Evaluating CFSv2 Subseasonal Forecast Skill with an Emphasis on Tropical Convection

NICHOLAS J. WEBER AND CLIFFORD F. MASS

Department of Atmospheric Sciences, University of Washington, Seattle, Washington

(Manuscript received 24 April 2017, in final form 6 July 2017)

ABSTRACT

This study examines the subseasonal predictive skill of CFSv2, focusing on the spatial and temporal distributions of error for large-scale atmospheric variables and the realism of simulated tropical convection. Errors in a 4-member CFSv2 ensemble forecast saturate at lead times of approximately 3 weeks for 500-hPa geopotential height and 5 weeks for 200-hPa velocity potential. Forecast errors exceed those of climatology at lead times beyond 2 weeks. Sea surface temperature, which evolves more slowly than atmospheric fields, maintains skill over climatology through the first month. Spatial patterns of error are robust across lead times and temporal averaging periods, increasing in amplitude as lead time increases and temporal averaging period decreases. Several significant biases were found in the CFSv2 reforecasts, such as too little convection over tropical land and excessive convection over the ocean. The realism of simulated tropical convection and associated teleconnections degrades with forecast lead time. Large-scale tropical convection in CFSv2 is more stationary than observed. Forecast MJOs propagate eastward too slowly and those initiated over the Indian Ocean have trouble traversing beyond the Maritime Continent. The total variability of simulated propagating convection is concentrated at lower frequencies compared to observed convection, and is more fully described by a red spectrum, indicating weak representation of convectively coupled waves. These flaws in simulated tropical convection, which could be tied to problems with convective parameterization and associated mean state biases, affect atmospheric teleconnections and may degrade extended global forecast skill.

1. Introduction

Subseasonal (weekly to monthly) prediction has great potential value for the energy sector, transportation, agriculture, and society in general. These temporal scales lie beyond the theoretical limit of deterministic weather predictability provided by the initial conditions (predictability of the first kind; Lorenz 1963, 1975), but are aided by predictability of the second kind, which originates from the slowly evolving boundary conditions, such as sea surface temperature (Chu 1999; Lorenz 1975). Thus, subseasonal forecast skill depends on both initial conditions and boundary forcing, with the latter increasingly dominating over the former as the period of projection lengthens. An important question is whether the signal from more predictable boundary forcing is consistently strong enough to outweigh atmospheric internal variability at subseasonal time scales (Kumar et al. 2011).

The deterministic and probabilistic performance of operational extended forecast systems targeting subseasonal time scales has been the topic of several recent studies. For example, Yuan et al. (2011) showed that the National Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2 (CFSv2; Saha et al. 2014), improves upon its predecessor (CFSv1) and compares favorably with the European Centre for Medium-Range Weather Forecasts (ECMWF) Variable-Resolution Ensemble Prediction System forecast system (Vitart 2004; Vitart et al. 2008) for temperature and hydrological prediction, although predictive skill is restricted to the first month. An intercomparison of three operational seasonal ensemble prediction systems (EPSs) revealed that, while ECMWF is the leader in terms of weekly averaged precipitation forecast skill, none of the models provide much skill after week 1, especially in the extratropics (Li and Robertson 2015). Roundy et al. (2015) highlighted the sensitivity of long-range forecast skill to temporal scale, spatial scale, geographic location, and lead time. The authors also found that predictive skill is dependent on

Corresponding author: Nicholas J. Weber, njweber2@atmos.washington.edu

DOI: 10.1175/MWR-D-17-0109.1

season and El Niño–Southern Oscillation (ENSO) phase. Other intraseasonal tropical phenomena, such as the Madden–Julian oscillation [MJO, after Madden and Julian (1971, 1972)], have also been shown to affect extended-range predictive skill (Jones et al. 2015). Other studies have evaluated extended model performance for specific temporal scales, geographic locations, atmospheric phenomena, and error metrics (e.g., Mo et al. 2012; Lang et al. 2014; Lynch et al. 2014), highlighting the need for a comprehensive forecast skill assessment.

There have been several attempts to improve the skill of operational seasonal/subseasonal forecast models through statistical postprocessing. Such efforts include applying multimodel ensemble means (Dutton et al. 2013; Becker et al. 2014; Wanders and Wood 2016), dynamical downscaling (Yuan and Liang 2011; Liu et al. 2016), and statistical downscaling (Li et al. 2009; Wu et al. 2012). While most of these techniques provide modest improvement upon global dynamical model skill, other work has shown that purely statistical forecasts are often comparable in skill to these global EPSs (Zheng and Frederiksen 2007; Seo et al. 2009). Together, the works described above suggest that operational subseasonal EPSs are unable to provide substantial deterministic skill at subseasonal to seasonal time scales, with postprocessing providing only modest improvement. There is either an innate lack of predictability at these time scales or model deficiencies undermine the potential to capitalize on the predictability supplied by slowly varying boundary conditions.

A potential failure mode of global models that might explain, at least in part, their lack of subseasonal skill is their unrealistic representation of tropical convection. It is well established that convection in the tropics has a considerable influence on global weather through atmospheric teleconnections (e.g., Trenberth et al. 1998; Alexander et al. 2002). Operational forecast systems, however, struggle to produce realistic distributions, intensities, and/or propagation characteristics of tropical convection. An example is shown in Fig. 1, which compares the derived tropical precipitation from the Tropical Rainfall Measuring Mission (TRMM) 3B42 product with that simulated by the NOAA/NWS Global Forecast System (GFS) model. The stark contrast between the observed and simulated precipitation demonstrates the well-documented tendency for models to produce more frequent rainfall of lower intensity than observed (e.g., Stephens et al. 2010).

Tropical convection and the associated teleconnections serve as an "atmospheric bridge" between relatively predictable tropical SSTs (and other slowly evolving phenomena like the MJO) and midlatitude

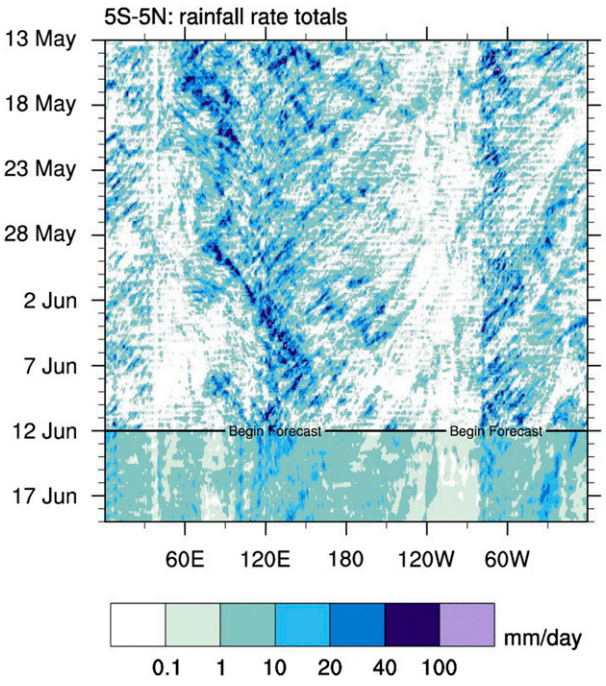

FIG. 1. Meridionally averaged (5°S–5°N) precipitation rate during the period 13 May–19 Jun 2017. Data above the black "Begin Forecast" line are from the TRMM 3B42 product and data below are from the 0000 UTC GFS forecast initialized on 29 Apr. (Figure courtesy of Michael Ventrice.)

circulations. Thus, with unrealistic convection, extratropical forecasts receive problematic low-frequency information from the tropics, limiting deterministic predictability to the period of influence from the initial conditions (roughly the first 1–2 weeks of the forecast).

This study assesses the limits of CFSv2 global forecast skill on several time scales and examines the structure and evolution of tropical convection produced by this modeling system. Biases in the model's mean state, and roles they might play in degrading the behavior of simulated convection and overall forecast quality, are also discussed.

2. Data and methods

a. The Climate Forecast System, version 2

The forecast model examined in this study, the NCEP CFSv2 (Saha et al. 2014), is the primary operational subseasonal/seasonal forecast system used in the United States. CFSv2 is a fully coupled atmosphere–ocean modeling system. The reforecast configuration (described below) for CFSv2 consists of the GFS at T126 (~0.937°) resolution for the atmospheric component and the Geophysical Fluid Dynamics Laboratory Modular Ocean Model at 0.25°–0.5° resolution for the ocean

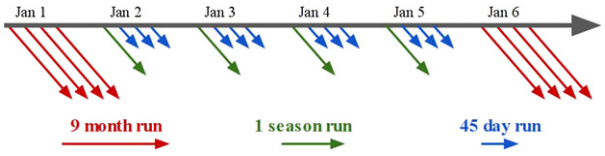

FIG. 2. CFSv2 reforecast configuration [adapted from Saha et al. (2014)].

component. The system also uses a two-layer sea ice model and the four-layer Noah land surface model. Saha et al. (2014) provide details of the operational and reforecast configurations of CFSv2.

CFSv2 was run over an extended reforecast period to provide long-term model performance statistics. While the operational configuration for CFSv2 includes 16 runs per day, only 4 runs were initialized daily during the reforecast period. This configuration, outlined in Fig. 2, consists of four 9-month runs initialized (at 0000, 0600, 1200, and 1800 UTC) every 5 days, and a 0000 UTC one-season (3 month) run with three subsequent 45-day runs (0600, 1200, and 1800 UTC) initialized during the intervening 4 days. CFSv2 reforecasts from the 1982–2008 period are evaluated. The verified forecasts in section 3 are 9-month, 4-member ensemble means initialized every 5 days from 1982 to 2008, providing 1951 forecasts. The CFS Reanalysis (CFSR; the initialization for the CFSv2 reforecasts) data are used to verify the reforecasts in this study.

b. *Verification metrics and temporal scales*

In this study, we apply three common error metrics to assess forecast performance: mean absolute error (MAE), anomaly correlation (AC), and bias. CFSR climatology (1982–2010) is used to compute anomalies for the forecasts and analyses. Persistence forecasts are created for various averaging periods by taking the preceding period's observed anomaly pattern (e.g., for a weekly forecast, the anomaly pattern from the past 7 days) and using that as the forecast for all future lead times. The CFSv2 reforecasts are evaluated for a range of averaging periods: 1 day, 1 week, 4 weeks, and 12 weeks. An X-week forecast is calculated at lead-Y (in days) by taking the average of all the daily forecasts from lead-Y to lead-$(Y + 7X)$. For example, the 4-week-averaged (''monthly'') forecast at a lead of 21 days is an average of all the daily forecasts from lead-21 to lead-49. The lead-dependent mean bias is removed from all forecasts evaluated in this study.

c. *Techniques for analyzing tropical convection*

Several approaches are used to evaluate CFSv2 tropical convection. The first is a simple compositing technique that determines extratropical teleconnection patterns associated with ''tropical convective events.'' These events are defined as days when some tropical region (e.g., the Indian Ocean), delimited by a 30° by 30° box, experiences anomalously active convection. To find such dates, daily outgoing longwave radiation (OLR) anomaly fields from CFSR are bandpass filtered (20–100-day periods) to highlight intraseasonal time scales, and then spatially averaged. Convective dates are recorded when a negative spatially averaged anomaly (for the box noted above) is in the lower quartile of the distribution. Only dates in boreal winter (DJF) are used, yielding a sample size of roughly 140 events. Analyzed and forecast patterns of tropical convection associated with these events are then determined by compositing CFSR and CFSv2 weekly averaged OLR anomalies about the recorded dates. Similarly, the extratropical teleconnection pattern is found by compositing weekly Northern Hemisphere (NH) Z500 anomalies for the same dates. Composites are made for several temporal lags from the convective dates to show the time evolution of the patterns. Running means from the preceding 120 days are removed from the weekly OLR and Z500 fields to filter out low-frequency variability [as in Wheeler and Hendon (2004) and Adames and Wallace (2014)]. Reanalysis data are used to compute these running means because the reforecasts have no information prior to initialization.

Lag regression is also employed to analyze propagation of tropical convection and related moisture anomalies; different anomaly fields (here, OLR and precipitable water) are regressed onto a standardized time series of area-averaged OLR at different temporal lags. The fields and time series use weekly averaged CFSR and CFSv2 grids with the low-frequency variability removed as detailed above. The spatial domain used to create the standardized OLR time series is the eastern Indian Ocean (5°S–5°N, 75°–85°E).

The methodology for calculating wavenumber–frequency OLR power spectra is adapted from the procedure in Yasunaga and Mapes (2012). Here, daily averaged OLR data from the NOAA Advanced Very High Resolution Radiometer (AVHRR; 1974–2013), CFSR reanalysis (1982–2008), and the last 90 days of each CFSv2 9-month reforecast (1982–2008) are used. First, the seasonal cycle is removed from the daily averaged OLR data by subtracting the CFSR climatology. Second, the data are partitioned into equatorially symmetric and antisymmetric components. The symmetric component is calculated simply by taking the mean OLR anomaly from 15°S to 15°N while the antisymmetric component is the NH mean minus the SH mean divided by 2 (where NH and SH are 0°–15°N and 15°S–0°, respectively). Third, the anomaly time series are

divided into 90-day segments with a 65-day overlap; the mean and linear trend of these segments are removed and time series are tapered to zero using a Hann taper algorithm. The CFSv2 segments are taken from the last 90 days of reforecasts initialized 25 days apart. Fourth, the power spectra are computed for each segment (via fast Fourier transform), averaged over all the segments, and smoothed with a 1–2–1 filter in the frequency dimension. Finally, to highlight the power peaks, these raw power spectra are normalized by dividing out the background spectrum as in Hendon and Wheeler (2008). This gives signal strength relative to the background spectrum, which is calculated by repeatedly applying a 1–2–1 filter to the raw power spectra, assuming a red noise process for each wavenumber.

3. Reforecast skill assessment

We examine the ability of CFSv2 to forecast several large-scale geophysical fields: 500-hPa geopotential height (Z500), 200-hPa velocity potential (CHI200), and sea surface temperatures (SST). These parameters were chosen for verification instead of surface fields because boundary layer details and resolution-dependent features like topography play less of a role in the quality of their forecast. Further, if large-scale flow patterns are skillfully predicted, then an improved forecast of a surface field, such as 2-m temperature, can be achieved through dynamical downscaling (Yuan and Liang 2011). Z500 is a useful analog for midlatitude circulation, CHI200 highlights large-scale tropical convective features, and SST is a lower boundary condition that drives many global-scale atmospheric circulations. Unless stated otherwise, the 1982–2008 mean bias was removed from all CFSv2 reforecasts.

Figure 3 shows the unconditional MAE (left column) and AC (right column) time series for the debiased Z500 (top row), CHI200 (middle row), and SST (bottom row) forecasts as a function of lead time and averaging period (colors). The Z500 MAE and AC were averaged over the NH midlatitudes (30°–70°N) while the skill metrics for CHI200 and SST were averaged over the tropics (10°S–10°N). The Z500 (solid lines) reforecast MAE plateaus roughly 2–3 weeks into the forecast for the shortest averaging periods and progressively earlier at longer time scales. This is reflected in the AC time series as well. Such flattening is an indication of error saturation: when the influence of the initial conditions fades because of the limitations of intrinsic predictability and systematic model errors. The ensemble-mean MAE of a perfect, infinitely large ensemble would saturate at the value of climatological forecast MAE (indicated by the broad dashed lines), while a single deterministic forecast would theoretically saturate at twice that value (Holton and Hakim 2012, 485–486). In this case—a small four-member ensemble with unconditional bias removed—the errors saturate at a value larger than the climatological MAE because of conditional biases within the forecasts (Murphy 1988; Goddard et al. 2013; Li and Robertson 2015). Thus, with this small ensemble, skill is lost when the forecast MAE (solid) exceeds the climatological MAE (broad-dashed), which is sooner than the forecast error saturates. For midlatitude Z500, the 1-day, 1-week, 4-week, and 12-week forecasts lose skill over climatology at leads of 14 days, 9 days, 5 days, and 0 days, respectively. The tropical CHI200 forecasts (Fig. 3, middle row) lose skill over climatology at similar leads, although errors saturate later in the forecast, indicating potential extended predictability from the initial conditions. The higher saturation AC values (nearly 0.3 for the 12-week forecasts) for CHI200 suggest a greater contribution from the boundary conditions to long-range predictability. Both types of predictability are likely higher in the tropics than in the extratropics, resulting in higher skill at long lead times (Zhu et al. 2014).

Tropical SST forecasts (bottom row) exhibit skill over climatology at longer lead times: 2–5 weeks, depending on the averaging period. Conditional (e.g., seasonal) SST biases are particularly strong (not shown), contributing to this loss of skill. Further, the SST MAE does not saturate in the time series shown here, implying that SST predictions from a larger ensemble would remain skillful for at least two months (for all averaging periods). This is expected as the slowly varying nature of SST renders it more innately predictable than atmospheric fields.

The "monthly" (4 week; purple lines) Z500 and SST forecast AC time series match well with the CFSv1 monthly Z200 and SST statistics presented in Kumar et al. (2011). While the magnitude of the error varies by season (e.g., higher Z500 MAE values in the winter hemisphere), a monthly breakdown of these time series (not shown) reveals that the error saturation times noted above are robust across all seasons. There is also a seasonal dependence for the AC scores, highlighting a slight increase in extended skill for all three parameters during the NH winter (not shown). The short-dashed lines in Fig. 3 represent the error statistics for persistence forecasts. For all parameters, averaging periods, and lead times, persistence forecasts are, on average, inferior to dynamical predictions.

The geographical distributions of MAE are similar across averaging periods (Fig. 4). MAE is only shown at a lead time of 3 weeks because the patterns only change in magnitude with lead time (i.e., larger errors at

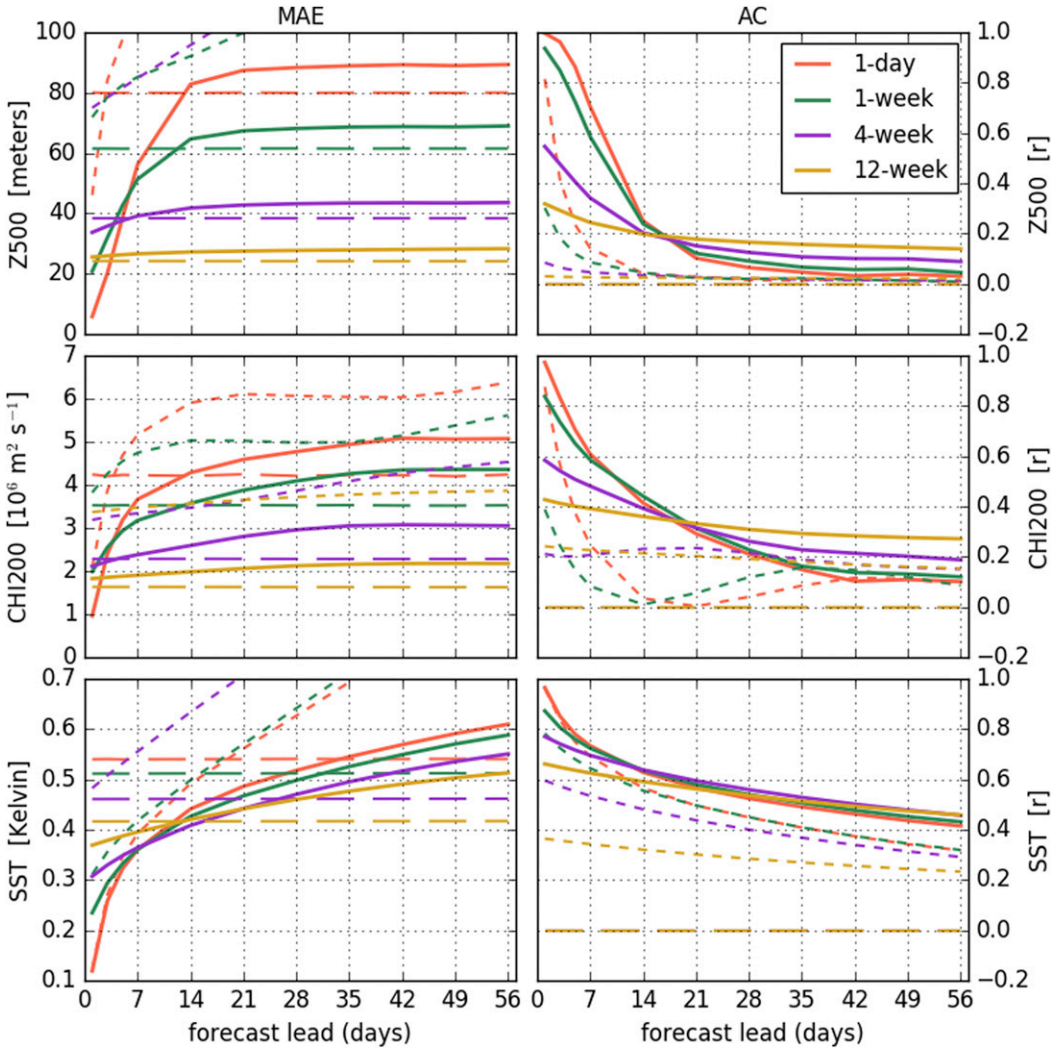

FIG. 3. Spatially averaged (left) MAE and (right) AC vs lead time for (top) NH midlatitude Z500, (middle) tropical CHI200, and (bottom) tropical SST forecasts during the 1982–2008 period. Solid lines are the CFSv2 9-month reforecasts, broad dashed lines are the CFSR climatology, and short dashed lines are persistence forecasts. Colors indicate the averaging time scale.

longer leads). For Z500 forecasts (top row), the largest errors are found in the midlatitude storm tracks. The CHI200 MAE distribution (middle row) highlights the tropics as the location of the largest errors, with maxima in the southeastern Pacific, west Pacific, and the Indian Ocean. The largest SST errors (bottom row) are found in the El Niño SST region of the east Pacific and in the western boundary current regions. For all three variables, the highest MAE values are found where observed internal variability is high.

Figure 5 shows the spatial distributions of the week-4 forecast AC (left column) and bias (right column). Here, only the weekly averaged forecast statistics are shown because the patterns of AC and bias vary only in magnitude for different averaging periods. In agreement with Fig. 3, the Z500 AC pattern reveals a general lack of skill in the midlatitudes for week 4. The only large positive AC values are in the tropics, where Z500 varies little, and in the central-eastern Pacific subtropics, perhaps in association with ENSO. In contrast, the week-4 CHI200 forecasts exhibit modest skill in the Maritime Continent region and in the central Pacific, likely associated with ENSO-related convection. SST forecasts produce much higher AC scores globally than the other parameters for week 4, due to the more slowly varying nature of SSTs. The equatorial central-eastern Pacific is a particularly skillful region because of the model's ability to reproduce ENSO variability.

Model bias (right column) is a useful skill metric since it informs about systematic forecast deficiencies. For

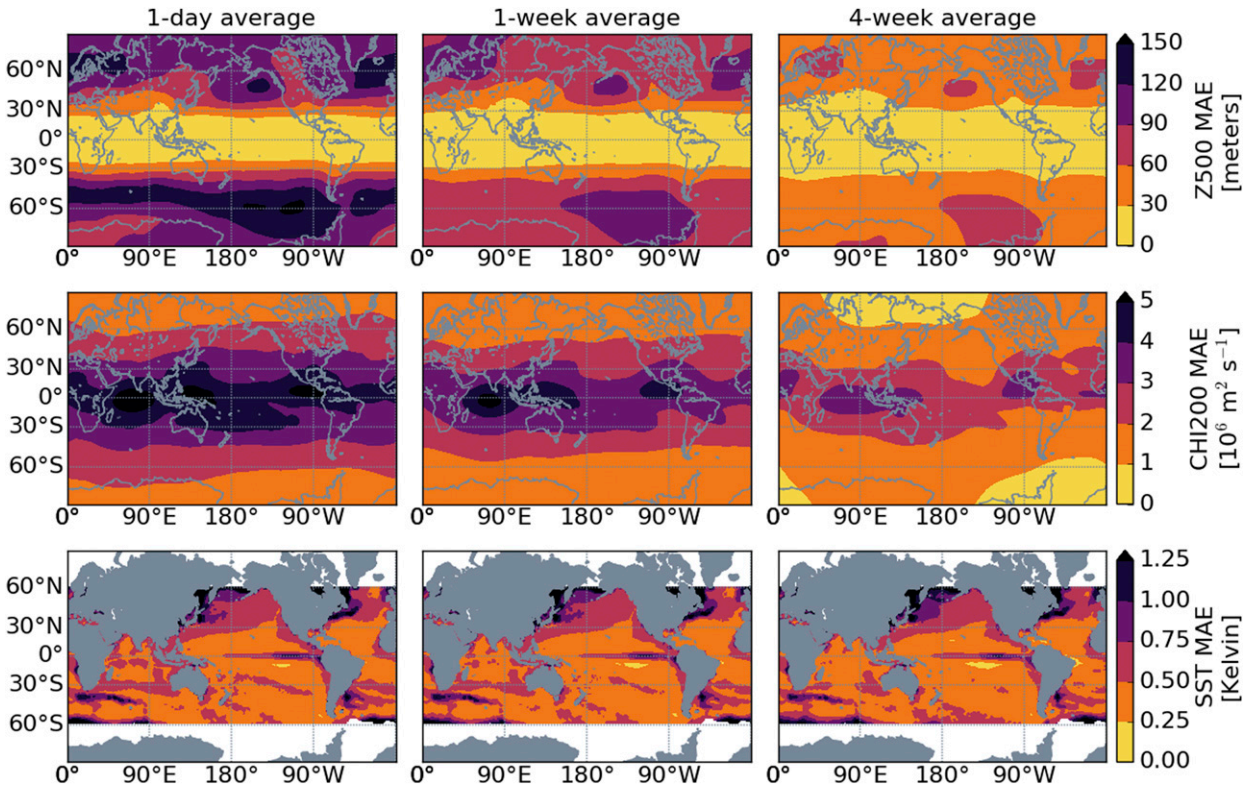

FIG. 4. CFSv2 MAE for the (left) daily, (middle) 1-week, and (right) 4-week averaged reforecasts of (top) Z500, (middle) CHI200, and (bottom) SST at a 3-week lead time. High latitudes are excluded from the SST charts because of noisy, high-amplitude errors from sea ice fluctuations.

Z500, the model tends to have a negative height bias in the tropics and midlatitudes, with near-zero or positive height biases near the poles. The forecasts also accentuate the Southern Ocean wave pattern and the Aleutian/Icelandic lows in the NH. These biases occur most strongly during the respective hemisphere's winter season (not shown). CFSv2 CHI200 exhibits a positive bias over equatorial landmasses (central Africa, Southeast Asia, and northern South America), reflecting underproduction of convection in these regions (Fig. 9). Further, the model produces too much convection in the Indian Ocean and equatorial west Pacific. The convective bias in the south-central Pacific is likely a manifestation of CFSv2's accentuated South Pacific convergence zone (Silva et al. 2014), a symptom of the double-intertropical convergence zone (ITCZ) issue common to many coupled climate models (Lin 2007).

The SST bias map (Fig. 5, bottom-right panel) reveals several interesting features. Equatorial SSTs are too cold in CFSv2, especially during boreal summer (not shown) when the Pacific cold bias extends west of the international date line. This overextension of the cold tongue is associated with the aforementioned double-ITCZ problem (Lin 2007). The warm SST biases off the west coasts of southeast North America, northern South America, and southern Africa reflect the model's inability to properly simulate persistent marine stratocumulus, which is confirmed by the excessive OLR (i.e., lack of clouds) in those regions (Fig. 9, top-left panel). Warm waters associated with western boundary currents are too close to shore in CFSv2. Finally, the alternating warm and cold biases in the Southern Ocean SSTs suggest an unrealistic representation of the climatological meridional meanders in the Antarctic Circumpolar Current (Chelton et al. 1990). An additional feature of these maps is the dissimilarity between the CHI200 biases and SST bias distributions. If errors in large-scale convection were forced primarily by misrepresentation of SST, warm (cold) SST biases would be superimposed with negative (positive) CHI200 biases. Because this is not the case, it appears that the biases in CFSv2's large-scale convection are not driven directly by errors in the ocean temperature.

While, on average, the skill of CFSv2 beyond two weeks is limited, there are circumstances under which the model's deterministic skill is enhanced.

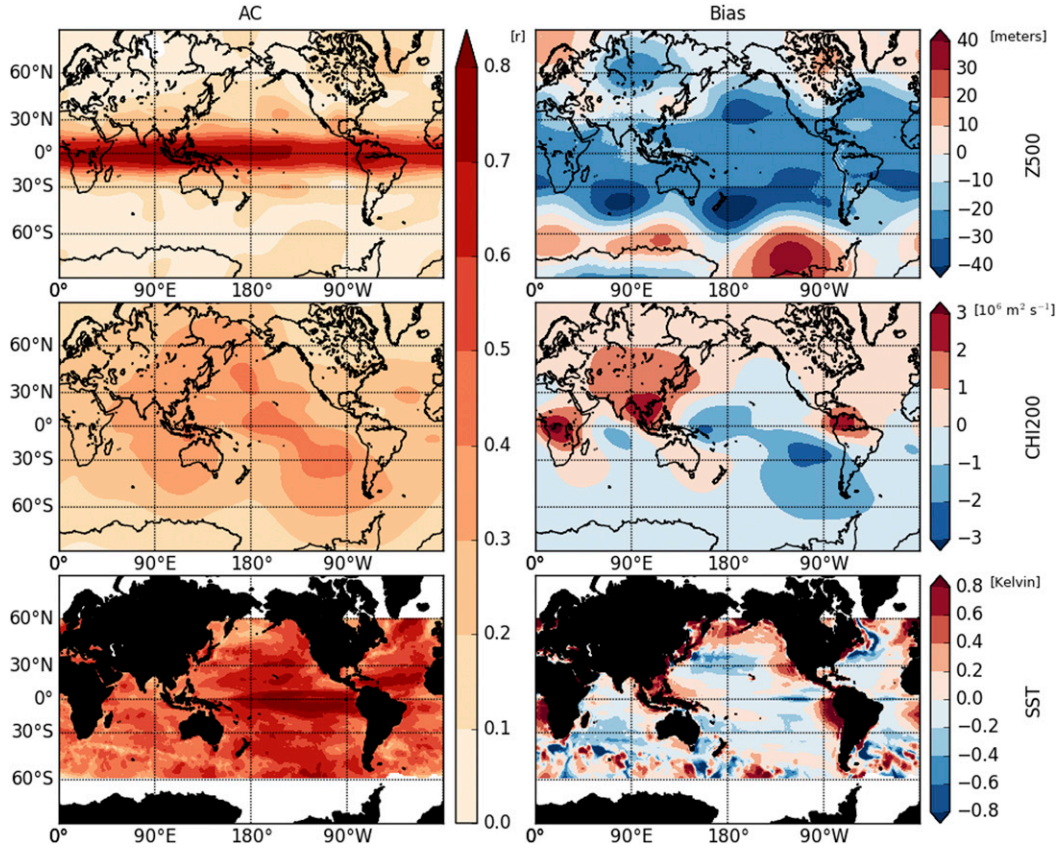

FIG. 5. (left) AC and (right) unconditional bias for the week-4 (top) Z500, (middle) CHI200, and (bottom) SST reforecasts.

The ENSO signal, for example, is well documented (e.g., Rasmusson and Wallace 1983) and can boost the accuracy of global forecasts (e.g., Kim et al. 2012). Time series and corresponding probability density functions (PDFs) of spatial AC scores for week-5 forecasts of NH Z500, tropical CHI200, and tropical SST as a function of ENSO phase are shown in Fig. 6. The strong 1982/83 and 1997/98 El Niño events are manifested as skill peaks in the three time series, suggesting that stronger ENSO events yield greater predictability. The PDFs reveal that, for the atmospheric parameters (Z500 and CHI200), El Niño events (when the CFSR Niño-3.4 index exceeds 1.0) significantly boost the skill of extended forecasts, especially in the tropics. Forecasts during La Niña events (when Niño-3.4 < -1.0), however, are virtually indistinguishable from other forecasts in terms of skill performance. SST skill, on the other hand, is improved during both El Niño *and* La Niña events. Thus, while both phases of ENSO are skillfully forecasted by the CFSv2, only El Niño events produce a predictable atmospheric signal that enhances global skill.

4. Forecast tropical convection

a. CFSv2 convective events and teleconnections

As noted previously, convection in the tropics can have a significant impact on global weather and climate through atmospheric teleconnections. Here, we investigate how the CFSv2 teleconnection patterns associated with convective anomalies change with lead time. As outlined in section 2, these teleconnection patterns are found by compositing weekly averaged boreal winter (DJF) Z500 anomalies using convective events in several tropical regions. An Indian Ocean (IO) convective event, for example, is recorded when the 20–100-day bandpass-filtered CFSR OLR anomaly averaged in the IO region (15°S–15°N, 70°–100°E) is within the lower quartile of the distribution. Weekly averaged OLR anomalies in the Indo-Pacific warm pool are also composited about the DJF convective dates.

The OLR composites and associated NH Z500 teleconnection patterns are shown for CFSR analyses (left column) and CFSv2 reforecasts (right column) initialized on the IO convective dates for different lead times (Fig. 7).

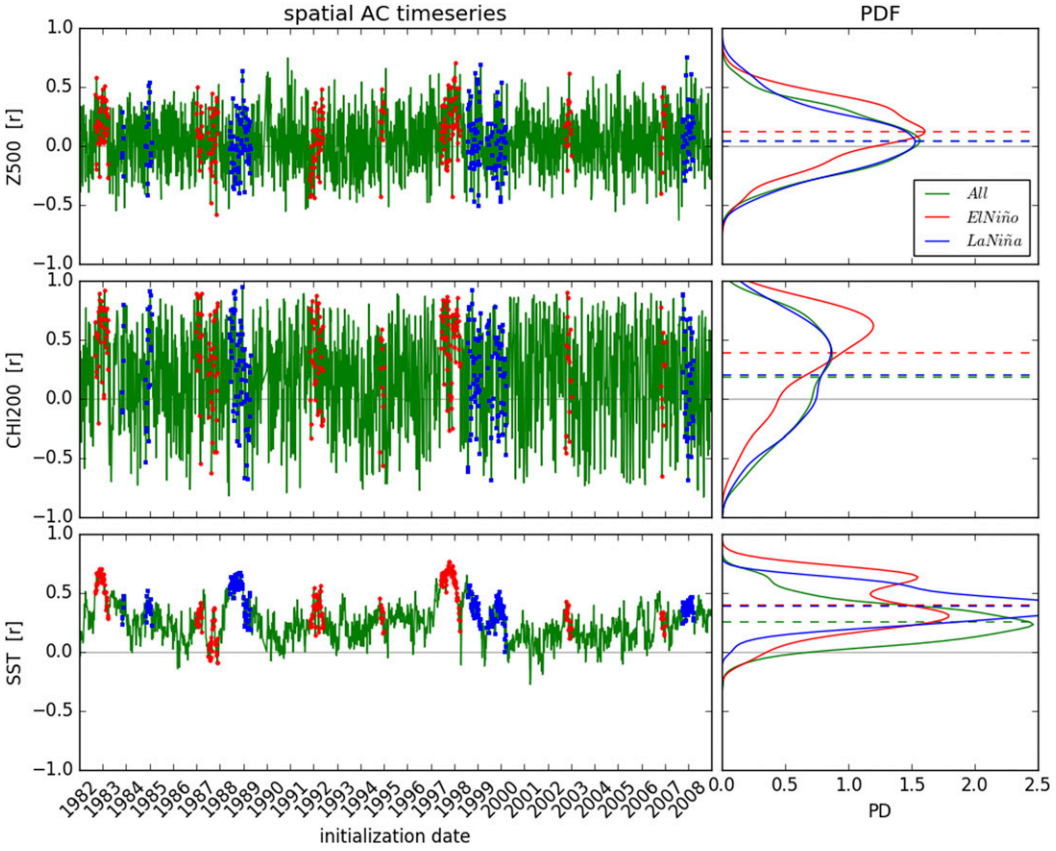

FIG. 6. (left) Spatial AC time series and (right) the associated probability densities (calculated with kernel density estimation) for week-4 (top) Z500, (middle) CHI200, and (bottom) SST reforecasts. Green represents all reforecasts in the 1982–2008 period, red dots are only the El Niño events, and blue dots are only the La Niña events. Horizontal dashed lines indicate the mean value of the PDF with the corresponding color.

The analyzed convection associated with these events (top-left panel, green box) looks as expected: convectively active in the IO region with a preceding suppressed anomaly in the west Pacific. Over time (left column, subsequent rows), the active convection slowly moves eastward over the Maritime Continent and into the west Pacific. The composited convective anomaly patterns from the CFSv2 reforecasts (right column, green boxes) reveal some important departures from the analyzed patterns. During weeks 1 and 2, the reforecasts exhibit much less equatorial Indian Ocean convection than observed in CFSR. Also, the eastward propagation of the anomalies in CFSv2 is slower than those in CFSR. The green r values show the steady decline of reforecast convective pattern skill with lead time.

The analyzed NH weekly Z500 teleconnection pattern (top-left panel, red box) consists of alternating ridges and troughs, reminiscent of the classic Pacific–North American (PNA) teleconnection pattern. Comparing the CFSR and CFSv2 teleconnections, we see that the reforecasts perform well through week 2 ($r > 0.8$), but fail to capture the week-3 and week-4 patterns.

The evolution of CFSv2 warm pool convection and NH teleconnection pattern correlation with lead time is shown in Fig. 8 for convective events in the IO, Maritime Continent (MC; 15°S–15°N, 110°–140°E), and west Pacific (WPac; 15°S–15°N, 150°E–180°). For all three regions, the reforecast skill at reproducing both the OLR pattern and Z500 pattern declines with lead time. The rate at which the r scores for the tropical and extratropical patterns decline is also dependent on where the convection occurs. For example, the teleconnection pattern for MC events degrades far more rapidly than for IO and WPac events.

b. *The tropical mean state in CFSv2*

Before examining the behavior of propagating tropical convection in CFSv2 forecasts, it is important to understand the systematic deficiencies in the model's mean state. These biases lend insights into physical processes that are handled improperly by the model.

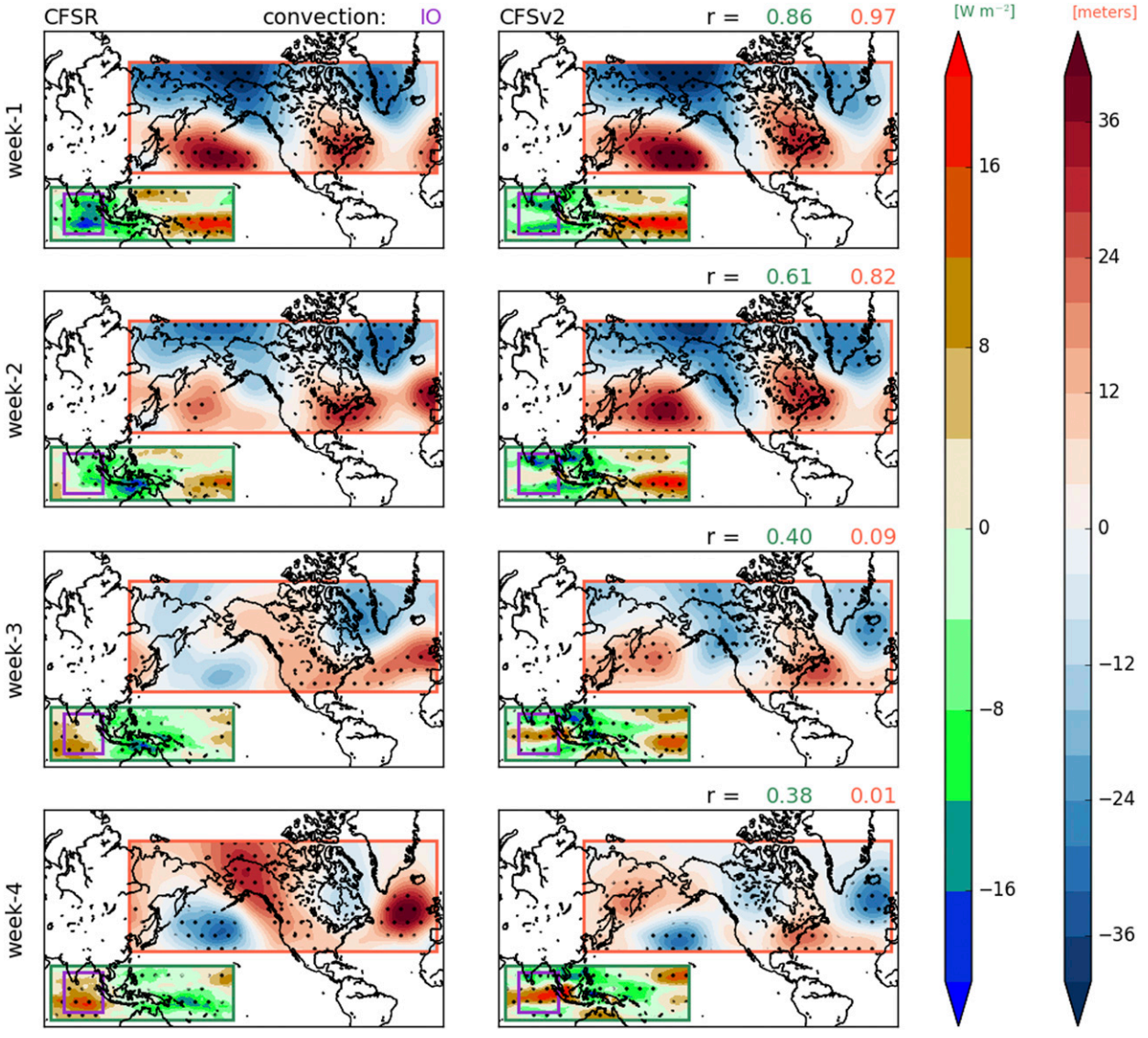

FIG. 7. Weekly averaged (left) CFSR and (right) CFSv2 OLR (green boxes) and Z500 (red boxes) DJF anomalies composited about IO convective events. Rows represent composites for the four weeks following IO convection onset. The purple box indicates the IO region over which the bandpass-filtered CFSR OLR anomalies were averaged to determine the convective dates. The green (red) *r* values above each CFSv2 panel is the pattern correlation coefficient between CFSR and CFSv2 for the composites in the green (red) boxes at each lead time. Stippling indicates a composite anomaly that is significantly different from zero at 95% confidence level.

The middle-right panel of Fig. 5 suggests that the forecasts produce too little convection over tropical land and too much over the ocean. The CFSv2 week-4 OLR and precipitation rate (PRATE) biases (Fig. 9; top two rows) mirror the idea of a systematic terrestrial/marine defect in simulated convection. From a large-scale perspective (left column), the forecasts generally produce too much convection/precipitation, except over tropical land regions and the equatorial east Pacific. The latter is associated with the exaggerated equatorial SST cold tongue and the accompanying double ITCZ, which is apparent in the excessive precipitation north and south of the equator. The Indo-Pacific warm pool appears to have a particularly large bias in convection/precipitation. An expanded view over this region (right column) reveals how stark the contrasting land–sea biases are; the individual islands of the Maritime Continent (as well as Southeast Asia and northern Australia) clearly exhibit a dry bias, with excessive precipitation over the surrounding waters.

The biases in tropical moisture are of opposite sign, as seen in the precipitable water (PWAT, or

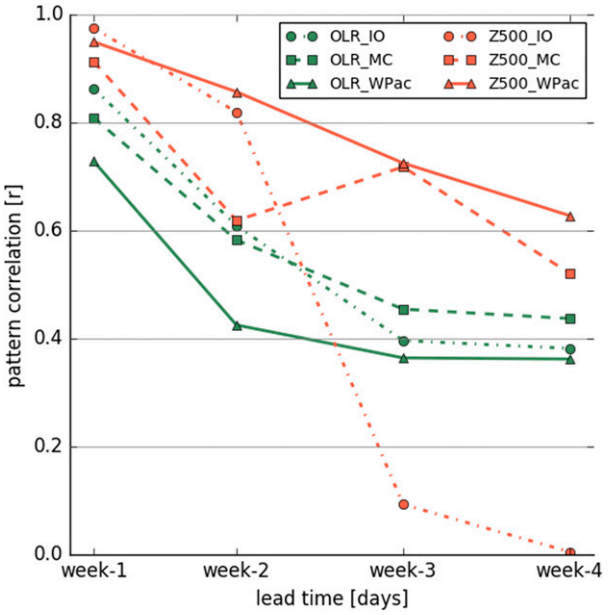

FIG. 8. Pattern correlation coefficients, as shown in Fig. 7, vs lead time for warm pool OLR (green) and NH Z500 (red) composited about IO (circles), MC (squares), and WPac (triangles) convective events.

column-integrated water vapor) bias maps (bottom row). CFSv2 reforecasts exhibit too little moisture over most of the deep tropics. The Maritime Continent region, where the climatological PWAT values are highest globally, is especially moisture-deficient over both land and sea. A possible mechanistic explanation for these biases in convection and moisture, and a discussion of their potential impacts on propagating convection and global circulation, is offered in section 5.

c. *Propagating tropical convection*

In this section, the nature of simulated CFSv2 tropical convection, and its dependence on lead time, is examined. For this analysis, only the 0000 UTC member of each 9-month reforecast is used to avoid "washing out" signals by ensemble averaging at long lead times. Averaging CFSv2 weekly CHI200 anomaly forecasts from 10°S to 10°N, we use Hovmöller diagrams to examine the fidelity of simulated, propagating convection evolution with lead time.

For a single forecast (initialized on 2 November 1987; Fig. 10), CFSv2 produces relatively realistic large-scale propagation early in the forecast, but slowly propagating or stationary anomalies at extended lead times. An examination of numerous forecast periods (not shown) reveals a tendency for propagating anomalies to weaken or halt when crossing the Maritime Continent region, a problem that has been previously documented in CFSv2 (e.g., Wang et al. 2014). At longer lead times, some of the previously highlighted convective biases, such as the dry bias over equatorial South America (~60°W), are evident in these individual forecasts.

The profound impact of ENSO on large-scale convection forecasts is apparent from the Hovmöller perspective. An example, the 1997/98 El Niño event, is shown in both CFSR analyses and the CFSv2 forecast initialized on 2 November 1997 (Fig. 11). The SST-forced signal of enhanced convection in the east Pacific and suppressed convection in the west Pacific is the dominant feature during this event. In this regard, the forecast performs quite well at all lead times. The few propagating features embedded within the background ENSO anomalies (e.g., the eastward-propagating negative anomaly during January–December) are weakly present or absent in the CFSv2 forecast. Since these transient anomalies are dwarfed by the ENSO signal, the model is able to produce a useful extended forecast of the CHI200 field during this El Niño event (and others, as shown in Fig. 6).

To secure a more general insight into the ability of CFSv2 to realistically simulate propagating convection, we constructed a reforecast composite about a common convective event: MJO initiation. The MJO onset dates are days when the AVHRR 20–100-day bandpass-filtered OLR anomaly in the Indian Ocean was more than one negative standard deviation from the mean, as in D. Kim et al. (2014). There were 163 such events during the 1982–2008 reforecast period. Weekly CFSR and bias-removed CFSv2 CHI200 Hovmöller composites about these MJO initiation dates (phases 2–3) are shown in Fig. 12. Reforecasts were composited based on their initialization date relative to the MJO convective dates described above. As the amount of time between forecast initialization and MJO onset increases, the forecast composite MJO signal becomes weaker. There is essentially no composite anomaly signal for the runs initialized more than 30 days before MJO initiation, consistent with the approximate 20-day CFSv2 MJO prediction skill limit suggested by Wang et al. (2014) and H.-M. Kim et al. (2014). In contrast, when the forecasts are initialized at the onset of MJO convection (0-day lag), the magnitude and zonal location of the enhanced Indian Ocean convection are well predicted. However, even at these very short lead times, the propagation speed of the convection is slow compared to the analyses, and the convection is unable to traverse the Maritime Continent into the west Pacific. These deficiencies in MJO propagation, the latter coined the "Maritime Continent barrier effect" (e.g., Inness and Slingo 2003), are not unique to CFSv2, having been

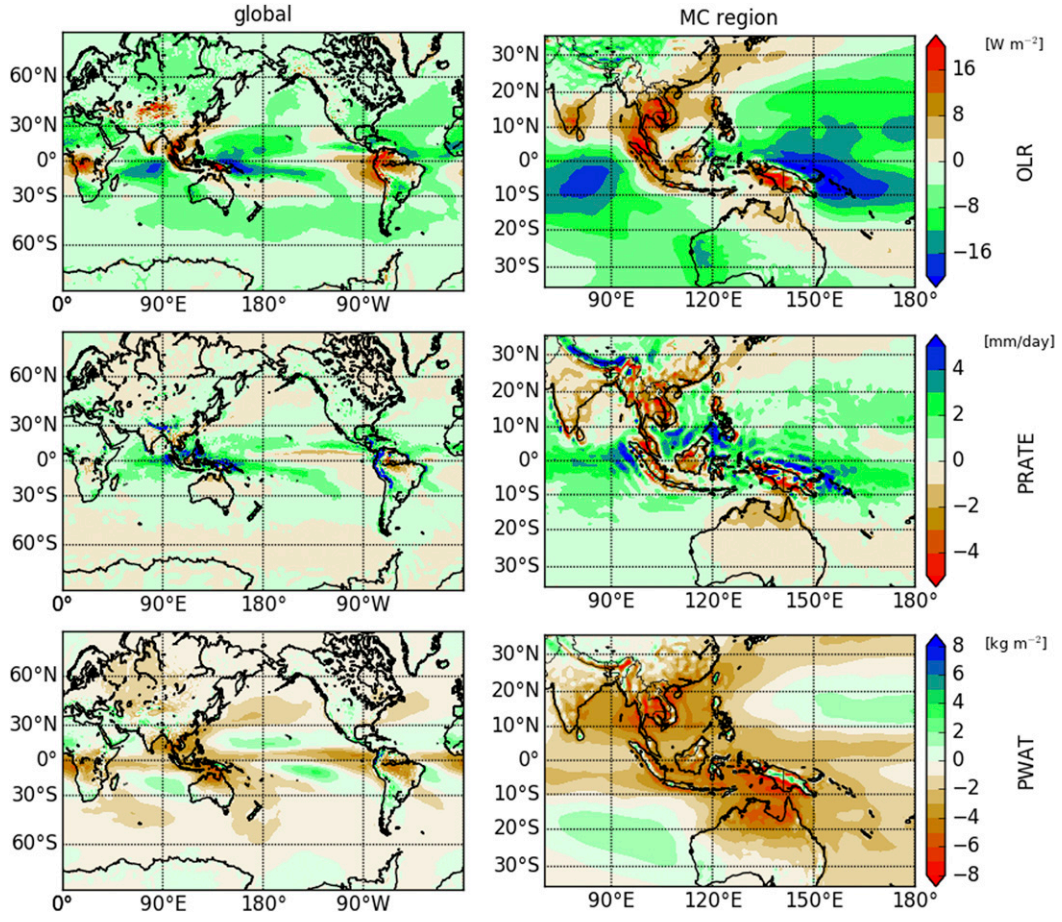

FIG. 9. Week-4 (top) OLR, (middle) PRATE, and (bottom) PWAT reforecast bias distributions for (left) the entire globe and (right) just the Maritime Continent region.

noted in several other global models (Fu et al. 2013; H.-M. Kim et al. 2014).

An alternative approach to analyzing MJO-like convection and propagation is performing lag regressions on the analyzed/predicted OLR and PWAT fields. The method, detailed in section 2, involves regressing, at different lags, weekly and meridionally averaged OLR and PWAT anomaly analyses and forecasts onto a time series of standardized weekly OLR anomalies averaged over the eastern Indian Ocean. The OLR and PWAT fields, and the corresponding standardized OLR time series, were taken from CFSR and CFSv2 reforecasts at various lead times. The OLR–OLR (shading) and OLR–PWAT (contours) lag regressions show a clear degradation of the temporal behavior of convection as lead time increases (Fig. 13). By week 2, the reforecast lag regressions indicate nearly stationary convection. At subsequent lead times, the regression values over the Indian Ocean increase further, with only slight indication of propagation eastward over and beyond the Maritime Continent (as in the CFSR regression map).

To estimate an "effective" MJO propagation phase speed, a linear fit was computed for three points, each indicating a local longitudinal maximum in the regression values temporally averaged over 10 lag days [adapted, though significantly altered, from Adames and Kim (2016)]. The slopes of the fit lines (recorded above each panel, in $\mathrm{m\,s^{-1}}$) indicate the propagation phase speeds computed from the OLR–OLR (green) and OLR–PWAT (pink) lag regression maxima. This method produces an MJO phase speed in CFSR that agrees with the observed value of $5\,\mathrm{m\,s^{-1}}$. The week-1 CFSv2 reforecasts exhibit a similar propagation speed, although slightly slower. At other lead times, the computed phase speed ($<1\,\mathrm{m\,s^{-1}}$) is drastically slower than observed because it is impacted not only by the change in propagation of MJO convection packets, but also the Maritime Continent barrier effect and the model's tendency to produce more stationary convection at longer lead times.

We can broaden the scope from MJO events to all propagating features by decomposing the analyzed and

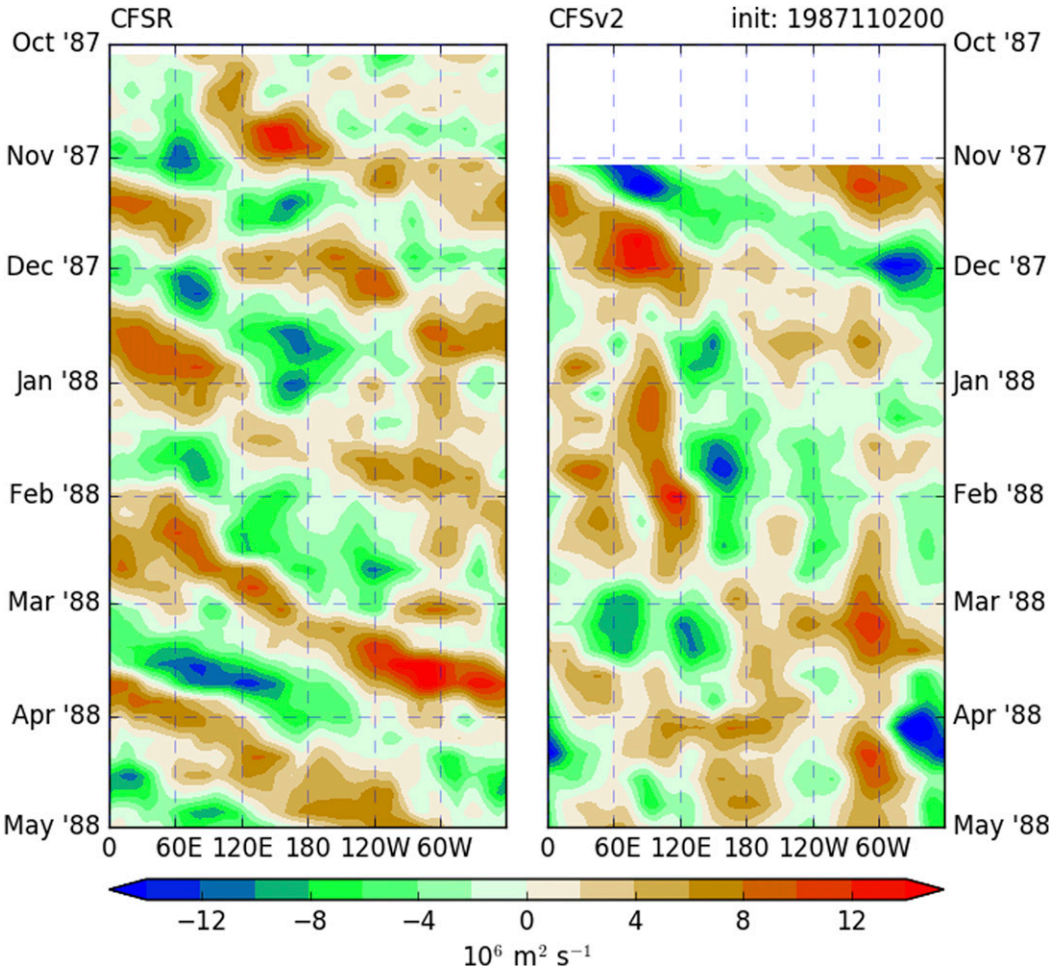

FIG. 10. Weekly CHI200 anomalies averaged from 10°S to 10°N during the October 1987–May 1988 period. CHI200 anomaly fields are from (left) CFSR and (right) one single CFSv2 reforecast initialized at 0000 UTC 2 Nov 1987.

simulated daily convection variance into space–time power spectra (as in Wheeler and Kiladis 1999). For the CFSv2 reforecasts, the last 90 days of each 9-month run was used for power spectrum computation, allowing an analysis of the spectral qualities of the model in ''free-running mode,'' well beyond the influence of the initial conditions. See section 2 for more details.

The raw power spectra for both the symmetric and antisymmetric component of the OLR (Fig. 14) show the differences in the spectral qualities of observed (AVHRR; top left), analyzed (CFSR; top right), and simulated (CFSv2; bottom left) convection. CFSR and CFSv2 both exhibit more overall variance in daily tropical OLR than observed, with power generally higher at all wavenumbers and frequencies compared to AVHRR. In contrast to both the observations and analyses, the free-running CFSv2 produces too much low-frequency, or slow-moving, tropical convection, as seen in the higher concentration of power in the lower portion of the diagram. To illustrate, the ratio between raw CFSv2 and CFSR power (bottom-right panel) indicates increased power at low frequencies and decreased power at high frequencies. The east–west ratio (EWR), a common metric used to gauge the presence of the MJO in general circulation models, is the ratio of the raw power in the 30–90-day period range between wavenumbers 1 through 6 (eastward propagating) and −1 through −6 (westward propagating). The observed OLR EWR is roughly 1.9, while the CFSR and CFSv2 each have an EWR just over 1.4. This implies that both the analyses and forecasts contain too much westward-propagating convection in the MJO wavenumber–frequency domain. In fact, Fig. 14 suggests that the analyses and forecasts exhibit too much convection at all wavenumbers and frequencies, especially for westward-propagating features. The differences between the observed (AVHRR) and analyzed (CFSR) convection could be caused by quickly developing

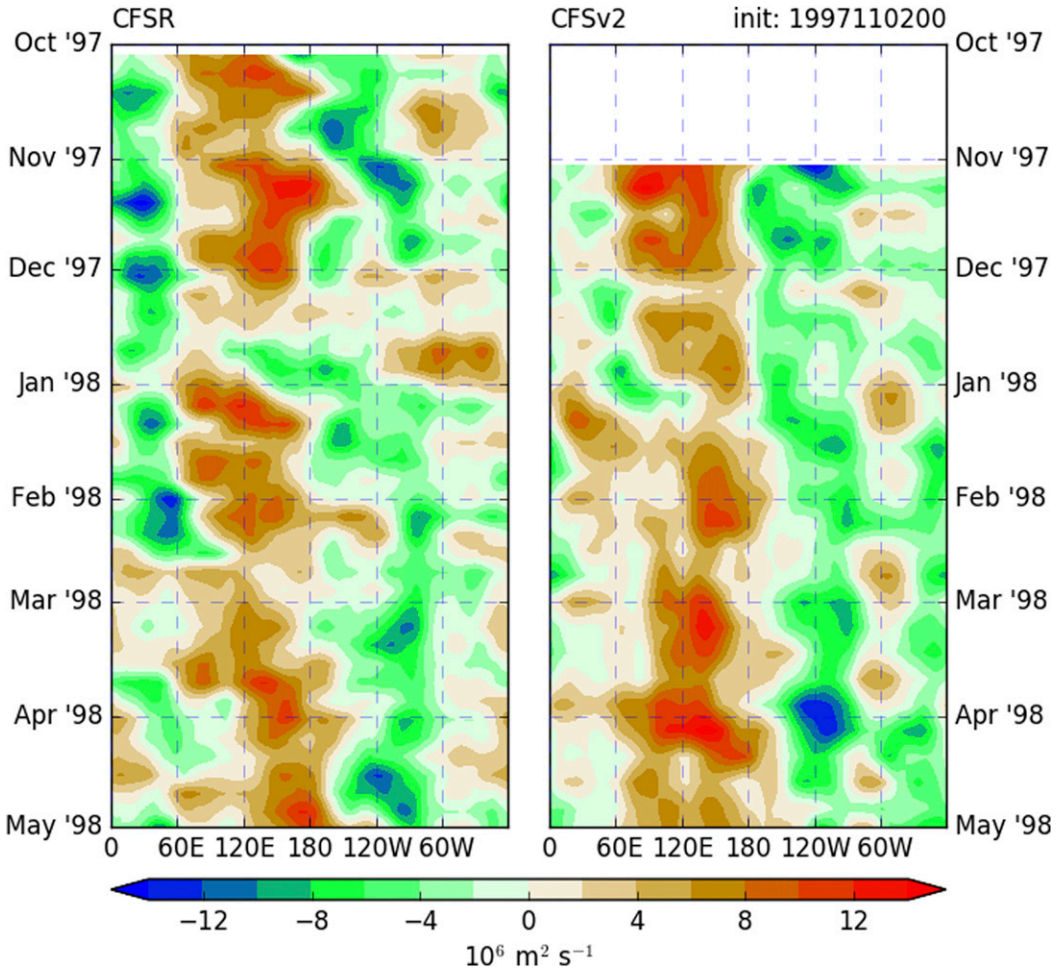

FIG. 11. As in Fig. 10, but for the October 1997–May 1998 period and (right) the CFSv2 reforecast was initialized at 0000 UTC 2 Nov 1997.

errors/biases imparted by parameterized processes in the model.

Figure 15 shows the power spectra of the symmetric component of the tropical OLR, normalized by the background "red" spectrum. This allows one to examine the convective phenomena that are distinct from the background noise (e.g., convectively coupled tropical waves). For example, prominent structures in the AVHRR spectrum include the MJO and Kelvin waves for eastward-propagating convection, and equatorial Rossby and tropical depression waves for westward-propagating convection. The main features that distinguish the CFSR spectrum from the AVHRR spectrum are slightly reduced power in the MJO and Kelvin waves (i.e., the eastward-propagating features) and increased power in the westward-propagating features, as mentioned above. The free-running CFSv2, in comparison to the CFSR, lacks distinct power peaks for the MJO and Kelvin waves. Moreover, both eastward- and westward-propagating features have greatly reduced signal strength compared to the background spectrum, indicating a convective spectrum more fully described by red noise than the analyzed/observed spectra. These deficiencies in the model's ability to produce realistically propagating tropical convection at longer lead times could be an important factor in the limited global subseasonal forecast skill described in section 3.

5. Discussion

a. Parameterized convection and model biases

The convective parameterizations used in global modeling systems have been shown to poorly represent certain aspects of convection, such as propagation and the diurnal cycle (Yang and Slingo 2001; Guichard et al. 2004). Such deficiencies in convection schemes could affect long-range teleconnections and thus degrade extratropical forecasts. Consistent with this hypothesis,

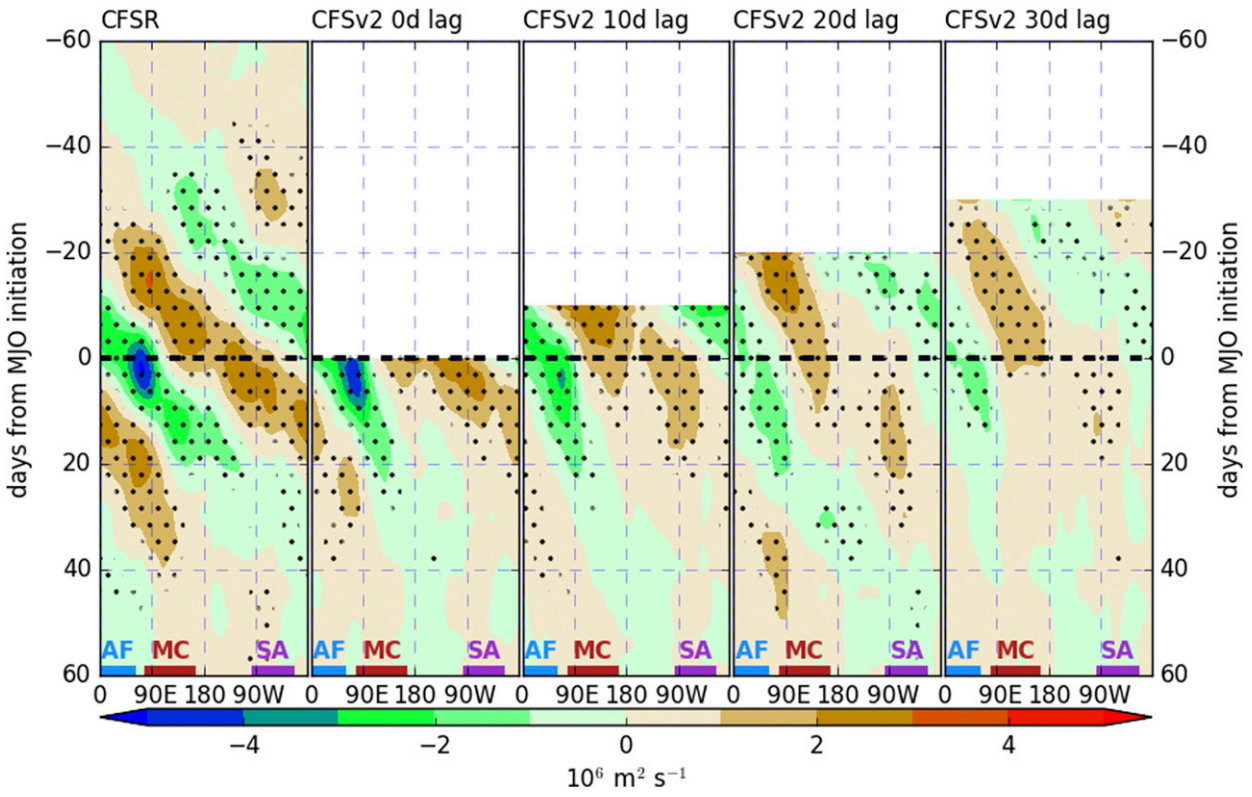

FIG. 12. Weekly CHI200 anomalies averaged from 10°S to 10°N composited about 163 MJO initiation dates (bold dashed line) during the 1982–2008 period. (left) CFSR reanalyses and subsequent columns are composites of CFSv2 forecasts initialized (from left to right) 0, 10, 20, and 30 days before MJO onset. Stippling indicates statistical significance at a 95% confidence level (via the Student's *t* test). Blue, red, and purple bars above the *x* axes denote the approximate locations of equatorial Africa, Maritime Continent, and South America, respectively.

Vitart (2014) suggested that improvements in extended ECMWF global forecast skill over the past decade were primarily due to changes in model physics (and the subsequent improvement of simulated MJOs) rather than resolution enhancements.

The convective biases shown in Fig. 5 may be rooted in deficiencies of cumulus parameterization in CFSv2. In their modeling study with the Met Office Unified Model (MetUM), Martin et al. (2017) proposed that the poor representation by convective parameterizations of small-scale processes (e.g., the terrestrial diurnal cycle) and intraseasonal variability produces the dry and wet biases over tropical land and sea, respectively. Specifically, the unrealistically small fractional entrainment/detrainment rate in convective parameterizations too readily permits the triggering of deep convection without the slow buildup of shallow cumulus; for the diurnal cycle, this produces an overly amplified convective peak that occurs too early in the day, compared to reality (e.g., Betts and Jakob 2002; Wang et al. 2007). In the case of CFSv2 and other models, such as MetUM, the issues with convective entrainment and the terrestrial diurnal cycle may be responsible for the robust dry bias over tropical land.

The convective biases in CFSv2 (Fig. 9, top two rows) are apparent at short leads and time scales, supporting the notion that they are rooted in parameterized processes. The temporal evolution of bias is analyzed by averaging the absolute value of the daily biases throughout the tropics (15°S–15°N) and normalizing by the value at an arbitrary lead time (here, 35 days). These time series of normalized mean absolute bias (Fig. 16) highlight a dichotomy between the evolution of convection-related biases (CHI200, OLR, PRATE) and that of other biases (Z500, SST, PWAT). Namely, we see that the biases in parameters used to diagnose tropical convection develop very early in the forecast and change little at longer lead times. Conversely, the Z500, SST, and PWAT biases evolve almost linearly as lead time increases. This suggests that cumulus parameterization has a significant, almost immediate, impact on the model's tropical mean state.

We hypothesize that the CFSv2 dry bias over tropical land (associated with downward motion), rooted in the

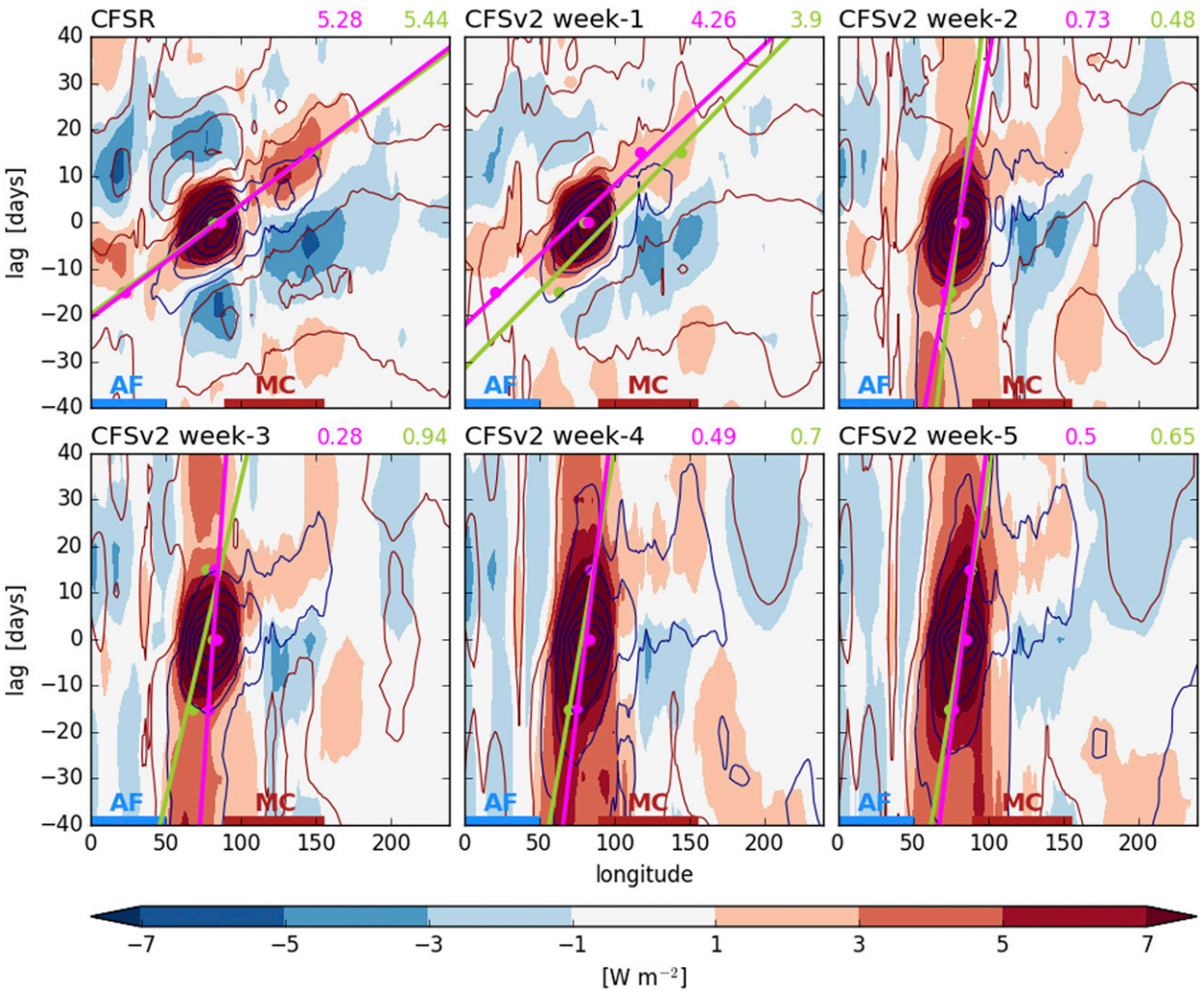

FIG. 13. Lag regression of 1982–2008 weekly and meridionally (5°S–5°N) averaged OLR (shading) and PWAT (contours; every 0.6 kg m^{-2} excluding 0.0) from CFSR and CFSv2 onto an IO-averaged standardized weekly OLR anomaly time series. Green (pink) lines and panel text indicate the phase speed (m s^{-1}) of local lag regression maxima (minima) for the OLR (PWAT) fields. See text for details.

cumulus parameterization's poor handling of entrainment and the terrestrial diurnal cycle, forces a compensating wet (upward motion) bias over tropical oceanic regions to conserve mass. A portion of this wet oceanic precipitation bias may also be attributed to the model's tendency to produce light tropical rain too frequently (Fig. 1), a well-documented issue that is also associated with the underestimated fractional entrainment rate in cumulus parameterizations (Demott et al. 2007; Stephens et al. 2010). The negative moisture bias is seemingly tied to the positive precipitation bias, based on the spatial coherence between the two (Fig. 9, bottom two rows). Physically, the increased total precipitation in the simulated tropical mean state creates a negative moisture tendency if the normalized gross moist stability, or the net export of moist energy per unit of convection, is positive in the CFSv2 atmospheric model. The normalized gross moist stability was found to be negative in the predecessor to the CFSv2, the CFS (Raymond and Fuchs 2009). If this applies to this generation's model as well, then increases in CFSv2 precipitation are not fully compensated by increases in vertical moisture advection, because of the unrealistic vertical mass flux profiles, which are primarily constrained by model physics, leading to a net gradual depletion of column moisture. This phenomenon is depicted in the slow evolution of the PWAT bias relative to the PRATE bias (Fig. 16, pink and purples lines). Hazra et al. (2016) suggest that this negative moisture bias could be linked to a cold tropospheric temperature bias caused by an upper-level cloud deficit in CFSv2. In any case, there is considerable evidence presented here and elsewhere

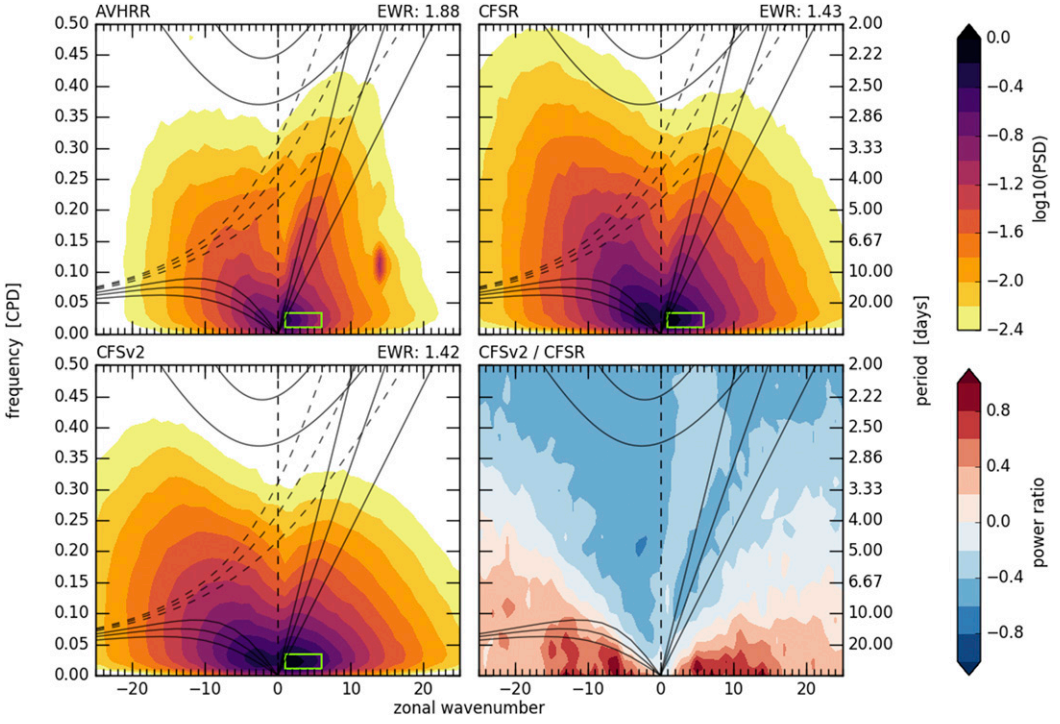

FIG. 14. Wavenumber–frequency diagrams for daily (top left) AVHRR, (top right) CFSR, and (bottom left) CFSv2 OLR data. Shading is the common logarithm of the raw power spectral density. The green box in each panel indicates the range of wavenumbers and frequencies used to calculate the EWR above the top-right corner. (bottom right) The raw symmetric CFSv2 power relative to that of CFSR (−1). Gray lines are the dispersion curves for Kelvin, equatorial Rossby, inertia gravity, and mixed Rossby–gravity (dashed) waves with equivalent depths of 50, 90, and 200 m. The AVHRR power peak at zonal wavenumber ~14 and frequency ~0.12 is an artifact of satellite aliasing.

suggesting that model biases in tropical convection and moisture are generated by deficiencies in convective parameterizations.

Apart from producing errors directly, the forecast biases outlined in this paper can have many negative impacts on atmospheric prediction. The stationary CHI200 anomalies associated with the biases, for instance, can act as Rossby wave sources and excite unrealistic stationary extratropical teleconnections, thus degrading midlatitude circulation forecasts. Tropical biases could also degrade the simulation of propagating convective features like the MJO. In their study, D. Kim et al. (2014) suggested that strongly propagating MJO events tend to follow positive OLR (suppressed convection) anomalies in the equatorial west Pacific, because the Rossby wave response induced by the dry anomaly is conducive to positive moisture advection. A negative west Pacific OLR bias, as noted in CFSv2, might impede the formation dry anomalies and thus inhibit the eastward propagation of MJO convection across the Maritime Continent. This theory was also posited by Kim et al. (2016) in their analysis of the ECMWF modeling system.

The biases in moisture outlined above could also have a detrimental impact on MJO development/propagation. In the presence of a tropical heating anomaly such as the MJO, the horizontal advection of mean moisture by anomalous winds comprises a substantial fraction of the overall moisture tendency, on which the growth and propagation of the MJO is dependent (Adames and Kim 2016). The aforementioned CFSv2 PWAT biases act to weaken the meridional and zonal gradients in mean moisture, thus reducing the magnitude of horizontal moisture advection by MJO circulation. The resultant decline in the overall moisture tendency degrades the growth and propagation of simulated MJOs. The combination of the effects listed above might explain the MJO's slow propagation speed and the Maritime Continent barrier effect in CFSv2 and other models.

b. *Moving forward: Convection-permitting simulations*

As noted above, model biases in precipitation, moisture, and tropical convection may be associated with flaws in cumulus schemes. It follows that, to

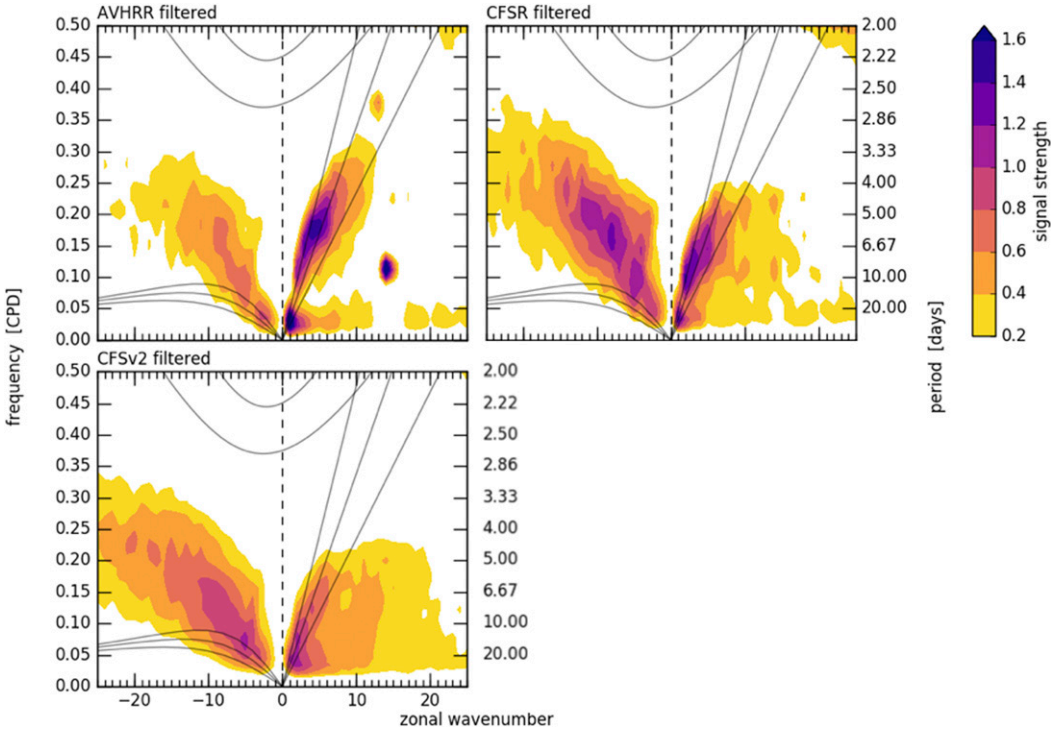

FIG. 15. As in Fig. 14, but for only the symmetric component of the OLR and the power spectra are normalized by the background red spectra. Contours are signal strength relative to the background spectrum.

alleviate these biases, convective parameterizations must either be significantly improved or relinquished altogether. Indeed, improving aspects of traditional convective parameterizations such as the fractional entrainment/detrainment rate does benefit the forecasts' representation of the tropical diurnal cycle (Wang et al. 2007). Superparameterization, or the embedding of 2D cloud-resolving models within each model grid box (Grabowski 2004), has been shown to improve the mean state of CFSv2 (Mukhopadhyay et al. 2016). Explicitly simulating convection by foregoing the use of cumulus parameterization at high resolution is a more direct way of resolving physical processes. Recent literature shows that, even at relatively coarse resolution (>10 km), the use of explicit convection can improve model precipitation distributions (Holloway et al. 2012) and produce MJOs with more fidelity than similar runs with parameterized convection (Pilon et al. 2016). At finer resolutions, explicit convection can more accurately reproduce the rainfall diurnal cycle over land (Sato et al. 2009), improve the distribution of tropical deep convection (Inoue et al. 2008), enhance the propagation characteristics of continental convection (Davis et al. 2003), and better represent propagating equatorial features like the MJO (Miura et al. 2007; Miyakawa et al. 2014; Wang et al. 2015). The apparent ability of convection-permitting simulations to mitigate the biases associated with parameterized convection makes them an attractive option for subseasonal forecasting. By alleviating tropical biases and improving representation of convective features like the MJO, such simulations may be able to more realistically transfer valuable information from relatively predictable, low-frequency tropical phenomena and SST anomalies into the extratropics through atmospheric teleconnections, thus improving subseasonal predictive skill globally.

Last, while many of the global models' shortcomings have been tied to cumulus parameterization, other parameterized processes may also contribute to poor extended forecast skill. For example, microphysics parameterizations (Van Weverberg et al. 2013) or subgrid mixing schemes (Holloway et al. 2012) can significantly impact the fidelity of the forecast. The improvement of both the microphysics and convection parameterizations in CFSv2 was shown to improve the predicted distribution of tropical convection (Abhik et al. 2017). We focus here on the issues with cumulus parameterization because convection, unlike microphysical or turbulent processes, is a phenomenon that we can reasonably resolve in global models using today's computing resources.

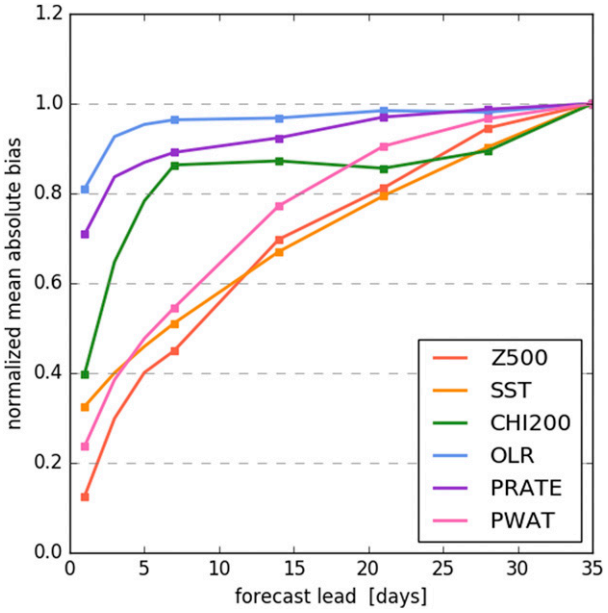

FIG. 16. Time series of absolute daily reforecast bias averaged from 15°S to 15°N for the tropical variables (SST, CHI200, OLR, PRATE, and PWAT) and from 30° to 70°N for Z500. Values are normalized by the value at lead 35.

6. Summary and conclusions

This paper presents an objective assessment of CFSv2 deterministic forecast skill on subseasonal time scales. The 500-hPa geopotential height (Z500) and 200-hPa velocity potential (CHI200) both maintain skill over climatology for the first 1–2 weeks of the forecasts, while monthly to seasonal forecasts offer little or no skill over climatology. Error saturation occurs later in the CHI200 forecasts than for Z500, indicating extended potential skill in the tropics. SST, which varies much more slowly than Z500 and CHI200, is skillful over climatology for several weeks and exhibits potential skill at lead times greater than 2 months. The discrepancy between the times of error saturation and skill loss to climatology can be explained by conditional biases [i.e., sources of error other than mean (unconditional) bias and internal variability (Goddard et al. 2013)].

The deterministic skill limits noted above are relatively insensitive to season, though forecasts initialized in boreal winter maintain slightly higher anomaly correlation (AC) values than those in boreal summer. Furthermore, there is increased forecast skill at extended lead times during active El Niño events. The global patterns of mean absolute error (MAE), AC, and bias do not change significantly with lead time or the temporal scale at which the forecast is averaged. The model's mean state is shown to be too dry over tropical land and too wet over the Indian Ocean and western tropical Pacific. The biases in simulated tropical convection (seen in the CHI200, OLR, and PRATE fields) do not coincide with SST biases, implying that mean state errors in tropical convection arise from issues originating within the atmospheric component of CFSv2.

As tropical convection and associated teleconnections are known to play an important role in global atmospheric circulation, we investigated the realism of simulated tropical convection and its evolution with lead time. For periods of active convection in the Indo-Pacific warm pool, the simulated extratropical teleconnection patterns, and associated patterns of tropical convection, degrade in predictive skill as lead time increases. Hovmöller diagrams of individual reforecast periods reveal that CFSv2 tends to produce slowly propagating or stationary convective anomalies, a problem that worsens with increasing lead time. This tendency is illustrated by compositing forecasts about MJO onset events. As found in other modeling studies, simulated MJOs exhibit a slower propagation speed than observed and are generally unable to move east of the Maritime Continent. The free-running (i.e., beyond the impact of the initial conditions) CFSv2 wavenumber–frequency power spectrum features more westward-propagating convection than observed, exhibits more power at lower frequencies compared to analyses and observations, is more "red" than the analyzed/observed spectra, and lacks distinct power peaks for the MJO and Kelvin waves.

Some of the forecast deficiencies noted above might be explained by deficiencies in the tropical mean state in CFSv2. An unrealistic entrainment/detrainment rate and perhaps other flaws with conventional convective parameterization induce a dry convective bias over tropical land, where the entrainment-sensitive diurnal cycle in precipitation is quite strong. The model conserves mass; thus, an accompanying wet bias over tropical oceanic regions develops concurrently. The model's tendency to produce unrealistically light/frequent tropical convection, an issue also rooted in deficiencies within cumulus parameterizations, likely contributes to this wet precipitation bias as well. Because tropical precipitation in the CFSv2 atmospheric model tends to deplete column-integrated moisture, the wet bias in precipitation forces the growth of a dry moisture bias in the deep tropics. These tropical biases influence the quality of global forecasts in several ways: by creating extratropical circulation biases through stationary wave teleconnections, impeding MJO development and propagation by reducing moisture advection, and deteriorating the structure of other convective phenomena and their associated extratropical teleconnections.

It is suggested that convection-permitting simulations, which have been shown to reduce biases in the tropical mean state and improve the realism of propagating convection in both the tropics and midlatitudes, may perform better on subseasonal time scales than current operational global models. Future work should include running extended, global simulations with convection-permitting resolution in the tropics, examining the resultant behavior of simulated tropical convection and teleconnections, and noting changes in the mean state and extended global forecast skill.

Acknowledgments. This research was funded by a NOAA Grant NA15NWS4680023. We acknowledge the extensive assistance of Professor Daehyun Kim, and the helpful comments of two anonymous reviewers and the publication editor, Dr. Paul Roundy.

REFERENCES

Abhik, S., R. P. M. Krishna, M. Mahakur, M. Ganai, P. Mukhopadhyay, and J. Dudhia, 2017: Revised cloud processes to improve the mean and intraseasonal variability of Indian summer monsoon in climate forecast system: Part 1. *J. Adv. Model. Earth Syst.*, **9**, 1002–1029, doi:10.1002/2016MS000819.

Adames, Á. F., and J. M. Wallace, 2014: Three-dimensional structure and evolution of the MJO and its relation to the mean flow. *J. Atmos. Sci.*, **71**, 2007–2026, doi:10.1175/JAS-D-13-0254.1.

——, and D. Kim, 2016: The MJO as a convectively coupled moisture wave: Theory and observations. *J. Atmos. Sci.*, **73**, 913–941, doi:10.1175/JAS-D-15-0170.1.

Alexander, M. A., I. Bladé, M. Newman, J. R. Lazante, N.-C. Lau, and J. D. Scott, 2002: The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate*, **15**, 2205–2231, doi:10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2.

Becker, E., H. van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, doi:10.1175/JCLI-D-13-00597.1.

Betts, A. K., and C. Jakob, 2002: Study of diurnal cycle of convective precipitation over Amazonia using a single column model. *J. Geophys. Res.*, **107**, 4732, doi:10.1029/2002JD002264.

Chelton, D. B., M. G. Schlax, D. L. Witter, and J. G. Richman, 1990: Geosat altimeter observations of the surface circulation of the Southern Ocean. *J. Geophys. Res.*, **95**, 17 877–17 904, doi:10.1029/JC095iC10p17877.

Chu, P. C., 1999: Two kinds of predictability in the Lorenz system. *J. Atmos. Sci.*, **56**, 1427–1432, doi:10.1175/1520-0469(1999)056<1427:TKOPIT>2.0.CO;2.

Davis, C. A., K. W. Manning, R. E. Carbone, S. B. Trier, and J. D. Tuttle, 2003: Coherence of warm-season continental rainfall in numerical weather prediction models. *Mon. Wea. Rev.*, **131**, 2667–2679, doi:10.1175/1520-0493(2003)131<2667:COWCRI>2.0.CO;2.

Demott, C. A., D. A. Randall, and M. Khairoutdinov, 2007: Convective precipitation variability as a tool for general circulation model analysis. *J. Climate*, **20**, 91–112, doi:10.1175/JCLI3991.1.

Dutton, J. A., R. P. James, and J. D. Ross, 2013: Calibration and combination of dynamical seasonal forecasts to enhance the value of predicted probabilities for managing risk. *Climate Dyn.*, **40**, 3089–3105, doi:10.1007/s00382-013-1764-2.

Fu, X., J.-Y. Lee, P.-C. Hsu, H. Taniguchi, B. Wang, W. Wang, and S. Weaver, 2013: Multi-model MJO forecasting during DYNAMO/CINDY period. *Climate Dyn.*, **41**, 1067–1081, doi:10.1007/s00382-013-1859-9.

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.*, **40**, 245–272, doi:10.1007/s00382-012-1481-2.

Grabowski, W. W., 2004: An improved framework for superparameterization. *J. Atmos. Sci.*, **61**, 1940–1952, doi:10.1175/1520-0469(2004)061<1940:AIFFS>2.0.CO;2.

Guichard, F., and Coauthors, 2004: Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Quart. J. Roy. Meteor. Soc.*, **130**, 3139–3172, doi:10.1256/qj.03.145.

Hazra, A., H. Chaudhari, and A. Dhakate, 2016: Evaluation of cloud properties in the NCEP CFSv2 model and its linkage with Indian summer monsoon. *Theor. Appl. Climatol.*, **124**, 31–41, doi:10.1007/s00704-015-1404-3.

Hendon, H. H., and M. C. Wheeler, 2008: Some space–time spectral analyses of tropical convection and planetary-scale waves. *J. Atmos. Sci.*, **65**, 2936–2948, doi:10.1175/2008JAS2675.1.

Holloway, C. E., S. J. Woolnough, and G. M. S. Lister, 2012: Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study. *Quart. J. Roy. Meteor. Soc.*, **138**, 1692–1708, doi:10.1002/qj.1903.

Holton, J. R., and G. Hakim, Eds., 2012: Numerical modeling and prediction. *An Introduction to Dynamic Meteorology*, 5th ed. Academic Press, 453–490.

Inness, P. M., and J. M. Slingo, 2003: Simulation of the Madden–Julian Oscillation in a coupled general circulation model. I: Comparison with observations and an atmosphere-only GCM. *J. Climate*, **16**, 345–364, doi:10.1175/1520-0442(2003)016<0345:SOTMJO>2.0.CO;2.

Inoue, T., M. Satoh, H. Miura, and B. Mapes, 2008: Characteristics of cloud size of deep convection simulated by a global cloud resolving model over the western tropical Pacific. *J. Meteor. Soc. Japan*, **86A**, 1–15, doi:10.2151/jmsj.86A.1.

Jones, C., A. Hazra, and L. M. V. Carvalho, 2015: The Madden–Julian Oscillation and boreal winter forecast skill: An analysis of NCEP CFSv2 reforecasts. *J. Climate*, **28**, 6297–6307, doi:10.1175/JCLI-D-15-0149.1.

Kim, D., J.-S. Kug, and A. H. Sobel, 2014: Propagating versus nonpropagating Madden–Julian oscillation events. *J. Climate*, **27**, 111–125, doi:10.1175/JCLI-D-13-00084.1.

Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dyn.*, **39**, 2957–2973, doi:10.1007/s00382-012-1364-6.

——, ——, V. E. Toma, and D. Kim, 2014: Predictability and prediction skill of the MJO in two operational forecasting systems. *J. Climate*, **27**, 5364–5378, doi:10.1175/JCLI-D-13-00480.1.

——, D. Kim, F. Vitart, V. E. Toma, J. Kug, and P. J. Webster, 2016: MJO propagation across the Maritime Continent in the ECMWF Ensemble Prediction System. *J. Climate*, **29**, 3973–3988, doi:10.1175/JCLI-D-15-0862.1.

Kumar, A., M. Chen, and W. Wang, 2011: An analysis of prediction skill of monthly mean climate variability. *Climate Dyn.*, **37**, 1119–1131, doi:10.1007/s00382-010-0901-4.

Lang, Y., and Coauthors, 2014: Evaluating skill of seasonal precipitation and temperature predictions of NCEP CFSv2 forecasts over 17 hydroclimatic regions of China. *J. Hydrometeor.*, **15**, 1546–1559, doi:10.1175/JHM-D-13-0208.1.

Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *J. Geophys. Res.*, **114**, D04114, doi:10.1029/2008JD010969.

Li, S., and A. W. Robertson, 2015: Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Mon. Wea. Rev.*, **143**, 2871–2889, doi:10.1175/MWR-D-14-00277.1.

Lin, J.-L., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean–atmosphere feedback analysis. *J. Climate*, **20**, 4497–4525, doi:10.1175/JCLI4272.1.

Liu, S., J. X. L. Wang, X.-Z. Liang, and V. Morris, 2016: A hybrid approach to improving the skills of seasonal climate outlook at the regional scale. *Climate Dyn.*, **46**, 483–494, doi:10.1007/s00382-015-2594-1.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

——, 1975: Climate predictability. *The Physical Basis of Climate and Climate Modelling,* GARP Publication Series, Vol. 16, World Meteorological Organization, 133–136.

Lynch, K. J., D. J. Brayshaw, and A. Charlton-Perez, 2014: Verification of European subseasonal wind speed forecasts. *Mon. Wea. Rev.*, **142**, 2978–2990, doi:10.1175/MWR-D-13-00341.1.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, doi:10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.

——, and ——, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29**, 1109–1123, doi:10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2.

Martin, G., N. P. Klingaman, and A. F. Moise, 2017: Connecting spatial and temporal scales of tropical precipitation in observations and the MetUM-GA6. *Geosci. Model Dev.*, **10**, 105–126, doi:10.5194/gmd-10-105-2017.

Miura, H., M. Satoh, T. Nasuno, A. Noda, and K. Oouchi, 2007: A Madden-Julian Oscillation event realistically simulated by a global cloud-resolving model. *Science*, **318**, 1763–1765, doi:10.1126/science.1148443.

Miyakawa, T., and Coauthors, 2014: Madden–Julian Oscillation prediction skill of a new-generation global model demonstrated using a supercomputer. *Nature Commun.*, **5**, 3769, doi:10.1038/ncomms4769.

Mo, K. C., S. Shukla, D. P. Lettenmaier, and L.-C. Chen, 2012: Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction? *Geophys. Res. Lett.*, **39**, L23703, doi:10.1029/2012GL053598.

Mukhopadhyay, P., R. Phani Murali Krishna, B. B. Goswami, S. Abhik, M. Ganai, M. Mahakur, M. Khairoutdinov, and J. Dudhia, 2016: Improvements of systematic bias of mean state and the intraseasonal variability of CFSv2 through superparameterization and revised cloud-convection-radiation parameterization. *Remote Sensing and Modeling of the Atmosphere, Oceans, and Interactions VI*, T. N. Krishnamurti and M. N. Rajeevan, Eds., International Society for Optical Engineering (SPIE Proceedings, Vol. 9882), 98820Z, doi:10.1117/12.2222982.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, **116**, 2417–2424, doi:10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Pilon, R., C. Zhang, and J. Dudhia, 2016: Roles of deep and shallow convection and microphysics in the MJO simulated by the Model for Prediction Across Scales. *J. Geophys. Res. Atmos.*, **121**, 10 575–10 600, doi:10.1002/2015JD024697.

Rasmusson, E. M., and J. M. Wallace, 1983: Meteorological aspects of the El Niño/Southern Oscillation. *Science*, **222**, 1195–1202, doi:10.1126/science.222.4629.1195.

Raymond, D. J., and Z. Fuchs, 2009: Moisture modes and the Madden–Julian Oscillation. *J. Climate*, **22**, 3031–3046, doi:10.1175/2008JCLI2739.1.

Roundy, J. K., X. Yuan, J. Schaake, and E. F. Wood, 2015: A framework for diagnosing seasonal prediction through canonical event analysis. *Mon. Wea. Rev.*, **143**, 2404–2418, doi:10.1175/MWR-D-14-00190.1.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Sato, T., H. Miura, M. Satoh, Y. N. Takayabu, and Y. Wang, 2009: Diurnal cycle of precipitation in the tropics simulated in a global cloud-resolving model. *J. Climate*, **22**, 4809–4826, doi:10.1175/2009JCLI2890.1.

Seo, K.-H., W. Wang, J. Gottschalck, Q. Zhang, J. E. Schemm, W. R. Higgins, and A. Kumar, 2009: Evaluation of MJO forecast skill from several statistical and dynamical forecast models. *J. Climate*, **22**, 2372–2388, doi:10.1175/2008JCLI2421.1.

Silva, G., L. Dutra, R. Rocha, T. Ambrizzi, and E. Leiva, 2014: Preliminary analysis on the global features of the NCEP CFSv2 seasonal hindcasts. *Adv. Meteor.*, **2014**, 695067, doi:10.1155/2014/695067.

Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211, doi:10.1029/2010JD014532.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14 291–14 324, doi:10.1029/97JC01444.

Van Weverberg, K., and Coauthors, 2013: The role of cloud microphysics parameterization in the simulation of mesoscale convective system clouds and precipitation in the tropical western Pacific. *J. Atmos. Sci.*, **70**, 1104–1128, doi:10.1175/JAS-D-12-0104.1.

Vitart, F., 2004: Monthly forecasting at ECMWF. *Mon. Wea. Rev.*, **132**, 2761–2779, doi:10.1175/MWR2826.1.

——, 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, doi:10.1002/qj.2256.

——, and Coauthors, 2008: The new VarEPS-monthly forecasting system: A first step towards seamless prediction. *Quart. J. Roy. Meteor. Soc.*, **134**, 1789–1799, doi:10.1002/qj.322.

Wanders, N., and E. F. Wood, 2016: Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environ. Res. Lett.*, **11**, 094007, doi:10.1088/1748-9326/11/9/094007.

Wang, S., A. H. Sobel, F. Zhang, Y. Q. Sun, Y. Yue, and L. Zhou, 2015: Regional simulation of the October and November MJO events observed during the CINDY/DYNAMO field campaign at gray zone resolution. *J. Climate*, **28**, 2097–2119, doi:10.1175/JCLI-D-14-00294.1.

Wang, W., M.-P. Hung, S. J. Weaver, A. Kumar, and X. Fu, 2014: MJO prediction in the NCEP Climate Forecast System version 2. *Climate Dyn.*, **42**, 2509–2520, doi:10.1007/s00382-013-1806-9.

Wang, Y., L. Zhou, and K. Hamilton, 2007: Effect of convective entrainment/detrainment on the simulation of the tropical precipitation diurnal cycle. *Mon. Wea. Rev.*, **135**, 567–585, doi:10.1175/MWR3308.1.

Wheeler, M., and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Sci.*, **56**, 374–399, doi:10.1175/1520-0469(1999)056<0374:CCEWAO>2.0.CO;2.

——, and H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Wu, W., and Coauthors, 2012: Statistical downscaling of climate forecast system seasonal predictions for the Southeastern Mediterranean. *Atmos. Res.*, **118**, 346–356, doi:10.1016/j.atmosres.2012.07.019.

Yang, G.-Y., and J. Slingo, 2001: The diurnal cycle in the tropics. *Mon. Wea. Rev.*, **129**, 784–801, doi:10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2.

Yasunaga, K., and B. Mapes, 2012: Differences between more divergent and more rotational types of convectively coupled equatorial waves. Part I: Space–time spectral analyses. *J. Atmos. Sci.*, **69**, 3–16, doi:10.1175/JAS-D-11-033.1.

Yuan, X., and X.-Z. Liang, 2011: Improving cold season precipitation prediction by the nested CWRF-CFS system. *Geophys. Res. Lett.*, **38**, L02706, doi:10.1029/2010GL046104.

——, E. F. Wood, L. Luo, and M. Pan, 2011: A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophys. Res. Lett.*, **38**, L13402, doi:10.1029/2011GL047792.

Zheng, X., and C. S. Frederiksen, 2007: Statistical prediction of seasonal mean Southern Hemisphere 500-hPa geopotential heights. *J. Climate*, **20**, 2791–2809, doi:10.1175/JCLI4180.1.

Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson, 2014: Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Wea. Rev.*, **142**, 1556–1569, doi:10.1175/MWR-D-13-00222.1.