Seasonal Forecasts of Category 4 and 5 Hurricanes and Landfalling Tropical Cyclones using a High-Resolution GFDL Coupled Climate Model

Hiroyuki Murakami[1, 2], Gabriel A. Vecchi[1, 2], Gabriele Villarini[3], Thomas L. Delworth[1, 2], Richard Gudgel[1], Seth Underwood[4], Xiaosong Yang[1,5], and Wei Zhang[1, 2], and Shian-Jiann Lin[1]

[1]National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

[2]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA

[3]IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA, USA

[4]Engility, NOAA/GFDL, Princeton, NJ, USA

[5]University Corporation for Atmospheric Research, Boulder, CO, USA

Submitted to *Geophysical Research Letters* on 4th March 2016

Key Points:
1. The new global model successfully predicts intense hurricanes a few months in advance.
2. The new global model shows significant skill in predicting landfalling hurricanes.
3. The skillful forecasts are obtained mainly by the increase of horizontal resolution.

Corresponding author address: Hiroyuki Murakami, NOAA/GFDL, 201 Forrestal Rd., Princeton, NJ 08540-6649
E-mail: hir.murakami@gmail.com
Tel: 609-452-5824

Abstract

Skillful seasonal forecasting of tropical cyclone (TC; wind speed ≥17.5 m s^{-1}) activity is challenging, even more so when the focus is on the most intense hurricanes (Category 4–5; wind speed ≥58.1 m s^{-1}) and landfalling TCs. Here we show that the 25-km mesh Geophysical Fluid Dynamics (GFDL) high-resolution climate model (HiFLOR) has improved skill in predicting the frequencies of Category 4–5 hurricanes in the North Atlantic and landfalling TCs over the United States and Caribbean Islands a few months in advance, relative to a more moderate resolution of 50-km mesh GFDL climate model (FLOR). HiFLOR also shows significant skill in predicting Category 4–5 hurricanes in the western North Pacific and eastern North Pacific, while both models show comparable skills in predicting basin-total TC frequency and landfall TC frequency in the basins.

Index Terms: 3372 Tropical cyclones, 1922 Forecasting, 3337 Global climate models

Key Words: Category 4 and 5 Hurricanes, Seasonal Forecasting, and High-Resolution Climate Model

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 intense hurricanes (Saffir-Simpson Categories 3, 4, and 5), even though they make up a very small fraction of overall TCs [*Pielke et al.* 2008]. Furthermore, even though non-landfalling TCs can cause damage (e.g., to off-shore energy platforms and ships), landfalliing TCs contribute substantially more to overall TC damages than do non-landfalling TCs. Therefore, predicting intense hurricanes and landfalling storms at seasonal time scales is a topic of large scientific and socio-economic interest [*Vecchi and Villarini* 2014]. Recent studies have reported that state-of-the-art dynamical models successfully predicted basin-total frequency of tropical storms and hurricanes a few months in advance [*Zhao et al.* 2010; *Chen and Lin* 2011, 2013; *Vecchi et al.* 2014]. Specifically, *Chen 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 seasons in advance for regional basin-wide TC activity across the Northern Hemisphere. However, prediction of category 4 and 5 (C45) hurricanes and landfall storm frequency remains challenging [*Vecchi and Villarini* 2014; *Camp et al.* 2015; *Murakami et al.* 2016], although there are some suggestive results with high-resolution 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]. *Murakami et al.* [2015] provided a preliminarily assessment of the predictability of C45 hurricanes in the new high-resolution Geophysical Fluid Dynamics Laboratory (GFDL) coupled model (HiFLOR). They evaluated retrospective seasonal forecasts initialized on July 1st in 1997 and 1998. The predictions showed the observed sharp contrast in terms of global TC activity due to the extreme El Niño and La Niña events of 1997–98 and 1998–99, respectively. Although HiFLOR could predict the contrast in the C45 hurricanes for 1997 and 1998 summer seasons, the results reported there should not be interpreted as applying broadly to predictive skill for all years.

In this study, we conduct retrospective seasonal forecasts initialized on July, April, and January between 1990 and 2015 to evaluate and quantify the skill HiFLOR has in predicting TC activity in the ocean basins of western North Pacific (WNP), eastern North Pacific (ENP), and North Atlantic (NAT) (see fig. 3 in *Murakami et al.* [2015] for regional boundaries). We especially focus on the prediction of intense hurricanes and regional TC frequency in NAT over the study period. We show for the first time that this high-resolution global atmosphere-ocean coupled model has significant skill in predicting the frequency of C45 hurricanes and regional TC activity as well as that of basin-total TCs a few months in advance in NAT as well as other ocean basins. We further show that both the FLOR and HiFLOR models can exhibit marginal skill at predictions of seasonal U.S. landfalling TC frequency. Section 2 provides a description of the models, seasonal forecasts, and TC detection method along with observed dataset. Section 3 presents the results. Finally, a summary is given in Section 4.

2. Methods

a. Dynamical Models, Seasonal Forecasts, TC Detection Methods

The dynamical models used here are the Forecast-oriented Low Ocean Resolution of GFDL Coupled Model version 2.5 (FLOR) [*Vecchi et al.* 2014] and the high-resolution version of FLOR (HiFLOR) [*Murakami et al.* 2015]. The atmosphere and land components of FLOR are taken from the Coupled Model version 2.5 (CM2.5) [*Delworth et al.* 2012] developed at GFDL, whereas the ocean and sea ice components are based on the GFDL Coupled Model version 2.1 (CM2.1) [*Delworth et al.* 2006; *Wittenberg et al.* 2006; *Gnanadesikan et al.* 2006]. FLOR comprises 50-km mesh atmosphere and land components, and 100-km mesh sea ice and ocean components. HiFLOR was developed from FLOR by decreasing the horizontal grid spacing of the atmospheric component to 25 km, while leaving most of the sub-grid physical parameterizations unchanged [*Murakami et al.* 2015]. HiFLOR yielded better simulations of the observed El Niño-Southern Oscillation (ENSO)-TC teleconnection in WNP, ENP, and NAT than FLOR does [*Murakami et al.* 2015; *Zhang et al.* 2016].

For each year and each month in the period 1990–2015, 12-month duration retrospective seasonal predictions were performed after initializing the model to the observed conditions for ocean components [*Murakami et al.* 2016]. The 12-member initial conditions for ocean and sea ice components were taken from GFDL's ensemble coupled data assimilation system using CM2.1 [*Zhang and Rosati* 2010; *Chang et al.* 2013], whereas those for atmosphere and land components were built from a suite of sea surface temperature (SST)-forced atmosphere-land-only simulations using the components for the FLOR predictions [*Vecchi et al.* 2014], and in the HiFLOR predictions using an arbitral year from a control climate simulation [*Murakami et al.* 2015]. Therefore, the predictability in these experiments comes entirely from the ocean and sea ice, and may be thought of as a lower bound on the potential prediction skill of the model, because predictability could also arise from 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. Forecasts from other initial months by FLOR will be shown for the comparisons of prediction skill in large-scale parameters among FLOR, CM2.1, and HiFLOR. We define forecasts from July (January) initial conditions as lead-month 0 or L0 (6 or L6) forecasts. Because North 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 *Murakami et al.* [2015]. 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 1K for FLOR and 2K for HiFLOR). The detection scheme also requires that the TC lasts for 36 consecutive hours while maintaining a warm core as well as a specified wind speed criterion (15.75 m s^{-1} for FLOR and 17.5 m s^{-1} for HiFLOR).

c. Observational datasets

The observed TC "best-track" data were obtained from the National Hurricane Center Best Track Database (HURDAT2) [*Landsea and Franklin* 2013] and Joint Typhoon Warning Center (JTWC) as archived in the International Best Track Archive for Climate Stewardship (IBTrACS v03r06) [*Knapp et al.* 2010] and used to evaluate the TC simulations in the retrospective seasonal predictions. Because the best-track data were not available for 2015 when this study started, we obtained observed global TC dataset compiled on the Unisys Corporation website [*Unisys* 2016]. We only used 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 1990–2015. We also removed TC tracks that reportedly transformed to extra-tropical cyclones (i.e., TC tracks after the time of its extra-tropical transition were removed). In this study, storms are categorized into three groups according to their lifetime maximum intensity: Tropical Cyclones (TC; ≥17.5 m s^{-1}); Hurricanes (HUR; ≥32.9 m s^{-1}); and Category 4 and 5 hurricanes (C45; ≥58.1 m s^{-1}). Note that although a hurricane is called "typhoon" in WNP, we describe hurricanes in WNP to represent typhoon in this study for convenience.

In order to address differences in forecast skill in NAT between FLOR and HiFLOR, we compared prediction skill in 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 (*SST*; 0–20°N, 10°–70°W), and relative humidity at 600 hPa over the tropical NAT (RH_{600}; 10°–20°N, 10°–90°W). *Murakami et al.* [2016] discussed the reason why Φ_{500} over the subtropical ENP shows a high correlation with TC frequency in NAT. When the anomaly of Φ_{500} is positive in the subtropical ENP, 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 through the so-called Pacific/North American (PNA) pattern. We use the UK Met Office Hadley Centre SST product (HadISST1.1) [*Rayner et al.* 2003] and the Japanese 55-year Reanalysis (JRA-55) [*Kobayashi et al.* 2015] for the period 1990–2015 as observed SST and atmospheric large-scale parameters, respectively.

d. Metrics for Evaluation of Forecast Skill

In this study, we used three metrics in order to evaluate prediction skill for interannual variation of TC activity relative to observed values: rank correlation coefficient, normalized root-mean-square error (NRMSE), and mean square skill score (MSSS) [*Kim et al.* 2012; *Li et al.* 2013]. NRMSE is defined as:

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

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

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

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].

3. Results

a. Retrospective forecast of basin-total TC activity

We first compare the retrospective forecast skill in basin-wide seasonal TC activity between FLOR and HiFLOR (Fig. 1) using scatterplots of rank correlation for interannual variation of seasonal mean value between observations and FLOR versus HiFLOR (see Supplemental Information Table S1 for more details). Here we compare basin-total frequencies of TCs, HURs, and C45s in addition to the basin-total values of accumulated cyclone energy (ACE) and power dissipation index (PDI). Note that we only show the values of the correlation coefficient for C45 for HiFLOR only because FLOR cannot simulate C45 hurricanes due to its low resolution. As expected, the shortest lead-month forecasts (L0) yield higher correlations than the longer lead months (L3 or L6) for most of the variables. In addition, both models show higher correlations in NAT than in the other two ocean basins. Most of the metrics forecasted from July initial conditions are statistically significant at more than 90% level for both HiFLOR and FLOR, and some of them are even significant for the forecasts initialized in April (Supporting Information Table S1). Overall, HiFLOR shows skill comparable to FLOR in both the WNP and ENP. On the other hand, compared to FLOR, HiFLOR shows higher correlations with observations in NAT than FLOR for most of variables. We also compared the NRMSE (Supporting Information Figure S1 and Table S2) and MSSS (Supporting Information Figure S2 and Table S3) [*Kim et al.* 2012; *Li et al.* 2013], resulting in the same conclusions as in the rank correlation.

Figure 2 shows time-series of the frequency of TCs, HUR, and C45 in NAT from the observations and HiFLOR/FLOR forecasts. Both models achieved high correlation coefficients (0.69–0.75) between observed 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 0.71 for C45 hurricanes in NAT. HiFLOR also yielded statistically significant correlations for C45 hurricanes even for lead-month 3 and 6 forecasts in NAT, highlighting that skillful forecasts of C45 hurricanes are feasible at least two seasons in advance (Supporting Information Table S1). Supporting Information Figure S3 shows the interannual variation of observed and predicted Accumulated Cyclone Energy (ACE) and Power Dissipation Index (PDI) in NAT. HiFLOR exhibits high correlation coefficients of 0.82 and 0.80 for ACE and PDI, respectively. This indicates that the GFDL dynamical model has significant skill predicting basin-total TC activity in NAT. In general, HiFLOR shows comparable skill for most of metrics relative to FLOR in WNP and ENP, although it depends on metrics. Specifically, for WNP, HiFLOR shows higher skill in predicting TCs and ACE for July initial forecasts than FLOR. For ENP, HiFLOR shows higher skill in predicting TCs for all initial forecast. Moreover, HiFLOR shows statistically significant correlations with observations for C45 hurricanes in both WNP and ENP for the lead-month 0 prediction. These results highlight practical use of HiFLOR to predict the most intense hurricanes of C45.

b. Retrospective forecast of regional TC activity

Predictions of regional TC activity are also investigated (Fig. 3). Both FLOR and HiFLOR show significant skill in predicting TCs over the tropical NAT and central Pacific (Figs. 3a,b), indicating potential predictability for landfalling TCs over the Caribbean Islands, the coastal Gulf of Mexico, and Hawaiian Islands. For hurricanes (Figs. 3c,d), both models show significant skill for the abovementioned regions, although the regions showing skillful predictions are smaller than the prediction for TCs. In addition, HiFLOR shows skill in predicting TC and hurricane frequency of occurrence over the coastline of the Bay of Bengal, Japan, Guam, Hawaiian Islands, and the eastern coast of the United States. Moreover, HiFLOR shows some skill in predicting C45 hurricanes in the Caribbean Sea, tropical central Atlantic, tropical eastern Pacific, and western Pacific, whereas FLOR cannot simulate/predict

C45 hurricanes due to the low horizontal resolution. These results highlight potential use of HiFLOR (or FLOR) to predict regional TC activity, especially for intense TC activity, before the summer season.

Figure 4 shows observed and predicted landfall TCs on 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.4). 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 300 km buffer. Both models show marked skill in predicting landfall TCs for these regions (correlation coefficient of 0.3–0.7, see Fig.1 and Supporting Information Table S1) for lead-month 0 prediction. Even for the lead-month 3 predictions, HiFLOR shows skill (statistically significant correlations of 0.5–0.6) in predicting landfall TCs over these regions (Fig. 1 and Supporting Information Table S1). Generally, HiFLOR shows higher skill than FLOR in predicting TC activity for these landfall regions (Fig.1, Supporting Information Figures S1 and S2). Overall, these results are very encouraging and provide 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

Jia et al. [2015] and *Murakami et al.* [2015] reported through multi-decadal SST-forcing experiments 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 Fig. 1c, HiFLOR yields higher skill than

FLOR in predicting various NAT TC metrics, although the predictability of TC activity in WNP and ENP are comparable between the models. To examine whether the higher skill in 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 NAT (correlation map and selected domain are shown in Supporting Information Figure S4).

Figure 5a compares the correlation coefficients 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 of points are located around the diagonal line, indicating similar skill between HiFLOR and FLOR. Similar results are obtained using NRMSE (Fig. 5b) and MSSS (Fig. 5c). In contrast to the previous studies of SST forcing experiments by *Jia et al.* [2015] and *Murakami et al.* [2015], these results indicate that the improvements in prediction of TC activity in NAT by HiFLOR relative to FLOR is not directly related to the improvements in prediction of the large-scale parameters. We hypothesize that the difference in prediction skill in TC activity in NAT between HIFLOR and FLOR 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 CM2.1 (Fig. 5, panels d–f). FLOR shows higher skill in predicting most of the large-scale parameters than CM2.1 (especially with respect to NRMSE and MSSS). The difference between CM2.1 and FLOR is mainly the horizontal resolution in the atmospheric component (i.e., 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 CM2.1

in simulating large-scale parameters although the model differences are mainly horizontal resolution in atmospheric component for those cases. Further study is 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 1990–2015 by the GFDL high-resolution coupled climate model (HiFLOR) and compared this to skill in the moderate resolution version of FLOR. HiFLOR yielded comparable or higher skill to FLOR in predicting TC activity in NAT and comparable skill in WNP and ENP. Both models show high correlation coefficients (0.69–0.75) between observed and simulated TC activity initialized from July (i.e., lead-month 0) in ENP and NAT. Moreover, HiFLOR obtained a high correlation coefficient of 0.71 for C45 hurricanes: this is the first time that a dynamical model shows such a high correlation for the extremely intense hurricanes through seasonal forecasts. Even the lead-month 3 and 6 forecasts show statistically significant skill in predicting C45 hurricanes in NAT. HiFLOR also showed high correlation coefficients (0.82 and 0.80) in predicting ACE and PDI in 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 forecast skill in predicting key large-scale parameters for TC genesis in NAT using FLOR and HiFLOR is compared. The results show comparable skill between them, suggesting that the improved skill of predicting TC activity in NAT by HiFLOR relative to FLOR are not obtained by improving the large-scale parameters.

We also examined the predictability of regional TC activity. HiFLOR and FLOR show significant skill in predicting TCs over the tropical NAT and the central Pacific. Both models show marked skill in predicting landfall TCs for the U.S. coastal regions, Caribbean and Hawaiian Islands (correlation coefficients of 0.3–0.7 for July initialized forecasts). Moreover, HiFLOR shows some skill in predicting C45 hurricanes in the Caribbean Sea, tropical eastern Pacific and western Pacific, while FLOR cannot simulate/predict C45 hurricanes due to the low horizontal resolution. These results highlight 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, the Award number NA14OAR4830101 to the Trustees of Princeton University (Gabriele Villarini). 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 USACE Institute for Water Resources. 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. The authors 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

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,** 1975–2470.

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 Dynamics, 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.

Chen, J.-H. and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate,* **26,** 380–398.

Delworth, T. L., and co-authors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19,** 643–674.

Delworth, T. L., and co-authors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25,** 2755–2781.

Gnanadesikan, A., and co-authors, 2006: GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation. *J. Climate*, **19,** 675–697.

Jia, L., and co-authors, 2015: Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. *J. Climate*, **28,** 2044–2062.

Jia, L., G.A. Vecchi, X. Yang, R. Gudgel, T. Delworth, W. Stern, K. Paffendorf, S. Underwood, and F. Zeng, 2016: The roles of radiative forcing, sea surface temperatures, and atmospheric and land initial conditions in U.S. summer warming episodes. *J. Climate*, submitted.

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.

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.

Kobayashi, S., and co-authors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48.

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.

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*., **118**, 12061–12074.

Murakami, H., G. A. Vecchi, S. Underwood, T. L. Delworth, A. T. Wittenberg, W. G. Anderson, J. -H. Chen, R. G. Gudgel, L. Harris, S. -J. Lin, and F. Zeng, 2015: Simulation and prediction of Category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *J. Climate*, **28**, 9058-9079.

Murakami, H., 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.*, in press.

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, *Natural Hazards Review,* **9(1),** 29–42.

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

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.

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(6171),** 618–619.

Vecchi, G. A., and co-authors, 2011: Statistical-dynamical predictions of seasonal North Atlantic hurricane activity. *Mon. Wea. Rev.*, **139,** 1070–1082.

Vecchi, G. A., and co-authors, 2013: Multiyear predictions of North Atlantic hurricane frequency: Promise and limitations. *J. Climate*, **26,** 7994–8016.

Vecchi, G. A., and co-authors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27,** 7994–8016.

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.

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.

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.

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.

Zhang, W., and co-authors, 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*, in press.

Zhao, M., I. M. Held, 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.

List of Figures

FIG. 1 Scatterplot of rank correlation coefficient between HiFLOR prediction and observations (*y* axis) vs FLOR prediction and observations (*x* axis) for (a) WNP, (b) ENP, and (c) NAT. Variables evaluated are basin-total frequency of TCs (TC), Hurricanes (HUR), categories 4 and 5 hurricanes (C45), basin-total values of accumulated cyclone energy (ACE), power dissipation index (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 C45 hurricanes, C45 plots for HiFLOR are located along the *y*-axis for convenience. A correlation coefficient above the diagonal lines indicates that HiFLOR shows higher correlation than FLOR.

FIG. 2 Retrospective forecasts of (a) basin-total TC frequency, (b) Hurricane frequency, and (c) C45 frequency in the NAT during the peak season of July–November for the period 1990–2015 for the retrospective forecasts initialized in July using HiFLOR. (d)–(f) As in (a)–(c), but for the retrospective forecasts using FLOR. The black lines refer to the observed quantities, the green lines to the mean forecast value, and shading indicates the confidence intervals computed by convolving inter-ensemble spread based on the Poisson distribution. The black dot indicates the forecast value from each ensemble member. The values of "R.Cor" and "RMSE" in each panel indicate the rank correlation coefficient and root-mean-square error between the black and green lines, respectively.

FIG. 3 Retrospective forecast skill of TC frequency of occurrence during July–November for the period 1990–2015 initialized in July. Shading indicates the retrospective rank correlation of predicted versus observed TC frequency of occurrence (1°×1° grid box), masked at a two-sided p=0.1 level. Results are shown for (top) TCs, (middle) HUR, and (bottom) C45, for (left) HiFLOR and (right) FLOR. Note that the results for C45 for FLOR are not shown due to its inability to simulate C45. Gray shading in all panels indicates that observed TC density is nonzero for at least 25% of years (i.e., 6 years).

FIG. 4 As in Fig. 2, but for landfalling TC frequency for U.S. (a, b), Caribbean Islands (c, d), and Hawaiian Islands (e, f). The panels to the left (right) refer to HiFLOR (FLOR).

FIG. 5 (a) Scatterplot of correlation coefficient between HiFLOR prediction and observations (*y* axis) vs FLOR prediction and observations (*x* axis). A correlation coefficient above the diagonal lines indicates that HiFLOR shows higher correlations than FLOR. (b), (c) As in (a), but for NRMSE and MSSS, respectively. A NRMSE (MSSS) below (above) the diagonal lines indicates that HiFLOR shows higher skill than FLOR. Variables evaluated are geopotential height at 500 hPa in the subtropical ENP (Φ_{500}), vertical wind shear in the tropical NAT (W_{shear}), SST anomaly over the tropical NAT (*SST*), and relative humidity at 600 hPa over the tropical NAT (RH_{600}). Different colors indicate different lead months. (d–f) As in (a–c), but for comparisons between FLOR (*y* axis) and CM2.1 (*x* axis).

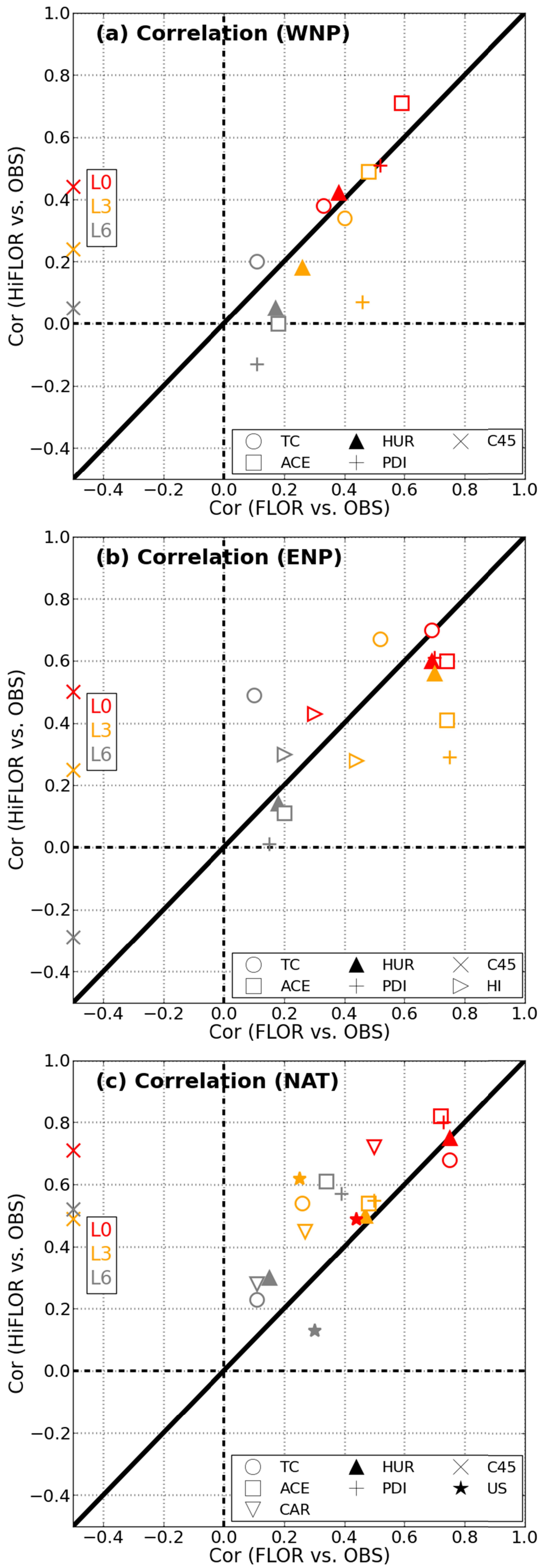
(a) Correlation (WNP)
Cor (HiFLOR vs. OBS)
Cor (FLOR vs. OBS)
L0
L3
L6
TC
ACE
HUR
PDI
C45
(b) Correlation (ENP)
Cor (HiFLOR vs. OBS)
Cor (FLOR vs. OBS)
L0
L3
L6
TC
ACE
HUR
PDI
C45
HI
(c) Correlation (NAT)
Cor (HiFLOR vs. OBS)
Cor (FLOR vs. OBS)
L0
L3
L6
TC
ACE
CAR
HUR
PDI
C45
US

FIG. 1 Scatterplot of rank correlation coefficient between HiFLOR prediction and observations (*y* axis) vs FLOR prediction and observations (*x* axis) for (a) WNP, (b) ENP, and (c) NAT. Variables evaluated are basin-total frequency of TCs (TC), Hurricanes (HUR), categories 4 and 5 hurricanes (C45), basin-total values of accumulated cyclone energy (ACE), power dissipation index (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 C45 hurricanes, C45 plots for HiFLOR are located along the *y*-axis for convenience. A correlation coefficient above the diagonal lines indicates that HiFLOR shows higher correlation than FLOR.

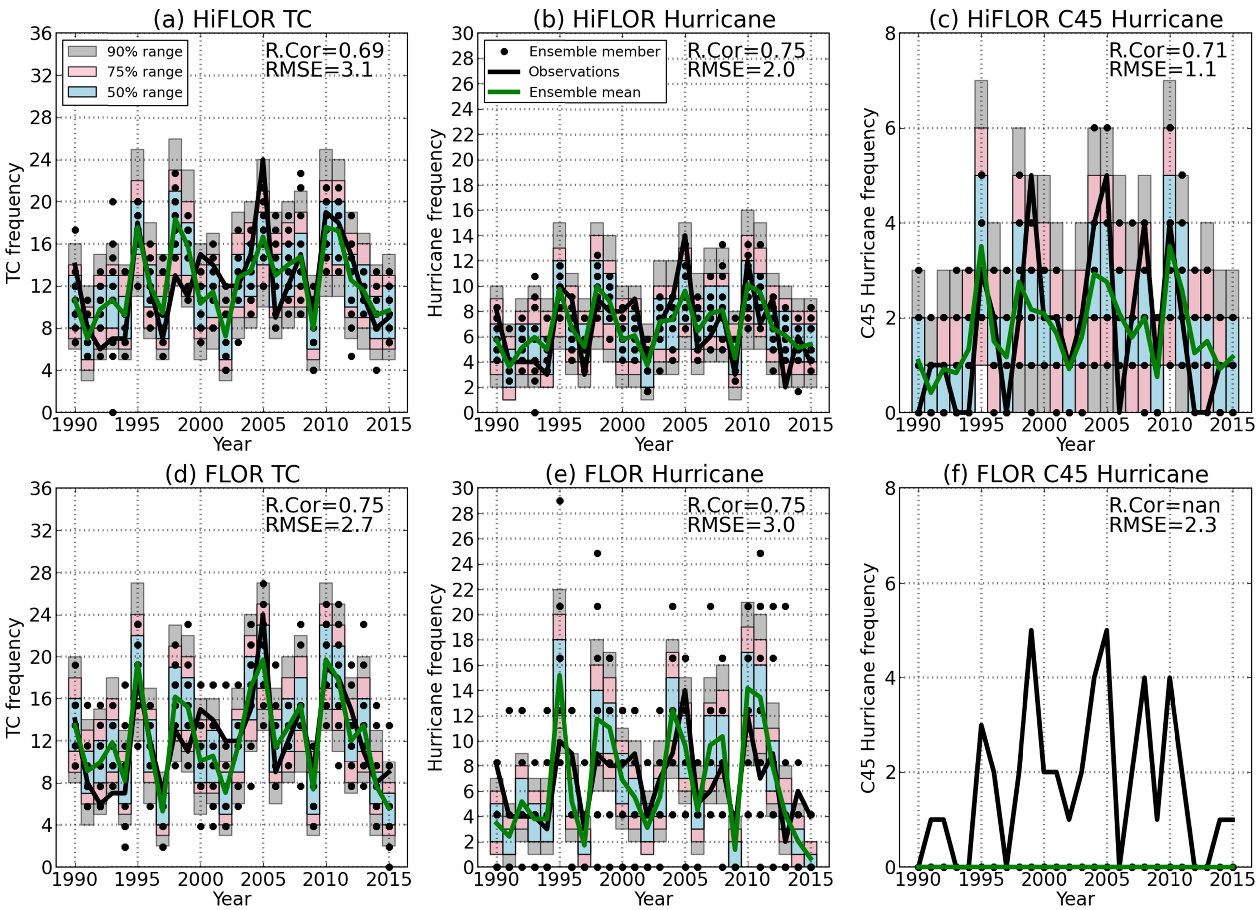

FIG. 2 Retrospective forecasts of (a) basin-total TC frequency, (b) Hurricane frequency, and (c) C45 frequency in the NAT during the peak season of July–November for the period 1990–2015 for the retrospective forecasts initialized in July using HiFLOR. (d)–(f) As in (a)–(c), but for the retrospective forecasts using FLOR. The black lines refer to the observed quantities, the green lines to the mean forecast value, and shading indicates the confidence intervals computed by convolving inter-ensemble spread based on the Poisson distribution. The black dot indicates the forecast value from each ensemble member. The values of "R.Cor" and "RMSE" in each panel indicate the rank correlation coefficient and root-mean-square error between the black and green lines, respectively.

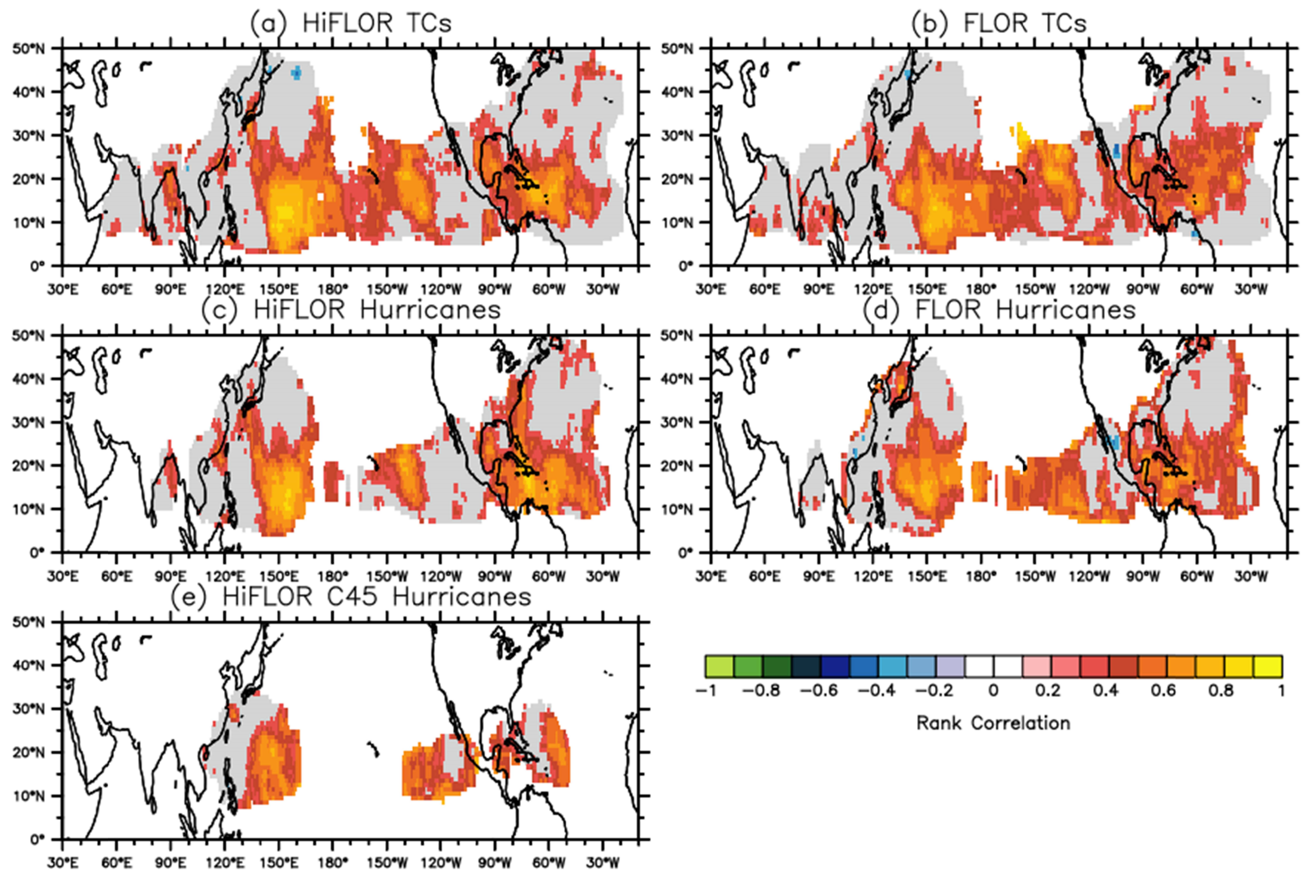

FIG. 3 Retrospective forecast skill of TC frequency of occurrence during July–November for the period 1990–2015 initialized in July. Shading indicates the retrospective rank correlation of predicted versus observed TC frequency of occurrence (1°×1° grid box), masked at a two-sided p=0.1 level. Results are shown for (top) TCs, (middle) HUR, and (bottom) C45, for (left) HiFLOR and (right) FLOR. Note that the results for C45 for FLOR are not shown due to its inability to simulate C45. Gray shading in all panels indicates that observed TC density is nonzero for at least 25% of years (i.e., 6 years).

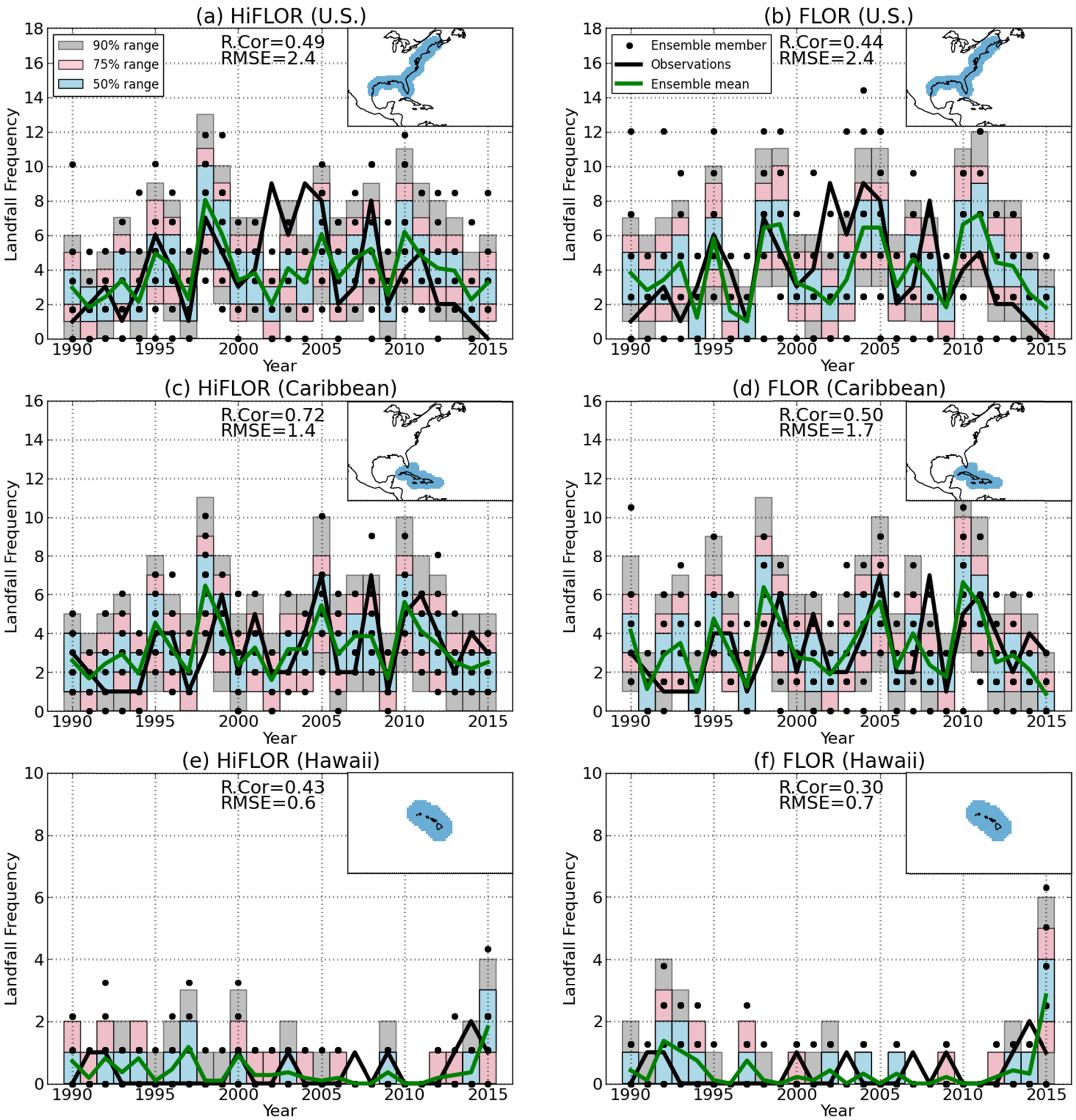

FIG. 4 As in Fig. 2, but for landfalling TC frequency for U.S. (a, b), Caribbean Islands (c, d), and Hawaiian Islands (e, f). The panels to the left (right) refer to HiFLOR (FLOR).

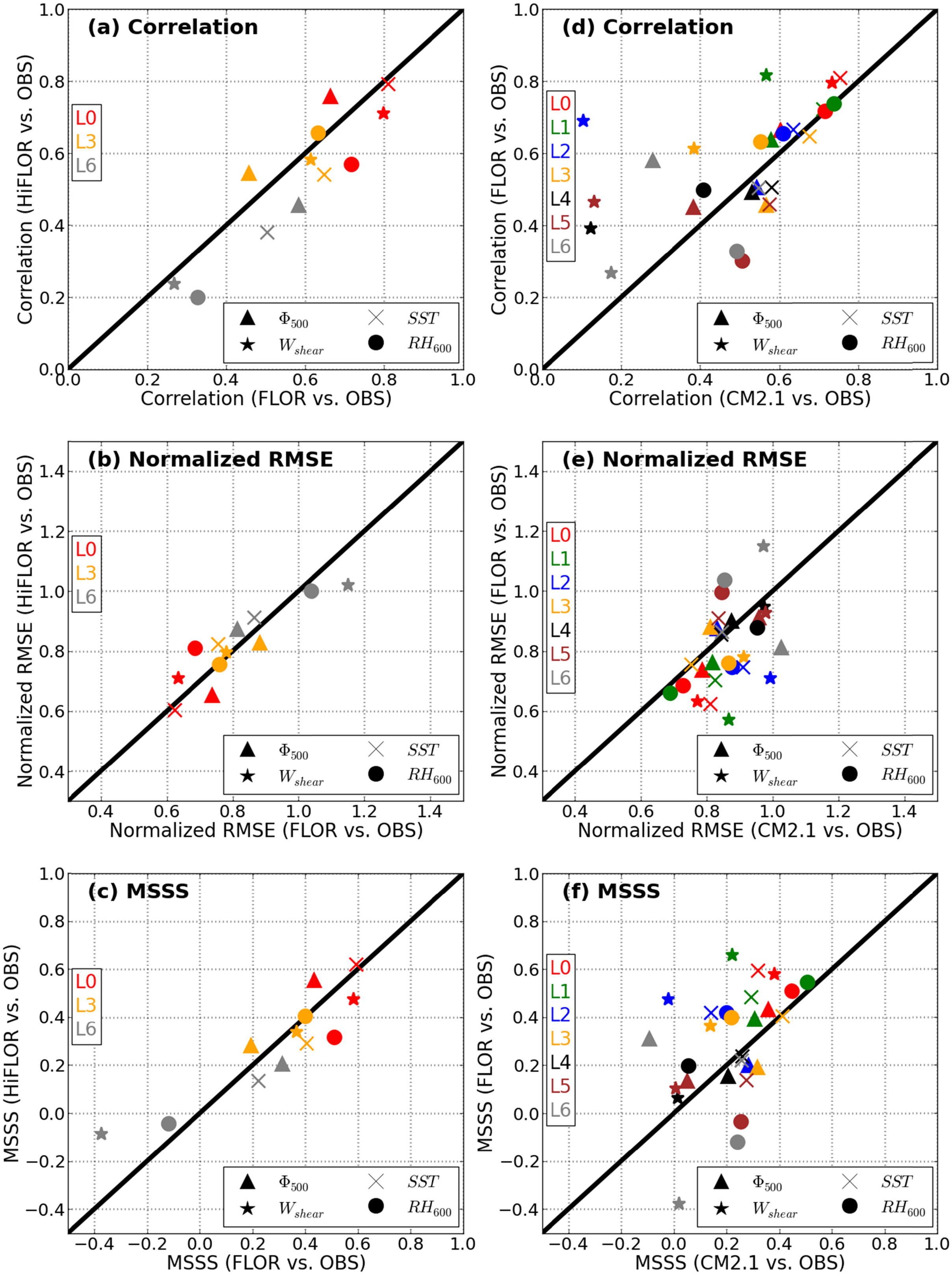
(a) Correlation
Correlation (HiFLOR vs. OBS)
Correlation (FLOR vs. OBS)
(b) Normalized RMSE
Normalized RMSE (HiFLOR vs. OBS)
Normalized RMSE (FLOR vs. OBS)
(c) MSSS
MSSS (HiFLOR vs. OBS)
MSSS (FLOR vs. OBS)
(d) Correlation
Correlation (FLOR vs. OBS)
Correlation (CM2.1 vs. OBS)
(e) Normalized RMSE
Normalized RMSE (FLOR vs. OBS)
Normalized RMSE (CM2.1 vs. OBS)
(f) MSSS
MSSS (FLOR vs. OBS)
MSSS (CM2.1 vs. OBS)
L0
L1
L2
L3
L4
L5
L6
Φ_{500}
SST
W_{shear}
RH_{600}

FIG. 5 (a) Scatterplot of correlation coefficient between HiFLOR prediction and observations (*y* axis) vs FLOR prediction and observations (*x* axis). A correlation coefficient above the diagonal lines indicates that HiFLOR shows higher correlations than FLOR. (b), (c) As in (a), but for NRMSE and MSSS, respectively. A NRMSE (MSSS) below (above) the diagonal lines indicates that HiFLOR shows higher skill than FLOR. Variables evaluated are geopotential height at 500 hPa in the subtropical ENP (Φ_{500}), vertical wind shear in the tropical NAT (W_{shear}), SST anomaly over the tropical NAT (*SST*), and relative humidity at 600 hPa over the tropical NAT (RH_{600}). Different colors indicate different lead months. (d–f) As in (a–c), but for comparisons between FLOR (*y* axis) and CM2.1 (*x* axis).