

Multimodel Ensembling of Subseasonal Precipitation Forecasts over North America

N. VIGAUD AND A. W. ROBERTSON

International Research Institute for Climate and Society, Earth Institute, Columbia University, Palisades, New York

M. K. TIPPETT

Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, and Department of Meteorology, Center of Excellence for Climate Change Research, King Abdulaziz University, Jiddah, Saudi Arabia

(Manuscript received 5 April 2017, in final form 13 June 2017)

ABSTRACT

Probabilistic forecasts of weekly and week 3–4 averages of precipitation are constructed using extended logistic regression (ELR) applied to three models (ECMWF, NCEP, and CMA) from the Subseasonal-to-Seasonal (S2S) project. Individual and multimodel ensemble (MME) forecasts are verified over the common period 1999–2010. The regression parameters are fitted separately at each grid point and lead time for the three ensemble prediction system (EPS) reforecasts with starts during January–March and July–September. The ELR produces tercile category probabilities for each model that are then averaged with equal weighting. The resulting MME forecasts are characterized by good reliability but low sharpness. A clear benefit of multimodel ensembling is to largely remove negative skill scores present in individual forecasts. The forecast skill of weekly averages is higher in winter than summer and decreases with lead time, with steep decreases after one and two weeks. Week 3–4 forecasts have more skill along the U.S. East Coast and the southwestern United States in winter, as well as over west/central U.S. regions and the intra-American sea/east Pacific during summer. Skill is also enhanced when the regression parameters are fit using spatially smoothed observations and forecasts. The skill of week 3–4 precipitation outlooks has a modest, but statistically significant, relation with ENSO and the MJO, particularly in winter over the southwestern United States.

1. Introduction

Predictions on subseasonal time scales, between medium-range weather (up to 2 weeks) and seasonal climate (from 3 to 6 months) forecasts, have recently received increasing interest owing to modeling advances (Vitart 2014) and a better understanding of climate phenomena on these time scales, particularly the MJO (Zhang 2013). Sources of predictability at subseasonal time scales include the inertia of sea surface temperature (SST) anomalies, the MJO (Waliser et al. 2003; Waliser 2011; Neena et al. 2014), but also stratospheric processes including the QBO (Baldwin and Dunkerton 2001; Scaife and Knight 2008; Yoo and Son 2016), memory in soil moisture (Koster et al. 2010), snow cover (Lin and Wu 2011), and sea ice (Holland et al. 2011).

Based on experience from probabilistic seasonal climate and medium-range forecasting, calibration of model probabilities is expected to be necessary to

account for model deficiencies and produce reliable forecasts (Goddard et al. 2001; Wilks 2002; Tippett et al. 2007). In comparison to seasonal hindcasts (reforecasts), submonthly hindcasts are often characterized by shorter length with fewer ensemble members, which represent additional challenges. While the value of the model output statistics (MOS) approach to improve weather probabilistic forecasts has been demonstrated (Hamill et al. 2004), no analysis has been yet done at subseasonal time scales. There is also a need to investigate to what extent skill can be enhanced by multimodel ensemble (MME) techniques, as has been demonstrated for seasonal (Robertson et al. 2004) and medium-range (Hamill 2012) forecasting.

Extended logistic regression (ELR), as used here, includes the quantile threshold along with the ensemble mean as predictor and produces mutually consistent quantile probabilities that sum to one (Wilks 2009; Wilks and Hamill 2007). In this respect, this study presents a first attempt to produce weekly and week 3–4 MME precipitation tercile probability forecasts. ELR is

Corresponding author: N. Vigaud, nicolas.vigaud@gmail.com

DOI: 10.1175/MWR-D-17-0092.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](http://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

applied at each grid point to the individual model forecasts, which are subsequently averaged together with equal weighting. The data and methods are presented in [section 2](#) together with diagnostics related to the regression model setup for weekly varying precipitation tercile categories. The skill of forecasts initialized during January–March (JFM, winter) and July–September (JAS, summer) seasons are examined over a North American continental domain in [section 3](#), first at weekly resolution. Improvements to skill through considering a week 3–4 outlook (instead of weeks 3 and 4 separately) and spatial smoothing of the regression model input are then discussed alongside their skill relationships to ENSO and the MJO. A summary and conclusions are presented in [section 4](#).

2. Data and methods

a. Observation and model datasets

Daily precipitation fields from the European Centre for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction (NCEP), and the China Meteorological Administration (CMA) forecasts for week 1, week 2, week 3, and week 4 leads of the reforecasts (i.e., the periods from $[d + 1, d + 7]$ to $[d + 22, d + 28]$ for a forecast issued on day d) were obtained from the Subseasonal-to-Seasonal (S2S) database ([Vitart et al. 2017](#)) through the IRI Data Library (IRIDL) portal. These ensemble prediction systems (EPSs) have different native resolutions (from 125 km at the equator with 40 vertical levels for CMA to 16/32 km and 91 vertical levels for ECMWF) and are archived on a common 1.5° grid in the S2S database. The number of ensemble members (51 for ECMWF, 4 for CMA and NCEP) and reforecasts length (between 44- and 60-day lead from the NCEP CFSv2 to CMA) depend on the modeling center as indicated in [Table 1](#); see [Vitart et al. \(2017\)](#) for further details. In particular, ECMWF is the only model for which reforecasts are generated on the fly twice a week (11 members every Monday and Thursday), while those from NCEP and CMA are generated four times daily from fixed model versions. We consider here weekly accumulated precipitation from ECMWF reforecasts that are generated for Monday starts from June 2015 to March 2016 and NCEP and CMA four-member daily reforecasts sampled from their respective 1999–2010 and 1994–2014 periods of issuance. The common period when all three EPS reforecasts are available is 1999–2010, and that period is used in our analysis. Subsequently, S2S data were spatially interpolated onto the GPCP 1° horizontal grid before the forecasted probabilities obtained from the three

TABLE 1. ECMWF, NCEP, and CMA forecasts attributes as archived in the S2S database at ECMWF.

Attributes	ECMWF	NCEP	CMA
Time range	Day 0–46	Day 0–44	Day 0–60
Resolution	Tco639/319 L91	T126L64	T106L40
Ensemble size	51	16	4
Frequency	Two per week	Daily	Daily
Reforecasts (RFC)	On the fly	Fix	Fix
RFC length	Past 20 yr	1999–2010	1994–2014
RFC frequency	Two per week	Daily	Daily
RFC size	11	4	4

individual models are averaged to form MME tercile precipitation forecasts from which the skill of starts during the winter and summer seasons is assessed over the continental North America (i.e., land points between 20° and 50°N).

The Global Precipitation Climatology Project (GPCP) version 1.2 ([Huffman et al. 2001](#); [Huffman and Bolvin 2012](#)) daily precipitation estimates on a 1° grid, available from October 1996 to October 2015, are used as observational data for the calibration and verification of the reforecasts over the 1999–2010 period of analysis.

b. Extended logistic regression model

Distributional regressions are 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 can be extended to produce the cumulative probability p of not exceeding the quantile q such as

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

by including an additional explanatory variable $g(q)$, which is a function of the quantile q as follows:

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

where f and g are linear functions of the EPS ensemble mean \bar{x}_{ens} and quantile q , respectively. A cube-root transformation of the precipitation amounts used in the regression model did not improve skill (cf. [Hamill 2012](#)). Because of the limited number of ensemble members available from the S2S reforecasts (4 members daily for CFS and CMA and 11 members twice weekly for ECMWF), the more familiar approach of estimating the forecast probabilities by counting how many members exceed a certain threshold leads to large errors. In the seasonal forecasting context, [Tippett et al. \(2007\)](#)

have shown that regression models outperform counting estimates, especially for small ensemble size. The definition in Eq. (1) leads to mutually consistent individual threshold probabilities (Wilks and Hamill 2007; Wilks 2009) as shown in section 2c. Ultimately, these allow flexible choice of threshold probabilities according to user's needs (Barnston and Tippett 2014). Extended logistic regression (referred to simply as regression in the following) is here used to produce precipitation tercile ($q = 1/3$ and $q = 2/3$) category probabilities referred to as "forecasts" in the following.

The observed climatological tercile categories corresponding to the 33rd and 67th percentiles of the GPCP weekly accumulated precipitation estimates are defined separately at each grid point for each start within the JFM (4 January–28 March, Monday start dates) and JAS (6 July–28 September, Monday start dates) seasons (i.e., 12 starts per season) and each lead (week 1–4) following a leave-one-year-out approach. Next, 1) the regression parameters are estimated separately for each model, grid point, calendar start date, and lead using all years except the one being forecast (leave-one-out cross validation), 2) the tercile probabilities of the left out year are forecast, and 3) MME probabilities are constructed by simple averaging of the three individual forecast probabilities.

c. Regression model setup

For forecasts of weekly averages, the climatological observed tercile categories are computed using 11-yr data periods following the leave-one-year-out methodology discussed above and 3-week windows formed by the forecast target week and a week either side. Wider windows were found to degrade the skill of the cross-validated forecasts contrasting with the findings of Wilks (2009). A "dry mask" is used, and forecasts are only produced when and where the 33rd percentile is non-zero. The MME probabilities are computed by simple averaging regardless of ensemble size.

In addition to weekly precipitation averages, we also considered forecasts for the week 3–4 target period (from $d + 15$ to $d + 28$ for forecasts issued on day d). This corresponds to a 2-week target at 2-week lead (Zhu et al. 2014). The tercile categories were derived using 6-week windows, including the 2-week target formed by week 3 and 4 leads and two weeks on either side, for which diagnostics are provided, while wider windows did not improve forecast skill.

Figure 1 shows an example of the ELR-based probabilities computed from ECMWF hindcasts for 3–9 August 1999, fitted using a 3-week window centered on the 3 August week and starts within the 2000–10 period at the grid point (13.5°N, 91.5°W) just off the Guatemala

Pacific coast, where there is some skill in summer. Regressions are based on observed terciles of GPCP observations over this 3-week window (27 July–17 August). Forecasted probabilities of nonexceedance of the 0.33 and 0.67 quantiles obtained from Eq. (1) for different values of the ensemble mean weekly accumulated precipitation forecasts (x axis) are characterized by parallel lines for different leads (week 1–4) in agreement with the regression formulated in Eq. (1) yielding logically consistent sets of forecasts, in the sense that cumulative probabilities for smaller predictand thresholds do not exceed those for larger thresholds (Wilks 2009).

GPCP observations along with forecasts from the three individual models and their average are shown in the Fig. 2 for week 1 forecasts made during JAS 1999 at this grid point. Once weekly terciles are defined (under cross validation), the regression model is then trained on the same pool of weeks (i.e., 11 years of 3-week windows centered on the target week) by fitting regression equations at each point, lead, and start separately for each model. The regression coefficients thus obtained are then used to produce weekly precipitation forecasts, as shown for week 1 from ECMWF, NCEP, and CMA weekly starts during the JAS 1999 season in Figs. 2b, 2e, and 2c, respectively. Overall, the individual category forecasts display highest weekly probabilities that are consistent to varying degrees with observed tercile categories (Fig. 2d). Forecasts from ECMWF are more skillful compared to those from NCEP and CMA. Finally, the forecasts from the three models are averaged with equal weighting to produce MME forecasts over the 12-week period (Fig. 2f).

d. Skill metrics

The skill of tercile precipitation forecasts obtained from the above regression model is evaluated using two statistical metrics. First, reliability diagrams are plotted to evaluate reliability, resolution, and sharpness (Wilks 1995; Hamill 1997) and are computed by pooling all land points over North America between 20° and 50°N. In addition, ranked probability skill scores (RPSS) (Weigel et al. 2007) complement the above diagnostics and quantify to which extent the predictions are improved upon climatological probabilities.

e. Significance testing

In section 3, RPSS computed only for starts during specific phases of the MJO (1 + 8 and 4 + 5) are tested for statistical significance at each grid point using Monte Carlo simulations based on random draws from the entire pool of forecasts in JAS and JFM separately. Statistical significance at the 5% level is assessed by

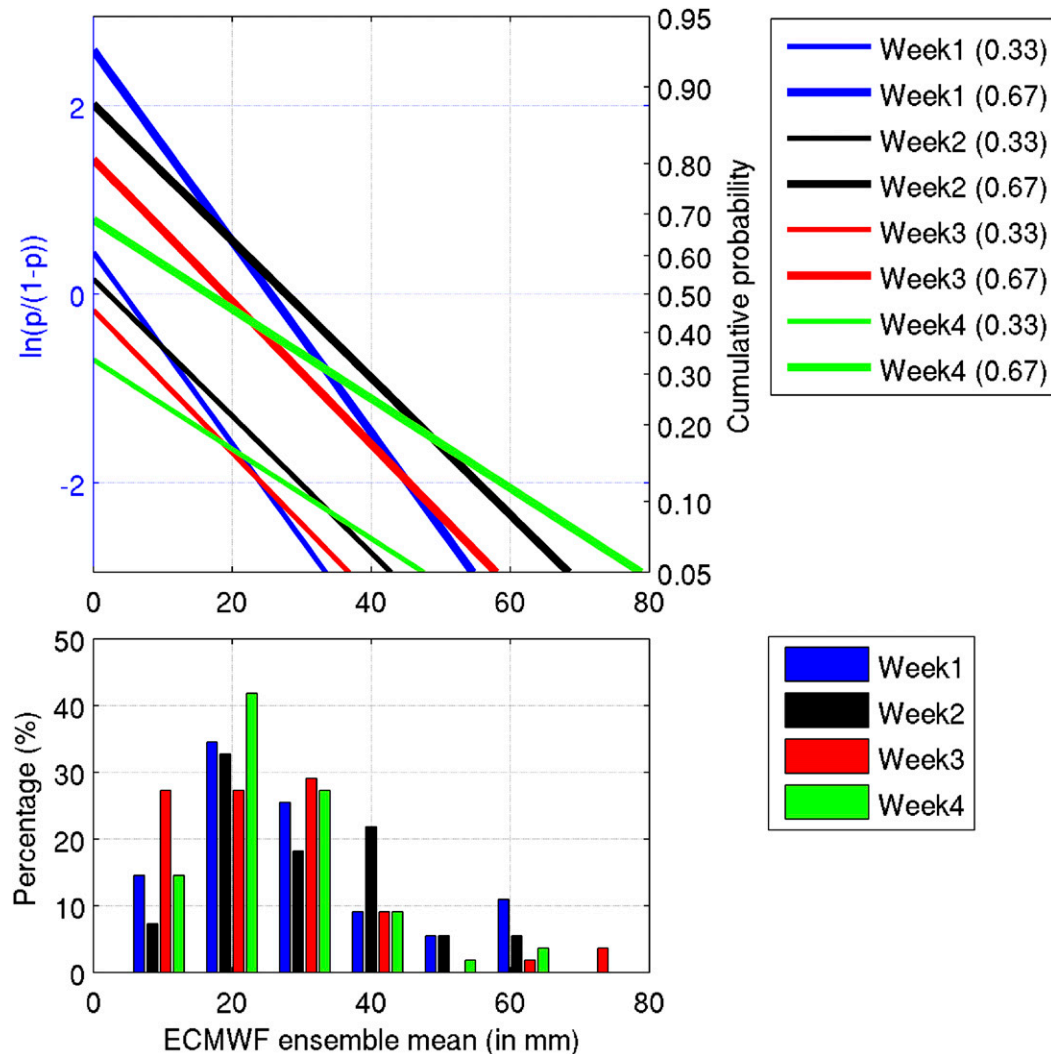


FIG. 1. (top) Extended logistic regressions plotted for ECMWF hindcasts issued 3 Aug 1999 at (13.5°N, 91.5°W) and fitted using 3-week windows over 11 yr for tercile definition and training. Forecasted probabilities of non-exceedance of the 0.33 (thin lines) and 0.67 (thick lines) quantiles computed from Eq. (1) for different values of the ensemble mean weekly accumulated precipitation forecasts (x axis, in mm) are shown by parallel lines at different leads (week 1–4) yielding to logically consistent sets of forecasts. (bottom) The distribution of ECMWF ensemble mean weekly rainfall over the 1999–2010 period at this grid point is plotted as bins centered on integer multiple of 10 for the respective leads.

comparing the 95th percentile RPSS values to that of those from specified MJO phases. Correlations are also tested using Monte Carlo simulations to assess the significance of the relationships between week 3–4 MME RPSS averaged over continental North America but also synthesized using a principal components analysis (PCA), and both observed Niño-3.4 index and MJO measured by Real-time Multivariate MJO 1 (RMM1) and RMM2 indices from Wheeler and Hendon (2004), as well as their best linear combination. Additional correlations are computed with the square of each index to examine skill associations with index amplitude.

3. Results

a. Weekly averages

Reliability diagrams for weekly ECMWF forecasts from all starts during the JFM seasons are displayed in Figs. 3a–c, with those from JAS starts in Figs. 3d–f. These exhibit reasonable skill for week 1 for both seasons in terms of reliability and resolution, as shown by the blue curves close to the diagonal and distant from the climatological 0.33 horizontal line (i.e., zero resolution line, not plotted) respectively. Corresponding histograms for week 1 ECMWF forecasts are spread

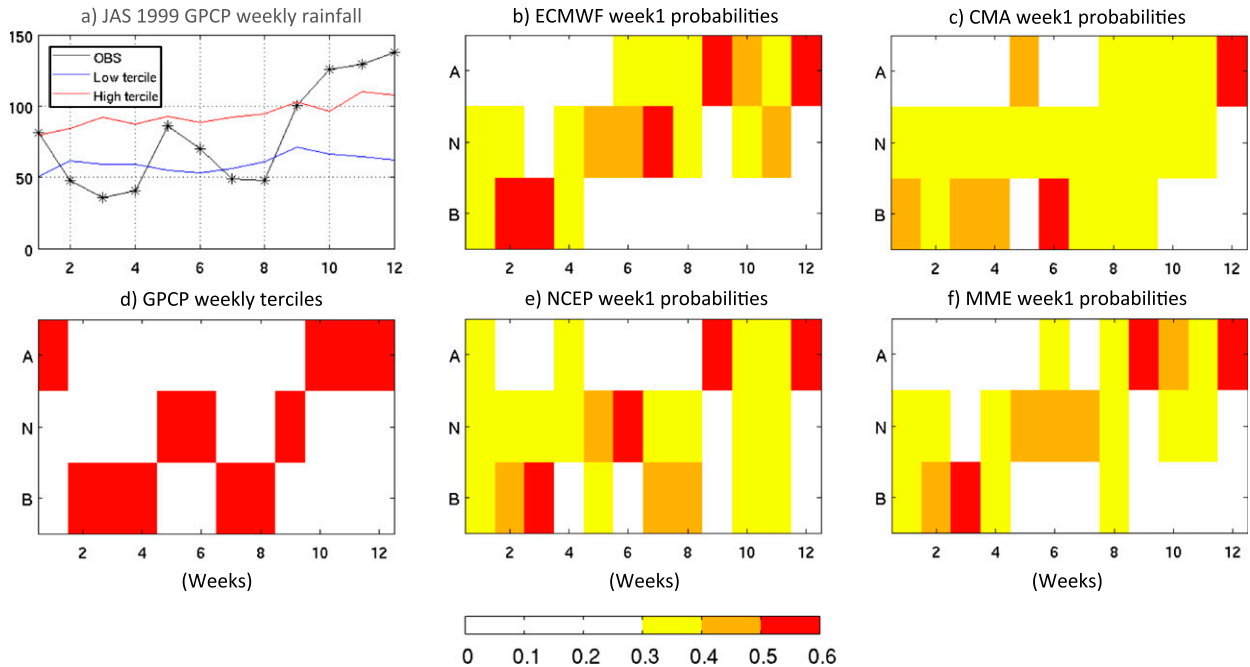


FIG. 2. Point statistics at (13.5°N, 91.5°W) showing (a) the mean GPCP cumulated precipitation for each week of the JAS 1999 season (x axis; i.e., from 6 Jul to 28 Sep), together with the low/high terciles (blue/red) computed from 3-week windows centered on the target week, and (d) associated GPCP weekly tercile probabilities; that is, above-normal (“A”), normal (“N”), and below-normal (“B”) categories. After training out-of-sample (11 yr) forecasted weekly tercile probabilities are issued for (b) ECMWF, (e) NCEP, and (c) CMA hindcasts, which are pooled together with equal weighting to produce (f) a multimodel ensemble (MME) weekly tercile precipitation forecasts.

across all bins, except for the normal category, characterizing high sharpness. As lead time increases, forecast issuance frequencies are centered around climatology (0.33; i.e., fourth bin). For the below- and above-normal terciles, distributions are also skewed toward equal odds with increasing leads, consistent with decreasing slopes from week 1 and week 2 onward when reliability and resolution sharply drop. Week 2 forecasts are characterized by higher skill in winter than summer, while little skill is visible in either season in week 3 and 4. NCEP and CMA forecasts exhibit qualitatively similar results (not shown) but are less skillful than ECMWF.

MME forecasts are characterized by slightly greater slopes for week 2 leads in particular, indicating increased reliability and resolution compared to individual models. Similar to individual model forecasts, sharpness remains low, while skill also decreases with increasing leads and from winter to summer (Fig. 4). The week 3 and week 4 MME forecasts show only small deviations from equal odds, and those lack reliability as displayed by even lower slopes than for ECMWF at similar leads (Fig. 3).

Maps of RPSS for individual models and their MME are shown in Figs. 5 and 6 for JFM and JAS starts, respectively. All three models and the MME show similar, positive RPSS values for week 1 forecasts starting during

JFM (Fig. 5), with maximum scores over land located toward the northwestern and eastern United States. Areas north of 30°N and south of 25°N exhibit high skill over the Pacific, while most skillful predictions are between 20° and 35°N and north of 40°N over the Atlantic. In week 2, these regions are still characterized by larger RPSS but with much lower magnitude. RPSS values for week 3 and week 4 forecasts are near zero or negative everywhere. Multimodel combination results in a marked increase in the RPSS of week 1 and week 2 forecasts almost everywhere compared to the most skillful individual models (ECMWF). The greatest benefit of multimodel ensembling is that it largely removes negative skill values. There are broad regions of positive skill in week 2 over the southwestern and eastern United States; over oceans there is skill north of 30°N and south of 25°N in the Pacific and north of 25°N in the Atlantic. From week 3, positive skill only remains over northeast regions of the Gulf of Mexico and along the U.S. East Coast where marginal skill remains in week 4. For starts within the JAS season, Fig. 6 shows skill during week 1 over northern and southern regions of the continental domain while it is maximum south of 24°N and north of 40°N in the Pacific, as well as over the intra-American seas and along the U.S. East Coast in the Atlantic. However, not much skill is found at longer

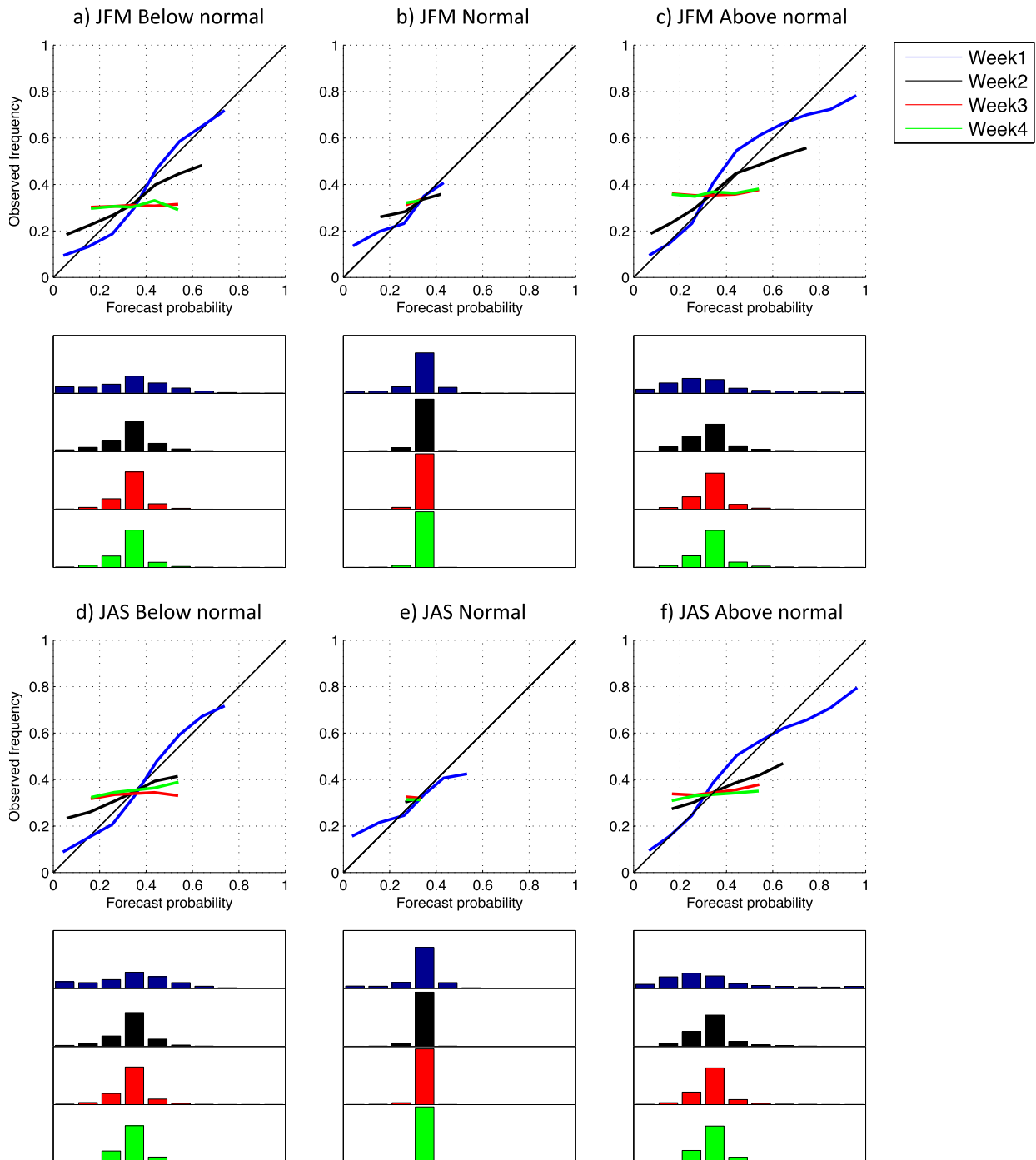


FIG. 3. Reliability diagrams for the below-normal, normal, and above-normal categories from ECMWF forecasts with starts in (a)–(c) JFM and (d)–(f) JAS with color coding based on week leads. The frequencies with which each category is forecasted are indicated as bins centered on an integer multiple of 0.10 in histograms plotted under the respective tercile category diagram. The bins are projected along the same x axis (forecast probabilities from 0 to 1) and scaled from 0% to 100%. Note that only bins with more than 1% of the total number of forecasts in each category are plotted in the relative diagrams for each lead. Diagrams are computed for all points over continental North America between 20° and 50°N latitudes.

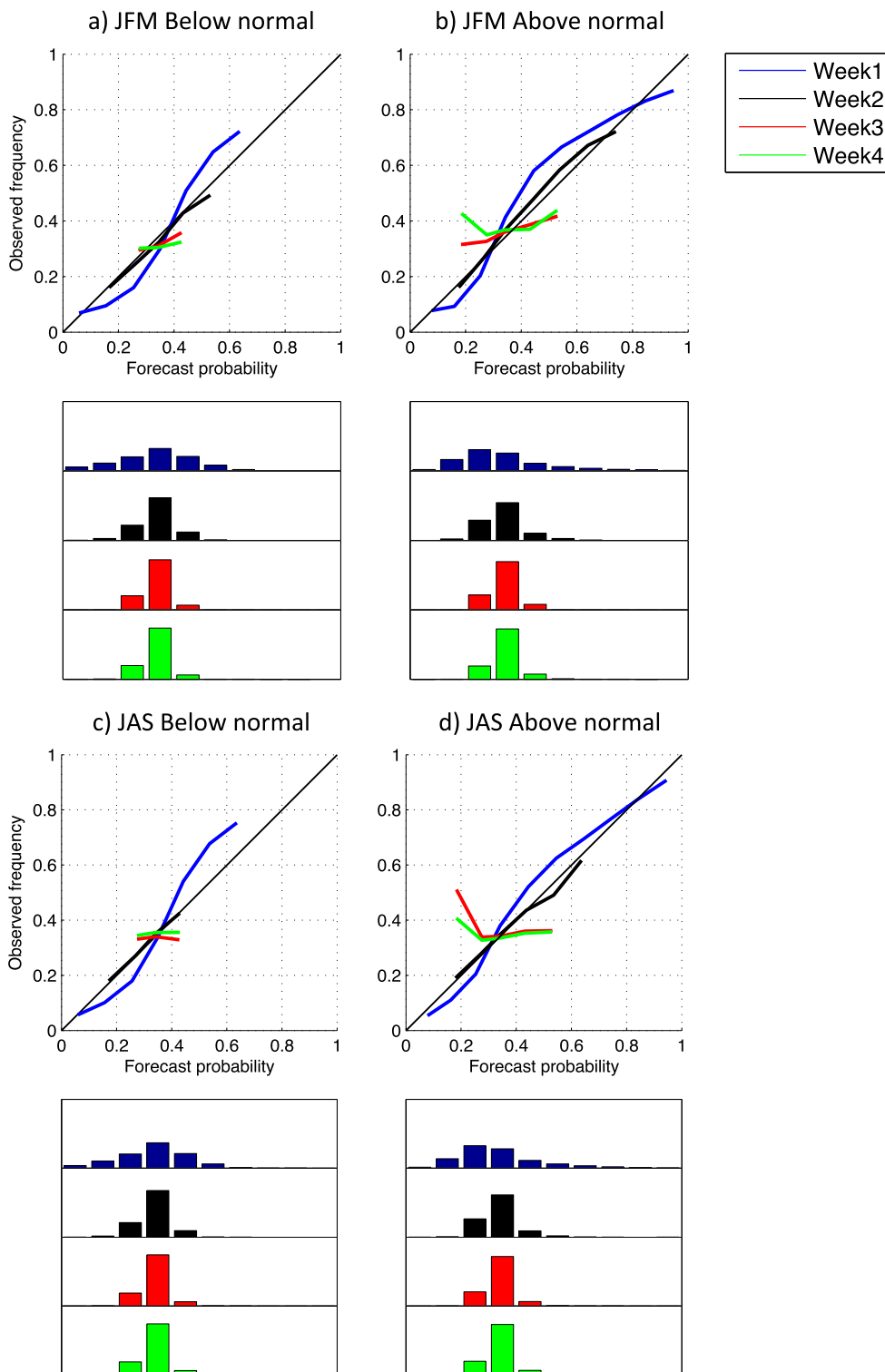


FIG. 4. Reliability diagrams for the below- and above-normal categories from the MME of ECMWF, NCEP, and CMA forecasts with starts in (a),(b) JFM and (c),(d) JAS with color coding based on week leads. The frequencies with which each category is forecasted are indicated as bins centered on an integer multiple of 0.10 in histograms plotted under the respective tercile category diagram. The bins are projected along the same x axis (forecast probabilities from 0 to 1) and scaled from 0% to 100%. Note that only bins with more than 1% of the total number of forecasts in each category are plotted in the relative diagrams for each lead. Diagrams are computed for all points over continental North America between 20° and 50°N latitudes.

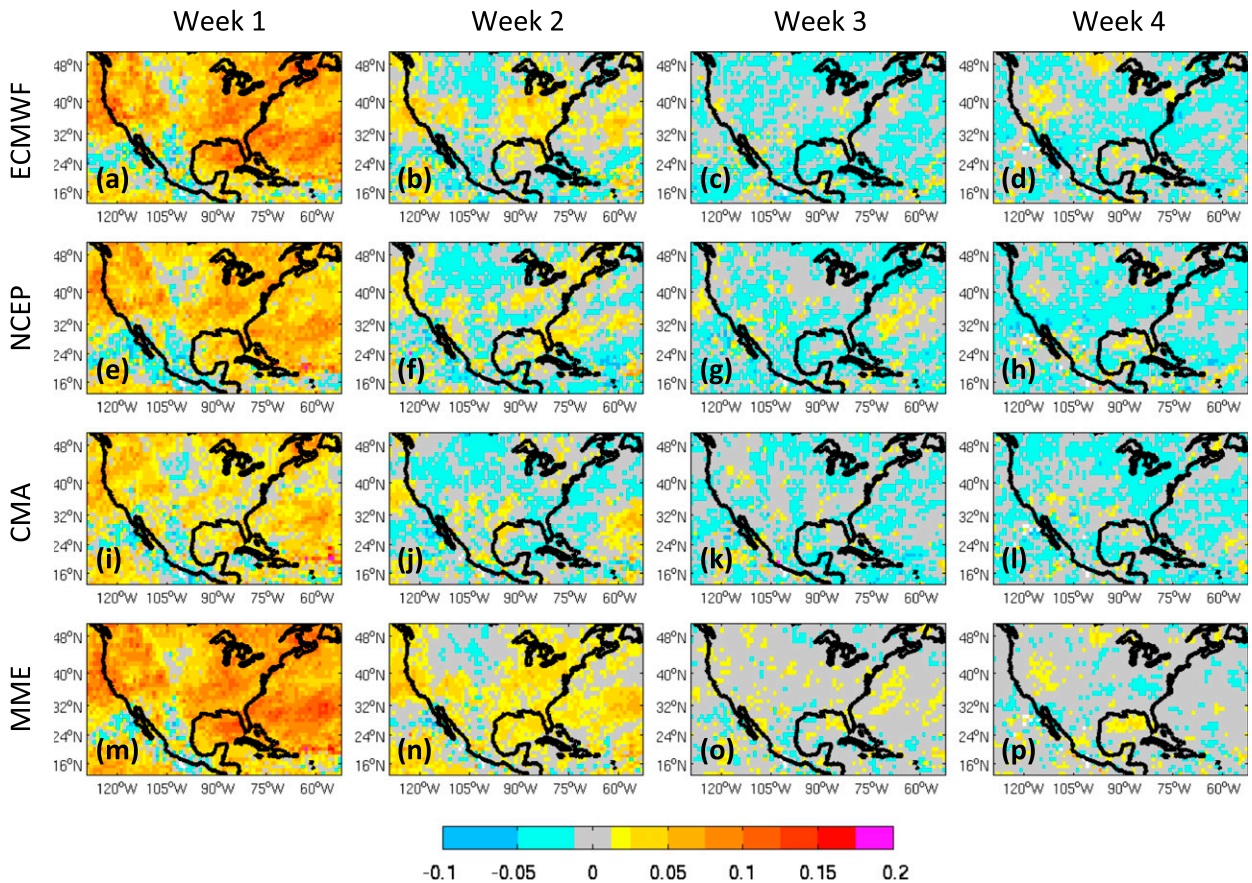


FIG. 5. RPSS for (a)–(d) ECMWF, (e)–(h) NCEP, and (i)–(l) CMA tercile precipitation forecasts as well as (m)–(p) their MME for starts during the JFM season. The different columns correspond to different leads from 1 to 4 weeks.

leads except in week 2 over the tropical Pacific and Caribbean basin. Consequently, the skill of the MME is low after week 1. Overall, the skill of the individual forecasts and their MME is lower in summer than winter, agreeing with the reliability diagrams in Figs. 3 and 4.

b. Week 3–4 outlooks

Figure 7 shows reliability diagrams for the below- (Fig. 7a) and above-normal (Fig. 7b) categories for 2-week, week 3–4, outlooks, from the individual models and their resulting MME with starts in JFM. Sharpness remains low but week 3–4 outlooks are characterized by greater slopes compared to weekly forecasts during this season (Figs. 3 and 4, top panels), indicating better reliability and resolution. The gain in reliability and resolution from multimodel ensembling is substantially increased (greater slopes) for week 3–4 compared to weekly averages. JAS starts exhibit similar results but with lower skill (not shown) agreeing with weekly forecasts.

In an attempt to further improve the skill, the input gridpoint observations and forecasts were spatially

smoothed before fitting the regression model with a bisquare weight function (Garcia 2010) using three points around each location in both latitude and longitude. However, spatial smoothing of both observation and forecasts leads to marginal improvements in overall sharpness, reliability, and resolution (not shown).

Maps of RPSS for raw and smoothed week 3–4 MME outlooks are shown in Fig. 8 for starts during the JFM (Figs. 8a,c) and JAS (Figs. 8b,d) seasons. Raw week 3–4 MME outlooks display more skill compared to weekly forecasts (Figs. 5m–p and 6m–p) along the U.S. East Coast and southwestern United States in JFM, and the west/central United States and the intra-American sea (IAS)/eastern Pacific in JAS. These translate into broader areas of skill for the smoothed week 3–4 outlooks. Note that both the raw and smoothed week 3–4 forecasts are verified against the same raw observations (i.e., unsmoothed) for a fair comparison.

c. Modulation of skill by ENSO and the MJO

Table 2 presents the temporal correlations between the RPSS of week 3–4 MME outlooks and the observed

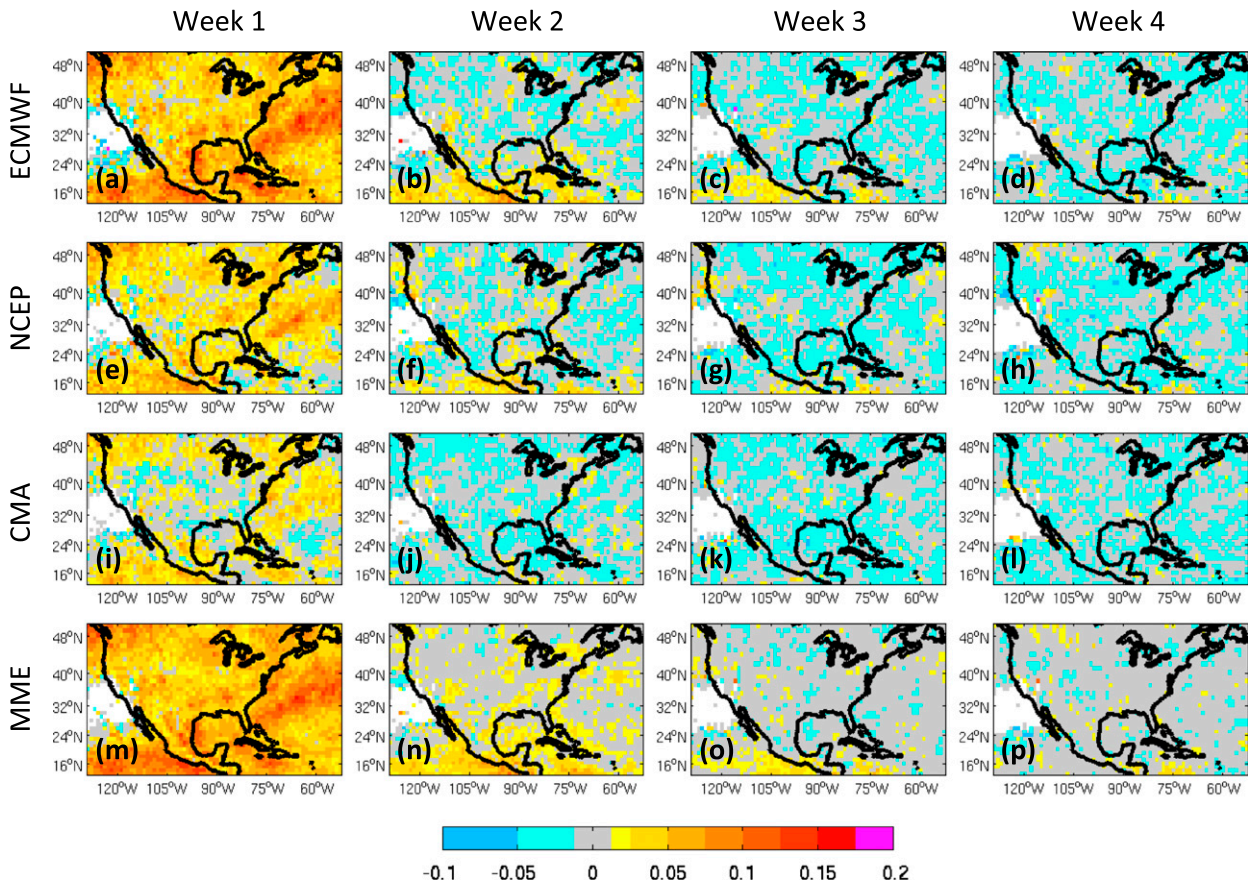


FIG. 6. As in Fig. 5, but for starts during the JAS season.

Niño-3.4 index as well as the MJO indices, RMM1 and RMM2, of Wheeler and Hendon (2004) and their best linear combination. Mean RPSS averaged over the continental North America between 20° and 50°N is significantly correlated with ENSO and MJO RMM2 for both raw and smoothed forecasts with starts during JFM. The best linear combination of MJO RMM1 and RMM2 exhibits relationships of the same magnitude as RMM2. In contrast, a significant relationship is found between RPSS and Niño-3.4 for starts during JAS but none with the MJO. Relationships to the Niño-3.4 modulus are not significant (second row for each seasons in Table 2), suggesting that ENSO polarity is important in regards to skill. The index and RPSS time series are plotted for the JFM season in Fig. 9. Interestingly, periods of higher week 3–4 RPSS appear to coincide with positive anomalies in Niño-3.4 and/or MJO RMM2. This further indicates skill increase during positive phases of ENSO in agreement with the prevalence of coherent patterns such as the tropical Northern Hemisphere (TNH) (Barnston and Livezey 1987), concomitant with northwest–southeast tilted negative height anomalies

over the North Pacific (Robertson and Ghil 1999) and more southerly and zonal storm tracks (Monteverdi and Null 1998) in winter during El Niño events, thus translating into precipitation anomalies over the western United States. Additional skill relationships to MJO RMM2 are consistent with MJO-induced modulations of the atmospheric river or “pineapple express,” leading to winter precipitation anomalies in the southwestern United States (Zhang 2013).

To investigate the skill of week 3–4 outlooks further, a PCA is applied to week 3–4 MME RPSS (total values; the mean is not removed) over continental North America between 20° and 50°N, at weekly resolution and for the JFM and JAS seasons separately. This approach allows us to examine if the regional structure of skill can be decomposed in geographically coherent patterns of variability. As shown in Figs. 10a and 10b, the spatial correlations typical of the first principal components (PCs) are mainly related to skill in the southwestern United States, where scores with highest magnitude contrast with opposite and less significant relationships in the northwestern United States in both

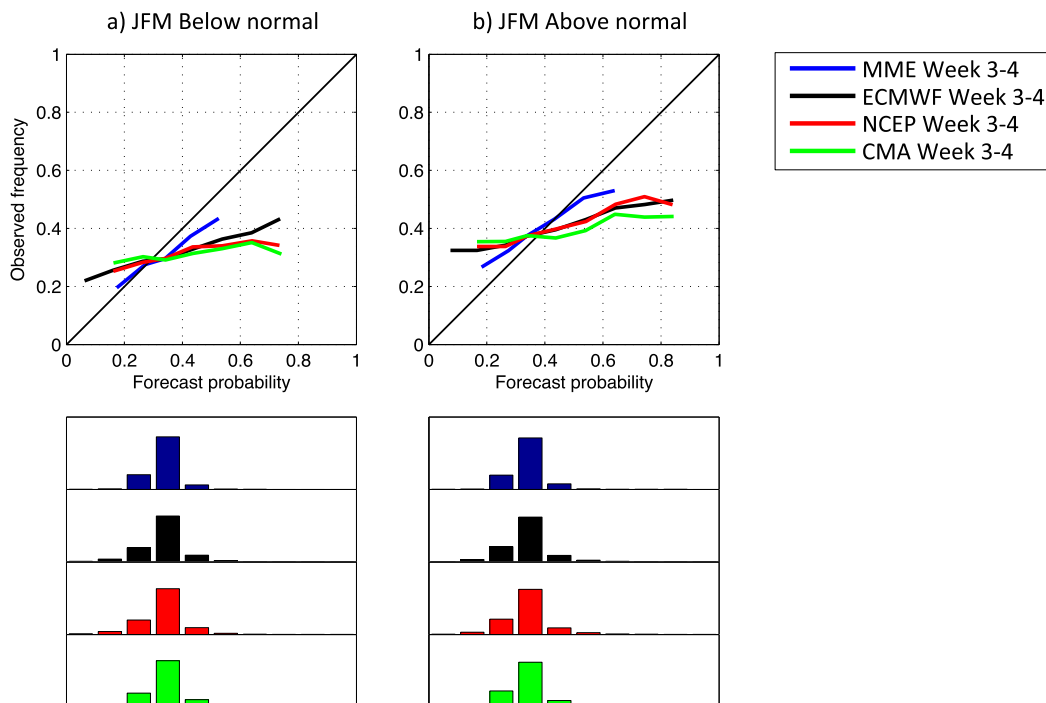


FIG. 7. Week 3–4 reliability diagrams for the below- and above-normal categories from ECMWF (black), NCEP (red), and CMA (green) forecasts with starts in JFM together with their multimodel ensemble (MME, in blue). The frequencies with which each category is forecasted are indicated as bins centered on an integer multiple of 0.10 in histograms plotted under the respective tercile category diagram for each forecast in their respective colors. The bins are projected along the same x axis (forecast probabilities from 0 to 1) and scaled from 0% to 100%. Note that only bins with more than 1% of the total number of forecasts in each category are plotted. Diagrams are computed for all points over continental North America between 20° and 50°N latitudes.

seasons, but also Florida to a lesser extent in JFM, all coinciding with regions of highest skill in week 3–4 outlooks (Fig. 8). Despite the small fraction of total variance explained (7% and 5% in JFM and JAS), RPSS PC1 is significantly correlated with mean RPSS over continental North America up to -0.25 and -0.48 in JFM and JAS, respectively, for the raw weekly forecasts, and -0.42 and -0.57 for the smoothed week 3–4 outlooks, suggesting that a significant amount of mean RPSS variability is represented by RPSS PC1 in both seasons. In JFM, the associated pattern characterized by maximum scores over the southwestern United States/Mexico alongside opposite and less significant loadings to the northwestern/eastern United States, is similar to the correlation maps between weekly GPCP precipitation and both observed Niño-3.4 and MJO RMM2 indices (Figs. 10c and 10g). Overall this suggests that skill is related to tropical forcing particularly toward the southwestern United States consistent with El Niño- and MJO-induced modulations of storm tracks and western U.S. winter precipitation (Monteverdi and Null 1998; Robertson and Ghill 1999; Zhang 2013). Less significant relationships with the indices are seen in JAS.

Both raw and smoothed RPSS PC1s exhibit significant correlations with ENSO in JFM (Table 3), agreeing with the significant correlations between week 3–4 mean RPSS and Niño-3.4 (Table 2) and indicating that skill depends on ENSO, particularly in the southwestern United States/Mexico (Fig. 10a) alongside opposite and less significant relationships in the northwestern/eastern United States. The predictability of week 3–4 precipitation appears to be significantly related to ENSO particularly during winter, when PC1 opposite loadings in the southwestern and northwestern/eastern United States resemble correlation patterns between weekly precipitation and Niño-3.4 (Fig. 10c). Significant correlations between RPSS PC1s and Niño-3.4 modulus are less than half the magnitude of those obtained for Niño-3.4 index, further illustrating that ENSO polarity is a determinant ingredient for skillful regional predictions. In JFM for instance, negative correlations between Niño-3.4 and RPSS PC1 (Table 3) suggest increased skill in the southwestern United States/Mexico during El Niño while it is decreased during La Niña events, but the reverse to the northwestern/eastern United States.

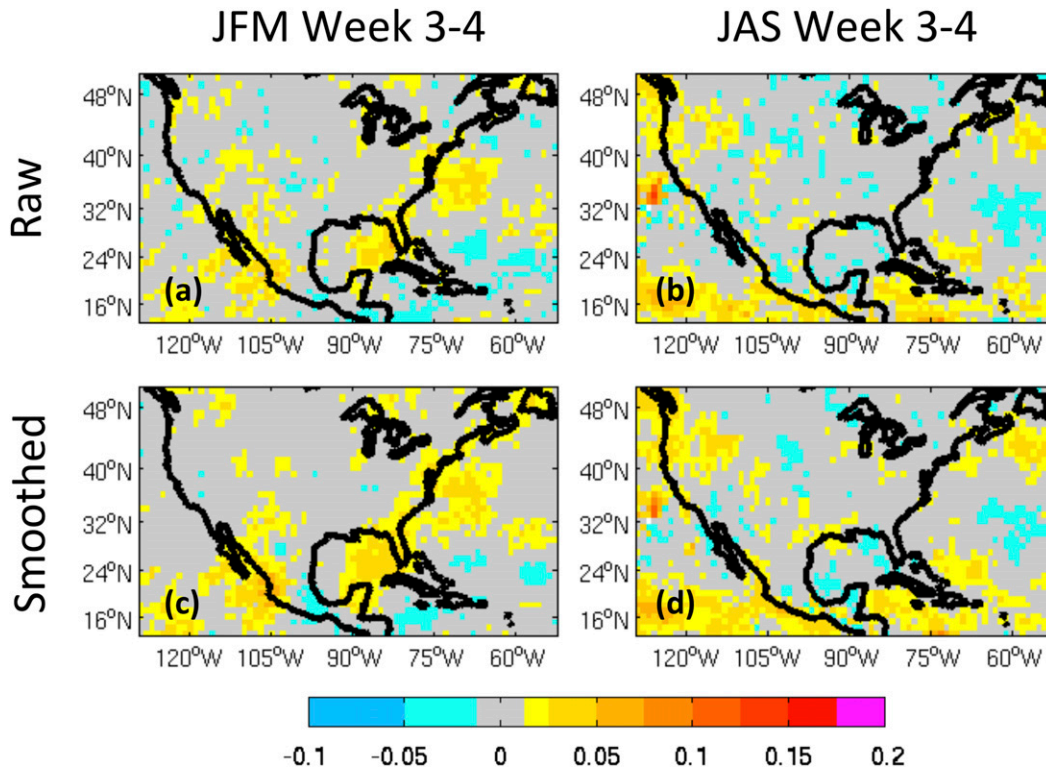


FIG. 8. RPSS for (a),(b) raw and (c),(d) smoothed week 3–4 outlooks from the MME of ECMWF, NCEP, and CMA tercile precipitation forecasts for all starts during the JFM and JAS seasons. Raw and smoothed forecasts are both verified against raw observation data (i.e., unsmoothed).

Table 3 displays significant relationships between the observed MJO and RPSS PC1s in both seasons with highest correlations in JFM, indicating that the skill of week 3–4 outlooks is also significantly correlated with the phase of the MJO, particularly in winter when the loadings of PC1 over the southwestern United States bear strong similarities to correlation patterns between weekly precipitation and MJO RMM2 (Fig. 10g). RPSS PC1 correlations with the modulus of both RMM2 and the best combination of RMMs are significant in the case of smoothed forecasts with winter starts but about half the magnitude of those with MJO indices. These suggest potential skill associations to MJO polarity recalling those obtained for Niño-3.4 and might partly reflect

known correlations between ENSO and MJO activity. Regressing out one signal from the other lowers the relationships with PC1; however, these remain significant (not shown). No spatially coherent structure is identified from RPSS composites conditioned on ENSO phases (not shown), perhaps owing to the small number of El Niño and La Niña events that can be sampled from the rather short 1999–2010 period of analysis.

Additional insights into MJO relationships to predictability and skill are given in Fig. 11, showing the RPSS of smoothed forecasts with starts during MJO phases 1 + 8 and 4 + 5, when convection is respectively increased and decreased over the Western Hemisphere/Africa (Wheeler and Hendon 2004; Ventrice et al. 2011).

TABLE 2. Correlations between JFM and JAS week 3–4 MME RPSS mean averaged over continental North America between 20° and 50°N latitudes and observed Niño-3.4 index (second column), MJO measured by the RMM1 (third column) and RMM2 (fourth column) indices of Wheeler and Hendon (2004), and their best linear combination (fifth column). The second line for each season corresponds to correlations with the modulus of each signal taken as the square of their respective time series. Scores for smoothed week 3–4 MME RPSS are in parentheses and those significant at 95% level of significance using Monte Carlo simulations are indicated with an asterisk.

Week 3–4 MME RPSS	Niño-3.4	RMM1	RMM2	MJO
JFM mean	0.15* (0.25*)	0.02 (0.01)	0.14* (0.15*)	0.14* (0.15*)
	−0.02 (−0.09)	0.01 (−0.02)	−0.02 (−0.01)	0.01 (0.03)
JAS mean	0.16* (0.18*)	0.07 (0.07)	0.03 (0.03)	0.07 (0.07)
	−0.12 (−0.13)	0.03 (−0.02)	0.10 (0.11)	0.12 (0.03)

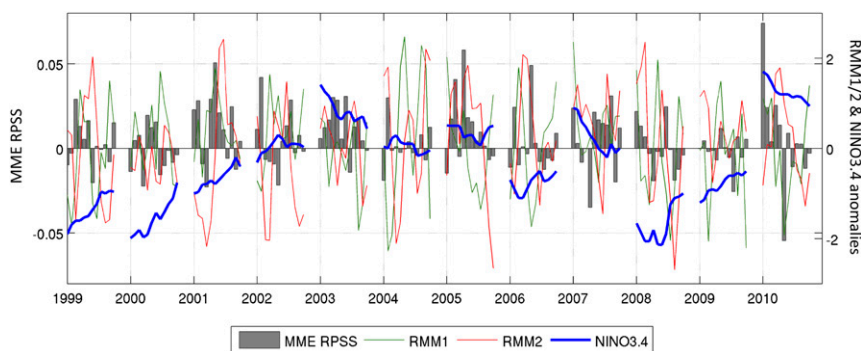


FIG. 9. Raw JFM week 3–4 MME RPSS averaged over continental North America between 20° and 50°N latitudes (bars) together with observed Niño-3.4 index (blue) and MJO measured by the RMM1 (green) and RMM2 (red) indices of Wheeler and Hendon (2004). Corresponding correlations can be found in Table 2.

Even if the 12 years of hindcasts available are limiting regarding statistical significance, the skill of smoothed week 3–4 outlooks over the U.S. East Coast in JFM (Figs. 11a and 11b) could be related to skillful predictions for starts during MJO phases 1 + 8, while skill in the southwestern United States could be drawn from skillful forecasts during MJO phases 4 + 5, as shown by scores significant at 95% level of significance locally using Monte Carlo simulations described in section 2e. These results illustrate the relationship evidenced between the MJO and week 3–4 RPSS PC1 (Table 3) with maximum correlations over the southwestern United States including also Florida (Figs. 10a and 10b). In JAS, the skill of week 3–4 outlooks over the west/central United States and IAS/east Pacific could be similarly related to significant skillful predictions for starts during MJO phases 1 + 8 and 4 + 5, respectively. Even if there is a huge range in how far the MJO would propagate from start dates, part of the skill of week 3–4 outlooks could be drawn from MJO predictability over the region. During MJO phase 5 in winter, for instance, Becker et al. (2011) emphasized a northward shift of the jet stream leading to fewer storms along the U.S. East Coast, while the pineapple express conveyor belt, transporting moisture from the tropical Pacific to the U.S. West Coast, is strengthened and yields to snow increase over the Sierra Nevada (Zhang 2013). Skillful predictions during the above MJO phases could thus be related to modulations of the jet stream and atmospheric rivers affecting east and west U.S. precipitation respectively.

4. Discussion and conclusions

The skill of weekly (week 1–4) and week 3–4 forecasts precipitation tercile probabilities has been examined for S2S forecasts. Probabilities are constructed by applying

extended logistic regression to ECMWF, NCEP, and CMA reforecasts over the common period of 1999–2010. A MME forecast is formed by averaging the individual model probabilities. The regression model can be considered as a reduced form of quantile regression in which the quantile q is one of the predictors, and it is particularly well suited for predicting a probability rather than a measurable physical quantity. As shown in Wilks (2009), this method consequently yields logically consistent sets of forecasts (Fig. 1). Terciles are defined using, for each start and lead, a 3-week window centered on the target week; the regression model is then trained using the same pool of weeks (Fig. 2). Cross validation (the year being forecast is left out) is used both in the definition of tercile categories and the estimation of regression parameters. To accommodate with the discontinuity between zero rain and rainy events in the observed precipitation PDFs, forecasts are only made for weeks where and when the lower tercile is nonzero.

The resulting weekly precipitation tercile forecasts for starts within the JFM and JAS seasons are characterized by low sharpness and decreasing skill with lead times. After weeks 1 and 2, reliability and resolution sharply drop over a broader continental North America domain for individual models (Fig. 3) as well as for their MME (Fig. 4). Predictions are more skillful in winter than summer; however, skill remains low after week 2 as reflected by RPSS in Figs. 5 and 6.

To improve skill and because it is sensible to increase the averaging window with increasing lead, week 3–4 forecasts are also considered. The terciles' definition has been adapted using 6-week windows centered on the 2-week target formed by week 3 and 4 leads in line with the concept of seamless predictions (Zhu et al. 2014). The regression model is subsequently trained on the same pool of weeks defined separately for each

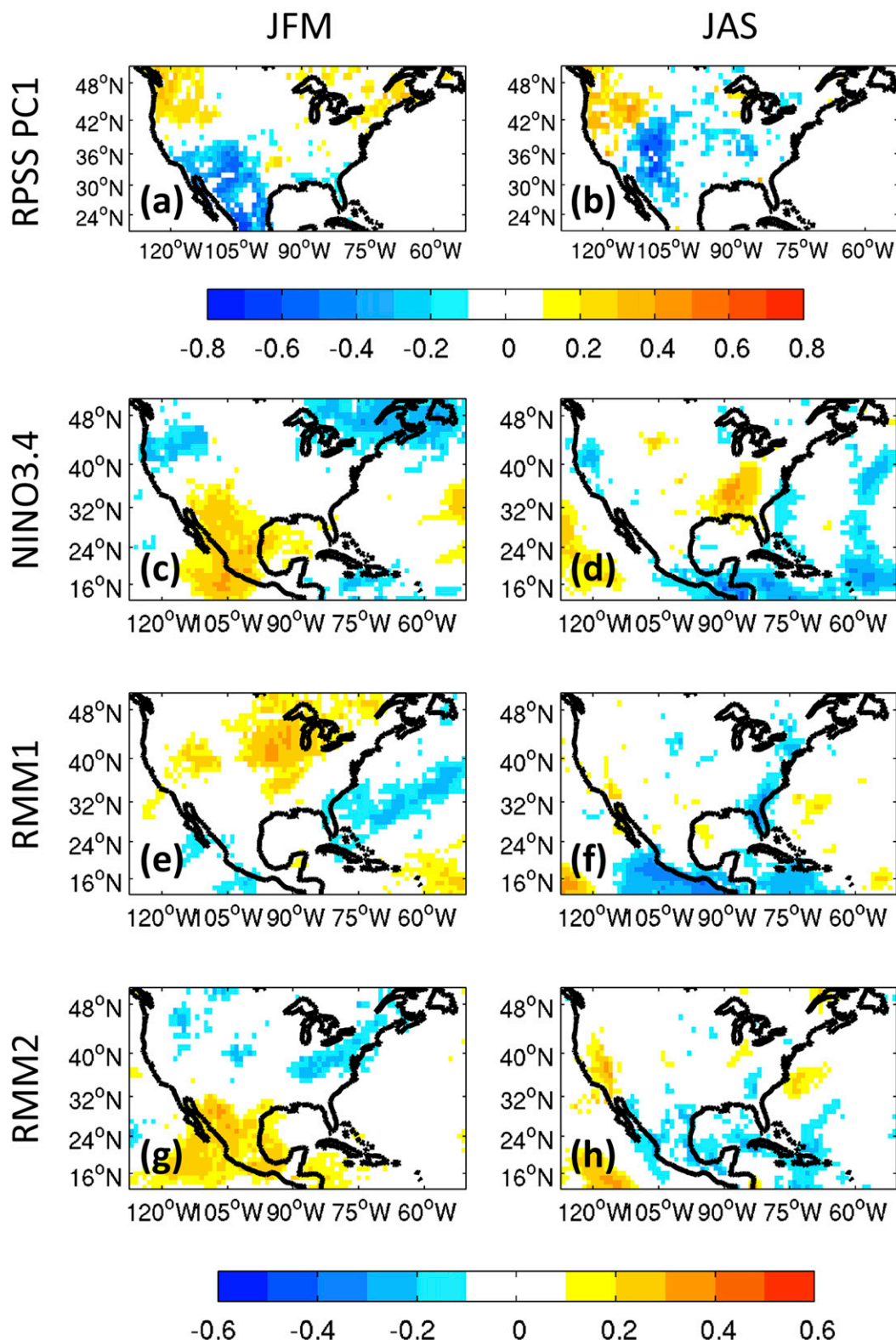


FIG. 10. (top) Spatial correlation patterns of raw week 3–4 multimodel ensemble (MME) RPSS leading principal component (PC1) for starts during the (a) JFM and (b) JAS seasons. (bottom three rows) Correlations between GPCP weekly precipitation and (c),(d) observed weekly Niño-3.4 index and (e),(f) RMM1 and (g),(h) RMM2 indices of [Wheeler and Hendon \(2004\)](#) for all starts in JFM and JAS, respectively. Only scores significant at 95% level of significance using Monte Carlo simulations are plotted.

TABLE 3. As in Table 2, but for week 3–4 MME RPSS PC1.

Week 3–4 MME RPSS	Niño-3.4	RMM1	RMM2	MJO
JFM PC1	−0.45* (−0.36*)	−0.13 (−0.05)	−0.28* (−0.29*)	0.32* (0.30*)
	0.19* (0.16*)	−0.11 (−0.06)	−0.08 (−0.15*)	−0.13 (−0.19*)
JAS PC1	0.02 (−0.04)	0.15* (−0.01)	−0.18* (−0.21*)	0.26* (0.21*)
	0.13 (0.14)	0.04 (0.06)	−0.05 (−0.04)	−0.02 (0.04)

start and week 3 leads in an out-of-sample manner. The forecasts obtained are still characterized by low sharpness but resolution and reliability are increased, with more gain for the MME when compared to weekly forecasts at week 3 and 4 leads separately (Fig. 7, top panels). Spatial smoothing of observation and forecast data used to fit the regression model does not lead to better sharpness nor reliability and resolution (not shown), but increases the extent of skillful areas for both winter and summer forecasts, as shown by RPSS in Figs. 8c,d. Raw and smoothed week 3–4 outlooks are more skillful along the U.S. East Coast and southwestern United States in JFM and the west/central United States and the IAS/eastern Pacific in JAS compared to week 3 and week 4 forecasts.

Relationships between skill and large-scale signals such as ENSO and the MJO (Tables 2–3 and Fig. 9) are examined by applying a PCA to week 3–4 MME RPSS for starts during the JFM and JAS seasons separately (Figs. 10a and 10b). In winter, the pattern of the leading PC is related to skill over the southwestern United States and bears similarities to the correlations between weekly precipitation and both Niño-3.4 and MJO RMM2 indices over North America (Fig. 10). This leading PC is in fact significantly correlated with RPSS averaged over continental North America and ENSO for both seasons, consistent with the forecast relationships to ENSO noted by DelSole et al. (2017), but also the MJO and most particularly RMM2 for winter starts (Table 3). However, increased RPSS for starts only

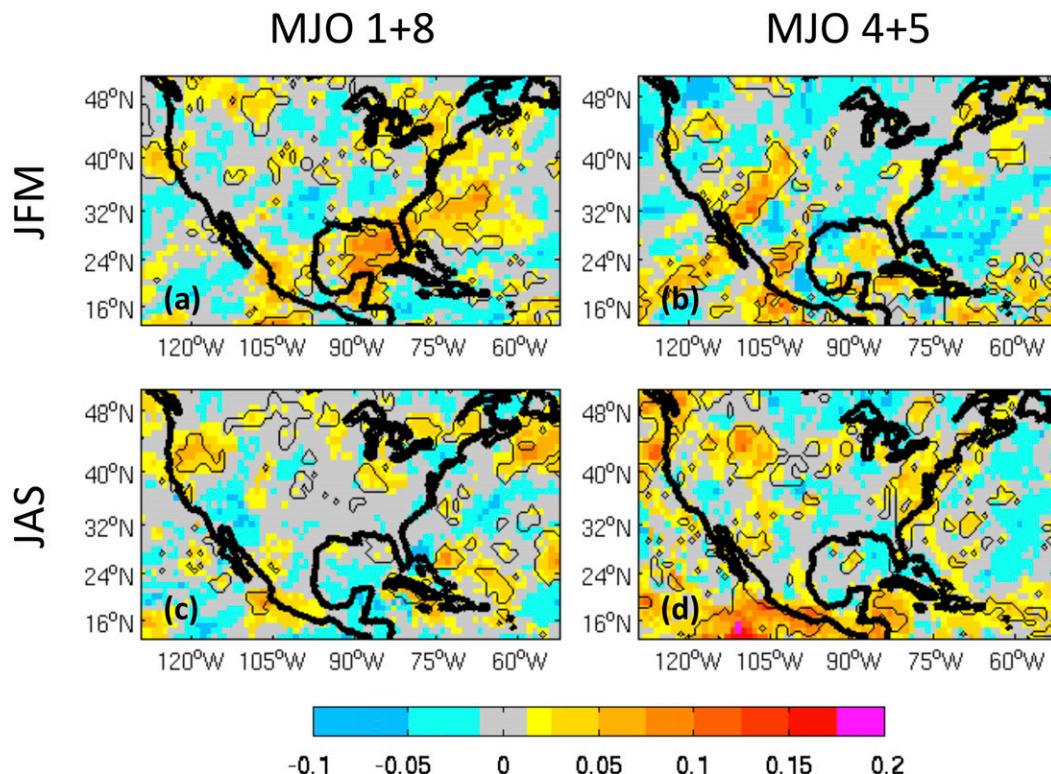


FIG. 11. RPSS for smoothed week 3–4 outlooks from the MME of ECMWF, NCEP, and CMA tercile precipitation forecasts with starts during MJO phases (a),(c) 1 + 8 and (b),(d) 4 + 5 in JFM and JAS. Contours indicate significant scores at 95% level of significance using Monte Carlo simulations.

during MJO phases 1 + 8 and 4 + 5 within the JFM and JAS seasons (Fig. 11) further suggest that some of the skill is drawn from MJO predictability over the region in association to its modulations of the jet stream and atmospheric river affecting U.S. East Coast storms and western U.S. precipitation, respectively (Becker et al. 2011; Zhang 2013). Despite the fact that skill remains low, opportunities of skillful predictions can be increased, as shown with both ENSO and specific MJO phases over the broader North American sector domain, and these need to be exploited further in future studies alongside those arising from other large-scale signals impacting local climate.

Acknowledgments. The authors are grateful to the editor and reviewers including Tom Hamill whose insightful comments resulted in substantial improvements to the manuscript. They would like to acknowledge the financial support of the NOAA Next Generation Global Prediction System (NGGPS) NA15NWS4680014 grant and are grateful to A. Kumar at NCEP Climate Prediction Center (CPC) for useful discussions. Calculations were performed using IRI resources and the S2S subset archived in the IRI Data Library (IRIDL, <http://iridl.ldeo.columbia.edu>) by M. Bell and L. Wang, to whom we are most grateful. The IRIDL was also used to access GPCP 1DD data provided by the NASA Goddard Space Flight Center's Mesoscale Atmospheric Processes Laboratory, which develops and computes the 1DD as a contribution to the GEWEX Global Precipitation Climatology Project.

REFERENCES

- Baldwin, M., and T. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584, doi:10.1126/science.1063315.
- Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126, doi:10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.
- , and M. K. Tippett, 2014: Climate information, outlooks, and understanding—Where does the IRI stand? *Earth Perspect.*, **1** (20), 1–17.
- Becker, E., E. H. Berbery, and R. Higgins, 2011: Modulations of cold-season U.S. daily precipitation by the Madden–Julian oscillation. *J. Climate*, **24**, 5157–5166, doi:10.1175/2011JCLI4018.1.
- DelSole, T., L. Trenary, M. K. Tippett, and K. Pegion, 2017: Predictability of week-3–4 average temperature and precipitation over the contiguous United States. *J. Climate*, **30**, 3499–3512, doi:10.1175/JCLI-D-16-0567.1.
- Garcia, D., 2010: Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. Data Anal.*, **54**, 1167–1178, doi:10.1016/j.csda.2009.09.020.
- Goddard, L., S. Mason, S. Zebiak, C. Ropelewski, R. Basher, and M. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152, doi:10.1002/joc.636.
- Hamill, T. M., 1997: Reliability diagrams for multicategory probabilistic forecasts. *Wea. Forecasting*, **12**, 736–741, doi:10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2.
- , 2012: Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States. *Mon. Wea. Rev.*, **140**, 2232–2252, doi:10.1175/MWR-D-11-00220.1.
- , J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, doi:10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2.
- Holland, M. M., D. A. Bailey, and S. Vavrus, 2011: Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model version 3. *Climate Dyn.*, **36**, 1239–1253, doi:10.1007/s00382-010-0792-4.
- Huffman, G. J., and D. T. Bolvin, 2012: Version 1.2 GPCP one-degree daily precipitation data set documentation. WDC-A, NCDC, 27 pp. [Available online at ftp://meso.gsfc.nasa.gov/pub/1dd-v1.2/1DD_v1.2_doc.pdf.]
- , 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 multisatellite observations. *J. Hydrometeorol.*, **2**, 36–50, doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.
- Koster, R., and Coauthors, 2010: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, **37**, L02402, doi:10.1029/2009GL041677.
- Lin, H., and Z. Wu, 2011: Contribution of the autumn Tibetan plateau snow cover to seasonal prediction of North American winter temperature. *J. Climate*, **24**, 2801–2813, doi:10.1175/2010JCLI3889.1.
- Monteverdi, J., and J. Null, 1998: A balanced view of the impact of the 1997/98 El Niño on Californian precipitation. *Weather*, **53**, 310–313, doi:10.1002/j.1477-8696.1998.tb06406.x.
- Neena, J. M., J. Y. Lee, D. 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, doi:10.1175/JCLI-D-13-00624.1.
- Robertson, A. W., and M. Ghill, 1999: Large-scale weather regimes and local climate over the western United States. *J. Climate*, **12**, 1796–1813, doi:10.1175/1520-0442(1999)012<1796:LSWRAL>2.0.CO;2.
- , 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, doi:10.1175/MWR2818.1.
- Scaife, A., and J. Knight, 2008: Ensemble simulations of the cold European winter of 2005–2006. *Quart. J. Roy. Meteor. Soc.*, **134**, 1647–1659, doi:10.1002/qj.312.
- Tippett, M., A. Barnston, and A. Robertson, 2007: Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *J. Climate*, **20**, 2210–2228, doi:10.1175/JCLI4108.1.
- Ventrone, M., C. Thorncroft, and P. Roundy, 2011: The Madden–Julian oscillation influence on African easterly waves and downstream cyclogenesis. *Mon. Wea. Rev.*, **139**, 2704–2722, doi:10.1175/MWR-D-10-05028.1.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, doi:10.1002/qj.2256.

- , and Coauthors, 2017: The Subseasonal to Seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, doi:10.1175/BAMS-D-16-0017.1.
- Waliser, D. E., 2011: Predictability and forecasting. *Intra-seasonal Variability of the Atmosphere-Ocean Climate System*, W. K. M. Lau and D. E. Waliser, Eds., Springer, 389–423.
- , K. M. Lau, W. Stern, and C. Jones, 2003: Potential predictability of the Madden–Julian oscillation. *Bull. Amer. Meteor. Soc.*, **84**, 33–50, doi:10.1175/BAMS-84-1-33.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124, doi:10.1175/MWR3280.1.
- Wheeler, M., and 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, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.
- Wilks, D., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.
- , 2002: Smoothing forecast ensembles with fitted probability distributions. *Quart. J. Roy. Meteor. Soc.*, **128**, 2821–2836, doi:10.1256/qj.01.215.
- , 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteor. Appl.*, **16**, 361–368, doi:10.1002/met.134.
- , and T. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 2379–2390, doi:10.1175/MWR3402.1.
- Yoo, C., and S.-W. Son, 2016: Modulation of the boreal wintertime Madden-Julian oscillation by the stratospheric quasi-biennial oscillation. *Geophys. Res. Lett.*, **43**, 1392–1398, doi:10.1002/2016GL067762.
- Zhang, C., 2013: Madden–Julian oscillation: Bridging weather and climate. *Bull. Amer. Meteor. Soc.*, **94**, 1849–1870, doi:10.1175/BAMS-D-12-00026.1.
- Zhu, H., M. Wheeler, A. Sobel, and D. Hudson, 2014: Seamless precipitation prediction skill in the tropics and extra tropics from a global model. *Mon. Wea. Rev.*, **142**, 1556–1569, doi:10.1175/MWR-D-13-00222.1.