Tropical Origins of Weeks 2–4 Forecast Errors during the Northern Hemisphere Cool Season

JULIANA DIAS,[a] STEFAN N. TULICH,[b] MARIA GEHNE,[b] AND GEORGE N. KILADIS[a]

[a] *NOAA/Physical Sciences Laboratory, Boulder, Colorado*
[b] *Cooperative Institute for Research in Environmental Sciences, NOAA/Physical Sciences Laboratory, University of Colorado Boulder, Boulder, Colorado*

(Manuscript received 8 February 2021, in final form 19 May 2021)

ABSTRACT: A set of 30-day reforecast experiments, focused on the Northern Hemisphere (NH) cool season (November–March), is performed to quantify the remote impacts of tropical forecast errors on the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS). The approach is to nudge the model toward reanalyses in the tropics and then measure the change in skill at higher latitudes as a function of lead time. In agreement with previous analogous studies, results show that midlatitude predictions tend to be improved in association with reducing tropical forecast errors during weeks 2–4, particularly over the North Pacific and western North America, where gains in subseasonal precipitation anomaly pattern correlations are substantial. It is also found that tropical nudging is more effective at improving NH subseasonal predictions in cases where skill is relatively low in the control reforecast, whereas it tends to improve fewer cases that are already relatively skillful. By testing various tropical nudging configurations, it is shown that tropical circulation errors play a primary role in the remote modulation of predictive skill. A time-dependent analysis suggests a roughly 1-week lag between a decrease in tropical errors and an increase in NH predictive skill. A combined tropical nudging and conditional skill analysis indicates that improved Madden–Julian oscillation (MJO) predictions throughout its life cycle could improve weeks 3–4 NH precipitation predictions.

KEYWORDS: Atmosphere; Extratropics; Northern Hemisphere; Tropics; Teleconnections; Hindcasts; Numerical weather prediction/forecasting; Subseasonal variability

1. Introduction

Tropical–extratropical teleconnections provide pathways for errors in global model forecasts at low and high latitudes to interact in ways that can ultimately degrade prediction skill in both regions at longer leads. Tropical–extratropical error relationships can be understood through Rossby wave theory, which shows how perturbations in upper-level tropical divergence tend to cause Rossby wave energy propagation from low to high latitudes (Hoskins and Karoly 1981; Sardeshmukh and Hoskins 1988). The effects of disturbances initiated near the equator take about 5–15 days to reach northern midlatitudes, depending on the source location and the basic state (Hoskins and Ambrizzi 1993; Newman and Sardeshmukh 1998; Branstator 2014). Local errors in the upper-level divergence field in the tropics can thus spread from their source region by altering the model's Rossby wave response, which can then impact weather predictions downstream. This theoretical expectation has been verified in numerical weather prediction (NWP) systems, which generally show that predictions of weeks 2–4 during Northern Hemisphere (NH) winter are improved in the midlatitudes when tropical errors are reduced (Ferranti et al. 1990; Hendon et al. 2000; Jung et al. 2010a; Dias and Kiladis 2019). The present study applies tropical nudging to evaluate and further characterize the relationship between tropical and extratropical subseasonal skill in the GFS, which is the medium range global model of NOAA's Unified Forecast System (UFS). The overarching goal is to better understand what types of model developments are needed in order to realize this potential source of skill from the tropics.

The approach of nudging, or relaxation, is a well-known strategy for investigating relationships between local and remote forecast errors. Briefly, local forecast errors are attenuated by nudging one or more prognostic variables toward model analyses, or reanalyses, within a limited area, while allowing the model to run freely everywhere else. Previous studies have shown that nudging of the tropical belt improves medium- to extended-range mean absolute errors in the midlatitudes, particularly over the North Pacific, North America, and the North Atlantic (Ferranti et al. 1990; Jung et al. 2010a; Jung 2011). Nudging has also been used to identify the tropical origin of specific cases of forecast errors (Jung et al. 2010b; Pohl and Douville 2011; Greatbatch et al. 2015). However, because it is generally hard to separate the predictable versus unpredictable sources of tropical error, it remains unclear as how much extratropical skill can be gained in practice, as well as what sort of model advancements are needed to realize such potential gains. Motivated by this issue, the present work further investigates the approach of tropical nudging, including its limitations, by exploring sensitivities to choices about which state variables are nudged, as well the region and duration over which the nudging is applied.

The extent to which tropical sources of remote subseasonal skill can be tapped remains unclear for a variety of reasons. For one, there are well known differences in the limits of subseasonal predictability in the tropics versus midlatitudes owing to the differing types of weather that prevail in each region. Weather disturbances in the extratropics are generally more constrained by a balance between gravitational and rotational

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

DOI: 10.1175/MWR-D-21-0020.1

restoring forces (e.g., baroclinic waves). Under weaker rotational effects and stronger insolation, tropical weather is more dependent on moist convective processes, a main driver of error growth (Selz and Craig 2015), and convectively coupled organized disturbances (e.g., convectively coupled equatorial waves, easterly waves and the MJO) emerge as the dominant players on subseasonal time scales. Recent tropical predictability studies based on cloud permitting models (Ying and Zhang 2017; Judt 2019), with some support from theory and observations (Li and Stechmann 2020), have suggested that tropical synoptic- to planetary-scale variability is more predictable than has previously been thought. Given the existence of tropical–extratropical teleconnections, this higher than expected tropical predictability suggests that we might not have reached the upper limit on the amount of subseasonal skill in the midlatitudes that can be drawn from the tropics, an idea that can be explored through tropical nudging.

Even if these recent tropical predictability estimates are overly optimistic, there is ample evidence that model advancements, particularly those dealing with the representation of subgrid-scale processes related to clouds and precipitation, can ultimately lead to more realistic simulations of organized tropical variability (Chikira and Sugiyama 2010; Frierson et al. 2011; Park 2014; Bengtsson et al. 2019). The fact that such improvements can, in turn, lead to improved tropical predictions is by now well established. By replacing reanalysis with an independent model prediction, tropical nudging experiments can provide a pathway for a more practical estimate of if and/or how much an improved model representation of tropical subseasonal variability should translate to improved midlatitude subseasonal predictions. This is potentially important at the current state of UFS development where recent versions of both medium and extended range predictions systems have been shown to underperform in the tropics in comparison to other state of the art prediction systems (Dias et al. 2018; Dias and Kiladis 2019). In addition, implementing nudging within the UFS could be useful for its development as it can be used as a diagnostic tool for relating local and remote error interdependencies when testing various model configurations.

While there are a large number of questions related to tropical–extratropical skill interdependencies that could be addressed with tropical nudging, here we focus on the nudging implementation in the GFS and initial results addressing a few practical questions:

1) Does tropical nudging applied to the GFS produce results consistent with previous studies using other models?
2) Do improvements in remote skill seen in dynamical variables extend to subseasonal precipitation predictions?
3) Which tropical fields need to be nudged in order to reduce remote errors, and where is nudging most effective?
4) How much of the remote error reduction originates from concurrent versus time-lagged tropical nudging?

In addition to addressing the questions above, we present a conditional skill analysis that tackles issues related to practical predictability originating from the tropics, including the role of the MJO on the high latitude error dependency. The set of tropical nudging reforecasts presented here is viewed as comprising the initial steps necessary to design experiments that more explicitly address such issues, as well as isolate what types of model developments would be helpful or most effective in realizing these sources.

2. Methodology

a. Description of subseasonal reforecasts experiments

The model used here is the NOAA/NCEP GFS v.15.1.1, which is an early version of NOAA's Unified Forecast Systems (UFS) that is currently used at NCEP for generating forecasts out to day 14. The effects of air–sea coupling are neglected, and all reforecasts are run out to day 30 at C128 resolution, which corresponds to a roughly $0.7° \times 0.7°$ horizontal grid with 64 vertical levels from the surface to 1 hPa. The treatment of the lower boundary is the same as in operations, where the initial sea surface temperature (SST) is taken from an analysis and then anomalies about a daily annual climatology are damped on a 90-day time scale. Details about the model formulation can be found at https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/implementations.php.

The focus of this study is on the extended NH cool season, from November to April. A series of 30-day reforecasts are generated once every 5 days, starting on 1 November of each year for the 20-yr period 1999–2018. The total number of reforecasts is 620, with 31 reforecasts per season. The spacing of 5 days between reforecasts is chosen as a compromise between computational/storage limitations and the need to adequately sample the full range of initial states. The model output is regridded to a 1° regular horizontal grid and all analyses are performed using that grid. The impact of tropical forecast errors is assessed by comparing a set of "free" reforecasts without nudging, also referred to as "control" (CNT), against several sets of tropical nudged reforecasts that are described below and summarized in Table 1.

1) NUDGING AND INITIALIZATION

The current GFS data assimilation (DA) system adopts an incremental analysis update (IAU) scheme, which was originally developed for the NASA Global Modeling and Assimilation Office (GMAO) GEOS model (Bloom et al. 1996) and is designed to reduce analysis-induced initial shocks during the model forecast step of the assimilation cycle. The IAU approach can be applied independently of the data assimilation cycle by calculating increments based on differences between a reanalysis dataset (as opposed to the model's own analysis) and a set of short-term forecasts. We use the IAU to effectively nudge the model to ERA interim reanalysis (Dee et al. 2011), as described below.

The IAU approach can be thought of as a refined form of nudging, in which a series of time-independent analysis increments or "updates" is applied to one or more of the model state variables, as part of a cycled forecast integration. Here the analysis increments (ΔI) are calculated generically as

$$\Delta I(t) = \frac{f_m(t - 3\,\mathrm{h}) - f_r(t)}{6\,\mathrm{h}}, \tag{1}$$

TABLE 1. Description of UFS reforecast datasets, where all reforecasts are initialized every 5 days with the same configuration, except for the nudging setup (details in the text).

Short name	Type of reforecast	Nudged variables	Extent of nudging region	Period (1 Nov–31 Mar)	Forecast/nudging length (days)
CNT	Control free	—	—	1999–2018	30/—
WTR	Wide tropical nudged	u, v, T, p, q	30°S–30°N	1999–2018	30/30
WTRuv	Wide tropical nudged	u, v	30°S–30°N	1999–2018	30/30
WTRqT	Wide tropical nudged	q, T	30°S–30°N	2016–18	30/30
NTR	Narrow tropical nudged	u, v, T, p, q	20°S–20°N	1999–2018	30/30
WTRwk1	Wide tropical nudged	u, v, T, p, q	30°S–30°N	1999–2018	30/7
WTRwk2	Wide tropical nudged	u, v, T, p, q	30°S–30°N	1999–2018	30/14

where $f_m = (t - 3\,\text{h})$ corresponds to a model forecast initialized 3 h prior to the valid time t, and $f_r(t)$ is the reanalysis field at t. The increment is then applied as a forcing term to another 6 h forecast cycle that is centered at time t (following the reanalysis output times) and is initialized from a $t - 3$ h restart file. This procedure is often referred to as "replay" (Orbe et al. 2017; Takacs et al. 2018) and is repeated for every 6-h period centered on the reanalysis output times, as illustrated in the schematic in Fig. 1.

Prior to generating the set of reforecasts, the model is first initialized on 1 November of each year from the ensemble mean of the Global Ensemble Forecast System (GEFSv12, https://www.emc.ncep.noaa.gov/users/meg/gefsv12/retros/), which provides initial data on the cubed sphere grid of the GFS v.15.1.1. A global nudged run is then performed from 1 November to 30 April, with model output saved every 6 h for verification and restart files saved every 5 days, as initial conditions for the set of reforecasts. Comparison of these global nudged runs against the original ERAi data showed only nominal differences for the fields of interest.

2) TROPICAL NUDGING FORMULATION

The tropical nudging experiments consist of repeating the global replay cycle described in the previous section, except that the ΔI are tapered to zero outside a specified latitudinal range centered at the equator. Two tapering functions are used, one where $\lambda = 1$ from 10°S to 10°N for the wide tropical nudging (WTR) and one where $\lambda = 1$ from 5°S to 5°N for the narrow tropical nudging (NTR). A hyperbolic tangent is used to smooth $\lambda = 1$ to $\lambda = 0$ over 20° in WTR and over 15° in NTR. The WTR tapering function is identical to the tapering function used in Jung et al. (2010a) where forecasts run free beyond 30°S/N. In NTR, forecasts are free beyond 20°S/N. The motivation for examining a narrower set up is to isolate the role of tropical circulation errors, in contrast to the WTR, where errors in the model's subtropical circulation are also partially reduced.

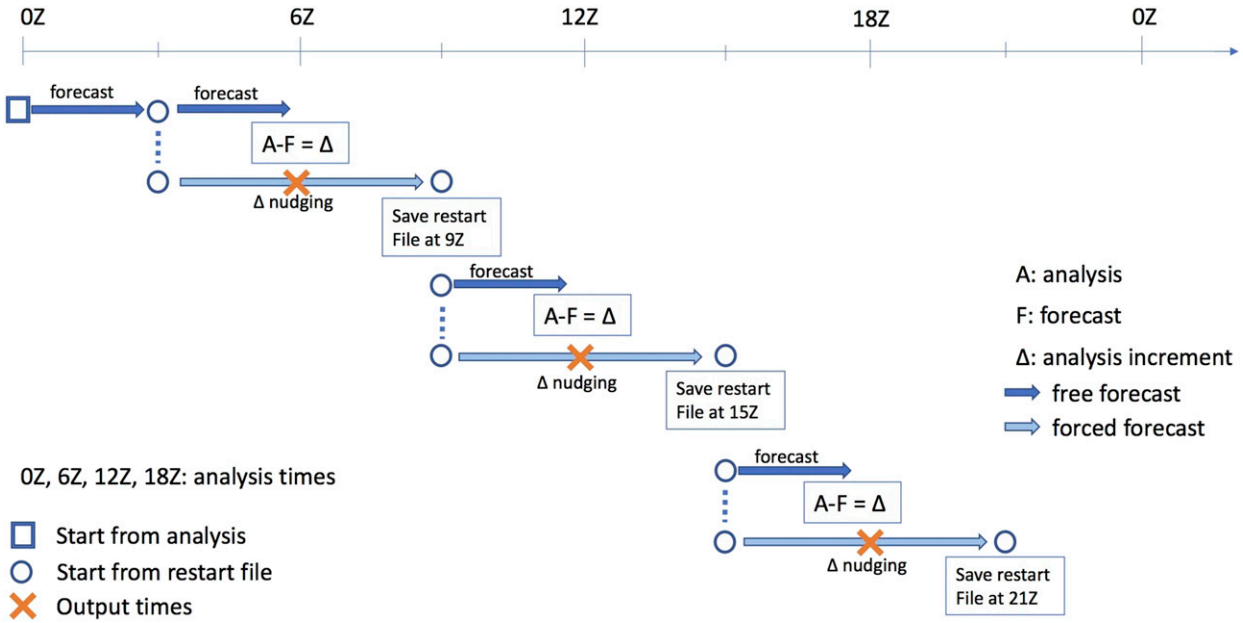

FIG. 1. Schematic of IAU approach to nudging (more details in the text).

In addition to varying the width of the tropical nudging belt, we also describe the impact of nudging different sets of model variables. The specific sets considered are: $[u, v, p, T, q]$, $[u, v, p, T]$, $[u, v]$, and $[T, q]$. The first set is chosen because, in principle, it should yield the ''best'' representation of the tropics as measured by the reanalysis. The second set enables more direct comparisons to results from Jung et al. (2010a), where specific humidity was not included in the relaxation scheme. It turns out, however, that exclusion of q in the GFS does not lead to significant differences in remote skill in comparison to when q is included, so results of this case are not reported. The third set, winds only, is motivated by theoretical considerations of the ''Rossby wave source'' (Sardeshmukh and Hoskins 1988), which implies that the upper-level flow is a primary driver of tropical error propagation from low to high latitudes. The final set $[T, q]$ is designed to contrast the winds-only set and spans just two seasons for testing purposes, as discussed further below. To address questions about the lag dependence of remote skill on tropical forecast errors, two additional sets of reforecasts are considered where tropical nudging is applied only during the first week or first two weeks of the forecast, as described further in section 3. The naming convention for each reforecast dataset and brief descriptions are listed in Table 1.

b. Verification metrics

We use standard verification metrics such as mean absolute errors (MAE), mean bias (BIAS), and anomaly pattern correlation (APC). Differences in skill among the reforecasts are assessed based on a random sampling method. Specifically, we first randomly resample with replacement each reforecast dataset 1000 times with a subsample size of 80% of the total number of initial times ($0.8 \times 620 = 496$), then we calculate the skill metric of interest in each of these subsamples. The reduced random sample size is somewhat arbitrary and is done to account for the fact that nearby initializations are not independent. We report the median skill over all random realizations, as opposed to the mean to minimize the influence of outliers. Confidence intervals (CI) correspond to the bounds defined as the 5th and 95th percentiles of the random sample distribution. Differences in skill among experiments are referred to as significant when the confidence intervals do not overlap. Skill over particular spatial domains is defined as the longitude and cosine-weighted latitude average. The APC is also area weighted, and anomalies are defined as full fields minus the forecast lead-dependent climatology. Climatology is defined for each reforecast dataset as the mean fields over the 20 realizations of each calendar day (every 5 days from 1 November to 31 March and 1999–2018) and for each lead day (1–30).

3. Results

a. The impact of tropical nudging on zonal mean biases

According to linearized Rossby wave theory, both the mean pattern and anomalies in the upper-level tropospheric circulation play a role in the triggering, pathway, and propagation characteristics of midlatitude Rossby waves originating from low latitudes (Hoskins and Karoly 1981; Sardeshmukh and Hoskins 1988). Therefore, if tropical nudging affects the model's mean state, changes in remote skill cannot necessarily be attributed exclusively to reduced errors in tropical subseasonal variability. To investigate this issue we first analyze how tropical nudging impacts the model's zonal mean state as function of lead time. Figure 2 illustrates that while week 1 upper-level zonal winds in all cases agree well with ERAi (Fig. 2a), the NH subtropical jets in CNT week 4 tend to be weaker than in ERAi, but stronger and closer to the equator in WTR and NTR (Fig. 2b). In contrast, when only winds are nudged, WTRuv week 4 NH subtropical jets tend to be weaker than in CNT, and slightly shifted toward the equator. Another difference in the zonal mean state can be seen in NTR week 4 Southern Hemisphere (SH) upper-level meridional winds. By analyzing latitude–longitude mean bias maps (not shown), we found that these differences in NTR meridional winds occur primarily over subtropical South America.

The differences in the mean upper-level zonal winds are consistent with the changes in the upper-level meridional temperature gradient that are imposed by the tropical nudging in NTR and WTR (red and blue curves, respectively, in Figs. 2c,d). Since pole to pole CNT upper-level temperatures display a cold bias when compared to ERAi, the net effect of nudging the model temperature to ERAi is to steepen the meridional temperature gradient in the subtropics and thus, strengthen the jets in both cases. In contrast to dynamical variables, model precipitation biases (Figs. 2e,f) generally develop earlier as illustrated by the week 1 averages. Comparing the profiles in WTR and WTRuv shows that the primary effect of nudging temperature and moisture in addition to winds is to change the model's tropical dry bias to a slight wet bias, whereas the model's midlatitude wet bias (CNT) remains largely unaffected in all cases. Importantly, because winds near the equator are very similar comparing WTR and WTRuv, differences in remote skill between the two nudging cases are likely related to differences in performance in the subtropics rather than in the tropics, and this is discussed further below.

b. The impact of tropical nudging on mean absolute errors

From previous nudging studies (Ferranti et al. 1990; Jung et al. 2010a; Jung 2011), we expect z500 MAE to be reduced over the NH when nudging the tropics and this is verified in Fig. 3a. The median MAE amplitude in all nudged reforecasts is less than in the free reforecast (CNT), particularly beyond week 2. This MAE attenuation also holds for precipitation rates (Fig. 3b), upper-level geopotential, winds and temperature, precipitable water and low level specific humidity (not shown). In the previous studies cited above, this systematic error reduction has been interpreted as indicating the potential for improving extended range NH predictions by reducing tropical errors. Figure 3b shows that nudging the tropics to reanalysis also lead to a decrease in weeks 2–4 precipitation errors in regions remote to the tropics. Figure 3 also illustrates how NH MAE varies depending on how the tropics are nudged. For instance, while errors are reduced less when nudging is applied to a narrower tropical region, the amount of error reduction in WTR and WTRuv is similar for both z500 and precipitation.

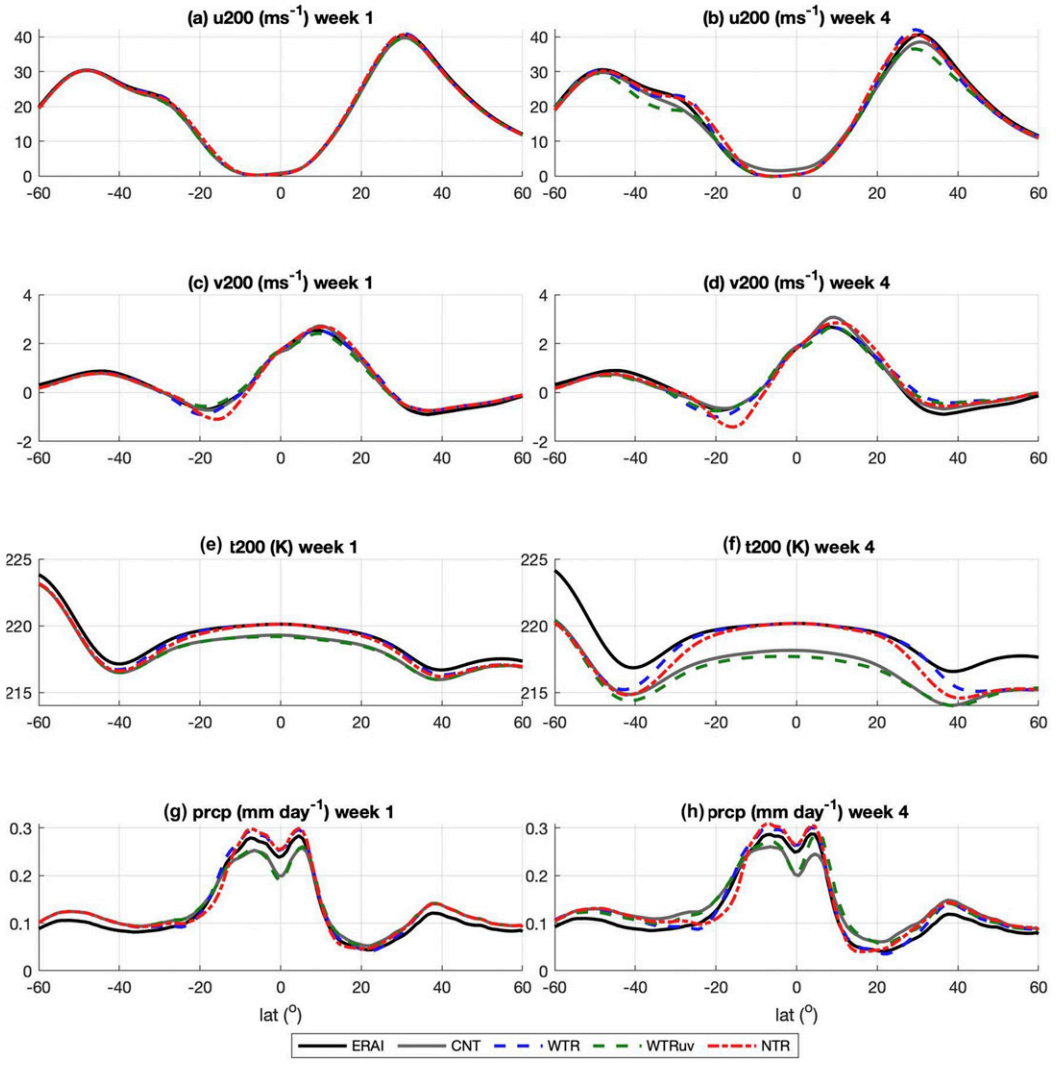

FIG. 2. Depiction of upper-troposphere (200 hPa) longitudinal averages of (a),(b) zonal and (c),(d) meridional winds, (e),(f) temperature, and (g),(h) mean precipitation rates. Averages including all initializations from November to March and reforecast lead days (left) 1–7 and (right) 22–28. ERAi averages match the reforecast valid times. Line color labels correspond to the experiments described in Table 1.

We have tested tropical nudging of only temperature and specific humidity with the WTR tapering and over a shorter period (the latest two seasons, not shown), which shows that errors are reduced more than in NTR, but not as much as in WTR and WTRuv. Taken together, these results support a stronger kinematic rather than thermodynamic link between tropical and NH midlatitude errors, consistent with expectations based on Rossby wave theory.

The first row in Fig. 4 displays maps of z500 MAE in CNT at weeks 1, 2, 3 and 4, which show that MAE increases most rapidly during weeks 1–2, with the largest values occurring at higher latitudes, as expected. Maps of the percent difference in MAE with respect to CNT in the nudged runs are plotted in Figs. 4e–4l, where the percent difference is calculated as: $\Delta\text{MAE} = 100 \times (\text{MAE}_{\text{NUD}} - \text{MAE}_{\text{CNT}})/\text{MAE}_{\text{CNT}}$. Extensive homogeneous regions with MAE amplitudes larger than 40% are shown in light gray, which roughly correspond to the tropical nudging zone. Maps of ΔMAE show that while errors are reduced in most regions outside the nudging region, MAE reductions are larger over the Northeast Pacific–western United States sector in the three cases displayed. It also appears that tropical nudging applied to winds only introduces a mass imbalance within the nudging region since the z500 ΔMAE is uniformly increased there. However, this imbalance does not have a strong remote imprint in the midlatitudes since WTR and WTRuv patterns are roughly similar. In contrast, at higher latitudes over East Asia, WTR and NTR are associated with an increase in ΔMAE that is not seen in WTRuv. Tropical nudging is not necessarily expected to reduce MAE everywhere because nudging affects the model's basic state (Fig. 2), and tropical–extratropical teleconnections are not uniform around the globe. In particular,

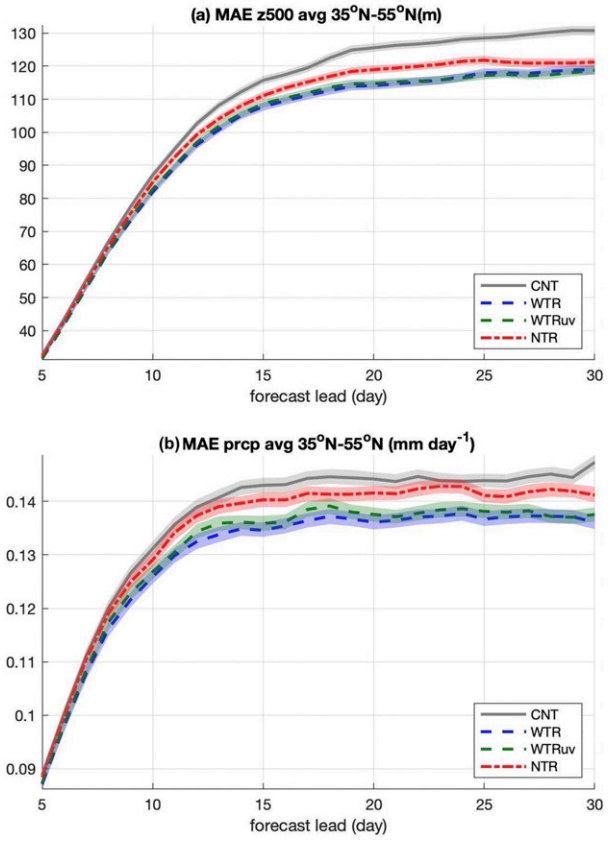

FIG. 3. The median MAE over the entire reforecast period and averaged from 35° to 55°N for geopotential height at (a) 500 hPa (z500) and (b) precipitation. The line color labels displayed in the legend correspond to the experiments described in Table 1. Shading displays the CI as described in the text.

by potentially distorting weather systems that move in and out of the nudging region, predictions of the remote response could be deteriorated.

The large-scale patterns where MAE is reduced are in qualitative agreement when comparing precipitation (Fig. 5) and z500. However, precipitation ΔMAE amplitude as well as the extent of the areas where differences are significant tend to be smaller. A remote region that stands out is the western United States where significant week 2–4 MAE reductions peaking around 30%–40%, with MAE decreases seen even in the NTR case. These results support the interpretation that reducing tropical forecast errors improves week 2–4 precipitation skill over the western United States, as discussed further below. Perhaps related to the tropical imbalance seen in z500, WTRuv MAE is slightly increased over the central United States, but not in WTR or NTR. While not the focus of the current work, we also note that precipitation weeks 3–4 MAE are significantly reduced along the South Pacific and Atlantic convergence zones, as well as over the southern Indian Ocean. These ''cloud band'' locations are well removed from the nudging region, and are known regions of significant tropical–extratropical interactions in the Southern Hemisphere. We also note that verification against satellite estimates of precipitation as opposed to reanalyses indicate similar patterns of ΔMAE, except that amplitudes are attenuated (not shown).

c. *The impact of tropical nudging on anomaly pattern correlations*

To further evaluate the impact of tropical nudging in predictions of the remote response, differences in the predicted anomaly patterns are assessed using APC. The APC is calculated within 35°–35°N and two longitudinal sectors: 180°–45°W (sector 1) for z500 and 125°–85°W for precipitation (sector 2, also referred to as the ''western United States''). The extended longitudinal sector used for z500 is motivated by its larger-scale patterns in comparison to precipitation, whereas the narrower sector used for precipitation is chosen because the western United States is both an important region for socioeconomic reasons and a region where MAEs remain substantially reduced over the northern midlatitudes. We note that the main conclusions are qualitative similar when using APC within 35° and 55°N with all longitudes included. As seen with MAE, week 2–4 NH APC tends to improve when tropical nudging is applied, which is illustrated in Fig. 6 for z500 and precipitation. This improvement is also found for anomaly patterns in upper-level winds, temperature and geopotential, as well as for precipitable water and specific humidity at 850 hPa (not shown). The difference between nudging all variables or winds only is negligible (blue and green bars), whereas the narrower tropical nudging region is associated with a smaller APC increase (red bars). The vertical black lines demonstrate that, based on the random resampling method described earlier, the changes in z500 week 2–4 APC between CNT and tropical nudged cases tend to be significant, and that is also the case for precipitation, with generally larger relative increases at longer lead times. Week 2 precipitation APC is an exception, where the slight APC increases in the nudged cases are not significant.

Histograms of weekly APCs (Fig. 7) show that tropical nudging is associated with changes in remote skill that extend well beyond the positive shift in the median shown in Fig. 6. Specifically, APC histograms are more left-skewed, with higher APC values more frequent when comparing tropical nudged to the CNT reforecasts. The increase/decrease in percentage of reforecasts with higher/lower skill is particularly noticeable during weeks 3 and 4 for WTR and WTRuv. Comparison of MAE histograms are analogous in the sense that tropical nudged cases are more right-skewed than in CNT (not shown). This result suggests that reducing tropical forecast errors might lead to an increase in the probability of having a skillful subseasonal forecast over sectors 1 and 2. These shifts are also seen when APC is calculated over all longitudes, although they are less pronounced in global precipitation.

The APC scatterplots shown in Fig. 8 demonstrate that tropical nudging does not always lead to improved remote predictive skill. The reforecasts marked by symbols lying below the black diagonal line are cases where APC is deteriorated by the nudging, and vice versa for reforecasts marked by symbols lying above the black diagonal. The distance between dashed lines corresponds to one standard deviation of the CNT APC,

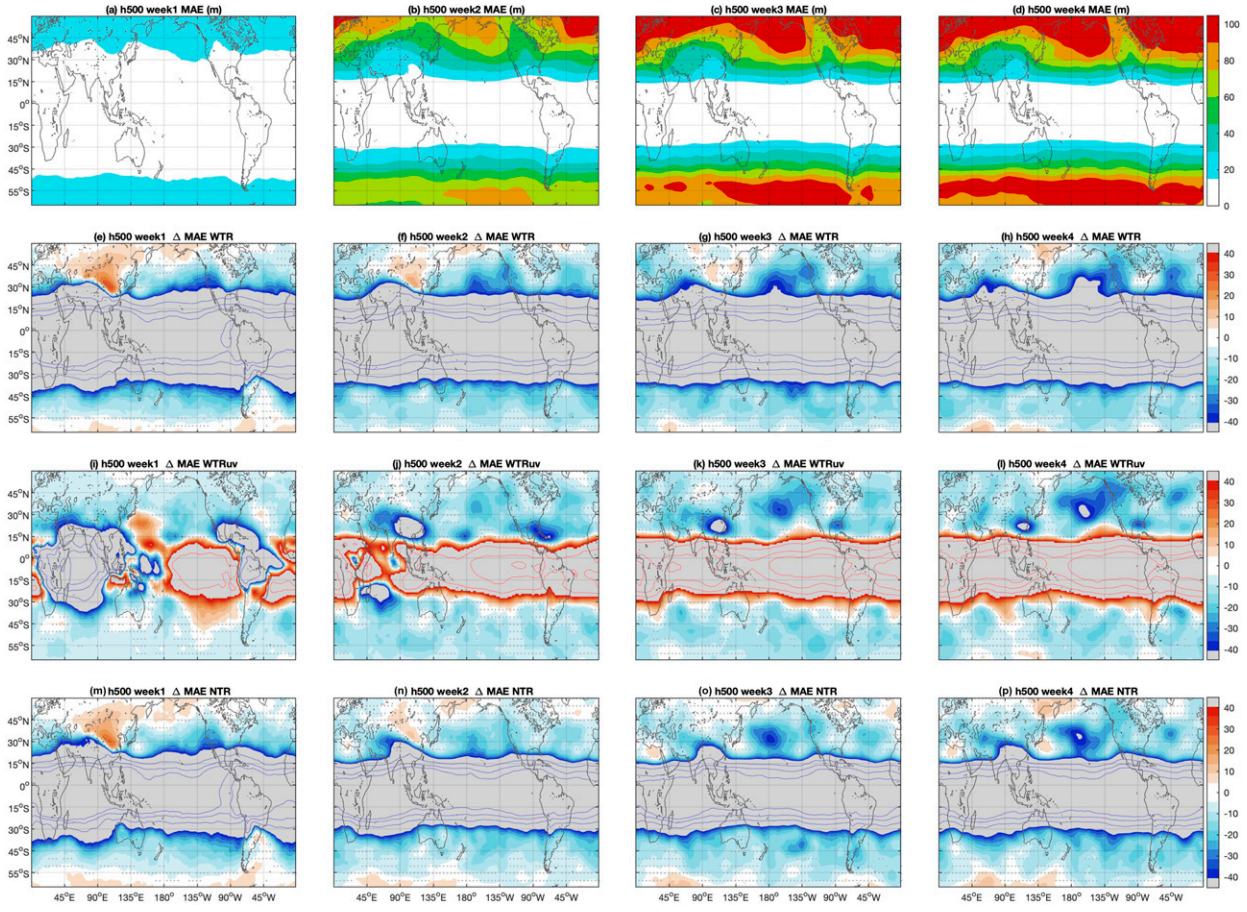

FIG. 4. (a)–(d) The z500 CNT MAE (m) for weeks 1–4, respectively. (e)–(p) As in (a)–(d), but for the MAE percentual difference (ΔMAE defined in the text) for WTR in (e)–(h), WTRuv in (i)–(l), and NTR in (m)–(p). Blue (red) shading denotes regions where errors are reduced (increased) in comparison to CNT. Regions where the amplitude of ΔMAE is larger than 45% are shaded gray and contoured with lines starting at 100% at 100% intervals. The gray symbols correspond to regions where the ΔMAE is not significantly different than zero.

and this is used as a rough estimate of the relative change in skill due to nudging in comparison to the expected spread in the free running forecasts. The darker symbols correspond to the median WTR/WTRuv/NTR APC within bins of the CNT APC showing that, more often than not, tropical nudging is associated with an increase in z500 APC over sector 1 and precipitation over sector 2, but there is certainly considerable spread in how much tropical nudging changes in comparison to CNT. A similar conclusion applies to APC extending over all longitudes and for other variables, and also for similar MAE scatterplots. An analogous depiction of week 1 shows symbols lying close to the diagonal suggesting that week 1 APC is not systematically shifted by tropical error reduction, a result in line with previous studies regarding the lag in the response to tropical forcing (Branstator 2014; Dias and Kiladis 2019). It is interesting that a CNT reforecast that has poor NH skill is more likely to be improved when tropical errors are reduced than a case where CNT APC is relatively good. The fact that the median nudged APC tends to lie close to the diagonal line when the CNT APC is larger implies that tropical nudging has less of a systematic positive impact on reforecasts that are already skillful. By week 3–4 there is some indication that, particularly for NTR, high CNT APC are lowered when applying tropical nudging; however, sample sizes are small at those ranges of CNT APC (see histograms in Fig. 7), therefore the significance of those median values is less clear.

The analogous APC scatter when all longitudes are included is qualitatively similar, and so is the SH APC scatter, except that median APC changes there tend to be less, with larger spread. The fact that WTR and WTRuv median APC's are similar, with NTR suggesting only slightly smaller shifts, implies that the changes in zonal mean state seen in Fig. 2 do not play a major role on how each tropical nudging formulation impacts remote predictive skill. While week 4 APCs are an exception, where there is more separation among WTR/WTRuv/NTR APCs, those results are harder to interpret because the version of the GFS here is not specifically tuned for its performance beyond week 2, as shown by the very low CNT APC. Similarly, because GFS weeks 3 and 4 precipitation skill is on average low, it is not clear that the potential increases

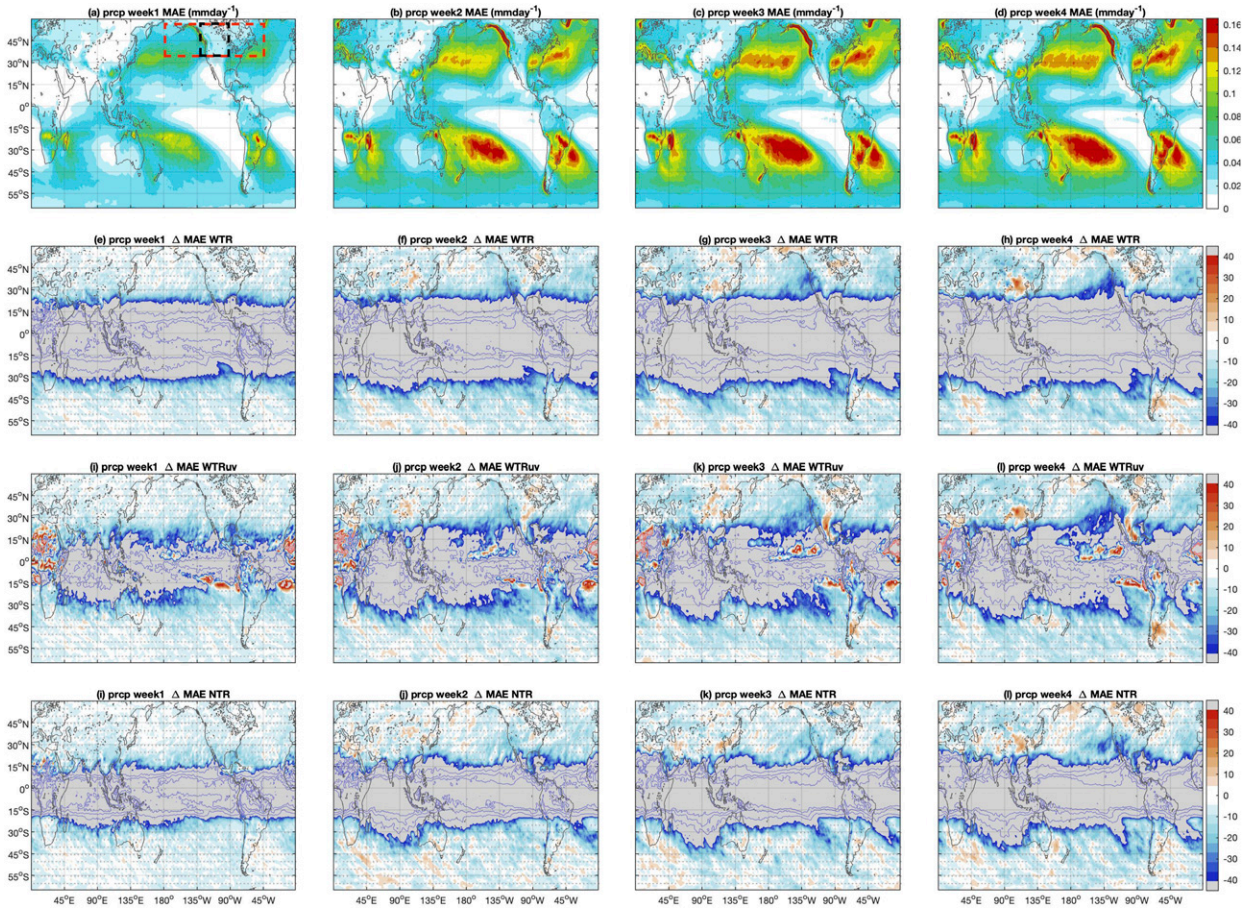

FIG. 5. As in Fig. 4, but for precipitation rate. The red box in (a) corresponds to sector 1 and the black box to sector 2, as defined in section 3c.

associated with the tropics would lead to useful skill, or more reliable predictions at those lead times.

d. *The lagged response between tropical and extratropical errors*

Aside from the lack of realism of a global prediction system where tropical predictions are nearly ''perfect'' throughout the forecast cycle, another limitation of tropical nudging is that remote errors could be reduced due to a ''forced boundary effect'' rather than because of an improved representation of the tropical source of midlatitude weather (e.g., the Rossby wave source). One indication that the former might be important is that the nudged median APCs tend to be constant when CNT APCs are below zero (Fig. 8). Said differently, because the northern edge of the nudging region carries information from both tropics and extratropics, the improved skill in cases that are otherwise not skillful might simply be a consequence of the fact that we are effectively weakly forcing the model with ''realistic'' information (reanalysis) at the latitudinal boundary between nudged and free forecasts, analogous to predictions based on regional models. Counter to this argument, the NTR setup indicates that tropical forecast errors might be the primary source of the change on weeks 2 and 3 because the NTR ''boundary'' is farther from the subtropics and we still observe an increase in remote APC that is comparable to WTR and WTRuv.

To test the role of concurrent nudging at the boundary as a potential driver of the improved remote skill, we generated two additional sets of reforecasts that are analogous to WTR, except that tropical nudging is switched off at day 7 (WTRwk1) and day 14 (WTRwk2). Upper-level zonal and meridional wind APCs within the tropics (Figs. 9a,b) illustrate the approach where, by design, correlation coefficients are nearly one throughout the forecast cycle in WTR as well as for WTRwk1 out to day 7 and WTRwk2 out to day 14. Comparison of the evolution of median z500 and precipitation APC over the NH extratropics in these cases points to a lag in the remote skill response to tropical nudging of about 4–5 days (Figs. 9c,d). Specifically, WTR, WTRwk1 and WTRwk2 yield median APC values that are not significantly different from CNT out to day 4–5 (gray lines), and from then on WTRwk1 and WTRwk2 APC are statistically lower than WTR's about 4–5 days after tropical nudging is switched off. The precise number of days is dependent on the variable and also the region where the APC is calculated, but tends to fall in the synoptic range of 3–8 days (not shown), again, consistent with

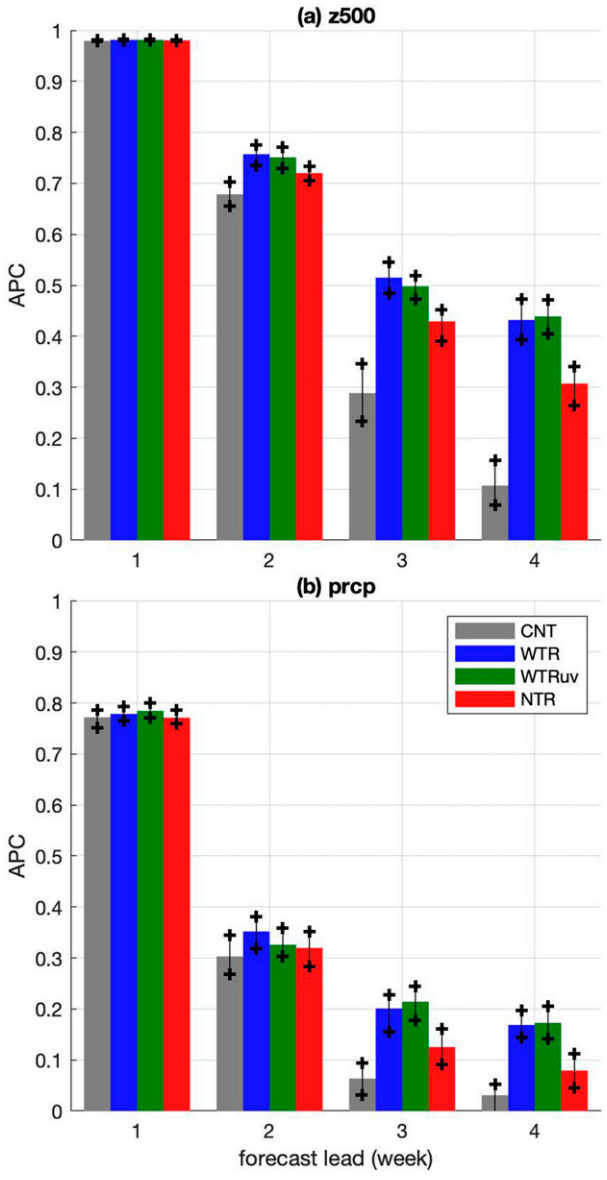

FIG. 6. (a) Sector 1 median APC for 500-hPa geopotential heights and (b) sector 2 APC for precipitation. Vertical bars correspond to the experiments defined in Table 1: CNT (gray), WTR (blue), WTRuv (green), and NTR (red). The vertical lines centered at the top of each bar display the 95% CI of the corresponding median APC value.

the time Rossby waves forced from the tropics take to reach midlatitudes (Branstator 2014).

The scatterplots in Fig. 10 are similar to the ones in Fig. 8, except that they compare WTR, WTRwk1 and WTRwk2. The nearly overlapping symbols on week 2 for both z500 and precipitation imply that the changes in remote APC at week 2 are primarily related to error reductions prior to that week, suggesting that the concurrent nudging is not playing a major role in the tendency for APC to increase in the WTR runs. A similar conclusion can be drawn from comparing WTR and WTRwk2 symbols for week 3 scatter (middle column). By week 4, there is a clear separation between the median WTR and WTRw1/WTRwk2 APCs; however, particularly for z500, WTRwk1 and WTRwk2 still show an improvement in comparison to CNT and when CNT APC values are less than zero. That could be because week 2–3 tropical circulation in WTRwk1 and WTRwk2 are improved in comparison to CNT (Figs. 9a,b) even though tropical nudging is applied only out to weeks 1 and 2, respectively.

The comparison between WTR, WTRwk1 and WTRwk2 strongly suggests that the concurrent effect of the boundary forcing plays only a minor role in the changes in remote skill. We note that a similar conclusion applies to potential concerns regarding errors introduced due to wave distortion at the nudging boundary. Overall, these results provide evidence of a synoptic lag between error reductions in the tropics and extratropics that is consistent with the Rossby wave mechanism underlying the tropical–extratropical prediction error relationship. One implication is that a model development that leads to improved week 1 tropical predictions is beneficial for week 2 NH cool season predictions, as well as for the subsequent weeks.

e. *Tropical to extratropical conditional skill*

As mentioned in the introduction, the issue of practical subseasonal predictability originating from the tropics is difficult to address with nudging experiments as applied here because all tropical predictions are kept skillful, regardless of any intrinsic predictability barriers. The question of how much remote skill can be gained from the tropics in practice is partially addressed in Dias and Kiladis (2019) where a conditional analysis was applied to subseasonal reforecasts to evaluate the relationship between tropical and extratropical predictive skill. Their results showed how reforecasts that perform well (poorly) in the short range over the tropics tend to perform better (worse) than average in the medium to extended range over the NH. This relationship is reproduced here for the GFS CNT following a similar approach as in Dias and Kiladis (2019). We first split reforecasts in two subsets: one where day 2–5 CNT APC from tropical (20°S–20°N) divergence at 200 hPa (D200) is above its upper tercile (''good tropical forecast''), and a second one where the same APC is below its lower tercile (''poor tropical forecast''). We then calculate the median NH (35°–55°N) CNT precipitation APC as a function of lead day in each of these subsets. The NH APC median values are compared to the median NH APC from random subsamples drawn from the entire reforecast period. The random subsampling is done 1000 times using 1/3 of the total number of reforecasts, which is consistent with the tercile thresholds used to define the cases where the tropical performance is good versus poor. The NH precipitation APC normalized difference D between these sets of reforecasts is defined as

$$D = 100 \times \frac{\overline{\mathrm{APC}}_{\mathrm{cond}} - \langle \overline{\mathrm{APC}}_{\mathrm{rnd}} \rangle}{\langle \overline{\mathrm{APC}}_{\mathrm{rnd}} \rangle}, \tag{2}$$

where $\overline{\mathrm{APC}}_{\mathrm{cond}}$ is the conditional (good or poor tropical performance) median NH precipitation APC and $\langle \overline{\mathrm{APC}}_{\mathrm{rnd}} \rangle$ is similar, except that we averaged the median APC over the 1000 random subsamples. The symbols in Fig. 11 highlight lead

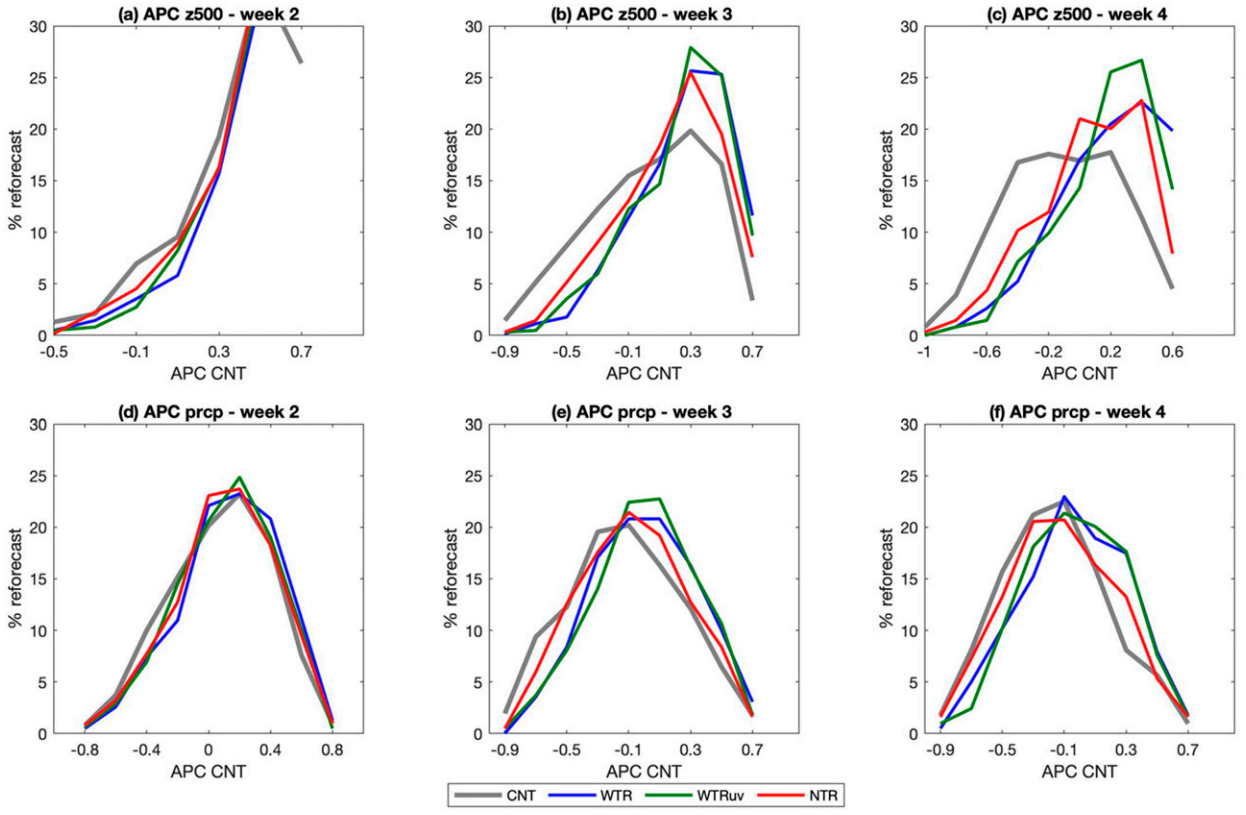

FIG. 7. (a)–(c) Histograms of sector 1 weekly 500-hPa geopotential height and (d)–(f) sector 2 precipitation APCs, using the entire reforecast period. The ticked gray lines correspond to CNT, and thinner colored lines correspond to tropical nudging APC, where the color convention is the same as in previous figures (blue, green, and red correspond to WTR, WTRuv, and NTR, respectively). Bin size is 0.2 with tick marks placed at every other bin center.

times where the sign of D is statistically significant. The sign significance is defined as lead times where 90% of the normalized differences using each $\overline{\text{APC}}_{\text{rnd}}$, as opposed to the average, agree in sign. While the significance test could be done in many different ways, the method here illustrates how even large values of D are not necessarily significant because at longer lead times the denominator in Eq. (2) becomes very small.

The NH weeks 2–3 positive/negative D values (blue/red bars) displayed in Fig. 11a imply that when the week 1 upper-level tropical divergent flow prediction is better (worse) than the median, then subsequent GFS precipitation predictions tend to perform above (below) average in the northern midlatitudes with a lag of about one week. This result is consistent with what was found for the two prediction systems studied in Dias and Kiladis (2019). However, as it was the case in Dias and Kiladis (2019), it is unclear how much of this relationship can be attributed to model errors versus an initial state dependence. The tropical nudging reforecasts allow us to investigate this issue, by comparing the conditional skill of the nudged realizations of the ''good'' versus ''poor'' CNT reforecasts. That is, if the improved or degraded precipitation skill seen at a lead-lag in Fig. 11a were solely a consequence of the initial conditions, then the corresponding WTR reforecasts should show the same behavior on average as their CNT counterparts. Conversely, if the changes in conditional skill at a lead–lag were instead due solely to model errors in the tropics, then the WTR conditional APC should show no statistical difference between the good/poor and random subsets because the WTR realizations are ''perfect'' in the tropics, by design. Figure 11c suggests a mix of these two possible scenarios, while emphasizing the likely role of model errors. The amplitudes of the normalized differences in the nudged realizations are generally smaller and less significant, indicating that model errors are indeed important in the linkage between tropical and extratropical skill. However, the signs of the D values are systematic and consistent with CNT, which suggests some influence of the initial state. This behavior is similar to what is seen in the WTRwk1 case shown in Fig. 11e, further emphasizing the lagged linkage between tropical and extratropical predictions. When using upper-level divergence as our metric of tropical performance, WTRuv, WTRwk2 and NTR yield similar conclusions (not shown).

The conditional analysis based on upper-level divergence indicates that model advancements resulting in improved week 1 predictions of tropical upper-level divergence circulation would be beneficial for week 2 NH precipitation predictions. Interestingly, this same statement does not appear to hold true when considering tropical predictions of lower-level divergence

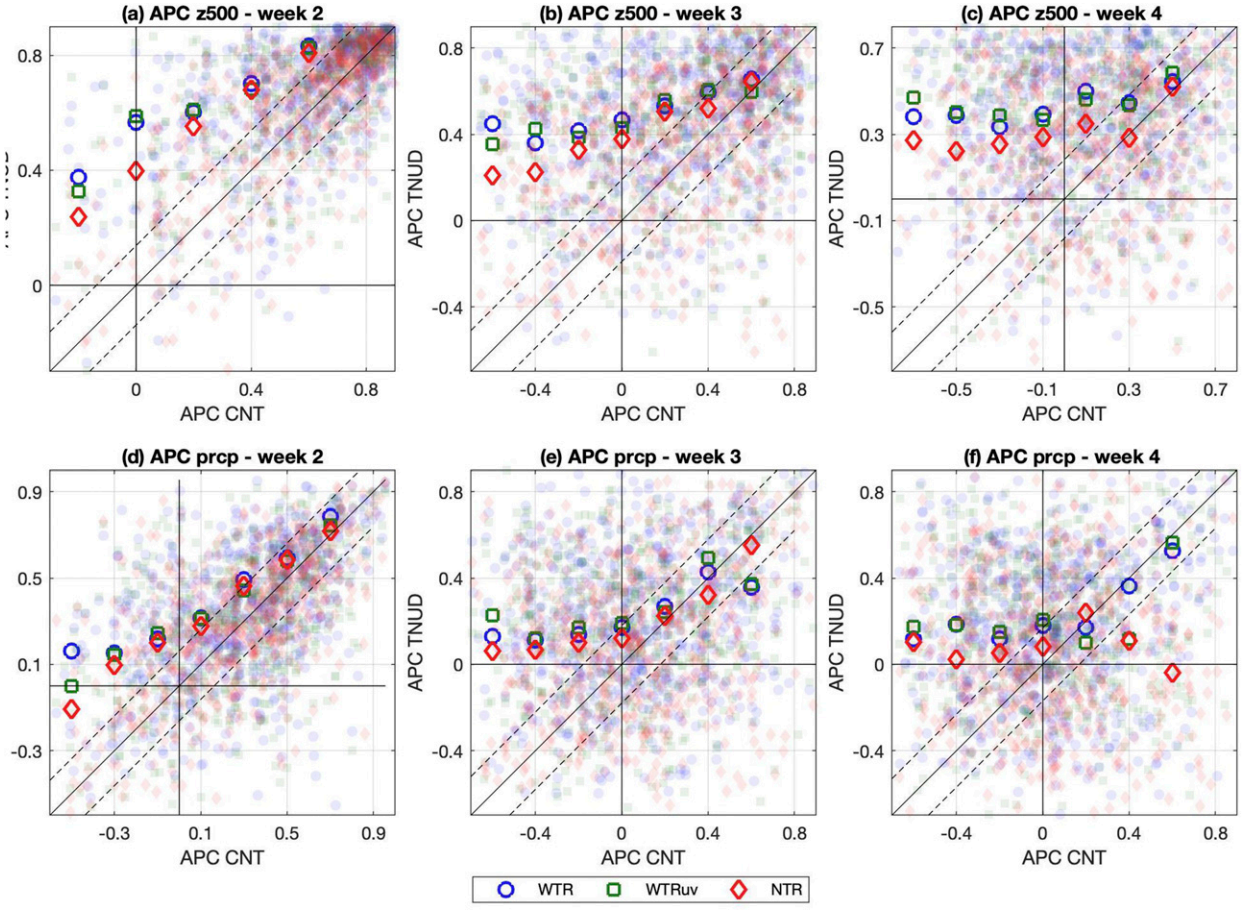

FIG. 8. (a)–(f) Light colored symbols denote the scatter of weekly western U.S. APC between CNT APC (x axis) and tropical nudging APC (y axis), where each light colored symbol denotes one reforecast initialized during the analyzed period. The dark colored symbols depict the median tropical nudged APC within bins of CNT APC. The bin width is 0.2, and bins are centered in 0.1 intervals from −1 to 1. The color convention is the same as in previous figures (blue, green, and red correspond to WTR, WTRuv, and NTR, respectively). Each panel is zoomed to tightly display all 620 reforecasts. Symbols lying on the slanted line indicate reforecasts where the APC is unchanged by tropical nudging. Vertical and horizontal lines highlight zero APC. The distance between the dashed lines corresponds to one CNT APC standard deviation for the corresponding forecast lead and variable.

(not shown) or upper-level vorticity. The right column in Fig. 11 shows that using 200-hPa vorticity as the metric for "good" versus "poor" tropical reforecasts yields no significant separation in remote performance at lag for any of the cases considered. These results are found to be unchanged when expanding the tropical APC area from 20°S–20°N to 30°S–30°N, or when separating the reforecasts based on the APC in NH subtropical bands (not shown). Because tropical–extratropical teleconnections originate with tropical precipitation systems that are the primary drivers of upper-level divergent flows in the tropics and subtropics, these results point to the well-known need for improving the representation of moisture–convection–circulation coupling in the tropics (Wolding et al. 2020a,b).

4. MJO dependence of remote skill

While the role of tropical synoptic waves as deterministic sources of tropical predictability are more uncertain, there are a number of studies that point to the MJO as a practical source of tropical and extratropical subseasonal predictability (Schreck et al. 2020 and references therein). Tropical nudging experiments can also be useful in revealing some aspects of the role of the MJO in tropical–extratropical error interdependencies because, in contrast to CNT, in WTR the MJO is by design well represented throughout its life cycle. To illustrate, here we apply a similar conditional skill analysis as shown in the previous section, except that we split reforecasts depending on the MJO amplitude at initial reforecast time. To focus on the MJO convective signal, the metric used for MJO amplitude is the OLR MJO index (OMI) (Kiladis et al. 2014), which was obtained from https://www.psl.noaa.gov/mjo/mjoindex/. Active/neutral/inactive MJO subsets are defined depending on OMI terciles for the period of November–March and 1999–2018. One question that this conditional analysis allows us to investigate is whether the effectiveness of tropical nudging in increasing NH precipitation APC depends on the initial state

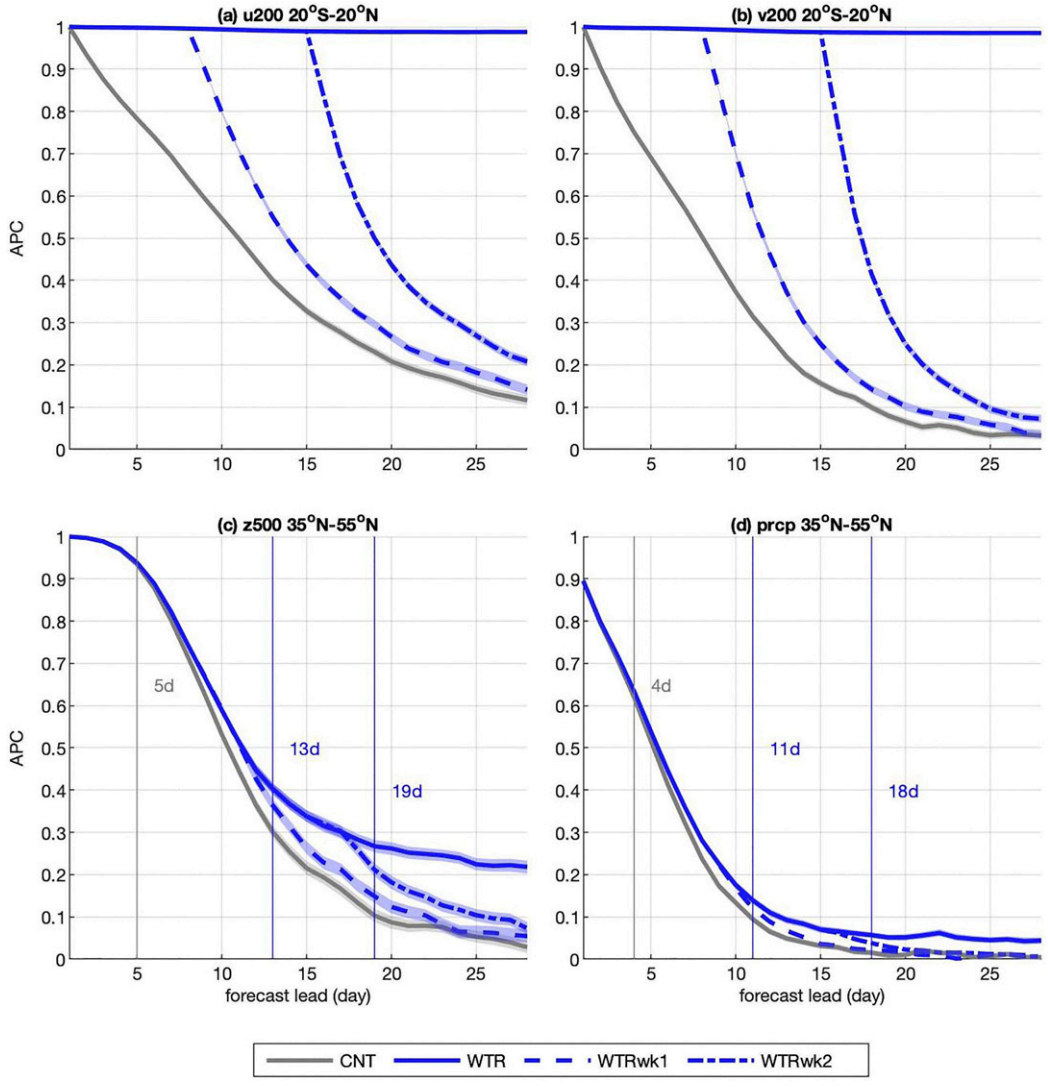

FIG. 9. (a) Tropical zonal and (b) meridional 200-hPa wind APCs. (c),(d) As in (a) and (b), but for midlatitude z500 and precipitation. The latitudinal bands are displayed in the panel titles. The gray lines correspond to CNT median APC, and the blue lines correspond to WTR (solid), WTRwk1 (dashed), and WTRwk2 (dash–dotted). The gray vertical lines highlight the earliest lead time where the tropical nudged APC is larger than CNT. From left to right, the blue vertical lines mark the earliest lead time where WTR APC is statistically larger then WTRwk1 and WTRwk2, respectively.

of the MJO. Figure 12a shows that, regardless of tropical nudging, the median weeks 2–3 NH precipitation CNT APCs tend to increase with initial MJO amplitude. Figure 12b shows that the absolute difference between WTR and CNT APC is positive and increases with lead time, independently of MJO amplitude. The only statistically significant difference is seen at week 4, where WTR shows a larger enhancement in APC with increasing MJO amplitude at initial time. Figure 12c indicates that most of the improvements in APC comparing WTR to CNT and any potential MJO modulation at week 4 largely rely on the tropical nudging being applied throughout the entire forecast cycle.

Figure 13a is similar to Fig. 11a, except that the vertical bars correspond to normalized differences of APC (D) using MJO active (blue) and inactive (red) as defined above. The increase/decrease in conditional skill with MJO activity is consistent with Fig. 12a because, aside from the normalization, the only difference is that we calculated APC on weekly averaged fields as opposed to daily. Note that when the MJO is inactive at initial time, week 2–3 NH precipitation APC is substantially reduced in comparison to random initializations. Comparing CNT (Fig. 11a) to WTR (Fig. 11b) shows that when the MJO is active and the tropics are nudged, remote skill is improved by a comparable amount on week 2 as in CNT, but it is also improved out to week 4. We note that APC here is based on all longitudes from 35° to 55°N, but when looking at different sectors and MJO phases, the lead time and amplitude of

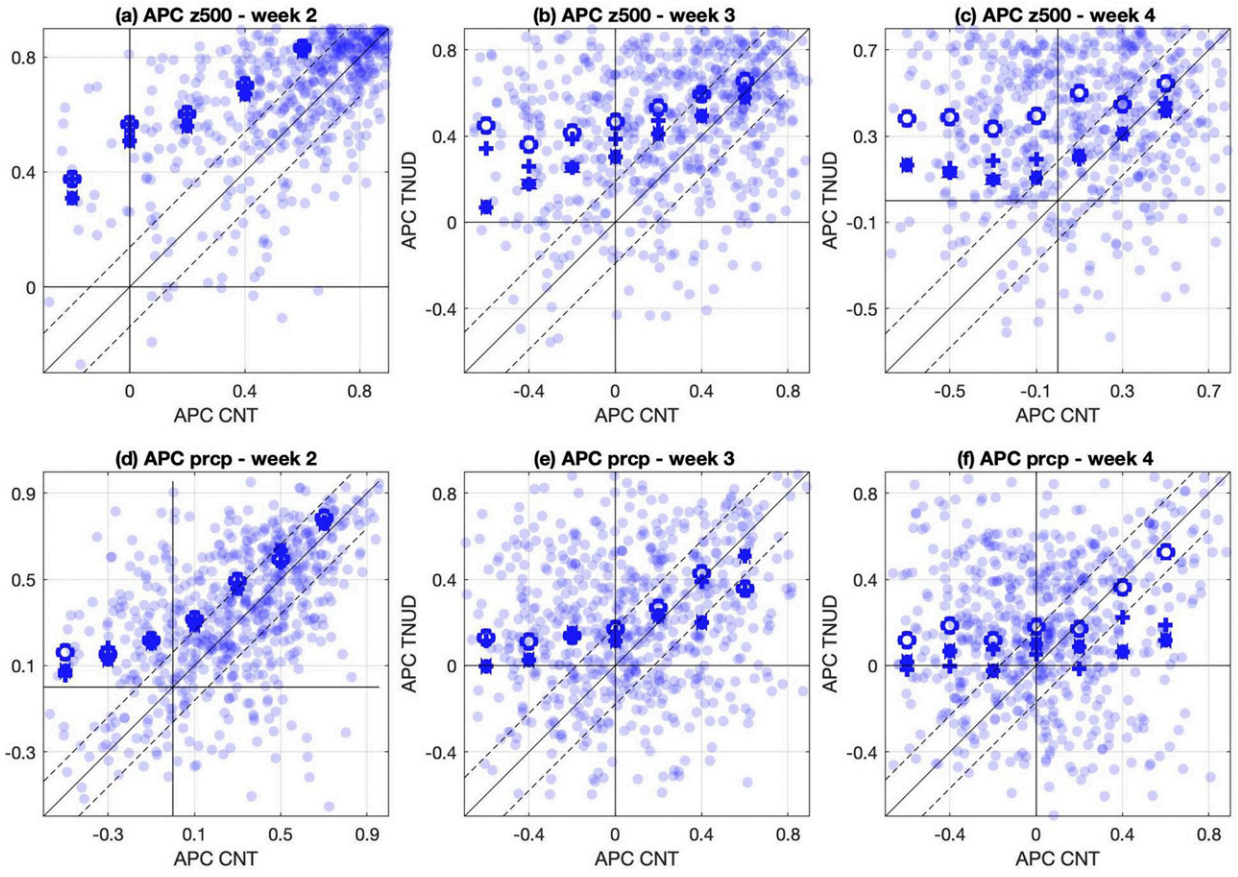

FIG. 10. As in Fig. 8, but the blue symbols correspond to WTR (circles), WTRwk1 (asterisks), and WTRwk2 (crosses).

those modulations varies (not shown). When the MJO is inactive and the tropics are nudged weeks 2–3 remote skill is not nearly as reduced as in CNT. Figure 11c suggests that nudging only week 1 in the tropics does not affect MJO related changes in subsequent lead times in comparison to CNT. Perhaps because the MJO in the GFS has a reasonable amount of skill in week 1, realizing remote improvements in week 2 suggested by Figs. 11c and 11e might rely on other sources of week 1 tropical predictability apart from the MJO, for example, the other convectively coupled equatorial waves. In contrast, at longer lead times (week 3–4) it is possible that an improved MJO would lead to improvements in remote skill. A thorough analysis of the role of the MJO in tropical to extratropical skill linkages, including seasonality, and MJO phase dependences would likely be very insightful, but is beyond the scope of the present manuscript.

Regarding El Niño–Southern Oscillation (ENSO), we first note that most reforecasts in the analyzed period correspond to weak to moderate La Niña conditions, where reforecasts initialized during strong El Niño are primarily coming from the 2015/16 season. This sampling distribution bias makes it difficult to untangle any potential skill modulation related to ENSO phases and, therefore, a conditional analysis similar to the MJO did not produce robust or conclusive results. This issue also merits further investigation because, particularly at the longer lead times, ENSO is certainly expected to play an important role on how tropical forecast errors propagate to higher latitudes.

5. Summary and conclusions

The tropical nudging experiments presented here demonstrate that GFS subseasonal NH predictions during the winter are improved when tropical forecast errors are reduced, with results broadly consistent with previous similar experiments applied to other NWP systems. A number of sensitivity tests are performed supporting the interpretation that much of the tropical to extratropical subseasonal forecast error is rooted to errors in the tropical atmospheric circulation, which is, in turn, known to be strongly coupled to moist convection. In addition to reductions in remote dynamical field errors, both MAE and APC analysis indicate the potential for week 2–4 *precipitation predictions* to be improved by reducing forecast errors in the tropics. Specifically, by fully nudging the tropics to reanalysis from 10°S to 10°N with no nudging beyond 30°S/30°N, weeks 2–4 precipitation predictions are improved by as much as 40% in some portions of the western United States. This remote skill improvement is only about 15%–20% weaker when fully nudging only 5°S–5°N, with no nudging beyond 20°S/20°N, once

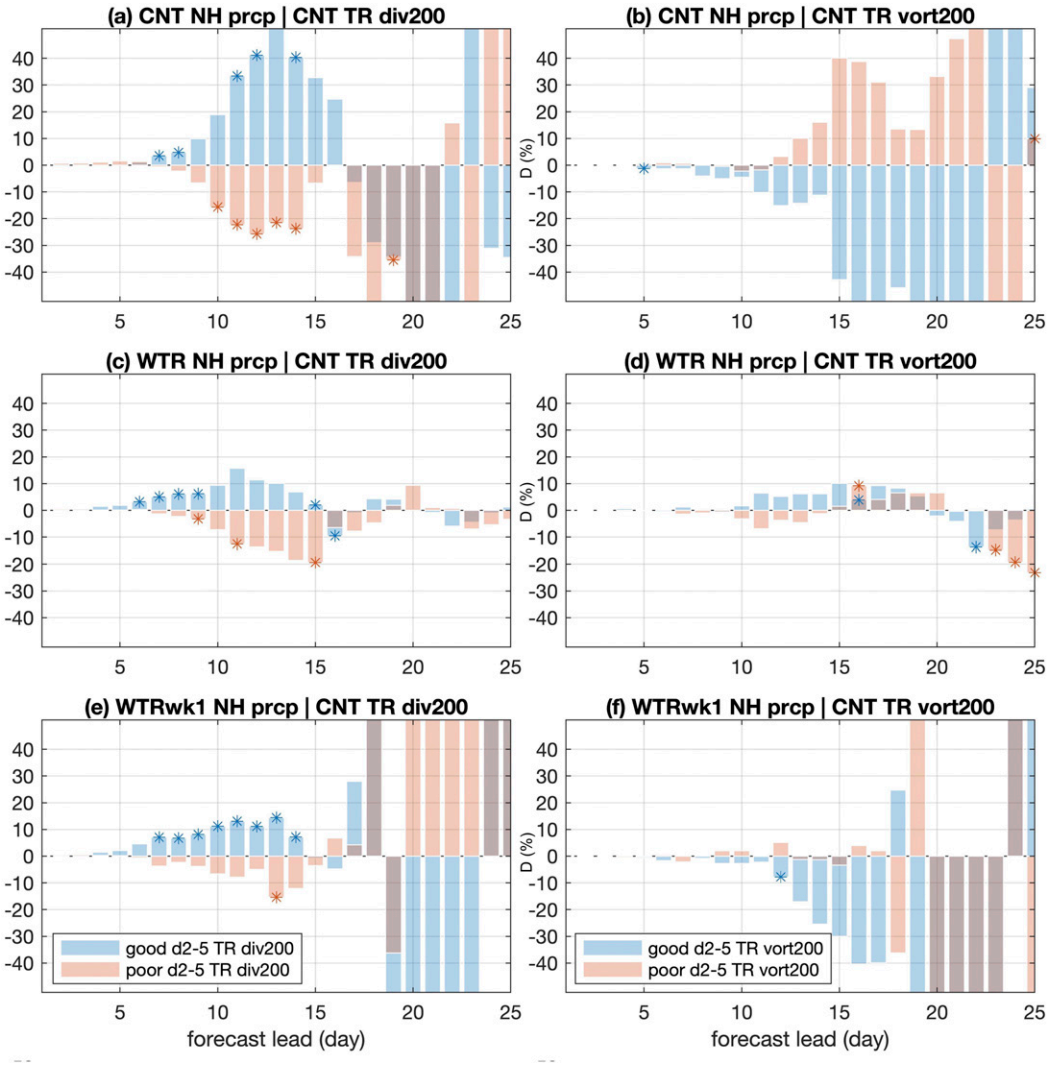

FIG. 11. Display of conditional normalized APC difference [D, defined in Eq. (2)] as a function of forecast lead in days, for (a),(b) CNT; (c),(d) WTR; and (e),(f) WTRwk1. The difference D when the conditional partitioning between "good" and "poor" tropical performance is based on the CNT APC from 200-hPa (left) tropical divergence and (right) vorticity. Symbols are plotted when the sign of D is statistically significant. Brown bars denote lead times where red and blue bars are overlaid.

again, highlighting the role of the deep tropics in driving extratropical circulation.

Interestingly, tropical nudging is more effective at improving predictions of the remote response is cases when the free reforecasts (CNT) are not skillful, whereas there is less of a systematic impact on reforecasts that are already skillful. This behavior is seen even when nudging is switched off at the end of week 1 or 2 (WTRwk1/WTRwk2), again indicating that the remote skill response does indeed originate from within the tropics. One interpretation of the differences in how effective tropical nudging is depending on the skill of the NH CNT prediction is that current GFS week 2–4 skillful forecasts do not strongly draw skill from the tropics, and, therefore, week 2–4 NH predictions would be skillful more often if tropical forecasts errors were reduced. This comes with the caution that the extent to which the decrease in tropical errors and subsequent gain in remote skill is achievable remains an open question, which could be further investigated, for example, by linking these modulations to known subseasonal tropical–extratropical teleconnection patterns (Stan et al. 2017).

Toward the practical goal of better understanding how much remote subseasonal skill can be gained from the tropics, we presented a conditional skill analysis. This analysis in conjunction with the week 1 or week 1–2 only nudging reforecasts, suggests that:

1) During NH winter, there is a lag of about one week between tropical forecast error growth and NH precipitation prediction skill; therefore, a model development that is beneficial for a given forecast range in the tropics will likely have a positive imprint on the following week's NH predictions.

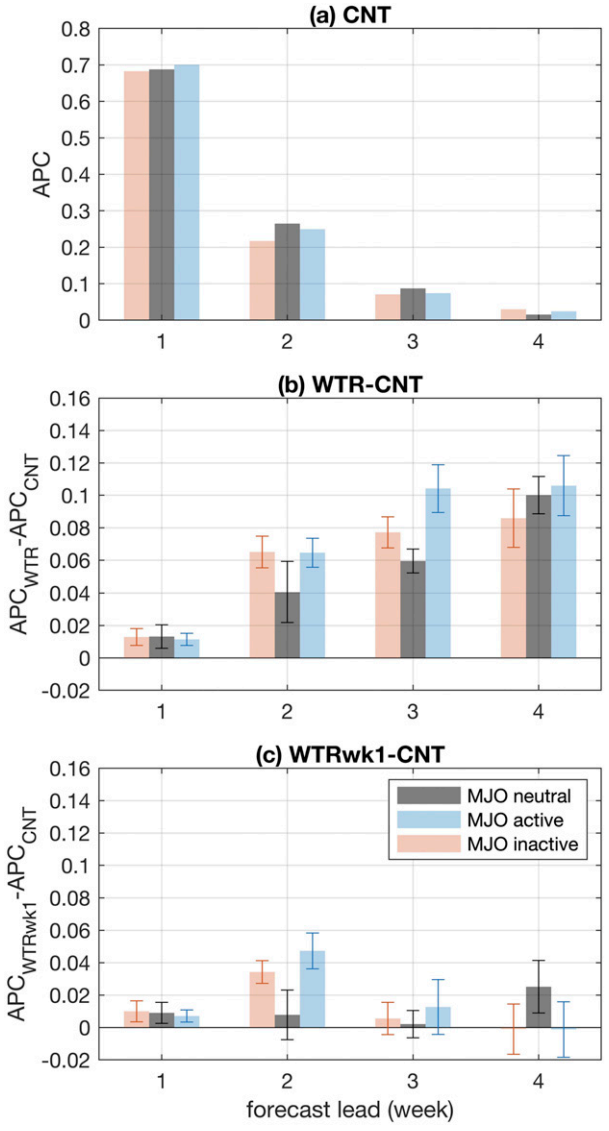

FIG. 12. (a) The weekly precipitation APC averaged over all longitudes and 35°–55°N split according to MJO amplitude at initialization (details in the text). (b) As in (a), but that the difference between WTR and CNT APC is displayed. (c) As in (b), but for WTRwk1. The CIs displayed in (b) and (c) (vertical bars) are calculated using the same resampling method described in the text and denote that 90% of the median APC differences between MJO conditional and random subsamples fall within that range.

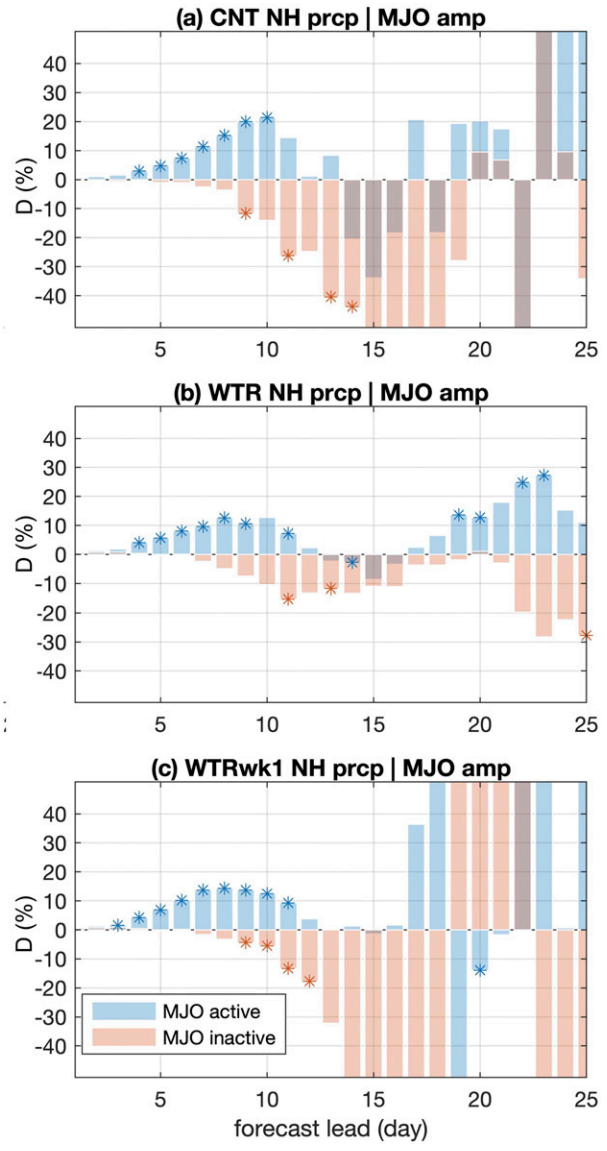

FIG. 13. As in Figs. 11a,c,e, but reforecasts subsets are based on MJO amplitude at initialization where blue (red) bars correspond to normalized APC difference [D, defined in Eq. (2)] for cases where the MJO is active (inactive) at initialization.

2) Tropical forecast errors in upper-level divergent circulations are of primary importance in the modulation of remote errors, and based on previous studies those errors are likely to be rooted in model deficiencies related to moisture–convection–circulation coupling.

The conditional analysis does not directly address the issue of practical predictability that originates from the tropics because, in nudging experiments, all tropical forecasts are made equally skillful. That being said, studies such as Wang et al. (2019); Bengtsson et al. (2019, 2021) have shown that advancements in the GFS model physics, and particularly those related to convective parameterizations, lead to improved week 1 tropical skill, which, in turn, tropical nudging suggests should lead to improved NH week 2 predictions. Because synoptic-scale convective variability in the tropics (e.g., Kelvin and easterly waves) is intermittent and localized, the linkage between tropical convection organization, beyond the MJO, to the potential boosts in remote skill estimated from the tropical nudged reforecasts might not be entirely realizable. This is a crucial issue that could be investigated by filtering and or introducing noise to the fields we nudge to. Because different GFS physics packages are currently being developed and tested, remote changes in predictions originating from the

tropics could also be tested by nudging the tropics to forecast target fields from different GFS versions. For example, if a particular model configuration performs better in the tropics, tropical nudging could be applied to the operational configuration using the better performing forecasts as the fields that are nudged to. A similar analysis as presented here would then allow an estimate of how much remote skill is gained from the tropics in isolation. We are currently working on the design and implementation of some of these alternative types of nudging experiments, and those results will be reported in the future.

In summary, as is the case in previous tropical nudging studies, the present work offers potential upper bounds on how much midlatitude subseasonal remote skill might be tapped from low latitudes. In addition, we hope to have demonstrated that tropical nudging is a productive diagnostic tool that can be beneficial to better characterize sources of tropical model errors, and which modes of tropical variability are important for NH subseasonal predictions. This diagnostic aspect of tropical nudging is particularly useful in the current stage of the UFS development (see https://ufscommunity.org/) because it might guide what types of model developments in the tropics are the most effective in improving remote subseasonal predictions.

Acknowledgments. We thank NCEP's Climate Prediction Center for providing HPC support for this work, Jeff Whitaker for suggesting the use of IAU method for nudging, and Phil Pegion for the helpful internal review. We are also grateful for the two anonymous reviewers' suggestions and comments that greatly improved our manuscript. This research was partially supported by the Physical Science Laboratory of NOAA's Earth System Research Laboratory, the California Department for Water Resources (CDWR) and NOAA WPO Grants NA19OAR4590151 and PGC02.FY20.1.

Data availability statement. The numerical model simulations upon which this study is based are too large (~ 70TB) to archive or to transfer. All data are stored on NOAA High Performance Storage System and subsets of it will be made more widely available on request.

REFERENCES

Bengtsson, L., and Coauthors, 2019: Convectively coupled equatorial wave simulations using the ECMWF IFS and the NOAA GFS cumulus convection schemes in the NOAA GFS model. *Mon. Wea. Rev.*, **147**, 4005–4025, https://doi.org/10.1175/MWR-D-19-0195.1.

——, J. Dias, S. Tulich, M. Gehne, and J.-W. Bao, 2021: A stochastic parameterization of organized tropical convection using cellular automata for global forecasts in NOAA's Unified Forecast System (UFS). *J. Adv. Model. Earth Syst.*, **13**, e2020MS002260, https://doi.org/10.1029/2020MS002260.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271, https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>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.

Chikira, M., and M. Sugiyama, 2010: A cumulus parameterization with state-dependent entrainment rate. Part I: Description and sensitivity to temperature and humidity profiles. *J. Atmos. Sci.*, **67**, 2171–2193, https://doi.org/10.1175/2010JAS3316.1.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Dias, J., and G. N. Kiladis, 2019: The influence of tropical forecast errors on higher latitude predictions. *Geophys. Res. Lett.*, **46**, 4450–4459, https://doi.org/10.1029/2019GL082812.

——, M. Gehne, G. N. Kiladis, N. Sakaeda, P. Bechtold, and T. Haiden, 2018: Equatorial waves and the skill of NCEP and ECMWF numerical weather prediction systems. *Mon. Wea. Rev.*, **146**, 1763–1784, https://doi.org/10.1175/MWR-D-17-0362.1.

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.

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.

Greatbatch, R. J., G. Gollan, T. Jung, and T. Kunz, 2015: Tropical origin of the severe European winter of 1962/1963. *Quart. J. Roy. Meteor. Soc.*, **141**, 153–165, https://doi.org/10.1002/qj.2346.

Hendon, H. H., B. Liebmann, M. Newman, J. D. Glick, and J. E. Schemm, 2000: Medium-range forecast errors associated with active episodes of the Madden–Julian oscillation. *Mon. Wea. Rev.*, **128**, 69–86, https://doi.org/10.1175/1520-0493(2000)128<0069:MRFEAW>2.0.CO;2.

Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2.

——, 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.

Judt, F., 2019: Atmospheric predictability of the tropics, middle latitudes, and polar regions explored through global storm-resolving simulations. *J. Atmos. Sci.*, **77**, 257–276, https://doi.org/10.1175/JAS-D-19-0116.1.

Jung, T., 2011: Diagnosing remote origins of forecast error: Relaxation versus 4D-var data-assimilation experiments. *Quart. J. Roy. Meteor. Soc.*, **137**, 598–606, https://doi.org/10.1002/qj.781.

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

——, T. N. Palmer, M. J. Rodwell, and S. Serrar, 2010b: Understanding the anomalously cold European winter of 2005/06 using relaxation experiments. *Mon. Wea. Rev.*, **138**, 3157–3174, https://doi.org/10.1175/2010MWR3258.1.

Kiladis, G. N., J. Dias, K. H. Straub, M. C. Wheeler, S. N. Tulich, K. Kikuchi, K. M. Weickmann, and M. J. Ventrice, 2014: A comparison of OLR and circulation-based indices for tracking the MJO. *Mon. Wea. Rev.*, **142**, 1697–1715, https://doi.org/10.1175/MWR-D-13-00301.1.

Li, Y., and S. N. Stechmann, 2020: Predictability of tropical rainfall and waves: Estimates from observational data. *Quart. J. Roy. Meteor. Soc.*, **146**, 1668–1684, https://doi.org/10.1002/qj.3759.

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.

Orbe, C., L. D. Oman, S. E. Strahan, D. W. Waugh, S. Pawson, L. L. Takacs, and A. M. Molod, 2017: Large-scale atmospheric transport in GEOS replay simulations. *J. Adv. Model. Earth Syst.*, **9**, 2545–2560, https://doi.org/10.1002/2017MS001053.

Park, S., 2014: A Unified Convection Scheme (UNICON). Part I: Formulation. *J. Atmos. Sci.*, **71**, 3902–3930, https://doi.org/10.1175/JAS-D-13-0233.1.

Pohl, B., and H. Douville, 2011: Diagnosing GCM errors over West Africa using relaxation experiments. Part I: summer monsoon climatology and interannual variability. *Climate Dyn.*, **37**, 1293–1312, https://doi.org/10.1007/s00382-010-0911-2.

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.

Schreck, C. J., M. A. Janiga, and S. Baxter, 2020: Sources of tropical subseasonal skill in the CFSv2. *Mon. Wea. Rev.*, **148**, 1553–1565, https://doi.org/10.1175/MWR-D-19-0289.1.

Selz, T., and G. C. Craig, 2015: Simulation of upscale error growth with a stochastic convection scheme. *Geophys. Res. Lett.*, **42**, 3056–3062, https://doi.org/10.1002/2015GL063525.

Stan, C., D. M. Straus, J. S. Frederiksen, H. Lin, E. D. Maloney, and C. Schumacher, 2017: Review of tropical-extratropical teleconnections on intraseasonal time scales. *Rev. Geophys.*, **55**, 902–937, https://doi.org/10.1002/2016RG000538.

Takacs, L. L., M. J. Suárez, and R. Todling, 2018: The stability of incremental analysis update. *Mon. Wea. Rev.*, **146**, 3259–3275, https://doi.org/10.1175/MWR-D-18-0117.1.

Wang, J.-W. A., P. D. Sardeshmukh, G. P. Compo, J. S. Whitaker, L. C. Slivinski, C. M. McColl, and P. J. Pegion, 2019: Sensitivities of the NCEP global forecast system. *Mon. Wea. Rev.*, **147**, 1237–1256, https://doi.org/10.1175/MWR-D-18-0239.1.

Wolding, B., J. Dias, G. Kiladis, F. Ahmed, S. W. Powell, E. Maloney, and M. Branson, 2020a: Interactions between moisture and tropical convection. Part I: The coevolution of moisture and convection. *J. Atmos. Sci.*, **77**, 1783–1799, https://doi.org/10.1175/JAS-D-19-0225.1.

——, ——, ——, E. Maloney, and M. Branson, 2020b: Interactions between moisture and tropical convection. Part II: The convective coupling of equatorial waves. *J. Atmos. Sci.*, **77**, 1801–1819, https://doi.org/10.1175/JAS-D-19-0226.1.

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.