



## Variability and Predictability of North Atlantic Hurricane Frequency in a Large Ensemble of High-Resolution Atmospheric Simulations

WEI MEI

*Department of Marine Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina*

YOUICHI KAMAE

*Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan*

SHANG-PING XIE

*Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California*

KOHEI YOSHIDA

*Meteorological Research Institute, Tsukuba, Japan*

(Manuscript received 24 August 2018, in final form 22 January 2019)

### ABSTRACT

Variability of North Atlantic annual hurricane frequency during 1951–2010 is studied using a 100-member ensemble of climate simulations by a 60-km atmospheric general circulation model that is forced by observed sea surface temperatures (SSTs). The ensemble mean results well capture the interannual-to-decadal variability of hurricane frequency in best track data since 1970, and suggest that the current best track data might underestimate hurricane frequency prior to 1966 when satellite measurements were unavailable. A genesis potential index (GPI) averaged over the main development region (MDR) accounts for more than 80% of the SST-forced variations in hurricane frequency, with potential intensity and vertical wind shear being the dominant factors. In line with previous studies, the difference between MDR SST and tropical mean SST is a useful predictor; a 1°C increase in this SST difference produces  $7.05 \pm 1.39$  more hurricanes. The hurricane frequency also exhibits strong internal variability that is systematically larger in the model than observations. The seasonal-mean environment is highly correlated among ensemble members and contributes to less than 10% of the ensemble spread in hurricane frequency. The strong internal variability is suggested to originate from weather to intraseasonal variability and nonlinearity. In practice, a 20-member ensemble is sufficient to capture the SST-forced variability.

### 1. Introduction

A good understanding and an accurate prediction of tropical cyclone (TC) activity can never be over-emphasized as these powerful storms cause tremendous damage to our society every year (Pielke and Landsea 1998; Pielke et al. 2008; Woodruff et al. 2013) in addition to their potentially important roles in the climate system (Emanuel 2001; Sriver and Huber 2007; Korty et al. 2008; Hart 2011; Mei et al. 2013). TC activity can be

characterized by various metrics, including annual frequency, tracks, intensity as well as their derivatives [e.g., the power dissipation index (PDI; Emanuel 2005) and the accumulated cyclone energy (ACE; Bell et al. 2000)]. In this study, we attempt to further our understanding of the variability and predictability of annual hurricane frequency in the North Atlantic (NA) basin.

While predicting genesis of individual TCs remains difficult (e.g., Pasch et al. 2006; Halperin et al. 2013), predictions of annual NA TC/hurricane frequency using large-scale environmental conditions are good in general with the explained variance ranging between 20%

---

*Corresponding author:* Wei Mei, wmei@email.unc.edu

DOI: 10.1175/JCLI-D-18-0554.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](https://www.ametsoc.org/PUBSReuseLicenses)).

and 80% (e.g., Klotzbach and Gray 2009; Knutson et al. 2010; Chen and Lin 2011, 2013; Murakami et al. 2016). A favorable environment promotes the probability of TC/hurricane genesis, and thus the environment averaged over the TC peak season is significantly correlated with annual TC/hurricane frequency. The favorable environmental conditions include but are not limited to below-normal sea level pressure, above-normal low-level vorticity and below-normal vertical wind shear over the subtropical NA, above-normal rainfall over the Sahel region of West Africa, and infrequent midlatitude Rossby wave breaking (e.g., McBride and Zehr 1981; Landsea and Gray 1992; Goldenberg and Shapiro 1996; Knaff 1997; Landsea et al. 1999; DeMaria et al. 2001; Elsner and Jagger 2006; Nolan and Rappin 2008; Kossin et al. 2010; Klotzbach 2011; Zhao and Held 2012; Daloz et al. 2012; Patricola et al. 2014; Zhang et al. 2016). These conditions can be linked to patterns of sea surface temperature anomalies (SSTAs), as demonstrated in the Atmospheric Model Intercomparison Project (AMIP)-type simulations in which atmospheric general circulation models (AGCMs) are subject to observed SSTs. Furthermore, both observational and modeling efforts show that differences between tropical NA SSTs and globally averaged tropical SSTs explain a considerable fraction of the interannual-to-decadal variability in NA TC/hurricane frequency (e.g., Knutson et al. 2008; Zhao et al. 2009; Vecchi et al. 2011).

Despite the fact that annual frequency is arguably the most predictable aspect of NA hurricanes at seasonal leads, prediction systems employed by meteorological agencies are still struggling from time to time. For example, Vecchi and Villarini (2014) show that six to nine hurricanes were expected to occur in year 2013 based on both statistical and dynamical predictions. In reality, however, only two hurricanes formed that year. The failure of the prediction systems may be due to either our limited understanding of the hurricane–climate interactions, the inherently limited predictability of the climate system or both.

While TC frequency is constrained by SSTs, it also exhibits randomness owing to the nonlinear processes and instabilities in the atmosphere (e.g., Jourdain et al. 2011; Wu et al. 2012; Chen and Lin 2013; Done et al. 2014). Ensemble simulations that differ in initial conditions are employed to incorporate information of internal variability, and have demonstrated that the internal variability can be substantial even in simulations with regional climate models where the internal variability is damped by the prescribed lateral boundary conditions (e.g., Wu et al. 2012; Done et al. 2014). For instance, using a regional model, Done et al. (2014) show that for year 1998 a 16-member ensemble simulates a

range of 6–12 TCs in a subregion of the NA with an ensemble mean of 8.8 TCs.

While previous studies have led to our recognition of the importance of internal variability in hurricane frequency, the simulations in these studies cover a relatively short period (typically no longer than 30 years) and have a small ensemble size (usually no more than 5 members for global simulations), which may hinder a full understanding of the internal variability. In this study, we shall use an unprecedentedly large ensemble of simulations to study both the forced and internal variability in NA hurricane frequency for the period of 1951–2010. Simulations from the early period make it possible to validate TC frequency in current best track data when satellite measurements were not available, and the large ensemble makes it possible to address questions that cannot be answered before regarding the internal variability, such as the number of ensemble members that are sufficient to capture the variability in observed hurricane frequency. After presenting the data and methods in use (section 2), we study the forced variability in simulated hurricane frequency using the ensemble mean, compare it with observations, and identify the controlling factors (section 3). We then investigate and discuss the internal variability and predictability of hurricane frequency in section 4.

## 2. Data and methods

### a. Observational and reanalysis data

The observed hurricane frequency is obtained from the National Hurricane Center best track dataset (McAdie et al. 2009; Landsea and Franklin 2013) and TC data with wind speed-dependent corrections by Professor K. Emanuel (<ftp://texmex.mit.edu/pub/emanuel/HURR/tracks/>), both of which provide the location and intensity of NA hurricanes at 6-h intervals since 1851. SSTs and atmospheric variables [including sea level pressure, temperature, specific and relative humidity, and 850- and 200-hPa winds] from three reanalysis datasets [including the Japanese 55-year Reanalysis (JRA-55; Kobayashi et al. 2015), the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) Reanalysis-1 (NCEP–NCAR-1; Kalnay et al. 1996), and the European Centre for Medium-Range Weather Forecasts (ECMWF) twentieth-century reanalysis (ERA-20C; Poli et al. 2016)] are used to compute a genesis potential index (GPI) defined in Emanuel (2010):

$$\text{GPI} = \frac{a|\eta|^3 [\max(V_{\text{PI}} - 35, 0)^2]}{\chi^{4/3} (25 + V_{\text{sh}})^4}, \quad (1)$$

where  $\eta$  is the 850-hPa absolute vorticity,  $V_{PI}$  is the TC potential intensity,  $\chi$  is the 600-hPa entropy deficit, and  $V_{sh}$  is the magnitude of the 250–850-hPa wind shear vector [see also Korty et al. (2012) and Tang and Emanuel (2012) for a detailed discussion of these contributing factors]. To be consistent with the simulations described below, only the observational data during 1951–2010 (1958–2010 for JRA-55) are used.

### b. Simulated hurricane frequency

We use the historical simulations from the Database for Policy Decision Making for Future Climate Change (d4PDF) (Mizuta et al. 2017), with the Meteorological Research Institute AGCM, version 3.2 (Mizuta et al. 2012), of 60-km resolution. The model is forced by observed monthly mean SST and sea ice concentration (COBE-SST2; Hirahara et al. 2014) and climatological monthly sea ice thickness, following the procedure of the AMIP. The simulations cover the period of 1951–2010, and consist of 100 members that differ in initial conditions and slightly in the imposed SSTs [“small perturbations of SST based on SST analysis error are added to the observed SSTs”; see the appendix of Mizuta et al. (2017) for more details on how the initial conditions and SSTs are perturbed]. The simulations reproduce interannual and decadal variability in large-scale atmospheric circulation related to SST variability in the Pacific and Atlantic Oceans (Kamae et al. 2017a,b; Ueda et al. 2018).

The 60-km model generates TC-like disturbances. They are detected and tracked using sea level pressure, 850-hPa relative vorticity, 850-hPa, 300-hPa and surface wind speed, warm-core temperature, and duration of the tracking, following Murakami et al. (2012); an examination of randomly selected events shows that they resemble TCs in observations. As illustrated in Yoshida et al. (2017), the simulations well capture many statistics and climatological characteristics of observed TCs, including relative probability distribution of annual TC genesis frequency over the global ocean and geographical distribution of climatological TC occurrence [Fig. 2 in Yoshida et al. (2017)]. We note, however, that Emanuel and Sobel (2013) document that atmospheric model simulations forced only with observed SSTs may not produce correct surface fluxes and surface wind speeds, and may thereby affect TC-related thermodynamic parameters (particularly potential intensity) and TC activity.

Hurricane frequency is commonly underestimated in both global and regional climate simulations when the criterion (i.e.,  $32.5 \text{ m s}^{-1}$  for the maximum surface wind speed during the lifetime of a TC) for observations is employed (Walsh et al. 2015), owing in part to different

average time periods and the relatively low resolution of the model (e.g., Bacmeister et al. 2018; Li and Srivier 2018). Here we adjust this threshold value for the simulations to be  $12 \text{ m s}^{-1}$  by matching the simulated TC number to the observed hurricane number during 1970–2010 (i.e.,  $\sim 6.15$  per year). In other words, TCs with a lifetime peak intensity greater than  $12 \text{ m s}^{-1}$  are termed hurricanes in the simulations. Using a different threshold value (e.g.,  $16 \text{ m s}^{-1}$ ; Walsh et al. 2007) produces similar results in all the aspects discussed later except the number of hurricanes.

Because the SSTs used to force the model differ slightly from one ensemble member to the next, we remove the effect of such SST differences on hurricane frequency and recover the internal variability due to differences in initial conditions using a statistical relationship obtained based on the ensemble mean of the simulations. Specifically, as shown in the next section, hurricane frequency is strongly correlated with the difference between SSTs averaged over the main development region (MDR;  $8^{\circ}$ – $20^{\circ}$ N,  $25^{\circ}$ – $80^{\circ}$ W; see the blue box in the inset of Fig. 2) of NA hurricanes and global tropical mean SSTs [we refer to this difference as relative SSTs; see Vecchi and Soden (2007) and Johnson and Xie (2010) for a discussion of the physical mechanisms]. We quantified the relationship between ensemble mean hurricane frequency and ensemble mean relative SSTs, and then for each individual year we used this relationship to remove the differences in hurricane frequency among the 100 simulations attributable to the differences in relative SSTs; note that this adjustment does not change the ensemble mean values. The effect of the differences in the imposed SSTs turns out to be minor relative to the effect of internal variability, and removing this effect slightly reduces the spread of the simulated hurricane frequency (not shown). We approximate the ensemble mean as the forced response in hurricane frequency to observed SSTs, and the deviation of individual member simulations after the adjustment from the ensemble mean as the internal variability.

### 3. Forced variability

The ensemble mean of the simulations skillfully reproduces the interannual-to-decadal variations in observed hurricane frequency since the late 1960s (Fig. 1a). The correlation coefficient between the red and black curves during 1970–2010 is 0.84 (this high level of correlation is largely from interannual variability and is insensitive to the length of the simulations, as shown in Fig. 1b), and the red curve is located within the gray area, which shows the range of one standard deviation of the ensemble simulations. In particular, the model

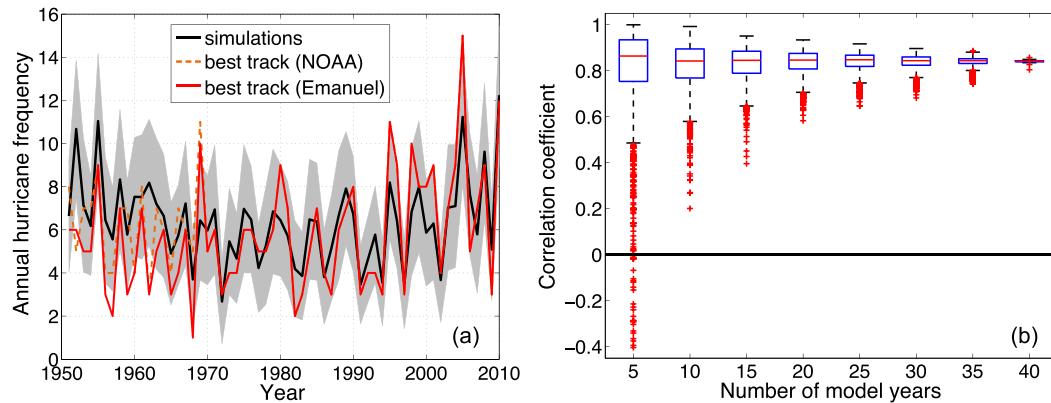


FIG. 1. (a) A comparison of simulated (ensemble mean; black curve) and observed (color curves) annual hurricane frequency in the North Atlantic (NA) basin between 1951 and 2010. The gray shading shows the spread of the model results, represented by the standard deviation of the results from the 100 ensemble members. The dashed brown curve shows the hurricane frequency from the NOAA/National Hurricane Center best track data, and the red curve shows the hurricane frequency from the best track data compiled by Prof. K. Emanuel with wind speed-dependent corrections. (b) Box-and-whisker plots of the distribution of the correlation coefficients between simulated and observed hurricane frequency as a function of the number of model years. Boxes show the interquartile range, whiskers show the 5%–95% interval, and outliers are marked with crosses. The calculations were performed by randomly selecting  $N$  ( $N = 5, 10, 15, \dots, 35, 40$ ) years out of the entire 41 years (i.e., during 1970–2010) for both the simulated and observed hurricane frequency, computing their correlation coefficient, and for each  $N$  value repeating this procedure 2000 times to obtain a distribution.

simulates the enhanced hurricane activity since the mid-1990s, and is able to capture hurricane frequency during years of extremely high activity like 2005 and 2010. These results demonstrate the strong SST control of NA hurricane activity, in line with previous studies (e.g., Zhao et al. 2009; Vecchi et al. 2011; Chen and Lin 2011, 2013; Camargo et al. 2013; Mei et al. 2014). It is worth pointing out that although environmental conditions in 2005 were highly favorable for hurricane occurrence, atmospheric internal variability contributed to a significant portion of the observed 15 hurricanes since model ensembles are rarely able to reproduce the observed high value (e.g., Vitart et al. 2007; Smith et al. 2010; Vecchi and Knutson 2011; Mei et al. 2014; Camp et al. 2015; Roberts et al. 2015; Manganello et al. 2016).

Prior to 1966 when satellite measurements were not available, however, hurricane frequency in the best track data is generally less than the ensemble mean hurricane frequency (i.e., the red and dashed brown curves are located below the black curve in Fig. 1a). This suggests that during the early part of the study period, current best track data might underestimate hurricane numbers by approximately one per year. [Further examinations of hurricanes with intensity of category 2 and greater and of category 3 and greater (not shown) suggest that the current best track data do not underestimate the frequency of hurricanes of category 2 and greater, and the missing storms have intensity around category 1.]

To test the robustness of the result regarding the underestimation of hurricane frequency in the current best track data prior to 1966, we computed GPI averaged over the MDR using the three reanalysis datasets described in section 2, and compared these indices with observed hurricane frequency (Fig. 2). The MDR GPI well captures the year-to-year variations in observed hurricane frequency during 1970–2010, with the correlation coefficient ranging between 0.72 and 0.79 (the correlation skill of the ensemble of the three GPIs is 0.81). But before the mid-1960s, hurricane frequency derived from GPI in all three datasets is consistently higher than hurricane frequency in the best track data. While we note that potentially considerable uncertainties exist in the reanalysis data during the early time period (e.g., Emanuel 2007; Saunders et al. 2017), the consistency of the results among the three reanalysis datasets and with our atmospheric ensemble simulations provides further evidence for the possible underestimate of hurricane frequency in the best track data during that time period (Fig. 1a). These results are also in line with Chang and Guo (2007) and Vecchi and Knutson (2011), which are based on analyses of hurricane tracks and reporting ship track density.

To understand the physical mechanisms underlying the variability in the simulated hurricane frequency, we computed the MDR GPI using simulated atmospheric fields together with the prescribed SSTs. The ensemble mean MDR GPI accounts for 81% of the variability in ensemble mean hurricane frequency ( $r = 0.90$ ) during the entire

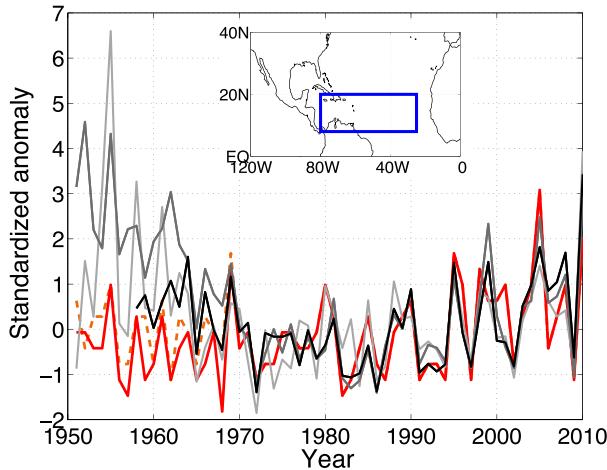


FIG. 2. A comparison of the standardized anomalies (base period: 1970–2010) in the genesis potential index (GPI) averaged over the main development region (MDR) of NA hurricanes (black and gray curves) and annual hurricane frequency in the best track datasets (red and dashed brown curves) between 1951 and 2010. The black and gray curves are obtained from three reanalysis datasets: JRA-55 (1958–2010; black), the NCEP–NCAR-1 (1951–2010; gray), and ERA-20C (1951–2010; light gray). The red and dashed brown curves correspond to the color curves in Fig. 1a. Blue box in the inset denotes the MDR.

study period (1951–2010). This, along with the above calculations on the basis of reanalysis datasets, suggests that in an *ensemble sense* seasonal-mean GPI in the MDR is a good measure of annual hurricane frequency. We further examined the individual factors involved in the GPI computation and found that all four components play important roles, with potential intensity and vertical wind shear dominating (Table 1). Similar conclusions were found in the calculations based on reanalysis datasets (Table 2), in line with previous studies (e.g., Bruyère et al. 2012). These results are also consistent with previous studies showing that in the NA thermodynamic and dynamical factors affect hurricane activity in a cooperative way (e.g., Emanuel 2007; Kossin and Vimont 2007; Vimont and Kossin 2007; Mei et al. 2014).

Both SSTs in the tropical NA and in the equatorial Pacific associated with El Niño–Southern Oscillation (ENSO) are important in modulating annual NA

hurricane frequency (Patricola et al. 2014): their correlation coefficients with hurricane frequency are 0.57 and  $-0.47$ , respectively, for the results based on the ensemble mean of the simulations (Table 1). As a result, the SSTs in the MDR relative to the global tropical mean SSTs (termed relative SSTs) appear to be a more useful predictor than SSTs in individual regions (Table 1), consistent with previous observational and modeling studies (e.g., Knutson et al. 2008; Zhao et al. 2009; Vecchi and Knutson 2011; Zhao and Held 2012). In our simulations, relative SSTs explain nearly the same portion of year-to-year variance in hurricane frequency ( $\sim 75\%$ ) as the GPI, achieved mainly through the effect of potential intensity. A regression analysis shows that the MDR needs to be  $1.57^\circ \pm 0.24^\circ\text{C}$  warmer than the entire tropics to generate hurricanes in the NA, and that a  $1^\circ\text{C}$  increase in the relative SSTs produces  $7.05 \pm 1.39$  more hurricanes (Fig. 3); the uncertainties (i.e.,  $0.24^\circ\text{C}$  and 1.39 hurricanes per  $^\circ\text{C}$  warming) are represented by the standard deviation of the results from the 100 ensemble members (not shown).

#### 4. Internal variability

In addition to the forced interannual-to-decadal variability, hurricane frequency also exhibits strong internal variability, as indicated by the model spread shown in Fig. 1a. On average, the spread is larger during more active years (the correlation coefficient between the spread and the ensemble mean is 0.84). Next we will use the 100-member ensemble to further explore the internal variability. Specifically, here we seek to answer the following two questions: (i) Are observations equivalent to one model realization? (ii) How many members are needed to capture the signal? We will also briefly discuss the sources of the internal variability in hurricane frequency.

##### a. Are observations equivalent to one model realization?

We examine this issue by comparing the properties of the observations and simulations. The year-to-year variability (represented by standard deviation) in observed hurricane frequency during 1970–2010 is 2.86,

TABLE 1. Correlation coefficients between ensemble-mean annual hurricane frequency and various ensemble-mean environmental factors averaged over the main development region (MDR) during 1951–2010 in the simulations.

	GPI	Potential intensity	Vertical wind shear	850-hPa vorticity	600-hPa entropy deficit	MDR SSTs	Niño-3.4 index	Relative SSTs
Hurricane frequency	0.90	0.85	$-0.84$	0.78	$-0.78$	0.57	$-0.47$	0.87
Relative SSTs	0.91	0.95	$-0.86$	0.64	$-0.82$	0.79	$-0.30$	1.00

TABLE 2. Correlation coefficients between observed annual hurricane frequency and various environmental factors averaged over the MDR in three reanalysis datasets during 1970–2010.

	GPI	Potential intensity	Vertical wind shear	850-hPa vorticity	600-hPa entropy deficit
JRA-55	0.78	0.76	−0.73	0.34	−0.66
NCEP–NCAR-1	0.79	0.73	−0.69	0.45	−0.33
ERA-20C	0.72	0.75	−0.74	0.57	−0.60

whereas the average variability in individual members of the simulations is 3.18; the variability for the ensemble mean is 1.99. This suggests that individual members exhibit slightly larger variance than observations, and correspondingly, the ensemble range of the simulations fully covers the observations and  $X\%$  ensemble range contains more than  $X\%$  of the observations ( $X = 1, 2, \dots, 98, 99$ ) (not shown); and that averaging across individual simulations reduces levels of noise and results in a smaller variance in the ensemble mean.

We then calculated the correlation coefficient between the observations and each of the individual ensemble members and obtained a probability density function (PDF) of the correlation coefficient being centered between 0.5 and 0.6 (red curve in Fig. 4a). Similarly, we calculated the correlation coefficient between one ensemble member and the other 99 members and obtained a PDF for each individual member (gray curves in Fig. 4a); the averaged PDF shown as the black curve is centered between 0.3 and 0.4. This implies that individual simulations are more similar to observations than to each other.

We also obtained a PDF of the correlation coefficient between the observations and an ensemble mean of randomly selected 50 members (out of the 100 members) by repeating this calculation 2000 times (magenta curve in Fig. 4a). Similar PDFs were obtained for the correlation coefficient between each individual ensemble member and an ensemble mean of randomly selected 50 members (out of the remaining 99 members) (light blue curves in Fig. 4a); blue curve shows the averaged PDF. The median values of the PDF for the observations and of the averaged PDF for model simulations are 0.84 and 0.6, respectively. In other words, the variance explained by a 50-member ensemble mean in the observations is twice that in one realization of model simulations.

The appreciable differences in the PDFs shown above between observations and model simulations indicate that the observations and simulations are quite different: the signal-to-noise ratio (SNR) is smaller in model simulations than in the observations, and the observations have a higher predictability than the model.

The discrepancy between the observations and model simulations can also be quantified using the ratio of predictable components (RPC; Eade et al. 2014) that compares levels of predictability in models and in observations:

$$\text{RPC} = \frac{\text{PC}_{\text{obs}}}{\text{PC}_{\text{model}}} \geq \frac{r}{\sqrt{\sigma_{\text{ensemble}}^2 / \sigma_{\text{member}}^2}}, \quad (2)$$

where the predictable component of the observations ( $\text{PC}_{\text{obs}}$ ) is defined as the correlation coefficient between the ensemble mean and the observations ( $r$ ), and the predictable component of the model ( $\text{PC}_{\text{model}}$ ) is derived from the ratio of the variance of the ensemble mean ( $\sigma_{\text{ensemble}}^2$ ) and the average variance of individual ensemble members ( $\sigma_{\text{member}}^2$ ). An RPC value greater than (smaller than) 1 indicates the model is underconfident (overconfident) or overdispersive (underdispersive). The RPC value for the simulations of hurricane frequency here is 1.35, and is insensitive to the number of model years and number of ensemble members (Fig. 5). This suggests that the model simulations are underconfident and hurricane frequency in the real world is

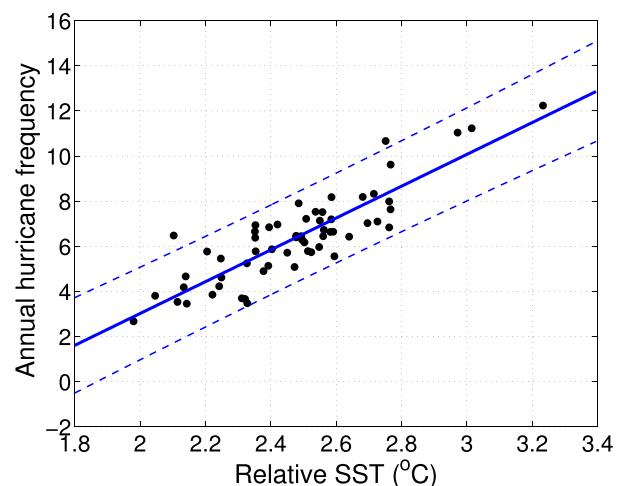


FIG. 3. Scatterplot of ensemble-mean annual hurricane frequency vs. relative SSTs (defined as the difference between the MDR SSTs and global tropical mean SSTs) during 1951–2010. Solid blue line shows the linear regression, with dashed blue curves showing the 95% confidence bands.

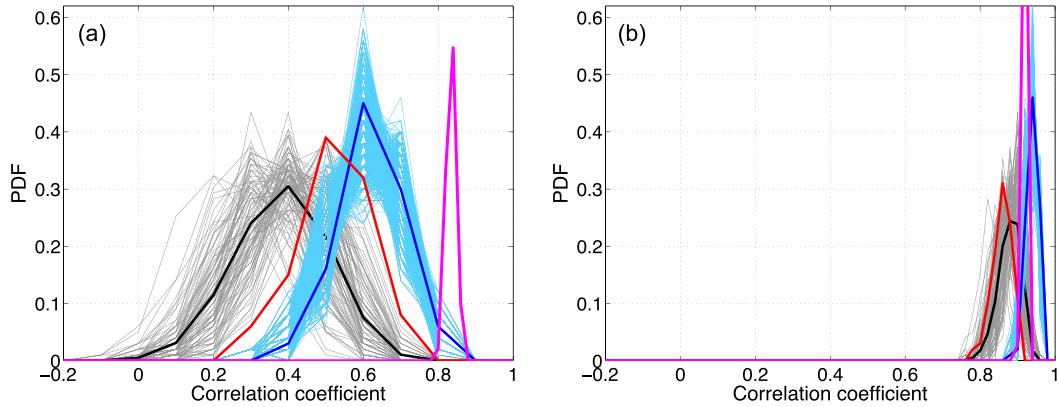


FIG. 4. (a) PDF of the correlation coefficients between annual hurricane frequency in one simulation and that in the other 99 simulations (gray curves), between annual hurricane frequency in observations and that in individual simulations (red curve), between annual hurricane frequency in one simulation and that in an ensemble mean of 50 randomly selected simulations (light blue curves), and between annual hurricane frequency in observations and that in an ensemble mean of 50 randomly selected simulations (magenta curve) during 1970–2010. Black curve shows the average of the gray curves, and blue curve shows the average of the light blue curves. (b) As in (a), but for GPI. Note that a bin interval of 0.1 is used in plotting curves in (a) except the magenta curve, and a bin interval of 0.02 is used in plotting the magenta curve in (a) and all curves in (b).

more predictable than that in the model world, which are consistent with the above findings based on a comparison of the PDFs of correlation coefficients.

*b. How many members are needed in order to capture the observed variability?*

1) RESULTS FROM THE LARGE ENSEMBLE AGCM SIMULATIONS

The correlation coefficient between individual ensemble members and the observations ranges between 0.3 and 0.7, as shown in Fig. 4a. Next we shall understand the impact of ensemble size on the skill of the ensemble mean in capturing the observed variability, by analyzing a large number of combinations of ensemble

members based on the entire 100 members. A sampling method was employed to randomly and independently select  $N$  ( $N = 5, 10, 20, 30, \dots, 90, 100$ ) members to form an ensemble, and the correlation coefficient between the resulting ensemble mean and the observations was then calculated. For each value of  $N$ , the sampling was repeated 2000 times to obtain a distribution of correlation values.

Figure 6a shows the distribution of correlation coefficients as a function of ensemble size as a box-and-whisker plot. As the ensemble size increases, the correlation coefficient increases and its range narrows since larger amounts of random variations are averaged out. The increase in the average correlation coefficient is rapid at the beginning as the ensemble size increases

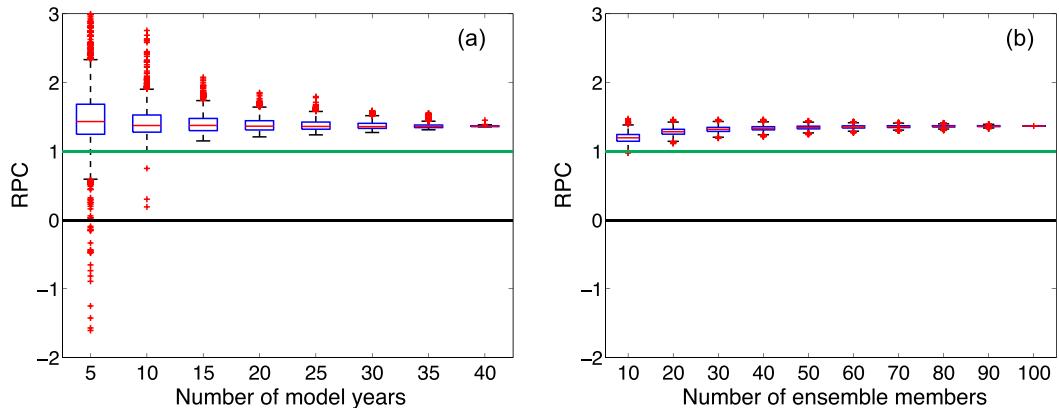


FIG. 5. Box-and-whisker plots of the distribution of the ratio of predictable components (RPCs) of the annual hurricane frequency as a function of the number of (a) model years and (b) ensemble size.

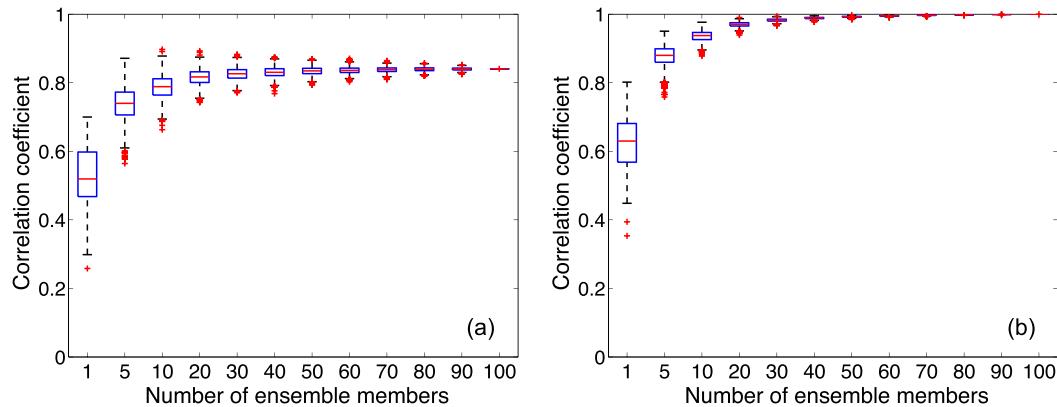


FIG. 6. Box-and-whisker plots of the distribution of the correlation coefficients (a) between annual hurricane frequency in simulations and that in observations, and (b) between annual hurricane frequency in simulations and that in the 100-member ensemble mean as a function of ensemble size.

from 1 to 20, and then the increase slows down with the correlation coefficient progressively converging toward 0.84, which is the correlation coefficient between the mean of all ensemble members and the observations.

We then replaced the observations with the 100-member ensemble mean and repeated the analysis since this ensemble mean might be closer to the signal than the observations. The distribution of correlation coefficients exhibits similar behaviors but with the correlation coefficient approaching 1 (Fig. 6b). It is evident from Fig. 6 that an ensemble of 20 simulations should be sufficient to skillfully simulate the forced/observed variability in hurricane frequency.

## 2) RESULTS FROM A TOY MODEL

We create a toy model described below to explain the behaviors of the correlation skill of the ensemble mean.

This toy model consists of two components: the signal and the noise, and is used to generate an ensemble of time series. Here, we define the signal as  $x = \sin(2t/\pi)/\text{std}[\sin(2t/\pi)] + 10$ , which has a mean value of 10 and unit variance, with  $t$  being the time ranging between 1 and 60. We then write the  $i$ th ensemble member  $y_i$  as  $y_i = x + \epsilon_i$  with  $\epsilon_i$ , an independent normal random variable representing the noise,  $\sim N(0, \sigma_\epsilon^2)$ . This last assumption is based on the result that the deviations from the ensemble mean of individual simulations nearly follow a normal distribution (not shown).

We set  $\sigma_\epsilon = 1.25$  and the corresponding SNR is 0.8. To be consistent with the AGCM simulations, we generated 100 members, which are shown as gray curves in Fig. 7a. The thick black curve shows their ensemble mean, and the red curve shows the signal; these two curves nearly overlap with each other with a correlation coefficient

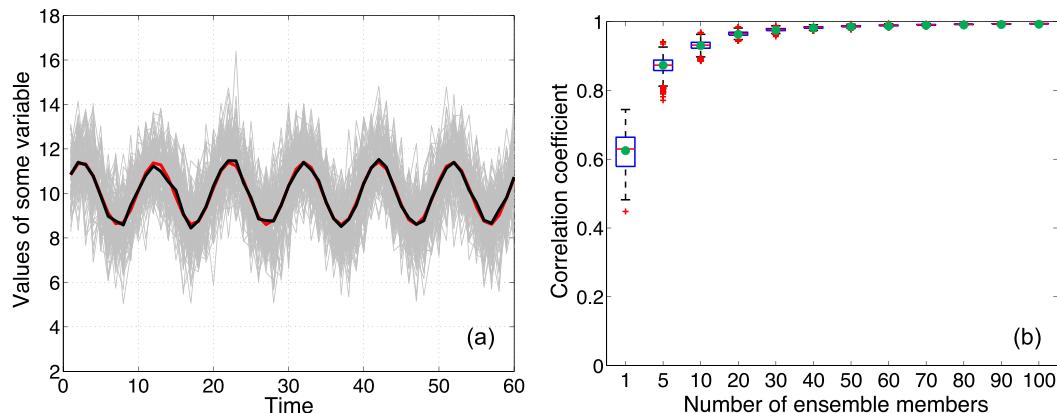


FIG. 7. (a) An example from the toy model described in the text with a signal-to-noise ratio (SNR) of 0.8: red curve shows the signal, gray curves show 100 time series that are randomly generated by the toy model, and black curve shows the average of the gray curves. (b) Box-and-whisker plots of the distribution of the correlation coefficients between the randomly generated time series and the signal as a function of ensemble size based on the toy model results. Green dots show the results from the theoretical arguments [i.e., Eq. (7)].

greater than 0.99. We then employed the sampling method described above to examine the dependence of the correlation skill on the ensemble size. The behavior of the correlation values shown in Fig. 7b (the boxes and whiskers) is similar to the results from the AGCM simulations shown in Fig. 6. The correlation skill improves quickly when the ensemble size increases from 1 to 20, accompanied by a reduction in the range.

### 3) THEORETICAL ARGUMENTS

In this subsection, we further study the behaviors of the correlation skill of the ensemble mean on the basis of theoretical arguments. Let  $x$  be the signal, and  $y_i$  the given variable (e.g., hurricane frequency in this study) in the  $i$ th ensemble member:  $y_i = x + \epsilon_i$  with  $\epsilon_i$ , an independent normal random variable representing the noise,  $\sim N[0, \text{var}(\epsilon_i)]$ . Then the correlation coefficient  $r_{\text{en}}$  between the ensemble mean of  $N$  members  $\bar{y}$  and the signal  $x$  can be written as

$$\begin{aligned} r_{\text{en}} &= \frac{\text{cov}(x, \bar{y})}{\sqrt{\text{var}(x)\text{var}(\bar{y})}} = \frac{\frac{1}{N} \sum_{i=1}^N \text{cov}(x, y_i)}{\sqrt{\text{var}(x)\text{var}(\bar{y})}} \\ &= \frac{1}{N} \sum_{i=1}^N \left[ \frac{\text{cov}(x, y_i)}{\sqrt{\text{var}(x)\text{var}(y_i)}} \sqrt{\frac{\text{var}(y_i)}{\text{var}(\bar{y})}} \right] \\ &= \frac{1}{N} \sum_{i=1}^N \left[ r_i \sqrt{\frac{\text{var}(y_i)}{\text{var}(\bar{y})}} \right] \approx \bar{r} \sqrt{\frac{\text{var}(y)}{\text{var}(\bar{y})}}, \end{aligned} \quad (3)$$

where  $\text{var}$  and  $\text{cov}$  denote variance and covariance, respectively;  $r_i$  is the correlation coefficient between the  $i$ th ensemble member and the signal; and  $\bar{r}$  represents the mean value of  $N$  correlation coefficients for individual ensemble members (i.e.,  $r_i$ ).

We can further write the variance of the ensemble mean as

$$\begin{aligned} \text{var}(\bar{y}) &= \text{var}(x) + 2\overline{\text{cov}(x, \epsilon)} + \frac{1}{N} \overline{\text{var}(\epsilon)} + \frac{1}{N^2} \sum_{i \neq j} \text{cov}(\epsilon_i, \epsilon_j) \\ &= \overline{\text{var}(y)} - \frac{N-1}{N} \overline{\text{var}(\epsilon)} + \frac{1}{N^2} \sum_{i \neq j} \text{cov}(\epsilon_i, \epsilon_j) \\ &\approx \overline{\text{var}(y)} - \frac{N-1}{N} \overline{\text{var}(\epsilon)}, \end{aligned} \quad (4)$$

where  $\overline{\text{var}(y)}$  is the mean of the variance of individual ensemble members, and  $\overline{\text{var}(\epsilon)}$  is the mean of the noise variance in individual members.

Substituting

$$\bar{r} = \frac{1}{N} \sum_{i=1}^N \left[ \frac{\text{cov}(x, y_i)}{\sqrt{\text{var}(x)\text{var}(y_i)}} \right] \approx \sqrt{\frac{\text{var}(x)}{\text{var}(y)}}, \quad (5)$$

$$\overline{\text{var}(y)} \approx \text{var}(x) + \overline{\text{var}(\epsilon)}, \quad (6)$$

and Eq. (4) into Eq. (3), we obtain

$$r_{\text{en}} \approx \bar{r} \sqrt{\frac{1}{1 - \frac{N-1}{N} \frac{\overline{\text{var}(\epsilon)}}{\overline{\text{var}(y)}}}} \approx \sqrt{\frac{1}{1 + \frac{1}{N} \frac{\overline{\text{var}(\epsilon)}}{\text{var}(x)}}}. \quad (7)$$

For a finite  $\overline{\text{var}(\epsilon)}/\text{var}(x)$ , when  $N \rightarrow +\infty$ ,  $r_{\text{en}} = 1$ . This is what we expected, since an infinite number of ensemble members can fully remove the noise, no matter how small the SNR is. We can also roughly estimate the number of ensemble members that we need to achieve a certain correlation skill ( $r_{\text{threshold}}$ ) as  $N \approx [r_{\text{threshold}}^2 / (1 - r_{\text{threshold}}^2)] [\overline{\text{var}(\epsilon)}/\text{var}(x)]$ .

Results based on Eq. (7) with  $\overline{\text{var}(\epsilon)}/\text{var}(x) \approx (1/0.8)^2 = 1.5625$  (i.e., SNR = 0.8) for various  $N$  values are shown as green dots in Fig. 7b. The theoretical values well match the toy model results. Figure 8a shows the correlation skill as a function of ensemble size for five selected SNR values obtained from the theoretical arguments. In all cases, the rate of the increase in correlation skill is large when the ensemble size is small. And in general, the correlation skill improves more quickly for a larger SNR when  $N$  increases from 1 to 20. For example, the correlation coefficient increases from  $\sim 0.4$  to  $\sim 0.9$  for a SNR of 0.5, and increases from  $\sim 0$  to  $\sim 0.1$  for a SNR of 0.02. This is consistent with the expectation that when the SNR is large, the ensemble mean of a small number of member simulations is sufficient to capture the signal, whereas a much larger number of ensemble members are needed to reproduce the signal for a very small SNR.

Figure 8b illustrates the dependence of the correlation skill on the inverse of SNR for a fixed ensemble size  $N$ . We set  $N$  to be 100, and for each SNR we repeated the calculations of the correlation skill 100 times. Black dots show the results for the correlation skill based on the toy model and red dots show the theoretical values. The mean values of the correlation coefficient between individual members and the signal in the toy model (i.e.,  $\bar{r}$ ) are shown as blue dots.  $\bar{r}$  quickly drops as SNR decreases from 1 to 0.05, and gradually approaches 0 afterward. Both the theoretical and toy model results demonstrate that using an ensemble mean of 100 members can considerably improve the correlation skill. For example, for a SNR of 0.1, a single member only captures 1% of the variability in the signal, whereas an ensemble mean of 100 members captures 50% of the variability.

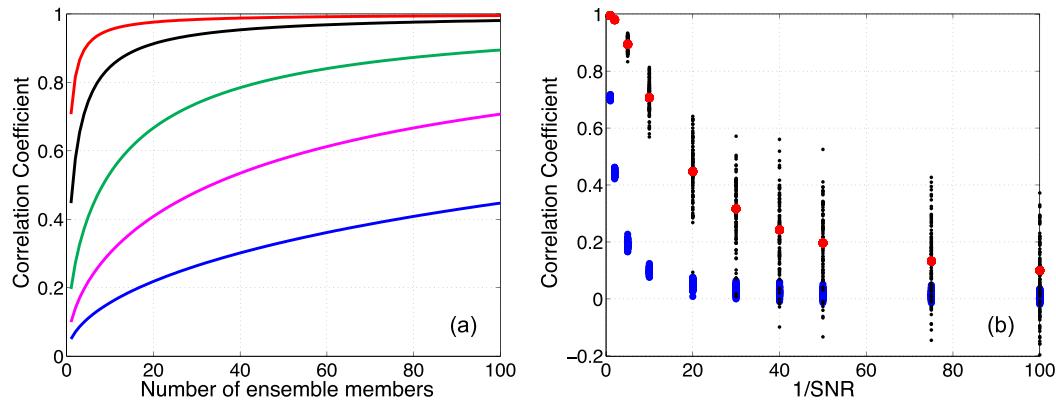


FIG. 8. (a) Correlation skill of the ensemble mean as a function of ensemble size for five selected SNR values (red: 1; black: 0.5; green: 0.2; magenta: 0.1; and blue: 0.05) obtained from the theoretical arguments [i.e., Eq. (7)]. (b) Correlation skill of the ensemble mean as a function of the inverse of SNR obtained from the theoretical arguments (red dots) and from an ensemble of 100 simulations using the toy model (black dots; repeated 100 times for each SNR value). Blue dots show the averaged correlation skill when only one member out of the 100-member ensemble generated by the toy model is used [i.e.,  $\bar{r}$  in Eq. (5)].

### c. Sources of internal variability in hurricane frequency

In AMIP-type GCM simulations, internal variability in NA hurricane frequency may originate from the following four aspects: 1) differences in seasonal-mean atmospheric environment, 2) intraseasonal variations in atmospheric environment, 3) differences in wave activity (including structure and amplitude) that is associated with the African easterly jet and midlatitude fronts, and 4) internal nonlinear processes associated with deep convection and interactions between disturbances and their synoptic environment. The contribution of differences in the imposed SSTs is small, as discussed before.

We address the first aspect by calculating the SNR used in Mei et al. (2014, 2015), shown in Table 3 for hurricane frequency, GPI, and its four components:

$$R = \frac{\sigma_F}{\sigma_I}, \quad (8)$$

where  $\sigma_F$  is the standard deviation of the ensemble mean and  $\sigma_I$  is the standard deviation of the departures from the ensemble mean in all ensemble members. A large value of  $R$  suggests relatively weak internal variability. It is evident that in general large-scale atmospheric state (particularly potential intensity and vertical wind shear) has weaker

internal variability than hurricane frequency, so does the GPI (Table 3). Adjustments of hurricane frequency and GPI by removing contributions of differences in imposed SSTs among members slightly reduce the internal variability, as expected. In both cases, the ratio of the  $R$  value for GPI to that for hurricane frequency is  $\sim 3$ . Consistently, correlation coefficients between simulated GPIs among ensemble members are much larger and much more narrowly distributed than hurricane frequency (cf. Fig. 4b and Fig. 4a).

We further quantified that on average in individual simulations 63% (37%) of the year-to-year variance in simulated annual hurricane frequency is due to internal (externally forced) variability, while for simulated GPI these two fractions are 14% and 86%, respectively. Thus, the internal variability in GPI may account for nearly 10% of the internal variability (or 6% of the total variability) in hurricane frequency. All these results suggest that seasonal-mean large-scale environment exhibits weak internal variability, and contributes little to the internal variability in hurricane frequency, consistent with Done et al. (2014).

In addition, we note that the  $R$  value of TC frequency is around 1.6 in Mei et al. (2014), whereas it is only 0.76 in the present study. In addition to differences in model configurations, the primary reasons for such a big

TABLE 3. Signal-to-noise ratio (SNR) of annual hurricane frequency and GPI and its four components calculated using Eq. (8) and model simulations for the period of 1951–2010; SNR values calculated using adjusted hurricane frequency and GPI are shown in the parentheses.

Hurricane frequency	GPI	Potential intensity	Vertical wind shear	850-hPa vorticity	600-hPa entropy deficit
0.74 (0.76)	2.04 (2.47)	2.41	1.37	0.66	1.16

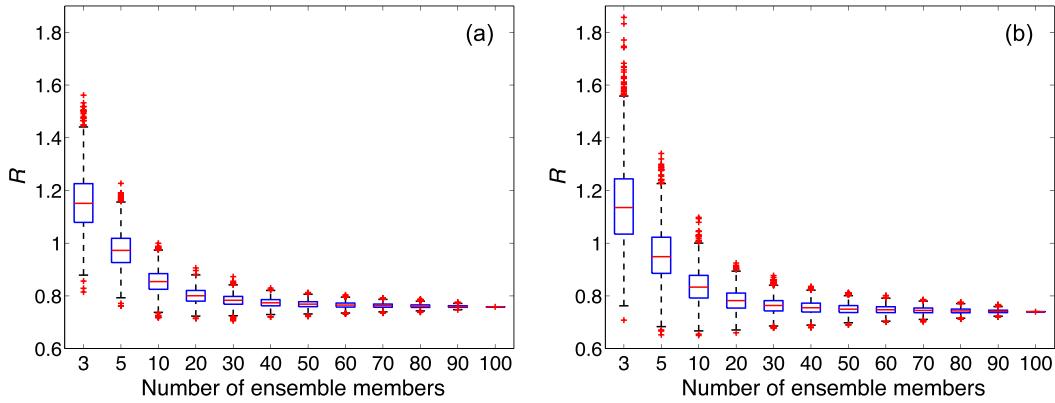


FIG. 9. Box-and-whisker plots of the distribution of SNR calculated using Eq. (8) as a function of ensemble size (a) during 1951–2010 and (b) during 1979–2008.

difference may include 1) the ensemble has a much larger size here than in Mei et al. (2014) (100 versus 3), and 2) the simulations cover a longer period of time in the present study (60 versus 30 years). As demonstrated in Fig. 9a, increasing the ensemble size reduces  $R$ , achieved by both decreasing  $\sigma_F$  and increasing  $\sigma_I$  in Eq. (8), and a small ensemble size may substantially overestimate the SNR. We also repeated the calculations of  $R$  using the d4PDF simulations during 1979–2008, which is the study period in Mei et al. (2014). We found that decreasing the length of the simulations does not significantly affect the median or mean value of  $R$ , although it increases the range of  $R$  owing to the sampling (Fig. 9b).

The above comparison of the  $R$  value between hurricane frequency and GPI suggests that the internal

variability in hurricane frequency is dominated by the other three aspects mentioned above rather than the seasonal-mean environment. This can be further illustrated by Fig. 10a, which shows PDFs of correlation coefficients between hurricane frequency and various large-scale variables including GPI in individual simulations. For the convenience of comparison, the results for the ensemble mean are shown as solid dots. In line with the ensemble mean, in general potential intensity and vertical wind shear are the controlling factors of the variability in hurricane frequency in individual simulations, and GPI and relative SST are relatively good predictors. However, the variance in hurricane frequency explained by GPI in individual simulations (~30%) is substantially smaller than in the ensemble mean (80%), suggesting that an identical GPI may

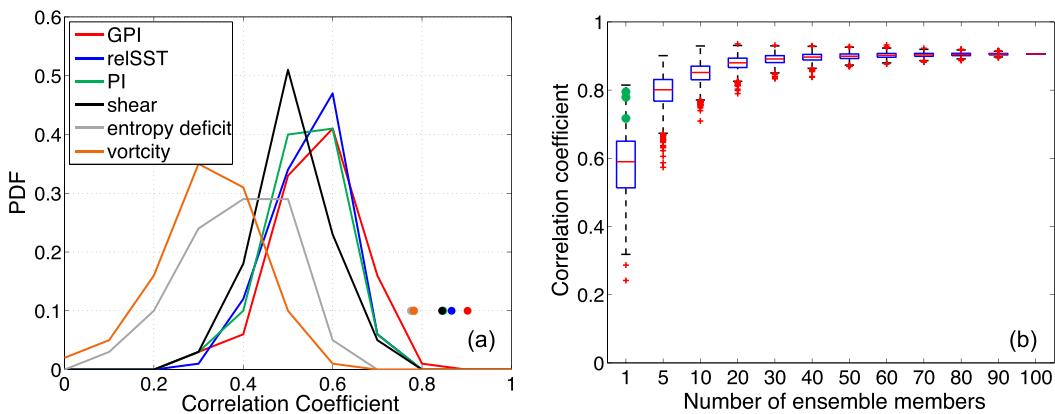


FIG. 10. (a) PDF of the correlation coefficients between annual hurricane frequency and MDR GPI (red), relative SST (blue), potential intensity (PI; green), vertical wind shear (black), 600-hPa entropy deficit (gray), and 850-hPa vorticity (brown) in individual simulations during 1951–2010. Color dots show the results for the ensemble mean; note the black and green dots almost overlap with each other. (b) Box-and-whisker plots of the distribution of the correlation coefficients between annual hurricane frequency and MDR GPI in simulations as a function of ensemble size during 1970–2010. Green dots show the results from the three reanalysis datasets (i.e., JRA-55, NCEP-NCAR-1, and ERA-20C).

produce different numbers of hurricanes from one member simulation to the next. Similar results apply to the relative SSTs and the four components of GPI. Increasing the size of ensemble members can quickly reduce the effect of noise and make it possible to build the connections between seasonal-mean environment and hurricane frequency (Fig. 10b). This has important implications for seasonal predictions of hurricane activity.

We also show in Fig. 10b the correlation coefficients between observed hurricane frequency and GPI in the three reanalysis datasets between 1970 and 2010 (green dots). The stronger correlations in observations suggest stronger climate control of hurricane activity in reality than in model simulations. We further notice that the RPC value for the simulations of MDR GPI is 0.98, significantly different from the value for hurricane frequency (i.e., 1.35; see section 4a). This suggests that the model is faithful in simulating the observed variability in seasonal-mean large-scale environment, but hurricane frequency in individual member simulations contains too much noise compared to the observations. Quantifying the respective contributions of the three aspects mentioned above may shed light on this issue.

## 5. Summary and conclusions

This study has examined the forced and internal variability in annual North Atlantic (NA) hurricane frequency between 1951 and 2010 using an ensemble of 100 simulations performed using an atmospheric general circulation model (AGCM) with a resolution of 60 km. Forced by observed sea surface temperatures (SSTs), the model is skillful at reproducing the observed interannual-to-decadal variability in hurricane frequency during 1970–2010, demonstrating the strong SST control of NA hurricane activity.

Prior to the mid-1960s when satellite measurements were unavailable, the ensemble mean hurricane frequency in the simulations is higher than frequency in current best track data, suggesting a possible underestimate of one hurricane per year in the latter. This is further corroborated by calculations of a genesis potential index (GPI) averaged over the main development region (MDR) using three reanalysis datasets and a comparison of the obtained GPI with hurricane frequency in the best track data.

Correlations of hurricane frequency with the MDR GPI and its four components in the simulations reveal the dominance of potential intensity and vertical wind shear, which is consistent with observations. Further calculations show that relative SSTs (defined as the difference between MDR SSTs and global tropical mean

SSTs) is a good predictor of annual hurricane frequency, with 1°C increase in relative SSTs producing  $7.05 \pm 1.39$  more hurricanes; the effect of relative SSTs is achieved primarily by potential intensity.

We then proceeded to investigate the internal variability in hurricane frequency using the unprecedented large ensemble size. By comparing correlations between observations and simulations with those between simulations and calculating the ratio of predictable components, we show that the modeled hurricane frequency differs from that observed. Specifically, model simulations appear to contain more noise, making the model overdispersive. As a result, 1) individual simulations are more similar to observations than each other, and 2) observations have a higher predictability than the model and the model is underconfident.

We also explored the impact of ensemble size on the correlation skill of the simulations, and found that 20 members are sufficient to capture the forced variability in NA hurricane frequency. The behaviors of the correlation skill of the ensemble mean in the AGCM simulations can be well explained using a toy model and theoretical arguments. The latter two tools allow us to further study the dependence of the correlation skill on signal-to-noise ratio (SNR). The results suggest that for other variables with a smaller SNR (e.g., hurricane lifetime peak intensity), many more members are needed. In addition, we note that the ensemble size also affects the estimation of SNR: a small ensemble size can substantially underestimate the internal variability.

The sources of internal variability in hurricane frequency were also briefly discussed. By comparing the SNR of hurricane frequency with that of seasonal-mean atmospheric environment, we show that seasonal-mean environment contributes little (~10%) to the internal variability in simulated hurricane frequency. This is further supported by the evidence that the correlation between GPI and hurricane frequency in individual simulations is much weaker than in ensemble mean. (This implies that it could be problematic to predict hurricane frequency using GPI in any one individual simulation; instead, GPI predicts the hurricane frequency averaged over 20 or more simulations.) The internal variability in hurricane frequency is thus primarily from 1) intraseasonal variations in atmospheric environment; 2) weather variability (including structure and amplitude) associated with the African easterly jet and midlatitude fronts; and 3) internal nonlinear processes associated with deep convection and interactions between disturbances and their synoptic environment (e.g., Reasor et al. 2005). A quantification of the respective contributions of these three factors, which may also gain insights into the understanding of why the

model is faithful in simulating the variability of seasonal-mean environment but produces excessive noise in hurricane frequency, is left for a future study. In particular, around 80% of NA hurricanes originate from African easterly waves (AEWs) in observations (e.g., Russell et al. 2017), whereas a recent modeling study suggests that a similar climatology of NA TCs can form even if all AEWs are removed (Patricola et al. 2018). It will be of great interest to examine the extent to which the activity of AEWs contributes to the strong internal variability in simulated hurricane frequency.

As mentioned in section 2, one caveat of this study is that prescribing SST perturbs the surface energy fluxes on which real hurricanes depend. It remains unclear how this affects model-generated storms. Future experiments using coupled models—or even those coupled to only a mixed layer ocean—would offer relief from this concern. In addition, it is desirable to test our results with global or regional models of higher resolutions [such as the Nonhydrostatic Icosahedral Atmospheric Model (NICAM; Satoh et al. 2014)].

*Acknowledgments.* This work was supported by the Department of Homeland Security's Coastal Resilience Center of Excellence, a startup fund from the University of North Carolina at Chapel Hill (W.M.), and the National Science Foundation (1637450; S.P.X.). This study used d4PDF produced with the Earth Simulator jointly by science programs (SOUSEI, TOUGOU, SI-CAT, DIAS) of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. We thank Prof. Kerry Emanuel for sharing the compiled tropical cyclone best track data, and we thank the editor and the anonymous reviewers for their comments that helped improve the manuscript.

#### REFERENCES

- Bacmeister, J. T., K. A. Reed, C. Hannay, P. Lawrence, S. Bates, J. E. Truesdale, N. Rosenbloom, and M. Levy, 2018: Projected changes in tropical cyclone activity under future warming scenarios using a high-resolution climate model. *Climatic Change*, **146**, 547–560, <https://doi.org/10.1007/s10584-016-1750-x>.
- Bell, G. D., and Coauthors, 2000: Climate assessment for 1999. *Bull. Amer. Meteor. Soc.*, **81**, S1–S50, [https://doi.org/10.1175/1520-0477\(2000\)81\[s1:CAF\]2.0.CO;2](https://doi.org/10.1175/1520-0477(2000)81[s1:CAF]2.0.CO;2).
- Bruyère, C. L., G. J. Holland, and E. Towler, 2012: Investigating the use of a genesis potential index for tropical cyclones in the North Atlantic basin. *J. Climate*, **25**, 8611–8626, <https://doi.org/10.1175/JCLI-D-11-00619.1>.
- Camargo, S. J., M. Ting, and Y. Kushnir, 2013: Influence of local and remote SST on North Atlantic tropical cyclone potential intensity. *Climate Dyn.*, **40**, 1515–1529, <https://doi.org/10.1007/s00382-012-1536-4>.
- Camp, J., M. Roberts, C. MacLachlan, E. Wallace, L. Hermanson, A. Brookshaw, A. Arribas, and A. A. Scaife, 2015: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 2206–2219, <https://doi.org/10.1002/qj.2516>.
- Chang, E. K. M., and Y. Guo, 2007: Is the number of North Atlantic tropical cyclones significantly underestimated prior to the availability of satellite observations? *Geophys. Res. Lett.*, **34**, L14801, <https://doi.org/10.1029/2007GL030169>.
- Chen, J.-H., and S.-J. Lin, 2011: The remarkable predictability of inter-annual variability of Atlantic hurricanes during the past decade. *Geophys. Res. Lett.*, **38**, L11804, <https://doi.org/10.1029/2011GL047629>.
- , and —, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, <https://doi.org/10.1175/JCLI-D-12-00061.1>.
- Daloz, A. S., F. Chauvin, K. Walsh, S. Lavender, D. Abbs, and F. Roux, 2012: The ability of general circulation models to simulate tropical cyclones and their precursors over the North Atlantic main development region. *Climate Dyn.*, **39**, 1559–1576, <https://doi.org/10.1007/s00382-012-1290-7>.
- DeMaria, M., J. A. Knaff, and B. H. Connel, 2001: A tropical cyclone genesis parameter for the tropical Atlantic. *Wea. Forecasting*, **16**, 219–233, [https://doi.org/10.1175/1520-0434\(2001\)016<0219:ATCGPF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0219:ATCGPF>2.0.CO;2).
- Done, J. M., C. L. Bruyère, M. Ge, and A. Jaye, 2014: Internal variability of North Atlantic tropical cyclones. *J. Geophys. Res. Atmos.*, **119**, 6506–6519, <https://doi.org/10.1002/2014JD021542>.
- Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620–5628, <https://doi.org/10.1002/2014GL061146>.
- Elsner, J. B., and T. H. Jagger, 2006: Prediction models for annual U.S. hurricane counts. *J. Climate*, **19**, 2935–2952, <https://doi.org/10.1175/JCLI3729.1>.
- Emanuel, K., 2001: Contribution of tropical cyclone to meridional heat transport by the oceans. *J. Geophys. Res.*, **106**, 14 771–14 781, <https://doi.org/10.1029/2000JD900641>.
- , 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688, <https://doi.org/10.1038/nature03906>.
- , 2007: Environmental factors affecting tropical cyclone power dissipation. *J. Climate*, **20**, 5497–5509, <https://doi.org/10.1175/2007JCLI1571.1>.
- , 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 A. Sobel, 2013: Response of tropical sea surface temperature, precipitation, and tropical cyclone-related variables to changes in global and local forcing. *J. Adv. Model. Earth Syst.*, **5**, 447–458, <https://doi.org/10.1002/jame.20032>.
- Goldenberg, S. B., and L. J. Shapiro, 1996: Physical mechanisms for the association of El Niño and West African rainfall with Atlantic major hurricane activity. *J. Climate*, **9**, 1169–1187, [https://doi.org/10.1175/1520-0442\(1996\)009<1169:PMFTAO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1169:PMFTAO>2.0.CO;2).
- Halperin, D. J., H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, 2013: An evaluation of tropical cyclone genesis forecasts from global numerical models. *Wea. Forecasting*, **28**, 1423–1445, <https://doi.org/10.1175/WAF-D-13-00008.1>.
- Hart, R. E., 2011: An inverse relationship between aggregate Northern Hemisphere tropical cyclone activity and subsequent winter climate. *Geophys. Res. Lett.*, **38**, L01705, <https://doi.org/10.1029/2010GL045612>.

- Hirahara, S., M. Ishii, and Y. Fukuda, 2014: Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate*, **27**, 57–75, <https://doi.org/10.1175/JCLI-D-12-00837.1>.
- Johnson, N. C., and S.-P. Xie, 2010: Changes in the sea surface temperature threshold for tropical convection. *Nat. Geosci.*, **3**, 842–845, <https://doi.org/10.1038/ngeo1008>.
- Jourdain, N. C., P. Marchesio, C. E. Menkes, J. Lefèvre, E. M. Vincent, M. Lengaigne, and F. Chauvin, 2011: Mesoscale simulation of tropical cyclones in the South Pacific: Climatology and interannual variability. *J. Climate*, **24**, 3–25, <https://doi.org/10.1175/2010JCLI3559.1>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kamae, Y., W. Mei, S.-P. Xie, M. Naoi, and H. Ueda, 2017a: Atmospheric rivers over the northwestern Pacific: Climatology and interannual variability. *J. Climate*, **30**, 5605–5619, <https://doi.org/10.1175/JCLI-D-16-0875.1>.
- , and Coauthors, 2017b: Forced response and internal variability of summer climate over western North America. *Climate Dyn.*, **49**, 403–417, <https://doi.org/10.1007/s00382-016-3350-x>.
- Klotzbach, P. J., 2011: El Niño–Southern Oscillation’s impact on Atlantic basin hurricanes and U.S. landfalls. *J. Climate*, **24**, 1252–1263, <https://doi.org/10.1175/2010JCLI3799.1>.
- , and W. M. Gray, 2009: Twenty-five years of Atlantic basin seasonal hurricane forecasts (1984–2008). *Geophys. Res. Lett.*, **36**, L09711, <https://doi.org/10.1029/2009GL037580>.
- Knaff, J. A., 1997: Implications of summertime sea level pressure anomalies in the tropical Atlantic region. *J. Climate*, **10**, 789–804, [https://doi.org/10.1175/1520-0442\(1997\)010<0789:IOSSLP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<0789:IOSSLP>2.0.CO;2).
- Knutson, T. R., J. J. Sirutis, S. T. Garner, G. A. Vecchi, and I. M. Held, 2008: Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions. *Nat. Geosci.*, **1**, 359–364, <https://doi.org/10.1038/ngeo202>.
- , and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.*, **3**, 157–163, <https://doi.org/10.1038/ngeo779>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Korty, R. L., K. A. Emanuel, and J. R. Scott, 2008: Tropical cyclone-induced upper-ocean mixing and climate: Application to equable climates. *J. Climate*, **21**, 638–654, <https://doi.org/10.1175/2007JCLI1659.1>.
- , S. J. Camargo, and J. Galewsky, 2012: Tropical cyclone genesis factors in simulations of the Last Glacial Maximum. *J. Climate*, **25**, 4348–4365, <https://doi.org/10.1175/JCLI-D-11-00517.1>.
- Kossin, J. P., and D. J. Vimont, 2007: A more general framework for understanding Atlantic hurricane variability and trends. *Bull. Amer. Meteor. Soc.*, **88**, 1767–1781, <https://doi.org/10.1175/BAMS-88-11-1767>.
- , S. J. Camargo, and M. Sitkowski, 2010: Climate modulation of North Atlantic hurricane tracks. *J. Climate*, **23**, 3057–3076, <https://doi.org/10.1175/2010JCLI3497.1>.
- Landsea, C. W., and W. M. Gray, 1992: The strong association between western Sahelian monsoon rainfall and intense Atlantic hurricanes. *J. Climate*, **5**, 435–453, [https://doi.org/10.1175/1520-0442\(1992\)005<0435:TSABWS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005<0435:TSABWS>2.0.CO;2).
- , 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>.
- , R. A. Pielke, A. M. Mestas-Nuñez, and J. A. Knaff, 1999: Atlantic basin hurricanes: Indices of climatic changes. *Climatic Change*, **42**, 89–129, <https://doi.org/10.1023/A:1005416332322>.
- Li, H., and R. L. Sriver, 2018: Tropical cyclone activity in the high-resolution Community Earth System Model and the impact of ocean coupling. *J. Adv. Model. Earth Syst.*, **10**, 165–186, <https://doi.org/10.1002/2017MS001199>.
- Manganello, J. V., and Coauthors, 2016: Seasonal forecasts of tropical cyclone activity in a high-atmospheric-resolution coupled prediction system. *J. Climate*, **29**, 1179–1200, <https://doi.org/10.1175/JCLI-D-15-0531.1>.
- McAdie, C. J., C. W. Landsea, C. J. Neumann, J. E. David, E. Blake, and G. R. Hammer, 2009: *Tropical Cyclones of the North Atlantic Ocean, 1851–2006*. Historical Climatology Series 6-2, National Climatic Data Center and National Hurricane Center, 238 pp., [https://www.nhc.noaa.gov/pdf/TC\\_Book\\_Atl\\_1851-2006\\_lowres.pdf](https://www.nhc.noaa.gov/pdf/TC_Book_Atl_1851-2006_lowres.pdf).
- McBride, J. L., and R. Zehr, 1981: Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems. *J. Atmos. Sci.*, **38**, 1132–1151, [https://doi.org/10.1175/1520-0469\(1981\)038<1132:OAOTCF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<1132:OAOTCF>2.0.CO;2).
- Mei, W., F. Primeau, J. C. McWilliams, and C. Pasquero, 2013: Sea surface height evidence for long-term warming effects of tropical cyclones on the ocean. *Proc. Natl. Acad. Sci. USA*, **110**, 15 207–15 210, <https://doi.org/10.1073/pnas.1306753110>.
- , S.-P. Xie, and M. Zhao, 2014: Variability of tropical cyclone track density in the North Atlantic: Observations and high-resolution simulations. *J. Climate*, **27**, 4797–4814, <https://doi.org/10.1175/JCLI-D-13-00587.1>.
- , —, —, and Y. Wang, 2015: Forced and internal variability of tropical cyclone track density in the western North Pacific. *J. Climate*, **28**, 143–167, <https://doi.org/10.1175/JCLI-D-14-00164.1>.
- Mizuta, R., and Coauthors, 2012: Climate simulations using MRI-AGCM with 20-km grid. *J. Meteor. Soc. Japan*, **90A**, 233–258, <https://doi.org/10.2151/jmsj.2012-A12>.
- , and Coauthors, 2017: Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bull. Amer. Meteor. Soc.*, **98**, 1383–1398, <https://doi.org/10.1175/BAMS-D-16-0099.1>.
- Murakami, H., R. Mizuta, and E. Shindo, 2012: Future changes in tropical cyclone activity projected by multi-physics and multi-SST ensemble experiments using the 60-km-mesh MRI-AGCM. *Climate Dyn.*, **39**, 2569–2584, <https://doi.org/10.1007/s00382-011-1223-x>.
- , G. Villarini, G. A. Vecchi, W. Zhang, and R. Gudgel, 2016: Statistical-dynamical seasonal forecast of North Atlantic and U.S. landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. *Mon. Wea. Rev.*, **144**, 2101–2123, <https://doi.org/10.1175/MWR-D-15-0308.1>.
- Nolan, D. S., and E. D. Rappin, 2008: Increased sensitivity of tropical cyclogenesis to wind shear in higher SST environments. *Geophys. Res. Lett.*, **35**, L14805, <https://doi.org/10.1029/2008GL034147>.
- Pasch, R. J., P. A. Harr, L. A. Avila, J. G. Jiing, and G. Elliot, 2006: An evaluation and comparison of predictions of tropical cyclogenesis by three global forecast models. *27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., 14B.5, <https://ams.confex.com/ams/pdfpapers/108725.pdf>.
- Patricola, C. M., R. Saravanan, and P. Chang, 2014: The impact of the El Niño–Southern Oscillation and Atlantic meridional

- mode on seasonal Atlantic tropical cyclone activity. *J. Climate*, **27**, 5311–5328, <https://doi.org/10.1175/JCLI-D-13-00687.1>.
- , —, and —, 2018: The response of Atlantic tropical cyclones to suppression of African easterly waves. *Geophys. Res. Lett.*, **45**, 471–479, <https://doi.org/10.1002/2017GL076081>.
- Pielke, R. A., Jr., and C. W. Landsea, 1998: Normalized hurricane damages 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](https://doi.org/10.1175/1520-0434(1998)013<0621:NHDITU>2.0.CO;2).
- , J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin, 2008: Normalized hurricane damage in the United States: 1900–2005. *Nat. Hazards Rev.*, **9**, 29–42, [https://doi.org/10.1061/\(ASCE\)1527-6988\(2008\)9:1\(29\)](https://doi.org/10.1061/(ASCE)1527-6988(2008)9:1(29)).
- Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097, <https://doi.org/10.1175/JCLI-D-15-0556.1>.
- Reasor, P. D., M. T. Montgomery, and L. F. Bosart, 2005: Mesoscale observations of the genesis of Hurricane Dolly (1996). *J. Atmos. Sci.*, **62**, 3151–3171, <https://doi.org/10.1175/JAS3540.1>.
- Roberts, M. J., and Coauthors, 2015: Tropical cyclones in the upscale ensemble of high-resolution global climate models. *J. Climate*, **28**, 574–596, <https://doi.org/10.1175/JCLI-D-14-00131.1>.
- Russell, J. O., A. Aiyyer, J. D. White, and W. Hannah, 2017: Revisiting the connection between African easterly waves and Atlantic tropical cyclogenesis. *Geophys. Res. Lett.*, **44**, 587–595, <https://doi.org/10.1002/2016GL071236>.
- Satoh, M., and Coauthors, 2014: The non-hydrostatic icosahedral atmospheric model: Description and development. *Prog. Earth Planet. Sci.*, **1**, 1–32, <https://doi.org/10.1186/s40645-014-0018-1>.
- Saunders, M. A., P. J. Klotzbach, and A. S. R. Lea, 2017: Replicating annual North Atlantic hurricane activity 1878–2012 from environmental variables. *J. Geophys. Res. Atmos.*, **122**, 6284–6297, <https://doi.org/10.1002/2017JD026492>.
- Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010: Skilful multi-year predictions of Atlantic hurricane frequency. *Nat. Geosci.*, **3**, 846–849, <https://doi.org/10.1038/ngeo1004>.
- Sriver, R. L., and M. Huber, 2007: Observational evidence for an ocean heat pump induced by tropical cyclones. *Nature*, **447**, 577–580, <https://doi.org/10.1038/nature05785>.
- Tang, B., and K. Emanuel, 2012: A ventilation index for tropical cyclones. *Bull. Amer. Meteor. Soc.*, **93**, 1901–1912, <https://doi.org/10.1175/BAMS-D-11-00165.1>.
- Ueda, H., K. Miwa, and Y. Kamae, 2018: Seasonal modulation of tropical cyclone occurrence associated with coherent Indo-Pacific variability during decaying phase of El Niño. *J. Meteor. Soc. Japan*, **96**, 381–390, <https://doi.org/10.2151/jmsj.2018-044>.
- 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, 2011: Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878–1965) using ship track density. *J. Climate*, **24**, 1736–1746, <https://doi.org/10.1175/2010JCLI3810.1>.
- , and G. Villarini, 2014: Next season’s hurricanes. *Science*, **343**, 618–619, <https://doi.org/10.1126/science.1247759>.
- , 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>.
- Vimont, D. J., and J. P. Kossin, 2007: The Atlantic Meridional Mode and hurricane activity. *Geophys. Res. Lett.*, **34**, L07709, <https://doi.org/10.1029/2007GL029683>.
- Vitart, F., and Coauthors, 2007: Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP. *Geophys. Res. Lett.*, **34**, L16815, <https://doi.org/10.1029/2007GL030740>.
- Walsh, K. J. E., and Coauthors, 2015: Hurricanes and climate: The U.S. CLIVAR Working Group on Hurricanes. *Bull. Amer. Meteor. Soc.*, **96**, 997–1017, <https://doi.org/10.1175/BAMS-D-13-00242.1>.
- , M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Climate*, **20**, 2307–2314, <https://doi.org/10.1175/JCLI4074.1>.
- Woodruff, J. D., J. L. Irish, and S. J. Camargo, 2013: Coastal flooding by tropical cyclones and sea-level rise. *Nature*, **504**, 44–52, <https://doi.org/10.1038/nature12855>.
- Wu, C.-C., R. Zhan, Y. Lu, and Y. Wang, 2012: Internal variability of the dynamically downscaled tropical cyclone activity over the western North Pacific by the IPRC regional atmospheric model. *J. Climate*, **25**, 2104–2122, <https://doi.org/10.1175/JCLI-D-11-00143.1>.
- Yoshida, K., M. Sugi, R. Mizuta, H. Murakami, and M. Ishii, 2017: Future changes in tropical cyclone activity in high-resolution large-ensemble simulations. *Geophys. Res. Lett.*, **44**, 9910–9917, <https://doi.org/10.1002/2017GL075058>.
- Zhang, G., Z. Wang, T. J. Dunkerton, M. S. Peng, and G. Magnusdottir, 2016: Extratropical impacts on Atlantic tropical cyclone activity. *J. Atmos. Sci.*, **73**, 1401–1418, <https://doi.org/10.1175/JAS-D-15-0154.1>.
- Zhao, M., and I. M. Held, 2012: Tc-permitting GCM simulations of hurricane frequency response to sea surface temperature anomalies projected for the late-twenty-first century. *J. Climate*, **25**, 2995–3009, <https://doi.org/10.1175/JCLI-D-11-00313.1>.
- , —, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, **22**, 6653–6678, <https://doi.org/10.1175/2009JCLI3049.1>.