A Decadal Climate Service for Insurance: Skillful Multiyear Predictions of North Atlantic Hurricane Activity and U.S. Hurricane Damage

JULIA F. LOCKWOOD,[a] NICK DUNSTONE,[a] LEON HERMANSON,[a] GEOFFREY R. SAVILLE,[b,c] ADAM A. SCAIFE,[a,d] DOUG SMITH,[a] AND HAZEL E. THORNTON[a]

[a] *Met Office Hadley Centre, Exeter, United Kingdom*
[b] *WTW Research Network, London, United Kingdom*
[c] *Canopius, London, United Kingdom*
[d] *University of Exeter, Exeter, United Kingdom*

(Manuscript received 14 September 2022, in final form 30 May 2023, accepted 5 June 2023)

ABSTRACT: North Atlantic Ocean hurricane activity exhibits significant variation on multiannual time scales. Advance knowledge of periods of high activity would be beneficial to the insurance industry as well as society in general. Previous studies have shown that climate models initialized with current oceanic and atmospheric conditions, known as decadal prediction systems, are skillful at predicting North Atlantic hurricane activity averaged over periods of 2–10 years. We show that this skill also translates into skillful predictions of real-world U.S. hurricane damage. Using such systems, we have developed a prototype climate service for the insurance industry giving probabilistic forecasts of 5-yr-mean North Atlantic hurricane activity, measured by the total accumulated cyclone energy (ACE index), and 5-yr-total U.S. hurricane damage (given in U.S. dollars). Rather than tracking hurricanes in the decadal systems directly, the forecasts use a relative temperature index known to be strongly linked to hurricane activity. Statistical relationships based on past forecasts of the index and observed hurricane activity and U.S. damage are then used to produce probabilistic forecasts. The predictions of hurricane activity and U.S. damage for the period 2020–24 are high, with ~95% probabilities of being above average. We note that skill in predicting the temperature index on which the forecasts are based has declined in recent years. More research is therefore needed to understand under which conditions the forecasts are most skillful.

SIGNIFICANCE STATEMENT: The purpose of this article is to explain the science and methods behind a recently developed prototype climate service that uses initialized climate models to give probabilistic forecasts of 5-yr-mean North Atlantic Ocean hurricane activity, as well as 5-yr-total associated U.S. hurricane damage. Although skill in predicting North Atlantic hurricane activity on this time scale has been known for some time, a key result in this article is showing that this also leads to predictability in real-world damage. These forecasts could be of benefit to the insurance industry and to society in general.

KEYWORDS: North America; Hurricanes/typhoons; Climate prediction; Forecasting

1. Introduction

When North Atlantic Ocean hurricanes reach land, the impacts can be devastating, including loss of life and widespread destruction of homes and infrastructure. They are also one of the leading causes of global insured losses: for example, in 2017, Hurricanes Harvey, Irma, and Maria contributed more than 60% of total global insured losses (including all human-made and natural disasters), and Hurricane Katrina in 2005 has the highest insured loss on record for a single event ($82 billion indexed to 2017; Swiss Re Institute 2018).

North Atlantic hurricane activity is known to exhibit multiannual time scale variation (e.g., Molinari and Mestas-Nuñez 2003; Chylek and Lesins 2008; Klotzbach and Gray 2008), for example, there was notably low hurricane activity in the 1970s–80s, with associated low U.S. hurricane damage followed by a very active and damaging period from the mid-1990s to mid-2000s (Pielke and Landsea 1998; Nyberg et al. 2007; Weinkle et al. 2018). Prediction of hurricanes on multiannual time scales is therefore of potential use to insurance companies, allowing them to better prepare for periods of high activity. The 5-yr time scale is chosen in particular as it smooths out annual peaks in activity, and, since legislation prevents large swings in pricing on interannual time scales, it can therefore be used to inform longer-term pricing strategies. It also captures decadal variability, informing insurers how the current climate may differ from their historical records. In fact, following the costly 2004 and 2005 hurricane seasons, three major global catastrophe model vendors developed near-term (5 yr) hurricane predictions for use by the insurance industry (Karen Clark and Company 2010; Pielke 2009; AIR Worldwide 2015), demonstrating the clear market interest in predictions on this time scale. These predictions used expert elicitation, statistical models, and sea surface temperature trends (e.g., Jewson et al. 2009; Toumi and Restell 2014; AIR Worldwide 2015), but when the high activity predicted

Denotes content that is immediately available upon publication as open access.

Corresponding author: Julia F. Lockwood, julia.lockwood@metoffice.gov.uk

DOI: 10.1175/JAMC-D-22-0147.1

for 2006–10 failed to materialize, two of the companies abandoned their efforts in this area (AIR Worldwide 2015).

To our knowledge, the insurance industry has not yet made use of *initialized* decadal predictions using physically based climate models, which in the following years were shown to be highly skillful in predicting North Atlantic tropical storm and hurricane activity (Smith et al. 2010; Caron et al. 2014, 2015, 2018). The skill comes from both external forcing and initialization with the current state of the ocean and atmosphere. For external forcing, the high anthropogenic aerosol levels in 1970s–80s are believed to have caused the hurricane "drought" in that period (Booth et al. 2012; Dunstone et al. 2013; Rousseau-Rizzi and Emanuel 2022). Initialization, on the other hand, particularly enhances sea surface temperature skill in the North Atlantic subpolar gyre, a key region known to be related to North Atlantic tropical cyclone activity (Goldenberg et al. 2001; Dunstone et al. 2011; Hermanson et al. 2014).

In this paper we describe how we have used a multimodel decadal prediction system ensemble to make probabilistic 5-yr forecasts of hurricane activity over the North Atlantic basin [measured by the accumulated cyclone energy (ACE)][1] and the associated U.S. damage (measured in U.S. dollars, adjusted to 2020). Rather than tracking tropical cyclones/hurricanes in the models directly, the forecasts use a relative temperature index known to be strongly linked to hurricane activity (e.g., Vecchi and Soden 2007; Swanson 2008; Vecchi et al. 2008; Ramsay and Sobel 2011; Dunstone et al. 2013). An important aspect of this work is that the index is predicted from the fundamental dynamics of the climate model, unlike the aforementioned near-term predictions developed by the insurance industry that used statistical methods to predict sea surface temperatures. Statistical relationships between past forecasts (hindcasts) of the index and observed hurricane activity and U.S. damage are then used to produce probabilistic forecasts of these measures.

The rest of the paper is structured as follows: In section 2 we describe the model and observational data used for this study, and in section 3 we outline the statistical verification methods used. Section 4 describes the methods used to generate the ACE and damage forecasts and gives an assessment of their skill, and describes development of the final forecast product. In section 5 we discuss the limitations of the predictions and present the conclusions.

2. Data

a. Model (forecast) and reanalysis data

A 40-member multimodel ensemble consisting of 4 dynamical decadal prediction systems (10 members each) was used to make the hurricane predictions. The systems used are based on fully coupled atmosphere–ocean dynamical models: HadGEM3-GC3.1 (Williams et al. 2018), CMCC-CM2-SR5 (Cherchi et al. 2019), EC-Earth3 (Bilbao et al. 2021), and MPI-ESM-HR (Müller et al. 2018).

Each ensemble member is initialized with the current state of the atmosphere, ocean, and land surface every November from 1960 to 2018 to form a set of retrospective forecasts (hindcasts), and run for 10 years. The forecast for 2020–24 is calculated from the runs initialized in November 2019. All models use the full-field initialization method for the atmosphere and ocean, with the exception of MPI-ESM-HR, which uses anomaly initialization in the ocean. All models are forced with the CMIP6 historical greenhouse gas, aerosol and natural forcings from 1960 to 2014, and the Shared Socioeconomic Pathway 2-4.5 (SSP2-4.5) scenario thereafter (O'Neill et al. 2016). [A full description of the models is given in https://climate.copernicus.eu/sites/default/files/2021-09/Technical_appendix_2020.pdf; see Müller et al. (2018) for details of the MPI-ESM-HR model.]

To calculate the relative temperature index described in section 3, we extract monthly mean near surface temperature (tas) for years 1–5 of the forecast for the North Atlantic hurricane season [June–November (JJASON)]. Ideally sea surface temperature would have been used but these data were not readily available at the time the analysis was undertaken. This has minimal impact on the results, since the sea surface temperatures and tas are strongly coupled on 5-yr time scales, especially in the tropics [their correlations in the Met Office decadal prediction system (DePreSys4) members over the North Atlantic hurricane main development region (MDR) and tropical region (TROP) (defined in section 4), where both variables were available, are greater than 0.99].

The ERA5 reanalysis dataset (Hersbach et al. 2020) is used as the "observations" to assess the skill in predicting tas and the relative temperature index. To assess the observed relationship between the relative temperature index with ACE and hurricane damage over a longer time period we use sea surface temperature data from the HadISST dataset (Rayner et al. 2003), for 1900–2020.

b. Hurricane and loss data

North Atlantic tropical cyclone and hurricane numbers and total ACE are from the Atlantic National Hurricane Center HURDAT2 database (accessed 2020; Landsea and Franklin 2013), for cyclones formed in the North Atlantic basin over the hurricane season (JJASON), over the period 1900–2020. We have attempted to correct for missing tropical cyclones in the presatellite era (pre-1965) using the estimated missing tropical cyclone numbers from Vecchi and Knutson (2008), and assuming the missing cyclones have average ACE.

U.S. hurricane damage, which include both insured and uninsured losses, is compiled from the NOAA NCEI U.S. Billion-Dollar Weather and Climate Disasters database (Smith and Katz 2013) for the period 1980–2020. Damage from 1961 to 1979 is compiled from the Annual Summaries of North Atlantic Storms (NOAA Miami Regional Library 2021).

[1] ACE is defined as $10^{-4}\sum_{i,t} v_{i,t}^2$, where $v_{i,t}$ is the estimated maximum 1-min wind speed of cyclone i at 6-hourly time interval t, when the cyclone is at tropical storm intensity ($v_{i,t} > 33$ kt) or greater.

The damage is adjusted to 2020 equivalent U.S. dollars, taking into account inflation and changes in wealth and population, using the method described in Pielke and Landsea (1998) and Weinkle et al. (2018, hereinafter W18). Assuming that the populations of the counties affected by hurricanes follow the same proportional changes as the population of the United States as a whole[2] and using gross domestic product the measure of wealth, the equation in Pielke and Landsea (1998) simplifies to

$$\mathrm{NL}_{2020} = L_y \times \frac{\mathrm{GDP}_{2020}}{\mathrm{GDP}_y},$$

where NL_{2020} is the annual damage normalized to 2020, L_y is the loss recorded in year y, and GDP_{2020} and GDP_y are the nominal gross domestic products (i.e., the values recorded at the time) in 2020 and year y, respectively. This simplification arises because the nominal GDP ratio implicitly includes the population and inflation increase. Gross domestic product data are from the Bureau of Economic Analysis (2021).

We also use the normalized loss data from W18, covering the period 1900–2017, to assess the relationship between damage and the relative temperature index in observations over this longer time period. The two damage datasets are very similar for the overlapping years (see section 4c). The largest discrepancy is for the damage of 2017, since the W18 dataset only includes the conterminous United States and thus excludes the damage from Hurricane Maria, which caused major damage to Puerto Rico.

3. Verification methods

Following the Copernicus Climate Change Service (C3S) Sectoral Application of Decadal Forecasts "Recommendations on forecast quality assessment for decadal predictions," we include assessment of both the deterministic and probabilistic forecast skill (see also Delgado-Torres et al. 2022). We use the Spearman rank correlation coefficient ρ to assess the deterministic skill of the forecasts so as to take into account the nonlinear relationship between U.S. damage and forecast median U.S. damage (see section 4c), although the standard Pearson correlation coefficient gives similar results.

We use the Brier skill score (BSS) for measuring the probabilistic skill of predicting above- or below-average ACE and U.S. damage, defined as

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}_{\mathrm{fc}}}{\mathrm{BS}_{\mathrm{ref}}},$$

where $\mathrm{BS}_{\mathrm{fc}}$ and $\mathrm{BS}_{\mathrm{ref}}$ are the Brier scores of the forecast and a reference forecast. The Brier score is defined as

$$\mathrm{BS} = \frac{1}{n}\sum_{k=1}^{n}(f_k - o_k)^2,$$

where n is the number of observed–forecast pairs, f_k is the forecast probability (range 0–1) of above-average ACE/damage for the 5-yr period k, and o_k the observed occurrence ($o_k = 1$ if the observed ACE/damage was above average and is 0 otherwise). A BSS of 1 indicates a perfect forecast, and BSS > 0 indicates that the forecast performs better than the reference. Here the reference used is 5-yr persistence, where the forecast probability of above-average ACE/damage is equal to 1 if the previous five years had above-average ACE/damage.

Contingency tables for forecasting above- or below-average ACE/U.S. damage are also shown. A forecast is considered to have predicted above-average ACE/loss if the forecast probability is greater than 0.5. The hit rate is the ratio of correct forecasts to the number of times this event occurred [hits/(hits + misses)] and the false alarm rate is the ratio of false alarms to number of nonoccurrences of the event [false alarms/(false alarms + correct rejections)]. When the hit rate is high and the false alarm rate is low, the forecast is said to exhibit discrimination, that is, the ability to distinguish situations leading to the occurrence of an event from those leading to the nonoccurrence of the event (Wilks 2019b). Forecasts with good discrimination can be useful for decision making, although required hit rate/false alarm thresholds depend on the application of the forecasts and the cost of action or inaction.

The correlations and Brier skill scores are tested for statistical significance using the nonparametric block bootstrapping approach as described in Smith et al. (2020), using a block length of 5 years to take into account autocorrelation of the data.

As the forecast probability distributions are based on a statistical relationship between past forecasts of the temperature index and observations, we use cross validation to assess out-of-sample skill. The ACE/damage for each 5-yr period y are estimated from the statistical relationship between observed ACE/damage and forecast temperature index for data points with no common years with the period y (see section 4b).

4. ACE and U.S. damage forecasts

To make the ACE and damage predictions, we first calculate a relative temperature index from the ensemble mean of the decadal prediction systems as an indicator of predicted hurricane activity. A description of the index and its predictability is described in section 4a. The statistical models described in sections 4b and 4c then transform the index into probabilistic 5-yr-mean ACE index and 5-yr-total total U.S. damage predictions. In section 4d we describe the final forecast product.

a. Relative temperature index

The relative temperature index used to indicate North Atlantic hurricane activity, MDR–TROP, is defined as

$$\mathrm{MDR\text{–}TROP} = T_{\mathrm{MDR}} - T_{\mathrm{TROP}},$$

[2] Use of U.S. population rather than that of the affected counties, and GDP as a measure of wealth instead of "fixed reproducible tangible wealth" as in Pielke and Landsea (1998), has little effect on the correction factor: the correlation coefficient between the 5-yr-total losses in this work and the updated losses presented in W18 [which use the original Pielke and Landsea (1998) equation] is 0.96 over the common time period.

where T_{MDR} is the 5-yr-mean (JJASON only) temperature anomaly in the North Atlantic hurricane MDR (10°–25°N, 80°–20°W), and T_{TROP} is the 5-yr-mean (JJASON only) temperature anomaly over the tropical oceans (averaged over 30°S–30°N).

This index, and similar variants, have often been used to estimate annual and multiannual North Atlantic hurricane activity at this time scale (e.g., Vecchi and Soden 2007; Swanson 2008; Vecchi et al. 2008, 2011; Dunstone et al. 2013; Caron et al. 2018). Its use is justified by considering the conditions necessary for hurricane formation: warm local sea surface temperatures and moist, unstable air, low vertical wind shear, and high vorticity (Gray 1998; Emanuel and Nolan 2004; Emanuel 2010). All of these conditions can be satisfied in the MDR, which lies on the northward flank of the intertropical convergence zone (ITCZ). Although high surface temperatures increase the probability of hurricane formation locally, high temperatures elsewhere in the tropics, for example, the Pacific or Indian Ocean, lead to increased atmospheric stability and increased wind shear in the tropical Atlantic due to changes in the Walker circulation, impeding hurricane development (Latif et al. 2007). In addition, Dunstone et al. (2013) showed that positive values of MDR–TROP are associated with a northward shift in the ITCZ and Hadley cell (and hence the subtropical jet), again reducing vertical wind shear in the MDR. The increase in vorticity with a more northerly ITCZ (due to the increased planetary contribution) also favors hurricane formation (Merlis et al. 2013; Sobel et al. 2021).

We note that there has been some debate in the literature on whether decadal hurricane predictions based on this relative SST index would have been able to predict the shift from low to high hurricane activity in the mid-1990s (Vecchi et al. 2013). However, the analysis present in Smith et al. (2014) shows that such systems would have provided clear evidence for an impending reversal to a period of above-average hurricane frequency had they been available in the mid-1990s, before the observed increase occurred.

The correlation between 5-yr-mean observed tas and ACE over the hindcast period is shown in Fig. 1a, showing a strong positive relationship between ACE and tas in the North Atlantic. The correlation in this region is still present after linearly detrending the data (Fig. 1c) and is particularly strong in the MDR region (marked by the black-outlined box) and the subpolar gyre region in the far North Atlantic (around Iceland and the southern tip of Greenland). Figures 1b and 1d show the correlation skill in predicting 5-yr-mean tas. Correlations are generally very high across the globe and part of this skill can be attributed to the warming trend, but Fig. 1d shows the much of the predictive skill in the North Atlantic remains after linearly detrending the data. The detrended skill is particularly strong in the North Atlantic MDR and subpolar gyre regions, which are the most important regions for predicting hurricane activity as shown in Figs. 1a and 1c.

Figure 1e shows the time series of observed 5-yr-mean MDR–TROP and North Atlantic ACE, showing an overall correlation of 0.70 (using ERA5 data; $p < 0.05$) over the hindcast period. Both measures show positive trends over the hindcast period, although the longer-term time series from 1900 to 2018 (using HadISST SST data to calculate MDR–TROP), show that the long-term trend is small. The correlation holds over this longer period ($\rho = 0.61$), with both time series showing troughs for approximately 1900–20 and 1965–95, and peaks for approximately 1920–65 and after 1995. This result adds further evidence that there is a relationship between MDR–TROP and hurricanes, rather than the correlation being caused by positive trends in both measures.

The correlation of the multimodel ensemble mean in predicting MDR–TROP is 0.67 ($p < 0.05$; Fig. 1f), demonstrating significant skill in predictions of this index. Given the high correlations shown in Figs. 1e and 1f, we expect the predicted MDR–TROP by the models to be a skillful predictor of observed ACE, which is indeed the case: The correlation between the predicted MDR–TROP and observed ACE is 0.77 ($p < 0.01$; see section 4b).

We note that Fig. 1f shows the MDR–TROP is poorly predicted at the end of the time series (central years 2014–18). During this period the observed MDR–TROP and observed ACE time series also diverge. This is discussed further in section 5.

We also investigated the skill of several other North Atlantic proxy hurricane indices that have previously been mentioned in the literature, including the lambda index (which is also based on the MDR–TROP relative temperature; Vecchi et al. 2011; Caron et al. 2014, 2015, 2018), the Atlantic Multidecadal Oscillation (AMO) index developed by Klotzbach and Gray (2008), and a relative temperature index based on the temperatures in the subpolar gyre region (SPG; from −50° to −10°E and from 50° to 60°N; Klotzbach and Gray 2008; Caron et al. 2018) and the tropics (SPG–TROP). We find little difference in the skill between the above indices in predicting North Atlantic ACE, with all model-predicted indices having a correlation with observed ACE of approximately 0.7–0.8. In addition, all indices show the same discrepancy between model and observed values for 2014–18 as found for MDR–TROP.

b. *ACE forecasts*

The relationship between observed 5-yr-mean ACE index and 5-yr-mean forecast MDR–TROP is quantified using ordinary least squares linear regression on past data. Rather than using the ensemble spread to estimate the forecast ACE prediction intervals, the prediction intervals are estimated from the residuals of the linear regression fit [Wilks 2019a, his Eq. (7.23)]. This method has successfully been used in seasonal forecasts (e.g., Bett et al. 2018, 2022) and has the advantage that it implicitly recalibrates the forecast probability distribution to the observed variability.

The linear regression fit is performed on nonoverlapping 5-yr-mean data, to avoid overconfidence in the prediction intervals. There are five sets of nonoverlapping 5-yr-means (e.g., the first set is the means of 1961–65, 1966–70, 1971–75, etc.; the second set is the means of 1962–66, 1967–71, 1972–76, etc.). Each set has 11–12 data points (10–11 when performing cross validation). The linear-regression is performed on each

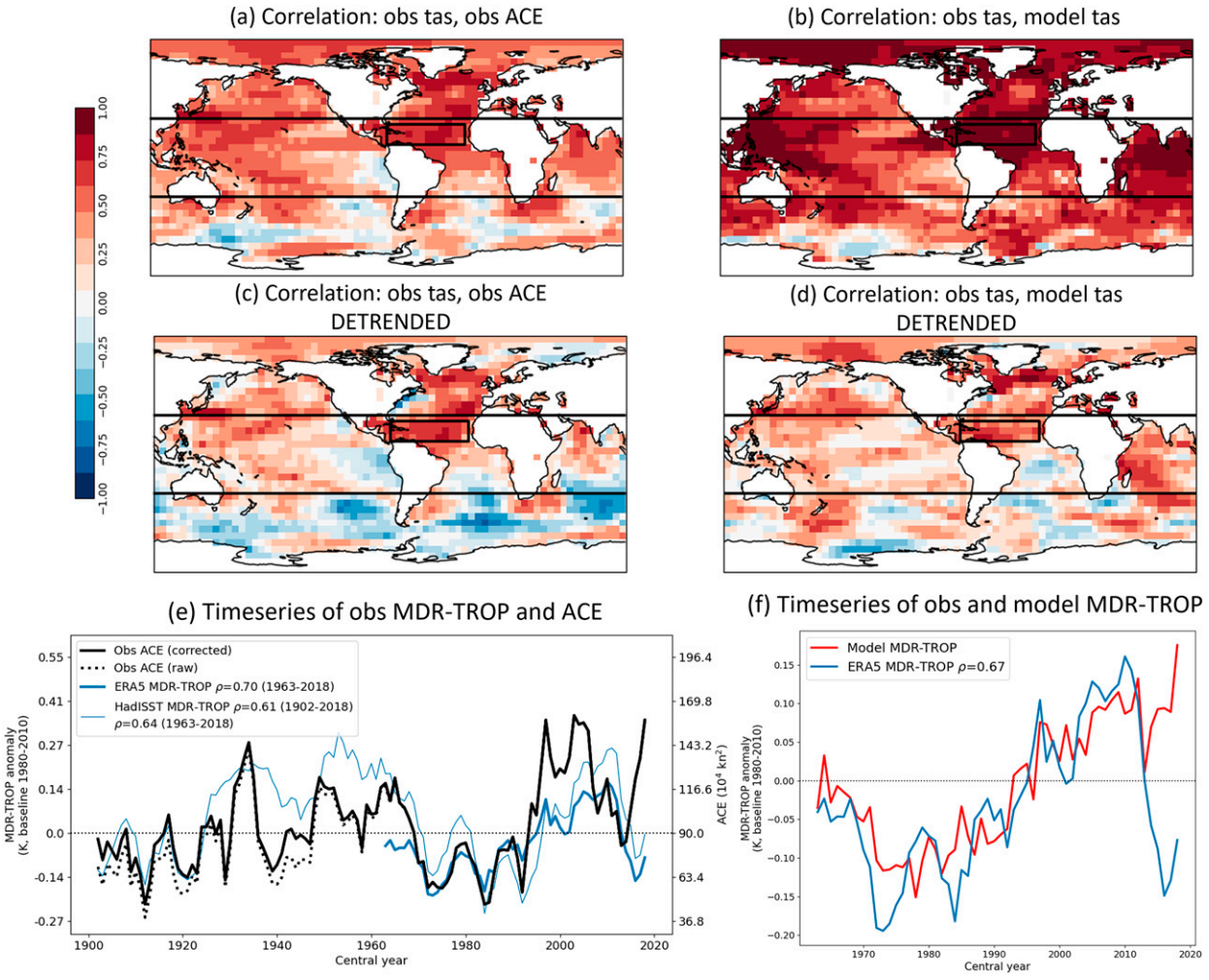

FIG. 1. Relative temperature index (MDR–TROP) as predictor of hurricane activity: (a) Correlation of observed 5-yr-running-mean tas and observed 5-yr-running-mean North Atlantic hurricane activity (ACE) over the model hindcast period. (b) Correlation of predicted 5-yr-running-mean tas (from multimodel ensemble mean) and observed 5-yr-running-mean ACE. (c),(d) As in (a) and (b), but with all time series linearly detrended before calculating correlations. In (a)–(d), the Pearson correlation coefficient is shown, and ERA5 is used for the observations. (e) Time series of observed 5-yr-running-mean ACE (black, showing the raw HURDAT data with the dotted line, and the corrected HURDAT data with the solid line), and 5-yr running mean MDR–TROP calculated from ERA5 tas (thick blue line), and HadISST SST (thin blue line). (f) Time series of observed (ERA5) and predicted 5-yr-running-mean MDR–TROP index (ensemble mean). The rank correlation coefficient between time series is shown in the legends in (e) and (f).

set separately, and we use the mean fit coefficients for the line of best fit and prediction interval estimation. The linear regression and 75% and 95% prediction intervals are shown in Fig. 2a, and the resulting predicted and observed ACE time series are shown in Fig. 2b.

The correlation between forecast and observed ACE is $\rho = 0.77$ ($p < 0.01$), and 0.72 ($p = 0.01$) using cross validation. The correlation remains significant after linearly detrending both variables ($\rho = 0.68$). The skill of persistence (using the last five years to predict the next five years) is $\rho = 0.42$, significantly lower than the model forecasts (the difference has p values of 0.02 and 0.12 with and without cross validation, respectively). The Brier skill score of predicting above- or below-average ACE, using 5-yr persistence as the reference forecast, is 0.62 ($p = 0.02$), and 0.56 ($p = 0.06$) using cross validation. The contingency table for predicting above-average ACE (Table 1) demonstrates the forecasts are able to accurately discriminate between above- and below-average ACE index, with a high hit rate (80%) and low false alarm rate (10%).

The forecast for 2020–24 (initialization date November 2019) is shown with the box-and-whisker plot in Figs. 2a and 2b. The forecast is for a very active period, with a 5-yr-mean prediction of 145×10^4 kt^2 (1 kt ≈ 0.51 m s^{-1}) (75% prediction intervals of 110×10^4–180×10^4 kt^2; 95% prediction intervals of 85×10^4–205×10^4 kt^2) and a 95% chance of above-average ACE. Note that there are some caveats to this forecast, discussed in section 5.

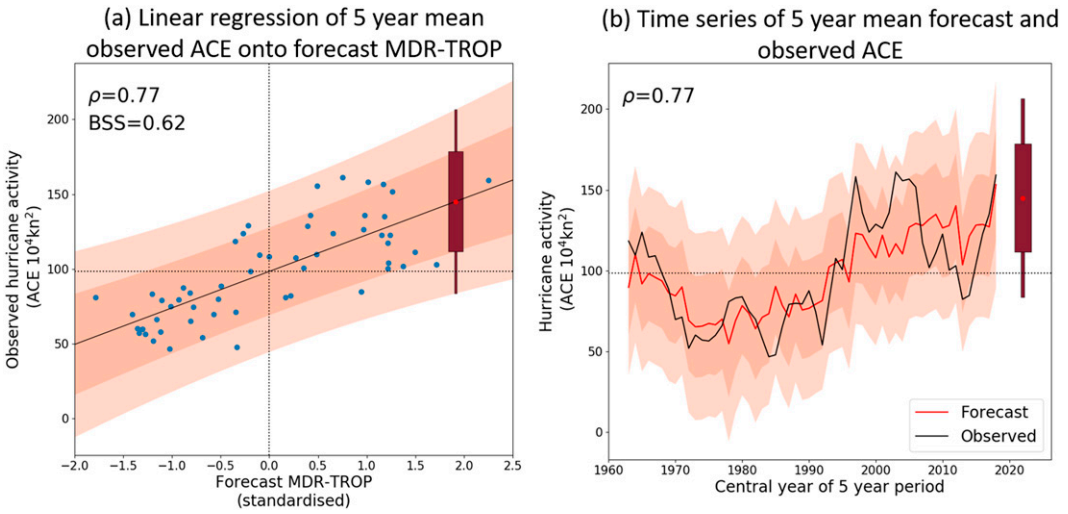

FIG. 2. ACE forecast: (a) Scatterplot showing observed 5-yr-mean hurricane activity (ACE) against the predicted 5-yr-mean MDR–TROP from the ensemble mean of the decadal prediction systems. The rank correlation and BSS are shown at the top left. The solid line shows the linear regression. (b) Time series of observed and forecast 5-yr-mean ACE. In (a) and (b), the shading shows the 75% and 95% prediction intervals. The forecast for 2020–24 is shown with the red dot, with the box and whiskers showing the 75% and 95% prediction intervals.

c. *U.S. damage forecasts*

As for the ACE forecast, we use a statistical relationship between observed damage and the forecast MDR–TROP index based on past data to make the damage prediction. The time series of 5-yr-total U.S. hurricane damage over the hindcast period is shown in Fig. 3a, along with the time series of observed and forecast MDR–TROP, and the scatterplot of 5-yr U.S. hurricane damage (on a logarithmic scale) against forecast MDR–TROP is shown in Fig. 3b. Although the rank correlation coefficient with forecast MDR–TROP is high (0.70; $p < 0.01$), the relationship is nonlinear, as can be seen in the scatterplot of 5-yr-total loss (on a linear scale) against forecast MDR–TROP in Fig. 3d. Two prominent jumps are seen in the loss time series, centered on the years 1992 and 2005. These peaks are caused by Hurricanes Andrew (1992) and Katrina (2005), and they also appear as outliers on the scatterplot in Fig. 3b, where points containing these years have been highlighted. It could be argued that data containing the 2017 hurricane season also have anomalously high damage, but in this case, it was not a single event causing the large losses.

The observed 5-yr U.S. damage and observed MDR–TROP are plotted over the long-term period 1902–2015 (central years) in Fig. 3c (using W18 U.S. damage data and HadISST MDR–TROP). Figure 3c confirms that the relationship holds over a longer time period and with alternative datasets, and that the high correlation is not simply due to positive trends over the hindcast period. In the prehindcast period, a further three jumps are seen in the damage time series, caused by the cyclones Galveston (1900), Galveston (1915), the great Miami storm (1926). The correlation between observed 5-yr U.S. damage and MDR–TROP is 0.34 for longer period 1902–2015, but this increases to 0.66 after removing the damage from the five outlier storms.

TABLE 1. Contingency table for predicting above-average ACE index. The average is defined as the mean observed ACE over the validation period (1960–2018), equal to 98.4×10^4 kt^2.

		Observed	
		Yes	No
Predicted	Yes	23 hits	3 false alarms
	No	5 misses	25 correct rejections
Hit rate			80%
False alarm rate			10%

Given the apparent strong relationship between log(damage) and predicted MDR–TROP index, a simple linear regression between these measures was attempted. However, the resulting prediction intervals appeared unrealistically large for high values of forecast MDR–TROP: For example, for a standardized forecast MDR–TROP value of 2.25 (the highest in the time series), the 95% prediction upper limit was $2.2 trillion, equivalent to 11 Hurricane Katrinas happening in a single 5-yr period. A generalized linear model assuming a gamma distribution was also fit to the data using the Python Statsmodels package (Seabold and Perktold 2010), but again a satisfactory fit and prediction intervals could not be obtained. This is possibly because of the constraint of a constant shape parameter in the gamma distribution with MDR–TROP.

The difficulty of fitting loss probability distributions for extreme hazards is well known (Toumi and Restell 2014) and is one of the reasons why insurers resort to using catastrophe modeling, where thousands of extreme events and their resulting losses are simulated. One issue with hurricanes is that hurricanes with similar ACE can inflict vastly differing amounts of damage, for example, small changes in paths or size could lead to large differences in exposure (Czajkowski

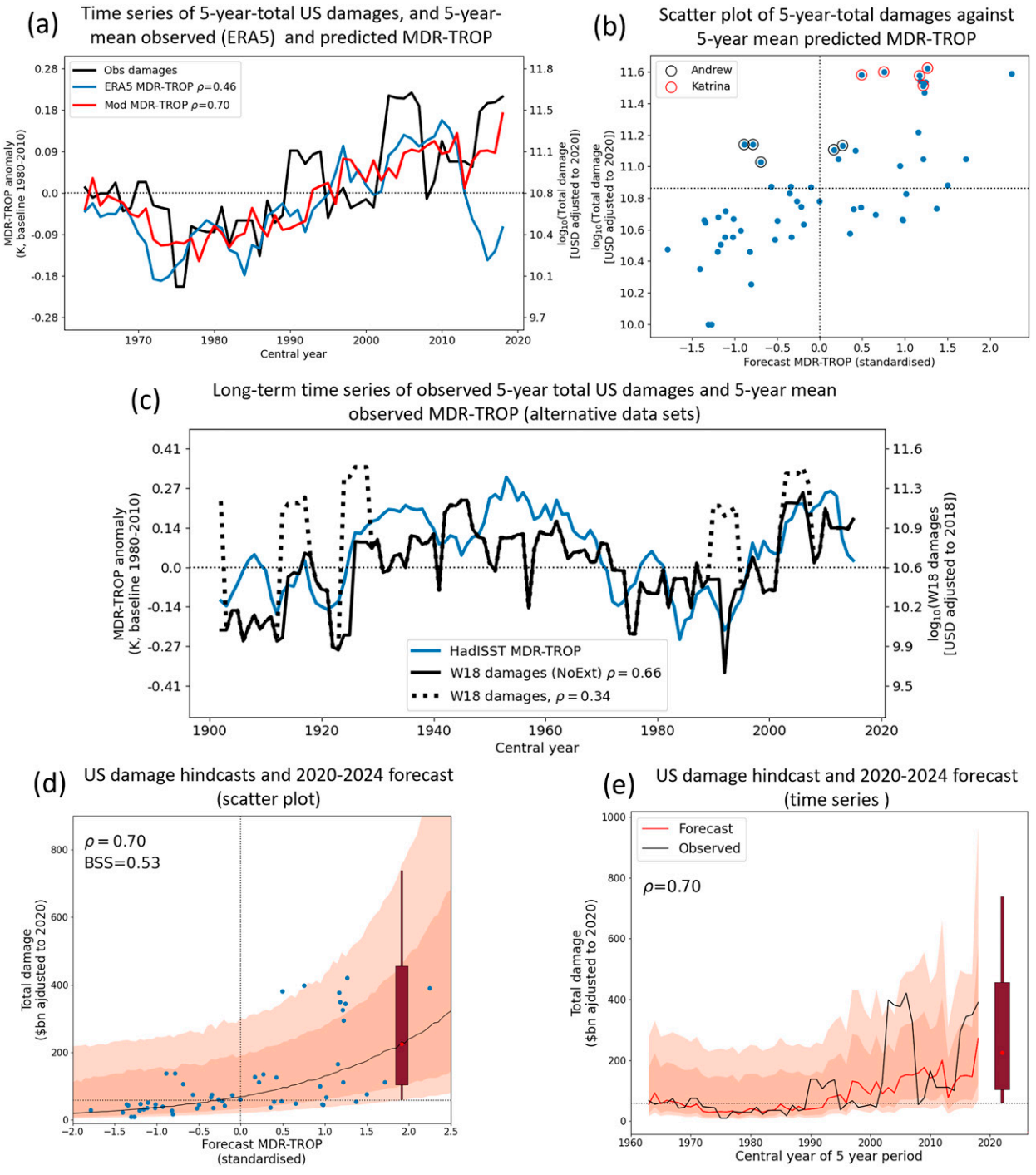

FIG. 3. U.S. damage forecast: (a) Time series of observed 5-yr-total U.S. damage (adjusted to 2020) and observed (ERA5) and predicted MDR–TROP. (b) Scatterplot of 5-yr-total U.S. damage (log scale) against forecast 5-yr-mean MDR–TROP, highlighting 5-yr periods that contain the damage from Hurricanes Andrew (points with black circles) and Katrina (points with red circles). (c) Long-term time series of observed MDR–TROP (calculated from HadISST) and 5-yr-total U.S. damage (adjusted to 2018; W18 data). The solid black line (NoExt) shows the data after damage from the extreme storms Galveston (1900), Galveston (1915), great Miami (1926), Andrew (1992), and Katrina (2005) have been removed. (d) Scatterplot of hindcast and 2020–24 forecast 5-yr-total U.S. damage [the data are the same as in (b) but on a linear scale]. (e) Time series of hindcast and 2020–24 forecast 5-yr-total U.S. damage. In (d) and (e), the solid line shows the median damage prediction, the shading shows the 75% and 95% prediction intervals, and the horizontal dashed line marks the long-term median damage. The forecast for 2020–24 is shown with the red dot, with the box and whiskers showing the 75% and 95% prediction intervals.

and Done 2014), or if one manages to breach flood defenses as in the case of Hurricane Katrina (Anderson et al. 2007). Such detailed catastrophe modeling is beyond the scope of this project. Instead, we have devised a method based on linear regression and random sampling to produce damage probability distributions for a given forecast MDR–TROP that take into account the unpredictable nature of extreme single events, such as those that cause the jumps in the damage time series, while retaining a dependency on MDR–TROP. The method is described below.

First, we remove the damage attributed to Andrew and Katrina from the data and perform linear regression of log(U.S. damage) onto forecast MDR–TROP. We then assume that such extreme single events occur randomly at the climatological frequency (two events in 60 years of data). To incorporate this into the predicted probability distributions, for a given MDR–TROP we estimate the "base" damage probability distribution from the linear regression, then randomly sample a "base" total damage for the 5-yr period. We then loop through each of the five years in the period, randomly sampling whether an extreme event occurs. If an extreme is selected, then its loss is added to the base damage. This gives the possibility of having more than one extreme-damage hurricane in a single 5-yr period. The value of damage for the extreme is sampled from a uniform probability distribution between \$90 billion[3] and \$210 billion (which are the approximate damage of Andrew and Katrina, respectively, after adjusting to 2020). This is repeated 5000 times for each MDR–TROP to build up a probability distribution as a function of MDR–TROP.

The resulting damage hindcast and forecast prediction intervals are shown in the scatterplot in Fig. 3d and time series in Fig. 3e. Although the method has involved estimations of unknown parameters such as the frequency and distribution of the extremes, it qualitatively reproduces the distributions we expect, that is, a low expected damage for low values of MDR–TROP, but with heavy tails indicating an ever-present risk of a random extreme, and high expected damage for high values of MDR–TROP, but without the upper limit of the distribution giving unrealistically high estimations. Note that the resulting predicted median loss and probability of above-average damage are insensitive to the choice extreme distribution: they remain relatively unchanged even after doubling the extreme upper limit to \$400 billion and doubling the frequency of extreme storms to 4 in 60 years. However, the predicted upper damage limits for low MDR–TROP years are sensitive to the choice of extreme distribution. This uncertainty is stated in the product and more research is needed to understand the probabilities of such extreme events.

The rank correlation coefficient of the median loss prediction is 0.70 ($p < 0.01$), and with the cross validation it is 0.66 ($p < 0.01$). The Brier skill score for predicting above-average loss is 0.53 ($p < 0.01$), and it is 0.50 with cross validation ($p = 0.01$). Note that the cross-validated skill scores have used knowledge of Hurricanes Andrew and Katrina for estimated the probability and distribution of extreme outlier events, but since the median and probability of above-average damage are insensitive to the extremes, this should have little effect on the cross-validated scores. Table 2 shows the contingency table for predicting above-average U.S. damage, demonstrating good discrimination with a high hit rate and low false alarm rate.

TABLE 2. Contingency table for predicting above-median 5-yr-total U.S. damage (adjusted to 2020). The median observed 5-yr-total U.S. damage (adjusted to 2020) is \$57.9 billion, calculated over the validation period (1960–2018).

		Observed	
		Yes	No
Predicted	Yes	21 hits	9 false alarms
	No	7 misses	19 correct rejections
Hit rate			75%
False alarm rate			30%

The prediction for 2020–24 is for total U.S. damage of \$230 billion (approximately \$46 billion yr^{-1}), with 75% and 95% prediction intervals of \$100–\$475 billion and \$55–\$770 billion, respectively, and greater than a 95% chance of being above average.

d. *Forecast product*

This work was part of the European Union–funded C3S project "Sectoral applications of decadal predictions," which was set up to develop climate services on decadal time scales (Dunstone et al. 2022) for users in agriculture (Solaraju-Murali et al. 2021, 2022), water management (Paxian et al. 2022), energy (Tsartsali et al. 2023), and insurance (this work). Up until this point there were very few services on this time scale, despite the skill of decadal predictions demonstrated in scientific literature and clear user needs (Kushnir et al. 2019).

This service was codeveloped with the insurance broker Willis Towers Watson. User needs were gathered in an initial meeting, and research results and drafts of the forecast product were shared throughout the project. Feedback on earlier versions of the forecast product included the need to make the forecast more accessible by reducing the use of meteorological jargon. The final product is a two-page document (available at https://climate.copernicus.eu/sites/default/files/2021-06/Insurance%20Product%20Sheet_2019.pdf) that follows the common template for all case studies in this project. Page 1 of the document gives the headline results, namely, that there will be 95% chance of above-average North Atlantic hurricane activity and >95% chance of above-average U.S. damage for the period 2020–24. This is followed by a more detailed forecast (giving the forecast ACE and U.S. damage median and 75% and 95% prediction intervals, and the time series shown in Figs. 2b and 3c to put the forecast into context). The second page briefly describes the method to generate the forecasts and

[3] For low forecast values of MDR–TROP, where the loss probability distribution from the linear regression is very narrow and centered on losses of ≪\$90 billion, the lower limit of the "extreme" is decreased to the 95% upper limit of damage from the linear regression. This avoids having a gap in the final probability distribution.

presents forecast verification measures (rank correlation, Brier skill score, and contingency tables). Note that the forecast statistics on the website differ slightly from those presented here, since more ensemble members became available during the writing of this paper.

The product was well received by the users, and they showed an interest in continuing the service and exploring ways to apply the information to business practices. In terms of usability, their view was that there would have to be a strong forecast signal, outweighing other market influences, to make a difference to business overall. Nevertheless, forecast products such as this can provide some guidance for how insurers may think about hurricane risk in the next few years, and can become part of the underwriting conversation in selecting which lines of business to grow or shrink. If predictions were available at higher resolutions, such that locations could be assessed across a portfolio of risks, further utility could be derived; however, this is currently beyond the scope of this product. More information on probabilities of extreme periods was also requested, although for the damage model in particular more research and verification are needed to be confident in the predicted upper percentiles, as these are highly dependent on extremes like Hurricanes Andrew and Katrina. Another suggestion for future versions of the forecast was to show predictions of the number of landfalling hurricanes or number of damaging events (defined as the number of events for which there is a recorded loss, which is similar, but not exactly equal to, the number of landfalling hurricanes). There is significant skill in predicting both of these measures ($\rho = 0.47$ with $p < 0.05$ and $\rho = 0.50$ with $p < 0.01$, respectively).

5. Discussion and conclusions

We have used decadal predictions systems to make probabilistic forecasts of North Atlantic hurricane activity and U.S. damage, and have demonstrated the predictions are skillful and able to discriminate between periods of below- and above-average activity. Figures 1e and 1f, however, show two issues: First, the time series of observed 5-yr-mean MDR–TROP and ACE diverge at the end of the time series, particularly for periods with central years 2014–18 (Fig. 3a shows the damage time series also diverges from observed MDR–TROP during this period); and second, the model predictions of MDR–TROP are poor during the same period. The model error in predicting MDR–TROP compensates for the divergence between observed MDR–TROP and ACE/damage, giving apparently skillful predictions during this period (Figs. 2b and 3e), which may indicate that the correlations between predicted MDR–TROP and observed ACE/U.S. damage overestimate the true skill. Assuming a two-stage framework where the observed ACE is linearly related to the observed MDR–TROP, which is linearly related to the predicted MDR–TROP, the final correlation between observed ACE and predicted MDR–TROP should be the product of the correlations of the individual stages [i.e., ρ(forecast MDR–TROP, observed ACE) = ρ(forecast MDR–TROP, observed MDR–TROP) $\times$ ρ(observed MDR–TROP, observed ACE); Kretschmer et al. 2021]. This would give estimates of correlation skill of 0.47 for ACE and 0.31 for damage. Both of these skill estimations are lower than the values obtained from modeling ACE/U.S. damage on predicted MDR–TROP directly, but note that there is large uncertainty in these values owing to the large uncertainty in the correlations for each stage.

To investigate the divergence between observed 5-yr-mean MDR–TROP and observed 5-yr-mean ACE, in Fig. 4a we plot the seasonal values of observed ACE alongside ACE predicted from the observed MDR–TROP (using linear regression). ACE is still well predicted by observed MDR–TROP on this time scale ($\rho = 0.56$), although a persistent negative bias appears in the last six seasons, with notably large errors in 2017 and 2018. This bias is also present when using the other hurricane proxy indices to predict ACE described in section 4b. The reasons for the bias are currently unclear, but for such a short time period it is insufficient evidence to definitively conclude nonstationarity in the relationship between ACE and MDR–TROP. However, the genesis of hurricanes and tropical cyclones is not fully understood (Sobel et al. 2021; Studholme et al. 2022), and the link between observed hurricane activity and proxy indices such as the simple relative temperature index used here, or even more complex genesis potential indices may change in a warming climate (e.g., Wang and Murakami 2020). It will be necessary to monitor the performance of the index in the coming years.

With regard to the poor predictions of MDR–TROP late in the time series, Figs. 4b and 4c show the observed and predicted temperature anomalies for the 5-yr period with the largest error, 2014–18. The observations show notable cool anomalies in the subpolar gyre region, which were driven by the years 2015 (Duchez et al. 2016) and 2018 (Dunstone et al. 2019). Temperatures in the subpolar gyre are strongly linked to the MDR (Smith et al. 2010), and it is indeed overprediction of the MDR temperatures that are causing the errors in the MDR–TROP index (not shown). Despite decadal prediction systems generally having high skill in predicting SPG temperatures (e.g., Smith et al. 2019), this recent period of cooling has clearly not been well predicted. Cool anomalies in the SPG can arise through various mechanisms (Robson et al. 2018) and therefore some SPG temperature transitions may be more predictable than others. The reasons for the cooling in this recent period and why it was not captured by the models are under investigation.

A further caveat to these predictions is that recent studies have shown climate change may be exacerbating the damage caused by individual hurricanes, through increased precipitation, slower moving systems, and higher sea levels leading to higher probabilities of coastal flooding (e.g., Risser and Wehner 2017; Trenberth et al. 2018; Zhang et al. 2020; Guzman and Jiang 2021; Strauss et al. 2021). This is difficult to account for in the damage time series, and may lead to errors in estimating the true skill in predicting hurricane damage. On the other hand, several studies have shown no trend in hurricane damage (after removing the effects of growth in population and wealth) when considering longer time periods (e.g., W18; Klotzbach et al. 2018), which can be seen in Fig. 3c. Furthermore, a positive correlation between forecast MDR–TROP and observed damage remains after linearly detrending both

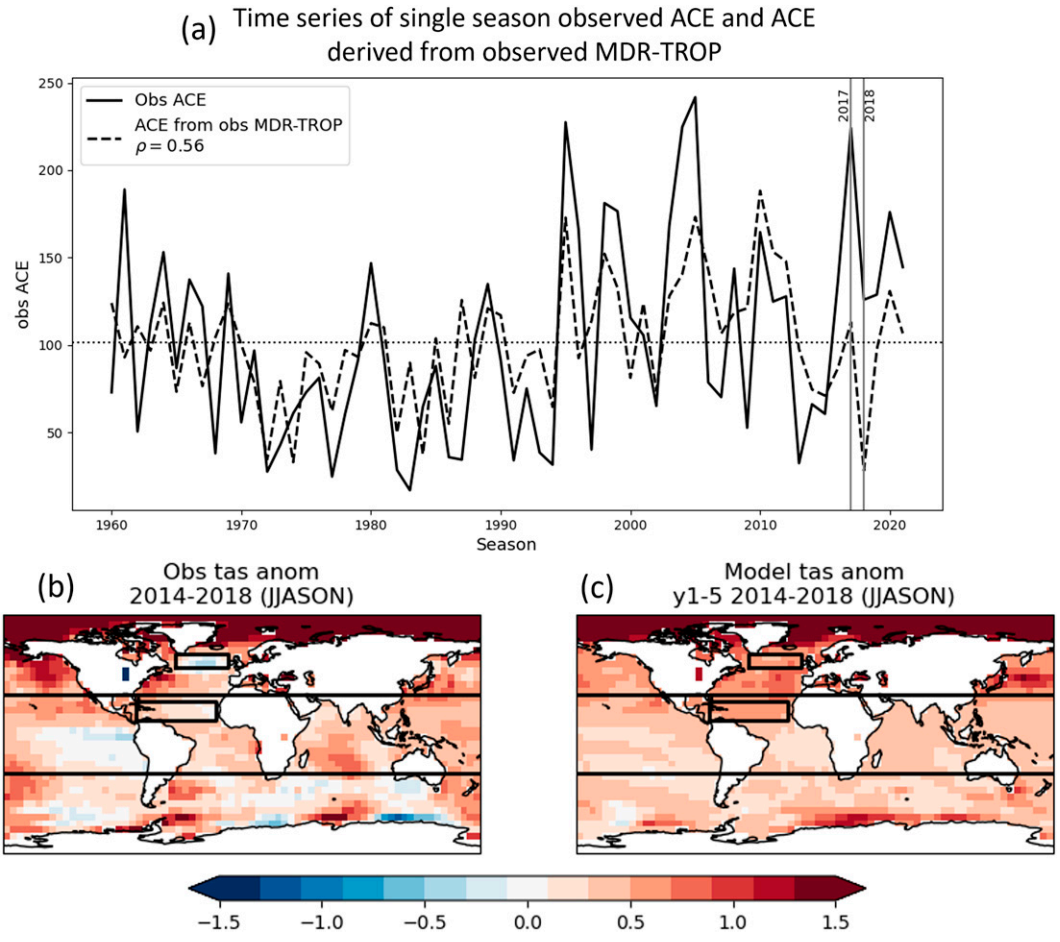

FIG. 4. Performance of the MDR–TROP index: (a) The time series of *seasonal* (JJASON) observed ACE (black line), and seasonal ACE predicted from the linear regression of observed ACE onto observed MDR–TROP (dashed line). The vertical lines mark the 2017 and 2018 seasons, and the rank correlation score is shown in the legend. Also shown are the (b) observed and (c) forecast 5-yr-mean tas anomaly for 2014–18 (JJASON). Anomalies are with respect to 1981–2010. The horizontal black lines and the outlined box inside them show the TROP and MDR regions, respectively, as in Fig. 1. The SPG region (from −50° to −10°E and from 50° to 60°N) is also shown, as the outlined box north of the top black line.

variables (ρ = 0.38, where the damage detrending has been performed on the logarithm of the damage). Although this correlation is no longer statistically significant, detrending will have removed any increase over the hindcast period due to decadal variability and thus could be removing genuine skill. It will be necessary to reevaluate this skill as time progresses and more data become available.

To conclude, we have demonstrated skillful 5-yr predictions of North Atlantic hurricane activity (as measured by the total ACE index) using ensembles of initialized climate predictions. We have shown that this translates into predictability of real-world U.S. damage. The predictions are made using a relative temperature index from the model ensemble mean, and then using linear regression of past observations onto the index to calculate probabilistic forecasts. The regression method is modified for the damage predictions to take into account the random nature of very extreme loss events such as Hurricanes Andrew and Katrina. A prototype climate service for the insurance industry has been codeveloped with users based on this work (Dunstone et al. 2022; available at https://climate.copernicus.eu/sectoral-applications-decadal-predictions). The predictions of ACE and U.S. damage for the period 2020–24 are high, both with ~95% probabilities of being above average. We note that skill in predicting the temperature index on which the forecasts are based has declined in recent years. More research is therefore needed to understand under which conditions the forecasts are most skillful.

Acknowledgments. The authors thank two anonymous referees whose comments led to considerable improvements to this paper. This study received support from the C3S_34c contract (ECMWF/COPERNICUS/2019/C3S_34c_DWD) of the C3S operated by ECMWF.

Data availability statement. The datasets used in this study are available as follows: ERA5 data are available at https://cds.climate.copernicus.eu/ (Hersbach et al 2020); HURDAT2 data are available at https://www.aoml.noaa.gov/hrd/hurdat/Data_Storm.html (Landsea and Franklin 2013); the Hurricane damage 1980–2019: NOAA NCEI U.S. Billion-Dollar Weather and Climate Disasters database is available at https://www.ncei.noaa.gov/access/billions/ (Smith and Katz 2013); the Hurricane damage 1961–79: Annual Summaries of North Atlantic Storms database is available at https://www.aoml.noaa.gov/general/lib/lib1/nhclib/mwreviews.html (NOAA Miami Regional Library 2021); Decadal prediction systems: Hindcast data are available from ESGF at https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/DCPP/index.html. Forecast data can be accessed by request to the corresponding author.

REFERENCES

AIR Worldwide, 2015: A retrospective on 10 years of modelling hurricane risk in a warm climate. AIR Issue Brief Rep., 6 pp., https://www.air-worldwide.com/siteassets/Publications/White-Papers/documents/A-Retrospective-on-10-Years-of-Modeling-Hurricane-Risk-in-a-Warm-Ocean-Climate.

Anderson, C., and Coauthors, 2007: The New Orleans hurricane protection system: What went wrong and why. American Society of Civil Engineers Hurricane Katrina Review Panel Rep., 92 pp., https://archive.org/details/erpreport/page/n1/mode/2up.

Bett, P. E., and Coauthors, 2018: Seasonal forecasts of the summer 2016 Yangtze River basin rainfall. *Adv. Atmos. Sci.*, **35**, 918–926, https://doi.org/10.1007/s00376-018-7210-y.

——, H. E. Thornton, A. Troccoli, M. De Felice, E. Suckling, L. Dubus, Y.-M. Saint-Drenan, and D. J. Brayshaw, 2022: A simplified seasonal forecasting strategy, applied to wind and solar power in Europe. *Climate Serv.*, **27**, 100318, https://doi.org/10.1016/j.cliser.2022.100318.

Bilbao, R., and Coauthors, 2021: Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth Syst. Dyn.*, **12**, 173–196, https://doi.org/10.5194/esd-12-173-2021.

Booth, B. B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**, 228–232, https://doi.org/10.1038/nature10946.

Bureau of Economic Analysis, 2021: Supplemental information & additional data. Accessed 31 March 2021, https://www.bea.gov/data/gdp/gross-domestic-product.

Caron, L.-P., C. J. Jones, and F. J. Doblas-Reyes, 2014: Multi-year prediction skill of Atlantic hurricane activity in CMIP5 decadal hindcasts. *Climate Dyn.*, **42**, 2675–2690, https://doi.org/10.1007/s00382-013-1773-1.

——, L. Hermanson, and F. J. Doblas-Reyes, 2015: Multi-annual forecasts of Atlantic U.S. tropical cyclone wind damage potential. *Geophys. Res. Lett.*, **42**, 2417–2425, https://doi.org/10.1002/2015GL063303.

——, ——, A. Dobbin, J. Imbers, L. Lledó, and G. A. Vecchi, 2018: How skillful are the multi-annual forecasts of Atlantic hurricane activity? *Bull. Amer. Meteor. Soc.*, **99**, 403–413, https://doi.org/10.1175/BAMS-D-17-0025.1.

Cherchi, A., and Coauthors, 2019: Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *J. Adv. Model. Earth Syst.*, **11**, 185–209, https://doi.org/10.1029/2018MS001369.

Chylek, P., and G. Lesins, 2008: Multidecadal variability of Atlantic hurricane activity: 1851–2007. *J. Geophys. Res.*, **113**, D22106, https://doi.org/10.1029/2008JD010036.

Czajkowski, J., and J. Done, 2014: As the wind blows? Understanding hurricane damage at the local level through a case study analysis. *Wea. Climate Soc.*, **6**, 202–217, https://doi.org/10.1175/WCAS-D-13-00024.1.

Delgado-Torres, C., and Coauthors, 2022: Multi-model forecast quality assessment of CMIP6 decadal predictions. *J. Climate*, **35**, 4363–4382, https://doi.org/10.1175/JCLI-D-21-0811.1.

Duchez, A., and Coauthors, 2016: Drivers of exceptionally cold North Atlantic Ocean temperatures and their link to the 2015 European heat wave. *Environ. Res. Lett.*, **11**, 074004, https://doi.org/10.1088/1748-9326/11/7/074004.

Dunstone, N. J., D. M. Smith, and R. Eade, 2011: Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophys. Res. Lett.*, **38**, L14701, https://doi.org/10.1029/2011GL047949.

——, ——, B. B. B. Booth, L. Hermanson, and R. Eade, 2013: Anthropogenic aerosol forcing of Atlantic tropical storms. *Nat. Geosci.*, **6**, 534–539, https://doi.org/10.1038/ngeo1854.

——, ——, S. Hardiman, R. Eade, M. Gordon, L. Hermanson, G. Kay, and A. Scaife, 2019: Skillful real-time seasonal forecasts of the dry northern European summer 2018. *Geophys. Res. Lett.*, **46**, 12 368–12 376, https://doi.org/10.1029/2019GL084659.

——, and Coauthors, 2022: Towards useful decadal climate services. *Bull. Amer. Meteor. Soc.*, **103**, E1705–E1719, https://doi.org/10.1175/BAMS-D-21-0190.1.

Emanuel, K., 2010: Tropical cyclone activity downscaled from NOAA-CIRES reanalysis, 1908–1958. *J. Adv. Model. Earth Syst.*, **2** (1), https://doi.org/10.3894/JAMES.2010.2.1.

——, and D. S. Nolan, 2004: Tropical cyclone activity and the global climate system. *26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 10A.2, http://ams.confex.com/ams/pdfpapers/75463.pdf.

Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nunez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, **293**, 474–479, https://doi.org/10.1126/science.1060040.

Gray, W. M., 1998: The formation of tropical cyclones. *Meteor. Atmos. Phys.*, **67**, 37–69, https://doi.org/10.1007/BF01277501.

Guzman, O., and H. Jiang, 2021: Global increase in tropical cyclone rain rate. *Nat. Commun.*, **12**, 5344, https://doi.org/10.1038/s41467-021-25685-2.

Hermanson, L., R. Eade, N. Robinson, N. J. Dunstone, M. B. Andrews, J. R. Knight, A. A. Scaife, and D. M. Smith, 2014: Forecast cooling of the Atlantic subpolar gyre and associated impacts. *Geophys. Res. Lett.*, **41**, 5167–5174, https://doi.org/10.1002/2014GL060420.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Jewson, S., and Coauthors, 2009: Five year prediction of the number of hurricanes that make United States landfall. *Hurricanes and Climate Change*, J. B. Elsner and T. H. Jagger, Eds., Springer-Verlag, 73–79.

Karen Clark & Company, 2010: Near term hurricane models: Performance update. Tech. Rep., 9 pp., https://s3.us-east-2.amazonaws.com/kcc-mainwebsite-dev/publications/KCCNearTermHurricaneModelsJanuary2010_2021-08-18-201411_ackc.pdf.

Klotzbach, P. J., and W. M. Gray, 2008: Multidecadal variability in North Atlantic tropical cyclone activity. *J. Climate*, **21**, 3929–3935, https://doi.org/10.1175/2008JCLI2162.1.

——, S. G. Bowen, R. Pielke Jr., and M. Bell, 2018: Continental U.S. hurricane landfall frequency and associated damage: Observations and future risks. *Bull. Amer. Meteor. Soc.*, **99**, 1359–1376, https://doi.org/10.1175/BAMS-D-17-0184.1.

Kretschmer, M., S. V. Adams, A. Arribas, R. Prudden, N. Robinson, E. Saggioro, and T. G. Shepherd, 2021: Quantifying causal pathways of teleconnections. *Bull. Amer. Meteor. Soc.*, **102**, E2247–E2263, https://doi.org/10.1175/BAMS-D-20-0117.1.

Kushnir, Y., and Coauthors, 2019: Towards operational predictions of the near-term climate. *Nat. Climate Change*, **9**, 94–101, https://doi.org/10.1038/s41558-018-0359-7.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, https://doi.org/10.1175/MWR-D-12-00254.1.

Latif, M., N. Keenlyside, and J. Bader, 2007: Tropical sea surface temperature, vertical wind shear, and hurricane development. *Geophys. Res. Lett.*, **34**, L01710, https://doi.org/10.1029/2006GL027969.

Merlis, T. M., M. Zhao, and I. M. Held, 2013: The sensitivity of hurricane frequency to ITCZ changes and radiatively forced warming in aquaplanet simulations. *Geophys. Res. Lett.*, **40**, 4109–4114, https://doi.org/10.1002/grl.50680.

Molinari, R. L., and A. M. Mestas-Nuñez, 2003: North Atlantic decadal variability and the formation of tropical storms and hurricanes. *Geophys. Res. Lett.*, **30**, 1541, https://doi.org/10.1029/2002GL016462.

Müller, W. A., and Coauthors, 2018: A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *J. Adv. Model. Earth Syst.*, **10**, 1383–1413, https://doi.org/10.1029/2017MS001217.

NOAA Miami Regional Library, 2021: *Monthly Weather Review*—Annual summaries of North Atlantic storms, 1872–2011. NOAA, accessed 1 June 2021, https://www.aoml.noaa.gov/general/lib/lib1/nhclib/mwreviews.html.

Nyberg, J., and Coauthors, 2007: Low Atlantic hurricane activity in the 1970s and 1980s compared to the past 270 years. *Nature*, **447**, 698–701, https://doi.org/10.1038/nature05895.

O'Neill, B. C., and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9**, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016.

Paxian, A., and Coauthors, 2022: Decadal drought predictions for German water boards: A case study for the Wupper catchment. *Front. Climate*, **4**, 867814, https://doi.org/10.3389/fclim.2022.867814.

Pielke, R. A., Jr., 2009: United States hurricane landfalls and damage: Can one-to five-year predictions beat climatology? *Environ. Hazards*, **8**, 187–200, https://doi.org/10.3763/ehaz.2009.0017.

——, and C. W. Landsea, 1998: Normalized hurricane damage in the United States: 1925–95. *Wea. Forecasting*, **13**, 621–631, https://doi.org/10.1175/1520-0434(1998)013<0621:NHDITU>2.0.CO;2.

Ramsay, H. A., and A. H. Sobel, 2011: Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model. *J. Climate*, **24**, 183–193, https://doi.org/10.1175/2010JCLI3690.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Risser, M. D., and M. F. Wehner, 2017: Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophys. Res. Lett.*, **44**, 12 457–12 464, https://doi.org/10.1002/2017GL075888.

Robson, J., I. Polo, D. L. R. Hodson, D. P. Stevens, and L. C. Shaffrey, 2018: Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model. *Climate Dyn.*, **50**, 921–937, https://doi.org/10.1007/s00382-017-3649-2.

Rousseau-Rizzi, R., and K. Emanuel, 2022: Natural and anthropogenic contributions to the hurricane drought of the 1970s–1980s. *Nat. Commun.*, **13**, 5074, https://doi.org/10.1038/s41467-022-32779-y.

Seabold, S., and J. Perktold, 2010: Statsmodels: Econometric and statistical modeling with python. *Proc. Ninth Python in Science Conf. (SciPy 2010)*, Austin, TX, SciPy, 92–96, http://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf.

Smith, A. B., and R. W. Katz, 2013: U.S. billion-dollar weather and climate disasters: Data sources, trends, accuracy and biases. *Nat. Hazards*, **67**, 387–410, https://doi.org/10.1007/s11069-013-0566-5.

Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010: Skillful multi-year predictions of Atlantic hurricane frequency. *Nat. Geosci.*, **3**, 846–849, https://doi.org/10.1038/ngeo1004.

——, and Coauthors, 2014: Comments on "Multiyear predictions of North Atlantic hurricane frequency: Promise and limitations." *J. Climate*, **27**, 487–489, https://doi.org/10.1175/JCLI-D-13-00220.1.

——, and Coauthors, 2019: Robust skill of decadal climate predictions. *npj Climate Atmos. Sci.*, **2**, 13, https://doi.org/10.1038/s41612-019-0071-y.

——, and Coauthors, 2020: North Atlantic climate far more predictable than models imply. *Nature*, **583**, 796–800, https://doi.org/10.1038/s41586-020-2525-0.

Sobel, A. H., A. A. Wing, S. J. Camargo, C. M. Patricola, G. A. Vecchi, C.-Y. Lee, and M. K. Tippett, 2021: Tropical cyclone frequency. *Earth's Future*, **9**, e2021EF002275, https://doi.org/10.1029/2021EF002275.

Solaraju-Murali, B., and Coauthors, 2021: Multi-annual prediction of drought and heat stress to support decision making in the wheat sector. *npj Climate Atmos. Sci.*, **4**, 34, https://doi.org/10.1038/s41612-021-00189-4.

——, D. Bojovic, N. Gonzalez-Reviriego, A. Nicodemou, M. Terrado, L.-P. Caron, and F. J. Doblas-Reyes, 2022: How decadal predictions entered the climate services arena: An example from the agriculture sector. *Climate Serv.*, **27**, 100303, https://doi.org/10.1016/j.cliser.2022.100303.

Strauss, B. H., and Coauthors, 2021: Economic damage from Hurricane Sandy attributable to sea level rise caused by anthropogenic climate change. *Nat. Commun.*, **12**, 2720, https://doi.org/10.1038/s41467-021-22838-1.

Studholme, J., A. V. Fedorov, S. K. Gulev, K. Emanuel, and K. Hodges, 2022: Poleward expansion of tropical cyclone latitudes in warming climates. *Nat. Geosci.*, **15**, 14–28, https://doi.org/10.1038/s41561-021-00859-1.

Swanson, K. L., 2008: Nonlocality of Atlantic tropical cyclone intensities. *Geochem. Geophys. Geosyst.*, **9**, Q04V01, https://doi.org/10.1029/2007GC001844.

Swiss Re Institute, 2018: Sigma 1/2018: Natural catastrophes and man-made disasters in 2017: A year of record-breaking losses. Swiss Re Institute, https://www.swissre.com/institute/research/sigma-research/sigma-2018-01.html.

Toumi, R., and L. Restell, 2014: Catastrophe modelling and climate change. Lloyd's Tech. Rep., 41 pp., https://assets.lloyds.com/media/d6b6597a-092e-4e13-a31c-6983e215504e/pdf-modelling-and-climate-change-CC-and-modelling-template-V6.pdf.

Trenberth, K. E., L. Cheng, P. Jacobs, Y. Zhang, and J. Fasullo, 2018: Hurricane Harvey links to ocean heat content and climate change adaptation. *Earth's Future*, **6**, 730–744, https://doi.org/10.1029/2018EF000825.

Tsartsali, E. E., P. J. Athanasiadis, S. Materia, A. Bellucci, D. Nicoli, and S. Gualdi, 2023: Predicting precipitation at decadal timescale—Developing a climate service for the energy sector in southern Europe. SSRN, 4358938, https://doi.org/10.2139/ssrn.4358938.

Vecchi, G. A., and B. J. Soden, 2007: Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature*, **450**, 1066–1070, https://doi.org/10.1038/nature06423.

——, and T. R. Knutson, 2008: On estimates of historical North Atlantic tropical cyclone activity. *J. Climate*, **21**, 3580–3600, https://doi.org/10.1175/2008JCLI2178.1.

——, K. L. Swanson, and B. J. Soden, 2008: Whither hurricane activity? *Science*, **322**, 687–689, https://doi.org/10.1126/science.1164396.

——, M. Zhao, H. Wang, G. Villarini, A. Rosati, A. Kumar, I. M. Held, and R. Gudgel, 2011: Statistical–dynamical predictions of seasonal North Atlantic hurricane activity. *Mon. Wea. Rev.*, **139**, 1070–1082, https://doi.org/10.1175/2010MWR3499.1.

——, and Coauthors, 2013: Multiyear predictions of North Atlantic hurricane frequency: Promise and limitations. *J. Climate*, **26**, 5337–5357, https://doi.org/10.1175/JCLI-D-12-00464.1.

Wang, B., and H. Murakami, 2020: Dynamic genesis potential index for diagnosing present-day and future global tropical cyclone genesis. *Environ. Res. Lett.*, **15**, 114008, https://doi.org/10.1088/1748-9326/abbb01.

Weinkle, J., C. Landsea, D. Collins, R. Musulin, R. P. Crompton, P. J. Klotzbach, and R. Pielke Jr., 2018: Normalized hurricane damage in the continental United States 1900–2017. *Nat. Sustainability*, **1**, 808–813, https://doi.org/10.1038/s41893-018-0165-2.

Wilks, D. S., 2019a: Statistical forecasting. *Statistical Methods in the Atmospheric Sciences*, 4th ed. D. S. Wilks, Ed., Elsevier, 235–312, https://doi.org/10.1016/B978-0-12-815823-4.00007-9.

——, 2019b: Forecast verification. *Statistical Methods in the Atmospheric Sciences*, 4th ed. D. S. Wilks, Ed., Elsevier, 369–483, https://doi.org/10.1016/B978-0-12-815823-4.00009-2.

Williams, K. D., and Coauthors, 2018: The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *J. Adv. Model. Earth Syst.*, **10**, 357–380, https://doi.org/10.1002/2017MS001115.

Zhang, G., H. Murakami, T. R. Knutson, R. Mizuta, and K. Yoshida, 2020: Tropical cyclone motion in a changing climate. *Sci. Adv.*, **6**, eaaz7610, https://doi.org/10.1126/sciadv.aaz7610.