

Predictability and Prediction Skill of the MJO in Two Operational Forecasting Systems

HYE-MI KIM

School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York

PETER J. WEBSTER AND VIOLETA E. TOMA

School of Earth and Atmospheric Science, Georgia Institute of Technology, Atlanta, Georgia

DAEHYUN KIM

Department of Atmospheric Science, University of Washington, Seattle, Washington

(Manuscript received 12 August 2013, in final form 4 April 2014)

ABSTRACT

The authors examine the predictability and prediction skill of the Madden–Julian oscillation (MJO) of two ocean–atmosphere coupled forecast systems of ECMWF [Variable Resolution Ensemble Prediction System (VarEPS)] and NCEP [Climate Forecast System, version 2 (CFSv2)]. The VarEPS hindcasts possess five ensemble members for the period 1993–2009 and the CFSv2 hindcasts possess three ensemble members for the period 2000–09. Predictability and prediction skill are estimated by the bivariate correlation coefficient between the observed and predicted Wheeler–Hendon real-time multivariate MJO index (RMM). MJO predictability is beyond 32 days lead time in both hindcasts, while the prediction skill is about 27 days in VarEPS and 21 days in CFSv2 as measured by the bivariate correlation exceeding 0.5. Both predictability and prediction skill of MJO are enhanced by averaging ensembles. Results show clearly that forecasts initialized with (or targeting) strong MJOs possess greater prediction skill compared to those initialized with (or targeting) weak or nonexistent MJOs. The predictability is insensitive to the initial MJO phase (or forecast target phase), although the prediction skill varies with MJO phases.

A few common model issues are identified. In both hindcasts, the MJO propagation speed is slower and the MJO amplitude is weaker than observed. Also, both ensemble forecast systems are underdispersive, meaning that the growth rate of ensemble error is greater than the growth rate of the ensemble spread by lead time.

1. Introduction

The Madden–Julian oscillation (MJO; [Madden and Julian 1971, 1972](#)) is a dominant mode of subseasonal variability in the tropical atmosphere and ocean that interacts with a wide range of weather and climate phenomena across the planet ([Han et al. 2001](#); [Zhang 2005](#); [Lau and Waliser 2011](#), and many others). Because of its impacts on the other components of the climate system, the MJO has been considered as a major potential source of the global climate predictability on subseasonal time scales. Benefiting from the significant

improvement in the representation of the MJO in numerical models that has been made in the past decades, contemporary operational dynamical prediction systems produce useful forecast of the MJO up to 20–25 days of forecast lead time ([Vitart and Molteni 2010](#); [Vitart et al. 2010](#); [Rashid et al. 2011](#); [Zhang and van den Dool 2012](#); [Zhang et al. 2013](#); [Vitart 2014](#); [Wang et al. 2014](#)). This is encouraging, but the prediction skill is lower than the theoretical estimates of the predictability, up to 25–40 days ([Waliser et al. 2003](#); [Reichler and Roads 2005](#); [Kim and Kang 2008](#)). Therefore, in order to guide further improvement of MJO prediction and to pinpoint particular weaknesses of dynamical prediction systems, it is crucial to understand the MJO predictability and to address the current status of the MJO prediction skill.

Previous studies have shown that the prediction skill depends strongly on amplitude and phase of the MJO, as

Corresponding author address: Hye-Mi Kim, School of Marine and Atmospheric Sciences, 119 Endeavour Hall, Stony Brook University, Stony Brook, NY 11794.
E-mail: hyemi.kim@stonybrook.edu

well as on the season in which the MJO event is occurring. The MJO prediction skill is distinctly better when the MJO is strong at the beginning of the forecast, irrespective of the phase, compared to those that are weak (Lin et al. 2008; Agudelo et al. 2009; Kang and Kim 2010; Rashid et al. 2011; Zhang and van den Dool 2012; Wang et al. 2014). Also, the MJO is more predictable in the boreal winter than summer, due to a more well-defined MJO with stronger amplitude and dominant eastward propagation (Lin et al. 2008; Agudelo et al. 2009; Rashid et al. 2011; Zhang and van den Dool 2012; Wang et al. 2014). Many operational dynamical predictions show difficulty in MJO propagation, especially when the MJO propagates over the Maritime Continent, known as the MJO “Maritime Continent prediction barrier” problem (Vitart et al. 2007; Lin et al. 2008; Seo et al. 2009; Vitart and Molteni 2010; Fu et al. 2011; Weaver et al. 2011; Zhang and van den Dool 2012; Wang et al. 2014). Also, the slow MJO propagation speed and the fast decrease of the predicted MJO amplitude limit the prediction skill as well, although this is model dependent (Vitart et al. 2007; Agudelo et al. 2009; Vitart and Molteni 2010; Matsueda and Endo 2011; Rashid et al. 2011; Wang et al. 2014). Another common weakness in operational ensemble MJO prediction is the underdispersive ensemble spread such that the ensemble spread in MJO prediction does not encompass the forecast error of ensemble mean (Rashid et al. 2011; Hudson et al. 2013).

Predictability and prediction skill of the MJO in dynamical models have been investigated individually in depth with various methodologies. However, progress in identifying and comparing MJO predictability, prediction skill, and ensemble dispersion in various operational models has been slow, perhaps in part because of the expense of integrating dynamical model simulation with a large number of different initial conditions. It has thus remained as an unexplored frontier. Operational forecast centers are now incorporating improved physics, ensemble generation methods, optimal initialization, and increased resolution to their coupled prediction systems. Therefore, a continuous systematic assessment of MJO predictability and prediction skill, as well as understanding the source and error of MJO prediction in current operational forecast system, is crucial to bridge the gap of skill between the weather forecasts and seasonal prediction.

In this study, we assess the MJO predictability and prediction skill in two current operational forecast systems that have been used for real-time subseasonal forecasts: the European Centre for Medium-Range Weather Forecasts (ECMWF) monthly forecasting system (Vitart 2014) and National Centers for Environmental Prediction (NCEP) Climate Forecasting System, version 2 (CFSv2; Saha et al. 2014). Vitart (2014) and

Wang et al. (2014) examined MJO prediction in each ECMWF and NCEP forecasting system, respectively. Vitart (2014) found that the MJO prediction skill has gradually improved in ECMWF monthly forecasting system since 2002, with an average gain of about 1 day of prediction skill per year until 2011. Also, the MJO amplitude has become more realistic although still weaker than observed (Vitart 2014). The NCEP CFSv2 also represents a substantial change from its previous version in all aspects of the forecast system including model components, the data assimilation system, and the ensemble configuration. These changes have led to improved MJO prediction. Wang et al. (2014) found that the CFSv2 has useful MJO prediction skill out to 20 days although the MJO amplitude decreases faster and the propagation speed is slower than observed.

In this study, we extend Wang et al. (2014) and Vitart (2014) analyses to also explore the prediction skill. In addition, we will compare the predictability and ensemble dispersion of the two operational forecasting systems systematically using large sets of ensemble reforecasts. Comparing the state-of-the-art operational models will lead to a better understanding of common model problems in MJO prediction. Section 2 introduces details of the hindcasts, reanalysis data, and verification methods. Section 3 examines the MJO predictability, prediction skill, and characteristics of ensemble dispersion. Model errors in MJO propagation and amplitude change will be investigated in section 4. Predictability and prediction skill for forecasts targeting MJO events are discussed in section 5 and results are summarized in section 6.

2. Data and verification methodology

The NCEP CFSv2 and the ECMWF monthly forecasting system are fully coupled model systems. Accompanying both systems are large sets of reforecasts (hindcasts) generated with the purpose of evaluating and calibrating the model simulations. CFSv2 hindcasts consist of fully coupled components of the ocean, atmosphere, and land (Saha et al. 2014). The oceanic component is the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 4 (MOM4) and the atmospheric component is the NCEP Global Forecast System (GFS) with a horizontal resolution of T126 spectral truncation and 64 vertical levels extending to 0.26 hPa (Saha et al. 2014). We analyze the 45-day hindcast runs commenced every 0600, 1200, and 1800 UTC (here we use three ensembles), 365 days a year over the 10-yr period from 2000 to 2009. Initial conditions come from NCEP CFS Reanalysis (Saha et al. 2010).

ECMWF combined its monthly forecasting system and the Variable Resolution Ensemble Prediction

System into a single system (VarEPS; <http://old.ecmwf.int/products/changes/vareps-monthly/>). The ECMWF VarEPS is projected to 32-day horizons every Monday and Thursday with the first 10 days at 30-km horizontal resolution (vertical resolution of 62 levels extending to 5 hPa) for the atmospheric model and forced by a persisted sea surface temperature (SST). Starting at day 11, the atmospheric model horizontal resolution changes to 60 km and is coupled to the ocean. In addition, starting on the same day and month as Thursday's monthly real-time forecast, a set of hindcasts including a 5-member ensemble of 32-day integrations for each of the past 18 years is generated. Reforecasts have been initialized from the ECMWF Interim Re-Analysis (ERA-Interim; Dee et al. 2011). More details can be found in Vitart (2014). For the present study, the VarEPS hindcasts for years 2010 and 2011 are used. For year 2010, the first instance selected is the hindcast associated with the real-time forecast starting on 7 January 2010. The hindcast is a five-member ensemble integrated with 18 different starting dates from 1992 until 2009. Similarly, for year 2011, the first set selected is the real-time forecast starting 6 January 2011. We select the 1993–2009 period to ensure two sets of five-member ensembles each week, resulting in a total of 52 weeks \times 2 sets of hindcasts, thus 108 cases per year.

In total, we have 9180 (108 cases $\text{yr}^{-1} \times 5$ ensemble members \times 17 yr) sets of 32-day integrations for VarEPS, 10950 (365 cases $\text{yr}^{-1} \times 3$ ensemble members \times 10 yr) sets of 45-day integrations for CFSv2, and 10950 (365 cases $\text{yr}^{-1} \times 30$ yr, 1981–2010) days of observations. Daily mean fields of outgoing longwave radiation (OLR) and zonal winds at 200 (U200) and 850 hPa (U850) are extracted from the hindcasts to obtain the predicted MJO index defined below. The ERA-Interim products and OLR from the National Oceanic and

Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR; Liebmann and Smith 1996) are used to create the observed MJO fields. A daily climatology of observed variables is calculated over the period from 1981 to 2010, from 2000 to 2009 for CFSv2, and from 1993 to 2009 for VarEPS.

To extract the MJO component, the Wheeler and Hendon (2004) real-time multivariate MJO index (RMM) is calculated following Gottschalck et al. (2010). RMM1 and RMM2 are the two leading modes of the combined empirical orthogonal functions (EOFs) of OLR, U200, and U850 averaged between 15°N and 15°S. Figure 1 represents the spatial patterns of observed MJO life cycle composite (OLR and U850) captured by the observed two RMM indices without discriminating for season or amplitude. The amplitude of the MJO is defined as the square root of RMM1^2 plus RMM2^2 . The total observed and predicted MJO cases are separated into "strong" and "weak/non" MJO cases based on the observed MJO amplitude. The MJO is defined as strong MJO for the amplitude larger than 1.5 (32.3% of total observation cases) and weak/non MJO for those less than 1.0 (32.4% of total observation cases) during the entire observation period from 1981 to 2010. The predicted RMM indices are obtained by projecting the ensemble hindcast anomalies of zonal winds and OLR onto the observed eigenvectors of combined EOF.

The prediction skill is examined by the bivariate anomaly correlation coefficient (ACC) and bivariate root-mean-square error (RMSE) developed by Rashid et al. (2011) as

$$\text{ACC}(\tau) = \frac{\sum_{t=1}^N [a_1(t)b_1(t, \tau) + a_2(t)b_2(t, \tau)]}{\sqrt{\sum_{t=1}^N [a_1^2(t) + a_2^2(t)]} \sqrt{\sum_{t=1}^N [b_1^2(t, \tau) + b_2^2(t, \tau)]}},$$

$$\text{RMSE}(\tau) = \sqrt{\frac{1}{N} \sum_{t=1}^N [|a_1(t) - b_1(t, \tau)|^2 + |a_2(t) - b_2(t, \tau)|^2]},$$

where $a_1(t)$ and $a_2(t)$ are the observed RMM1 and RMM2 at time t , and $b_1(t, \tau)$ and $b_2(t, \tau)$ are the respective forecasts for time t with a lead time of τ days or lag time of τ days. Also, N is the number of predictions, and $\text{ACC}(\tau)$ is equivalent to a spatial pattern correlation between observation and forecast when they are reconstructed from the two leading EOFs (Lin et al. 2008). We use $\text{ACC}(\tau) = 0.5$ as a threshold for skillful prediction. Because of the large sets of hindcasts, the correlation coefficient is significant at 99% level when it

exceeds about 0.1. To examine the propagation speed error in predictions, we calculate the phase angle difference (error) between the observed and predicted RMMs following Rashid et al. (2011) as

$$\text{ERR}(\tau) = \frac{1}{N} \sum \tan^{-1} \left[\frac{a_1(t)b_2(t, \tau) - a_2(t)b_1(t, \tau)}{a_1(t)b_1(t, \tau) + a_2(t)b_2(t, \tau)} \right].$$

Negative angle indicates the slower propagation in predictions compared to the observation.

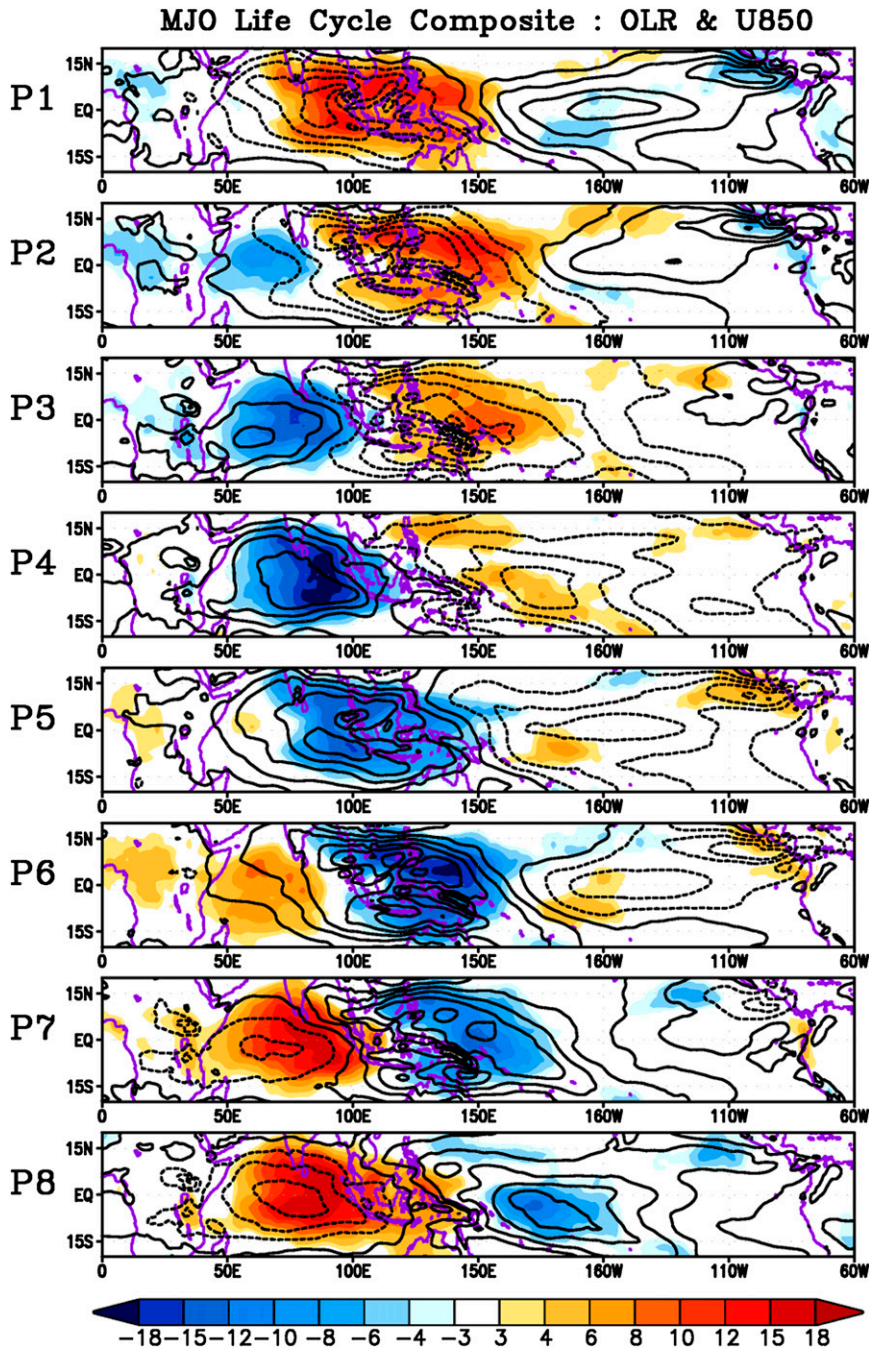


FIG. 1. MJO life cycle composite maps for OLR ($W m^{-2}$, shading) and 850-hPa zonal wind (ms^{-1} , contour interval is 0.5) anomalies calculated for each of the eight MJO phases.

3. MJO predictability and prediction skill

a. Predictability and prediction skill

Predictability is the skill that is theoretically achievable with a perfect model for a given set of equations, whereas prediction skill is what is actually achievable in a given prediction systems that contains model errors.

Therefore, the difference between the predictability and the prediction skill will provide an estimate of how much skill we can expect to increase by reducing the model error and by improving initial conditions. In this section, we compare the MJO predictability and prediction skill in both hindcasts. As recent studies have clearly shown that averaging ensembles can enhance the subseasonal

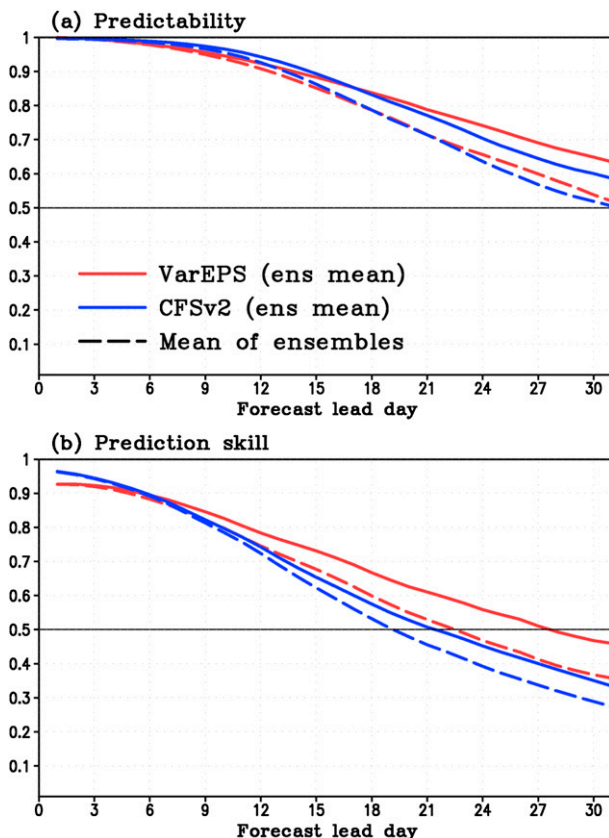


FIG. 2. (a) Predictability (bivariate correlation coefficient) and (b) prediction skill (bivariate correlation coefficient) for ensemble mean (solid) and mean of individual ensemble members (dashed) as a function of forecast lead day in VarEPS (red) and CFSv2 (blue).

prediction skill (Vitart et al. 2007; Fu et al. 2013), the effect of ensemble average on both predictability and prediction skill is also investigated.

First, we examine the predictability using ensemble mean and individual ensemble members. To estimate predictability using ensemble mean, the bivariate correlation coefficient $ACC(\tau)$ between one ensemble member (considered as “truth”) and the ensemble mean calculated from the rest of ensemble members is computed. Predictability is assessed by the average of $ACC(\tau)$ for each of the ensemble subsamples. Figure 2a shows the predictability of total MJOs, irrespective of the initial MJO amplitude, as a function of forecast lead time from 0 to 32 days for both VarEPS and CFSv2 hindcasts. Similar to Rashid et al. (2011), predictability remains around 0.6 at 32-day lead time in CFSv2 and even higher in VarEPS. The dashed line in Fig. 2a represents the predictability measured using individual ensemble. This is calculated by the $ACC(\tau)$ between one ensemble member and the rest of ensemble members and then

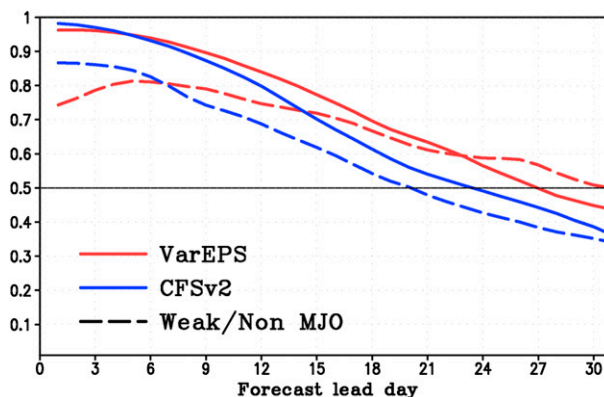


FIG. 3. Prediction skill (bivariate correlation coefficient) for initially strong (solid) and weak/non (dashed) MJOs as a function of forecast lead day in VarEPS (red) and CFSv2 (blue).

averaged over the subsamples. The predictability of individual ensembles is lower than that of the ensemble mean.

The prediction skill is far below the predictability (Fig. 2b). The ensemble mean prediction skill (solid line), defined as the forecast lead day when the ACC is 0.5, is about 27 days in VarEPS and 21 days in CFSv2, similar to the skill explored in recent studies for each system (Vitart 2014; Wang et al. 2014). The mean prediction skill of ensemble members is 3–5 days lower than that of the ensemble mean. The enhancement of skill in ensemble mean over individual ensembles is greater in VarEPS (~5 days) than in CFSv2 (~3 days), probably due to the slightly larger number of ensemble members or the smaller ensemble spread and error ratio. This issue will be explored later in this section. It is obvious that the MJO prediction skill has been gradually increased in each forecasting system compared to its previous version and the skill is now higher than that of statistical models (Maharaj and Wheeler 2005; Seo et al. 2009; Kang and Kim 2010; Rashid et al. 2011). However, the gap between the predictability and prediction skill still remains about 10 days.

Previous studies have demonstrated that the prediction skill depends strongly on the initial MJO amplitude (Lin et al. 2008; Agudelo et al. 2009; Kang and Kim 2010; Rashid et al. 2011; Zhang and van den Dool 2012; Wang et al. 2014). Figure 3 compares the prediction skill $ACC(\tau)$ for forecasts initialized with strong and weak/non MJO cases in two hindcasts. It clearly shows that forecasts initialized with strong MJO (solid line) possess greater prediction skill compared to those initialized with weak/non MJO (dashed line) in both models, probably due to the disorganized anomalies in the initial condition in the weak/non MJO. In CFSv2, the prediction skill of the MJO is greater by about 4–5 days

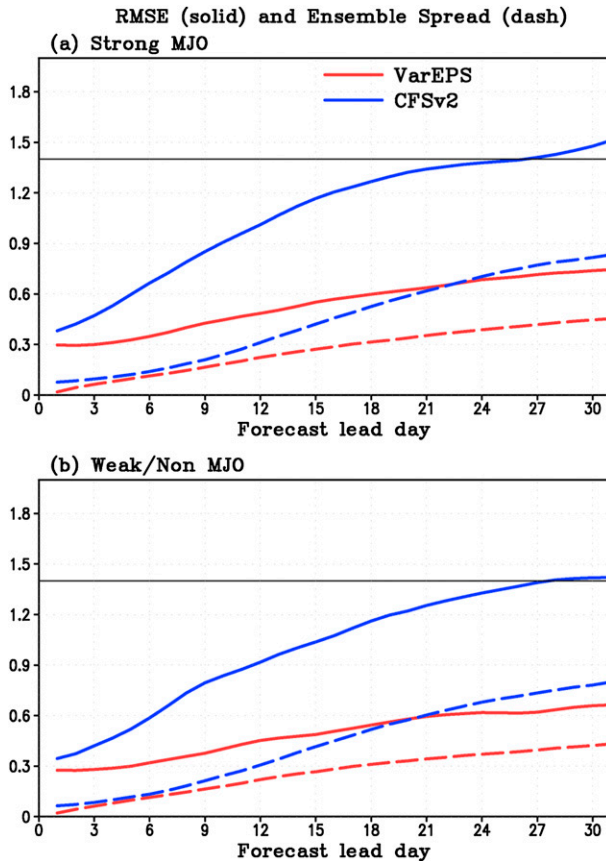


FIG. 4. RMSE of ensemble mean (solid) and ensemble spread (dashed) for (a) strong and (b) weak/non MJO as a function of forecast lead day in VarEPS (red) and CFSv2 (blue).

in forecast initialized in the strong MJO compared to the weak/non MJO for the entire forecast lead time. The VarEPS exhibits greater prediction skill when the MJO is initially strong compared to the weak/non MJO until day 23. The lower skill in the initially strong MJO after day 23 compared to the weak/non MJO might be related to the faster error growth in strong MJO cases (Fig. 4).

b. Ensemble spread and error

To estimate the uncertainties in ensemble predictions, characteristics of the ensemble spread and error are examined. While the error and the spread diagnostics should not be viewed as a complete assessment of reliability, they are important in identifying at what lead time the ensemble predictions are overdispersive or underdispersive and hence unreliable. In a perfect ensemble system, over a large sample of forecasts, the ensemble spread would equal the error of the ensemble mean (Weisheimer et al. 2011). However, current ensemble predictions for the MJO are in general underdispersive, meaning that there is a lack of spread around

the ensemble mean (Rashid et al. 2011; Hudson et al. 2013).

Ensemble spread and error of the ensemble mean in the VarEPS and CFSv2 are compared as a function of forecast lead days for forecasts initialized with strong (Fig. 4a) and weak/non MJO cases (Fig. 4b). Ensemble spread is defined as a standard deviation of the ensemble members about the ensemble mean, and error is defined as RMSE of ensemble mean about the observation. Results of both initially strong (Fig. 4a) and weak/non MJO cases (Fig. 4b) are almost identical. In both initially strong and weak/non MJO cases, the error in the VarEPS and CFSv2 is similar at the beginning, while it grows faster in CFSv2 as lead time increase. The VarEPS shows relatively smaller error and error growth rate by time than the CFSv2. The error exceeds the ensemble spread in both systems from the beginning, indicating that the ensemble prediction systems are underdispersive. For both initially strong and weak/non MJO cases, the growth rate of ensemble error by time is similar to the growth rate of ensemble spread by time in the VarEPS, while the growth rate of error is larger than the growth rate of spread by time in the CFSv2.

Previously, relative to Fig. 3, we mentioned that the skill of initially weak/non MJO exceeds the skill of initially strong MJO after 23 day in the VarEPS. The error for the strong MJO becomes larger than that of weak/non MJO after 23 day, while the spread is almost the same. More specifically, the largest error for initially strong MJO in the VarEPS results from forecasts initiated during MJO phase 5 where the MJO convection is located around the Maritime Continent. We will discuss this issue later in section 3.

c. Source of predictability

Although previous studies have examined the MJO prediction skill in depth, the source of the MJO predictability has yet to be fully addressed. By definition, the RMMs consist of three component variables (OLR, U200, and U850), making it possible to assess the contribution of each variable to the total prediction skill. The particular interest we have in this regard is whether it is the convective anomalies (OLR) or the lower- and upper-level large-scale circulation anomalies that provide the predictability of the MJO. We apply the measure of prediction skill (bivariate correlation coefficient of ensemble mean) using RMM indices constructed with OLR, U850, and U200, separately and in combination. Predictions of the individual components of the RMM index (i.e., RMM_OLR, RMM_U850, RMM_U200, RMM_U850 + U200) are compared against the observed total RMM. For example, the RMM_OLR is obtained by projecting only the predicted OLR anomaly onto the

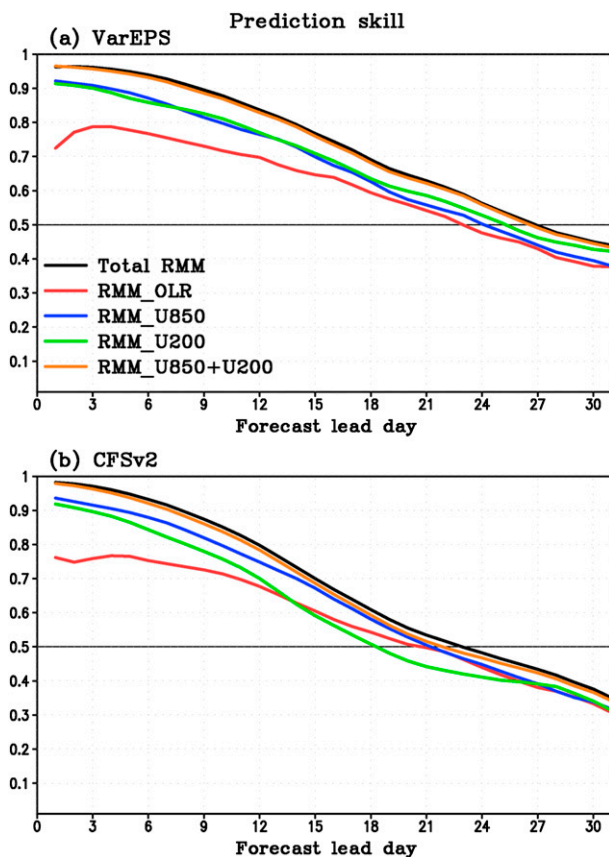


FIG. 5. Prediction skill for RMM indices constructed with OLR (red), U850 (blue), and U200 (green) separately, and U850 and U200 combined (brown), for initially strong MJO cases in (a) VarEPS and (b) CFSv2.

OLR component from the observed combined EOF eigenvector. By comparing predictions of RMM_OLR, RMM_U850, RMM_U200, and RMM_U850 + U200 with the observed total RMM, it might be possible to determine the contributions of each variable to RMM prediction skill and to find which variable tends to erode the skill.

Figure 5 shows the $ACC(\tau)$ between the observed total RMM and predicted RMM comprised of each variable separately, only for initially strong MJO cases. In both hindcasts, it is obvious that the MJO prediction skill relies on the prediction of large-scale circulation fields (RMM_U850, RMM_U200, and RMM_U850 + U200) compared to that of convection (RMM_OLR). In both hindcasts, wind components contribute mostly to the prediction skill from about 15-day lead, while OLR erodes the skill in these forecast lead times, consistent with previous studies (Waliser et al. 2003; Agudelo et al. 2009). For lead times less than 3 days, the VarEPS correlations for OLR show a dip in the correlation. This might be related to the OLR data used for verification

(NOAA AVHRR), but this issue needs to be examined further. In CFSv2, beyond a 15-day lead time, the correlation for the RMM_OLR starts to match or beat that for the individual wind components. These results suggest that OLR is not the major limitation for the predictions beyond 15 days.

It is natural to assume that the error emanating from the convective regions affects the error in the zonal wind, thus eroding, in turn, the prediction skill of the circulation. However, the prediction skill of RMM_U850 + U200 shows similar skill to the total RMM in both hindcasts, indicating that wind fields are not particularly affected by the large errors in OLR. It might be a result of a weak coupling between convection and large-scale circulations in both hindcasts. Therefore, a better representation of convection, circulation, and their interaction in dynamical models is crucial to improve MJO prediction. There may be a further factor: the contribution of each variable to the total RMM. It needs to be acknowledged that the OLR contributes less to the total RMM than the wind fields in observation. The fractional contribution of OLR to the variance of total RMMs is only 14.7%, compared to 43.9% for U850 and 41.4% for U200 (Ventrice et al. 2013). Therefore, the lower prediction skill in RMM_OLR can be expected.

d. Dependency of predictability and prediction skill on MJO phases

To assess the dependency of MJO predictability and prediction skill on the initial phase of the MJO, we compare first the prediction skill of each of the eight different MJO phases, as defined in Fig. 1. The ensemble mean prediction skill $ACC(\tau)$ for forecasts initialized with strong MJOs as a function of lead time and phase is assessed. In the VarEPS (Fig. 6a), the prediction skill is relatively high for the forecasts initialized with the MJO in phases 4 and 7 while a sharp decrease appears in phase 5. The relatively low skill is found in phases 1, 2, 5, and 8 when the anomalous convective signal is located around the Indian Ocean, the Maritime Continent, and the date line in the initial conditions. The skill in NCEP CFSv2 (Fig. 6b) is relatively lower than in VarEPS and does not vary as much between phases as does VarEPS. Skill decreases relatively quickly in phase 2, consistent with Wang et al. (2014). Knowing that the MJO is propagating eastward at a speed of 5 m s^{-1} , after about 15 days, the convective anomaly from the Indian Ocean (phase 2) is likely to be located near the Maritime Continent or about to enter the western Pacific (phase 5). The low MJO prediction skill for CFSv2 for cases initialized with the MJO in initial phase 2, therefore, can be related to the Maritime Continent predictability barrier, which was also apparent in the previous version (Seo and Wang 2010; Fu

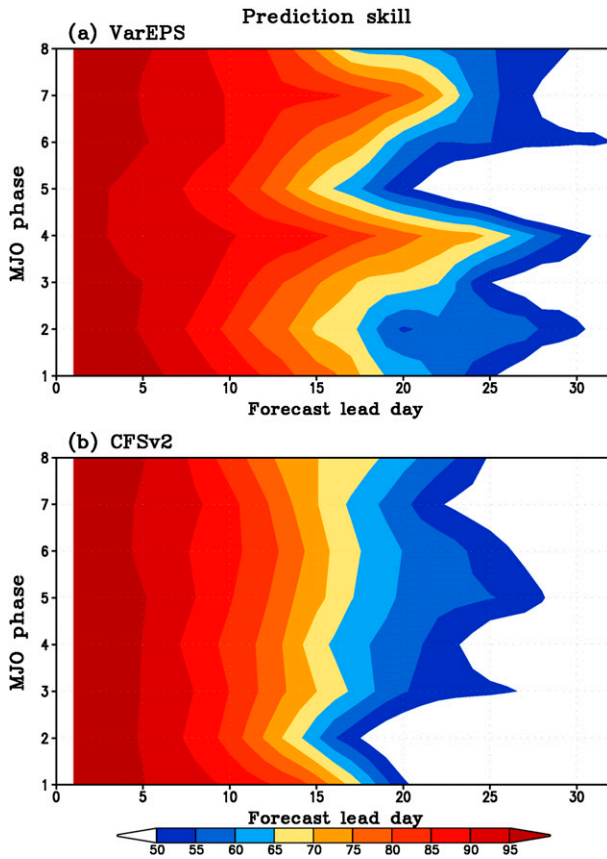


FIG. 6. Prediction skill for forecasts initialized in the MJO phase given on the left (*y* axis) for initially strong MJO cases in (a) VarEPS and (b) CFSv2. Correlation coefficients are multiplied by 100.

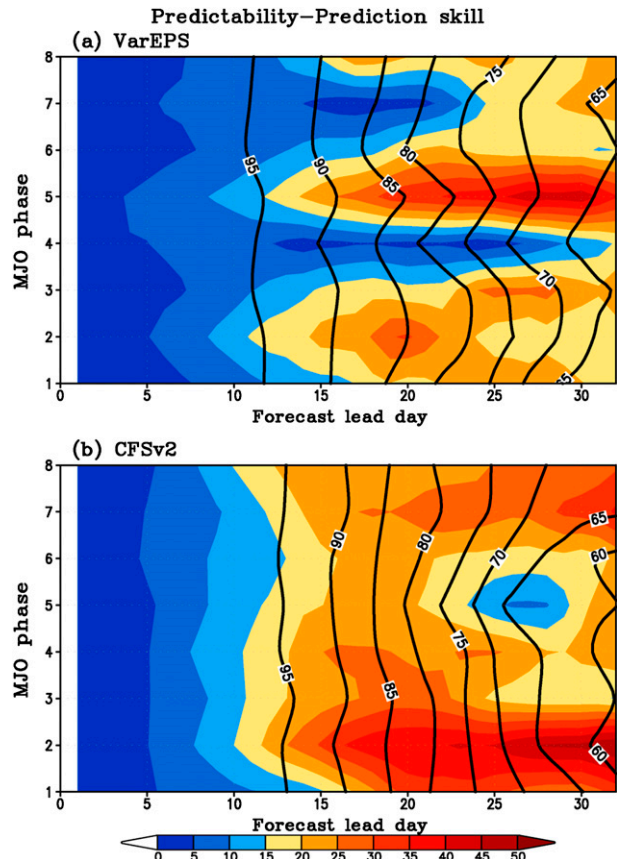


FIG. 7. Predictability (bivariate correlation coefficient; contours) and differences between predictability and prediction skill (bivariate correlation coefficient; shading) as a function of initial MJO phase and forecast lead day for (a) VarEPS and (b) CFSv2. The values are multiplied by 100.

et al. 2011; Weaver et al. 2011; Zhang and van den Dool 2012; Wang et al. 2014). The relatively low prediction skill for VarEPS for forecasts initialized with the MJO in initial phase 2 is found as well, but is not as clear as that of the CFSv2. Noting the inconsistency of the prediction skill by phases between two hindcasts, it can be concluded that the sensitivity of prediction skill to the initial MJO phases is forecast system dependent.

While the prediction skill differs by initial MJO phases, predictability is not sensitive to the phase of the initial MJO. Figure 7 shows the predictability (contour) and the differences between predictability and prediction skill for ensemble mean in bivariate correlation coefficients. In general, the difference increases as forecast lead day increases. In VarEPS, the largest difference is shown in phases 2 and 5. It is interesting that the prediction skill reaches close to the predictability in phases 4 and 7. The predictability measured with the CFSv2 hindcasts is similar to the VarEPS before about 25 days. However, the difference between the predictability and prediction skill in the CFSv2 is large in all phases,

especially in phase 2. The difference between the predictability and prediction skill indicates that the overall MJO prediction skill can be enhanced by focusing on the error growth in specific MJO phases with detailed analysis.

4. MJO propagation and amplitude

The errors in MJO propagation speed and amplitude have a direct impact on prediction skill. Figure 8 shows composite maps of OLR and U850 anomaly averaged in the band 5°S–5°N for both observational fields and ensemble mean hindcasts for initially strong MJOs. Initial phases 2 and 5 are selected for comparison, based on the prediction skill results (Figs. 6 and 7). Phase 2 is the phase that both hindcasts show large differences between the predictability and prediction skill. For the forecast cases initialized with the MJO in phase 2, both hindcasts are able to represent the propagation to some extent (Figs. 8a–c). However, the propagation speed is

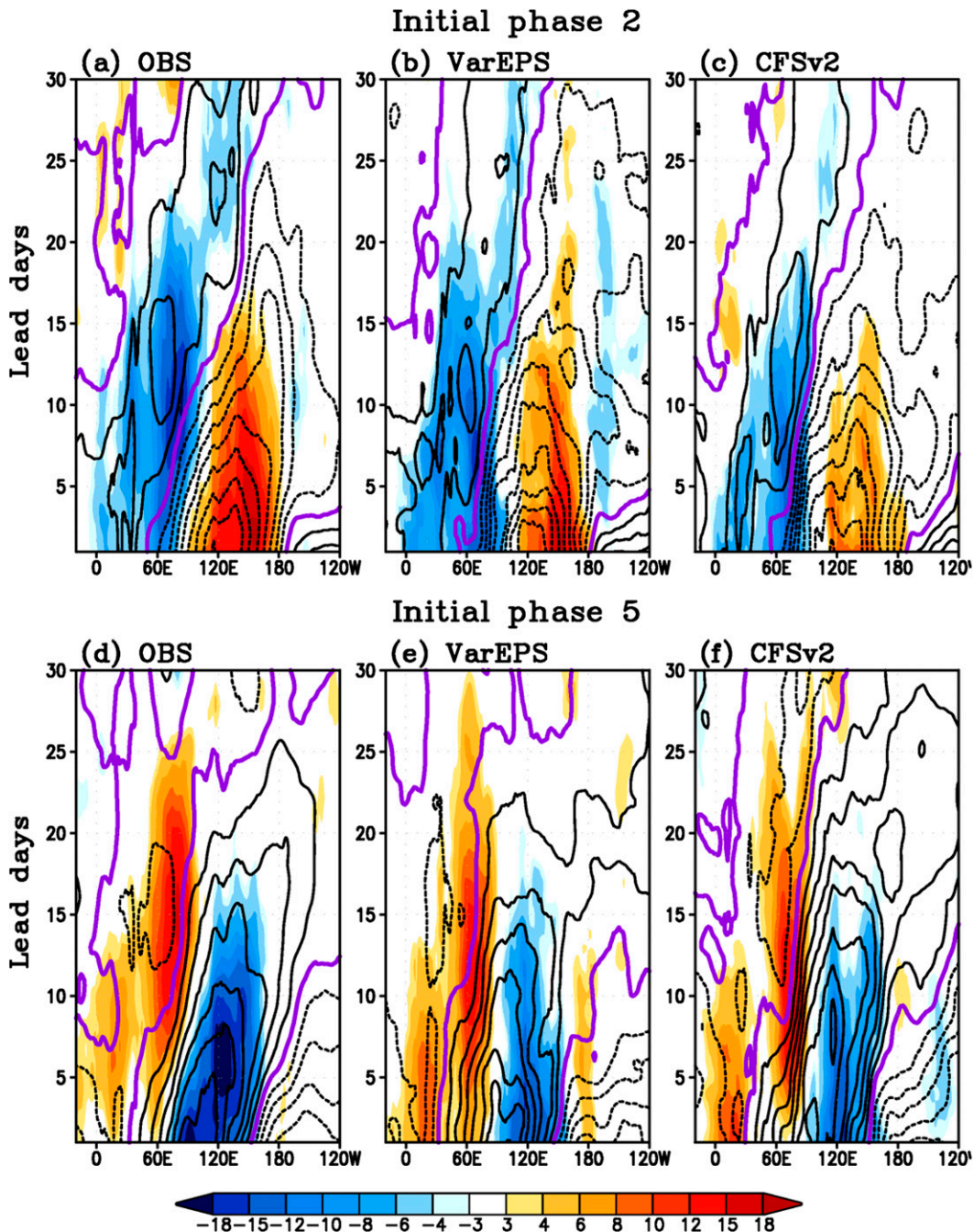


FIG. 8. Longitude–time composites of the OLR (shading) and U850 (contour interval of 0.5 m s^{-1}) anomaly averaged over 5°S – 5°N initiated at (a)–(c) MJO phase 2 and (d)–(f) phase 5 in observation, VarEPS, and CFSv2. The purple line indicates the zero line for U850 anomaly.

slower than that observed and the amplitude of convective anomaly is not as strong after about 10–15 days in both hindcasts. The amplitude of zonal wind anomaly is maintained, while it shows slower eastward propagation compared to the observed. The weaker signal of the suppressed convection over the western Pacific in the

predicted anomalies might contribute to the weaker eastward propagation of the convective anomaly over the Indian Ocean (Kim et al. 2014). Figures 8d–f show the composite maps for the forecasts initialized with the strong MJO in phase 5. This is the phase in which VarEPS shows the largest difference between the predictability

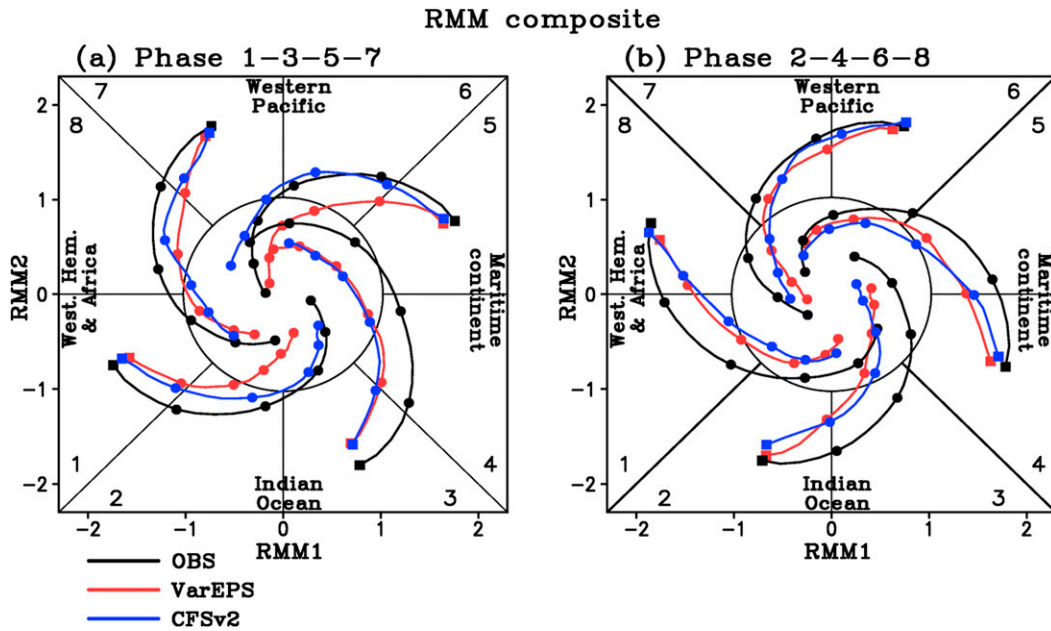


FIG. 9. RMM composite phase–space diagram for observation (black) and the ensemble mean of VarEPS (blue) and CFSv2 (red) hindcasts initialized in strong MJO phases. The dots represent every 5 days from the forecast starting date (square).

and prediction skill, while CFSv2 has relatively better prediction skill (Fig. 7). The skill difference between the two systems mainly results from the wind field. The VarEPS does not predict the propagation signal, while CFSv2 predicts the amplitude realistically with propagation of zonal wind field anomaly.

To condense the characteristics of propagation and amplitude in difference phases, a phase–space diagram for the predicted MJO is compared with the observed as a function of lead time (Fig. 9). It represents a composite of the RMM indices starting with an initially strong MJO for lead time up to 25 days. Overall, the propagation speed is slower and the amplitude is weaker than the observation. To examine quantitatively the propagation error in both hindcasts, we calculate the phase angle error $ERR(\tau)$ based on the composite phase diagram (Fig. 9) averaged over a forecast lead time of 1–25 days (Fig. 10). Negative values indicate slower propagation in hindcasts relative to the observation. Consistent with previous results (Fig. 9), the predicted MJO is indeed slower than the observed. The average propagation speed occurring in all phases is about 14.7° slower in VarEPS and 16.1° slower in CFSv2 than the observed for the first 25 days. The phase angle error varies among the phases. For VarEPS, the predicted propagation speed shows the largest difference from the observation in phase 1, and for CFSv2 in phase 8.

Next, we compare the change of the predicted MJO amplitude by forecast lead day. Figure 11 shows the

change of MJO amplitude after an initially strong MJO. Amplitude is defined here as the average amplitude of individual ensemble members. Error bars represent the ranges of one standard deviation of the MJO amplitude in ensemble members. Predicted MJO amplitude is weaker than the observed fields at the beginning of both hindcasts. Amplitude decreases gradually in both observation and CFSv2 as lead day increases, saturating at similar amplitude after 23 days. The amplitude in VarEPS is much less than that observed. The amplitude

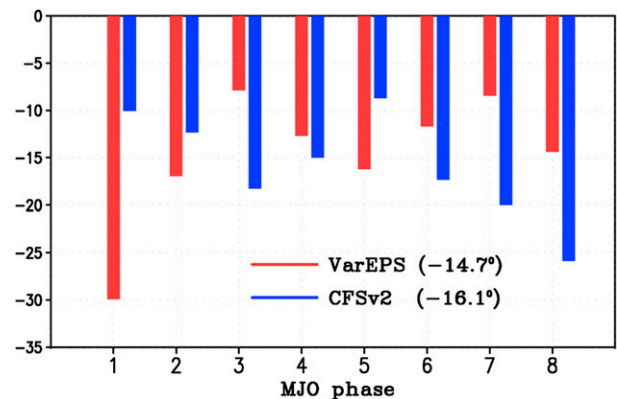


FIG. 10. The phase angle error averaged over 1–25-day forecast lead time for forecasts initialized in each MJO phase in VarEPS (red) and CFSv2 (blue). Numbers in parentheses indicate the average phase angle error over eight phases.

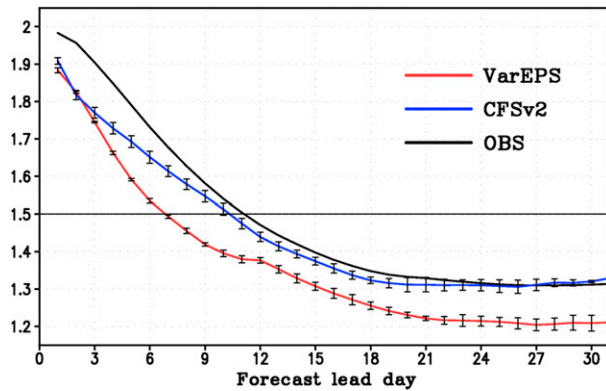


FIG. 11. The MJO amplitude for forecasts initialized with strong MJO averaged over individual ensemble members in VarEPS (red), CFSv2 (blue), and observation (black) as a function of forecast lead day. Error bars represent the ranges of one standard deviation of the ensemble members.

reaches the threshold of MJO amplitude (defined as 1.5) at 7 days in VarEPS, and at 10–11 days in the observation and CFSv2. Figure 12 shows the MJO amplitude averaged for 1–25 days and sorted by MJO phases. It has to be mentioned that the amplitude in Fig. 12 (computed by the mean of the ensemble members) is not the same as that of Fig. 9 (computed by the ensemble mean), as the amplitude computation is not linear. The averaged amplitude over eight phases is 1.63 for observation, 1.49 for VarEPS, and 1.57 for CFSv2. The observed MJO amplitude is relatively high when the strong MJO starts at initial phases 2 and 3, and 6 and 7, where the MJO is generally well organized. Phases 1 and 5 show the lowest amplitude in the observation. Those phases are where the convective signal is initially over the African continent or over the Maritime Continent (Fig. 1).

In both model sets, the MJO amplitude is smaller compared to observations at phases 2 and 3. Although the MJO convection is well organized as a strong MJO at the beginning, it becomes rapidly weaker during its eastward propagation, especially before the MJO enters the Maritime Continent (Figs. 8 and 9). VarEPS consistently predicts weaker amplitude over all phases, while CFSv2 overestimates the amplitude in several phases, especially phases 5, 6, and 7. These differences may be related to the mean SST biases in each model. Kim et al. (2012a,b) compared the mean SST biases in both sets of seasonal hindcasts, and found that the CFSv2 possesses a warm bias from the Maritime Continent to the western Pacific. On the other hand, the ECMWF model shows negative biases over the entire tropics, probably restricting the development of well-organized convection anomalies (Seo and Wang 2010). These key factors that modulate the MJO propagation and amplitude, and

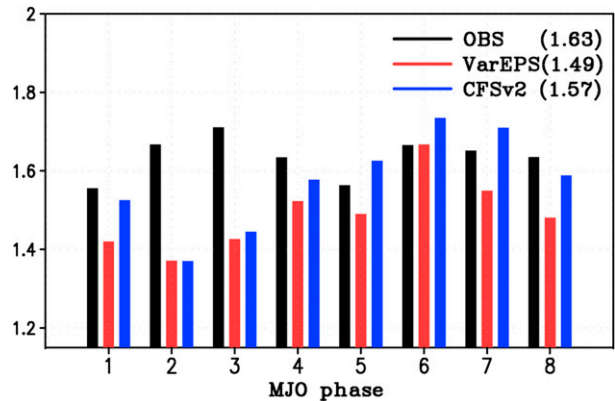


FIG. 12. MJO amplitude averaged over 25-day forecast lead time in eight MJO phases in VarEPS (red) and CFSv2 (blue). Numbers in parentheses indicate the average MJO amplitude over eight phases.

thus in turn the prediction skill, need to be investigated further with more detailed analysis.

5. Predictability and prediction skill for forecasts targeting MJO events

Many previous studies have assessed MJO prediction skill based on the initial amplitude and phases of the MJO. That is, the prediction skill is based on the existence of an MJO. But this procedure is limited. It does not provide information of how well the MJO would be predictable prior to its occurrence. In this study, we assess the predictability and prediction skill targeting MJO events relative to forecast lag days. Figure 13 shows the ensemble mean $ACC(\tau)$ for forecasts targeting strong and weak/non MJOs. The 0 day represents the day of the occurrence of strong or weak/non MJO and negative numbers are the lag day τ prior to the MJO events. As one might expect, the prediction skill increases as the time approaches toward zero. Both systems have clearly higher predictability and prediction skill when predictions are targeting a strong MJO rather than a weak/non MJO. In both hindcasts, the predictability is above 0.7 at 32 days ahead of the strong MJO and about 21 days (correlation 0.5) ahead of weak/non MJO (Fig. 13a). For prediction skill (Fig. 13b), the strong MJO is predictable 30 days ahead of time in CFSv2 and earlier than 32 days ahead in VarEPS. Both systems have prediction skill about 10 days when forecasts are targeting a weak/non MJO.

Figure 14 shows the prediction skill of the ensemble mean forecasts targeting strong MJOs in different MJO phases. In VarEPS, skill does not show clear differences between phases, while the CFSv2 shows sharp decrease

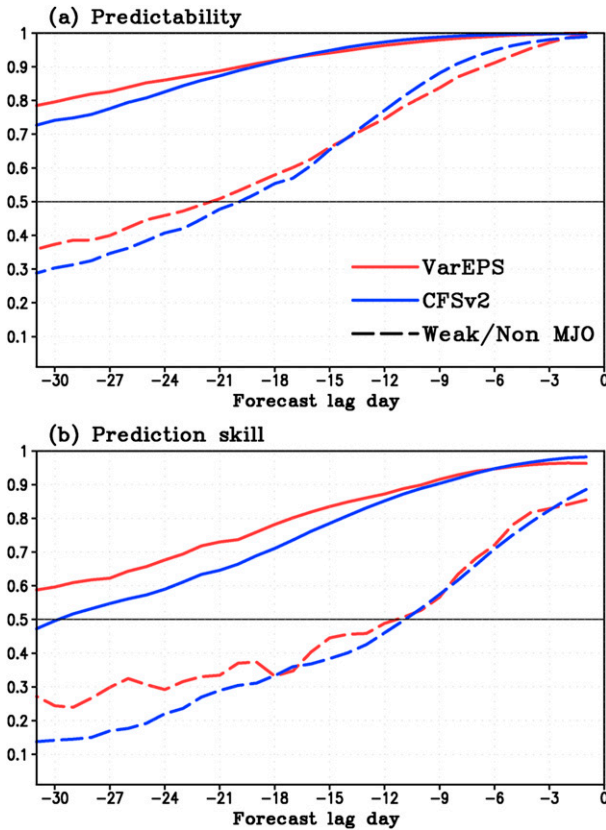


FIG. 13. (a) Predictability and (b) prediction skill for ensemble mean forecasts targeting strong (solid) and weak/non (dashed) MJOs as a function of forecast lag day in VarEPS (red) and CFSv2 (blue).

in skill at a specific phase 1 and 5. Skill decrease in the forecasts targeting MJO in phase 1 and 5 is related to the deficiency of the CFSv2 in predicting the enhanced (or suppressed) convective signal associated with the MJO over the Maritime Continent (Fig. 1) as mentioned before. The Maritime Continent prediction barrier in CFSv2 is found in Wang et al. (2014) as well. However, this barrier is not clearly represented in VarEPS. An operational model analyzed in Rashid et al. (2011) shows no existence of the Maritime Continent barrier either. Therefore, noting the discrepancy of skill in various operational models, it can be concluded that the sensitivity of prediction skill, including the Maritime Continent barrier, strongly depends on the forecast system.

Figure 15 shows the predictability (contour) and the differences between predictability and prediction skill (shading) for forecasts targeting strong MJOs. The predictability does not vary significantly among the phases in both hindcasts. The difference between predictability and prediction skill is apparent when forecasts are targeting MJO in phase 5 in CFSv2, indicating that prediction of the MJO over the Maritime Continent

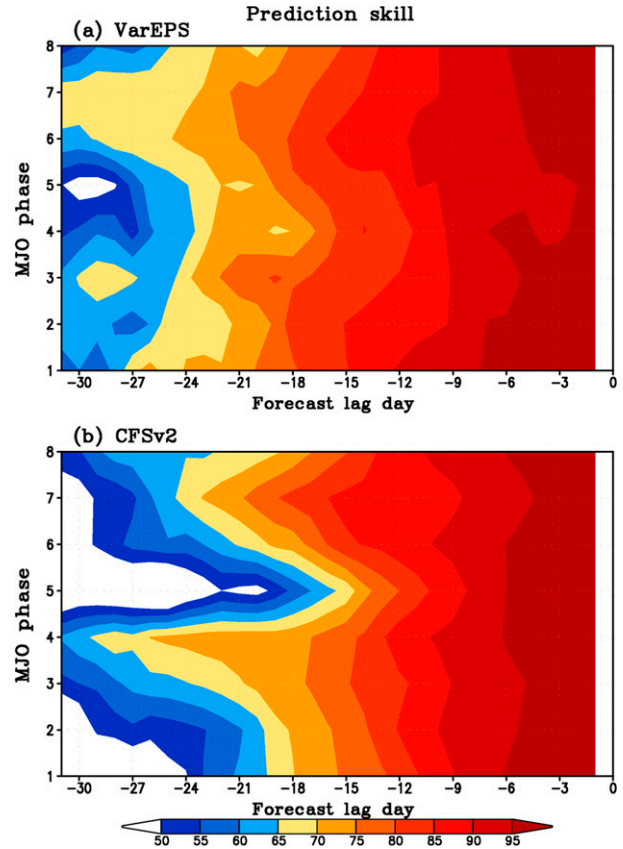


FIG. 14. Prediction skill for forecasts targeting strong MJOs as a function of MJO phase and forecast lag day in (a) VarEPS and (b) CFSv2. Correlation coefficients are multiplied by 100.

has room for further improvement by reducing the error in model physics and by improving initial conditions. In CFSv2, the difference in phase 1 and 2 is obvious as well.

We have compared the predictability and prediction skill for forecasts targeting the MJO. However, this analysis of skill that targets MJO events does not provide any information on how well a specific MJO event may be predicted, since the prediction skill analysis averages together the results of many different events. We have found that there are differences in the predictability of strong MJOs compared to weak/non MJOs. But we have not demonstrated any physical understanding of whether the next MJO will be strong (and presumably more predictable) or weak. Clearly, analysis focused on specific MJO events needs to be examined in detail. Investigation of the precursor signal of specific MJO (its formation and propagation) may lead us to better understand the source of predictability as well as the limitation of current models, and thus bring MJO prediction closer to its theoretical limits. Forecasting of the initiation of an MJO and its probable intensity remains a critical issue as well.

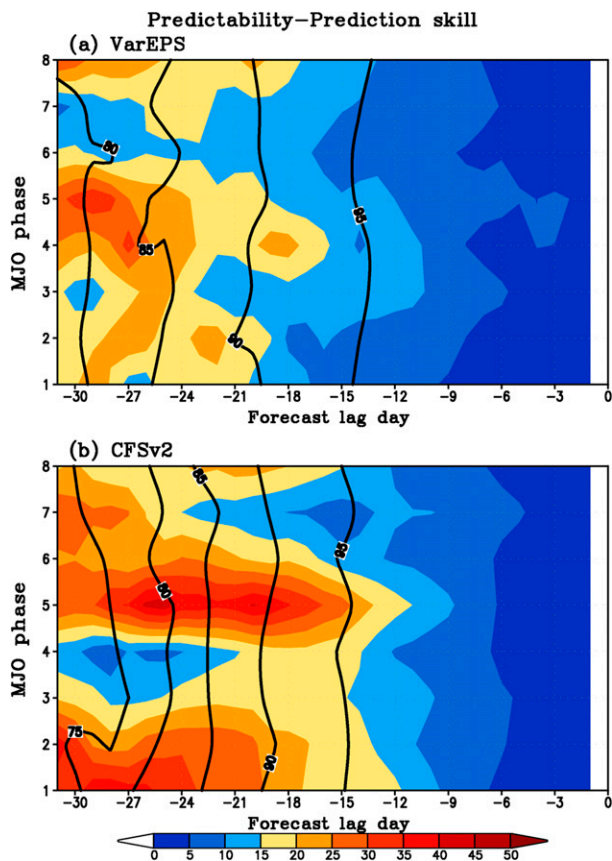


FIG. 15. Predictability (contours) and differences between predictability and prediction skill (shading) for forecasts targeting strong MJOs as a function of MJO phase and forecast lag day in (a) VarEPS and (b) CFSv2. Correlation coefficients are multiplied by 100.

6. Summary and discussion

This study has analyzed the current status of the MJO predictability and prediction skill by applying systematic verification methods to two state-of-the-art operational model hindcasts, the VarEPS and CFSv2. Predictability and prediction skill of the MJO are estimated using the bivariate anomaly correlation coefficient and the root-mean-square error between the observed and predicted RMM indices.

MJO predictability remains above 32 days lead time in both hindcasts, while the prediction skill is about 27 days in VarEPS and 21 days in CFSv2. Results show that forecasts initialized with strong MJOs possess greater prediction skill compared to those initialized with weak/non MJOs in both models, probably due to the disorganized anomalies in the initial condition in weak/non MJOs. By comparing ensemble spread and error in both hindcasts, it is shown that error exceeds the ensemble spread from the beginning of lead time, indicating that both ensemble prediction systems are underdispersive.

The error grows faster than the spread by lead time in CFSv2 while the ratio between error and spread is almost consistent across by forecast lead time in VarEPS. To investigate the source of predictability, we compared the prediction skill using RMM indices constructed with different variables separately. Both wind components contribute mostly to the prediction skill, while skill erodes after 15 days. The results imply that a better representation of convection, circulation, and interaction between those in dynamical models is crucial for improvement of the MJO prediction.

Dependency of MJO prediction skill and predictability on the initial MJO phase has been assessed as well. Prediction skill in CFSv2 is relatively lower than the VarEPS and skill does not vary as much between phases as does the VarEPS. The differences of skill between two hindcasts indicate that the sensitivity of prediction skill to the initial MJO phases depends strongly on the forecast system. While the prediction skill varies with initial MJO phases, predictability is not sensitive to the phase of the MJO. The discrepancy between the predictability and prediction skill in both hindcasts implies that the MJO prediction skill can be enhanced by reducing the model error and by improving initial conditions, especially by focusing on specific MJO phases.

The error in propagation speed and amplitude of the MJO directly impacts the prediction skill. By quantitatively examining the propagation speed error in both hindcasts, we found that the propagation speed is about 14.7° slower in VarEPS and 16.1° slower in CFSv2 than observed. The MJO amplitude decreases gradually and results in averaged amplitude of 1.63 for observation, 1.49 for VarEPS, and 1.57 for CFSv2 over 25-day lead time.

We have also assessed the predictability and prediction skill for forecasts targeting MJO events. Both hindcasts possess higher predictability and prediction skill when predictions are targeting strong MJOs than weak/non MJOs. For the prediction skill, a strong MJO is predictable 30 days ahead of time in CFSv2 and earlier than 32 days in VarEPS. Both systems have prediction skill of 10 days ahead of a weak/non MJO. It is seen clearly that skill decreases when forecast is targeting strong MJO in phase 5 in CFSv2, indicating the Maritime Continent prediction barrier. Characteristics of precursor signals for MJO prediction need to be investigated in depth to enhance the prediction skill closer to its predictability.

We have compared the predictability and prediction skill in the current operational model hindcasts focusing on total MJO cases. It needs to be emphasized that the analysis in this study uses mixed cases of primary and successive MJOs as well as propagating and nonpropagating MJOs. Given that primary and

successive MJOs have different characteristics as well as distinct precursor signals (Matthews 2008), further analysis needs to classify these two MJO types and examine their predictability and prediction skill separately. Another important issue that needs to be considered in the study of MJO predictability is the classification of MJO by its propagation type. Recent observational studies have shown that almost half of the observed MJOs located in the Indian Ocean propagate over the Maritime Continent, although half of those weaken before they reach the Maritime Continent (Lawrence and Webster 2002; Hirata et al. 2013; Kim et al. 2014). Distinguishing between these MJO events, propagating and nonpropagating, may provide insights into the overall predictability of MJO. The predictability for each type of MJO (primary and successive, propagating and nonpropagating) should be addressed further.

Acknowledgments. The constructive and valuable comments of three anonymous reviewers are greatly appreciated. We thank ECMWF and NCEP for providing the data to make this analysis possible. The Climate Dynamics Division of the National Science Foundation under Grant NSF-AGS 0965610 and the Korea Meteorological Administration Research and Development Program under Grant APCC2013-3141 provided funding support for this research. DK was supported by the NASA Grant NNX13AM18G.

REFERENCES

- Agudelo, P. A., C. D. Hoyos, P. J. Webster, and J. A. Curry, 2009: Application of a serial extended forecast experiment using the ECMWF model to interpret the predictive skill of tropical intraseasonal variability. *Climate Dyn.*, **32**, 855–872, doi:10.1007/s00382-008-0447-x.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Fu, X., B. Wang, J.-Y. Lee, W. Wang, and L. Gao, 2011: Sensitivity of dynamical intraseasonal prediction skills to different initial conditions. *Mon. Wea. Rev.*, **139**, 2572–2592, doi:10.1175/2011MWR3584.1.
- , J.-Y. Lee, P.-C. Hsu, H. Taniguchi, B. Wang, W. Wang, and S. Weaver, 2013: Multi-model MJO forecasting during DYNAMO/CINDY period. *Climate Dyn.*, **41**, 1067–1081, doi:10.1007/s00382-013-1859-9.
- Gottschalck, J., and Coauthors, 2010: A framework for assessing operational Madden–Julian oscillation forecasts: A CLIVAR MJO Working Group project. *Bull. Amer. Meteor. Soc.*, **91**, 1247–1258, doi:10.1175/2010BAMS2816.1.
- Han, W., D. Lawrence, and P. J. Webster, 2001: Dynamical response of equatorial Indian Ocean to intraseasonal winds: Zonal flow. *Geophys. Res. Lett.*, **28**, 4215–4218, doi:10.1029/2001GL013701.
- Hirata, F. E., P. J. Webster, and V. E. Toma, 2013: Distinct manifestations of austral summer tropical intraseasonal oscillations. *Geophys. Res. Lett.*, **40**, 3337–3341, doi:10.1002/grl.50632.
- Hudson, D., A. G. Marshall, Y. Yin, O. Alves, and H. Hendon, 2013: Improving intraseasonal prediction with a new ensemble generation strategy. *Mon. Wea. Rev.*, **141**, 4429–4449, doi:10.1175/MWR-D-13-00059.1.
- 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, doi:10.1175/2010JCLI3288.1.
- Kim, D., J. S. Kug, and A. H. Sobel, 2014: Propagating versus nonpropagating Madden–Julian oscillation events. *J. Climate*, **27**, 111–125, doi:10.1175/JCLI-D-13-00084.1.
- Kim, H.-M., and I.-S. Kang, 2008: The impact of ocean–atmosphere coupling on the predictability of boreal summer intraseasonal oscillation. *Climate Dyn.*, **31**, 859–870, doi:10.1007/s00382-008-0409-3.
- , P. J. Webster, and J. A. Curry, 2012a: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dyn.*, **39**, 2957–2973, doi:10.1007/s00382-012-1364-6.
- , —, —, and V. Toma, 2012b: Asian summer monsoon prediction in ECMWF system 4 and NCEP CFSv2 retrospective seasonal forecasts. *Climate Dyn.*, **39**, 2975–2991, doi:10.1007/s00382-012-1470-5.
- Lau, W. K. M., and D. E. Waliser, Eds., 2011: *Intraseasonal Variability of the Atmosphere–Ocean Climate System*. 2nd ed. Springer, 613 pp.
- Lawrence, D., and P. J. Webster, 2002: The boreal summer intraseasonal oscillation and the South Asian monsoon. *J. Atmos. Sci.*, **59**, 1593–1606, doi:10.1175/1520-0469(2002)059<1593:TBSIOR>2.0.CO;2.
- Liebmann, B., and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.
- 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, doi:10.1175/2008MWR2459.1.
- Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, doi:10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.
- , and —, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29**, 1109–1123, doi:10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2.
- Maharaj, E. A., and M. C. Wheeler, 2005: Forecasting an index of the Madden–Julian oscillation. *Int. J. Climatol.*, **25**, 1611–1618, doi:10.1002/joc.1206.
- Matsueda, M., and H. Endo, 2011: Verification of medium-range MJO forecasts with TIGGE. *Geophys. Res. Lett.*, **38**, L11801, doi:10.1029/2011GL047480.
- Matthews, A. J., 2008: Primary and successive events in the Madden–Julian oscillation. *Quart. J. Roy. Meteor. Soc.*, **134**, 439–453, doi:10.1002/qj.224.
- Rashid, H. A., H. H. Hendon, M. C. Wheeler, and O. Alves, 2011: Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. *Climate Dyn.*, **36**, 649–661, doi:10.1007/s00382-010-0754-x.
- Reichler, T., and J. O. Roads, 2005: Long-range predictability in the tropics. Part II: 30–60-day variability. *J. Climate*, **18**, 634–650, doi:10.1175/JCLI-3295.1.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

- , and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.
- Seo, K.-H., and W. Wang, 2010: The Madden–Julian oscillation simulated in the NCEP Climate Forecast System model: The importance of stratiform heating. *J. Climate*, **23**, 4770–4793, doi:10.1175/2010JCLI2983.1.
- , —, J. Gottschalck, Q. Zhang, J.-K. Schemm, W. Higgins, and A. Kumar, 2009: Evaluation of MJO forecast skill from several statistical and dynamical forecast models. *J. Climate*, **22**, 2372–2388, doi:10.1175/2008JCLI2421.1.
- Ventrice, M. J., M. C. Wheeler, H. H. Hendon, C. J. Schreck III, C. D. Thorncroft, and G. N. Kiladis, 2013: A modified multivariate Madden–Julian oscillation index using velocity potential. *Mon. Wea. Rev.*, **141**, 4197–4120, doi:10.1175/MWR-D-12-00327.1.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2256, in press.
- , 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, doi:10.1002/qj.623.
- , S. Woolnough, M. A. Balmaseda, and A. Tompkins, 2007: Monthly forecast of the Madden–Julian oscillation using a coupled GCM. *Mon. Wea. Rev.*, **135**, 2700–2715, doi:10.1175/MWR3415.1.
- , A. Leroy, and M. C. Wheeler, 2010: A comparison of dynamical and statistical predictions of weekly tropical cyclone activity in the Southern Hemisphere. *Mon. Wea. Rev.*, **138**, 3671–3682, doi:10.1175/2010MWR3343.1.
- Waliser, D. E., 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.
- 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, doi:10.1007/s00382-013-1806-9.
- Weaver, S. J., W. Wang, M. Chen, and A. Kumar, 2011: Representation of the MJO variability in the NCEP Climate Forecast System. *J. Climate*, **24**, 4676–4694, doi:10.1175/2011JCLI4188.1.
- Weisheimer, A., T. N. Palmer, and F. J. Doblas-Reyes, 2011: Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophys. Res. Lett.*, **38**, L16703, doi:10.1029/2011GL048123.
- 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, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.
- Zhang, C., 2005: Madden–Julian oscillation. *Rev. Geophys.*, **43**, RG2003, doi:10.1029/2004RG000158.
- , J. Gottschalck, E. D. Maloney, M. Moncrieff, F. Vitart, D. E. Waliser, B. Wang, and M. C. Wheeler, 2013: Cracking the MJO nut. *Geophys. Res. Lett.*, **40**, 1223–1230, doi:10.1002/grl.50244.
- Zhang, Q., and H. van den Dool, 2012: Relative merit of model improvement versus availability of retrospective forecasts: The case of Climate Forecast System MJO prediction. *Wea. Forecasting*, **27**, 1045–1051, doi:10.1175/WAF-D-11-00133.1.