Precipitation Prediction Skill for the West Coast United States: From Short to Extended Range

BAOXIANG PAN

Center for Hydrometeorology and Remote Sensing, University of California, Irvine, Irvine, California

KUOLIN HSU

Center for Hydrometeorology and Remote Sensing, University of California, Irvine Center of Excellence for Ocean Engineering, National Taiwan Ocean Engineering, Taiwan

AMIR AGHAKOUCHAK AND SOROOSH SOROOSHIAN

Center for Hydrometeorology and Remote Sensing, University of California, Irvine, Irvine, California

WAYNE HIGGINS

Climate Program Office, National Oceanic and Atmospheric Administration, Silver Spring, Maryland

(Manuscript received 12 June 2018, in final form 17 October 2018)

ABSTRACT

Precipitation variability significantly influences the heavily populated West Coast of the United States, raising the need for reliable predictions. We investigate the region's short- to extended-range precipitation prediction skill using the hindcast database of the Subseasonal-to-Seasonal Prediction Project (S2S). The prediction skill–lead time relationship is evaluated, using both deterministic and probabilistic skill scores. Results show that the S2S models display advantageous deterministic skill at week 1. For week 2, prediction is useful for the best-performing model, with a Pearson correlation coefficient larger than 0.6. Beyond week 2, predictions generally provide little useful deterministic skill. Sources of extended-range predictability are investigated, focusing on El Niño–Southern Oscillation (ENSO) and the Madden–Julian oscillation (MJO). We found that periods of heavy precipitation associated with ENSO are more predictable at the extended range period. During El Niño years, Southern California tends to receive more precipitation in late winter, and most models show better extended-range prediction skill. On the contrary, during La Niña years Oregon tends to receive more precipitation in winter, with most models showing better extended-range skill. We believe the excessive precipitation and improved extended-range prediction skill are caused by the meridional shift of baroclinic systems as modulated by ENSO. Through examining precipitation anomalies conditioned on the MJO, we verified that active MJO events systematically modulate the area's precipitation distribution. Our results show that most models do not represent the MJO or its associated teleconnections, especially at phases 3–4. However, some models exhibit enhanced extended-range prediction skills under active MJO conditions.

1. Introduction

The West Coast of the United States receives a majority of its precipitation during the cold season (from October to March). This precipitation supports the water requirements of approximately 15.7% of the nation's population (OECD 2018), generates approximately 52.6% of the domestic hydroelectricity (Uría-Martínez et al. 2017), and waters approximately 21.7% of the country's irrigated farm land (Vilsack and Reilly 2013). Occasional extended wet or dry periods, which are strongly linked to the presence or absence of winter storms, threaten the area's ecological and economic security. Additionally, extremes of droughts and floods can end or occur abruptly, posing challenges to public safety and many other aspects of the society. To better plan for and respond to both the

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-18-0355.s1.

Corresponding author: Baoxiang Pan, baoxianp@uci.edu

DOI: 10.1175/JCLI-D-18-0355.1

beneficial and destructive impacts of the precipitation variations, it is imperative to understand the accuracy and extent of the predictions.

There has been substantial progress in day-to-day precipitation predictions in recent decades (Bauer et al. 2015). The skill improvements are largely due to 1) more realistic estimations of initial atmosphere conditions and 2) improvements in the ability of numerical prediction models to simulate the dynamics and physics of the weather systems. While these advances have led to improved forecasts at longer lead times, it is also true that small-scale errors roughly double in 1–2 days, leading to a rapid loss of useful skill within about 2 weeks (Lorenz 1965).

Provided with loosened requirements on spatial temporal resolutions, deterministic and ensemble-based forecasts occasionally provide useful prediction beyond the synoptic time range (Weber 2015), which has the potential for significant economic value. Generally, the skill depends on 1) the existence of sources of predictability at corresponding temporal ranges and 2) the model's ability to represent the dynamics associated with these modes of variability (National Academies of Sciences, Engineering, and Medicine 2016). For regions with distinct dynamics and sources of predictability, the prediction skill at the extended range is reported to be different (Zhu et al. 2014; Wheeler et al. 2017).

Most winter precipitation events along the West Coast are driven by moisture convergence associated with passing extratropical cyclones (Bao et al. 2006; Dacre et al. 2015). At short to medium ranges, due to the coherent life cycle of cyclone events, cyclogenesis is highly predictable. At the extended range, the prediction skill decreases rapidly (Kumar et al. 2011; Zhu et al. 2014; Robertson et al. 2015), because of the chaotic nature of the baroclinic systems. Teleconnections between extratropical cyclone activities and tropical disturbances offer the potential for extending forecast lead time. For instance, semiperiodic tropical variations, including El Niño–Southern Oscillation (ENSO) and the Madden–Julian oscillation (MJO), often trigger quasi-stationary Rossby wave trains that propagate into midlatitudes, which in turn influence the precipitation distribution (Hoskins and Karoly 1981; Sardeshmukh and Hoskins 1988; Matthews 2004; Hoskins 2013). Such effects are expected to be more significant for the West Coast, given its proximity to the Pacific and the associated sources of potential predictability (Weber 2015; Baggett et al. 2017; Mundhenk et al. 2018).

While much research focuses on explaining teleconnections (Whitaker and Weickmann 2001; Mundhenk et al. 2018; Mamalakis et al. 2018), it is important to keep investigating how forecasts of opportunities are expressed in general circulation models (GCMs), since GCMs remain the most important tool for testing potential sources of predictability. Numerous studies have evaluated the ability of GCMs to predict intraseasonal variability at global and regional scales (Neena et al. 2014; Vitart 2014; Li and Robertson 2015; Tian et al. 2017; Weber and Mass 2017). However, a systematic evaluation of the prediction skill for precipitation at the short to extended range along the West Coast has not been completed.

The purpose of this study is to investigate short- to extended-range precipitation prediction skill for the West Coast during its rainy season. In particular, we explore the impact of the leading modes of intraseasonal to seasonal variability on the distribution and prediction skill of precipitation. Our intention is to use the results as a baseline for follow-on investigations of seamless weather and climate prediction.

The evaluation is based on extended-range retrospective forecast (hereafter referred to as hindcast) experiments developed by 11 operational centers and hosted by the World Weather Research Programme (WWRP)–World Climate Research Programme (WCRP) Subseasonal to Seasonal (S2S) Prediction Project science plan (Vitart et al. 2017). The abundance of hindcast cases and model diversity offer an unprecedented opportunity for investigation of the potential predictability and prediction skill of precipitation. Our specific experiments are as follows:

1) Evaluate the prediction skill for West Coast precipitation during the cold season in each GCM on time scales from the short range to extended range.
2) Investigate the influence of intraseasonal and seasonal variability on precipitation prediction skill in the GCMs at the extended range, with emphasis on ENSO and the MJO.

The rest of the paper is organized as follows. In section 2, we introduce the data used in this study. Section 3 describes the methodology, including evaluation strategy and skill scores. The evaluation results are presented in section 4. Section 5 focuses on the impact of ENSO and the MJO. A discussion and conclusions are provided in section 6.

2. Data and materials

a. Study area

The study area is restricted to the heavily populated coastal region of the western United States, which includes California, western Oregon, and western Washington (to the west of 120°W, as roughly divided by the Cascade Range). The western Cascade Range is considered separately from the eastern range due to their

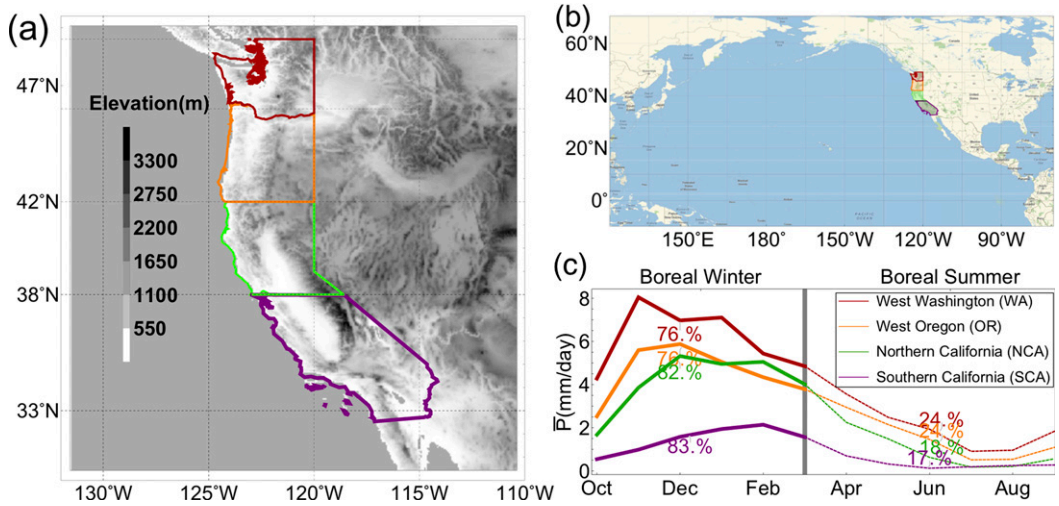


FIG. 1. (a) The geographic map of the West Coast. The elevation data are provided by United States Geological Survey (Gesch et al. 2002). The four subdivisions, namely Southern California (SCA), Northern California (NCA), western Oregon (OR), and western Washington State (WA), are outlined with colored polygons. (b) Geoposition of the study area in a larger scale. (c) The monthly mean precipitation rate for the four subdivisions, based on the CPC precipitation dataset. The boreal winter (October–March) precipitation ratio is labeled.

distinct synoptic and precipitation regimes (Bond and Vecchi 2003). The proximity of this region to the neighboring Pacific Ocean suggests that it is likely to have significant predictability on the extended range. Considering the climate variation within the study area, we further divide the region into four subdivisions, namely Southern California (SCA), Northern California (NCA), western Oregon (OR), and western Washington State (WA). The study area is highlighted in Fig. 1.

b. *CPC Unified Gauge-Based Analysis of Daily Precipitation*

The Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation database (Xie et al. 2007) is used as ''ground truth'' for assessing the performance of the GCMs. This database is constructed by merging various precipitation information sources, including gauge observations, satellite estimates, and numerical model predictions. It provides daily precipitation records covering the contiguous United States from 1948 to 2017 with spatial resolution of 0.20° × 0.25°. Data for the West Coast in cold season (from October to March) are considered. The data are spatially averaged to a 1.5° latitude × 1.5° longitude grid to match the resolution of the GCM hindcast precipitation products.

c. *Climate indices*

The leading patterns of intraseasonal to interannual variability considered in this study are ENSO and the MJO. ENSO is measured by the Niño-3.4 index (Trenberth and Stepaniak 2001), which is the mean monthly sea surface temperature anomalies (SSTA) averaged over 5°S–5°N, 170°–120°W. The MJO is quantified by the Real-Time Multivariate MJO Index (RMM), which consists of the two leading principal components (PCs) of the field that combines average outgoing longwave radiation, zonal wind at 850 hPa, and zonal wind at 200 hPa from 15°S to 15°N (Wheeler and Hendon 2004).

d. *S2S hindcast database*

As a key component of the S2S Prediction Project, the S2S hindcast database offers a large number of hindcast cases to investigate the forecast skill and potential predictability at the extended time range. The database consists of extended-range hindcast cases implemented by 11 operational centers:

- The Australian Bureau of Meteorology (BoM; Alves et al. 2003)
- The China Meteorological Administration (CMA; Wu et al. 2014)
- The European Centre for Medium-Range Weather Forecasts (ECMWF; Vitart et al. 2008)
- The Environment and Climate Change Canada (ECCC; Gagnon et al. 2013)
- The Institute of Atmospheric Sciences and Climate of the National Research Council (ISAC-CNR; Malguzzi et al. 2011)
- The Hydrometeorological Centre of Russia (HMCR; Courtier and Geleyn 1988)
- The Japan Meteorological Agency (JMA; JMA 2013)
- The Korea Meteorological Administration (KMA; Best et al. 2011)

TABLE 1. Model configurations in the S2S hindcast database.

Model	Time range (days)	Resolution	Hindcast coverage	Hindcast frequency	Ensemble size
BoM	0–62	T47L17	1981–2013	6 per month	33
CMA	0–60	T106L40	1994–2014	Daily	4
ECCC	0–32	0.45° × 0.45° L40	1995–2014	Weekly	4
ECMWF	0–46	Tco639/319 L91	Past 20 years	2 per week	11
HMCR	0–61	1.1° × 1.4° L28	1985–2010	Weekly	10
ISAC-CNR	0–32	0.75° × 0.56° L54	1981–2010	Every 5 days	1
JMA	0–33	Tl479/Tl319L100	1981–2010	3 per month	5
KMA	0–60	N216L85	1991–2010	4 per month	3
Météo-France	0–32	T255L91	1993–2014	2 per month	15
NCEP	0–44	T126L64	1999–2010	Daily	4
UKMO	0–60	N216L85	1993–2015	4 per month	3

- The Météo-France/Centre National de Recherche Meteorologiques (Météo-France; Voldoire et al. 2013)
- The National Centers for Environmental Prediction (NCEP; Saha et al. 2014)
- The Met Office (UKMO; Wood et al. 2014)

Model configurations are listed in Table 1.

For each hindcast case, each model is initialized with realistic estimates of the atmosphere, land surface, and ocean states. After initialization, the model iteratively predicts the weather for a preset extension without any boundary constrains. It should be noted that there are large differences between the models when one considers the initialization strategy, dynamics core, parameterization schemes, resolution, ensemble generation scheme, hindcast extensions, ocean and sea ice coupling, and so on. This diversity may offer an opportunity to determine best practices for subseasonal predictions (Vitart et al. 2017; White et al. 2017). In this study, daily precipitation hindcasts for the West Coast during the cold season (from October to March) were used for assessment.

3. Methodology

a. *Evaluation strategy*

A basic fact associated with precipitation prediction is that the position, timing, and intensity of the precipitation forecast diverge from reality as forecast lead time increases. A gridscale, day-to-day deterministic prediction generally holds little efficacy beyond the synoptic range. However, predictions might still have skill if assessed at regional scales or over a range of lead times.

Given this, we implemented evaluations at both stringent and loosened spatial temporal scales. The stringent scale evaluation refers to evaluating each grid's day-to-day prediction skill. In addition, the evaluation is also carried out for regional average predictions and predictions that span specific windows of lead time. The regional average predictions are calculated by averaging predictions within the geographical divisions shown in Fig. 1. Small deviations in predicting cyclone trajectories and their associated precipitation positions are likely to be averaged out in this way. Windows of lead time are defined following Zhu et al. (2014): For a lead time of n days, the subsequent n days average precipitation prediction is evaluated. Thus, an ndnd evaluation refers to evaluating the prediction of the mean precipitation rate from the $(n + 1)$th day to the $(2n)$th day. This strategy offers a fair comparison across a range of time scales from the short range to the extended range, since the deviation in predicting the timing of precipitation at longer lead times will be averaged out for the wider evaluation windows. Using different spatial and temporal scales, we carry out the following four experiments:

1) Daily grid-point-scale evaluation: Evaluate the nth day prediction skill at each grid point, n ranges for the entire period of forecast. The overall skill for each climate division is calculated by averaging skill scores for all the grid points within this division.
2) Daily regional-scale evaluation: Evaluate daily regional average forecasts for each geographical division.
3) Variable temporal windows, grid-point-scale evaluation: Evaluation is carried out at each grid point for various windows of lead time, following the strategy of Zhu et al. (2014).
4) Variable temporal windows, regional-scale evaluation: For each geographical division, the regional average precipitation forecasts are evaluated for variable windows of lead time.

b. *Skill metrics*

1) DETERMINISTIC SKILL METRICS

Two deterministic skill metrics, namely the Pearson correlation coefficient r and the Nash–Sutcliffe model

efficiency coefficient (NSE; Nash and Sutcliffe 1970), are used to assess the performance of the ensemble-mean forecasts. Their formulas are given as follows:

$$r = \frac{E[(P_{\text{obser}} - \overline{P_{\text{obser}}})(P_{\text{simu}} - \overline{P_{\text{simu}}})]}{\sigma_{P_{\text{obser}}} \sigma_{P_{\text{simu}}}}, \quad (1)$$

$$\text{NSE} = 1 - \frac{\sum (P_{\text{obser}} - P_{\text{simu}})^2}{\sum (P_{\text{obser}} - \overline{P_{\text{obser}}})^2}. \quad (2)$$

Here $\overline{P_{\text{obser}}}$ denotes mean value of the precipitation observations and P_{simu} denotes the ensemble mean prediction. Operator E denotes the expectation taken over all available samples, σ denotes standard deviation, r quantifies the linear correlation, and NSE quantifies the relative magnitude of the mean square error compared to the climatology variance.

2) Probabilistic skill metrics

Generally, forecasts beyond 10 days are no longer deterministic (Vitart 2014); each ensemble member from the ensemble forecast system offers useful information in predicting the real-world weather evolution. To account for the information provided by each ensemble member, we also evaluate the probabilistic prediction skill based on all ensemble members. Here, the relative operating characteristics (ROC) score and the continuous ranked probability score (CRPS) are adopted.

The ROC score provides a complete summary of the hit ratio and false alarm ratio for different observation intervals. To calculate the ROC score for each model, we construct a sample space that consists of all ensemble members starting at different dates. For instance, for ECMWF, there are 1482 hindcast starts; each start has 11 ensemble members, so together we have 1482 × 11 samples. For this sample space, the hit ratios and false alarm ratios for observation intervals of (x, ∞) (here x is set as 10 deciles of observation range) are calculated and scatterplotted (hit ratio on the vertical axis and false alarm ratio on the horizontal axis; Fawcett 2004). The points construct the ROC curve, which should be above the 1:1 line if the model has positive skill. The ROC score is defined as the area under the ROC curve. An ROC score closer to 1 indicates higher skill. A no-skill forecast has an ROC score of 0.5 (Vitart 2004). It should be noted that the ROC curve and ROC score are constructed by sorting the elements of the joint distribution of observations and predictions. Thus, the actual numerical values are immaterial, and the final score is insensitive to prediction biases (Wilks 2011). The score reflects the potential performance that can be achieved if the forecasts were correctly calibrated. We acknowledge that there are considerable biases that are introduced by the precipitation-related parameterization schemes. However, since the objective of the paper is to investigate the potential precipitation prediction skills achieved by the dynamic modules of the models, we believe applying the ROC is justified.

The CRPS measures the ensemble forecast skill by comparing the probability distribution of the ensemble predictions and the observations (Hersbach 2000). As is shown in Fig. 2, it is represented as the integrated squared difference between the cumulative probability distribution function (cdf) of the forecasts and the observation.

To evaluate the general performance of the ensemble forecast systems, we apply the mean CRPS ($\overline{\text{CRPS}}$). The formulas are given as follows:

$$\overline{\text{CRPS}} = \frac{1}{n}\text{CRPS} = \frac{1}{n}\int_{\mathbb{R}} [F_{\text{obser}}(x) - F_{\text{simu}}(x)]^2 dx$$
$$\approx \frac{1}{n}\hat{\text{CRPS}} = \frac{1}{n}\int_{\mathbb{R}} [F_{\text{obser}}(x) - \hat{F}_{\text{simu}}(x)]^2 dx. \quad (3)$$

Here n represents the ensemble forecast case count, and F_{obser} (F_{simu}) is the cdf of the precipitation observation (simulation) as shown in Fig. 2; $\hat{F}_{\text{simu}}$ could be estimated by assigning equal probability to each ensemble member.

4. Evaluation results

a. Deterministic skills

1) Pearson correlation coefficient

The estimated Pearson correlation coefficient r between ensemble mean predictions and observations for the four experiments is displayed in Fig. 3.

For day-to-day evaluation (first two columns in Fig. 3), as is expected, each model shows a rapid decrease of r skill with forecast lead time. We labeled the extent to which model's r skill is greater than 0.2 (this threshold is subjective and should be customized regarding specific application purposes). Generally, because of the model performance differences, r falls below 0.2 within 8–15 days (10–16 days) for experiment 1 (experiment 2). A comparison between columns 1 and 2 shows that with a lead time of as much as 2 weeks, regional average predictions generally have higher r skill compared to gridscale predictions. The skill improvements through spatial averaging are most obvious for SCA, which is attributable to the uneven precipitation distribution for this region.

For the temporal interval evaluation (experiments 3 and 4 in the last two columns of Fig. 3), we also give the statistics of best and mean performances at different windows of lead time in Table 2. Within the synoptic

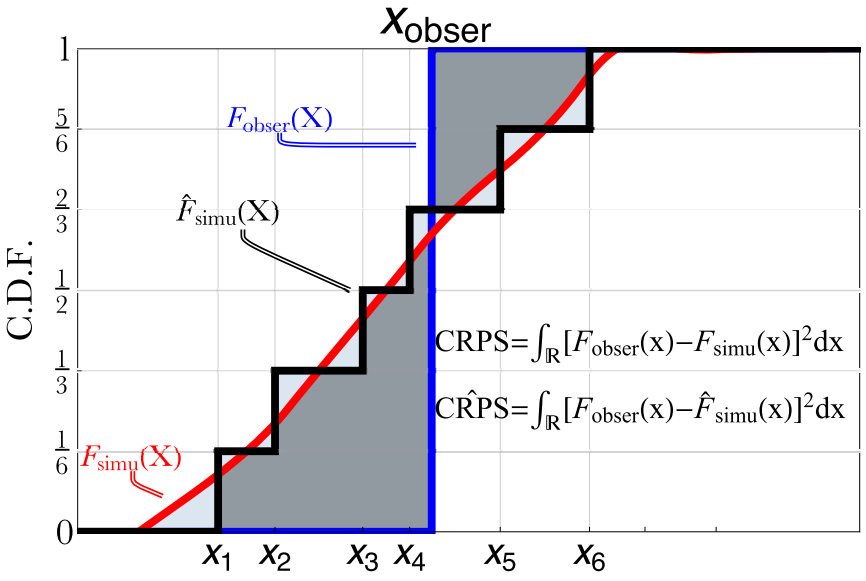


FIG. 2. Explanation of the CRPS using a six-ensemble-member forecast case ($x_1, x_2, \ldots, x_6$). The red (black) line represents the theoretical (empirical) cdf of the precipitation forecast, which is denoted as F_{simu} ($\hat{F}_{simu}$); the blue line represents the cdf of the observation F_{obser}. Since the observation is deterministic, F_{obser} is in the Heaviside function form; i.e., if $x < x_{obser}$, $F_{obser}(x) = 0$, otherwise $F_{obser}(x) = 1$. The CRPS is defined as the integrated squared difference (shaded area) between the cumulative distribution function of the forecasts and the observations. The definition and estimation method are given on the bottom right of the figure.

range, the day 2 (1d1d), day 3–4 (2d2d), and day 5–8 (4d4d) *r* skills are generally of the same order of magnitude (above 0.6 at grid scale and 0.7 at regional scale). This indicates that the decrease of prediction skill as lead time increases is compensated by the expanding of evaluation windows following the *n*d*n*d temporal averaging strategy. The JMA, KMA, ECCC, and ECMWF models have the best performance at this temporal range. It is important to note that these models are of higher resolution compared to the others. For week 2 (1w1w), there is large variability in the models' *r* skills. The best-performing model (ECMWF) achieves *r* skill of approximately 0.5 at grid scale and 0.6 at regional scales. The average performance for all models is of the order of 0.4 for both grid and regional scales. Beyond 2 weeks, the models generally show little usable skill. However, it is noteworthy that some models show unexpectedly good performance at this time range, such as BoM for SCA and HMCR for WA.

The results above suggest that models' *r* skills are distinct regarding different regions and forecast lead time. For the same region and lead time, the informative predictable range may differ by up to 6–7 days due to the model performance differences. The huge sample size of the S2S dataset offers opportunity to test the significance of model performance differences at critical forecast lead time periods. The results would benefit model selections for practical forecasts and multimodel ensemble predictions. Below we carry out the significance test on models' *r* skill differences for the day 2, day 7, week 2, and week 3–4 periods. These periods are selected since they represent critical lead time and scales in weather forecasts. To perform the test, we first applied the Fisher *r*-to-*z* transformation (Fisher 1921) on the *r* estimations. Later, we applied significance test on the *z* statistics to assess the significance of the difference between the models' *r* skills. Results are shown in Fig. 4.

For the day 2 forecast (row 1 in Fig. 4), ECCC, ECMWF, ISAC-CNR, JMA, and KMA generally show significant advantages over the other models, while the BoM and CMA models show significantly lower skill. For the day 7 forecast (row 2), the ECCC, ECMWF, JMA, and KMA models still lead the performance, while ISAC-CNR loses its advantage over most models. This might be because the ISCA-CNR model is applying deterministic rather than ensemble forecasts here. For week 2 forecast (row 3), the best-performing models are ECCC, ECMWF, and JMA. ECMWF shows significant advantage over all the other models except for the WA prediction when compared against JMA. For the week 3–4 forecast (row 4), although there is essentially no useful *r* skill,

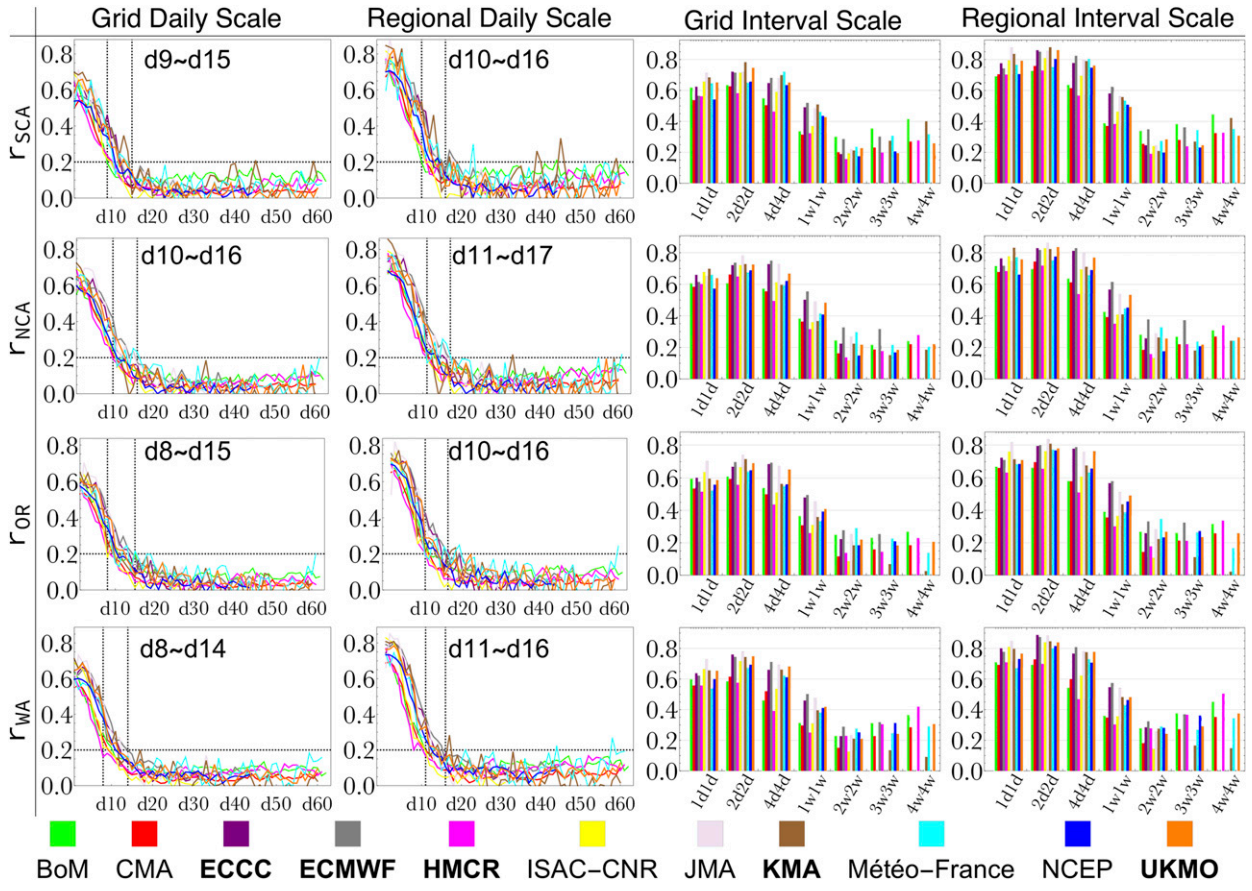


FIG. 3. The Pearson correlation coefficient between the ensemble mean of precipitation predictions and the observations for the four experiments defined in section 3. The evaluation results for the four divisions are shown in rows 1–4. The columns represent different experiments. (first column) The daily grid-point-scale evaluation results, (second column) the daily regional-scale evaluation results, (third column) the variable temporal windows, grid-point-scale evaluation results, and (fourth column) the variable temporal windows, regional-scale evaluation results. For the first and second columns, the extents to which $r > 0.2$ for different models are labeled.

the ECMWF model still shows advantage over the rest of the models.

2) NASH–SUTCLIFFE MODEL EFFICIENCY

Generally a large value of r relates to positive NSE, which indicates that the model outperforms the baseline of climatology. However, evaluation of the original precipitation ensemble mean predictions shows negative NSE in most experiments. Given this discrepancy, we performed a linear bias correction for each scale. Results after the correction are shown in Fig. 5. The best and mean performances are given in Table 3.

Overall, the NSE results are similar to those shown earlier. Daily-scale NSE reaches 0.2 within approximately 5–10 days. Models lose their advantage over climatology after approximately 2 weeks (NSE $\approx$ 0). Through spatial averaging, short-range NSE could be improved by 0.2, and the range of NSE > 0.2 is extended by 1 day.

Considering evaluations at different windows of forecast lead time, for day 2 and day 3–4 predictions, the JMA, KMA, and ECCC models still have the best performance. For the medium range, ECMWF achieves the largest NSE. Week 2 predictions are of considerable value, with NSE around 0.25 at grid scale and 0.35 at regional scale. Beyond two weeks, skills decrease rapidly.

b. Probabilistic skill

1) ROC SCORE

This section evaluates the models' probabilistic skill using the ROC score. To illustrate the idea of ROC, we draw the ROC curves for ECMWF regional predictions in Fig. 6. Each labeled point represents the hit ratio (HR) and false alarm ratio (FAR) for a corresponding interval of precipitation observations. For instance, the point labeled 0.9 represents the (HR, FAR) of prediction for the $(P_{90\%}, \infty)$ interval, where $P_{90\%}$ represents the 90%

TABLE 2. Correlation coefficient of precipitation predictions at temporal interval scales.

	SCA		NCA		OR		WA	
Scale	Best	Mean	Best	Mean	Best	Mean	Best	Mean
Day 2 grid	0.71 (JMA)	0.62	0.7 (KMA)	0.63	0.71 (JMA)	0.59	0.73 (JMA)	0.62
Day 2 regional	0.88 (JMA)	0.76	0.83 (KMA)	0.74	0.82 (JMA)	0.71	0.85 (JMA)	0.76
Day 3–4 grid	0.78 (KMA)	0.69	0.78 (JMA)	0.7	0.74 (JMA)	0.66	0.78 (JMA)	0.7
Day 3–4 regional	0.88 (KMA)	0.81	0.86 (JMA)	0.79	0.84 (JMA)	0.76	0.89 (ECCC)	0.81
Day 5–8 grid	0.72 (Météo-France)	0.62	0.75 (ECMWF)	0.63	0.7 (ECMWF)	0.58	0.71 (ECMWF)	0.6
Day 5–8 regional	0.82 (ECMWF)	0.73	0.83 (ECMWF)	0.7	0.79 (ECMWF)	0.67	0.81 (ECMWF)	0.69
Week 2 grid	0.52 (ECMWF)	0.43	0.56 (ECMWF)	0.42	0.5 (ECMWF)	0.38	0.5 (ECMWF)	0.38
Week 2 regional	0.62 (ECMWF)	0.5	0.61 (ECMWF)	0.47	0.58 (ECMWF)	0.44	0.58 (ECMWF)	0.45
Week 3–4 grid	0.3 (BoM)	0.22	0.33 (ECMWF)	0.22	0.29 (Météo-France)	0.2	0.29 (ECMWF)	0.22
Week 3–4 regional	0.35 (ECMWF)	0.26	0.38 (ECMWF)	0.25	0.35 (Météo-France)	0.24	0.32 (ECMWF)	0.26
Week 4–6 grid	0.35 (BoM)	0.26	0.32 (ECMWF)	0.2	0.25 (ECMWF)	0.19	0.32 (ECMWF)	0.26
Week 4–6 regional	0.38 (BoM)	0.29	0.37 (ECMWF)	0.24	0.33 (ECMWF)	0.24	0.38 (BoM)	0.31
Week 5–8 grid	0.41 (BoM)	0.32	0.28 (HMCR)	0.23	0.27 (BoM)	0.18	0.42 (HMCR)	0.29
Week 5–8 regional	0.45 (BoM)	0.36	0.34 (HMCR)	0.28	0.34 (HMCR)	0.23	0.51 (HMCR)	0.36

quantile of observed precipitation. The farther a point is above the 1:1 line, the more likely a $P > P_{90\%}$ forecast is true. As is shown, points in all subfigures here are above the 1:1 line, showing a larger HR than FAR at different evaluating thresholds for all scales. Small precipitation cases generally appear on the top-right part of the ROC curve (larger HR but also larger FAR), while large precipitation cases appear on the bottom-left part of the ROC curve (smaller HR but also smaller FAR). For short to medium range, the ROC curves of day 2 (1d1d), days 3–4 (2d2d), and days 5–8 (4d4d) show considerable overlap. The ROC curve for week 2 (1w1w) prediction falls below the previous three cases. For weeks 3–4 (2w2w) and weeks 4–6 (3w3w), again the curves overlap, and they fall below all the previous cases. The ROC score is defined as the area below the ROC curve, which summarizes the probabilistic prediction skill for different evaluating intervals. The scores are given at bottom right in each subfigure. The days 3–4 prediction achieves a slightly better ROC score compared to day 2 and days 5–8. The week 2 (1w1w) prediction achieves a ROC score above 0.7 in all divisions. For weeks 3–4 (2w2w) and weeks 4–6 (3w3w), the score is around 0.6, showing better performance than random guess.

Based on this same approach, we calculated the ROC scores for all models in different evaluation experiments. Results are shown in Fig. 7. For day-to-day evaluation (first two columns), we labeled the extent to which the ROC score is larger than 0.6 (again, this threshold is subjective and should be adjusted if necessary). Generally, the daily ROC scores begin to fall below 0.6 in the second week. They reach 0.5 at approximately 20 days, which means that beyond 20 days day-to-day estimations show no advantage over climatology.

Considering evaluations at different windows of lead time (last two columns in Fig. 7), for short to medium range, the best-performing models achieve ROC scores above 0.85 for day 2 (1d1d), days 3–4 (2d2d), and days 5–8 (4d4d). For week 2, the ROC score is around 0.65 for grid scale and 0.7 for regional scale. Models still hold positive probabilistic prediction skills beyond week 2. The best and mean performances are summarized in Table 4.

2) CONTINUOUS RANKED PROBABILITY SCORE

Compared to the previous skill metrics of r, NSE, and the ROC score, which are dimensionless and roughly bounded within certain ranges, the CRPS is of the same dimension as the predictand (Hersbach 2000; Vogel et al. 2018); also, it has a lower bound of 0 but is not restricted by an upper bound. Because of these characteristics, evaluations using $\overline{\text{CRPS}}$ at different spatiotemporal scales are sensitive to the distribution variation of the predictand. Thus, the results for difference scales cannot be compared directly. Given this deficiency, we do not present the evaluation results here, but show them in the online supplemental material. The information and implications of the evaluation are discussed in the following section.

For the day-to-day evaluation, as can be expected, all models show increase of $\overline{\text{CRPS}}$ with forecast lead time until $\overline{\text{CRPS}}$ becomes relatively steady after approximately 2 weeks. Compared to the evaluation results using r, NSE, and the ROC score, the $\overline{\text{CRPS}}$-forecast lead day curves show considerable oscillation. By examining the mean precipitation rate on different forecast lead days, we confirmed that the oscillations can be attributed to the variation of precipitation rates on these days. Considering the model performance differences,

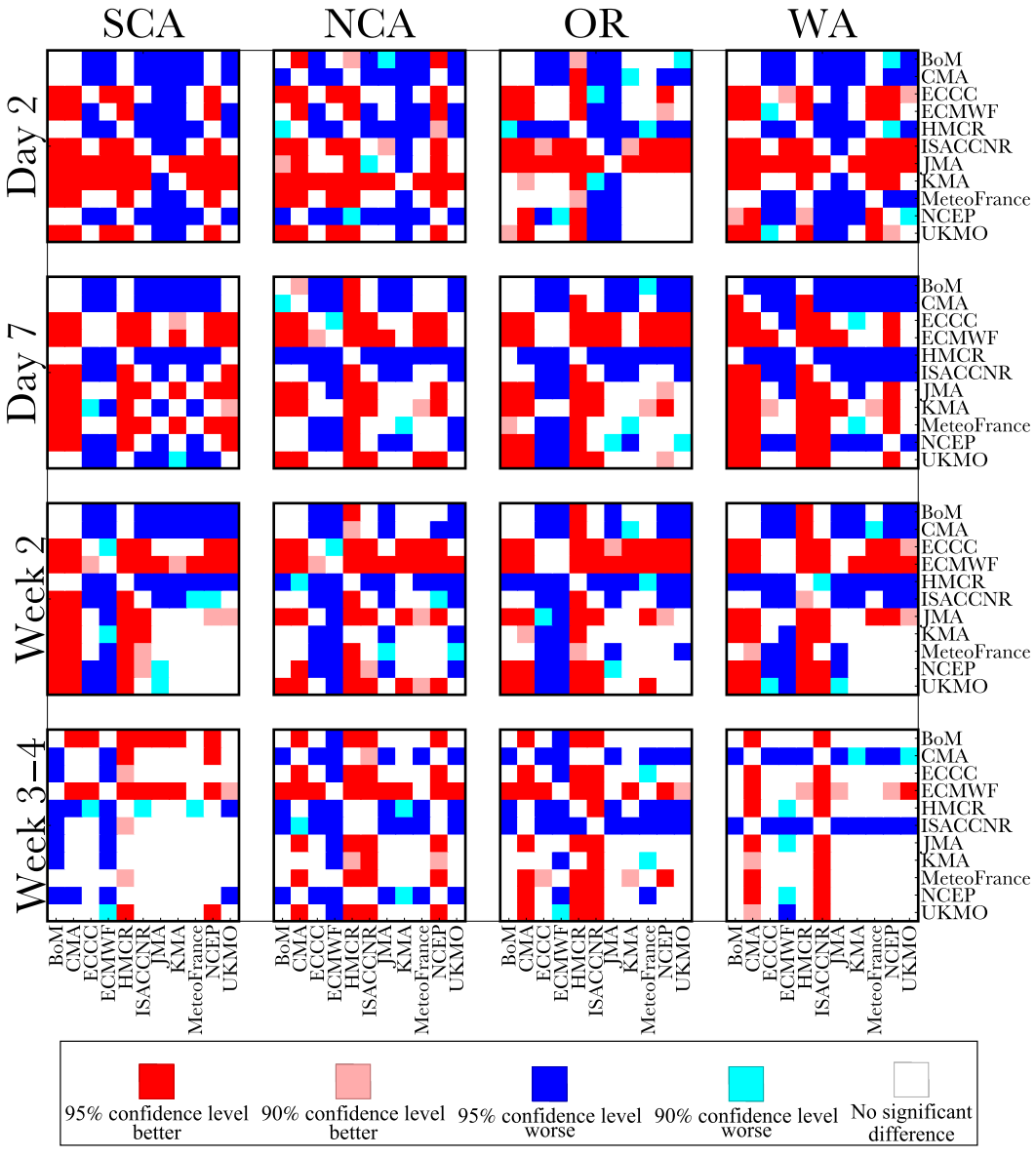


FIG. 4. The significance test result for regional average predictions in the four divisions for day 2, day 7, week 2, and weeks 3–4. The rows represent the forecast period; the columns represent the geographic divisions. For each matrix, the grid at row m, column n is labeled red if the r skill for the model at row m is better than the model at column n at the 95% confidence level, and so on for other colored labels.

the BoM model shows best (lowest) $\overline{\text{CRPS}}$ in most daily evaluation cases; also, for longer forecast lead time, the $\overline{\text{CRPS}}$ for BoM does not increase as significantly as for the other models. By examining the model configuration, we found that the BoM precipitation product is of lower spatial resolution compared to the rest of the models. Since evaluations using other skill scores suggest no significant advantage of BoM, we believe the low $\overline{\text{CRPS}}$ for BoM is due to its low spatial resolution rather than advantageous performance. On the other hand, the ISAC-CNR model shows significant worse (higher) $\overline{\text{CRPS}}$. This is because the ISAC-CNR model provides deterministic forecast with single ensemble member, while the other models have multiple ensemble members. Results here suggest the advantage of applying an ensemble forecast rather than a deterministic forecast, especially for the extended-range period. For the rest of the nine models, which are of same spatial resolution, the $\overline{\text{CRPS}}$-forecast lead day curves show considerable overlap, with the ECMWF model showing slightly better performance.

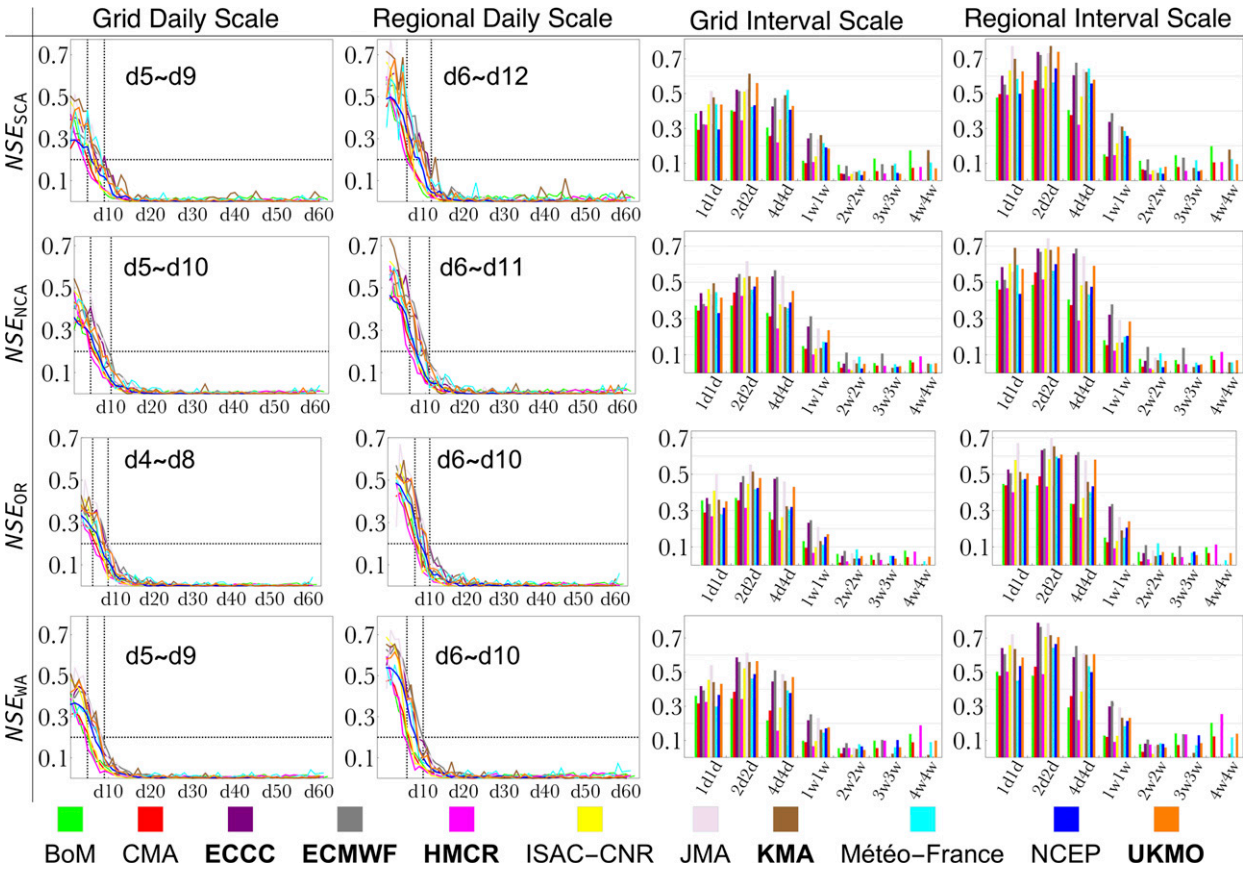


FIG. 5. As in Fig. 3, but using Nash–Sutcliffe efficiency of ensemble mean predictions (linear bias corrected). For day-to-day evaluation, the extent to which models have NSE > 0.2 is labeled.

Considering the evaluation results at different windows of lead time, larger $\overline{\text{CRPS}}$ values are achieved for larger evaluation time windows. This is because the $\overline{\text{CRPS}}$ has the same dimension as the predictand and is scaled up as the aggregated precipitation amount increases with forecast time window width. In the online supplemental material, we show the rescaled $\overline{\text{CRPS}}$ by dividing the $\overline{\text{CRPS}}$ by the corresponding time window width. Results show no apparent skill variation for different evaluation windows.

TABLE 3. Nash–Sutcliffe model efficiency of precipitation predictions (linearly bias corrected) at temporal interval scales.

	SCA		NCA		OR		WA	
Scale	Best	Mean	Best	Mean	Best	Mean	Best	Mean
Day 2 grid	0.51 (JMA)	0.39	0.49 (KMA)	0.41	0.5 (JMA)	0.35	0.54 (JMA)	0.4
Day 2 regional	0.77 (JMA)	0.59	0.69 (KMA)	0.54	0.67 (JMA)	0.5	0.72 (JMA)	0.58
Day 3–4 grid	0.61 (KMA)	0.48	0.62 (JMA)	0.5	0.55 (JMA)	0.44	0.61 (JMA)	0.49
Day 3–4 regional	0.77 (KMA)	0.65	0.74 (JMA)	0.63	0.7 (JMA)	0.58	0.79 (ECCC)	0.66
Day 5–8 grid	0.52 (Météo-France)	0.4	0.57 (ECMWF)	0.41	0.49 (ECMWF)	0.35	0.51 (ECMWF)	0.37
Day 5–8 regional	0.68 (ECMWF)	0.54	0.69 (ECMWF)	0.5	0.62 (ECMWF)	0.45	0.65 (ECMWF)	0.49
Week 2 grid	0.27 (ECMWF)	0.19	0.31 (ECMWF)	0.19	0.25 (ECMWF)	0.15	0.25 (ECMWF)	0.16
Week 2 regional	0.39 (ECMWF)	0.25	0.38 (ECMWF)	0.23	0.34 (ECMWF)	0.2	0.33 (ECMWF)	0.21
Week 3–4 grid	0.09 (BoM)	0.05	0.11 (ECMWF)	0.05	0.09 (Météo-France)	0.05	0.09 (ECMWF)	0.05
Week 3–4 regional	0.12 (ECMWF)	0.07	0.14 (ECMWF)	0.07	0.12 (Météo-France)	0.06	0.11 (ECMWF)	0.07
Week 4–6 grid	0.13 (BoM)	0.07	0.11 (ECMWF)	0.05	0.07 (ECMWF)	0.04	0.1 (ECMWF)	0.08
Week 4–6 regional	0.15 (BoM)	0.09	0.14 (ECMWF)	0.06	0.11 (ECMWF)	0.06	0.14 (BoM)	0.1
Week 5–8 grid	0.18 (KMA)	0.11	0.09 (HMCR)	0.06	0.08 (BoM)	0.05	0.19 (HMCR)	0.1
Week 5–8 regional	0.2 (BoM)	0.13	0.12 (HMCR)	0.08	0.11 (HMCR)	0.06	0.26 (HMCR)	0.14

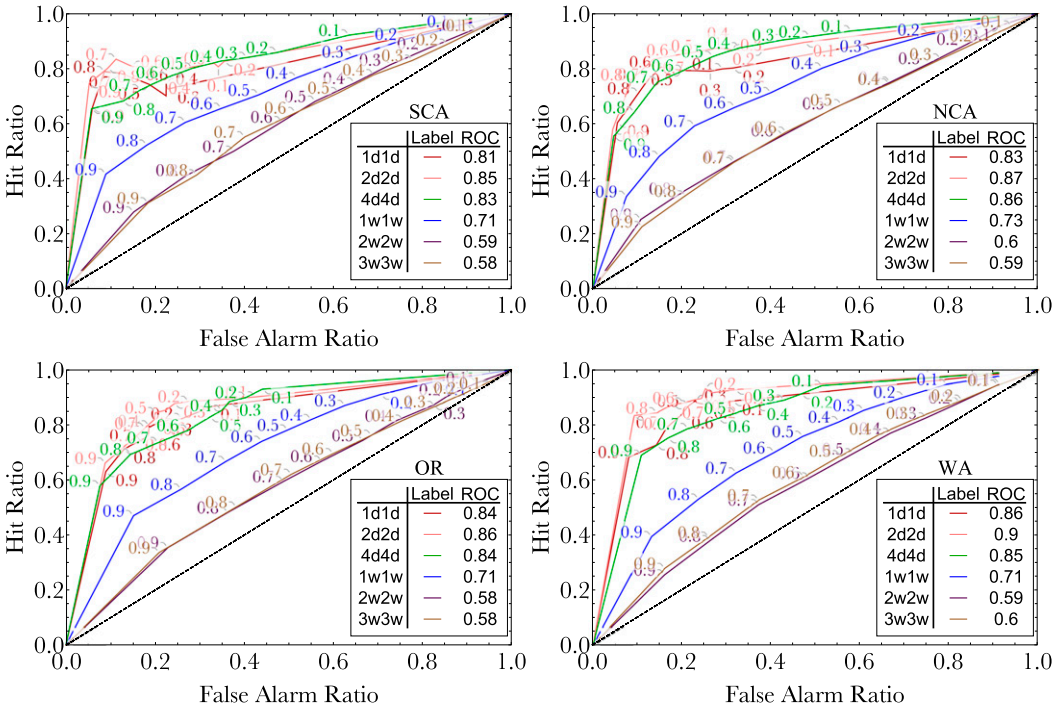


FIG. 6. ROC curves for ECMWF regional precipitation predictions for the four regions at various windows of lead time. The points with labeled numbers show the hit ratio and false alarm ratio of a corresponding threshold. The ROC scores for different intervals are given in the inset tables.

Overall, the $\overline{\text{CRPS}}$ evaluation results confirm the advantage of applying ensemble forecasts rather than deterministic forecasts, especially for the extended range. However, the scaling issue of the $\overline{\text{CRPS}}$ restricts us from comparing model performance differences at different spatiotemporal scales.

5. The impacts of ENSO and MJO

Our evaluation results show a sharp drop in prediction skill after week 1. Beyond this time range, predictions rely heavily on the existence of sources of predictability and the model's ability to realize them as prediction skill. In this section, we explore the impact of key sources of intraseasonal to seasonal predictability on precipitation distribution and prediction skill at the extended range. In particular, the focus is on ENSO and the MJO.

a. ENSO

ENSO is a semiperiodic variation in winds and sea surface temperatures over the tropical eastern Pacific Ocean. Generally, it is represented by the corresponding area's sea surface temperature anomaly, as shown in Fig. 8. ENSO influences the seasonal variability across the tropical Pacific and in much of the extratropics, including the North Pacific and North America. Precipitation anomalies in the regions along the West Coast are believed to be influenced by ENSO through its influence on the Aleutian low (Bjerknes 1969; Schonher and Nicholson 1989) and subtropical jets (Rasmusson and Wallace 1983; Trenberth et al. 1998). The connection has been investigated extensively using observations (Schonher and Nicholson 1989; Jong et al. 2016), models (Wang et al. 2009; Dettinger 2011), and composite approaches (Seager et al. 2015; Madadgar et al. 2016). However, results suggest that the connections, such as the magnitude and sign of precipitation anomalies, are not robust (Yarnal and Diaz 1986).

In the following part of this section, we explore ENSO's impact on the precipitation distribution and extended-range prediction skill. We focus on evaluations at the weekly scale, as this is in accordance with the rough time scale for the life cycle of cyclone events.

1) INFLUENCE ON PRECIPITATION DISTRIBUTION

To investigate the weekly precipitation statistics conditioned on the ENSO phases, we first constructed the wintertime weekly precipitation time series by averaging precipitation records for consecutive 7-day windows (from day 1 to day 7, from day 2 to day 8, etc.). Next, we removed the impact of seasonal cycle by

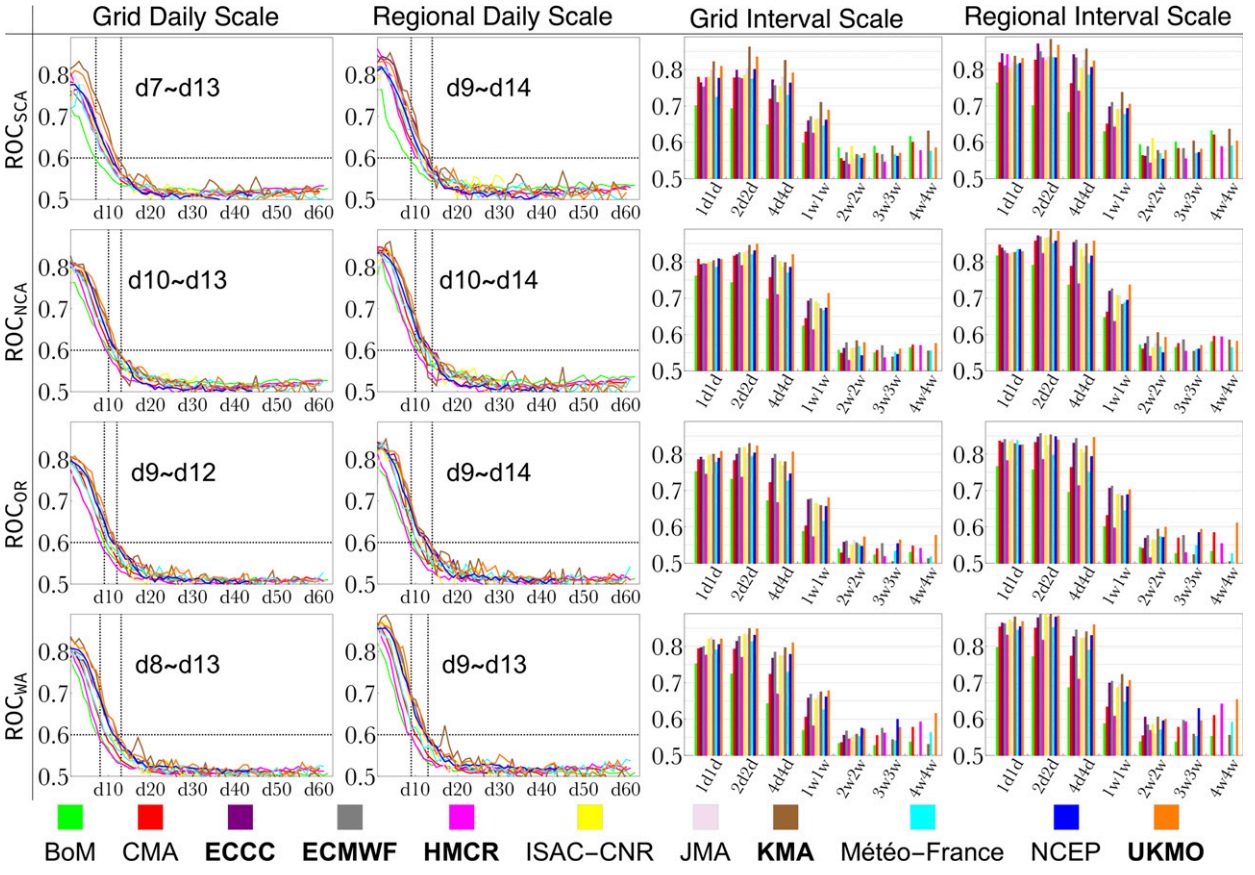


FIG. 7. As in Fig. 3, but using the ROC score of ensemble precipitation predictions. For day-to-day evaluation, the extent to which models have a ROC score > 0.6 is labeled.

subtracting the leading two harmonics of the weekly series, using a fast Fourier transform (Welch 1967). Finally, we constructed the empirical distribution of weekly precipitation anomalies conditioned on different ENSO phases for early winter season [October–December (OND)], late winter season [January–March (JFM)], and the entire winter season [October–March (O–M)]. The distributions were estimated using the smooth kernel method, built on 5000 bootstrap samples. Samples were selected randomly (with replacement) from weekly precipitation anomaly time series for the corresponding season and ENSO phase. We applied the Kolmogorov–Smirnov test to determine whether the distributions differ significantly due to ENSO.

Figure 9 shows the empirical distributions of weekly precipitation anomalies. In early winter (OND), the El Niño phase tends to favor negative precipitation anomalies, while La Niña phase tends to favor positive precipitation anomalies. This pattern is more obvious for NCA, OR, and WA, as compared to SCA. SCA has a higher probability of receiving abnormally high precipitation events during La Niña, as shown in the tail of the distribution. For NCA, there tends to be more precipitation during ENSO-neutral phases, on average. For OR and WA, there tends to be more precipitation during La Niña.

In late winter (JFM), with the onset of the rainy season in SCA, ENSO's influence is flipped and strengthened. For El Niño (La Niña) phases, SCA receives 0.35 mm day^{-1} more (0.3 mm day^{-1} less) precipitation than climatology. For NCA, the El Niño phase also tends to favor more precipitation, but the negative influence of La Niña is not as obvious as it is for SCA. For OR, and WA, like early winter, the La Niña phase tends to favor more precipitation compared to climatology; however, El Niño is not accompanied by less precipitation, on average.

The last column shows the weekly precipitation anomaly distribution for the entire winter season. As a summary of the two cases analyzed above, we see that during El Niño years, SCA tends to receive more precipitation, while NCA, OR, and WA tend to receive less precipitation. During La Niña years, SCA tends to receive less precipitation while the others tend to receive more. However, the variance for the distribution of precipitation anomalies conditioned on ENSO phases is

TABLE 4. ROC score at temporal interval scales.

Scale	SCA		NCA		OR		WA	
	Best	Mean	Best	Mean	Best	Mean	Best	Mean
Day 2 grid	0.82 (KMA)	0.77	0.81 (NCEP)	0.8	0.81 (UKMO)	0.79	0.83 (JMA)	0.8
Day 2 regional	0.84 (ECCC)	0.82	0.85 (CMA)	0.83	0.84 (JMA)	0.82	0.88 (KMA)	0.86
Day 3–4 grid	0.86 (KMA)	0.79	0.85 (UKMO)	0.82	0.83 (KMA)	0.8	0.85 (KMA)	0.81
Day 3–4 regional	0.88 (KMA)	0.83	0.89 (KMA)	0.86	0.86 (ECMWF)	0.83	0.9 (KMA)	0.87
Day 5–8 grid	0.83 (KMA)	0.75	0.82 (UKMO)	0.78	0.81 (UKMO)	0.75	0.81 (UKMO)	0.75
Day 5–8 regional	0.86 (KMA)	0.8	0.86 (ECMWF)	0.81	0.85 (UKMO)	0.79	0.86 (UKMO)	0.8
Week 2 grid	0.71 (KMA)	0.66	0.71 (UKMO)	0.67	0.68 (UKMO)	0.64	0.68 (UKMO)	0.64
Week 2 regional	0.74 (KMA)	0.69	0.74 (UKMO)	0.69	0.71 (ECMWF)	0.67	0.72 (KMA)	0.67
Week 3–4 grid	0.59 (ISACCNR)	0.56	0.58 (KMA)	0.56	0.57 (UKMO)	0.55	0.58 (NCEP)	0.56
Week 3–4 regional	0.61 (ISACCNR)	0.57	0.61 (KMA)	0.57	0.6 (UKMO)	0.57	0.61 (KMA)	0.58
Week 4–6 grid	0.59 (KMA)	0.57	0.57 (ECMWF)	0.55	0.56 (UKMO)	0.54	0.6 (NCEP)	0.56
Week 4–6 regional	0.6 (KMA)	0.58	0.59 (ECMWF)	0.57	0.59 (UKMO)	0.56	0.63 (NCEP)	0.58
Week 5–8 grid	0.63 (KMA)	0.6	0.58 (UKMO)	0.57	0.58 (UKMO)	0.54	0.62 (UKMO)	0.57
Week 5–8 regional	0.64 (KMA)	0.61	0.6 (CMA)	0.58	0.61 (UKMO)	0.55	0.66 (UKMO)	0.6

considerably large, making ENSO a less robust indicator for precipitation predictions.

2) INFLUENCE ON PRECIPITATION PREDICTION SKILLS

The previous section discussed ENSO's influence on weekly precipitation statistics. Ideally, we would like to see these effects be well simulated in the GCMs. If so, then realistic representation of ENSO in models could facilitate useful boundary forcings in conditioning the precipitation distribution, and predictions at the extended range could be improved.

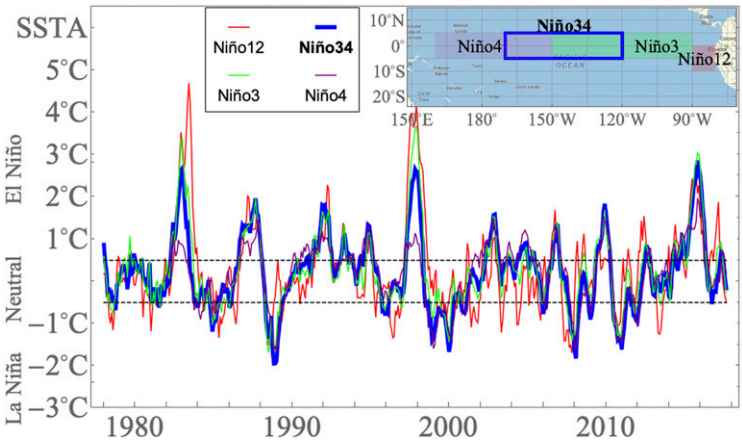
To examine the influence of ENSO on extended-range prediction skill, we first clustered the hindcast cases by the ENSO phase at the model's start time. Next, we computed and compared the prediction skills for the different clusters. Four clusters were adopted here, namely the ENSO active phase (El Niño/La Niña), El Niño phase, La Niña phase, and ENSO-neutral phase. Their union also was evaluated as a reference. Sample sizes are listed in Fig. 8. Using the Fisher r-to-z transformation (Fisher 1921), the correlation skills are transformed to the z statistics, which are then applied to assess the significance of the difference between two r skills.

We focus on the regional average predictions for week 2 and weeks 3–4. Figure 10 shows the week 2 and weeks 3–4 r skills conditioned on different ENSO phases. The NSE skill score and ROC score generally show similar results and are not shown here. For the week 2 forecast (column 1 of Fig. 10), in SCA, all 11 models show improved r skill during the El Niño phase compared to during the La Niña or ENSO-neutral phase, with 8 of them showing better r skill to a 90% significance level during El Niño compared to La Niña. Particularly, for the better-performing models (i.e., ECCC and ECMWF), r could differ by up to 0.2 comparing El Niño prediction and La Niña prediction. Correspondingly, for OR, most models show better r skill during La Niña phase compared to during El Niño phase, with seven of them showing better r skill to a 90% significance level during La Niña compared to El Niño. For ECCC and ECMWF, r could differ by up to 0.2 comparing La Niña prediction and El Niño prediction. For NCA and WA, the results for ENSO phase do not agree between the models, although for NCA the better-performing models generally have higher r skill scores during La Niña phase.

For the week 3–4 forecast (column 2 of Fig. 10), in SCA more of the models have improved performance during El Niño than during La Niña, with five of them showing better r skill to a 90% significance level during El Niño. For ECMWF, r is 0.42 for El Niño, whereas for La Niña r is 0.24. The advantage for ECMWF also occurs for NCA, with r being 0.46 during the El Niño phase. This indicates that the El Niño phase might allow better extended-range prediction for California. For OR and WA, most of the models show improved r skill during La Niña when compared to El Niño, with five of them showing better r skill to a 90% significance level during La Niña compared to El Niño.

b. *MJO*

The MJO is a traveling pattern characterized by a coherent eastward-propagating perturbation over the tropical Indian and Pacific Oceans (Madden and Julian 1972). Previous studies have found that MJO-related variability in the tropical Pacific convection modifies West Coast precipitation regimes through the propagation of extratropical wave trains (Mo and Higgins 1998; Higgins et al. 2000; Bond and Vecchi 2003).



Model	El Niño/La Niña	El Niño	La Niña	ENSO-Neutral	All
BoM	744	360	384	444	1188
CMA	2488	1065	1423	1334	3822
ECCC	497	192	305	283	780
ECMWF	1004	440	564	478	1482
HMCR	1088	550	538	569	1657
ISAC-CNR	699	363	336	381	1080
JMA	348	180	168	192	540
KMA	152	76	76	88	240
MétéoFrance	172	78	94	92	264
NCEP	1579	640	939	605	2184
UKMO	368	180	188	184	552

FIG. 8. (left) Definition and time series plot of ENSO. ENSO is quantified based on the SSTA of certain Pacific tropical regions, as delineated on the top right. The time series plot for Niño indexes from 1978 to 2016 is shown. To investigate ENSO impact on extended-range prediction skill, hindcasts for each model are clustered into different groups based on the ENSO phase at model start time. (right) Case counts for each cluster.

Provided with a realistic representation of the MJO and its teleconnections, this forecast opportunity could be realized as improvements in models' extended-range prediction skills. However, it is also noted that poor representation results in systematic worsened forecast during the MJO active periods, as compared to quiescent periods (Chen et al. 1993; Hendon et al. 2000).

The prediction skill for the MJO and its teleconnections have been improved significantly in recent decades, reaching a useful forecast range beyond 20 days (Kang and Kim 2010; Wang et al. 2014; Lim et al. 2018) and producing realistic teleconnections with the large-scale circulation (Vitart and Molteni 2010). Here, we make a dedicated examination of how MJO modifies precipitation distribution and extended-range prediction skills for the West Coast.

To quantify the MJO, The Real-Time Multivariate MJO Index is adopted. As shown in Fig. 11, the RMM index is composed of the two leading PCs of the field that combines average outgoing longwave radiation and zonal wind at 850 and 200 hPa from 15°S to 15°N (Wheeler and Hendon 2004). We define active MJO events using the criteria of Bond and Vecchi (2003):

1) there should be at least 30 days during which the amplitude ($\sqrt{PC_1^2 + PC_2^2}$) exceeds 0.5 [for Bond and Vecchi (2003), the threshold is 0.7 for U_{850hPa} field PCs], and
2) the MJO phase [$\tanh^{-1}(PC_1/PC_2)$] should move eastward for the entire period.

The detected active MJO events also are displayed in Fig. 11. Based on this classification, we explore the MJO's impact on precipitation distribution and extended-range prediction skill below.

1) INFLUENCE ON PRECIPITATION DISTRIBUTION

To investigate the MJO's influence on precipitation distribution, we follow the method in Mundhenk et al. (2018) to examine the average weekly precipitation anomalies conditioned on the MJO status. The method is described as follows. We first derived the weekly precipitation anomalies using same approach in section 5a(1). Next, for the early/late winter season and the whole winter season, we clustered the weekly precipitation anomalies based on the MJO status. Both the MJO phase and the lag days after the MJO phase are considered for constraining precipitation anomalies, since it takes time for the MJO-related variability to exert influence. Finally, we computed the mean value of the clusters and drew them in Fig. 12.

The most obvious pattern in Fig. 12 is the angled bands of precipitation anomalies, which generally stretch from top right to bottom left, following the MJO phase transitions. For instance, in late winter (JFM), abnormally high precipitation favors WA at week 2 following the onset of an active MJO event from phase 1. The enhanced and suppressed precipitation bands are separated more distinctly for NCA, OR, and WA as compared to SCA. A comparison between results for early winter season (column 1) and late winter season (column 2) shows that the phases of the MJO that promote enhanced or suppressed precipitation are substantially different during these two seasons. This is in agreement with the findings of Bond and Vecchi (2003).

Generally, results confirmed the impact of the MJO on modulating precipitation regimes for the West Coast, especially for NCA, OR, and WA. The time lag for MJO

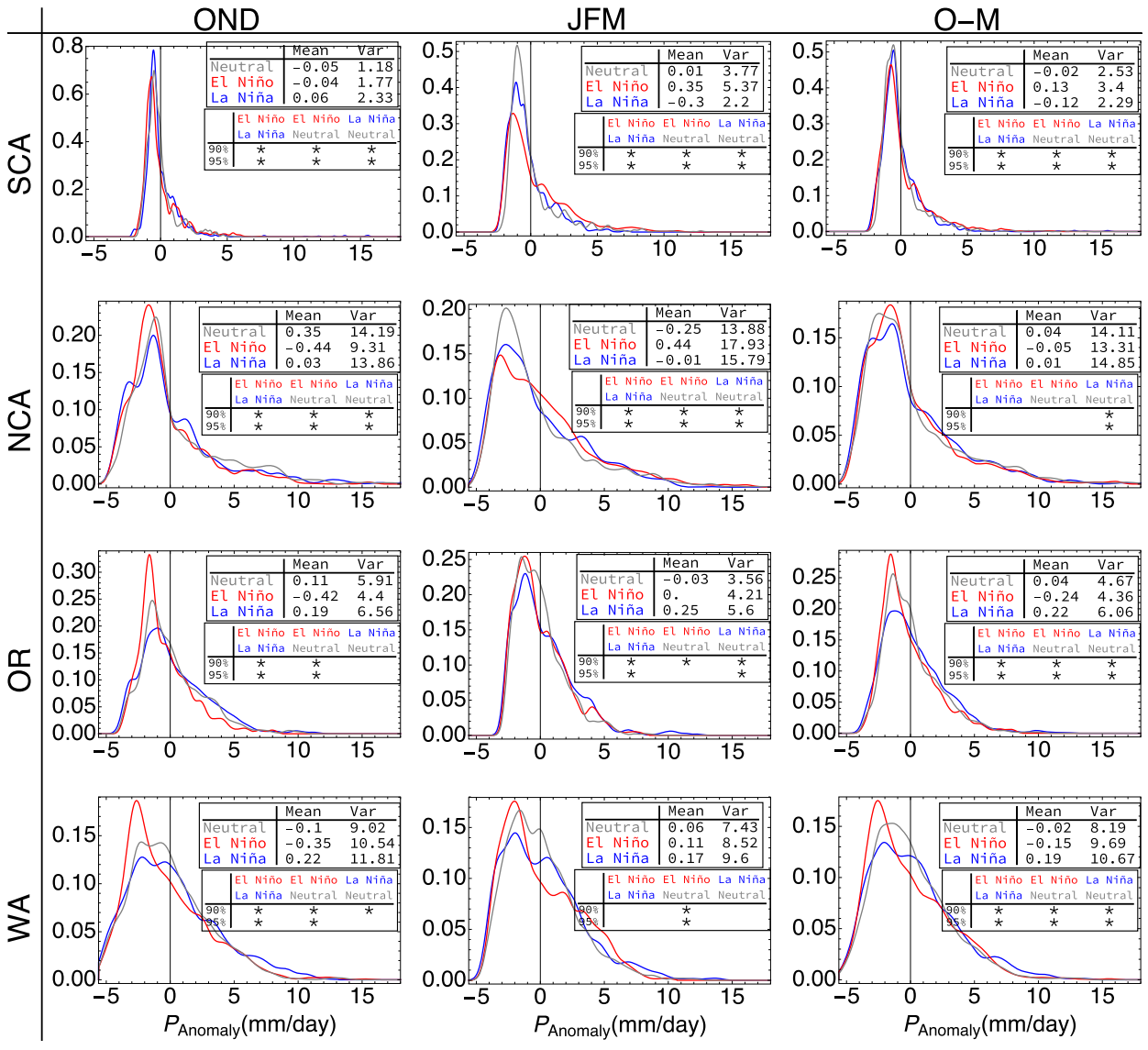


FIG. 9. Distribution of weekly precipitation anomalies conditioned on ENSO phases. (left) The early winter season (OND); (center) the late winter season (JFM); (right) the entire winter season (O–M). The rows represent results for different geographic divisions. For each panel, we listed the mean and variance of the distribution conditioned on ENSO phases. The comparison between two distributions is labeled with an asterisk if the Kolmogorov–Smirnov statistic lies out of the 90% or 95% confidence interval, indicating the two distributions are statistically significantly different.

to manifest its effects provides valuable potential for extending the range of skillful predictions.

2) Influence on precipitation prediction skills

To investigate the impact of the MJO on precipitation prediction skills, we first grouped the hindcast cases for each GCM according to their start-time MJO status. Five groups are adopted here following Jones et al. (2015): phases 1–2, phases 3–4, phases 5–6, and phases 7–8 for the active MJO period, as well as the MJO quiescent period. The sample size is listed in Fig. 11. Next, we evaluated the week 2 and week 3–4 prediction skill for each group. The statistical significance of r skill differences between MJO-active groups and MJO-quiescent group were determined using the z test. Results for r skills and significant tests are shown in Fig. 13.

While results confirmed the former findings that the prediction skill varies for different MJO groups and models (Matsueda and Endo 2011), there are some common patterns here. First, for hindcasts initialized during active MJO in phases 3–4, most models show

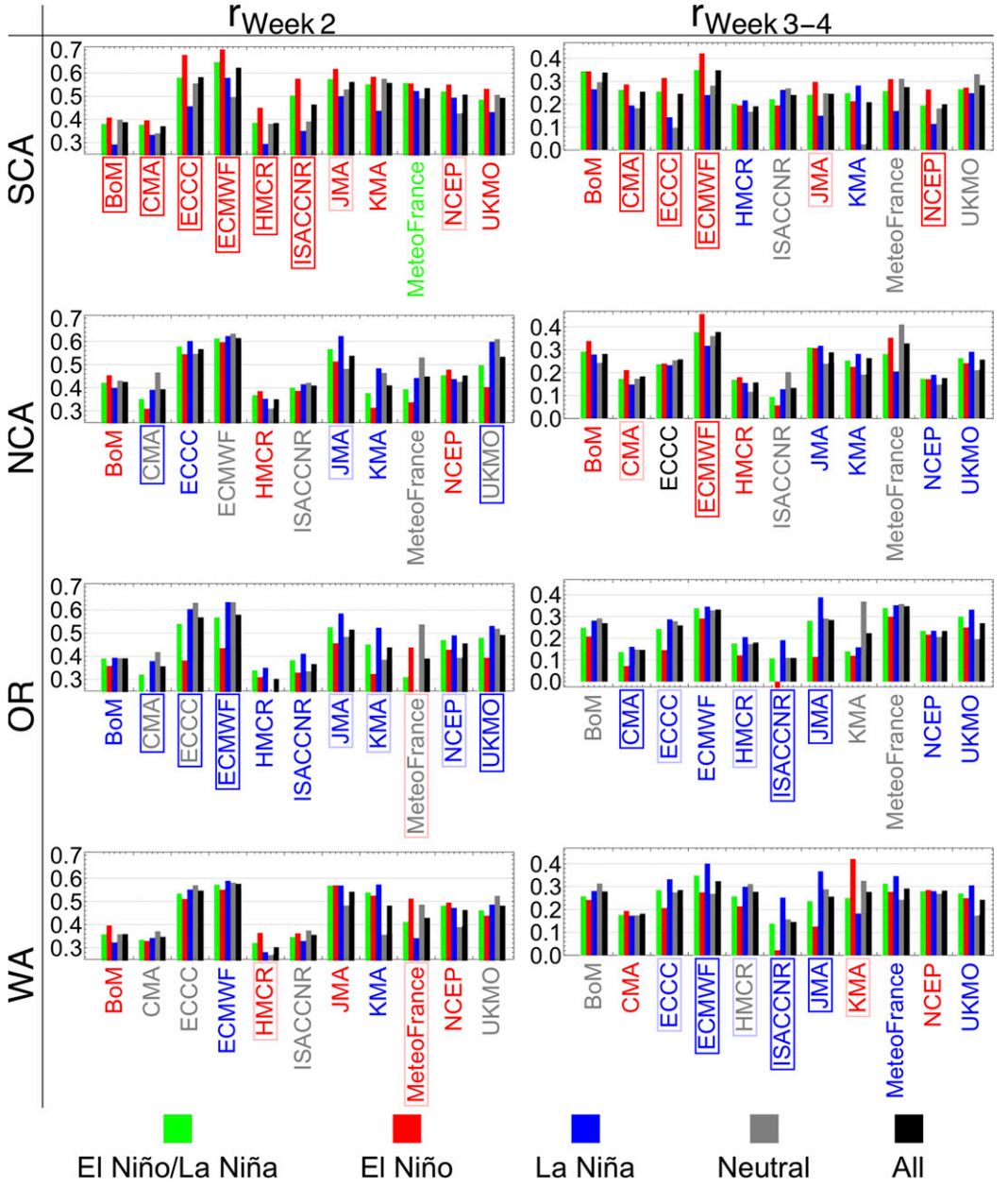
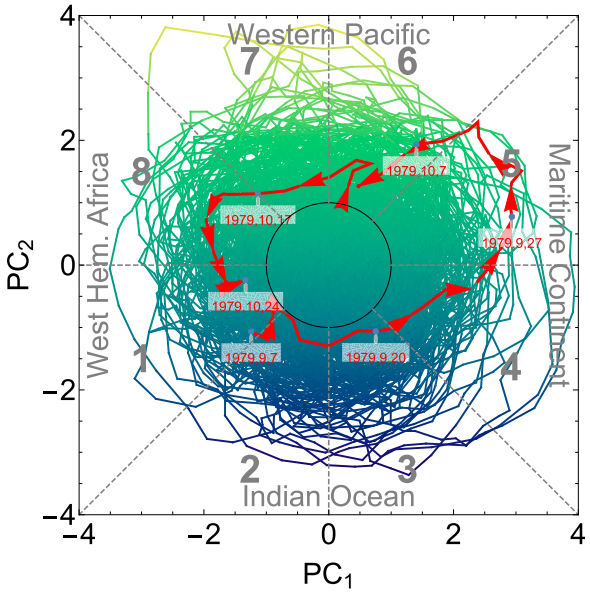


FIG. 10. (left) Week 2 and (right) week 3–4 precipitation prediction skills for different ENSO phases. The rows represent results for the four geographic divisions. For each subfigure, 11 models are evaluated; each model is colored depending on the phase in which model has highest score. Models with a significant r skill difference between El Niño and La Niña phases are framed: a red (blue) frame indicates that model shows significantly better r skill for El Niño (La Niña) phase based on the z test, and a light (dark) frame indicates the difference is statistically significant at the 90% (95%) confidence level.

lower extended-range prediction skills as compared to MJO-quiescent cases, except for the week 2 prediction in SCA. This skill drop was also found in Jones et al. (2015) when studying the impact of the MJO on intraseasonal predictability in the midlatitudes of the Northern Hemisphere. The skill drop might be attributed to the fact that many models cannot represent well the propagation of the MJO across the Maritime Continent (Lin et al. 2008; Vitart and Molteni 2010; Wang et al. 2014). If so, models cannot produce the MJO-associated extratropical response revealed in Fig. 12, resulting in systematic forecast biases. Second, for hindcasts initialized during active MJO in



Model	MJO Active				MJO Non Active	All
	Phase 1-2	Phase 3-4	Phase 5-6	Phase 7-8		
BoM	95	123	130	103	737	1188
CMA	252	387	424	286	2473	3822
ECCC	57	70	86	55	512	780
ECMWF	100	142	178	110	952	1482
HMCR	137	183	197	151	989	1657
ISACCNR	83	112	121	93	671	1080
JMA	43	53	63	45	336	540
KMA	19	20	22	19	160	240
MeteoFrance	24	24	29	21	166	264
NCEP	150	232	253	173	1376	2184
UKMO	37	52	67	36	360	552

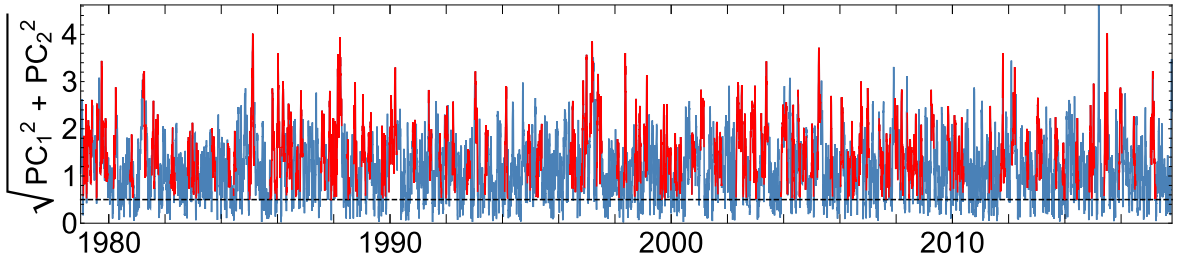


FIG. 11. (top left) The leading two PCs of the field that combines average outgoing longwave radiation and zonal wind at 850 and 200 hPa from 15°S to 15°N. The phase and amplitude of MJO are defined based on the position of (PC_1, PC_2) and its distance to the origin. For instance, the red arrowed line represents an MJO event that starts from 7 Sep 1979, goes counterclockwise (eastward when reprojected to geographic map), and ends on 24 Oct 1979. On most days, (PC_1, PC_2) lies out of the middle circle, whose radius is 1, indicating a strong MJO event. (bottom) The time series of MJO amplitude, as represented by $\sqrt{PC_1^2 + PC_2^2}$. Active MJO events are labeled with red lines. (top right) Hindcasts for 11 GCMs are labeled as "MJO Active" and grouped into corresponding clusters if the start time is within an active MJO period.

phases 1–2, phases 5–6, and phases 7–8, forecasts from the better-performing models (i.e., ECCC, ECMWF, Météo-France, and UKMO) are generally more skillful compared to MJO-quiescent cases. For instance, for SCA and NCA, the prediction skill during active MJO in phases 1–2 and phases 7–8 is generally higher than in MJO-quiescent days; for OR and WA, the prediction skill during active MJO in phases 5–6 and phases 7–8 is generally higher than in MJO-quiescent days.

6. Discussion and conclusions

We evaluated the precipitation prediction skills at the short to extended range for the West Coast, where precipitation variation significantly influences the local ecology and economy. The evaluation is based on the extended-range hindcast dataset of the WWRP–WCRP S2S project. For the 11 models used here, the hindcast sample size ranges from 240 to 3822 across more than 20 winters, covering various climate circumstances categorized by intraseasonal to seasonal variability. This guarantees that the evaluation is less prone to biases from limited sample size or model diversity.

Since there is inevitable deviation of prediction along forecast lead time, the evaluation is implemented at both stringent and loosened spatial temporal scales, measured by both deterministic scores (r and NSE) and probabilistic scores (ROC and $\overline{\text{CRPS}}$). We further examined the impact of extended-range predictability sources, focusing on ENSO and the MJO. Our key findings are listed as follows:

1) We investigated the S2S models' prediction skill–forecast lead time relationship for the four divisions in the West Coast.
   - For week 1, the S2S models show advantageous precipitation prediction skills. The r, NSE, and ROC scores are approximately of the order of 0.8, 0.7, and

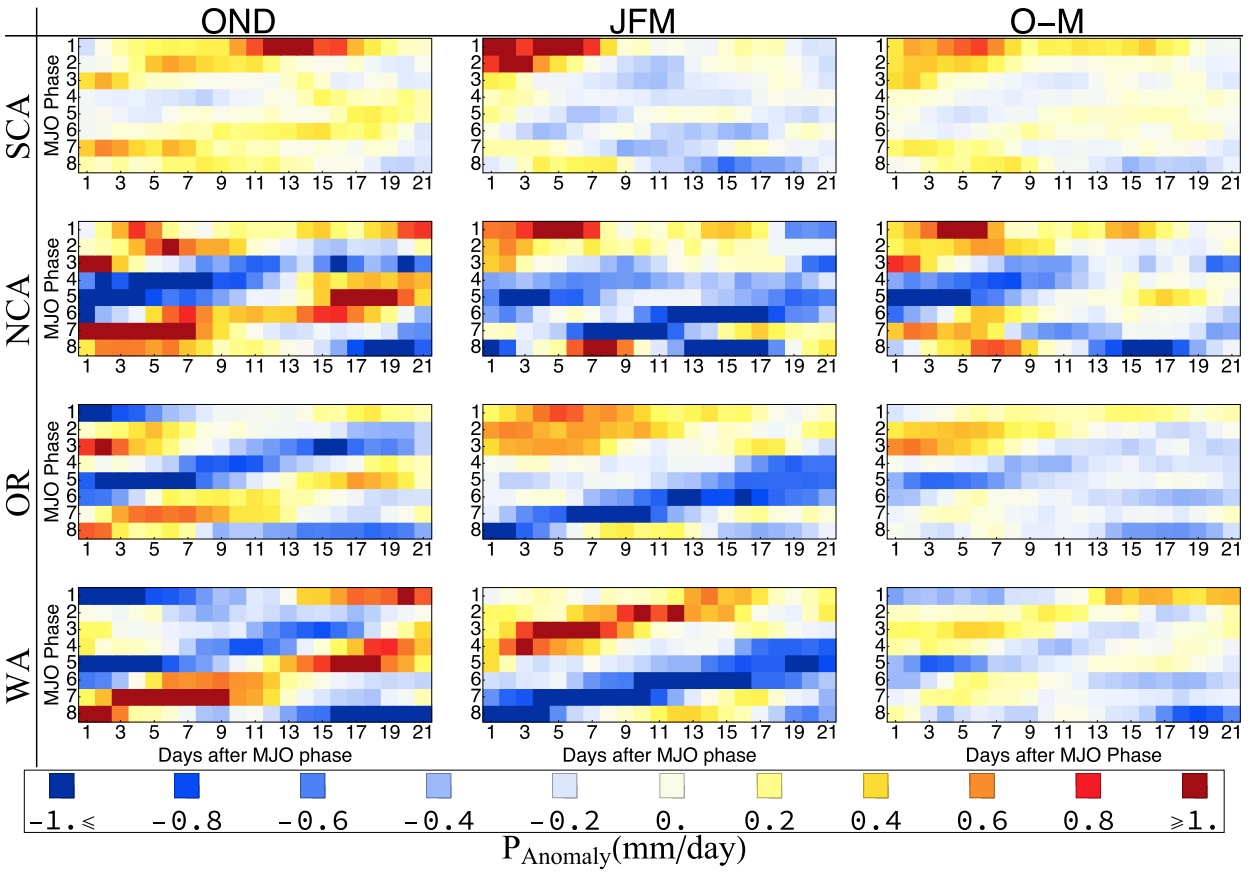


FIG. 12. Mean value of weekly precipitation anomalies conditioned on active MJO events' phase and number of days after MJO phase. The four rows represent results for four geographic divisions, and the columns represent results for different seasons. In each subfigure, the grid color represents the mean weekly precipitation anomalies for day m after MJO phase n; here m ranges from day 1 to day 21, as labeled on the x axis, and n ranges from phase 1 to phase 8, as labeled on the y axis.

0.8, respectively, for this period. By spatial averaging, the skill score can be further improved.

- For week 2, models show large variations regarding their performances. The week-2 mean precipitation forecast from the best-performing model (i.e., ECMWF) is of considerable value, with $r > 0.6$, NSE > 0.35, and ROC score > 0.7.
- Beyond week 2, predictions generally provide little deterministic skill. For this range period, probabilistic evaluation of ensemble forecasts using the $\overline{\text{CRPS}}$ shows the significant advantage of ensemble forecasts over deterministic forecast.
- Considering the performance difference of the S2S models, the informative predictable range may differ by up to 6–7 days comparing different models. For the short range, models with higher resolution tend to have better performances (JMA, KMA, ECCC, and ECMWF). For the medium to extended range, ensemble mean predictions show significant better performance compared to deterministic predictions. The best-performing models for this range period are the ECCC, ECMWF, and JMA models. For the week 3–4 forecast, although there is essentially no useful deterministic forecast skill, the ECMWF model still shows an advantage over the rest of the models. Results here can benefit model selections for practical forecasts and multimodel ensemble predictions.

2) Through investigating the impact of ENSO on the West Coast precipitation distribution and models' prediction skill, we found a spatial seesaw effect for ENSO to modulate precipitation distribution and GCMs' prediction skills:

- During El Niño years, Southern California tends to receive more precipitation in late winter, as compared to La Niña years. Also, we found that most models show significantly better extended-range prediction skills in El Niño years, as compared to La Niña years.

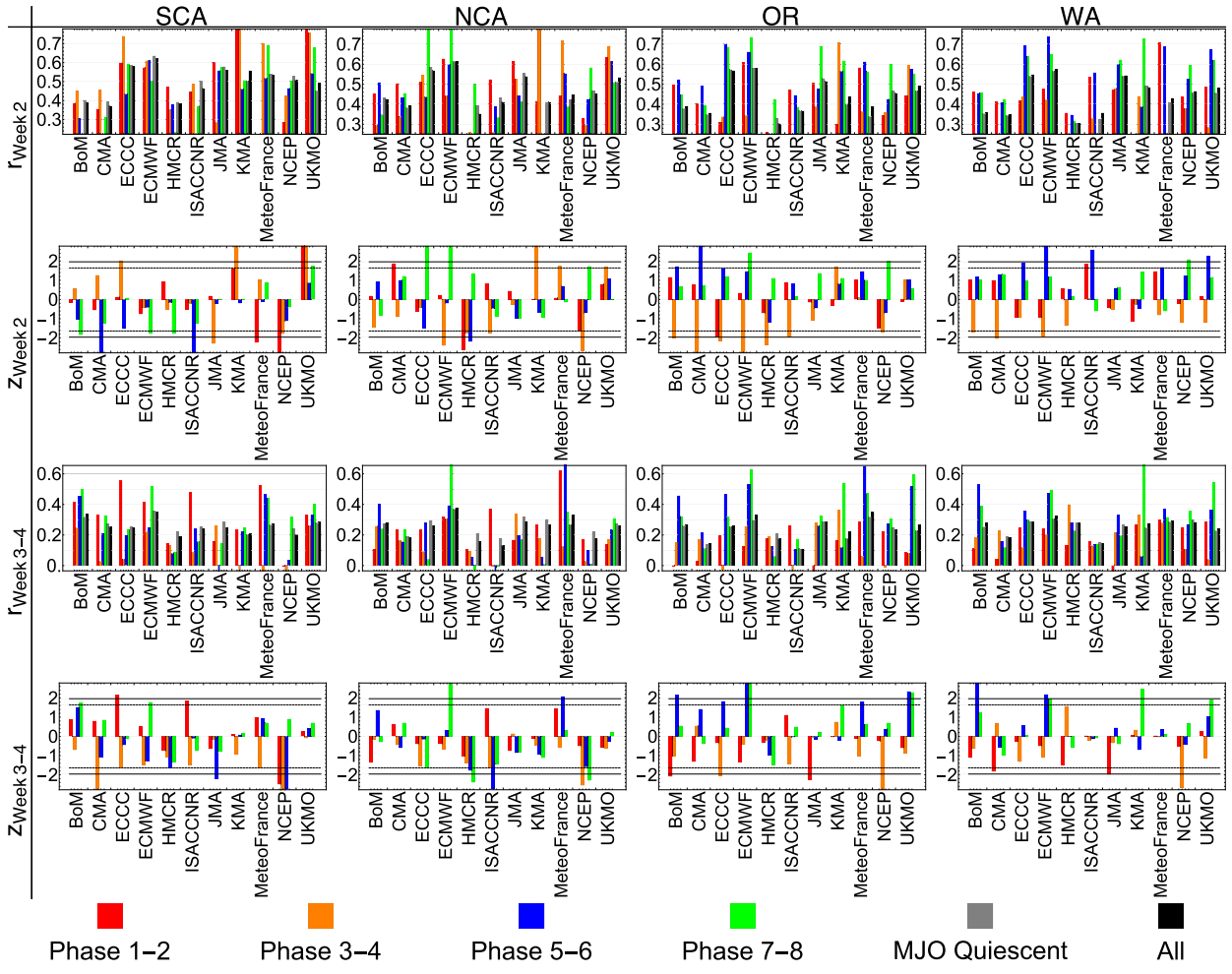


FIG. 13. (first row) Week 2 and (third row) week 3–4 *r* skills for different MJO groups. (second row) Week 2 and (fourth row) week 3–4 *z* statistics of differences between *r* skills for MJO-active groups and MJO-quiescent group, and the dashed (solid) grid lines indicate statistical difference at the 90% (95%) confidence level.

- During La Niña years, Oregon tends to receive more precipitation in the winter season, as compared to El Niño years. Also, we found that most models show significantly better extended-range prediction skills in La Niña years, as compared to El Niño years.

For Northern California and Washington, ENSO influences the precipitation distribution, but specific models may either have higher or lower prediction skills depending on the ENSO phases. We assume the excessive precipitation and improved extended-range prediction skills accompany the meridional shift of baroclinic systems as modulated by ENSO. This predictability difference related to ENSO phases will be useful for extended-range prediction applications.

3) We investigated the impact of the MJO on the West Coast precipitation distribution and GCMs' prediction skills.

- Regarding the impact of MJO on precipitation distributions, we examined the average precipitation anomalies conditioned on the MJO phases and days after MJO phases. Results show a systematic below/above climatology precipitation pattern following certain MJO phases at corresponding lead time. The time lag for MJO to manifest its effects provides valuable potential for skillful predictions at the extended range.
- Regarding the impact of the MJO on GCMs' extended-range precipitation prediction skills, we verified that for certain MJO phases (especially, phases 5–6 and 7–8), some S2S models can capture the MJO-associated teleconnection in improving week 3–4 prediction skills. However, for hindcasts initialized during active MJO in phases 3–4, many models show lower extended-range prediction skills

as compared to MJO-quiescent cases, suggesting that the forecast opportunity may also be a curse if models have deficiencies in capturing its influences.

Results here suggest the potential for predictability across a range of time scales (Hoskins 2013; Zhu et al. 2014). We hope the baseline provided here can foster practical subseasonal prediction applications and facilitate further research on improving midlatitude subseasonal precipitation forecasts.

Acknowledgments. This research is supported by National Aeronautics and Space Administration (NASA) Grant NNX16AO56G, and NOAA Modeling, Analysis, Predictions, and Projections (MAPP) Program Grant NA14OAR4310222. The CPC Unified Gauge-Based Analysis of Daily Precipitation database is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at https://www.esrl.noaa.gov/psd/. We appreciate the comments and suggestions from Professor Fuqing Zhang from Pennsylvania State University. The two anonymous reviewers offered comprehensive and valuable comments in improving the paper.

REFERENCES

Alves, O., and Coauthors, 2003: POAMA: Bureau of Meteorology operational coupled model seasonal forecast system. *Proc. National Drought Forum*, Brisbane, Queensland, Australia, 49–56.

Baggett, C. F., E. A. Barnes, E. D. Maloney, and B. D. Mundhenk, 2017: Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophys. Res. Lett.*, **44**, 7528–7536, https://doi.org/10.1002/2017GL074434.

Bao, J., S. Michelson, P. Neiman, F. Ralph, and J. Wilczak, 2006: Interpretation of enhanced integrated water vapor bands associated with extratropical cyclones: Their formation and connection to tropical moisture. *Mon. Wea. Rev.*, **134**, 1063–1080, https://doi.org/10.1175/MWR3123.1.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Best, M., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, https://doi.org/10.5194/gmd-4-677-2011.

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172, https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2.

Bond, N. A., and G. A. Vecchi, 2003: The influence of the Madden–Julian oscillation on precipitation in Oregon and Washington. *Wea. Forecasting*, **18**, 600–613, https://doi.org/10.1175/1520-0434(2003)018<0600:TIOTMO>2.0.CO;2.

Chen, S.-C., J. O. Roads, and J. C. Alpert, 1993: Variability and predictability in an empirically forced global model. *J. Atmos. Sci.*, **50**, 443–463, https://doi.org/10.1175/1520-0469(1993)050<0443:VAPIAE>2.0.CO;2.

Courtier, P., and J.-F. Geleyn, 1988: A global numerical weather prediction model with variable resolution: Application to the shallow-water equations. *Quart. J. Roy. Meteor. Soc.*, **114**, 1321–1346, https://doi.org/10.1002/qj.49711448309.

Dacre, H. F., P. A. Clark, O. Martinez-Alvarado, M. A. Stringer, and D. A. Lavers, 2015: How do atmospheric rivers form? *Bull. Amer. Meteor. Soc.*, **96**, 1243–1255, https://doi.org/10.1175/BAMS-D-14-00031.1.

Dettinger, M., 2011: Climate change, atmospheric rivers, and floods in California—A multimodel analysis of storm frequency and magnitude changes. *J. Amer. Water Resour. Assoc.*, **47**, 514–523, https://doi.org/10.1111/j.1752-1688.2011.00546.x.

Fawcett, T., 2004: ROC graphs: Notes and practical considerations for researchers. *Mach. Learn.*, **31**, 1–38.

Fisher, R. A., 1921: On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.

Gagnon, N., and Coauthors, 2013: Improvements to the Global Ensemble Prediction System (GEPS) from version 2.0.3 to version 3.0.0. Tech. Rep., Meteorological Research Division, Canadian Meteorological Center, Environment Canada, 22 pp.

Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler, 2002: The National Elevation Dataset. *Photogramm. Eng. Remote Sensing*, **68** (1), 5–11.

Hendon, H. H., B. Liebmann, M. Newman, J. D. Glick, and J. Schemm, 2000: Medium-range forecast errors associated with active episodes of the Madden–Julian oscillation. *Mon. Wea. Rev.*, **128**, 69–86, https://doi.org/10.1175/1520-0493(2000)128<0069:MRFEAW>2.0.CO;2.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Higgins, R., J. E. Schemm, W. Shi, and A. Leetmaa, 2000: Extreme precipitation events in the western United States related to tropical forcing. *J. Climate*, **13**, 793–820, https://doi.org/10.1175/1520-0442(2000)013<0793:EPEITW>2.0.CO;2.

Hoskins, B. J., 2013: The potential for skill across the range of the seamless weather–climate prediction problem: A stimulus for our science. *Quart. J. Roy. Meteor. Soc.*, **139**, 573–584, https://doi.org/10.1002/qj.1991.

——, and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2.

JMA, 2013: Outline of the operational numerical weather prediction at the Japan Meteorological Agency. Japan Meteorological Agency, accessed 1 June 2018, https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2013-nwp/index.htm.

Jones, C., A. Hazra, and L. M. Carvalho, 2015: The Madden–Julian oscillation and boreal winter forecast skill: An analysis of NCEP CFSv2 reforecasts. *J. Climate*, **28**, 6297–6307, https://doi.org/10.1175/JCLI-D-15-0149.1.

Jong, B.-T., M. Ting, and R. Seager, 2016: El Niño's impact on California precipitation: Seasonality, regionality, and El Niño intensity. *Environ. Res. Lett.*, **11**, 054021, https://doi.org/10.1088/1748-9326/11/5/054021.

Kang, I.-S., and H.-M. Kim, 2010: Assessment of MJO predictability for boreal winter with various statistical and dynamical models. *J. Climate*, **23**, 2368–2378, https://doi.org/10.1175/2010JCLI3288.1.

Kumar, A., M. Chen, and W. Wang, 2011: An analysis of prediction skill of monthly mean climate variability. *Climate Dyn.*, **37**, 1119–1131, https://doi.org/10.1007/s00382-010-0901-4.

Li, S., and A. W. Robertson, 2015: Evaluation of submonthly precipitation forecast skill from global ensemble prediction

systems. *Mon. Wea. Rev.*, **143**, 2871–2889, https://doi.org/10.1175/MWR-D-14-00277.1.

Lim, Y., S.-W. Son, and D. Kim, 2018: MJO prediction skill of the subseasonal-to-seasonal prediction models. *J. Climate*, **31**, 4075–4094, https://doi.org/10.1175/JCLI-D-17-0545.1.

Lin, H., G. Brunet, and J. Derome, 2008: Forecast skill of the Madden–Julian oscillation in two Canadian atmospheric models. *Mon. Wea. Rev.*, **136**, 4130–4149, https://doi.org/10.1175/2008MWR2459.1.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17A**, 321–333, https://doi.org/10.3402/tellusa.v17i3.9076.

Madadgar, S., A. AghaKouchak, S. Shukla, A. W. Wood, L. Cheng, K.-L. Hsu, and M. Svoboda, 2016: A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States. *Water Resour. Res.*, **52**, 5095–5110, https://doi.org/10.1002/2015WR018547.

Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29**, 1109–1123, https://doi.org/10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2.

Malguzzi, P., A. Buzzi, and O. Drofa, 2011: The meteorological global model GLOBO at the ISAC-CNR of Italy assessment of 1.5 yr of experimental use for medium-range weather forecasts. *Wea. Forecasting*, **26**, 1045–1055, https://doi.org/10.1175/WAF-D-11-00027.1.

Mamalakis, A., J.-Y. Yu, J. T. Randerson, A. AghaKouchak, and E. Foufoula-Georgiou, 2018: A new interhemispheric teleconnection increases predictability of winter precipitation in southwestern US. *Nat. Commun.*, **9**, 2332, https://doi.org/10.1038/s41467-018-04722-7.

Matsueda, M., and H. Endo, 2011: Verification of medium-range MJO forecasts with TIGGE. *Geophys. Res. Lett.*, **38**, L11801, https://doi.org/10.1029/2011GL047480.

Matthews, A. J., 2004: Atmospheric response to observed intraseasonal tropical sea surface temperature anomalies. *Geophys. Res. Lett.*, **31**, L14107, https://doi.org/10.1029/2004GL020474.

Mo, K. C., and R. Higgins, 1998: Tropical convection and precipitation regimes in the western United States. *J. Climate*, **11**, 2404–2423, https://doi.org/10.1175/1520-0442(1998)011<2404:TCAPRI>2.0.CO;2.

Mundhenk, B. D., E. A. Barnes, E. D. Maloney, and C. F. Baggett, 2018: Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian oscillation and quasi-biennial oscillation. *npj Climate Atmos. Sci.*, **1**, 20177, https://doi.org/10.1038/s41612-017-0008-2.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, **10**, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6.

National Academies of Sciences, Engineering, and Medicine, 2016: *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*. National Academies Press, 350 pp.

Neena, J., 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, https://doi.org/10.1175/JCLI-D-13-00624.1.

OECD, 2018: Population. Organisation for Economic Co-operation and Development, accessed 1 May 2016, https://data.oecd.org/pop/population.htm.

Rasmusson, E. M., and J. M. Wallace, 1983: Meteorological aspects of the El Niño–Southern Oscillation. *Science*, **222**, 1195–1202, https://doi.org/10.1126/science.222.4629.1195.

Robertson, A. W., A. Kumar, M. Peña, and F. Vitart, 2015: Improving and promoting subseasonal to seasonal prediction. *Bull. Amer. Meteor. Soc.*, **96**, ES49–ES53, https://doi.org/10.1175/BAMS-D-14-00139.1.

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.

Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.*, **45**, 1228–1251, https://doi.org/10.1175/1520-0469(1988)045<1228:TGOGRF>2.0.CO;2.

Schonher, T., and S. Nicholson, 1989: The relationship between California rainfall and ENSO events. *J. Climate*, **2**, 1258–1269, https://doi.org/10.1175/1520-0442(1989)002<1258:TRBCRA>2.0.CO;2.

Seager, R., M. Hoerling, S. Schubert, H. Wang, B. Lyon, A. Kumar, J. Nakamura, and N. Henderson, 2015: Causes of the 2011–14 California drought. *J. Climate*, **28**, 6997–7024, https://doi.org/10.1175/JCLI-D-14-00860.1.

Tian, D., E. F. Wood, and X. Yuan, 2017: CFSv2-based subseasonal precipitation and temperature forecast skill over the contiguous United States. *Hydrol. Earth Syst. Sci.*, **21**, 1477–1490, https://doi.org/10.5194/hess-21-1477-2017.

Trenberth, K. E., and D. P. Stepaniak, 2001: Indices of El Niño evolution. *J. Climate*, **14**, 1697–1701, https://doi.org/10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2.

——, G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14 291–14 324, https://doi.org/10.1029/97JC01444.

Uría-Martínez, R., M. Johnson, and P. O'Connor, 2017: U.S. hydropower market report 2017 update (April). Oak Ridge National Laboratory, 12 pp.

Vilsack, T., and J. T. Reilly, 2013: Census of Agriculture: 2013 Farm and Ranch Irrigation Survey. United States Department of Agriculture, National Agricultural Statistics Survey, https://www.nass.usda.gov/Publications/AgCensus/2012/Online_Resources/Farm_and_Ranch_Irrigation_Survey/.

Vitart, F., 2004: Monthly forecasting at ECMWF. *Mon. Wea. Rev.*, **132**, 2761–2779, https://doi.org/10.1175/MWR2826.1.

——, 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, https://doi.org/10.1002/qj.2256.

——, and F. Molteni, 2010: Simulation of the Madden–Julian oscillation and its teleconnections in the ECMWF forecast system. *Quart. J. Roy. Meteor. Soc.*, **136**, 842–855, https://doi.org/10.1002/qj.623.

——, and Coauthors, 2008: The new VAREPS-monthly forecasting system: A first step towards seamless prediction. *Quart. J. Roy. Meteor. Soc.*, **134**, 1789–1799, https://doi.org/10.1002/qj.322.

——, 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.

Vogel, P., P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting, 2018: Skill of global raw and postprocessed ensemble predictions of rainfall over northern tropical Africa. *Wea. Forecasting*, **33**, 369–388, https://doi.org/10.1175/WAF-D-17-0127.1.

Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dyn.*, **40**, 2091–2121, https://doi.org/10.1007/s00382-011-1259-y.

Wang, B., and Coauthors, 2009: Advance and prospectus of seasonal prediction: Assessment of the APCC/CliPAS 14-model

ensemble retrospective seasonal prediction (1980–2004). *Climate Dyn.*, **33**, 93–117, https://doi.org/10.1007/s00382-008-0460-0.

Wang, W., M.-P. Hung, S. J. Weaver, A. Kumar, and X. Fu, 2014: MJO prediction in the NCEP Climate Forecast System version 2. *Climate Dyn.*, **42**, 2509–2520, https://doi.org/10.1007/s00382-013-1806-9.

Weber, N., 2015: Subseasonal prediction over the western United States. 2015 Fall Meeting, San Francisco, CA, Amer. Geophys. Union, Abstract A33M-0393.

——, and C. F. Mass, 2017: Evaluating CFSv2 subseasonal forecast skill with an emphasis on tropical convection. *Mon. Wea. Rev.*, **145**, 3795–3815, https://doi.org/10.1175/MWR-D-17-0109.1.

Welch, P., 1967: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.*, **15**, 70–73, https://doi.org/10.1109/TAU.1967.1161901.

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.

——, H. Zhu, A. H. Sobel, D. Hudson, and F. Vitart, 2017: Seamless precipitation prediction skill comparison between two global models. *Quart. J. Roy. Meteor. Soc.*, **143**, 374–383, https://doi.org/10.1002/qj.2928.

Whitaker, J. S., and K. M. Weickmann, 2001: Subseasonal variations of tropical convection and week-2 prediction of wintertime western North American rainfall. *J. Climate*, **14**, 3279–3288, https://doi.org/10.1175/1520-0442(2001)014<3279:SVOTCA>2.0.CO;2.

White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, https://doi.org/10.1002/met.1654.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* Vol. 91, International Geophysics Series, Academic Press, 346 pp.

Wood, N., and Coauthors, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quart. J. Roy. Meteor. Soc.*, **140**, 1505–1520, https://doi.org/10.1002/qj.2235.

Wu, T., and Coauthors, 2014: An overview of BCC climate system model development and application for climate change studies. *J. Meteor. Res.*, **28**, 34–56, https://doi.org/10.1007/s13351-014-3041-7.

Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeor.*, **8**, 607–626, https://doi.org/10.1175/JHM583.1.

Yarnal, B., and H. F. Diaz, 1986: Relationships between extremes of the Southern Oscillation and the winter climate of the Anglo-American Pacific coast. *Int. J. Climatol.*, **6**, 197–219, https://doi.org/10.1002/joc.3370060208.

Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson, 2014: Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Wea. Rev.*, **142**, 1556–1569, https://doi.org/10.1175/MWR-D-13-00222.1.