Seasonal Forecasts of Major Hurricanes and Landfalling Tropical Cyclones using a High-Resolution GFDL Coupled Climate Model

HIROYUKI MURAKAMI,[a,b] GABRIEL A. VECCHI,[a,b] GABRIELE VILLARINI,[c] THOMAS L. DELWORTH,[a,b] RICHARD GUDGEL,[a] SETH UNDERWOOD,[d] XIAOSONG YANG,[a,e] WEI ZHANG,[a,b,c] AND SHIAN-JIANN LIN[a]

[a] *NOAA/GFDL, Princeton, New Jersey*
[b] *Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey*
[c] *IIHR—Hydroscience and Engineering, The University of Iowa, Iowa City, Iowa*
[d] *Engility, NOAA/GFDL, Princeton, New Jersey*
[e] *University Corporation for Atmospheric Research, Boulder, Colorado*

(Manuscript received 19 March 2016, in final form 1 August 2016)

ABSTRACT

Skillful seasonal forecasting of tropical cyclone (TC; wind speed $\geq 17.5\ m\ s^{-1}$) activity is challenging, even more so when the focus is on major hurricanes (wind speed $\geq 49.4\ m\ s^{-1}$), the most intense hurricanes (category 4 and 5; wind speed $\geq 58.1\ m\ s^{-1}$), and landfalling TCs. This study shows that a 25-km-resolution global climate model [High-Resolution Forecast-Oriented Low Ocean Resolution (FLOR) model (HiFLOR)] developed at the Geophysical Fluid Dynamics Laboratory (GFDL) has improved skill in predicting the frequencies of major hurricanes and category 4 and 5 hurricanes in the North Atlantic as well as landfalling TCs over the United States and Caribbean islands a few months in advance, relative to its 50-km-resolution predecessor climate model (FLOR). HiFLOR also shows significant skill in predicting category 4 and 5 hurricanes in the western North Pacific and eastern North Pacific, while both models show comparable skills in predicting basin-total and landfalling TC frequency in the basins. The improved skillful forecasts of basin-total TCs, major hurricanes, and category 4 and 5 hurricane activity in the North Atlantic by HiFLOR are obtained mainly by improved representation of the TCs and their response to climate from the increased horizontal resolution rather than by improvements in large-scale parameters.

1. Introduction

Tropical cyclones (TCs) are one of the most costly natural disasters to affect coastal regions all over the world (e.g., Pielke et al. 2008; Smith and Katz 2013). In recent history, about 85% of the total TC damage has been caused by major hurricanes [Saffir–Simpson categories 3, 4, and 5 (C345)], even though they make up a very small fraction of overall TCs (e.g., Pielke et al. 2008). Furthermore, even though nonlandfalling TCs can cause damage (e.g., to offshore energy platforms and ships), landfalling TCs contribute substantially more to overall TC damages than nonlandfalling TCs do. Therefore, predicting intense hurricanes and landfalling storms at seasonal time scales is a topic of large scientific and socioeconomic interest (Vecchi and Villarini 2014).

Since Gray (1984a,b), there has been a large body of literature on statistical seasonal forecasts of TC activity, in which observed large-scale climate ahead of the TC season has been used for predictions of TC activity [see Camargo et al. (2007) or Murakami et al. (2016) for a review]. On the other hand, operational dynamical seasonal TC forecasts began at the European Centre for Medium-Range Weather Forecasts (ECMWF) in 2001 (Vitart and Stockdale 2001) and at the International Research Institute for Climate and Society (IRI) in 2003 (Camargo and Barnston 2009). Since then, a number of high-resolution dynamical models, along with multimodel ensemble techniques, have demonstrated reasonably skillful forecasting of the basin-total frequency of tropical storms and hurricanes (e.g., Vitart 2006; Vitart et al. 2007; LaRow et al. 2008, 2010; Zhao et al. 2010; Alessandri et al. 2011; Chen and Lin 2011, 2013; Vecchi et al. 2014; Camp et al. 2015; Manganello et al. 2016) and regional TC activity (Vecchi et al. 2014; Camp et al. 2015; Manganello et al. 2016). Specifically, Chen

Corresponding author address: Hiroyuki Murakami, NOAA/GFDL, 201 Forrestal Rd., Princeton, NJ 08540-6649.
E-mail: hir.murakami@gmail.com

DOI: 10.1175/JCLI-D-16-0233.1

and Lin (2011) reported a correlation coefficient of 0.96 between observed and predicted year-by-year variation in hurricanes (i.e., storms with maximum wind speed greater than 32.9 m s^{-1}), and Vecchi et al. (2014) reported skillful forecasting months in advance for regional basin-wide hurricane activity across the Northern Hemisphere. However, the prediction of C345 hurricanes and landfall tropical storm frequency remains challenging (Camargo et al. 2007; Vecchi and Villarini 2014; Camp et al. 2015; Murakami et al. 2016), although there are some promising results with high-resolution dynamical models for U.S. landfalling frequency (Murakami et al. 2016). Therefore, the limitations of dynamical forecasts have been alleviated using empirical statistical–dynamical methods (e.g., Wang et al. 2009; Zhao et al. 2010; Vecchi et al. 2011, 2013, 2014; Villarini and Vecchi 2013; Murakami et al. 2016).

Given the limitations of dynamical forecasts, much of the literature on operational dynamical predictions has focused on basin-total frequency of tropical storms and hurricanes as well as basin-total values of the accumulated cyclone energy (ACE; Bell et al. 2000) [see Table II in Camargo et al. (2007)]. The skill in predicting landfalling TCs and C345 hurricanes at seasonal time scale has not been reported in the literature so far, although they are of paramount societal and scientific importance (Vecchi and Villarini 2014). On the other hand, Murakami et al. (2015, henceforth M15) provided a preliminary assessment of the predictability of Saffir–Simpson category 4 and 5 (C45) hurricanes in the high-resolution global coupled climate model [High-Resolution Forecast-Oriented Low Ocean Resolution (FLOR) model (HiFLOR)] that was developed at the Geophysical Fluid Dynamics Laboratory (GFDL). M15 evaluated the skill of a couplet of seasonal forecasts initialized on 1 July 1997 and 1998. Those retrospective predictions captured the observed sharp contrast of global TC activity driven by the extreme El Niño event of 1997/98 and La Niña of 1998/99. Although HiFLOR could predict the contrast in the C45 hurricanes for 1997 and 1998 summer seasons because of the extreme nature of the large-scale climate forcing from ENSO, it is unclear whether the skill reported in M15 would apply more broadly to predictions over a larger number of years.

In this study, we conduct a suite of seasonal retrospective predictions initialized on 1 July, 1 April, and 1 January between 1980 and 2015 to evaluate and quantify the skill of HiFLOR in predicting TC activity in the western North Pacific (WNP), eastern North Pacific (ENP), and North Atlantic (NAT) tropical cyclone basins (see Fig. 3 in M15 for regional boundaries). We especially focus on the prediction of intense hurricanes (i.e., C345 and C45 hurricanes) and landfalling TC frequency over the study period in addition to the conventional metrics of basin-total frequencies of tropical storms and hurricanes as well as basin-total values of the ACE and power dissipation index (PDI; Emanuel 2005, 2007). We show for the first time that this high-resolution global atmosphere–ocean coupled model has significant skill in predicting the frequencies of C345 and C45 hurricanes and the frequency of landfalling TCs along with that of basin-total TCs a few months in advance in the NAT as well as in other ocean basins. We further show that both FLOR and HiFLOR can exhibit marginal skill at predicting seasonal U.S. landfalling TC frequency. Section 2 provides a description of the models, seasonal forecasts, and TC detection method along with the observed dataset. Section 3 shows the results. Finally, section 4 gives a summary.

2. Methods

a. *Dynamical models, seasonal forecasts, TC detection methods*

The dynamical models used here are the Forecast-Oriented Low Ocean Resolution version of the Geophysical Fluid Dynamics Laboratory Climate Model, version 2.5 (GFDL CM2.5) (Vecchi et al. 2014), and the high atmospheric resolution version of FLOR (M15). The atmospheric and land components of FLOR are based on an approximately 50-km-resolution cubed-sphere grid, taken from GFDL CM2.5 (Delworth et al. 2012). HiFLOR was developed from FLOR by increasing the horizontal resolution of the atmosphere/land components to a cubed-sphere grid of approximately 25-km resolution, keeping the subgrid physical parameterizations unchanged (M15). The ocean and ice components in both FLOR and HiFLOR are at approximately 1° resolution. HiFLOR yields better simulations of the mean large-scale patterns of surface temperature and rainfall as well as modes of variability such as the El Niño–Southern Oscillation (ENSO) phenomenon and its TC teleconnections in the WNP, ENP, and NAT than FLOR does (M15; Zhang et al. 2016).

For each year and each month in the period 1980–2015, 12-month-duration retrospective seasonal predictions were generated by initializing each model to observationally constrained conditions for the ocean and sea ice components (Vecchi et al. 2014; M15; Murakami et al. 2016). The 12-member initial conditions for the ocean and sea ice were generated with GFDL's ensemble coupled data assimilation system (Zhang and Rosati 2010; Chang et al. 2013) using Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1 (GFDL CM2.1) (Delworth et al. 2006; Gnanadesikan et al. 2006; Wittenberg et al. 2006). For FLOR, the atmosphere and land components were initialized from a suite of sea surface temperature (SST)-forced atmosphere–land-only simulations (Vecchi

et al. 2014), and in the HiFLOR predictions the atmosphere and land were initialized using an arbitrary year from a control climate simulation (M15). Therefore, the HiFLOR forecasts are constructed so the predictability comes entirely from the ocean and sea ice state, and we therefore estimate that they may serve as a lower bound estimate on the prediction skill of these models if they were also to include an observationally constrained atmospheric and land initialization (Jia et al. 2016). HiFLOR has forecasts only from July, April, and January at this moment, whereas FLOR has forecasts starting from every month. Therefore, we mainly focus on the forecasts from July, April, and January initial conditions for the predictions of TC activity in the boreal summer season (i.e., July–November) for the comparisons between FLOR and HiFLOR. Vecchi et al. (2014) showed the prediction skill in basin-total frequency of hurricanes in the NAT by FLOR as compared to the other prediction systems [e.g., Vitart et al. 2007; Klotzbach and Gray 2009; Zhao et al. 2009; LaRow et al. 2010; Wang et al. 2009; Chen and Lin 2013; see Fig. 9 in Vecchi et al. (2014)], highlighting comparable or higher prediction skill in FLOR relative to the other prediction systems. Therefore, showing the prediction skill by HiFLOR relative to FLOR is useful to indicate relative prediction skill by HiFLOR compared to the other prediction systems, although direct comparisons between them are still difficult because of the different evaluation periods and definitions of scores. Forecasts from other initial months by FLOR will be shown for the comparisons of prediction skill in large-scale parameters among FLOR, GFDL CM2.1, and HiFLOR. We define forecasts from July (January) initial conditions as lead-month-0 (L0) [lead month 6 (L6)] forecasts. Because the northern Indian Ocean has one of the two peaks of TC activity before July, we only focus on prediction skill in the WNP, ENP, and NAT during July–November.

Model-generated TCs were detected following Harris et al. (2016) and M15. Briefly, the tracking scheme applies the flood fill algorithm to find closed contours of some specified negative sea level pressure (SLP) anomaly with a warm core (temperature anomaly higher than 1 K for FLOR and 2 K for HiFLOR). The detection scheme also requires that a TC have a duration of at least 36 h while maintaining its warm core in addition to a specified surface wind speed criterion ($15.75\,m\,s^{-1}$ for FLOR and $17.5\,m\,s^{-1}$ for HiFLOR). The model-dependent thresholds for the warm core and surface wind speed are necessary because TC structure critically depends on horizontal resolution; FLOR simulates weaker storms relative to HiFLOR/observations (M15).

b. Observational datasets

The observed TC "best track" data for the period 1980–2015 were obtained from the National Hurricane Center "best track" hurricane database (HURDAT2) (Landsea and Franklin 2013) and Joint Typhoon Warning Center (JTWC) as partially archived in the International Best Track Archive for Climate Stewardship (IBTrACS, version 03 release 07) (Knapp et al. 2010). Because the best track data for 2015 were not fully available for the central Pacific Ocean (CPO; Northern Hemisphere 180°–140°W) at this moment, we obtained the 2015 CPO dataset from the Unisys Corporation website (Unisys 2016). We derived TCs with tropical storm intensities or stronger (i.e., TCs possessing 1-min sustained surface winds of $17.5\,m\,s^{-1}$ or greater) during the period 1980–2015.

To address the differences in the forecast skill in the NAT between FLOR and HiFLOR, we will compare the prediction skill in the four key large-scale parameters relative to observations. The four parameters are geopotential height at 500 hPa over the subtropical ENP Φ_{500} (20°–40°N, 130°–170°W), vertical wind shear over the tropical NAT W_{shear} (10°–20°N, 30°–90°W), SST anomaly over the tropical NAT (SSTA; 0°–20°N, 10°–70°W), and relative humidity at 600 hPa over the tropical NAT RH_{600} (10°–25°N, 10°–90°W). The time series of observed TC frequency in the NAT is highly correlated with those of the area mean values of the key variables in observations (Fig. 1) and predictions [see Fig. 3 in Murakami et al. (2016)]. Murakami et al. (2016) discussed the reason why Φ_{500} over the subtropical ENP shows a high correlation with TC frequency in the NAT. When the anomaly of Φ_{500} is positive in the subtropical ENP, the geopotential height is negative in the subtropical NAT (30°–50°N, 55°–75°W) through a series of wave trains along the subtropical westerly jet associated with a part of the so-called Pacific–North American (PNA) pattern. We use the Hadley Centre Sea Ice and Sea Surface Temperature dataset, version 1.1 (HadISST1.1) (Rayner et al. 2003), and the Japanese 55-year Reanalysis (JRA-55) (Kobayashi et al. 2015) for the period 1980–2015 as observed SST and atmospheric large-scale parameters, respectively.

c. Metrics for evaluation of forecast skill

In this study, storms are categorized into three groups according to their lifetime maximum intensity: tropical cyclones (or tropical storms; wind speed $\geq 17.5\,m\,s^{-1}$), hurricanes (HUR; wind speed $\geq 32.9\,m\,s^{-1}$), and category 3–5 (or major) hurricanes (wind speed $\geq 49.4\,m\,s^{-1}$). In addition, we examine the prediction skill for category 4 and 5 hurricanes (wind speed $\geq 58.1\,m\,s^{-1}$) because our previous study showed potential skill in predicting C45 hurricanes by HiFLOR (M15). Note that

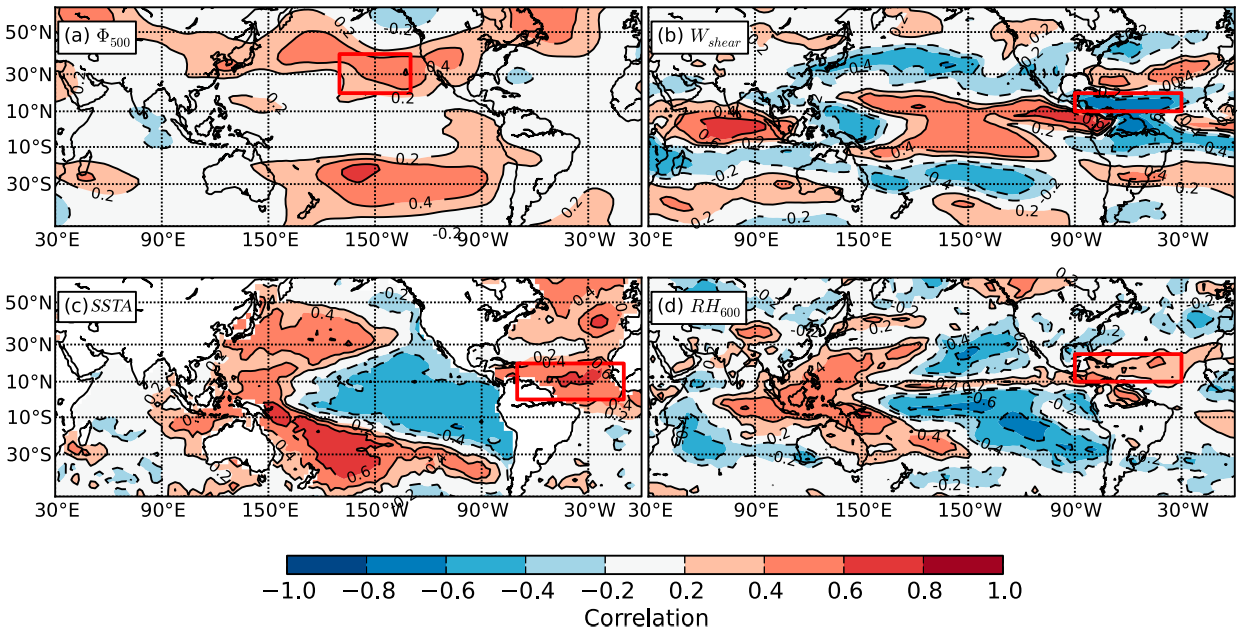

FIG. 1. Correlation maps between the time series of observed TC frequency in the NAT and observed mean large-scale parameters during July–November 1980–2015 for each 2.5° × 2.5° grid box: (a) Φ_{500}, (b) W_{shear} (200–850 hPa), (c) SSTA, and (d) RH_{600}. Rectangles indicate key domains for TC frequency in the NAT.

although a hurricane is called a ''typhoon'' in the WNP, for convenience we will refer to WNP typhoons in this study as hurricanes.

We examined the prediction skill in interannual variation of the basin-total frequencies for the four intensity groups (i.e., TC, HUR, C345, and C45), basin-total values of the ACE and PDI, and basin-total frequency of landfalling TCs. Because the models have systematic model biases in the predicted values, the predicted values are calibrated using observed data; predicted values for each ensemble member are scaled by the ratio of the observed and predicted ensemble-mean values for the period 1980–2015. We used five scores to evaluate prediction skill for the above TC activity relative to observed values: correlation coefficient (COR), rank correlation coefficient (RCOR), root-mean-square error (RMSE) or normalized root-mean-square error (NRMSE), and mean square skill score (MSSS) (Kim et al. 2012; Li et al. 2013). NRMSE is defined as follows:

$$\text{NRMSE} \equiv \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i^{\text{obs}} - f_i)^2}}{\sigma^{\text{obs}}} = \frac{\text{RMSE}}{\sigma^{\text{obs}}}, \qquad (1)$$

where n is the total number of years, f_i^{obs} and f_i are the values from observations and predictions for the i^{th} year, respectively, and σ^{obs} is the observed standard deviation. The RMSE is normalized by the observed standard deviation because we want to compare variables in different units. The MSSS is defined by the following equation:

$$\text{MSSS} \equiv 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(f_i^{\text{obs}} - f_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(f_i^{\text{obs}} - f^{\text{obs}})^2}, \qquad (2)$$

where f^{obs} is the observational mean value. The MSSS is a metric that compares the skill of the model against climatological forecasts, with high values indicating a good predictive skill (Kim et al. 2012; Li et al. 2013).

For the computations of COR and RCOR, we first computed the anomaly from the climatological mean for each year and then computed correlations between the predicted and observed anomalies. In the case of one constant climatological mean (i.e., mean of 1980–2015), there is no difference in the correlations from the anomalies and those from the raw values. However, climatological mean could be varying year by year in the so-called cross-validation mode, in which the anomaly for evaluating year is computed from the time-varying climatological mean data for which the raw data are averaged except for the evaluating year (or multiple years including the evaluating year). We preliminarily computed cross-validated RCORs by excluding 5-year data from the original data to compute time-varying climatological mean and compared the results with the RCORs using the constant climatological mean. It turned out that the

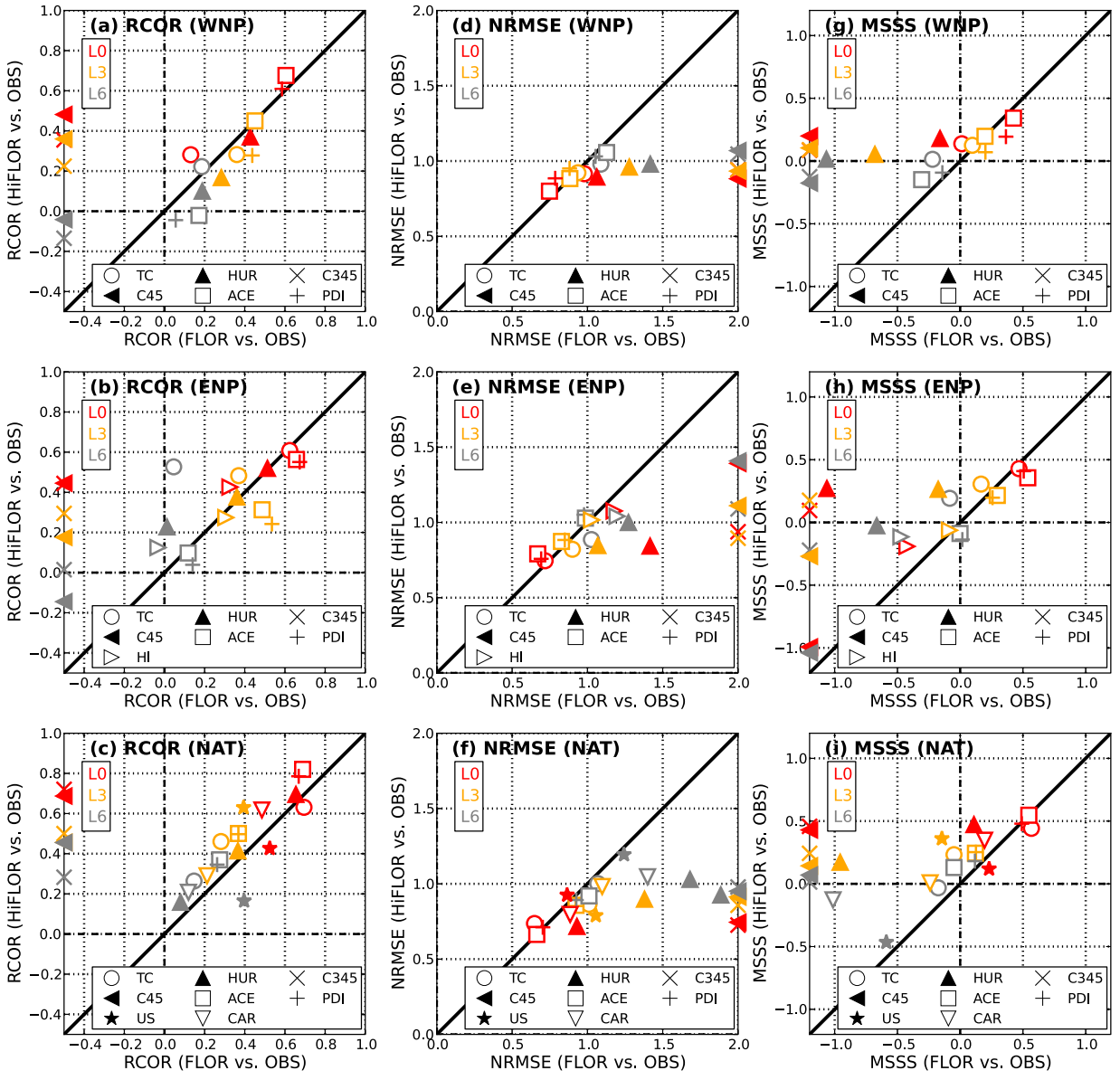

FIG. 2. Scatterplots of RCOR between HiFLOR prediction and observations (*y* axis) and FLOR prediction and observations (*x* axis) for the (a) WNP, (b) ENP, and (c) NAT. A correlation coefficient above the diagonal lines indicates that HiFLOR shows higher correlation than FLOR. (d)–(f) As in (a)–(c), but for NRMSE. An error below the diagonal lines indicates that HiFLOR shows smaller error than FLOR. (g)–(i) As in (a)–(c), but for MSSS. A value of MSSS above the diagonal lines indicates that HiFLOR shows higher skill than FLOR. Variables evaluated are basin-total frequency of TC, HUR, C345, and C45 as well as basin-total values of ACE, PDI, the regional TC frequency for the United States (US), Caribbean islands (CAR), and Hawaiian Islands (HI). Different colors indicate different lead months (L0, L3, and L6). Because FLOR cannot predict C345 and C45 hurricanes, those plots for HiFLOR are located along the *y* axis for convenience.

differences between them were very small (about 0.02, and 0.2 at most), indicating that this cross-validation mode does not change the correlations significantly from those using one constant climatological mean in this study. Therefore, we will show the results of CORs and RCORs using one constant climatological mean of 1980–2015 rather than applying the cross-validation mode in this study.

3. Results

a. Retrospective forecast of basin-total TC activity

We first compare the retrospective forecast skill in basinwide seasonal TC activity between FLOR and HiFLOR using scatterplots of RCOR, NRMSE, and MSSS for interannual variation of seasonal mean value between observations and FLOR versus HiFLOR

TABLE 1. RCORs between observed and predicted TC activity for each initialization month. The metrics for TC activity are the frequency of basin-total TC, HUR, C345, and C45 as well as basin-total ACE (ACE) and frequency of landfalling TCs in the Hawaiian Islands (HI), United States (US), and Caribbean islands (CAR). Boldface number indicates that the RCOR is statistically significant at more than the 90% level.

		Initial month (lead month)							
Category	Model	Jul (0)	Jun (1)	May (2)	Apr (3)	Mar (4)	Feb (5)	Jan (6)	Dec (7)
WNP									
TC	FLOR	+0.13	**+0.30**	−0.03	**+0.36**	+0.26	+0.12	+0.19	**+0.32**
	HiFLOR	**+0.28**			**+0.28**			+0.22	
HUR	FLOR	**+0.43**	**+0.32**	+0.25	**+0.28**	+0.09	+0.10	+0.19	+0.19
	HiFLOR	**+0.37**			+0.17			+0.10	
C345	FLOR	—	—	—	—	—	—	—	—
	HiFLOR	**+0.36**			+0.22			−0.13	
C45	FLOR	—	—	—	—	—	—	—	—
	HiFLOR	**+0.48**			**+0.36**			−0.04	
ACE	FLOR	**+0.61**	**+0.65**	**+0.49**	**+0.45**	**+0.29**	+0.19	+0.17	+0.11
	HiFLOR	**+0.68**			**+0.45**			−0.02	
ENP									
TC	FLOR	**+0.62**	**+0.64**	**+0.62**	**+0.37**	+0.15	+0.01	+0.05	−0.02
	HiFLOR	**+0.61**			**+0.48**			**+0.53**	
HUR	FLOR	**+0.51**	**+0.60**	**+0.46**	**+0.36**	+0.12	+0.06	+0.01	+0.08
	HiFLOR	**+0.52**			**+0.38**			+0.23	
C345	FLOR	—	—	—	—	—	—	—	—
	HiFLOR	**+0.44**			**+0.29**			+0.01	
C45	FLOR	—	—	—	—	—	—	—	—
	HiFLOR	**+0.45**			+0.18			−0.14	
ACE	FLOR	**+0.66**	**+0.67**	**+0.66**	**+0.49**	**+0.28**	−0.01	+0.12	−0.04
	HiFLOR	**+0.56**			**+0.31**			+0.10	
HI	FLOR	**+0.33**	+0.28	**+0.54**	**+0.31**	+0.19	+0.23	−0.03	+0.12
	HiFLOR	**+0.43**			+0.27			+0.12	
NAT									
TC	FLOR	**+0.69**	**+0.66**	**+0.38**	**+0.28**	+0.23	+0.24	+0.15	+0.10
	HiFLOR	**+0.63**			**+0.46**			+0.26	
HUR	FLOR	**+0.65**	**+0.59**	**+0.32**	**+0.37**	+0.19	+0.20	+0.08	+0.17
	HiFLOR	**+0.69**			**+0.41**			+0.16	
C345	FLOR	—	—	—	—	—	—	—	—
	HiFLOR	**+0.72**			**+0.50**			**+0.28**	
C45	FLOR	—	—	—	—	—	—	—	—
	HiFLOR	**+0.69**			**+0.46**			**+0.46**	
ACE	FLOR	**+0.69**	**+0.70**	**+0.46**	**+0.37**	+0.28	**+0.29**	+0.27	+0.24
	HiFLOR	**+0.82**			**+0.50**			**+0.37**	
US	FLOR	**+0.52**	**+0.56**	**+0.48**	**+0.39**	+0.18	+0.26	**+0.40**	**+0.33**
	HiFLOR	**+0.43**			**+0.63**			+0.17	
CAR	FLOR	**+0.49**	**+0.48**	**+0.29**	+0.21	+0.08	+0.17	+0.12	+0.08
	HiFLOR	**+0.62**			**+0.29**			+0.21	

(Fig. 2). Table 1 shows the values of RCOR for more details. Here we compare basin-total frequencies of TCs, HURs, C345s, and C45s in addition to the basin-total values of ACE and PDI. Note that we show the values of the correlation coefficient for C345 and C45 for HiFLOR along the *y* axis for convenience because FLOR cannot simulate these hurricanes owing to its low resolution. As expected, the shortest lead-month forecasts (L0) yield higher correlations, smaller NRMSEs, and higher MSSSs than the longer lead months [lead month 3 (L3) or L6] for most of the variables. In addition, both models show higher correlations in the NAT than in the other two ocean basins. Most of the RCORs forecasted from July initial conditions are statistically significant at more than the 90% level for both HiFLOR and FLOR, and some of them are even significant for the forecasts initialized in April (Table 1). Overall, HiFLOR shows skill comparable (or higher for some variables) relative to FLOR in both the WNP and ENP. On the other hand, HiFLOR shows higher correlations with respect to the observations in the NAT than FLOR for most of the variables. We also evaluated the NRMSE (Figs. 2d–f) and MSSS (Figs. 2g–i), resulting in the same conclusions as in the RCOR. We also preliminarily

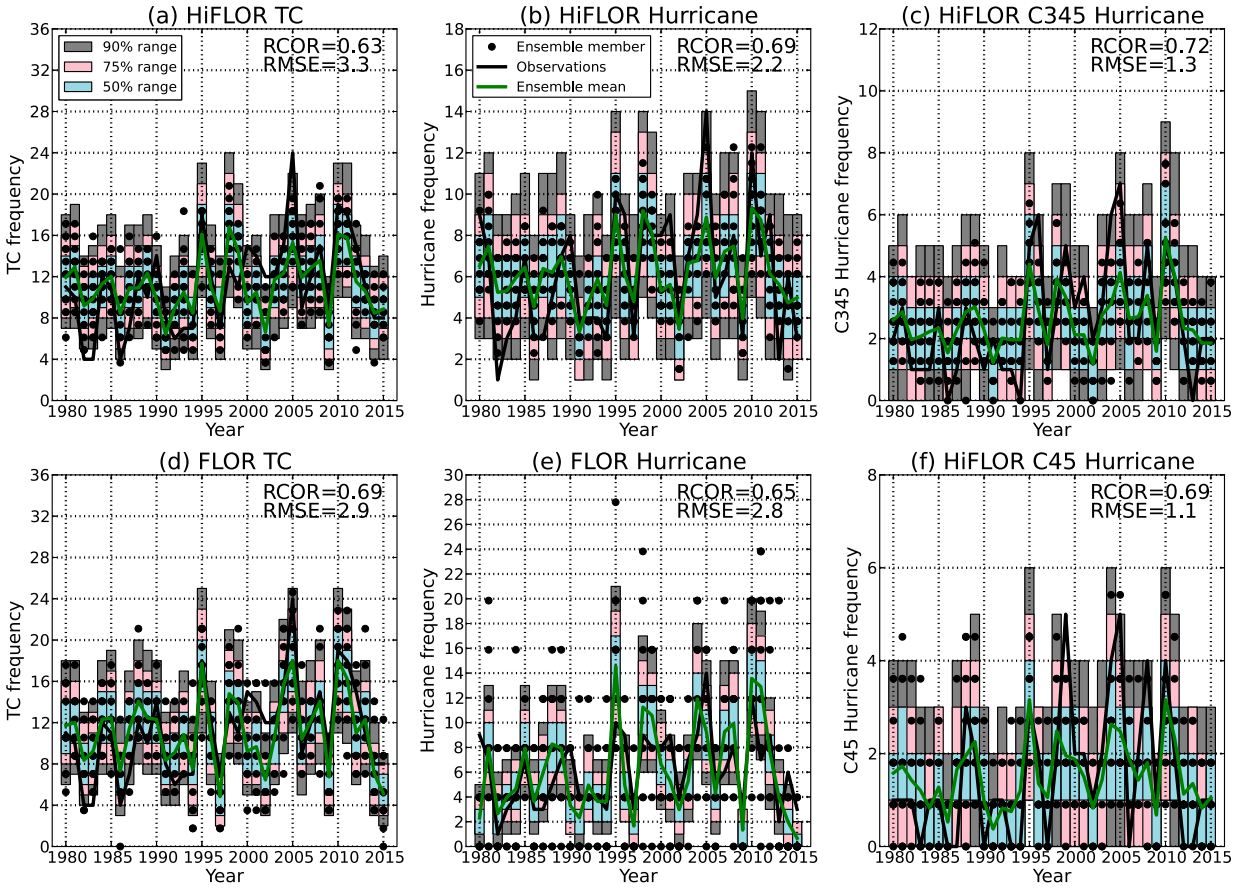

FIG. 3. Frequency of (a) basin-total TC, (b) HUR, (c) C345, and (f) C45 in the NAT during July–November 1980–2015 for the retrospective forecasts initialized in July using HiFLOR. (d),(e) As in (a),(b), but for the retrospective forecasts using FLOR. The black lines refer to the observed quantities, the green lines refer to the mean forecast value, and shading indicates the confidence intervals computed by convolving interensemble spread based on the Poisson distribution. The black dot indicates the forecast value from each ensemble member. The values of RCOR and RMSE in each panel indicate the rank correlation coefficient and root-mean-square error between the black and green lines, respectively.

evaluated the prediction skill for the shorter period 1990–2015, yielding the same conclusions (figure not shown). For the shorter period, HiFLOR shows more clearly the higher skill in predicting the variables over the NAT than FLOR does.

Figure 3 shows time series of the frequency of TC, HUR, C345, and C45 in the NAT from the observations and HiFLOR/FLOR forecasts. Both models achieved high correlation coefficients (0.63–0.69) between observations and simulations initialized from July (i.e., L0) for NAT TCs and hurricanes. This skill is comparable to previous studies in which equivalent or higher correlations have been already reported using a dynamical model for the basin-total frequency of TCs and hurricanes (e.g., Chen and Lin 2011). On the other hand, HiFLOR yielded a high correlation coefficient value of 0.72 (0.69) for C345 (C45) hurricanes in the NAT (Figs. 3c,f and Table 1). HiFLOR also yielded statistically significant correlations for C345 and C45 hurricanes even for lead-month-3 and lead-month-6 forecasts in the NAT (Table 1), highlighting that skillful forecasts of C345 and C45 hurricanes are feasible at least six months in advance.

Figure 4 shows the interannual variation of observed and predicted ACE and PDI in the NAT. HiFLOR exhibits high correlation coefficients of 0.82 and 0.79 for ACE and PDI, respectively. This indicates that the GFDL dynamical model has significant skill predicting basin-total TC activity in the NAT.

As discussed before, HiFLOR generally shows comparable skill for most of the scores relative to FLOR in the WNP and ENP, although there is some dependence on the scores of interest. For the ENP, HiFLOR shows higher skill in predicting TCs for all initial forecasts except for the July initial forecast (Table 1). Moreover, HiFLOR shows statistically significant correlations with observations for C345 and C45 hurricanes in both the

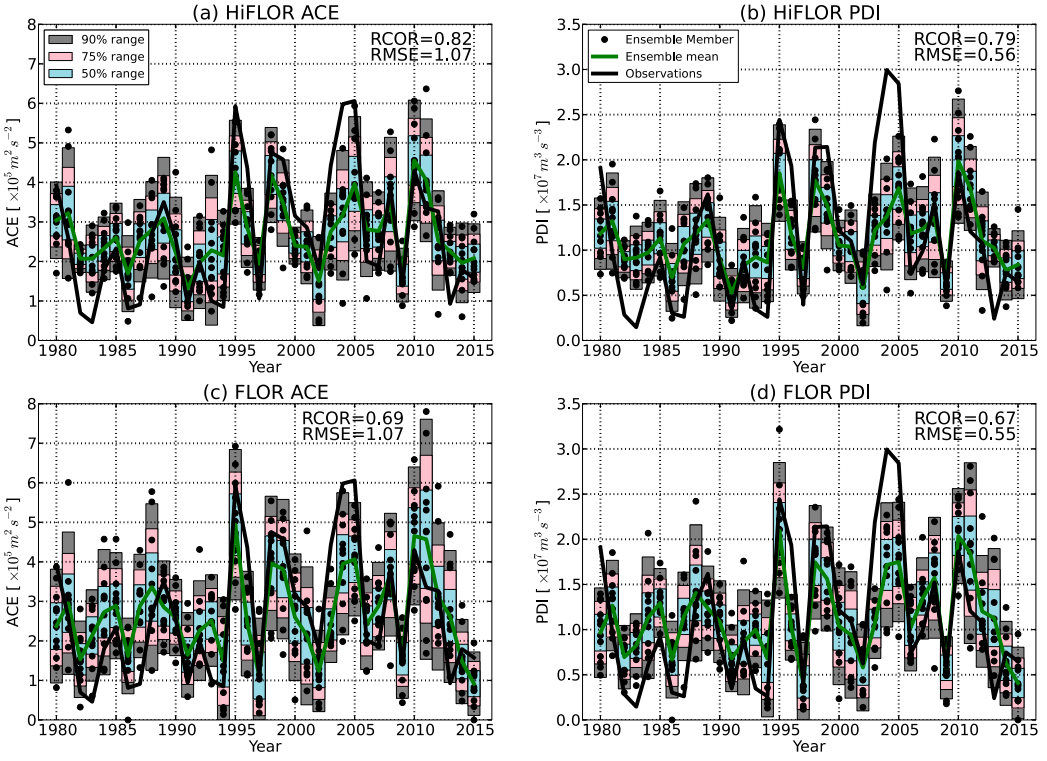

FIG. 4. (a) Basin-total ACE and (b) PDI in the NAT during July–November 1980–2015 for the retrospective forecasts initialized in July using HiFLOR. (c),(d) As in (a),(b), but for the retrospective forecasts using FLOR. The black lines refer to the observed quantities, the green lines refer to the mean forecast value, and shading indicates the confidence intervals computed by convolving interensemble spread based on the normal distribution. The values of RCOR and RMSE in each panel indicate the rank correlation coefficient and root-mean-square error between the black and green lines, respectively.

WNP and ENP for the July initialized prediction (Table 1). These results highlight the practical use of HiFLOR to predict the most intense hurricanes not only for the NAT but also for the WNP and ENP.

b. *Retrospective forecasts of regional TC activity*

Predictions of regional TC activity are also investigated (Fig. 5). Both FLOR and HiFLOR show significant skill in predicting TCs over the tropical NAT and CPO (Figs. 5a,d), indicating potential predictability for landfalling TCs over the Caribbean islands, the Gulf of Mexico coast, and the Hawaiian Islands. For hurricanes (Figs. 5b,e), both models show significant skill for the above-mentioned regions, although the regions showing skillful predictions are slightly smaller than those obtained for the prediction of TCs. In addition, HiFLOR shows skill in predicting TC and hurricane frequency of occurrence over the coastline of Japan, Guam, the Hawaiian Islands, and the eastern coast of the United States. Moreover, HiFLOR shows some skill in predicting C345 and C45 hurricanes in the Caribbean Sea, tropical central Atlantic, tropical eastern Pacific, and tropical western Pacific, whereas FLOR cannot simulate/predict C345 and C45 hurricanes owing to the low horizontal resolution. These results highlight potential use of HiFLOR (or FLOR) to predict regional TC activity, especially for the most intense TC activity, before the summer season.

Figure 6 shows observed and predicted landfall TCs in the United States, Caribbean islands, and Hawaiian Islands for HiFLOR and FLOR. Here, we define landfall TCs as those storms propagating within a 300-km buffer zone from the coastline (see blue domains in Fig. 6). We investigated the dependence of the skill scores on the width of the buffer zone and found only a small variation over the range (0–500 km), even though the skill is the highest for the 300-km buffer. Both models show marked skill in predicting landfall TCs for these regions (correlation coefficients of 0.3–0.6) for the lead-month-0 prediction. Even for the lead-month-3 predictions, HiFLOR shows skill (statistically significant correlations of about 0.3–0.6) in predicting landfall TCs over the United States and Caribbean islands (Table 1). Overall, these results are very encouraging and provide

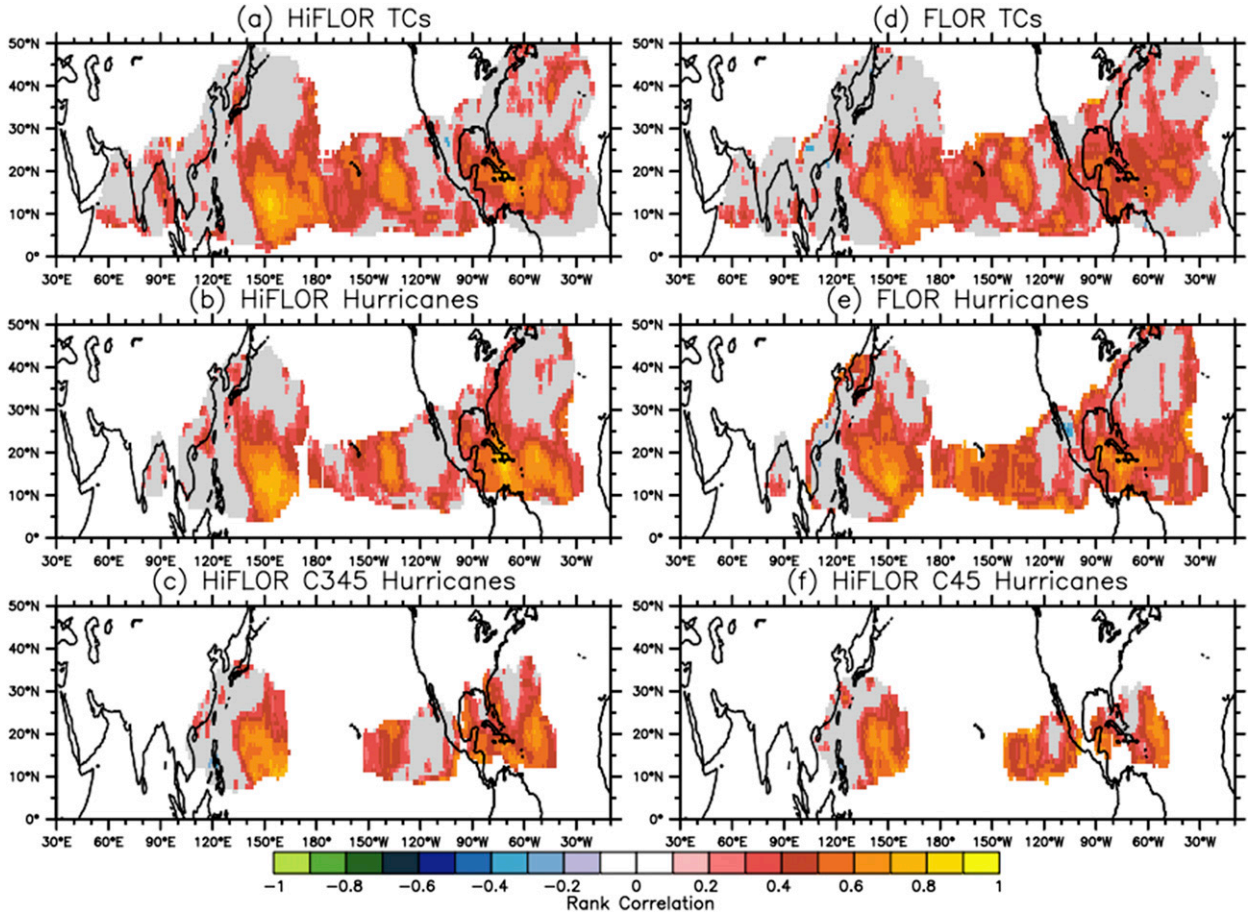

FIG. 5. Skill of frequency of occurrence during July–November 1980–2015 for the retrospective forecasts initialized in July. Shading indicates the retrospective rank correlation of predicted vs observed TC frequency of occurrence (1° × 1° grid box), masked at a two-sided $p = 0.1$ level. Results are shown for (a) TCs, (b) HUR, (c) C345, and (f) C45, for HiFLOR. (d),(e) As in (a),(b), but for FLOR. Note that the results for C345 and C45 for FLOR are not shown owing to its inability to simulate C345 and C45 hurricanes. Gray shading in all panels indicates that observed TC density is nonzero for at least 25% of years (i.e., 9 years).

empirical evidence to support the use of dynamical models for prediction of regional TC activity as well as basin-total TC activity.

c. *Retrospective forecast of large-scale parameters*

Through multidecadal SST-forcing experiments, Vecchi et al. (2014), Jia et al. (2015), Krishnamurthy et al. (2016), and M15 reported that the high-resolution model improves the simulation of large-scale parameters relative to the low-resolution model, leading to improved predictions of TC activity in the high-resolution model. As shown in Figs. 2c, 2f, and 2i, HiFLOR yields higher skill than FLOR in predicting various NAT TC metrics, although the predictability of TC activity in the WNP and ENP are mostly comparable between the models. To examine whether the higher skill in the NAT by HiFLOR is obtained by the higher skill in predicting large-scale parameters, we compare the forecast skill in FLOR and HiFLOR in predicting large-scale parameters. Here we consider four parameters (see section 2c), which appear to be highly correlated to the observed TC frequency for the NAT (correlation map and selected domain are shown in Fig. 1).

Figure 7a compares the CORs between the observed and predicted large-scale parameters in the key domains by FLOR (x axis) and between observed and predicted by HiFLOR (y axis). Most points are located around the diagonal line, indicating similar skill between HiFLOR and FLOR. HiFLOR even shows lower skill than FLOR for some variables (e.g., SSTA; Fig. 7a). Similar results are obtained using NRMSE (Fig. 7b) and MSSS (Fig. 7c). In contrast to the previous studies of SST-forcing experiments by Jia et al. (2015) and M15, these results indicate that the improvements in predicting TC activity in the NAT by HiFLOR relative to FLOR is not directly related to the

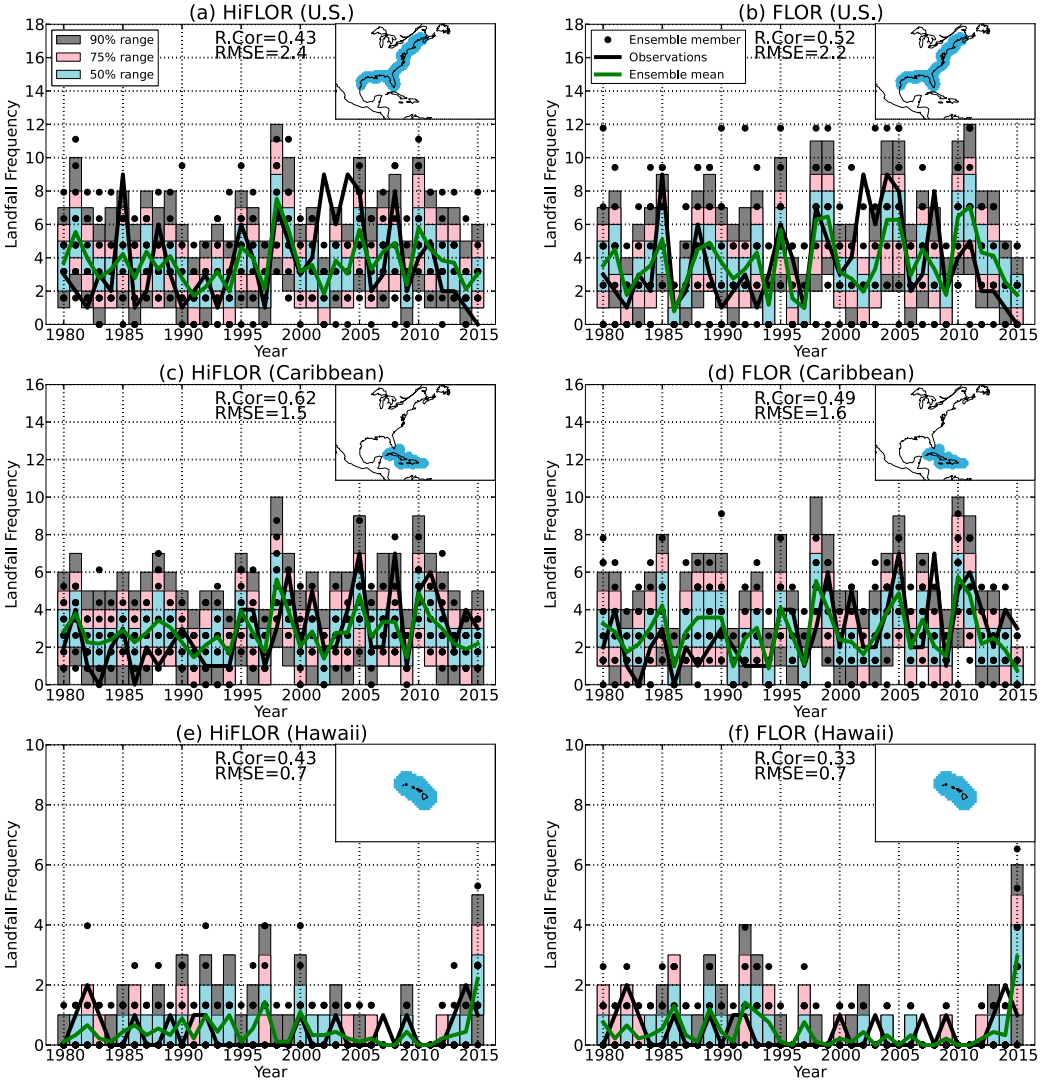

FIG. 6. As in Fig. 3, but for landfalling TC frequency for the (a),(b) United States; (c),(d) Caribbean islands; and (e),(f) Hawaiian Islands, for (left) HiFLOR and (right) FLOR.

improvements in prediction of the large-scale parameters. We hypothesize that the difference in the prediction skill in TC activity between HIFLOR and FLOR in the NAT may be due to the difference in the simulation of TCs themselves and the response of TC climatology to the same large-scale conditions. On the other hand, we also compare the prediction skill between FLOR and GFDL CM2.1 (Figs. 7d–f). FLOR shows higher skill in predicting most of the large-scale parameters than GFDL CM2.1. The difference between GFDL CM2.1 and FLOR is mainly the horizontal resolution in the atmospheric component (i.e., GFDL CM2.1: 250 km; FLOR: 60 km). It is not clear why HiFLOR has comparable skill in simulating large-scale parameters to FLOR, whereas FLOR has better skill than GFDL CM2.1 in simulating large-scale parameters, although the model differences are mainly horizontal resolution in atmospheric component for those cases. Further investigations are needed to address this question.

4. Summary

In this study, we have evaluated the retrospective seasonal forecasts of TC activity during the boreal summer (July–November) for the period 1980–2015 by the GFDL high-resolution coupled climate model (HiFLOR) and compared this to the skill in the moderate-resolution version of FLOR. HiFLOR yielded skill comparable to or higher than FLOR in predicting TC

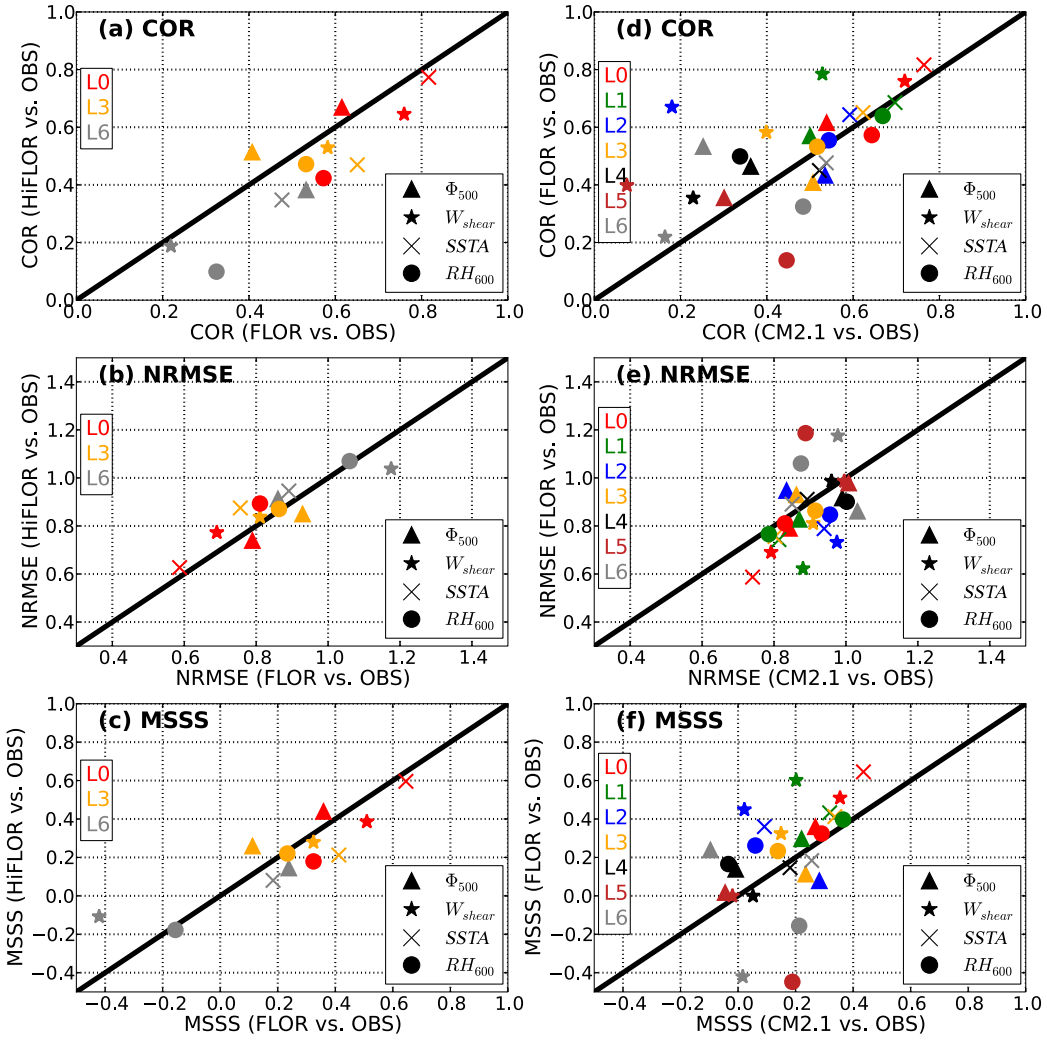

FIG. 7. (a) Scatterplot of COR between HiFLOR prediction and observations (y axis) and FLOR prediction and observations (x axis). A COR above the diagonal lines indicates that HiFLOR shows higher correlation than FLOR. (b),(c) As in (a), but for NRMSE and MSSS, respectively. An NRMSE (MSSS) below (above) the diagonal lines indicates that HiFLOR shows higher skill than FLOR. Variables evaluated are Φ_{500}, W_{shear}, SSTA, and RH_{600}. Different colors indicate different lead months. (d)–(f) As in (a)–(c), but for comparisons between FLOR (y axis) and GFDL CM2.1 (x axis).

activity in the NAT and comparable skill in the WNP and ENP. Both models show high correlation coefficients (0.63–0.69) between observed and simulated TC activity initialized from July (i.e., lead month 0) in the NAT. Moreover, HiFLOR obtained a high correlation coefficient of 0.72 (0.69) for C345 (C45) hurricanes; this is the first time that a dynamical model shows such a high correlation for the most intense hurricanes through seasonal forecasting. Even the lead-month-3 and lead-month-6 forecasts show statistically significant skill in predicting C345 and C45 hurricanes in the NAT. HiFLOR also showed high correlation coefficients (0.82 and 0.79) in predicting ACE and PDI in the NAT. These encouraging results indicate that the GFDL's dynamical model has significant skill in forecasting basin-total TC activity a few months in advance. The skill in predicting key large-scale parameters for NAT TC genesis using FLOR and HiFLOR are comparable, indicating that the improved skill of predicting TC activity in the NAT by HiFLOR relative to FLOR is obtained not by improving the predictions of large-scale conditions but likely by improving predictions of the relationship of TCs to large-scale climate.

We also examined the predictability of regional TC activity. HiFLOR and FLOR show significant skill in predicting TCs over the tropical NAT, the tropical

CPO, and the tropical WNP. Both models show marked skill in predicting landfall TCs for the U.S. coastal regions, Caribbean islands, and Hawaiian Islands (correlation coefficients of 0.3–0.6 for July initialized forecasts). Moreover, HiFLOR shows some skill in predicting C345 and C45 hurricanes in the Caribbean Sea, tropical ENP, and WNP, while FLOR cannot simulate/predict C345 and C45 hurricanes owing to the low horizontal resolution. These results highlight the potential use of HiFLOR to predict regional TC activity, especially high-intensity storms, before the onset of the summer season.

Acknowledgments. This material is based in part upon work supported by the National Science Foundation under Grant AGS-1262099. This report was prepared by Hiroyuki Murakami under Award NA14OAR4830101 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Gabriele Villarini also acknowledges financial support from the U.S. Army Corps of Engineers Institute for Water Resources and Award NA14OAR4830101 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration, the U.S. Department of Commerce, or the U.S. Army Corps of Engineers. We thank Dr. Baoqiang Xiang and Dr. Liping Zhang for useful comments and suggestions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Hiroyuki Murakami.

REFERENCES

Alessandri, A., A. Borrelli, S. Gualdi, E. Scoccimarro, and S. Masina, 2011: Tropical cyclone count forecasting using a dynamical seasonal prediction system: Sensitivity to improved ocean initialization. *J. Climate*, **24**, 2963–2982, doi:10.1175/2010JCLI3585.1.

Bell, G. D., and Coauthors, 2000: The 1999 North Atlantic and eastern North Pacific hurricane season [in ''Climate Assessment for 1999'']. *Bull. Amer. Meteor. Soc.*, **81** (5), S19–S22.

Camargo, S. J., and A. G. Barnston, 2009: Experimental seasonal dynamical forecasts of tropical cyclone activity at IRI. *Wea. Forecasting*, **24**, 472–491, doi:10.1175/2008WAF2007099.1.

——, ——, P. J. Klotzbach, and C. W. Landsea, 2007: Seasonal tropical cyclone forecasts. *WMO Bull.*, **56**, 297–309.

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, doi:10.1002/qj.2516.

Chang, Y.-S., S. Zhang, A. Rosati, T. L. Delworth, and W. F. Stern, 2013: An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dyn.*, **40**, 775–803, doi:10.1007/s00382-012-1412-2.

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, doi:10.1029/2011GL047629.

——, and ——, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, doi:10.1175/JCLI-D-12-00061.1.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, doi:10.1175/JCLI3629.1.

——, and Coauthors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25**, 2755–2781, doi:10.1175/JCLI-D-11-00316.1.

Emanuel, K., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688, doi:10.1038/nature03906.

——, 2007: Environmental factors affecting tropical cyclone power dissipation. *J. Climate*, **20**, 5497–5509, doi:10.1175/2007JCLI1571.1.

Gnanadesikan, A., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation. *J. Climate*, **19**, 675–697, doi:10.1175/JCLI3630.1.

Gray, W. M., 1984a: Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668, doi:10.1175/1520-0493(1984)112<1649:ASHFPI>2.0.CO;2.

——, 1984b: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683, doi:10.1175/1520-0493(1984)112<1669:ASHFPI>2.0.CO;2.

Harris, L. M., S.-J. Lin, and C. Y. Tu, 2016: High resolution climate simulations using GFDL HiRAM with a stretched global grid. *J. Climate*, **29**, 4293–4314, doi:10.1175/JCLI-D-15-0389.1.

Jia, L., and Coauthors, 2015: Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. *J. Climate*, **28**, 2044–2062, doi:10.1175/JCLI-D-14-00112.1.

——, and Coauthors, 2016: The roles of radiative forcing, sea surface temperatures, and atmospheric and land initial conditions in U.S. summer warming episodes. *J. Climate*, **29**, 4121–4135, doi:10.1175/JCLI-D-15-0471.1.

Kim, H.-S., C.-H. Ho, J.-H. Kim, and P.-S. Chu, 2012: Track-pattern-based model for seasonal prediction of tropical cyclone activity in the western North Pacific. *J. Climate*, **25**, 4660–4678, doi:10.1175/JCLI-D-11-00236.1.

Klotzbach, P. J., and W. M. Gray, 2009: Twenty-five years of Atlantic basin seasonal hurricane forecasts (1984–2008). *Geophys. Res. Lett.*, **36**, L09711, doi:10.1029/2009GL037580.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neuman, 2010: The international best track archive for climate stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, doi:10.1175/2009BAMS2755.1.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, doi:10.2151/jmsj.2015-001.

Krishnamurthy, L., G. A. Vecchi, R. Msadek, H. Murakami, A. Wittenberg, and F. Zeng, 2016: Impact of strong ENSO on regional tropical cyclone activity in a high-resolution climate model. *J. Climate*, **29**, 2375–2394, doi:10.1175/JCLI-D-15-0468.1.

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, doi:10.1175/MWR-D-12-00254.1.

LaRow, T. E., Y.-K. Lim, D. W. Shin, E. P. Chassgnet, and S. Cocke, 2008: Atlantic basin seasonal hurricane simulations. *J. Climate*, **21**, 3191–3206, doi:10.1175/2007JCLI2036.1.

——, L. Stefanova, D.-W. Shin, and S. Cocke, 2010: Seasonal Atlantic tropical cyclone hindcasting/forecasting using two sea surface temperature datasets. *Geophys. Res. Lett.*, **37**, L02804, doi:10.1029/2009GL041459.

Li, X., S. Yang, H. Wang, X. Jia, and A. Kumar, 2013: A dynamical–statistical forecast model for the annual frequency of western Pacific tropical cyclones based on the NCEP Climate Forecast System version 2. *J. Geophys. Res. Atmos.*, **118**, 12 061–12 074, doi:10.1002/2013JD020708.

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, doi:10.1175/JCLI-D-15-0531.1.

Murakami, H., and Coauthors, 2015: Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *J. Climate*, **28**, 9058–9079, doi:10.1175/JCLI-D-15-0216.1.

——, 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, doi:10.1175/MWR-D-15-0308.1.

Pielke, R. A., J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin, 2008: Normalized hurricane damage in the United States: 1990–2005. *Nat. Hazards Rev.*, **9**, 29–42, doi:10.1061/(ASCE)1527-6988(2008)9:1(29).

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

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

Unisys, 2016: Unisys weather: Hurricane/tropical data. [Available online at http://weather.unisys.com/hurricane/.]

Vecchi, G. A., and G. Villarini, 2014: Next season's hurricanes. *Science*, **343**, 618–619, doi:10.1126/science.1247759.

——, and Coauthors, 2011: Statistical–dynamical predictions of seasonal North Atlantic hurricane activity. *Mon. Wea. Rev.*, **139**, 1070–1082, doi:10.1175/2010MWR3499.1.

——, and Coauthors, 2013: Multiyear predictions of North Atlantic hurricane frequency: Promise and limitations. *J. Climate*, **26**, 7994–8016, doi:10.1175/JCLI-D-12-00464.1.

——, and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, doi:10.1175/JCLI-D-14-00158.1.

Villarini, G., and G. A. Vecchi, 2013: Multiseason lead forecast of the North Atlantic power dissipation index (PDI) and accumulated cyclone energy (ACE). *J. Climate*, **26**, 3631–3643, doi:10.1175/JCLI-D-12-00448.1.

Vitart, F., 2006: Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quart. J. Roy. Meteor. Soc.*, **132**, 647–666, doi:10.1256/qj.05.65.

——, and T. N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Wea. Rev.*, **129**, 2521–2537, doi:10.1175/1520-0493(2001)129<2521:SFOTSU>2.0.CO;2.

——, and Coauthors, 2007: Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP. *Geophys. Res. Lett.*, **34**, L16815, doi:10.1029/2007GL030740.

Wang, H., J.-K. E. Schemm, A. Kumar, W. Wang, L. Long, M. Chelliah, G. D. Bell, and P. Peng, 2009: A statistical forecast model for Atlantic seasonal hurricane activity based on the NCEP dynamical seasonal forecast. *J. Climate*, **22**, 4481–4500, doi:10.1175/2009JCLI2753.1.

Wittenberg, A. T., A. Rosati, N.-C. Lau, and J. J. Ploshay, 2006: GFDL's CM2 global coupled climate models. Part III: Tropical Pacific climate and ENSO. *J. Climate*, **19**, 698–722, doi:10.1175/JCLI3631.1.

Zhang, S., and A. Rosati, 2010: An inflated ensemble filter for ocean data assimilation with a biased coupled GCM. *Mon. Wea. Rev.*, **138**, 3905–3931, doi:10.1175/2010MWR3326.1.

Zhang, W., and Coauthors, 2016: Improved simulation of tropical cyclone responses to ENSO in the western North Pacific in the high-resolution GFDL HiFLOR coupled climate model. *J. Climate*, **29**, 1391–1415, doi:10.1175/JCLI-D-15-0475.1.

Zhao, M., I. M. Held, 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**, 333–363, doi:10.1175/2009JCLI3049.1.

——, ——, and G. A. Vecchi, 2010: Forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Mon. Wea. Rev.*, **138**, 3858–3868, doi:10.1175/2010MWR3366.1.