


A Multimodel Real-Time System for Global Probabilistic Subseasonal Forecasts of Precipitation and Temperature

ANDREW W. ROBERTSON^{},^a JING YUAN,^a MICHAEL K. TIPPETT,^b RÉMI COUSIN,^a KYLE HALL,^a NACHIKETA ACHARYA,^a BOHAR SINGH,^a ÁNGEL G. MUÑOZ,^a DAN COLLINS,^c EMERSON LAJOIE,^c AND JOHNNA INFANTI^c

^a *International Research Institute for Climate and Society, Columbia University, Palisades, New York*

^b *Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York*

^c *Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland*

(Manuscript received 11 September 2022, in final form 27 March 2023, accepted 1 April 2023)


ABSTRACT: A global multimodel probabilistic subseasonal forecast system for precipitation and near-surface temperature is developed based on three NOAA ensemble prediction systems that make their forecasts available publicly in real time as part of the Subseasonal Experiment (SubX). The weekly and biweekly ensemble means of precipitation and temperature of each model are individually calibrated at each grid point using extended logistic regression, prior to forming equal-weighted multimodel ensemble (MME) probabilistic forecasts. Reforecast skill of week-3–4 precipitation and temperature is assessed in terms of the cross-validated ranked probability skill score (RPSS) and reliability diagram. The multimodel reforecasts are shown to be well calibrated for both variables. Precipitation is moderately skillful over many tropical land regions, including Latin America, sub-Saharan Africa and Southeast Asia, and over subtropical South America, Africa, and Australia. Near-surface temperature skill is considerably higher than for precipitation and extends into the extratropics as well. The multimodel RPSS skill of both precipitation and temperature is shown to exceed that of any of the constituent models over Indonesia, South Asia, South America, and East Africa in all seasons. An example real-time week-3–4 global forecast for 13–26 November 2021 is illustrated and shown to bear the hallmarks of the combined influences of a moderate Madden–Julian oscillation event as well as weak–moderate ongoing La Niña event.

SIGNIFICANCE STATEMENT: This paper develops a system for forecasting of precipitation and temperatures globally over land, several weeks in advance, with a focus on biweekly averaged conditions between three and four weeks ahead. The system provides the likelihood of biweekly and weekly conditions being below, near, or above their long-term averages, as well the probability of exceeding (or not exceeding) any threshold value. Using historical data, the precipitation forecasts are demonstrated to have skill in many tropical regions, and the temperature forecasts are more widely skillful. While weather and seasonal range forecasts have become quite generally available, this is one of the first examples of a publicly available, calibrated multimodel probabilistic real-time forecasting system for the subseasonal range.

KEYWORDS: Climate prediction; Ensembles; Forecast verification/skill; Forecasting; Forecasting techniques; Probability forecasts/models/distribution

1. Introduction

The development of skillful, well-calibrated, multimodel subseasonal probabilistic forecast products is still in its infancy compared to the weather and seasonal forecast ranges

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

Hall's current affiliation: NOAA/Physical Sciences Laboratory, Boulder, Colorado.

Acharya's current affiliation: Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, and NOAA/Physical Sciences Laboratory, Boulder, Colorado.

Muñoz's current affiliation: Barcelona Supercomputing Center, Barcelona, Spain.

Corresponding author: Andrew W. Robertson, awr@iri.columbia.edu.

(Robertson and Vitart 2019). Previous work has examined whether probabilistic forecast skill can be enhanced through effective calibration and by multimodel ensemble techniques, as has been demonstrated for seasonal (Robertson et al. 2004; Hagedorn et al. 2005) and medium range (Hamill 2012) forecasting. Extended logistic regression (Wilks 2009) has previously been applied to three models from the WWRP/WCRP Subseasonal to Seasonal (S2S) project database (Vitart et al. 2012, 2017) with forecast-start dates on Mondays, to construct probabilistic tercile-category forecasts of weekly and week-3–4 precipitation (Vigaud et al. 2017b,a; Robertson et al. 2019), and temperature (Vigaud et al. 2019b). In these studies, the regression parameters were fitted separately at each gridpoint and lead time (from week 1 to 4) for the three models' ensemble-mean reforecasts following a leave-one-year-out approach. The fitted model was then used to produce tercile-category forecast probabilities for each model that were then averaged across models with equal weighting to form a multimodel ensemble (MME) probabilistic forecast. Subseasonal reforecast

DOI: 10.1175/WAF-D-22-0160.1

© 2023 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (<https://www.ametsoc.org/PUBSReuseLicenses>).

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 08/01/23 07:28 PM UTC

ensembles generally contain fewer ensemble members than in the seasonal forecasting case, so a straightforward counting of ensemble members exceeding a chosen threshold can lead to large errors, further motivating the regression approach based on the models' ensemble means; [Tippett et al. \(2007\)](#) have shown that regression models outperform counting estimates in the seasonal forecasting context, especially for small ensemble size. Logistic regression provides forecast probabilities directly, without the need to assume a parametric form for the forecast distribution.

In this paper, we expand the use of the extended logistic regression (ELR) for subseasonal forecasting to create a system for *real-time* multimodel subseasonal probabilistic forecasts of precipitation and temperature based on models available in real time from the Subseasonal Experiment "SubX" Project ([Pegion et al. 2019](#)) via the IRI Data Library. The intent is to further document the performance of ELR calibration and multimodel combination of subseasonal forecasts, and to illustrate the real-time forecasts from such a model.

The models and datasets are described in [section 2](#), followed by the methodology in [section 3](#). The results are presented in [section 4](#), including an assessment of reforecast skill and an example real-time forecast, which is interpreted in terms of concurrent MJO and ENSO conditions. The paper concludes with a summary and conclusions in [section 5](#).

2. Data and models

This paper makes use of the Subseasonal Experiment "SubX" Project ([Pegion et al. 2019](#)) database, which provides public access to 17 years of historical reforecasts (1999–2016), plus several years of real-time forecasts from seven U.S. and Canadian modeling groups.

Three general circulation models (GCMs) from SubX were selected, each having reforecasts for the period 1999–2016, and real-time forecasts from August 2017 to the present, initialized on Wednesdays. Only three SubX models have common Wednesday "start dates," which facilitates the construction of the MME combination used here. The SubX protocol ([Pegion et al. 2019](#)) encouraged forecast initialization on Wednesdays for operational relevance to NOAA Climate Prediction Center (CPC) which issues week-3–4 forecasts on Fridays. The three models are as follows: National Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2 (CFSv2) ([Saha et al. 2014](#)), the NCEP Environmental Modeling Center Global Ensemble Forecast System (GEFSv12) ([Zhou et al. 2016, 2017; Zhu et al. 2018](#)), and the National Oceanic and Atmospheric Administration (NOAA), Earth System Research Laboratory (ESRL) Flow-Following Icosahedral Model (FIM; [Sun et al. 2018a,b](#)). The FIM and CFSv2 models are coupled atmosphere–ocean GCMs with prognostic sea ice models, while the GEFSv12 is an uncoupled atmospheric GCM with a prescribed observed estimates of sea surface temperatures (SST) and sea ice concentrations ([Zhu et al. 2018](#)). The FIM and CFSv2 reforecast ensembles each contain 4 ensemble members, while the GEFSv12 has 11 members; the real-time forecast ensemble sizes are 4, 16, and 31,

respectively. All SubX model data are provided on a uniform $1^\circ \times 1^\circ$ latitude–longitude grid.

Observational precipitation data are taken from the 1996–present Global Precipitation Climatology Project (GPCP) daily precipitation with 1° resolution globally, derived from station observations and satellite measurements ([Huffman et al. 2001](#)). For temperature, we use CPC Global Unified Temperature daily surface temperature estimates available from 1979 to the present on a 0.5° grid (land only), averaged onto the SubX 1° grid. All datasets were obtained via the IRI Data Library.

3. Calibration and multimodel combination methodology

Forecast calibration of each GCM is based on extended logistic regression (ELR). Distributional regression is well suited to probability forecasting (i.e., when the predictand is a probability of cumulative exceedance rather than a measurable physical quantity) allowing the conditional distribution of a response variable to be derived given a set of explanatory predictors. In this context, logistic regression is extended to produce the cumulative probability p of not exceeding the observed climatological quantile value q :

$$p = \Pr\{V \leq q\},$$

by including the observed climatological quantile q as a second explanatory variable, in addition to the model-forecast ensemble mean \bar{x}_{ens} :

$$\ln\left(\frac{p}{1-p}\right) = f(\bar{x}_{\text{ens}}) + g(q). \quad (1)$$

Following [Vigaud et al. \(2017b, 2019b\)](#), we train such a model for each GCM at every grid point, start date and lead time (weekly or biweekly), using leave-one-year-out cross-validation over the 1999–2016 reforecast period, by regressing ensemble mean precipitation or temperature \bar{x}_{ens} against the tercile quantiles of the observations q . Note that \bar{x}_{ens} is not expressed as an anomaly, while the observed climatological quantile value q varies by calendar day and geographical location, expressing the climatology; it is included as a second predictor variable to ensure consistent results for different values of q ([Wilks 2009](#)). The tercile probabilities of the left-out year are then forecast, and the MME probabilities are constructed by simple averaging of the three individual forecast probabilities. Forecasts and reforecasts are produced for both weekly averages, as well as for week-3–4 biweekly averages, of precipitation and temperature. For forecasts of weekly averages, the climatological observed tercile categories are computed from the reforecast years following a leave-one year-out methodology, and over 3-week windows formed by the forecast target week and a week either side; a 6-week window is used for week-3–4 forecasts. The model is trained under cross validation using the remaining 17 reforecast years and these same 3-forecast sample windows; the training set thus contains $17 \times 3 = 51$ samples ([Vigaud et al. 2017b](#)). For precipitation, a "dry mask" is used, such that the forecasts are only produced when and where the

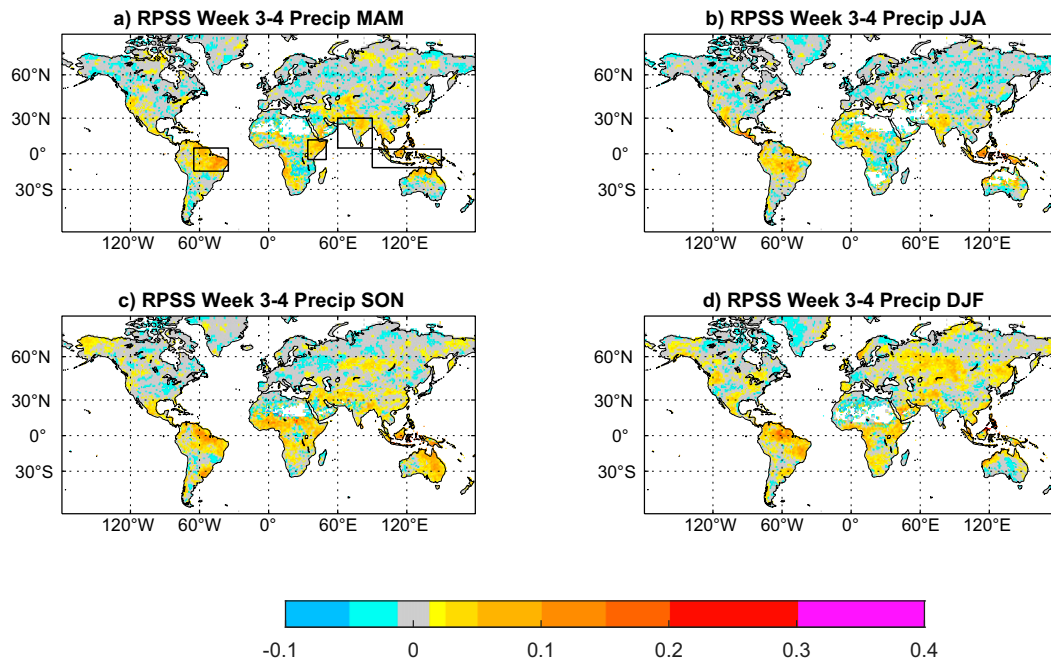


FIG. 1. Ranked probability skill score (RPSS) for week-3-4 MME precipitation reforecasts 1999–2016, stratified by season: (a) March–May (MAM), (b) June–August (JJA), (c) September–November (SON), and (d) December–February (DJF). The boxes used for the spatial averages in Fig. 2 are shown in (a).

33rd percentile of the observed climatology is nonzero (see Fig. 9a below for an example).

We follow the definitions of forecast weeks used by NOAA CPC, with each forecast week beginning on a Saturday and with the forecast issued one day before on the Friday. For our SubX forecasts initialized on Wednesdays, this means that week 1 is defined to be days 4–10 of the forecast. Weeks 2–4 correspond to days 11–17, 18–24, and 25–31, respectively, and weeks 3–4 to days 18–31. Note that these target window time ranges correspond to slightly longer lead times than those defined by Vigaud et al. (2017b, 2019b).

The ranked probability skill score (RPSS) has been widely used to describe the quality of categorical probabilistic forecasts and expresses the amount of error in the forecast probabilities, compared to categorical climatological probabilities (0.33 for terciles); it is a generalization of the Brier skill score to multiple categories which penalizes forecasts more if the category with the highest forecast probability is not adjacent to the observed one (Weigel et al. 2007). RPSS is used to measure the skill of the ELR-calibrated multimodel reforecasts under cross-validation. Each year of the 1999–2016 reforecast set was withheld for verification in turn, with the tercile categories and the ELR parameters determined from the remaining 17 years. The RPSS maps are computed for all Wednesday reforecast starts in a 3-month period. This yields a sample size of about 12 starts per year over 18 years, thus about 216. RPSS is computed relative to a reference of climatological forecast probabilities of (0.33, 0.33, 0.33), so that positive RPSS corresponds to exceeding the performance of assigning equiprobable forecast outcomes.

It should be noted, following Vigaud et al. (2017b, 2019b), that the tercile breaks include seasonality, so that the forecasts indicate the probability of conditions being below, near, or above the climatology normal for the forecast 1- or 2-week target period. This follows the conventional seasonal forecasting practice, in contrast to weather forecasts which are usually expressed as total fields.

4. Results

a. Reforecast skill

Seasonally averaged maps of RPSS for week-3-4 MME precipitation are shown in Fig. 1, for all reforecasts initialized in the corresponding season (1999–2016). Skill is largest in the tropics and subtropics and varies between regions and seasons, consistent with previous subseasonal (Li and Robertson 2015; Vigaud et al. 2017b,a, 2019a) and seasonal (e.g., Robertson et al. 2004; Barnston et al. 2010) predictability studies. Relatively few grid points exhibit negative RPSS (blue shading), and large negative RPSS values are rare (less than -0.05 ; dark blue). Large negative RPSS is an indication of miscalibration, and thus their absence indicates that there are no gross miscalibration issues. Positive RPSS values reach 0.1 to 0.2. It can be shown (Tippett et al. 2010) that in the idealized case of well-calibrated normally distributed forecasts:

$$RPSS \sim 1 - \sqrt{1 - r^2}, \tag{2}$$

where r is the anomaly correlation, indicating that an anomaly correlation value of 0.4 equates to an RPSS of 0.083.

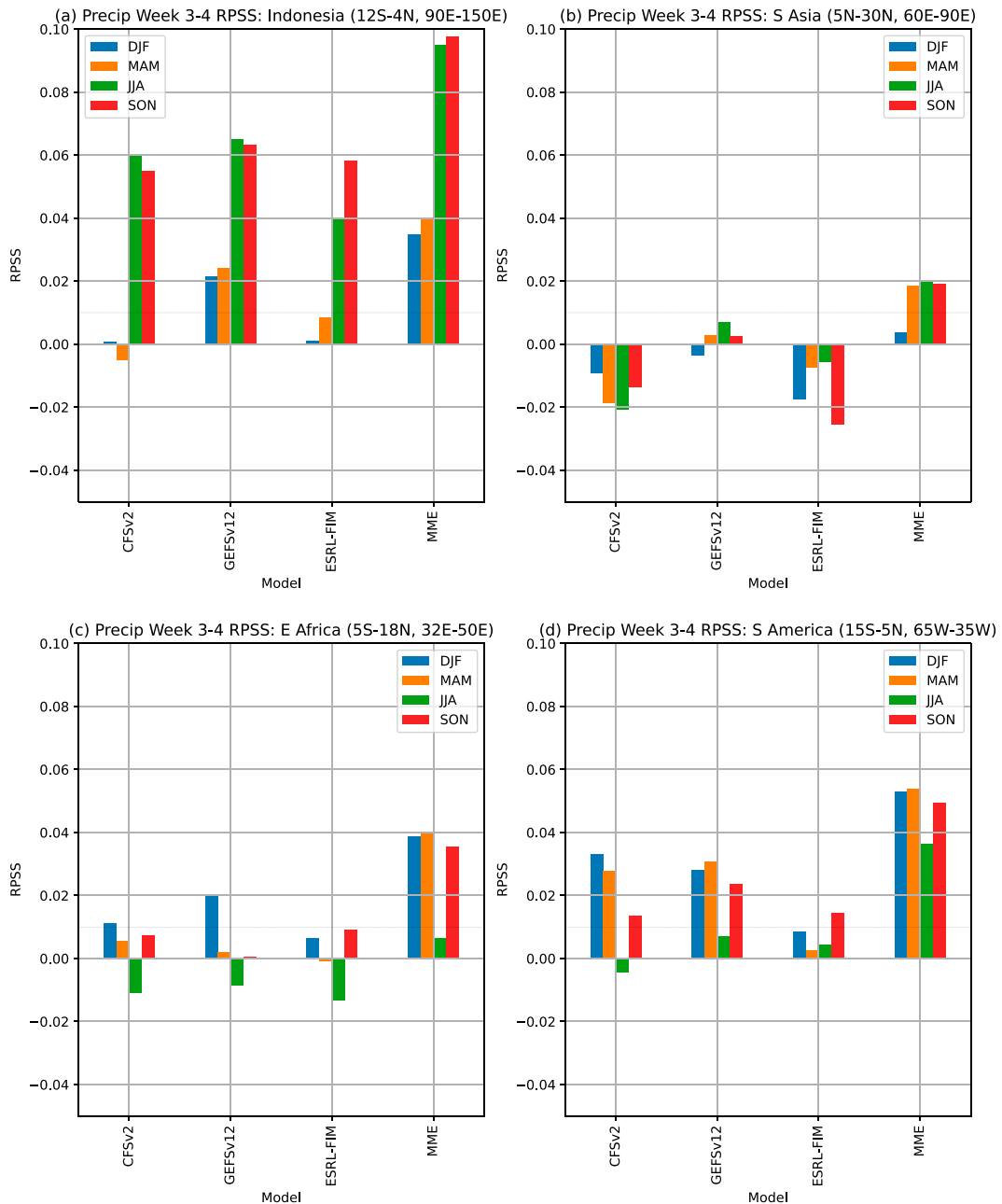


FIG. 2. Regionally averaged week-3-4 MME precipitation RPSS over land points stratified by season, for each model and the MME. (a) Indonesian region (12°S – 4°N , 90° – 150°E), (b) South Asia (5° – 30°N , 60° – 90°E), (c) East Africa (5°S – 18°N , 32° – 50°E), and (d) northern South America (15°S – 5°N , 65° – 35°W). The 0.01 RPSS value that corresponds approximately to the 5% significance threshold level is indicated by a horizontal line.

For reforecasts initialized in boreal spring (Fig. 1a), there is appreciable week-3-4 precipitation skill over northeast South America and around Uruguay, the Greater Horn of Africa, southwest and South Asia, and over Southeast Asia. For boreal summer reforecasts (Fig. 1b), skill is found over Central America and the Caribbean, Amazonia, parts of the Sahel, and South and Southeast Asia. During boreal fall (Fig. 1c), there is again high skill over Central America and the Caribbean, Northern, northeast and southeast South America, with

high skill levels across much of the Sahel, and central and east Africa, Southeast Asia, and eastern Australia. During boreal winter (Fig. 1d), the highest skill levels are found over northern South America and parts of the Maritime Continent, with positive skill extending over large parts of sub-Saharan and southern Africa, parts of southwest Asia, central Eurasia and China, as well as some parts of North America and Europe.

The skills of the individual models are compared with the MME in Fig. 2, using spatially averaged RPSS over four regions:

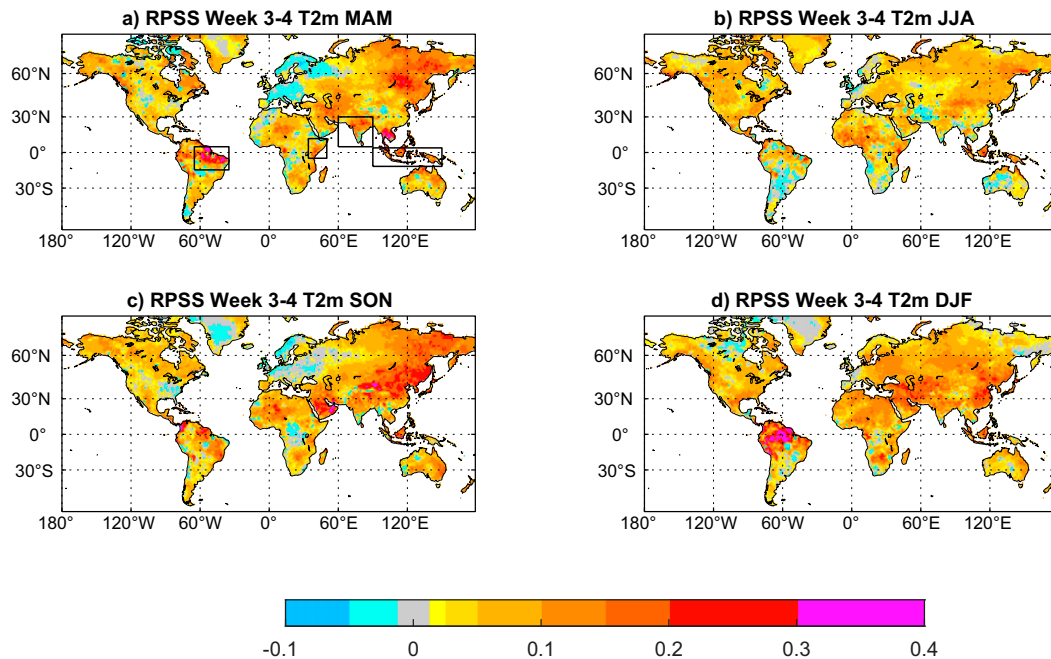


FIG. 3. Ranked probability skill score (RPSS) for week-3-4 MME 2-m temperature, stratified by season. Details as in Fig. 1. The boxes used for the spatial averages in Fig. 4 are shown in (a).

the Maritime Continent (Indonesia), South Asia, East Africa, and northern South America. Statistical significance of these regional averages is estimated using the following rule of thumb. The threshold for anomaly correlation significance at the 5% level can be approximate by $2/\sqrt{N}$, where N is the number of degrees of freedom (DelSole and Tippett 2022). Substituting this expression for r into the approximation in Eq. (2), yields

$$RPSS_{5\%} = 1 - \sqrt{\frac{N-4}{N}}. \quad (3)$$

For the 3-month season calculations of RPSS from the hindcasts 1999–2016, $N \approx 12 \times 18 = 216$ assuming independent weekly issued reforecasts, which yields $RPSS_{5\%} \approx 0.01$.

The MME spatially averaged skill exceeds that of any of the individual models in all four regions and in all seasons and is statistically different from zero in all four regions. The highest skill levels are found over the Indonesian region in boreal summer and fall, dipping over boreal winter and spring, mirroring the seasonal evolution of seasonal forecast skill there, which is controlled by characteristics of the Maritime Continent monsoon (e.g., Robertson et al. 2011). The individual models have no skill during boreal winter and spring over Indonesia yet the multimodel combination recovers some skill. Multimodel precipitation skill is also substantial over tropical South America in all seasons, and over East Africa except in JJA; it is lower though still statistically significant over South Asia, except in DJF.

These results provide additional evidence that multimodel combination is an effective means of boosting the skill of sub-seasonal forecasts, as has been demonstrated in many studies

on the seasonal forecasting time scale (e.g., Robertson et al. 2004; Hagedorn et al. 2005), and confirms the results of Vigaud et al. (2017b,a) who used the S2S project models. However, there is implicitly a larger ensemble size in the MME (4 + 11 + 4 members), so the improved skill could be due both to the use of multiple models and more members.

Turning to temperature, maps of the RPSS week-3-4 cross-validated reforecast skill of the MME by season are plotted in Fig. 3. Skill levels are generally much higher than for precipitation, consistent with higher persistence of the thermodynamic field, and with the presence of the anthropogenic warming trend, and confirm previous findings on the subseasonal scale (Wang and Robertson 2019; Vigaud et al. 2019b). As in the case of precipitation, the temperature skill is largely nonnegative, indicative that there are no gross miscalibration issues; however, there are some contiguous areas of small negative RPSS such as over Europe in MAM. Figure 4 shows the same regional averages in Fig. 2, for the three individual models and MME. Again, in all four regions and in all four seasons, the MME is always more skillful than any individual model. Individual model temperature skills are always positive.

While our focus is on biweekly week-3-4 averages, Fig. 5 shows the RPSS for individual weeks 1, 2, 3, and 4, for both precipitation (left) and temperature (right), averaged over the Indonesian region. The expected monotonic decrease in weekly precipitation skill is clearly evident from week 1 to week 4. While this decrease is also seen in temperature in JJA and SON—the seasons with the highest precipitation skill—it is less evident in DJF and MAM between weeks 2 and 4, which may reflect the stronger role of temperature persistence

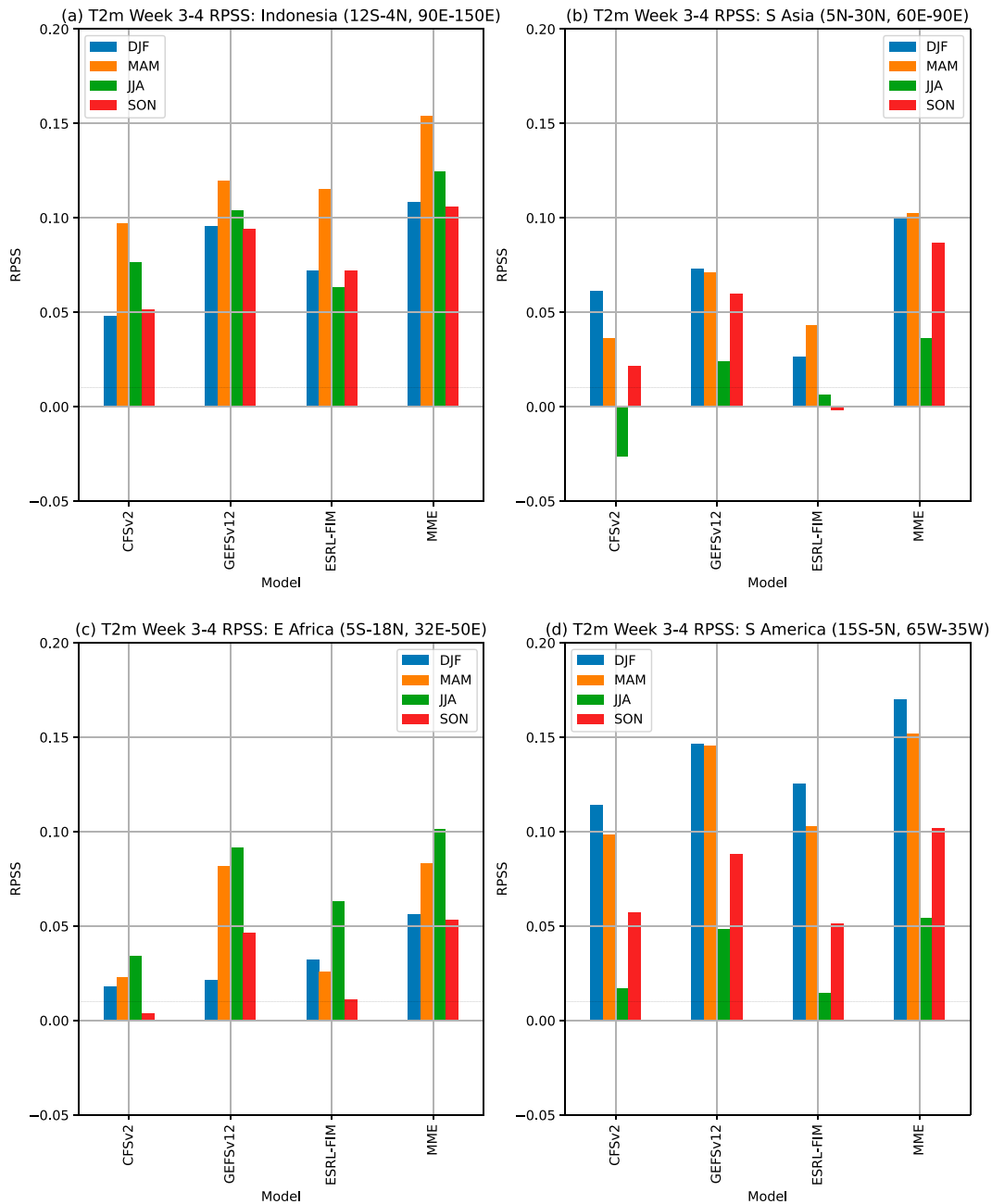


FIG. 4. Regionally averaged week-3-4 MME 2-m temperature RPSS over land points. Details as in Fig. 2.

or the trend during DJF and MAM when rainfall predictability is low. The other notable feature of Fig. 5 is the higher skill of the biweekly week-3-4 average compared to the week 3, in all seasons and for both variables. The longer 2-week averaging period can be expected to enhance signals that persist beyond a week, such as those associated with SST and MJO forcing, and to damp daily weather noise more than in a weekly average. The resulting increase in the signal-to-noise ratio in the biweekly average in Fig. 5 is seen to outweigh the reduction in signal between the week-3 and week-4 weekly averages.

Reliability diagrams are plotted in Fig. 6, for both variables, pooling the reforecasts over all tropical (blue curves) and extratropical land grid points (red curves), for each tercile category. Tropical precipitation exhibits a high degree of reliability (the curves fall along the diagonal), as well as sharpness (the curves extend over a significant range of forecast probabilities); this latter attribute is clear in the forecast histograms for the below-normal and above-normal categories, as is typical in seasonal precipitation forecasts (Barnston et al. 2010). Extratropical precipitation, exhibits much less sharpness and is also less reliable, as expected from the RPSS maps. The temperature reforecasts

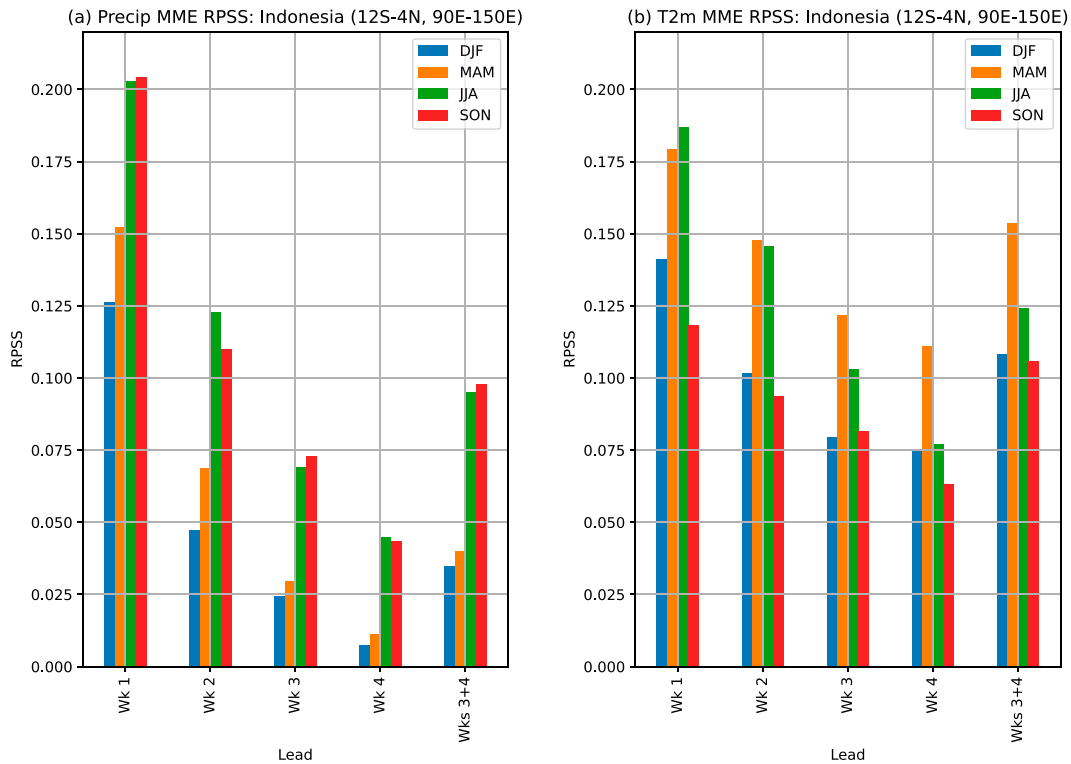


FIG. 5. Regionally averaged MME RPSS over the Indonesian region for individual weeks 1, 2, 3, and 4, and for weeks 3–4. (a) Precipitation and (b) 2-m temperature.

are considerably sharper than for precipitation, especially in the extratropics, and are generally reliable, except for higher probabilities of above-normal temperature which are overconfident (a forecast probability of 0.6 verifies in only 40% of cases). Unlike in the case of precipitation, near-normal temperature forecasts are both reliable and sharper.

b. Probabilistic forecasts in real time

The IRI has been issuing calibrated global probabilistic forecasts of biweekly precipitation based on SubX experimentally in real time every Friday since August 2018.¹ The regression parameters for the real-time forecasts are estimated using all the reforecast years 1999–2016 (no cross validation). Figure 7 shows an example of an MME forecast of precipitation and temperature in terciles format, issued in real time on 29 October 2021, for the 13–26 November 2021 week-3–4 period. This particular forecast was one of the more skillful ones and was thus chosen for illustration.

The forecast maps (Figs. 7a,c) show the probability of the dominant tercile category—the forecasted category with the highest probability—whenever that exceeds 35%. White areas on the maps correspond to grid points where the MME forecast does not deviate (within 2%) from climatological equal-odds probabilities. Figures 7b and 7d show the observed

percentile of the biweekly average, computed over the 1999–2016 period, obtained by ranking the observed biweekly average of precipitation and temperature against the 18 reforecast years, providing an indication how anomalous the 13–26 November 2021 conditions were compared with past 13–26 November periods.²

Below-normal precipitation is the dominant category of the forecast (Fig. 7a) over equatorial and eastern Africa, Uruguay, Chile, Colombia, and Sumatra, as well as broadly over Asia and the southern United States. The above-normal category dominant over much of northeastern South America, and parts of Southeast Asia and Australia. This general pattern of forecasted categories matches fairly well with the observed ones (Fig. 7b) over South America, eastern Africa and Sumatra, while the dry forecasts over central equatorial Africa and wet forecast over northern Australia are not matched by dry-observed percentiles.

The forecasted dominant category of temperature (Fig. 7c) also matches the observed category (Fig. 7d) in some areas, but not others, with near-normal temperatures indicated as most likely over much of the Northern Hemisphere land. Above-normal temperatures are forecasted over much of equatorial and southern Africa, western/southern South America,

¹ https://iridl.ldeo.columbia.edu/maproom/Global/ForecastsS2S/precip_subx.html.

² More precisely, the 6-week period centered on 13–26 November, 1999–2016 is used to define the reforecast climatology, to match that used in the ELR training. This results in less-noisy observed percentile maps than 2-week periods.

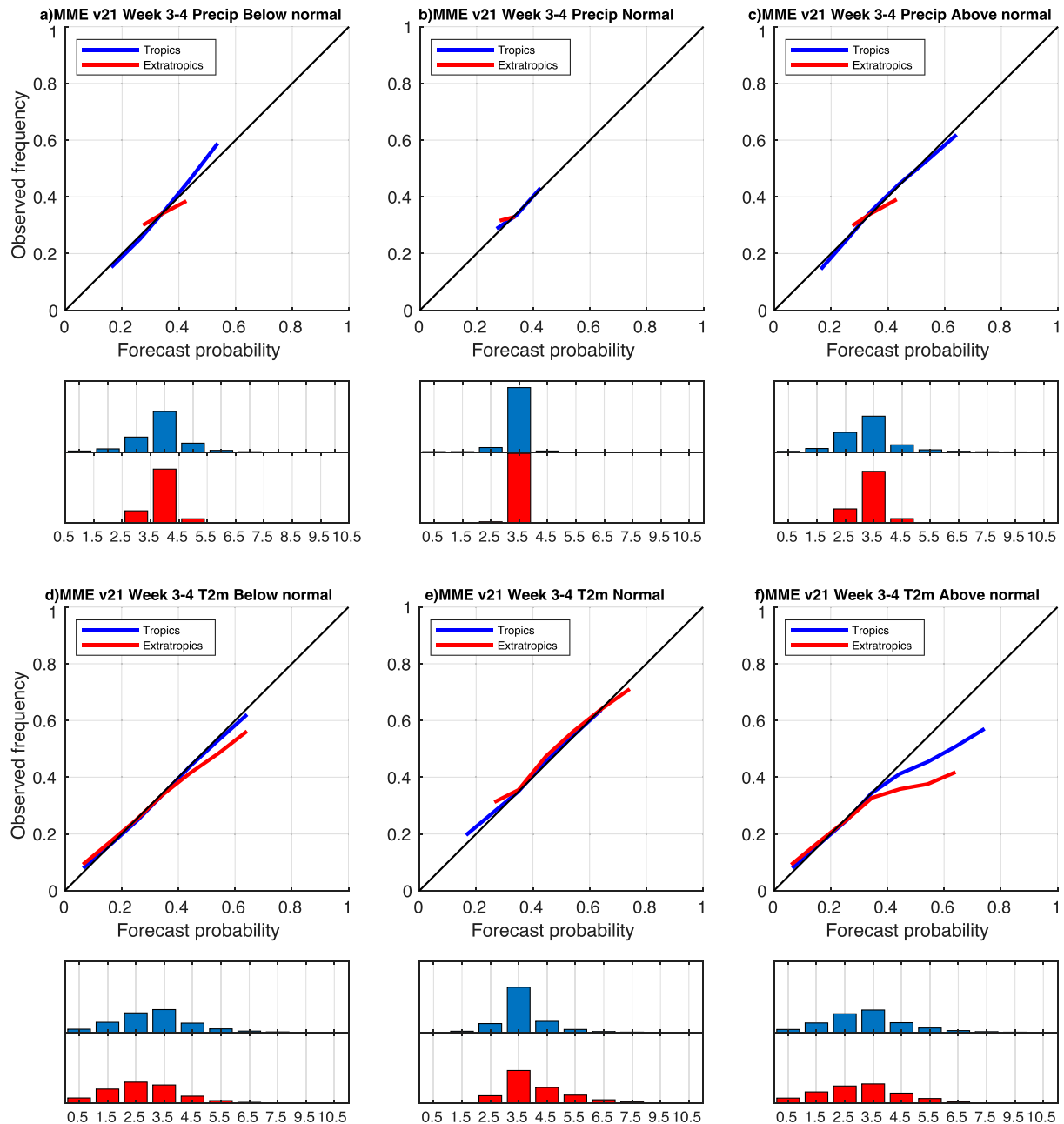


FIG. 6. Reliability diagrams for week-3-4 MME reforecasts [(a)–(c) precipitation, (d)–(f) 2-m temperature] pooled over the whole calendar year, and over all tropical land points (30°S – 30°N ; blue), and all extratropical land points (red), for reforecast probabilities of the (left) below-normal, (center) near-normal, and (right) above-normal tercile categories. The histograms show the frequency of reforecasts in each 0.1 probability bin (the 10 bin centers are labeled), with the ordinate scaled from 0% to 100%.

and parts of the Maritime Continent, with below-normal temperatures over much of Brazil, South/Southeast Asia, and Southwest Australia.

Verification of the single case in Fig. 7 provides intuition on the forecast's quality and interpretation in terms of the patterns of S2S climate drivers discussed for precipitation in the next subsection. However, to set the performance of this

particular forecast in context, the time series of RPSS integrated spatially over land 60°N – 60°S , is plotted in Fig. 8, with the score for the 29 October 2021 highlighted. Just as RPSS is typically used to sum the rank probability score (RPS) over time (normalized by its climatology, e.g., Robertson et al. 2004), the summation can instead be made over grid points in space for a specific date in time, to illustrate the time

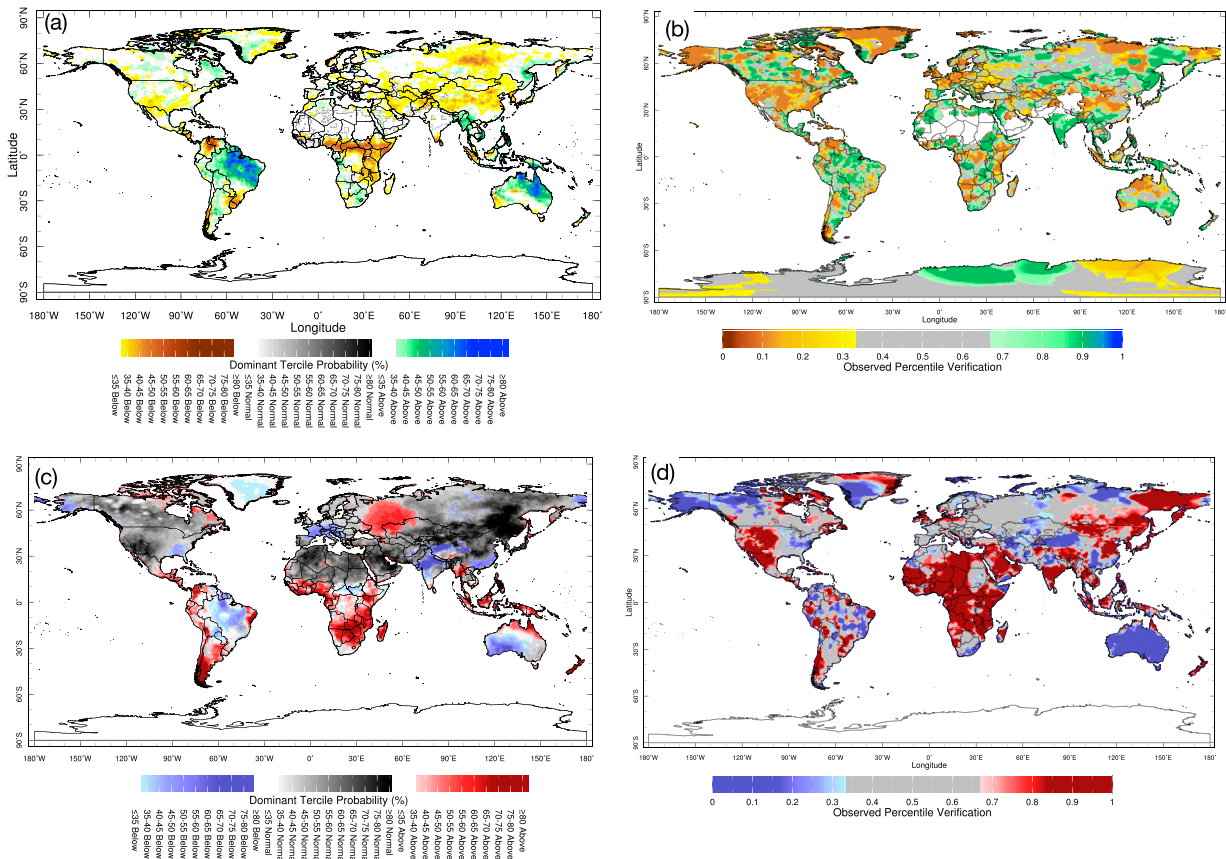


FIG. 7. Real-time week-3-4 MME probability forecast maps for the 13–26 Nov 2021 period, issued 29 Oct 2021, for (a) precipitation and (c) 2-m temperature. (b),(d) The observed percentile as verification. See text for further details. Plotted from the IRI Data Library.

evolution of the skill of probabilistic forecasts. Though not an outlier, this particular forecast is seen to be one of the more globally skillful over real-time forecasting period, particularly for precipitation.

c. MJO and ENSO impacts on the precipitation forecast

A moderate MJO event occurred in November 2021, propagating from phase 1 (Wheeler and Hendon 2004) on the forecast initialization date (27 October 2021) and persisting in MJO phase 4 during the 2-week forecast target period, 13–26 November 2021, with an amplitude of about one standard deviation (Fig. 9c). During phase 4, MJO convection is located over the Maritime Continent and western Indian Ocean, with anomalously dry conditions over the eastern Indian Ocean, tropical Africa, central America, and northern South America (Fig. 9b). Many of these features are consistent with the precipitation forecast map (Fig. 7a), here plotted over ocean as well as land to aid physical interpretation (Fig. 9a).³ However, the MJO state cannot account for the wet forecast over northern South America,

³ Unlike in Fig. 7a, the 35%–40% probability bins in Fig. 9a are also shaded in color to highlight the dry mask in this figure.

over northern Australia, nor the dryness forecasted over Uruguay.

November 2021 was also characterized by a moderate La Niña event (Fig. 10a), that also strongly impacted the week-3-4 forecast. A map of the correlations between the Niño-3.4 index and November precipitation anomalies (Fig. 10b) also resembles many aspects of the forecast, with negative correlations over the Maritime Continent, northern Australia, the eastern Indian Ocean, as well as over northern South America. Positive correlations extend over the tropical Pacific, the Caribbean, and the western Indian Ocean and East Africa. Figures 9 and 10 point to the importance of the combined impacts of both the MJO and ENSO on the week-3-4 forecast: over much of the tropics the impacts of both phenomena are seen to reinforce each other, while opposing each other over South America and northern Australia where La Niña’s impact dominates.

d. Flexible probability format forecasts

While tercile categories are the most commonly used presentation format for probabilistic seasonal climate forecasts and which we have used for our subseasonal forecasts too, probability of exceedances for specific user-chosen thresholds are often more relevant to particular forecast users (Barnston

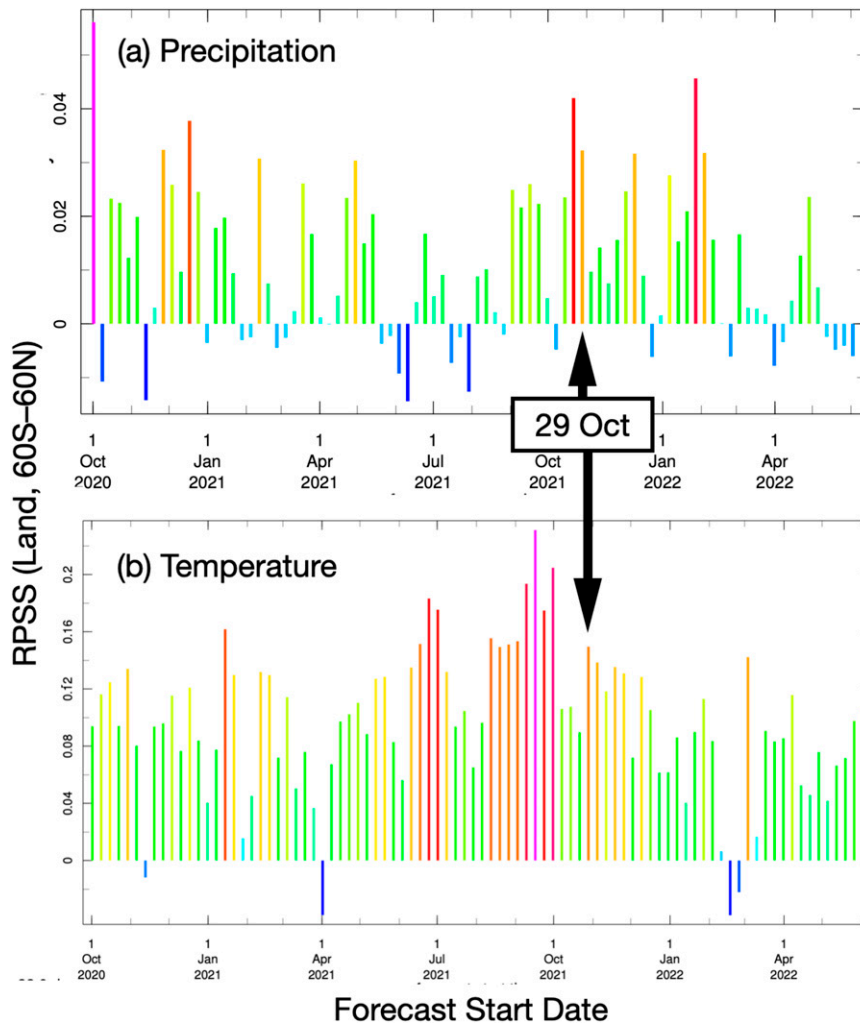


FIG. 8. Time series of RPSS score for the real-time week-3–4 MME forecasts issued 2 Oct 2020–2 Jun 2022 for (a) precipitation and (b) temperature. The forecast issue date is plotted on the abscissa with RPSS integrated over land points, 60°S–60°N on the ordinate (scale differs between panels). The 29 Oct 2021 forecast is highlighted. Colors are only indicative and are relative in each panel.

and Tippett 2014). Extended logistic regression enables the full forecast distribution to be derived and plotted at any grid-point. Figure 11 provides an example of the 13–26 November 2021 forecast for a location over Indonesia, showing the probability of exceedance (top) and the probability distribution (bottom), for both precipitation and temperature. These graphs can be plotted interactively via the online “Flexible Forecast” map rooms⁴ by clicking any land point on the map. Details of the computations are provided in the appendix.

The forecast at this location is wetter and warmer than the climatology, so that both forecast distributions are shifted to the right of their climatological counterparts. The mode of the precipitation forecast distribution is more than 10 mm week⁻¹

wetter than the climatology. The precipitation forecast distribution is visibly narrower than the precipitation climatological distribution (i.e., more confident), while the temperature forecast does not reduce the climatological uncertainty, while shifting it by about 1°C to the right. This may be associated with warmer SSTs around Indonesia during November 2021 (Fig. 10a). Figure 11 reflects the Gaussian nature of the temperature distributions, while the biweekly precipitation distributions are truncated at zero and asymmetrical; the latter resemble truncated Gaussian distributions or gamma distributions with large shape parameters.

5. Summary and conclusions

A global multimodel probabilistic subseasonal forecast system for precipitation and near-surface temperature has been

⁴ <https://iridl.ldeo.columbia.edu/maproom/Global/ForecastsS2S/index.html>.

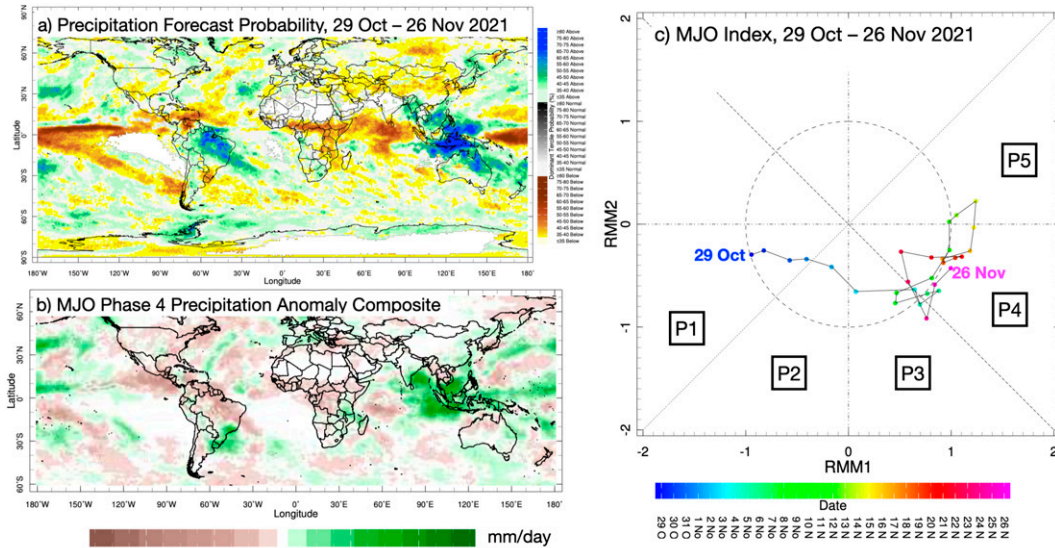


FIG. 9. MJO diagnostics. (a) Precipitation forecast map as in Fig. 7a, but showing both land and ocean areas; (b) MJO phase-4 anomaly composite of precipitation constructed over the October–December (OND) season, 1996–2021, using GPCP precipitation; and (c) observed bivariate Wheeler–Hendon real-time multivariate MJO (RMM) index from the forecast initialization date (27 Oct 2021) to the end of the week-3–4 forecast period (13–26 Nov 2021), with MJO phases 1–4 indicated as P1–P4. The forecast dry mask is indicated in white in (a). Plotted from the IRI Data Library.

developed based on three NOAA ensemble prediction systems that make their forecasts available publicly in real time, as part of the Subseasonal Experiment (SubX). The Saturday–Friday weekly and biweekly time-averaged raw model ensemble means of precipitation and temperature of each model are individually calibrated against historical data at each gridpoint using extended logistic regression, prior to forming an equal-weighted MME. The system has been implemented at IRI and has been run on a weekly basis, every Thursday, since mid-2018, and served publicly through a suite of virtual map rooms⁵ available in the IRI Data Library where weekly forecasts from week 1 to 4 can also be accessed. Reforecast skill of weeks 3–4 is assessed in terms of the cross-validated RPSS and reliability.

The multimodel reforecasts are shown to be well-calibrated for both variables. Precipitation is moderately skillful in many tropical land regions, including Latin America, sub-Saharan Africa, and Southeast Asia, and over subtropical South America, Africa, and Australia in some seasons (Fig. 1). Near surface temperature skill is considerably higher than for precipitation and it extends into the extratropics as well (Fig. 3). The multimodel combination RPSS skill of both precipitation and temperature is shown to exceed that of any of the constituent models for spatially averaged RPSS over the Maritime Continent, South Asia, South America, and East Africa, in all seasons (Figs. 2 and 4). The week-3–4 MME spatially averaged RPSS over the Maritime Continent is shown to exceed

that of the week-3 forecast in all seasons (Fig. 5); the longer 2-week averaging period will act to enhance signals that persist beyond a week, such as those associated with SST and MJO forcing, and to damp daily weather noise more than in a weekly average, both acting to enhance the signal-to-noise ratio. On average, week-3–4 tropical precipitation and temperature globally exhibit a high degree of reliability and sharpness, while extratropical precipitation sharpness is poor (Fig. 6). The temperature reforecasts are considerably sharper than for precipitation, especially in the extratropics, and are generally reliable, except for higher probabilities of above-normal temperature which are overconfident. Unlike in the case of precipitation, near-normal temperature forecasts are both reliable and sharper; this finding is unexpected from seasonal forecasting and deserves further research.

An example real-time week-3–4 forecast for 13–26 November 2021 is presented in terciles format (Fig. 7) and shown to bear the hallmarks of a moderate Madden–Julian oscillation together with a moderate La Niña. Active MJO convection conditions were observed over the Maritime Continent and western Indian Ocean during 13–26 November 2021 (Fig. 9a), with anomalously dry conditions inferred over the eastern Indian Ocean, tropical Africa, central America and northern South America (Fig. 9b). Many of these features are consistent with the precipitation forecast map (Fig. 7a), with the exception of much of South America and northern Australia. However, November 2021 was also characterized by a moderate La Niña event (Fig. 10a), whose canonical impacts (Fig. 10b) also resemble the week-3–4 forecast, with positive precipitation anomalies expected over the Maritime Continent, northern Australia, the eastern Indian Ocean, as well as much of northern South

⁵ <https://iridl.ldeo.columbia.edu/maproom/Global/ForecastsS2S/index.html>.

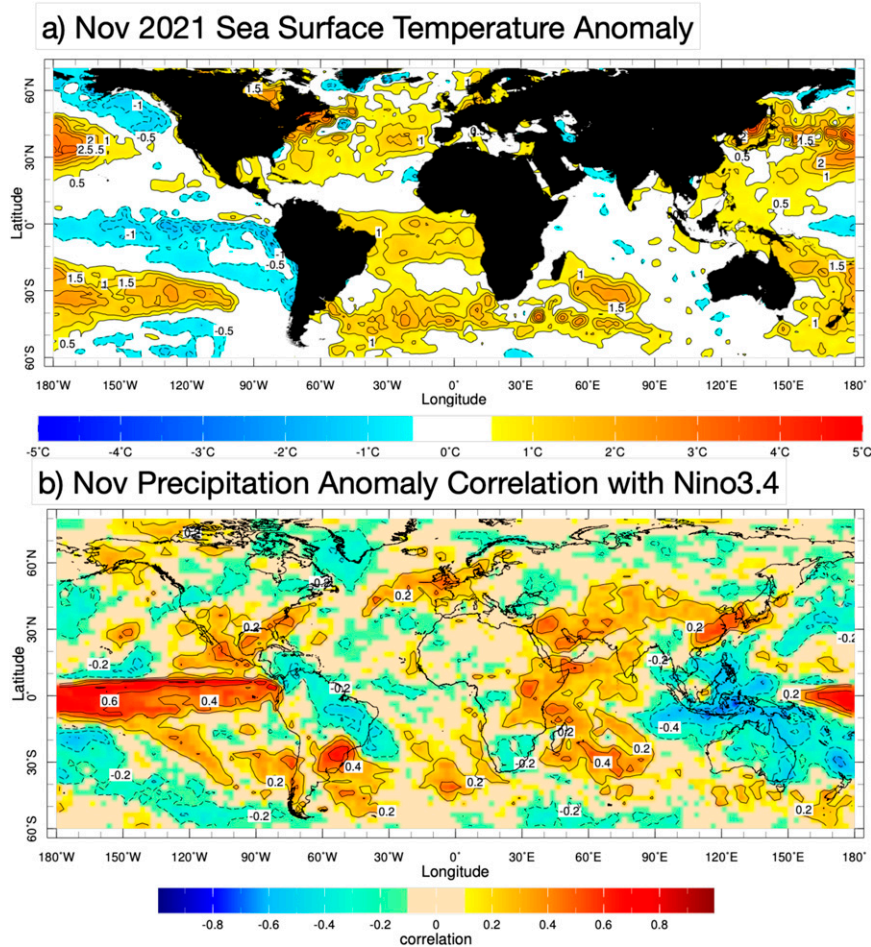


FIG. 10. (a) November 2021 sea surface temperature anomaly with respect to the 1991–2020 climatology and (b) anomaly correlation between observed Niño-3.4 SST (Reynolds et al. 2002) and GPCP November precipitation data (1979–2021). Plotted from the IRI Data Library.

America, and negative ones over the Caribbean and the western Indian Ocean, east Africa and in particular over southeast South America. These MJO and ENSO teleconnections are constructive over many tropical land regions where the week-3–4 forecast is sharp (Maritime Continent, eastern Africa, and the Caribbean), while they oppose each other over South America where La Niña’s impact appears to dominate the forecast, emphasizing the importance of the impact of SST anomalies on subseasonal forecasts. This particular week-3–4 precipitation forecast example was one of the most skillful over the 2-yr period of real-time forecasts analyzed in terms of globally averaged RPSS (Fig. 8); the analysis in section 4c helps understand why, illustrating the potential for skillful subseasonal “windows of opportunity” when multiple sources of S2S predictability are active (Mariotti et al. 2020). The well-calibrated probabilistic subseasonal forecast system developed here is designed to automatically provide forecast probability distributions that deviate from their climatological expectations when such spatiotemporal windows of opportunity arise (Fig. 11).

However, further work is required to confirm the generality of this analysis of a single case regarding the roles of MJO

and ENSO, and to investigate interactions between additional sources of S2S predictability (Muñoz et al. 2015, 2016) which also include the quasi-biennial oscillation and sudden stratospheric warmings (Domeisen et al. 2020a,b), and land surface interactions (Dirmeier et al. 2019), among others. More work is also required to further develop subseasonal forecast processing techniques, and to include additional skillful models beyond the three chosen here. It is hoped that the availability of the IRI SubX-based forecast map rooms described in this paper will aid such research, and further the applications of subseasonal forecasting.

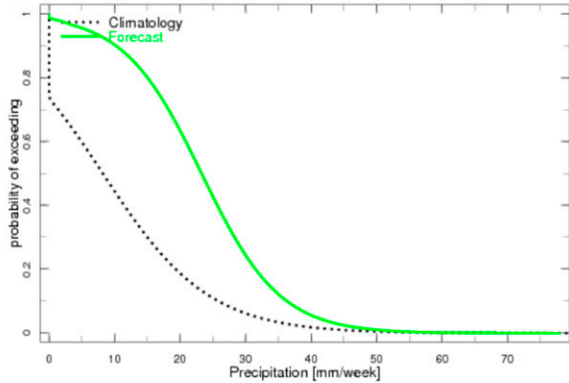
Acknowledgments. AWR wishes to thank Gilbert Brunet for originally suggesting that IRI develop subseasonal forecasts in real time, and to Vincent Moron for illuminating discussions. We are thankful for the detailed comments of two anonymous reviewers that considerably improved the manuscript. This work was supported by NOAA’s Office of Water and Air Quality, Awards NA18OAR4310295 and NA19OAR4590159, as well as by a Columbia University Climate and Life Fellowship to AWR. We acknowledge the

(a) Precipitation



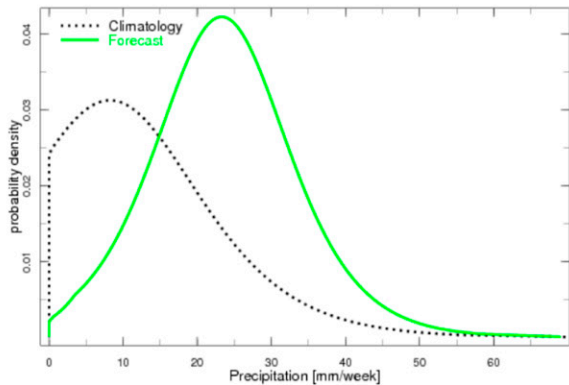
Forecast made for [126E-127E, 9.5S-8.5S]
located in or near **Baguia, Baucau, Timor-Leste**

Probability of Exceedance



13-26 Nov 2021 Flexible SubX Precipitation forecast issued 29 Oct 2021

Probability Distribution



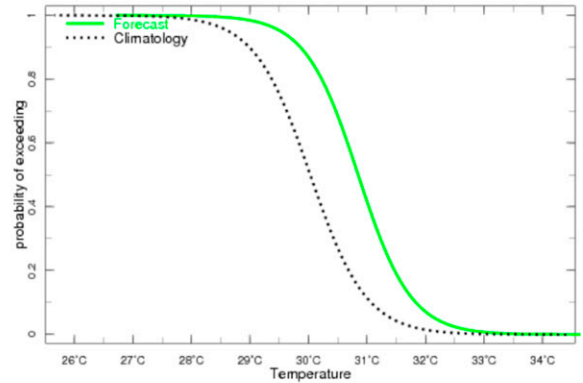
13-26 Nov 2021 Flexible SubX Precipitation forecast issued 29 Oct 2021

(b) Temperature



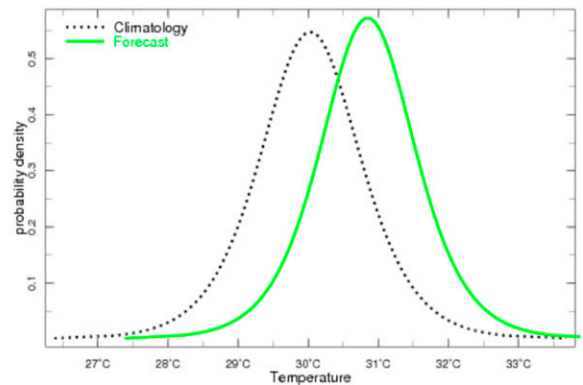
Forecast made for [126E-127E, 9.5S-8.5S]
located in or near **Baguia, Baucau, Timor-Leste**

Probability of Exceedance



13-26 Nov 2021 Flexible SubX Temperature forecast issued 29 Oct 2021

Probability Distribution



13-26 Nov 2021 Flexible SubX Temperature forecast issued 29 Oct 2021

FIG. 11. Forecast probability of exceedances and probability distributions for a point located over East Timor, Indonesia, for (a) precipitation and (b) 2-m temperature, for the 13–26 Nov 2021 period, issued 29 Oct 2021. The historical distributions (1999–2016) are indicated as dotted curves. Plotted from the IRI Data Library.

agencies that support the SubX system, and we thank the climate modeling groups (Environment Canada, NASA, NOAA/NCEP, NRL, and University of Miami) for producing and making available their model output. AGM was partially supported by NOAA Grants NA18OAR4310275, NA18OAR4310339, and the FORMAS Arbo-Prevent Project. NOAA/MAPP, ONR, NASA, and NOAA/NWS jointly provided coordinating support and led development of the SubX system. The SubX data were obtained from IRI Data Library <https://iridl.ldeo.columbia.edu/SOURCES/Models/SubX/>. CPC Global Unified Temperature data were provided by the NOAA Physical Sciences Laboratory (PSL), Boulder, Colorado, from their website at <https://psl.noaa.gov>.

Data availability statement. All data used in this study are available via the IRI Climate Data Library <https://iridl.ldeo.columbia.edu>.

APPENDIX

Flexible Forecast Methodology

The logistic regression Eq. (1) for the forecast probability p may be written as

$$p = \frac{e^f}{1 + e^f}, \tag{A1}$$

where

$$f = b_0 + b_1 \bar{x}_{\text{ens}} + b_2 q. \quad (\text{A2})$$

Once the regression coefficients b have been estimated from reforecasts, the forecast probability p can be obtained for any historical percentile P of the observed climatology. To obtain the corresponding climatological value of q , the regression model is fit to the reforecast data interval but without including the model's ensemble mean predictor, again using tercile categories:

$$P = \frac{e^F}{1 + e^F}, \quad (\text{A3})$$

where

$$F = B_0 + B_2 q. \quad (\text{A4})$$

Rearranging for q then yields

$$q = \frac{1}{B_2} \ln \left[\frac{P}{(1-P)\exp(B_0)} \right]. \quad (\text{A5})$$

Substituting q back into (A2), we then arrive at the forecast probability p as a function of any historical percentile P :

$$p = \frac{\exp \left[b_0 + b_1 \bar{x}_{\text{ens}} + \frac{b_2}{B_2} \ln(P) \right]}{\exp \left\{ \frac{b_2}{B_2} [\ln(1-P) + B_0] \right\} + \exp \left[b_0 + b_1 \bar{x}_{\text{ens}} + \frac{b_2}{B_2} \ln(P) \right]}. \quad (\text{A6})$$

To create the probability of exceedance curves in Fig. 11, the quantile value for each climatological percentile P (500 samples from zero to one) is calculated from (A5) yielding the climatological exceedances curve. The forecast probability curve is then derived from (A6) as a function of P and the regression parameters.

To plot the probability density function (pdf), we again begin with p , which always lies between 0 and 1. We introduce an intermediary independent variable (sv) that both the pdf and q depends on isomorphically:

$$\text{sv} = \ln \left(\frac{p}{1-p} \right). \quad (\text{A7})$$

Substituting sv into (A5) yields

$$q = \frac{1}{B_2} (\text{sv} - B), \quad (\text{A8})$$

where

$$B = B_0 \text{ for climatology, and} \\ = b_0 + b_1 \bar{x}_{\text{ens}} \text{ for the forecast.}$$

The pdf is then given by

$$\frac{dp}{dq} = B_2 \frac{dp}{d\text{sv}}. \quad (\text{A9})$$

REFERENCES

- Barnston, A. G., and M. K. Tippett, 2014: Climate information, outlooks, and understanding—Where does the IRI stand? *Earth Perspect.*, **1**, 20, <https://doi.org/10.1186/2194-6434-1-20>.
- , S. Li, S. J. Mason, D. G. DeWitt, L. Goddard, and X. Gong, 2010: Verification of the first 11 years of IRI's seasonal climate forecasts. *J. Appl. Meteor. Climatol.*, **49**, 493–520, <https://doi.org/10.1175/2009JAMC2325.1>.
- DelSole, T. M., and M. K. Tippett, 2022: *Statistical Methods for Climate Scientists*. Cambridge University Press, 542 pp.
- Dirmeyer, P. A., P. Gentine, M. B. Ek, and G. Balsamo, 2019: Land surface processes relevant to sub-seasonal to seasonal (S2S) prediction. *Sub-Seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting*, A. W. Robertson and F. Vitart, Eds., Elsevier, 165–181, <https://doi.org/10.1016/B978-0-12-811714-9.00008-5>.
- Domeisen, D. I. V., and Coauthors, 2020a: The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *J. Geophys. Res. Atmos.*, **125**, e2019JD030920, <https://doi.org/10.1029/2019JD030920>.
- , and Coauthors, 2020b: The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *J. Geophys. Res. Atmos.*, **125**, e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus*, **57A**, 219–233, <https://doi.org/10.3402/tellusa.v57i3.14657>.
- Hamill, T. M., 2012: Verification of TIGGE multi-model and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States. *Mon. Wea. Rev.*, **140**, 2232–2252, <https://doi.org/10.1175/MWR-D-11-00220.1>.
- Huffman, G., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multi-satellite observations. *J. Hydrometeorol.*, **2**, 36–50, [https://doi.org/10.1175/1525-7541\(2001\)002<0036:GPAODD>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2).
- Li, S., and A. W. Robertson, 2015: Evaluation of sub-monthly precipitation forecast skill from global ensemble prediction systems. *Mon. Wea. Rev.*, **143**, 2871–2889, <https://doi.org/10.1175/MWR-D-14-00277.1>.
- Mariotti, A., and Coauthors, 2020: Forecasts of opportunity: Opening windows of skill, subseasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101**, 597–601, <https://doi.org/10.1175/BAMS-D-18-0326.A>.
- Muñoz, Á. G., L. Goddard, A. W. Robertson, Y. Kushnir, and W. Baethgen, 2015: Cross-time scale interactions and rainfall extreme events in southeastern South America for the austral summer. Part I: Potential predictors. *J. Climate*, **28**, 7894–7913, <https://doi.org/10.1175/JCLI-D-14-00693.1>.
- , —, S. J. Mason, and A. W. Robertson, 2016: Cross-time scale interactions and rainfall extreme events in southeastern South America for the austral summer. Part II: Predictive skill. *J. Climate*, **29**, 5915–5934, <https://doi.org/10.1175/JCLI-D-15-0699.1>.
- 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>.
- 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:AIISAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2).
- Robertson, A. W., and F. Vitart, Eds., 2019: *Sub-Seasonal to Seasonal Prediction: The Gap between Weather and Climate Forecasting*. 1st ed. Elsevier, 588 pp.
- , U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Wea. Rev.*, **132**, 2732–2744, <https://doi.org/10.1175/MWR2818.1>.
- , V. Moron, J.-H. Qian, C.-P. Chang, F. Tangang, E. Aldrian, T. Y. Koh, and L. Juneng, 2011: The maritime continent monsoon. *The Global Monsoon System: Research and Forecast*, 2nd ed. C.-P. Chang et al., Eds., World Scientific Publishing Co., 85–98.
- , —, N. Vigaud, N. Acharya, A. M. Greene, and D. S. Pai, 2019: Multi-scale variability and predictability of Indian summer monsoon rainfall. *MAUSAM*, **70**, 277–292, <https://doi.org/10.54302/mausam.v70i2.172>.
- 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>.
- Sun, S., R. Bleck, S. G. Benjamin, B. W. Green, and G. A. Grell, 2018a: Subseasonal forecasting with an icosahedral, vertically quasi-Lagrangian coupled model. Part I: Model overview and evaluation of systematic errors. *Mon. Wea. Rev.*, **146**, 1601–1617, <https://doi.org/10.1175/MWR-D-18-0006.1>.
- , B. W. Green, R. Bleck, and S. G. Benjamin, 2018b: Subseasonal forecasting with an icosahedral, vertically quasi-Lagrangian coupled model. Part II: Probabilistic and deterministic forecast skill. *Mon. Wea. Rev.*, **146**, 1619–1639, <https://doi.org/10.1175/MWR-D-18-0007.1>.
- Tippett, M. K., A. G. Barnston, and A. W. Robertson, 2007: Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *J. Climate*, **20**, 2210–2228, <https://doi.org/10.1175/JCLI4108.1>.
- , —, and T. DelSole, 2010: Comments on “Finite samples and uncertainty estimates for skill measures for seasonal prediction.” *Mon. Wea. Rev.*, **138**, 1487–1493, <https://doi.org/10.1175/2009MWR3214.1>.
- Vigaud, N., A. W. Robertson, M. K. Tippett, and N. Acharya, 2017a: Subseasonal predictability of boreal summer monsoon rainfall from ensemble forecasts. *Front. Environ. Sci.*, **5**, 67, <https://doi.org/10.3389/fenvs.2017.00067>.
- , —, and —, 2017b: Multi-model ensembling of subseasonal precipitation forecasts over North America. *Mon. Wea. Rev.*, **145**, 3913–3928, <https://doi.org/10.1175/MWR-D-17-0092.1>.
- , M. K. Tippett, and A. W. Robertson, 2019a: Deterministic skill of subseasonal precipitation forecasts for the East Africa-West Asia sector from September to May. *J. Geophys. Res. Atmos.*, **124**, 11 887–11 896, <https://doi.org/10.1029/2019JD030747>.
- , —, J. Yuan, A. W. Robertson, and N. Acharya, 2019b: Probabilistic skill of subseasonal surface temperature forecasts over North America. *Wea. Forecasting*, **34**, 1789–1806, <https://doi.org/10.1175/WAF-D-19-0117.1>.
- Vitart, F., A. W. Robertson, and D. L. T. 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, 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>.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124, <https://doi.org/10.1175/MWR3280.1>.
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an 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).
- Wilks, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteor. Appl.*, **16**, 361–368, <https://doi.org/10.1002/met.134>.
- 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: Towards the improvement of subseasonal prediction in the National Centers for Environmental Prediction Global Ensemble Forecast System (GEFS). *J. Geophys. Res. Atmos.*, **123**, 6732–6745, <https://doi.org/10.1029/2018JD028506>.