Equatorial Waves and the Skill of NCEP and ECMWF Numerical Weather Prediction Systems

JULIANA DIAS AND MARIA GEHNE

Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and NOAA/Earth System Research Laboratory, Boulder, Colorado

GEORGE N. KILADIS

NOAA/Earth System Research Laboratory, Boulder, Colorado

NAOKO SAKAEDA

School of Meteorology, University of Oklahoma, Norman, Oklahoma

PETER BECHTOLD AND THOMAS HAIDEN

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received 1 December 2017, in final form 15 February 2018)

ABSTRACT

Despite decades of research on the role of moist convective processes in large-scale tropical dynamics, tropical forecast skill in operational models is still deficient when compared to the extratropics, even at short lead times. Here we compare tropical and Northern Hemisphere (NH) forecast skill for quantitative precipitation forecasts (QPFs) in the NCEP Global Forecast System (GFS) and ECMWF Integrated Forecast System (IFS) during January 2015–March 2016. Results reveal that, in general, initial conditions are reasonably well estimated in both forecast systems, as indicated by relatively good skill scores for the 6–24-h forecasts. However, overall, tropical QPF forecasts in both systems are not considered useful by typical metrics much beyond 4 days. To quantify the relationship between QPF and dynamical skill, space–time spectra and coherence of rainfall and divergence fields are calculated. It is shown that while tropical variability is too weak in both models, the IFS is more skillful in propagating tropical waves for longer lead times. In agreement with past studies demonstrating that extratropical skill is partially drawn from the tropics, a comparison of daily skill in the tropics versus NH suggests that in both models NH forecast skill at lead times beyond day 3 is enhanced by tropical skill in the first couple of days. As shown in previous work, this study indicates that the differences in physics used in each system, in particular, how moist convective processes are coupled to the large-scale flow through these parameterizations, appear as a major source of tropical forecast errors.

1. Introduction

Deterministic short- and medium-range weather forecasts have steadily become more skillful over the past five decades, owing to a combination of improvements in observation systems, data assimilation, and model development, as well as computing advances (Bauer et al. 2015). Despite this progress, low-latitude forecasts remain less skillful than those at higher latitudes, and this is likely related to differences in their limits of deterministic predictability (Shukla 1989; Reynolds et al. 1994; Boer 1995; Straus and Paolino 2008). One main difference between tropical versus extratropical predictability results from the ''memory'' carried by propagating disturbances, the properties of which depend on the balance of restoring forces due to gravity and Earth's rotation. In the tropics, wave dynamics rely more strongly on cloud–radiation and latent

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/MWR-D-17-0362.s1.

Corresponding author: Juliana Dias, juliana.dias@noaa.gov

DOI: 10.1175/MWR-D-17-0362.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

heating feedbacks than in the extratropics where moist convection is to a greater extent tied to the large-scale stronger rotational flow (Mapes et al. 2008; Zhu et al. 2014; Wheeler et al. 2017). This means that predictions of tropical variability are more dependent on model physics related to convection and cloud–radiation feedbacks, which is one of the major sources of uncertainty in current predictions of both weather and climate (Flato 2013; Webb et al. 2015).

One potential source of tropical skill from short to extended ranges derives from tropical disturbances such as the Madden–Julian oscillation (MJO; Madden and Julian 1971, 1994; Zhang 2005) and convectively coupled equatorial waves (CCEWs; Takayabu 1994; Wheeler and Kiladis 1999; Wheeler et al. 2000), which are also well known to impact weather well beyond the tropics (Kiladis et al. 2009; Zhang 2013). Since operational numerical weather prediction (NWP) models often struggle to initiate and correctly propagate equatorial waves, it is thought that there is room for reducing the differences between deterministic short- and medium-range weather forecasts at low versus high latitudes by improving the representation of these wavelike disturbances in NWP models (Matsueda and Endo 2011; Ying and Zhang 2017, 2018). To further understand the relationship between the ability of models to simulate equatorial waves and forecast skill, this paper analyzes two global NWP systems: the National Centers for Environmental Prediction Global Forecast System (NCEP GFS) and the European Centre for Medium-Range Weather Forecasts Integrated Forecast System (ECMWF IFS) during January 2015–March 2016. The two models are chosen to contrast the link between equatorial waves and forecast skill in a model that is known to perform relatively well in the tropics (IFS; Kerns and Chen 2014; Li and Robertson 2015; Taraphdar et al. 2016) to a model with lower tropical skill (GFS).

This study was initially motivated by the El Niño Rapid Response field campaign (ENRR; Dole et al. 2018), during which some of the authors participated in the planning of flight paths aimed at encircling deep tropical convection in the eastern Pacific. Perhaps not surprisingly, this turned out to be a rather difficult task given that flight plans had to be submitted 48 h in advance and that the spread in tropical forecasts across various NWPs is large, even at the short range. Common issues with tropical rainfall forecasts can be easily seen in Fig. 1, which shows precipitation between 5°S and 5°N during ENRR from two satellite estimates (Figs. 1a,d) and for the GFS and IFS 12- (Figs. 1b,c) and 120-h (Figs. 1e,f) forecasts, valid at the same time as the observations (a description of these datasets is in the next section). The period featured several coherent eastward- and westward-propagating precipitation envelopes. The eastward-moving envelopes are predominately Kelvin waves, which are much more common in boreal winter over the equatorial Pacific during El Niño, and propagate at around 15 m s^{-1} while strongly modulating precipitation in their path (Straub and Kiladis 2002; Kiladis et al. 2009). Despite the fact that satellite precipitation estimates are not directly assimilated into either model, the GFS and IFS performance at a 12-h lead time is remarkably good in terms of representation of zonally propagating and even some of the mesoscale features. This demonstrates that both models are able to produce realistic large-scale precipitation features from their initialized wind and thermodynamic fields. This very short-range forecast differs strikingly from the 5-day forecasts (Figs. 1e,f) where precipitation appears to be much less organized, indicating that the models have difficulty in maintaining the zonal propagation of the eastward Kelvin waves in particular.

This situation for equatorial forecasts is in sharp contrast to that in midlatitudes as shown in Fig. 2, which is similar to Fig. 1 except for the 35°S–50°N latitude band. Figure 2 shows that disturbances in the North Pacific and Atlantic storm track are not only well represented in both models in comparison to satellite estimates, but their timing, location, and intensity are still quite well predicted with 5-day lead time. Figures 1 and 2 also illustrate the well-known model tendency to overestimate light rain rates (Dai 2006; Stephens et al. 2010), although we note that while the two satellite estimates are in very good agreement, they also likely underestimate light precipitation to some extent (Huffman et al. 2007; Ellis et al. 2009; Behrangi et al. 2012). It turns out that this misrepresentation of tropical precipitation during ENRR is not uncommon, as the present study demonstrates in detail.

While the basic dynamical mechanisms for the existence of CCEWs is well understood (Matsuno 1966), the manner in which they couple with moisture is much less settled (Raymond and Fuchs 2007; Kiladis et al. 2009). This is particularly true for the MJO, which is not a solution to Matsuno's shallow-water system on an equatorial beta plane, and whose very existence likely depends on feedback between moisture and circulation (Raymond and Fuchs 2007; Majda and Stechmann 2009). A number of modeling studies (Lin et al. 2008; Frierson et al. 2011; Khouider et al. 2011; Hirons et al. 2013a, b; Hirota et al. 2014) have indicated that convective parameterization is crucial to a model's ability to represent organized tropical disturbances, which is indicative of the fundamental role of convective

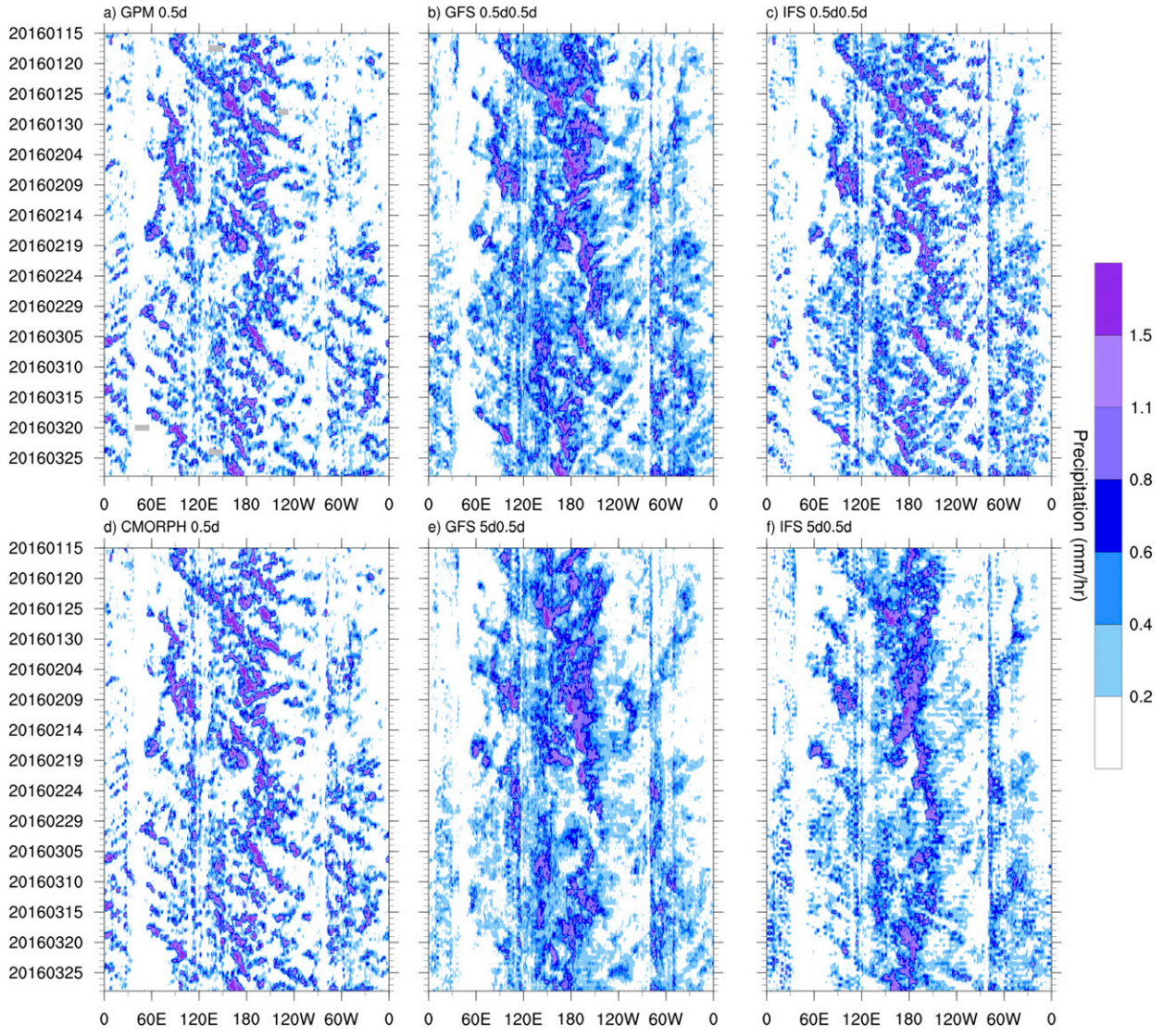

FIG. 1. Time–longitude section of 12-h average rain rates from 15 Jan to 25 Mar 2016 averaged from 5°S to 5°N. Satellite estimates are shown for (a) GPM and (d) CMORPH, and 12-h forecasts are shown for (b) the NCEP GFS and (c) the ECMWF IFS, and 120-h forecasts are shown for (e) the NCEP GFS and (f) the ECMWF IFS.

coupling to the maintenance of the disturbances. Because of the potential societal impacts of better predicting the tropical atmosphere from weather to climate scales, improving tropical variability has been a main focus of model development at the ECMWF, which has resulted in significant advances in tropical forecast skill over the last decade (Bechtold et al. 2008; Vitart 2014). For example, improvement in the IFS simulation of the MJO has been explicitly associated with changes in their convective parameterization (Hirons et al. 2013b; Hirota et al. 2014). That the IFS overall outperforms other operational models including the GFS is generally well established (Kerns and Chen 2014; Li and Robertson 2015; Taraphdar et al. 2016); however, most recent studies regarding the tropics have been focused on the MJO and medium- to extended-range forecast skill (Matsueda and Endo 2011; Kerns and Chen 2014). Here we concentrate on a detailed characterization of 1–10-day global forecast skill in the GFS and IFS, as well as on the contrasts and linkages between tropical and extratropical skill and the relationship between equatorial waves and tropical skill.

We emphasize rainfall because of its inherent impact, but also to facilitate the comparison between lower- and higher-latitude skill since, for example, temperature and

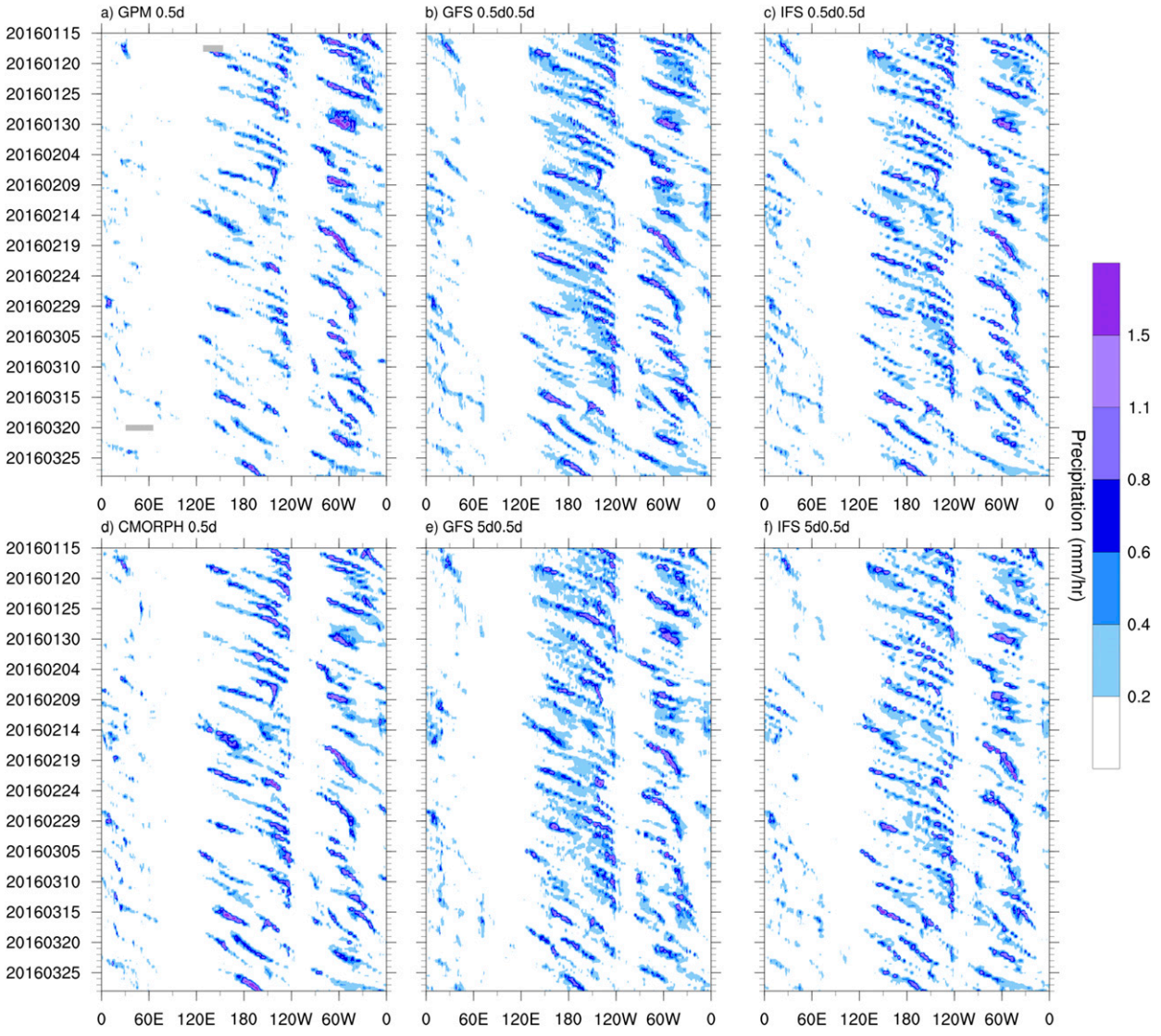

FIG. 2. As in Fig. 1, but for rain rates averaged from 35° to 50°N.

pressure vary much less in the tropics than in the extratropics. Because most of the tropics is covered by oceans where rain gauge measurements are sparse in time and space, global quantitative precipitation forecasts (QPFs) are obtained by verifying modeled precipitation over land and ocean against satellite estimates of precipitation. We refer to the study by Haiden et al. (2012) for a global land QPF skill analysis in NWP systems that utilizes rain gauge data. We choose two sets of observations because of the large uncertainties involved in estimates of precipitation because of its highly variable nature, where the main sources of errors derive from relatively sparse satellite observations as well as biases in retrieval algorithms (Ebert et al. 2007; Lu et al. 2010). For instance, because satellite estimates of precipitation rely on infrared imagery, errors in location and amounts of precipitation are more likely in highly sheared environments where cirrus clouds might be advected in a different direction from the convective cores below them. To test how precipitation is related to the large-scale flow, dynamical fields are also analyzed and verified against each model's own analysis system. The data used in this manuscript are reviewed in section 2, which also provides a description of the methodology used. Results are presented in section 3 and the summary and conclusions are presented in section 4.

2. Data and methodology

a. Deterministic forecast data

1) NCEP GFS

GFS data are from NCEP operational model version GSM v12.0.0, which was operational from 15 January 2015 to 11 May 2016. (Details on the GFS implementation can be found online at http://www.emc.ncep.noaa.gov/GFS/doc.php.) The native horizontal resolution is T1534 (~13km) out to 10-day lead time. Model data are available from the Research Data Archive (RDA) at the National Center for Atmospheric Research (NCAR). GFS output is provided at 3-hourly intervals, with initial times at 0000, 0600, 1200, and 1800 UTC.

2) ECMWF IFS

Integrated Forecast System (IFS) data are mainly from ECMWF operational model version CY41R1, which was operational from 12 May 2015 to 8 March 2016. From January 2015 to 11 May 2015, model version CY40R1 is used. (Details on the IFS model implementation can be found online at https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation.) The native resolution for the IFS is a reduced Gaussian grid at T1279 (~15 km; www.ecmwf.int/en/what-horizontal-resolution-data). Model data were obtained at 6-hourly intervals, with initial times at 0000 and 1200 UTC.

b. Observations

1) GPM

The Global Precipitation Measurement (GPM) *Core Observatory* satellite 3B42 product (Huffman et al. 2007) is available at 3-hourly resolution as area averages of precipitation rate on a 0.25° grid between 50°S and 50°N. While they are meant to be interpreted as instantaneous measurements valid at each time stamp, passive microwave data during the 90 min before and after the observation time can be included (Huffman et al. 2007). Each 3-hourly field is bias corrected using a monthly satellite-gauge analysis. (GPM data are available for download at https://mirador.gsfc.nasa.gov/cgi-bin/mirador/presentNavigation.pl?tree=project&project=TRMM.) Throughout the manuscript we refer to the 3B42 rainfall product as GPM.

2) CMORPH

The Climate Prediction Center morphing technique (CMORPH; Joyce et al. 2004) precipitation estimates are available on the same 0.25° grid as GPM between 60°S and 60°N. The native resolution for CMORPH is 8 km, 30 min, and the 0.25° estimates are box averages of the native resolution estimates. The instantaneous half-hourly values are averaged to generate the 3-hourly estimates valid for the 3-h interval after the observation time. CMORPH is bias corrected over land using a rain gauge analysis and over the ocean using Global Precipitation Climatology Project pentad data. (CMORPH data are available for download at http://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/0.25deg-3HLY/.)

c. Data postprocessing

All datasets are first mapped to a common 0.25° horizontal grid. This is done using bilinear interpolation from the GFS and IFS model output, which means that the data are subsampled and the interpolation is not conservative. Thus, the interpolated values on the 0.25° grid represent a mix of points and area-average data, representing a smaller area than the 0.25° grid box for which it is valid. Because CMORPH and GPM grids are offset by 0.125° from the GFS grid, both datasets are linearly interpolated to the GFS grid. To test for the impact of horizontal resolution on skill, coarse gridded data at 1° × 1° and 2.5° × 2.5° are obtained by the NCL area conserving algorithm using the 0.25° × 0.25° data.

Forecasts of precipitation are provided as a rain-rate average between forecast times, whereas both GPM and CMORPH represent instantaneous rain rates at 3- and 0.5-hourly increments, respectively. To partially compensate for this difference, we test skill using twice-daily and/or daily averages. For the twice-daily GFS and IFS rain rates, both initializations from 0000 and 1200 UTC are used with an averaging window of 12 h. We adopt the terminology from Zhu et al. (2014) and others: 0.5d0.5d stands for the 12-h forecast time of the subsequent 12-h average rain rate, 1d0.5d stands for the 24-h forecast time of the subsequent 12-h rain rate average, and so on. The time series of rain rates at their time of validation is referred to as the forecast time series. For example, forecast time series illustrated in Figs. 1 and 2 are for 0.5d0.5d and 5d0.5d. The GPM and CMORPH twice-daily time series are calculated by averaging rain rates from each day from 0300 to 1200 UTC and then from 1500 to 0000 UTC. We have also tested using half weights at the endpoints (0000 and 1200 UTC) and then averaging, and the results are very similar. Daily forecast time series are analogous where the terminology in this case is 1d1d for the 1-day forecast of the subsequent 1-day rain-rate average. Results shown here use the 0000 UTC initializations for daily estimates, with the results using 1200 UTC nearly identical. To evaluate the shortest available forecasts, we also use the 6-h GFS and IFS forecasts initialized at

either 0000 or 1200 UTC, which are compared to 6-hourly estimates of GPM and CMORPH calculated using half-weights at the endpoints. Forecasts of dynamical variables are instantaneous and are verified against each respective model's analysis at the same valid time.

d. Verification methods

Forecast verification is done for the period of January 2015–March 2016 since this is the longest recent period where only minor updates were made to each operational system described above. We have computed a wide number of different verification measures including spatial pattern correlations, gridpoint correlations, root-mean-square error, bias, equitable threat score, and fractions skill score (FSS) and found that our conclusions in terms of relative skill are not overly sensitive to the metric used. Because our interest is mostly on the relative skill between the two models and between tropical and extratropical regions, we report our results using either gridpoint or spatial pattern correlations of total rain rates. To test the impact of horizontal resolution, we report some results using the FSS in appendix A and to analyze the diurnal variability in the two models, a comparison of their rainfall diurnal cycles is presented in appendix B. To test sensitivity within the 15-month period shown here, skill is also assessed for independent 3-month-long subperiods (not shown), where we find that our main conclusions regarding relative skill are robust. We also note that while the IFS model version changed between January–March 2015 and 2016, we do not find significant differences in relative skill (not shown).

For a given lead time, the gridpoint correlation is the temporal correlation between the forecast time series (IFS or GFS) and the corresponding observed time series (GPM or CMORPH) valid at the same time, using the entire 15-month sample. These gridpoint correlations are given as area-weighted averages over the four regions defined in Table 1, where the ocean versus land partition is based on the GFS land mask. Pattern correlations are calculated using all grid points within the equatorial band from 10°S to 10°N and separately for an extratropical band from 35° to 50°N. The statistical significance of the differences between GFS and IFS correlations is assessed by bootstrapping the total number of forecasts, where at each grid point, we use 1000 random subsamples with replacement and the confidence interval is defined by the lower and upper fifth percentile of the correlation distribution. Over the regions shown in Table 1, differences in correlations are significant when the mean confidence intervals do not overlap.

TABLE 1. Average regions used in Figs. 5, 7, and 8.

Region	Latitude band	Ocean	Land
TR-O	10°S–10°N	×	
NH-O	35°–50°N	×	
TR-L	10°S–10°N		×
NH-L	35°–50°N		×

Rain-rate space–time power spectra calculations using the forecast time series are compared to satellite estimates following the methodology of Wheeler and Kiladis (1999), except that we use 64-day segments overlapping by 15 days to increase our sample size since only a 15-month period is analyzed. This calculation involves first detrending and tapering to zero all gridpoint time series within each 64-day segment. Then, for each time segment and latitude from 50°S to 50°N, spectral coefficients in time and longitude are obtained by calculating a two-dimensional Fourier transform. The mean spectral coefficients over all time segments are denoted $a_{k\omega}$, where k is the planetary wavenumber and ω is the frequency in cycles per day (cpd). The mean raw power spectrum is the average of $|a_{k\omega}|^2$ between 15°S and 15°N for the tropics, where the symmetric and antisymmetric components with respect to the equator are calculated separately. For the extratropics, we display the mean $|a_{k\omega}|^2$ between 30° and 50°N. The squared coherence between rain rate and divergence are similarly calculated where first the divergence spectral coefficients $b_{k\omega}$ are calculated in an analogous way as $a_{k\omega}$. Squared coherence spectra and the corresponding significance levels are computed using the NCL function mjo_cross.ncl where the 95% confidence level is 0.18 for a window length of 64 days and 1.25 years of data. In all space–time power and coherence plots, Matsuno's equatorial wave dispersion curves (Matsuno 1966) are displayed following the notation from Kiladis et al. (2009) where $n = 1$ ER stands for equatorial Rossby waves, and $n = 1$ IG stands for inertio-gravity waves.

3. Results

a. Mean rainfall patterns

The top panels in Fig. 3 show the GPM and CMORPH mean daily rain rates during the analyzed period. These two plots highlight the inherent uncertainty in precipitation datasets in general (e.g., Gehne et al. 2016). As shown further below, while there are some obvious differences, the satellite estimates agree better with each other than with the individual models' precipitation rates. The middle panels show that the mean patterns

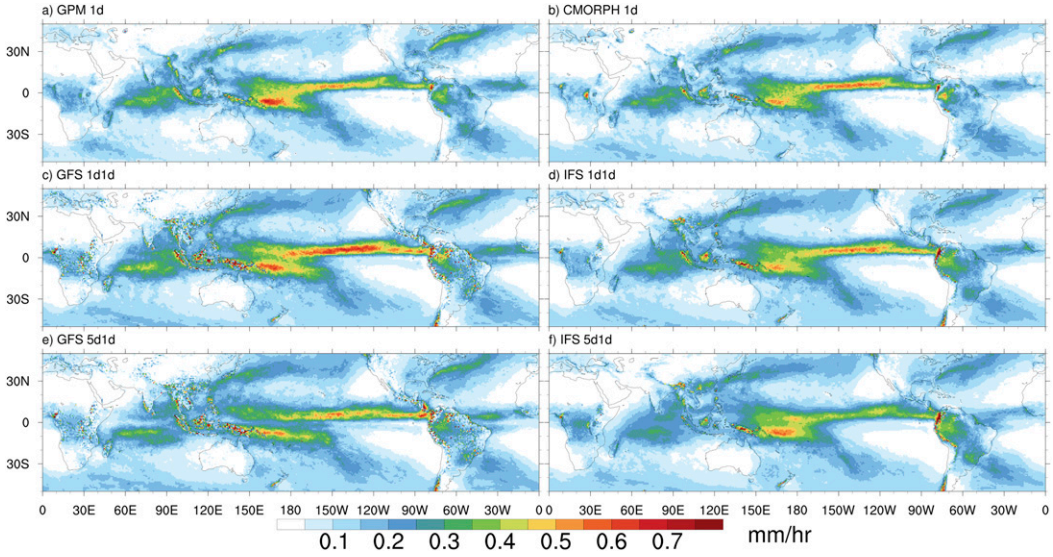

FIG. 3. Mean daily precipitation rate for the period January 2015–March 2016 from the (a) GPM, (b) CMORPH, (c) NCEP GFS day-1 forecast (1d1d), (d) ECMWF IFS day-1 forecast (1d1d), (e) NCEP GFS day-5 forecast (5d1d), and (f) IFS day-5 forecast (5d1d).

and intensity of rainfall from 1d1d forecasts are in reasonable agreement with observations, as well as with one another, although some biases are clearly evident. In contrast, the bottom panels indicate a clear dry bias along the equator for the GFS at a 5-day lead time and, although this bias in the IFS is much weaker, the equatorial zone is still too dry over the Indian, Pacific, and Atlantic Oceans. Contrasts in mean rain rates are more clear by looking at the relative differences shown in Fig. 4, defined as

$$r(\mathrm{lat},\mathrm{lon})^{\mathrm{LT}}_{D_iD_j} = \frac{\overline{\mathrm{RR}_{D_i} - \mathrm{RR}_{D_j}}}{0.5\left(\mathrm{RR}_{D_i} + \mathrm{RR}_{D_j}\right)}, \qquad (1)$$

where D_i and D_j stand for datasets GPM, CMORPH, GFS, or IFS; RR is the rain rate at each grid point; and

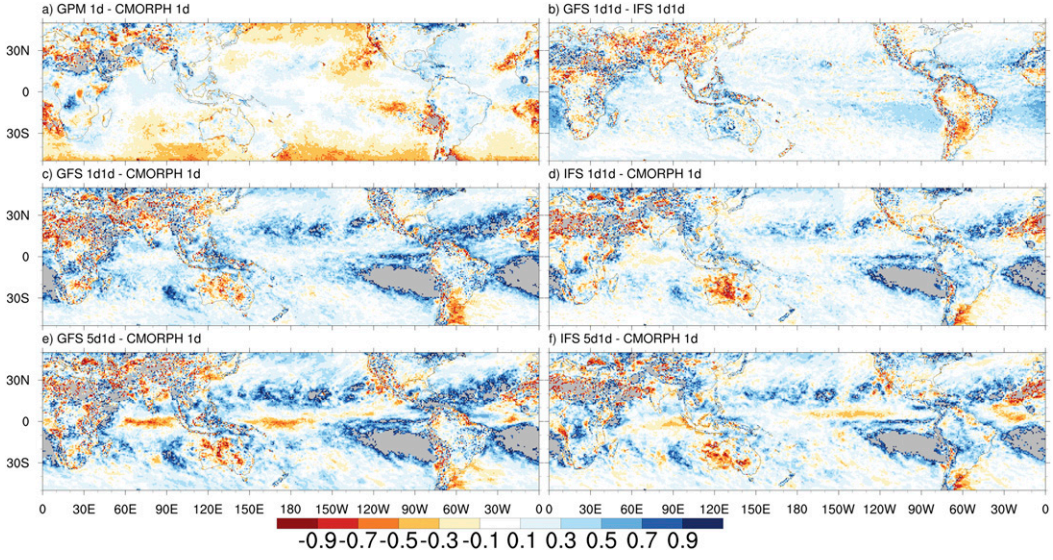

FIG. 4. Normalized differences in precipitation for the period January 2015–March 2016 [using Eq. (1)] for (a) GPM and CMORPH, (b) day-1 NCEP GFS and ECMWF IFS (1d1d), (c) day-1 NCEP GFS (1d1d) and CMORPH, and (d) day-1 ECMWF IFS (1d1d) and CMORPH. (e),(f) as in (c),(d), but for day 5 (5d1d). Gray shading corresponds to regions where the amplitude of the normalized difference is larger than 1.

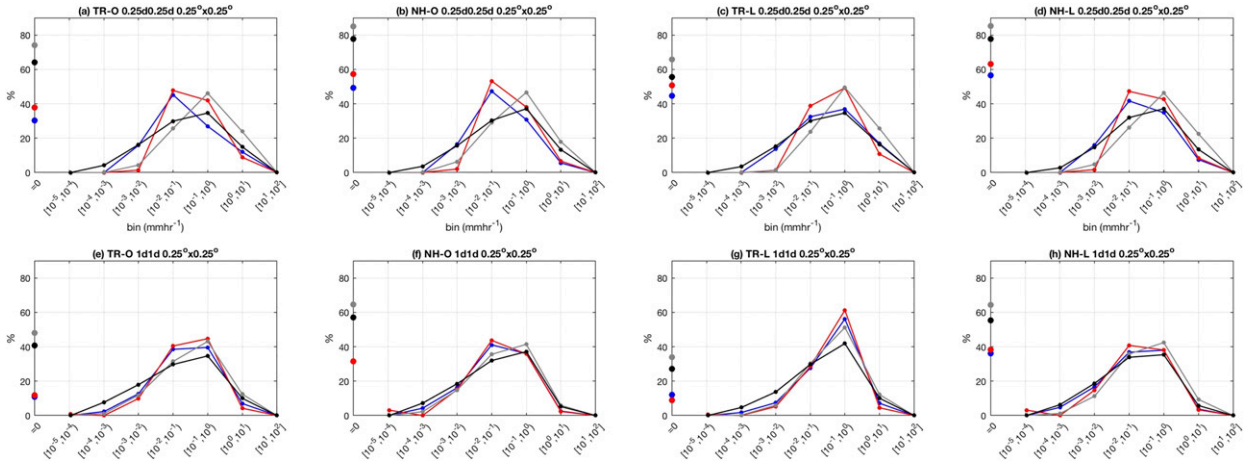

FIG. 5. Probability density functions (pdfs) of mean rain rates from (a)–(d) 0–6-h and (e)–(h) 0–24-h forecast periods for the GFS (blue) and IFS (red) overlaid with the pdfs from GPM (gray) and CMORPH (black) using mean rain rates valid at the same times. (from left to right) four different regions are shown: tropical oceans, NH oceans, tropical land, and NH land (defined as in Table 1 and in section 4c). Filled circles on the left side of each panel display the percentage of no rain occurrences for each dataset.

the overbar denotes an average over all forecasts for a fixed forecast lead time (LT). Differences that are larger than the mean are shaded in gray. Figure 4 reveals many interesting features concerning biases, such as the following:

- The largest differences between the satellite estimates occur over Asian and African land regions, and at higher oceanic latitudes (Fig. 4a). Over the oceans, GPM mean rain rates are weaker at high latitudes and in the dry subtropical regions of the eastern Pacific and Atlantic. The relatively good agreement over most of the tropics is notable.
- The amplitude of the differences between 1d1d IFS and GFS forecasts are comparable to the differences between CMORPH and GPM, with the GFS tending to be wetter over the tropical and subtropical oceans and drier over the continents than the IFS (Figs. 4a,b).
- Forecasts by both the GFS and IFS exhibit generally wet biases compared to CMORPH (Figs. 4c,d) and GPM (not shown) with overall larger amplitude in the GFS, particularly over the tropical oceans. These biases grow larger with time (Figs. 4e,f).
- By 5-day lead times, both the GFS and IFS have developed dry biases along the equator (Figs. 4e,f), with this bias larger in the GFS.
- Over land there is more geographic variability in biases. For instance, GFS rains less over most of Asia, but it rains more over North America, and the IFS has a larger dry bias over Australia.

b. *Rainfall histograms*

While mean rain-rate patterns, particularly at short forecast lead times, are in reasonable agreement with observations, probability density distributions (pdfs) of rain rate shown in Fig. 5 reveal more substantial discrepancies between modeled and observed rainfall. These pdfs are estimated based on gridpoint histograms of forecasts summed and area weighted over two latitude bands: 10°S–10°N (Figs. 5a,c) and 35°S–50°N (Figs. 5b,d) as defined in Table 1. The plots are normalized such that the binned rain rates add up to 100%. We excluded no rain occurrences from the pdfs but show their percentage contributions as circles on the left ordinate. We also use logarithmic bin widths to mitigate the dry side skewness of rain-rate histograms, making it easier to see differences between the pdfs at higher rain rates. Focusing on the top row, all of the panels display the tendency for GPM and CMORPH to report zero rain rates much more frequently than in the models, as shown by the filled circles along the left ordinate. This is certainly in part because they report instantaneous measurements as opposed to accumulated totals from the models. These panels also show that even the observed distributions differ substantially from each other, with CMORPH displaying a heavier tail at the lowest rain rates ($<0.01\,\mathrm{mm\,h^{-1}}$). In addition, weak rain rates ($0.01–0.1\,\mathrm{mm\,h^{-1}}$) are more likely in both models than in either satellite estimate. In fact, modeled rain rates peak within this bin ($0.01–0.1\,\mathrm{mm\,h^{-1}}$) whereas the satellite estimates peak in the $0.1–1.0\,\mathrm{mm\,h^{-1}}$ range. Stronger rain rates ($>1.0\,\mathrm{mm\,h^{-1}}$) are generally underestimated by the GFS and IFS and these are more frequent in the GPM in comparison to CMORPH. In summary, the pdfs from 6-h forecasts of both the IFS and GFS are narrower than the observed CMORPH and GPM pdfs.

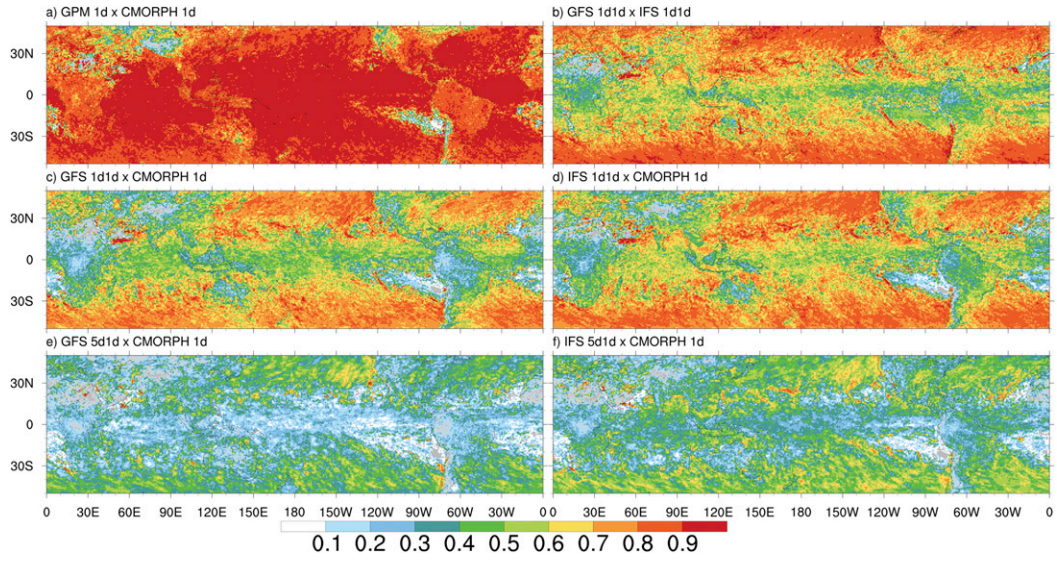

FIG. 6. As in Fig. 3, but for gridpoint temporal correlations.

Interestingly, the disagreement among 6-h forecast pdfs does not seem to depend on the geographic regions displayed, that is, there are large differences among the pdfs shown in each panel on the top row of Fig. 5. This result is in contrast with the pdfs of daily averaged rain rates shown on the bottom row of Fig. 5, where the NH pdfs are much closer to one another than their tropical counterparts. This further supports the conclusion that at least part of the 6-hourly differences are due to comparing instantaneous to average rain rates. Another striking difference between modeled and observed pdfs mentioned above regards the much lower frequency of no rain occurrences in models compared to observations, particularly over the tropical ocean. These differences are even larger when using daily averages, thus cannot be entirely because satellite observations are instantaneous. Spatial averaging tends to reduce the differences in no rain occurrences, but then larger differences between observed and modeled pdfs are seen at low rain rates (not shown). While excessive drizzle in models is a common problem (Sun et al. 2006; Stephens et al. 2010; Ahlgrimm and Forbes 2014), contrasting CMORPH and GPM it is clear that there are also large uncertainties in the observations regarding fractions of no to low rain rates. In particular, there is ample evidence that the algorithm used in the GPM estimates tends to underrepresent low precipitation amounts and occurrence while overrepresenting no rain occurrences (Ellis et al. 2009; Behrangi et al. 2012). The better agreement of daily pdfs in the NH than in the tropics is consistent with the generally better NH QPF skill than the tropical QPF skill that is discussed next.

c. *Forecast skill*

Figure 6 displays QPF skill based on gridpoint temporal correlations between GPM, CMORPH, and the 1d1d and 5d1d forecasts, providing a broad view of errors in rainfall forecasts. Consistent with their small normalized differences (Fig. 4a), temporal variability between CMORPH and GPM is quite highly correlated (Fig. 6a), with the exception of high topographic or very dry regions. In contrast to the observations, there is a strong meridional gradient in correlations when comparing the 1d1d precipitation forecasts from the two models, with correlations generally increasing with distance from the equator, especially over the ocean (Fig. 5b). This meridional gradient is also clear between GFS1d1d and CMORPH (Fig. 6c) and IFS1d1d and CMORPH (Fig. 6d), as well as when models are compared to GPM (not shown). Thus, the skill in both systems is much higher, and their forecasts agree with each other much more, over the extratropics than at low latitudes. Interestingly, while IFS skill is higher than the GFS over most regions of the globe, their correlation patterns are similar, even down to detailed features, when comparing the two models. The cause of these mutual signals is not clear, but could be related to specific events in the relatively short record that both systems were more (or less) skillful in forecasting. As expected, global skill drops everywhere at the 5-day lead time (Figs. 6e,f), but tropical skill in the IFS remains higher than in the GFS.

To further investigate the regional differences in QPF skill between the models, Fig. 7 shows the gridpoint temporal correlations averaged over the regions defined

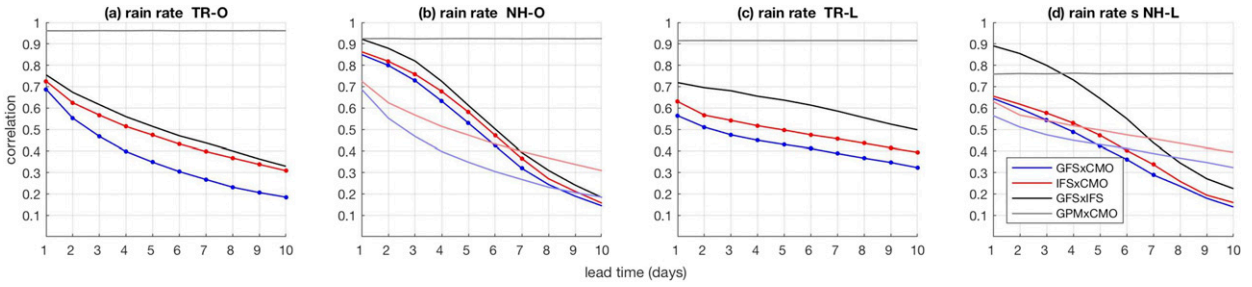

FIG. 7. Gridpoint temporal correlations averaged over the same four regions shown in Fig. 5 and defined in Table 1. Filled circles correspond to forecast lead times where differences between NCEP GFS and ECMWF IFS skill are significant (see section 4c for details). The mean gridpoint temporal correlation is shown for the GFS (blue) and IFS (red) against CMORPH, as well as between the GFS and IFS (black) and CMORPH and GPM (light gray). For better comparison between tropics and NH, the tropical IFS and GFS skill curve is shown again in lighter colors in the NH panels.

in Table 1 as a function of forecast lead time, with symbols displayed where differences between models are significant. The upper thin gray curves show the temporal correlations between GPM and CMORPH, which are in much closer agreement with each other than each model is with observations. The black lines show the correlations between the models, which in general is higher than each model is with CMORPH (as well as GPM, not shown). Overall, the IFS has higher skill than the GFS, and while the skill scores shown here are calculated using 2.5° × 2.5° averages, this relative relationship is not sensitive to the various metrics or resolutions we have extensively tested, such as the FSS results shown in appendix A. It is clear from Fig. 7 that QPF in the IFS outperforms the GFS from day 1, particularly in the tropics, which suggest that the IFS initialization is better than the GFS, possibly because of the IFS advanced four-dimensional variational data assimilation approach. The GFS QPF skill also tends to decay more rapidly than the IFS, once again, particularly over the tropics. As shown by many others (Haiden et al. 2012; Zhu et al. 2014; Wheeler et al. 2017), QPF skill decays more rapidly in the tropics than in the extratropics, but beyond week 1, forecast skill scores in the tropics are higher than in the extratropics. We note that a very similar qualitative relationship between QPF skill between models and regions is found when forecasts are verified against each model's respective 1d1d forecast (not shown), as opposed to satellite estimates. Last, verification against GPM also reveals a very similar picture, except that skill is overall slightly lower, possibly because of the lower frequency of light rain rates in comparison to CMORPH mentioned in the previous section (Fig. 3).

Figure 8 depicts 850-hPa divergence (D850) and 500-hPa geopotential (Z500) forecast skill, where each model is verified against its own analysis. Over the tropical oceans, low-level divergence skill is very similar for the two models, but in all other regions the IFS again shows higher skill. While in the extratropics differences in the IFS versus GFS D850 skill are smaller than with QPF, the differences in D850 skill over tropical land are more comparable to those differences for QPF. Z500 skill is a common metric for model performance in the extratropics and it also indicates higher skill for the IFS at most lead times. The D200 skill is similar to D850, except that skill is even more similar for the two models, including land points (not shown). Since the forecast skill in tropical divergence is more comparable between the GFS and IFS than for QPF, this suggests that model physics rather than model dynamics plays a larger role in the QPF performance differences between the two systems. We further discuss this issue in the next section.

d. *Space–time spectral analysis*

Space–time spectral analysis has become a standard approach to quantify properties of equatorial waves in observations (Hayashi 1982; Takayabu 1994; Wheeler and Kiladis 1999; Hendon and Wheeler 2008; Kiladis et al. 2009), and models (Lin et al. 2006; Straub et al. 2010; Guo et al. 2015). We apply the technique here to forecast time series detailed in section 2d. For conciseness, results are shown only for the equatorial symmetric component since conclusions are qualitatively similar for the antisymmetric signals. Figures 9a and 9d show that the main features of the climatological equatorial rainfall space–time spectra are well represented even when using only a 15-month period of satellite rainfall estimates. These observed space–time spectra are red in both wavenumber and frequency, with clear MJO peaks between 30 and 60 days and eastward wavenumbers 1–3. Enhanced power also follows CCEW dispersion curves such as Kelvin and ER (Matsuno 1966) and at higher frequencies (>0.3 cpd) there is a clear westward bias where enhanced power is associated with

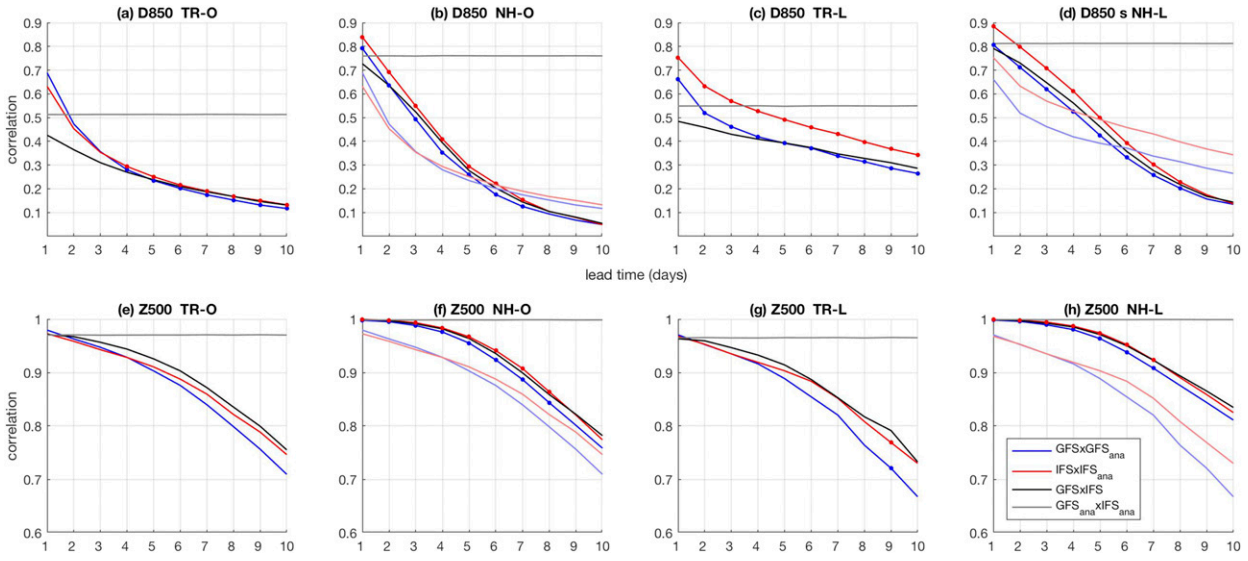

FIG. 8. As in Fig. 7, but for 850-hPa divergence (D850) and 500-hPa geopotential (Z500).

westward-propagating IGs (WIGs; see Kiladis et al. 2009; Tulich and Kiladis 2012). It is clear that GPM and CMORPH (Figs. 9a,d) are in very close agreement and that the power spectra from both model forecast time series (shown on panels to the right) are too red, with too little power at higher frequencies in comparison to observations. The IFS (Figs. 9e,f) appears to have less spectral power than the GFS (Figs. 9b,c), but the

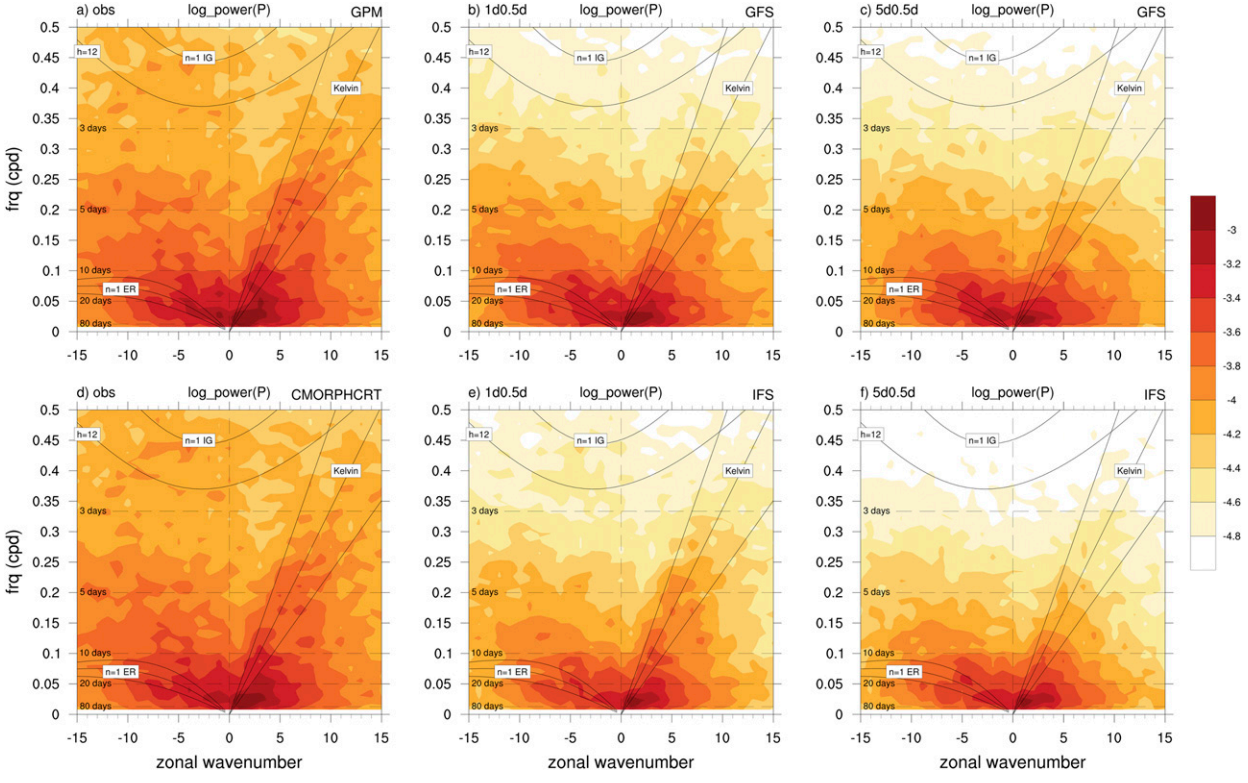

FIG. 9. Zonal wavenumber and frequency power spectra of symmetric rain rates about the equator averaged from 15°S to 15°N for the period January 2015–March 2016 of (a) GPM, (b) GFS 1d0.5d, (c) GFS 5d0.5d, (d) CMORPH, (e) IFS 1d0.5d, and (f) IFS 5d0.5d. Dispersion curves shown are for equatorial waves as in Wheeler and Kiladis (1999), for equivalent depths of 12, 25, and 50 m.

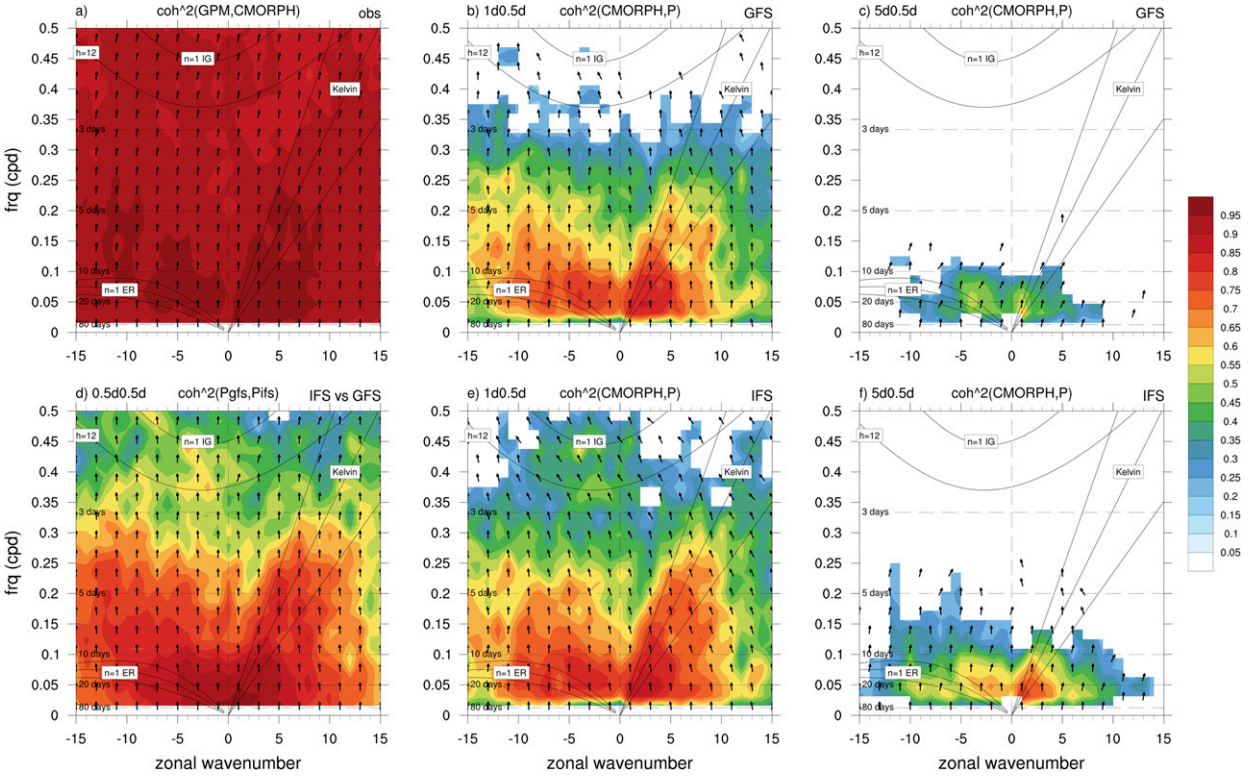

FIG. 10. As in Fig. 9, but for coherence squared between (a) GPM and CMORPH, (b) CMORPH and GFS 1d0.5d, (c) CMORPH and GFS 5d0.5d, (d) IFS and GFS 0.5d0.5d, (e) CMORPH and IFS 1d0.5d, and (f) CMORPH and IFS 5d0.5d. Statistically significant signals at the 95% level are color shaded (white areas are not significant). Arrows show the phase between the two variables where upward-pointing arrows are in phase, leftward-pointing arrows mean, for example, that at higher frequencies GPM lags CMORPH by 1/8 cycle (see phase convention in Table 2).

distribution of power is closer to observations in the IFS than the GFS. For instance, at intraseasonal time scales, the GFS ratio of eastward to westward power is smaller and closer to 1 than for the IFS, where a larger ratio indicates better MJO model performance (Waliser 2009).

Another way to diagnose CCEWs is to calculate space–time cross spectra between variables (Wheeler and Kiladis 1999; Hendon and Wheeler 2008; Waliser 2009). These cross spectra can be interpreted as decompositions of the lag-correlation between variables for each wavenumber and frequency, and they provide insights on the space–time scales that make up the overall skill. The cross-spectral approach is applied here to investigate the scales that contribute to QPF skill and then to assess the relationship between rain rates and low-level divergence, which is chosen here as a proxy for the large-scale mass circulation.

Figure 10a shows that coherence between GPM and CMORPH is very high (> 0.9) across space–time scales, which is not surprising given the strong gridpoint temporal correlations between them (Fig. 6a). According to our phase convention (see Table 2), GPM and CMORPH are in phase at most scales, except that GPM slightly lags CMORPH around the WIG spectral region where slightly leftward-pointing arrows are seen. The patterns of coherence and phase between GFS and IFS in the short range (0.5d0.5d, Fig. 10d) are similar to GPM and CMORPH in the sense that models tend to be in phase with the highest coherence located at the lowest frequencies and wavenumbers. Higher coherence extends to synoptic scales along the Kelvin, WIG, and ER dispersion curves as well as to the westward spectral region that corresponds to tropical depression (TD) disturbances (Takayabu 1994; Dunkerton and Baldwin 1995; Wheeler and Kiladis 1999). Within the first

TABLE 2. Display of the phase sign convention for any two variables X and Y.

↑	X in phase with Y
→	X leads Y by a 1/4 of a cycle
↓	X out of phase with Y
←	X lags Y by a 1/4 of a cycle

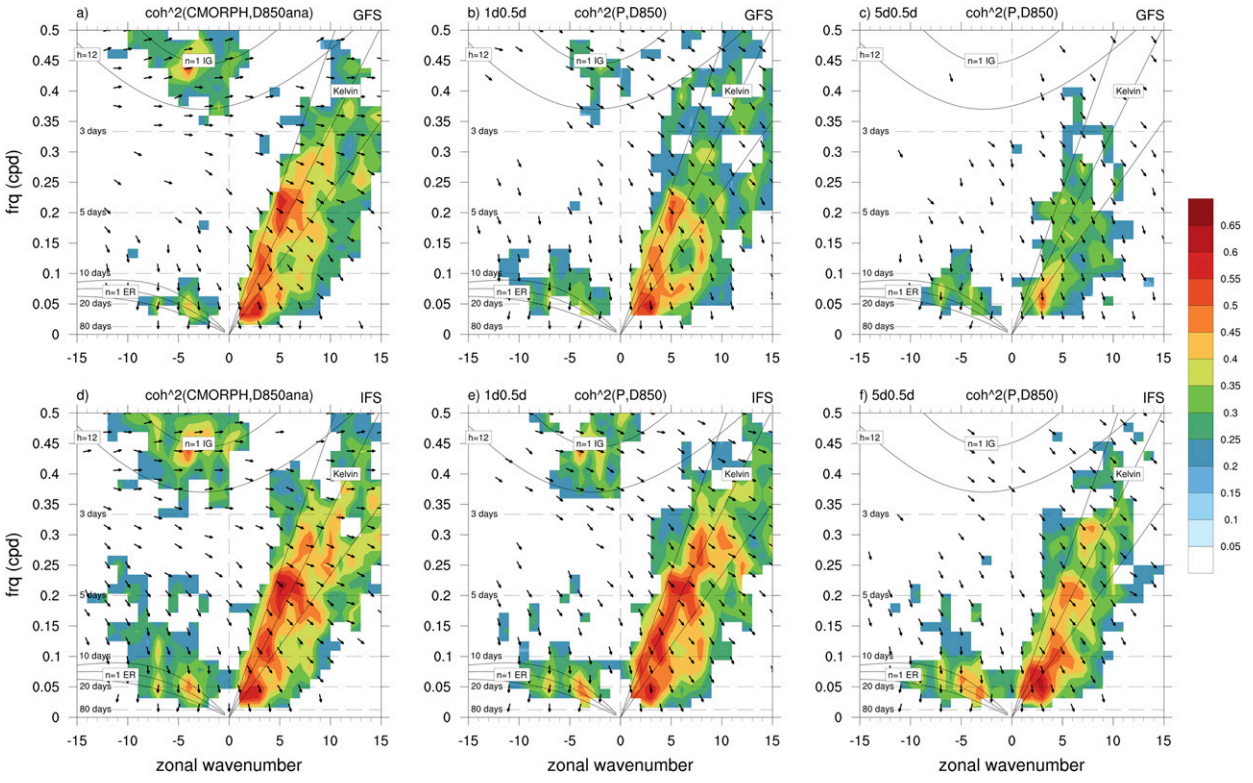

FIG. 11. As in Fig. 10, but for coherence between precipitation *P* and 850-hPa divergence (D850). (a) CMORPH and GFS D850 analysis, (b) GFS *P* 1d0.5d and D850 1d, (c) GFS *P* 5d0.5d and D850 5d, (d) CMORPH and IFS D850 analysis, (e) IFS *P* 1d0.5d and D850 1d, and (f) IFS *P* 5d0.5d and D850 5d. Arrows pointing down and to the right mean that precipitation leads low-level divergence, or equivalently, precipitation lags low-level convergence (see phase convention in Table 2).

forecast day, coherence between models and observations is already much weaker than between GPM and CMORPH (Figs. 10b,e), but the level of coherence associated with the MJO is similar in the GFS and IFS. In contrast to the intraseasonal signal, the IFS is more coherent with CMORPH and GPM (not shown) at synoptic scales than the GFS, particularly along Matsuno's dispersion curves. Rainfall from both models is in phase with observations at most scales, except that the IFS phasing at higher frequencies implies that IFS rainfall tends to lead the CMORPH and GPM (not shown), suggesting that IFS Kelvin waves and WIGs propagate too fast. Coherence along WIGs and higher-frequency Kelvin waves is much lower in the GFS, implying that this model has a difficult time propagating these small-scale waves, even when they are reasonably well initialized. At the medium range (5d0.5d), synoptic coherence drops dramatically in both models, but more so for the GFS (Figs. 10c,f). For example, while low, the IFS still shows significant coherence extending to higher frequencies than the GFS, with phases consistent with those in the observations. In addition, the ER and MJO coherence is now substantially stronger in the IFS than in the GFS. These results are consistent with the IFS having an overall superior tropical QPF performance, and suggest that the IFS draws skill from its better ability to propagate initialized equatorial waves. It is important to note that the lack of spectral power or coherence does not necessarily imply that the model is not able to initiate a particular disturbance, but rather that their triggering is not consistent when comparing forecasts that are initialized 12 h apart.

Figure 11 displays the coherence and phase between precipitation and low-level divergence (D850) and reveals patterns consistent with studies such as Hendon and Wheeler (2008) where much longer records of analyzed and observed data are used. Figures 11a and 11d show that, for both models, coherence between the respective analyses of low-level divergence and CMORPH estimates are very similar, peaking along the MJO and Matsuno's dispersion curves, with nearly identical results obtained when using GPM (not shown). The phase depends on the wave type, so for example, low-level convergence leads precipitation by about 1/4 and 1/8 of a cycle for WIG and Kelvin waves, respectively, with a smaller lag seen for the MJO (Figs. 11a,d). In addition,

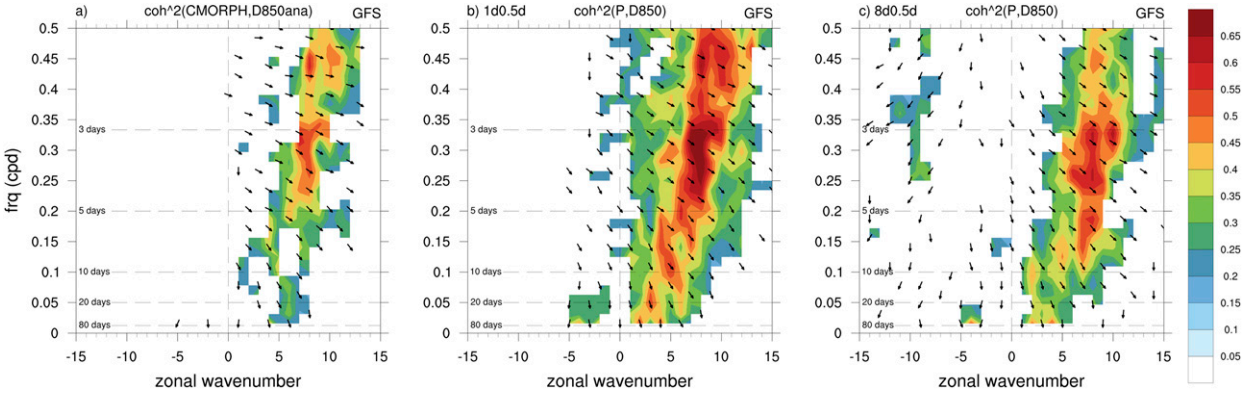

FIG. 12. As in Fig. 11 (top), but that the coherence is averaged from 35° to 50°N and the analysis of (a) CMORPH and GFS D850, (b) GFS *P* 1d0.5d and D850 1d, and (c) GFS *P* 8d0.5d and D850 8d is shown.

ER low-level convergence tends to be more in phase with precipitation than in the other equatorial disturbances. At 1-day lead time (Figs. 11b,e), both models tend to show low-level convergence and precipitation along Matsuno's dispersion curves more in phase than in the observations (Figs. 11a,d) with this tendency stronger in the GFS. This phasing implies that rainfall associated with low-level convergence develops too quickly in both models. Phase leads and coherences deteriorate further at 5-day lead time in both systems (Figs. 11c,f), although the IFS compares notably better to observations than the GFS. Another important point is that, while both models have similar skill in low-level divergence (shown in Fig. 8), the IFS is better able to simulate the relationship between low-level divergence and precipitation than the GFS.

Taken together, these results suggest that coupling between thermodynamics and large-scale circulation is stronger in the IFS than in the GFS, which likely stems from differences in the convective parameterizations (Hirons et al. 2013a,b). In particular, it is known that some convective parameterizations tend to be active too often and too early depending on the entrainment rate, which affects CCEW propagation (Lin et al. 2008; Frierson et al. 2011; Hirons et al. 2013b). Importantly, the coherence analysis shown here provides a process-based diagnostic that could be useful for model development aimed at improving short- to medium-range forecasts in the tropics.

The space–time coherence analysis can be applied to the extratropics as well, which is illustrated in Fig. 12 for coherence averaged from 35° to 50°N between GPM and GFS D850. The main difference between tropics and extratropics is that coherence and spectral power (not shown) peak along synoptic scales associated with the storm track, as opposed to intraseasonal scales, with a lag of about a quarter of cycle between D850 and rainfall. Both models are very similar in coherence and phase (IFS not shown), and both are able to sustain these spectral relationships for much longer lead times than in the tropics (see day 8 in Fig. 12c). Unlike the tropical QPF, the differences in extratropical QPF model performance between the GFS and IFS are not likely associated with differences in their ability to propagate midlatitude baroclinic waves.

e. *Relationship between tropical and extratropical QPF*

The atmospheric response to variations in latent heating in the tropics is known to extend well beyond the tropics (Sardeshmukh and Hoskins 1988; Grimm and Silva Dias 1995; Branstator 2014) and past studies have shown that a reduction of tropical forecast errors leads to improved medium to extended forecast skill over distant regions such as North America and Europe (Ferranti et al. 1990; Jung et al. 2010). One general mechanism for tropical–extratropical teleconnections is that latent heating in the tropics generates subtropical "Rossby wave sources," which then give rise to Rossby wave trains that propagate to higher latitudes (Sardeshmukh and Hoskins 1988). The details of the wave patterns and their paths depend on interactions between the horizontal and vertical distribution of the tropical heat source due to precipitation and the large-scale basic-state flow (Sardeshmukh and Hoskins 1988; Grimm and Silva Dias 1995; Newman and Sardeshmukh 1998). One potential way to test if these mechanisms are present in the GFS and IFS is to verify whether there is a relationship between tropical short-range QPF skill and extratropical skill at later lead times, which would then suggest that when the modeled tropical heat source is well represented, extratropical skill is gained because of improvements in the generated Rossby responses.

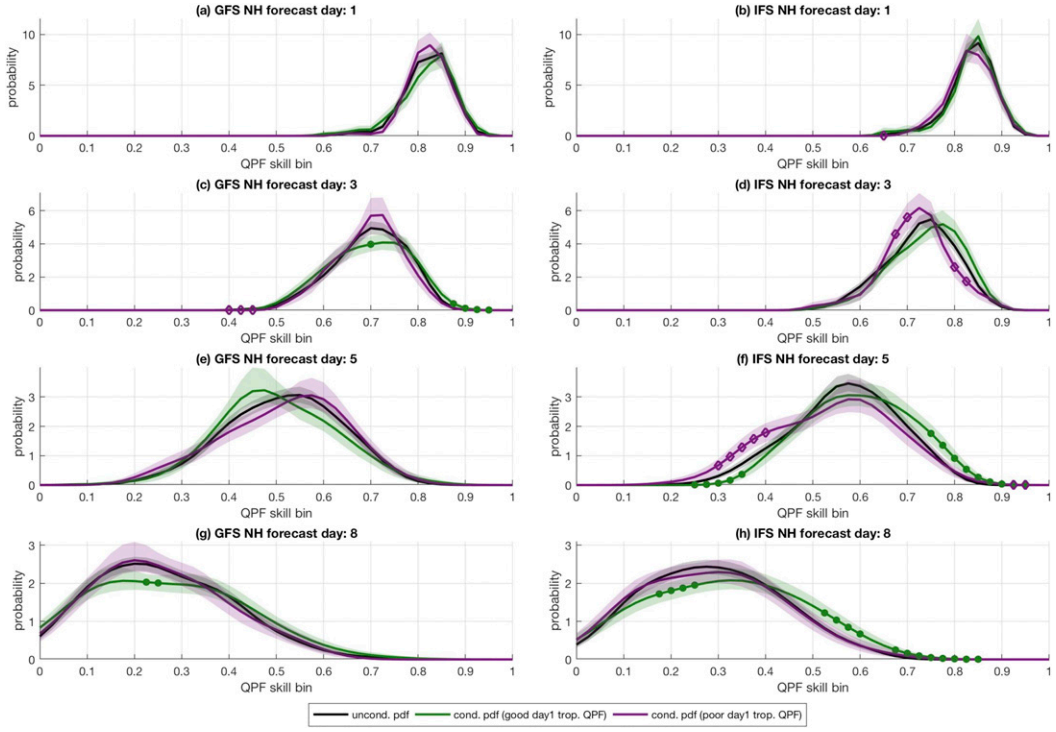

FIG. 13. Pdfs of NH QPF forecast skill score (from top to bottom) for forecast day 1, 3, 5, and 8 using daily pattern correlations for the period January 2015–March 2016. (left) NCEP GFS and (right) ECMWF IFS pdfs. The pdfs over the entire period are shown in black and the conditional pdf on good (poor) day-1 tropical QPF skill is shown in green (purple). Shading represents the 90% CI of the probability in each bin and symbols highlight bins where the CI of the conditional pdf does not overlap with the unconditional pdf.

To assess the relationship between tropical QPF and extratropical forecast skill in the GFS and IFS, we first estimate the probability density function (pdf) of all the daily rainfall skill scores using pattern correlations (defined in section 2d) over the band extending from 35° to 55°N ($\text{pdf}_{\text{NH,QPF}}$). These unconditional pdfs are shown by the black curves in Fig. 13, where the pdfs become broader and shift toward the left (lesser skill) at longer lead times. We then condition similar pdfs on 1-day lead time tropical QPF, where we use the 1d1d pattern correlation for the region between 10°S and 10°N and define good (poor) tropical forecasts as days where scores are above (below) its upper (lower) quartile. The pdfs are generated using a kernel estimation method (MATLAB's ksdensity function) and, while more noisy, results are similar using a histogram calculation. Confidence intervals for the pdfs are estimated using a bootstrap method with 1000 random samples, which is used to calculate the 5th and 95th percentiles of the pdf at each bin. The conditional $\text{pdf}_{\text{NH,QPF}}$ for various lead times and for each model are superimposed on the unconditional pdfs in Fig. 13.

As mentioned above, our underlying hypothesis is that if the model is able to represent the dynamical response to tropical heating, then the $\text{pdf}_{\text{NH,QPF}}$ should shift to higher skill when the 1-day lead tropical forecast is good and to lower skill when the tropical QPF is poor. This is precisely the tendency shown in Fig. 13 for the IFS (right panels), where the good tropical forecasts are associated with a shift to the right (green curves) when compared to the poor tropical forecasts (purple curves) at all leads out to 8 days. However, this is not true for the GFS (left panels). The shift in the conditional pdfs is also seen when using NH geopotential at 500 hPa and divergence at 200-hPa skill scores as opposed to QPF, but it is not seen in the equivalent Southern Hemisphere (SH) latitudinal band (not shown). The lack of a signal in the SH is interesting and could be due to differences in wave guiding (Hoskins and Ambrizzi 1993) or their interactions with the ENSO basic state, which we plan on investigating in a future study using more extensive reforecast datasets. Importantly, when tropical forecast skill is quantified in terms of the upper-level divergence (D200), the analogous conditional $\text{pdf}_{\text{NH,QPF}}$s are shifted in a similar way as the IFS in Fig. 13, except that the shift toward better (worse) scores when the tropical D200 skill is good (poor) is seen for both IFS and GFS (see the

online supplemental material). This indicates that if the GFS short-range forecast of tropical upper-level divergence flow is good, then there is a downstream positive impact in extratropical skill. As with the coherence analysis shown in the previous section, this result strongly suggests that GFS forecast errors are more likely associated with model physics rather than dynamics, because tropical rainfall and circulation are not as realistically coupled as in the IFS. This view is also consistent with the results from the previous sections that equatorial waves are a key mechanism linking tropical moist convective processes and large-scale dynamics.

4. Summary and conclusions

In this study the skill of the GFS and IFS operational forecasts out to 10-day lead time is analyzed, with a focus on the relationship between equatorial waves and quantitative precipitation forecasts. Our underlying hypothesis is that models that are able to initialize and propagate disturbances that are inherently deterministic such as Kelvin waves and the MJO will perform better in terms of tropical QPF, which will also lead to improvements in extratropical forecasts. Forecasts from the operational IFS and GFS forecast systems are compared, using twice-daily initializations from January 2015 to March 2016. The QPF performance is evaluated against GPM and CMORPH precipitation satellite estimates or the models' respective analysis in the case of dynamical variables.

Both systems produce good very short-range (12 h–1 day) forecasts of rainfall with similar mean patterns of daily precipitation rates when compared to each other and to satellite observations. This indicates that the initialized fields are reasonably good, especially when considering that precipitation data are not directly assimilated into either system. At longer lead times, the GFS develops a much stronger dry bias at the equator than the IFS (see Figs. 3 and 4). We note that, except for the driest regions, the amplitude of the normalized differences between the two models at 1-day lead are of the same order of magnitude as the differences between GPM and CMORPH mean rain rates, although the geographic patterns of these differences are very distinct. More substantial differences are found in the 1-day lead probability density distribution of rain rates in the models versus observations, particularly when pdfs are estimated using rain rates at $0.25° \times 0.25°$ and within the initial 6 h (top row in Fig. 5). Both models tend to overestimate light to moderate rain rates and underestimate no rain occurrences, very light, and strong rain rates. This is true even when accounting for the differences between GPM and CMORPH, which also highlights the issue of large uncertainties in estimates of rainfall (Gehne et al. 2016). The differences between the models and observations are partially due to the fact that the standard output from the models reflects the average rain rate over the forecast period whereas both GPM and CMORPH report instantaneous rain rates. Nevertheless, even when calculating daily averages, the pdfs still suggest that modeled rain rates vary less than in satellite observations. Interestingly, by calculating daily averages, we also find that pdfs are in much closer agreement over the midlatitudes than in the tropics (bottom row in Fig. 5), which is in line with the overall increase in forecast skill with distance from the equator.

Maps of gridpoint temporal correlations show that, similarly to mean rain rates, the day-1 forecast time series of rainfall are well correlated with observations, but much more so in the extratropical regions and over the oceans (Fig. 6). The meridional gradient in QPF skill is seen when comparing models with one another or with observations, whereas the two observational datasets are better correlated with each other regardless of the latitude except over regions of elevated terrain (see Dai et al. 2007). QPF skill decays very quickly with lead time everywhere, and the geographic contrasts seen at 1-day remain intact out to longer lead times. By comparing averaged gridpoint correlations over the tropics versus NH, it is clear that the IFS performs better than the GFS in both regions. This can be seen at very short lead times and, therefore, could be partially due to the differences in their data assimilation systems. There are, however, some interesting contrasts in how skill decays over the tropics versus over the NH. For example, while the differences between IFS and GFS skill continue to increase with lead time from days 1 to 10, over the NH this difference is nearly constant for the first 2 days, then it increases until day 8 when it starts to decrease. Moreover, in agreement with a number of other studies (Haiden et al. 2012; Zhu et al. 2014; Wheeler et al. 2017), we find that skill deteriorates much faster at earlier forecasts in the tropics than in the extratropics, but then at around day 3 extratropical forecast skill starts to decay faster than tropical skill so that by day 7 for the IFS and day 8 for the GFS, tropical QPF skill scores are higher than NH QPF. Similar behavior is seen in dynamical variables such as D850 and Z500 (Fig. 8), except that tropical D850 skill in the IFS and GFS is very similar over the tropics, suggesting that physical parameterizations related to precipitation might explain the differences in QPF between GFS and IFS.

To investigate the relationship between tropical QPF and equatorial waves, we apply longitude–time spectral

analysis of forecast time series. From this analysis it is clear that in the GFS and IFS most of the coherence between modeled and observed precipitation at the 1-day lead comes from low-frequency variability, but the IFS shows more coherence at synoptic scales, particularly along WIG and Kelvin wave dispersion curves (Fig. 10). The phase between modeled and observed precipitation shows that while the two fields are in phase at low frequencies, there is a tendency for forecast precipitation to occur too soon compared to observations at higher frequencies. The coherence between analyzed low-level divergence and observed rainfall aligned with the MJO and Matsuno's dispersion curves is similarly represented using the IFS or the GFS (Figs. 11a,d). However, the coherence between IFS forecast low-level divergence and rainfall is closer to the observed relationship for longer lead times than in the GFS, implying that the IFS is more able to realistically propagate such disturbances. The enhanced coherence at synoptic scales between IFS modeled rainfall, its low-level divergence, and observed rainfall imply that CCEWs contribute substantially to tropical short- to medium-range QPF deterministic skill, which is consistent with Ying and Zhang (2017, 2018). The relative roles of the quality of tropical initial conditions versus physics in the ability of the model to propagate CCEWs is an interesting topic that we are currently investigating.

To relate the quality of forecasts in the tropics to the extratropics, we applied a conditional pdf approach showing that when very short-range tropical forecast are good, NH extratropical medium-range forecasts are more likely to be good and vice versa. When the quality of the tropical forecast is measured from its QPF skill, this relationship is found in the IFS beyond day 3, but not in the GFS. The timing is roughly consistent with the time it takes for Rossby wave trains excited in the tropics to propagate to midlatitudes (Newman and Sardeshmukh 1998; Branstator 2014). When the quality of the tropical forecast is measured from its D200 very short-range skill, this relationship is found in both forecast systems. The simplest interpretation of these results is that the GFS short lead time tropical forecast errors related to horizontal and vertical distribution of latent heating quickly lead to errors in the tropical atmospheric circulation response; therefore, any potential extratropical skill deriving from wave trains excited in the tropics is deteriorated. In contrast, if the GFS is initialized with good tropical circulation (based here on the quality of D200), then the tropical wave source associated with tropical divergence is better represented and, therefore, extratropical skill is drawn from wave patterns forced from the tropics. In the IFS, because circulation and precipitation are more tightly related, errors in precipitation and divergence are more strongly related, therefore, both initial QPF and D200 tropical skills modulate extratropical skill at later lead times. The pdf results shown here indicate that the difference in midrange extratropical skill in the IFS and GFS at least partially originates from tropical short-range forecast errors.

By contrasting the GFS and IFS, this study suggests that a NWP model's ability to initialize and propagate equatorial waves is related to global skill in short- to medium-range forecasts. Therefore, diagnostics such as the coherence and phase between low-level divergence and precipitation, which are relatively easy to calculate, are potentially useful targets for model development aimed at reducing tropical forecast errors and or understanding sources of errors. However, it is important to note that while the relationship between equatorial waves and QPF is compelling, we cannot assess a causal relationship using the diagnostics applied here. For instance, both improved QPF skill and representation of equatorial waves could be unrelated when both are the result of an improved representation of the background state or vertical profiles of latent heating. We plan on further investigating these issues in a future modeling study.

One aspect of model performance that we excluded from the main analysis presented here is related to the diurnal cycle of rainfall. Using an earlier version of the IFS, Kidd et al. (2013) found that the modeled amplitude of the diurnal cycle is too strong and the peak rainfall occurs too early over most of the tropics in comparison to satellite estimates. These errors have been reduced in a more recent model version (Bechtold et al. 2014). We have carried out a similar analysis (see appendix B) that demonstrates similar results for the IFS version used here. In particular, the phase difference between IFS and observations is strongest over tropical land, a result that was confirmed using both CMORPH and GPM. As a positive note regarding the GFS performance, the diurnal cycle phasing is one example where the GFS is more realistic than the IFS. As shown in appendix B, the phase of the GFS diurnal cycle of rainfall is similar to the IFS over most of the globe, except that early afternoon peak of the GFS diurnal rainfall cycle over tropical land agrees with observations, as opposed to the IFS peak that often occurs a few hours earlier. The causes and implications of NWP errors associated with the diurnal cycle of tropical rainfall is fundamental to the problem of modeling precipitation, and is certainly a topic that merits further investigation.

Acknowledgments. We thank Lisa Bengtsson, Pedro Silva Dias, and three anonymous reviewers for their comments and suggestions, which greatly helped to improve the manuscript. This study was funded by NOAA/Earth System Research Laboratory Physical Sciences Division.

APPENDIX A

Fractions Skill Score

To investigate the role of spatial scales in QPF skill we use the fractions skill score (FSS; Roberts 2008; Roberts and Lean 2008), which among other useful properties, yields an intuitive view of what spatial scales a forecast can be considered useful. The FSS calculation is relatively simple and we summarize this here, and refer to Roberts and Lean (2008) for greater detail.

The first step in the FSS calculation is to convert the data to a binary field by choosing an absolute (i.e., a given value) or relative (i.e., quartile) threshold. Here the threshold used is the upper quartile of the CMOPRH twice-daily averaged rain rates during the analyzed period within the two regions (i.e., tropics or NH); however, our main conclusions are not sensitive to the choice of a reasonable threshold. The next step is to generate the fractions associated with each scale, which is calculated at each grid point in the region of verification. The scale refers to the choice of number of grid points surrounding the target grid point and the fraction is the number of such grid points where rain rates exceeds the threshold divided by the total number of grid points. In our case, we use squared areas and because we start with data at 0.25° × 0.25°, a scale of 0.25° means that there is only one grid point in that region; therefore, fractions can be only 0 or 1. A scale of 0.75° means that the 8 grid points surrounding the target grid point are used to calculated the fraction and fractions now can be 0, 1/9, 2/9, . . . , 1 and similarly to larger scales. The FSS is then calculated following Eqs. (6) and (7) in Roberts and Lean (2008), noting that we calculate the FSS for two latitudinal bands: 10°S–10°N and 35°S–50°N.

Each panel in Fig. A1 corresponds to the mean FSS at a given scale and these show the evolution of tropical (darker lines) and NH (lighter lines) FSS with lead time. The flat lines show that the FSS between CMORPH and GPM in the tropics (black) and NH (gray) is very high, but the agreement at finer scales is better in the tropics than in NH. As scale increases, the FSS tends to be the same in both regions. The difference between NH and tropical QPF at short- to medium-range forecasts at the smallest scale (Fig. A1a) is consistent with the mean gridpoint correlations shown in Fig. 6, but it decreases with increasing scale. In fact, the very short-range QPF skill is nearly the same for both models and regions at 1.75°, and tropical skill is higher than extratropical skill beyond this scale. Interestingly, the IFS tropical skill is larger than that in the NH for most scales after day 6 or 7, but for the GFS, the transition only occurs for the largest scales shown in Fig. A1. When the switch from higher NH to higher tropical skill occurs, the timing of the transition is consistent with the correlation skill. The overall FSS decay with lead time is also consistent with the correlation skill decay. The FSS analysis shown here implies that the skill differences between IFS and GFS and tropics and extratropics are not overly sensitive to spatial scales.

APPENDIX B

Analysis of the Diurnal Cycle of Rainfall

To compare the performance of the IFS and GFS regarding the diurnal rainfall cycle, we first calculate the amplitude and phase of rain rates using the 0.25° × 0.25° and 6-hourly outputs. For models, the day-1 amplitude and phase are defined by fitting the first harmonics of diurnal cycle to rain rate among forecast lead times 6–24. Day-2 amplitude and phase are similar, except that amplitude and phase are calculated based on the 30–48 forecast time window, and similarly for all forecast days. Times at each grid point are converted to local solar time (LST). For observations, we derive daily time series of amplitude and phase in an analogous way after converting 3-hourly GPM and CMORPH rain rates to 6-hourly means centered at the same UTC hours as the models.

Mean diurnal phase at each grid point is calculated by averaging the daily time series of diurnal phase about its mode [see Sakaeda et al. (2017) for more details], which calculates the mean diurnal phase by equally weighting all days independent of the amplitude of the diurnal cycle. This method is slightly different from the method used in Bechtold et al. (2014), where the mean diurnal phase was calculated by fitting diurnal harmonics to mean hourly rain rates, which would preferentially weight the diurnal phase of days with relatively strong rain rates. Figure B1 shows that CMORPH and GPM are in very good agreement regarding the timing of peak rainfall and phases during the 15 months analyzed, consistent with similar analysis over extended periods (Dai et al. 2007). This analysis also indicates that over ocean, rainfall tends to peak a few hours earlier in IFS and GFS compared to GPM or

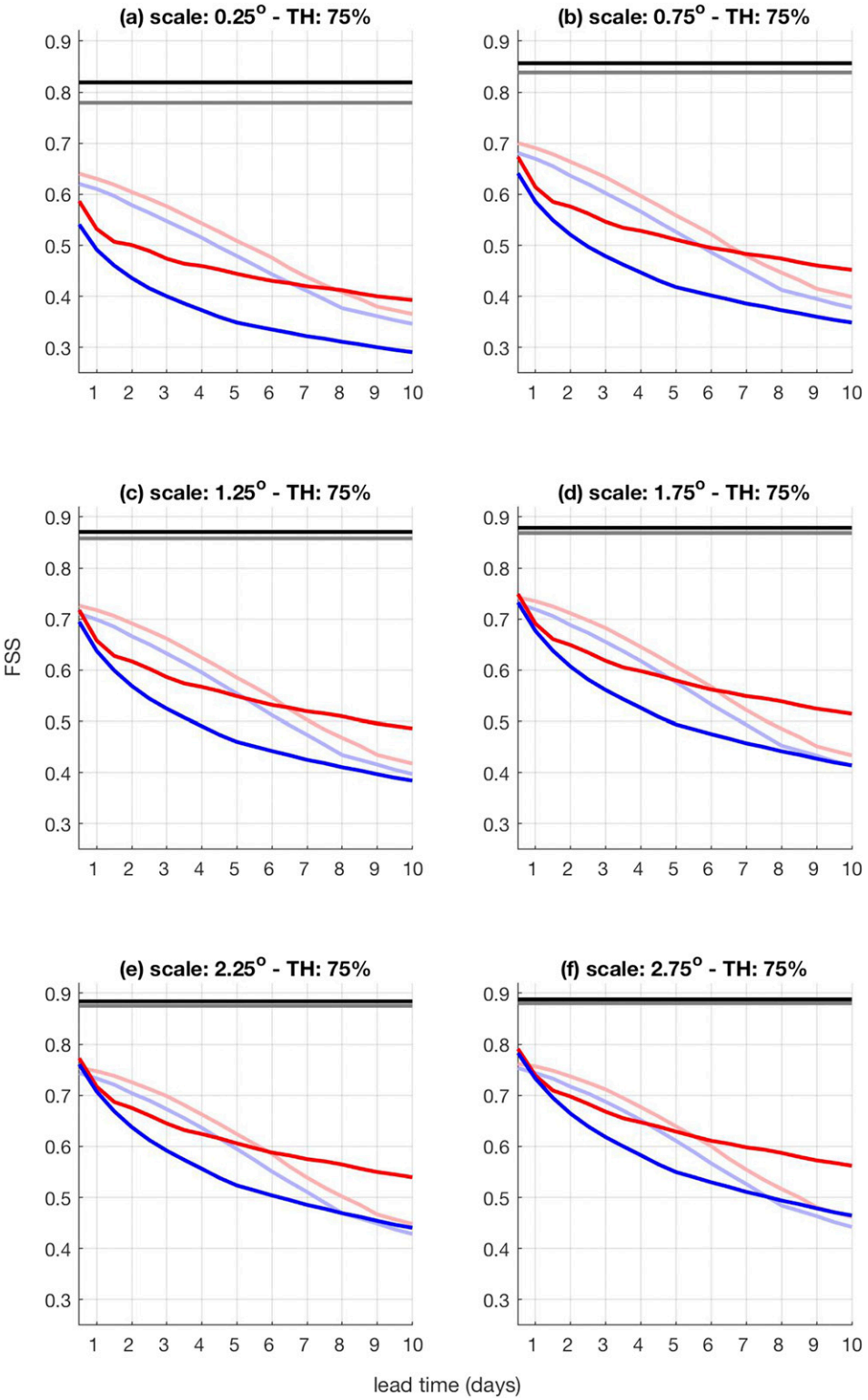

FIG. A1. FSS for the tropics (darker lines) and NH (lighter lines). NCEP GFS in blue and ECMWF IFS in red are verified against CMORPH. Gray and black correspond to FSS between GPM and CMORPH in the NH and tropics, respectively.

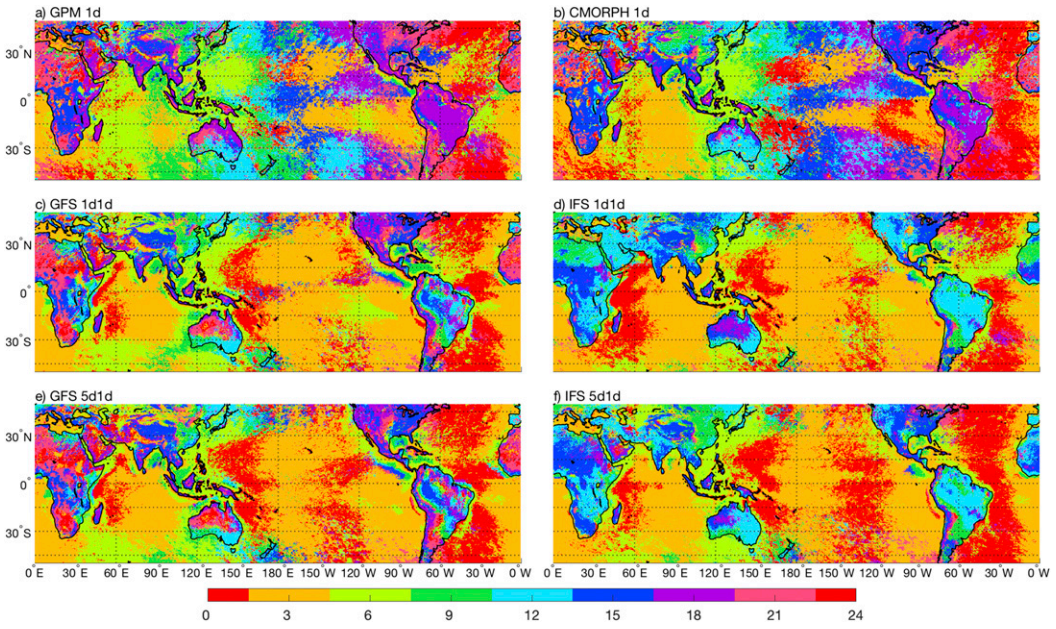

FIG. B1. Panels display the diurnal phase (LST hour of day) of (a) GPM, (b) CMORPH, (c) GFS 1d1d, (d) IFS 1d1d, (e) GFS 5d1d, and (f) IFS 5d1d.

CMORPH, whereas over tropical land, the IFS tends to have peak rainfall earlier than the GFS, GPM, and CMORPH.

The phase difference over tropical land between IFS and GPM has been previously reported (Kidd et al. 2013; Bechtold et al. 2014). In Bechtold et al. (2014) it is demonstrated that in a new version of the ECMWF model the difference in phase is less than in Kidd et al. (2013). While the IFS rainfall over tropical land is likely to peak in the tropics only a few hours earlier than in GPM, since we are analyzing 6-hourly output as opposed to hourly in Bechtold et al. (2014), we cannot assess what the precise phase difference is between models and observations. A similar analysis of the diurnal cycle was carried out with 3-hourly GFS and GPM, which shows that the GFS and GPM peak tropical rainfall over land occurs at the same time. We also find that the timing of peak rainfall does not depend strongly on lead time, which can be seen comparing the day-1 to day-5 phases in Fig. B1.

Similar to the map of the mean rain rates shown in Fig. 1, the IFS tends to underestimate the diurnal amplitude over the ocean (not shown). However, the IFS tends to overestimate the diurnal amplitude over land in contrast to its underestimation of mean rain rates (Figs. 2d,f). The GFS tends to overestimate the diurnal amplitudes over both land and ocean even though it tends to underestimate the mean rain rates (Fig. 2). This result indicates that the total rainfall variability explained by the diurnal cycle is too large in the GFS, perhaps related in part to its lack of rainfall variability associated with the convectively coupled equatorial waves.

REFERENCES

Ahlgrimm, M., and R. Forbes, 2014: Improving the representation of low clouds and drizzle in the ECMWF model based on ARM observations from the Azores. *Mon. Wea. Rev.*, **142**, 668–685, https://doi.org/10.1175/MWR-D-13-00153.1.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Bechtold, P., M. Köhler, T. Jung, F. Doblas-Reyes, M. Leutbecher, M. J. Rodwell, F. Vitart, and G. Balsamo, 2008: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quart. J. Roy. Meteor. Soc.*, **134**, 1337–1351, https://doi.org/10.1002/qj.289.

——, N. Semane, P. Lopez, J.-P. Chaboureau, A. Beljaars, and N. Bormann, 2014: Representing equilibrium and non-equilibrium convection in large-scale models. *J. Atmos. Sci.*, **71**, 734–753, https://doi.org/10.1175/JAS-D-13-0163.1.

Behrangi, A., M. Lebsock, S. Wong, and B. Lambrigtsen, 2012: On the quantification of oceanic rainfall using spaceborne sensors. *J. Geophys. Res.*, **117**, D20105, https://doi.org/10.1029/2012JD017979.

Boer, G. J., 1995: Analyzed and forecast large-scale tropical divergent flow. *Mon. Wea. Rev.*, **123**, 3539–3553, https://doi.org/10.1175/1520-0493(1995)123<3539:AAFLST>2.0.CO;2.

Branstator, G., 2014: Long-lived response of the midlatitude circulation and storm tracks to pulses of tropical heating. *J. Climate*, **27**, 8809–8826, https://doi.org/10.1175/JCLI-D-14-00312.1.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630, https://doi.org/10.1175/JCLI3884.1.

——, X. Lin, and K.-L. Hsu, 2007: The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes. *Climate Dyn.*, **29**, 727–744, https://doi.org/10.1007/s00382-007-0260-y.

Dole, R. M., and Coauthors, 2018: Advancing science and services during the 2015/16 El Niño: The NOAA El Niño rapid response field campaign. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-16-0219.1, in press.

Dunkerton, T. J., and M. P. Baldwin, 1995: Observation of 3–6-day meridional wind oscillations over the tropical Pacific, 1973–1992: Horizontal structure and propagation. *J. Atmos. Sci.*, **52**, 1585–1601, https://doi.org/10.1175/1520-0469(1995)052<1585:OODMWO>2.0.CO;2.

Ebert, E. E., J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 47–64, https://doi.org/10.1175/BAMS-88-1-47.

Ellis, T. D., T. L'Ecuyer, J. M. Haynes, and G. L. Stephens, 2009: How often does it rain over the global oceans? The perspective from CloudSat. *Geophys. Res. Lett.*, **36**, L03815, https://doi.org/10.1029/2008GL036728.

Ferranti, L., T. N. Palmer, F. Molteni, and E. Klinker, 1990: Tropical–extratropical interaction associated with the 30–60 day oscillation and its impact on medium and extended range prediction. *J. Atmos. Sci.*, **47**, 2177–2199, https://doi.org/10.1175/1520-0469(1990)047<2177:TEIAWT>2.0.CO;2.

Flato, G. A., 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866, https://doi.org/10.1017/CBO9781107415324.020.

Frierson, D. M. W., D. Kim, I.-S. Kang, M.-I. Lee, and J. Lin, 2011: Structure of AGCM-simulated convectively coupled Kelvin waves and sensitivity to convective parameterization. *J. Atmos. Sci.*, **68**, 26–45, https://doi.org/10.1175/2010JAS3356.1.

Gehne, M., T. M. Hamill, G. N. Kiladis, and K. E. Trenberth, 2016: Comparison of global precipitation estimates across a range of temporal and spatial scales. *J. Climate*, **29**, 7773–7795, https://doi.org/10.1175/JCLI-D-15-0618.1.

Grimm, A. M., and P. L. Silva Dias, 1995: Analysis of tropical–extratropical interactions with influence functions of a barotropic model. *J. Atmos. Sci.*, **52**, 3538–3555, https://doi.org/10.1175/1520-0469(1995)052<3538:AOTIWI>2.0.CO;2.

Guo, Y., D. E. Waliser, and X. Jiang, 2015: A systematic relationship between the representations of convectively coupled equatorial wave activity and the Madden–Julian oscillation in climate model simulations. *J. Climate*, **28**, 1881–1904, https://doi.org/10.1175/JCLI-D-14-00485.1.

Haiden, T., M. J. Rodwell, D. S. Richardson, A. Okagaki, T. Robinson, and T. Hewson, 2012: Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Wea. Rev.*, **140**, 2720–2733, https://doi.org/10.1175/MWR-D-11-00301.1.

Hayashi, Y., 1982: Space-time spectral analysis and its applications to atmospheric waves. *J. Meteor. Soc. Japan*, **60** (1), 156–171.

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, https://doi.org/10.1175/2008JAS2675.1.

Hirons, L. C., P. Inness, F. Vitart, and P. Bechtold, 2013a: Understanding advances in the simulation of intraseasonal variability in the ECMWF model. Part I: The representation of the MJO. *Quart. J. Roy. Meteor. Soc.*, **139**, 1417–1426, https://doi.org/10.1002/qj.2060.

——, ——, ——, and ——, 2013b: Understanding advances in the simulation of intraseasonal variability in the ECMWF model. Part II: The application of process-based diagnostics. *Quart. J. Roy. Meteor. Soc.*, **139**, 1427–1444, https://doi.org/10.1002/qj.2059.

Hirota, N., Y. N. Takayabu, M. Watanabe, M. Kimoto, and M. Chikira, 2014: Role of convective entrainment in spatial distributions of and temporal variations in precipitation over tropical oceans. *J. Climate*, **27**, 8707–8723, https://doi.org/10.1175/JCLI-D-13-00701.1.

Hoskins, B. J., and T. Ambrizzi, 1993: Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci.*, **50**, 1661–1671, https://doi.org/10.1175/1520-0469(1993)050<1661:RWPOAR>2.0.CO;2.

Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8**, 38–55, https://doi.org/10.1175/JHM560.1.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.*, **5**, 487–503, https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Jung, T., M. J. Miller, and T. N. Palmer, 2010: Diagnosing the origin of extended-range forecast errors. *Mon. Wea. Rev.*, **138**, 2434–2446, https://doi.org/10.1175/2010MWR3255.1.

Kerns, B. W., and S. S. Chen, 2014: ECMWF and GFS model forecast verification during DYNAMO: Multiscale variability in MJO initiation over the equatorial Indian Ocean. *J. Geophys. Res. Atmos.*, **119**, 3736–3755, https://doi.org/10.1002/2013JD020833.

Khouider, B., A. St-Cyr, A. J. Majda, and J. Tribbia, 2011: The MJO and convectively coupled waves in a coarse-resolution GCM with a simple multicloud parameterization. *J. Atmos. Sci.*, **68**, 240–264, https://doi.org/10.1175/2010JAS3443.1.

Kidd, C., E. Dawkins, and G. Huffman, 2013: Comparison of precipitation derived from the ECMWF operational forecast model and satellite precipitation datasets. *J. Hydrometeor.*, **14**, 1463–1482, https://doi.org/10.1175/JHM-D-12-0182.1.

Kiladis, G. N., M. C. Wheeler, P. T. Haertel, K. H. Straub, and P. E. Roundy, 2009: Convectively coupled equatorial waves. *Rev. Geophys.*, **47**, RG2003, https://doi.org/10.1029/2008RG000266.

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

Lin, J.-L., and Coauthors, 2006: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals. *J. Climate*, **19**, 2665–2690, https://doi.org/10.1175/JCLI3735.1.

——, M. I. Lee, D. Kim, I. S. Kang, and D. M. W. Frierson, 2008: The impacts of convective parameterization and moisture triggering on AGCM-simulated convectively coupled equatorial waves. *J. Climate*, **21**, 883–909, https://doi.org/10.1175/2007JCLI1790.1.

Lu, C., H. Yuan, E. I. Tollerud, and N. Wang, 2010: Scale-dependent uncertainties in global QPFs and QPEs from NWP model and satellite fields. *J. Hydrometeor.*, **11**, 139–155, https://doi.org/10.1175/2009JHM1164.1.

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, https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.

——, and ——, 1994: Observations of the 40–50-day tropical oscillation—A review. *Mon. Wea. Rev.*, **122**, 814–837, https://doi.org/10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2.

Majda, A. J., and S. N. Stechmann, 2009: The skeleton of tropical intraseasonal oscillations. *Proc. Natl. Acad. Sci. USA*, **106**, 8417–8422, https://doi.org/10.1073/pnas.0903367106.

Mapes, B., S. Tulich, T. Nasuno, and M. Satoh, 2008: Predictability aspects of global aqua-planet simulations with explicit convection. *J. Meteor. Soc. Japan*, **86A**, 175–185.

Matsueda, M., and H. Endo, 2011: Verification of medium-range MJO forecasts with TIGGE. *Geophys. Res. Lett.*, **38**, L11801, https://doi.org/10.1029/2011GL047480.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44**, 25–43, https://doi.org/10.2151/jmsj1965.44.1_25.

Newman, M., and P. D. Sardeshmukh, 1998: The impact of the annual cycle on the North Pacific/North American response to remote low-frequency forcing. *J. Atmos. Sci.*, **55**, 1336–1353, https://doi.org/10.1175/1520-0469(1998)055<1336:TIOTAC>2.0.CO;2.

Raymond, D. J., and Z. Fuchs, 2007: Convectively coupled gravity and moisture modes in a simple atmospheric model. *Tellus*, **59**, 627–640, https://doi.org/10.1111/j.1600-0870.2007.00268.x.

Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281–1305, https://doi.org/10.1175/1520-0493(1994)122<1281:REGING>2.0.CO;2.

Roberts, N. M., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169, https://doi.org/10.1002/met.57.

——, and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Sakaeda, N., G. Kiladis, and J. Dias, 2017: The diurnal cycle of tropical cloudiness and rainfall associated with the Madden–Julian oscillation. *J. Climate*, **30**, 3999–4020, https://doi.org/10.1175/JCLI-D-16-0788.1.

Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, **45**, 1228–1251, https://doi.org/10.1175/1520-0469(1988)045<1228:TGOGRF>2.0.CO;2.

Shukla, J., 1989: Tropical forecasting: Predictability perspective. *Aust. Meteor. Mag.*, **141–153** (37), 9158–9165.

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

Straub, K. H., and G. N. Kiladis, 2002: Observations of a convectively coupled Kelvin wave in the eastern Pacific ITCZ. *J. Atmos. Sci.*, **59**, 30–53, https://doi.org/10.1175/1520-0469(2002)059<0030:OOACCK>2.0.CO;2.

——, P. T. Haertel, and G. N. Kiladis, 2010: An analysis of convectively coupled Kelvin waves in 20 WCRP CMIP3 global coupled climate models. *J. Climate*, **23**, 3031–3056, https://doi.org/10.1175/2009JCLI3422.1.

Straus, D., and D. Paolino, 2008: Intermediate time error growth and predictability: Tropics versus mid-latitudes. *Tellus*, **61**, 579–586, https://doi.org/10.1111/j.1600-0870.2009.00411.x.

Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934, https://doi.org/10.1175/JCLI3672.1.

Takayabu, Y. N., 1994: Large-scale cloud disturbances associated with equatorial waves. Part I: Spectral features of the cloud disturbances. *J. Meteor. Soc. Japan*, **72**, 433–449, https://doi.org/10.2151/jmsj1965.72.3_433.

Taraphdar, S., P. Mukhopadhyay, L. R. Leung, and K. Landu, 2016: Prediction skill of tropical synoptic scale transients from ECMWF and NCEP ensemble prediction systems. *Math. Climate Wea. Forecasting*, **2** (1), https://doi.org/10.1515/mcwf-2016-0002.

Tulich, S. N., and G. N. Kiladis, 2012: Squall lines and convectively coupled gravity waves in the tropics: Why do most cloud systems move westward? *J. Atmos. Sci.*, **69**, 2995–3012, https://doi.org/10.1175/JAS-D-11-0297.1.

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

Waliser, D., 2009: MJO simulation diagnostics. *J. Climate*, **22**, 3006–3030, https://doi.org/10.1175/2008JCLI2731.1.

Webb, M. J., and Coauthors, 2015: The impact of parametrized convection on cloud feedback. *Philos. Trans. Roy. Soc. A*, **373** (2054), https://doi.org/10.1098/rsta.2014.0414.

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, https://doi.org/10.1175/1520-0469(1999)056<0374:CCEWAO>2.0.CO;2.

——, ——, and P. J. Webster, 2000: Large-scale dynamical fields associated with convectively coupled equatorial waves. *J. Atmos. Sci.*, **57**, 613–640, https://doi.org/10.1175/1520-0469(2000)057<0613:LSDFAW>2.0.CO;2.

——, H. Zhu, A. H. Sobel, D. Hudson, and F. Vitart, 2017: Seamless precipitation prediction skill comparison between two global models. *Quart. J. Roy. Meteor. Soc.*, **143**, 374–383, https://doi.org/10.1002/qj.2928.

Ying, Y., and F. Zhang, 2017: Practical and intrinsic predictability of multiscale weather and convectively coupled equatorial waves during the active phase of an MJO. *J. Atmos. Sci.*, **74**, 3771–3785, https://doi.org/10.1175/JAS-D-17-0157.1.

——, and ——, 2018: Potentials in improving predictability of multiscale tropical weather systems evaluated through ensemble assimilation of simulated satellite-based observations. *J. Atmos. Sci.*, **75**, 1675–1698, https://doi.org/10.1175/JAS-D-17-0245.1.

Zhang, C., 2005: Madden-Julian Oscillation. *Rev. Geophys.*, **43**, RG2003, https://doi.org/10.1029/2004RG000158.

——, 2013: Madden–Julian oscillation: Bridging weather and climate. *Bull. Amer. Meteor. Soc.*, **94**, 1849–1870, https://doi.org/10.1175/BAMS-D-12-00026.1.

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, https://doi.org/10.1175/MWR-D-13-00222.1.