Key Points:
- The skill of tropical short range and N.H. forecasts at subseasonal leads are positively correlated
- The tropical-extratropical skill association is modulated by the tropical low frequency state
- The extent to which tropical forecasts errors influence subsequent N.H. predictions is model dependent

Correspondence to:
J. Dias,
juliana.dias@noaa.gov

The Influence of Tropical Forecast Errors on Higher Latitude Predictions

Juliana Dias[1,2] and **George N. Kiladis**[2]

[1]CIRES, University of Colorado, Boulder, CO, USA, [2]NOAA Earth System Research Laboratory, Boulder, CO, USA

Abstract The atmospheric response to variations in tropical latent heating extends well beyond its source region, and therefore it is thought that a reduction of tropical forecast errors should also benefit subsequent forecasts over the extratropics. This relationship is evaluated using a conditional skill analysis applied to subseasonal reforecasts from the National Centers for Environmental Prediction Coupled Forecast System and the European Centre for Medium-Range Weather Forecasts Integrated Forecast System. It is shown that there is enhanced or attenuated skill in Northern Hemisphere Weeks 2–4 forecasts when tropical short range precipitation forecasts are "good" or "poor," respectively. This conditional skill is modulated by both El Niño Southern Oscillation and the Madden and Julian Oscillation, particularly in the Integrated Forecast System. The results presented here indicate that midlatitude Weeks 2–4 predictive skill would benefit from improvements in Week 1 tropical performance, particularly for the National Centers for Environmental Prediction system.

1. Introduction

Tropically forced extratropical teleconnections are of practical interest because they represent a pathway for tropical forecast errors in numerical models to propagate to higher latitudes, where they can affect subsequent midlatitude forecasts. The fact that operational extratropical forecasts are substantially more skillful than those for the tropics is well documented (e.g., Dias et al., 2018; Zhu et al., 2014). Over the last few decades the inference, through observations and models, of a remote impact of tropical forecast errors has motivated prediction centers around the world to improve their tropical performance (e.g., Vitart, 2014; Xiang et al., 2015), where errors from the short (synoptic) to the extended (weekly to monthly) range have been largely attributed to physics related to clouds and precipitation (Goswami et al., 2017; Hirons et al., 2013). These tropical forecast improvements, in principle, hold the promise that midlatitude extended forecasts can draw more skill from lower latitudes than they could in the past. While midlatitude extended range forecasts have improved substantially over the last couple of decades, these advancements stem from many factors beyond remote tropical influence including, for example, improved model physics, increased model resolution, better initialization, and the use of ensembles (Bauer et al., 2015). Therefore, it might be argued that the skill originating from the tropics still lies in the noise when compared to the current state of midlatitude predictive skill. The main goal of this study is to investigate, through the use of existing hindcast data sets, to what extent current extended midlatitude forecasts are able to draw skill from the tropics, including characterizing typical lead times involved as well as potential dependencies on the tropical low frequency basic state.

Tropical-to-extratropical teleconnections originate with tropical precipitation, where the associated convective heating is balanced by upward motion leading to upper level divergent flow. This divergence anomaly is seen to be the primary driver of the so-called "Rossby wave source" through vortex stretching, and planetary and relative vorticity advection by the divergent horizontal flow (Hoskins & Karoly, 1981; Sardeshmukh & Hoskins, 1988). These studies showed that subtropical Rossby wave sources are the most effective way to excite propagation of Rossby wave energy from low to high latitudes where they can impact weather, and that the sources are most efficiently produced within regions of strong relative vorticity gradients associated with the extratropical westerly flow, such as found in the vicinity of the wintertime subtropical jets. The midlatitude response to tropical heating is very rapid in both theory and models, with a substantial response even after 2 days, with its character depending on the season and frequency of the tropical forcing (Branstator, 2014; Matthews et al., 2004; Newman & Sardeshmukh, 1998). Over time the response initially spreads poleward and eastward, with lower frequency forcing such as that associated with the Madden and Julian

Table 1
Summary of ECMWF and NCEP Model Configuration

	NCEP	ECMWF
Model	CFSv2	IFS CY43R3
Range (days)	44	46
Horizontal resolution	~100 km	16 km up to day 15 and 31 km after day 15
Vertical levels	64	91
Top of model (hPa)	0.02	0.02
Frequency	daily	2× a week
Length	1999–2010	past 20 years
Ocean Coupling	yes	yes

Note. ECMWF = European Centre for Medium-Range Weather Forecasts; NCEP = National Centers for Environmental Prediction; CFSv2 = Climate Forecast System version 2; IFS = Integrated Forecast System.

Oscillation (MJO) and El Niño Southern Oscillation (ENSO) being particularly effective at producing teleconnections (Berbery & Nogues-Paegle, 1993; Hsu, 1996; Liu et al., 2016; Roundy et al., 2010). Of particular relevance is that even short-lived tropical convective pulses might also have a systematic imprint on subsequent midlatitude weather days and weeks later (Branstator, 2014, their Figures 3 and 5). A relationship between low and high latitude forecast skill is, therefore, expected across a wide range of timescales based on our understanding of these teleconnections. On the other hand, skillful prediction of the midlatitude response to tropical forcing also relies strongly on accurate representation of tropical heating and its vertical structure, along with nonlinearities and interactions with the background flow (Ambrizzi & Hoskins, 1997; Sardeshmukh & Hoskins, 1988).

Besides theoretical arguments, another reason to anticipate a relationship between tropical and extratropical forecast errors stems from what are referred to as "relaxation experiments" using forecast models (Ferranti et al., 1990; Hansen et al., 2017; Jung et al., 2010; Klinker, 1990). This approach involves nudging forecasts toward analyses or reanalyses over a tropical region, while allowing the model to run freely elsewhere. By comparing nudged to global free running forecasts, these studies have generally shown that midlatitude forecasts are improved in association with reducing tropical forecast errors. Jung et al. (2010), for example, showed that Weeks 2–4 forecast errors over the North Pacific and North America in particular are reduced by tropical nudging. While relaxation seems to be an effective technique to demonstrate global impacts of poor representation of the tropics, it is limited because it artificially accounts for both the predictable and unpredictable portion of tropical phenomena. In addition, these experiments tend to be performed at relatively low resolution in comparison to what is used in current subseasonal prediction systems. For these reasons, the estimates of contributions from tropical errors to extratropical errors from relaxation experiments must be interpreted as an upper bound that is unlikely to be achieved. This further motivates our goal here of estimating the impacts of tropical forecasts on midlatitude subseasonal predictive skill in existing forecast model runs.

The Subseasonal to Seasonal (S2S) prediction project database (Vitart et al., 2017) offers an unprecedented opportunity to statistically assess the relationship between tropical and extratropical forecast errors. For conciseness, we focus on the National Centers for Environmental Prediction (NCEP) and European Centre for Medium-Range Weather Forecasts (ECMWF) subseasonal reforecasts and on the October–March period. The two systems were chosen because of their contrasting behavior in the tropics (see section 3). The analysis presented here builds on the methods used in Dias et al. (2018) and could be easily expanded to other regions, seasons, and models.

2. Data
2.1. S2S Reforecasts and Verification Data Sets
NCEP and ECMWF reforecasts are archived in the S2S database (Vitart et al., 2017) and detailed model information can be found online (at https://confluence.ecmwf.int/display/S2S/Models). Table 1 displays a summary of the ECMWF and NCEP model configurations used in this study. At NCEP, reforecasts were

produced from a fixed Climate Forecast System version 2 (CFSv2) configuration for the period of 1999–2010 with daily initializations. At ECMWF, reforecasts are produced "on the fly" using the latest version of Integrated Forecast System (IFS) with twice a week initializations. Another difference, as shown in Table 1, is that the NCEP reforecasts are produced at much coarser resolution than ECMWF reforecasts.

We primarily use the 624 initial times that are common to both models, which correspond to twice a week initializations from 1999 to 2010 during October through March. The entire NCEP (ECMWF) S2S data set includes 4,380 (1,976) reforecasts. All model output are regridded to 1.5° × 1.5° before storage in the S2S database, which is the resolution we use here. The variables included in this analysis are daily average precipitation and instantaneous daily values of geopotential heights at 500 hPa (z500) and meridional winds at 200 hPa (v200).

Precipitation forecast skill is evaluated against the Global Precipitation Measurement (GPM) satellite 3B42 product (Huffman et al., 2007), which is available at 3-hourly resolution as area averages of precipitation rate on a 0.25° grid between 50°S and 50°N. ERA-Interim reanalysis (Dee et al., 2011) is used to verify z500 and v200. GPM is not directly assimilated in either systems and the results shown here are not overly sensitive to the specific reanalysis product used for verification. Both ERA-Interim reanalysis and GPM verification data sets are regridded to the S2S grid.

2.2. MJO and ENSO Indexes

To characterize the MJO, we use the OLR-based MJO index (OMI), which is designed to identify the MJO convective signal (Kiladis et al., 2014). We define MJO active periods as cases where OMI is above its upper tercile (1.5) and inactive when it is below the lower tercile (0.85), where thresholds are calculated using all October–March OMI values from 1979 to 2017. Sensitivity tests of these thresholds and MJO index used suggest that our main conclusions are robust to how we define MJO activity (not shown). The ENSO state is assessed according to the Oceanic Nino index (http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php) where warm/cold periods are based on a threshold of ± 0.5 °C. The conditional sample sizes using the thresholds mentioned above are the following: 221/198 MJO active/inactive cases and 192/165/267 ENSO neutral/warm/cold. We note that our mean results might be overly influenced by cold ENSO events since they dominate the 1999–2010 period.

3. Measures of Predictive and Conditional Skill

We report our conditional skill results using anomaly pattern correlation coefficients (APC), defined as

$$APC(IN, LT) = \frac{cov(X'_f, X'_o)}{std(X'_f)std(X'_o)}. \tag{1}$$

The covariance (*cov*) and standard deviation (*std*) are calculated using reforecasts and verification anomalies (X'_f, X'_o) from all grid points that lie within specified regions, at a fixed initial time (*IN*) and at each forecast lead time (*LT*). Anomalies are calculated by removing the lead time dependent first three harmonics of the mean seasonal cycle from each model. The mean seasonal cycle is calculated by filling any missing calendar day with the monthly mean value. Additional testing indicate that the seasonal cycle is reasonably well defined given the available reforecast record. We apply the same method to calculate the seasonal cycle and anomalies for all data sets used here.

The tropical region where the APC is calculated is defined as all longitudes between 15°S and 15°N, which includes most of the regions of intense tropical convection during October–March. Reducing the tropical region to 10°S–10°N does not affect our conclusions. The NH verification region is defined as all longitudes between 30 and 50°N. The APC calculation yields a distribution of APC values for each region and at each lead from Day+1 to +44. For a more continuous transition from short to extended lead times, we also adopt the methodology from Zhu et al. (2014) and others, where the forecast lead time at Day "x" uses an averaging window of "x" days and those are denoted 1d1d, 2d2d, and so forth. Here 2d2d skill corresponds to the skill of the average Day+3/Day+4 prediction and 1w1w is the skill of the average Week 2 prediction (see schematic in Figure 1 from Zhu et al., 2014). We have computed other verification measures including root mean square error, bias, equitable threat score, and fractions skill score and found that our conclusions in terms of relative and conditional skill regarding the CFSv2 versus IFS and tropical versus extratropical regions are not overly sensitive to the metric used.
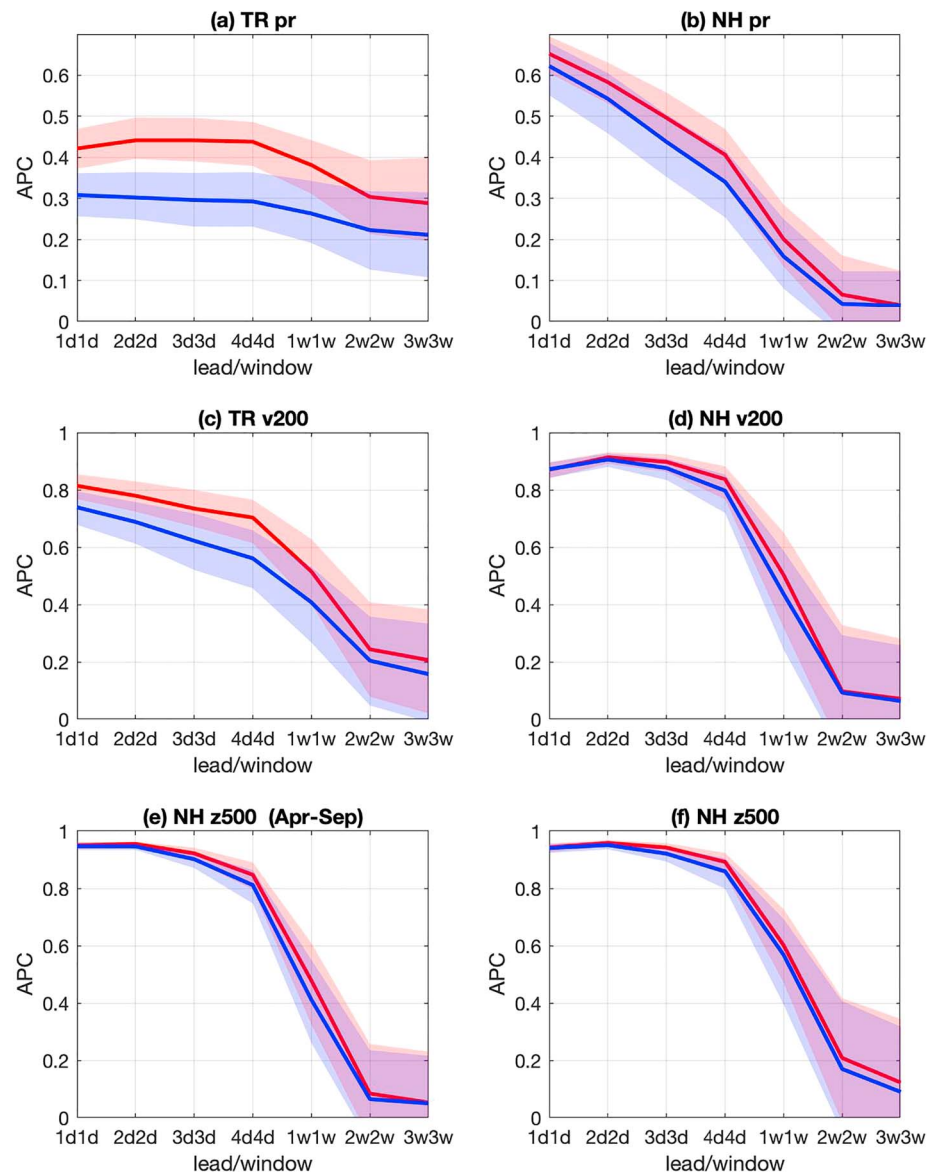
Figure 1. Seamless anomaly pattern correlation (APC) during October–March for (a) tropical precipitation (TR pr), (b) NH precipitation (NH pr), (c) tropical 200 hPa meridional wind (TR v200), (d) NH 200 hPa meridional wind (NH v200), and (f) NH 500 hPa geopotential height (NH z500). Panel (e) shows APC during April–September for NH 500 hPa geopotential height (NH z500). Thick lines display the APC median over all common CFSv2 (blue) and IFS (red) reforecasts. Shading is similar, except that for the upper and lower quartiles.

Reforecast initial times are split depending on whether the tropical APC at a fixed lead/window is below (above) the lower (upper) quartile ("poor" vs "good" tropical forecasts) following the approach from Dias et al. (2018). We apply this method to "seamless" APC values where we compare conditioned to mean values. Since we are interested in tropical forecast errors associated with misrepresentation of latent heating sources, the skill of the tropical forecast is measured by the tropical precipitation APC. The skill of NH forecasts is measured by precipitation, z500, and v200. Throughout the paper, confidence intervals are constructed using bootstrapping, where the lower and upper intervals of each measure are defined as the 5th and 95th percentiles of their distributions calculated from 1,000 resampled realizations.

4. Conditional Skill Analysis

An overview of the predictive skill of each model out to 3w3w (roughly Week 4) over the tropics for precipitation and v200 and over the NH for precipitation, v200 and z500 can be seen in Figure 1. As demonstrated
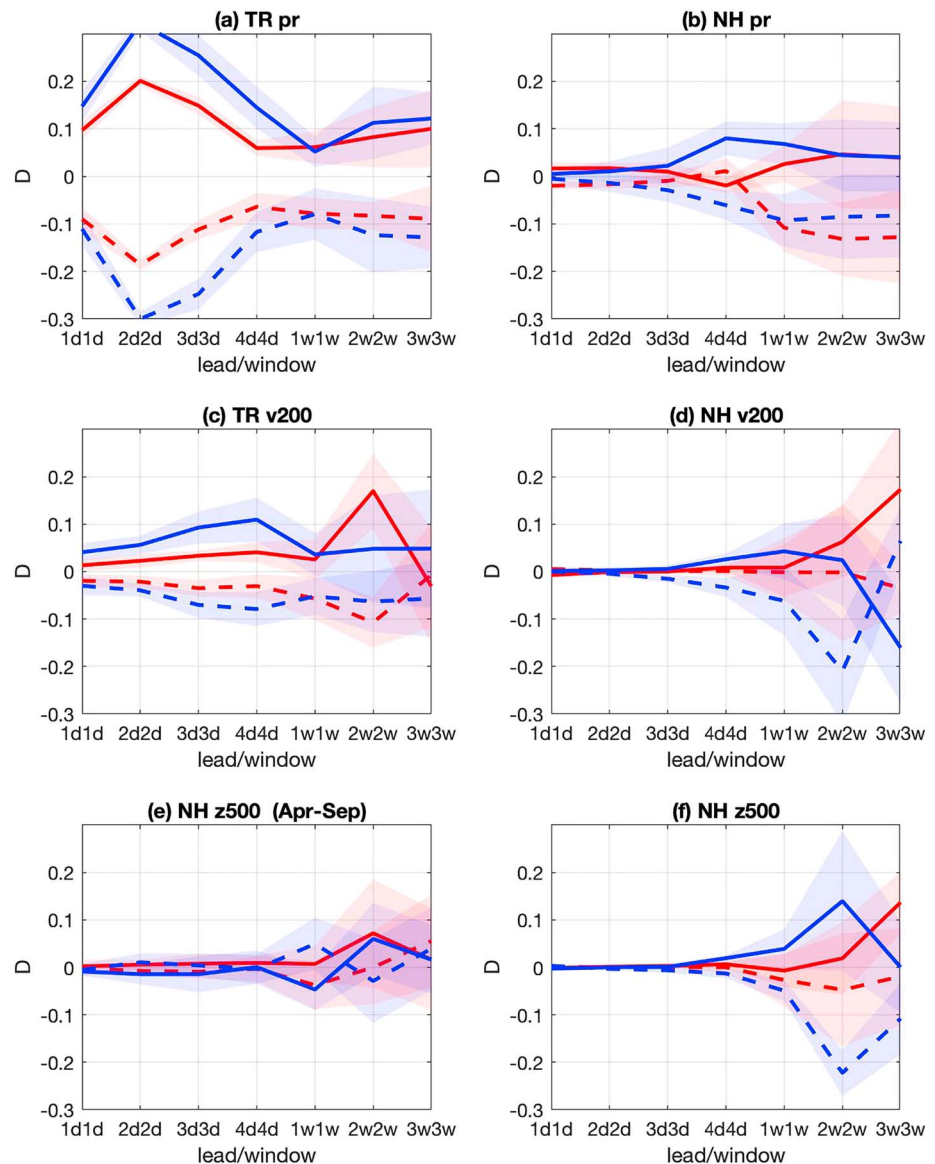
Figure 2. As in Figure 1, except that panels show normalized conditional skill (*D*; see definition in section 4) for reforecasts where 2d2d tropical precipitation APC is above the upper quartile (solid) and below the lower quartile (dashed). The solid and dashed lines represent the median *D*, over 1,000 realizations using random sampling with replacement and shading shows the 5% and 95% percentiles of the distribution. Only common dates common CFSv2 (blue) and IFS (red) reforecasts are used.

in a recent study (Janiga et al., 2018), Figures 1a and 1c show substantially higher tropical skill in the IFS than in the CFSv2. This is also the case for daily APC values (not shown), where the tropical precipitation APC decorrelation, as measured by the number of days to drop to half of the Day+1 APC, is 3.5 in the CFSv2 versus 9 days in the IFS. Although the IFS also tends to outperform the CFSv2 skill by 15% to 20% in the NH, the differences are much less accentuated than in the tropics (Figures 1b, 1d, and 1f). These relative differences between tropics and extratropics are seen in the period of April–September as well (Figure 1e). The contrast in model performance makes the comparison between tropical and NH skill more intriguing because the IFS has a better representation of tropical convective variability than the CFSv2, which might have implications for how tropical-extratropical interactions are handled. We also note that the differences between the IFS and CFSv2 are larger when APC values are averaged over the Northern Pacific and North America (not shown), which are regions known to be sensitive to tropical forecast errors (Jung et al., 2010).
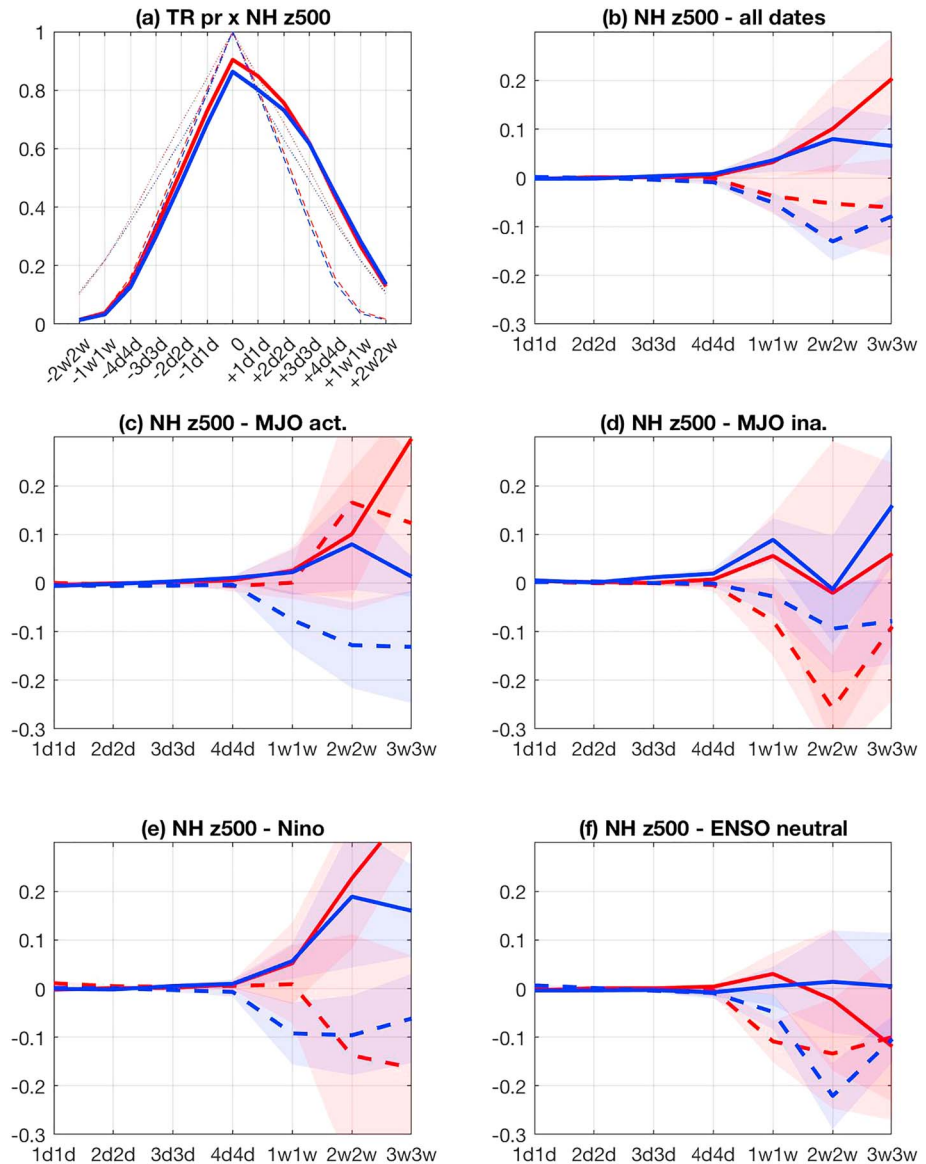
Figure 3. (a) shows the lead/lag cross-correlation for tropical precipitation and NH z500 (thick lines) and the lead/lag autocorrelation for NH z500 (dashed) and tropical precipitation (dotted). Positive (negative) lead/lag mean that tropical precipitation leads (lags) NH z500. (b)–(f) are similar to Figure 2c, except that all reforecasts available are included where (c)–(f) are further conditioned on periods of active/inactive MJO (c and d) and warm/neutral ENSO (e and f). MJO = Madden and Julian Oscillation; ENSO = El Niño Southern Oscillation.

Figure 2 shows how tropical and NH APC evolve with lead/window when the 2d2d tropical forecasts are good or poor in comparison to the mean. The conditional skill is displayed as a normalized difference (*D*) between conditional and unconditional seamless APCs defined as

$$D = \frac{APC_{cond} - APC}{APC}. \tag{2}$$

Therefore, if $D < 0$ ($D > 0$), the conditional forecast skill is worse (better) than the mean skill over the analyzed period. The choice of conditioning on tropical precipitation APC for lead/window 2d2d is based on the compromise between short range skill and its dependence on initial conditions, but results are not too sensitive to small changes on this parameter. Figures 2a and 2c show uniformly positive (negative) *D* values in the case of good (poor) 2d2d tropical precipitation and v200, where *D* in the case of tropical precipitation (Figure 2a) peaks at 2d2d by design. This is not too surprising since one would expect 2d2d skill to

be positively correlated with nearby lead/windows and upper level tropical flow errors to be correlated with errors in precipitation. These differences are also seen when conditioning on 1d1d and when using only the overlapping dates of good and poor forecasts across the two models (not shown). In the case of both precipitation and v200, the CFSv2 conditional forecast skill is further from the unconditional skill out to 4d4d. This initial skill difference could be due to the fact that since the IFS has a better representation of the tropics, there is both less of an absolute difference between good and poor tropical performance, and less skill dependence on the initial conditions. The peak at 2w2w in the IFS v200 might be a sampling issue since it does not appear when forecasts from 2011 to 2017 are included (not shown).

The evolution of precipitation, v200, and z500 skill in the NH is especially interesting because in both models, the pick up in conditional skill is delayed, starting at about lead/window 3d3d (Figures 2b, 2d, and 2f), which would roughly coincide with the early stages of the estimated Rossby response time discussed above. While the amplitude of the pick up/drop off in skill depends on the thresholds that define the performance of the 2d2d tropical forecast, its timing and the relative amplitude between models are reasonably robust. Based on Figure 2, it appears that CFSv2 NH predictions are more strongly modulated by short range tropical precipitation forecasts than in the IFS. Assuming that teleconnections patterns are handled similarly in both systems, this difference in amplitude can be interpreted as an indication that the IFS is already tapping its superior tropical skill for tropical-to-extratropical teleconnections, whereas the CFSv2 could do better if the tropics were improved. While this might be true for the period of 1999–2010, when the entire record of forecasts is included, the pickup in the amplitude of NH skill in both models is similar (Figure 3b). Because CFSv2 reforecasts are not available for the period of 2011–2017, it is unclear whether this is a sampling issue. Inclusion of all available forecasts as opposed to only the overlapping period between IFS and CFSv2 also makes the pick up/drop off at 4d4d in z500 even more pronounced (Figure 3b), and the drop off at 3w3w for z500 and v200 (not shown) in the CFSv2 disappears. Note also that the increase/decrease in skill during the NH warm season is not seen in Figure 2e, which confirms the expectation that the basic state influences the Rossby wave source (e.g., Newman & Sardeshmukh, 1998).

The differences in NH conditional skill across variables are larger at longer lead/windows, when the uncertainty in the estimates is also larger. In general, while the tendencies of the conditional skill at longer lead/window are suggestive, those results have to be interpreted with caution because forecasts are not particularly skillful based on APCs in these ranges (Figures 1b and 1d–1f), and also samples are relatively small. We have calculated the reciprocal conditional skill to investigate NH forecasts influences in the subsequent tropical skill, which shows that early lead NH APCs are not associated with changes in later lead tropical APCs. Similarly, we have not found an analogous relationship in the Southern Hemisphere cool season. However, that the relationship is not revealed by the hemispheric APC analysis, does not mean that there are not particular regions in the extratropics that influence particular tropical regions. For example, there is evidence that tropical Eastern Pacific variability is modulated by higher latitudes when the upper tropospheric "Westerly duct" (Webster & Holton, 1982) is active (e.g., Matthews & Kiladis, 2000). A remarkable feature of Figure 2 is the relative symmetry between conditional skill on either good or poor tropical performance, and these symmetries are even stronger when entire record of IFS and CFSv2 reforecasts are used. This further supports the hypothesis that tropical skill systematically modulates NH skill at extended lead times.

We also calculate the lead-lag correlations between tropical and extratropical skill and contrast those values to the NH lead-lag autocorrelation. The idea is that at a fixed lead-lag, a cross-correlation that is larger than the autocorrelation suggests that NH skill at that lead is more related to processes originating within the tropics than in the NH. We first examine the NH z500 APC autocorrelation values (dashed lines in Figure 3a), and note that they are very similar for the IFS and CFSv2. The solid lines show the cross-correlation between NH z500 APC and tropical precipitation APC, and in both cases the cross-correlation with tropical precipitation APC is larger than the autocorrelation at longer positive lead times. The crossing occurs earlier for the IFS than for the CFSv2, and this is also seen when looking at daily APCs (not shown). The asymmetry of the lead-lag correlation implies that the converse argument is not true, such that, at longer lead/windows, the correlations between tropical precipitation and NH z500 APCs are less than the tropical precipitation or z500 autocorrelation (dotted and dashed lines in Figure 3a). We found that these relationships are similar when using tropical and NH v200. Overall, the lead-lag correlation is consistent with the conditional skill, again suggesting that in both models NH predictive skill is related to the performance of earlier lead tropical forecasts.

One way to investigate how tropical low frequency variability influences the conditional skill seen in Figures 2c and 2d is to further condition good and poor forecasts on MJO activity or ENSO phase. The main limitation of this analysis is the relatively short extent of the overlapping reforecast record, which already leads to small sample sizes even without considering MJO and ENSO. Since the uncertainty of our estimates when also controlling for MJO activity and ENSO phase is extremely large when using only common dates, the conditional skill estimates in Figures 3c–3f include all available October–March reforecasts. The caveat is that comparisons between the models are complicated by the fact that the CFSv2 period includes only a subset of the MJO and ENSO events from the IFS period. Despite the unavoidable limitations regarding sample sizes and periods, there are a number of interesting features related to how ENSO and MJO affect conditional skill:

- During active MJO periods, the IFS Weeks 2–4 NH z500 APCs are better than average regardless of the 2d2d tropical precipitation performance. In contrast, the CFSv2 enhanced/attenuated skill depending on 2d2d tropical APC is nearly insensitive to MJO active periods.
- MJO inactive periods do not seem to affect the tropical/extratropical Weeks 1–2 skill relationship in either systems. Weeks 3–4 NH z500 performance is worse than average when 2d2d tropical precipitation skill is poor. During MJO inactive periods, Week 3 NH z500 skill is not sensitive to when 2d2d tropical precipitation skill is good.
- ENSO warm phases generally amplify positive/negative differences seen in Figure 3b. When ENSO is neutral, conditional skill during Weeks 1–2 behaves similarly to when all ENSO phases are included. In contrast, Weeks 3–4 conditional skill is asymmetric in that changes are primarily seen when 2d2d tropical precipitation skill is poor. It also appears that, when ENSO is neutral, the IFS Week 4 NH z500 forecasts are less skillful regardless of the 2d2d tropical precipitation APC.

Overall, when comparing different variables and thresholds, it appears that the evolution of conditional skill in the IFS is more sensitive to the MJO and ENSO than in the CFSv2, possibly because the IFS tropical variability is better represented, and also as indicated by its better overall tropical skill in comparison to the CFSv2.

5. Summary and Conclusions

As discussed in section 1, there is an expectation from both theory and idealized modeling for extended (weekly to monthly) extratropical predictive skill in numerical forecast models to be influenced by tropical forecast errors at earlier lead times. We apply a conditional skill analysis to evaluate this relationship in S2S predictions from the NCEP-CFSv2 and ECMWF-IFS. In terms of absolute skill, the CFSv2 tends to underperform when compared to the IFS at both low and high latitudes, but these differences are much larger in the tropics (Figure 1). With respect to conditional skill, when comparing the same set of initial times, our analysis suggests that in both systems, NH cool season predictions beyond a few days lead time tend to be better when the short range tropical forecast is good, and vice-versa. CFSv2 NH Weeks 2–4 predictions are more sensitive to short range precipitation forecasts in the tropics in comparison to the IFS (Figures 2b, 2d, and 2f). These differences in sensitivity could be due to the fact that the IFS performs better in the tropics and therefore there is less of a difference between good and poor tropical skill than in the CFSv2; however, other factors such as model resolution and physics could be important as well because they could affect the relative importance of tropical-to-extratropical forecast error propagation.

The systematic relationships found between tropical precipitation and NH cool season predictive skill are seen in z500, v200, and even in precipitation. Therefore, our results suggest that subseasonal forecasts could benefit from model improvements leading to a reduction of tropical forecast errors. Interestingly, based on the tropical predicability limits estimated by Ying and Zhang (2017) such improvements might be possible. One caveat of the present analysis is the limited size of overlapping reforecasts between IFS and CFSv2. This sample size sensitivity is illustrated by comparing the IFS conditional skill using 1999–2010 (Figure 2f) to 1998–2017 (Figure 3b), which shows that conditional skill is sensitive to the period of analysis. By testing thresholds to characterize tropical performance and comparing different variables, we found that at least the timing of the delayed skill pick up/drop off, if not its amplitude, is reasonably robust across different periods.

Despite issues with sample sizes, the MJO/ENSO modulation of conditional skill seems consistent with the fact that the IFS performs better in the tropics. For instance, when the MJO is active, IFS Weeks 3–4 z500

APCs are better than average, regardless of tropical precipitation earlier lead performance. That could be because, as shown by many previous studies (see Stan et al., 2017), the NH draws extended range skill from the MJO persistent large-scale convective signal, therefore the impact of tropical short range forecast performance is less noticeable when the MJO is active. The MJO is a known weakness of the CFSv2 (Hendon et al., 2000; Janiga et al., 2018), where we found a relative insensitivity of conditional skill during MJO active periods. While not the focus here, we did not find an analogous relationship during April–September between tropics and Southern Hemisphere, but it is possible that such relationship would appear over particular longitudinal sectors. Another interesting result is that modulations of subsequent NH skill, particularly at Week 2, are seen during MJO inactive and ENSO neutral phases. That is, other tropical processes besides the MJO and ENSO, such as higher frequency tropical waves, might also play a role on how much remote skill from the tropics can be tapped.

Detailed predictability studies are certainly needed to better understand and characterize how extratropical predictive skill depends on multiscale tropical variability. The relaxation experiments discussed in section 1 are one way to further evaluate these dependencies; however, they are computationally expensive in comparison to our conditional skill analysis of long records of reforecasts. Our hope is that we have provided an alternate approach here that might be helpful as an initial investigative tool of predictive skill associated with tropical-to-extratropical teleconnections that can then be used to guide the design of predictability experiments.

References

Ambrizzi, T., & Hoskins, B. J. (1997). Stationary Rossby-wave propagation in a baroclinic atmosphere. *Quarterly Journal of the Royal Meteorological Society*, *123*(540), 919–928.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*, 47–55.

Berbery, E. H., & Nogues-Paegle, J. (1993). Intraseasonal interactions between the tropics and extratropics in the southern hemisphere. *Journal of the Atmospheric Sciences*, *50*(13), 1950–1965.

Branstator, G. (2014). Long-lived response of the midlatitude circulation and storm tracks to pulses of tropical heating. *Journal of Climate*, *27*(23), 8809–8826.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

Dias, J., Gehne, M., Kiladis, G. N., Sakaeda, N., Bechtold, P., & Haiden, T. (2018). Equatorial waves and the skill of NCEP and ECMWF numerical weather prediction systems. *Monthly Weather Review*, *146*(6), 1763–1784.

Ferranti, L., Palmer, T. N., Molteni, F., & Klinker, E. (1990). Tropical-extratropical interaction associated with the 30–60 day oscillation and its impact on medium and extended range prediction. *Journal of the Atmospheric Sciences*, *47*(18), 2177–2199.

Goswami, B. B., Khouider, B., Phani, R., Mukhopadhyay, P., & Majda, A. J. (2017). Improved tropical modes of variability in the NCEP climate forecast system (version 2) via a stochastic multicloud model. *Journal of the Atmospheric Sciences*, *74*(10), 3339–3366.

Hansen, F., Greatbatch, R. J., Gollan, G., Jung, T., & Weisheimer, A. (2017). Remote control of North Atlantic Oscillation predictability via the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 706–719.

Hendon, H. H., Liebmann, B., Newman, M., Glick, J. D., & Schemm, J. E. (2000). Medium-range forecast errors associated with active episodes of the Madden-Julian Oscillation. *Monthly Weather Review*, *128*(1), 69–86.

Hirons, L. C., Inness, P., Vitart, F., & Bechtold, P. (2013). Understanding advances in the simulation of intraseasonal variability in the ECMWF model. Part I: The representation of the MJO. *Quarterly Journal of the Royal Meteorological Society*, *139*(675), 1417–1426.

Hoskins, B. J., & Karoly, D. J. (1981). The steady linear response of a spherical atmosphere to thermal and orographic forcing. *Journal of the Atmospheric Sciences*, *38*(6), 1179–1196.

Hsu, H.-H. (1996). Global view of the intraseasonal oscillation during northern winter. *Journal of Climate*, *9*(10), 2386–2406.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*(1), 38–55.

Janiga, M. A., J. Schreck, C., Ridout, J. A., Flatau, M., Barton, N. P., Metzger, E. J., & Reynolds, C. A. (2018). Subseasonal forecasts of convectively coupled equatorial waves and the MJO: Activity and predictive skill. *Monthly Weather Review*, *146*(8), 2337–2360.

Jung, T., Miller, M. J., & Palmer, T. N. (2010). Diagnosing the origin of extended-range forecast errors. *Monthly Weather Review*, *138*(6), 2434–2446.

Kiladis, G. N., Dias, J., Straub, K. H., Wheeler, M. C., Tulich, S. N., Kikuchi, K., et al. (2014). A comparison of OLR and circulation-based indices for tracking the MJO. *Monthly Weather Review*, *142*(5), 1697–1715.

Klinker, E. (1990). Investigation of systematic errors by relaxation experiments. *Quarterly Journal of the Royal Meteorological Society*, *116*(493), 573–594.

Liu, F., Li, T., Wang, H., Deng, L., & Zhang, Y. (2016). Modulation of boreal summer intraseasonal oscillations over the western North Pacific by ENSO. *Journal of Climate*, *29*(20), 7189–7201.

Matthews, A. J., Hoskins, B. J., & Masutani, M. (2004). The global response to tropical heating in the Madden and Julian Oscillation during the northern winter. *Quarterly Journal of the Royal Meteorological Society*, *130*(601), 1991–2011.

Matthews, A. J., & Kiladis, G. N. (2000). A model of Rossby waves linked to submonthly convection over the eastern tropical Pacific. *Journal of the Atmospheric Sciences*, *57*(23), 3785–3798.

Newman, M., & Sardeshmukh, P. D. (1998). The impact of the annual cycle on the North Pacific/North American response to remote low-frequency forcing. *Journal of the Atmospheric Sciences*, *55*(8), 1336–1353.

Roundy, P. E., MacRitchie, K., Asuma, J., & Melino, T. (2010). Modulation of the global atmospheric circulation by combined activity in the Madden-Julian Oscillation and the El Nino Southern Oscillation during boreal winter. *Journal of Climate*, *23*(15), 4045–4059.

Sardeshmukh, P. D., & Hoskins, B. J. (1988). The generation of global rotational flow by steady idealized tropical divergence. *Journal of the Atmospheric Sciences*, *45*(7), 1228–1251.

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

Vitart, F. (2014). Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1889–1899.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173.

Webster, P. J., & Holton, J. R. (1982). Cross-equatorial response to middle-latitude forcing in a zonally varying basic state. *Journal of the Atmospheric Sciences*, *39*(4), 722–733.

Xiang, B., Zhao, M., Jiang, X., Lin, S.-J., Li, T., Fu, X., & Vecchi, G. (2015). The 3–4-week MJO prediction skill in a GFDL coupled model. *Journal of Climate*, *28*(13), 5351–5364.

Ying, Y., & Zhang, F. (2017). Practical and intrinsic predictability of multiscale weather and convectively coupled equatorial waves during the active phase of an MJO. *Journal of the Atmospheric Sciences*, *74*(11), 3771–3785.

Zhu, H., Wheeler, M. C., Sobel, A. H., & Hudson, D. (2014). Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Monthly Weather Review*, *142*(4), 1556–1569.