observed values. Then the MDR–IP SST samples are used as predictors for the chain of model components—genesis, track path, and intensity—that form the track model described in section 3.

Let us consider the time series of observed MDR–IP values for the period 1950–2011. We turn the 62 data pairs into an empirical distribution from which we obtain 100 000 MDR–IP samples—on average the same MDR–IP pair will appear 100 000/62 times. Each of the MDR–IP samples is then used to generate a year’s worth of TC events. The collection of all simulated tropical storm tracks forms a stochastic catalog that is assumed to represent the climatological distribution of TC events in the period 1950–2011. It is useful to note here that
these very long simulations are not relevant to either future or past conditions on the 100 000-yr time scale. In fact, this simulation strategy can be viewed as a way to model the fractional quantiles of the empirical distribution of tropical storm tracks. Given a pair of MDR–IP SST values, we obtain samples of tropical storms that represent a collection of events that are likely to happen in a year whose climate is defined by the MDR–IP values. Owing to the stochastic nature of models used in this work, any two tracks sampled from identical MDR–IP conditions would differ and a large number of samples is required to properly explore the probability space defined by a long chain of regression models—and their error terms—that describe the dynamical evolution of TC events in the Atlantic.

This work focuses on the comparison between two kinds of stochastic catalogs of hurricane tracks: (i) a LT catalog obtained from a simulation whose climate conditions are sampled from the observed MDR–IP SSTs in the period 1950–2011 and (ii) a set of 58 MT catalogs representing 5-yr (overlapping) subsets of LT, for example, 1950–54, 1951–55, ..., 2007–11. MT simulations differ from LT only with respect to the start and end date of the SST time series used to sample pairs of MDR–IP SST values. Alternatively, the LT catalog can be viewed as being composed of 100 000 years sampled randomly from the chained MT simulation of total length of 100 000 × 58 yr.

a. Tropical storms in the Atlantic basin

The genesis component of the model describes the overall frequency of tropical storm occurrences in the Atlantic basin and the spatial locations of their origins. Figure 2 shows a comparison between the counts of events in the Atlantic basin between HURDAT and MT model simulations. Observed counts shown include all storms classified as tropical storm, hurricane, or subtropical storm in the HURDAT record. MDR and IP SSTs are the only drivers of the interannual variability resolved by the Poisson process model and are sufficient to capture the major trends in the observed TC frequency. An initial decrease in the number of Atlantic basin events is observed throughout the 1950s, and then a period a relative calm follows until the end of the 1980s when an upward trend is observed. It is important to note that the historical record of Atlantic basin events is considered to be incomplete particularly in the pre-satellite era when the detection of tropical cyclones was dependent on whether a ship was passing close to the storm at that time. Several estimates of the number of tropical storms missing from the record have been made by, for example, Chang and Guo (2007), Landsea (2007), and Vecchi and Knutson (2008), who used shipping records to estimated that up to one tropical cyclone is missing per year from the HURDAT record in the 1950s and early 1960s. Improvements in observing practices over time have led to an increase in the count of short-duration storms that were previously missed from or wrongly classified in the record (Landsea et al. 2010; Villarini et al. 2011). The high bias of the MT model in comparison to HURDAT in the 1950s and 1960s is likely a result of the incompleteness of the HURDAT record in this era. The very strong upward trend in observed tropical cyclones from the late 1990s onward is related to the increased detection of short-duration storms. This is most likely the reason for the tendency for the modeled trend to sit below that of the observations in this period.

Regionalization patterns in the genesis locations vary sharply between cold (MDR SST below 1950–2011 average) and warm (MDR SST above 1950–2011 average) years. Genesis density maps for MT1990–94, MT2003–07, and the control run are shown in Fig. 3. The development of storms in the Gulf of Mexico is less dependent on local sea surface temperatures—the Gulf is always extremely warm during hurricane season—and more dependent on large-scale atmospheric conditions (Bracken and Bosart 2000; Hulme and Martin 2009). During warm conditions (MT03–07), genesis is most notably increased in the eastern MDR, with additional contributions from the western Gulf of Mexico and the eastern subtropical Atlantic.
Although hurricane tracks depend on highly variable synoptic conditions, there is an influence from tropical SSTs on the atmosphere. This is true for the MDR, where patterns like the AMM are associated with both atmospheric circulation and SST anomalies. The tropical Pacific, through modulations of the Walker circulation, also influences Atlantic weather patterns and can affect hurricane development. The mean of the distribution from which incremental track displacements are simulated in the model depends on both the MDR and IP SSTs.

In fact, warm- and cold-year MT simulations differ in terms of their track density maps (not shown here). High density in the MDR is achieved during warm conditions, whereas during cold years the majority of tracks are confined in the western portion of the Atlantic basin. Inside the MDR, we observe a modulation between warm years of different intensity. During extreme SST conditions, as observed after 2005, we observed an increase in genesis to the eastern part of the MDR. Storms having genesis in the eastern MDR are historically more likely to recurve before encountering the American continent, and so this shift in density away from land and toward the open ocean is expected from the shifts in genesis (Kossin et al. 2010). As a consequence, although there is more genesis in the eastern Atlantic in very warm conditions, these storms are somewhat less likely to make landfall. Figure 4 shows the time series of modeled hurricane landfall proportions in overlapping 5-yr bins, that is, the percentage ratio of category (CAT) 1–5 storms that make landfall along the U.S. coastline over the number of category 1–5 storms in the Atlantic basin. There is a clear decrease in the simulated proportion of hurricanes making landfall from 1950 to present in response to changing SSTs. While a direct comparison with the observed proportion of hurricanes making landfall is complicated by the question of how many hurricanes are missing from the HURDAT record (e.g., Vecchi and Knutson 2011), this simulated decrease in landfall proportion does agree with the observed decrease.

b. U.S. landfalls

In the MDR, we observe a modulation between warm years of different intensity. During extreme SST conditions, as observed after 2005, we observed an increase in genesis to the eastern part of the MDR. Storms having genesis in the eastern MDR are historically more likely to recurve before encountering the American continent, and so this shift in density away from land and toward the open ocean is expected from the shifts in genesis (Kossin et al. 2010). As a consequence, although there is more genesis in the eastern Atlantic in very warm conditions, these storms are somewhat less likely to make landfall. Figure 4 shows the time series of modeled hurricane landfall proportions in overlapping 5-yr bins, that is, the percentage ratio of category (CAT) 1–5 storms that make landfall along the U.S. coastline over the number of category 1–5 storms in the Atlantic basin. There is a clear decrease in the simulated proportion of hurricanes making landfall from 1950 to present in response to changing SSTs. While a direct comparison with the observed proportion of hurricanes making landfall is complicated by the question of how many hurricanes are missing from the HURDAT record (e.g., Vecchi and Knutson 2011), this simulated decrease in landfall proportion does agree with the observed decrease.
with the work of Wang et al. (2011), who showed that the possibility of a hurricane making landfall is decreased when an eastward expansion of the Atlantic warm pool causes an eastward shift in hurricane genesis location and induces a northeast steering flow that steers hurricanes away from the U.S. coastline.

Given that economic losses—and fatalities—are most likely to occur in the case of a landfalling hurricane, hereafter we will focus on the landfall frequency and severity. A landfall is defined as the crossing of a “gate”—a latitude, longitude line—by a track segment. Given that track data are available—and simulated—at 6-hourly intervals, landfall intensity is taken as $V_{\text{max}}$ at the closest track point before landfall to avoid any contamination of the intensity measure due to wind–land interactions.

The U.S. coastline is divided into gates of uniform length, roughly 100 km, from just below the Mexico–Texas border to just above the Maine–Canada border. Albeit gates do not exactly follow the coastline, we consider it extremely likely that any track crossing a gate will eventually make landfall into the real U.S. coastline. As a result, at every gate we obtain 58 different catalogs of landfall events (one for every MT model), plus the control run. Each landfall event is described by the track parameters—$V_{\text{max}}$, $R_{\text{max}}$, and the extratropical transition (ET) flag—and its rate. Storms are classified by their $V_{\text{max}}$ at the 6-hourly point before landfall. This removes the need to make assumptions about the development of each storm as it approaches land. The very long simulations performed in this study allow us to model the probability distribution of landfalls at fine intensity/frequency thresholds. However, for validation purposes, it is most convenient to focus on a limited number of intensity thresholds using the Saffir–Simpson scale.

Fitting procedures used to estimate the parameters of the track model components described in section 3 ensure that mean squared errors are minimized (or likelihood maximized), but chaining components together would likely result in biased estimates of combined metrics, such as landfall counts by category. The stochastic track model used in this work is able to replicate the overall spatial distribution of landfalls across the domain. The HURDAT average landfall rates across the model domain for the period 1950–2011 are shown in Fig. 5 together with the average rates inferred from the LT catalog. For any practical purposes, it is most convenient to focus on a limited number of intensity thresholds using the Saffir–Simpson scale.

Extreme value statistics could also be used to define multivariate target distributions as described in Bonazzi et al. (2011). Alternatively, methods such as those used in Cook (2011) and Serinaldi (2009) can be employed to define nonparametric target distributions. An optimization procedure would then be devised to adjust the stochastic catalog accordingly to predefined targets. As an in-depth discussion of calibration methodologies and results is beyond the scope of this work, in the following we limit our attention to the relative scores of MT catalogs versus their overall mean, the LT catalog. In section 5, we show how the relativity between MT and LT models could be used in the context of a realistic, albeit simplified, insurance application.

Observed and modeled anomalies of the U.S. landfall rates for overlapping 5-yr intervals are shown in Fig. 6. The correlation coefficients between observed and modeled rates of U.S. landfalls is at best 0.45 (CAT2–5), always lower than that we found for the correlation of genesis counts in the entire basin. For instance, MT simulations fail to mimic the spike in the CAT 1–5 number observed between 1980 and 1985 and the spike of CAT 3–5 events in the early 1960s. To successfully hindcast interannual variability, one would need a detailed knowledge of the specific genesis locations and steering flow variability in each specific season. The reduced variability in the MT simulations in comparison to the historic landfall data is in large part due to the omission of this very specific information.
For a given gate $g$, year interval $y$, and category $c$, we score MT models with respect to our LT view using the following metrics:

$$LD_{g,y,c} = \ln(p(n_{g,y,c}^{\text{MT}} | \lambda_{g,c}^{\text{MT}})) - \ln(p(n_{g,y,c}^{\text{LT}} | \lambda_{g,c}^{\text{LT}})).$$

(1)

The likelihood difference (LD) is positive when the MT models outperform the LT view. We assume that landfall occurrences follow a Poisson distribution, such that $p(n, \lambda) = \Pr[N = n, \lambda] = e^{-\lambda} \lambda^n / n!$. The skill score is computed for all combinations of gates ($g$) and 5-yr periods ($y$) and is subsequently summed up to several aggregation levels. Likelihood differences are heavily influenced by observations that fall in the tail of the modeled probability. As a result, LDs have similar magnitudes across all intensity thresholds.

We use numerical simulation to estimate $p$ values corresponding to LD realizations under the null hypothesis that HURDAT landfalls are samples from the MT distributions. We construct 1000 artificial time series randomly sampling years from individual MT runs and then aggregate the results in 5-yr averages. By chance the LT model will score more favorably then MTs for some of the synthetic time series. Positive LD scores are also observed when considering all CAT 3–5 landfalls. Albeit we see some negative terms in the split LD results, especially for the Gulf region and the inactive years, the associated $p$ value is rather large, suggesting that equal or worse results can easily be obtained by chance sampling from the MT distributions.

At the very end of the intensity spectrum, MT simulations score negatively and associated $p$ values are between 1.1% and 5.4%, suggesting that we could reject the null hypothesis. The three category 5 events that made U.S. landfall since 1950 happened during inactive
hurricanes when MT landfall rates are lower than the LT view. Note that we include in this count Hurricane Beulah that struck just below the United States–Mexico border in 1967. Conversely, there were no CAT 5 landfalls in active years when MT landfall rates were higher than LT’s. We note that underestimation of the probability CAT5 landfalls during inactive years proves to be a bigger contributor to the negative LD score than overestimation during active years.

5. A primary insurance perspective on model results

In this section, we use a simple loss proxy to assess whether a hypothetical PC insurer would benefit from using MT simulations over LT simulations of hurricane activity to plan its reinsurance coverage requirements.

The Carvill hurricane index (CHI) is a regression model used to mimic insured property losses due to tropical cyclones (Kantha 2006). It is parameterized by the physical properties of hurricanes that are available from the National Hurricane Center and that are also resolved by our track model. The index is calculated as

$$I_{CHI} = \left( \frac{v}{74} \right)^2 + \frac{3}{2} \left( \frac{R}{60} \right) \left( \frac{v}{74} \right)^2,$$

where $v$ is the maximum 1-min sustained wind speed measured in miles per hour, and $R$ is the radius to hurricane winds measured in miles. The Willoughby et al. FIG. 6. HURDAT and modeled anomalies of the U.S. landfall rates for CAT 1–5, CAT 2–5, CAT 3–5, and CAT 4–5 with respect to the HURDAT 1950–2011 average and LT catalog, respectively.
For the purposes of this analysis, we consider the CHI as a currency of risk. In catastrophe modeling, “hazard values” that represent the strength and size of a tropical cyclone are converted into damage estimations by “vulnerability functions.” Albeit the definition of these functions is a complex process that accounts for the multitude of building practices and construction types, it is fair to describe the vulnerability functions as monotonically increasing with hazard intensity and generally exponential. These two properties are clearly satisfied by the CHI definition.

In the previous section, we showed that the MT simulations are better predictors of hurricane landfall rates than our long-term view when all years and intensities are considered. However, we noted that extreme events have occurred during inactive periods resulting in negative skill scores for the MT simulations for CAT 4–5 (see Table 2). The revenue of a hypothetical primary PC insurer is used here to investigate whether trusting MT simulations would eventually translate into an economic benefit and hence summarize its value.

We apply similar simplifications as in Emanuel et al. (2012). We assume that the only losses incurred by the primary insurance are due to hurricanes and that the only revenue stream is the premium leveraged on the policyholders. We ignore inflation and assume that profits are not invested.

The primary insurer is able to buy stop-loss2 reinsurance coverage at a price—the reinsurance premium—that is based on the LT model. The reinsurance premium

---

1 The expected frequency above threshold $T$ is the sum of the rates of landfall events that have CHI greater than $T$. Stochastic events are assumed to have rates equal to $1/N$, where $N$ is the number of simulated years.

2 A stop-loss contract is a form of nonproportional reinsurance contract that can be used to put a cap on the annual aggregate loss. Under nonproportional reinsurance the reinsurer only pays out when the total claims suffered by the insurer in a given period exceed an amount called the “retention.”

---

Table 2. Log LD for U.S. landfalls. Western Florida coastline is assigned to Gulf and Atlantic regions include eastern Florida up to the northeast United States. Inactive years are between 1965 and 1994. All other years are considered active. LDs are summed in 5 different year groups of nonoverlapping simulations to preserve independence. Group 1 is formed by years in sequence 1950, 1955, . . . , 2005; group 2 by 1951, 1956, . . . , 2006; and similarly for groups 3, 4, and 5. Minimum–maximum LD values (of the 5 LD groups) are used here to summarize results.

<table>
<thead>
<tr>
<th>CAT 1–5</th>
<th>CAT 2–5</th>
<th>CAT 3–5</th>
<th>CAT 4–5</th>
<th>CAT 5–5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LD (p value)</td>
<td>LD (p value)</td>
<td>LD (p value)</td>
<td>LD (p value)</td>
</tr>
<tr>
<td>Total</td>
<td>Min 0.41 (19.2%)</td>
<td>1.34 (37.44%)</td>
<td>0.12 (25.5%)</td>
<td>−1.53 (2.21%)</td>
</tr>
<tr>
<td></td>
<td>Max 4.12 (82.8%)</td>
<td>3.83 (91.99%)</td>
<td>2.16 (86.5%)</td>
<td>0.34 (48.65%)</td>
</tr>
<tr>
<td>Gulf region</td>
<td>Min −1.58 (4.3%)</td>
<td>−0.35 (16.12%)</td>
<td>−0.56 (14.4%)</td>
<td>−0.86 (6.22%)</td>
</tr>
<tr>
<td></td>
<td>Max 1.57 (58.3%)</td>
<td>2.12 (88.79%)</td>
<td>1.5 (84.4%)</td>
<td>0.65 (72.22%)</td>
</tr>
<tr>
<td>Atlantic region</td>
<td>Min 1.23 (71.2%)</td>
<td>1.03 (75.98%)</td>
<td>0.66 (61.6%)</td>
<td>−0.9 (2.01%)</td>
</tr>
<tr>
<td></td>
<td>Max 2.55 (92.7%)</td>
<td>2.67 (97.2%)</td>
<td>1.17 (86.8%)</td>
<td>−0.31 (17.65%)</td>
</tr>
<tr>
<td>Inactive years</td>
<td>Min −0.29 (21.5%)</td>
<td>−0.07 (27.2%)</td>
<td>−1.12 (7.4%)</td>
<td>−1.91 (0.5%)</td>
</tr>
<tr>
<td></td>
<td>Max 2.21 (78.8%)</td>
<td>0.74 (48.4%)</td>
<td>−0.78 (10.3%)</td>
<td>−0.65 (10.32%)</td>
</tr>
<tr>
<td>Active years</td>
<td>Min −0.2 (18.6%)</td>
<td>1.41 (48.95%)</td>
<td>1.24 (58.5%)</td>
<td>−0.34 (25.85%)</td>
</tr>
<tr>
<td></td>
<td>Max 2.35 (85.5%)</td>
<td>3.09 (97.6%)</td>
<td>3.18 (99.6%)</td>
<td>1.09 (92.09%)</td>
</tr>
</tbody>
</table>

(2006) wind profile is used to obtain $R$ from the available track parameters; first, a full wind profile is computed at each landfall, and then the radius to hurricane winds is extracted from the wind snapshot.

Figure 7a shows the expected frequency1 (EF) of CHI at landfall for the entire U.S. coastline based on the LT simulation, considering category 1–5 hurricanes. The 58 MT simulations are used to compute the mean and 5–95 percentiles of CHI that are exceeded at discrete frequencies ranging from 1 to $10^{-3}$ yr$^{-1}$. The envelope of MT curves can be interpreted as the uncertainty of losses due to climate variability—as resolved by our modeling framework.

CHI values for old historical storms are somewhat difficult to estimate. Storm radius measurements are available in the extended best-track dataset for storms from 1988 onward, and accuracy in the detection of maximum sustained wind has increased enormously since the satellite era began (Landsea and Franklin 2013). The track model has been used to fill missing $R_{max}$ values with expected values from a linear regression on central pressure and latitude for storms between 1950 and 1988. An alternative approach would be to use the observational $R_{max}$ dataset of Ho et al. (1987). CHI historical reconstructions of storms before 1988 are then assumed to be of lower quality than more recent events. Quantitative tests on the quality of MT simulations are performed using the higher quality tier of historical reconstructions from 1988 onward; the full set of historical events is only used for the qualitative comparison between stochastic and historical exceedance frequencies shown in Fig. 7a.
FIG. 7. (a) Event EF for LT model (mean), envelope of MT simulations (5th and 95th quantiles) and historical events between 1950 and 2011. Most intense event in historical record is Camille 1969. (b) The reinsurance premium for a stop-loss annual contract as a function of retention. (c) Mean surplus between 1988 and 2011: at zero retention, mean profit is zero since the same price function is used to price insurance and reinsurance. (d) Standard deviation of surplus for the same period. Dashed lines in (b), (c), and (d) highlight price, mean surplus, and volatility when coverage is set to 20.5 [CHI]. (e), (f) Relative increase/decrease of surplus mean and standard deviation are shown as a function of $\delta C$. Quantiles of distribution of random samples are shown as gray lines.
is defined here to be a function of the mean $E$ and standard deviation $\text{var}^{1/2}$ of aggregate yearly loss $L$ above a retention level $C$:

\[ P(C) = \alpha E(L - C \mid L > C) + \beta \text{var}(L - C \mid L > C)^{1/2}, \]

where the parameters have been arbitrarily set to $\alpha = 1.2$ and $\beta = 0.25$ (see Fig. 7b). For simplicity’s sake, the insurance premium is set to $P(0)$. We note that this setting leads to an unrealistic condition whereby the insurance can pass all claims to the reinsurance without making a loss. On the other hand, doing so we remove one layer of complexity: the definition of the minimum retention for which insurance is profitable. It is also important to point out that albeit our catalogs are formed by a collection of individual events, it is straightforward to produce aggregate estimates of year losses using timestamps associated with stochastic—and historical—events.

The primary’s surplus (i.e., profit) for year $y$ is given by the difference between earnings and costs:

\[ U_y = P(0) - \min \left( \sum_{i=1}^{m_y} L_{y,i} C \right) - P(C), \quad (3) \]

where $P(0)$ is the insurance premium—the only positive term in Eq. (3). The value $C$ is the retention level, and $P(C)$ is the reinsurance premium for retention $C$. The summation inside the minimum function is taken over the $m_y$ loss events in year $y$.

Using historical CHI reconstructions from 1988 to 2011, we compute sample mean and standard deviation of net revenues achieved by the primary with reinsurance coverage between 0 and 30 CHI (see Figs. 7c,d). We see that in buying reinsurance, the primary accepts to reduce its profit to contain volatility. While reducing volatility is favorable to maintain a stable functioning business, increasing profit is also desirable. A reasonable balance between the two must therefore be sought. We assume here that the primary—without any knowledge of MT simulations—would set its retention level to 20.50 CHI, a value equivalent to an aggregate year loss exceeded every 20 yr on average, hereafter RP20. Any losses above this retention level will be paid by the reinsurer. The corresponding reinsurance premium is 1.08 CHI. This simple framework is chosen to avoid specifying any arbitrary decision on the insurance capitalization and solvency requirements. Insurance companies must retain a certain amount of capital in order to reduce the risk of becoming insolvent. We are assuming that the primary has enough initial capital to meet solvency requirements under this reinsurance program.

The insurance company may adjust its retention level, for year $y$, according to this rule:

\[ C_y = \hat{C} - \delta C, \quad \text{if} \quad L_y^{100} > q_{L_y}^{100}(0.6) \]
\[ C_y = \hat{C} + \delta C, \quad \text{if} \quad L_y^{100} < q_{L_y}^{100}(0.4) \]
\[ C_y = \hat{C}, \quad \text{otherwise}. \quad (4) \]

The term $L_y^{100}$ is the aggregate year loss expected to be exceeded once in 100 yr according to MT simulations. A distribution of $L_y^{100}$ is obtained from MT simulations, and quantiles $q_{L_y}^{100}(x)$ are readily estimated. In years when $L_y^{100}$ exceeds $q_{L_y}^{100}(0.6)$, more coverage is bought by the primary, hence reducing its retention level. The opposite choice is taken when $L_y^{100}$ is lower than $q_{L_y}^{100}(0.4)$. The insurance is assumed to use the most up-to-date 5-yr view of hurricane landfalls to set the retention level for year $y$. Here, we are interested in mean and standard deviation of profits as a function of $\delta C$.

The mean surplus between 1988 and 2011 for 100 different $\delta C$ between 0 in 10 CHI is shown in Fig. 7e. To benchmark MT results—and isolate the added value of our simulations from baseline effects due to the strategy framework and pricing assumptions—we consider 1000 synthetic surplus time series from a randomization of the strategy described in Eq. (4), where the reinsurance premium is set according to samples from $U(0, 1)$ being higher (or lower) than 0.6 (0.4). Using MT simulations to inform the reinsurance program results in an increased annual surplus for the primary. For $\delta C$ between 1 and 9 CHI, the increase is between the 75th and 85th quantiles of the surplus distribution obtained with randomized strategies.

The interdecadal oscillation in the IP and MDR SSTs is translated by the track model into time-varying landfall probabilities that capture a meaningful portion of the observed variability of the loss index, sufficient to produce a positive marginal revenue when MT simulations are considered.

The overall surplus is affected by some erroneous choices implied by the MT simulations. During quiet hurricane seasons after 2008, our strategy is suggesting to increase the retention level at the cost of a higher reinsurance premium. On the other hand, the same strategy proves successful during the extremely active 2004 and 2005 seasons, lowering the incurred loss.

The surplus volatility also decreases with $\delta C$ when random strategies are considered. In Fig. 7f, we see that the sample standard deviation of the cash flow increases

---

3 In the case of a stop-loss contract, the retention level $C$ is the maximum amount of total claims suffered by the primary insurer in a given period. If the total loss is bigger than retention, the reinsurer will pay out the difference $L - C$.

4 For years between 2007 and 2011, we contrast historical events with simulation MT07–11.
less steeply for retentions above RP20 than below, and it flattens out above the maximum aggregate loss observed in the period 1988–2011. For this reason, even when positive and negative $\delta C$ are equally probable, the net effect would be to lower the surplus volatility. This is a positive result for business stability.

When MT simulations are used to inform the reinsurance coverage strategy, the volatility decreases with $\delta C$ faster than in most random simulations; MT results fall between the 5th and 10th percentile of the distribution of volatility obtained with random strategies.

We already assumed that the insurance is solvent with respect to the baseline retention level; that is, the primary insurer retains enough capital to pay out claims up to that level. We now see that the adoption of MT simulations to inform the reinsurance strategy results in an increased surplus and decreased volatility, therefore improving both the profitability and stability of the business and its solvency prospects.

6. Conclusions

In this analysis, we quantify the risk of hurricane landfalls in the United States in varying climate conditions. A set of 58 simulations, conditioned on 5-yr bins of measured IP and MDR SSTs, form our medium-term views of hurricane activity in the Atlantic basin from 1950 to 2011. A likelihood-based metric has been applied to score MT simulations relative to a long-term reference view that is conditioned on the entire SST time series from 1950 to 2011. A clear positive signal emerges when all landfall intensities and regions are considered suggesting that the medium-term simulations are better able to predict landfalls than the long-term view. We note however that the timing of CAT 5 landfalls is not captured by our simulations. The observed CAT 5 landfalls during the last 60 yr occurred in relatively cold SST conditions when MT estimates are lower than the reference view. The work presented here assumes a perfect knowledge of MDR and IP SST conditions. The ability to reproduce the result herein when using predicted SSTs would depend on the accuracy of the SST predictions. The impact of using hindcast SSTs instead of historic data will be the subject of future investigations.

Finally a simple—but fully replicable—experiment was implemented to assess whether model results would score positively in a simplified insurance application. The results are encouraging as they suggest that medium-term catastrophe models could help a hypothetical user improve his financial resiliency in varying climate conditions.

The quantification of a climate signal in damage losses has been the object of several studies that do not rely on stochastic catalogs. For example, Jagger et al. (2011) applied a Bayesian framework to estimate generalized Pareto distributions of United States–wide damage loss that are conditioned on different climate scenarios. Although these methodologies provide robust estimates of key loss metrics, they generally lack the flexibility of stochastic catalogs in considering different portfolios of insured properties and the ability to account for the many financial structures that are commonly used in the insurance industry. On the other hand, stochastic catalogs produced by long model runs, as the one presented in this study, generally require to be adjusted according to some externally defined targets. Merging these apparently diverging approaches would then be required to enable resilient risk management strategies.

REFERENCES


