Identifying limitations when deriving probabilistic views of North Atlantic hurricane hazard from counterfactual ensemble NWP re-forecasts

Book or Report Section

Published Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Chapter 10
Identifying Limitations when Deriving Probabilistic Views of North Atlantic Hurricane Hazard from Counterfactual Ensemble NWP Re-forecasts

Tom J. Philp, Adrian J. Champion, Kevin I. Hodges, Catherine Pigott, Andrew MacFarlane, George Wragg, and Steve Zhao

Abstract Downward counterfactual analysis – or quantitatively estimating how our observed history could have been worse – is increasingly being used by the re/insurance industry to identify, quantify, and mitigate against as-yet-unrealised "grey-swan" catastrophic events. While useful for informing site-specific adaptation strategies, the extraction of probabilistic information remains intangible from such downside-only focused analytics. We hypothesise that combined upward and downward counterfactual analysis (i.e., how history could have been either better or worse) may allow us to obtain probabilistic information from counterfactual research if it can be applied objectively and without bias.

Here we test this concept of objective counterfactual analysis by investigating how initial-condition-driven track variability of events in our North Atlantic Hurricane (NAHU) record may affect present-day probabilistic views of US landfall risk. To do this, we create 10,000 counterfactual NAHU histories from NCEP GEFS v2 initial-condition ensemble reforecast data for the period 1985-2016 and compare the

T. J. Philp (✉)
Maximum Information, London, UK
e-mail: tom@maximuminformation.com

A. J. Champion
University of Exeter, Exeter, UK
e-mail: adrian.champion@aon.com

K. I. Hodges
National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK
e-mail: k.i.hodges@reading.ac.uk

C. Pigott · G. Wragg · S. Zhao
AXA XL, London, UK
e-mail: catherine.pigott@axaxl.com; george.wragg@axaxl.com; steve.zhao@axaxl.com

A. MacFarlane
AXA XL, Hamilton, Bermuda
e-mail: Andrew.macfarlane@axaxl.com

statistics of these counterfactual histories to a model-based version of our single observational history.

While the methodology presented herein attempts to produce the histories as objectively as possible, there is clear – and, ultimately, intuitively understandable – systematic underprediction of US NAHU landfall frequency in the counterfactual histories. This limits the ability to use the data in real-world applications at present. However, even with this systematic under-prediction, it is interesting to note both the magnitude of volatility and spatial variability in hurricane landfalls in single cities and wider regions along the US coastline, which speaks to the potential value of objective counterfactual analysis once methods have evolved.

Keywords Counterfactual · Ensemble · Landfall · Probabilistic · Risk

10.1   Introduction

For centuries, the global re/insurance industry has estimated future risk by applying statistical methods to historical loss data (Halley 1693), with the early methods growing into an expansive research discipline now known as actuarial science.

While the advent of catastrophe modelling in the late 1980s and early 1990s saw an evolution of traditional actuarial loss-based methods toward the incorporation of explicit scientific information from a wide variety of non-loss focused sources, probabilistic views of catastrophe risk have largely remained driven by historical observations.

In very recent years, methods to derive views of probabilistic atmospheric risk directly from climate models have begun to be developed (Carozza and Boudreault 2021; Jones et al. 2020). However, it is well known that ingesting climate model information coherently into the historical, observation-based, and usually site-specific, views of risk that are widely prevalent in the industry is likely to bring its own suite of problems, both scientific and philosophical (Frigg et al. 2015).

Although extremely non-trivial, there is substantial demand to overcome this challenge given increasing societal concern surrounding the need to quantify the impact of a shifting climate on the frequency and intensity of catastrophic atmospheric perils, from both acute and chronic onset perspectives. This demand is evidenced, for example, by recent climate disclosure guidance issued by the Bank of England to financial institutions (Bank of England 2019).

It is important to note that although the challenge of building the connection between forward-looking climate and catastrophe models is yet to be fully overcome, the re/insurance industry already looks to alternative methods to help improve its ability to use information about extreme weather events from climate projections as coherently and appropriately as possible. For example, for many years the industry has employed the concept of "downward counterfactual" thinking to sensitivity test observationally derived views of risk; the same type of thinking exists in other financial markets, usually under the auspice of more generically-termed

"financial stress tests". As this downward counterfactual term is not yet widely used in catastrophe risk management, we explicitly define it here as "a thought about the past where the outcome was worse than what actually happened", with the definition taken directly from the introductory discussion provided in Woo (2019), itself following the formative definition provided by Roese (1997).

While this language is relatively new to the risk management industry, methods to incorporate the underlying ideas in risk modelling are not. The industry understands well that – for rare and high impact events in particular – historical observations may only paint a partial picture of future risk. Thus, deterministic "what-if?" scenario analyses have been employed, which ultimately attempt to foresee those "grey swans" that do not appear in our observed loss record, yet can easily be imagined, rationally estimated, and thus mitigated against.

At present, this "downward counterfactual" or "what-if" thinking is applied through such practical actions as formal reporting on Realistic Disaster Scenarios (RDSs), which help to steer re/insurers on capital requirements, as well as quantifying estimates such as maximum possible/probable loss in the case of an extreme event happening. As an example, in the context of North Atlantic Hurricanes (NAHUs), one of the Lloyd's RDSs for 2021 poses the scenario: "A North-East US hurricane, immediately followed by a South Carolina hurricane" (Lloyd's 2021).

While already embedded in risk management processes, in a joint report on the topic, Lloyd's and RMS concluded that including more "downward counterfactuals" in analyses of risk – that is, specifically re-imagining "how historical near misses might have become major disasters" – is likely to "bring benefits to insurers" (Lloyd's 2017). For example, what if Hurricanes Matthew (2016) or Dorian (2019) had rolled onshore in downtown Miami, as opposed to only grazing the South Florida coastline? Would the building codes and practices put in place since the highly destructive Hurricane Andrew (1992) work to limit outsized losses in this region? And how have population dynamics in the past twenty years altered the shape of risk in that area? While the Lloyd's RDSs begin to touch on questions like this, they are often less focused on re-imagining actual historical events. Other industry-led research is beginning to unpick the potential value of this type of targeted question (Chap. 9).

Thus, while more focus on downward counterfactual analysis is certainly likely to be useful, it is not necessarily novel. Arguably, it relegates the potential value of the work to deterministic applications only and doesn't help to address the problem of extracting potentially valuable and untapped probabilistic information from such types of analyses.

At present, probabilistic information for catastrophe risk quantification in the re/insurance industry is driven by stochastic modelling, itself underpinned almost solely by historical event data. The stochastic modelling process acts to fill in holes such that our spatial picture of risk is smoother (and more realistic) than if drawn from raw observations. However, the statistics of the historical dataset are preserved during the stochastic modelling process, at least to some extent. While this preservation of underlying statistics may be highly desirable for well-observed perils and long historical records, it is likely less desirable in places with sparse observational

datasets (as is the case with many extreme weather perils). The preservation of statistics in these sparsely observed areas may lead to an erroneous view of risk purely by virtue of random chance or luck in the observed historical period. Thus, identifying further information sources that can help to remove random luck in risk quantification and to ultimately facilitate more optimized risk selection may prove highly valuable. It is hypothesized here that counterfactual analysis may provide an opportunity to build additional reliability on baseline stochastic modelling.

For probabilistic counterfactual applications to be achieved, we must move away from the single-sided "downward" philosophy of counterfactual analysis, toward one where both "upward" and "downward" counterfactuals are considered simultaneously. If it were possible that this upward and downward analysis could be done objectively, we may be able to extract information such that our stochastic, observation-based methods for risk quantification become more robust and reliable. At present, questions about validity of stochastic modelling output can only be asked from a subjective "model completeness" perspective – i.e., do we believe that the stochastic process is filling in the gaps in history appropriately?

To practically combine the two worlds of future focused climate-catastrophe modelling with objective upward and downward counterfactual analysis, here we propose to use NWP ensemble re-forecasts to create multiple counterfactual NAHU histories, and to compare them to the observed historical record of NAHUs. At present, the authors are only aware of one other study that attempts to extract probabilistic information from re-forecast data in the case of Tropical Cyclones (Ng and Leckebusch 2021). This study, while similar in fundamental philosophy, is somewhat tangential to the applications hypothesized here. Ng and Leckebusch (2021) utilize a multi-model archive to create basin-wide counterfactual climatologies of key catastrophe risk related variables, such as return periods of windspeeds. Here we intend to accomplish something slightly different, focusing only on identifying historical NAHUs in re-forecast data and attempting to use these data to evaluate – however loosely – the probabilities of intersection with the US coastline of these historical storms. If the methods were to prove successful, it would allow targeted adjustment of the stochastic track sets that typically drive contemporary NAHU catastrophe models.

Thus, beyond the counterfactual aspect of the analysis, we also hope this work will help to lay the foundation to connect traditional historical observation derived views of catastrophe risk to weather and climate model derived views of risk, and highlight where limitations currently exist that may limit the ability to do this optimally.

10.2   Potential Real-World Applications

While beyond the scope of this experimental study, it is envisaged that a truly objective implementation of counterfactual thinking would allow us to quantifiably deconstruct our observational history, and begin to ask questions such as:

1. Do counterfactual histories suggest that a specific country or city has been particularly lucky or unlucky given our single historical record of landfalling hurricanes, in a way that the hole-filling of traditional stochastic catastrophe modelling may not reveal?
2. Was any single highly impactful historical event extremely unusual, even after the stochastic process attempted to fill in gaps, or do we see these events regularly in our NWP modelling?
3. Were highly active years such as 2005 and 2020 extreme outliers, as a traditional observation-derived stochastic view would suggest, or are they more common than they appear in our historical record?
4. Were particularly active/inactive periods in our history (e.g., Hall and Hereid, 2015) random or structural in a way that dynamical climate and NWP models may reveal? Relatedly, are there correlations in active/inactive seasons following one another, and is the assumption that individual hurricane seasons are independent, as is often derived during stochastic modelling of NAHUs, a good one?
5. Considering the US coastline in aggregate, where does our observed history sit in the distribution of possible outcomes, and what does the relative uncertainty between stochastic modelling and counterfactual analysis look like?

Some of the above questions can be seen to fall under the more general area of climate attribution research; it is important to note that it is necessary to use counterfactual analysis in the climate attribution process to allow questions on non-events to be asked, and thus for probabilistic views on events to be gleaned (van Oldenborgh et al. 2021). For example, without explicit counterfactual data, it is possible to ask "was the 2012 Superstorm Sandy event made more likely by climate change?", but it is impossible to ask the philosophical reverse of that question - i.e., "was the non-occurrence of event X made more likely by climate change?" – because there is no concrete event to begin to quantify changes from.

Further than facilitating the derivation of probabilistic information from climate attribution research, the first four examples in the list above speak to investigating the potential to "hedge" risk – with (1) and (2) focused on spatial hedging, and (3) and (4) focused on hedging of temporal aspects, or certain parts of an Exceedance-Probability (EP) curve.

The spatial distribution of risk is an important business consideration when a (re)insurer constructs its portfolio geographically. Understanding a portfolio's areas of high or low levels of accumulation are key to determining whether to take on or hedge additional risk in a specific geographic area, be that a city, county, or state. The assessment of portfolio shape is typically determined using stochastic catastrophe models and calculating key statistics such as the average annual loss (mean expected loss) and a range of return periods for the aggregate loss distribution. Deterministic scenarios, such as the RDSs mentioned previously, also provide context against (re)insurers' risk appetite statements regarding the maximum downside they wish to be exposed to under given conditions, but are limited in the sense that they fail to give statistics relating to the loss distribution. Identification of bias or error within the stochastic tools used to make this assessment is important in

determining the underwriting strategy and capital management and influences decisions regarding geographic portfolio mix, technical pricing assessment and any mechanisms used to mitigate risk.

(Re)insurers will typically mitigate their gross risk and manage their capital with the use of ceded reinsurance protections. Various structures and vehicles can be used to take the risk carried by the original insurer and pass it further up the risk chain to another party in exchange for a premium. Some of these transactions focus on the severity of potential losses, looking to cap the downside from a single event. In this case, understanding the probability of large events occurring is crucial. Whether a particular geography, or indeed an area in aggregate, has been "lucky" or "unlucky" historically feeds directly into the real-world decision regarding how best to manage capital through risk reduction; any objective information that can help better inform this concept of historical luck would be greatly welcomed.

In addition to spatial conditions, the frequency distribution is another important parameter for managing capital in the form of ceded reinsurance. An example of this in the context of a core business decision is deciding when to buy cover that will pay out dependent on the number of landfalling storms in a hurricane season. There is a cost and benefit to the number selected, and determining the preferred choice relies on assessing the probability of these protections being required. Hence, the application of counterfactual analysis in this area could have very tangible use in the (re)insurance market.

Finally, this type of analysis may seek to present opportunities to (re)insurers to take on risk that – when looked at with traditional stochastic methods – appears not to fit within their target risk profile, due to price or aggregation of risk in a particular area. In turn, the analysis could help to benefit consumers of (re)insurance in areas that have been "unlucky" in the past, by allowing risk to be reassessed and viewed with a counterfactual perspective.

10.3   Methods & Data

Ten thousand counterfactual NAHU histories are generated from ensemble reforecast NWP data for the period 1985–2016, along with one "best-estimate" reforecast history that represents our historical baseline, from which to draw fair comparisons. This section details the data selection and processing methods.

To circumvent any issues regarding tropical cyclogenesis biases in NWP output (e.g., Halperin et al. 2013), the selection of the counterfactual ensemble data is restricted to finding alternative tracks only of the NAHUs that are recorded in our observational history. This immediately imposes a limitation on the output – and is a key divergent point from the setup of Ng and Leckebusch (2021) – as we will not capture certain types of counterfactual realities of NAHUs. For example, we are unable to capture NAHUs that didn't form in our observed reality but may have done in a counterfactual one. Therefore, each one of the counterfactual histories will always contain the same number of tracks as has been observed in our history.

Ultimately, this means that the study is restricted to looking at track uncertainty of historical storms only, as opposed to a more complete track uncertainty coupled with genesis uncertainty. This means that, for the time being, the results would be unable to unpick some of the broader questions presented earlier. The authors acknowledge that this is likely to severely limit the real-world application of the data at present, and stress that the results should be seen as experimental at this point. However, given that the study is exploratory, and primarily aims to uncover potential limitations with this type of application, we feel that adding in the cyclogenesis aspect at this point has the potential to add a needless layer of complexity to experimental methods and results.

10.3.1   Data Selection

The counterfactual histories are generated from NWP reforecast data, as opposed to historical operational NWP forecast ensemble data, because a reforecast allows for a consistent dataset whilst also using a contemporary NWP model across the entire historical period of the study.

The NOAA PSL Global Ensemble Forecast System (GEFS) reforecast v2 (Hamill et al. 2013) is utilized as the reference reforecast dataset. This initializes daily (00z timestep) an 11-member (1 control + 10 initial condition ensemble members) forecast, with a running period of 16 days, approximately 40–54-km spatial horizontal resolution, and 6-hourly temporal resolution. The data used in the study runs from 1985 to 2016.

Given the need to match to historically observed NAHUs, two more datasets are employed, namely:

 i. the observation-based International Best Track Archive for Climate Stewardship (IBTrACS), which enables the identification of historical NAHUs and their real-world intensities.
ii. the National Center for Environmental Protection's Climate Forecast System Reanalysis (NCEP-CFSR) (Saha et al. 2010), for allowing track matching between the model world of the NWP reforecasts and the observation world of the IBTrACS data.

Importantly, both datasets are available for the entire period of the GEFS reforecast data.

10.3.2   Tracking and Storm Matching Part 1: Reanalysis to Observations

Tropical Cyclones are identified and tracked in the NCEP-CFSR reanalysis using the TRACK algorithm (Hodges 1994, 1995; Hoskins and Hodges 2002). Initially all

cyclonic systems are tracked in the NCEP-CFSR and the tracks are then matched to, and filtered by, the IBTrACS tracks using mean separation matching (Hodges et al. 2017). Any amount of temporal overlap and a mean separation distance of 5 degrees (geodesic) for the overlap periods causes a track match.

10.3.3 Tracking and Storm Matching Part 2: Reforecast to Matched Reanalysis

The TRACK algorithm is then applied to the GEFS reforecast. These tracks are subsequently matched to the previously matched and filtered reanalysis tracks using mean separation matching: the reforecast tracks are matched to the reanalysis tracks using the first day of the forecast track that overlaps with the analysis track to within a 4-degree (geodesic) radius, and the reforecast tracks have their first point within the first 3 days of the forecast (Hodges and Klingaman 2019; Froude et al. 2007).

   This allows for the TCs to be found in the NWP reforecast data before they are identified in the observations. The resultant combined data files contain up to 12 tracks per historical track per day (due to the daily GEFS initialization) – one reanalysis track of a historical storm, one control track of that storm from the GEFS reforecast, and up to 10 GEFS initial condition ensemble members (the number of GEFS ensemble members being dependent on whether the perturbed ensemble members continue to develop the storm or not).

10.3.4 Tracking and Storm Matching Part 3: Reforecast Tracks to Observational Tracks

A final matching and filtering step of the reanalysis and reforecast tracks to the IBTrACS data is undertaken to confirm the tracking and matching process has been successful, and to ensure that a historical hurricane "name" is attached to the reforecast and reanalysis track data. First, a match occurs if there is at least one timestamp that the NCEP-CFSR and the IBTrACS track are within 1 degree (geodesic) of each other. The variable used from the NCEP CFSR tracks to define the center of the storm is the latitude & longitude of the maximum 850hPa vorticity center. Using this method across the entire study period, there are ten unmatched IBTrACS storms with the criteria at 1 degree. The spatial-matching region is therefore relaxed successively: five more NAHUs are matched within 2 degrees, and two more are matched within 5 degrees. Three storms remain unmatched, namely Matthew (2004), Zeta (2005), and Barbara (2013). These storms are thus, in effect, filtered from the historical dataset, both for our "observational" model history, and also for our alternative history creation.

At this step in the process the combined track files contain up to 12 tracks per historical storm and, because of the daily initialization of the 11-member GEFS reforecast, there is a new combined track file created each day that an observational track exists. For example, in our observed record, Hurricane Andrew (1992) formed on August 16th, 1992, and dissipated on August 29th, 1992. Thus, with 14 days of existence, and with a maximum of 12 storm tracks per day, we have a maximum of 168 tracks across the 14 combined files for Hurricane Andrew. However, the reanalysis track for a single storm will be identical in each of the daily files for a single historical storm, as it is merely a reference track from a single model run that will have been truncated to start on the date of the GEFS initialization. Thus, the number of different GEFS tracks that are theoretically available for selection into the counterfactual histories is 11 multiplied by the number of days a single storm is active. In the case of Hurricane Andrew, this would be a maximum of 154 distinct tracks that could be selected from for addition into the counterfactual histories.

10.3.5 Creation of Extended Landmasses for Track Selection into Counterfactual Histories

There are many potential ways to construct the counterfactual histories from the track files, and it is at minimum difficult, but arguably impossible, to completely remove all levels of subjectivity from this process. While it would be possible to collate all of the daily GEFS tracks for a single storm and simply randomly sample from them, we realize that this may introduce structural issues. For example, it is likely that sampling for Hurricane Andrew (1992) from GEFS data initialized at a point shortly before US landfall, versus Hurricane Katrina (2005) initialized in the Atlantic basin's Main Development Region, would introduce structural biases. Conversely, if generating counterfactual histories by selecting an ensemble member at only the time and date that the storm first appears in the IBTrACS observed data, we limit ourselves to very little data (i.e., we remove the potential to use all of the ensemble reforecasts created after the start date of the storm in the record, and thus only have the ability to select from a maximum of 11 GEFS versions of Andrew), while at the same time we may also introduce steering biases that are present in the more cyclogenesis prone regions (e.g., Main Development Region) of the Atlantic. And further, if we introduced a single "spatial barrier" that the storm would have to cross for it to be included in selection (for example, if we stated that the storm would have to cross the 55th Meridian West for it to be included in the sample selection), we would still be constrained both by limiting data and by introducing a potentially difficult to untangle structural issue.

To attempt to combat any region-specific model bias issues, utilize as much data as possible, and have the ability to piece together any unforeseen structural issues, we introduce a novel methodology in which multiple theoretical "extended" land-masses are generated at various distances from the US coastline. Once a historical storm crosses the line of the theoretical extended landmass, the GEFS tracks for that

storm will be available for an ensemble member selection to create the alternative histories. For example, on the date that Hurricane Andrew crosses the 300-km extended landmass in the observational data, the GEFS reforecast data initialized on that day becomes available for selection into the counterfactual histories. While this means that we will only retain the storms that have occurred in our history, it's important to note that this method still allows for large divergence of the tracks – for example, because the GEFS reforecast data is free-running and not constrained by observations, some or all of the Hurricane Andrew tracks in the forecast may not make actual landfall in the US. It is this aspect of the analysis that we hope will begin to allow us to probabilistically re-evaluate our history, and even probabilistically re-evaluate specific historical events.

Ideally the number of landmasses should be so numerous as to use as much of the ensemble data as possible, without being so exhaustive as to cause overly cumbersome re-selection of data that has already been used. Four different "extended landmasses" are created using QGIS and converted to a grid with a resolution of 0.1 degrees. These landmasses are generated at 300-km intervals at distances between 400-km and 1300-km from the North American coastline. While the choice of extended landmass distances will always carry some level of arbitrariness, they are here chosen given knowledge about NAHU translation speeds and daily initialization limitation of the GEFS data. With a mean NAHU translation speed between 18 and 25-km/h (Kim et al. 2020), it is likely that the average NAHU will travel approximately 432–600 km distance in 24 h, which thus represents the maximum distance that would be reasonable to employ between extended landmasses. Given the further reality that NAHUs neither travel perpendicularly to extended landmass contours, nor in straight lines, the distance between extended landmasses is reduced to 300 km.

10.3.6 Counterfactual History Creation

2,500 histories per extended landmass are created, producing a total of 10,000 counterfactual histories. With the period of analysis, this creates 320,000 years (or NAHU seasons) of data. It is important to note that, because of the finite number of GEFS ensemble storms per observed historical event, generating this many histories is likely to cause multiple selections of the same ensemble storm on occasion. Thus, the 2500 histories at each extended landmass cannot be considered entirely independent of one another. However, the method remains desirable for risk management because:

 i. this re-selection of the same ensemble member multiple times potentially better allows the "worst-case scenario" of what the continuous chain of the most deleterious events in a single season could have been.
ii. The method produces 10,000 versions of each individual historical year. This is a somewhat standard number for the minimum number of years desirable for a

catastrophe model stochastic set (Jewson et al. 2019), and allows us to delve deeply into key loss years, such as 2005, while retaining probabilistic rigor.

To create the histories, the reanalysis tracks that make landfall with respect to the theoretical extended landmass (i.e., by crossing the imaginary line of the extended landmass) and have been successfully matched by name to a historical IBTrACS Tropical Cyclone are identified. Tracks that form whilst already over the extended landmass are also included so as not to impose a filtering of the storm number by virtue of their point of genesis. Thus, the date on which they cross the line of, or first form on, the extended landmass is used for ensemble selection. For example, for the 1300 km landmass, Hurricane Andrew's (1992) GEFS ensemble member will be selected for the initialization date that it crosses the line of said extended landmass. Hurricane Wilma (2005), however, formed in the Caribbean Sea, and thus technically never crosses the 1300 km landmass line because it already forms *on* the extended landmass. Therefore, Wilma (2005) is kept in the 1300 km landmass selection on the date that it forms.

Thus, an ensemble track is randomly selected from the GEFS reforecast data for each storm name on the date that it forms (if that formation point exists on the extended landmass), or on the date that it first crosses the line of the extended landmass.

A further filtering of the data occurs at this point: only an initial condition ensemble member (i.e., not the GEFS control member) can be selected for inclusion in the counterfactual histories. This decision was taken to remove the concept that some member selections may have been "better-estimate" (in the case of picking the control vs an initial condition member), and thus have introduced a probabilistic bias for some alternative history tracks.

10.3.7 GEFS Based Observational History

For comparison of the counterfactual histories to "reality", it is obvious that a direct evaluation between the counterfactual alternative histories and the IBTrACS data would be unfair; the limitations imposed by resolution and, relatedly, incomplete physics will make the GEFS model NAHUs, both in terms of track and intensity, look different from the observational IBTrACS history. Thus, the differences between the reforecast and observational data are minimized by creating a GEFS reforecast model view of our observed reality from which to make these comparisons.

However, this again is not a trivial task. The reforecast data are not constrained by observations while the forecasts are running, and thus the tracks of the GEFS data are likely to vary from the orientation of both the tracks seen in the observational history and in the reanalysis. We therefore use a "0km" landmass – in effect, just the US coastline – to generate the GEFS model-based history at as close to a timestep as possible from the GEFS initialization, and we only take the control run (i.e., the

unperturbed model run, which in this instance could be considered to be the model best estimate) from the reforecast for this landmass. This is because, with the GEFS model initialized so close to land (i.e., at the timestep before it makes landfall), the GEFS model does not have a chance to materially impact the track or intensity of the storm. Thus, we would likely end up with a cluster of events at landfall in most situations. Further, this 0-km data is only intended to act as the benchmark data from which to understand relativities in the counterfactual histories, so we only need a single best estimate from which to do this.

While the creation of the 0-km benchmark GEFS data means that IBTrACS-derived observational US landfall locations should match fairly well with the 0-km reforecast data, the storm could still be some way away from landfall because of the timestep limitation enforced by the reforecast data. A cubic-spline interpolation is thus applied to the GEFS and IBTrACS data to up-sample the tracks to 15-minute temporal resolution, which allows for close temporal matching of the two datasets at the precise point of landfall. This same interpolation is later applied to landfalling storms in the alternative histories to allow for fair comparisons. The choice of cubic-spline interpolation here follows similar temporal resampling studies (e.g., Baudouin et al. 2019).

10.3.8   Intensity Downscaling

As mentioned previously, resolution and incomplete model physics data mean that windspeeds are likely to be systematically different between the model data and the IBTrACS observational data. While the potential issue is analytically negated by creating a model-based observational history from which to draw direct comparisons, reporting coherently on impact-based narratives for risk-focused communities is difficult to accomplish without attachment to easily understood intensity metrics, such as the Saffir-Simpson scale.

Thus, a simple statistical downscaling is applied to the histories to bias-correct them toward the usually higher observational intensities. The mean windspeed of all category 1+ hurricanes at landfall in the IBTrACS data is calculated. Using the names as matched in Sect. 10.3.4, the same hurricanes are then extracted from the control run of GEFS from the 0-km landmass. The hurricanes in GEFS are then interpolated to match the timing of the landfall in IBTrACS. The mean windspeed of these GEFS landfalls is then calculated. The percentage difference between the two is calculated, and this single factor scaling uplift is then applied to all GEFS data. The authors acknowledge that this is an overly simplistic method for generating accurate intensities across all Saffir-Simpson categories, but we purposefully keep the method simplistic here so that we can better focus on questions relating to the counterfactual analysis methods.

In Fig. 10.1, the red bars show the counts of US landfalls per Saffir-Simpson category from the raw GEFS 0-km history, while the blue bars show the counts per
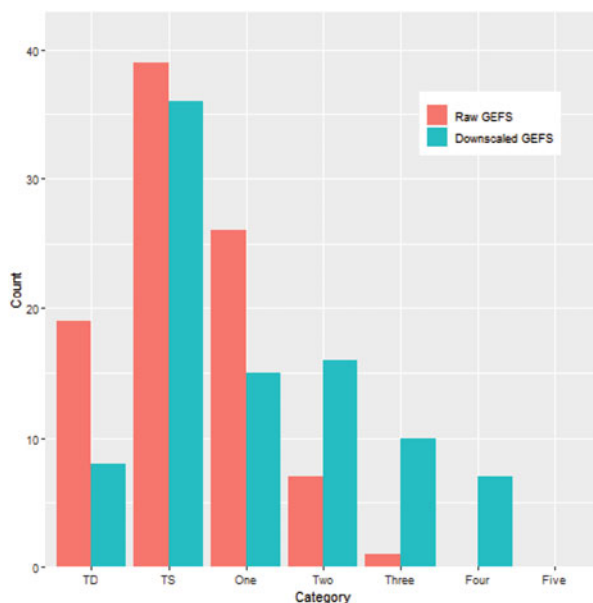
category after the downscaling has been applied. While crude, the overall results and narratives are unlikely to be negatively impacted given that we are primarily looking at the relative analytics of model-derived data for both our observational history and our alternative history.

10.4   Results

10.4.1   GEFS vs IBTrACS Observational History Differences

Before analyzing the differences between the statistics of the counterfactual histories and the GEFS-derived "observational" history, it is also important to note that the matching of these two landfalling data are not perfect. For example, 10 Tropical Cyclones that appeared as being US landfalling in the GEFS data did not have a corresponding landfall in the IBTrACS data. Upon unpicking, there were two key reasons for the mismatches:

1. IBTrACS has a human element to the track distance recorded. For example, it is up to a human forecaster to decide when a Tropical Cyclone has come into being, and when it has dissipated, with the human forecaster usually having to decide during an ongoing live event (though this can be corrected/adjusted later – but with no less subjectivity). Conversely, the GEFS tracks are constrained by the

objective nature of the TRACK algorithm, which defines the center of the storm as the local vorticity maximum.

2. The eye of a hurricane coming very close to land and only crossing the coastline in one of the datasets (e.g., Tropical Storm Cristobal, 2008). The relative diameter of the eye in the different datasets may add complicating consequences here that are not investigated in the analysis.

While this does not represent an issue given our GEFS model created history based on observations, being constrained by the same criteria as the alternative histories, it is important for us to stress again that the results presented here should not be compared directly to raw IBTrACS data.

Figure 10.2 shows how the intensities of the matched and downscaled storms in GEFS compare to their equivalents in IBTrACS. It can be seen that while the downscaling has uplifted the GEFS intensities to being comparable to the IBTrACS data, there are clear differences, such as the IBTrACS data having a skewed distribution with a heavier tail than the downscaled GEFS data. IBTrACS consequently sees more Tropical Storms and category 1 hurricanes, as well as more category 5s. Given the simplicity of the downscaling, the effects of which can be seen in these results, it was decided that only hurricane intensity storms would be analyzed. No further sub-division by intensity (e.g., between minor and major hurricane) is made in any of the analyses.

Fig. 10.2 Comparison of Counts, split by Saffir-Simpson Category, of Matched NAHU US Landfalls (0-km landmass) in the IBTrACS data and the Downscaled GEFS data. It can be seen that while overall counts per category are somewhat comparable, the statistical downscaling is not perfect, and leads to a less peaked + quicker decaying distribution than has been observed. For the relativity comparison of this study however, it is envisaged that this will not have significant impacts
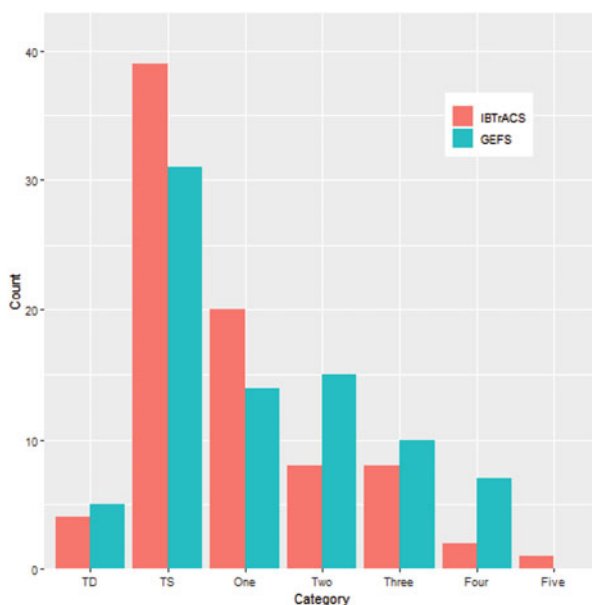
10.4.2   Extended Landmass Histories: US NAHU Landfalls, All Categories

Table 10.1 shows the counts of NAHU landfalls in the US from the 0-km landmass data, and the mean number of landfalls per history per extended landmass distance. Figure 10.3 shows extended landmass histograms for the frequency of alternative histories split by the number of US landfalling storms.

What is immediately obvious from the table is that the mean landfalls for each of the extended landmasses are between approximately 41–52% of the observed (0-km landmass distance) landfall number. Secondly, the table and histogram plots show a consistent drop in US landfalling rates as the extended landmass is pushed further and further away from the US coastline. At 400 km, the mean number of US landfalling hurricanes in the histories is 28.5, and this drops away gradually until the mean of 22.8 at 1300 km. The range of the data is quite stable throughout, usually approximately +/− 8 hurricanes. It can therefore be said that this is a systematic effect that occurs consistently across extended landmass distances.

While finding such a systematic effect was an unexpected result, it is an intuitive one. On average, as one moves further from a coastline, the probability of a hurricane making landfall decreases (Brettschneider 2008). Thus, this is purely a probabilistic artefact that is imposed by the ensemble member selection from the extended landmass generation method. This has wide-ranging impacts on reliable information and potential conclusions that can be drawn from the results. Further discussion of this is therefore picked up in Sect. 10.5.

10.4.3   City Specific Investigation

The box and whisker plots in Fig. 10.4 show the number of hurricanes making landfall in the metropolitan areas of the cities Houston, Miami, New Orleans, New York City and Tampa over the historical period used in the study for all extended landmass distances aggregated together. While this figure was intended to facilitate the unpicking of the relative risk of the different cities – both from a mean activity and an extreme activity perspective – it is clear from the histograms in Fig. 10.4 that probabilistic impacts of the extended landmass selection method will be distorting these results. For example, Miami and New York City, both being

Table 10.1 No. of US NAHU landfalls (Cat 1–5) per landmass distance from the GEFS histories. It should be noted that the 0-km landmass is the actual US coastline and is our GEFS based version of the observational history. It is therefore a single history and not a mean count. The 400, 700, 1000 and 1300 km landmasses are all mean numbers of US NAHU landfalls generated from the counterfactual GEFS histories at each of these distances

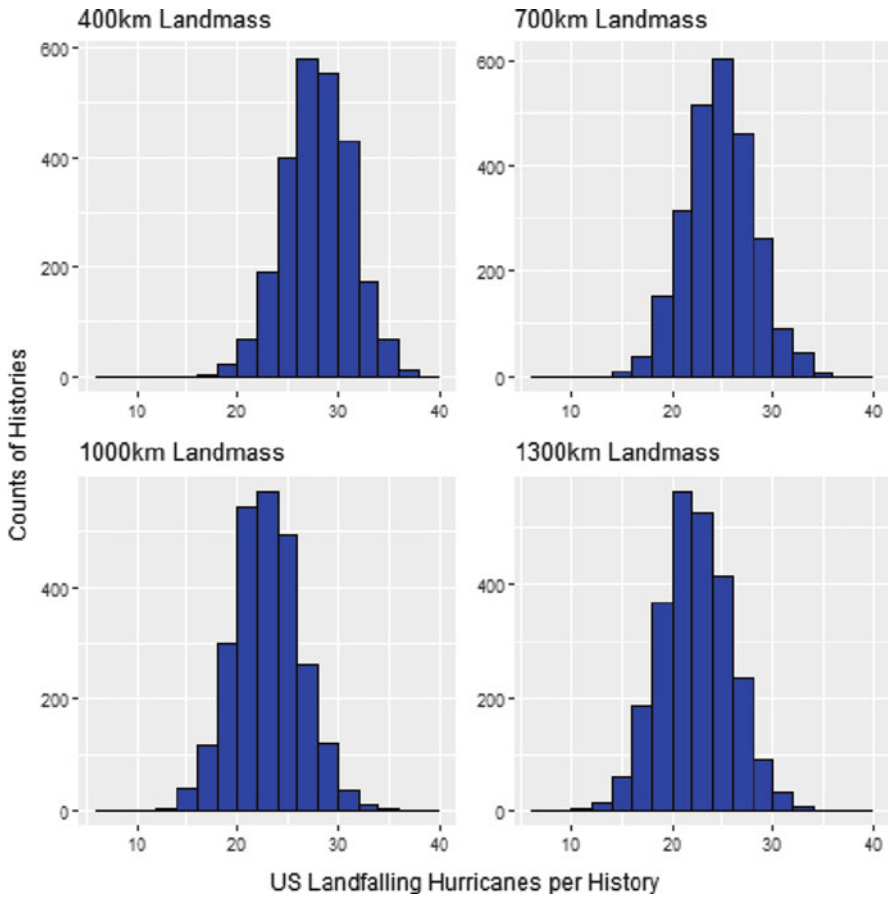| Landmass distance	0-km	400-km	700-km	1000-km	1300-km
Hurricane count	48	28.5	25.2	23.4	22.8

Fig. 10.3 Histograms for the counts of US hurricane landfalls in each of the alternative histories for the period 1985–2016, split by distance of the extended landmasses (top left: 400-km extended landmass, top right: 700-km extended landmass, bottom left: 1000-km extended landmass, bottom right: 1300-km extended landmass). The 0-km history is not shown because it is a single number, as opposed to a distribution

situated on the East Coast of the US, are much closer to the extended landmass ensemble selection lines than the cities of Houston and New Orleans in the Gulf of Mexico. Thus, there are likely to be probabilistic impacts that weight the east coast cities to seeing more landfalls in the alternative histories than the Gulf coast cities. Further, it is likely that there will be impacts even between the two east coast cities.

There are also likely region-specific issues with the simplistic downscaling when it comes to the North-East region. The sub/post/extra-tropical structure of the cyclones that make landfall here tend to have a broader wind field than their tropical counterparts, which weather forecast models more adequately capture the upper bounds of intensities of (Hodges & Emerton, 2015). Thus, the number of hurricanes in these regions may be over-inflated compared to the cities in the more tropical regions.
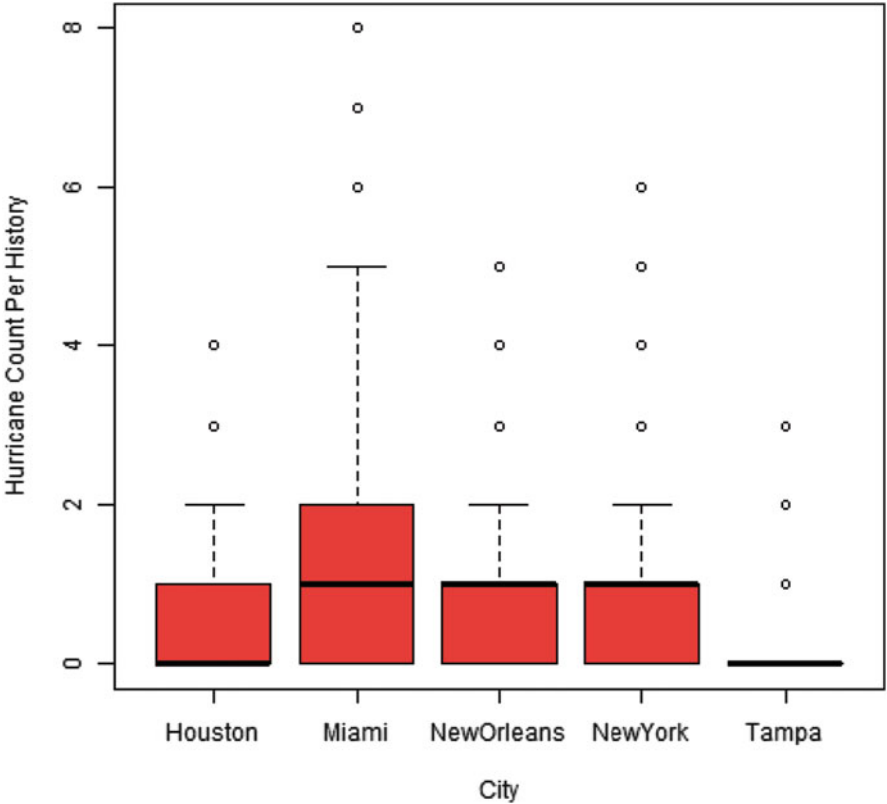
Fig. 10.4 Box and whisker plots for counts of hurricanes landfalling in the cities of Houston, Miami, New Orleans, New York City, and Tampa from the counterfactual GEFS histories. Data has been aggregated across all extended landmasses

We therefore encourage the cities in this part of the analysis to be viewed from a single-city variability perspective, as it is hypothesized that this aspect of the results still has the potential to be informative, if only from a fairly subjective basis. For example, it is very intriguing to notice that there seems to have been the potential for Miami to have been struck by up to 8 hurricanes in the historical period 1985–2016 – and this is for histories constructed from data that could be considered probabilistically weighted low, given what we can see for the numbers of landfalls from the counter-factual extended landmass histories relative to the observational history. The historical number of hurricanes observed over this period landfalling in Miami was just one.

10.4.4   Gate-Rate Maps

A key reason for stochastic modelling of Tropical Cyclone tracks is to appropriately fill gaps and extend the variability of our observational history. This type of

variability is highlighted in the landfalling data that can be seen in Fig. 10.5, with various hotspots of landfalling activity in stochastic TC event sets likely driven by events that have occurred in a short historical period. While the stochastic track modelling process does help to overcome this hotspot issue, historical statistics are still preserved – at least to some extent – by the modelling process, and so it is always unknown if the issue is fully overcome. It is hypothesized that the counterfactual modelling process introduced here could help to unpick this issue because it offers the potential to include an independent data source in identification of potential hotspots. However, before this could be achieved, the probabilistic artefact from the extended landmass method would need to be overcome. At present, we know that the extended landmass selection method introduces a low bias toward landfalling US NAHUs.

To attempt to investigate whether the alternative histories can help to glean a better picture on local variability, Fig. 10.6 shows hurricane frequency "gate-rate" maps for the US that are generated by the alternative histories aggregated together. This type of gate-rate map is a relatively standard output in evaluations of catastrophe model output. While difficult to glean anything concrete in this instance, it is interesting to note that parts of south Florida and parts of the Gulf of Mexico look comparatively much higher hazard than some of their immediate adjoining gates. Additionally, when comparing back to the observed landfalls in Fig. 10.5, the peak landfall gates in Fig. 10.6 are often slightly displaced from the peak regions of observed landfalls. For example, the highest landfalling gate in the Gulf of Mexico would likely be centered around New Orleans if derived from observations, but from
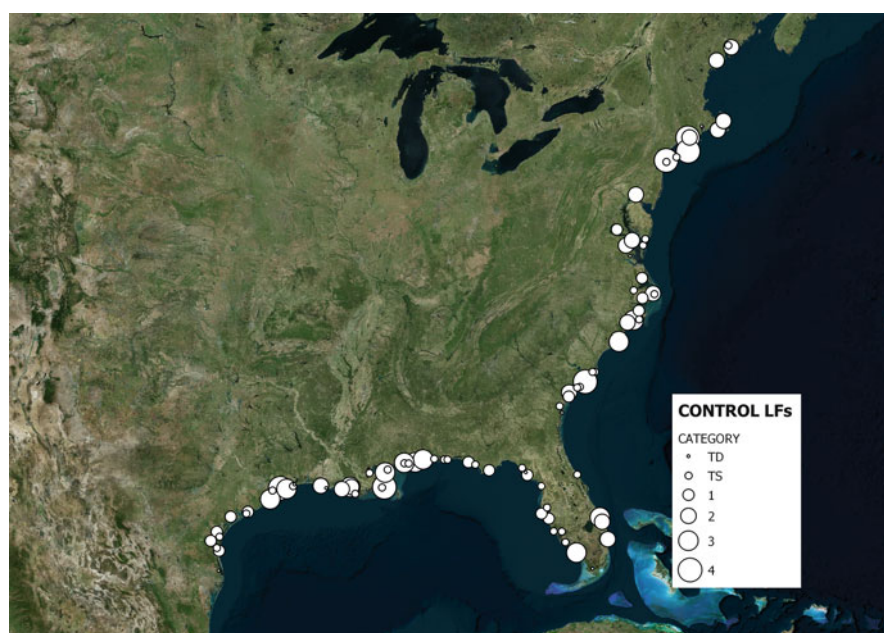

Fig. 10.5 Observed hurricane landfalls from the 0-km history (i.e., the best-estimate GEFS model-based observed history) for the period 1985–2016
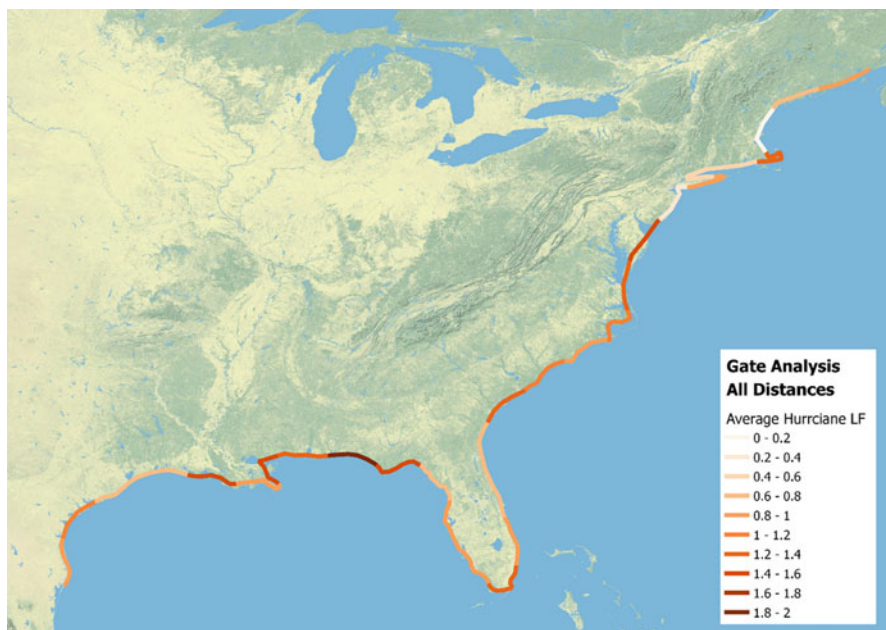
Fig. 10.6 Average number of NAHU landfalls per gate per counterfactual history for all extended landmasses

the alternative histories seems to be shifted further east toward the Florida panhandle. Similarly, from observations, the highest landfalling region in Florida looks like it would be on the mid-east coast of the state, while the alternative histories would suggest the highest hazard gate is the southern-most gate.

While it is difficult to know whether these conclusions hold, they raise important questions about how much trust should be put in probabilities derived from our single observed history, and thus in the stochastically derived risk estimates. This was the premise from which we hypothesized that these analyses could have real-world value. However, we have to stress at this point that the results should not be used practically given the biases apparent from the sampling issues driven by the extended landmass track selection process. Having said this, we believe that, from the foundational work presented herein, if the biases can be overcome there is large potential value in the use of these data from which to build probabilistic views of risk that are independent from traditional stochastic modelling techniques.

10.5   Discussion

Most importantly, there is a simple probabilistic US landfall artefact that falls out of the alternative history building from the method employed herein. As we move further and further from land, there is, on average, a lower and lower chance of a

storm making landfall. For example, if we were to select a random storm near the Caribbean, it might have a 60% chance of US landfall, but if we selected a storm off the coast of Africa, it might have a 30% chance of making landfall. Thus, when adding ensemble members to the counterfactual histories from further and further away, the probability of US landfall, on average, is weighted further and further down. It is therefore currently impossible to use the data to compare to our 0-km data and objectively make statements such as "across the entire US, we have been unlucky/lucky in our observational history". The probabilistic artefact imposes an inability to reliably infer spatial information across the US coastline from this analysis.

A possible evolution of the method to overcome this issue could be to construct extended landmasses that have a consistent objective probability of US landfall, and to attempt to normalize the event rates of the ensemble history at a single "probability" distance to the US, as opposed to an arbitrary geographical/spatial distance. However, the spatial pattern of probabilistic landfalls is neither simple nor static in time, and it itself would have to be driven by historical data; this would negate the benefit of doing this type of analysis in the first place, especially because we are aiming to move beyond historical data to increase our understanding of uncertainty in observationally derived stochastic hazard modelling.

This therefore represents an important question: how can we construct alternative histories from model data in such a way to allow us to derive novel probabilistic information? While this question may seem somewhat narrow in scope with this single application (i.e., generating counterfactual NAHU histories) in mind, the impacts of this are much more wide-ranging, and likely extend to the modelling of any other atmospheric peril, as well as more longer-term climate change risk-oriented questions.

Thus, while these limitations currently exist in the results, it seems likely that overcoming them has the potential to unlock many opportunities for enhancing views of atmospheric hazard for many risk-focused practitioners.

10.6   Suggestions for Future Research

As the discussion section presented, if the valuable applied aspects of this work are to be achieved, it is the probabilistic challenge presented in previous sections that needs significant attention. Further than this, there are two other immediate avenues of research that would be valuable to address given the limitations already imposed.

The first is in addressing the tropical cyclogenesis aspect to include non-observed hurricanes in alternative histories. While it could easily be included in the methodology above, it will introduce another complex issue that will need to be overcome before the data are fully coherent. The second would be to add alternative NWP models into the study to see how much model difference drives variability in the results.

References

Bank of England (2019) The 2021 biennial exploratory scenario on the financial risks from climate change. Accessed Online May 19th, 2021, from https://www.bankofengland.co.uk/-/media/boe/files/paper/2019/the-2021-biennial-exploratory-scenario-on-the-financial-risks-from-climate-change.pdf

Baudouin JP, Caron L-P, Boudreault M (2019) Impact of reanalysis boundary conditions on downscaled Atlantic hurricane activity. Clim Dyn 52:3709–3727

Brettschneider B (2008) Climatological hurricane landfall probability for the United States. J Appl Meteorol Climatol 47:704–716

Carozza D, Boudreault M (2021) A global flood risk modeling framework built with climate models and machine learning. J Adv Model Earth Syst 13:e2020MS002221. https://doi.org/10.1029/2020MS002221

Frigg R, Thompson E, Werndl C (2015) Philosophy of climate science part II: modelling climate change. Philos Compass 10(12):965–977. https://doi.org/10.1111/phc3.12297

Froude LSR, Bengtsson L, Hodges KI (2007) The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. Mon Weather Rev 135:315–333

Hall T, Hereid K (2015) The frequency and duration of U.S. hurricane droughts. Geophys Res Lett 42:3482–3485. https://doi.org/10.1002/2015GL063652

Halley E (1693) An estimate of the degrees of mortality of mankind, drawn from the curious tables of the births and funerals at the City of Breslaw, with an attempt to ascertain the price of annuities upon lives. Philos Trans 17:596–610

Halperin DJ, Fuelberg HE, Hart RE, Cossuth JH, Sura P, Pasch RJ (2013) An evaluation of tropical cyclone genesis forecasts from global numerical models. Weather Forecast 28(6):1423–1445

Hamill TM, Bates GT, Whitaker JS, Murray DR, Fiorino M, Galarneau TJ Jr, Zhu Y, Lapenta W (2013) NOAA's second-generation global medium-range ensemble forecast dataset. Bull Am Meteorol Soc 94:1553–1565. https://doi.org/10.1175/BAMS-D-12-00014.1

Hodges KI (1994) A general method for tracking analysis and its application to meteorological data. Mon Weather Rev 122:2573–2586

Hodges KI (1995) Feature tracking on the unit sphere. Mon Weather Rev 123:3458–3465

Hodges KI, Emerton R (2015) The Prediction of Northern Hemisphere tropical cyclone extended life cycles by the ECMWF ensemble and deterministic prediction systems. Part I: tropical cyclone stage. Mon Weather Rev 143:5091–5114

Hodges KI, Klingaman NP (2019) Prediction errors of tropical cyclones in the western north pacific in the met office global forecast model. Weather Forecast 34:1189–1209

Hodges KI, Cobb A, Vidale PL (2017) How well are tropical cyclones represented in reanalysis datasets? J Clim 30:5243–5264

Hoskins BJ, Hodges KI (2002) New perspectives on the Northern Hemisphere winter storm tracks. J Atmos Sci 59:1041–1061

Jewson S, Barnes C, Cusack S, Bellone E (2019) Adjusting catastrophe model ensembles using importance sampling, with application to damage estimation for varying levels of hurricane activity. Meteorol Appl. https://doi.org/10.1002/met.1839

Jones S, Raven E, Toothill J (2020) Assessing the global risk of climate change to re/insurers using catastrophe models and hazard maps, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-5323. https://doi.org/10.5194/egusphere-egu2020-5323

Kim S-H, Moon I-J, Chu P-S (2020) An increase in global trends of tropical cyclone translation speed since 1982 and its physical causes. Environ Res Lett 15:094084

Lloyd's (2017) Reimagining history: counterfactual risk analysis. Accessed Online May 19th, 2021, from https://assets.lloyds.com/assets/reimagining-history-report/1/Reimagining-historyreport.pdf

Lloyd's (2021) Realistic disaster scenarios, scenario specification 2021. Accessed Online May 19th, 2021, from https://assets.lloyds.com/media/e73cc2f7-a535-4eaf-8196-fedcf5e1432c/2%20RDS%20Scenario%20Specification%20%20January%202021.pdf

Ng KS, Leckebusch GC (2021) A new view on the risk of typhoon occurrence in the western North
    Pacific. Nat Hazards Earth Syst Sci 21:663–682
Roese NJ (1997) Counterfactual thinking. Psychol Bull 121:131–148
Saha S, Moorthi S, Pan H-L et al (2010) The NCEP climate forecast system reanalysis. Bull Am
    Meteorol Soc 91:1015–1057
van Oldenborgh GJ, van der Wiel K, Kew S et al (2021) Pathways and pitfalls in extreme event
    attribution. Clim Change 166. https://doi.org/10.1007/s10584-021-03071-7
Woo G (2019) Downward counterfactual search for extreme events. Front Earth Sci 7. https://doi.
    org/10.3389/feart.2019.00340