Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL082529

Dynamical Seasonal Prediction of Tropical Cyclone Activity: Robust Assessment of Prediction Skill and Predictability

Gan Zhang[1,2], Hiroyuki Murakami[2,3], Rich Gudgel[2], and Xiaosong Yang[2,3]

[1]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA, [2]Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton, NJ, USA, [3]University Corporation for Atmospheric Research, Boulder, CO, USA

Key Points:
- Robust analysis reveals when and where the seasonal prediction of tropical cyclone activity is skillful
- Predictability for tropical cyclone activity is lower in the extratropics and the coastal region
- Further improvements of seasonal prediction need to better constrain atmospheric variability

Supporting Information:
- Supporting Information S1

Correspondence to:
G. Zhang,
ganzhang@princeton.edu

Citation:
Zhang, G., Murakami, H., Gudgel, R., & Yang, X. (2019). Dynamical seasonal prediction of tropical cyclone activity: Robust assessment of prediction skill and predictability. *Geophysical Research Letters*, *46*, 5506–5515. https://doi.org/10.1029/2019GL082529

Received 19 FEB 2019
Accepted 29 APR 2019
Accepted article online 2 MAY 2019
Published online 16 MAY 2019

Abstract Improving the seasonal prediction of tropical cyclone (TC) activity demands a robust analysis of the prediction skill and the inherent predictability of TC activity. Using the resampling technique, this study analyzes a state-of-the-art prediction system and offers a robust assessment of when and where the seasonal prediction of TC activity is skillful. We found that uncertainties of initial conditions affect the predictions and the skill evaluation significantly. The sensitivity of predictions to initial conditions also suggests that landfall and high-latitude activity are inherently harder to predict. The lower predictability is consistent with the relatively low prediction skill in these regions. Additionally, the lower predictability is largely related to the atmospheric environment rather than the sea surface temperature, at least for the predictions initialized shortly before the hurricane season. These findings suggest the potential for improving the seasonal TC prediction and will help the development of the next-generation prediction systems.

Plain Language Summary Dynamical seasonal prediction systems have recently shown great promises in predicting tropical cyclone (TC) activity. This study takes advantage of the new tool and evaluates the realized and the potential skills of the prediction. The evaluation reveals a caveat in earlier examinations of the TC prediction, which may significantly inflate or deflate the prediction skill. Our analysis takes account of the issue and offers a detailed view of when and where the TC prediction is skillful. The analysis also suggests that the realized and the potential prediction skill are relatively low for landfall activity and high-latitude activity. By analyzing simulated TC activity and its environmental controls, the study highlights the potential for improving seasonal TC prediction. The information is valuable to decision makers who manage risk in these regions, as well as the improvement of dynamical prediction systems.

1. Introduction

Dynamical models were recently adopted in the seasonal prediction of tropical cyclone (TC) activity and have shown remarkable skill (e.g., Camargo & Barnston, 2009; Camp et al., 2015; Chen & Lin, 2013; Harnos et al., 2017; Manganello et al., 2016, 2017; Murakami et al., 2018, 2016; Vecchi et al., 2014; Vitart et al., 2003; Vitart, 2006; Vitart et al., 2007; Vitart et al., 2010; Vitart & Stockdale, 2001). A key to the success of these systems is running multiple simulations with perturbed initial conditions. This approach, known as ensemble prediction, accounts for the sensitivity of model predictions to uncertainties in initial conditions. An average of the ensemble predictions reduces the noisiness associated with uncertainties in initial conditions and underlines the robust response among predicted scenarios. For a given model, the skill gain from the ensemble approach depends on how well uncertainties in initial conditions are sampled, which is partly indicated by the number of perturbed initial conditions used to initialize the ensemble prediction. For example, the skill gain from adding perturbed initial conditions does not start to saturate until the ensemble size reaches 25–30, at least in a TC-permitting prediction system developed at the European Centre for Medium-Range Weather Forecasts (Manganello et al., 2016). However, such an ensemble size is difficult to attain in the operational prediction of TC activity, as a high spatial resolution is required to simulate TCs but is limited by the computational resource available within narrow time frames.

This study aims at exploring the implications of uncertainties in initial conditions for the seasonal prediction of TC activity. We will start with testing the hypothesis that prediction uncertainties, which can be contributed by uncertainties in initial conditions, can introduce substantial uncertainties in the skill evaluation of predictions. More specifically, such evaluation is usually applied to the hindcasts, which are retrospective forecasts using only the initial conditions that are practically available at the time of prediction. However, inadequate or biased sampling of uncertainties in initial conditions might introduce errors that have a random component. Such errors would exist in predictions even when a prediction system perfectly represents the climate dynamics. Therefore, the true skill of a prediction system may not be accurately evaluated if the role of uncertainties in initial conditions is not properly considered. Nonetheless, the issue is not always well recognized in the skill evaluation, as will be further discussed in section 2.2.

Following the skill evaluation, we will examine the prediction's sensitivity to the initial condition and explore the predictability of TC activity. The predictability is the extent to which future states of a dynamic system can be predicted. As an intrinsic characteristic of a dynamic system, the predictability is associated with the upper bound of potential prediction skill, which might not have been realized by prediction systems. The predictability of basin-wide TC activity has been extensively studied (e.g., Gray, 1984; Goldenberg et al., 2001; Dunstone et al., 2011; Vimont & Kossin, 2007; Wang et al., 2015). Recent findings from dynamical predictions suggested that TC activity is predictable on the subbasin scale, even though the prediction skill varies across regions (e.g., Camp et al., 2015; Vecchi et al., 2014). In particular, Mei et al. (2014, 2015) suggested that near-land TC activity was not strongly constrained by the sea surface temperature (SST) and might be inherently less predictable. However, the findings by Mei et al. (2014, 2015) were based on a three-member ensemble of Atmospheric General Circulation Model simulations. Aside from the relatively small ensemble size, their SST-forced experiments did not consider uncertainties in initial conditions of the ocean or the ensuing error growth in the coupled climate system. Therefore, the findings by Mei et al. (2014, 2015) may not directly apply to the predictability estimate in the seasonal TC prediction.

This study will explore the two foregoing issues with the aim of informing decision makers of when and where the seasonal TC prediction is reliable. Our investigation will leverage the Forecast-Oriented Low Ocean Resolution (FLOR) model developed at Geophysical Fluid Dynamics Laboratory (GFDL; Vecchi et al., 2014). The model is a state-of-the-art model in predicting the seasonal activity of TCs, and its variants have been adopted in the real-time North American Multi-Model Ensemble prediction. Previous studies reported that FLOR is skillful in predicting seasonal TC activity at regional scale (Murakami et al., 2015, 2016; Vecchi et al., 2014). The model setup, as well as the data and methodology, will be further described in section 2. Section 3 will present the results, followed by a summary and discussion in section 4.

2. Data and Methodology

2.1. Observation Data and Model Setup

We use the International Best Track Archive for Climate Stewardship (Knapp et al., 2010) as the reference TC data set. Our analysis focuses on TC activity during July–November, which approximately corresponds to the peak TC seasons of the northern hemisphere basins, including the North Atlantic, the Northeastern Pacific, and the Northwestern Pacific (Schreck et al., 2014). Basin-wide TC activity is characterized using the tropical cyclone frequency (TCF) and the accumulated cyclone energy (ACE; Bell et al., 2000). The ACE is defined as the sum of the squares of maximum sustained surface wind speed for all the records of named TCs. To analyze TC activity on the subbasin scale, we calculate the number of TC days during each hurricane season in $5° \times 5°$ grid boxes within the global domain. The data of TC days are smoothed using a 9-point spatial filter to reduce its noisiness.

The GFDL FLOR model used in the study has been described in detail by Vecchi et al. (2014). We use a version with the flux adjustment (FA) correction (Magnusson et al., 2013), which reduces the SST biases and improves the TC prediction (Vecchi et al., 2014). In the FLOR-FA configuration, the horizontal grid spacing of the atmosphere and the ocean components are about 50 km and 1°, respectively. The relatively high resolution of the atmosphere component helps simulate TCs, which are tracked using an algorithm described by Murakami et al. (2015) and Harris et al. (2016). The prediction model also outputs the large-scale environmental variables, such as the SST and the precipitable water, which are closely associated with TC activity. The environmental variables are regridded to a 5° grid to facilitate the computationally intensive postprocessing described in section 2.2.

2.2. Model Initialization and Hindcast Resampling

The FLOR-FA hindcasts are 12-month predictions initialized on the first day of each calendar month of 1981–2014. Our analysis focuses on the predictions initialized during January–July, which precede the peak TC seasons of the Northern Hemisphere basins. To initialize the predictions, 12 sets of the ocean-ice initial conditions are generated using a coupled ensemble data assimilation system (Zhang et al., 2007), while three sets of atmosphere-land initial conditions are generated using SST-constrained simulations by the FLOR model (Vecchi et al., 2014). Each set of the atmosphere-land initial conditions is combined with 4 of the 12 sets of ocean-ice initial conditions to acquire 12 complete sets of initial conditions. More details of the model settings and initialization are available in the supporting information (Delworth et al., 2006; Delworth et al., 2012; Mulholland et al., 2015). As noted by Vecchi et al. (2014), the initialization procedure is a technical choice and is "suboptimal" for sampling uncertainties in initial conditions.

Based on the initialization procedure, we can denote the FLOR-FA hindcasts as $S_{i,j,k}$, where i represents an arbitrarily assigned ensemble index ranging from 1 to 12, j represents the initialization month ranging from 1 to 7, and k represents the initialization year ranging from 1981 to 2014. We consider the 12-member predictions initialized at each time step as essentially independent simulations. Furthermore, the predictions do not provide the initial condition for any predictions initialized at a later time step. Therefore, a prediction initialized in a given month of one year ($S_{i,\ j\ =\ J,\ k\ =\ K}$) is independent of the prediction initialized in the same month of the next year ($S_{i,j\ =\ J,k\ =\ K+1}$). For hindcasts initialized in a specific month, the independence suggests that a 34-year simulation ($k = 1981, \ldots, 2014$) can be assembled by randomly selecting i for each year.

To evaluate the skill of an ensemble prediction system, an important metric is the Pearson correlation coefficient (r) between the observation and the ensemble average. Manganello et al. (2016) noted that the correlation coefficient varies when different combinations of ensemble members are used, especially when the ensemble size is smaller than 25–30. The variations indicate that uncertainties in initial conditions can introduce a random component when the skill metric is evaluated. However, the skill metric is usually evaluated for an *arbitrary* combination of hindcast members (e.g., Chen & Lin, 2013; Vecchi et al., 2014), suggesting that the skill evaluation might have unrecognized uncertainties. To explore if such skill evaluation are robust, we resample i in predictions $S_{i,j,k}$ with replacement and draw 12-member samples. By conducting the resampling for $k = 1981, \ldots, 2014$, we can generate an artificial set of 34-year hindcasts that consists of 12 ensemble members. For each initialization month ($j = 1, \ldots, 7$), we repeat the procedure and generate 10,000 artificial sets of 12-member and 34-year hindcasts. Each artificial ensemble is equal in representing the model-simulated interannual variability. Using the resampling approach, we evaluate the model performance by calculating the correlation coefficients between the observation and the ensemble averages. Results of some other skill metrics are available in Supplementary Information.

2.3. Metrics of Estimating Predictability

This study estimates the seasonal predictability of TC activity using the concept of "predictable component" (PC). Eade et al. (2014) proposed that predictions within a time range can be decomposed into (i) unpredictable noise related to the chaos of the climate system and (ii) a "PC" that is constrained by predictable dynamic processes. The study also defines the PCs as the square root of the predictable fraction within the total variance. The PC in the observation (PC_{obs}) can be estimated using the Pearson correlation (r) between the observation and the ensemble average, while the PC in model hindcasts (PC_{mod}) can be estimated using the following expressions:

$$PC_{\mathrm{mod}} = \sqrt{\sigma_{\mathrm{sig}}^2 / \sigma_{\mathrm{tot}}^2} \tag{1}$$

Where σ_{sig}^2 is the temporal variance of the model ensemble mean and σ_{tot}^2 is the average of the temporal variance of individual ensemble members (Eade et al., 2014). We calculate PC_{obs} and PC_{mod} for each of the artificial ensembles, which helps outline the uncertainties in estimating PCs (Shi et al., 2015). The estimates of PC_{obs} and PC_{mod} are subject to prediction biases, and the predictability indicated by the metrics is also model dependent. Furthermore, the mathematical formulations of PC_{obs} and PC_{mod} allow the two metrics to characterize prediction skills. For example, PC_{obs}, by its correlation definition, is essentially a commonly used skill metric; a comparison of PC_{obs} and PC_{mod} can suggest whether a prediction system is overconfident ($PC_{\mathrm{obs}} < PC_{\mathrm{mod}}$) or underconfident ($PC_{\mathrm{obs}} > PC_{\mathrm{mod}}$; Eade et al., 2014).

The predictability can also be estimated by evaluating the sensitivity of predictions to initial conditions using the perfect model approach (e.g., Lorenz, 1982). By assuming that a model perfectly represents the real-world dynamic system, the approach affords an opportunity to put aside model errors and focus on the sensitivity of predictions to uncertainties in initial conditions. For an N-member ensemble, the sensitivity can be indicated by the Pearson correlation between the prediction by one ensemble member and the average of the other ensemble members ($N - 1$). Following section 3.1 in Pegion et al. (2017), we calculate the correlation coefficient, rotate through each ensemble member, and average the coefficients from the 12 calculations. Together with the resampling described in section 2.2, the approach helps to examine a large number of possible combinations and represent the sensitivity in the predictions. The average of correlation coefficients indicates the consistency among ensemble members, with weaker intraensemble correlations suggesting that the prediction is more sensitive to the initial condition. This procedure is repeated for each of the artificial ensemble (section 2.2) to ensure a robust estimate of the sensitivity.

3. Results

3.1. Prediction Skill and PCs

We first examine the prediction skill of the basin-wide TCF and the basin-wide ACE. Figure 1 shows the correlation coefficients between the observation and the ensemble averages. The correlation coefficients (denoted with red) suggest that the FLOR-FA model is skillful in predicting TC activity, especially for the North Atlantic and the northeastern Pacific. The relatively poor skill in predicting the TCF in the northwestern Pacific basin has been noted and can be improved using statistical methods (e.g., Murakami et al., 2016; W. Zhang et al., 2016). The prediction skills also vary with the lead time of the initialization, with the predictions initialized in later months generally being more skillful. However, there are some noteworthy behaviors or even exceptions. For example, there is a skill jump of the ACE prediction between March and April in the northwestern Pacific (Figure 1f), which might be related to the spring barrier of predicting the Pacific climate (Webster & Yang, 1992). Another interesting example occurs in the northeastern and the northwestern Pacific basin, where the June prediction of TCF is more skillful than the July TCF prediction (Figure 1b). This behavior might be partly related to model drifts and biases that depend on the initialization time (Figures S3–S7 in the supporting information).

Figure 1 also shows that the range in correlation coefficients for basin-wide TC activity (PC_{obs}) vary between ~0.10 and ~0.25. The range could widen to ~0.40 if the ensemble size is reduced from 12 to 6 (not shown). The large spreads are consistent with the hypothesis that uncertainties in initial conditions affect the skill evaluation of TC predictions. For example, the 2.5th to 97.5th percentile range of the correlation coefficients is approximately 0.40 to 0.60 in the April prediction of the North Atlantic TCF. Without considering the uncertainties using the resampling technique, the skill evaluation of an arbitrary ensemble combination could deem the prediction as marginally skillful ($r \approx 0.40$) or highly skillful ($r \approx 0.60$) by chance. The issue becomes more severe with a smaller ensemble size. For example, the correlation range in Figure 1 can reach ~0.40 when we reduce the ensemble size from 12 to 6 (not shown).

The blue box plots in Figure 1 show the PCs of basin-wide TC metrics estimated using the FLOR-FA hindcasts (PC_{mod}). Similar to the PC_{obs} (viz., the correlation coefficients), the PC_{mod} also varies with the lead time of the initialization, with higher values appearing when the lead time is small. However, the variations of PC_{mod} with the lead time are generally weaker, and the values of PC_{mod} remain above 0.5. In nearly all the cases, PC_{obs} is significantly smaller than PC_{mod}. As noted in section 2.3 and by Eade et al. (2014), the relation suggests that the predictions have a relatively high signal-to-noise ratio and are likely overconfident. In other words, the predictions by ensemble members show relatively high consistency even when they are not particularly skillful. In the Pacific basins, for example, the July predictions of the TCF have a relatively low skill despite of the high intraensemble consistency suggested by the PC_{mod} (Figure 1b and 1c). The behavior also suggests that prediction errors might make PC_{obs} indicate unrealistically low predictability.

Besides the basin-wide metrics, the prediction skill of spatial characteristics of TC tracks also carries important values. Figure 2 shows the PC_{obs} and the PC_{mod} estimated using the July prediction of TC

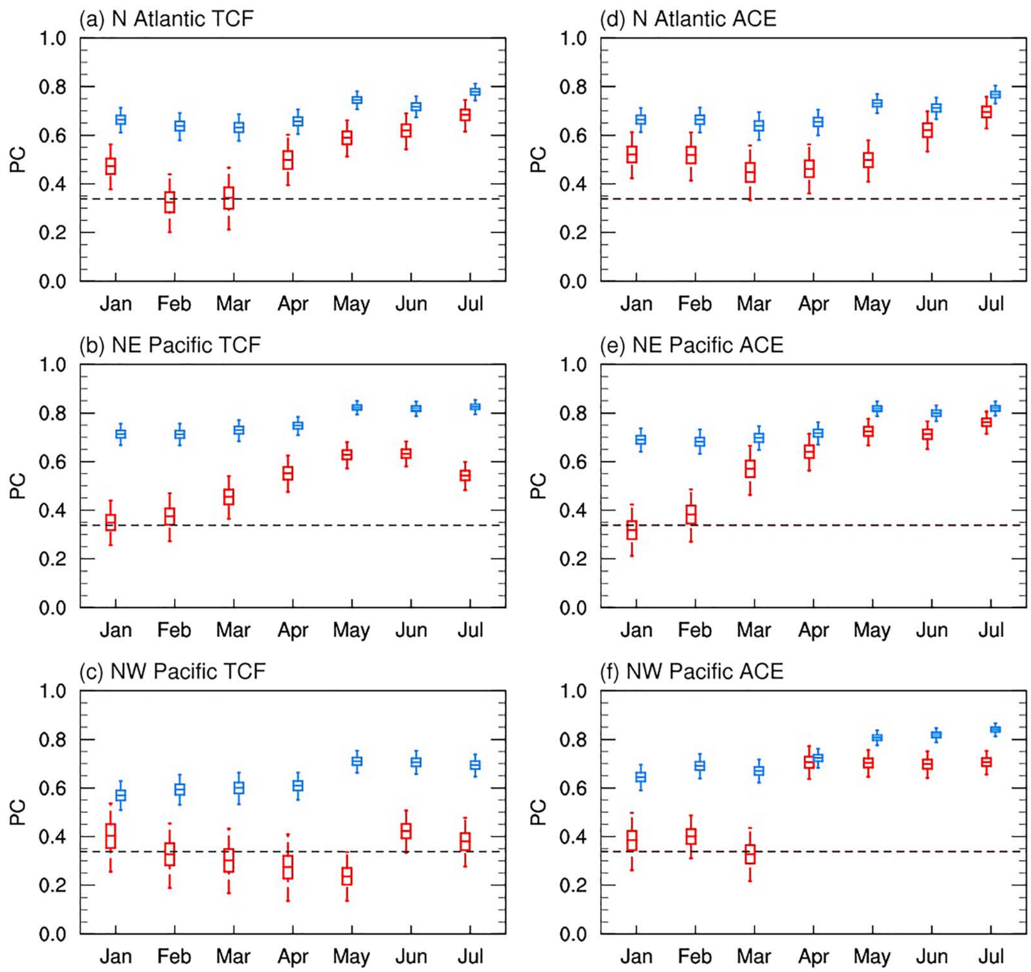

Figure 1. Estimates of predictable components (PCs) of the observation (red) and the Forecast-Oriented Low Ocean Resolution-flux adjustment hindcasts (blue). By its definition, PC_{obs} also indicates the prediction skill. (a–c) Tropical cyclone frequency (TCF) for the North Atlantic, northeastern Pacific, and the northwestern Pacific. (d–f) The same as (a)–(c) but for the accumulated cyclone energy (ACE). The box plots show 2.5th, 25th, 50th, 75th, and 97.5th percentiles. The dashed line represents the 95% confidence level for the correlation.

days (section 2.1). The PC_{obs}, namely, the correlation coefficients, show a pattern that is similar to the correlation map of the FLOR predictions described by Vecchi et al. (2014). In particular, the prediction of TC days near the Caribbean Islands, and Central America, Hawaii, and Micronesia is relatively skillful ($r > 0.40$; Figure 2a). The regions with high values of the PC_{obs} roughly correspond to the regions with high values of the PC_{mod}. The consistency suggests the relatively skillful predictions may be related to a relatively large PC, which exists in the real-world climate system and is successfully captured by the FLOR-FA. Nonetheless, the region where the PC_{obs} and the PC_{mod} are comparable is patchy and limited. The prediction is persistently underdispersive and overconfident in many coastal regions of Asia and North America.

3.2. Sensitivity to the Initial Condition and Predictability Estimate

To further illustrate the predictability of TC distribution on the subbasin scale, we now examine the intraensemble correlations (section 2.3) of TC days in the FLOR-FA hindcasts. Figure 3 shows the intraensemble correlation (section 2.3) of the predictions initialized in January, April, and July. The three predictions show similar spatial patterns of the intraensemble correlation, with the strongest intraensemble correlation appearing over the open ocean near 15°N. The correlation strengthens with the reduction of the lead time of initialization. For example, the strongest correlation over the northwestern Pacific increases from 0.5–

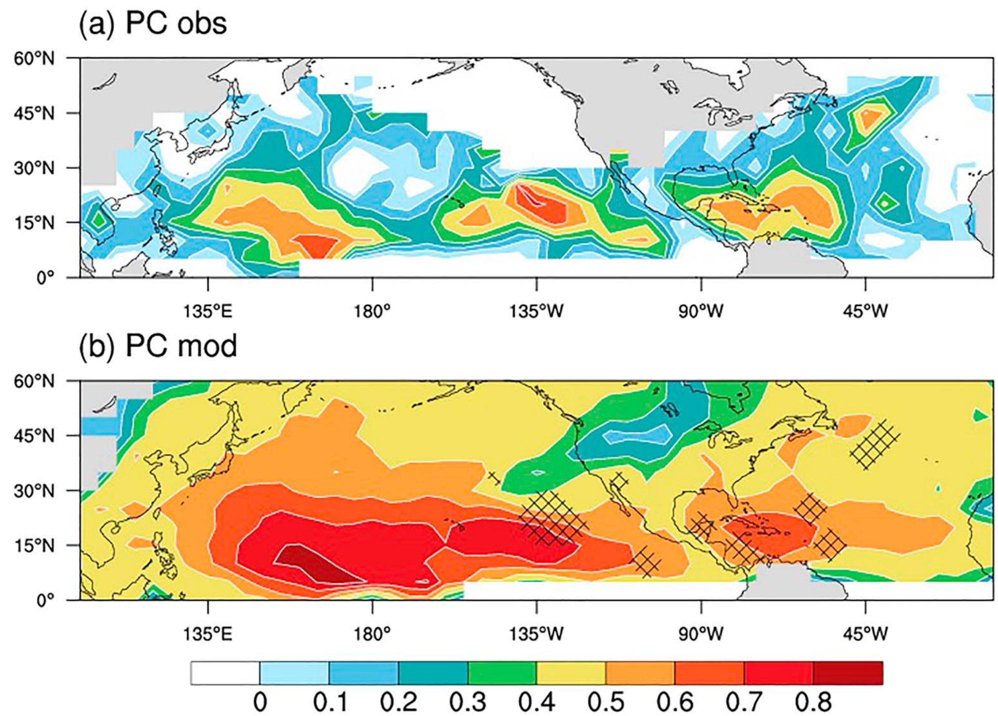

Figure 2. Estimates of predictable components (PCs) of tropical cyclone days in (a) the observation and (b) the hindcasts. The predictions are initialized in July and target at July–November. Color shading shows the mean of resampled predictable components. Hatching in (b) highlights the regions where the difference between PC_{obs} and PC_{mod} is not statistically significant at 95% confidence level. The confidence level is determined with the resampling.

0.6 in January to 0.7–0.8 in July. In parts of the extratropics and most coastal regions, the correlation also increases from the January prediction but remains <0.40 even in the July prediction. The results suggest that the prediction of TC activity in the coastal regions is more sensitive to uncertainties in initial conditions and thus less predictable for the FLOR-FA model.

On the subbasin scale, the predictability indicated by the sensitivity to initial conditions appears to correspond to the prediction skill. For example, the regions where the prediction is the least sensitive to the initial condition are nearly identical to the regions with high prediction skills ($r > 0.40$ in Figure 2a). Therefore, the prediction in these regions might have benefited from higher inherent predictability of the dynamical system. Interestingly, the northwestern Pacific prediction appears less sensitive to the initial condition than the North Atlantic prediction does. The difference is consistent with that a larger portion of TC development in the Atlantic basin involve baroclinic processes (McTaggart-Cowan et al., 2013), which is associated with lower predictability on the synoptic time scale (Wang et al., 2018). However, the seasonal prediction of northwestern Pacific TC activity is only comparable to or even worse than the North Atlantic prediction (Figures 1 and 2). The result suggests that the perfect prediction skill, or the potential skill indicated by the "perfect" model, does not always translate to the actual prediction skill.

To explore what might affect the sensitivity of TC distribution to initial conditions, we examine the intraensemble correlation of some TC-related environmental variables (Figure 4). For the July prediction, the SST is the least sensitive to uncertainties in initial conditions. The ensemble members are nearly perfectly correlated ($r > 0.8$) except in some extratropical regions. The prediction of the precipitation and the vertical wind shear show strong intraensemble correlations in the deep tropics, but the correlations degrade substantially poleward where the atmosphere gradually becomes baroclinically unstable. The results are consistent with the notion that the tropical ocean is closely coupled with the tropical atmosphere (Trenberth et al., 1998), but the upper ocean does not strongly constrain the interannual variability of the extratropical atmosphere (Kushnir et al., 2002; Lee et al., 2011; Sutton & Hodson, 2007). A comparison of the sensitivity of the environmental variables to the initial condition suggests that the short-range predictability of TC days might be

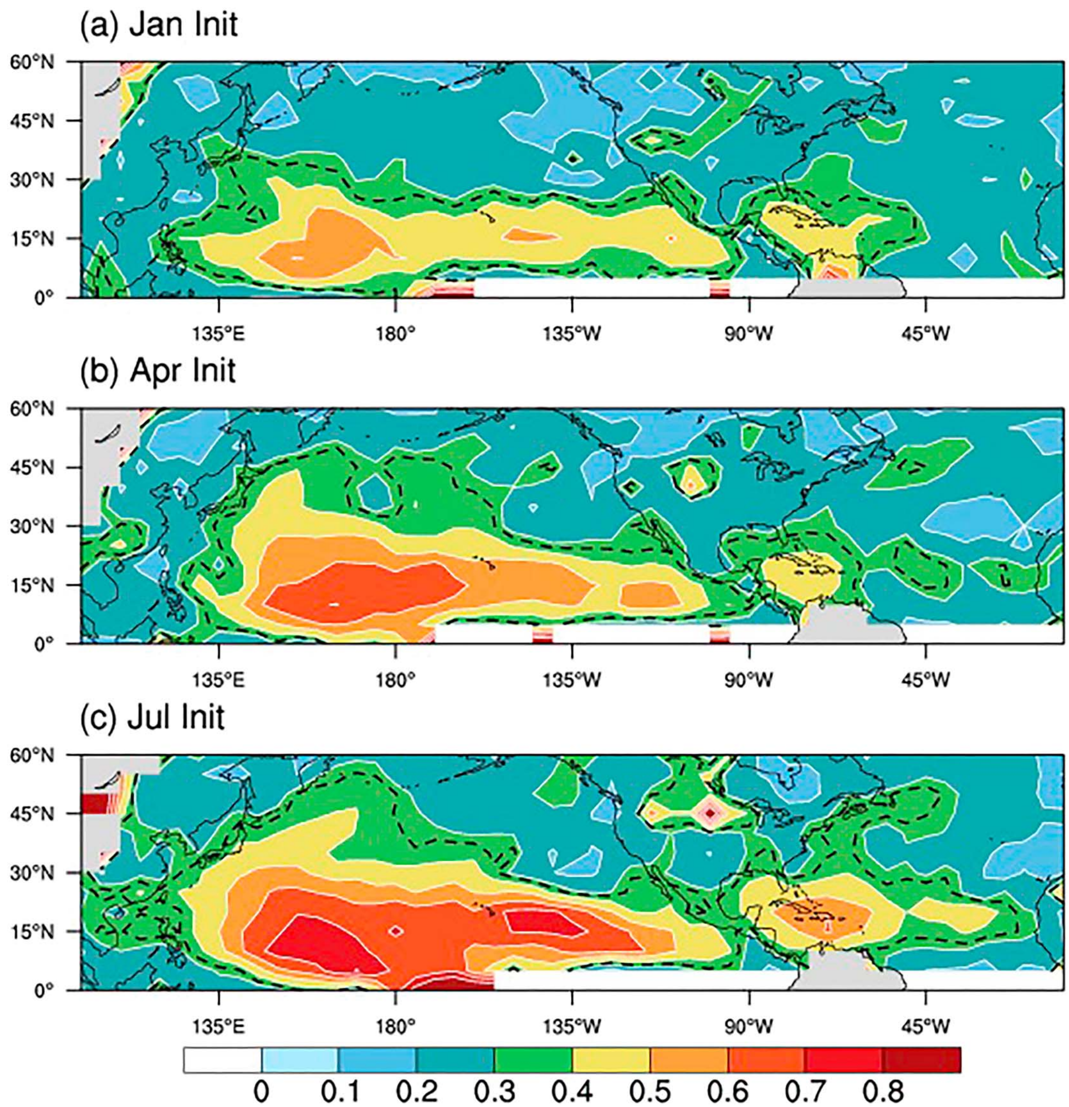

Figure 3. Intraensemble correlation of predicted tropical cyclone days (July–November). Predictions are initialized in (a) January, (b) April, and (c) July. The black dashed line denotes the Pearson correlation that is at the 95% confidence level.

more strongly limited by the atmospheric variables than by the oceanic forcing. For long-range predictions initialized earlier, the limitation by oceanic processes also becomes apparent (not shown) and further lowers the predictability of TC days (Figure 3).

4. Summary and Discussion

This study uses the GFDL FLOR-FA prediction system to examine how uncertainties in initial conditions affect the seasonal prediction of TC activity. We show that a "suboptimal" sampling of uncertainties in initial conditions (Vecchi et al., 2014) affects the skill evaluation of the seasonal prediction. For the 12-member hindcasts of 1981–2014, the uncertainties can make the correlation between the observed and the predicted TC metrics vary by ~0.10 to ~0.25. These variations can affect the model skill evaluation by making predictions from the same model appear marginally or highly skillful. The uncertainty in the skill evaluation is greater when the ensemble size is small. A skill evaluation with the resampling suggests that the prediction by FLOR-FA tends to be more skillful when the lead time of initialization is relatively short and in tropical regions with extensive ocean coverage, such as the Caribbean Islands, and Central America, Hawaii, and Micronesia.

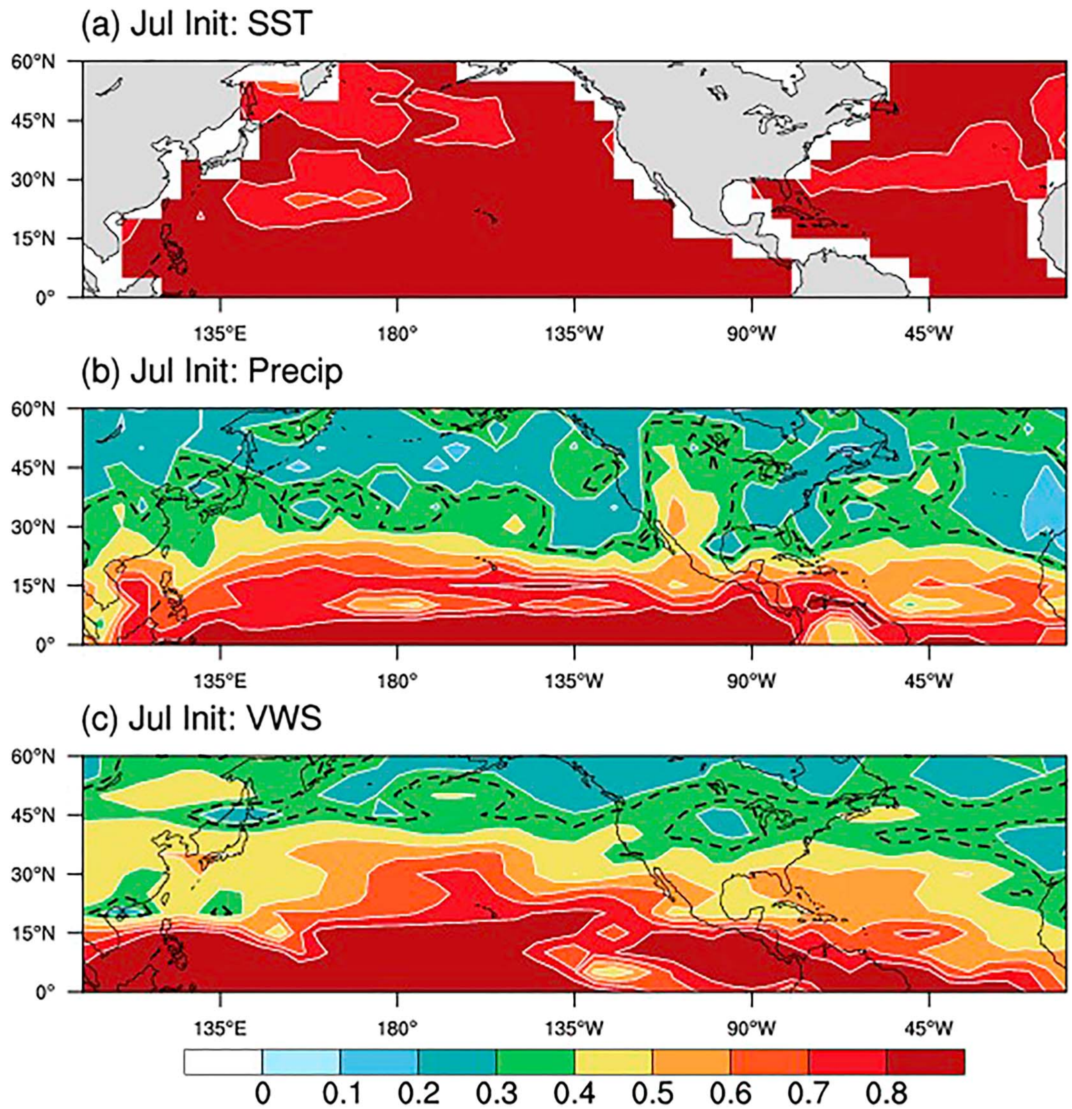

Figure 4. Intraensemble correlation of predicted large-scale environment (July–November). (a) Sea surface temperature (SST), (b) precipitation, and (c) 200- to 850-hPa vertical wind shear. The predictions are initialized in July. The black dashed line denotes the Pearson correlation that is at the 95% confidence level.

A comparison of PCs in the observation and the modeled climate suggests the prediction tends to be overconfident, especially when the prediction is less skillful. The issue might arise because of an underestimate of the PC in the observation, since model biases and a suboptimal initialization may lower the correlation between the observation and the prediction (Eade et al., 2014). Using the perfect model approach, we examine the sensitivity of the prediction to the initial condition and find that the predictability of TC activity is lower in the extratropics and the coastal region. The predictability, at least for short-range predictions, appears largely limited by the atmospheric environment variables rather than the SST. Consistent with this finding, earlier studies suggest that the variability of extratropical atmosphere is not strongly constrained by the oceanic forcing (Kushnir et al., 2002; Lee et al., 2011; Sutton & Hodson, 2007) and is associated with lower prediction skills of TC activity (Bell et al., 2014; Fitzpatrick et al., 1995; Li et al., 2018; Wang et al., 2018; G. Zhang, et al., 2016). Overall, the skill of the FLOR-FA prediction appears higher in regions where the potential predictability is higher.

The study outlines when and where the FLOR-FA prediction is the most skillful. Although findings might be partially model dependent, they should be useful for decision makers who wish to take advantage of the dynamical predictions. The study also highlights some intriguing issues in the seasonal prediction of TC activity. For example, there is a significant gap between the realized prediction skill and the potential predictability indicated by the FLOR-FA hindcasts. Although the gap does not necessarily indicate room for skill improvements (Kumar et al., 2014), it does suggest significant differences between the modeled and the real-world climate. Another issue is the lower predictability of near-land TC activity. The cause of the low predictability is not immediately clear. It may be related to the crude initialization of the land-atmosphere component of the FLOR-FA model (Vecchi et al., 2014) or the land-atmosphere-ocean interaction that might be inherently less predictable. These issues suggest new opportunities for improving the dynamical prediction and physical understanding. For example, refining the model initialization will likely improve predictions. Our efforts that attempt to address these issues will be introduced in upcoming studies.

Acknowledgments
G. Zhang is supported by Princeton University's Cooperative Institute for Modeling the Earth System, through the predicting and explaining extremes Initiative. The IBTrACS data are acquired from National Centers for Environmental Information (https://www.ncdc.noaa.gov/ibtracs/index.php). The model output used to generate the plots is available at GFDL data portal (ftp://data1.gfdl.noaa.gov/users/Gan.Zhang/FLOR/GFDL-CM2p5/). The model code is publicly available at https://www.gfdl.noaa.gov/cm2-5-and-flor/, and part of the FLOR predictions is available at the https://www.cpc.ncep.noaa.gov/products/NMME/ website. We thank Drs. Mitch Bushuk, Baoqiang Xiang, Suzana Camargo, and two reviewers for thoughtful suggestions.

References

Bell, G. D., Halpert, M. S., Schnell, R. C., Higgins, R. W., Lawrimore, J., Kousky, V. E., et al. (2000). Climate Assessment for 1999. *Bulletin of the American Meteorological Society*, *81*(6), s1–s50. https://doi.org/10.1175/1520-0477(2000)81[s1:CAF]2.0.CO;2

Bell, G. D., Landsea, C. W., Goldenberg, S. B., Pasch, R. J., Blake, E. S., Schemm, J., & Kimberlain, T. B. (2014). The 2013 North Atlantic hurricane season: A climate perspective. In State of the climate in 2013. *Bulletin of the American Meteorological Society*, *95*(7), S1–S279. https://doi.org/10.1175/2014BAMSStateoftheClimate.1

Camargo, S. J., & Barnston, A. G. (2009). Experimental seasonal dynamical forecasts of tropical cyclone activity at IRI. *Weather Forecasting*, *24*(2), 472–491. https://doi.org/10.1175/2008WAF2007099.1

Camp, J., Roberts, M., MacLachlan, C., Wallace, E., Hermanson, L., Brookshaw, A., et al. (2015). Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, *141*(691), 2206–2219. https://doi.org/10.1002/qj.2516

Chen, J.-H., & Lin, S.-J. (2013). Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *Journal of Climate*, *26*(2), 380–398. https://doi.org/10.1175/JCLI-D-12-00061.1

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., et al. (2006). GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate*, *19*(5), 643–674. https://doi.org/10.1175/JCLI3629.1

Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., et al. (2012). Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *Journal of Climate*, *25*(8), 2755–2781. https://doi.org/10.1175/JCLI-D-11-00316.1

Dunstone, N. J., Smith, D. M., & Eade, R. (2011). Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophysical Research Letters*, *38*, L14701. https://doi.org/10.1029/2011GL047949

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*, 5620–5628. https://doi.org/10.1002/2014GL061146

Fitzpatrick, P. J., Knaff, J. A., Landsea, C. W., & Finley, S. V. (1995). Documentation of a systematic bias in the aviation model's forecast of the Atlantic tropical upper-tropospheric trough: Implications for tropical cyclone forecasting. *Weather and Forecasting*, *10*(2), 433–446. https://doi.org/10.1175/1520-0434(1995)010<0433:DOASBI>2.0.CO;2

Goldenberg, S. B., Landsea, C. W., Mestas-Nuñez, A. M., & Gray, W. M. (2001). The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, *293*(5529), 474–479. https://doi.org/10.1126/science.1060040

Gray, W. M. (1984). Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Monthly Weather Review*, *112*(9), 1649–1668. https://doi.org/10.1175/1520-0493(1984)112<1649:ASHFPI>2.0.CO;2

Harnos, D. S., Schemm, J.-K. E., Wang, H., & Finan, C. A. (2017). NMME-based hybrid prediction of Atlantic hurricane season activity. *Climate Dynamics*. https://doi.org/10.1007/s00382-017-3891-7

Harris, L. M., Lin, S.-J., & Tu, C. (2016). High-resolution climate simulations using GFDL HiRAM with a stretched global grid. *Journal of Climate*, *29*(11), 4293–4314. https://doi.org/10.1175/JCLI-D-15-0389.1

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The International Best Track Archive for Climate Stewardship (IBTrACS). *Bulletin of the American Meteorological Society*, *91*(3), 363–376. https://doi.org/10.1175/2009BAMS2755.1

Kumar, A., Peng, P., & Chen, M. (2014). Is there a relationship between potential and actual skill? *Monthly Weather Review*, *142*(6), 2220–2227. https://doi.org/10.1175/MWR-D-13-00287.1

Kushnir, Y., Robinson, W. A., Bladé, I., Hall, N. M. J., Peng, S., & Sutton, R. (2002). Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation. *Journal of Climate*, *15*(16), 2233–2256. https://doi.org/10.1175/1520-0442(2002)015<2233:AGRTES>2.0.CO;2

Lee, J.-Y., Wang, B., Ding, Q., Ha, K.-J., Ahn, J.-B., Kumar, A., et al. (2011). How predictable is the northern hemisphere summer upper-tropospheric circulation? *Climate Dynamics*, *37*(5-6), 1189–1203. https://doi.org/10.1007/s00382-010-0909-9

Li, W., Wang, Z., Zhang, G., Peng, M. S., Benjamin, S. G., & Zhao, M. (2018). Subseasonal variability of Rossby wave breaking and impacts on tropical cyclones during the North Atlantic warm season. *Journal of Climate*, *31*(23), 9679–9695. https://doi.org/10.1175/JCLI-D-17-0880.1

Lorenz, E. N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, *34*(6), 505–513. https://doi.org/10.1111/j.2153-3490.1982.tb01839.x

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, *41*(9-10), 2393–2409. https://doi.org/10.1007/s00382-012-1599-2

Manganello, J. V., Cash, B. A., Hodges, K. I., & Kinter, J. L. (2017). Seasonal forecasts of North Atlantic tropical cyclone activity in the North American Multi-Model Ensemble. *Climate Dynamics*. https://doi.org/10.1007/s00382-017-3670-5

Manganello, J. V., Hodges, K. I., Cash, B. A., Kinter, J. L., Altshuler, E. L., Fennessy, M. J., et al. (2016). Seasonal forecasts of tropical cyclone activity in a high-atmospheric-resolution coupled prediction system. *Journal of Climate*, *29*(3), 1179–1200. https://doi.org/10.1175/JCLI-D-15-0531.1

McTaggart-Cowan, R., Galarneau, T. J. Jr., Bosart, L. F., Moore, R. W., & Martius, O. (2013). A global climatology of baroclinically influenced tropical cyclogenesis. *Monthly Weather Review*, *141*(6), 1963–1989. https://doi.org/10.1175/MWR-D-12-00186.1

Mei, W., Xie, S.-P., & Zhao, M. (2014). Variability of tropical cyclone track density in the North Atlantic: Observations and high-resolution simulations. *Journal of Climate*, *27*(13), 4797–4814. https://doi.org/10.1175/JCLI-D-13-00587.1

Mei, W., Xie, S.-P., Zhao, M., & Wang, Y. (2015). Forced and internal variability of tropical cyclone track density in the western North Pacific. *Journal of Climate*, *28*(1), 143–167. https://doi.org/10.1175/JCLI-D-14-00164.1

Mulholland, D. P., Laloyaux, P., Haines, K., & Balmaseda, M. A. (2015). Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Monthly Weather Review*, *143*(11), 4631–4644. https://doi.org/10.1175/MWR-D-15-0076.1

Murakami, H., Levin, E., Delworth, T. L., Gudgel, R., & Hsu, P.-C. (2018). Dominant effect of relative tropical Atlantic warming on major hurricane occurrence. *Science*, *362*, 794–799. https://doi.org/10.1126/science.aat6711

Murakami, H., Vecchi, G. A., Underwood, S., Delworth, T. L., Wittenberg, A. T., Anderson, W. G., et al. (2015). Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *Journal of Climate*, *28*(23), 9058–9079. https://doi.org/10.1175/JCLI-D-15-0216.1

Murakami, H., Villarini, G., Vecchi, G. A., Zhang, W., & Gudgel, R. (2016). Statistical–dynamical seasonal forecast of North Atlantic and U.S. landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. *Monthly Weather Review*, *144*(6), 2101–2123. https://doi.org/10.1175/MWR-D-15-0308.1

Pegion, K., DelSole, T., Becker, E., & Cicerone, T. (2017). Assessing the fidelity of predictability estimates. *Climate Dynamics*. https://doi.org/10.1007/s00382-017-3903-7

Schreck, C. J., Knapp, K. R., & Kossin, J. P. (2014). The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Monthly Weather Review*, *142*(10), 3881–3899. https://doi.org/10.1175/MWR-D-14-00021.1

Shi, W., Schaller, N., MacLeod, D., Palmer, T. N., & Weisheimer, A. (2015). Impact of hindcast length on estimates of seasonal climate predictability. *Geophysical Research Letters*, *42*, 1554–1559. https://doi.org/10.1002/2014GL062829

Sutton, R. T., & Hodson, D. L. R. (2007). Climate response to basin-scale warming and cooling of the North Atlantic Ocean. *Journal of Climate*, *20*(5), 891–907. https://doi.org/10.1175/JCLI4038.1

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C., & Ropelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research*, *103*(C7), 14,291–14,324. https://doi.org/10.1029/97JC01444

Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A. T., et al. (2014). On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, *27*(21), 7994–8016. https://doi.org/10.1175/JCLI-D-14-00158.1

Vimont, D. J., & Kossin, J. P. (2007). The Atlantic Meridional Mode and hurricane activity. *Geophysical Research Letters*, *34*, L07709. https://doi.org/10.1029/2007GL029683

Vitart, F., Huddleston, M. R., Déqué, M., Peake, D., Palmer, T. N., Stockdale, T. N., et al. (2007). Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP. *Geophysical Research Letters*, *34*, L16815. https://doi.org/10.1029/2007GL030740

Vitart, F., Leroy, A., & Wheeler, M. C. (2010). A comparison of dynamical and statistical predictions of weekly tropical cyclone activity in the southern hemisphere. *Monthly Weather Review*, *138*(9), 3671–3682. https://doi.org/10.1175/2010MWR3343.1

Vitart, F. D. (2006). Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quarterly Journal of the Royal Meteorological Society*, *132*(615), 647–666. https://doi.org/10.1256/qj.05.65

Vitart, F. D., Anderson, D., & Stockdale, T. (2003). Seasonal forecasting of tropical cyclone landfall over Mozambique. *Journal of Climate*, *16*(23), 3932–3945. https://doi.org/10.1175/1520-0442(2003)016<3932:SFOTCL>2.0.CO;2

Vitart, F. D., & Stockdale, T. N. (2001). Seasonal forecasting of tropical storms using coupled GCM integrations. *Monthly Weather Review*, *129*(10), 2521–2537. https://doi.org/10.1175/1520-0493(2001)129<2521:SFOTSU>2.0.CO;2

Wang, Z., Li, W., Peng, M. S., Jiang, X., McTaggart-Cowan, R., & Davis, C. A. (2018). Predictive skill and predictability of North Atlantic tropical cyclogenesis in different synoptic flow regimes. *Journal of the Atmospheric Sciences*, *75*(1), 361–378. https://doi.org/10.1175/JAS-D-17-0094.1

Wang, Z., Zhang, G., Peng, M. S., Chen, J.-H., & Lin, S.-J. (2015). Predictability of Atlantic tropical cyclones in the GFDL HiRAM model. *Geophysical Research Letters*, *42*, 2547–2554. https://doi.org/10.1002/2015GL063587

Webster, P. J., & Yang, S. (1992). Monsoon and ENSO: Selectively interactive systems. *Quarterly Journal of the Royal Meteorological Society*, *118*(507), 877–926. https://doi.org/10.1002/qj.49711850705

Zhang, G., Wang, Z., Dunkerton, T. J., Peng, M. S., & Magnusdottir, G. (2016). Extratropical impacts on Atlantic tropical cyclone activity. *Journal of the Atmospheric Sciences*, *73*(3), 1401–1418. https://doi.org/10.1175/JAS-D-15-0154.1

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, *135*(10), 3541–3564. https://doi.org/10.1175/MWR3466.1

Zhang, W., Vecchi, G. A., Villarini, G., Murakami, H., Gudgel, R., & Yang, X. (2016). Statistical–dynamical seasonal forecast of western North Pacific and East Asia landfalling tropical cyclones using the GFDL FLOR coupled climate model. *Journal of Climate*, *30*(6), 2209–2232. https://doi.org/10.1175/JCLI-D-16-0487.1