

RESEARCH ARTICLE

Adjusting catastrophe model ensembles using importance sampling, with application to damage estimation for varying levels of hurricane activity

Stephen Jewson  | Clair Barnes | Stephen Cusack | Enrica Bellone

Risk Management Solutions Ltd,
London, UK

Correspondence

Stephen Jewson, Risk Management
Solutions Ltd, London, UK.
Email: stephen.jewson@gmail.com

Funding information

RMS Ltd

Abstract

Risk modellers in the insurance industry use catastrophe models to estimate the distribution of possible damage from natural catastrophes. The output from catastrophe models is often adjusted to create alternative risk scenarios. These adjustments are made for many reasons, such as to reflect different scientific hypotheses, different interpretations of historical data or different scenarios related to climate variability and climate change. Models that present the output in a list of simulated synthetic events with their associated damage (so-called event loss tables) can be adjusted rather easily, since information about desired adjustments is typically expressed in terms of changes in the properties of events. Models that present the output in a list of simulated synthetic years (so-called year loss tables) are harder to adjust, however, because the occurrences of the events are hard-wired into the simulated years. A method is described that allows the adjustment of the results in a year loss table by the application of weights to the years. The weights are calculated in such a way as to capture the specified changes in properties of the underlying events. The method is demonstrated by applying it to output from a catastrophe model and using it to quantify the changes in US hurricane wind damage due to shifts between long-term average, active and inactive levels of hurricane activity. It is shown that the method works well by comparing the results with more accurate results derived directly from the underlying event loss table.

KEYWORDS

catastrophe model, hurricane activity, hurricane damage, importance weighting, insurance, natural catastrophe, natural disaster

1 | INTRODUCTION

Catastrophe (cat) models are complex computer models that are used in both private and public sectors to estimate the distribution of possible damage (hereafter known as loss) that might be caused by future natural disasters. According

to the review by Friedman (1972), the first such models were developed in the 1960s by the US insurance company Travelers to quantify the amounts of money that the company might have to pay out due to hurricanes, earthquakes, floods and other perils. The models Friedman describes used ensembles of statistically simulated synthetic natural

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Meteorological Applications published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

disasters. Simulated values of a particular hazard, such as wind speeds of hurricanes, were combined with datasets of the properties of insured buildings (known as exposure) and estimates of how those buildings might be damaged by the hazard (known as vulnerability) to create estimates of both regional and total loss. Cat models are now used throughout the insurance industry for pricing of insurance and reinsurance, as well as allocation of capital and calculations related to solvency. Today's cat models, described in textbooks such as Dong (2001), Mitchell-Wallace *et al.* (2017) or Michel (2018), follow a similar design to those described by Friedman (1972), albeit with higher resolution, more realistic simulations and larger ensembles.

Two particular variations of model design have emerged and both are commonly used in today's insurance industry. They differ in terms of how they calculate and present results: event loss table (ELT) models calculate and report losses by simulated event, and year loss table (YLT) models calculate and report losses by simulated year.

1.1 | ELT catastrophe models

An ELT model for a given peril (such as hurricane winds or earthquake) creates output that consists of a set, or ensemble, of simulated synthetic catastrophic events, with information about the estimated frequency and loss of each event in the set. These event sets may contain somewhere from 10,000 to 1,000,000 events for each peril and region combination, in an attempt to simulate all possible significant events within some notional level of tolerance. In a commonly used version of the ELT design, the events are considered independent with frequencies that are Poisson distributed, with each event having its own Poisson frequency for occurrence rates *per* year. The event losses are given by either a single fixed value for each event or a probability density function (PDF) for the conditional loss given the occurrence of that event, known as the distribution of secondary uncertainty. The distribution of secondary uncertainty is different for different events: intense events hitting highly populated regions create a large loss, and weak events hitting thinly populated regions create a small loss. The secondary uncertainty introduces some variability around those losses. Only these kinds of Poisson ELTs are considered in the present study. However, it is worth noting that ELTs can be extended to support other frequency distributions such as the negative binomial (motivated by the work on winter storm clustering by, for example, Brady (2000), Mailier *et al.* (2006) and Cusack (2016), and the work on hurricane clustering by, for example, Jagger and Elsner (2012)) and certain limited types of dependences between events, as described by Khare *et al.* (2015).

The events in an ELT are typically created using complex statistical or dynamical models that simulate large numbers

of synthetic events with as much physical realism as possible. These simulations may take months or years to create. The output from these simulations consists of maps of hazard values (such as wind speed) for each event, and the hazard events and vulnerability data are then stored in a database and distributed to the users of the cat model. The users calculate losses for their own portfolios of exposures in a run-time calculation that combines hazard, vulnerability and exposure and that may take a few hours or days on a computing cluster running cat modelling software. The output is a table of events with losses that can be used to derive standard summary statistics such as the average annual loss (AAL) (i.e. the expected annual damage) and the aggregate exceedance probability (AEP) (i.e. the probability that the total annual damage exceeds a certain level). AEP curves are presented either in terms of annual probabilities or in terms of annual return periods, defined as 1 divided by the annual probability.

There are two main benefits to using ELT-based models. First, the summary statistics listed above, and certain other summary statistics, can be calculated very precisely given the event set. Second, adjustments can be made to the output, after losses have been calculated, by adjusting either the Poisson frequencies or the individual distributions of secondary uncertainty, event by event, and recalculating the summary statistics. This second benefit is significant because calculating losses from cat models can be computationally expensive, particularly for large portfolios of exposure, and if adjustments to model results can be made without having to go back to calculating event losses from the hazard, vulnerability and exposure that is more efficient. These benefits are counterbalanced by the disadvantage that the ELT structure does not allow for the inclusion of arbitrarily complicated temporal dependence among cat events. Although Khare *et al.* (2015) were able to introduce some simple temporal dependence into the ELT framework, more complex temporal dependences between events such as the influence of one flood event on the next (Villarini *et al.*, 2013), or the influence of the level of repair from one event on the level of damage of the next event, cannot typically be reproduced.

1.2 | YLT catastrophe models

The YLT approach is a more general approach that overcomes the limitations of the ELT structure by simulating large sets, or ensembles, of coherent years, rather than sets, or ensembles, of separate events. YLT ensembles may have between 10,000 and 1,000,000 members. Any simulation methodology can be used to create the sequence of hazard in each simulated year, and any temporal dependences can in principle be accommodated. In practice, however, most YLT

approaches still assume the existence of discrete events, and in such cases a YLT consists of a list of years, a list of events that occur within each year and a single realization of loss for each event. Only these kinds of event-based YLTs are considered in the present study.

Although the YLT approach is ultimately more general than the ELT approach, and hence preferable in many complex applications of cat models, it has two potential shortcomings. The first is that summary statistics are now calculated from simulated years and so are subject to an additional source of simulation error relative to the ELT approach. For large portfolios of exposure (e.g. a set of 1 million or more buildings) and loss results at relatively high levels of probability (i.e. short return periods), such as a probability of 1 in 10 years, this additional simulation error is typically small. For individual buildings or for loss results at relatively low levels of probability (i.e. long return periods), such as a probability of 1 in 200 years, this additional simulation error can be large, and further simulations may be required to achieve adequately stable and converged results (see Kaczmariska *et al.* (2018) for a detailed study of these convergence issues in the context of a cat model for European flooding).

The second shortcoming of the YLT approach is that making adjustments to a YLT is more complex than it is for an ELT. This is for two main reasons. The first is that many potential adjustments are expressed in terms of the user's expectations about frequencies and losses (or changes in frequencies and losses) of types of events, rather than in terms of information about types of years that might occur. Examples of event-based information that might motivate adjustments include changes in rates due to (a) climate change and its impacts, e.g. changes in the clustering of tornado frequency (Elsner *et al.*, 2014), (b) long-term climate variability, such as the Atlantic multidecadal oscillation forcing changes in hurricane numbers (Goldenberg *et al.*, 2001) or Europe winter storm frequency (Peings and Magnusdottir, 2014), and (c) predictable interannual variability, such as for US tornadoes and hail (Allen *et al.*, 2015) and European winter storms (Wang *et al.*, 2017). Such event-based information can be applied readily to an ELT, since results for each event are presented separately, but is harder to apply to a YLT in which events are hard-wired into simulated years. The methods described below solve this issue by converting event-based statements like 'there will be $x\%$ more flood events' into year-based statements like 'years with flooding will become $y\%$ more likely'. It is possible to apply event-based adjustments to a YLT by resimulating the YLT based on the updated information about the frequencies and losses of events. However, if the YLT simulation process is itself very complex this may not be practical. For example, the flood model described by Kaczmariska *et al.* (2018) and

Zanardo *et al.* (2019) simulates the years of the YLT using the output of continuous temporal simulations of the shallow water equations. This requires months of computation time and cannot be easily repeated. The second reason that adjusting YLTs directly is complex is that YLTs only contain hazard information via the events they contain. For instance, to adjust sea level (which is a property of a year, not of any particular event) in a YLT model for storm surge risk, one would still have to consider the sea level in each of the events in each year in that model.

1.3 | YLT weighting

In the present study, the difficulty of adjusting YLTs is addressed by presenting an algorithm for using event-based information to adjust YLTs without having to resimulate the years that make up the YLT. The method works by applying weights to the years in the YLT. The weights are calculated using a version of the statistical method of importance sampling in order to adjust the frequencies and losses of events within the simulated years in the desired way. The result is a weighted YLT, or WYLT.

The use of weights in a YLT, however, can be problematic. For instance, whereas unweighted YLTs (UYLTs) from independent perils can easily be combined to create estimates of total losses due to all perils by adding them together year by year, WYLTs are more difficult to combine. To address this a second stage of the algorithm is described which consists of a method for approximating the output WYLT with a new UYLT, which is referred to as the output UYLT.

1.4 | Testing YLT weighting

The YLT adjustment method is tested on a YLT from a hurricane loss model. The model is a commercial system, widely used in the insurance industry, which produces both ELT and YLT output, where the YLT is generated from the ELT using statistical simulation. The commercial product has a number of settings and outputs. The initial output used here consists of an ELT and a YLT that capture a baseline view of hurricane risk for industry exposure, in which the annual rates of hurricane occurrence are determined from long-term averages using data from 1900 to 2017. These results are referred to as the 'long-term rates' (LTR) view of risk. In the last 70 years, however, hurricane activity has fluctuated between active and inactive phases, with fewer hurricanes observed during the period 1970–1995 than either before or after that period (Goldenberg *et al.*, 2001). The causes of these fluctuations have been much debated, but without scientific consensus as to what they reveal about hurricane activity now or in the future. For instance Kim

et al. (2018) and Booth *et al.* (2012) present arguments for different possible mechanisms that could be driving these fluctuations, which would in turn suggest different methods of making predictions. Given this uncertainty, and the importance of understanding future hurricane risk for the insurance industry, various methods have been developed by the industry for making predictions of hurricane risk based on different possible future scenarios (Jewson *et al.*, 2008). These methods have been tested and shown to improve predictions of hurricane landfall rates and loss in hindcasts (Bonazzi *et al.*, 2014). A simple version of the logic used to make scenarios would be that, if the inactive phase is considered unlikely to return, then future risk would be better estimated using a scenario developed by eliminating inactive years from the statistics. However, if the inactive phase is considered likely to return, then future risk would be better estimated using a scenario in which those years are included in the analysis. Losses calculated from two scenarios (described in more detail below) designed to capture these possible variations in hurricane activity will be called the active and inactive views of risk, and are designed to be reasonable bounds for the range of hurricane losses in the near future (i.e. in the next few years).

As a straightforward test of the impact of different possible hurricane activity scenarios, the LTR view of risk as calculated from the YLT output from the cat model is adjusted using the YLT adjustment method that is presented here to create the active and inactive views of risk. This example, however, is unusual in that the losses in the active and inactive scenarios can also be calculated using a more accurate calculation. Because the LTR YLT to which the adjustment is applied is itself generated from an underlying Poisson ELT (which is not always the case for other loss models, either for hurricane or for other perils), the losses for the active and inactive views of risk can also be calculated by adjusting the ELT directly and recalculating summary statistics without simulation of years. The changes calculated in this way from active and inactive ELTs will be more accurate since they avoid convergence errors due to annual simulation that arise when using YLTs. In fact, in this particular case, using ELTs would be the preferable way to calculate the changes in losses for the active and inactive scenarios. For our current purposes, however, being able to calculate the results both ways (from adjusted ELTs and from adjusted YLTs) provides a useful opportunity to test the accuracy of the YLT adjustment method.

In more general settings in which YLTs are not generated from underlying ELTs this method for testing the YLT adjustment method is not available. This is typically the case for any cat model that attempts to move beyond a simple statistical representation of the frequency of events and instead uses physically based simulations to capture frequencies. Of

the various perils represented by cat models, physically based simulations of frequency are currently most commonly used in flood cat modelling, although one might imagine that ultimately all weather-related cat models would be created this way. In these flood models, events are derived from the output of continuous simulations of differential equations, and are locked into place in their simulated years in order to preserve the overall character and properties of that year (e.g. see the European flood model described by Kaczmariska *et al.* (2018) and Zanardo *et al.* (2019)). For these models the YLT adjustment method is a way to adjust the model while still conserving the properties of the individual simulation years.

Figure 1 illustrates the YLT adjustment method, and the relationships between the ELTs and YLTs used in this paper to evaluate the YLT adjustment method. In Section 2 the cat model and the YLT adjustment methodology are described in more detail, and results based on the LTR, active and inactive views of risk, all based on ELTs, are presented. These results are the benchmarks against which the results of the YLT adjustment method are validated. In Section 3 basic US nationwide results from the YLT adjustment method are shown and compared with the benchmarks. In Section 4 further results are presented, looking at (a) changes in regional modelled losses, which are a tougher test of the method, and (b) the impact of using different underlying YLT simulation sets. In Section 5 the method and results are summarized.

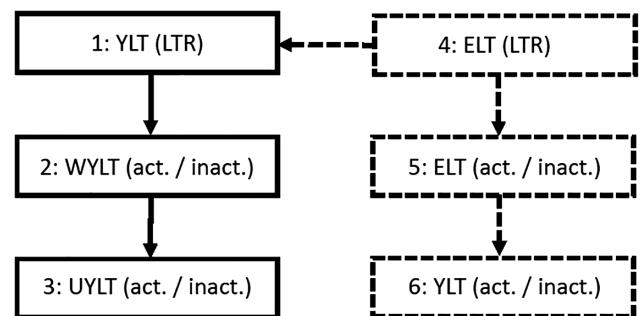


FIGURE 1 The relationships between the datasets used in this paper. The left column illustrates the year loss table (YLT) adjustment method being presented and tested in this study, which converts a YLT for the long-term rates (LTR) view of risk into weighted YLTs (WYLTs) for the active and inactive views of risk, and then converts those into unweighted YLTs (UYLTs) for the active and inactive views of risk. The right column illustrates the testing performed. The adjustment method is tested by starting with an event loss table (ELT) for the LTR view of risk and uses that to create the LTR YLT. The LTR ELT is also used to create ELTs for the active and inactive views of risk, and YLTs for the active and inactive views of risk directly from those ELTs. The results from steps 1, 2 and 3 in the diagram are evaluated by comparing with results from 4, 5 and 6

2 | CATASTROPHE MODEL AND REWEIGHTING METHODOLOGY

2.1 | Hurricane catastrophe model

The cat model used in this study is a commercial model produced by Risk Management Solutions Ltd for estimating wind damage in the mainland United States due to hurricanes. It is one of a number of such models that are used in the insurance industry. Its structure follows a fairly standard template for the construction of such models, as described in, for example, Friedman (1972), Michel (2018) or Mitchell-Wallace *et al.* (2017). The hurricane wind hazard is simulated from several components, including models that simulate locations of hurricane genesis, tracks, intensities and wind fields. The hazard model contains 29,693 events. Exposure is represented in a database that captures building locations, values and attributes, and the vulnerability of different building types is based largely on empirical damage data from previous hurricanes. Various aspects and versions of the model have been published in academic journals, such as the basic formulation of hurricane track modelling (Hall and Jewson, 2007), aspects of the wind field model (Khare *et al.*, 2009) and dynamical-modelling-derived components for hurricane landfall (Colette *et al.*, 2010) and hurricane transitioning (Loridan *et al.*, 2015). Successive versions of the model have been peer-reviewed (annually or biannually from 1996 to the present) by scientists employed by the Florida government as part of the Florida Commission on Hurricane Loss Projection Methodology, and many details of the model are published as part of that process (see www.sbafla.com).

In the present study generic exposures that capture the entire US residential building stock, known as a residential industry exposure database, are used throughout. Since output from the model is in dollars, which quickly become relatively meaningless because of inflation and other changes in the values of exposures, all results are normalized so that the nationwide AAL for the LTR view of risk calculated from the ELT is 100.

2.1.1 | LTR, active and inactive views of risk

The LTR view of risk in the model is based on the average number of US landfalling hurricanes *per year* during the period 1900–2017, which gives 3.89 hurricanes *per year* (where the definition of landfalling includes storms which cause damage over land but for which the eye does not make landfall). The active and inactive views of risk can be considered as predictions for how many hurricanes might make landfall *per year*, over the next 5 years, under active and inactive climate scenarios. The predictions are based on the methods described by Jewson *et al.* (2008) and Bonazzi

et al. (2014). These methods do not simply average the number of landfalling hurricanes during the active and inactive periods, since landfalling hurricane numbers are strongly affected by the ‘noise’ of weather variability, which makes signals of climate variability hard to detect. Instead, in an attempt to make more accurate predictions, they take into account a number of factors that are considered relevant for estimating US hurricane landfall rates including sea surface temperatures in the Indian, Pacific and Atlantic oceans, the relationships between these sea surface temperatures and Atlantic basin hurricane numbers (Emanuel, 2005) and estimates of the proportion of Atlantic basin storms that make landfall (Coughlin *et al.*, 2009). After taking these factors into account the predictions result in different changes in the rates of storms by strength and by region, and this leads to changes in the rates of every storm in the event set. The active view contains 4.16 hurricanes *per year* (an increase of 7% relative to the LTR view) and the inactive view contains 3.49 hurricanes *per year* (a decrease of 10% relative to the LTR view).

2.1.2 | ELT outputs

The cat model produces ELT output for all three views of risk (LTR, active and inactive), consisting of frequency and loss for each event. For each of these three ELTs, the AAL is calculated as the weighted average of the expected loss from each event, where the weights are the Poisson rates. The normalized AAL values are 100 for the LTR view of risk (by definition), 121.3 for the active view of risk and 82.7 for the inactive view of risk. The changes in loss are larger than the changes in the number of hurricanes because loss is disproportionately caused by the major hurricanes, which vary more between the three scenarios than the overall number of hurricanes. The AEPs are calculated from each ELT using the mathematical result that the Fourier transform of a sum of independent random variables is the product of the Fourier transforms of the individual random variables (see for example Wang, 1998): the individual random variables are the losses for each event in a particular year, and their sum is the annual loss. The differences between the AEP curves for the three views of risk, calculated from the respective ELTs, are shown in Figure 2a. The vertical zero line corresponds to the LTR view of risk, the right curve to the active view of risk, and the left curve to the inactive view of risk.

These AAL and AEP values derived from ELTs are not subject to simulation error due to annual simulation (although they are subject to various other sources of simulation error, such as errors in the simulation of the events themselves) and are used as synthetic ‘truth’, or exact values, in the comparisons below. The AAL values and the

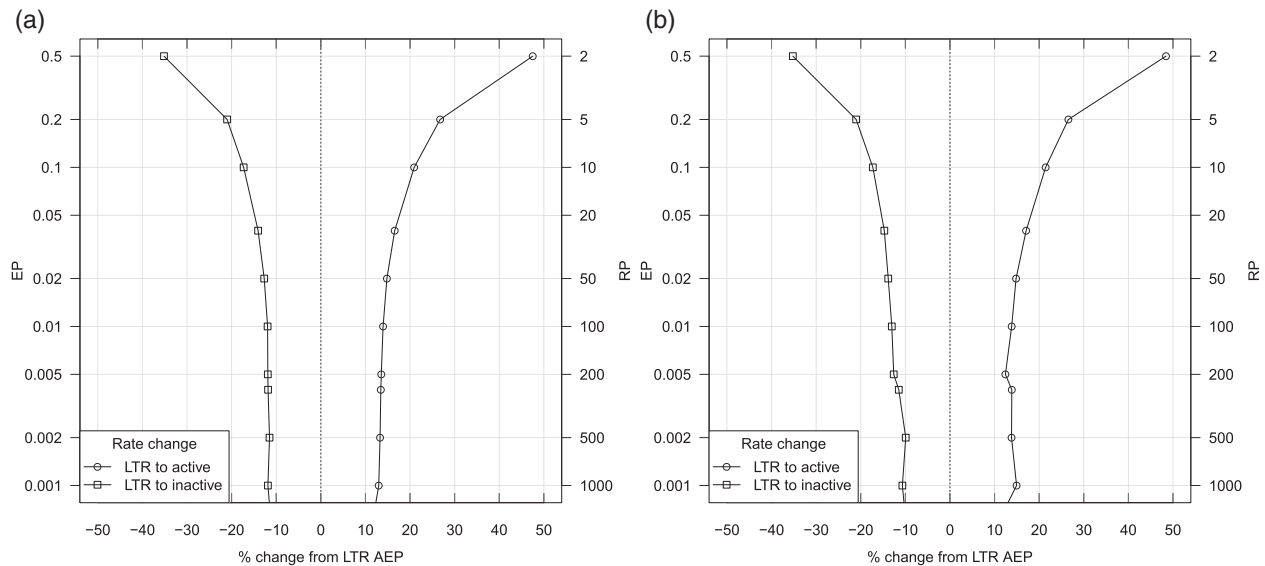


FIGURE 2 (a) Differences in annual US hurricane wind loss (aggregate exceedance probability, AEP), calculated using the Risk Management Solutions model, and derived from event loss table (ELT) output from the model. The left vertical axis shows Exceedance Probability (EP) and the right vertical axis shows Return Period (RP). Values are shown as changes relative to the losses calculated from hurricane rates derived from long-term historical rates (the long-term rates [LTR] view of risk), which therefore correspond to the vertical zero line. The right curve (open circles) shows the changes in losses due to hurricane rates derived from hurricane activity during active historical periods (the active view of risk) and the left curve (open squares) shows the changes in losses due to hurricane rates derived from hurricane activity during inactive historical periods (the inactive view of risk). (b) As in (a) but all based on a single realization of 50,000 years of simulation. The vertical zero line corresponds to the LTR view of risk derived from this single simulation, and the right and left curves (open circles and squares) are approximations to the right and left curves in (a) derived by applying the year loss table adjustment method described in the text. Comparison with (a) suggests that the adjustment method is capturing the differences between the curves well, with some variability at long return periods

AEP curves show that, not surprisingly, the losses are higher during periods of higher hurricane activity and lower during periods of lower hurricane activity.

2.1.3 | YLT outputs

YLT output from the cat model is also generated for all three views of risk from the corresponding ELT in the following two-step process. In the first step, 800,000 years of simulation are generated by sampling from the Poisson distribution for the occurrence of each event and from the distribution of secondary uncertainty for the loss of each event. 800,000 is a sufficiently large number of years that the mean numbers of hurricanes *per* year in these YLTs match very closely with the expected numbers of hurricanes implied by the underlying ELT (3.89, 4.16 and 3.49 hurricanes *per* year, as described in Section 2.1.1. above). Different occurrences of the same event within the 800,000 years are given different samples of secondary uncertainty. Since all events occur multiple times within the 800,000 years (for instance, in the LTR YLT events occur on average 105 times each) the distribution of secondary uncertainty for each event is well sampled. The second step then seeks to reduce the size of the simulation set while still preserving the loss statistics reasonably well. This is achieved by ranking by loss the

800,000 simulated years generated in the first step and selecting 50,000 years from the 800,000 at equal intervals in the ranking. This reduced set of 50,000 years of simulation gives similar results to the 800,000 year set and gives a more accurate estimate of the AEP than using unreduced simulations to generate 50,000 years directly. This works because the 800,000 year sets contain many near duplicates of years, with similar numbers of hurricanes causing similar losses. The number of near duplicates is reduced in the 50,000 year sets.

The AAL and AEP curves derived from the reduced set of 50,000 years differ slightly from the values generated directly from the ELT, because of annual simulation error. This variability is quantified below by creating multiple random versions of each type of YLT, based on different random versions of the underlying 800,000 year simulation. In Section 4.2 the impact on the results of using other methods for the creation of the YLT that use different numbers of years of simulation and eliminate the reduction step is also explored.

For each set of YLT output, the AAL from the model is calculated as the average of the annual losses from each simulated year, and the AEP is calculated as 1 minus the empirical cumulative distribution function based on the simulated annual losses. The empirical cumulative distribution

function is calculated in the usual way by ranking the annual losses and assigning equally spaced cumulative probabilities to each.

2.2 | Reweighting scheme stage 1: from YLT to WYLT

Figure 2a shows the difference between the losses for the LTR view of risk and the losses for the active and inactive views of risk, with all calculations based on the corresponding ELTs. The goal of the present study is to attempt to replicate the active and inactive results by starting from a simulated YLT for the LTR view of risk and applying the YLT adjustment method. If by doing so it can be demonstrated that the results from the YLT adjustment method are reasonably accurate in capturing the impacts of rate changes, then in other cases where an ELT is not available (such as most flood cat models: see the discussion in Section 1.4 above) one could apply the adjustment method with relative confidence.

The YLT adjustment method works as follows. In a standard YLT, the ensemble members (the individual years of simulation) have equal weighting (equal to 1 over the ensemble size): this is an UYLT. A generalization of UYLTs is to weight the ensemble members differently, with weights that sum to 1: this is a WYLT.

The addition of weights can be used to adjust the results of an UYLT. To give a simple example, by increasing the weights on years that contain large losses and decreasing the weights on years that contain small losses (such that the weights still sum to 1), the distribution of loss results will be shifted towards larger losses. Weighting the years in an UYLT based on event-level information, however, is not straightforward, because individual years may contain some events for which the frequency needs to increase and some events for which the frequency needs to decrease. The first stage of the YLT adjustment method is an algorithm that determines yearly weights that adjust the frequencies of all events as desired, while accounting for the possible occurrence of events of different types within each year. The algorithm is based on the statistical method of importance sampling, which is a method that allows samples from one distribution to be used as samples from a different distribution. Given samples from a PDF or probability mass function (PMF) $g(x)$ (referred to as the proposal distribution), then importance sampling allows the creation of weighted samples from a different PDF or PMF $f(x)$ (referred to as the target distribution) by weighting each of the samples from $g(x)$ using a weight $f(x)/g(x)$. This method is discussed in detail in many places in the statistics literature such as Todkar and Kass (2010) and Wasserman (2003). Intuitively, the method works because it puts weights greater than 1 on values of x which are more likely

in the target distribution f than in the proposal distribution g , and weights of less than 1 on values of x which are less likely in f than g .

In our case, the random variable x describes a year of hurricane activity by listing how many times each event from the event set occurs in that year. The proposal distribution is the PDF or PMF for the LTR version of the model, while the target distribution is the PDF or PMF for either the active version or the inactive version. The weights $f(x)/g(x)$ are calculated for each year in the YLT simulation and are then applied to the YLT to create WYLTs that approximate the distribution of hurricane properties, and hence loss, for the active and inactive views from the model.

The details of the method for calculating the probabilities g and f for the proposal and target distributions work as follows. Since the changes being specified between the LTR and active/inactive views are changes in frequencies of events, the probabilities can be calculated using the frequency component of the statistical model given by the ELT, which specifies a Poisson distribution for the occurrence of each event. Since the Poisson distribution takes only discrete values for the random variable, it uses a PMF rather than a PDF. The PMF value of a year in the YLT can be calculated as the product of the PMF values of the occurrence of each event in the event set, using the Poisson PMF.

To illustrate how this would work in a simple case, imagine an event set with just three hurricanes, A, B and C, and that year 1 in the LTR YLT contains just a single occurrence of hurricane A, year 2 contains just a single occurrence of hurricane B and year 3 contains single occurrences of both hurricanes A and B. The weight on year 1 would be based on the product of the target distribution probabilities of hurricane A occurring and hurricanes B and C not occurring, divided by the product of the proposal distribution probabilities of hurricane A occurring and hurricanes B and C not occurring. Similarly, the weight on year 2 would be based on the ratio of probabilities for hurricane B occurring and hurricanes A and C not occurring, while the weight on year 3 would be based on the ratio of probabilities for hurricanes A and B both occurring and hurricane C not occurring. Over many years of simulation this leads to convergence onto the target distribution for the rates of occurrence of each of the hurricanes A, B and C.

In fact there are 29,693 events in the ELT and so the calculation of the PMF for each individual year involves the product of 29,693 Poisson PMF values. For the proposal distribution this gives a probability mass g_i for year i of:

$$g_i = \prod_{j=1}^N \text{Poisson}(n_{ij}; \lambda_j) \\ = \prod_{j=1}^N \frac{e^{-\lambda_j} \lambda_j^{n_{ij}}}{n_{ij}!}$$

where the index j runs over all N events in the model (where $N = 29,693$), $\text{Poisson}(a; b)$ indicates the PMF of the Poisson distribution for random variable a and parameter b , n_{ij} is the number of occurrences of hurricane j in year i of the LTR YLT, and λ_j is the Poisson parameter for hurricane j in the proposal distribution (the LTR model). Similarly for the target distribution, the PMF f_i for year i is given by:

$$f_i = \prod_{j=1}^N \text{Poisson}(n_{ij}; \mu_j)$$

$$= \prod_{j=1}^N \frac{e^{-\mu_j} \mu_j^{n_{ij}}}{n_{ij}!}$$

The only difference between the expressions for the proposal and target distributions is the Poisson rate, where μ_j is now the Poisson parameter for hurricane j in the target distribution (which in our example comes from either the active or inactive scenarios).

The ratio of these probability masses for year i gives the weight for year i as:

$$w_i = \frac{f_i}{g_i} = \frac{\prod_{j=1}^N e^{-\mu_j} \mu_j^{n_{ij}}}{\prod_{j=1}^N e^{-\lambda_j} \lambda_j^{n_{ij}}}$$

$$= \prod_{j=1}^N \frac{e^{-\mu_j} \mu_j^{n_{ij}}}{e^{-\lambda_j} \lambda_j^{n_{ij}}}$$

The numerical values are very small, and everything is calculated using logs to avoid underflow.

2.3 | Reweighting scheme: From WYLT to adjusted UYLT

The second stage of the YLT adjustment method is an algorithm that approximates the output WYLT produced in the first stage with a new unweighted YLT that is called the output UYLT. There are various ways this could be done. Perhaps the simplest way would be to sample years randomly from the WYLT, taking the weights into account in the sampling. It was found, however, that this leads to poorer results than other methods tested because it introduces considerable randomness. To illustrate this problem, consider a member in the WYLT with a weight of a 2.3/50,000. In a random sampling scheme to create a 50,000 year UYLT from this WYLT, this year would be expected to occur 2.3 times on average over many repeats of 50,000 years. Since UYLTs only allow an integer number of occurrences for each year, the WYLT will be approximated most closely, for this year, if it occurs exactly twice. However, in an individual

50,000 year sample created by random sampling it might occur 0, 1, 2, 3, 4 or more times, depending on the random sampling for that realization. If it occurs anything other than twice, the simulated UYLT is losing information unnecessarily relative to the WYLT. To reduce this randomness and loss of information that occurs in random sampling schemes, the WYLT is converted to an UYLT using a simple non-random sampling scheme. The years in the WYLT are ordered by the annual loss and used to create an empirical cumulative distribution function. Fifty thousand regularly spaced values are drawn from the interval from 0 to 1, and the inverse of the empirical cumulative distribution function is used to convert these values from probabilities into years. Numerical comparisons of this method vs. random sampling were performed, and this method was found to be more accurate (not shown).

3 | RESULTS FOR US LOSSES FOR ACTIVE AND INACTIVE VIEWS OF RISK

Results from applying and testing the YLT adjustment method described above are now presented. In this section all results represent US nationwide losses. Regional losses are discussed in Section 4.1. Figure 2b shows curves that correspond to the curves in Figure 2a but are now based on simulation and application of the adjustment method. The vertical zero line represents the AEP for the LTR view, now based on a single realization of the LTR YLT. The right (left) curve in Figure 2b is the difference from this LTR AEP for the active (inactive) view, generated from the output UYLT created from the LTR simulation by applying the YLT adjustment method. By comparing with Figure 2a it can be seen that, qualitatively, the YLT adjustment method is working: the differences in the AEP curves for the active and inactive views based on 50,000 years of simulation (Figure 2b) agree well with the differences generated from the ELTs (Figure 2a).

A further 49 random realizations of the LTR YLT are then created (each of 50,000 years, and each created from a new 800,000 year simulation, but all based on the same LTR ELT) to understand the level of variability in the results due to annual simulation. This leads to 50 versions of the LTR AEP curve, which are shown relative to the exact LTR AEP as the narrow central envelope in Figure 3a. It can be seen that the results based on different realizations show some slight spread around the benchmark, because of variability due to the use of random simulations.

Each of the 50 LTR YLT realizations can then be separately converted using the YLT adjustment method to generate active and inactive output UYLTs and corresponding AEP curves. The 50 resulting active and

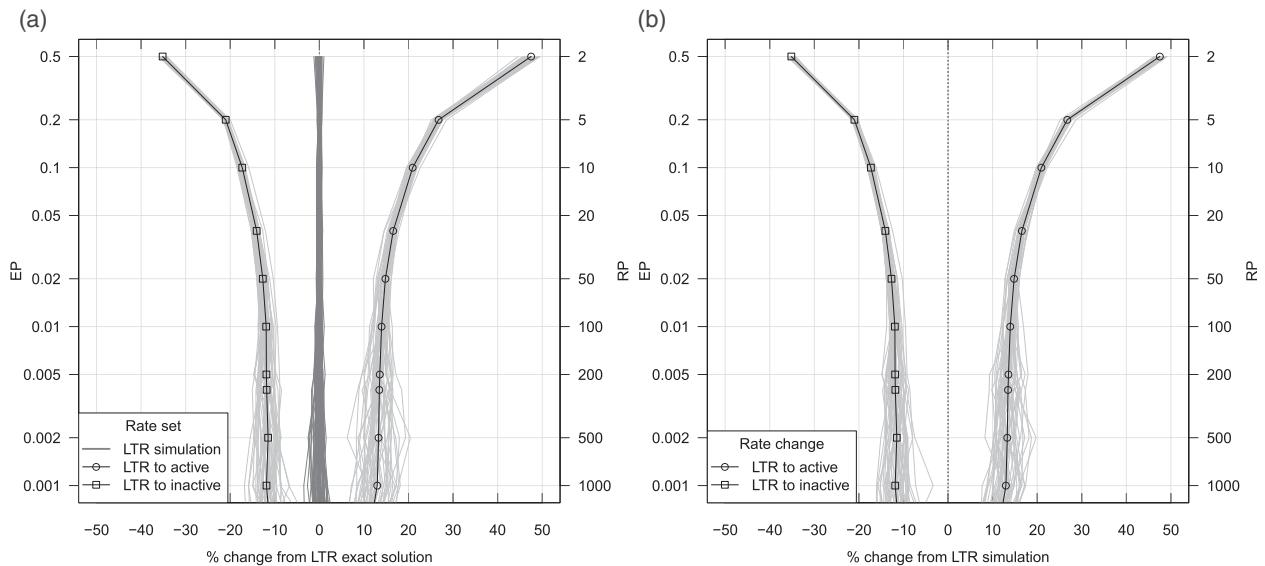


FIGURE 3 In (a) the two darker lines (with circles and squares) are the same as in Figure 2a. The 50 fine lines around the vertical zero line show changes in losses for the long-term rates (LTR) view of risk when generated from 50 realizations of 50,000 years of simulation, relative to the exact LTR view of risk. The 50 fine lines around the right (left) dark line show changes in losses generated by applying the adjustment method separately to each of the 50 central lines. One from each set was shown as the right and left lines in Figure 2b. In (b) the 50 fine lines on the right (left) of zero show differences between corresponding pairs of fine lines in (a), and thus show 50 realizations of the size of change from the LTR view of risk to the active (inactive) view of risk as estimated by the year loss table adjustment method. The solid lines show the same changes derived from the benchmark results from Figure 2a, for comparison

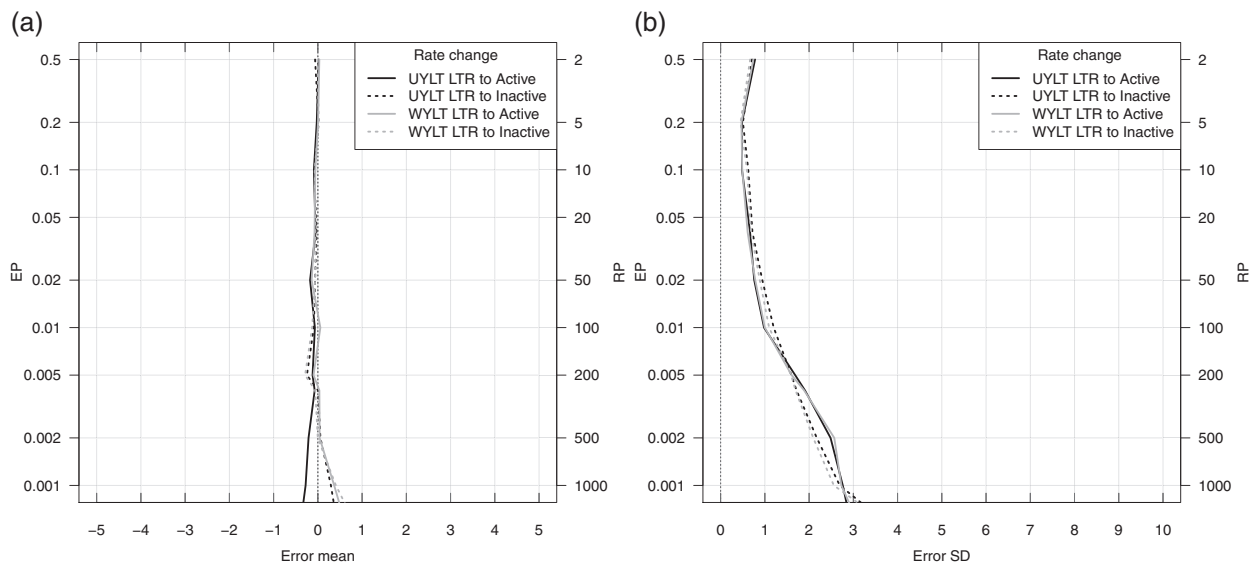


FIGURE 4 The mean and the standard deviation of the percentage errors due to the year loss table (YLT) adjustment method, corresponding to the results shown in the previous figures. Errors are calculated by comparing the results from the adjustment method with exact results derived from active and inactive versions of the event loss table

inactive AEP curves, their variability and how accurately they replicate the benchmarks can be illustrated in various ways. First, in Figure 3a, the 50 active and inactive AEP curve changes are shown (as the right and left envelopes), along with the benchmark active and inactive AEP changes copied from Figure 2a (as the right and left bold

lines). It can be seen that there is some spread of losses around the exact values, but that all the simulated curves capture well the basic impact of shifting rates from LTR to active or inactive.

Second, the changes from each of the 50 LTR AEP curves to each of the active and inactive AEP curves derived by

applying the YLT adjustment method can be calculated. These 50 changes are shown in Figure 3b, along with the exact changes derived from the benchmarks (bold lines). This figure shows more clearly that the changes calculated via the adjustment method agree with the benchmark changes, with some variability, although the variability is significantly less than the changes themselves. The variability around the adjusted results is largest at long return periods.

From the changes shown in Figure 3b various diagnostics can be derived to investigate the performance of the method more quantitatively. First the mean and standard deviations of the percentage errors in annual losses for each probability level, which are referred to as the bias and the standard deviation of the errors, are calculated. These are shown in Figure 4. There is relatively little bias, and the standard deviation increases with return period. In addition, the ratio of the absolute value of the mean change to the standard deviation can be calculated, which is referred to as the signal-to-noise ratio and summarizes the performance of the method. Figure 5 (dark lines) shows this signal-to-noise ratio: values above 5 for all return periods up to 500 years show that, consistent with Figure 3, the change is captured cleanly.

Corresponding results for the AAL are also shown. Figure 6 shows the AALs of the 50 YLTs generated for the LTR view of risk, and also the AALs of the 50 output UYLTs derived from them using the YLT adjusted method

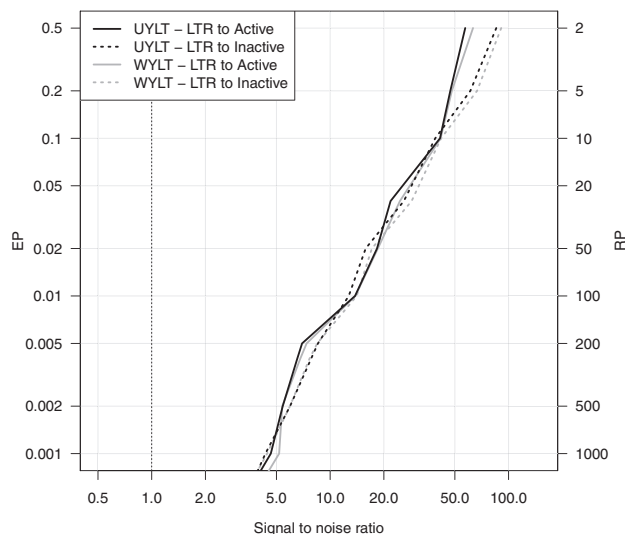


FIGURE 5 The signal-to-noise ratio for the errors in the weighting method described in the text, derived as the ratio of the mean change (shown by the solid lines in Figure 2b) to the standard deviation shown in Figure 4b. These are shown both for the first stage of the method alone (from the year loss table [YLT] to the weighted YLT [WYLT]) and for both stages of the method together (from the YLT to the output unweighted YLT [UYLT]). It can be seen that the ratio is similar for both stages, and hence it can be concluded that the second stage does not significantly degrade the results

(Figure 6c for active and 6e for inactive), along with the true AALs determined directly from the corresponding ELTs (dashed lines). It can be seen that the change in AAL is well captured by the adjustment method, with only a little variability relative to the size of the changes.

Figure 6a also shows the range of AAL values that are given by simulating 50 realizations of 50,000 years directly from active and inactive ELTs (Figure 6b and 6d), rather than via the YLT adjusted method. The variability in the AAL in these curves is roughly half the variability in the AAL calculated for the output UYLT created using the YLT adjustment method. This shows that the YLT adjustment method introduces extra variability relative to simulating from adjusted ELTs, in this case roughly doubling the variability. It can be concluded that, if simulating from adjusted ELTs is an option, it gives more accurate results than simulating from an unadjusted ELT and applying the YLT adjustment method to the resulting YLT. In many models, however, an ELT is not available and adjusting the YLT is the only option.

3.1 | Results for WYLT to adjusted UYLT: impact of the second stage

The separate impacts of the two stages of the adjustment method on the variability in the results are now considered, to understand whether the variability in the results arises from the first stage of the method (calculating the weights that convert the YLT to the output WYLT) or the second stage of the method (approximating the output WYLT with the output UYLT), and whether the second stage degrades the results to any significant extent. Figure 5 shows the signal-to-noise ratio for AEP curves calculated from the results of both stage 1 on its own (output WYLT: grey lines) and stages 1 and 2 together (output UYLT: darker lines). For both active and inactive views the two curves are very close, suggesting that although the output UYLT results are presumably slightly less accurate because of the approximation involved this is not creating noticeable extra variability in the results. The conclusion can be drawn that approximating the output WYLT with the output UYLT does not noticeably degrade the results.

4 | SENSITIVITY AND UNCERTAINTY

4.1 | Regional results

It has been shown above that the YLT adjustment method works well for US nationwide loss results, based on weights that adjust the frequencies of hurricanes at a nationwide level. A tougher test is whether the weights derived at a

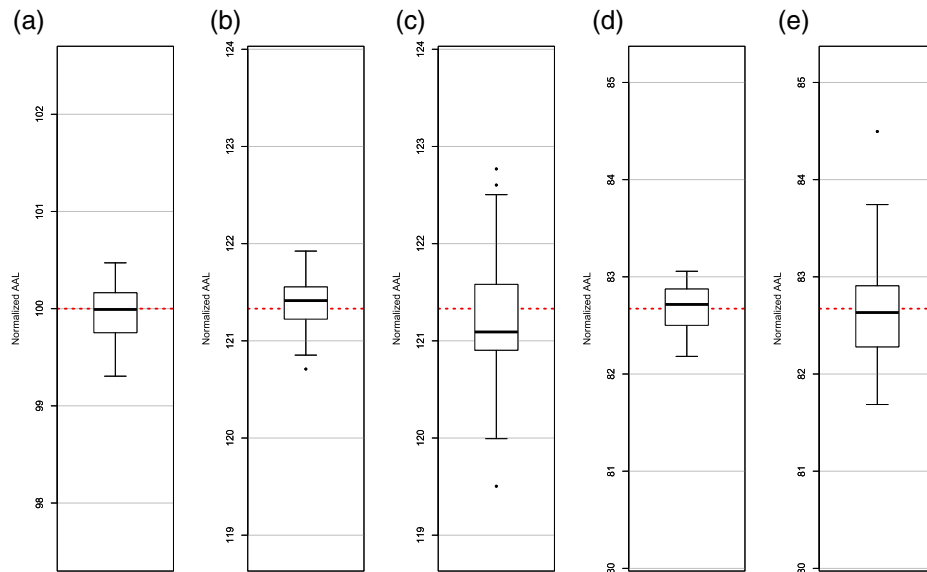


FIGURE 6 Average annual loss (AAL) values based on 50 realizations of 50,000 years of simulation for (a) the long-term rates (LTR) view, (b) the active view, simulated directly from an active view event loss table (ELT), (c) the output unweighted year loss tables (UYLTs) derived by applying the YLT adjustment method for active rates to the LTR simulations, (d) as (b) but for the inactive view, (e) as (c) but for the inactive view. By comparing (b) and (c), or (d) and (e), it can be seen that the results from the adjustment method suffer from greater variability than from just calculating the AAL directly from simulations from an adjusted ELT, but that the variability is still small enough, by a large margin, to distinguish the signal from the variability

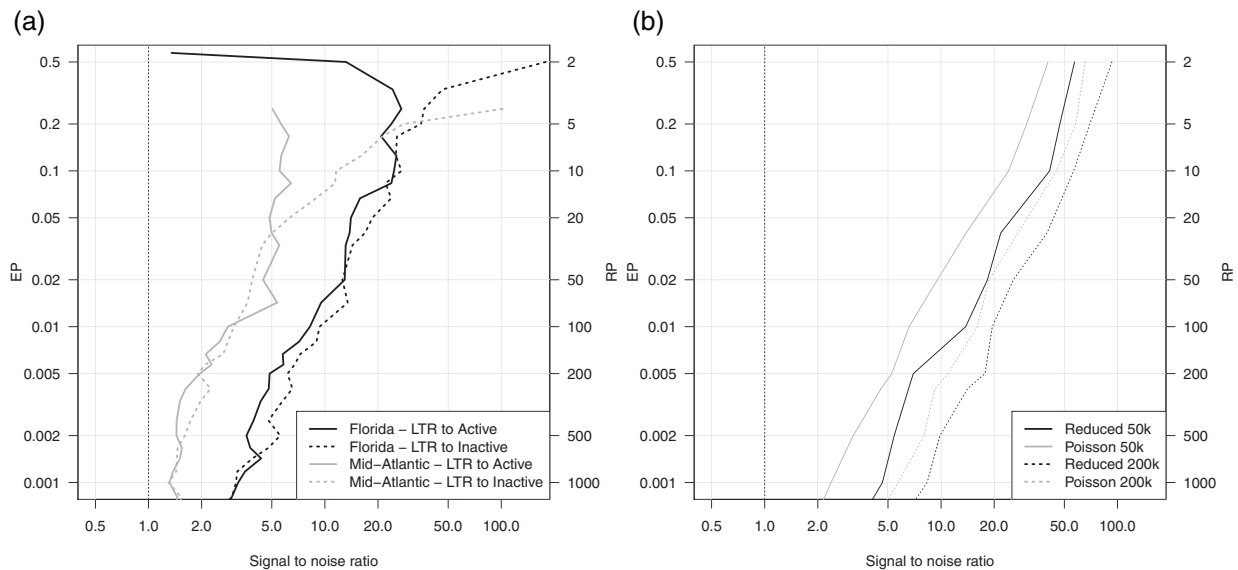


FIGURE 7 (a) As Figure 5 but now for two regions in the United States (Florida and mid-Atlantic). The values for the mid-Atlantic are lower (i.e. worse) than for Florida, as would be expected since the mid-Atlantic experiences fewer hurricanes and so signals are less well resolved. Comparison with Figure 5 shows that for both regions the signal-to-noise values are lower than for the United States as a whole, again as would be expected based on overall numbers of hurricanes in the different regions. (b) As Figure 5 but now for alternative simulation sources, as described in the text

nationwide level also give good regional loss results. The coastal United States is split into six regions and each region is considered separately. In Figure 7a, results are shown for the regions with the best and the worst results, ranked according to the level of signal-to-noise ratio, with best meaning the highest ratio and worst meaning the lowest

ratio. Results are only shown for the change from LTR to active, since the results for the change from LTR to inactive show a similar pattern. Comparison with the nationwide results shown in Figure 5 shows that the regional results are poorer: the signal-to-noise ratio is smaller than for the nationwide results. However, the signal-to-noise ratios are

all still above 1, even at the longest return periods, indicating that the method is still identifying the correct sign and approximate magnitude of change clearly above the noise.

4.2 | Effect of the method used to generate the initial YLT

All the above results were derived from a YLT based on 50,000 years of simulation created by reduction from 800,000 years, as described in Section 2.1.3 above. The question of whether the use of a different underlying simulation methodology, and the size of the YLT ensemble, makes a difference to the variability in the results is now investigated. To this end 50 realizations of each of three alternative LTR YLTs were created from the same LTR ELT, and the YLT adjustment method was applied to each. For the first of these sensitivity tests the alternative YLTs were generated using 50,000 years of unreduced simulation. For the second test the YLTs were generated from 800,000 years of simulation, reduced to 200,000 years using the same reduction algorithm as was used to generate the main 50,000 year set, and for the third test the YLTs were generated from 200,000 years of unreduced simulation.

Figure 7b shows the signal-to-noise ratios for these three sensitivity tests, along with values for the main case, which are the same as those shown in Figure 5 (solid dark lines). Only results for the LTR to active adjustment are shown. It can be seen that the signal-to-noise ratio depends heavily on the underlying simulation set used: the first sensitivity test, based on 50,000 years of unreduced simulation, gives the worst results, while the second sensitivity test, based on 200,000 years of reduced simulation, gives the best results, with the highest signal-to-noise ratios.

5 | CONCLUSIONS

The most general way to present the distribution of possible future disaster losses is to use large ensembles of simulated future years (or any other time period of interest), containing simulated disasters and their losses. These types of ensembles are known as year loss tables (YLTs). Often, it is of interest to be able to understand the sensitivity of the results in a YLT to various changes, e.g. possible changes due to climate change or changes in modelling methodology. However, YLTs are time-consuming to produce, involving months or years of modelling work, and it is generally not possible to create multiple alternative versions of a YLT based on differing assumptions. There is therefore a need for methods that allow users of YLTs to be able to adjust the output to estimate sensitivities without having to rerun the entire modelling calculation. However, many adjustments that users might consider making come in the form of

information about changes in the frequency or severity of individual events or types of events, rather than in the form of information about the changes in likely frequency of occurrence of certain types of years. Also, even when adjustments come in the form of information about years, they may be difficult to apply directly since YLTs consist of lists of events. For these two reasons, applying adjustments to YLTs can be difficult. A method based on importance sampling that addresses this challenge and that allows event-based adjustments to be applied to a YLT has been described. The method works by applying weights to the years in the YLT. The weights are determined according to the probability density function or probability mass function of the years before and after adjustment, which can be calculated from knowledge of the model and the adjustments required. Applying weights calculated in this way allows for the adjustment of rates of events, even in situations where there are multiple events within a year that need to be adjusted in different ways. By using this method, the losses in different scenarios can be calculated without having to recalculate event losses, or regenerate ensemble members, or recalculate annual losses, which makes the calculation very fast and vastly more feasible than rebuilding the model from scratch with new assumptions.

The first stage of the method produces a weighted YLT, which may be sufficient for some applications. For cases in which unweighted YLTs are preferred, a second stage consisting of a method by which the weighted YLT can be approximated with an unweighted YLT has been described, and shown to work well. An unweighted YLT created in this way is similar to the original unweighted YLT but with some years repeated and others deleted in such a way as to capture the desired adjustment.

The YLT adjustment method has been applied to output from a hurricane loss model and used to understand the impact on hurricane losses of modelling using US hurricane landfall rates based on either active or inactive periods of hurricane behaviour instead of the average behaviour over long historical records. Because of the way the hurricane loss model is constructed, it is also possible to calculate the exact results (where exact means exact given a fixed event set). This allows the weighting method to be evaluated in a way which would not be possible in most loss models, for which exact results would not be available. It was found that the weighting method performs well overall. The best results are for short return periods, for the largest ensembles, for reduced simulation ensembles and for US nationwide results. For these results, the estimated changes in loss calculated by applying the YLT adjustment method are highly accurate, with little sampling variability due to the use of annual simulations. The worst results are for long return periods, for the smallest ensembles tested, for unreduced

simulations and for regional results. However, even in the worst results presented (for 1,000 year return period regional losses, based on a 50,000 member unreduced YLT) the method successfully identifies the sign and approximate magnitude of the changes due to periods of active and inactive hurricane activity, albeit with significant variability.

The example presented is for a cat model that uses Poisson frequencies. However, the method could equally well be applied to cat models with more complex frequency models, such as negative binomial. For models in which the temporal variability is determined from dynamical simulations, rather than from an explicit statistical model, empirical frequency distributions could be used. Changes expressed in terms of changes in the severity of events can also be applied simply by representing them in terms of changes in frequency at different levels of severity.

This method is proposed as an efficient and general purpose scheme for applying changes to cat model results, for situations in which the model results are presented in YLTs. In particular, the method is a promising candidate for addressing the urgent need for methods to evaluate the impact of climate change and climate variability on damage due to natural catastrophes, in that it allows scientific hypotheses relating to the changing frequencies and severities of events to be applied to cat models without having to rebuild the model, which is rarely feasible.

CONFLICT OF INTEREST

The authors are all employed by Risk Management Solutions Ltd, and this paper presents and discusses results from a catastrophe model that is made and sold by Risk Management Solutions Ltd.

ORCID

Stephen Jewson  <https://orcid.org/0000-0002-6011-6262>

REFERENCES

- Allen, J., Tippet, M. and Sobel, A. (2015) Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. *Nature Geoscience*, 8, 278–283.
- Bonazzi, A., Dobbin, A.L., Turner, J.K., Wilson, P.S., Mitás, C. and Bellone, E. (2014) A simulation approach for estimating hurricane risk over a 5-yr horizon. *Weather, Climate, and Society*, 6(1), 77–90.
- Booth, B., Dunstone, N.J., Halloran, P.R., Andrews, T. and Bellouin, N. (2012) Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, 484, 228–232.
- Brady, C. (2000) *A dynamical and statistical investigation into serial storms*. Reading University Masters Thesis.
- Colette, A., Leith, N., Daniel, V., Bellone, E. and Nolan, D.S. (2010) Using mesoscale simulations to train statistical models of tropical cyclone intensity over land. *Monthly Weather Review*, 138, 2058–2073.
- Coughlin, K., Bellone, E., Laepple, T., Jewson, S. and Nzerem, K. (2009) A relationship between all Atlantic hurricanes and those that make landfall in the USA. *Quarterly Journal of the Royal Meteorological Society*, 135, 371–379.
- Cusack, S. (2016) The observed clustering of damaging extratropical cyclones in Europe. *Natural Hazards and Earth System Sciences*, 16, 1–13.
- Dong, W.-M. (2001) *Building a More Profitable Portfolio-Modern Portfolio Theory with Application to Catastrophe Insurance*. London: Reactions Publications.
- Elsner, J., Elsner, S. and Jagger, T. (2014) The increasing efficiency of tornado days in the United States. *Climate Dynamics*, 45(3–4), 651–659.
- Emanuel, K. (2005) Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, 436, 686–688.
- Friedman, D. (1972) Insurance and the natural hazards. *ASTIN Bulletin*, 7, 4–58.
- Goldenberg, S., Landsea, C., Mestas-Nunez, A. and Gray, W. (2001) The recent increase in Atlantic hurricane activity: causes and implications. *Science*, 293, 474–479.
- Hall, T. and Jewson, S. (2007) Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus*, 59A, 486–498.
- Jagger, T. and Elsner, J. (2012) Hurricane clusters in the vicinity of Florida. *Journal of Applied Meteorology and Climatology*, 51, 869–877.
- Jewson, S., Bellone, E., Laepple, T., Nzerem, K., Khare, S., Lonfat, M., O'Shay, A., Penzer, J. and Coughlin, K. (2008) Five year prediction of the number of hurricanes that make United States landfall. In: Elsner, J., and Jagger, T. (Eds.) *Hurricanes and Climate Change*. Boston: Springer, pp. 73–99.
- Kaczmarek, J., Jewson, S. and Bellone, E. (2018) Quantifying the sources of simulation uncertainty in natural catastrophe models. *Stochastic Environmental Research and Risk Assessment*, 32(3), 591–605.
- Khare, S., Bonazzi, A., Mitás, C. and Jewson, S. (2015) Modelling clustering of natural hazard phenomena and the effect on re-insurance loss perspectives. *Natural Hazards and Earth System Sciences*, 15(6), 1357–1370.
- Khare, S., Bonazzi, A., West, N., Bellone, E. and Jewson, S. (2009) On the modelling of over-ocean hurricane surface winds and their uncertainty. *Quarterly Journal of the Royal Meteorological Society*, 135, 1350–1365.
- Kim, W., Yeager, S. and Danabasoglu, G. (2018) Key role of internal ocean dynamics in Atlantic multidecadal variability during the last half century. *Geophysical Research Letters*, 45, 13449–13457.
- Loridan, T., Khare, S., Scherer, E., Dixon, M. and Bellone, E. (2015) Parametric modeling of transitioning cyclone wind fields for risk assessment studies in the Western North Pacific. *Journal of Applied Meteorology and Climatology*, 54(3), 624–642.
- Mailier, P., Stephenson, D., Chris, F. and Kevin, H. (2006) Serial clustering of extratropical cyclones. *Monthly Weather Review*, 134, 2224–2240.
- Michel, G. (2018) *Risk Modelling for Hazards and Disasters*. Amsterdam: Elsevier.
- Mitchell-Wallace, K., Jones, M., Hillier, J. and Foote, M. (2017) *Natural Catastrophe Risk Management and Modelling*. London: Wiley Blackwell.

- Peings, Y. and Magnusdottir, G. (2014) Forcing of wintertime atmospheric circulation by the multidecadal fluctuations on the North Atlantic Ocean. *Environmental Research Letters*, 9(3), 1–8.
- Todkar, S. and Kass, R. (2010) Importance sampling: a review. *Wiley Interdisciplinary Reviews*, 2(1), 54–60.
- Villarini, G., Smith, J., Vitolo, R. and Stephenson, D. (2013) On the temporal clustering of US floods and its relationship to climate teleconnection patterns. *International Journal of Climatology*, 33, 629–640.
- Wang, S. (1998) Aggregation of correlated risk portfolios: models and algorithms. *Proceedings of the Casualty Actuarial Society*, 85(163).
- Wang, L., Ting, M. and Kushner, P. (2017) A robust empirical seasonal prediction of winter NAO and surface climate. *Scientific Reports*, 7 (279), 848–939.
- Wasserman, L. (2003) *All of Statistics*. Boston: Springer.
- Zanardo, S., Nicotina, L., Hilberts, A. and Jewson, S. (2019) Modulation of economic losses from European floods by the North Atlantic Oscillation. *Geophysical Research Letters*, 46(5), 2563–2572.

How to cite this article: Jewson S, Barnes C, Cusack S, Bellone E. Adjusting catastrophe model ensembles using importance sampling, with application to damage estimation for varying levels of hurricane activity. *Meteorol Appl.* 2020;27:e1839. <https://doi.org/10.1002/met.1839>