# US Tropical Cyclone Activity in the 2030s Based on Projected Changes in Tropical Sea-Surface Temperature

Timothy M. Hall

NASA Goddard Institute for Space Studies, New York, NY

timothy.m.hall@nasa.gov

James P. Kossin

NOAA National Centers for Environmental Information, Madison, WI

Terence Thompson

The Climate Service, Durham, NC

James McMahon

The Climate Service, Durham, NC

**Abstract**

We use a statistical tropical cyclone (TC) model, the North Atlantic Stochastic Hurricane Model (NASHM), in combination with sea-surface temperature (SST) projections from climate models, to estimate regional changes in US TC activity into the 2030s. NASHM is trained on historical variations in TC characteristics with two SST indices: global-tropical mean SST and the difference between tropical North-Atlantic (NA) SST and the rest of the global tropics, often referred to as "relative SST." Testing confirms the model's ability to reproduce historical US TC activity, as well as to make skillful predictions. When NASHM is driven by SST projections into the 2030s, overall NA annual TC counts increase, and the fractional increase is the greatest at the highest wind intensities. However, an eastward anomaly in mean TC tracks and an eastward shift in TC formation region result in a geographically-varied signal in US coastal activity. Florida's Gulf coast is projected to see significant increases in TC activity, compared to the long-term historical mean, and these increases are fractionally greatest at the highest intensities. By contrast, the northwestern US Gulf and the US East Coast will see little change.

## 1. Introduction

Tropical cyclones (TCs) are among the most hazardous natural catastrophes to impact the US (Smith and Katz, 2013). How this hazard evolves with climate variation, both natural and anthropogenic, is of crucial importance to coastal populations and is an extremely active area of research. Other factors constant, coastal storm-surge hazard is certain to rise with sea level (Tebaldi et al., 2012), TC rainfall is projected to increase with high confidence (Kossin et al., 2017; Knutson et al., 2020), and the interaction of fresh-water flooding and storm surge will be exacerbated (Moftakhani et al, 2017). Theory indicates that a warming climate causes TC wind intensity to increase, especially at the highest intensity levels, a result borne out by numerical model simulations (Villarini and Vecchi, 2013; Sobel et al., 2016; Knutson et al., 2020) and observations (Elsner et al., 2008; Kossin et al., 2013; Kossin et al., 2020; Knutson et al., 2019).

At least as important as intensity for coastal hazard is evolution in the geographic distribution of TC landfall via changes in formation and storm track, though the nature and degree of these changes are considerably more uncertain than sea-level rise, rainfall, and intensity (Knutson et al., 2020). Most model studies find little change or decreases in global TC annual frequency (Yoshida et al., 2017; Murakami et al, 2020; Knutson et al., 2020), though the variation across studies is greater in individual ocean basins. However, several studies find an eastward shift in North-Atlantic (NA) TC formation region toward the eastern tropics (Murakami et al., 2010; Wang et al., 2011; Colbert et al., 2013; Yoshida et al., 2017). There is also evidence for eastward or northeastward TC track anomalies (Wang et al., 2011; Colbert et al., 2013; Baldini et al,

2016). In addition to changes in basin-wide TC rates, both the formation-region and track changes would play a crucial role in determining regional changes in US landfall rates.

Regional and local spatial scales are of primary importance to disaster and adaptation planning, and it is of considerable interest to estimate the impact at these scales of global and basin-wide relationships between climate and TC activity. Here, we use the North Atlantic Stochastic Hurricane Model (NASHM), a statistical-stochastic model of NA TCs and their climate variation (Hall and Jewson, 2007; Hall and Yonekura, 2013), along with projections of climate variation from an ensemble Coupled Model Inter-comparison Project Phase 5 (CMIP5) climate models (Taylor et al, 2012), to estimate US regional TC activity into the 2030s. This time-scale is highly relevant to a number of sectors, such as infrastructure development and the insurance and reinsurance industry. NASHM is able to replicate historical US coastal TC activity on regional scales, as well as climate-induced variation of that activity (see Methods). It therefore provides a method to project large-scale climate variations onto regional TC activity.

We represent climate variation by two annually-varying sea-surface temperature (SST) indices: 1., tropical SST (GSST), averaged over all longitudes and during the NA hurricane season; and 2., "relative SST" (RSST), the NA main development region (MDR) SST minus the tropics over all longitudes outside the Atlantic. Historical values of GSST and RSST are derived from the HadSSTI product (Raynor et al., 2003), while the future projected values are taken from CMIP5 climate model data (Taylor et al., 2012). (See Methods for further details.) GSST exhibits a secular increase since the 1970s driven largely by anthropogenic climate change (Santer et al., 2006;

Gillette et al., 2008; Knutson et al., 2013). RSST can be considered as a proxy during this period for potential intensity (PI), which is a function of SST and atmospheric thermodynamic factors (Bister and Emanuel, 2002). RSST has been shown to be correlated with NA TC activity (Vecchi and Soden, 2007), as well as NA major hurricane activity (Murakami et al., 2018).

Large-scale SST indices have an advantage of being relatively insensitive to heterogeneities in the reanalysis data compared to historical estimation of PI (Kossin, 2015). Moreover, SST data are available decades earlier than the reanalysis data required for PI construction, allowing the use of longer historical time periods in the construction of NASHM. In addition, GSST and RSST are relatively robust across climate model future projections compared to the ingredients needed to construct PI. However, these SST indices have the disadvantage that they are only indirectly related to the specific meteorological mechanisms responsible for evolving TC characteristics. Thus, we use these large-scale SST variables largely as proxies for specific meteorological mechanisms. For example, reduction in the tropical overturning circulation in a warming climate (Held and Soden, 2012) and a weakening of the summer mid-latitude circulation (Coumou et al., 2015) both impact TC steering winds and would be represented here only indirectly via increasing large-scale SST. As proxies, the SST variables have limitations. Emanuel and Sobel (2013) have argued that TC characteristics cannot be expressed as any unique function of SST, other than the approximate relationship between PI and RSST, and RSST is not projected to increase with climate warming (Vecchi et al., 2008; and see below). This leaves the increase in PI projected by climate models (Sobel et al., 2016) to be captured here by the increase in GSST, which is an incomplete representation (Emanuel and Sobel, 2013).

Therefore, our results may be conservative, as they do not capture other mechanisms that may increase PI as greenhouse gas concentrations increase.

We partly buffer the results here against the limitations of solely using SST-based climate variables by restricting the time horizon of the projections. NASHM is a statistical model that exploits historical relationships between aspects of TC activity and the SST indices. Once the relationships are established in the training and calibrating, we use NASHM to estimate TC activity at projected values of the SST indices. Such projections become increasingly less reliable as GSST or RSST move increasingly outside the historical range of variation, and, as mentioned above, projections in SST may not adequately capture longer-term trends in PI. Here, we make future projections only up to the year when the GSST anomaly, as determined from the ensemble-mean of CMIP5 models, reaches 0.75°C, or 3.2 standard deviations above the 1900-2017 mean. For RCP4.5 this occurs in 2038; for RCP8.5 it occurs in 2030. Prior to these years NASHM's TC-activity projections via its GSST dependence can be tested. Out-of-sample tests on the highest historical GSST years verify NASHM's forecast skill compared to a naïve SST-independent model (see below).

In Section 2 we describe NASHM, evaluate its historical coastal TC activity and perform out-of-sample forecast tests, quantify model uncertainty, and present the historical and climate-model generated SST indices. In Section 3 we analyze the TC sensitivity to the SST indices and present the forecast TC activity in the 2030s, driving NASHM with SST projections. In Section 4 we

discuss mechanism by isolating the SST dependence of NASHM's formation, track, and intensity components.

## 2. Methods

### *2.1.* NASHM

The North Atlantic Stochastic Hurricane Model (NASHM) is a statistical-stochastic model that simulates the life cycle of TCs from formation through termination. NASHM was developed by Hall and Jewson (2007) and Hall and Yonekura (2013) and has been applied to study coastal TC activity and its relationship to climate (e.g., Hall and Hureid, 2015; Hall and Tippett, 2017). The model is trained on SST data and HURDAT2 archives (Landsea and Franklin, 2013), in this application from 1900 to 2017. The SST indices are discussed below. In the construction of NASHM, attention has been paid to avoid overfitting. The spatial and temporal scales of HURDAT2 data used to estimate local rates (formation at a point, propagation from a point, intensity along a track) is determined objectively by maximizing out-of-sample likelihood. This procedure strikes the balance between resolving true structure, which pushes for smaller scales, and including sufficient data to avoid sampling error, which pushing for larger scales.

NASHM shares many characteristics with other statistical TC track models (e.g., Vickery et al, 2000; James and Mason, 2005; Bleomendaal et al., 2020), many of which are used as part of commercial hurricane risk assessment and are calibrated to historical observations. Unlike many such models, NASHM is constructed to be sensitive to indices of climate variation, here GSST and RSST, via its three primary components---TC formation, track, and maximum-

sustained wind speed ($V_{max}$). Thus, it can employ information from global climate-model projections to forecast future TC activity. In this regard, NASHM is similar to statistical-dynamical TC models (Emmanuel et al., 2008; Lee et al., 2018), though such models require environmental fields as input that are more directly related to TC mechanisms than NASHM's SST indices.

Fig 1, top row, shows the 1900-2017 HURDAT2 data on which NASHM is trained and an example stochastic simulation of the period. The basin-wide structure, westward storm movement in the subtropics, re-curvature, and eastward return at mid-latitudes, is well captured. This agreement between the basic basin-wide structure in HURDAT2 and NASHM is further revealed in the track-density plots in the bottom row of Fig 1. Once model training is performed on HURDAT2 and the SST indices, a calibration step is applied to long-term average landfall rates on several North American regions. The calibrated model is then used to estimate climate-driven variations on coastal TC activity. Fig 2 shows regional landfall rates as functions of $V_{max}$. A necessary criterion for lack of model landfall bias is that the stochastic spread in rates bound the historical rates, a criterion that NASHM meets, as illustrated in the figure.

HURDAT2 is known to have under-reporting biases in the earlier part of the record prior to the satellite era, and even more so prior to era of routine aircraft reconnaissance (Landsea, 2007; Landsea et al., 2010; Vecchi and Knudson, 2011). In order to buffer NASHM from these biases, at least in part, in the training data we only use TCs that last at least two days, reach at least tropical-storm status (at least 63 km/hr one-minute-sustained $V_{max}$, referred to throughout as

165 "Cat0+"), and the origin of tracks is taken as the point at which Cat0 intensity is achieved, rather
166 than the first reported latitude-longitude fix. To examine further the sensitivity to these
167 potential early-record HURDAT2 biases, much of the analysis reported below is subsequently
168 repeated for a version of NASHM trained on 1970-2017 HURDAT2, and the results lead to the
169 same conclusions about evolution regional US TC activity.
170
171 *2.2.* SST Indices
172 The climate indices used here are derived from sea-surface temperature (SST): 1., global
173 tropical SST (GSST), averaged over at all longitudes, latitudes 23°S to 23°N, and each year from
174 July through October (JASO); and 2., "relative SST" (RSST), the NA MDR (20°-60°W & 6°-18°N)
175 JASO SST minus the JASO global 23°S-23°N region outside the Atlantic. As noted in the
176 Introduction, SST indices have the advantage of being relatively insensitive to heterogeneities in
177 the historic record and fairly robust across climate model projections. A disadvantage of simple
178 SST indices is that their physical link to key TC characteristics is not direct. SST itself plays a
179 direct role in formation and intensification via the PI (Bister and Emanuel, 2002), but RSST likely
180 under-represents long-term trends in PI and thus may provide conservative estimate of future
181 trends in TC activity. For example, ensemble-mean RSST does not increase in CMIP5
182 projections, as shown below, even though PI increases (Sobel et al. 2016). SST variation are
183 also associated with steering wind variation and, therefore, TC tracks, but the relationship is
184 indirect and mediated by intermediate mechanisms, e.g., Wang et al (2011) and other
185 references from the discussion in Section 4. Similarly, SST indices have indirect relationships
186 with wind shear and spatial SST patterns, important influences on TC activity, for example the

187 Atlantic Meridional Mode (Vimont and Kossin, 2007). There is no guarantee that the
188 intermediate mechanisms providing a relationship between SST, steering winds, and wind shear
189 in the historical period will continue to operate in a future climate. These limitations are the
190 reason we limit our future projections to the 2030s. Beyond this time, GSST will likely be well
191 above the highest value in the historic training period.
192
193 The projected SSTs are computed from the eight CMIP5 climate models (CCSM4, CNRM-CM5,
194 NorESM1-M, ACCESS1.0, GISS-E2-R, HadGEM2-CC, HadGEM2-E, MRI-GCM3S) that were ranked
195 highest in a model-observation comparison of tropical SST anomalies (Brown et al., 2015). We
196 use one ensemble member for each model and each emission scenarios RCP4.5 and RCP8.5 and
197 an additional RCP4.5 and RCP8.5 ensemble member for each of the models ranked top three
198 (CCSM4, CNRM-CM5, NorESM1-M) by Brown et al (2015). This results in a total of 11 ensemble
199 members for each emission scenario. The CMIP5 SST projections start in 2006 (Taylor et al.,
200 2012), thereby providing 12 years of overlap (2006-2017) with the historical SST. We use this
201 overlap period to adjust the projected GSST series such that each ensemble member's 2006-
202 2017-mean anomaly matches the historical 2006-2017 mean anomaly, resulting in an
203 ensemble-mean adjustment down of 0.07°C. For RSST no adjusting was necessary, as the
204 ensemble-mean 2006-2017-mean RSST already closely matched the historical mean value over
205 this period.
206
207 Fig 3 shows the historical GSST and RSST and the CMIP5 ensemble-mean projections for the
208 two climate-change emission scenarios, RCP8.5 and RCP4.5. The ensemble-mean projected

GSST continues the historical increase that began in the 1970s, with RCP8.5 rising faster than RCP4.5. For RSST, the RCP4.5 and RCP8.5 ensemble-mean projections remain essentially constant, equal to the historical 2006-2017 value, which is elevated compared to the 1900-2017 RSST mean. Although individual ensemble members display considerable interannual variability in RSST, the ensemble means do not. The low-frequency variation in historical RSST, closely related to the Atlantic-Multidecadal Oscillation (AMO), is not projected to continue in a coherent fashion by these models, possibly because the models replicate most of the historical AMO signal as an externally forced variation, rather than a natural mode of variability (Mann et al., 2020).

### *2.3.* NASHM-SST Forecast-Skill Assessment

We use out-of-sample tests to evaluate NASHM's ability to forecast TC rates, given values of GSST and RSST. The changes projected for the 2030s depend primarily on GSST, and so we focus on this covariate. We select as a subset of historical data (the in-sample years) all the years whose GSST is less than 1.5 standard deviations above the 1900-2017 GSST mean. There are 109 such in-sample years out of the 118 total years. The remaining 9 years (the out-of-sample years: 1987, 1997, 1998, 2003, 2009, 2014, 2015, 2016, and 2017) have GSST greater than 1.5 standard deviation above the mean. We then train two versions of NASHM on the in-sample years: 1. a version that depends on GSST and RSST; and, 2., a naive version, whose components are trained on 1900-2017 HURDAT2 means and variances assuming an annually repeating mean state, as described in Hall and Jewson (2007), independent of interannual variation in GSST and

RSST. For each model version, we perform 118,000 simulations for each of the out-of-sample years.

We evaluate the skill of these out-of-sample forecasts by comparing the rates to the historical events in the following manner: A 0.5° grid is defined covering the western North Atlantic and coastal North America, spanning 65°W-100°W and 17°N-46°N, 50km-radius circles are drawn around each grid point, and TC tracks passing through the circles are counted. Because of the small area of the circles, passage rates are low and to a good approximation equivalent to the annual probability of TC passage through the circles. We take the passage rate averaged over the 118,00 simulations of each out-of-sample year as the mean rate of a Poisson distribution and evaluate that distribution at the historical count in the circle for the year. This results in a model-predicted probability for the historical event in that year and circle.

We then evaluate the skill in two ways:

1. The log probabilities are summed over all the out-of-sample years and over specified evaluation regions and intensity levels. We repeat the procedure using the naïve model and compare the sums of the log probabilities. The climate-dependent model is deemed more skillful if its total log probability is greater than that of the naïve model. This procedure is sensitive to the overall counts in a region, as well as the geographic distribution of counts within a region.
2. An F-score is computed from calculations of true positive (TP), false positives (FP), and false negatives (FN). The climate-sensitive and naïve model Poisson distributions are

sampled at each location. A model forecast of a TC occurrence in the intensity category of interest is recorded TP if such an event occurred in the out-of-sample historical record, FP if such an event did not occur. A model forecast of no TC occurrence in the intensity category of interest is recorded FN if such an event did occur. From these, the precision, P=TP/(TP+FP), and recall, R=TP/(TP+FN), are computed, followed by the F score, F = 2PR/(P+R). We then document the percent difference in F between the climate-sensitive and naïve models. Higher F indicates better performance. When counting TP, FP, and FN the Poisson sampling is random. Values are accumulated over many repeated realizations of the procedure until the ratio of the climate-sensitive to naïve F scores converges at the 1% level.

We choose five regions within 350km of the coast to evaluate skill: Northwest Gulf, Florida, the US Southeast, the US Northeast, and the full US Gulf and Atlantic coast (Fig 1b). We also evaluate skill on the complete 65°W-100°W and 17°N-46°N domain. The evaluations are performed for Saffir-Simpson Cat0, Cat1-2 and Cat3-5 intensity levels. (The regions and intensity ranges are chosen to balance the desire to assess skill regionally, but still have historical counts available for scoring.) Table 1 summarizes the results of the out-of-sample evaluation. In all but one case (15 of 16 region-intensity combinations), the climate-dependent model is more skillful than the naïve model at forecasting TC occurrence in high GSST years according to both the log-likelihood comparison and the F-score comparison, though the US Southeast Cat0 difference is marginal. The exception is Cat1-2 storms in the US Northeast. Many such storms have transitioned from tropical to extra-tropical cyclones, which derive their energy from different meteorological mechanisms, and therefore may be less sensitive to the

tropically-based RSST and GSST indices used here. In general, the increase in skill is greatest for high-intensity TCs. Skill scores have the greatest potential for differentiating models where the differences between the out-of-sample and in-sample data are greatest, and it is at high intensity where the high-GSST years are most different fractionally from the in-sample years.

Although testing forecast skill for high GSST years is our primary focus, we have also performed similar out-of-sample tests on the five years (1908, 1909, 1910, 1911, and 1916) whose GSST anomaly is lower than -1.5 standard deviations. In this case, all of region-intensity combinations show increased total log likelihoods using the climate-sensitive model. The F-score tests cannot be performed on all of the region-intensity combinations, because there are not always historical events available to allow for TPs. In all cases where there is the possibility of a TP, the climate-sensitive model's F-score is higher than the naïve model. For cases where there is not the possibility of a TP, the climate-sensitive model has lower FN+FP than the naïve model.

To evaluate further NASHM's skill, we compute the landfall rates on the same six regions of Fig 2, but now only for the nine out-of-sample years of the skill testing. The results are shown in Fig 4. The stochastic spread is here larger than in Fig 2, because individual simulations are only nine years long, rather than 118 years. It's common for a 20km/hr $V_{max}$ bin not to be populated over the nine years. Nonetheless, the historical values fall inside the inner 90% of the simulations, further evidence for NASHM's skill in making probabilistic forecasts.

*2.4.* Uncertainty Estimates

In addition to testing skill, we want to estimate the uncertainty of our best-estimate occurrence rates for the historical baseline and the projections shown below. To do this, we perform a bootstrap analysis (Efron, 1979). The 118 years (1900-2017) of HURDAT2 are resampled with replacement 100 times. NASHM is then trained on each of the 100 bootstrap resamples, and each NASHM-bootstrap model is used to simulate 1000 realizations of the 118-year period (118,000 years). Mean TC occurrence rate maps (annual passage rates through the set of 50km circles) are computed from each bootstrap simulation, resulting in bootstrap distributions of the mean occurrence rates at each location. We then compute quantiles from the bootstrap distributions for use in assessing uncertainty of the best-estimate rates. For uncertainties on the mean rates, we compute the 5%-95% quantile range of the bootstrap distribution.

To quantify the significance of differences between rates in one period and another period, a different quantile-selection scheme is used. We want to estimate how significant is the difference between two best-estimate values, given their individual bootstrap distributions. What is the probability that a value from the lower distribution is greater than a value from the higher distribution? The probability that a value in the lower distribution is in the upper 0.224 fraction (the 77.6% quantile) and the value in the higher distribution is in the lower 0.224 fraction (the 22.4% quantile) is $(0.224)^2 = 0.050$, assuming the distributions are independent. Therefore, if the 22.4%-77.6% ranges of the distributions do not overlap, then the best estimates are significantly different to at least 95% confidence. This estimate of the significance of the differences is conservative, because the distributions are not, in fact, independent. In the simulations, if a 1950-1980 value of a bootstrap-resample member is on the low/high end of its

bootstrap distribution, then the 2030 value also tends to be on the low/high end of its distribution, because the long-term averaged rates from that bootstrap member are lower/higher.

## 3. Results

We now discuss the impact of GSST and RSST variation on TC activity and make projections into the 2030s. To analyze the impact of various levels of GSST and RSST on TC activity, both individually and in combination, long NASHM simulations are performed (118,000 years) with GSST and RSST held fixed in various combinations. We choose 30 combinations of fixed-anomaly values with respect to 1900-2017, using the six values -0.5°C, -0.25°C, 0°C, +0.25°, +0.5°C, and +0.75°C for GSST; and the five values -0.5°C, -0.25°C, 0°C, +0.25°, and +0.5°C for RSST. This represents a range of -2.1 to +3.2 standard deviations for GSST and -1.9 to +1.9 standard deviations for RSST. We then compute from the simulated TCs for each GSST-RSST combination the average annual rates of TC passing through the set of 50km circles on the gridded field defined previously. Fig 5 shows the annual Cat1+ (Saffir Simpson category 1 and higher) TC occurrence rates for 9 of the 30 GSST-RSST combinations. Increasing RSST (bottom to top) is primarily associated with an overall basin-wide increase in rate, with relatively little spatial variation in the increase. By contrast, increasing GSST (left to right) results in changes in geographic distribution, but comparatively little basin-wide change in TC annual rates. Of particular interest is a shift eastward in Gulf and Atlantic occurrence rate as GSST is increased from a neutral to an elevated level. For example, in the Gulf the hotspot near Texas and Louisiana for GSST=0C° anomaly (for any RSST) shifts eastward near the Florida panhandle at

GSST=+0.5°C anomaly. Similarly, the hotspot near North Carolina shifts eastward off the US east coast.

The rate maps of Fig 5 are bi-linearly interpolated in the two-dimensional GSST-RSST space to the GSST and RSST values of each year of the CMIP5 projections, 2020-2038 for RCP4.5 and 2020-2030 for RCP8.5. (We could have simulated directly the response to the projected CMIP5 RSST and GSST values. We chose the approach of simulating the GSST-RSST combinations then interpolating in order to document the separate dependence on each SST covariate. As the CMIP5 ensemble-mean RSST and GSST are smoothly varying, interpolating versus direct simulation yields very similar results year-by-year through the projected time horizon.) We then compare the rate maps at the end of the projections to NASHM's rate map from a set of 1950-1980 historical baseline simulations. Note that the rates depend on time only via the SST indices, and thus the RCP4.5 rate map in 2038 is identical to the RCP8.5 rate map in 2030. The SST values at these final dates are GSST=0.75°C and RSST=0.20°C, expressed as anomalies from the 1900-2017 means. It may appear at first that a single projected year (2030 or 2038) is being compared to an average over a range of years (1950-1980), which would be problematic due to internal variability. However, the single projected year is in fact derived from 118,000 years of stochastic simulation variability, and the driving GSST and RSST values are derived from an ensemble-mean of climate model simulations, each with its individual variability.

Fig 6 shows the TC occurrence rate maps for the 1950-1980 baseline, the 2038 RCP4.5 (equivalent to 2030 RCP8.5), and the difference, for intensities of Cat1+, Cat3+, and Cat4+. Basic

climatological features of the maps are peaks in rates in the Gulf and US mid-Atlantic, as well as a rapid decrease with intensity of rates at mid latitudes, seen by comparing Cat3+ (middle row) and Cat4+ (bottom row) to Cat1+ (top row). The difference maps reveal a complex geographic pattern of change that is approximately coherent across the TC intensities. In the eastern Gulf, extending from the eastern Yucatan, through western Cuba, to the Florida panhandle, change in TC occurrence rates is either positive or neutral. Over that region, the fractional rate increase is greatest for intensity Cat4+ (bottom row). By contrast, in the western Gulf and the Atlantic off the US eastern seaboard, change in TC occurrence rates is mostly negative or neutral, except for a positive region at higher intensities near the Texas-Mexico border. In both of these regions, the fractional decrease is smallest for intensity Cat4+ (bottom row). Further east, into the central Atlantic, the change is again negative or neutral. This change pattern can in part be summarized as a shift eastward in rates for both the Gulf and the Atlantic, resulting in a west-to-east negative-to-positive pattern in both the Gulf and the western Atlantic. Superposed on this spatial pattern is a broad increase at the highest intensity levels.

For US Gulf-coast states, the eastward shift results in a projected decrease in TC occurrence rates to the west, particularly on eastern coastal Texas and western Louisiana, and a projected increase to the east, particularly Florida's panhandle and Gulf coasts. In between, centered roughly on Mississippi and Alabama, is a node, with little or no change.

Fig 7 shows the TC occurrence rates averaged over four coastal regions: Texas through Alabama, Florida's Gulf Coast, Florida's Atlantic Coast, and Georgia through Cape Hatteras, NC.

Rates are shown for three time periods: 1950-1980, 2010-2019, and 2030 for RCP8.5 (identical to 2038 for RCP4.5). Consistent with the change maps of Fig 6, rates increase on Florida's Gulf Coast and decrease Georgia to Cape Hatteras for all intensity thresholds, and the changes from 1950-1980 to 2030 are significant according to the bootstrap analysis described in Section 2.4, except for Cat3+ on Florida Gulf. Rates decrease Texas-Alabama for Cat1+ and Cat3+, but the change for Cat4+ is not significant.

We note that RSST displays considerable year-to-year and decadal variability over the historic period (Fig 3). This in turn drives year-to-year and decadal variability in TC rates. The RSST projections based on ensemble-mean CMIP5 results, however, show no variability. The resulting projected TC rates therefore reflect only the secular increase of the ensemble-mean GSST. Natural climate variability will certainly continue to drive variability in TC rates. Using ensemble-mean projected SST, however, we have no basis to project such variability, and we focus instead on the change driven by the secular variation in GSST.

## 4. Discussion

To explore the causes for the pattern of increased and decreased TC rates, we look at the GSST and RSST dependence of the three major components of NASHM: formation, track, and intensity. In NASHM, TC formation occurs by sampling Poisson densities whose mean rates are dependent on GSST and RSST (Hall and Yonekura, 2013). Fig 8 (bottom row) shows spatial maps of change in mean rates as a function of GTSST and RSST compared to the neutral state of zero GSST and RSST anomalies. GSST and RSST induce different formation responses. Elevating GSST

406 by 0.75°C (the 2030-RCP8.5 & 2038-RCP4.5 value, 3.2 standard deviations above the 1900-2017
407 mean) with RSST held neutral, results in a modest increase in overall annual formation count
408 (10.4 to 12.6). However, there is a distinct eastward shift in formation, illustrated by the large
409 increase in the eastern NA MDR. By contrast, overall storm count is more sensitive to RSST.
410 Elevating RSST by the much smaller 0.20°C (the 2030-RCP85 & 2038-RCP45 value, roughly one
411 standard deviation above the 1900-2017 mean), with GSST held neutral, results in about the
412 same increased storm count (10.4 to 12.5). However, in contrast to GSST, the formation
413 increase induced by RSST is spatially more uniform throughout the tropical North Atlantic. The
414 combined GSST and RSST increases, reflecting the 2030-RCP8.5 and 2038-RCP4.5 values, results
415 in a basin-wide count increase 10.4 to 15.2, combined with the eastward shift induced by RSST.
416
417 TC mean track in NASHM also depend on the SST covariates (Hall and Yonekura, 2013). In order
418 to better visualize the impact of the SST covariates on the TC tracks, we turn off the stochastic
419 component of NASHM and release mean tracks from a fixed set of locations. Fig 8 (top row)
420 shows the results for the same set of GSST and RSST as the formation results. There is an
421 eastward anomaly in the mean track for elevated GSST, but little change for RSST. For example,
422 the mean track that landfalls in Louisiana in the neutral case, instead landfalls in the Florida
423 panhandle in the 0.75°C & 0.0°C and 0.75°C & 0.20°C cases. In addition, in mid-latitudes, mean
424 tracks near the US east coast recurve away from the US Northeast more in the elevated GSST
425 cases.
426

427 TC intensity in NASHM is modeled by a weighted resampling scheme with perturbation of $V_{max}$
428 time series (Hall and Yonekura, 2013). Weights are applied such that it is more likely to
429 resample a $V_{max}$ time series from a historical year whose SST covariate values are similar to the
430 simulation year. The weights are selected objectively to minimize out-of-sample $V_{max}$ forecast
431 error. Fig 8 (middle row) shows histograms of simulation lifetime maximum intensity (LMI) in
432 the elevated GSST & RSST states compared to the neutral state. Elevated RSST induces an
433 increase at all intensities, with the fractional change increasing gradually with intensity. By
434 contrast, elevated GSST induces a sizeable increase only at Cat5, with smaller or no increases at
435 lower intensities. The combined GSST-RSST increase, representing 2030 (RCP8.5) and 2038
436 (RCP4.5), results in an approximate factor 2.5 increase in Cat5 occurrence and a 30%-60%
437 increase at lower intensity thresholds.
438
439 An important point here is that large-scale changes in TC characteristics (formation,
440 propagation, intensity) in response to these global and basin SST covariates can result in
441 regional coastal changes in TC activity that are highly heterogeneous, even in sign. Fig 9 is an
442 idealized illustration of this point that helps to understand the increased occurrence rate in the
443 northeast Gulf. Consider two sets of idealized mean tracks both originating from the same
444 fixed set of formation points at 15°N every 0.5° in longitude. One set differs from the other only
445 in a constant uniform eastward propagation anomaly of 5km eastward for every 25km
446 northward. Despite the uniformity of the anomaly, the coastal impact varies regionally. Fig 9
447 shows the example of Florida. Without the eastward anomaly, Florida's eastern Panhandle,
448 Great-Bend region, and much of is Gulf Coast are shielded from direct landfalls. With the

eastward anomaly, parts of these regions now get direct landfalls. Thus, there is an increase in activity on western Florida. By contrast, the same track anomaly results in decreased activity on Florida's east coast, especially the northeast, because the mean tracks now curve roughly parallel to the coast without making landfalls. Clearly, with realistic tracks (stochastic in NASHM), any region can be struck directly. However, the idealization illustrates that mean track shifts, even if uniform, can increase probability in some regions and decrease it in other regions. The NASHM mean track changes of Fig 6 have similar character to these idealized changes and result in the increased occurrence rates in the northeast Gulf.

NASHM is trained on HURDAT2 1900-2017. However, there are known biases in the earlier pre-satellite record, approximately prior to 1970 (Landsea, 2007; Landsea et al., 2010; Vecchi and Knudson, 2011), especially for weak storms short that don't landfall. We have attempted to buffer against these biases in the model training by only using HURDAT2 tracks that reach at least TS status and starting the tracks at the point where TS status is reached. Nonetheless, to test sensitivity to the data record period, we have reconstructed NASHM using 1970-2017 HURDAT2. Vecchi and Knutson (2011) estimate that after 1965 undercounts in annual TC frequency due to sparce sampling are negligible, hence 1970-2017 is inside the well-sampled satellite era. Fig 10 repeats Fig 8's illustration of formation, mean track, and LMI sensitivity to GSST and RSST, but now for the 1970-2017-trained model. The GSST and RSST anomalies for 2030/2038 (RCP85/RCP45) are here reduced to 0.57°C and 0.19°C from 0.72°C and 0.20°C because they are now computed with respect to a 1970-2017 baseline.

471 The qualitative behavior in TC sensitivity is similar for all three model components: a shift
472 eastward in formation toward the eastern MDR, an eastward propagation anomaly, and an
473 increase in the LMI distribution at the highest intensities. The resulting changes in TC passage
474 rates, shown in Fig 11, are also similar to those of the 1900-2017 training shown in Fig 6.
475 However, there are quantitative differences: the basin-wide increase with respect to the zero-
476 anomaly state for GSDST and RSST in annual TC count is reduced from 4.8 to 3.0, and the
477 eastward track anomaly is enhanced. The change in baseline period explains much of the
478 change in basin-wide formation; the 1970-2017 period has 1.5 more TC $yr^{-1}$ than 1900-2017
479 (11.9 compared to 10.4). Undercounts in the early HURDAT2 record (Landsea, 2007) may be
480 responsible for some of this difference. The increased eastward track anomaly could be an
481 indication of an acceleration in the signal or simply statistical variability, which is exaggerated in
482 the shorter data training record. These quantitative differences have an impact on occurrence
483 rates in certain coastal regions. Fig 6 shows rates based on 1970-2017 analysis, as well as 1900-
484 2017. The uncertainty ranges for the shorter period are significantly larger, and the Western
485 Gulf and US Southeast do not change significantly. However, the increase in rates on Florida's
486 Gulf Coast is significantly amplified, likely due to the increased eastward track anomaly. In
487 addition, the 1970-207-based analysis projects and increase on Florida's Atlantic coast for
488 Cat1+and Cat3+. Southern Florida is near a node between decreased and increased rates (Figs 6
489 and 11), and in the 1970-2017.
490
491 Similar SST-dependent formation, track, and intensity features have been reported in other
492 studies, which have examined a range of time periods and driving mechanisms. The NASHM

493 intensity variations are consistent with a number of modeling and observational studies finding
494 that positive trends in TC activity are fractionally greatest at the highest intensities (Sobel et al.,
495 2016; Gutmann et al., 2018; Knutson et al., 2020; Elsner et al., 2008; Kossin et al., 2013; Kossin
496 et al., 2020). Meanwhile, an eastward shift in formation has been seen by Kossin et al. (2010)
497 during the warm phase of the Atlantic Meridional Mode (AMM), due not only to increased
498 subtropical SST, but also to covariation with vertical wind shear. Eastward formation shifts have
499 also been reported by Wang et al. (2011) in the context of expanded Atlantic warm pools, by
500 Murakami et al (2010) in a 20-km-mesh atmospheric model driven by SST projections, and by
501 Colbert et al (2013) in a CMIP3 downscaling study. Chauvin et al (2019) has also seen an
502 eastward formation shift using high-resolution global modeling.
503
504 A subtropical eastward track anomaly similar to that shown here (Figs 8 and 10, top row) has
505 also been reported in modeling and observational studies. Explanations of mechanisms vary,
506 but all the studies have in common an association between increased SST and changes in
507 tropospheric wind patterns that steer TCs. Wang et al (2011) argue that an expanded Atlantic
508 warm pool, which could be associated with changes in either GSST or RSST, results in a
509 weakened NA subtropical high-pressure system, thereby weakening westward tropical mid-
510 tropospheric winds that steer TCs. In statistically downscaled CMIP3 climate-change
511 simulations, Colbert et al (2013) observe an eastward TC track anomaly, due largely to a
512 reduction in westward steering winds. A study of Hurricane Sandy under enhanced SST
513 conditions found a tendency for the storm to recurve eastward away from the coast (Lau et al,
514 2016), similar to the eastward anomaly seen here in mid-latitude. Liu et al (2017), using 4km-

515 resolution regional modeling, detected a northeastward anomaly in the NA mid-tropospheric
516 TC steering winds. Based on reconstructions from proxy records, Baldini et al (2016) presented
517 evidence for a northeastward shift over the industrial era in TC tracks that originate in NA MDR.
518 Recently, Hassanzadeh et al (2020) have found in analysis of CMIP5 climate projections a
519 positive meridional anomaly in steering winds in the eastern Gulf, which would direct tracks
520 northward before reaching the western Gulf, consistent with the eastward anomaly we find in
521 the Gulf. All of these changes project onto US landfall rates.
522
523 The projected basin-wide increase in annual NA TC frequency shown here runs counter to
524 many, though not all, previous results. Climate model results of Murakami et al (2020) indicate
525 that increased NA TC frequency of recent decades will wane in future, as greenhouse-gas
526 forcing dominates aerosol forcing, and TC frequency will decline in the NA starting around 2030
527 (their Fig S3). By contrast, Bhatia et al (2018) show an increase in global TC frequency, including
528 in the NA. Increased frequency has also been seen in CMIP5 models (Villarini and Vecchi, 2012).
529 The work of Vecchi et al (2019) indicates that projected TC frequency change has complex
530 dependence on numerical model resolution and the spatial pattern of warming, with decreases
531 in some experiments and increases in others. Analyzing the distribution of results across a large
532 number of studies, Knutson et al (2020) show that the median North Atlantic TC frequency
533 change is negative, but that the inter-quartile range across the studies spans positive change.
534 We deduce an increase of 4.6 $yr^{-1}$ in 2038 (RCP4.5) compared to a 1900-2017 baseline of 10.4
535 $yr^{-1}$, which is an increase of 2.6 $yr^{-1}$ compared to the 1986-2005 baseline of 12.4 $yr^{-1}$ used by
536 Knutson et al (2020). The 21% increase (2.6/12.4) falls within the 5%-95% quantile spread show

by Knutson et al (2020). Similarly, the mean time trend of 0.6 TCs per decade 1995-2038 (RCP45) falls within the spread across CMIP5 models shown by Villarini and Vecchi (2012) for the first half 21$^{st}$ century. For the model trained on 1970-2017 HURDAT2, we deduce an increase by 2038 (RCP4.5) of 2.0 yr$^{-1}$ compared to 1986-2005, a 17% increase (2.0/12.4) that also falls within the 5%-95% quantile spread shown by Knutson et al (2020).

## 5. Conclusions

In summary, we have used a statistical model trained on historical TCs and large-scale SST indices to project US hurricane activity into the 2030s, given ensemble-mean climate-model projections of the SST indices under warming scenarios. We find a complex signal of changing hazard along the US coast driven by changes in basin-wide TC frequency and changes in track characteristics. Some areas are forecast to experience increased hazard, such as the Gulf coast of Florida, and other areas are forecast to experience decreased hazard, such as parts of the western Gulf and the US mid-Atlantic.

In our analysis, changes in U.S. TC activity are driven by projected secular increase in tropical SST and, indirectly, changes in other factors impacting TC activity, such as steering winds, wind shear, and spatial SST patterns, via their covariance with the large-scale SST indices. We stress that, in addition to these changes, there is large, well-documented interannual TC variation driven by internal modes of climate variability, in particular the AMO (e.g., Goldenberg et al., 2001) and El Nino/Southern Oscillation (ENSO) (e.g., Gray, 1984). Ensemble-mean climate-model SST projections provide no prediction for these important modes of variability, and we

559 do not include their effects here. From year to year, variations in TC activity driven by strong
560 ENSO events can be larger than the secular variations discussed here, and on a multi-year
561 timescale, a changing AMO phase may have an even larger impact. It has not been our goal
562 here to estimate the impact on TCs of natural variability, such as the AMO and ENSO, important
563 as they are. There are likely additional variations at small scales in wind shear, steering winds,
564 and SST patterns affecting TC activity that are not correlated with GSST and RSST, and the
565 impact of these variations is also not directly represented here, except through their impact on
566 uncertainty estimates. Instead, our aim has been to estimate the impact on TCs into the 2030s
567 of projected large-scale secular climate trends that exist in addition to the natural variability.
568 Estimation of trends in TC hazard in the coming decades, as well has the impact of natural
569 variability, has important applications to estimating the changing risk to populations and
570 infrastructure in harm's way.
571
572 **Data Availability:** The HURDAT2 data on which NASHM is trained is freely available for
573 download at the National Hurricane Center website, as is the historical SST data at the Hadley
574 center website. CMIP5 model-projected SST data are also freely available on several online
575 platforms.
576
577 **Acknowledgments:** This research was supported by NASA (National Aeronautics and Space
578 Administration), NOAA (National Oceanic and Atmospheric Administration), and The Climate
579 Service. TMH's effort was funded by NASA Modeling, Analysis, and Prediction (MAP) Program

580 Award 19-MAP19-0005. We thank three anonymous reviewers for comments and ideas that led
581 to an improved manuscript.
582

583 **6. References**

584 Baldini, L. M., et al., Persistent northward North Atlantic tropical cyclone migration over the
585 past fine centuries. *Nature Sci. Rep.*, 6, 37522 (2016).

586 Bhatia, K., et al., Projected response of tropical cyclone intensity and intensification in a global
587 climate model. J. Clim., 31, 8281-8303 (2018).

588 Bister, M., K. A. Emanuel, Low frequency variability of tropical cyclone potential intensity, 1.
589 Interannual to interdecadal variability. *J. Geophys. Res*., 107, 4801 (2002).

590 Bleomendaal, N., et al., Generation of a global synthetic tropical cyclone hazard dataset using
591 STORM. Sci Data, 7, 40, (2020).

592 Brown, J. N., C. Langlais, A. S. Gupta, Projected sea surface temperature changes in the
593 equatorial Pacific relative to the Warm Pool edge. *Deep-Sea Res II*, 113, 47-58 (2015).

594 Chauvin, F., P. Romain, P. Palany, A. Belmadani, Future changes in Atlantic hurricanes with the
595 rotated-stretched ARPEGE-Climat at very high resolution. *Clim. Dyn.,* doi:10.1007/s00382-
596 019-05040-4 (2019).

597 Colbert, A. J., B. J. Soden, G. A. Vecchi, B. P. Kirtman, The impact of anthropogenic climate
598 change on North Atlantic tropical cyclone tracks. *J. Climate*, 26, 4088-4095 (2013).

599 Coumou, D., et al., The weakening summer circulation in the Northern Hemisphere mid-
600 latitudes. *Science*, 348, 324-327 (2015).

601 Efron, B., Bootstrap methods: Another look at the jackknife. *Ann Stat*, 7, 1-26 (1979).

602 Elsner, J. B., J. P. Kossin, T. H. Jagger, The increasing intensity of the strongest tropical cyclones,
603 *Nature*, 455, 92-95 (2008).

604 Emanuel, K., R. Sundarajan, J. Williams, Hurricanes and global warming: Results from
605 downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc*., 89, 347-368 (2008).
606 Emanuel, K., A. H. Sobel, Response of tropical sea surface temperature, precipitation, and
607 tropical cyclone-related variables to changes in global and local forcing. *J. Adv. Model. Earth*
608 *Sys*., 5, 1-12 (2013).
609 Gillett, N. P., P. A. Stott, B. D. Santer, Attribution of cyclogenesis region sea surface temperature
610 change to anthropogenic influence. *Geophys. Res. Lett*., 35, L09707,
611 doi:10.1029/2008GL033670 (2008).
612 Goldenberg, S. B., et al., The recent increases in Atlantic hurricane activity. *Science*, 293, 474-
613 479 (2001).
614 Gray, W. M., Atlantic seasonal hurricane frequency. Part I: El Nino and the 30-mb quasi-biennial
615 oscillation influences. *Mon. Wea. Rev*., 112, 1649-1668 (1984).
616 Gutmann, D., et al., Changes in hurricanes from a 13-yr convection-permitting pseudo-global
617 warming simulation. *J. Climate*, 31, 3643-3657 (2018).
618 Hall, T. M., S. Jewson, Statistical modeling of North Atlantic tropical cyclone tracks. *Tellus*, 59A,
619 486-498 (2007).
620 Hall, T. M., E. Yonekura, North-American tropical cyclone landfall and SST: A statistical model
621 study. *J. Climate*, 26, 8422-8439 (2013).
622 Hall, T. M., K. Hureid, The frequency and duration of U.S. hurricane droughts. *Geophys. Res.*
623 *Lett.*, 42, 3482-3485 (2015).
624 Hall, T. M., M. Tippett, Pacific hurricane landfalls on Mexico and SST. *J. Appl. Meteorol.*
625 *Climatol*., 56, 667-676 (2017).

626 Hassanzadeh, P., et al., Effects of climate change on the movement of future landfalling Texas
627 tropical cyclones. *Nature Comm*., 11:3319, doi.org/10.1038/s41467-020-17130-7 (2020).
628 Held, I. M., and B. J. Soden, Robust changes of the hydrological cycle to global warming. *J. Clim*.,
629 19, 5686-5699 (2006).
630 James, M. K., L. B. Mason, Synthetic tropical cyclone database. *J. Waterw. Port C Div*, 131, 181-
631 192 (2005).
632 Kossin, J. P., S. J. Camargo, M. Sitkowski, Climate modulation of North Atlantic hurricane tracks.
633 *J. Climate*, 23, 3057-3076 (2010).
634 Kossin, J. P., T. L. Olander, K. R. Knapp, Trend analysis with a new global record of tropical
635 cyclone intensity. *J. Climate*, 26, 9960-9976 (2013).
636 Kossin, J. P., Validating atmospheric reanalysis data using tropical cyclones as thermometers.
637 *Bull. Amer. Meteorol. Soc*. ,1089-1096 (2015).
638 Kossin, J. P., T. M. Hall, T. Knutson, K. E. Kunkel, R. J. Trapp, D. E. Waliser, M. F. Wehner.
639 Extreme Storms, *Climate Science Special Report: Fourth Annual National Climate*
640 *Assessment, Vol.* I. D. J. Weubbles, et al, Eds., U.S. Global Climate Change Research Program,
641 pp. 256-276 (2017).
642 Kossin, J. P., K. R. Knapp, T. L. Olander, Emergence of trends in global tropical cyclone intensity.
643 *J. Climate*, submitted (2020).
644 Knutson, T. R., F. Zeng, A. T. Wittenberg, Multimodel assessment of regional surface
645 temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *J. Climate*, 26, 8709-
646 8743 (2013).

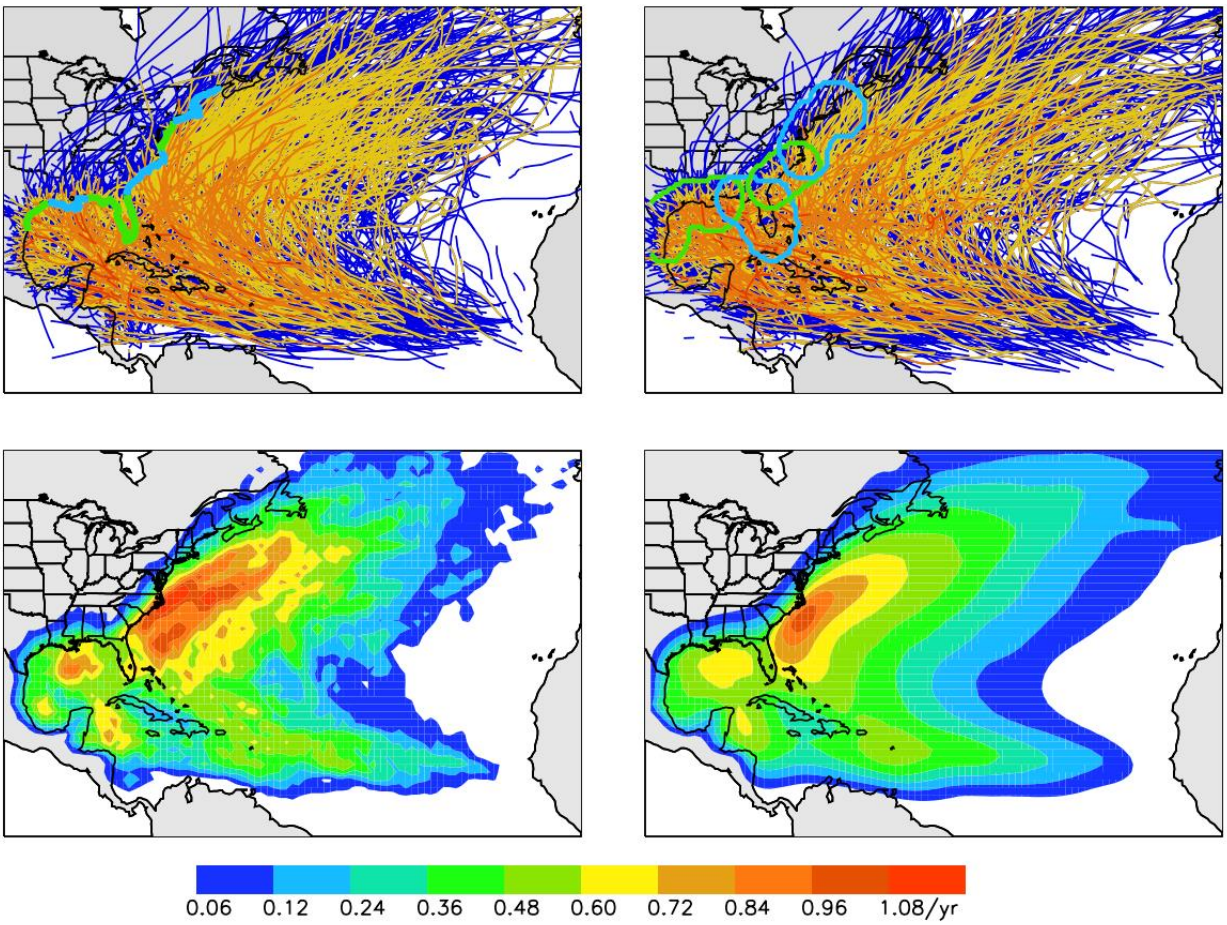
647 Knutson, T. R., S. J. Camargo, C. L. Chan, K. Emanuel, C.-H. Ho, J. P. Kossin, M. Mohapatra, M.
648 Satoh, M. Sugu, K. Walsh, L. Wu, Tropical cyclones and climate change assessment: Part I:
649 Detection and attribution. *Bull. Amer. Meteor. Soc*., 100, 1987-2007 (2019).
650 Knutson, T. R., S. J. Camargo, C. L. Chan, K. Emanuel, C.-H. Ho, J. P. Kossin, M. Mohapatra, M.
651 Satoh, M. Sugu, K. Walsh, L. Wu, Tropical cyclones and climate change assessment: Part II.
652 Projections. *Bull. Amer. Meteor. Soc*., 101, E303-E322 (2020).
653 Landsea, C. W., J. L. Franklin, Atlantic hurricane database uncertainty and presentation of a new
654 database format. *Mon. Wea. Rev*., 141, 3576-3592 (2013).
655 Landsea, C. W., Counting Atlantic tropical cyclones back to 1900. *EOS Trans., Amer. Geophys.*
656 *Union*, 88, 197-208 (2007).
657 Landsea, C. W., et al., Impact of duration thresholds on Atlantic tropical cyclone counts. *J.*
658 *Climate*, 23, 2508-2519 (2010).
659 Lau, W. K. M., J. J. Shi, W. K. Tao, K. M. Kim, What would happen to Superstorm Sandy under
660 the influence of a substantially warmer Atlantic. *Geophys. Res. Lett*., 43, 802-811,
661 doi:10.1002/2015GL067050 (2016).
662 Lee, C.-Y., M. K. Tippett, A. H. Sobel, S. J. Camargo, An environmentally forced tropical cyclone
663 hazard model. J. Adv. Earth Syst., 10, 223-241 (2018)
664 Liu, C., et al., Continental-scale convection-permitting modeling of the current and future
665 climate of North America. *Climate Dyn*., 49, 71-95 (2017).
666 Mann, M. E., B. A. Steinman, S. K. Miller, Absence of internal multidecadal and interdecadal
667 oscillations in climate model simulations. *Nature Comm*, 11, 49 (2020).

668 Moftakhani, H. R., G. Salvadori, A. AghaKouchak, B. F. Sanders, R. A. Matthew, Compounding
669 effects of sea level rise and fluvial flooding. *Proc. Nat. Acad. Sci*, 114, 9785-9790 (2017).
670 Murakami, H., E. Levin, T. L. Delworth, R. Gudgel, P.-C. Hsu, Dominant effect of relative tropical
671 Atlantic warming on major hurricane occurrence. *Science*, 362, 794-799 (2018).
672 Murakami, H., B. Wang, Future change of North Atlantic tropical cyclone tracks: projection by a
673 20-km-mesh global atmospheric model. *J. Climate*, 23, 2699-2721 (2010).
674 Murakami, H., et al., Detected climatic change in global distribution of tropical cyclones. Proc.
675 Natl. Acad. Sci, 117, 10706-10714 (2020).
676 Raynor, N. A., et al., Global analysis of sea surface temperature, sea ice, and night marine air
677 temperature since the late nineteenth century. *J. Geophys. Res*., 108, 4407,
678 doi:10.10239/2002JD002670 (2003).
679 Santer, B. D., et al., Forced and unforced ocean temperature changes in Atlantic and Pacific
680 tropical cyclogenesis regions. *Proc. Natl. Acad. Sci*., 103, 13905-13910 (2006).
681 Sobel, A. H., S. J. Camargo, T. M. Hall, C.-Y. Lee, M. K. Tippett, A. A. Wing, Human influence on
682 tropical cyclone intensity. *Science*, 353, 242-246 (2016).
683 Taylor, K. E., R. J. Stouffer, G. A. Meehl, An overview of CMIP5 and the experimental design.
684 *Bull. Amer. Meteor. Soc*., 93, 485-498 (2012).
685 Tebaldi, C., B. H. Strauss, C. E. Zervas, Modelling sea level rise impacts on storm surges along US
686 coasts. *Environ. Res. Lett*., 7, 014032 (2012).
687 Vecchi, G. A., B. J. Soden, Effect of remote sea surface temperature change on tropical cyclone
688 potential intensity. *Nature*, 450, 1066-1070 (2007).

689 Vecchi, G. A., K. L. Swansom, B. J. Soden, Whither hurricane activity. Science, 322, 687-689
690 (2008).
691 Vecchi, G. A., and T. R. Knutson, Estimating annual numbers of Atlantic hurricanes missing from
692 the HURDAT database (1878-1965) using ship track density. *J. Clim*., 24, 1736-1746 (2011).
693 Vecchi, G., et al, Tropical cyclone sensitivity to $CO_2$ doubling: roles of atmospheric resolution,
694 synoptic variability, and background climate changes. *Climate Dynamics*, 59, 5999-6033
695 (2019).
696 Vickery, P. J. P. F. Skerlj, L. A. Twisdale, Simulation of hurricane risk in the U.S. using empirical
697 track model*. J. Struct. Eng.*, 126, 1222-1237 (2000).
698 Villarini, G., and G. A. Vecchi, Projected increases in North Atlantic tropical cyclone intensity
699 from CMIP5 models, J. Clim., 26, 3231-3240 (2013).
700 Wang, C., H. Liu, S.-K. Lee, R. Atlas, Impact of the Atlantic warm pool on United States
701 landfalling hurricanes, *Geophys. Res. Lett*., 38, L19702 (2011).
702 Yoshida, K., M. Sugi, R. Mizuta, H. Murakami, M. Ishii, Future changes in tropical cyclone activity
703 in high-resolution large-ensemble simulations. *Geophys. Res. Lett*., 44, 9910-9917 (2017).
704
705
706
707
708

| | NW Gulf | Florida | Southeast | Northeast | US Total | N Atlantic |
|---|---|---|---|---|---|---|
| **Cat0** | **+, 5%** | **+, 8%** | **+, 0%** | **+, 8%** | **+, 6%** | **+, 8%** |
| **Cat1-2** | **+, 15%** | **+, 15%** | **+, 10%** | **-, -6%** | **+, 4%** | **+, 12%** |
| **Cat3-5** | **+, 23%** | **+, 51%** | **+, 79%** | **+, 92%** | **+, 71%** | **+, 56%** |

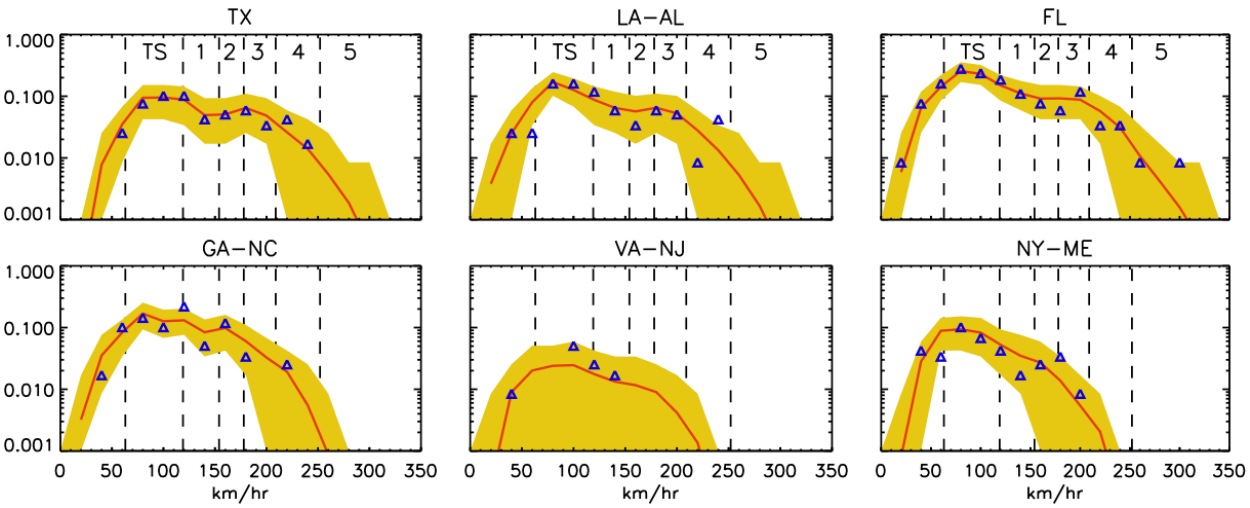
**Table 1:** Out-of-sample predictive skill assessment on the nine years with GSST anomaly above 1.5 sigma by coastal region and intensity range according to two measures: accumulated log likelihood and F-score based on precision and recall. In each entry, a plus indicates the climate-dependent model has a higher total log likelihood that the naïve model at predicting rates of TS occurrence within 350km of the indicated region and intensity range, while minus indicates the climate-dependent model is not more skillful than the naïve model. Numbers for F-scores are the percentage increase or decrease in the climate-sensitive F-score compared to the naïve F-score.




720

721 **Fig 1:** Illustration of the historical TC data used to train NASHM, NASHM simulations, and
722 evaluation of NASHM's US TC landfall characteristics. Top left: HURDAT2 1900-2017 TC tracks
723 upon which the stochastic model is trained. Tracks are color coded by maximum sustained wind
724 speed ($V_{max}$): TS and higher (blue), Saffir-Simpson Cat1 and higher (yellow), Cat3 and higher
725 (orange), Cat5 (red). Also shown are landfall regions used in subsequent analysis. Top right: An
726 example NASHM stochastic simulation of the 1900-2017 period. Also shown are skill-
727 assessment regions used in subsequent analysis. Bottom: Annual track density (occurrence per
728 year) in 1° boxes for TS and higher TCs. Left is HURDAT2 1900-2017, corresponding to the tracks
729 of the upper left panel. Right is the average over 1000 1900-2017 simulations, one of which is

730 shown in the upper left panel. White regions correspond to rates below the indicated dark blue
731 contour.
732
733




734
735

736 **Fig 2:** Distribution of annual landfall rates as functions of $V_{max}$ in 20km/hr bins on six US coastal
737 regions, as labeled by US states and as mapped in the top left panel: Texas, Louisiana through
738 Alabama, Florida, Georgia through North Carolina, Virginia through New Jersey, and New York
739 through Maine. Red is the NASHM mean from 1000 simulations of 1900-2017, yellow is the
740 inner 90% about the mean, and blue is computed directly from 1900-2017 HURDAT2. Vertical
741 dashed lines indicate Saffir-Simpson category boundaries, as labeled in the TX, LA-AL, and FL
742 panels: Tropical Storm (TS) and categories 1, 2, 3, 4, and 5. Note that for bins with zero
743 historical events a blue symbol does not appear.
744

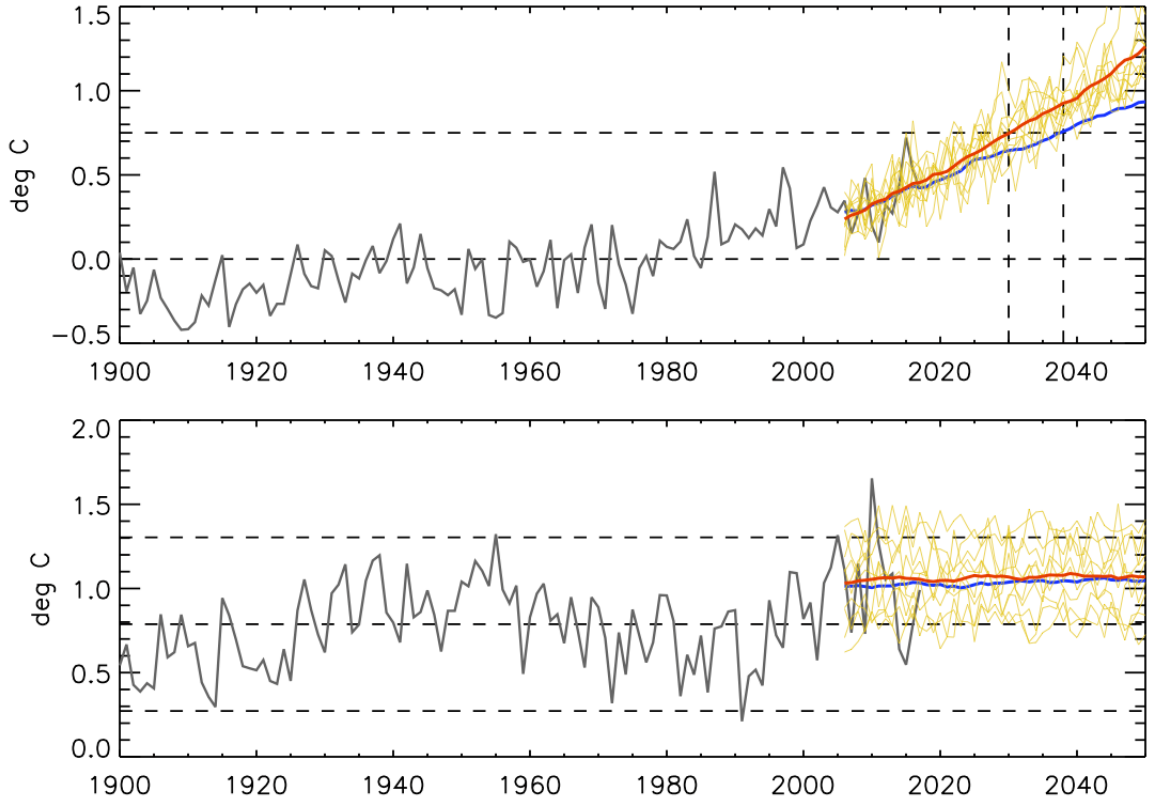



**Fig 3:** The SST-index covariates. Black represents the historical annual time series. Blue represents the RCP4.5 CMIP5 ensemble annual mean and red the RCP8.5 CMIP5 ensemble annual mean. The top panel is GSST (anomaly compared to the 1900-2017 mean), and the bottom panel is RSST. The yellow lines are individual CMIP5 ensemble members for RCP8.5. In the top panel, the lower dashed line indicates the zero value, and the upper dashed line indicates the highest value over the historical period. The vertical dashed lines indicate the years 2030 and 2038 at which the ensemble means exceed the highest historical value for RCP8.5 and RCP4.5, respectively. In the bottom panel, the center dashed line is the 1900-2017 mean, and the lower and upper dashed lines indicate -2 and +2 standard deviations about the mean.

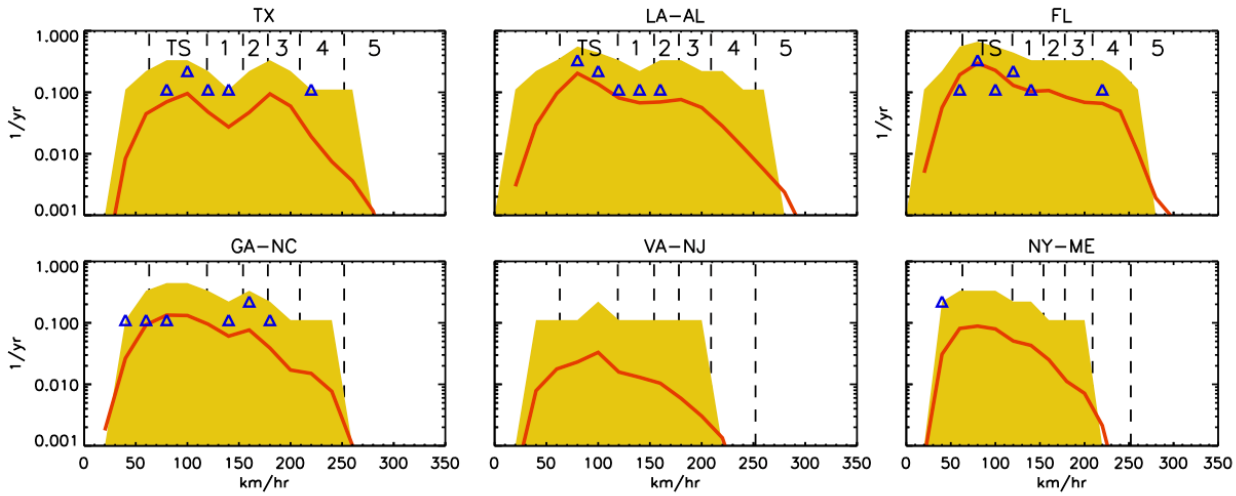
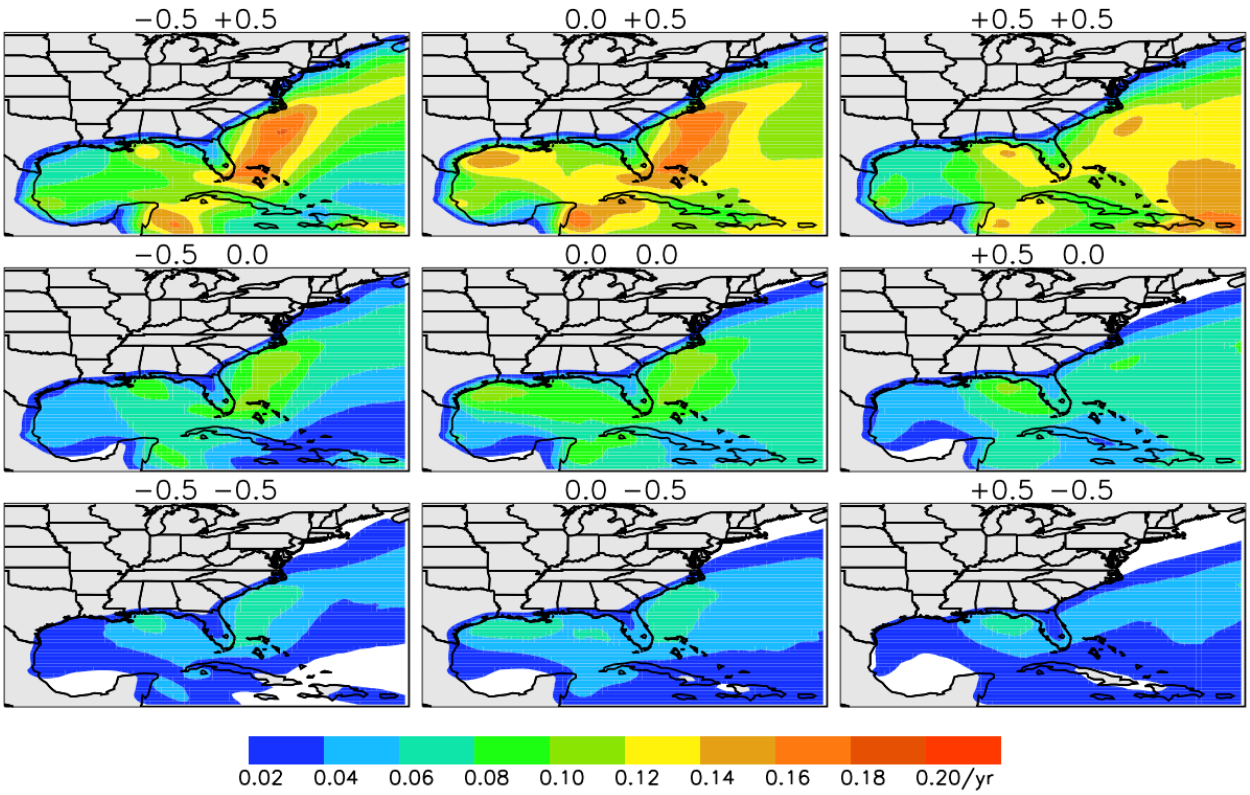



**Fig 4:** As in Fig 2, but here using the for simulations of the nine out-of-sample years (1987, 1997, 1998, 2003, 2009, 2014, 2015, 2016, 2017) of the forecast skill test. Note that for many Vmax bins (and all of VA-NJ), there are zero events in the nine out-of-sample years, an event count falls inside the inner 90% model spread.




765

766 **Fig 5:** Cat1+ annual occurrence rate maps as a function GSST and RSST anomaly, as labeled.
767 Anomaly labels represent degrees C difference from the 1900-2017 means. GSST is the first
768 label and RSST the second. White regions correspond to rates below the indicated dark blue
769 contour.

770

771

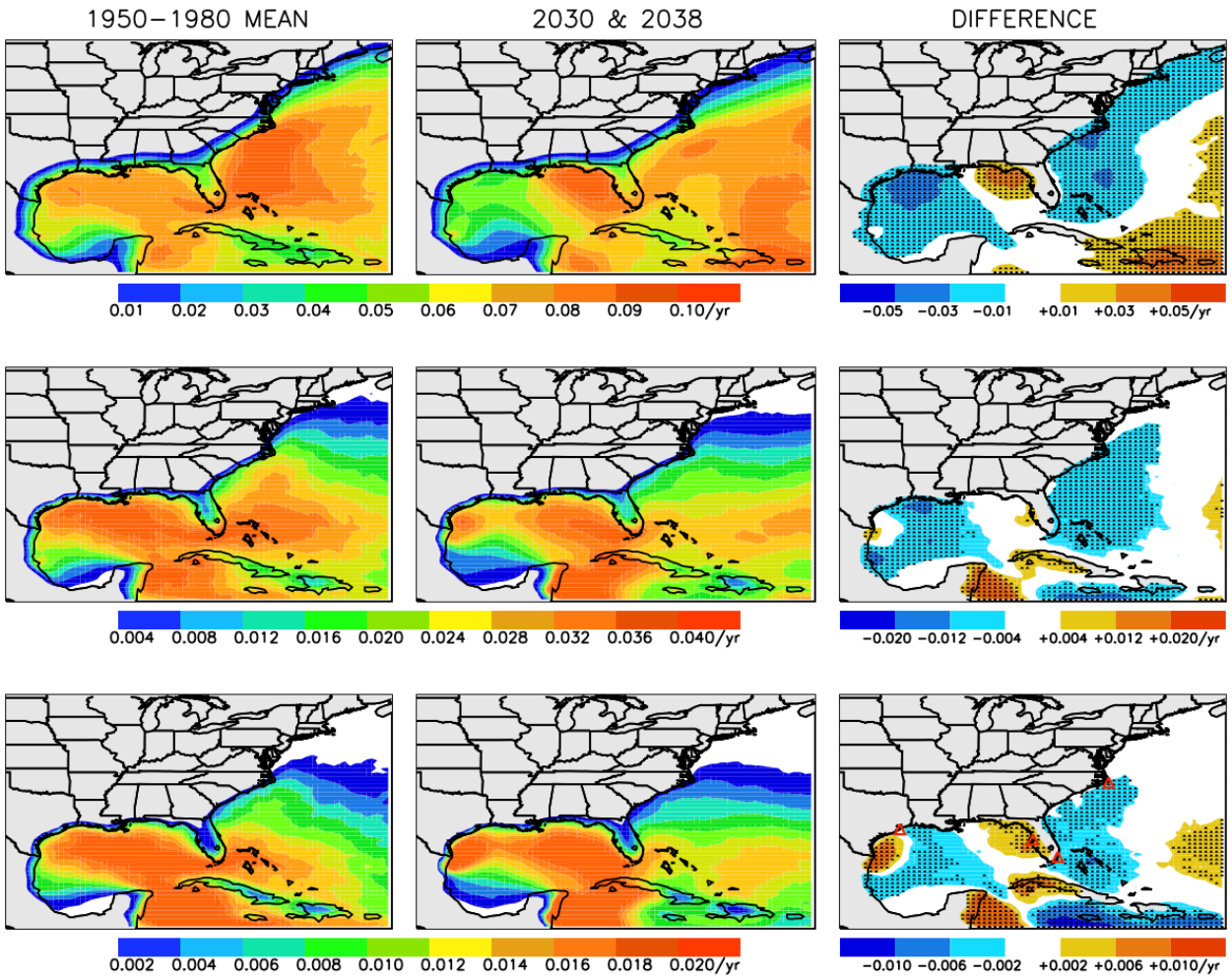
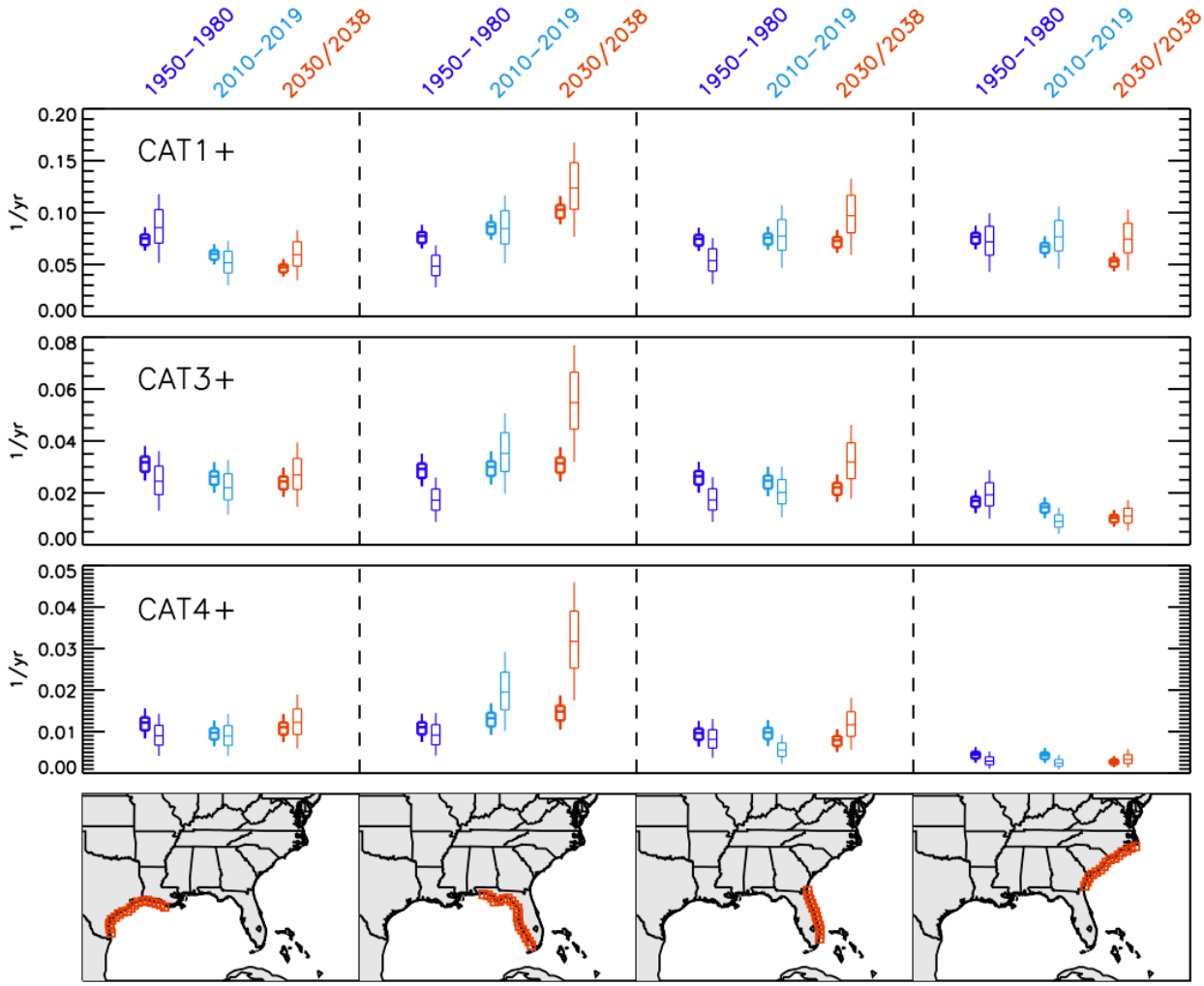



**Fig 6:** Maps of annual TC annual occurrence rates in units of $yr^{-1}$ for 1950-1980 (left), 2030 (RCP85) and 2038 (RCP45) (center), and the difference (right). Top row represents TCs of intensity Cat 1 and greater, middle row Cat 3 and greater, and the bottom row Cat 4 and greater. The contour scales change row to row, as indicated by the color scales. In the left and center columns, white regions have rates below the dark blue scale indicated on the color bars. In the right column, symbols indicate points where the differences are estimated to be significant at 95% confidence by a bootstrap analysis.




780

781 **Fig 7:** Annual occurrence rates averaged over 50km-radius circles spanning four coastal
782 locations, as shown in bottom row. The top row represents Cat1+, second row Cat3+, and third
783 row Cat4+. Dark blue represents the period 1950-1980, light blue the period 2010-2019, and
784 red the RCP85 projected rates for 2030 (identical to the RCP45 projected rates for 2038). Box
785 and whiskers indicate the best-estimate rates and the 22%-78% and 5%-95% confidence ranges
786 about the best estimate determined from a bootstrap re-sampling analysis. Heavy-lines
787 indicates results based on 1900-2017 HURDAT2, and light lines indicate results based on 1970-
788 2017 HURDAT2.

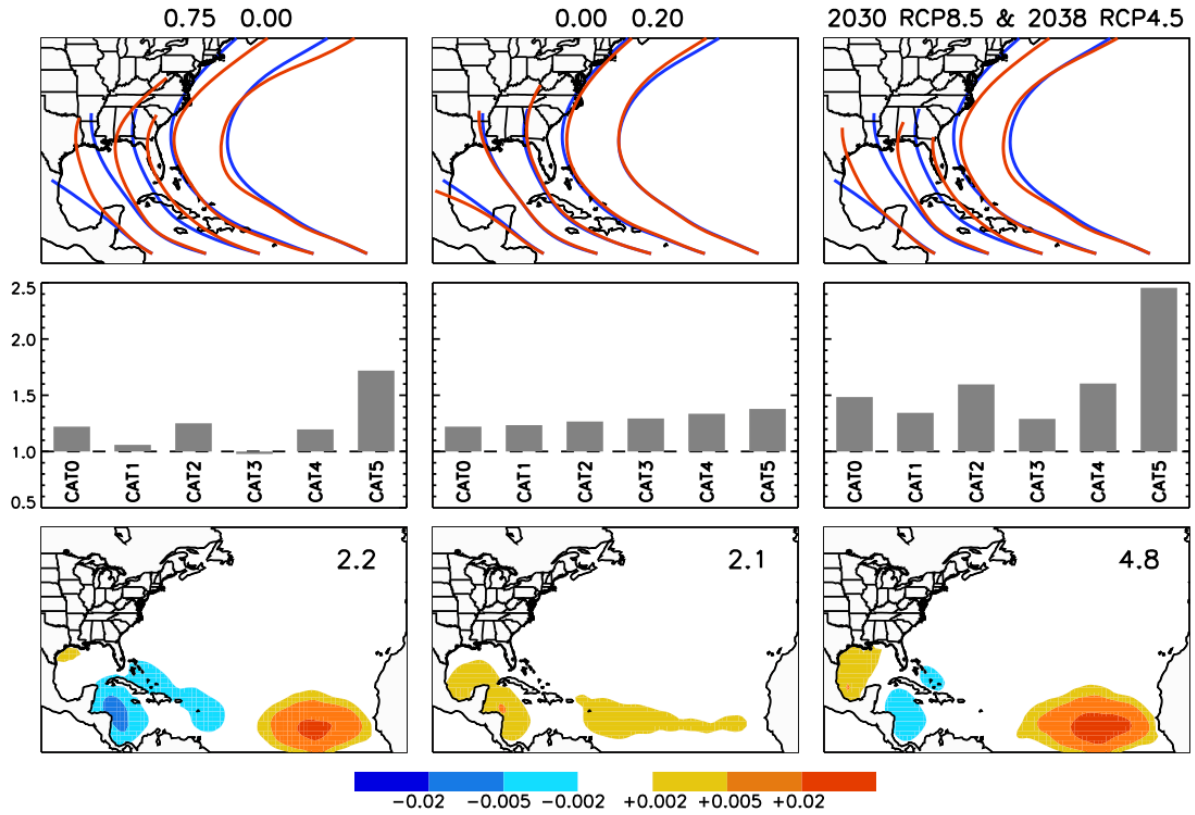
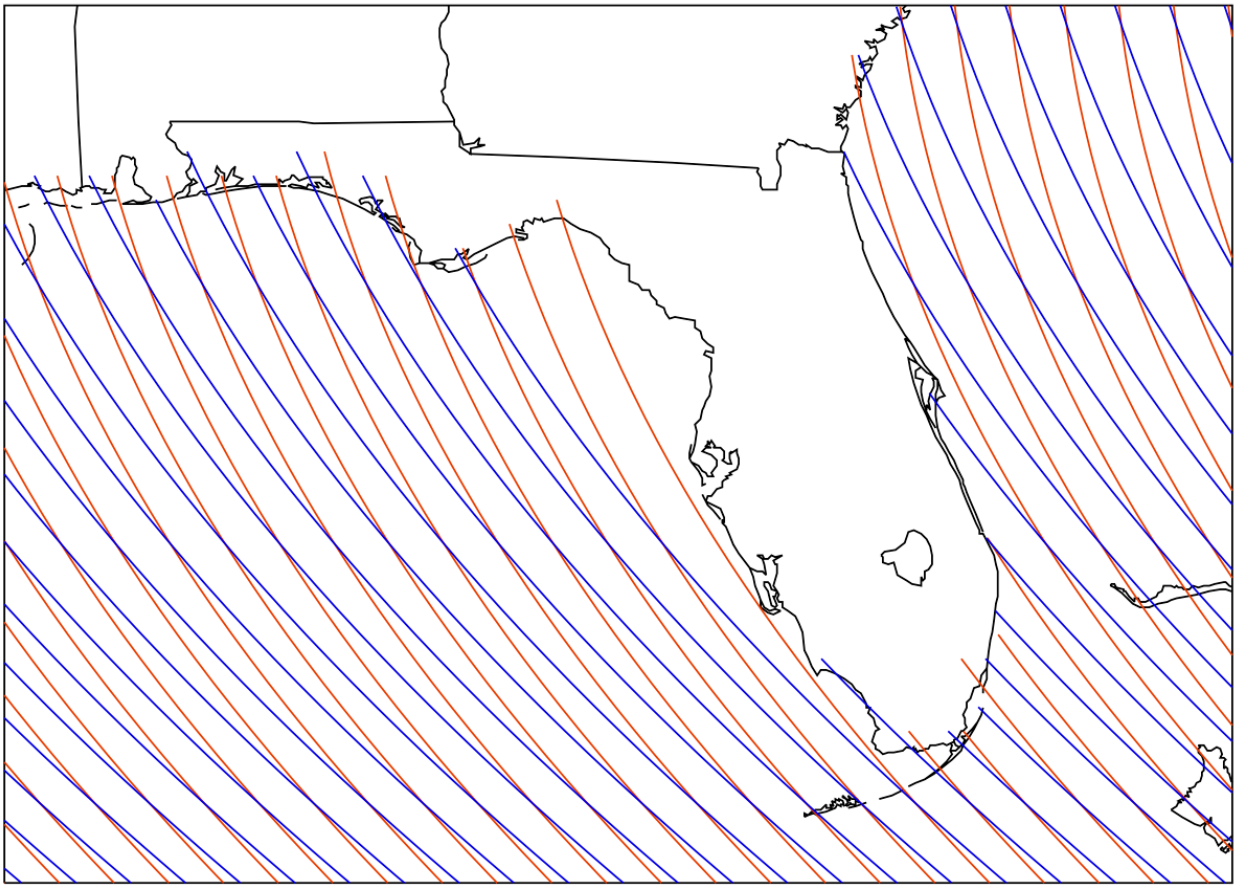



**Fig 8:** Illustration of influence of covariates GSST and RSST on TC tracks (top row), lifetime maximum intensity (LMI, middle row), and formation rate (bottom row). GSST-RSST anomaly values with respect to 1900-2017 are 0.75°C & 0.00°C (left column), 0.00°C & 0.20°C (center column), and 0.75°C & 0.20°C (right column). The 0.75°C & 0.20°C combination corresponds to the 2030 RCP8.5 & 2038 RCP4.5 values, as labeled at top. For tracks, the stochastic component of NASHM is turned off for clarity, leaving only mean tracks, which are launched from five specified locations. In each panel, blue is the neutral state and red the GSST-RSST combination. For LMI, intensity is binned in Saffir-Simpson categories, and the ratios with respect to the neutral state are plotted. For TC formation, the difference between the GSST-RSST combination and the neutral state is mapped, with red positive and blue negative. Contour units $yr^{-1}$ per 1°

801 latitude-longitude grid box. Differences in basin-wide annual TC formation are inset top right of

802 each panel.

803



804

805 **Fig 9:** Idealized set of tracks illustrating the regional impacts of basin-wide propagation changes.

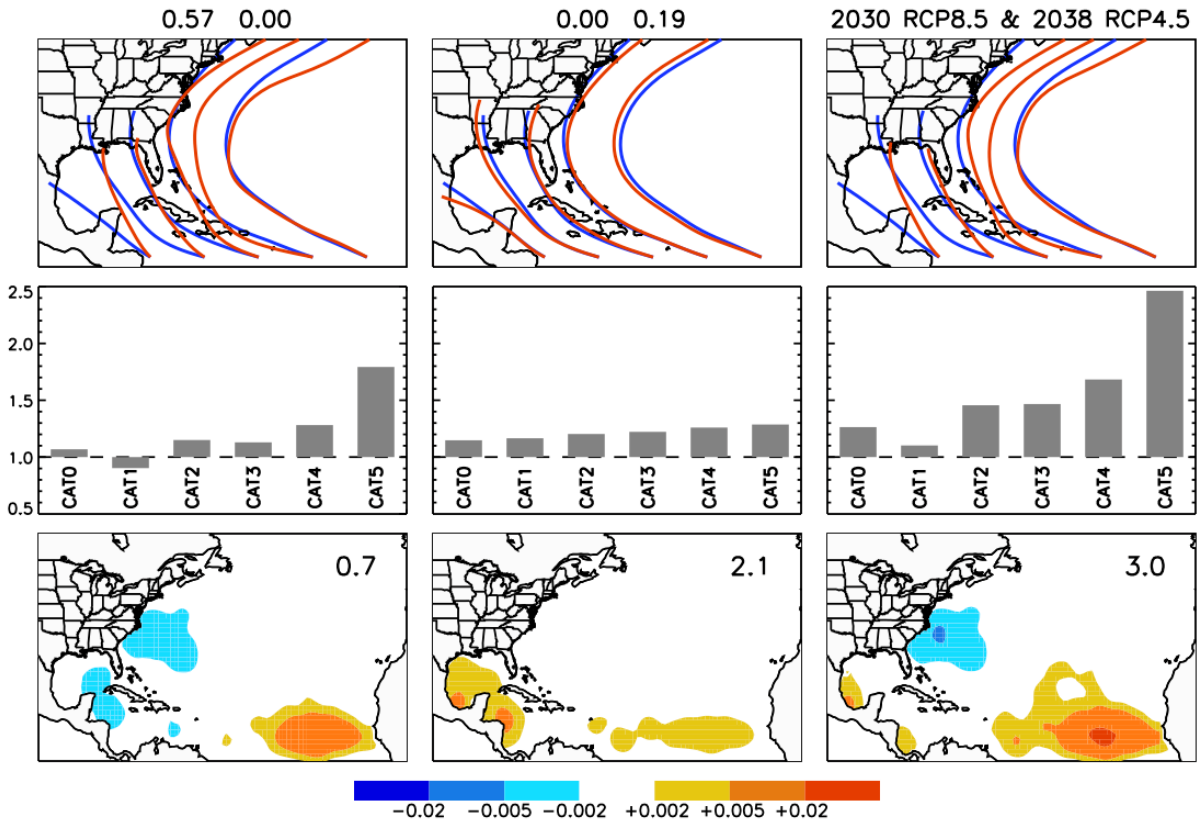
806 Two sets of mean tracks start from the same origin points. Red differs from blue only in a

807 constant uniform propagation anomaly of an 5km eastward for every 25km north. Florida's Gulf

808 coast is more exposed to the red tracks, while its Atlantic coast is less exposed.

809

810




811

812 **Fig 10:** As in Fig 8, but here for the model trained only on HURDAT2 from 1970 to 2017.

813

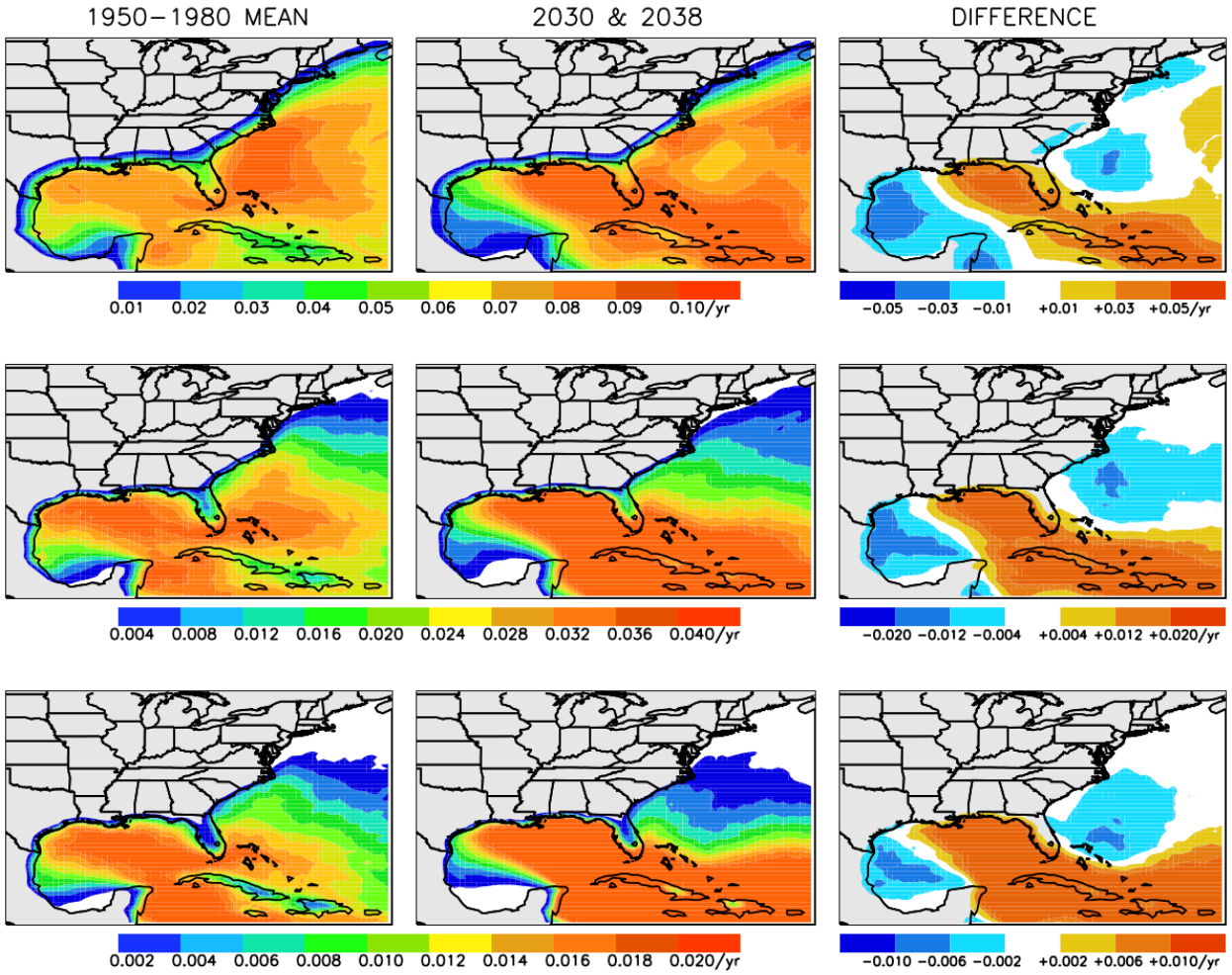



**Fig 11:** As in Fig 6, but here the projected maps are based on NASHM trained on HURDAT2 1970-2017, instead of 1900-2017.