

# *Subseasonal prediction performance for austral summer South American rainfall*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Klingaman, N. P. ORCID: <https://orcid.org/0000-0002-2927-9303>, Young, M., Chevuturi, A. ORCID: <https://orcid.org/0000-0003-2815-7221>, Guimaraes, B., Guo, L., Woolnough, S., Coelho, C. A. S., Kubota, P. Y. and Holloway, C. E. ORCID: <https://orcid.org/0000-0001-9903-8989> (2021) Subseasonal prediction performance for austral summer South American rainfall. *Weather and Forecasting*, 36 (1). pp. 147-169. ISSN 0882-8156 doi: <https://doi.org/10.1175/WAF-D-19-0203.1>  
Available at <http://centaur.reading.ac.uk/94704/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1175/WAF-D-19-0203.1>

Publisher: American Meteorological Society

including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## Subseasonal Prediction Performance for Austral Summer South American Rainfall

NICHOLAS P. KLINGAMAN,<sup>a,b</sup> MATTHEW YOUNG,<sup>a,b</sup> AMULYA CHEVUTURI,<sup>a,b</sup> BRUNO GUIMARAES,<sup>c</sup> LIANG GUO,<sup>a,b</sup> STEVEN J. WOOLNOUGH,<sup>a,b</sup> CAIO A. S. COELHO,<sup>c</sup> PAULO Y. KUBOTA,<sup>c</sup> AND CHRISTOPHER E. HOLLOWAY<sup>b</sup>

<sup>a</sup> National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom

<sup>b</sup> Department of Meteorology, University of Reading, Reading, United Kingdom

<sup>c</sup> Centre for Weather Forecast and Climate Studies, National Institute for Space Research, Cachoeira Paulista, São Paulo, Brazil

(Manuscript received 4 October 2019, in final form 5 October 2020)

**ABSTRACT:** Skillful and reliable predictions of week-to-week rainfall variations in South America, two to three weeks ahead, are essential to protect lives, livelihoods, and ecosystems. We evaluate forecast performance for weekly rainfall in extended austral summer (November–March) in four contemporary subseasonal systems, including a new Brazilian model, at 1–5-week leads for 1999–2010. We measure performance by the correlation coefficient (in time) between predicted and observed rainfall; we measure skill by the Brier skill score for rainfall terciles against a climatological reference forecast. We assess unconditional performance (i.e., regardless of initial condition) and conditional performance based on the initial phase of the Madden–Julian oscillation (MJO) and El Niño–Southern Oscillation (ENSO). All models display substantial mean rainfall biases, including dry biases in Amazonia and wet biases near the Andes, which are established by week 1 and vary little thereafter. Unconditional performance extends to week 2 in all regions except for Amazonia and the Andes, but to week 3 only over northern, northeastern, and southeastern South America. Skill for upper- and lower-tercile rainfall extends only to week 1. Conditional performance is not systematically or significantly higher than unconditional performance; ENSO and MJO events provide limited “windows of opportunity” for improved S2S predictions that are region and model dependent. Conditional performance may be degraded by errors in predicted ENSO and MJO teleconnections to regional rainfall, even at short lead times.


**KEYWORDS:** South America; ENSO; Madden-Julian oscillation; Rainfall; Forecast verification/skill; Intraseasonal variability

### 1. Introduction

Subseasonal to seasonal (S2S) variations in local- and regional-scale rainfall present considerable hazards in the tropics, through floods and meteorological droughts that reduce agricultural yields, limit hydropower generation, and degrade human and ecosystem health. In monsoonal regions where the seasonal cycle is strong and assumed to be predictable, crop sowing dates are tied to climatological rainfall onset. Delays to the onset, or “false onsets” in which breaks in the rains immediately follows onset, cause seeds to fail to germinate and lead to substantial agricultural losses (e.g., Marteau et al. 2011). Conversely, flooding after planting substantially reduces yields; heavy rain during harvest can delay harvests or damage crops (e.g., Coomes et al. 2016).

Historically, S2S forecasts, usually made from two to eight weeks in advance, were judged to be less useful than numerical weather predictions (NWP)—which provide shorter-range (1–15 days) initial-condition driven predictions at daily scales—or seasonal forecasts—which provide longer-range (3–6 months) boundary-condition driven predictions at monthly scales (e.g., Hudson et al. 2011; Vitart et al. 2012). The perceived lack of

utility stemmed from poor S2S performance for weekly averages required by forecast users and targeted by producing centers, to fill the lead-time and prediction-scale gap between NWP and seasonal forecasts. Weekly rainfall variations have proven difficult to predict at the 2–3-week lead times required by users (e.g., farmers, hydroelectric dam managers) to mitigate damage (e.g., Laux et al. 2008; Moron et al. 2009). S2S prediction difficulties have been ascribed to the failure of forecast models to represent key subseasonal phenomena, such as the Madden–Julian oscillation (MJO) or the related boreal summer intraseasonal oscillation (Neena et al. 2014; Lee et al. 2015), and their teleconnections to tropical and midlatitude rainfall and circulation. However, recent advances in model resolution, physics and data assimilation have improved S2S prediction quality, including for the MJO and its global teleconnections (Vitart 2017; Vitart and Robertson 2018) and for the onset and cessation of major monsoons (Bombardi et al. 2017), such that many sectors are reconsidering the potential social and economic benefits of S2S predictions. The successful application of S2S forecasts requires careful evaluation of contemporary S2S prediction performance. Such evaluation includes whether there are “windows of opportunity” for improved S2S performance, based on regional- or large-scale atmospheric circulations, as in seasonal forecasts during

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Nicholas P. Klingaman, nicholas.klingaman@ncas.ac.uk

DOI: 10.1175/WAF-D-19-0203.1

© 2021 American Meteorological Society



This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

El Niño–Southern Oscillation (ENSO) events (e.g., [Goddard and Dilley 2005](#)).

In South America, rainfall extremes have devastating costs for human lives and livelihoods. In Brazil alone, extreme rainfall events in 1979–2013 are estimated to have caused approximately 5000 deaths and billions of dollars in damage ([Vörösmarty et al. 2013](#); [Hirata and Grimm 2018](#)). In Amazonia, approximately 70% of annual precipitation falls in 3–5 spells of intense precipitation that last 4–15 days ([Rao and Hada 1990](#)), associated with intense moisture convergence ([de Oliveira Vieira et al. 2013](#)). These heavy rains are linked to atmospheric convection across scales, but particularly to the presence, strength and location of the South Atlantic convergence zone (SACZ; [Carvalho et al. 2010](#)). In turn, SACZ variability is linked on decadal scales to Atlantic and Pacific sea surface temperature variability, modulated by surface–atmosphere feedbacks (e.g., [Robertson and Mechoso 2000](#); [Grimm et al. 2007](#); [Grimm and Saboia 2015](#)); on interannual scales to ENSO (e.g., [Grimm and Tedeschi 2009](#)); on intraseasonal scales to the MJO (e.g., [Grimm 2019](#)); and on synoptic scales to midlatitude Rossby wavetrains (e.g., [Hirata and Grimm 2017](#)), among other phenomena. The number and diversity of these relationships illustrate the challenge in understanding and predicting South American rainfall variability. Of particular relevance to this study is the ENSO teleconnection, which typically suppresses rainfall across equatorial South America in El Niño years and enhances rainfall in La Niña years. These signals are often of the opposite sign in subtropical South America, including heavily populated regions in southeastern and southern Brazil, Uruguay, and northeastern Argentina. The ENSO teleconnection is a key source of seasonal predictability and prediction skill, particularly for northern and southeastern South America ([Bombardi et al. 2018](#)), but its influence on S2S predictions remains unclear.

Subseasonal rainfall variability in South America has been connected to tropical and midlatitude influences. The MJO ([Madden and Julian 1971](#); [Zhang 2005](#)) is the leading tropical influence on intraseasonal scales: large-scale (zonal wavenumber 1–3), quasi-periodic (30–70 day) variability in tropical convection and associated zonally overturning circulation that propagates east along the equator, typically from the Indian Ocean through the Pacific to the Western Hemisphere. Active phases of the MJO over tropical South America cause stronger and more persistent SACZ rainfall extremes, but also suppress rainfall over subtropical South America; active MJO phases over subtropical South America cause opposite-signed signals (e.g., [Carvalho et al. 2004](#); [Grimm 2019](#)). Subseasonal midlatitude influences on rainfall come primarily through Rossby waves that propagate equatorward into tropical South America, draw moisture from Amazonia and initiate SACZ convection over land ([Grimm and Silva Dias 1995](#); [Liebmann et al. 1999](#)). The most intense rainfall extremes are linked to coincidence and superposition of these tropical and midlatitude influences ([Hirata and Grimm 2017](#)). For example, MJO convection in the Pacific can initiate an extratropical wave train that propagates into the midlatitudes, then around South America, and which eventually triggers heavy SACZ rainfall ([Grimm 2019](#)). Many of these influences modulate the

regional-scale meridional overturning circulation connecting tropical and subtropical South America, creating opposite-signed rainfall anomalies between these regions (e.g., [Gan et al. 2004](#); [Cavalcanti et al. 2017](#)).

Despite much research into mechanisms of South American subseasonal rainfall variability with observations, reanalysis data and model simulations, few studies have evaluated contemporary S2S forecasts of South American weekly rainfall and its variability, or the teleconnections from major large-scale phenomena such as the MJO. In a global-scale analysis, [Li and Robertson \(2015\)](#) found that the European Centre for Medium-range Weather Forecasts (ECMWF) S2S model showed high performance—measured by correlation coefficients with observed rainfall above 0.2—for forecasts 1–3 weeks ahead over northeastern Brazil. [Coelho et al. \(2018\)](#) evaluated forecasts of autumn rainfall and proposed a verification framework for South American precipitation subseasonal predictions, finding that ECMWF performed well over northeastern Brazil. [de Andrade et al. \(2018\)](#) evaluated the ability of all S2S project models to reproduce global austral summer subseasonal rainfall variability and identified biases associated with model deficiencies in representing atmospheric teleconnections. [Hirata and Grimm \(2017\)](#) estimated that the U.S. National Centers for Environmental Prediction (NCEP) model could represent rainfall extremes up to two weeks ahead, but this result was based on only a handful of case studies in 2010–11, and achieving useful skill required statistical calibration.

We investigate S2S prediction quality for South American weekly rainfall in four recent forecast models, including conditional performance evaluation based on MJO and ENSO phases to understand whether large-scale variability improves S2S forecasts. Our conditional evaluation is distinct from [de Andrade et al. \(2018\)](#), as we evaluate total forecast rainfall in ENSO and MJO phases, rather than removing the linear effect of those phenomena first. Our regional focus on South America, our focus on austral summer, the major wet season in most of South America, and our inclusion of a new Brazilian model distinguish our study from previous studies that evaluated S2S forecasts for austral autumn ([Coelho et al. 2018](#)), or for all seasons ([Pegion et al. 2019](#)), or at global scales at which regional features are difficult to distinguish ([Li and Robertson 2015](#); [de Andrade et al. 2018](#)). We describe the S2S models, verifying rainfall dataset and analysis techniques ([section 2](#)); assess unconditional and conditional performance ([section 3](#)); discuss the broader context and limitations of our findings ([section 4](#)); and summarize our conclusions ([section 5](#)).

## 2. Data and methods

### a. Subseasonal reforecasts

We use subseasonal reforecasts from the S2S Prediction Project database ([Vitart et al. 2017](#)), as well as from the Brazilian Global Atmospheric Model version 1.2 (BAM-1.2; [Guimarães et al. 2020](#)) developed at the Centre for Weather Forecast and Climate Studies (CPTEC). As the S2S database comprises models with various reforecast start dates, lengths, ensemble sizes and periods, we focus on three models to reduce

TABLE 1. For each model analyzed, the reforecast period, reforecast ensemble size, and reforecast initialization frequency. A list of BAM initialization dates can be found in Table 1 of [Guimarães et al. \(2020\)](#).

Model	Period	Original ensemble	Frequency
ECMWF	1997–2016 <sup>a</sup> or 1998–2017 <sup>b</sup>	11 members	Every Monday and Thursday
NCEP	1999–2010	4 members	Daily
UKMO	1993–2015	7 members	1st, 9th, 17th, and 25th of month
BAM	1999–2010	11 members	On the ECMWF initial date closest to the 1st and 15th of the month

<sup>a</sup> For reforecasts performed in 2017.

<sup>b</sup> For reforecasts performed in 2018.

the effects of these variations on our results: ECMWF ([Vitar 2014](#)), the Met Office (UKMO; [MacLachlan et al. 2015](#)), and NCEP ([Saha et al. 2014](#)). ECMWF and UKMO perform reforecasts “on the fly” (i.e., in near-real time, alongside the operational forecasts). We use reforecasts performed during May 2017–April 2018, which corresponds to ECMWF Cycles 43r1 and 43r3 and the UKMO Global Coupled 2.0 configuration. NCEP and BAM perform a “fixed” (frozen) reforecast set. We use NCEP reforecasts from the Climate Forecast System, version 2; we use BAM reforecasts from BAM-1.2 ([Guimarães et al. 2020](#)). We analyze reforecasts valid during extended austral summer, November–March (NDJFM), for November 1999–March 2010. We use only this common period, even though some models have longer reforecasts available (see below and [Table 1](#)).

Each modeling center employs a different strategy for its reforecast ensemble ([Table 1](#)). For example, NCEP has more frequent initializations (daily) but a smaller ensemble (four members), whereas UKMO has fewer initializations (every 6–8 days) but a larger ensemble (seven members). There are two main approaches to compare models with variations in ensemble size and initialization frequency; neither approach is perfect or fair. The first is to evaluate each model for its own initialization dates. For models with frequent initializations, but small ensembles, this approach produces a larger sample of forecasts but smaller ensembles, which may reduce deterministic or probabilistic forecast performance (e.g., [de Andrade et al. 2018](#)). The second approach is to create lagged ensembles, by combining, for a model with more frequent initializations, some or all reforecasts made between the dates of a model with less frequent initializations. For instance, to create an NCEP lagged ensemble corresponding to the UKMO ensemble initialized on 9 November, we would combine the NCEP reforecasts initialized 2–9 November, to create a 32-member ensemble (8 initializations  $\times$  4 members per initialization). We would then evaluate the NCEP lagged ensemble and the UKMO ensemble from 9 November onward. Relative to the first approach, using lagged ensembles grows the NCEP ensemble considerably (from 4 to 32 members), but at the potential cost of performance, because “day 1” corresponds to leads of 1–8 days in the lagged ensemble.

We create ECMWF and NCEP lagged ensembles that correspond to the UKMO initialization dates, by combining reforecasts made between the UKMO initialization dates (e.g., between 2 and 9 November for the 9 November UKMO

initialization). We use the UKMO dates, rather than the less-frequent BAM dates, because the  $\approx$ 15-day spacing of the BAM dates would cause ECMWF and NCEP to lose up to two weeks’ lead time, a considerable disadvantage for S2S reforecasts. UKMO and BAM are analyzed with respect to their own initialization dates. Thus, ECMWF, NCEP and UKMO are analyzed for common validity periods (as weekly means), while BAM is not. This results in a different verification sample size for BAM (88–110 samples, depending on lead time) than the other three models (220 samples). We discuss this issue further in [section 4](#). We create the ECMWF and NCEP lagged ensembles using an 8-day window prior to and including the UKMO initialization date—rather than, for example, using a range of dates centered on the UKMO initialization date—to mimic a real-time operational procedure. For ECMWF, we use the last three initializations on or before the UKMO date to form a 33-member lagged ensemble; for NCEP, we use the last eight initializations on or before the UKMO date to form a 32-member lagged ensemble.

Our four ensembles have different sizes: UKMO (seven members), ECMWF (33 members), NCEP (32 members) and BAM (11 members). The optimum size of lagged ensembles is an area of active research (e.g., [DelSole et al. 2017](#); [Trenary et al. 2017, 2018](#)). At short leads (e.g., 1–2 weeks) when the signal-to-noise ratio is high and predictability arises from initial conditions, ensembles lagged over shorter windows may outperform those lagged over longer windows. Conversely, at long leads (e.g., >2 weeks) when the signal-to-noise ratio is low and predictability arises from slowly evolving atmospheric or oceanic conditions, ensembles lagged over longer windows (and hence with more members) may outperform those lagged over shorter windows (and hence with fewer members). S2S requires a balance between short- and long-term performance. Our 8-day window for ECMWF and NCEP mimics the strategy used for NCEP in the operational North American Multimodel Ensemble ([Kirtman et al. 2014](#)). Similarly, the Subseasonal Experiment (SubX) project compared models by lagging ensembles over a 7-day window ([Pegion et al. 2019](#)). [Trenary et al. \(2017\)](#) and [Trenary et al. \(2018\)](#) found that NCEP S2S ensembles lagged over a 5–10-day window showed optimum performance for MJO and ENSO, respectively.

We test the sensitivity to our lagged-ensemble strategy by building alternative NCEP and ECMWF ensembles, using for ECMWF only the initialization closest to (but not later than) the UKMO date, while for NCEP we use a 2-day window to

create an eight-member lagged ensemble. This produces more similarly sized ensembles for UKMO (seven members), NCEP (eight members) and ECMWF (11 members). The verification methods for these alternative ensembles are identical to those described below for the primary NCEP (32 members) and ECMWF (33 members) lagged ensembles. Unless otherwise mentioned, all analysis uses the primary, larger lagged ensembles.

Note that our NDJFM analysis period refers to the validity time of the forecasts, not the initialization time. All model data are provided and analyzed on a common  $1.5^\circ \times 1.5^\circ$  horizontal grid, although the original resolutions differ considerably. Many diagnostics and metrics are computed as a function of forecast lead time, expressed in weeks. For the ECMWF and NCEP lagged ensembles, lead time refers to the time since the UKMO initialization date to which the lagged ensemble is referenced (e.g., to days since 9 November for the 2–9 November lagged example, using the example above). “Week 1” refers to lead times 1–7 days, “week 2” to lead times 8–14 days, and so on.

### b. Indices

To analyze conditional performance based on ENSO phase, we divide the 1999–2010 period into quartiles based on the monthly oceanic Niño index from the NCEP Climate Prediction Center ([https://origin.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ONI\\_v5.php](https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php)). We use the upper quartile (14 months) for El Niño, the middle two quartiles (28 months) for Neutral and the lower quartile (14 months) for La Niña. We refer to El Niño and La Niña together as “strong ENSO.” This differs from the “official” NOAA definition, by which 16 months in our period were classified as El Niño, 24 as La Niña and 15 as Neutral. We use our quartile-based definition to ensure equally sized strong and neutral ENSO samples.

To analyze conditional performance based on MJO phase, we use the real-time multivariate MJO (RMM) indices from Wheeler and Hendon (2004), which are derived from projections onto a pair of empirical orthogonal functions (EOFs) of intraseasonal anomalies in latitude-averaged ( $15^\circ\text{S}$ – $15^\circ\text{N}$ ) outgoing longwave radiation (OLR) and zonal winds at 850 and 200 hPa. We use observed RMM indices to compute performance conditioned on the MJO phase on the reforecast initialization date. The observed RMM indices are calculated from National Oceanic and Atmospheric Administration (NOAA) OLR data and NCEP/National Center for Atmospheric Research (NCAR) reanalysis winds; data are available from <http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealttime.txt>. We use RMM indices computed from the reforecast OLR and zonal winds to evaluate variations with lead time in the MJO teleconnection to South American rainfall. RMM indices for all reforecasts were computed following Gottschalck et al. (2010), which is a modified form of the original Wheeler and Hendon (2004) procedure. There are eight RMM phases. We pair MJO phases to increase sample size: phases 8 + 1, 2 + 3, 4 + 5, and 6 + 7. Phases 8 + 1 combine the wettest two phases over tropical South America, whereas phases 4 + 5 are the driest (Gottschalck et al. 2010; Grimm 2019). For these phases, we include only days with RMM amplitude  $\geq 1$ ; we refer to days with RMM amplitude  $< 1$ , regardless of phase, as “weak MJO.”

TABLE 2. For each model, the number of reforecast initializations in each ENSO and MJO category considered. The number of initialization is the same in ECMWF, NCEP, and UKMO, since we use lagged ensembles for ECMWF and NCEP referenced to the UKMO dates. The number of El Niño and La Niña initializations is exactly one-half the number of “strong ENSO” initializations, by definition.

Condition	ECMWF, UKMO, and NCEP	BAM
Neutral ENSO	108	54
Strong ENSO	112	56
Weak MJO	72	38
MJO phases 8 + 1	26	14
MJO phases 4 + 5	38	22

Table 2 shows the reforecast sample sizes for each ENSO and MJO category in our conditional performance analysis.

### c. Rainfall data

We validate reforecast rainfall against land-only rainfall estimates from the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS; Funk et al. 2015) dataset. CHIRPS blends station observations of rainfall with infrared-based satellite estimates that uses cold-cloud duration as a proxy for rainfall. Most of our analyses are performed on weekly average rainfall, following similar studies (Li and Robertson 2015; de Andrade et al. 2018), for which we average daily CHIRPS values. Certain analyses are performed on daily rainfall, which use the CHIRPS daily data directly; these are noted in the text. Several recent studies have found that CHIRPS compares well to gauge estimates of South American rainfall, particularly in the northeast and southeast (e.g., Paredes-Trejo et al. 2017; Nogueira et al. 2018). CHIRPS data are spatially interpolated to the models’  $1.5^\circ \times 1.5^\circ$  grid.

### d. Diagnostics and metrics

We evaluate reforecasts at lead times of 1–5 weeks. We compute biases in the mean reforecast rainfall, including biases conditioned on MJO and ENSO phases, as a function of lead time. We compute root-mean-squared errors (RMSE) by comparing weekly rainfall anomalies from each ensemble member (i.e., of the lagged ensembles for ECMWF and NCEP) to CHIRPS weekly rainfall anomalies, then computing the RMSE across all ensemble members (i.e., this is the RMSE of all members, not the RMSE of the ensemble mean). This provides a less biased comparison of model ensembles of different sizes, as used here. Using the RMSE of the anomalies, rather than of the total value, excludes the RMSE contribution from the lead-time-dependent systematic model bias. This mimics an operational procedure, in which the systematic bias in the real-time forecasts is removed using the reforecast climatology. For each year in 1999–2010, we compute anomalies relative to the weekly 1999–2010 climatology, excluding the year for which the anomaly is computed; again, this mimics a real-time procedure.

We measure model performance by the correlation coefficient (in time, hereafter “CC”) between the reforecast and CHIRPS anomalies; we measure skill by the Brier skill score

(BSS). CC and BSS are computed using on the models'  $1.5^\circ \times 1.5^\circ$  grid. To determine if CCs are statistically significant, we compute the critical CC value at the 5% significance level using an effective sample size, which accounts for the lag-1 autocorrelation between consecutive forecast initializations at the same lead time (e.g., the lag-1 autocorrelation of week-1 rainfall at a given grid point), following [Zwiers and von Storch \(1995\)](#).

BSSs are computed for terciles of weekly mean rainfall:

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}}}, \quad (1)$$

where BS is the Brier score of the reforecast and  $\text{BS}_{\text{ref}}$  is the Brier score of the reference forecast: a climatological forecast in which the probability of each tercile is always 1/3. The Brier score for a given tercile, at a given grid point, is

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2, \quad (2)$$

where  $N$  is the number of forecasts (i.e., the sample size in section a);  $f_i$  is the probability that the weekly rainfall forecast lies within this tercile of the reforecast distribution for this grid point and this week of the year (i.e., number of forecast ensemble members in this tercile, divided by ensemble size); and  $o_i$  is 0 if the observed rainfall was not in this tercile of the observed rainfall distribution for this week of the year, at this grid point, and 1 if it was.  $\text{BSS} > 0$  indicates skill above the climatological reforecast. Tercile boundaries are computed separately for CHIRPS and the reforecasts; for the latter, tercile boundaries are computed separately for each week of lead time, using 1999–2010 data, excluding the year for which probabilities are computed (as for the rainfall climatology above).

To measure conditional performance by MJO or ENSO phase, we compute the difference in CC between reforecasts initialized in a strong phase and those initialized in a neutral or weak phase (e.g., in MJO phases 4 + 5 relative to weak MJO). MJO phase and amplitude are determined from the observed RMM indices; these may differ slightly in models not initialized the NCEP/NCAR product used for the observed RMM indices (e.g., ECMWF is initialized from ERA-Interim), but we do not account for this. For ECMWF and NCEP, we use the observed MJO amplitude and phase on the UKMO initialization date, since this the date to which the lagged ensembles are referenced. This may bias the results, because members of the lagged ensemble may have been initialized in a different phase or with a different amplitude. The evaluation for UKMO and BAM uses the RMM amplitude and phase on the respective initialization dates for those models. To estimate the significance of the conditional performance results, we randomly resample (with replacement) from the distributions of available reforecasts 1000 times, for both the strong phase (e.g., phases 4 + 5) and the weak phase. We declare the difference to be statistically significant if the CC for 70% of the distributions for the strong phase lie outside the 90% confidence interval for the CC value of the weak phase. This is equivalent to stating that we are 70% confident that the strong phase has a skill that

is distinct from the weak phase (itself known with 90% confidence).

Some diagnostics and metrics are averaged across six relatively homogeneous rainfall regions (shown in [Fig. 1a](#)): southern Amazonia (AMZ;  $5^\circ\text{--}15^\circ\text{S}$ ,  $67^\circ\text{--}47^\circ\text{W}$ ), Andes (AND;  $15^\circ\text{--}40^\circ\text{S}$ ,  $75^\circ\text{--}67^\circ\text{W}$ ), northeastern South America (NDE;  $5^\circ\text{--}15^\circ\text{S}$ ,  $47^\circ\text{--}34^\circ\text{W}$ ), northern South America (NSA;  $0^\circ\text{--}12^\circ$ ,  $80^\circ\text{--}50^\circ\text{W}$ ), Patagonia (PAT;  $40^\circ\text{--}50^\circ\text{S}$ ,  $75^\circ\text{--}60^\circ\text{W}$ ), and southeastern South America (SESA;  $22^\circ\text{--}35^\circ\text{S}$ ,  $60^\circ\text{--}48^\circ\text{W}$ ). Regions are adapted from [de Souza Custodio et al. \(2017\)](#); we added NSA to cover South America more completely. The relative importance of austral summer precipitation in the annual cycle for each region can be seen in [Fig. 1](#) of [Grimm \(2011\)](#). All statistics are computed on the original  $1.5^\circ \times 1.5^\circ$  model grid and then averaged (i.e., these are the area-average statistics, not the statistics of the area-average rainfall). To determine the statistical significance of the regionally averaged CC, we compare it to the regionally averaged critical correlation value, adjusted for effective sample size.

### 3. Results

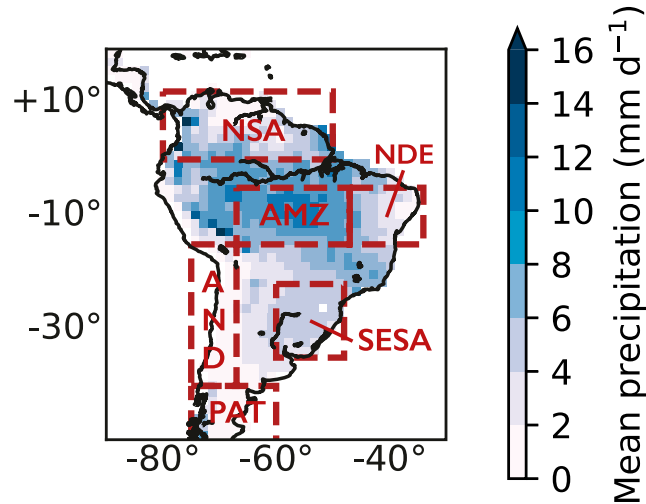
First, we examine S2S rainfall biases and RMSEs to characterize the representation of climatological rainfall in extended austral summer ([section 3a](#)). We then analyze unconditional performance and skill for weekly rainfall using CC and BSS, respectively ([section 3b](#)). To understand the potential for conditional prediction of rainfall based on large-scale tropical variability, we compute mean biases, errors and performance conditioned on the phases of ENSO ([section 3c](#)) and the MJO ([section 3d](#)). All analysis uses the larger lagged ensembles for ECMWF (33 members) and NCEP (32 members), except for the sensitivity test in [section 3b](#).

#### a. Mean biases

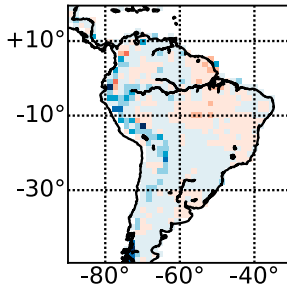
Extended austral summer (NDJFM) climatological rainfall in CHIRPS shows maxima in Amazonia and along the eastern slopes of the Andes, with a northwest–southeast-oriented band of high rainfall extending to the Atlantic coast of Brazil ([Fig. 1a](#)). There are local minima over the northern coast of South America, northeastern Brazil and the western slopes of the Andes. The four S2S models develop mean rainfall biases by week 1 ([Figs. 1b–e](#)). All models underpredict mean rainfall near the CHIRPS maximum over Amazonia, with biases largest in NCEP ([Fig. 1c](#)) and smallest in ECMWF ([Fig. 1b](#)). All models also overpredict rainfall near the Andes, with biases higher in NCEP and UKMO ([Fig. 1d](#)) and smaller in ECMWF and BAM ([Fig. 1e](#)). Mean rainfall drifts remarkably little with lead time: week 5 ([Figs. 1f–i](#)) biases are highly similar to those in week 1. Notable exceptions include the growth of the BAM Amazonian dry bias ([Fig. 1i](#)) and of the UKMO wet bias near the Andes ([Fig. 1f](#)). Overall, biases are lowest in ECMWF, moderate in UKMO and strongest in NCEP and BAM.

We compare RMSEs of ensemble-member weekly rainfall anomalies against CHIRPS to the standard deviation of weekly anomalies from CHIRPS ([Fig. 2a](#)). The latter is equivalent to the RMSE of a climatological reference forecast; it is fairly

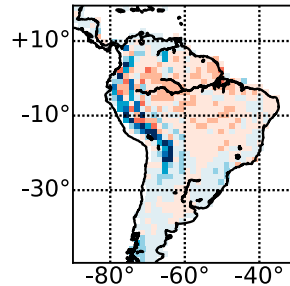
## a. CHIRPS climatology



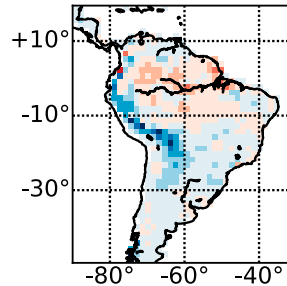
## b. ECMWF Week 1



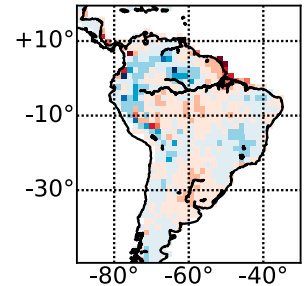
## c. NCEP Week 1



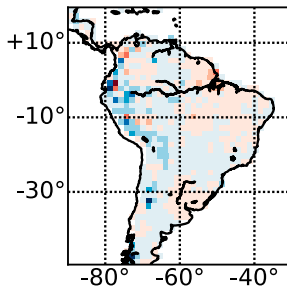
## d. UKMO Week 1



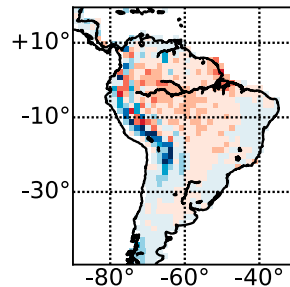
## e. BAM Week 1



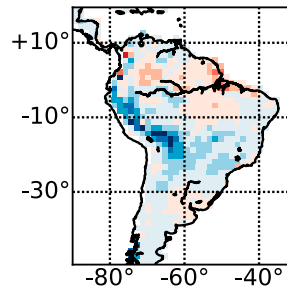
## f. ECMWF Week 5



## g. NCEP Week 5



## h. UKMO Week 5



## i. BAM Week 5

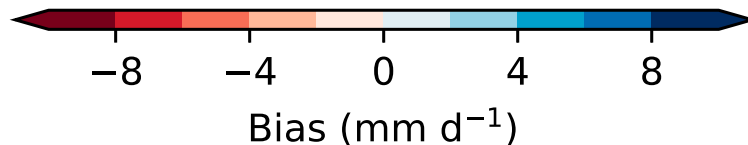
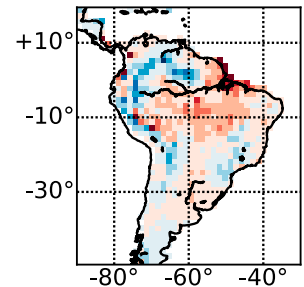


FIG. 1. NDJFM (a) mean rainfall ( $\text{mm day}^{-1}$ ) from CHIRPS and biases in mean rainfall ( $\text{mm day}^{-1}$ ) with respect to CHIRPS from S2S reforecasts from (b),(f) ECMWF; (c),(g) NCEP; (d),(h) UKMO; and (e),(i) BAM at week-1 lead time in (b)–(e) and week-5 lead time in (f)–(i). Note that (a) uses a separate colorbar, to the right of the panel. Panel (a) identifies the six regions used elsewhere in the manuscript.

spatially uniform, though higher near the mouth of the Amazon River. RMSEs of model anomalies represent the “random” component of model error, separate from the mean bias; both bias and RMSE vary with lead time. All models show

RMSEs much larger than the CHIRPS standard deviation nearly everywhere (Figs. 2b–m), with higher RMSEs in southern and eastern Brazil, particularly in NCEP (Figs. 2c,g,k) and UKMO (Figs. 2d,h,l). These errors are striking



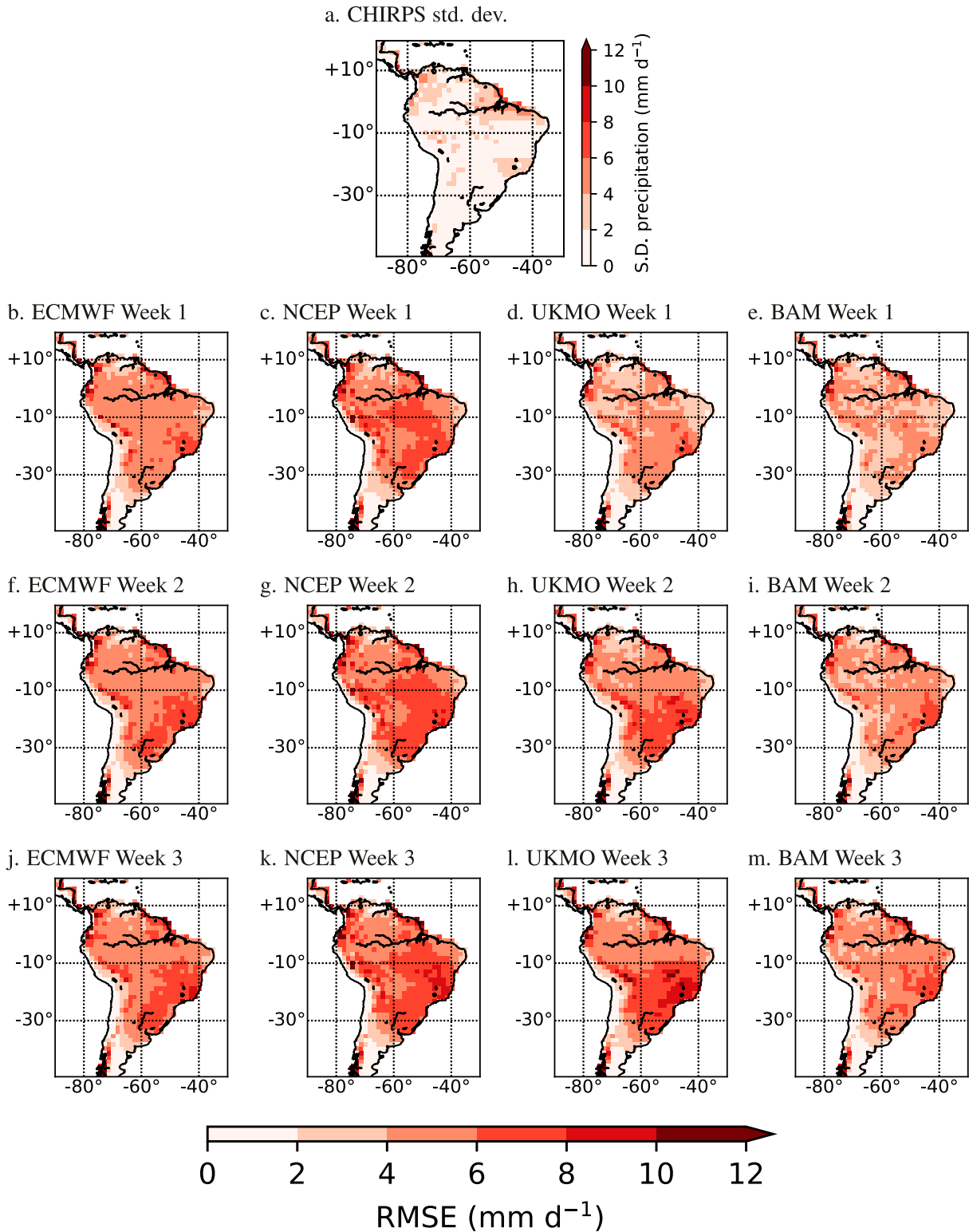


FIG. 2. (a) Standard deviation in weekly mean CHIRPS rainfall anomalies for NDJFM, equivalent to the root-mean-squared error (RMSE) of a climatological reference forecast; RMSEs in reforecast NDJFM weekly mean rainfall anomalies (mm day<sup>-1</sup>) against CHIRPS for (b),(f),(j) ECMWF; (c),(g),(k) NCEP; (d),(h),(l) UKMO; and (e),(i),(m) BAM at lead times of week 1 in (b)–(e), week 2 in (f)–(i), and week 3 (j)–(m). Note that (a) uses a separate color bar, to the right of the panel.

given that climatological rainfall is moderate and CHIRPS variance is similar to elsewhere. Week-1 RMSEs are highest in NCEP (Fig. 2c) and lowest in BAM (Fig. 2e). RMSEs grow substantially with lead time in all models (Figs. 2b–m), particularly in the south and southeast. By week 3, models show a similar pattern of RMSEs that are slightly lower in ECMWF (Fig. 2j) and BAM (Fig. 2m) and higher and more biased toward southeast Brazil in NCEP (Fig. 2k) and UKMO (Fig. 2l). Regional-scale RMSEs are similar in AMZ (Fig. 3a), NDE (Fig. 3c), NSA (Fig. 3d) and SESA (Fig. 3f) despite substantial differences in mean rainfall (Fig. 1): AMZ has relatively high mean rainfall; SESA has moderate mean rainfall and NSA and NDE have lower mean rainfall. This demonstrates that models have as much difficulty, if not more, predicting rainfall variations in wet regions as in dry regions. High RMSEs in densely populated SESA are particularly concerning. Regional-mean RMSEs are generally highest for NCEP and lowest for BAM and ECMWF.

### b. Unconditional performance assessment

At week 1, all models have statistically significant CCs (at 5% level) across most of South America (Figs. 4a–d). CCs are highest in northeastern Brazil, where mean rainfall is relatively low, and in southeastern South America; CCs are lowest over southern Amazonia and across central-eastern Brazil, near the climatological SACZ position, where mean rainfall is relatively high. CCs decrease with lead time, as expected. At week 2, CCs decline most strongly in NCEP (Fig. 4f) and BAM (Fig. 4h), particularly in central-eastern Brazil, such that significant CCs are limited to northern, northeastern, and southeastern South America; CCs are near zero in southern Amazonia and Argentina. ECMWF (Fig. 4e) and UKMO (Fig. 4g) maintain significant CCs across most of South America. At week 3, in all models CCs are significant only in northern South America, northeastern Brazil, and a small region of southeastern South America (Figs. 4i–l). At weeks 4 and 5, only isolated regions of significant CCs remain in northeastern Brazil (not shown).

Regional-mean CCs suggest similar performance, with the highest CCs in the climatologically drier NDE (Fig. 3i) and in SESA (Fig. 3l) and lowest CCs in extratropical PAT (Fig. 3k) and the climatologically wetter tropical AMZ (Fig. 3g). The four models perform similarly, although CCs are slightly higher in ECMWF and UKMO and slightly lower in NCEP and BAM. Statistically significant CCs in most models extend to week 1 in PAT, to week 2 in AMZ and AND, to week 3 in NSA and SESA, and to week 4 in NDE.

To explore the relationship between CCs and climatological rainfall, for each model and lead time we produce distributions of gridpoint CCs binned by CHIRPS NDJFM mean rainfall (Figs. 5a–c). We use values at all grid points in the domain (shown in Fig. 4), regardless of whether the CC is statistically significant. CCs are higher at grid points with low or moderate mean rainfall and lower at grid points with high mean rainfall, for weeks 1–3 and for all models. A similar analysis for CCs binned by the standard deviation of weekly CHIRPS rainfall for NDJFM (shown in Fig. 2a) shows that all models show higher CCs at points with low or moderate subseasonal variability and lower CCs in regions of high

subseasonal variability (Figs. 5d–f). We return to these results in section 4.

For regional-mean CCs and RMSEs, we assess the sensitivity of forecast performance to our lagged-ensemble strategy for NCEP (32 members) and ECMWF (33 members) using an 8-day window (section 2a). Figure 6 compares CCs and RMSEs for these ensembles to CCs and RMSEs for alternative smaller NCEP (8 members, 2-day lagged ensemble) and ECMWF (11 members, one initialization) ensembles similar in size to UKMO (7 members) and BAM (11 members). Variations in CC and RMSE between the two sets of ensembles are generally small, but for some regions and lead times they are meaningful. At weeks 1 and 2, the larger ensembles perform slightly worse than the smaller ensembles, with higher RMSEs in NDE (Fig. 6c) and SESA (Fig. 6f) but similar CCs. Using a longer window to create the lagged ensembles degrades performance most at short leads, when predictability arises mainly from initial conditions. Beyond week 2, the larger ensembles slightly outperform the smaller ensembles, with higher CCs for NDE (weeks 4 and 5), SESA (weeks 3–5), and AMZ (Fig. 6a; weeks 4 and 5). This is likely because using larger ensembles improves the signal-to-noise ratio at longer leads. There is little change in performance in AND, NSA, or PAT (Figs. 6b,d,e, respectively). Since S2S forecasts are used mostly for lead times of two weeks and beyond, we continue to use the larger NCEP and ECMWF ensembles that perform slightly better at these lead times.

BSSs for upper-tercile rainfall demonstrate that all models show poor skill beyond week 1, across most of South America (Fig. 7). At week 1, ECMWF outperforms a climatological forecast in eastern Brazil, southern Brazil, and northern Amazonia (Fig. 7a). By week 2, however, BSSs are only slightly above zero (Fig. 7e); by week 3 the model performs similarly to or worse than the climatological forecast (Fig. 7i). NCEP shows similar skill in eastern and southern Brazil at week 1 (Fig. 7b), but fails to maintain skill at week 2 (Fig. 7f). NCEP also shows negative BSS in Amazonia and near the Andes, even at week 1. UKMO and BAM fail to outperform a climatological forecast across most of South America even at week 1 (Figs. 7c,d, respectively), with only isolated areas of positive BSS in eastern and southern Brazil. Lower performance for BSS than for CC suggests that while the ensemble-mean may capture the sign of week-to-week rainfall variations at 1–2 weeks ahead, the ensemble members struggle to capture shifts in the distributions of the anomalies.

Regional-mean BSSs (Fig. 8) confirm that ECMWF consistently outperforms a climatological forecast at week 1 for most regions, with skill extending to week 2 over NDE (Fig. 8c) and SESA (Fig. 8f). NCEP exhibits skill above a climatological forecast in NDE at weeks 1 and 2, but shows little useful skill elsewhere. UKMO and BAM show useful skill only in NDE at week 1. Results for both gridscale and regional-mean BSSs are similar for lower-tercile weekly rainfall (not shown). For normal (middle-tercile) rainfall, no model outperforms the climatological forecast at any lead.

### c. Conditional biases and performance based on ENSO

Before examining conditional performance by ENSO phase, we first verify the predicted ENSO–rainfall relationship by

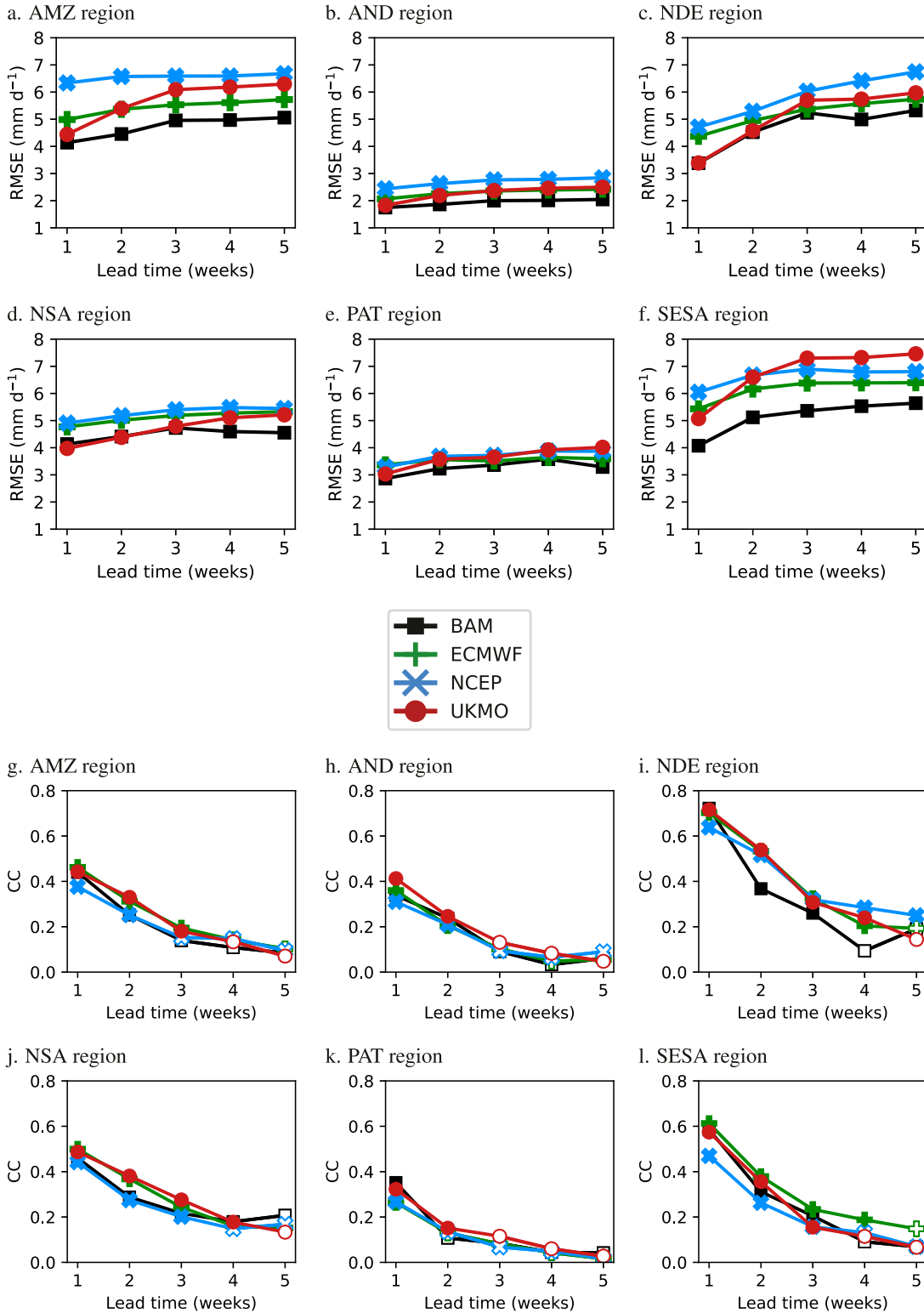


FIG. 3. Regional-mean (a)–(f) root-mean-squared errors (RMSE;  $\text{mm day}^{-1}$ ) and (g)–(l) correlation coefficients (CC) between reforecast and observed (CHIRPS) anomalies for NDJFM weekly mean rainfall as a function of lead time from all models, compared to CHIRPS. Regions are Amazonia (AMZ) in (a) and (g); Andes (AND) in (b) and (h); northeastern South America (NDE) in (c) and (i); northern South America (NSA) in (d) and (j); Patagonia (PAT) in (e) and (k); and southeastern South America (SESA) in (f) and (l). Metrics are computed on the original  $1.5^\circ$  grid, then averaged over the region. The regions are shown in Fig. 1a. In (g)–(l), filled symbols show statistically significant CCs at the 5% level, based on the regionally averaged critical CC threshold, adjusted for effective sample size.

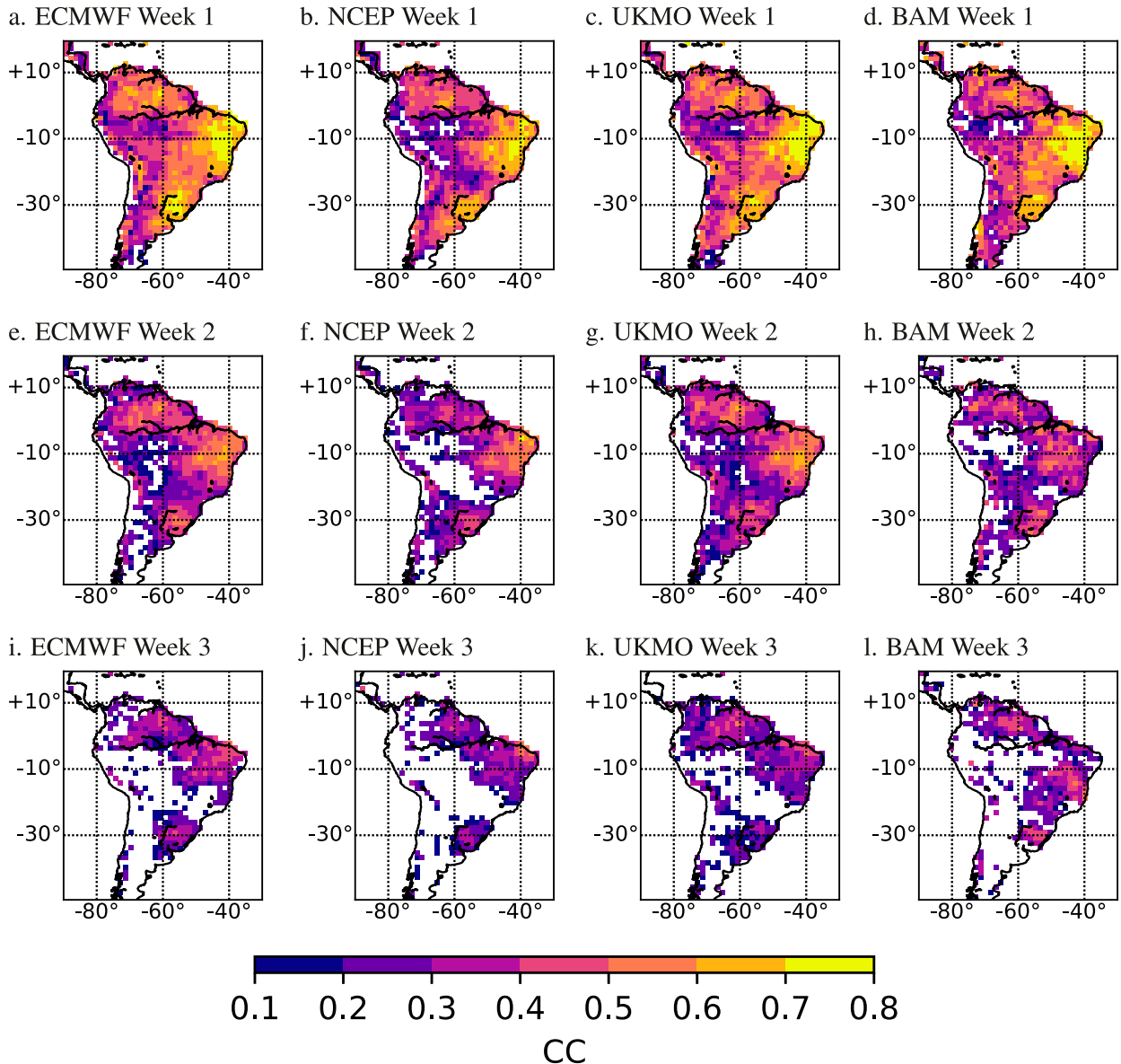


FIG. 4. Correlation coefficients (CCs) between the reforecast and observed (CHIRPS) anomalies for NDJFM weekly mean rainfall for (a),(e),(i) ECMWF; (b),(f),(j) NCEP; (c),(g),(k) UKMO; and (d),(h),(l) BAM at (top) week-1, (middle) week-2, and (bottom) week-3 lead times. Only statistically significant CCs at the 5% level (adjusted for effective sample size) are shown.

compositing NDJFM rainfall anomalies by ENSO phase. A realistic representation of the ENSO–rainfall teleconnection, and hence also of the large-scale ENSO-driven seasonal circulation, may be necessary for ENSO conditional performance to exceed unconditional performance, particularly if models can persist the initialized seasonal-scale ENSO-associated circulation throughout the S2S forecast. To fairly compare S2S and CHIRPS El Niño rainfall anomalies, we create a separate CHIRPS composite for each week of S2S lead time, because the forecast validity period shifts with lead time (i.e., the validity period for week 3 differs from that for week 1). We composite CHIRPS weekly mean rainfall for the same period over which the S2S reforecasts are valid. We show the

CHIRPS composites for the validity windows common to UKMO, ECMWF and NCEP. The validity windows differ for BAM, but the CHIRPS composites are qualitatively similar (not shown).

In El Niño, CHIRPS shows the expected pattern of lower rainfall over Amazonia, northern South America, and northeastern Brazil, with enhanced rainfall over southern Brazil, Uruguay and northeastern Argentina (Figs. 9a–c). The CHIRPS “week 5” (Fig. 9c) and “week 1” (Fig. 9a) composites differ somewhat, particularly in northeastern Brazil, due to sampling a different set of weekly rainfall data, but the broad pattern remains similar. At week 1, all S2S models reproduce this broad pattern (Figs. 9d–g), but with variations in amplitude: ECMWF shows weaker anomalies over land than

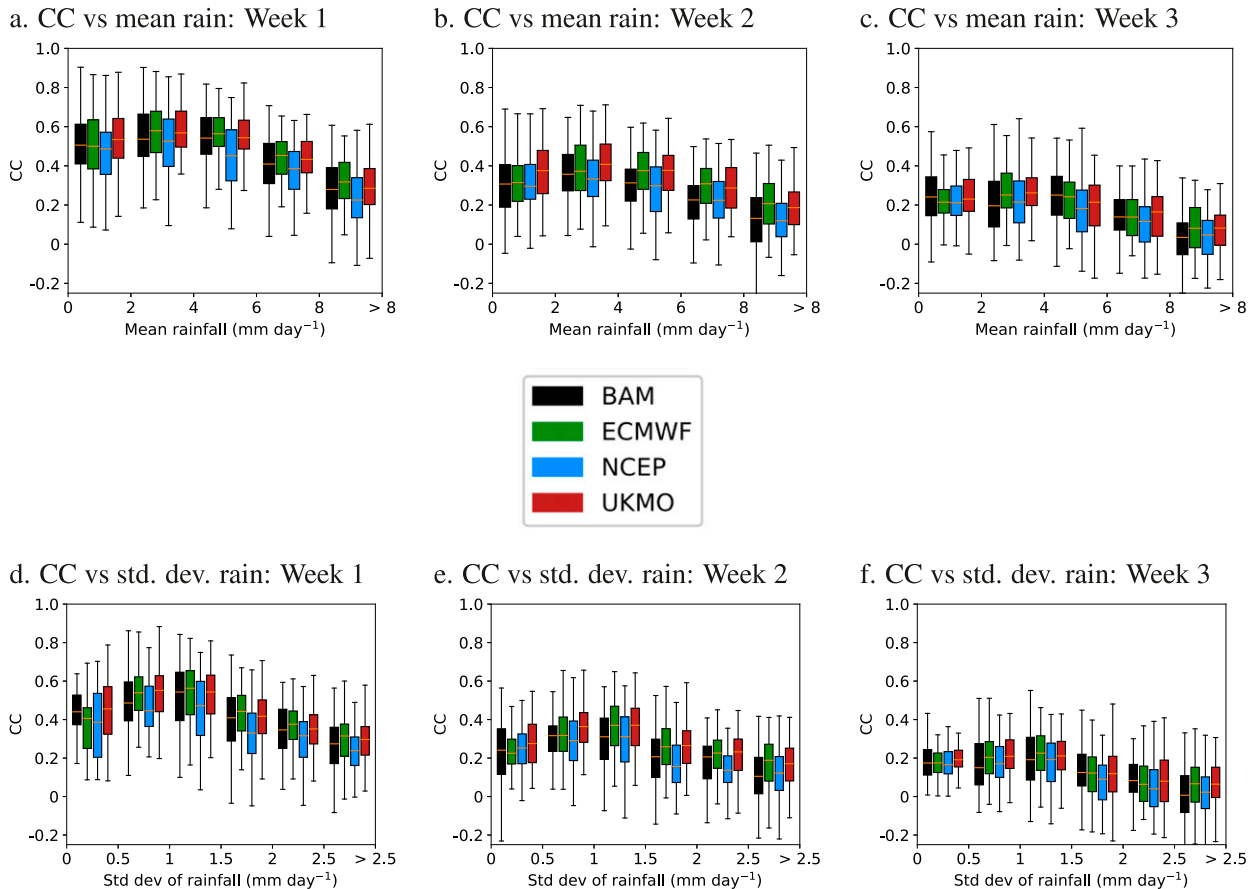


FIG. 5. Box-and-whisker diagrams of the distribution of the CCs shown in Fig. 4, binned by either (a)–(c) the NDJFM mean CHIRPS rainfall from Fig. 1a or (d)–(f) the standard deviation of weekly CHIRPS NDJFM rainfall from Fig. 2a. The diagrams each show CCs from all S2S models for a given week of lead time: (left) week 1; (center) week 2; and (right) week 3. The filled color of the box indicates the model (see legend). For each distribution, the yellow line shows the median, the box shows the interquartile range and the whiskers show the range between the 5th and 95th percentiles of the distribution. The bins were chosen to give approximately equal sample sizes of grid points.

CHIRPS in both northern and southern Brazil (Fig. 9d); NCEP shows weaker anomalies in southern Brazil and Uruguay, with dry anomalies that extend too far south from northern Brazil (Fig. 9e); UKMO is similar to NCEP in northern South America, but with stronger positive anomalies in the south, similar to CHIRPS (Fig. 9f); BAM overestimates the reduced rainfall in northern Brazil (Fig. 9g).

Unlike the unconditional rainfall biases in Fig. 1, which changed very little with lead time, the ENSO–rainfall teleconnection drifts moderately by week 3 (Figs. 9h–k) and substantially by week 5 (Figs. 9l–o). In all models, by week 3 the enhanced rainfall in southeastern South America weakens substantially. In NCEP (Fig. 9i) and UKMO (Fig. 9j), enhanced rainfall stretches into central-eastern Brazil and encroaches into the region of dry anomalies in CHIRPS. These results suggest that as these models drift from their initial conditions, the region of anomalous subtropical ascent stretches meridionally across the continent. The dry anomalies in near-equatorial South America remain fixed spatially at week 5, relative to week 3, but weaken considerably in BAM

(Fig. 9o)—reducing the dry bias from week 1—and weaken slightly in UKMO (Fig. 9n), but are maintained in ECMWF (Fig. 9l) and NCEP (Fig. 9m). BAM provides the best spatial pattern of El Niño anomalies at week 5. In La Niña phases, the anomalies are slightly weaker than, and opposite in sign to, the El Niño anomalies in CHIRPS and the S2S models (not shown).

Next, we assess conditional performance for reforecasts started in strong ENSO phases—either El Niño or La Niña—relative to performance for reforecasts started in neutral ENSO phases (see section 2d for method and Table 2 for sample sizes). We evaluate conditional performance only until week 3, as this is the limit of both conditional and unconditional (Fig. 4) performance. In most regions, performance in strong ENSO does not significantly differ from performance in neutral ENSO (Fig. 10). However, ECMWF, NCEP, and UKMO show significantly lower performance in parts of central and eastern Brazil, where the models struggle to capture the sign of the observed ENSO rainfall anomaly (Fig. 9). BAM is the only model to develop a coherent region

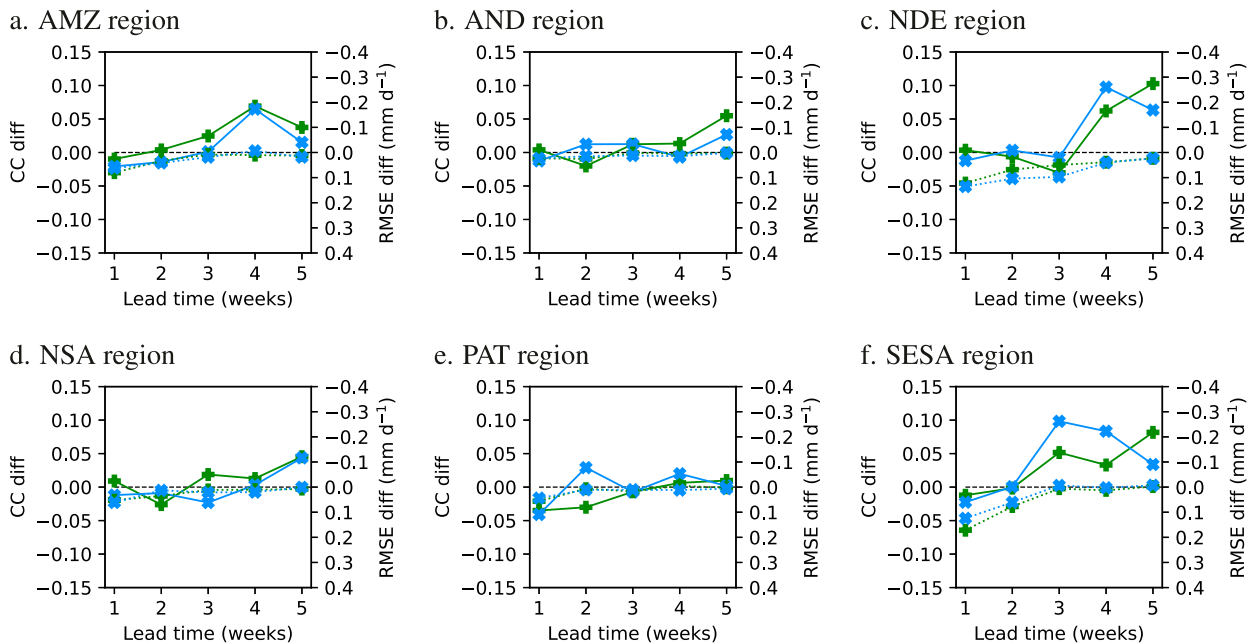


FIG. 6. Differences in regional-mean CCs (solid lines; left-hand vertical axis) and RMSEs (dotted lines; right-hand vertical axis) between two sets of NCEP (blue) and ECMWF (green) ensembles: the larger lagged ensembles constructed using an 8-day window, to produce a 32-member NCEP ensemble and a 33-member ECMWF ensemble; and the smaller ensembles that are more similar in size to the 7-member UKMO ensemble, with an 8-member NCEP ensemble (two initialization dates) and an 11-member ECMWF ensemble (one initialization date). Differences are taken as the CC or RMSE for the larger ensemble minus that for the smaller ensemble, for each model. Note that the right-hand axis for RMSE is inverted so that degradations in CC (lower) or RMSE (higher) in the larger ensembles lie below the dashed black line (at zero difference) and that improvements in CC (higher) or RMSE (lower) in the larger ensembles lie above the dashed black line.

of higher performance in strong ENSO, at weeks 2 and 3 in northeastern Brazil (Figs. 10h,i) where the model also captures the sign and magnitude of the ENSO-related anomalous rainfall (Fig. 9k).

Differences in regional-mean CC between strong ENSO and neutral ENSO phases confirm that there are few statistically significant differences in prediction performance at regional scale (Fig. 11). Thus, ENSO events provide only limited “windows of opportunity” for improved S2S rainfall predictions over South America, which are region and model dependent.

#### d. Conditional biases and performance based on MJO

Before examining conditional performance based on the MJO, we first analyze the MJO–rainfall teleconnection in CHIRPS and the S2S models (Fig. 12). We composite daily rainfall during strong (RMM amplitude  $\geq 1$ ) MJO days in two phase pairs: phases 8 and 1 (8 + 1), when MJO convection is enhanced over the tropical Western Hemisphere; and phases 4 and 5 (4 + 5), when MJO convection is enhanced over the Maritime Continent and suppressed in the tropical Western Hemisphere. We then average rainfall over all days in each week of lead time in each pair of MJO phases, using the MJO phase and amplitude on the validity date, to examine the instantaneous MJO–rainfall teleconnection as a function of lead time. We composite CHIRPS rainfall based on the observed RMM indices, but composite the S2S models based on their

predicted RMM indices, as the aim is to study the simulated MJO–rainfall teleconnection, not to evaluate RMM predictions. We composite CHIRPS daily rainfall, using all days in each pair of MJO phases during the S2S validity window for a given week of lead time. This produces a separate CHIRPS MJO composite for each week of S2S lead time. As for the ENSO analysis in section 3c, we show the CHIRPS composites for the ECMWF, UKMO and NCEP validity window; the composites for the BAM validity window (not shown) are qualitatively similar.

In CHIRPS, phases 8 + 1 have enhanced rainfall across northern South America, particularly in Peru and northeastern and central Brazil, with reduced rainfall in southeastern South America, including southern Brazil, Uruguay and northeastern Argentina (Fig. 12a). The magnitudes of these anomalies change with the shift in validity period from week 1 to week 5 (Fig. 12b), due to sampling a different validity period and hence a different set of MJO events, but the spatial pattern remains similar. Phases 4 + 5 display the opposite pattern (Figs. 12c,d). The CHIRPS composites differ from the gauge-based MJO rainfall composites in (Grimm 2019), particularly over northwest Brazil and parts of Peru and Bolivia. We attribute these to differences in observation density between datasets and in the compositing method, for example Grimm (2019) uses bandpass filtering to isolate the MJO signal, while we do not, due to the short length of the S2S forecasts; Grimm (2019) also analyses a different period

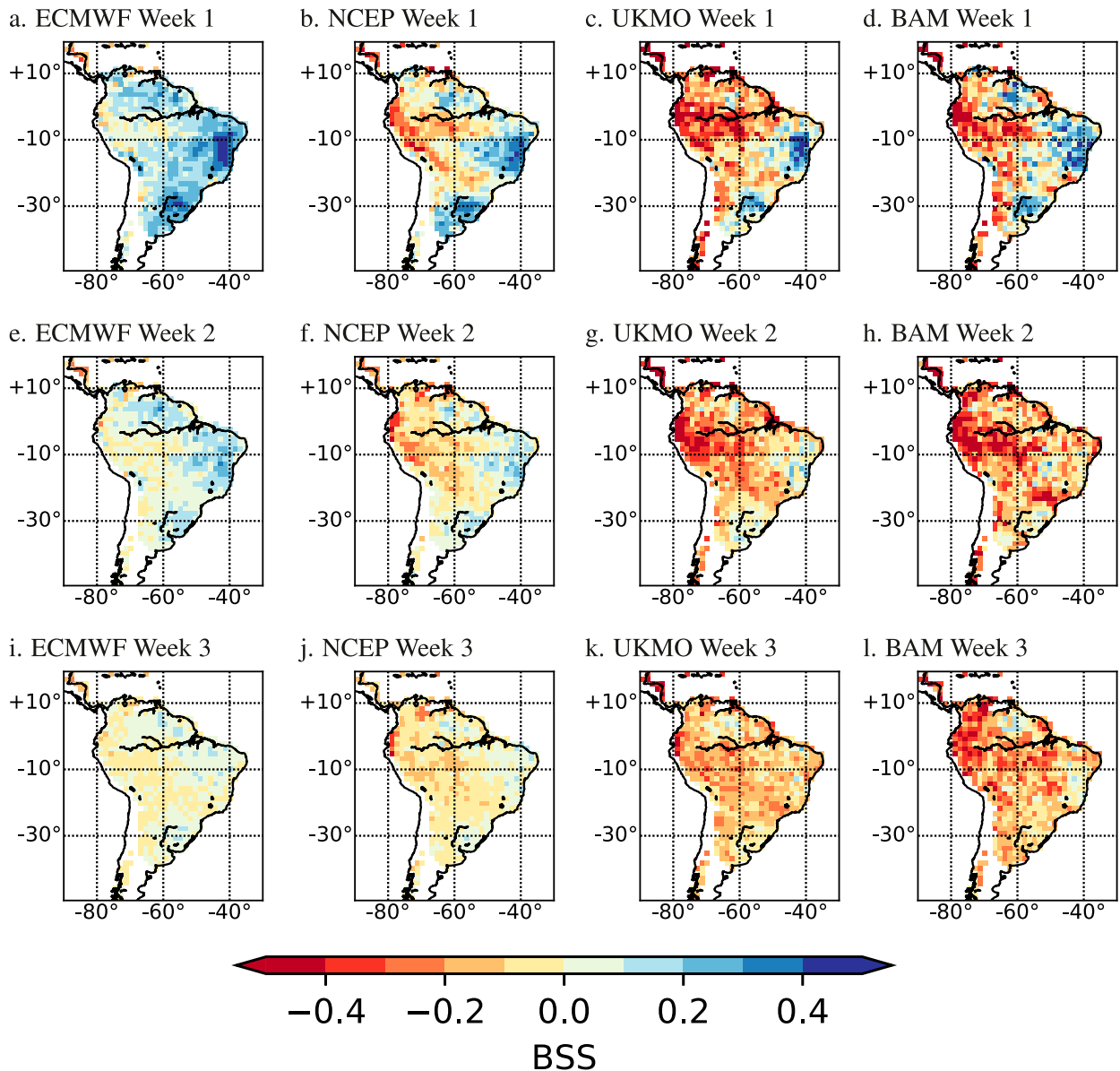


FIG. 7. Brier skill scores for NDJFM weekly mean rainfall above the upper tercile of each model's reforecast distribution, computed relative to a climatological forecast, for (a),(e),(i) ECMWF; (b),(f),(j) NCEP; (c),(g),(k) UKMO; and (d),(h),(l) BAM at lead times of (top) week 1, (middle) week 2, and (bottom) week 3. BSS values above zero indicate skill above a climatological forecast. BSS is not shown where CHIRPS rainfall is less than  $1 \text{ mm day}^{-1}$ .

(December–February, 1979–2009) than ours (November–March, 1999–2010).

All models show substantial errors in location and amplitude of the MJO–rainfall teleconnection, which grow with lead time. At week 1 in phases 8 + 1, all models show the band of maximum anomalous rainfall farther south and east than in CHIRPS, toward southern Amazonia and central-eastern Brazil rather than over the Amazon River (Figs. 12e,i,m,q), where predicted rainfall anomalies are weak or near-zero. The area of observed reduced rainfall in southeastern and southern Brazil and northeastern Argentina is weak in ECMWF and NCEP and almost absent in

BAM, perhaps linked to the southward contraction of the enhanced convection away from the equator. Only UKMO captures the dry anomalies. At week 5, the MJO–rainfall teleconnection weakens in ECMWF (Fig. 12f), NCEP (Fig. 12j) and BAM (Fig. 12q). In UKMO, the region of enhanced rainfall intensifies and shifts north, while the area of reduced rainfall disappears (Fig. 12n). NCEP develops a dry anomaly near the equator, opposite to the observed positive anomaly, which suggests an off-equatorial shift in MJO convection.

All models show similar biases in phases 4 + 5 at week 1, with southward (away from the equator) contractions in the

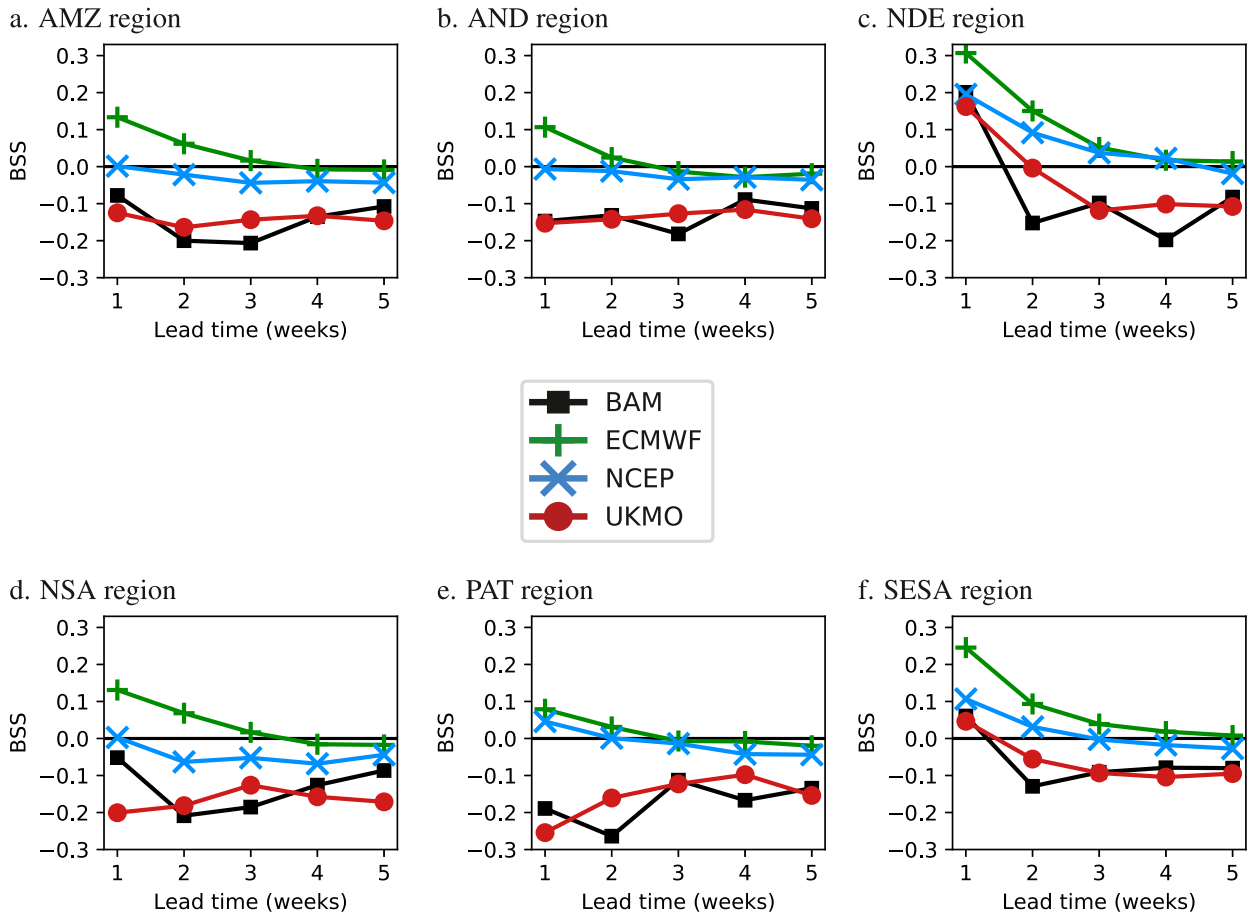


FIG. 8. Regional-mean Brier skill scores for NDJFM weekly mean rainfall above the upper tercile of each model's reforecast distribution, computed relative to a climatological forecast, for (a) Amazonia (AMZ), (b) Andes (AND), (c) northeastern South America (NDE), (d) northern South America (NSA), (e) Patagonia (PAT), and (f) southeastern South America (SESA). Metrics are computed on the original  $1.5^\circ$  grid, then averaged over the region. The regions are shown in Fig. 1a.

regions of suppressed rainfall in the deep tropics. ECMWF (Fig. 12g), UKMO (Fig. 12o) and BAM (Fig. 12s) produce weak positive or near-zero anomalies over the Amazon River, opposite to the observed negative anomalies. ECMWF and NCEP (Fig. 12k) underestimate the positive anomalies in southeastern South America, while UKMO overestimates them; only BAM captures the amplitude. The rainfall teleconnection weakens considerably by week 5, particularly in ECMWF (Fig. 12h) and NCEP (Fig. 12l). UKMO (Fig. 12p) exhibits the same northward shift of the main band of rainfall anomalies seen for phases  $8 + 1$  at week 5 (Fig. 12n). BAM also shows much weaker anomalies at week 5 than at week 1, particularly over southeastern South America (Fig. 12t).

The week-1 results demonstrate that even when initialized with a strong MJO circulation, all models quickly develop strong biases in the spatial rainfall pattern, particularly over the deep tropics. UKMO performs best for southeastern South America, particularly at week 1, but worst for near-equatorial rainfall. All models except UKMO strongly damp MJO-associated anomalies with lead time and erroneously contract the MJO-associated convection south of the equator.

To assess MJO conditional performance, for each MJO phase pair we compute changes in regional-mean CC between reforecasts started on days of observed strong MJO amplitude and reforecasts started on days of observed weak MJO amplitude (see section 2d for method and Table 2 for sample sizes). The interpretation of conditional performance based on CC may be complicated because (i) CC measures the ability to predict variations between samples relative to the mean, and so measures the ability of a model to predict the variation in rainfall between reforecasts with a given initial MJO phase, not the mean rainfall in that MJO phase; (ii) although the reforecasts in each sample have the same initial phase, the MJO phases and amplitudes diverge with lead time, making the sample less consistent and less distinguishable from the control set of weak MJO events; (iii) the true initial MJO phases and amplitudes are not consistent among the ECMWF and NCEP ensembles, due to the lagged ensemble approach (section 2d).

With a few exceptions, variations in performance with MJO phase are small and not statistically significant. For phases  $8 + 1$ , performance increases significantly in NDE for all models in week 1 (Fig. 13c), in NSA for BAM at all lead times except



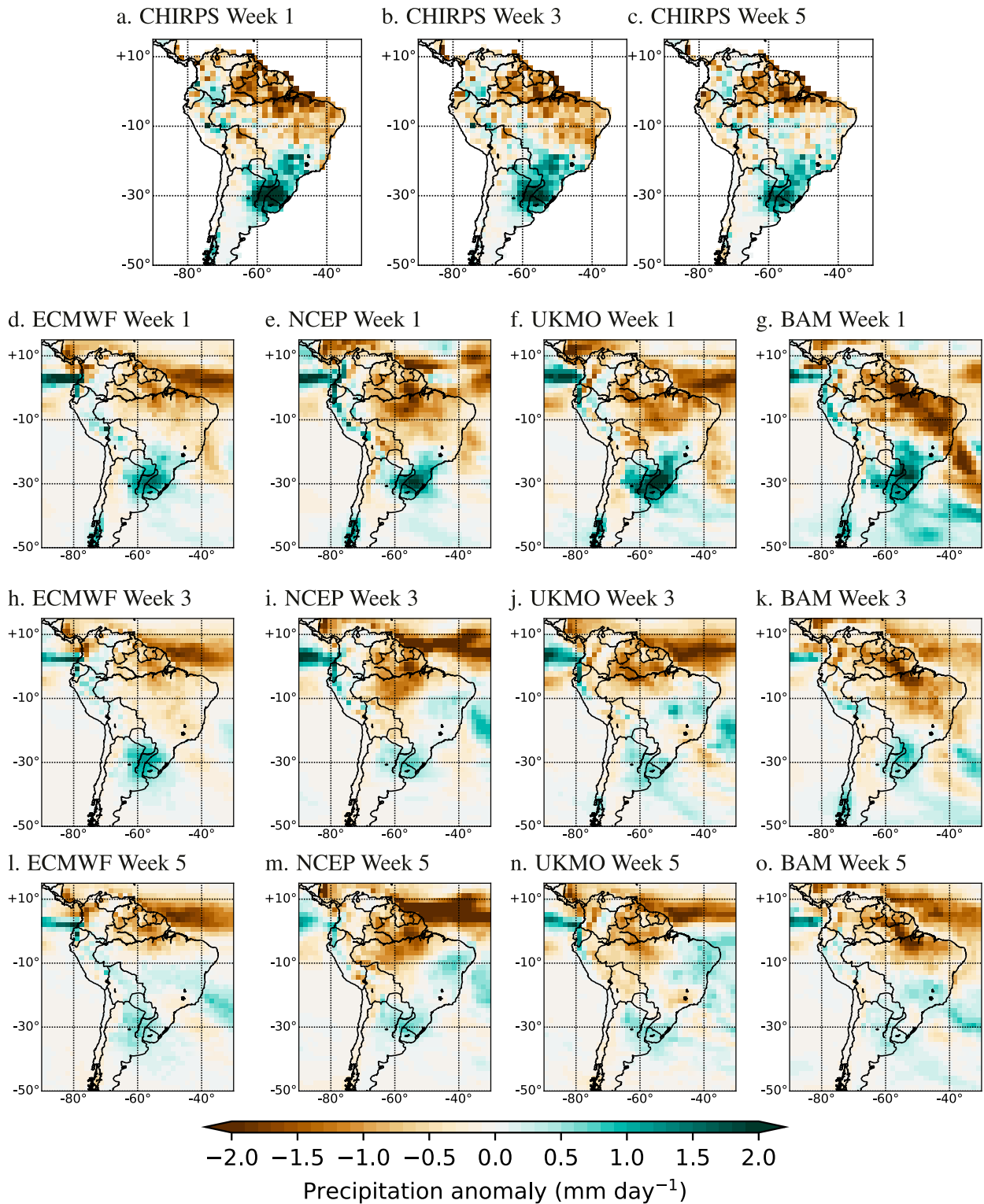


FIG. 9. Mean NDJFM rainfall anomaly (mm day<sup>-1</sup>) in El Niño [upper quartile of the oceanic Niño index (ONI)]. For CHIRPS, we show composites for validity windows corresponding to (a) week-1, (b) week-3, and (c) week-5 reforecasts, based on UKMO initialization dates. For S2S models, we show composites based on the observed ONI at initialization, for (d)–(g) week-1, (h)–(k) week-3, and (l)–(o) week-5 lead times. Anomalies are computed relative to 1999–2010. Note that CHIRPS is available only over land, but model anomalies are shown also over the ocean, to give larger-scale context.

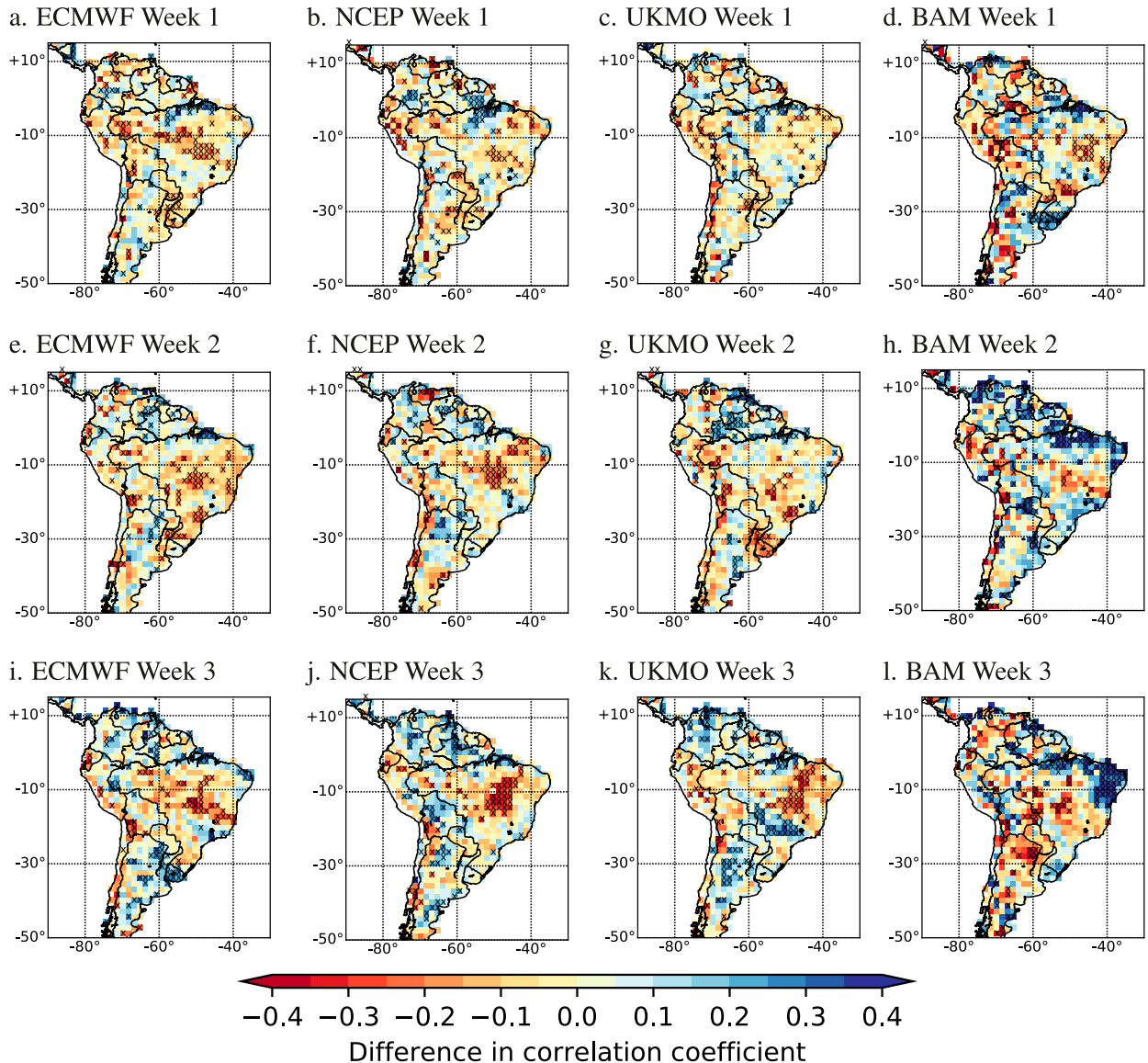


FIG. 10. Difference in S2S prediction performance between strong and neutral ENSO phases, defined as the difference in CC (between CHIRPS and S2S weekly rainfall anomalies) between NDJFM reforecasts started in strong ENSO phases and reforecasts started in neutral ENSO phases for (a),(e),(i) ECMWF; (b),(f),(j) NCEP; (c),(g),(k) UKMO; and (d),(h),(l) BAM at (top) week-1, (middle) week-2, and (bottom) week-3 lead times. Positive values indicate higher performance in strong ENSO phases than in neutral ENSO phases. Crosses indicate where 70% of strong ENSO reforecast samples have a CC outside the 90% confidence interval for the CC of neutral ENSO reforecast samples, based on resampling each dataset 1000 times (with replacement).

week 3 (Fig. 13d) and in SESA for NCEP in weeks 1 and 2 and for ECMWF in week 1 (Fig. 13f). Performance increases significantly for NCEP in NSA in weeks 3 and 4 (Fig. 13d), despite an incorrect sign of the MJO-associated rainfall anomaly (Fig. 12j).

For phases 4 + 5, the only notable significant changes are declines in performance in UKMO in AMZ (Fig. 13g) and SESA (Fig. 13l). In AMZ, UKMO predicts the MJO-associated reduced rainfall anomalies well at week 1, whereas in SESA the sign of the rainfall anomaly is correct but the magnitude is too strong (Fig. 12o). NCEP also shows

reduced performance in SESA for week 1, where the MJO-associated enhanced rainfall is reasonably well predicted (Fig. 12k). We find little relationship between performance for the MJO-associated rainfall anomaly and MJO conditional rainfall prediction performance.

#### 4. Discussion

S2S performance, measured by CC, and skill, measured by BSS, for weekly extended austral summer rainfall across South America is modest at best. While CCs are significant

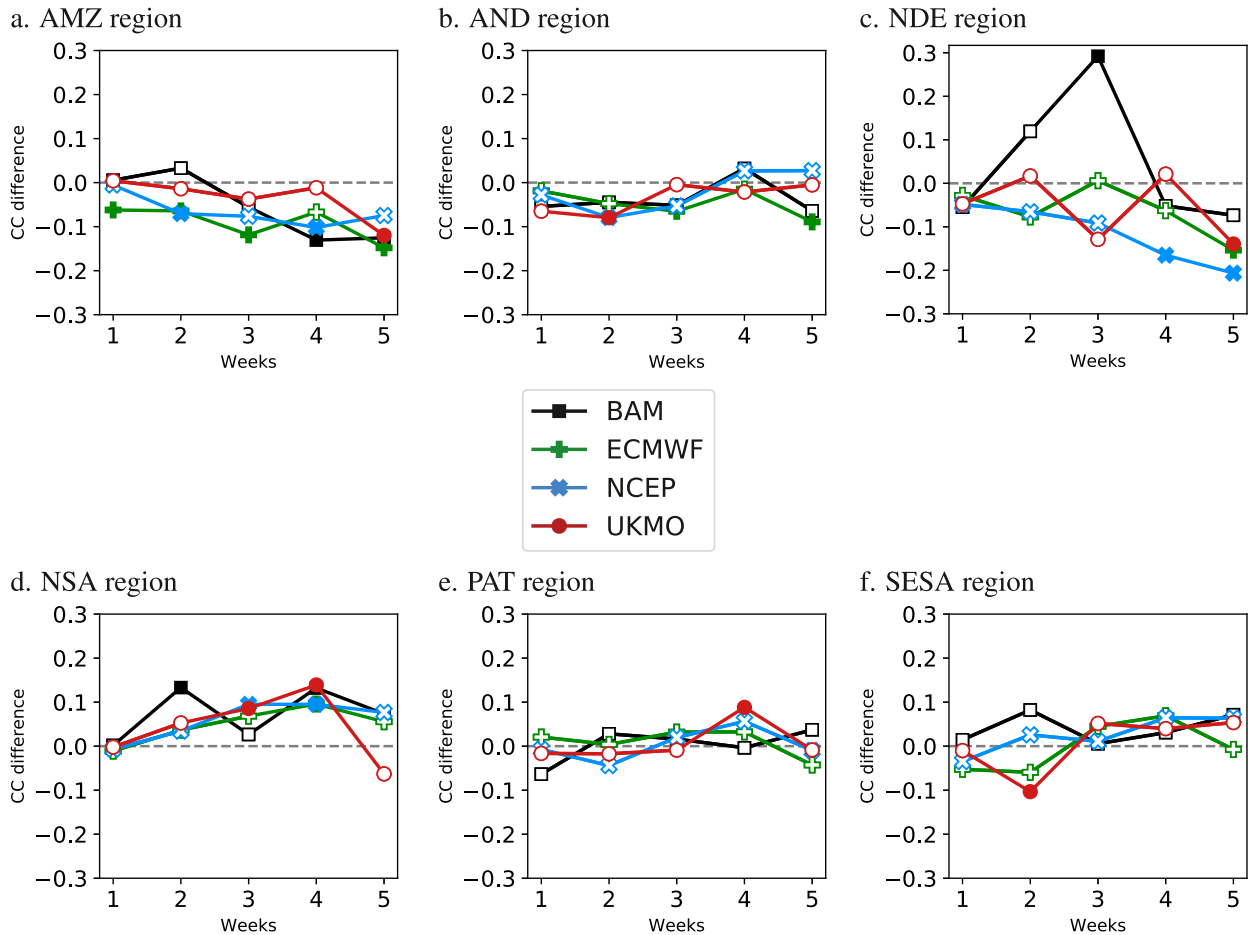


FIG. 11. Difference in regionally averaged S2S performance between strong and neutral ENSO, defined as the difference in regionally averaged CCs (between CHIRPS and S2S weekly rainfall anomalies) between NDJFM reforecasts started in strong ENSO phases and reforecasts started in neutral ENSO phases. Filled symbols show where 70% of strong ENSO reforecast samples have a CC outside the 90% confidence interval for the CC of neutral reforecast samples, based on resampling each dataset 1000 times (with replacement). Regions are (a) Amazonia (AMZ), (b) Andes (AND), (c) northeastern South America (NDE), (d) northern South America (NSA), (e) Patagonia (PAT), and (f) southeastern South America (SESA). The regions are shown in Fig. 1a.

(at 5%) in most regions for weeks 1 and 2, by week 3 CCs are significant only in northern, northeastern, and southeastern South America. The BSS metric shows that models are unable to skillfully predict upper-tercile or lower-tercile rainfall beyond week 1, except for ECMWF. Encouragingly, the regions of highest performance by CC in eastern Brazil are reasonably densely populated and agriculturally productive, such that in these regions S2S forecasts may be useful for agricultural applications or dam management (e.g., to manage water for hydropower or human consumption). However, skill, measured by BSS, is relatively low in the most densely populated and agriculturally productive regions, for example in the northeastern coastal region of Brazil and southern Brazil.

Our performance estimates agree with past assessments for South America, including Hirata and Grimm (2017), who found that NCEP could usefully predict extreme rain events in 2010–11 two weeks ahead. Coelho et al. (2018) found useful S2S prediction performance in austral autumn over northeastern Brazil and southeastern South America, three to four

weeks ahead, as did de Andrade et al. (2018) for austral summer in the ECMWF model three weeks ahead. Pegion et al. (2019) considered reforecasts from the SubX models initialized in all months, demonstrating that several models had high CCs for weekly rainfall over eastern Brazil three weeks ahead. SubX performance over southeastern South America was typically lower than that of the S2S models evaluated here, which may be due to our focus on austral summer or to differences in the models considered.

In all models analyzed here, unconditional performance is highest in relatively dry regions and lowest in relatively wet regions (Fig. 5). Unconditional performance is also higher in regions of low to moderate subseasonal variability and lower in regions of higher subseasonal variability. Models may be better able to predict rainfall where observed subseasonal variability is low or moderate (e.g., in northeast Brazil and southeast South America; Fig. 2a), as models may gain performance by persisting an initialized circulation anomaly and the associated rainfall anomaly. Comparing models to a persistence forecast,

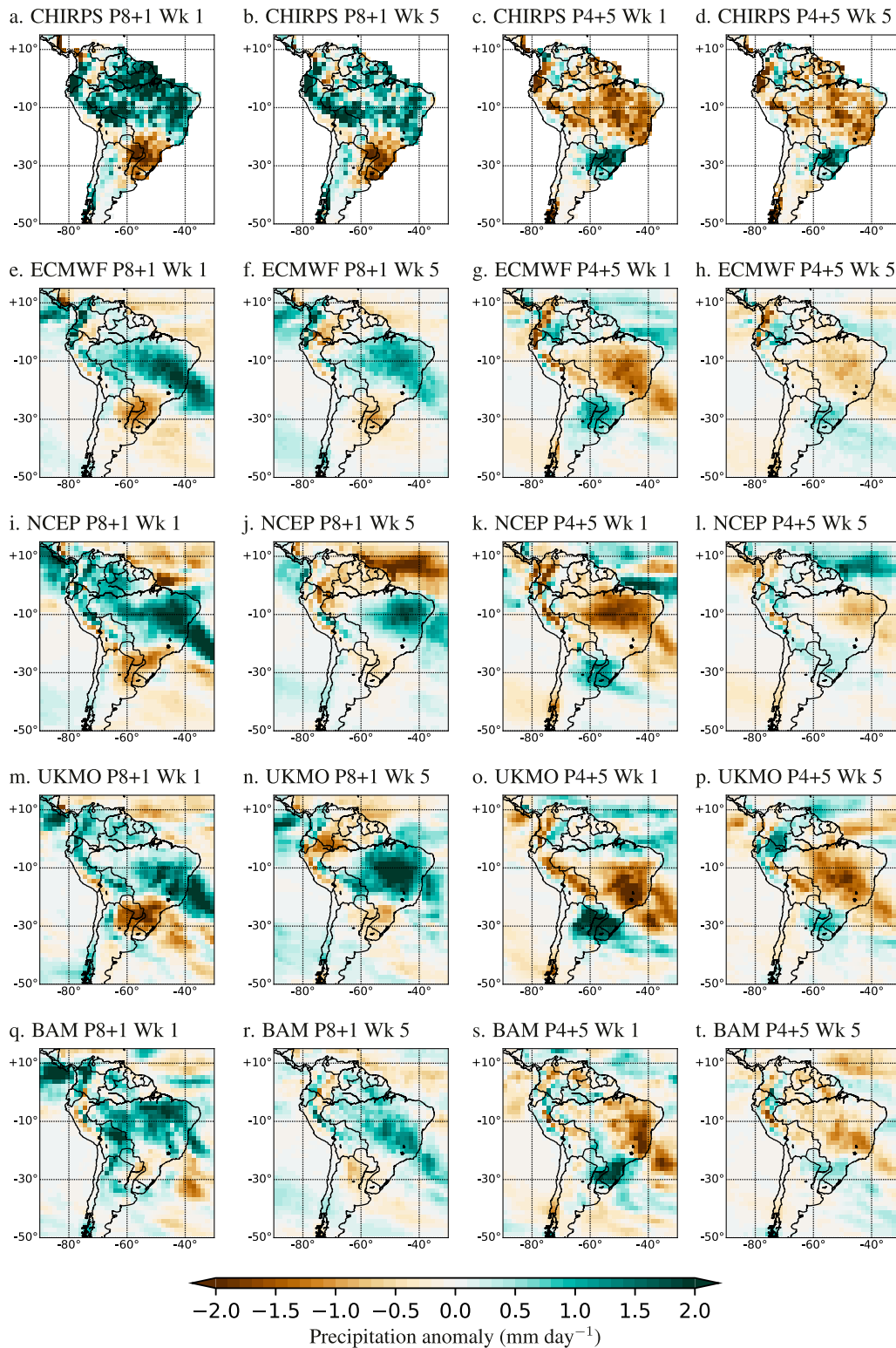


FIG. 12. NDJFM mean rainfall anomaly ( $\text{mm day}^{-1}$ ) for each pair of MJO phases, using only strong MJO days. For CHIRPS, we show composites of strong MJO in phases (a),(b) 8 + 1 and (c),(d) 4 + 5, for validity windows corresponding to week-1 reforecasts in (a) and (c) and week 5 reforecasts in (b) and (d), based on UKMO initialization dates. For S2S models, we show composites of strong MJO during week 1 in (c),(g),(k),(o) phases 8 + 1 and (e),(i),(m),(q) phases 4 + 5, as well as during week 5 in (d),(h),(l),(p) phases 8 + 1 and (f),(j),(n),(r) phases 4 + 5. MJO phases are based on observations for CHIRPS and on model output for the S2S models; the latter are composited on MJO phase at the validity time, not the initialization time.

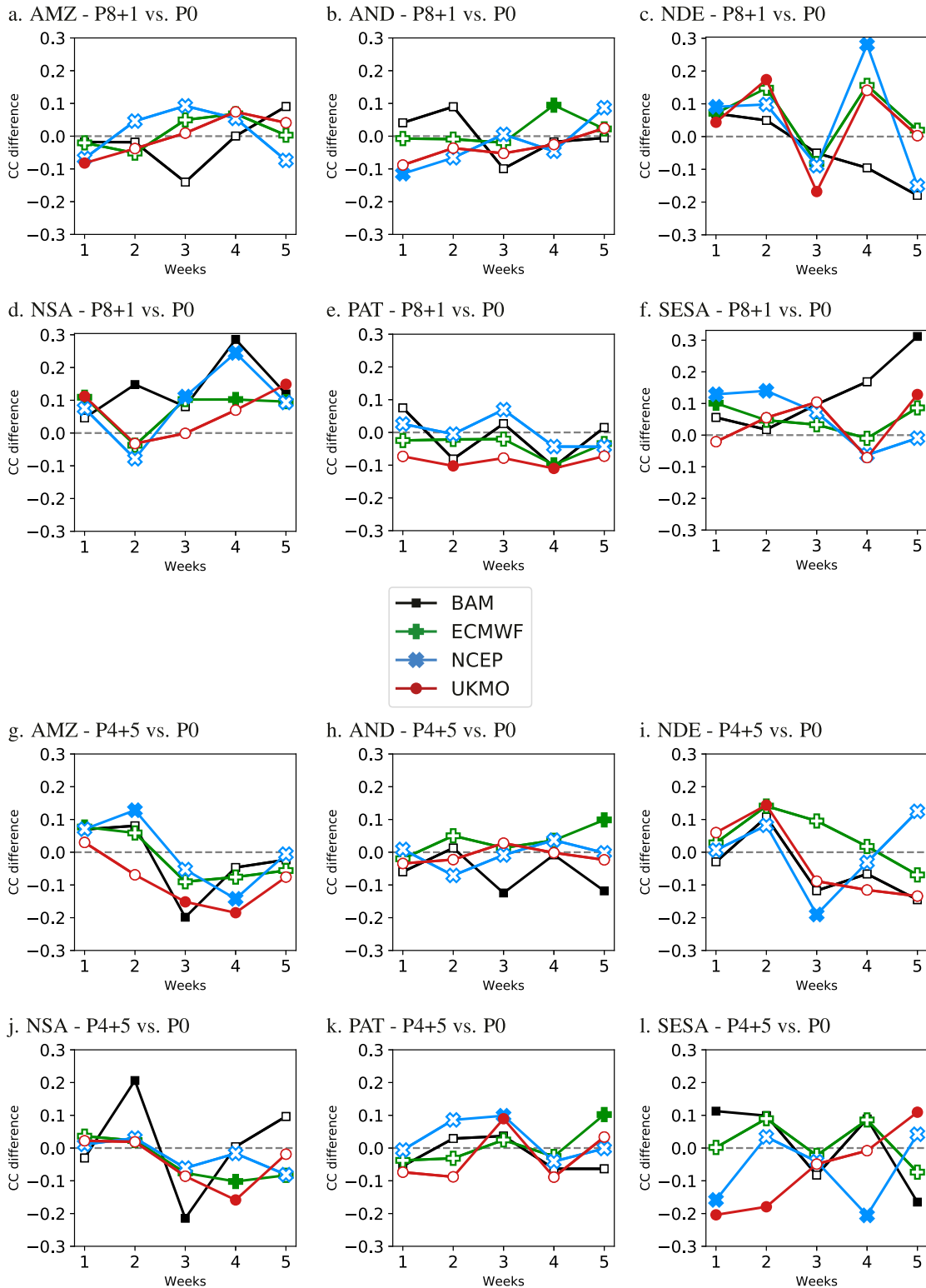


FIG. 13. Difference in regionally averaged S2S performance between strong and weak MJO, defined as the difference in regionally averaged CCs (between CHIRPS and S2S weekly rainfall anomalies) between NDJFM reforecasts started on strong MJO days in (a)–(f) phases 8 + 1 and (g)–(l) phases 4 + 5, compared to reforecasts started on weak MJO days in any phase (phase 0). Filled symbols show where 70% of strong MJO reforecast samples have a CC outside the 90% confidence interval for the CC of weak MJO reforecast samples, based on resampling each dataset 1000 times (with replacement). Regions are Amazonia (AMZ) in (a), Andes (AND) in (b), northeastern South America (NDE) in (c), northern South America (NSA) in (d), Patagonia (PAT) in (e), and southeastern South America (SESA) in (f). The regions are shown in Fig. 1a.

rather than to a climatological forecast, would test this hypothesis. We chose to use a climatological reference to avoid altering the reference forecast between models initialized on different dates (i.e., to use the same reference forecast for all models). BSSs would almost certainly be even lower if evaluated against a persistence forecast, as persistence is typically more skillful than climatology. An alternative evaluation strategy would compare conditional S2S performance by MJO and ENSO to a conditional climatological forecast, constructed using the probabilities of each rainfall category conditional on MJO or ENSO phase. A conditional climatological forecast would likely be more skillful than an unconditional climatological forecast, reducing S2S skill estimates. As our results show mostly insignificant differences between conditional and unconditional S2S performance, we expect that using a conditional climatological reference would degrade conditional performance.

Alternatively, models may perform more poorly in climatologically wetter regions due to the high contributions to mean rainfall from intense, short-lived events. Southern Amazonia is a prime example, where up to 70% of annual rainfall comes from a handful of events that last 4–15 days (e.g., Rao and Hada 1990). If these events are unpredictable at S2S lead times, then S2S forecasts will suffer from the essentially random nature of these events. Even at week 1, CCs in the southern Amazonia are below 0.4 (Fig. 4) and BSSs are near zero (Fig. 7) in all models. Yet another explanation for spatial variability in performance is the density of verifying observations. CHIRPS calibrates satellite rainfall against gauge measurements, but gauge density is much higher in northeastern and southeastern South America—where performance is highest—and lowest in southern Amazonia and near the Andes—where performance is also lowest. Thus, model performance may be artificially degraded by observational uncertainty, raising the possibility that the relationship between performance and climatological rainfall is spurious.

Our characterization of model performance is based on a very limited period: 1999–2010. We chose to verify the four models over their common period, which is limited by the NCEP and BAM reforecast periods. This may penalize ECMWF and UKMO, which have longer reforecasts (Table 1), but we chose a clean comparison over a larger sample size. A longer period would allow a more robust performance estimate, particularly for the conditional analysis. The limited sample size affects BAM most severely (Table 2). Further, we chose to construct lagged ensembles for ECMWF and NCEP relative to the UKMO initialization dates, using an 8-day window. Our sensitivity test in Fig. 6, in which we use only one ECMWF initialization and only a 2-day window for NCEP, shows that using larger lagged ensembles penalizes ECMWF and NCEP at short lead times (i.e., weeks 1 and 2, when the lag is long compared to the lead time) but benefits those models at longer lead times through an increased signal-to-noise ratio. We recommend further, more comprehensive study of the effects of ensemble size and lagged ensembles on perceptions of S2S prediction quality. There is likely no optimal method for comparing models with variations in reforecast period, ensemble size and initialization frequency, but it is important to

document the choices made and consider their effect on our conclusions.

## 5. Conclusions

South American rainfall variability during the main wet season (austral summer, November–March) is driven locally by the SACZ, modulated by large-scale phenomena such as ENSO on interannual scales, the MJO on subseasonal scales and midlatitude Rossby waves on synoptic scales (e.g., Grimm and Tedeschi 2009; Hirata and Grimm 2017; Grimm 2019). Despite considerable research into the mechanisms of subseasonal rainfall variability in South America, there are few comprehensive assessments of contemporary prediction systems, although S2S forecasts may have useful performance out to two to four weeks ahead in austral summer (Hirata and Grimm 2017; de Andrade et al. 2018) and autumn (Coelho et al. 2018). Successful prediction at these lead times would allow users, such as farmers and dam managers, to take effective action to mitigate damage and protect lives, livelihoods and ecosystems (e.g., Laux et al. 2008; Moron et al. 2009). Recent advances in S2S prediction suggest such performance may be possible (Vitart et al. 2016; White et al. 2017). We evaluate mean biases, errors and prediction quality for weekly November–March rainfall at 1–5-week lead times, using reforecasts of 1999–2010 from four S2S models: ECMWF, UKMO, NCEP and BAM. For prediction quality, we evaluate both unconditional performance and skill (i.e., using all reforecast data) and performance conditioned on the phase of the MJO and ENSO. Conditional evaluations are essential to identify potential “windows of opportunity”: certain conditions, such as local or large-scale circulation regimes, under which forecast performance increases. We measure performance by CC and skill by BSS (against a climatological reference forecast). “Useful performance” is defined as a CC statistically significantly different from zero (at 5%); “useful skill” is defined as  $BSS \geq 0$ .

All four models show biases in mean South American rainfall, most of which are established by week 1 and vary little thereafter (Fig. 1). All models underestimate mean Amazonian rainfall, where observed rainfall is high, and overestimate rainfall near Andean topography. Root-mean-squared errors grow more strongly with lead time and show smaller spatial variations than mean biases, suggesting models benefit from compensating errors in regions of low bias (Fig. 2). Biases are smallest in ECMWF and largest in NCEP; errors are smaller in ECMWF and BAM and larger in NCEP and UKMO.

When measured by CC, performance is useful in most models and regions at week 1 and week 2, although performance is lower over southern Amazonia and near the Andes (Figs. 4 and 7). By week 3, useful performance remains only over northern, northeastern, and southeastern South America; there is no useful performance beyond week 3. Performance is higher in areas with low to moderate rainfall, or low to moderate subseasonal rainfall variability, than in areas with high rainfall or high subseasonal variability (Fig. 5). When measured by BSS, skill declines more quickly: only ECMWF shows

skill in week 2, and then only for northern, northeastern, and southeastern South America. Higher CC performance in eastern South America is encouraging for potential application of S2S forecasts to the agricultural and hydropower sectors in those regions, though skill (measured by BSS) is low in many regions, including those with highest population density and greatest agricultural production. BSS for upper-tercile rainfall is higher in ECMWF and NCEP and lower in UKMO and BAM (Figs. 7 and 8). UKMO and BAM show BSS < 0 at week 1 for upper-tercile and lower-tercile (not shown) rainfall in almost all regions, even at week 1. No model has BSS > 0 for middle-tercile rainfall (not shown). Higher values of CC than BSS suggests that models capture the variability of weekly rainfall anomalies better than the intensity distribution of rainfall anomalies.

At week 1, all models represent well the spatial pattern and magnitude of ENSO-driven rainfall anomalies (Fig. 9). By week 5, these anomalies weaken substantially, suggesting inability to maintain the ENSO-driven anomalous meridional overturning circulation. Even at week 1, models struggle to capture the spatial pattern of MJO-driven rainfall anomalies: the observed equatorial signal is poorly represented and contracted to the south; the opposite-signed subtropical anomalies are too weak (Fig. 12). These anomalies weaken further by week 5, particularly in ECMWF and NCEP. UKMO performs relatively well over southeastern South America, but poorly for tropical South America. NCEP generates an equatorial rainfall signal opposite in sign to observations. BAM and ECMWF strongly damp MJO-associated anomalies with lead time. With few exceptions, conditional performance by ENSO (Fig. 11) and MJO (Fig. 13) phase does not substantially differ from unconditional performance, which may be linked to the errors in associated teleconnections.

Our results may be sensitive to limited common reforecast period of the S2S models (1999–2010) and limited rainfall observations in interior South America. The former particularly affects the conditional performance results; the latter particularly affects the perceived low performance over southern Amazonia and the Andes. S2S performance may be lower if assessed against a persistence forecast, or a conditional climatological forecast, rather than an unconditional climatological forecast. Our results may also be affected by our choices to compare models over their common period, not the full period of each model. Our results are affected by our choice to construct larger lagged ensembles for ECMWF and NCEP, rather than ensembles of similar size to UKMO. This choice slightly increases NCEP and ECMWF performance for lead times beyond week 2, but slightly reduces performance for weeks 1 and 2 (Fig. 6). Further research is needed to understand the effects of analysis choices on comparisons of performance in heterogeneous multimodel databases, such as the S2S database.

*Acknowledgments.* NPK was supported by an Independent Research Fellowship from the Natural Environment Research Council (NE/L010976/1). MY, AC, LG, SJW, and CEH were supported by the Climate Science for Services Partnership Brazil Project (CSSP-Brazil) funded by the Newton Fund. NPK and MY were also supported by the National Centre for

Atmospheric Science Atmospheric Hazards in Developing Countries: Risk Assessment and Early Warning (ACREW) Programme. CASC thanks CNPq, Process 305206/2019-2, and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Process 2015/50687-8 (CLIMAX Project) for the support received.

## REFERENCES

- Bombardi, R. J., K. V. Pegion, J. L. Kinter, B. A. Cash, and J. M. Adams, 2017: Sub-seasonal predictability of the onset and demise of the rainy season over monsoonal regions. *Front. Earth Sci.*, **5**, 14, <https://doi.org/10.3389/feart.2017.00014>.
- , L. Trenary, K. Pegion, B. Cash, T. DelSole, and J. L. Kinter, 2018: Seasonal predictability of summer rainfall over South America. *J. Climate*, **31**, 8181–8195, <https://doi.org/10.1175/JCLI-D-18-0191.1>.
- Carvalho, L. M. V., C. Jones, and B. Liebmann, 2004: The South Atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall. *J. Climate*, **17**, 88–108, [https://doi.org/10.1175/1520-0442\(2004\)017<0088:TSACZI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0088:TSACZI>2.0.CO;2).
- , A. E. Silva, C. Jones, B. Liebmann, P. L. Silva Dias, and H. R. Rocha, 2010: Moisture transport and intraseasonal variability in the South American monsoon system. *Climate Dyn.*, **36**, 1865–1880, <https://doi.org/10.1007/s00382-010-0806-2>.
- Cavalcanti, I. F. A., J. A. Marengo, L. M. Alves, and D. F. Costa, 2017: On the opposite relation between extreme precipitation over west Amazon and southeastern Brazil: Observations and model simulations. *Int. J. Climatol.*, **37**, 3606–3618, <https://doi.org/10.1002/joc.4942>.
- Coelho, C. A., M. A. Firpo, and F. M. de Andrade, 2018: A verification framework for South American sub-seasonal precipitation predictions. *Meteor. Z.*, **27**, 503–520, <https://doi.org/10.1127/metz/2018/0898>.
- Coomes, O. T., M. Lapointe, M. Templeton, and G. List, 2016: Amazon River flow regime and flood recessional agriculture: Flood stage reversals and risk of annual crop loss. *J. Hydrol.*, **539**, 214–222, <https://doi.org/10.1016/j.jhydrol.2016.05.027>.
- de Andrade, F. M., C. A. S. Coelho, and I. F. A. Cavalcanti, 2018: Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dyn.*, **52**, 5451–5475, <https://doi.org/10.1007/s00382-018-4457-z>.
- DelSole, T., L. Trenary, and M. K. Tippett, 2017: The weighted-average lagged ensemble. *J. Adv. Model. Earth Syst.*, **9**, 2739–2752, <https://doi.org/10.1002/2017MS001128>.
- de Oliveira Vieira, S., P. Satyamurty, and R. V. Adreoli, 2013: On the South Atlantic convergence zone affecting southern Amazonia in austral summer. *Atmos. Sci. Lett.*, **14**, 1–6, <https://doi.org/10.1002/asl2.401>.
- de Souza Custodio, M., R. Porfirio da Rocha, T. Ambrizzi, P. L. Vidale, and M.-E. Demory, 2017: Impact of increased horizontal resolution in coupled and atmosphere-only models of the HadGEM1 family upon the climate patterns of South America. *Climate Dyn.*, **48**, 3341–3364, <https://doi.org/10.1007/s00382-016-3271-8>.
- Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, <https://doi.org/10.1038/sdata.2015.66>.
- Gan, M. A., V. E. Kousky, and C. F. Ropelewski, 2004: The South America monsoon circulation and its relationship to rainfall

- over west-central Brazil. *J. Climate*, **17**, 47–66, [https://doi.org/10.1175/1520-0442\(2004\)017<0047:TSAMCA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0047:TSAMCA>2.0.CO;2).
- Goddard, L. G., and M. Dille, 2005: El Niño: Catastrophe or opportunity. *J. Climate*, **18**, 651–665, <https://doi.org/10.1175/JCLI-3277.1>.
- Gottschalck, J., and Coauthors, 2010: A framework for assessing operational Madden–Julian oscillation forecasts: A CLIVAR MJO Working Group project. *Bull. Amer. Meteor. Soc.*, **91**, 1247–1258, <https://doi.org/10.1175/2010BAMS2816.1>.
- Grimm, A. M., 2011: Interannual climate variability in South America: Impacts on seasonal precipitation, extreme events, and possible effects of climate change. *Stochastic Environ. Res. Risk Assess.*, **25**, 537–554, <https://doi.org/10.1007/s00477-010-0420-1>.
- , 2019: Madden–Julian Oscillation impacts on South American summer monsoon season: Precipitation anomalies, extreme events, teleconnections, and role in the MJO cycle. *Climate Dyn.*, **53**, 907–932, <https://doi.org/10.1007/s00382-019-04622-6>.
- , and P. L. Silva Dias, 1995: Analysis of tropical–extratropical interactions with influence functions of a barotropic model. *J. Atmos. Sci.*, **52**, 3538–3555, [https://doi.org/10.1175/1520-0469\(1995\)052<3538:AOTIWI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<3538:AOTIWI>2.0.CO;2).
- , and R. G. Tedeschi, 2009: ENSO and extreme rain events in South America. *J. Climate*, **22**, 1589–1609, <https://doi.org/10.1175/2008JCLI2429.1>.
- , and J. P. J. Saboia, 2015: Interdecadal variability of the South American precipitation in the monsoon season. *J. Climate*, **28**, 755–775, <https://doi.org/10.1175/JCLI-D-14-00046.1>.
- , J. S. Pal, and F. Giorgi, 2007: Connection between spring conditions and peak summer monsoon rainfall in South America: Role of soil moisture, surface temperature, and topography in eastern Brazil. *J. Climate*, **20**, 5929–5945, <https://doi.org/10.1175/2007JCLI1684.1>.
- Guimarães, B. S., C. A. S. Coelho, S. J. Woolnough, P. Y. Kubota, C. F. Bastarz, S. N. Figueroa, J. P. Bonatti, and D. C. de Souza, 2020: Configuration and hindcast quality assessment of a Brazilian global sub-seasonal prediction system. *Quart. J. Roy. Meteor. Soc.*, **146**, 1067–1084, <https://doi.org/10.1002/qj.3725>.
- Hirata, F. E., and A. M. Grimm, 2017: The role of synoptic and intraseasonal anomalies on the life cycle of rainfall extremes over South America: Non-summer conditions. *Climate Dyn.*, **49**, 313–326, <https://doi.org/10.1007/s00382-016-3344-8>.
- , and —, 2018: Extended-range prediction of South Atlantic convergence zone rainfall with calibrated CFSv2 forecast. *Climate Dyn.*, **50**, 3699–3710, <https://doi.org/10.1007/s00382-017-3836-1>.
- Hudson, D., O. Alves, H. H. Hendon, and A. G. Marshall, 2011: Bridging the gap between weather and seasonal forecasting: Intraseasonal forecasting for Australia. *Quart. J. Roy. Meteor. Soc.*, **137**, 673–689, <https://doi.org/10.1002/qj.769>.
- Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Laux, P., H. Kunstmann, and A. Bardossy, 2008: Predicting the regional onset of the rainy season in West Africa. *Int. J. Climatol.*, **28**, 329–342, <https://doi.org/10.1002/joc.1542>.
- Lee, S. S., B. Wang, D. E. Waliser, J. M. Neena, and J. Y. Lee, 2015: Predictability and prediction skill of the boreal summer intraseasonal oscillation in the Intraseasonal Variability Hindcast Experiment. *Climate Dyn.*, **45**, 2123–2135, <https://doi.org/10.1007/s00382-014-2461-5>.
- 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>.
- Liebmann, B., G. N. Kiladis, J. A. Marengo, T. Ambrizzi, and P. Glick, 1999: Submonthly convective variability over South America and the South Atlantic convergence zone. *J. Climate*, **12**, 1877–1891, [https://doi.org/10.1175/1520-0442\(1999\)012<1877:SCVOSA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<1877:SCVOSA>2.0.CO;2).
- MacLachlan, C., and Coauthors, 2015: Global seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, <https://doi.org/10.1002/qj.2396>.
- Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, [https://doi.org/10.1175/1520-0469\(1971\)028<0702:DOADOI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2).
- Marteau, R., B. Sultan, V. Moron, A. Alhassane, C. Baron, and S. B. Traore, 2011: The onset of the rainy season and farmers’ sowing strategy for pearl millet cultivation in southwest Niger. *Agric. For. Meteorol.*, **151**, 1356–1369, <https://doi.org/10.1016/j.agrformet.2011.05.018>.
- Moron, V., A. R. Robertson, and R. Boer, 2009: Spatial coherence and seasonal predictability of monsoon onset over Indonesia. *J. Climate*, **22**, 840–850, <https://doi.org/10.1175/2008JCLI2435.1>.
- Neena, J. M., J. Y. Lee, D. E. Waliser, B. Wang, and X. Jiang, 2014: Predictability of the Madden–Julian oscillation in the Intraseasonal Variability Hindcast Experiment (ISVHE). *J. Climate*, **27**, 4531–4543, <https://doi.org/10.1175/JCLI-D-13-00624.1>.
- Nogueira, S. C., M. Moreira, and M. L. Volpato, 2018: Evaluating precipitation estimates from Eta, TRMM and CHIRPS data in the south-southeast region of Minas Gerais state—Brazil. *Remote Sens.*, **10**, 313, <https://doi.org/10.3390/rs10020313>.
- Paredes-Trejo, F. J., H. Barbosa, and T. L. Kumar, 2017: Validating CHIRPS-based satellite precipitation estimates in northeast Brazil. *J. Arid Environ.*, **139**, 26–40, <https://doi.org/10.1016/j.jaridenv.2016.12.009>.
- 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>.
- Rao, V. B., and K. Hada, 1990: Characteristics of rainfall over Brazil: Annual variations and connections with Southern Oscillation. *Theor. Appl. Climatol.*, **42**, 81–91, <https://doi.org/10.1007/BF00868215>.
- Robertson, A. W., and C. R. Mechoso, 2000: Interannual and interdecadal variability of the South Atlantic convergence zone. *Mon. Wea. Rev.*, **128**, 2947–2957, [https://doi.org/10.1175/1520-0493\(2000\)128<2947:IAIVOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2947:IAIVOT>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>.
- Trenary, L., T. DelSole, M. K. Tippett, and K. Pegion, 2017: A new method for determining the optimal lagged ensemble. *J. Adv. Model. Earth Syst.*, **9**, 291–306, <https://doi.org/10.1002/2016MS000838>.
- , —, —, and —, 2018: Monthly ENSO forecast skill and lagged ensemble size. *J. Adv. Model. Earth Syst.*, **10**, 1074–1086, <https://doi.org/10.1002/2017MS001204>.
- 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>.
- , and A. W. Robertson, 2018: The subseasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate Atmos. Sci.*, **1**, 3, <https://doi.org/10.1038/s41612-018-0013-0>.
- , —, and D. L. T. Anderson, 2012: Subseasonal to seasonal prediction project: Briding 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>.
- Vörösmarty, C. J., and Coauthors, 2013: Extreme rainfall, vulnerability and risk: A continental-scale assessment for South America. *Philos. Trans. Roy. Soc. London*, **371A**, 20120408, <https://doi.org/10.1098/rsta.2012.0408>.
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development and index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>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>.
- Zhang, C., 2005: Madden–Julian oscillation. *Rev. Geophys.*, **43**, RG2003, <https://doi.org/10.1029/2004RG000158>.
- Zwiers, F. W., and H. von Storch, 1995: Taking serial correlation into account in tests of the mean. *J. Climate*, **8**, 336–351, [https://doi.org/10.1175/1520-0442\(1995\)008<0336:TSCIAI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<0336:TSCIAI>2.0.CO;2).