

Attribution of North American Subseasonal Precipitation Prediction Skill

LANTAO SUN,^a MARTIN P. HOERLING,^b JADWIGA H. RICHTER,^c ANDREW HOELL,^b ARUN KUMAR,^d
AND JAMES W. HURRELL^a

^a *Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado*

^b *NOAA/Physical Sciences Laboratory, Boulder, Colorado*

^c *Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

^d *NOAA/Climate Prediction Center, College Park, Maryland*

(Manuscript received 20 April 2022, in final form 15 July 2022)

ABSTRACT: The skill of NOAA's official monthly U.S. precipitation forecasts (issued in the middle of the prior month) has historically been low, having shown modest skill over the southern United States, but little or no skill over large portions of the central United States. The goal of this study is to explain the seasonal and regional variations of the North American subseasonal (weeks 3–6) precipitation skill, specifically the reasons for its successes and its limitations. The performances of multiple recent-generation model reforecasts over 1999–2015 in predicting precipitation are compared to uninitialized simulation skill using the atmospheric component of the forecast systems. This parallel analysis permits attribution of precipitation skill to two distinct sources: one due to slowly evolving ocean surface boundary states and the other to faster time-scale initial atmospheric weather states. A strong regionality and seasonality in precipitation forecast performance is shown to be analogous to skill patterns dictated by boundary forcing constraints alone. The correspondence is found to be especially high for the North American pattern of the *maximum* monthly skill that is achieved in the reforecast. The boundary forcing of most importance originates from tropical Pacific SST influences, especially those related to El Niño–Southern Oscillation. We discuss physical constraints that may limit monthly precipitation skill and interpret the performance of existing models in the context of plausible upper limits.

SIGNIFICANCE STATEMENT: Skillful subseasonal precipitation predictions have societal benefits. Over the United States, however, NOAA's official U.S. monthly precipitation forecast skill has been historically low. Here we explore origins for skill of North American week-3 to week-6 precipitation predictions. Skill arising from initial weather states is compared to that arising from ocean surface boundary states alone. The monthly and seasonally varying pattern of U.S. monthly precipitation skill is appreciably derived from boundary constraints, linked especially with El Niño–Southern Oscillation. Forecasts of opportunity are identified, despite the low skill of monthly precipitation forecasts on average. Potential limits of monthly precipitation skill are explored that provide insight on the juxtaposition of “skill deserts” over the central United States with high skill regions over western North America.

KEYWORDS: North America; ENSO; Precipitation; Climate prediction; Forecast verification/skill

1. Introduction

There is great demand for information on the expected behavior of weather and climate many weeks into the future (National Academies of Sciences, Engineering and Medicine 2016). This demand can be met by skillful subseasonal forecasts, defined broadly as forecasts from two weeks to three months (Lucas 2017), which can guide effective decision-making intended to mitigate the negative effects of adverse environmental conditions or take advantage of favorable ones (White et al. 2017). Recognizing the importance of subseasonal predictions, there have been coordinated research and operational efforts to understand the temporal and spatial

variations of prediction skill, including their sources and prospects for improvement. These efforts include the Subseasonal-to-Seasonal (S2S) Project (Robertson et al. 2015; Vitart et al. 2017) and the Subseasonal Experiment (SubX; Pegion et al. 2019).

Given the importance of forecast guidance in various aspects of society, including droughts, floods, and their modulation of near surface temperature, understanding subseasonal precipitation prediction skill, sources of prediction skill, and the potential for improvement continues to be a principal area of research. This understanding is even more important considering poor historical precipitation prediction skill on subseasonal time scales from leading forecast centers like the NOAA Climate Prediction Center (CPC). Figure 1 shows the precipitation prediction skill of CPC's probabilistic monthly precipitation outlooks since 1995 issued in the middle of the prior month. Little to no skill is apparent across the majority of the continental United States with comparably higher skill over the southern tier and negative skill over parts of the Midwest. The time series of forecast skill over 1995–2021 (not shown) also indicates that although forecast practices have

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-22-0076.s1>.

Corresponding author: Martin P. Hoerling, martin.hoerling@noaa.gov

January 1995 - May 2022 0.5-Month Lead Monthly Precipitation Forecast Skill

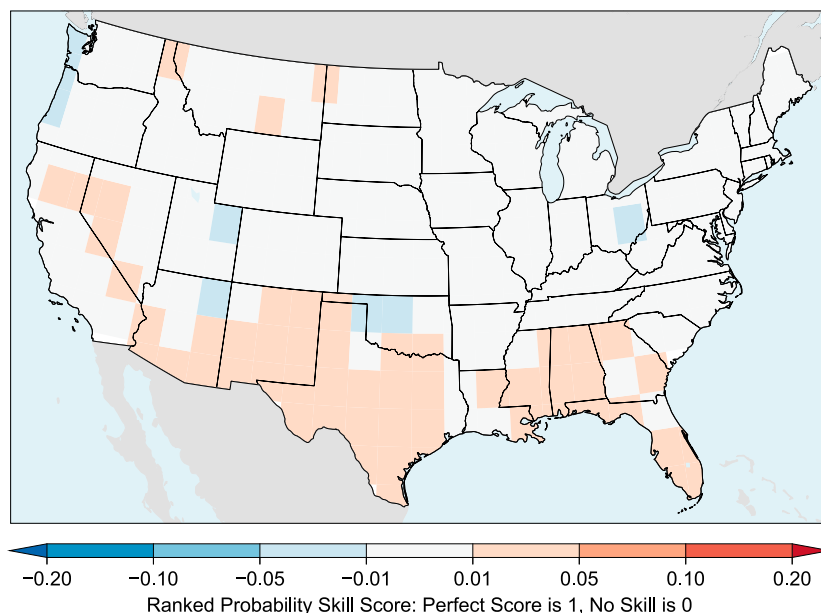


FIG. 1. Rank probability skill score for CPC's probabilistic monthly precipitation outlooks issued in the middle of prior month (i.e., 0.5-month lead). See <https://www.cpc.ncep.noaa.gov/products/verification/summary/index.php?page=map>.

evolved over time, this evolution has not led to appreciable improvements in prediction skill. Changes in forecast practices include an increasing reliance on initialized predictions based on newer coupled models with improved representation of physical processes and advances in data assimilation (e.g., Kwon et al. 2018; Mariotti et al. 2018).

The needs for skillful precipitation prediction, and lack thereof over the continental United States, motivated NOAA to establish the Precipitation Prediction Grand Challenge (PPGC) in 2020 with the strategic objective to “provide more accurate, reliable, and timely precipitation forecasts across time scales from weather to subseasonal-to-seasonal to seasonal-to-decadal through the development and application of a fully Earth system prediction model.” Toward achieving this strategic objective, the PPGC advocates for identifying and understanding the key physical processes related to precipitation prediction. In response to the PPGC, we perform a systematic attribution of subseasonal precipitation prediction skill over North America by considering the following four topics: 1) the regional patterns of forecast skill and their attribution to particular predictability sources, 2) the temporal characteristics of skill including causes for its seasonality and reasons for its overall decay as a function of lead time, 3) the robustness of time/space variations in skill across different prediction systems, and 4) the interpretation of actual forecast skill in a theoretical framework of potential skill.

This study therefore seeks to explain the time/space variations in the skill of North American subseasonal precipitation predictions using recent-generation forecast systems. It proceeds to demonstrate that there are better prospects, under circumstances, than suggested in the gross annually averaged,

multidecade averaged perspective of skill revealed in Fig. 1. Here we quantify the skill of dynamical predictions arising from two sources: one linked explicitly with slowly evolving boundary states and the other due to faster time-scale initial states of the Earth system. Many studies, either via empirical case studies or analyses of prediction models, have linked precipitation skill to tropical intraseasonal convection (Whitaker and Weickmann 2001; Zhou et al. 2012; Li and Robertson 2015; Vitart 2017; Zheng et al. 2018; Nardi et al. 2020; Dias et al. 2021; Stan et al. 2022), stratosphere–troposphere dynamical coupling (Sigmond et al. 2013; Nardi et al. 2020; Albers and Newman 2021), land surface processes (Koster et al. 2011; Dirmeyer et al. 2018), atmospheric rivers (Mundhenk et al. 2018; DeFlorio et al. 2019), variations of atmospheric low-frequency modes such as the North Atlantic Oscillation (NAO)/Arctic Oscillation (AO) and the Pacific–North American pattern (PNA; see the review article by Stan et al. 2017). More recently, Krishnamurthy et al. (2021) evaluated the S2S forecast skill of precipitation over CONUS using the Unified Forecast System and showed that during boreal summer sources of predictability are associated with an intraseasonal oscillation with a period of about 50 days and a warming trend mode. Studies have also attempted to isolate subseasonal precipitation skill to boundary forcing alone especially related to El Niño–Southern Oscillation (ENSO; DelSole et al. 2017; Wang and Robertson 2019). Here we seek to quantify the separate effects of boundary forcing and initial weather states on S2S variability, and thereby attribute the origins of the precipitation skill specifically.

An early effort on forecast skill attribution was conducted by Kumar et al. (2011) based on 1981–2006 hindcasts from an

TABLE 1. Summary of reforecast datasets.

Reforecast model	Ensemble members	Initialization frequency	Reforecast length (days)	Reforecast period	References
IFS	11	Twice per week	46	1997–2016	Vitart (2014), S2S project: Vitart et al. (2012, 2017)
CFSv2	4	Daily	45	1999–2015	Saha et al. (2014), SubX project: Pegion et al. (2019)
GEFSv12	11	Once per week	35	2000–19	Zhou et al. (2016, 2017), Zhu et al. (2018)
CESM1	11	Once per week	45	1999–2015	Richter et al. (2020)
CESM2	10	Once per week	45	1999–2020	Richter et al. (2022)

early generation initialized coupled model. Their analysis explored the skill dependence of monthly means for different prediction lead times, the roles of initial versus boundary conditions in determining prediction skill, and the spatial and seasonal dependence in prediction skill. The main results of their analyses at the global scale include the following: (i) the time scale for the influence of initial conditions was inferred to be approximately 30–40 days at which time the skill levels were comparable to those attributed to sea surface temperature anomalies; and (ii) over land area, precipitation skill beyond 20 days was low. Similar conclusions were found by Quan et al. (2012) in their study of the subseasonal skill sources for precipitation attending drought over the contiguous United States.

In this study, we build on methods of Kumar et al. (2011) and conduct a parallel diagnosis of recent generation reforecast predictions and Atmospheric Model Intercomparison Project (AMIP; Gates et al. 1999) simulations that utilize identical dynamical models. We use four different prediction systems in order to address robustness. Our analysis, while not providing a granularity on the specific processes associated with skill contributions from initial weather variability, separates the overall contribution of initial state information on skill from boundary state information. This approach is applied to the time evolving skill from weeks 1 to 6, but will focus principally on the monthly average, taken here as weeks 3–6 (corresponding to CPC's 2-week-lead monthly forecast). This time scale of predictions continues to be a frontier science challenge for disentangling the forward stretch of initialized weather skill from the backward reach of climate-based boundary condition skill. Our study focuses on the physical origins of North American precipitation prediction skill. It does not seek to optimize the prediction skill in forecast systems, nor does it explore the skill improvements gained from statistical methods and postprocessing calibrations.

The outline of this paper is as follows. In section 2, we describe the tools and methods, notably the initialized reforecasts and uninitialized AMIP simulations and the diagnostic methods. In section 3, we present the seasonality and regionality of the precipitation skill and the connection to the lower boundary conditions. A summary and discussion are provided in section 4.

2. Data and methodology

a. Reforecasts and AMIP simulations

Five recent-generation reforecast datasets are utilized in this study to evaluate precipitation skill. These include reforecasts

from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS), National Centers for Environmental Prediction (NCEP) Coupled Forecast System version 2 (CFSv2), NCEP Environmental Modeling Center (EMC) Global Ensemble Forecast System version 12 (GEFSv12), and Community Earth System Model (CESM) Version 1 (CESM1) and Version 2 (CESM2). These reforecasts are generally initialized in various ways at different frequencies with different ensemble sizes and forecast lead time (Table 1). Because the primary purpose is prediction from weeks to decades, CESM1 and CESM2 do not employ data assimilation and incorporate initialization procedures used for their systems' decadal predictions (Yeager et al. 2018), unlike IFS, CFSv2, and GEFSv12. The CESM atmospheric model is initialized using the reanalysis (ERA-Interim for CESM2 and NCEP CFSv2 reanalysis for CESM2). The land initial conditions are produced by running a stand-alone land model spinup forced with observed atmospheric forcing and the ocean–sea ice initial conditions come from an ocean sea ice coupled configuration forced with the adjusted Japanese 55-year Reanalysis (JRA-55) products. Details can be found in Richter et al. (2020, 2022) for CESM1 and CESM2, respectively.

To facilitate the comparison of these datasets, we use the common period of 1999–2015 (with the exception of 2000–15 for GEFSv12) for the majority of the analysis. In addition, we examine the sensitivity of the precipitation prediction skill to the length of the analysis period by looking at an extended reforecast period for IFS (i.e., including years 1997/98 and 2015/16). Recognizing that prediction skill is sensitive to forecast ensemble size (e.g., Kumar and Hoerling 2000; Sun et al. 2020; Meehl et al. 2021), we ensure that comparable ensembles are used for skill evaluation of each of the five reforecast datasets. We follow Wang and Robertson (2019) to construct a 12-member lag ensemble for CFSv2 so that all reforecasts have ensemble size close to 11, which is generally available for the other systems. Last, the length of GEFSv12 reforecasts is only 35 days, in contrast to 45 days in other forecast systems. Since the focus of this study is the weeks-3–6 reforecast, we only use the GEFSv12 to evaluate the weekly precipitation skill.

To isolate the role of monthly surface boundary forcing in the precipitation prediction skill, we utilize four uninitialized simulations with the atmospheric component of the IFS, CFSv2, CESM1 and CESM2, which are IFS-AMIP (ECMWF 2014a,b,c), Global Forecast System version 2 (GFSv2; Saha et al. 2014), Community Atmosphere Model version 5 (CAM5; Neale et al. 2012, used in CESM1), and version 6 (CAM6; Danabasoglu et al. 2020, used in CESM2), respectively

TABLE 2. Details of the AMIP simulations. The SST/SIC forcings include Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST2; Titchner and Rayner 2014), merged products of HadISST2 and NOAA NOAA-Optimum Interpolation version 2 (OI; Reynolds et al. 2002) (Had-OI; Hurrell et al. 2008), and NOAA Extended Reconstructed SST V5 (ERSSTv5; Huang et al. 2017).

Atmosphere model	Corresponding reforecast model	Ensemble members	SST/SIC forcing	Period	References
IFS-AMIP	IFS	13	Had-OI or HadISST2	1999–2014	Haarsma et al. (2016)
GFSv2	CFSv2	12	Had-OI	1999–2015	Murray et al. (2020)
CAM5	CESM1	11	Had-OI	1999–2015	Murray et al. (2020)
CAM6	CESM2	10	ERSSTv5	1999–2015	NCAR CVCWG ^a

^a NCAR Climate Variability and Change Working Group (CVCWG): https://www.cesm.ucar.edu/working_groups/CVC/simulations/cam6-prescribed_sst.html.

(Table 2). The atmosphere model historical simulations follow the AMIP protocol (Gates et al. 1999) and are forced by observed greenhouse gases, sea surface temperature (SST) and sea ice conditions (SIC). We use nearly the same ensemble size and period between the reforecast and atmosphere model simulations.

The reforecasts and their corresponding atmosphere model simulations are verified against the Global Precipitation Climatology Project (GPCP) daily precipitation analysis (Adler et al. 2017). GPCP covers the global ocean and thus can help evaluate the precipitation skill over the North American continent and over the tropical oceans.

b. Skill evaluation

We use two metrics to evaluate the precipitation prediction skill against observations: anomaly correlation coefficient (ACC) that measures the temporal correlation at each grid point and uncentered, area-weighted anomaly pattern correlation (APC) that measures the spatial correlation for each forecast (e.g., Newman et al. 2003). They can be expressed by

$$[\text{ACC}(x, y), \text{APC}(t)] = \left[\frac{\sum_{t=1}^T (X_{\text{mod}} - C_{\text{mod}})(X_{\text{obs}} - C_{\text{obs}})}{\sqrt{\sum_{t=1}^T (X_{\text{mod}} - C_{\text{mod}})^2 (X_{\text{obs}} - C_{\text{obs}})^2}}, \frac{\sum_{x,y} (X_{\text{mod}} - C_{\text{mod}})(X_{\text{obs}} - C_{\text{obs}})}{\sqrt{\sum_{x,y} (X_{\text{mod}} - C_{\text{mod}})^2 (X_{\text{obs}} - C_{\text{obs}})^2}} \right], \quad (1)$$

where X_{mod} and X_{obs} represent the values for the model ensemble mean and GPCP reanalysis, respectively. The term

C_{mod} is the lead-time dependent model climatology, calculated using methodology presented in Pegion et al. (2019), and is over the same period of time as the reanalysis climatology C_{obs} . The ACC and APC are diagnosed for the weekly and monthly (weeks 3–6) averages to be the corresponding skill. The APC is calculated over the North American continent (i.e., 10°–55°N, 67°–129°W with the exclusion of small areas in South America; see Fig. 4) and skill time series can be generated for each weekly reforecast and corresponding AMIP simulations. These skill time series are then smoothed by biseasonal (26 week) running mean (i.e., we use 26-point unweighted running average for those reforecasts initialized once per week and 52-point running average for those initialized twice per week). ACC is aggregated over four seasons: June–July–August (JJA), September–October–November (SON), December–January–February (DJF), and March–April–May (MAM). For each season, we also subsample the skill conditioned onto the ENSO phases. Specifically, the El Niño (La Niña) states are defined by NOAA’s oceanic Niño index (ONI) rising above 0.5°C (falling below −0.5°C) for that season. Table 3 lists the ENSO-active and neutral summers and winters during 1999–2015. The statistical significance of the skill is estimated using the nonparametric bootstrapping method (Mudelsee 2010). While yielding similar results to those based a standard Student’s t test, the bootstrapping method makes no assumptions on the distribution of the statistics, contrary to the Gaussianity assumption of the Student’s t test. For each model, we resample the reforecast (or AMIP) cases and their corresponding observations 1000 times with replacement and define the ACC to be statistically significant if the 2.5% value across the 1000 bootstrapping samples is above zero.

We also evaluate the perfect prognostic (hereafter perfect-prog; Kumar and Hoerling 1998a) ACC skill for the reforecasts

TABLE 3. ENSO active and neutral summers and winters during 1999–2015 defined in this study.

Summer (JJA)			Winter (DJF)		
El Niño	Neutral	La Niña	El Niño	Neutral	La Niña
2002, 2004, 2009, 2015	2001, 2003, 2005, 2006, 2008, 2012, 2013, 2014	1999, 2000, 2007, 2010, 2011	2002/03, 2004/05, 2006/07, 2009/10, 2014/15, 2015/16 (only Dec)	2001/02, 2003/04, 2012/13, 2013/14	1998/99 (only Jan, Feb), 1999/2000, 2000/01, 2005/06, 2007/08, 2008/09, 2010/11, 2011/12

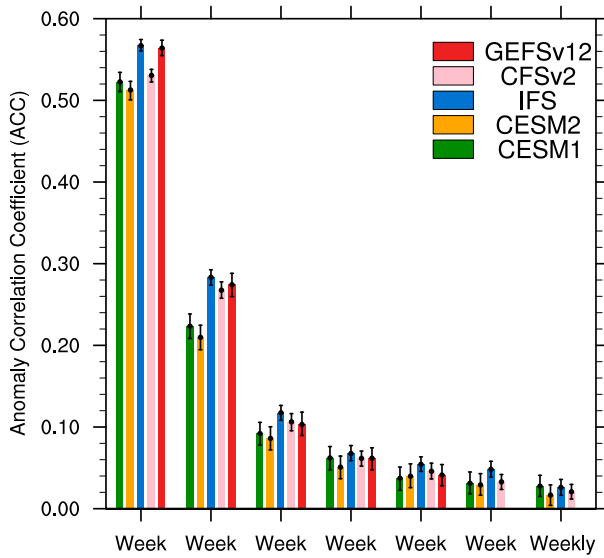


FIG. 2. Weekly precipitation skill averaged over North American land (10° and 55°N) for five reforecast systems (Table 1) and four AMIP ensembles (Table 2). Error bars indicate the 95% confidence interval (2.5%–97.5%) based on bootstrapping.

and AMIP simulations, which approximates the upper limit of prediction skill. For any individual member of an ensemble, the perfect-prog skill can be diagnosed by using Eq. (1), but replacing the model ensemble-mean ($X_{\text{mod}} - C_{\text{mod}}$) and observational anomaly ($X_{\text{obs}} - C_{\text{obs}}$) with the anomaly from one individual member and model ensemble-mean anomaly (but excluding that individual member). We note the caveat that as a result of model biases the predictable signals in models may be less than in observations (Kumar et al. 2014; Scaife and Smith 2018), thus the forecast skill upper limit may be higher than the perfect prognostic skill. However, as will be shown, such an apparent paradox of actual versus perfect-prog performance that has been noted for seasonal sea level pressure over the Atlantic basin especially (Scaife and Smith 2018) does not emerge for North American monthly precipitation skill studied herein.

3. Results

a. North American averaged and annually averaged precipitation skill

We begin by summarizing the evolution of weekly mean precipitation skill averaged over North America and over the entire year for the period 1999–2015. Reforecasts from all models exhibit a nearly identical exponential decay in anomaly correlation skill (Fig. 2). Average anomaly correlations of between 0.5 and 0.6 in week 1 fall to roughly 0.04 by week 6. Predictions generated from reforecast models that do not include a data assimilation system for their initialization (CESM1 and CESM2) are less skillful in the first several weeks but become indistinguishable from the more advanced prediction systems by about week 4. Nonetheless, even these somewhat more primitive initialized systems capture the vast

majority of skill achieved by their more advanced counterparts (IFS, CFSv2, GEFSv12) during the first several weeks.

The reforecast anomaly correlation skill from weeks 1 to 6 asymptotes to the weekly-mean AMIP precipitation skill (Fig. 2, rightmost bars), rather than to a zero-skill condition. It is fair to question if there is any “useful skill” at week 6, at least for the North American average shown here. However, the reason of the nonzero skill is the boundary-forced constraint as given in the AMIP simulations, which subsequently are shown to have appreciable (and arguably useful) spatial/temporal variations. This boundary condition influence on skill originates principally from tropical SST variations, to be established below. Suffice it to note here that the weekly evolving skill shows the week-3–6 window to be a transition period, one in which the capacity for initial weather states to constrain forecast evolution earlier is handed over to boundary constraining influences later. Within this window, week-3 skill originates mainly from initial weather state information while week-6 skill originates mainly from boundary state information.

A spatial pattern correlation metric for North American precipitation skill (Fig. 3) exhibits similar sharp declines over the first several weeks, mimicking the rate of skill drop-off in the temporal anomaly correlation metric (see Fig. 2). Figure 3 presents a time series of the anomaly pattern correlation for 1999–2015 for week-1 (green), week-2 (pink), and the average of week-3–4 (brown) and week-5–6 forecasts (blue). Evident from this skill metric is a considerable seasonal and interannual variability in skill mostly in weeks 1 and 2, the analysis of which is pursued in subsequent sections. Importantly, from an attribution perspective, it is apparent that periods of high forecast skill during week 3 to week 4 and week 5 to week 6 coincide with periods of high skill in AMIP simulations (red curve in Fig. 3). The correlation between the reforecast and AMIP skill time series increases from 0.1 to 0.2 in the first two weeks to 0.6 in weeks 3–4 and 0.8 in weeks 5–6. Figure 3 thus provides an initial indication that an important monthly precipitation skill source is linked to boundary constraints on precipitation, and that these create considerable “forecasts of opportunity.”

b. Regionality and seasonality of North American monthly precipitation skill

We next consider precipitation skill (temporal anomaly correlation) averaged for the weeks-3–6 window (day 15–42), henceforth to be referred to as the monthly skill. This is effectively the skill of monthly averaged precipitation for predictions made at 2-week leads.

Figure 4 displays the reforecast skill for each of the modeling systems (rows) for the four seasons (columns): JJA (summer), SON (fall), DJF (winter), and MAM (spring). Colors denote areas with statistically significant correlation skill at a 95% significance threshold. The variability in the monthly forecast skill by season and region is considerable, such that important forecast capabilities emerge under scrutiny of their space and time fluctuations which were not evident from inspection of North American averaged and annually averaged skill (Fig. 1).

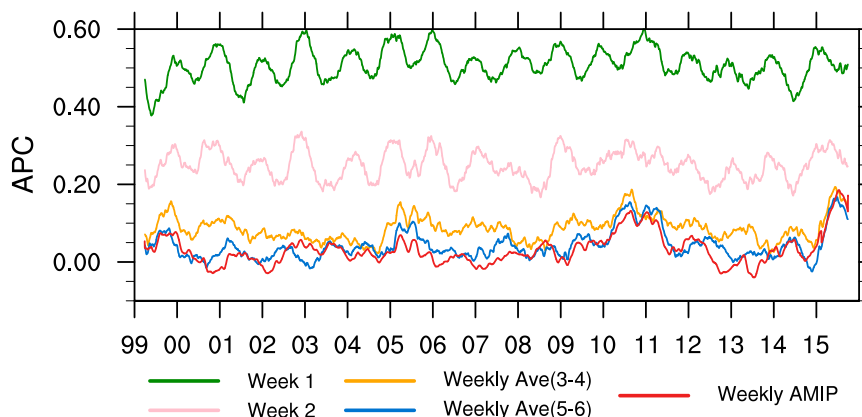


FIG. 3. Time series of the uncentered, area-weighted anomaly pattern correlation [APC; Eq. (1)] of North American precipitation skill for week-1 forecasts (green), week-2 forecasts (pink), and the average of the week-3 to week-4 (brown), and week-5 to week-6 weekly forecasts (blue). The forecast skill is based on an equal-weighted average skill of IFS, CFSv2, CESM1, and CESM2 models. The corresponding AMIP simulation skill is shown for the week-3 to week-6 period (red). The curves span 1999–2015, and have been smoothed with a moving 26-week average (see the text for details).

Beginning with predictions verifying during summer (Fig. 4, left column), there is an absence of significant skill over much of the contiguous United States in all models. By contrast, monthly rainfall over Central America and southern Mexico is skillfully predicted with correlations near 0.5, indicating an ability to anticipate important variations during this region's major rainy season. Predictions verifying in fall show significant skill over much of the southern United States and central Mexico, while skill wanes over Central America. Wintertime sees a farther northward displacement of high skill region, and a maximum of 0.4 centered over northern Mexico and the southwest United States. Other regions of significant skill during winter months occur over the Pacific Northwest (its rainy season) and also over the upper Ohio Valley and Great Lakes regions. For predictions verifying in spring (Fig. 4, right column), skill is confined to western North America. A particularly striking feature of the springtime skill pattern are the high correlations along the Pacific coastal sections of Central America and Mexico, a feature seen in all models.

Some features in the time and space variations of monthly precipitation reforecast skill are reproduced in the AMIP simulations (Fig. 5). In common with the reforecasts, skill in AMIP simulations is greatest over Central America in summer, with correlations exceeding 0.3 in all models. Fall skill is more spatially coherent being widespread over the southern United States and Gulf Coast regions—a seasonal change analogous to the seasonal progression of reforecast skill. Wintertime skill approaches 0.5 correlation in several models over northern Mexico and the southwest United States, values on par with the correlation skill in reforecast, despite the lack of skill in the northwest United States. Finally, the pattern of springtime AMIP skill, though less coherent over the western United States than in the reforecasts, does share the common feature of significant correlations along the Pacific coastal regions of Central America and Mexico particularly in IFS

and CAM models. Wang and Robertson (2019) found that weeks-3–4 spring precipitation skill over the northwest United States can be largely explained by its interannual component associated with AO. Thus, the smaller AMIP skill in this region may imply that the AO variations responsible for monthly precipitation skill in reforecasts are unconstrained by boundary forcing.

Having summarized the *average monthly* precipitation skill, we next present the *maximum value* of monthly skill for any consecutive 3-month window. The pattern of maximum monthly reforecast skill (Fig. 6, far left column) and its seasonal preference (second left most column) are significantly dictated by sensitivity to boundary constraints as revealed by the similarity with its AMIP counterpart. A reproducible pattern emerges from the four reforecast systems—maximum skill occurs in each over the western and southern United States and over Mexico/Central America (leftmost column). The value of maximum reforecast skill is generally between 0.3 and 0.4, though values exceeding 0.5 occur in some portions of Mexico and Central America. Also robust is the region characterized by an absence of significant skill—effectively a “skill desert”—over the central United States that includes the Great Plains and portions of the middle and lower Ohio Valley. Each of these primary features of maximum monthly forecast skill emerge from analysis of the performance in the AMIP simulations, including a western and southern North America maximum in the boundary-forced skill and a skill desert over the central United States. The absence of either reforecast or AMIP skill over the Great Plains in our multi-model analysis is consistent with a similar finding of a Great Plains skill desert by Quan et al. (2012) in their assessment of subseasonal dynamical predictions of precipitation that focused on droughts. It is especially noteworthy that the magnitude of maximum AMIP skill for monthly precipitation is on par with that in the reforecasts, attesting to the importance of

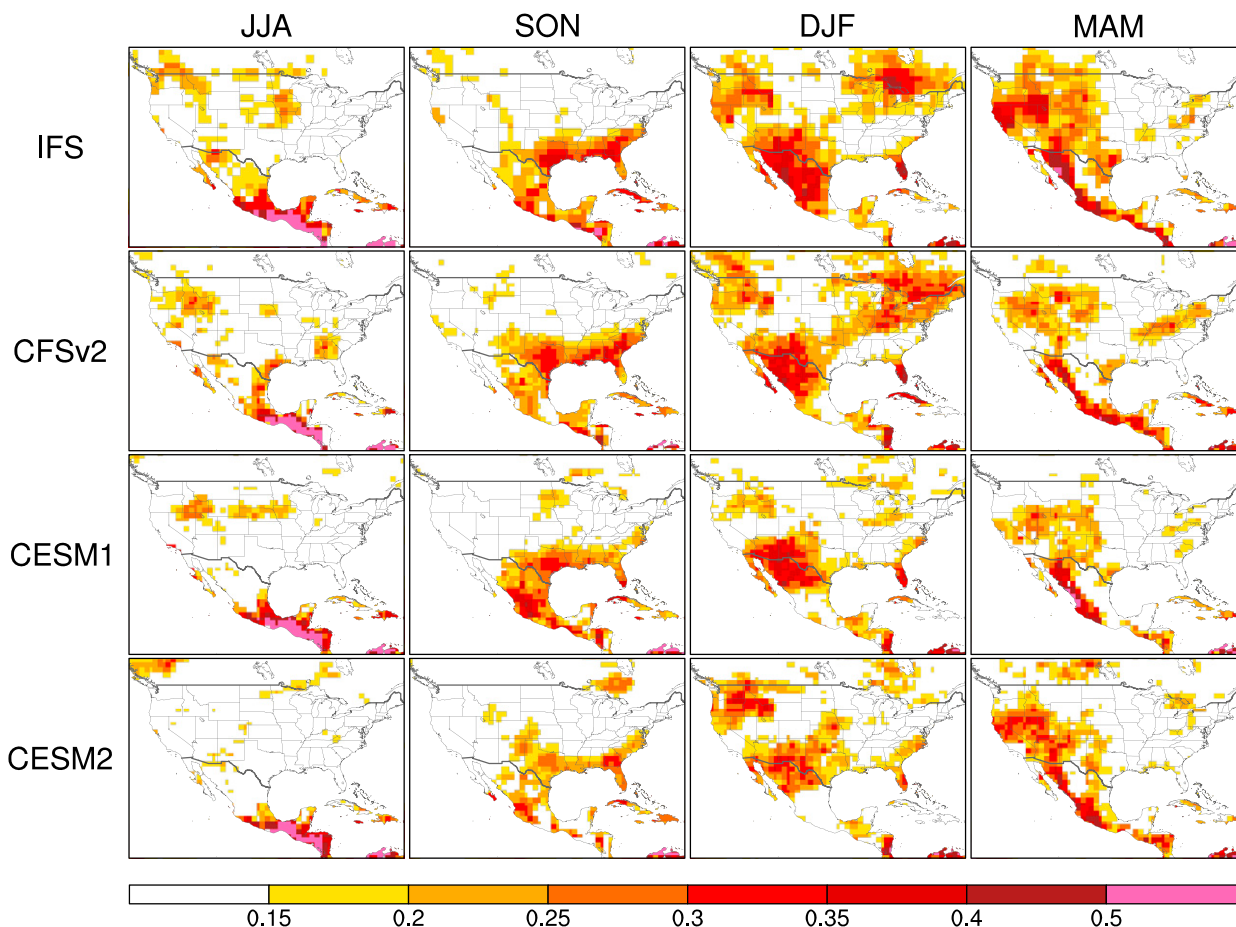


FIG. 4. Monthly (weeks 3–6) precipitation skill for the reforecast systems. Nonwhite grids show skill that is 95% statistically significant based on 1000 times bootstrapping.

boundary forcing for the maximum achieved skill – and generally where AMIP skill is low so too is the reforecast skill.

The two right-side columns of Fig. 6 illustrate the season during which maximum monthly correlation skill occurs. Over the contiguous United States the most skillful predictions of monthly precipitation tend to occur during the colder portion of the annual cycle, from about October to April in both reforecasts and AMIP, even though the exact season may differ slightly. Few contiguous U.S. locations experience their maximum monthly skill during summer, whereas summer is the season of maximum monthly precipitation forecast skill over Central America. In contrast, winter tends to be the maximum skill season over Mexico. These time/space variations align with those occurring in boundary forced AMIP simulations, partly reflecting a timing and regionality of ENSO-related teleconnections to be discussed further below. Analogously, while we cannot exclude the possibility of model bias (e.g., land–atmosphere interactions; Dirmeyer et al. 2018), the absence of significant skill maxima over much of the central United States in the reforecasts appears attributable to an absence of appreciable boundary forced skill source over that region.

c. Tropical effects related to North American monthly precipitation skill

Boundary forced contributions to North American monthly precipitation forecast skill, and to the magnitude and pattern of maximum achieved forecast skill in particular, is mostly attributable to tropical SST effects (e.g., Johnson et al. 2020). Several lines of evidence demonstrate that tropical SST variations, especially those linked to ENSO, are the major attributable boundary source for North American monthly precipitation forecast skill. First is the result that both reforecasts and AMIP have large monthly precipitation skill over the tropics (Fig. 7). Skill maxima occur over the equatorial central Pacific region where ENSO-related SST variability is large—the area-averaged monthly precipitation skill over the Niño-3.4 region is 0.6 in both reforecasts and AMIP (Fig. 8, left). Precipitation skill in reforecasts is appreciably greater than in AMIP simulations over the warmer portions of the tropical oceans such as the west Pacific and Indian Ocean (Fig. 8, right). It is plausible that the greater reforecast skill in these latter warm pool regions reflects the initialized system’s ability to predict intraseasonal time-scale rainfall variations (e.g., those associated with the Madden–Julian oscillation

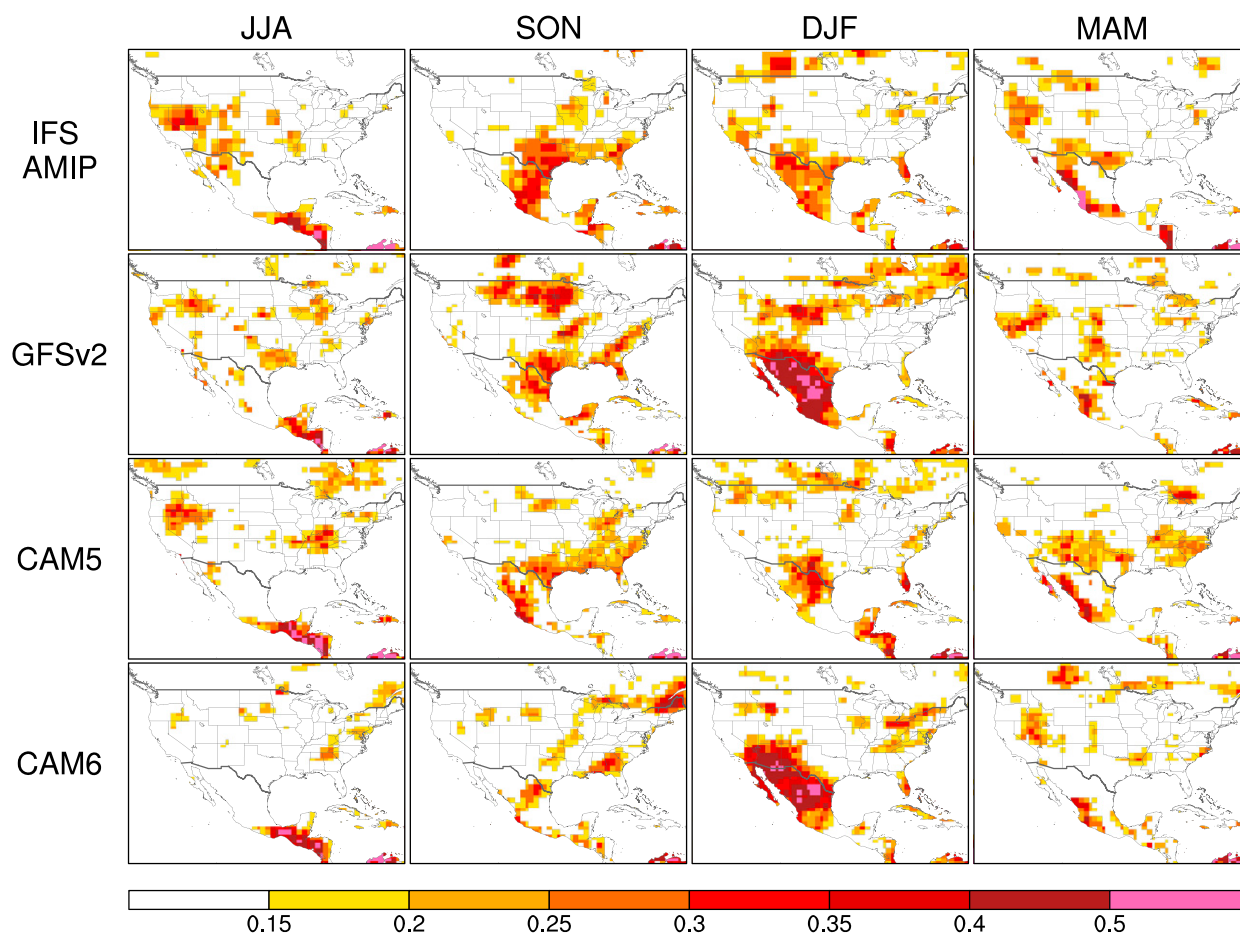


FIG. 5. As in Fig. 4, but for the skill in the corresponding AMIP simulations.

(MJO; Kim et al. 2019; Richter et al. 2020, 2022; Wang et al. 2022). It can also be due to AMIP biases, arising from the prescription of SST conditions, in misrepresenting air–sea interactions over the tropical warm pool regions (Kumar and Hoerling 1998b; Kumar et al. 2013). Nonetheless, the outstanding skill source, equally realized in reforecasts and AMIP, is over the cooler ocean regions of the central and eastern equatorial Pacific where ENSO is especially influential.

A second line of evidence comes from analysis of skill within subsamples of the verification period that are disaggregated according to the magnitude of NOAA’s ONI (3-month average of the SST anomalies in the Niño-3.4 region). Figure 9 shows the monthly skill during the summer and winter season for ENSO-active and ENSO-neutral years (see Table 3). The reforecast skill (top) is compared to the AMIP skill (bottom) derived from the multimodel mean, with similar results found for individual models (not shown). For both seasons, the monthly precipitation reforecast skill conditioned on ENSO-active situations is largely determined by the AMIP simulation skill. The ENSO contribution to skill is especially large in winter over the American Southwest, and also in summer over Central America. There are, nonetheless, indications that ENSO may not be the sole boundary forced skill source.

Note especially the evidence of some AMIP skill in summer over Central America during ENSO-neutral years, for which there is an analogous skill pattern in the reforecasts. For winter, it is also evident that monthly reforecasts are skillful over a large portion of western North America and central Plains when ENSO is inactive and the AMIP skill is nonexistent. While there is considerable complexity to these seasonally varying skill pattern and their conditionality on the nature of ENSO, the results for all four seasons are nonetheless robust in indicating that forecast skill during ENSO-active years is largely dictated by the seasonally and regionally varying signal of ENSO impacts.

These results suggest forecasts of opportunity during which greater skill can be expected owing to occasional large contributions from ENSO-related boundary constraints on monthly precipitation. Such is also the impression gained from the time series of anomaly pattern correlation skill for North American shown in Fig. 3. To get a local sense of such a forecast of opportunity, we present in Fig. 10 a time series of observed winter precipitation anomalies averaged over California during 1997–2016 (top) and its corresponding reforecast anomalies (middle panel) and AMIP simulated anomalies (lower panel). The largest observed anomaly

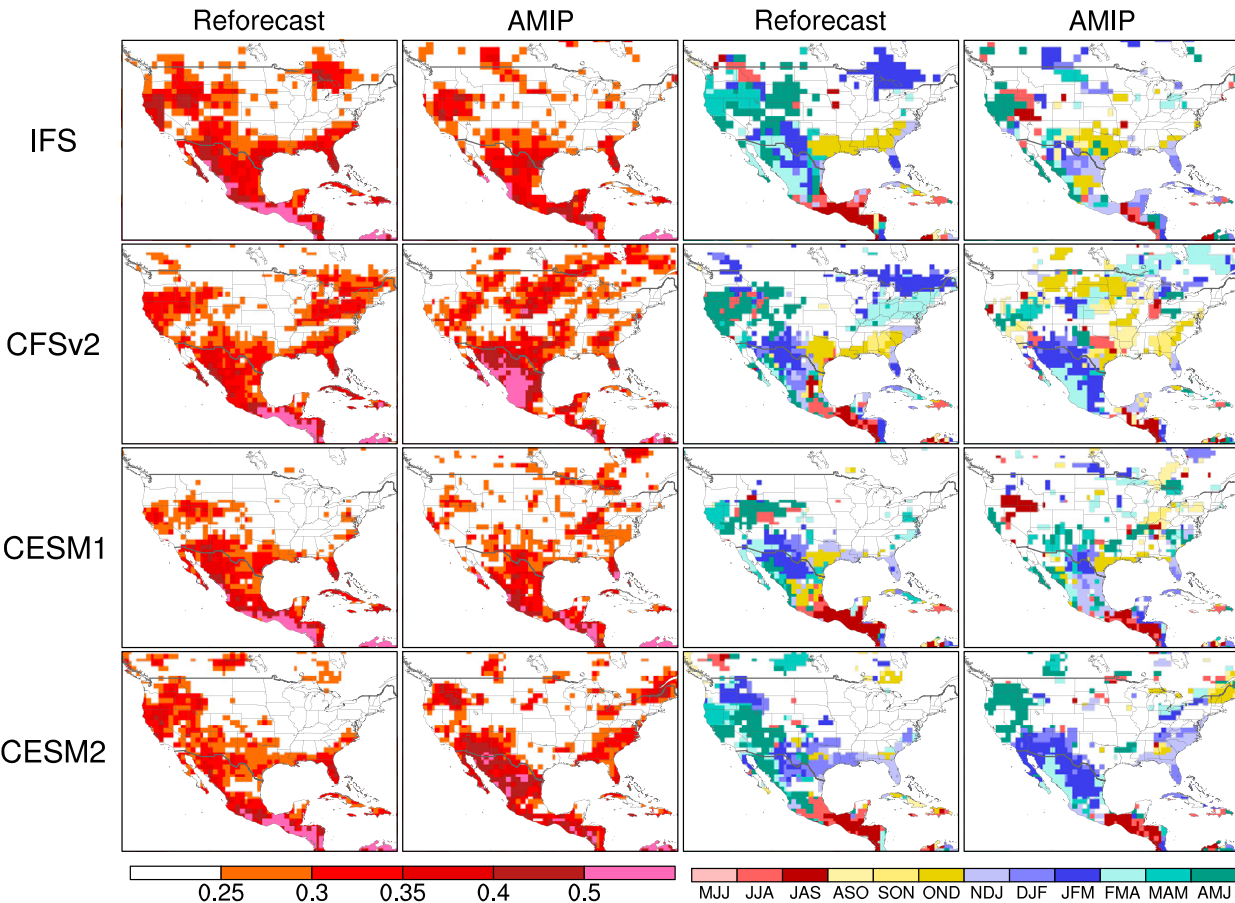


FIG. 6. (left) Maximum of the week-3-6 precipitation skill over any three consecutive months for the four reforecast systems and their corresponding AMIP simulations. (right) As in the left columns, but for the season with the maximum of the skill. The season is only calculated for the grid points where the maximum of the skill is above 0.25.

occurred in association with wet conditions during the strong 1997/98 El Niño, conditions well captured in model reforecasts and AMIP simulations. During most winters, however, the ensemble-mean anomalies in reforecasts and AMIP simulations are much smaller than those observed, indicating that the majority of observed monthly precipitation variance during California’s climatological wet season are not well constrained by either initial weather information (at 2-week lead) or by boundary condition information. These qualitative aspects of the variability are consistent

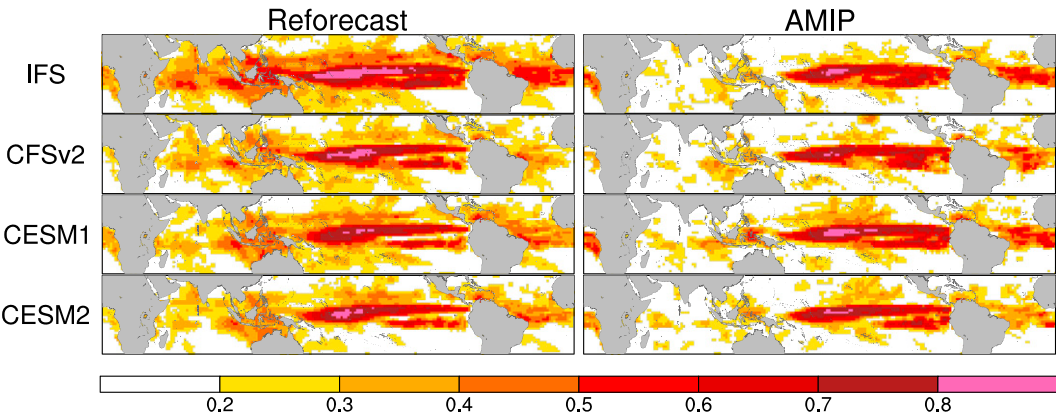


FIG. 7. As in Figs. 4 and 5, but for the annual tropical monthly precipitation skill.

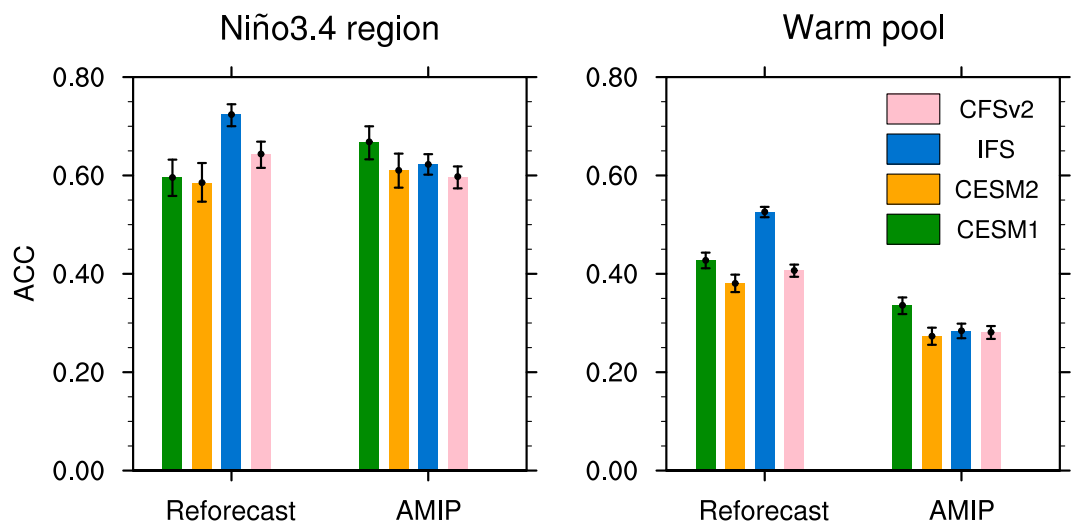


FIG. 8. Monthly precipitation skill averaged over (left) the Niño-3.4 region and (right) the Pacific warm pool region for the four reforecast systems and their corresponding AMIP simulations. Error bars indicate the 95% confidence interval (2.5%–97.5%) based on bootstrapping.

with the modest overall correlation skill in this region of about 0.2 in winter (see Figs. 4 and 5). Previous studies also found that California state-averaged precipitation is not well correlated with Niño-3.4 SSTs, and is better correlated with shifts in the position of the jet stream over the northeast Pacific (Wang et al. 2017; L’Heureux et al. 2021). Moreover, California precipitation is also influenced by MJO on S2S time scales, which can mask the ENSO signal (Arcodia et al. 2020).

There is a further implication of such forecasts of opportunity—namely, that assessments of precipitation skill

can be sensitive to the verification period used. This could be especially the case when the verification period experiences unusual ENSO variability. Figure 11 illustrates the sampling uncertainty of the IFS precipitation skill for reforecasts (top) and AMIP (bottom). Skill is greater when the analysis period includes the 1997/98 ENSO events (center column) compared to when that particular ENSO cycle is excluded in verifications (left column). This is particularly the case in regions where skill is elevated during ENSO-conditioned years alone (see Fig. S1 in the supplemental material), and is consistent with the qualitative impression gained from the Fig. 10

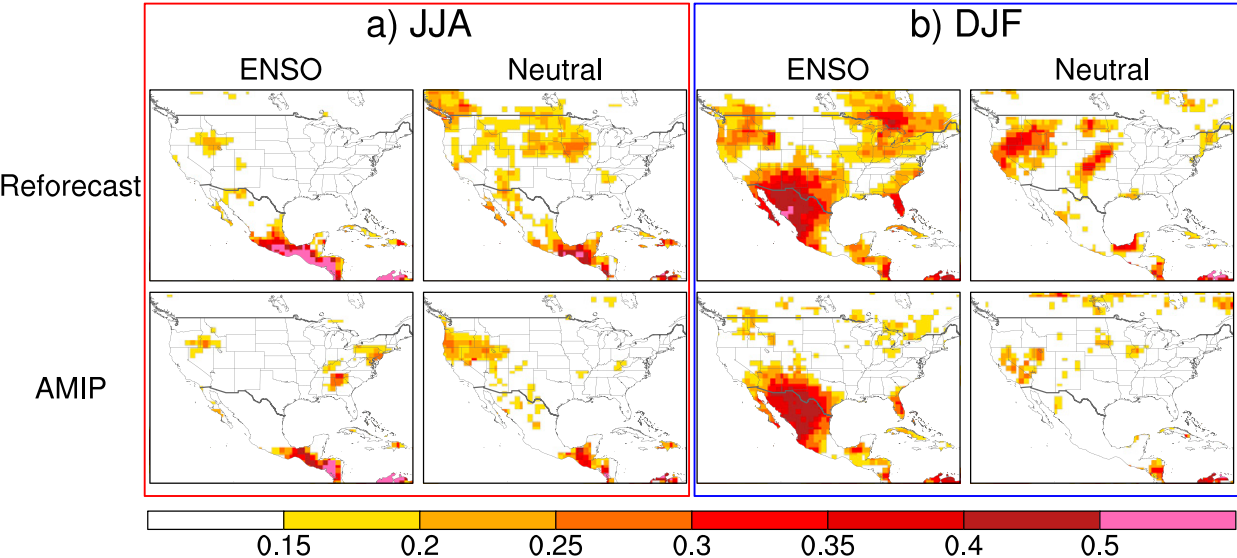


FIG. 9. Monthly precipitation skill in (a) JJA and (b) DJF over ENSO and ENSO-neutral states. The skill is calculated based on equally weighted multimodel mean over (top) four reforecasts and (bottom) their corresponding AMIP simulations. ENSO states are defined based on the NOAA ONI. See section 2 for details.

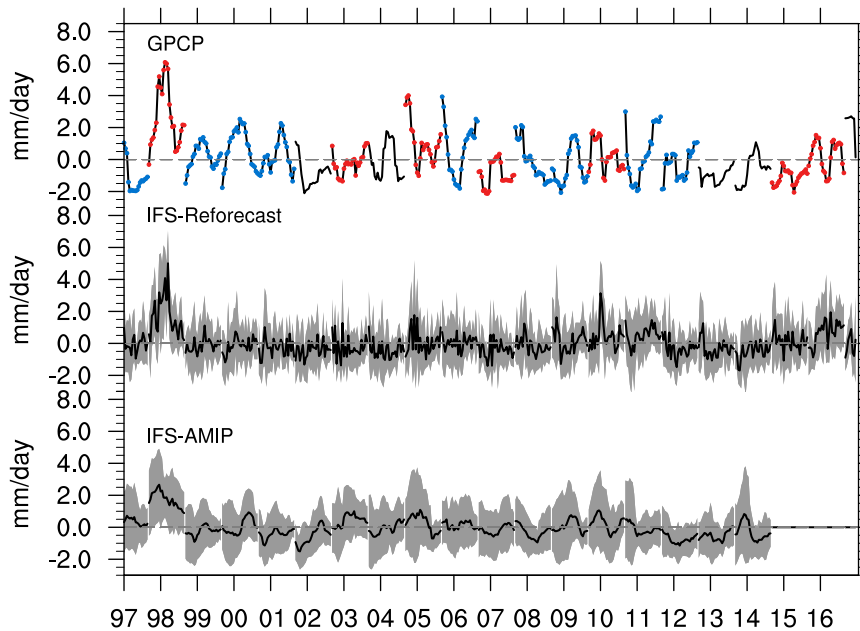


FIG. 10. The 1997–2016 time series of the DJF precipitation anomalies (mm day^{-1}) averaged over the state of California for observations (GPCP), IFS reforecasts, and IFS AMIP simulations (through 2014). Shading indicates ± 1 standard deviation among model individual members. Red and blue dots indicate the reforecast cases during El Niño and La Niña years, respectively.

precipitation time series showing a strong effect of the 1997/98 winter El Niño on California rainfall in both observations and model predictions. However, not all ENSO events, even when large in magnitude, are accompanied by enhanced monthly

forecast skill. For instance, if the verification period is augmented to include 2015/16 rather than 1997/98 (right column), comparatively little change in average skill is found. [Lybarger et al. \(2020\)](#) suggested that during the El Niño event in

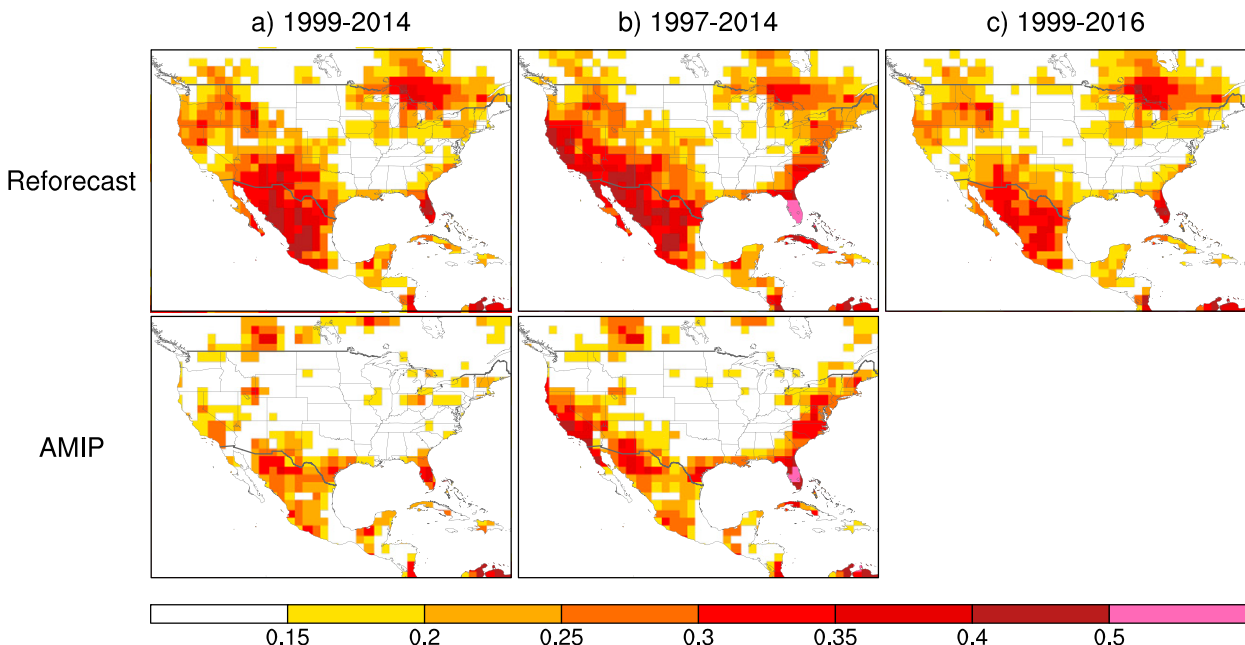


FIG. 11. Sensitivity of the winter (initialized in DJF) monthly precipitation skill to the reforecast period: (left) 1999–2014, (center) 1997–2014, and (right) 1999–2016 for the (top) IFS reforecast and (bottom) AMIP simulations. Note AMIP simulations available only thru 2014. Nonwhite grids indicate 95% statistical significance based on bootstrapping.

1997/98 there was a constructive interference between MJO and ENSO signal, while such interference did not occur during the El Niño event in 2015/16. Indicated hereby is that monthly forecast precipitation skill is augmented during ENSO events on average, and not necessarily during each and every ENSO year. That result is a consequence of the well-known modest signal-to-noise ratio of North American precipitation variability, even when ENSO forcing is strong (e.g., [Kumar and Hoerling 1995, 1998a](#)).

4. Conclusions

a. Summary

The principal goal of this study was to attribute the time/space variations in the skill of North American monthly precipitation predictions to slowly evolving ocean surface boundary states and to faster time-scale initial atmospheric weather states. The analysis focused on the week-3 to week-6 window, which bridges a predictability gap spanning atmospheric weather-driven sources in the first two weeks to boundary-related sources and climate-driven variations beyond about a month. Our study was also motivated by the hitherto unexplained but distinctive pattern of historical operational skill of *monthly precipitation forecasts* (issued at 2-week lead) at NOAA's Climate Prediction Center ([Fig. 1](#)). The performance averaged over all months has been characterized by modest skill over the southern United States and an absence of skill over large portions of the central United States. Our study thus addressed the questions whether the particular skill pattern achieved in operational forecasts is characteristic of all seasons, whether it is conditional on particular boundary states such as ENSO, and whether regions exhibiting no forecast skill are inherently unpredictable.

The attribution of skill sources was pursued via a side-by-side comparison of initialized ensemble prediction systems having reforecasts spanning 1999–2015, with a parallel set of uninitialized AMIP simulations of the same dynamical models but constrained only by SST, sea ice, and atmospheric chemical composition variations. For North American and annual averages, weekly precipitation reforecast skill was shown to decline rapidly during the first several weeks. It was concluded that skill for week-3 verifications were largely determined by initial weather state information whereas skill for week-6 verifications arose mostly from boundary condition information. Precipitation skill averaged for this week-3 to week-6 period (referred to as the monthly skill herein) thus originated from a blend of both sources. Importantly, our results demonstrated that the strong regionality and seasonality in forecast performance were significantly influenced by boundary forcing constraints. This was especially evident for the maximum monthly skill achieved in the reforecasts. This maximum skill pattern, having greatest magnitudes over western North America and mostly occurring during the cold half of the annual cycle (except over Central America, which exhibited a summer maximum in skill) were reproduced in AMIP simulations. Collectively, these results provide evidence for a physical basis for the reforecast skill patterns and

suggest, despite sources of modeling and sampling uncertainty, that these patterns are unlikely to originate from random noise.

This important role of boundary forcing in determining monthly precipitation skill was shown to originate mostly from tropical SST influences, especially those related to ENSO. Indeed, our result indicating a strong seasonal cycle for North American monthly precipitation skill can be at least partly explained by the known seasonality in the ENSO-related North American teleconnections (e.g., [Kumar and Hoerling 1998a](#)). Reforecast skill for tropical Pacific monthly rainfall was found to be high (correlations of 0.6 compared to 0.3 for local North American maxima), and many skill features can be replicated by the skill of AMIP simulations. These results alone indicated that teleconnections linked to ENSO forcing (e.g., [Ropelewski and Halpert 1986](#)) represent an important attributable source for North American monthly precipitation skill. Indeed, we constructed the composite precipitation rate with ENSO (El Niño minus La Niña) for the observations and IFS reforecast model over four seasons ([Fig. S3](#)) and found that the precipitation high skill regions in [Fig. 4](#) generally collocate with the ENSO teleconnections, the latter being well simulated by the models. These teleconnections are initiated by the tropical precipitation response to the state of ENSO, thus indicating that a skillful prediction of the tropical precipitation renders a more skill prediction of North American precipitation. This link was further verified via analysis of reforecast skill calculated for subsamples of ENSO-active years only, the skill patterns of which were nearly identical to those in AMIP simulations. Not all the reforecast skill originated from such boundary forcing, however, and considerable skill in North American precipitation was shown to also exist during ENSO-neutral years. The causes for these latter sources were not further explored, though it was noted that reforecasts, contrary to AMIP, had considerable tropical precipitation skill over the Indo-west Pacific warm pool regions. Suggested hereby was an ability to skillfully capture aspects of monthly rainfall variability linked to tropical intraseasonal variability, which in turn may contribute to North American precipitation skill as suggested in previous studies (e.g., [Nardi et al. 2020](#)). Land surface processes have been suggested to enhance the S2S predictability especially in summer (e.g., [Koster et al. 2011](#); [Dirmeyer et al. 2018](#)) and could be an additional skill source. It is beyond the scope of this study to isolate the role of land surface conditions. In that regard, it should be noted that our finding of low overall reforecast skill over large portions of the Great Plains in spring and summer is not necessarily evidence for an absence of land surface constraints on precipitation. Prediction of the atmospheric rivers ([Mundhenk et al. 2018](#); [DeFlorio et al. 2019](#)) is also important for the precipitation skill over U.S. West Coast. Additionally, recent studies have suggested other oceanic skill sources for the summer U.S. precipitation related to a Pacific–Atlantic interbasin SST anomaly contrast ([Kim et al. 2020](#)). [Malloy and Kirtman \(2020\)](#) found that a strengthened Caribbean low-level jet (LLJ), negative PNA teleconnection, El Niño, and a negative Atlantic multidecadal variability each have a relatively strong relationship with a

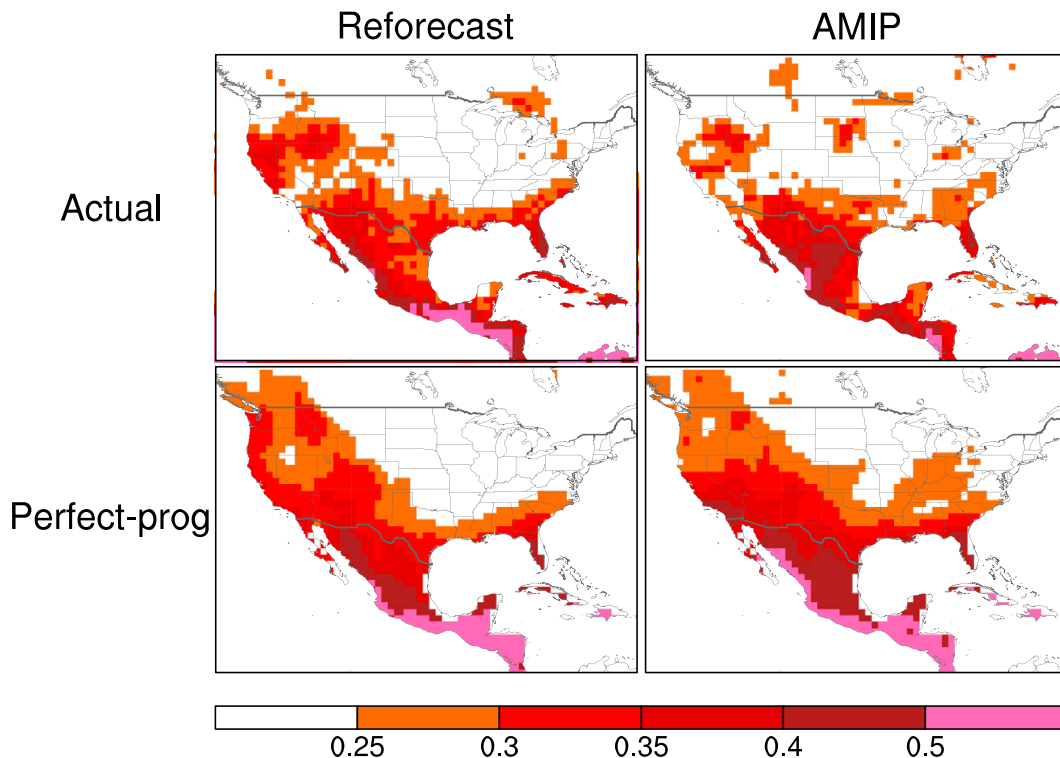


FIG. 12. (top) Maximum of the weeks-3–6 precipitation skill over 12 consecutive 3-month windows for (left) the reforecast systems and (right) their corresponding AMIP simulations. (bottom) As in the top panels, but from the perfect-prognostic approach. The skill is calculated based on equally weighted multimodel mean. For the perfect-prognostic, maximum skill is first calculated based on individual members of each mode, then averaged over all ensemble members of each model and over four models.

strengthened Great Plain LLJ and its associated precipitation. [Jong et al. \(2021\)](#) revealed an observed summer relationship between tropical Pacific El Niño transitioning to La Niña events and decreased precipitation over the Midwest via wave trains. Such a relationship, however, is absent in the SST-forced experiments and is weaker in the North American Multimodel Ensemble models ([Jong et al. 2021](#)). While these factors are in principal operative in the reforecasts examined herein, further study is needed to understand how much each of these may contribute to S2S precipitation skill.

b. Discussion

Does monthly precipitation reforecast skill achieved in the dynamical systems studied herein represent an upper bound—for which the maximum skill in any particular region and season was shown to explain only 30% (corresponding to a correlation of about 0.5; [Fig. 6](#)) of the monthly precipitation variance?

We discuss this issue both in the context of the overall North American precipitation skill for week 3 to week 6, and more specifically, for the spatial pattern of maximum skill whose striking feature is a “skill desert” over the central United States. It should be recognized that inquiries into skill limits are not entirely well posed, for a variety of reasons. For instance, the results of this paper demonstrated the considerable sampling variability in skill that arose from using

different verification periods alone. Further, the models used herein themselves have limitations, linked in part to biases and errors in the assimilation of initial states from which the predictions are launched, in their representation of physical processes relevant for precipitation (e.g., convection overall and the mesoscale organization of rain systems during the warm season over the central United States especially), and likely also in their sensitivity to boundary forcing. Additionally, before discussing skill limits, it is important to reiterate that the focus of this paper is on the *physical origins* in particular ENSO for North American precipitation forecast skill. It has not sought to optimize the skill for the systems diagnosed, which alone can enhance skill as has been shown to arise via multimodel ensemble weighting and logistic regression methods ([Vigaud et al. 2017](#)). Nor has this study explored skill improvements that can arise from other statistical methods such as postprocessing and calibration (e.g., [Bauer et al. 2015](#); [Hamill et al. 2017](#)). Rather, the spirit of the question posed above concerns the underlying nature of possible limits on skill and its spatial structure, and the physical reasons why they exist in this generation of forecast systems.

Concerning the appearance of large portions of the continental United States devoid of statistically significant skill, we synthesize the results of [Fig. 6](#) by generating a multimodel average, shown in [Fig. 12](#) (top row), which compares the multimodel maximum skill of reforecasts (left) and AMIP (right).

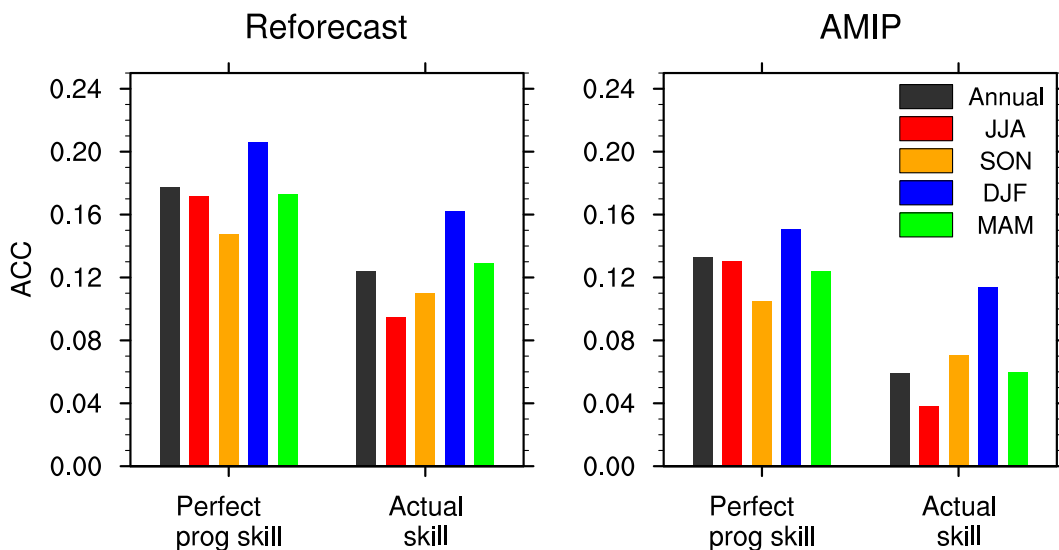


FIG. 13. Monthly precipitation skill averaged over North American land based on (left) perfect prognostics and actual skill for the reforecast and (right) corresponding AMIP simulations.

As seen for the individual models in section 3, the absence of central U.S. reforecast skill in the multimodel average is mimicked by the pattern of AMIP simulation skill. The latter skill originates from a sensitivity to SSTs, the patterns of which indicate high sensitivity over southern and western North America but low sensitivity over the interior continent. Thus, the apparent absence of appreciable ocean constraints on precipitation over significant portions of the United States constitutes one physical argument for skill deserts, combined with the prior indications that skill tends to be more limited (though not entirely absent) during ENSO-neutral years. But the dynamical systems used herein are not without their biases, and though we tried to reduce such effects by diagnosing skill in different systems, the overarching question remains: how does one understand the skill of monthly precipitation predictions in a theoretical framework of maximum achievable skill? To address this, we estimate the maximum achievable skill based on a perfect-prognostic (perfect-prog) approach.

Figure 12 (bottom row) also presents an analysis of the so-called perfect-prog skill for the reforecast (left) and the AMIP (right). In this analysis, each member of a particular model ensemble is selected as a proxy “observation,” and the ensemble mean of the remaining members serves as the prediction of that single model member. This is permuted such that each member serves as an observation; identical procedures are then applied to the reforecasts and AMIP. Perhaps not surprising, the maximum value of the perfect-prog skill (for both reforecasts and AMIP) is greater than the maximum value of the actual skill, though we note that the differences are small and are within a sampling distribution of the permuted perfect-prog samples (not shown). The key result is that actual and perfect-prog skill patterns are nearly identical. In particular, there is very little skill over large portions of the central United States. This skill desert thus is a robust feature

of the actual skill, the perfect-prog skill, and its forecast manifestation is reproduced in AMIP simulations. The interpretation thus is that a skill desert arises over the central United States, on average, owing to an absence of a boundary forced sensitivity of monthly precipitation, together with limited skill from contributions by initial weather states (though with recognition of possible land surface effects alluded to previously).

To further discuss the question of maximum achievable skill, Fig. 13 summarizes the North America average for the reforecast and AMIP actual skills for each season, with a comparison to their perfect-prog counterpart. There is modest increase in skill for the latter, which is broadly consistent between reforecasts and AMIP. It is also interesting to note that the difference between reforecast and AMIP actual skill is about the same as the corresponding difference between the perfect prog skill. In other words, for both estimates of skill, the increased skill arising from contributions of initial conditions is about the same. Irrespective of different approaches for skill estimates, it is important to recognize the small magnitudes of these correlations, which in most instances are less than 0.2 correlation skill.

The perfect-prog skill for reforecasts provides an estimate of the maximum achievable skill. Notwithstanding caveats regarding possible model dependency of these theoretical skill estimates, the spatial structure of perfect-prog skill in reforecasts (Fig. 12, bottom left), indicates a regional preference for skill that is largely constrained by the ENSO influence, and therefore, the skill desert in the Midwest may not be an artifact of model bias. An interesting question to pursue further would be that although we know the physical basis for spatial variations in ENSO influences over the United States (which does not dictate high skill over the Midwest), why does the influence of initial conditions also not yield skill in reforecasts over the Midwest? Is it because the influence of initial

conditions for longer-lead forecasts is conditioned toward improving skill over the regions that are influenced by the boundary conditions to begin with?

What then are the implications for operational practices and for prospects of monthly precipitation forecasts over the North American region? The NOAA operational practice of monthly forecasts consists of two products—one released at 15-day lead the other at 0 lead. This paper has probed the skill of the former by addressing dynamical model prediction skill averaged over week 3 to week 6. The latter product would be equivalent to a prediction of week 1 to week 4. Though not shown in detail, our analysis of weekly evolving skill (see Fig. 2) clearly indicated much greater precipitation skill during weeks 1 and 2 than during weeks 5 and 6. Substantial improvement in monthly precipitation skill would thus arise from reducing the lead time of issuance (Fig. S2). This should be compared to the only modest theoretical skill improvements that are implied by the perfect prog analysis for week 3 to week 6. Finally, we stress that the monthly precipitation skill is highly conditional, being particularly elevated when ENSO boundary forcings are operative. Likewise, skill is highly regional, and is again heavily constrained by the spatial footprint of the ENSO signal and its time/space variability.

Acknowledgments. We thank Jon Eischeid, Xiao-Wei Quan, Lesley Smith, Yan Wang, and Tao Zhang at NOAA's Physical Sciences Laboratory (PSL) for the help with downloading the reforecasts and AMIP data. We also appreciate the helpful comments from Xiao-Wei Quan (NOAA/PSL) and Kirsten Mayer (CSU) on an earlier version of this paper. Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological and Environmental Research (BER) via National Science Foundation IA 1947282, and under Award DE-SC0022070. This work also was supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement 1852977.

Data availability statement. IFS reforecast data are available from the ECMWF's international S2S database at <https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/origin=ecmf/type=cf/> (Vitart et al. 2012, 2017). CESM1 and CFSv2 reforecasts are available from the International Research Institute (IRI)'s SubX data library (<http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/>). GEFSv12 reforecast data are available at AWS through <https://registry.opendata.aws/noaa-gefs-reforecast/>. CESM2 reforecast outputs are available for download from the NCAR Climate Data Gateway and can be accessed via the DOIs: <https://doi.org/10.5065/0s63-m767>. CAM6 simulations are available for download from the NCAR Climate Data Gateway at https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm2.cam6.prescribed_sst_amip.html. IFS-AMIP simulations are obtained from the HighResMIP (Haarsma et al. 2016). GFSv2 and CAM5 simulations are available at NOAA Physical Science Laboratory (PSL)'s FACility for Weather and Climate

Assessments (FACTS; Murray et al. 2020) at <https://psl.noaa.gov/repository/facts/>. GPCP daily analysis data are available at NOAA's NCEI (<https://www.ncei.noaa.gov/products/climate-data-records/precipitation-gpcp-daily>) via doi: <https://doi.org/10.7289/V5RX998Z>. NOAA's ONI data are available at <https://psl.noaa.gov/data/correlation/oni.data>.

REFERENCES

- Adler, R. F., G. Gu, M. Sapiiano, J.-J. Wang, and G. J. Huffman, 2017: Global precipitation: Means, variations and trends during the satellite era (1979–2014). *Surv. Geophys.*, **38**, 679–699, <https://doi.org/10.1007/s10712-017-9416-4>.
- Albers, J. R., and M. Newman, 2021: Subseasonal predictability of the North Atlantic Oscillation. *Environ. Res. Lett.*, **16**, 044024, <https://doi.org/10.1088/1748-9326/abe781>.
- Arcodia, M. C., B. P. Kirtman, and L. S. P. Siqueira, 2020: How MJO teleconnections and ENSO interference impacts U.S. precipitation. *J. Climate*, **33**, 4621–4640, <https://doi.org/10.1175/JCLI-D-19-0448.1>.
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, <https://doi.org/10.1038/nature14956>.
- Danabasoglu, G., and Coauthors, 2020: The Community Earth System Model Version 2 (CESM2). *J. Adv. Model. Earth Syst.*, **12**, e2019MS001916, <https://doi.org/10.1029/2019MS001916>.
- DeFlorio, M. J., and Coauthors, 2019: Experimental subseasonal-to-seasonal (S2S) forecasting of atmospheric rivers over the western United States. *J. Geophys. Res. Atmos.*, **124**, 11 242–11 265, <https://doi.org/10.1029/2019JD031200>.
- DelSole, T., L. Trenary, M. K. Tippett, and K. Pegion, 2017: Predictability of week-3–4 average temperature and precipitation over the contiguous United States. *J. Climate*, **30**, 3499–3512, <https://doi.org/10.1175/JCLI-D-16-0567.1>.
- Dias, J., S. N. Tulich, M. Gehne, and G. N. Kiladis, 2021: Tropical origins of weeks 2–4 forecast errors during the Northern Hemisphere cool season. *Mon. Wea. Rev.*, **149**, 2975–2991, <https://doi.org/10.1175/MWR-D-21-0020.1>.
- Dirmeyer, P. A., S. Halder, and R. Bombardi, 2018: On the harvest of predictability from land states in a global forecast model. *J. Geophys. Res. Atmos.*, **123**, 13 111–13 127, <https://doi.org/10.1029/2018JD029103>.
- ECMWF, 2014a: IFS documentation— CY40r1. Part III: Dynamics and numerical procedures. ECMWF, 29 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2014/9203-part-iii-dynamics-and-numerical-procedures.pdf>.
- , 2014b: IFS documentation— CY40r1. Part IV: Physical processes. ECMWF, 190 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2014/9204-part-iv-physical-processes.pdf>.
- , 2014c: IFS documentation— Cy40r1. Part VII: ECMWF wave model. ECMWF, 79 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2014/9207-part-vii-ecmwf-wave-model.pdf>.
- Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–56, [https://doi.org/10.1175/1520-0477\(1999\)080<0029:A0OTRO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0029:A0OTRO>2.0.CO;2).
- Haarsma, R. J., and Coauthors, 2016: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, <https://doi.org/10.5194/gmd-9-4185-2016>.
- Hamill, T., E. Engle, D. Myrick, M. Peroutka, C. Finan, and M. Scheuerer, 2017: The U.S. National Blend of Models for

- statistical postprocessing of probability of precipitation and deterministic precipitation amount. *Mon. Wea. Rev.*, **145**, 3441–3463, <https://doi.org/10.1175/MWR-D-16-0331.1>.
- Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, <https://doi.org/10.1175/JCLI-D-16-0836.1>.
- Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate*, **21**, 5145–5153, <https://doi.org/10.1175/2008JCLI2292.1>.
- Johnson, C. N., L. Krishnamurthy, A. T. Wittenberg, B. Xiang, G. A. Vecchi, S. B. Kapnick, and S. Pascale, 2020: The impact of sea surface temperature biases on North American precipitation in a high-resolution climate model. *J. Climate*, **33**, 2427–2447, <https://doi.org/10.1175/JCLI-D-19-0417.1>.
- Jong, B. T., M. Ting, and R. Seager, 2021: Assessing ENSO summer teleconnections, impacts, and predictability in North America. *J. Climate*, **34**, 3629–3643, <https://doi.org/10.1175/JCLI-D-20-0761.1>.
- Kim, D., S.-K. Lee, H. Lopez, G. Foltz, V. Misra, and A. Kumar, 2020: On the role of Pacific–Atlantic SST contrast and associated Caribbean Sea convection in August–October U.S. regional rainfall variability. *Geophys. Res. Lett.*, **47**, e2020GL087736, <https://doi.org/10.1029/2020GL087736>.
- Kim, H., M. A. Janiga, and K. Pegion, 2019: MJO propagation processes and mean biases in the SubX and S2S reforecasts. *J. Geophys. Res. Atmos.*, **124**, 9314–9331, <https://doi.org/10.1029/2019JD031139>.
- Koster, R. D., and Coauthors, 2011: The second phase of the Global Land–Atmosphere Coupling Experiment: Soil moisture contributions to subseasonal forecast skill. *J. Hydrometeorol.*, **12**, 805–822, <https://doi.org/10.1175/2011JHM1365.1>.
- Krishnamurthy, V., and Coauthors, 2021: Sources of subseasonal predictability over CONUS during boreal summer. *J. Climate*, **34**, 3273–3294, <https://doi.org/10.1175/JCLI-D-20-0586.1>.
- Kumar, A., and M. Hoerling, 1995: Prospects and limitations of seasonal atmospheric GCM predictions. *Bull. Amer. Meteor. Soc.*, **76**, 335–345, [https://doi.org/10.1175/1520-0477\(1995\)076<0335:PALOSA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0335:PALOSA>2.0.CO;2).
- , and —, 1998a: Annual cycle of Pacific–North American seasonal predictability associated with the different phases on ENSO. *J. Climate*, **11**, 3295–3308, [https://doi.org/10.1175/1520-0442\(1998\)011<3295:ACOPNA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<3295:ACOPNA>2.0.CO;2).
- , and —, 1998b: Specification of regional sea surface temperatures in atmospheric general circulation model simulations. *J. Geophys. Res.*, **103**, 8901–8907, <https://doi.org/10.1029/98JD00427>.
- , and —, 2000: Analysis of a conceptual model of seasonal climate variability and implications for seasonal predictions. *Bull. Amer. Meteor. Soc.*, **81**, 255–264, [https://doi.org/10.1175/1520-0477\(2000\)081<0255:AOACMO>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0255:AOACMO>2.3.CO;2).
- , M. Chen, and W. Wang, 2011: An analysis of prediction skill of monthly mean climate variability. *Climate Dyn.*, **37**, 1119–1131, <https://doi.org/10.1007/s00382-010-0901-4>.
- , —, and —, 2013: Understanding prediction skill of seasonal mean precipitation over the tropics. *J. Climate*, **26**, 5674–5681, <https://doi.org/10.1175/JCLI-D-12-00731.1>.
- , P. Peng, and M. Chen, 2014: Is there a relationship between potential and actual skill? *Mon. Wea. Rev.*, **142**, 2220–2227, <https://doi.org/10.1175/MWR-D-13-00287.1>.
- Kwon, I., S. English, W. Bell, R. Potthast, A. Collard, and B. Ruston, 2018: Assessment of progress and status of data assimilation in numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **99**, ES75–ES79, <https://doi.org/10.1175/BAMS-D-17-0266.1>.
- L’Heureux, M. L., M. K. Tippett, and E. J. Becker, 2021: Sources of subseasonal skill and predictability in wintertime California precipitation forecasts. *Wea. Forecasting*, **36**, 1815–1826, <https://doi.org/10.1175/WAF-D-21-0061.1>.
- Li, S., and A. W. Robertson, 2015: Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Mon. Wea. Rev.*, **143**, 2871–2889, <https://doi.org/10.1175/MWR-D-14-00277.1>.
- Lucas, F. D., 2017: Weather Research and Forecasting Innovation Act of 2017. Public Law 115–25, H.R. 353, <https://www.congress.gov/bill/115th-congress/house-bill/353/text>.
- Lybarger, N. D., C. S. Shin, and C. Stan, 2020: MJO wind energy and prediction of El Niño. *J. Geophys. Res. Oceans*, **125**, e2020JC016732, <https://doi.org/10.1029/2020JC016732>.
- Malloy, K. M., and B. P. Kirtman, 2020: Predictability of midsummer Great Plains low-level jet and associated precipitation. *Wea. Forecasting*, **35**, 215–235, <https://doi.org/10.1175/WAF-D-19-0103.1>.
- Mariotti, A., P. M. Ruti, and M. Rixen, 2018: Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *npj Climate Atmos. Sci.*, **1**, 4, <https://doi.org/10.1038/s41612-018-0014-z>.
- Meehl, G. A., and Coauthors, 2021: Initialized Earth System prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.*, **2**, 340–357, <https://doi.org/10.1038/s43017-021-00155-x>.
- Mudelsee, M., 2010: *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*. 1st ed. Springer, 474 pp.
- Mundhenk, B. D., E. A. Barnes, E. D. Maloney, and C. F. Baggett, 2018: Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian Oscillation and quasi-biennial oscillation. *npj Climate Atmos. Sci.*, **1**, 20177, <https://doi.org/10.1038/s41612-017-0008-2>.
- Murray, D., and Coauthors, 2020: Facility for Weather and Climate Assessments (FACTS): A community resource for assessing weather and climate variability. *Bull. Amer. Meteor. Soc.*, **101**, E1214–E1224, <https://doi.org/10.1175/BAMS-D-19-0224.1>.
- Nardi, K., C. Baggett, E. Barnes, E. Maloney, D. Harnos, and L. Ciasto, 2020: Skillful all-season S2S prediction of U.S. precipitation using the MJO and QBO. *Wea. Forecasting*, **35**, 2179–2198, <https://doi.org/10.1175/WAF-D-19-0232.1>.
- National Academies of Sciences, Engineering, and Medicine, 2016: *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*. The National Academies Press, 350 pp., <https://doi.org/10.17226/21873>.
- Neale, R. B., and Coauthors, 2012: Description of the NCAR Community Atmosphere Model (CAM 5.0). NCAR Tech. Note NCAR/TN-486+STR, 274 pp., http://www.cesm.ucar.edu/models/cesm1.0/cam/docs/description/cam5_desc.pdf.
- Newman, M., P. D. Sardeshmukh, C. R. Winkler, and J. S. Whitaker, 2003: A study of subseasonal predictability. *Mon. Wea. Rev.*, **131**, 1715–1732, <https://doi.org/10.1175/2558.1>.
- Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100**, 2043–2060, <https://doi.org/10.1175/BAMS-D-18-0270.1>.
- Quan, X.-W., M. Hoerling, B. Lyon, A. Kumar, M. Bell, M. Tippett, and H. Wang, 2012: Prospects for dynamical prediction of meteorological drought. *J. Appl. Meteor. Climatol.*, **51**, 1238–1252, <https://doi.org/10.1175/JAMC-D-11-0194.1>.

- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AIHSAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIHSAS>2.0.CO;2).
- Richter, J. H., and Coauthors, 2020: Subseasonal prediction with and without a well-represented stratosphere in CESM1. *Wea. Forecasting*, **35**, 2589–2602, <https://doi.org/10.1175/WAF-D-20-0029.1>.
- , and Coauthors, 2022: Subseasonal Earth system prediction with CESM2. *Wea. Forecasting*, **37**, 797–815, <https://doi.org/10.1175/WAF-D-21-0163.1>.
- Robertson, A. W., A. Kumar, M. Peña, and F. Vitart, 2015: Improving and promoting subseasonal to seasonal prediction. *Bull. Amer. Meteor. Soc.*, **96**, ES49–ES53, <https://doi.org/10.1175/BAMS-D-14-00139.1>.
- Ropelewski, C. F., and M. S. Halpert, 1986: North American and precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, [https://doi.org/10.1175/1520-0493\(1986\)114<2352:NAPATP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2).
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Scaife, A. A., and D. Smith, 2018: A signal-to-noise paradox in climate science. *npj Climate Atmos. Sci.*, **1**, 28, <https://doi.org/10.1038/s41612-018-0038-4>.
- Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 98–102, <https://doi.org/10.1038/ngeo1698>.
- 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>.
- , and Coauthors, 2022: Advances in the prediction of MJO teleconnections in the S2S forecast systems. *Bull. Amer. Meteor. Soc.*, **103**, E1426–E1447, <https://doi.org/10.1175/BAMS-D-21-0130.1>.
- Sun, L., J. Perlwitz, J. H. Richter, M. P. Hoerling, and J. W. Hurrell, 2020: Attribution of NAO predictive skill beyond 2 weeks in boreal winter. *Geophys. Res. Lett.*, **47**, e2020GL090451, <https://doi.org/10.1029/2020GL090451>.
- Titchner, H. A., and N. A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864–2889, <https://doi.org/10.1002/2013JD020316>.
- Vigaud, N., A. Robertson, and M. Tippett, 2017: Multimodal ensembling of subseasonal precipitation forecasts over North America. *Mon. Wea. Rev.*, **145**, 3913–3928, <https://doi.org/10.1175/MWR-D-17-0092.1>.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, <https://doi.org/10.1002/qj.2256>.
- , 2017: Madden–Julian Oscillation prediction and teleconnections in the S2S database. *Quart. J. Roy. Meteor. Soc.*, **143**, 2210–2220, <https://doi.org/10.1002/qj.3079>.
- , A. Robertson, and D. Anderson, 2012: Sub-seasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *WMO Bull.*, **61**, 23–28.
- , and Coauthors, 2017: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Wang, J., H. Kim, and M. J. DeFlorio, 2022: Future changes of PNA-like MJO teleconnections in CMIP6 models: Underlying mechanisms and uncertainty. *J. Climate*, **35**, 3459–3478, <https://doi.org/10.1175/JCLI-D-21-0445.1>.
- Wang, L., and A. W. Robertson, 2019: Week 3–4 predictability over the United States assessed from two operational ensemble prediction systems. *Climate Dyn.*, **52**, 5861–5875, <https://doi.org/10.1007/s00382-018-4484-9>.
- Wang, S., A. Anichowski, M. K. Tippett, and A. H. Sobel, 2017: Seasonal noise versus subseasonal signal: Forecasts of California precipitation during the unusual winters of 2015–2016 and 2016–2017. *Geophys. Res. Lett.*, **44**, 9513–9520, <https://doi.org/10.1002/2017GL075052>.
- Whitaker, J. S., and K. M. Weickmann, 2001: Subseasonal variations of tropical convection and week-2 prediction of wintertime western North American rainfall. *J. Climate*, **14**, 3279–3288, [https://doi.org/10.1175/1520-0442\(2001\)014<3279:SVOTCA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<3279:SVOTCA>2.0.CO;2).
- White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, <https://doi.org/10.1002/met.1654>.
- Yeager, S. G., and Coauthors, 2018: Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **99**, 1867–1886, <https://doi.org/10.1175/BAMS-D-17-0098.1>.
- Zheng, C., E. Kar-Man Chang, H. Kim, M. Zhang, and W. Wang, 2018: Impacts of the Madden–Julian Oscillation on storm-track activity, surface air temperature, and precipitation over North America. *J. Climate*, **31**, 6113–6134, <https://doi.org/10.1175/JCLI-D-17-0534.1>.
- Zhou, S., M. L’Heureux, S. Weaver, and A. Kumar, 2012: A composite study of the MJO influence on the surface air temperature and precipitation over the continental United States. *Climate Dyn.*, **38**, 1459–1471, <https://doi.org/10.1007/s00382-011-1001-9>.
- Zhou, X., Y. Zhu, D. Hou, and D. Kleist, 2016: A comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP Global Ensemble Forecast System. *Wea. Forecasting*, **31**, 2057–2074, <https://doi.org/10.1175/WAF-D-16-0109.1>.
- , —, —, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, <https://doi.org/10.1175/WAF-D-17-0023.1>.
- Zhu, Y., and Coauthors, 2018: Toward the improvement of sub-seasonal prediction in the National Centers for Environmental Prediction Global Ensemble Forecast System. *J. Geophys. Res. Atmos.*, **123**, 6732–6745, <https://doi.org/10.1029/2018JD028506>.