

Dynamical Seasonal Predictions of Tropical Cyclone Activity: Roles of Sea Surface Temperature Errors and Atmosphere–Land Initialization

GAN ZHANG,^{a,b,d} HIROYUKI MURAKAMI,^{b,c} XIAOSONG YANG,^{b,c} KIRSTEN L. FINDELL,^b
ANDREW T. WITTENBERG,^b AND LIWEI JIA^{b,c}

^a *Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey*

^b *Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton, New Jersey*

^c *University Corporation for Atmospheric Research, Boulder, Colorado*

(Manuscript received 25 March 2020, in final form 22 October 2020)

ABSTRACT: Climate models often show errors in simulating and predicting tropical cyclone (TC) activity, but the sources of these errors are not well understood. This study proposes an evaluation framework and analyzes three sets of experiments conducted using a seasonal prediction model developed at the Geophysical Fluid Dynamics Laboratory (GFDL). These experiments apply the nudging technique to the model integration and/or initialization to estimate possible improvements from nearly perfect model conditions. The results suggest that reducing sea surface temperature (SST) errors remains important for better predicting TC activity at long forecast leads—even in a flux-adjusted model with reduced climatological biases. Other error sources also contribute to biases in simulated TC activity, with notable manifestations on regional scales. A novel finding is that the coupling and initialization of the land and atmosphere components can affect seasonal TC prediction skill. Simulated year-to-year variations in June land conditions over North America show a significant lead correlation with the North Atlantic large-scale environment and TC activity. Improved land–atmosphere initialization appears to improve the Atlantic TC predictions initialized in some summer months. For short-lead predictions initialized in June, the potential skill improvements attributable to land–atmosphere initialization might be comparable to those achievable with perfect SST predictions. Overall, this study delineates the SST and non-oceanic error sources in predicting TC activity and highlights avenues for improving predictions. The nudging-based evaluation framework can be applied to other models and help improve predictions of other weather extremes.

KEYWORDS: Atmosphere-land interaction; Atmosphere-ocean interaction; Tropical cyclones; Climate prediction; Coupled models; Model evaluation/performance

1. Introduction

The recent development of high-resolution climate models has led to remarkable success in simulating and predicting tropical cyclone (TC) activity (e.g., Vitart and Stockdale 2001; Vitart 2006; Vitart et al. 2007; Camargo and Barnston 2009; Zhao et al. 2009; Vitart 2009; Chen and Lin 2013; Vecchi et al. 2014; Camp et al. 2015; Manganello et al. 2016, 2019; Murakami et al. 2016, 2018). However, many current dynamic models still have notable errors in simulating TC activity, such as biases in the spatial distribution of simulated TCs (e.g., Vecchi et al. 2014; Walsh et al. 2016; Camargo and Wing 2016) and occasional failures of seasonal predictions (e.g., Bell et al. 2014; Vecchi and Villarini 2014; Zhang et al. 2016). Notably, there is possibly a gap between the potential skill and the actual skill of TC seasonal predictions, especially in the coastal regions (Zhang et al. 2019). A better understanding of model errors and their sources could help to improve predictions of TC activity and facilitate applications in the economic and policy sectors.

Model errors in simulating TC activity stem from complex origins. Early global climate simulations of TC activity contained notable errors, such as a lack of TC-strength storms, and

these errors were mainly attributed to the low horizontal resolution (Walsh et al. 2016; Camargo and Wing 2016). As resolution increased, atmospheric models forced by observed sea surface temperature (SST) became skillful in simulating TC activity (e.g., Zhao et al. 2009; Manganello et al. 2012; Murakami et al. 2012). However, prescribing the SST forcing disrupts the atmosphere–ocean coupling, altering the large-scale convection patterns and the energy budget (Wittenberg and Anderson 1998; Emanuel 2010; He et al. 2018). Coupled climate models avoid these problems but tend to show climatological biases and/or poor simulations of regional TC activity (e.g., Camp et al. 2015; Manganello et al. 2016; Li and Srivier 2018), which are associated with systematic biases in simulating the upper ocean and large-scale atmospheric environment (e.g., Camargo 2013; Kim et al. 2014; Vecchi et al. 2014; Hsu et al. 2019). Analyses of high-resolution simulations also pointed to the role of parameterized atmospheric physics in simulations of TC frequency (Zhao et al. 2012) and TC intensity (Murakami et al. 2012; Kim et al. 2018; Wing et al. 2019). Yet the relative roles of atmosphere/ocean resolution, convective and cloud parameterizations, and emergent climatological biases in simulations of TC activity remain poorly understood.

Recent studies also point to the importance of land–atmosphere process and initialization for TC predictions. Idealized modeling studies have identified land conditions as a contributor to summertime variability of the extratropical atmosphere on both regional and global scales (Koster et al. 2016;

^d Current affiliation: Citadel Americas, LLC, Chicago, Illinois.

Corresponding author: Gan Zhang, ganzhang@princeton.edu

Xue et al. 2018; Teng et al. 2019). Land–atmosphere initial conditions affect predictions of the large-scale environment via land–atmosphere coupling and the seasonal-scale memory induced by land conditions such as albedo, snow cover, soil water, and soil temperature (e.g., Delworth and Manabe 1988, 1989; Koster et al. 2004; Santanello et al. 2018). That land–atmosphere initial conditions matter for seasonal predictions of the summertime extratropics is also supported by prediction experiments conducted at the National Centers for Environmental Prediction (NCEP) (Dirmeyer et al. 2018), the Geophysical Fluid Dynamics Laboratory (GFDL) (Jia et al. 2016), and several European modeling centers (Ardilouze et al. 2019). Such findings have interesting implications for predicting TC activity, as a growing body of evidence suggests that extratropical atmosphere variability modulates Atlantic TC activity, including storm tracks and genesis frequency (Kossin et al. 2010; Murakami et al. 2016; Zhang et al. 2016, 2017; Li et al. 2018; Zhang and Wang 2019). Murakami et al. (2016) analyzed retrospective seasonal predictions by a high-resolution GFDL model (Vecchi et al. 2014) and found that the coupled model lacked skill in predicting a key mode of the extratropical atmosphere in the Atlantic basin, possibly due to the use of a crude initialization procedure for the land and atmosphere components (Vecchi et al. 2014; Zhang et al. 2019).

To help understand error sources and explore avenues for improving predictions of TC activity, this study examines retrospective seasonal predictions (“hindcasts”) from a state-of-the-art operational model at GFDL and proposes an evaluation framework for coupled climate models. We use a “nudging” technique to relax model solutions toward specified observations of the climate system (e.g., the atmosphere or the ocean) while retaining relatively realistic interactions among model components. This technique helps to address the following research questions:

- 1) How do errors in individual model components affect simulations and predictions of TC activity?
- 2) How do non-oceanic sources of predictability—such as land–atmosphere initial conditions—affect simulations and predictions of TC activity?
- 3) What error sources should the community prioritize to improve the seasonal prediction of TC activity?

While the answers to these questions may be model dependent, our approach should aid analyses of other climate models, especially those used in operational predictions and climate projections, and inspire improvements of models and understanding. The rest of the paper is organized as follows. Section 2 introduces the model experiments, the data, and the methodology used in this study. Section 3 investigates errors in the atmosphere–ocean coupling and the parameterized physics, and their links to errors in TC predictions. Section 4 explores the potential impact of land–atmosphere coupling on predictions of TC activity. Section 5 concludes with a summary and discussion.

2. Data and methodology

a. Seasonal prediction experiments

The Forecast-Oriented Low Ocean Resolution (FLOR) model used in this study is a variant of the GFDL Climate

Model, version 2.5 (Delworth et al. 2012), with a lower-resolution ocean component (Vecchi et al. 2014). The horizontal grid spacings are $1^\circ \times 1^\circ$ for the ocean–ice components (telescoping to 0.33° meridional spacing near the equator), and about $0.5^\circ \times 0.5^\circ$ for the atmosphere–land components. These resolution choices help the model to simulate weather extremes (van der Wiel et al. 2016) while making real-time seasonal predictions feasible as part of the North American Multimodel Ensemble (Kirtman et al. 2014). The model configuration (e.g., parameterized physics) has been documented by Delworth et al. (2012) and Vecchi et al. (2014), and a detailed description of the land model (LM3) component has been provided by Milly et al. (2014). The model generates realistic simulations of tropical climate (Wittenberg et al. 2018; Ray et al. 2018a,b; Newman et al. 2018). It also skillfully predicts the El Niño–Southern Oscillation (ENSO), continental precipitation and temperature (Jia et al. 2015, 2016), mountain snowpack (Kapnick et al. 2018), the seasonal statistics of extratropical storms (Yang et al. 2015), and seasonal TC activity (Vecchi et al. 2014; Murakami et al. 2016, 2018; Liu et al. 2018).

This study focuses on retrospective predictions conducted with FLOR-FA, a version of FLOR that includes artificial air–sea “flux adjustments” to help reduce model drift and emergent biases (Stockdale 1997; Magnusson et al. 2013). The flux adjustments apply climatological corrections to the momentum, turbulent heat fluxes, and freshwater fluxes received by the ocean component, so that the model’s climatological wind stress, SST, and sea surface salinity better resembles the observed climatology. It is true that flux adjustments can distort tropical atmosphere–ocean interactions (e.g., Neelin and Dijkstra 1995) and cannot substitute for a long-term commitment to improving model physics (Shackley et al. 1999). However, flux adjustments can be a powerful tool for exploring the sources of model biases, and for understanding the impacts of model biases on simulated climate variability and extremes (e.g., Spencer et al. 2007; Manganello and Huang 2009; Ray et al. 2018b). Relative to FLOR, FLOR-FA shows improved simulations and predictions of TC activity (Vecchi et al. 2014; Krishnamurthy et al. 2016) and the ENSO’s teleconnections to North America (Krishnamurthy et al. 2015). At the specified model resolution, FLOR-FA delivers the most skillful seasonal TC predictions at GFDL.

Three sets of FLOR simulations are used in this study (Table 1). The first set (“FA-Basic”) consists of FLOR-FA hindcast experiments following the flux-adjusted and retrospective seasonal predictions described by Vecchi et al. (2014). A total of 12 different realizations of FLOR-FA are started with 12 sets of initial conditions and integrated for 12 months each. The ocean–ice initial conditions are provided by a coupled data assimilation system with an ensemble Kalman filter (Zhang et al. 2007). The land–atmosphere initial conditions are acquired offline from a three-member ensemble of SST-forced simulations, which include land–atmosphere coupling but are not otherwise constrained by land–atmosphere observations. For FA-Basic, the initial SST states generally differ much less than 0.5 K across the ensemble members, while the initial near-surface air temperatures can differ by more than 5 K over land (supplementary information in Zhang et al. 2019). FA-Basic is designed to test the model response to accurate three-dimensional

TABLE 1. List of FLOR experiments, each of which spans the 1981–2014 time period.

Expt	Nature of simulation	Ocean–ice nudging and initial conditions	Atmosphere–land initial conditions	Ensemble size	Analyzed simulation years
Flux adjustment (FA-Basic)	Retrospective seasonal predictions, initialized in each of 7 months (January, . . . , July)	12 initializations from ensemble data assimilation, no nudging	Three initializations from offline SST-forced simulations	12	$12 \times 7 \times 34 = 2856$
Land–atmosphere initial conditions (FA-AL)	Same as FA-Basic	Same as FA-Basic	One initialization from a multiyear simulation with atmosphere nudged toward MERRA	12	$12 \times 7 \times 34 = 2856$
Restore SST (ReSST)	Multiyear climate simulation with SST restoring, initialized in January 1980	Two simulations with continuous nudging of SST toward HadISST.v1, using either a 5- or 10-day restoring time scale	Three initializations in January 1980 as in FA-Basic	6	$6 \times 34 = 204$

initialization of the ocean surface and subsurface, in the absence of atmospheric and land initialization.

The second set of FLOR-FA hindcast experiments (“FA-AL”) is identical to FA-Basic, except for the land–atmosphere initial conditions. As documented by Jia et al. (2016), the land and atmosphere initial conditions are acquired by conducting a multi-decade FLOR simulation that 1) restores the model SST toward the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST v1; Rayner et al. 2003) and 2) nudges the surface pressure and three-dimensional atmospheric variables (winds and temperature) toward the 6-hourly Modern Era-Retrospective Analysis for Research and Applications (MERRA) reanalysis (Rienecker et al. 2011). The nudging of the SST and atmosphere gradually adjusts the land conditions toward a relatively realistic state (section 4c), which would otherwise be difficult to acquire due to a lack of reliable long-term observational records of land properties. The nudging simulation, which consists of a single realization for 1979–2014, provides the land–atmosphere initial conditions for each ensemble member of FA-AL; i.e., for each prediction all the ensemble members of FA-AL share the same land–atmosphere initial conditions. Comparing FA-AL to FA-Basic highlights the potential role of accurate land–atmosphere initialization in the seasonal predictions.

The third set of FLOR simulations (“ReSST”) removes the flux adjustments but restores the SST toward interannually varying monthly mean observations and restores the sea surface salinity toward the monthly mean climatology from the *World Ocean Atlas 2005* (Antonov et al. 2006). The nudging *e*-folding time is set to either 5 or 10 days for the top 10-m ocean layer, so that the simulated SST closely follows observations. Compared to an SST-forced atmosphere-only simulation, the coupled nudging allows weather systems (e.g., TCs) to interact with the ocean in a more realistic way. The choice of two nudging time scales (5 or 10 days), combined with three different choices for land–atmosphere initial conditions as in FA-Basic, generates a six-member ensemble initialized on 1 January 1980 that is run continuously afterward. As in FA-Basic, the land–atmosphere

initial conditions used by the ReSST simulations are unconstrained by observations. ReSST is designed to estimate the upper limit of atmospheric and land prediction skill, given perfectly predicted SSTs. Comparing FA-Basic and ReSST with observations can help to disentangle TC errors arising from the ocean–ice evolution and other model processes (e.g., parameterized atmospheric physics).

The FA-Basic and FA-AL retrospective predictions each consist of 12 ensemble members that are initialized at 0000 UTC on the first day of each calendar month and then run for 12 months. Unless otherwise specified, our analysis is based on monthly mean fields from each of the simulations in Table 1 and focuses on July–November, when the Northern Hemisphere produces the majority of its TCs (Schreck et al. 2014). We examine the FA-Basic and FA-AL predictions that are initialized from January–July of 1981–2014, respectively. The six ReSST simulations are not initialized predictions, and so we simply examine those simulations during July–November of 1981–2014.

b. Observational and model data

We acquire the observational data of TC activity (1981–2014) from the International Best Track Archive for Climate Stewardship (IBTrACS v03r09), which is organized based on individual basins (Knapp et al. 2010; Schreck et al. 2014). The simulated TCs are tracked using an algorithm that detects warm-core storms that meet certain criteria of duration and near-surface wind speed, as described by Murakami et al. (2015) and Harris et al. (2016). In short, the algorithm identifies closed contours of negative sea level pressure anomalies that have a warm-core structure. To qualify as a TC, a storm candidate must maintain a warm core and sufficiently strong wind ($>16.5 \text{ m s}^{-1}$) for at least 36 consecutive hours. To analyze the spatial distributions of TCs in the predictions, during the July–November TC season we calculate the number of days when TCs are present within a $5^\circ \times 5^\circ$ grid box in the Northern Hemisphere. We also characterize basinwide TC activity using the seasonal TC number and the accumulated cyclone energy

(ACE), which is an approximation of the wind energy over the lifetime of a TC (Bell et al. 2000). FLOR is skillful in predicting the year-to-year variations of TC number and ACE, but this model substantially underestimated the climatological values of these two metrics (Vecchi et al. 2014; Zhang et al. 2019). For applications, these negative biases are a posteriori and can be “corrected” using statistical methods; but here we focus on the original model outputs to address the model deficiency directly and keep the result interpretation straightforward. Interested readers can find the TC number and ACE from selected FLOR simulations in appendix A (Figs. A1 and A2).

We use the first-generation MERRA reanalysis (Rienecker et al. 2011) as the nudging target for atmospheric fields when generating the initial conditions for the FA-AL predictions. For all the FLOR experiments, the ocean–ice targets for nudging and evaluation are from the HadISST v1 (Rayner et al. 2003), and the National Oceanic and Atmospheric Administration (NOAA) weekly optimum interpolation SST analysis (OISST v2) (Hurrell et al. 2008). The FLOR simulations and predictions are also evaluated against the MERRA-Land reanalysis (Reichle et al. 2011) and atmospheric data from the ERA-Interim reanalysis (Dee et al. 2011). All the observational and reanalysis products, except for the land data analyzed in section 4, are averaged on a $2.5^\circ \times 2.5^\circ$ grid for comparison with FLOR.

c. Resampling analysis of ensemble simulations

Following Zhang et al. (2019), we use a bootstrap resampling technique to characterize the role of initial condition uncertainty in the predictions. We denote 12-month prediction segments as $S_{i,j,k}$, where i is an ensemble index (1, . . . , 12), j is the initialization month (1, . . . , 7), and k is the initialization year (1981, . . . , 2014). Since these 12-month segments with different initialization months and years are approximately independent predictions, we can randomly select i for each year to assemble a 34-yr model prediction. Repeating this selection N times produces an N -member ensemble prediction, which we use to account for the role of ensemble size in prediction skill (Manganello et al. 2016; Mei et al. 2019). Although both the FA-Basic and the FA-AL ensembles consist of 12 members, we set $N = 6$ to match the size of the ReSST ensemble. For FA-Basic and FA-AL, we repeat the random selection 1000 times to generate 1000 six-member 34-yr bootstrap ensembles, each of which is analyzed separately. This bootstrap resampling helps to illustrate the uncertainty ranges related to the unforced variability in model predictions.

The bootstrap resampling also helps with the test of statistical significance. For example, the difference between two populations of climatology (or correlation coefficient) can be estimated using the bootstrapping ensembles; if the distribution of this difference suggests that a null hypothesis (H_0 : the difference is zero) can be rejected within the 2.5th–97.5th percentile range, the difference is considered to be statistically significant at the 95% confidence level. This significance test differs from some variance-based parametric methods, which are sensitive to assumptions of data statistical distributions and tends to underestimate the statistical significance of sample differences (DelSole and Tippett 2014). While alternative tests are available (Hamill 1999; DelSole and Tippett 2014, 2016),

those tests generally emphasize square errors that can be dominated by climatological biases, and thus work better with predictions that have few biases or have been bias-corrected (appendix B). Given our interest in year-to-year variations of the original model predictions, we illustrate uncertainty ranges and conduct significance tests by consistently using the bootstrap resampling approach. The potential limitations of this approach are further discussed in appendix B.

3. Impacts of oceanic and non-oceanic errors

a. Climatological biases in TC activity and large-scale environment

Figure 1 shows the climatological TC days of the observation, FA-Basic, and ReSST. Compared to the observations the July–November TC activity predicted by FA-Basic is too high in the northwest tropical Pacific, especially in the monsoon trough region near 20°N . Meanwhile, weaker but notable negative biases are present in the northeast tropical Pacific. Although some of these biases may be associated with SST biases, similar TC biases are also present in the northwest tropical Pacific in the ReSST simulation (Fig. 1b), where the SST conditions are nudged toward observations. In the northwest tropical Pacific, the positive TC biases in ReSST are even larger than in FA-Basic (Fig. 1c), suggesting that these biases stem mostly from atmospheric processes. In the northeast tropical Pacific, ReSST’s warmer and more realistic SST increases the TC activity relative to FA-Basic, eliminating the negative biases near 20°N but resulting in positive biases at lower latitudes. Overall, the climatological biases in FA-Basic involve compensating errors related to oceanic and non-oceanic factors. The role of SST biases is not necessarily dominant but certainly important.

Figure 2 shows time–longitude Hovmöller plots of the near-equatorial SST. Partly owing to the flux adjustments (Vecchi et al. 2014), FA-Basic produces a realistic annual cycle of equatorial SST in all three ocean basins. However, cold biases develop in the eastern Pacific shortly after the model is initialized in July. By early September, the cold bias exceeds 1.2 K and becomes comparable to the year-to-year variations of local SST. The relatively cold equatorial SST affects the atmospheric environment in nearby tropical regions, reducing TC activity on the central-Pacific flank of the northwest tropical Pacific (approximately $5^\circ\text{--}20^\circ\text{N}$, $150^\circ\text{E--}180^\circ$) (e.g., Wang and Chan 2002; Camargo et al. 2007) and northeast tropical Pacific (approximately $5^\circ\text{--}20^\circ\text{N}$, $110^\circ\text{--}150^\circ\text{W}$) (e.g., Camargo et al. 2007; Jien et al. 2015). The differences between FA-Basic and ReSST (Fig. 1c) are consistent with a suppression of TC activity by the cold SST biases in FA-Basic. Compared to observations, FA-Basic also shows a slight westward displacement of the interannual variability of equatorial SST toward the central Pacific (Figs. 2a,b), which then affects the Pacific and Atlantic TC activity (Kim et al. 2009, 2011; Patricola et al. 2018).

Figure 3 shows the climatological biases of the simulated large-scale environment relative to observations. Perhaps surprisingly, notable SST biases are present in both FA-Basic and ReSST. Relatively large SST biases occur where strong atmosphere–ocean interactions take place, such as in the equatorial Pacific in FA-Basic (Fig. 3a), and near the

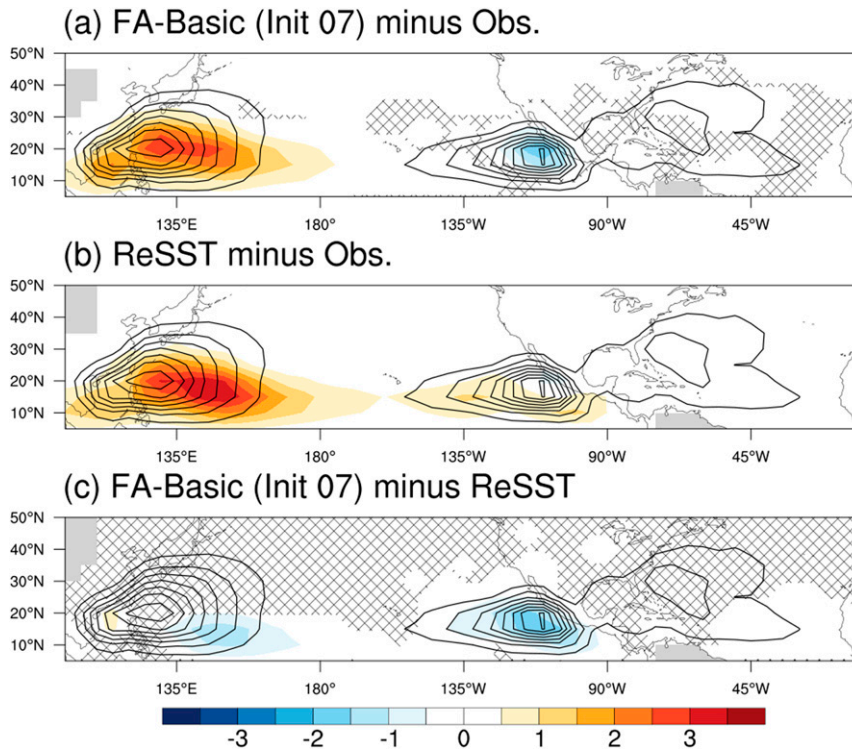


FIG. 1. Climatology of observed and the simulated TC days during July–November. (a) FA-Basic biases, (b) ReSST biases, and (c) their difference (FA-Basic minus ReSST, which indicates the impact of SST biases in FA-Basic). The black contour is the observed TC days; the contour interval is 0.5 days during a 5-month period. Color shading represents biases in (a) and (b) or differences of simulations in (c). The FA-Basic prediction examined here is initialized in July. In (a) and (c), hatching indicates that the differences are below the 95% confidence level tested using a bootstrapping test (section 2c). Statistical significance was not tested for ReSST because the inter-year dependence in each ReSST ensemble makes it improper to resample the climatology using the method described in section 2c. Gray shading indicates regions where TC observations are unavailable.

midlatitude western boundaries of the oceans in both sets of simulations (Figs. 3a,b). In contrast to the FA-Basic predictions, the free-running FLOR-FA climate simulation does not show strong climatological biases along the equator (Fig. 2 of Vecchi et al. 2014; Fig. 1 of Ray et al. 2018b). Furthermore, the SST biases in ReSST suggest that some error sources in this model are incompletely compensated by ReSST's 5–10 day relaxation of the SST toward observations. FA-Basic and ReSST show similar precipitation biases, some of which arise from SST biases. Compared to ReSST, the tropical Pacific precipitation in FA-Basic shows a subtle westward shift that leads to larger precipitation biases near 135°E. For 500-hPa geopotential height, ReSST reduces the weak negative biases of FA-Basic in the tropics, but the biases associated with the midlatitude circulation mostly remain the same. For vertical wind shear, the impact of the SST biases is most notable in the low-latitude parts of the eastern Pacific basin but is weak elsewhere.

An analysis of the predictions initialized in the earlier months reveals climatological biases that are highly similar to those in Figs. 1–3. For brevity, we omit those results except for adding a remark on the North Atlantic. Compared to predictions initialized in July (Fig. 3), the tropical North Atlantic in

predictions initialized in earlier months have slightly warmer SST (~ 0.5 K) and weaker vertical wind shear (~ 2 m s $^{-1}$) (not shown). These weak differences introduce a positive bias in TC days to the western part (45°–90°W) of the tropical North Atlantic. Nonetheless, the magnitude of this positive bias is comparable to the eastern Pacific TC bias in Fig. 1, underlining the sensitivity of TC climatology to environmental biases. Overall, the results suggest that SST biases contribute to the climatological biases of the tropical large-scale environment, implying that an improved simulation of SST could potentially contribute to a more realistic simulation of TC climatology. However, SST errors are not the primary cause of some of FA-Basic's most notable biases, such as its excessive precipitation and TC activity in the northwest tropical Pacific. In other words, these biases appear to stem from non-oceanic error sources, such as the parameterized atmospheric physics. Pinpointing the underlying physical processes of these model errors will be pursued in future studies.

b. Skill gaps associated with SST errors

This section examines to what extent FA-Basic's SST prediction errors affect its skill in predicting seasonal TC activity. Figures 4a–c show the seasonal prediction's correlations with

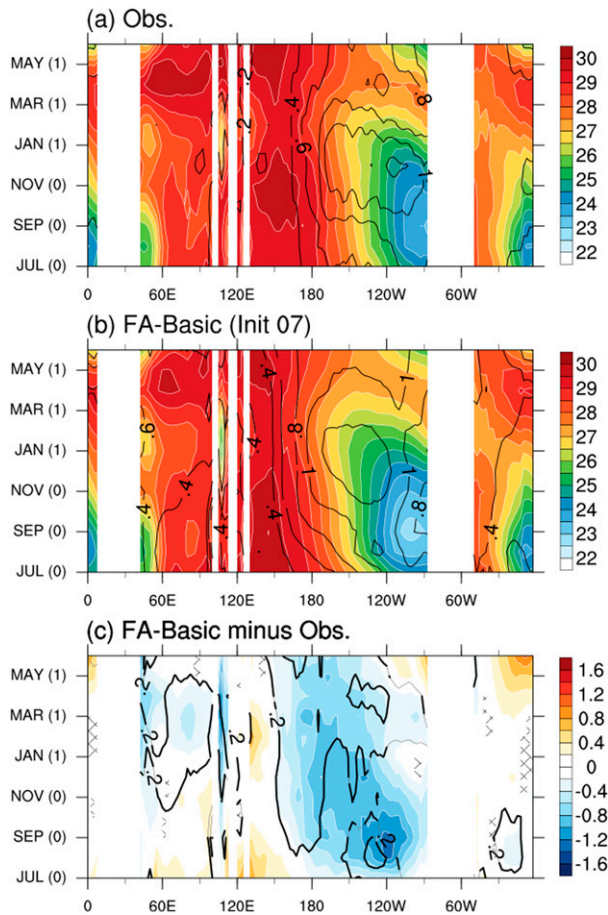


FIG. 2. Hovmöller diagram of climatological SST (K) averaged over 5°S–5°N. (a) Observations, (b) FA-Basic initialized in July, and (c) their difference (FA-Basic minus observations). The ordinate shows the verification month, with JUL (0) indicating July in the year of initialization, and MAY indicating the May of the following year. SST means and standard deviations are represented with shading and black contours, respectively. In (c), differences are tested with bootstrapping (section 2c), and the parts below the 95% confidence level are marked with hatching (SST means) or thin gray lines (SST standard deviations).

the observation (“skill” hereafter) for two metrics of TC activity, the basinwide TC number, and accumulated cyclone energy (ACE). The skill tends to be lower than that in a similar analysis by Zhang et al. (2019), as here the ensemble size has been reduced from 12 to 6 (section 2c). The prediction skill varies among individual basins, TC metrics, and prediction lead times. For FA-Basic, predictions are generally more skillful for ACE than TC number, especially in the Pacific. The high skill of FA-Basic’s short-lead ACE predictions is consistent with the potential skill indicated by ReSST, which is represented by long dashed lines in Figs. 4a–c and shows strong correlations with observations in all the three basins ($r^2 \approx 0.5$). It is unclear whether in the Pacific the TC number is inherently harder to predict than ACE, as the model has pronounced biases in the Pacific that likely affect TC genesis (section 3a). Figures 4a–c also suggests that biases in the SST predictions limit the skill of predicted TC activity, since the FA-Basic predictions are less skillful than the

ReSST simulations. However, as the prediction lead time shortens, the skill difference between FA-Basic and ReSST narrows or vanishes (e.g., North Atlantic predictions initialized in July), suggesting that an improved SST prediction might increase TC prediction skill more at long leads than short leads. But due to chaotic dispersion, the SSTs are inherently more difficult to predict at long leads, potentially limiting the attainable prediction skill for TC activity.

Figures 4d–f show the root-mean-square error (RMSE) of FA-Basic. Compared to the correlations in Figs. 4a–c, the RMSE is not strongly dependent on the initialization month, especially in the Pacific basin. A notable exception is the North Atlantic, the RMSE of which increases as the lead time of predictions decreases. As shown by previous analyses of FA-Basic (Zhang et al. 2019), the short-lead predictions of the North Atlantic TC metrics have larger negative biases. The climatological biases can dominate the RMSE if these biases exceed the magnitude of year-to-year variations. We found this is indeed the case for the short-lead predictions of the North Atlantic. Such large negative biases also exist in the Pacific basins, consistent with the fact that the atmospheric resolution and parameterized physics of FLOR-FA are inadequate to simulate intense TCs. While this bias–RMSE issue is less severe for the long-lead predictions of the North Atlantic TCs (Fig. 4d and Zhang et al. 2019), a comparison between FA-Basic and ReSST reveals a caveat for interpreting these low RMSE values. With nearly perfect SST, ReSST produces much larger RMSE values than FA-Basic. This counterintuitive result suggests that the small RMSE in the long-lead predictions of the North Atlantic TCs arises by chance from compensating model errors.

Figure 5 shows the skill of regional July–November TC predictions initialized in January, April, and June. The prediction skill for FA-Basic increases at the shorter lead times, drawing closer to the potential skill of ReSST. This skill increase tends to be greatest over the open ocean, consistent with previous findings that TC tracks over the open ocean are easier to predict than those in the coastal regions (e.g., Zhang et al. 2019). As SST errors have relatively strong impacts on the short-lead North Atlantic predictions initialized in June (Figs. 4a and 5c), we show the skill of the prediction initialized in this month rather than July. Figure 5c suggests that SST errors in the June-initialized prediction still undermines the prediction of TC activity in certain regions (e.g., the subtropical east Atlantic and east Pacific). But interestingly, the more realistic SST in the ReSST significantly worsens the short-lead predictions of TC activity near Taiwan, the western coast of Mexico, and the U.S. East Coast (Fig. 5c). While some skill decrease might arise from randomness (5% probability) and could change if the ensemble size or the analysis period increases, we speculate that the skill decrease in coastal regions may indicate a deficiency in the model or TC tracking. Overall, these results suggest that the relatively low skill of FA-Basic in predicting TC days in coastal regions does not simply arise from SST errors. This suggests that future improvements in FA-Basic’s seasonal predictions of landfall activity will require not just an improvement in the predicted SSTs, but also an improvement in the atmospheric response to those SSTs.

We further examine the SST prediction errors and their impacts on the prediction of TC-related environmental

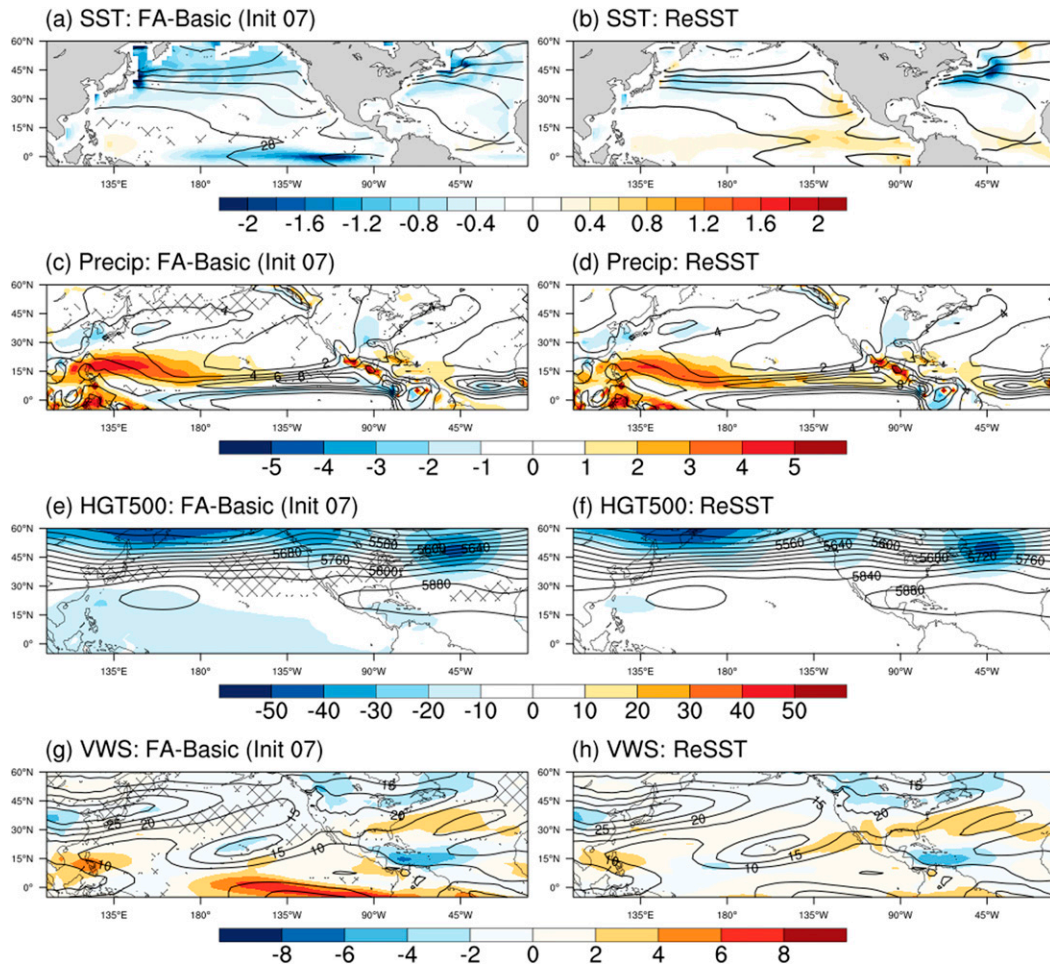


FIG. 3. Climatological (July–November) biases in (left) FA-Basic and (right) ReSST. (a),(b) SST (K); (c),(d) precipitation (mm day^{-1}); (e),(f) 500-hPa geopotential height (m); and (g),(h) vertical wind shear (m s^{-1}) between 200 and 850 hPa. Black contours show observations (or reference), and color shading shows the simulation biases. The FA-Basic hindcasts are initialized in July. The FA-Basic biases marked with hatching are below the 95% confidence level based on a bootstrapping test (section 2c). This significance test was not conducted for ReSST because the inter-year dependence in each ReSST ensemble makes it improper to resample the climatology using the method described in section 2c.

variables (Fig. 6). While FA-Basic is generally skillful at predicting tropical SSTs, its predictions of the northwest Pacific warm pool, the equatorial Atlantic, and the midlatitudes is much less skillful ($r < 0.6$; Fig. 6a). Figure 6b suggests that if the SST could be better predicted (ReSST), that could improve FA-Basic's prediction of tropical precipitation, even for the short-lead predictions initialized in June. But for extratropical precipitation, the impact of reducing SST errors is mixed and often not robust. Nonetheless, reducing the SST errors may benefit the predictions of the large-scale circulation, as suggested by 500-hPa geopotential height (Fig. 6c) and vertical wind shear (VWS; Fig. 6d). Most of the potential skill gains are expected in the tropics and subtropics. For example, the potential improvement in the VWS prediction is substantial in the North Atlantic basin, especially in the main development region of TCs, the Caribbean, and part of the subtropics.

Significant skill changes also appear in the extratropics (e.g., geopotential height near 45°N , 50°W), possibly associated with tropical–extratropical teleconnections and the wave–mean flow feedback in the midlatitudes. If paired with realistic parameterized atmospheric physics, the potential improvements of the large-scale environment might benefit the prediction of TC activity (e.g., lifetime and intensity). When FA-Basic is initialized at longer leads (figures not shown), the prediction of the tropical SST and related atmospheric environment is less skillful (e.g., due to the well-known spring barrier for ENSO predictions), underlining the importance of understanding and reducing SST prediction errors (Fig. 4).

Overall, the results in this section suggest that better predictions of SST could aid seasonal predictions of TC activity, especially at longer leads. However, better SST predictions might not substantially improve short-lead predictions, partly

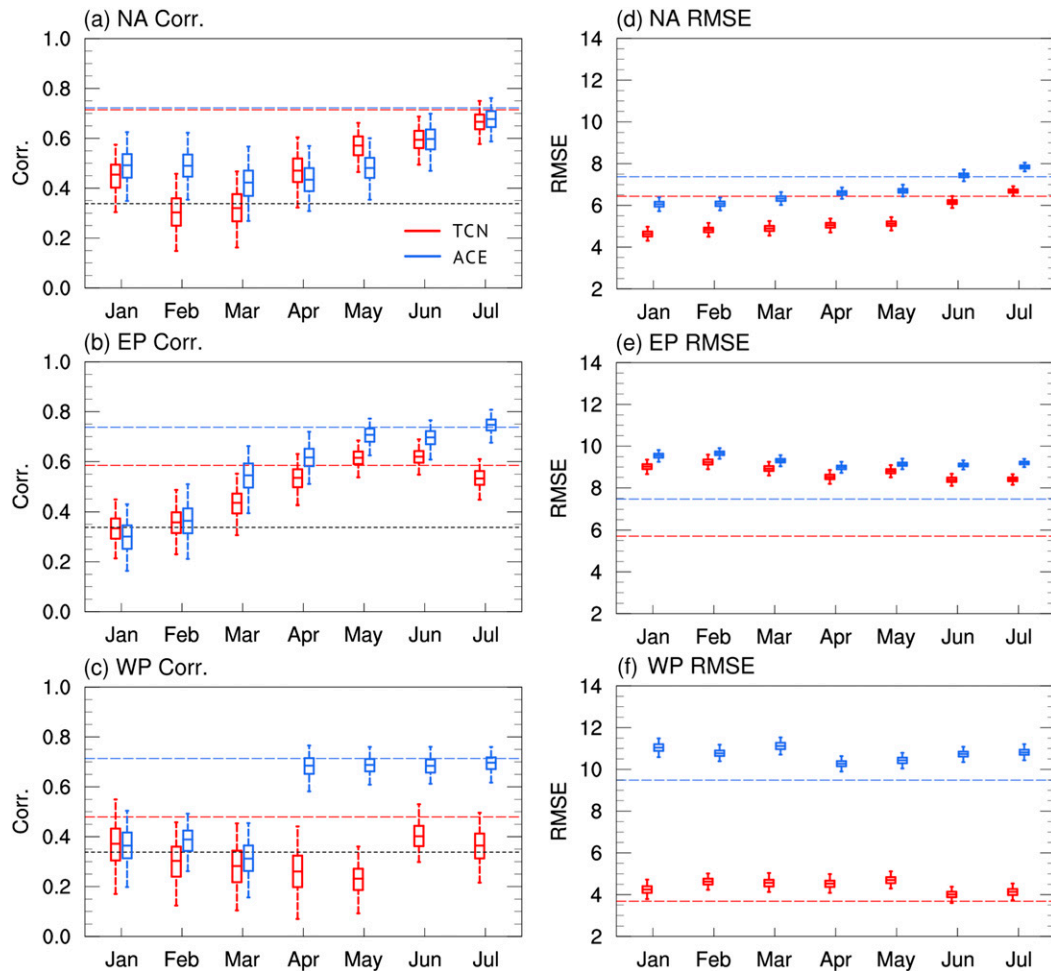


FIG. 4. Correlations of TC activity metrics (July–November) between observations and FA-Basic predictions initialized in January, February, . . . , and July, for the (a) North Atlantic, (b) northeast Pacific, and (c) northwest Pacific. (d)–(f) As in (a)–(c), but for root-mean-square error (RMSE). The basin definitions follow the IBTrACS convention (Fig. 1 in Schreck et al. 2014), and the northeast and northwest Pacific are separated at the date line. The boxplots denote 2.5th, 25th, 50th, 75th, and 97.5th percentiles of the bootstrap ensembles. Colors indicate results for TC number (TCN; red) or accumulated cyclone energy (ACE, blue). The RMSE of ACE is scaled with a factor of 1/10 for the convenience of illustration. Black short-dashed lines in (a)–(c) indicate a reference correlation coefficient at the 95% confidence level based on a two-tailed Student's t test. Long-dashed horizontal lines (blue and red) show the metrics of ReSST. The ReSST metrics do not vary with the initialization month because each ensemble member of ReSST is a continuously integrated climate simulation rather than predictions initialized in each month (Table 1). This model characteristic also prevents estimating the uncertainty range of the ReSST metrics using the bootstrap reampling technique described in section 2c. The statistical significance of the differences between ReSST and FA-Basic can be inferred by comparing the long-dashed horizontal lines and the whiskers of boxplots.

due to the remarkable SST prediction skill that has already been achieved. Perhaps more importantly, improved SST predictions do not seem to guarantee better predictions of TC activity in coastal regions either. To further improve seasonal TC predictions, it might be necessary to look beyond SST.

4. Impacts of land–atmosphere initial conditions

While the benefit of an accurate representation of atmospheric initial conditions is easy to anticipate, the potential impact of land initial conditions has not received much

attention in the context of TC prediction. Nonetheless, recent idealized modeling studies suggest that the land state of the United States affects precipitation and extratropical circulation on both local scales (Koster et al. 2014, 2016) and global scales (Teng et al. 2019). Meanwhile, an increasing number of studies suggest that the extratropical circulation can affect Atlantic TC behavior, including the TC tracks (Kossin et al. 2010; Murakami et al. 2016), motion speed (Zhang et al. 2019, 2020), and seasonal counts (Zhang et al. 2016, 2017). It appears plausible that land–atmosphere coupling over the continental United States may affect Atlantic TC activity and its seasonal

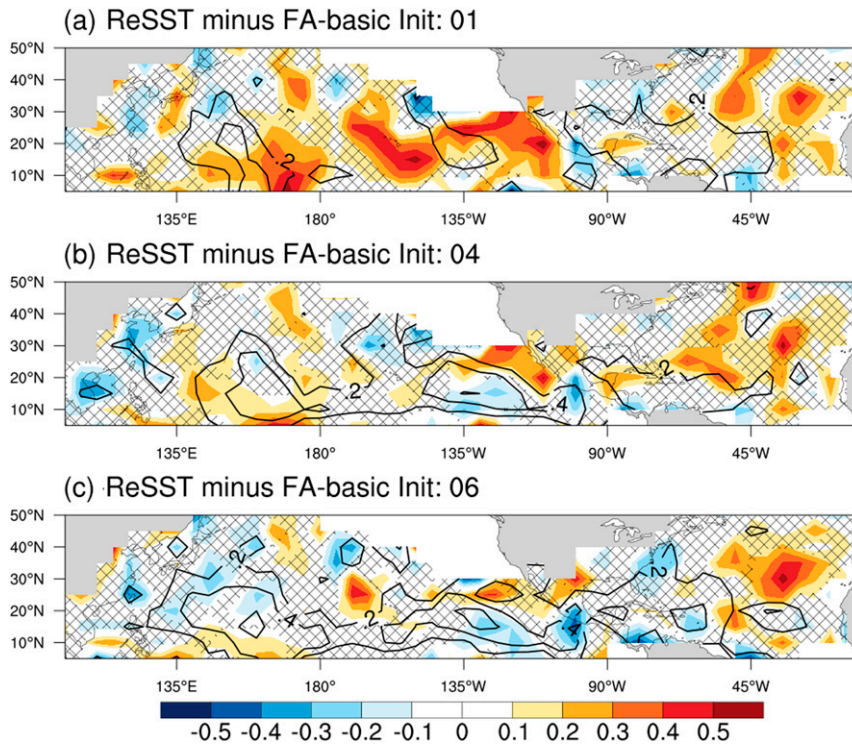


FIG. 5. Correlations between TC days (July–November) in the observation and FLOR simulations. Contours indicate the local correlation between the July–November TC days predicted by FA-Basic and those observed, for predictions initialized in (a) January, (b) April, and (c) June. Shading indicates the potential improvement in correlation skill achievable with a perfect SST prediction, estimated from the ReSST skill minus the FA-Basic skill. Hatching indicates correlation skill differences that are below the 95% confidence level based on a bootstrapping test (section 2c).

prediction. We next explore this possibility by analyzing the reanalysis data and FLOR hindcasts.

a. Potential associations between land conditions and Atlantic TC activity

To better leverage the previously discussed knowledge of land–atmosphere coupling and TC activity, the ensuing discussion focuses on North America and the North Atlantic. The focus is also motivated by the fact that the land–atmosphere coupling is particularly strong in North America during boreal summer (Koster et al. 2004; Santanello et al. 2018). We first characterize year-to-year variations of U.S. continental land conditions, using an empirical orthogonal function (EOF) analysis. Figure 7 shows the first two EOFs of land surface temperature (T-EOF) and root-zone soil moisture (Q-EOF), for the MERRA-land reanalysis data. For brevity, we only present the analysis for June, a month that features strong land–atmosphere coupling and corresponds to the early TC season for the Northern Hemisphere ocean basins. EOF1 shows a band-like pattern of temperature and moisture anomalies that extend from the western to the eastern coast of the United States, while EOF2 shows a dipole pattern associated with a “cool and wet” versus “warm and dry” contrast between the Pacific coast and the Gulf coast. However, the activity centers of these EOF modes are not perfectly aligned. For example, the

activity center of T-EOF1 is in the western United States, but the activity center of Q-EOF1 is east of the Rocky Mountains. Nonetheless, the soil moisture content east of the Rockies is much higher than that of the arid west (not shown), so the fractional changes of soil moisture are substantial in the west. All the EOFs have strong year-to-year variations (Figs. 7c,f). The pattern of Q-EOF2 appears consistent with a mode of moisture changes that contributes to drought development in the Southern Great Plains (Seager et al. 2019). Interestingly, Q-EOF2 shows a statistically significant upward trend during 1981–2014, which might be associated with land use changes or interdecadal climate variability.

Table 2 shows correlations among these North American land EOFs, an extratropical Atlantic variability index, and the number of North Atlantic TCs. The temperature and moisture EOFs show a strong positive correlation with each other, suggesting a coupling leading to either “warm and dry” or “cool and wet” conditions. EOF2 is also anticorrelated with a July–September index of extratropical variability, namely the number of anticyclonic Rossby wave breaking (RWB) events over the northwestern Atlantic (Zhang et al. 2016, 2017). RWB events tend to suppress Atlantic TC activity (e.g., Zhang et al. 2016, 2017; Li et al. 2018), so the RWB wave index RWBw is significantly anticorrelated with TC number in July–September ($r = -0.51$ during 1981–2014; also see Zhang et al. 2017). On synoptic time scales, RWB events are associated with

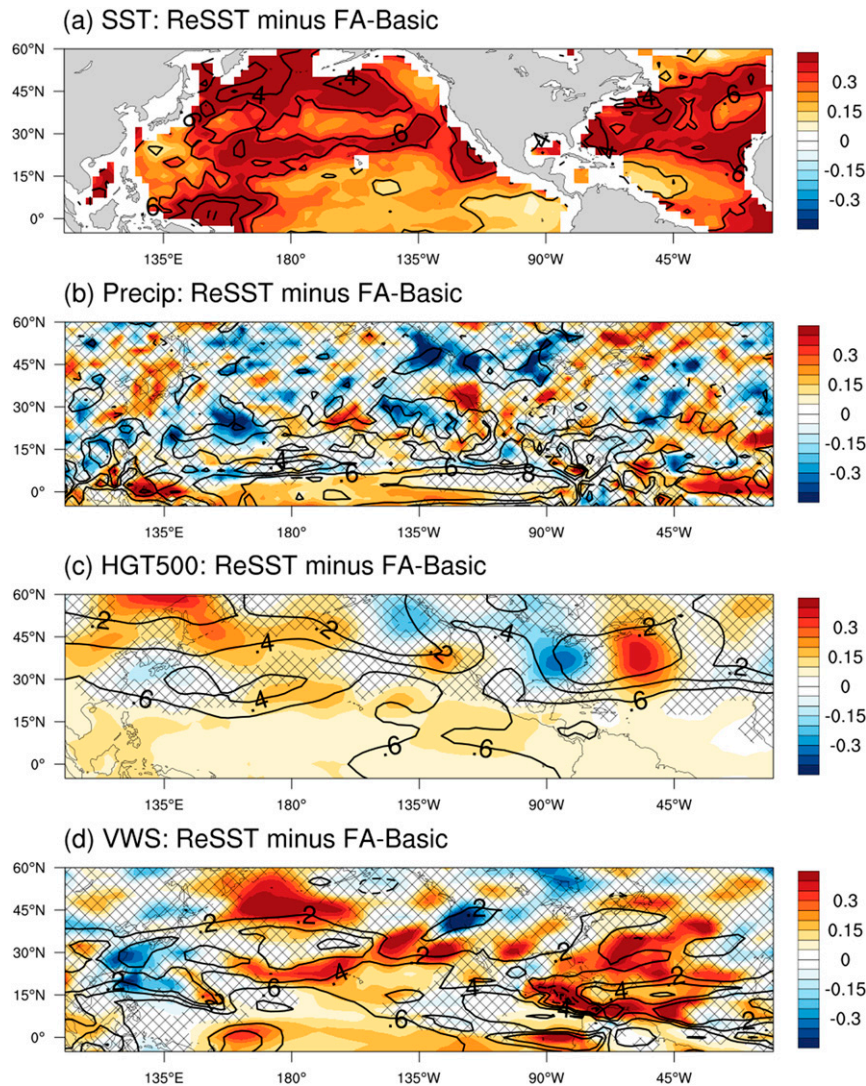


FIG. 6. Local correlations (black contours) of large-scale environment variables (averaged July–November) between observations and the FA-Basic predictions initialized in June for (a) SST, (b) precipitation, (c) 500-hPa geopotential height, and (d) vertical wind shear between the 200- and 850-hPa levels. The solid and dashed lines show positive and negative values, respectively; the correlation contour interval is 0.2, except near the zero line that is omitted. Color shading indicates the potential improvement in correlation skill achievable with a perfect SST prediction, estimated from the ReSST skill minus the FA-Basic skill. Hatching indicates correlation skill differences that are below the 95% confidence level based on a bootstrapping test (section 2c).

equatorward propagation of extratropical Rossby waves, which can be traced back to convection east of the Rocky Mountains ($\sim 90^\circ\text{W}$) (Zhang and Wang 2018). Such convection tends to be less active when land conditions are warm and dry (Findell and Eltahir 2003a,b; Findell et al. 2011; Koster et al. 2016; Santanello et al. 2018), and its impacts on the downstream extratropical circulation have been demonstrated in modeling studies (Jia et al. 2016; Teng et al. 2019). Consistent with this land–atmosphere coupling, T-EOF2 and Q-EOF2 are significantly correlated with the TC number in July–September ($r = 0.44$ and 0.63 , respectively), suggesting that a warm and dry June near the

Gulf coast tends to precede an active North Atlantic TC season. The variability of the extratropical atmosphere may also be subject to influences from land conditions of the western United States (Koster et al. 2016; Teng et al. 2019), despite the local land–atmosphere coupling being relatively weak (Findell and Eltahir 2003b; Dirmeyer 2011).

We further examine the association between the land EOFs and the large-scale circulation characterized by geopotential height (Fig. 8). As the memory of land conditions is roughly three months (section 4b), the analysis here focuses on June–September and does not include later months. Given the strong

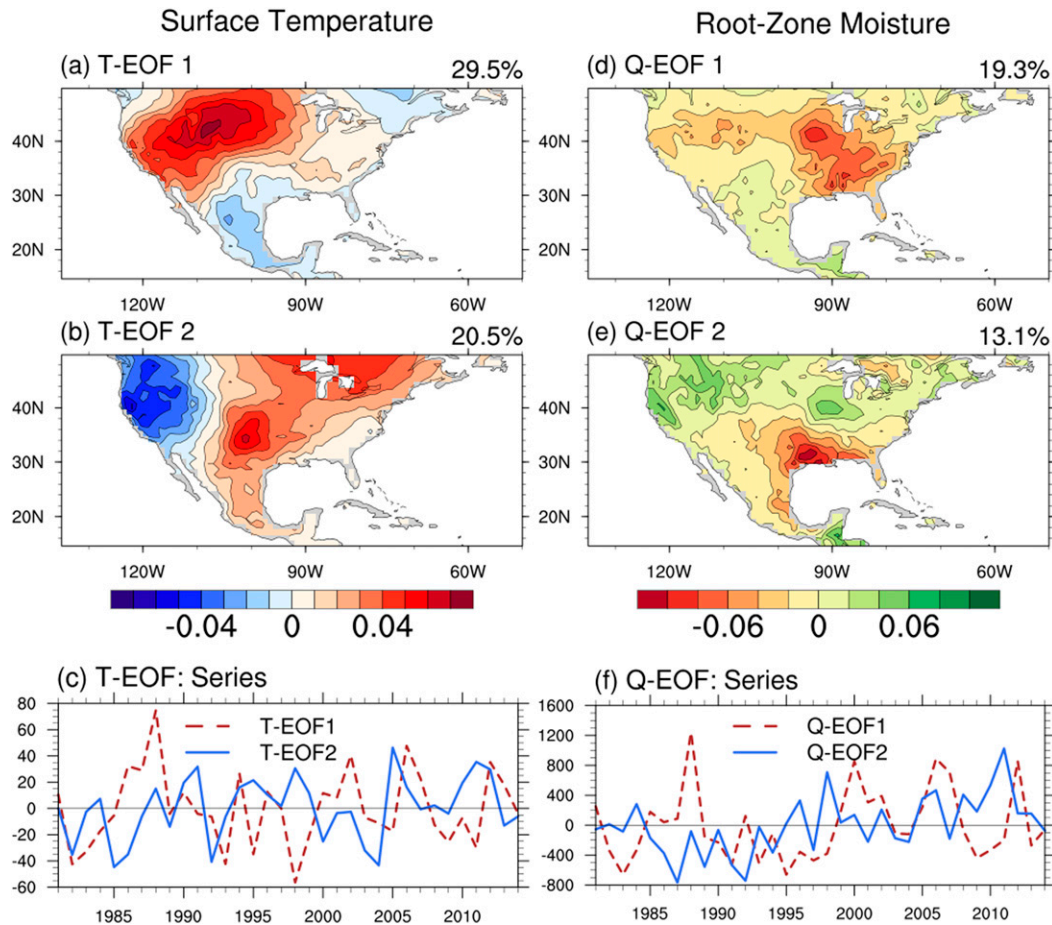


FIG. 7. Leading EOFs of June land surface temperature (K) and June root-zone soil moisture content ($\text{m}^3 \text{m}^{-3}$; water volume divided by soil volume) in the MERRA-Land reanalysis. The EOF patterns of temperature and moisture are denoted as T-EOF and Q-EOF, respectively. (a) T-EOF1, (b) T-EOF2, (d) Q-EOF1, and (e) Q-EOF2. The EOF patterns are normalized, and the year-to-year variance explained by each EOF is labeled in the top-right corner of each panel. The associated amplitude time series are also shown for (c) temperature and (f) moisture, and the means of each time series have been subtracted to better illustrate year-to-year variations. The trend in the Q-EOF2 time series is statistically significant at the 99% confidence level based on a bootstrapping test (section 2c).

correlations between the T-EOFs and Q-EOFs (Table 2), the correlation map for T-EOF1 resembles that for Q-EOF1 at the same geopotential height, and similarly for T-EOF2 and Q-EOF2. For brevity, we present and discuss the correlations between the T-EOFs and 850-hPa geopotential height, and the correlations between the Q-EOFs and 200-hPa geopotential

height. The 850-hPa geopotential heights are significantly correlated with T-EOF1 over the western United States and adjacent regions, and with T-EOF2 across the western hemisphere. This correlation pattern suggests a possible interbasin connection linking the North Pacific to the North Atlantic, though without a clear wave train pattern. Additionally, 200-hPa

TABLE 2. Correlations of land EOFs, extratropical variability, and TC activity. Correlation coefficients in bold exceed the 95% confidence level based on a two-sided t -statistics test. RWBw is an index of Rossby wave breaking over the northwestern Atlantic (Zhang et al. 2017), and TC Num is the number of North Atlantic TCs. The EOFs are derived using June data, while RWBw and TC Num are calculated using July–September data.

	T-EOF1 (June)	T-EOF2 (June)	Q-EOF1 (June)	Q-EOF2 (June)
T-EOF2 (June)	0.00			
Q-EOF1 (June)	0.73	−0.01		
Q-EOF2 (June)	−0.25	0.50	0.00	
RWBw (JAS)	0.06	−0.57	−0.21	−0.41
TC Num (JAS)	−0.05	0.44	0.20	0.63

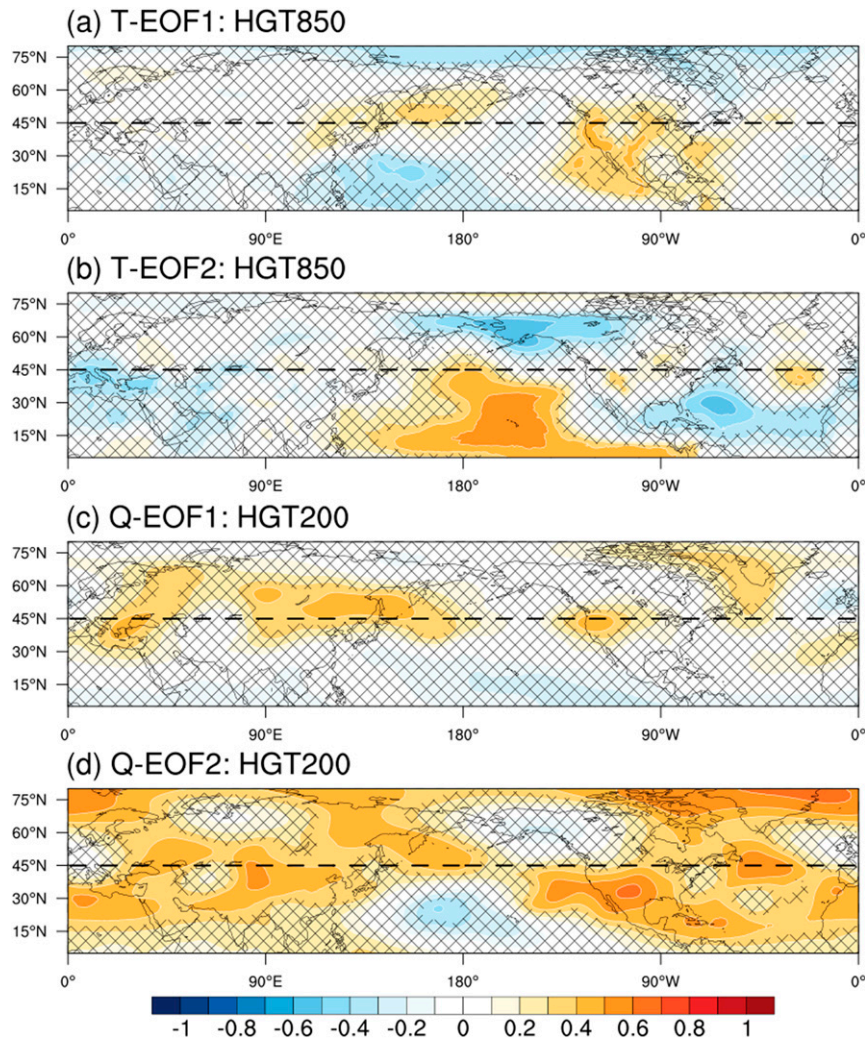


FIG. 8. Correlations between geopotential height (June–September) and the land EOFs of Fig. 7. The panels show correlations (a) between T-EOF1 and 850-hPa geopotential height, (b) between T-EOF2 and 850-hPa geopotential height, (c) between Q-EOF1 and 200-hPa geopotential height, and (d) between Q-EOF2 and 200-hPa geopotential height. Hatching indicates correlations below the 95% confidence level based on a two-tailed t -statistics test. The black dashed line at 45°N indicates the southern boundary in Fig. 5 of Teng et al. (2019).

geopotential heights are positively correlated with Q-EOF1 over the U.S. Pacific Northwest, consistent with simulated atmospheric responses when dry anomalies are prescribed over the central United States (Koster et al. 2016; Teng et al. 2019). Q-EOF1 is also significantly correlated with 200-hPa geopotential height in some remote regions. The significant correlations are much more extensive for Q-EOF2, where strong midlatitude correlations appear over the western United States, North Atlantic, eastern Europe, and central–northeastern Asia. This pattern appears similar to the simulated summertime circumglobal patterns that can be excited by idealized drying of the central and southern United States (Teng et al. 2019). The correlations weaken when the EOF time series are detrended, but the overall correlation patterns remain similar, especially for 850-hPa geopotential height (figure not shown). Overall, the lead–lag

correlations suggest that land–atmosphere coupling may play some active role in modulating the observed atmospheric variability.

b. Land initialization and land–atmosphere coupling

If land–atmosphere coupling modulates the large-scale environment, there are at least two necessary conditions to realize any related benefits in dynamical seasonal predictions. First, the model should be initialized with useful land information and retain the information for some additional time. Second, the model should characterize the land–atmosphere coupling in a relatively realistic way. Here we will explore whether these conditions are true in the FLOR hindcasts. Before introducing the results, we emphasize that the variables from the MERRA-Land reanalysis and the FLOR model are not perfectly comparable, partly because of different settings

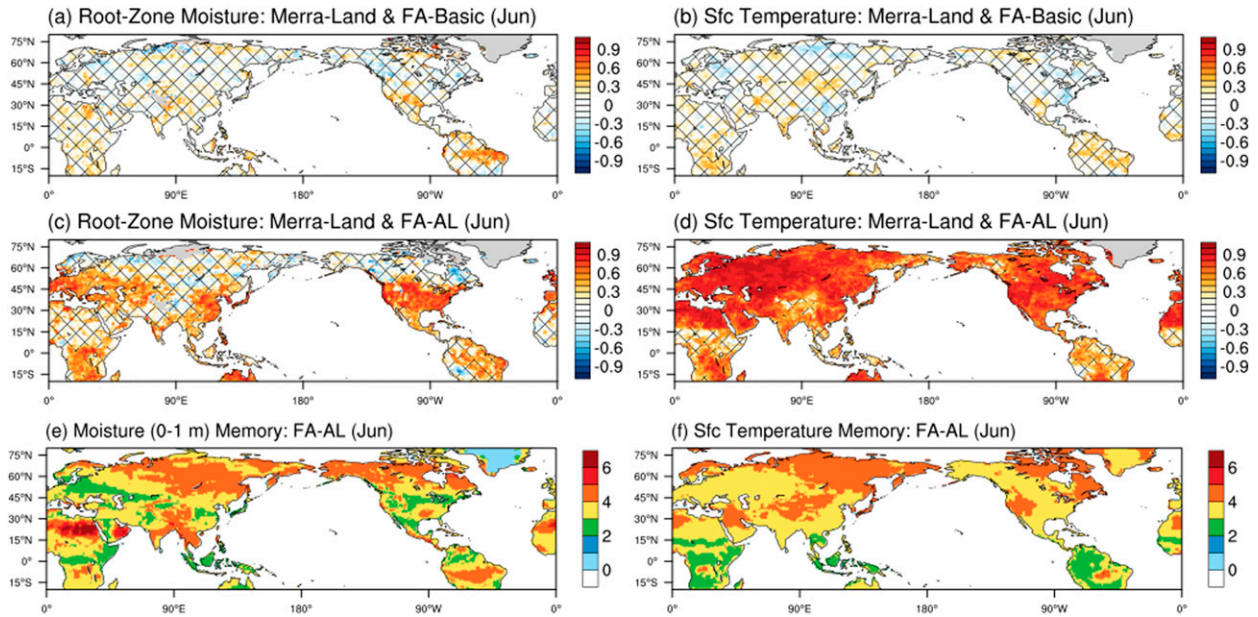


FIG. 9. The initial state and memory of land conditions in the FA-Basic and FA-AL hindcasts. (a) Correlation between the soil moisture content of MERRA-Land and FA-Basic at around 0000 UTC 1 Jun. (b) As in (a), but for land surface temperature. (c),(d) As in (a) and (b), but for FA-AL. (e),(f) Memory of land moisture and surface temperature, defined as the e -folding decay time (month) for the local autocorrelation function calculated from the monthly mean prediction data. The hatching in (a)–(d) denotes correlations below the 95% confidence level based on a two-tailed t -statistics test. All the calculations are conducted using data on the same $1^\circ \times 1^\circ$ grid.

of their land models and data output routines. For example, the root zone in MERRA-Land is a nominal 1-m layer that can be shallower in some regions with bedrock; since this layer is treated differently by the FLOR, its moisture content is approximated using the liquid soil moisture in the 0–1-m layer. Nonetheless, we expect the following analysis to reveal qualitatively useful information about the FLOR hindcasts.

Figure 9 explores the consistency between MERRA-Land and the initial conditions for the FLOR hindcasts. For FA-Basic, the year-to-year variations of land initial conditions are not well correlated with MERRA-Land. Since the FLOR initial conditions are generated offline using SST-forced simulations, the poor correlations with the MERRA-Land suggest that the SST does not completely dictate the land state in the FLOR predictions, consistent with earlier studies (e.g., Dirmeyer et al. 2003; Seager et al. 2019). Therefore, the land conditions can potentially serve as a source of predictability that is relatively independent of the SST forcing. In comparison with FA-Basic, the year-to-year variations of FA-AL's land initial conditions are much more closely correlated with the MERRA-Land. For example, significant correlations of soil moisture content appear in North America, Europe, East Asia, and some regions in the Southern Hemisphere. Strong correlations of land surface temperature also appear in most regions of the Northern Hemisphere ($r > 0.6$). Additionally, the mean states of the FA-AL initial conditions also appear realistic, except that the western United States is substantially warmer than in MERRA-Land (not shown). Figures 9e and 9f suggest that the land initial conditions persist in the FA-AL predictions; the e -folding memory of June-mean land conditions is generally around three months in the Northern Hemisphere. Longer memory of soil

moisture is present in desert regions, possibly arising from persistent dry conditions. The land memory in the FA-Basic hindcasts is nearly identical, and the memory length also qualitatively consistent with the MERRA-Land data (not shown). The 3-month time scale suggests that land initial conditions in June may persist to September, possibly contributing to predictions of early season Atlantic TC activity.

A comparison of the land–atmosphere coupling in MERRA-Land and FA-AL is available in Fig. 10. Here we assess the two-legged coupling index proposed by Dirmeyer (2011), using the monthly data in July and September to represent the transition from summer to autumn. The coupling index (CI) is a product of the standard deviation (σ) of a predictor variable (e.g., soil moisture Q) times the linear regression slope of a predictand variable [e.g., surface latent heat flux (LHF)] with respect to the predictor variable:

$$CI = \sigma_Q \frac{dLHF}{dQ}. \quad (1)$$

Due to limited data availability, the analysis here focuses on the impact of soil moisture on surface latent heat flux (“terrestrial leg”), and the impact of surface latent heat flux on local precipitation (“atmospheric leg”). MERRA-Land suggests that the land–atmosphere coupling in the Northern Hemisphere is generally stronger in July than September. Some exceptions include the Pakistan–India border, near-equatorial Africa, and the southern part of North America, where the coupling strength is comparable in July and September. However, not all the strong coupling indicated by MERRA-Land is similarly represented by FA-AL. A notable example for the atmospheric leg is in Central Africa, where the coupling is likely too weak in FA-AL. This issue might

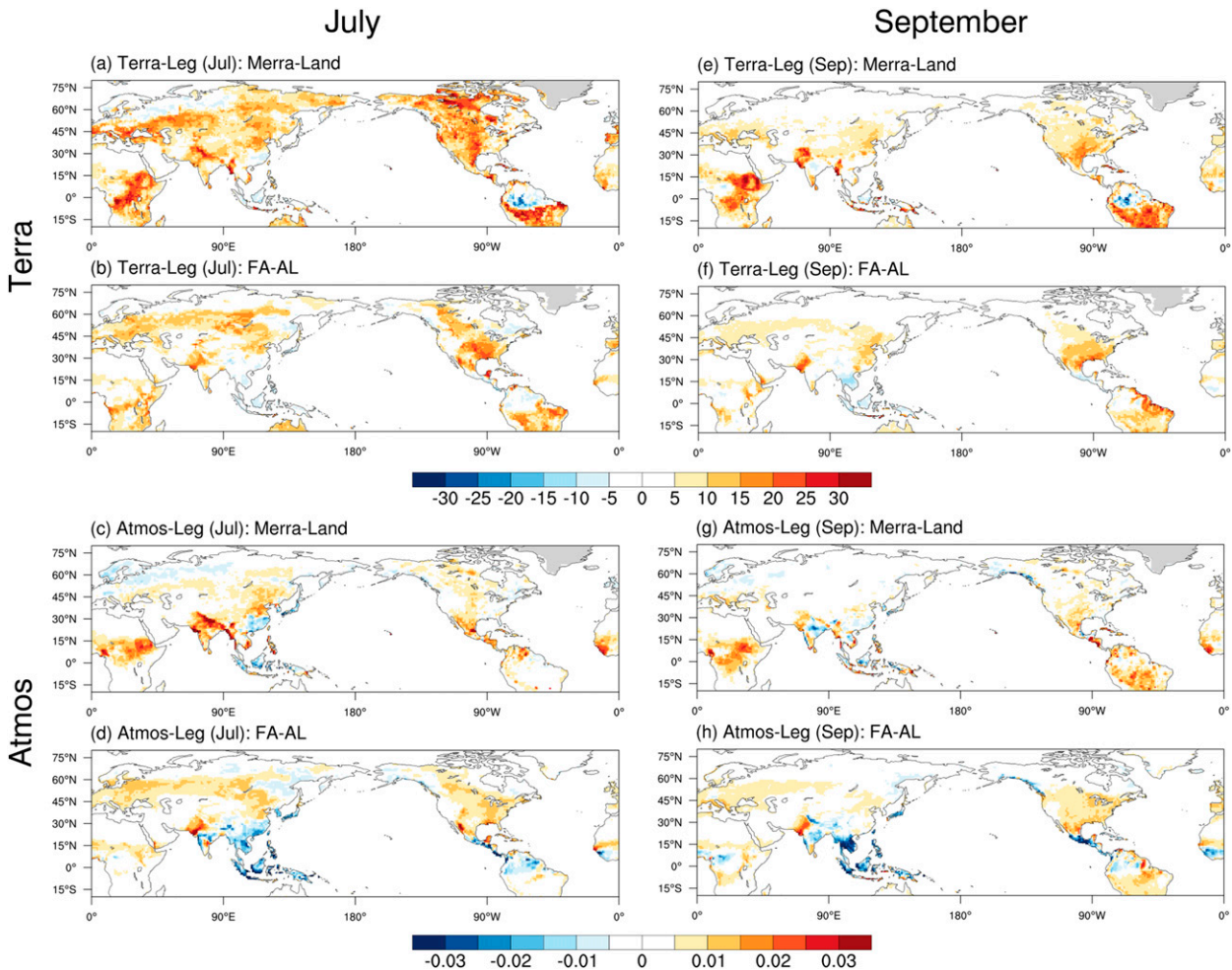


FIG. 10. Land–atmosphere coupling in the June-initialized FA-AL predictions as characterized by two-legged metrics (Dirmeyer 2011). (a),(b),(e),(f) The terrestrial leg (“Terra”) shows the product of the standard deviation of soil moisture content (kg m^{-3}) and the local linear regression of the latent heat flux (W m^{-2}) onto soil moisture content (kg m^{-3}). (c),(d),(g),(h) The atmospheric leg (“Atmos”) shows the product of the standard deviation of surface latent heat flux (W m^{-2}) and the local linear regression of precipitation (10^{-3}kg m^{-2}) onto surface latent heat flux (W m^{-2}). The two columns show the results for (left) July and (right) September, respectively. Due to the limited data availability for FA-AL, the calculation uses monthly mean data from MERRA-Land in (a), (c), (e), and (g) and FA-AL in (b), (d), (f), and (h) (June-initialized).

negatively affect FA-AL’s ability to predict Atlantic TC activity, as the land conditions in this region are significantly correlated with Atlantic TC activity (not shown). Additionally, fine-structure differences between MERRA-Land and FA-AL are evident in some regions, including North America (e.g., Figs. 10a,b,g,h). Finally, the land–atmosphere coupling indices of FA-AL and FA-Basic share nearly identical patterns (not shown), suggesting that the coupling characterized by the two-legged analysis is mostly a function of the model physics, not the land initialization.

Overall, the analyses in this section suggest that FA-AL has relatively realistic land initial conditions and land–atmosphere coupling over North America. FA-AL also skillfully represents the atmospheric initial conditions (Jia et al. 2016). With these advantages over FA-Basic, we next examine whether FA-AL shows improved skill in predicting Atlantic TC activity and the large-scale environment.

c. Impacts of land–atmosphere initial conditions

As the TC climatology of FA-Basic and FA-AL are highly similar (not shown), this section focuses on *predictions* of the large-scale environment and TC activity. The land memory is around three months in North America (Figs. 9e,f), so we mainly discuss June–September predictions initialized in June, when the land–atmosphere coupling and the land impact on prediction are relatively strong (e.g., Dirmeyer 2011; Guo et al. 2011). We shall emphasize the predictions for North America and Atlantic TC activity, along with a brief discussion of skill changes in other regions.

Figure 11 shows the seasonal skill for predicted TC-related environmental variables during June–September. Compared to FA-Basic, the SST predictions by FA-AL worsen in the west Pacific and tropical west Atlantic but improve in the subtropical northeast Atlantic. These skill changes are moderate but

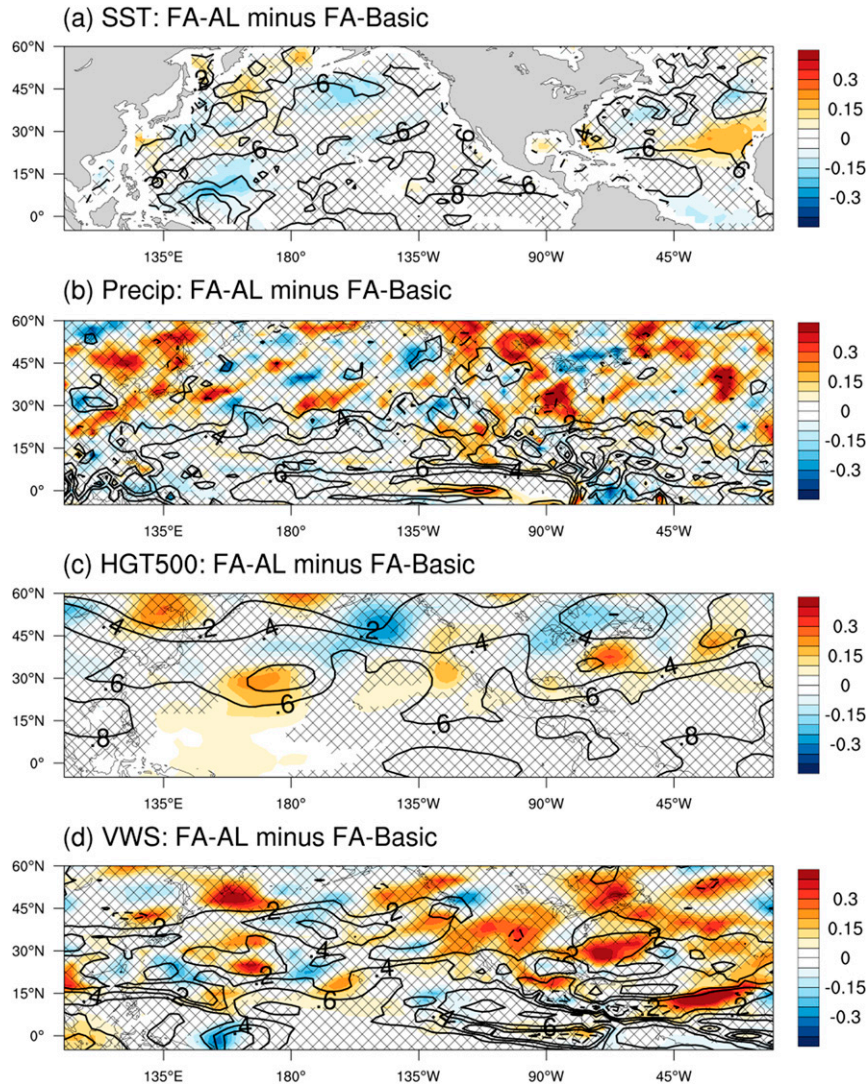


FIG. 11. Correlations of large-scale environment variables (June–September) between observations and the FA-Basic or FA-AL predictions initialized in June. (a) SST, (b) precipitation, (c) 500-hPa geopotential height, and (d) vertical wind shear between the 200- and 850-hPa pressure levels. Black contours show the skill for FA-Basic. Color shading shows the skill difference, FA-AL minus FA-Basic. Hatching indicates that differences are below the 95% confidence level based on a bootstrapping test (section 2c).

statistically significant. The changes in precipitation skill are noisy but mostly improve in FA-AL in and around extratropical land regions. For example, FA-AL shows improved precipitation skill ($\Delta r > 0.2$) in some regions with strong land–atmosphere coupling, such as the southeastern United States near 30°N, 90°W. This increase of precipitation skill over land and downstream regions is particularly extensive in the first month after initialization (Fig. 1 in Jia et al. 2016). The skill for 500-hPa geopotential height improves near the southern and eastern coasts of the United States and in the western Pacific, though it decreases in parts of the midlatitudes. FA-AL also shows improved skills for wind shear over the southwestern United States and North Atlantic. Overall, the improved land–atmosphere initial conditions in FA-AL (Fig. 9) help to

improve its predictions of the large-scale environment near the U.S. coasts and the North Atlantic.

We next examine whether FA-AL’s improved prediction of the large-scale environment also improves its predictions of TC activity (Fig. 12). To facilitate comparison with FA-Basic (Fig. 4), we analyze the predictions for the period of July–November and denote the median values of FA-Basic’s values. When initialized in June, FA-AL shows better skill than FA-Basic in predicting year-to-year variations of Atlantic TC number. The skill increase is ~ 0.10 and is comparable to perfecting the SST. Meanwhile, the RMSE of TC number and ACE decreases below the values of FA-Basic and ReSST. The prediction of TC tracks (and thus landfalls) by the FA-AL also improves, though the improvements are localized and small in all the basins (not shown). But when

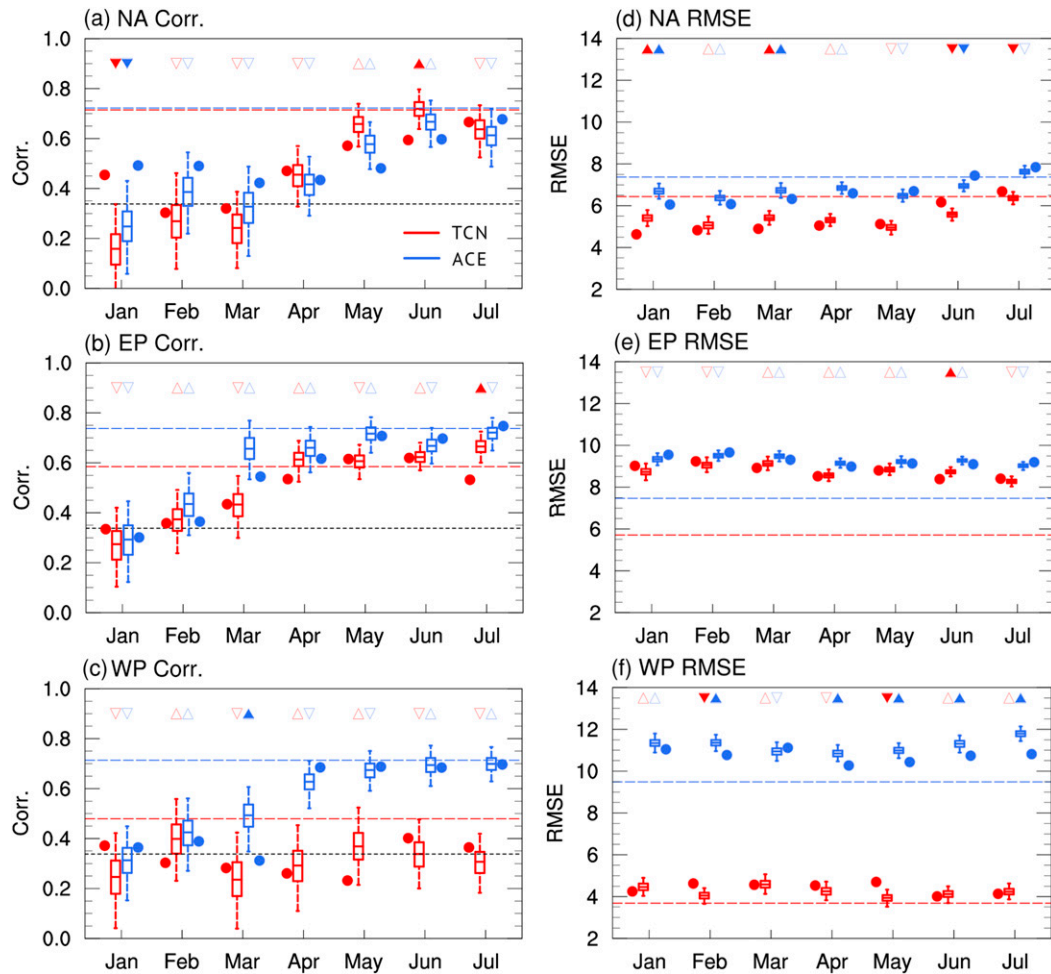


FIG. 12. Correlations and root-mean-square error (RMSE) of FA-AL (July–November). The plot settings are identical to Fig. 4, but with additional markers to facilitate comparisons with FA-Basic. Triangles indicate either an increase (upward triangles) or decrease (downward triangles) in the metrics of FA-AL relative to FA-Basic. The triangles are color filled if the metric differences are at the 90% confidence level based on a bootstrapping test (section 2c). The dots beside boxplots show the median value of the correlation from FA-basic.

initialized in January–March, FA-AL shows reduced skill in predicting Atlantic TC number and ACE relative to FA-Basic. The skill decrease in the January initialization is ~ 0.20 and is not well understood, but we speculate that land processes (e.g., snow accumulation and melting) and simulation drifts play some role. Overall, the improvements due to FA-AL’s atmosphere and land initialization are most evident for TC activity in the North Atlantic, possibly because the basin is free of large biases in marine precipitation (section 3a) and is downstream of regions with strong land–atmosphere coupling (section 4b).

To illuminate the TC prediction skill changes in FA-AL, Fig. 13 shows 3-month predictions of 200- and 850-hPa geopotential height initialized in June and July. In both cases, FA-AL shows widespread significant improvements relative to FA-Basic. For the lower troposphere (850 hPa), skill improvements appear over or downstream of North America and East Asia, with substantial spatial variations. For June initializations, the correlation skill nearly doubles near the southern and the eastern coasts of the

United States (Fig. 13b). In comparison, July initializations show little change in skill (Fig. 13c) or even degraded skill near the U.S. East Coast (Fig. 13d) despite extensive improvements elsewhere. The fact that FA-AL’s Atlantic TC predictions improve for June but not July initializations (Fig. 12d) is consistent with the changes in geopotential height prediction skill (Fig. 13), since the large-scale circulation near the U.S. East Coast is important for TC activity (e.g., Kossin et al. 2010; Murakami et al. 2016; Zhang et al. 2016; Zhang and Wang 2019). Interestingly, the skill in predicting 850-hPa geopotential height of the northeastern Pacific shows a more apparent increase in the July-initialized prediction than in the June-initialized prediction (Figs. 13b,d), consistent with an improvement of TC number prediction in the July-initialized prediction (Fig. 12b). While it is not fully clear why the prediction skill responds to land–atmosphere initialization in such a complex manner, the limited verification period (1981–2014) and hindcast ensemble size (12) may contribute to sampling variability of the diagnosed skill, especially at smaller spatial scales. Additional

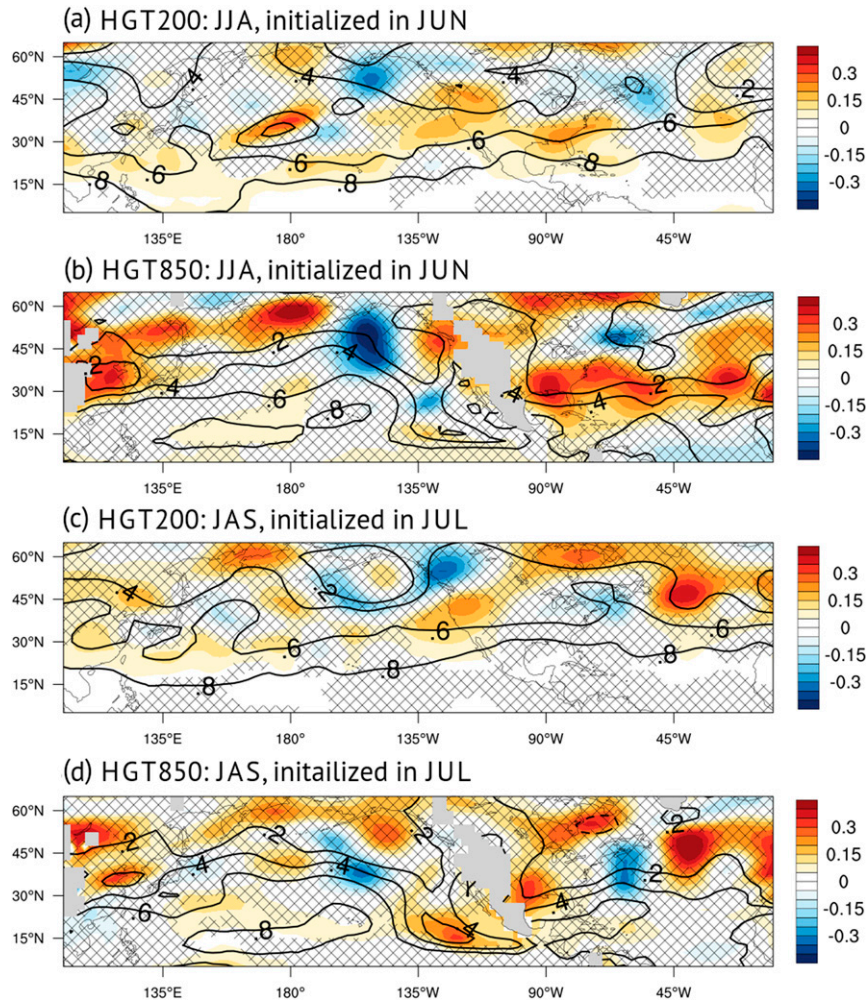


FIG. 13. Local correlations between the observations and the FA-Basic or FA-AL predictions, for (a) 200-hPa geopotential height (June–August) for predictions initialized in June, (b) 850-hPa geopotential height (June–August) for predictions initialized in June, (c) 200-hPa geopotential height (July–September) for predictions initialized in July, and (d) 850-hPa geopotential height (July–September) for predictions initialized in July. Black contours show the correlation skill of FA-Basic. Color shading shows the skill difference, FA-AL minus FA-Basic. Hatching indicates that differences are below the 95% confidence level based on a bootstrapping test (section 2c).

uncertainty could arise from the model initialization of FA-AL, which uses only a single realization of the land–atmosphere initial conditions. Nonetheless, the results here suggest that the impact on TC predictions of land–atmosphere initialization can be comparable to that of SST errors (e.g., for the predictions initialized in June, Figs. 4 and 12).

5. Summary and discussion

This study analyzes three sets of FLOR experiments and seeks to explore avenues for future improvements in simulating and predicting TC activity. The results highlight several existing sources of error:

- SST biases and prediction errors influence the simulation and prediction of TC activity, even in a flux-adjusted pre-

diction model. The impacts of these SST errors are reduced at short leads or near coasts.

- Non-oceanic sources (e.g., parameterized atmospheric physics) strongly contribute to biases in FLOR’s simulated TC activity. The impacts of these non-oceanic errors have notable regional variations, and their dependence on SST-related errors can be assessed through SST-restoring experiments.
- Realistic land–atmosphere initialization might affect and potentially improve seasonal predictions of TC activity, depending on the basin and initialization month. For some short-lead predictions (e.g., June-initialized predictions for the North Atlantic), land–atmosphere initialization might improve the predicted TC numbers as much as perfecting the SST.

Some of these findings may be model dependent. For example, in a different model with more severe SST biases, those

SST biases might be the main driver of that model's errors in simulated TC activity. Rather than conducting an exhaustive survey of various error sources, this study instead aims to illustrate an evaluation framework for identifying intertwined errors in a coupled model, assisting model development, and facilitating scientific discoveries. The framework can be applied to other climate models, toward improving simulations and predictions of TCs and other extreme weather events. It is also worth noting that the statistical significance test in this study has some caveats (appendix B), so the robustness of skill changes should be interpreted with caution.

This study does not separately attribute the impacts of land initialization and atmospheric initialization. Jia et al. (2016) used the same model to conduct predictions for the boreal summer of 2006. Their study showed that both land and atmospheric initialization contributed to improved environmental predictions for the United States and downstream regions. Unfortunately, such experiments are computationally expensive to conduct across multiple years.

The representation of the land–atmosphere coupling is not optimal in the FLOR model. In particular over tropical Africa and South America, FLOR's land–atmosphere coupling is weaker than that suggested by MERRA-Land reanalysis (Fig. 10). The shortcomings of land–atmosphere coupling are not unique to the FLOR (Dirmeyer et al. 2018; Ardilouze et al. 2019) and may hinder the identification of land impacts on TC activity. Another challenge is data availability, since only a subset of variables was saved when the original hindcast experiments were conducted. More carefully designed model experiments, perhaps with an updated climate model (e.g., Delworth et al. 2020), could further advance understanding of how land conditions may affect TC activity (Zhou et al. 2019; Baldwin et al. 2019).

Overall, the findings in this study are consistent with Zhang et al. (2019), suggesting that a gap exists between the actual and potential skill of seasonal TC predictions. To improve predictions of TC activity, it will be helpful to continue improving SST predictions, especially at longer leads. Nonetheless, seasonal TC predictions may already have reached a stage where *further* improvements in skill may need to draw from new sources—such as improved atmospheric physics parameterizations, and land–atmosphere initialization—which will require intensified collaboration across the research and modeling communities.

Acknowledgments. G. Zhang acknowledges the support from the NOAA's Predicting and Explaining Extremes Initiative through Princeton University's Cooperative Institute for Modeling the Earth System (CIMES). We thank Tom Delworth, Rich Gudgel, Zhi Liang, and Ahmed Tawfik for discussion and/or technical support that facilitated this study.

Data availability statement. The IBTrACS data are acquired from the National Centers for Environmental Information (<https://www.ncdc.noaa.gov/ibtracs/index.php>). We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA-Interim data (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim>), and the National Aeronautics and Space Administration (NASA) for providing the

MERRA-Land data (<https://gmao.gsfc.nasa.gov/reanalysis/MERRA-Land/>). The FLOR model code is publicly available (<https://www.gfdl.noaa.gov/cm2-5-and-flor/>), and the FLOR predictions are available on the North American Multi-Model Ensemble website (<https://www.cpc.ncep.noaa.gov/products/NMME/>).

APPENDIX A

Simulations of TC Metrics

To aid evaluation of the simulations, we present scatterplots of simulated TC number (Fig. A1) and ACE (Fig. A2) against their observed counterparts. Since the predictions initialized in January–July have similar characteristics (not shown), the subplots of FA-Basic and FA-AL show only the predictions initialized in July for brevity. Figure A1 suggests that the FLOR simulations underestimate the TC number in the North Atlantic and the northeastern Pacific. While the TC number in the Northwestern Pacific is reasonable, the simulated variations of TC number underestimate the observed contrast between active and inactive years. Figure A2 suggests FLOR underestimates ACE in all the basins, especially when the observed TC activity is high. The underestimation of ACE is contributed by the aforementioned TC number biases and FLOR's inability to simulate intense TCs. Consistent with Fig. 1 and the related discussion in the main text, the biases of TC activity in individual basins are overall similar across ReSST, FA-Basic, and FA-AL. The scatterplots also suggest that the relationship between the simulated and observed variables is roughly linear. Consistent with this relationship, the Pearson correlation used in the main text and the Spearman rank correlation used in Figs. A1 and A2 produce similar results. Many of these aspects of FLOR have been identified previously (e.g., Vecchi et al. 2014; Murakami et al. 2016; Zhang et al. 2019). Readers interested in other aspects of the model performance can refer to the references in section 2a.

APPENDIX B

Additional Considerations of Significance Tests

The skill evaluation in this study emphasizes models' capability in predicting year-to-year variations. This emphasis was motivated by two factors: 1) dynamical predictions of TC activity have substantial climatological biases; and 2) the year-to-year variations are much more relevant to the real-world needs related to seasonal predictions. Without conducting bias corrections, the skill in predicting year-to-year variations is better characterized by correlation coefficients rather than RMSEs.

When testing the significance level of correlation differences, a conventional way is to use Fisher's Z transformation and proceed with an assumption of the normal distribution. However, DelSole and Tippett (2014) showed that the predictions from different models are not independent in the sense that the predictand is the same observation. This sample dependence undermines the distribution assumptions used by the parametric-based significance tests that assume independent samples. In comparison, the bootstrapping resampling approach in this study directly evaluates the

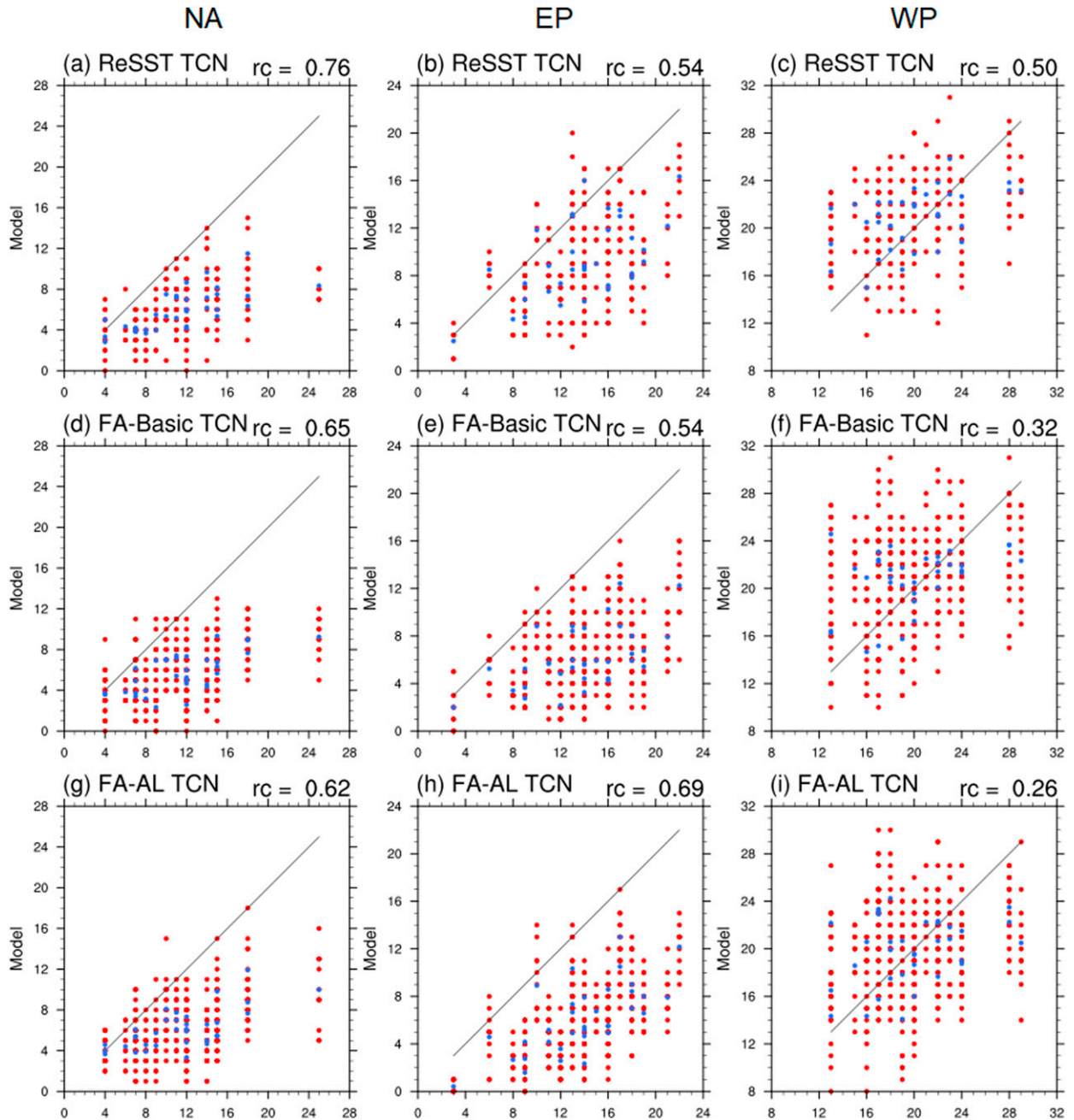


FIG. A1. TC number in the observations and simulations (ReSST, FA-Basic, and FA-AL). The predictions of both FA-Basic and FA-AL are initialized in July. The horizontal axis shows the observed values, and the vertical axis shows the simulated values. The red dots denote the results from individual ensemble members, and the blue dots show the ensemble means. The black line corresponds to the condition with equal observed and simulated values. (a)–(c) The ReSST subplots include 6 ensemble members, while the subplots of (d)–(f) FA-Basic and (g)–(i) FA-AL each include 12 ensemble members. The Spearman rank correlation between the ensemble means and the observation is denoted in the top right of subplots.

distribution of the difference between correlation coefficients. This approach circumvents Fisher's Z transformation and the vulnerable assumption about statistical distributions. Meanwhile, the bootstrapping method accounts for unforced variability/noises in the dynamical system, which might cause statistically significant but physically insignificant differences when an ensemble

prediction does not adequately sample the system's nonlinear evolution.

An alternative significance test is the random walk test employed by [DelSole and Tippett \(2014\)](#). This test uses a score metric to evaluate predictions for each observation as a single event. Also referred as the sign test, the random walk test

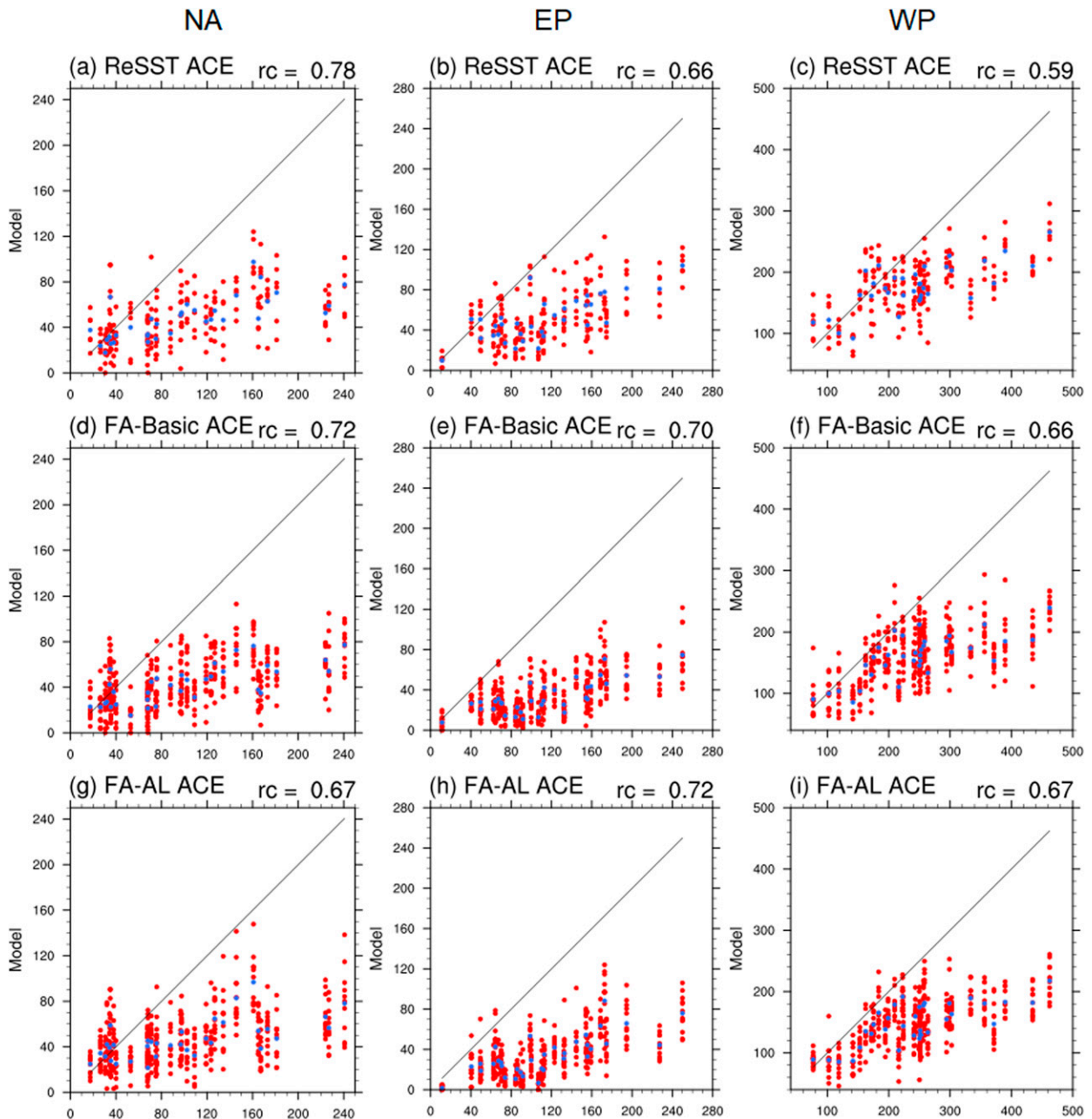


FIG. A2. As in Fig. A1, but showing ACE instead of TC number.

compares predictions in a series of equal-probability Bernoulli trials and evaluate the statistical significance using a binomial distribution. This approach differs from the two aforementioned significance tests as it does not assume or evaluate the distributions of forecast errors. However, some score metrics (e.g., squared errors used by DelSole and Tippett (2014)) can be sensitive to forecast errors and heavily penalize predictions that have large climatological biases but are otherwise skillful in predicting variations. In this particular circumstance, such score metrics are less relevant to evaluating models' capability in predicting year-to-year variations.

For the sake of completeness, Fig. B1 shows the statistical significance determined by the random walk test. The random walk test produces notably inconsistent results when applied to the differences between two sets of six-member ensembles (Fig. B1). We speculate that the underlying issue is model-related and similar to the CCSM3 results examined by DelSole and Tippett (2016). Compared to the results from the bootstrapping (Fig. 12), the random walk test is more generous in granting high statistical significance. The high-significance results from these two types of significance tests are not entirely consistent, either. These issues suggest the statistical significance

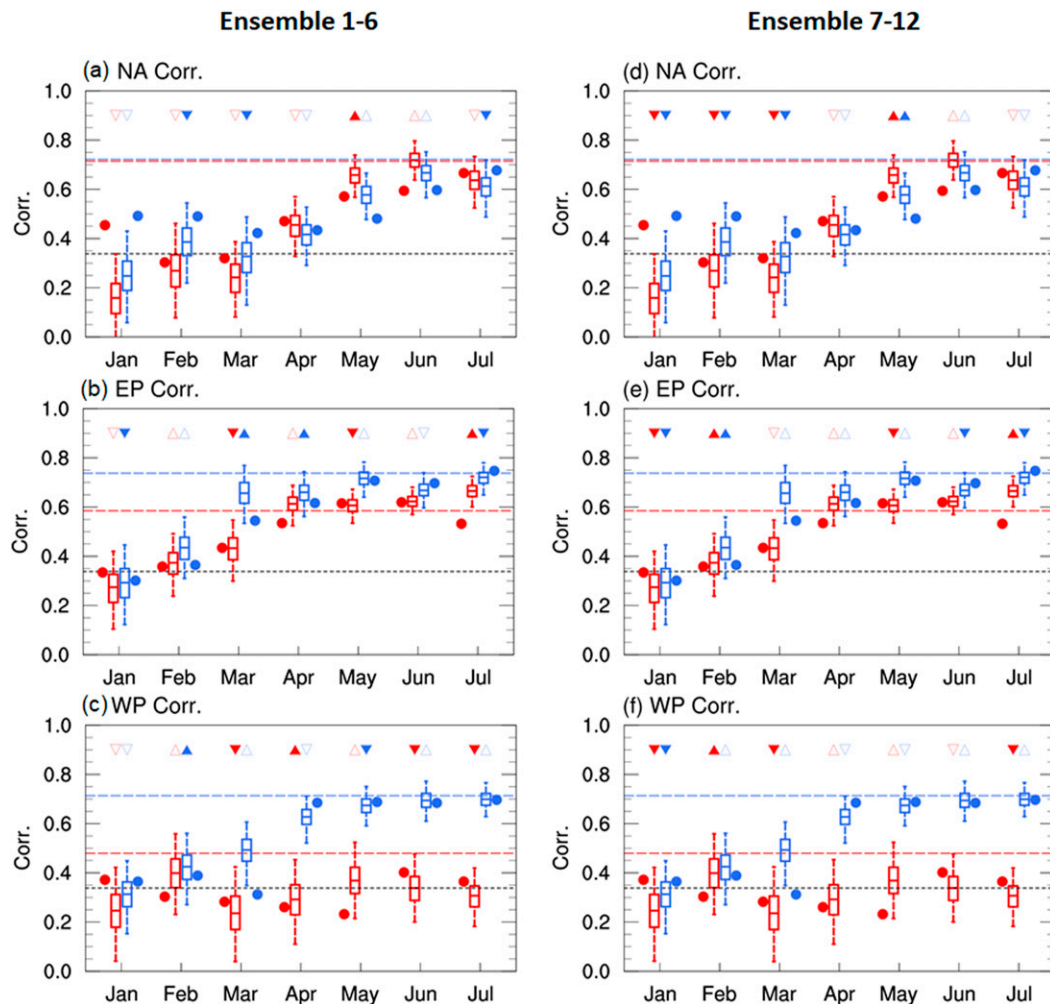


FIG. B1. As in Figs. 12a–c, but that the filled triangles denote the statistical significance determined from the random walk test. The comparison between FA-Basic and FA-AL is conducted using the means of six-member ensembles. Without resampling, the available hindcast data of a specific model can be simply divided into two nonoverlapping six-member subsets. We label the subsets as (a)–(c) ensemble 1–6 and (d)–(f) ensemble 7–12 and conduct the significance test separately (i.e., FA-AL 1–6 minus FA-Basic 1–6; FA-AL 7–12 minus FA-Basic 7–12). No resampling or bias corrections are involved.

of the skill changes described in the main text is subject to method-related uncertainties.

REFERENCES

- Antonov, J. I., R. A. Locarnini, T. P. Boyer, A. V. Mishonov, and H. E. Garcia, 2006: *Salinity*. Vol. 2, *World Ocean Atlas 2005*, NOAA Atlas NESDIS 62, 182 pp.
- Ardilouze, C., L. Batté, M. Déqué, E. van Meijgaard, and B. van den Hurk, 2019: Investigating the impact of soil moisture on European summer climate in ensemble numerical experiments. *Climate Dyn.*, **52**, 4011–4026, <https://doi.org/10.1007/s00382-018-4358-1>.
- Baldwin, J. W., G. A. Vecchi, and S. Bordoni, 2019: The direct and ocean-mediated influence of Asian orography on tropical precipitation and cyclones. *Climate Dyn.*, **53**, 805–824, <https://doi.org/10.1007/s00382-019-04615-5>.
- Bell, G. D., and Coauthors, 2000: Climate assessment for 1999. *Bull. Amer. Meteor. Soc.*, **81**, S1–S50, [https://doi.org/10.1175/1520-0477\(2000\)81\[s1:CAF\]2.0.CO;2](https://doi.org/10.1175/1520-0477(2000)81[s1:CAF]2.0.CO;2).
- , C. W. Landsea, S. B. Goldenberg, R. J. Pasch, E. S. Blake, J. Schemm, and T. B. Kimberlain, 2014: The 2013 North Atlantic hurricane season: A climate perspective [in “State of the Climate in 2013”]. *Bull. Amer. Meteor. Soc.*, **95** (7), S86–S90.
- Camargo, S. J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902, <https://doi.org/10.1175/JCLI-D-12-00549.1>.
- , and A. G. Barnston, 2009: Experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. *Wea. Forecasting*, **24**, 472–491, <https://doi.org/10.1175/2008WAF2007099.1>.
- , and A. A. Wing, 2016: Tropical cyclones in climate models. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 211–237, <https://doi.org/10.1002/wcc.373>.

- , K. A. Emanuel, and A. H. Sobel, 2007: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, **20**, 4819–4834, <https://doi.org/10.1175/JCLI4282.1>.
- Camp, J., M. Roberts, C. MacLachlan, E. Wallace, L. Hermanson, A. Brookshaw, A. Arribas, and A. A. Scaife, 2015: Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 2206–2219, <https://doi.org/10.1002/qj.2516>.
- 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, <https://doi.org/10.1175/JCLI-D-12-00061.1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- DelSole, T., and M. K. Tippett, 2014: Comparing forecast skill. *Mon. Wea. Rev.*, **142**, 4658–4678, <https://doi.org/10.1175/MWR-D-14-00045.1>.
- , and —, 2016: Forecast comparison based on random walks. *Mon. Wea. Rev.*, **144**, 615–626, <https://doi.org/10.1175/MWR-D-15-0218.1>.
- Delworth, T., and S. Manabe, 1988: The influence of potential evaporation on the variabilities of simulated soil wetness and climate. *J. Climate*, **1**, 523–547, [https://doi.org/10.1175/1520-0442\(1988\)001<0523:TIOPEO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1988)001<0523:TIOPEO>2.0.CO;2).
- , and —, 1989: The influence of soil wetness on near-surface atmospheric variability. *J. Climate*, **2**, 1447–1462, [https://doi.org/10.1175/1520-0442\(1989\)002<1447:TIOSWO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1989)002<1447:TIOSWO>2.0.CO;2).
- , and Coauthors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25**, 2755–2781, <https://doi.org/10.1175/JCLI-D-11-00316.1>.
- , and Coauthors, 2020: SPEAR—The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001895, <https://doi.org/10.1029/2019MS001895>.
- Dirmeyer, P. A., 2011: The terrestrial segment of soil moisture-climate coupling. *Geophys. Res. Lett.*, **38**, L16702, <https://doi.org/10.1029/2011GL048268>.
- , M. J. Fennessy, and L. Marx, 2003: Low skill in dynamical prediction of boreal summer climate: Grounds for looking beyond sea surface temperature. *J. Climate*, **16**, 995–1002, [https://doi.org/10.1175/1520-0442\(2003\)016<0995:LSIDPO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0995:LSIDPO>2.0.CO;2).
- , S. Halder, and R. Bombardi, 2018: On the harvest of predictability from land states in a global forecast model. *J. Geophys. Res. Atmos.*, **123**, 13 111–13 127, <https://doi.org/10.1029/2018JD029103>.
- Emanuel, K., 2010: Tropical cyclone activity downscaled from NOAA-CIRES Reanalysis, 1908–1958. *J. Adv. Model. Earth Syst.*, **2**, 1, <https://doi.org/10.3894/JAMES.2010.2.1>.
- Findell, K. L., and E. A. B. Eltahir, 2003a: Atmospheric controls on soil moisture–boundary layer interactions. Part I: Framework development. *J. Hydrometeorol.*, **4**, 552–569, [https://doi.org/10.1175/1525-7541\(2003\)004<0552:ACOSML>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<0552:ACOSML>2.0.CO;2).
- , and —, 2003b: Atmospheric controls on soil moisture–boundary layer interactions. Part II: Feedbacks within the continental United States. *J. Hydrometeorol.*, **4**, 570–583, [https://doi.org/10.1175/1525-7541\(2003\)004<0570:ACOSML>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<0570:ACOSML>2.0.CO;2).
- , P. Gentine, B. R. Lintner, and C. Kerr, 2011: Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation. *Nat. Geosci.*, **4**, 434–439, <https://doi.org/10.1038/ngeo1174>.
- Guo, Z., P. A. Dirmeyer, and T. DelSole, 2011: Land surface impacts on subseasonal and seasonal predictability. *Geophys. Res. Lett.*, **38**, L24812, <https://doi.org/10.1029/2011GL049945>.
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, [https://doi.org/10.1175/1520-0434\(1999\)014<0155:HTFENP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2).
- Harris, L. M., S.-J. Lin, and C. Tu, 2016: High-resolution climate simulations using GFDL HiRAM with a stretched global grid. *J. Climate*, **29**, 4293–4314, <https://doi.org/10.1175/JCLI-D-15-0389.1>.
- He, J., N. C. Johnson, G. A. Vecchi, B. Kirtman, A. T. Wittenberg, and S. Sturm, 2018: Precipitation sensitivity to local variations in tropical sea surface temperature. *J. Climate*, **31**, 9225–9238, <https://doi.org/10.1175/JCLI-D-18-0262.1>.
- Hsu, W.-C., C. M. Patricola, and P. Chang, 2019: The impact of climate model sea surface temperature biases on tropical cyclone simulations. *Climate Dyn.*, **53**, 173–192, <https://doi.org/10.1007/s00382-018-4577-5>.
- Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate*, **21**, 5145–5153, <https://doi.org/10.1175/2008JCLI2292.1>.
- 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, <https://doi.org/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, <https://doi.org/10.1175/JCLI-D-15-0471.1>.
- Jien, J. Y., W. A. Gough, and K. Butler, 2015: The influence of El Niño–Southern Oscillation on tropical cyclone activity in the eastern North Pacific basin. *J. Climate*, **28**, 2459–2474, <https://doi.org/10.1175/JCLI-D-14-00248.1>.
- Kapnick, S. B., and Coauthors, 2018: Potential for western US seasonal snowpack prediction. *Proc. Natl. Acad. Sci. USA*, **115**, 1180–1185, <https://doi.org/10.1073/pnas.1716760115>.
- Kim, D., and Coauthors, 2018: Process-oriented diagnosis of tropical cyclones in high-resolution GCMs. *J. Climate*, **31**, 1685–1702, <https://doi.org/10.1175/JCLI-D-17-0269.1>.
- Kim, H.-M., P. J. Webster, and J. A. Curry, 2009: Impact of shifting patterns of Pacific Ocean warming on North Atlantic tropical cyclones. *Science*, **325**, 77–80, <https://doi.org/10.1126/science.1174062>.
- , —, and —, 2011: Modulation of North Pacific tropical cyclone activity by three phases of ENSO. *J. Climate*, **24**, 1839–1849, <https://doi.org/10.1175/2010JCLI3939.1>.
- Kim, H.-S., G. A. Vecchi, T. R. Knutson, W. G. Anderson, T. L. Delworth, A. Rosati, F. Zeng, and M. Zhao, 2014: Tropical cyclone simulation and response to CO₂ doubling in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **27**, 8034–8054, <https://doi.org/10.1175/JCLI-D-13-00475.1>.
- Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.*, **91**, 363–376, <https://doi.org/10.1175/2009BAMS2755.1>.

- Kossin, J. P., S. J. Camargo, and M. Sitkowski, 2010: Climate modulation of North Atlantic hurricane tracks. *J. Climate*, **23**, 3057–3076, <https://doi.org/10.1175/2010JCLI3497.1>.
- Koster, R. D., and Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, <https://doi.org/10.1126/science.1100217>.
- , Y. Chang, and S. D. Schubert, 2014: A mechanism for land–atmosphere feedback involving planetary wave structures. *J. Climate*, **27**, 9290–9301, <https://doi.org/10.1175/JCLI-D-14-00315.1>.
- , —, H. Wang, and S. D. Schubert, 2016: Impacts of local soil moisture anomalies on the atmospheric circulation and on remote surface meteorological fields during boreal summer: A comprehensive analysis over North America. *J. Climate*, **29**, 7345–7364, <https://doi.org/10.1175/JCLI-D-16-0192.1>.
- Krishnamurthy, L., G. Vecchi, R. Msadek, A. Wittenberg, T. Delworth, and F. Zeng, 2015: The seasonality of the Great Plains low-level jet and ENSO relationship. *J. Climate*, **28**, 4525–4544, <https://doi.org/10.1175/JCLI-D-14-00590.1>.
- , —, —, H. Murakami, A. Wittenberg, and F. Zeng, 2016: Impact of strong ENSO on regional tropical cyclone activity in a high-resolution climate model in the North Pacific and North Atlantic. *J. Climate*, **29**, 2375–2394, <https://doi.org/10.1175/JCLI-D-15-0468.1>.
- Li, H., and R. L. Sriver, 2018: Tropical cyclone activity in the high-resolution community earth system model and the impact of ocean coupling. *J. Adv. Model. Earth Syst.*, **10**, 165–186, <https://doi.org/10.1002/2017MS001199>.
- Li, W., Z. Wang, G. Zhang, M. S. Peng, S. G. Benjamin, and M. Zhao, 2018: Subseasonal variability of Rossby wave breaking and impacts on tropical cyclones during the North Atlantic warm season. *J. Climate*, **31**, 9679–9695, <https://doi.org/10.1175/JCLI-D-17-0880.1>.
- Liu, M., G. A. Vecchi, J. A. Smith, H. Murakami, R. Gudgel, and X. Yang, 2018: Towards dynamical seasonal forecast of extra-tropical transition in the North Atlantic. *Geophys. Res. Lett.*, **45**, 12 602–12 609, <https://doi.org/10.1029/2018GL079451>.
- Magnusson, L., M. Alonso-Balmaseda, S. Corti, F. Molteni, and T. Stockdale, 2013: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dyn.*, **41**, 2393–2409, <https://doi.org/10.1007/s00382-012-1599-2>.
- Manganello, J. V., and B. Huang, 2009: The influence of systematic errors in the Southeast Pacific on ENSO variability and prediction in a coupled GCM. *Climate Dyn.*, **32**, 1015–1034, <https://doi.org/10.1007/s00382-008-0407-5>.
- , and Coauthors, 2012: Tropical cyclone climatology in a 10-km global atmospheric GCM: Toward weather-resolving climate modeling. *J. Climate*, **25**, 3867–3893, <https://doi.org/10.1175/JCLI-D-11-00346.1>.
- , and Coauthors, 2016: Seasonal forecasts of tropical cyclone activity in a high-atmospheric-resolution coupled prediction system. *J. Climate*, **29**, 1179–1200, <https://doi.org/10.1175/JCLI-D-15-0531.1>.
- , B. A. Cash, K. I. Hodges, and J. L. Kinter, 2019: Seasonal forecasts of North Atlantic tropical cyclone activity in the North American multi-model ensemble. *Climate Dyn.*, **53**, 7169–7184, <https://doi.org/10.1007/s00382-017-3670-5>.
- Mei, W., Y. Kamae, S.-P. Xie, and K. Yoshida, 2019: Variability and predictability of North Atlantic hurricane frequency in a large ensemble of high-resolution atmospheric simulations. *J. Climate*, **32**, 3153–3167, <https://doi.org/10.1175/JCLI-D-18-0554.1>.
- Milly, P. C. D., and Coauthors, 2014: An enhanced model of land water and energy for global hydrologic and earth-system studies. *J. Hydrometeor.*, **15**, 1739–1761, <https://doi.org/10.1175/JHM-D-13-0162.1>.
- Murakami, H., R. Mizuta, and E. Shindo, 2012: Future changes in tropical cyclone activity projected by multi-physics and multi-SST ensemble experiments using the 60-km-mesh MRI-AGCM. *Climate Dyn.*, **39**, 2569–2584, <https://doi.org/10.1007/s00382-011-1223-x>.
- , 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, <https://doi.org/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, <https://doi.org/10.1175/MWR-D-15-0308.1>.
- , E. Levin, T. L. Delworth, R. Gudgel, and P.-C. Hsu, 2018: Dominant effect of relative tropical Atlantic warming on major hurricane occurrence. *Science*, **362**, 794–799, <https://doi.org/10.1126/science.aat6711>.
- Neelin, J. D., and H. A. Dijkstra, 1995: Ocean–atmosphere interaction and the tropical climatology. Part I: The dangers of flux correction. *J. Climate*, **8**, 1325–1342, [https://doi.org/10.1175/1520-0442\(1995\)008<1325:OAIATT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1325:OAIATT>2.0.CO;2).
- Newman, M., A. T. Wittenberg, L. Cheng, G. P. Compo, and C. A. Smith, 2018: The extreme 2015/16 El Niño, in the context of historical climate variability and change. *Bull. Amer. Meteor. Soc.*, **99**, S16–S20, <https://doi.org/10.1175/BAMS-D-17-0116.1>.
- Patricola, C. M., S. J. Camargo, P. J. Klotzbach, R. Saravanan, and P. Chang, 2018: The influence of ENSO flavors on western North Pacific tropical cyclone activity. *J. Climate*, **31**, 5395–5416, <https://doi.org/10.1175/JCLI-D-17-0678.1>.
- Ray, S., A. T. Wittenberg, S. M. Griffies, and F. Zeng, 2018a: Understanding the equatorial Pacific cold tongue time-mean heat budget. Part I: Diagnostic framework. *J. Climate*, **31**, 9965–9985, <https://doi.org/10.1175/JCLI-D-18-0152.1>.
- , —, —, and —, 2018b: Understanding the equatorial Pacific cold tongue time-mean heat budget. Part II: Evaluation of the GFDL-FLOR coupled GCM. *J. Climate*, **31**, 9987–10 011, <https://doi.org/10.1175/JCLI-D-18-0153.1>.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, <https://doi.org/10.1175/JCLI-D-10-05033.1>.
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.
- Santanello, J. A., and Coauthors, 2018: Land–atmosphere interactions: The LoCo perspective. *Bull. Amer. Meteor. Soc.*, **99**, 1253–1272, <https://doi.org/10.1175/BAMS-D-17-0001.1>.
- Schreck, C. J., K. R. Knapp, and J. P. Kossin, 2014: The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Mon. Wea. Rev.*, **142**, 3881–3899, <https://doi.org/10.1175/MWR-D-14-00021.1>.
- Seager, R., J. Nakamura, and M. Ting, 2019: Mechanisms of seasonal soil moisture drought onset and termination in the

- southern Great Plains. *J. Hydrometeorol.*, **20**, 751–771, <https://doi.org/10.1175/JHM-D-18-0191.1>.
- Shackley, S., J. Risbey, P. Stone, and B. Wynne, 1999: Adjusting to policy expectations in climate change modeling. *Climatic Change*, **43**, 413–454, <https://doi.org/10.1023/A:1005474102591>.
- Spencer, H., R. Sutton, and J. M. Slingo, 2007: El Niño in a coupled climate model: Sensitivity to changes in mean state induced by heat flux and wind stress corrections. *J. Climate*, **20**, 2273–2298, <https://doi.org/10.1175/JCLI4111.1>.
- Stockdale, T. N., 1997: Coupled ocean–atmosphere forecasts in the presence of climate drift. *Mon. Wea. Rev.*, **125**, 809–818, [https://doi.org/10.1175/1520-0493\(1997\)125<0809:COAFIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0809:COAFIT>2.0.CO;2).
- Teng, H., G. Branstator, A. B. Tawfik, and P. Callaghan, 2019: Circumglobal response to prescribed soil moisture over North America. *J. Climate*, **32**, 4525–4545, <https://doi.org/10.1175/JCLI-D-18-0823.1>.
- van der Wiel, K., and Coauthors, 2016: The resolution dependence of contiguous U.S. precipitation extremes in response to CO₂ forcing. *J. Climate*, **29**, 7991–8012, <https://doi.org/10.1175/JCLI-D-16-0307.1>.
- Vecchi, G. A., and G. Villarini, 2014: Next season's hurricanes. *Science*, **343**, 618–619, <https://doi.org/10.1126/science.1247759>.
- , and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, <https://doi.org/10.1175/JCLI-D-14-00158.1>.
- Vitart, F., 2006: Seasonal forecasting of tropical storm frequency using a multi-model ensemble. *Quart. J. Roy. Meteor. Soc.*, **132**, 647–666, <https://doi.org/10.1256/qj.05.65>.
- , 2009: Impact of the Madden–Julian Oscillation on tropical storms and risk of landfall in the ECMWF forecast system. *Geophys. Res. Lett.*, **36**, L15802, <https://doi.org/10.1029/2009GL039089>.
- , and T. N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Wea. Rev.*, **129**, 2521–2537, [https://doi.org/10.1175/1520-0493\(2001\)129<2521:SFOTSU>2.0.CO;2](https://doi.org/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, <https://doi.org/10.1029/2007GL030740>.
- Walsh, K. J. E., and Coauthors, 2016: Tropical cyclones and climate change. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 65–89, <https://doi.org/10.1002/wcc.371>.
- Wang, B., and J. C. L. Chan, 2002: How strong ENSO events affect tropical storm activity over the western North Pacific. *J. Climate*, **15**, 1643–1658, [https://doi.org/10.1175/1520-0442\(2002\)015<1643:HSEEAT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1643:HSEEAT>2.0.CO;2).
- Wing, A. A., and Coauthors, 2019: Moist static energy budget analysis of tropical cyclone intensification in high-resolution climate models. *J. Climate*, **32**, 6071–6095, <https://doi.org/10.1175/JCLI-D-18-0599.1>.
- Wittenberg, A. T., and J. L. Anderson, 1998: Dynamical implications of prescribing part of a coupled system: Results from a low-order model. *Nonlinear Processes Geophys.*, **5**, 167–179, <https://doi.org/10.5194/npg-5-167-1998>.
- , and Coauthors, 2018: Improved simulations of tropical Pacific annual-mean climate in the GFDL FLOR and HiFLOR coupled GCMs. *J. Adv. Model. Earth Syst.*, **10**, 3176–3220, <https://doi.org/10.1029/2018MS001372>.
- Xue, Y., and Coauthors, 2018: Spring land surface and subsurface temperature anomalies and subsequent downstream late spring–summer droughts/floods in North America and East Asia. *J. Geophys. Res. Atmos.*, **123**, 5001–5019, <https://doi.org/10.1029/2017JD028246>.
- Yang, X., and Coauthors, 2015: Seasonal predictability of extratropical storm tracks in GFDL's high-resolution climate prediction model. *J. Climate*, **28**, 3592–3611, <https://doi.org/10.1175/JCLI-D-14-00517.1>.
- Zhang, G., and Z. Wang, 2018: North Atlantic extratropical Rossby wave breaking during the warm season: Wave life cycle and role of diabatic heating. *Mon. Wea. Rev.*, **146**, 695–712, <https://doi.org/10.1175/MWR-D-17-0204.1>.
- , and —, 2019: North Atlantic Rossby wave breaking during the hurricane season: Association with tropical and extratropical variability. *J. Climate*, **32**, 3777–3801, <https://doi.org/10.1175/JCLI-D-18-0299.1>.
- , —, T. J. Dunkerton, M. S. Peng, and G. Magnusdottir, 2016: Extratropical impacts on Atlantic tropical cyclone activity. *J. Atmos. Sci.*, **73**, 1401–1418, <https://doi.org/10.1175/JAS-D-15-0154.1>.
- , —, M. S. Peng, and G. Magnusdottir, 2017: Characteristics and impacts of extratropical Rossby wave breaking during the Atlantic hurricane season. *J. Climate*, **30**, 2363–2379, <https://doi.org/10.1175/JCLI-D-16-0425.1>.
- , H. Murakami, R. Gudgel, and X. Yang, 2019: Dynamical seasonal prediction of tropical cyclone activity: Robust assessment of prediction skill and predictability. *Geophys. Res. Lett.*, **46**, 5506–5515, <https://doi.org/10.1029/2019GL082529>.
- , —, T. Knutson, R. Mizuta, and K. Yoshida, 2020: Tropical cyclone motion in a changing climate. *Sci. Adv.*, **6**, eaaz7610, <https://doi.org/10.1126/sciadv.aaz7610>.
- Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon. Wea. Rev.*, **135**, 3541–3564, <https://doi.org/10.1175/MWR3466.1>.
- 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**, 6653–6678, <https://doi.org/10.1175/2009JCLI3049.1>.
- , —, and —, 2012: Some counterintuitive dependencies of tropical cyclone frequency on parameters in a GCM. *J. Atmos. Sci.*, **69**, 2272–2283, <https://doi.org/10.1175/JAS-D-11-0238.1>.
- Zhou, W., M. Zhao, and D. Yang, 2019: Understand the direct effect of CO₂ increase on tropical circulation and TC activity: Land surface warming versus direct radiative forcing. *Geophys. Res. Lett.*, **46**, 6859–6867, <https://doi.org/10.1029/2019GL082865>.