Climate Dynamics
https://doi.org/10.1007/s00382-018-4423-9

Evaluating the MJO prediction skill from different configurations of NCEP GEFS extended forecast

Wei Li[2] · Yuejian Zhu[1] · Xiaqiong Zhou[2] · Dingchen Hou[1] · Eric Sinsky[2] · Christopher Melhauser[2] · Malaquias Peña[3] · Hong Guan[4] · Richard Wobus[2]

Received: 11 May 2018 / Accepted: 29 August 2018

Abstract
NOAA is accelerating its efforts to improve the numerical guidance and prediction capability for extended range (weeks 3 and 4) prediction in its seamless forecast system. Madden Julian Oscillation (MJO) is the dominant mode of sub-seasonal variability in tropics and prediction skill of MJO is investigated in this paper. We used different configurations of the NCEP Global Ensemble Forecast System (GEFS) to perform the experiments. The configurations include (1) the operational version of the stochastic perturbation forced with operational Sea Surface Temperatures (SSTs); (2) an updated stochastic physics forced with operational SSTs; (3) an updated stochastic physics forced with bias-corrected SSTs that are from Climate Forecast System (Version 2); and (4) as in (3) but with the addition of a scale aware-convection scheme. We evaluated MJO prediction skill from the experiments using Wheeler–Hendon indices and also examined the performance of the forecast system on prediction of key MJO components. We found that using the updated stochastic scheme improved the MJO prediction lead-time by about 4 days. Further updating the underlying SSTs with the bias corrected CFSv2 forecast increased the MJO prediction lead time by another 1.7 days. The best configuration of the four experiments is the last configuration which extends forecast lead time by ~9 days. Further investigation shows that upper and lower level zonal wind over the tropics has larger improvement than the outgoing longwave radiation (OLR). The improvement of the MJO prediction skill appears to be related primarily to the improvement in the representation associated circulations and OLR over the tropical West Pacific.

1 Introduction

A skillful forecast for the sub-seasonal time scale (3–4 weeks) is valuable in socio-economic context but poses substantial challenges. The limited prediction skill for this time window is mainly related to its relatively weak dependence on the initial conditions (an important source of predictability for the short term weather forecasts) and the insufficient time for the forecast system to 'feel' the effects of the lower boundary forcing that provide predictability on seasonal and longer timescales (Vitart 2009; Johnson et al. 2014; Liu et al. 2016; Troccoli 2010; Tian et al. 2017). Thus improving sub-seasonal forecasts is likely to come from substantial efforts on the model development, with a focus on improving representation of the sources of the sub-seasonal predictability.

As a dominant mode in tropical variability on the sub-seasonal timescale, the Madden–Julian Oscillation (MJO), which features as a 30–60 day oscillation of convection and circulation in the tropics has been a focal point of the research community and operational centers that are looking to improve sub-seasonal prediction. Numerical studies have found that improvement in tropical and extratropical prediction on sub-seasonal time scales can be linked to improved prediction of the MJO (Ferranti et al. 1990; Waliser et al. 2003; Lin and Brunet 2009; Pegion and Sardeshmukh 2011; Vitart 2014; Liu et al. 2017). With increasing interest and demand for skillful sub-seasonal forecasts, better representation of the MJO is of particular importance in operational numerical weather prediction (NWP) centers. In recent years, progress in MJO forecasting as a result of NWP developments has been quite promising. For

✉ Wei Li
Wei.Li@noaa.gov

[1] EMC, NCEP, NWS, NOAA, College Park, MD 20740, USA

[2] IMSG at EMC, NCEP, NWS, NOAA, College Park, MD 20740, USA

[3] Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT 06269, USA

[4] SRG at EMC, NCEP, NWS, NOAA, College Park, MD 20740, USA

Published online: 07 September 2018

example, the European Centre for Medium-Range Weather Forecasts (ECMWF) has mitigated the temporal decline in MJO prediction skill by about 1 day per year since 2002 (Vitart 2014). It was found that the large improvement before 2009 was mostly attributed to the change in convective parameterization (Vitart 2009). The progress of MJO prediction skill offers promise for corresponding potential improvements in the sub-seasonal forecast for most other phenomena.

Concerning the MJO simulation in weather and climate models, researchers and model developers mainly focus on these key areas for improvement of MJO representation (1) model physics parameterization: particularly, the MJO propagation was found to be sensitive to the convection scheme or diabatic heating profile (Wang and Schlesinger 1999; Maloney and Hartmann 2001; Liu et al. 2005; Lin et al. 2008; Li et al. 2009; Zhang and Song 2009; Vitart 2009; Zhou et al. 2012; Xiang et al. 2015). (2) Ocean impact. The MJO is mainly an atmospheric phenomenon but is largely impacted by the ocean, particularly accurate sea surface temperatures and atmospheric ocean coupling are believed to be critically important for prediction (Wang et al. 2015; Liu et al. 2017). (3) Ensemble forecast using either single model (Vitart and Molteni 2010; Hudson et al. 2013) or multi-model approaches to effectively sample model uncertainty (Gottschalck et al. 2010; Fu et al. 2013). The National Center for Environmental Prediction (NCEP) Global ensemble Forecast System (GEFS) provides numerical guidance for probabilistic forecast with the lead time up to 16 days. To align with the NOAA effort to accelerate sub-seasonal prediction in a seamless ensemble forecast system, several experiments that extend the GEFS integration time to cover weeks 3 and 4 lead time were performed. An early investigation indicated that MJO prediction skill was improved by using an optimal SST scheme in 35-day run of GEFS (Zhu et al. 2017). Following that investigation, more comprehensive experiments were performed to test the impact of GEFS configurations on the MJO prediction skill.

Section 2 describes the details of the experimental design and data used in this study. The evaluation of the prediction skill of the MJO and its associated key components is demonstrated in Sect. 3. Conclusions and discussion are provided in Sect. 4.

2 Experimental design and data

In this study, four experiments were conducted to investigate the key factors in the GEFS configurations that impact the MJO forecast prediction. The scientific areas include different stochastic perturbation schemes, the underlying SST forcing and the convection scheme.

2.1 35-day run using operational version of GEFS

The NCEP Global Ensemble Forecast System (GEFS version 11, Zhou et al. 2017), based on the Global Forecast System version 12 (GFS, i.e. Global spectral model GSM + land surface model LSM) is used to perform the experiments. For each experiment, a 21-member (1 control run and 20 perturbed members) ensemble was used to integrate the forecast system that is integrated up to 35 days. Considering the computational cost and the relatively small impact of the resolution on longer lead times (Tracton and Kalnay 1993; Ma et al. 2012), T_L 574 (semi-Lagrangian dynamics) horizontal resolution (roughly equivalent to 33 km grid spacing) for the first 8 lead days and T_L 382 (~ 55 km) for lead days 8–35 was used. 64 vertical levels were used for all lead times.

The initial analyses and perturbations of the 21-member GEFS are taken from the NCEP GFSv13 data assimilation system (using 4D Hybrid Ensemble-Var), which was implemented in May 2016. Because GFS is an uncoupled atmosphere ocean forecast system, a prescribed SST initialized from Real Time Global (i.e. RTG) analysis that damped towards the observed climatology (with a 90-day e-folding rate) was used to force the model (Gemmill et al. 2007). The sea ice concentration is initialized from daily 0000UTC sea ice analysis from the Interactive multisensor snow and ice mapping system (Ramsay 1998). The SST and sea ice forcing are identical for all GEFS members.

The current operational GEFS uses the Stochastic Total Tendency Perturbation method (STTP; Hou et al. 2006, 2008) that adds a random combination of the total tendency from other ensemble members every 6 h to represent the random model error. Based on the investigation of the past GEFS experiment, the effect of the STTP is to increase the spread and represent the model error more effectively over the extratropical region in boreal winter but minor effect over the tropical region (Zhu et al. 2018). Since STTP is the scheme used in the operational GEFS, the 35-day run of this GEFS configuration is taken as a reference for the other three experiments and is named 'STTP' hereafter.

2.2 New stochastic physics scheme

The first experiment that was performed is to replace the STTP with a new stochastic physics which includes three schemes (1) stochastic perturbed parameterization tendency (SPPT, Buizza et al. 1999; Palmer et al. 2009), (2) stochastic kinetic energy backscatter (SKEB; Shutts and Palmer 2004; Shutts 2005; Berner et al. 2009), and (3) stochastic perturbed humidity (SHUM; Tompkins and Berner 2008). The

suite of the three stochastic physics perturbation schemes has already been implemented in GFS hybrid-EnKF data assimilation system and will be used in the next configuration of the operational GEFS. As such, it is naturally considered as the first experiment to be tested. Specifically, SPPT perturbs the tendencies of temperature, wind and moisture, thus it represents the structural uncertainty related to sub-grid physics parameterizations (excluding clear-sky radiation). In practice, five random patterns with spatial correlation scale/time scales of 500 km/6 h, 100 km/3 day, 200 km/30 day, 2000 km/90 day, 2000 km/1 year are used to determine the perturbations. The patterns are applied to all model levels except it is reduced in magnitude near the surface to avoid instability caused by the perturbation in boundary layer and is tapered to zero near and above the tropopause to avoid perturbations on to the well-constrained clear-sky radiation (Ollinaho et al. 2017). SKEB simulates the transferring of the unresolved subgrid-scale energy to resolved scale. This is done via a stream function forcing based on the dissipation rate. Unlike other implementations of SKEB that considered the numerical dissipation, dissipation related to the mountain gravity wave drag, and estimated kinetic energy generated by updrafts and detrainment within sub-grid deep convection (Shutts 2005; Leutbecher et al. 2017), GFS implementation of SKEB only considers numerical dissipation (diffusion). In practice, perturbations are generated independently on each vertical level, and then vertically smoothed to provide some vertical coherence. The inclusion of SKEB improves the power spectrum of the global model, which otherwise exhibits damped power near the truncation frequency (Zhu et al. 2018). SHUM simulates the uncertainty related to the sub-grid scale humidity variability on the triggering of convection (Tompkins and Berner 2008). In practice, SHUM uses the same random pattern generator as SPPT but only one single spatial–temporal scale. The perturbation is maximum in the lowest model level and decreases exponentially with height within the boundary layer (Zhu et al. 2018).

Compared to STTP, the effect of SPPT and SHUM is to produce additional spread over the tropics especially upon the mid-upper level of the troposphere while the effect of the SKEB is similar to STTP (figure not shown, Zhu et al. 2018). The combined effect of SPPT, SHUM and SKEB (SPs hereafter) thus lead to an increase in the spread over the tropics. The impact of the SPs on the prediction skill of key MJO associated variables will be discussed in Sect. 3.

2.3 Updated SST forcing

In the operational version of GEFS, the prescribed SST uses analysis SST at the initial time and then damps to the climatological SST with increasing lead time. This kind of boundary SST is valid for the short term forecast which is when the initial condition dominates. With the increase in lead time, the contribution of the boundary SST becomes more significant. As such, in the third experiment, we replaced the SST forcing with a weighting average of the climatological SST and bias-corrected SST from a coupled model (NCEP CFSv2, Saha et al. 2014). i.e.

$$SST^{t} = (1-w)\left(SST_a^{t0} - \overline{SST}_a^{t0} + \overline{SST}_a^{t}\right) + w\left[SST_m^{t} - \left(\overline{SST}_m^{t} - \overline{SST}_a^{t}\right)\right], \quad w = \frac{t-t0}{T} \tag{1}$$

where SST_a^{t0} is the initial analysis SST, $\overline{SST}_a^{t0}$ is the initial climatological SST from analysis data, $\overline{SST}_a^{t}$ is the climatological SST from analysis data at lead time t, SST_m^{t} is the CFSv2 forecast SST at lead time t, $\overline{SST}_m^{t}$ is the CFSv2 climatological SST at lead time t. w is the weighting with t = [0,35], $t0 = 0$, T = 35. The idea of new SST scheme is using model forecast SST to represent the day-to-day SST variability. Since even the bias-corrected SST has difference from the analysis SST at the initial time, we weighted the analysis SST and model SST to reduce the initial shock so that to obtain a smoothed SST curve for all lead days. We named this scheme as the "2-tiered" SST scheme. The first tier refers to the coupled model forecast and second tier refers to the input for the uncoupled model. We considered the "2-tiered" SST scheme as an interim configuration between the uncoupled GEFS and a future fully coupled version. The merit of 2-tiered SST scheme is it considered more realistic SST variation and more computationally economical than the coupled model.

2.4 Updated convection scheme

NCEP operational GFS uses the Simplified Arakawa-Schubert convective parameterization scheme for deep convection (SAS, Arakawa and Schubert 1974; Grell 1993; Pan and Wu 1995; Han and Pan 2011) and the Mass-flux shallow convection scheme for shallow convection. Since July 2017, GFS updated the cumulus convection scheme with a scale- and aerosol-aware convection scheme (Han et al. 2017). The change in the convection scheme is associated with the increasing model resolution and the major updates included (1) the cloud base mass flux decreases with the increase of the fractional updraft area; (2) the convective adjustment time is proportional to the ratio of convective time to the convective turnover time; (3) cloud base mass flux in both deep and shallow convection now a function of mean updraft velocity for a grid size smaller than a threshold; (4) the rain conversion rate decreases with decreasing air temperature above the freezing level but also increase with the decreasing aerosol number concentration. (5) Convective inhibition in the sub-cloud layer is an additional trigger condition to suppress unrealistic noisy rainfall (Han

et al. 2017). The updated cumulus convection scheme demonstrated significant improvement on reducing the precipitation bias (too-much convective rain and light rain) especially during summer time.

It should be noted that although we proposed three scientific areas that could be focused on, in the actual experimental design, we didn't separate the individual impact of SST forcing and convection scheme in our experiments due to the limit of computational resources that we can hardly afford. In fact, except the second experiment which replaces the STTP with the SPs, the third experiment is designed to include both updated SST and SPs, and the fourth experiment includes an updated SST, SPs and the updated convection scheme. The combination of SST and new convection scheme is based on the early stage investigation (Zhu et al. 2017) and short term examination.

The configuration of all experiments is summarized in Table 1. To make readers more easily identify the experiments, we reinstate the experiments abbreviation here. Experiment 1: STTP; experiment 2: SPs; experiment 3: SPs + CFSBC, and experiment 4: SPs + CFSBC + CNV. All experiments were initialized every 5 days at 00Z, starting May 1st 2014 and ending May 26, 2016.

The analysis data used in this study is the NCEP GFS analysis. To avoid the impact of possible missing data in the forecast, a 7-point daily mean centered at 00Z of each lead day (e.g. using 06Z, 12Z, 18Z 24Z, 30Z, 36Z and 42Z average to represent the daily mean of the lead day 1) is used to represent the daily mean in both forecast and analysis data. Both forecast and analysis data are on 2.5° resolution.

3 Evaluation results

3.1 MJO prediction

We evaluated the MJO prediction skill following Gottschalck et al. (2010) using Wheeler Hendon MJO indices (Wheeler and Hendon 2004), The MJO prediction skill is defined as the bivariate anomaly correlation between the analysis and forecast Realtime Multivariate MJO (RMM1) and (RMM2) calculated using the outgoing longwave radiation (OLR), zonal wind at 200 hPa (U200) and 850 hPa (U850). i.e.

$$AC(\tau) = \frac{\sum_{i=1}^{N}\left[a_1(t)f_1(t,\tau) + a_2(t)f_2(t,\tau)\right]}{\sqrt{\sum_{i=1}^{N}\left[a_1^2(t) + a_2^2(t)\right]}\sqrt{\sum_{i=1}^{N}\left[f_1^2(t,\tau) + f_2^2(t,\tau)\right]}}, \quad (2)$$

where f_1(t,τ) and f_2(t,τ) are the RMM1 and RMM2 of the forecast at lead day τ initialized at day t. $a_1(t)$and $a_2(t)$ are the RMM1 and RMM2 of the analysis data corresponding to the forecast at day t.

To obtain the RMM1 and RMM2 for each experiment, we first calculated the anomaly of the OLR, U200 and U850 of both forecast ensemble mean and analysis data. Since we only have 18 years reforecast for one of the four experiments and the GDAS data is only available since 2011, the forecast and analysis anomaly were calculated by removing the 1979–2001 climatology using NCEP NCAR reanalysis and prior 120-day mean using GDAS for each forecast initial date and the corresponding analysis data. The anomaly data was then averaged over 15°S–15°N and normalized using 15.1 W/m for OLR, 1.81 m/s for U850 and 4.81 m/s for U200 (Gottschalk et al. 2010), to represent the tropical mean anomaly. The RMM1 and RMM2 were calculated by projecting the forecast and analysis anomaly onto the Wheeler Hendon Empirical Orthogonal Functions (EOFs) for each field and normalizing the RMMs by the given coefficient (Gottschalk et al. 2010). The RMM1 and RMM2 calculated this way should be considered as removing the interannual and seasonal variability so that the impact of El Niño and Sothern Oscillation (ENSO) can be neglected.

The MJO prediction skill with the updated stochastic physics schemes, SST scheme and convection scheme showed improvement from the operational GEFS with each successive enhancement (Fig. 1). A big improvement in MJO skill occurred by only changing the stochastic schemes from STTP to SPs, with the skill improved from ~ 12.5 days (defined as anomaly correlation reached 0.5) to ~ 16.8 days. Further updating the underlying SST improved the MJO prediction skill by another 1.7 days and updating all three areas of the GEFS configuration (configuration 4) improved the MJO skill to 22 days (Fig. 1a). In the GEFS forecast,

Table 1 Summary of the GEFS configurations

Abbreviation	Stochastic physics scheme	SST	Convection scheme
STTP	STTP	Relax to Climatology	SAS
SPs	SPPT + SHUM + SKEB	Relax to Climatology	SAS
SPs + CFSBC	SPPT + SHUM + SKEB	Initial analysis + bias corrected CFS forecast	SAS
SPS + CFSBC + CNV	SPPT + SHUM + SKEB	Initial analysis + bas corrected CFS forecast	Updated SAS

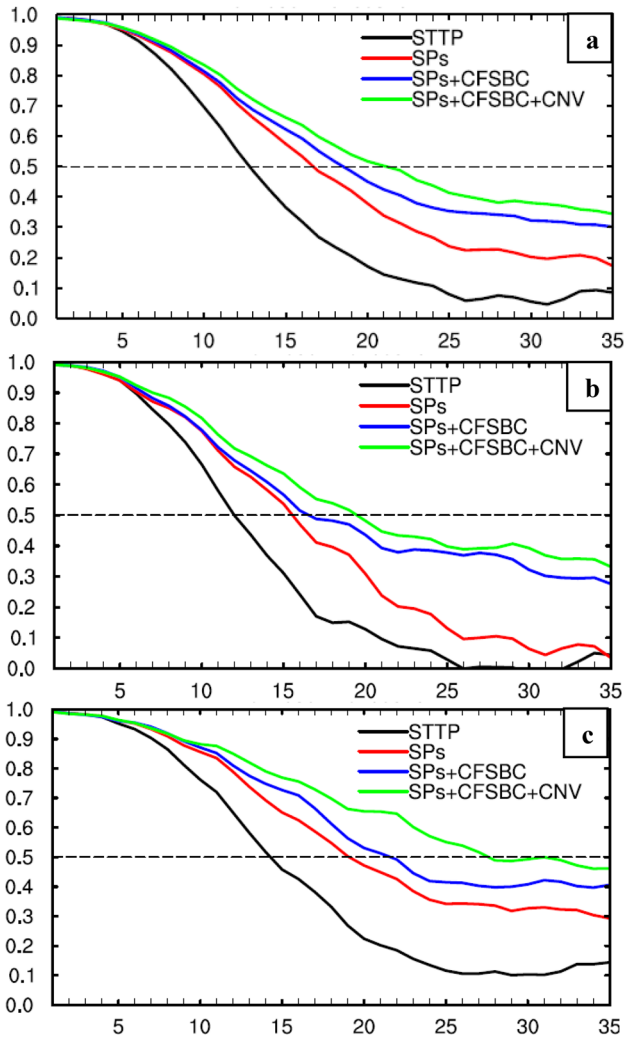

Fig. 1 MJO prediction skill for **a** RMM1 and 2, **b** RMM1, **c** RMM2 for different experiments during the period of May 1st 2014–May 26, 2016 (5 day-inteval). The MJO prediction skill is defined as the bivariate correlation of the RMM1 and RMM2 between ensemble mean forecast anomaly and anomaly of the analysis data. Climatology and 120-day running mean are removed from the forecast and analysis data when calculating the RMMs. A dash line of anomaly correlation equals to 0.5 is added in the plot to indicate the MJO skill

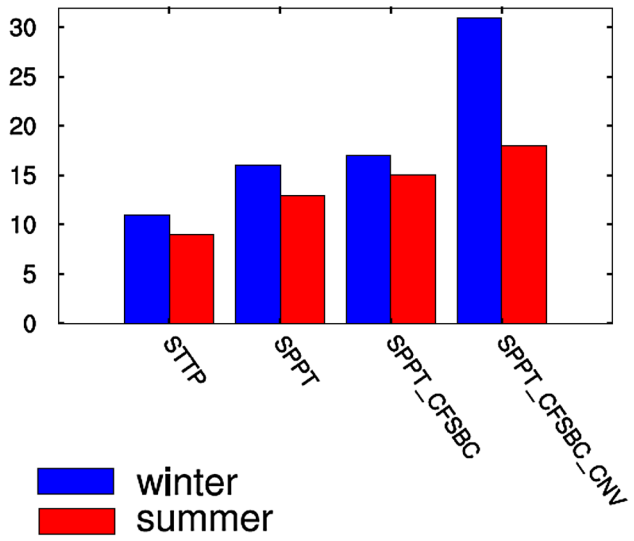

Fig. 2 Lead days that the MJO skill score reaches 0.5 for the four experiments in winter and summer seasons during the period of May 1st 2014-May 26 2016. 36 cases in November, December 2014-January 2015 and November, December 2015-January 2016 was used to calculate the MJO skill score for winter season. 36 cases in June–August, 2014 and 2015 was used to calculate the MJO skill score for summer season

the ranking of the experiments in RMM1 is consistent with RMM2 and they both are consistent with the skill of combined RMM1 and RMM2. The skill of RMM2 is greater than RMM1 (Fig. 1b, c), leaving an open question whether the forecast over the Maritime continent and tropical Africa is more skillful than that over tropical West Pacific and Indian Ocean in GEFS.

The noteworthy part of this result is the significant effect of the stochastic physics scheme. Although improvement on SST forcing and model physics also have positive impact on the MJO prediction skill, the improvement of the MJO skill in experiments 3 and 4 should be considered as the combined effect with the stochastic physics perturbation. Neena et al. (2014) by examining the relationship between ensemble spread and ensemble mean RMSE for MJO prediction from different ensemble prediction systems (EPS) found that the EPS with better spread is associated with better ensemble mean prediction skill. Vitart (2017) mentioned the ensemble generation method is also an important factor in the MJO prediction skill in addition to the initialization, model physics and resolution. We put a short note on this factor here and leave detailed discussion of impact of the stochastic physics perturbation on the prediction of ensemble mean temperature, zonal wind and humidity in Sect. 3.3.

The MJO forecast skill shows seasonality (Fig. 2). Over the 2-year experimental period in this work, the MJO prediction skill in winter seasons (data collected during November, December and January of 2014–2015 and 2015–2016) are all higher than summer seasons (data collected during June, July and August of 2014 and 2015). The 0.5 anomaly correlation score in winter season is about 1–2 days higher than summer season except for the SPs + CFSBC + CNV experiment, which shows the winter score is more than 10 days higher than the summer score.

In addition to the MJO prediction skill that describes the model performance on MJO propagation, another two useful perspectives on MJO evaluation are the biases on

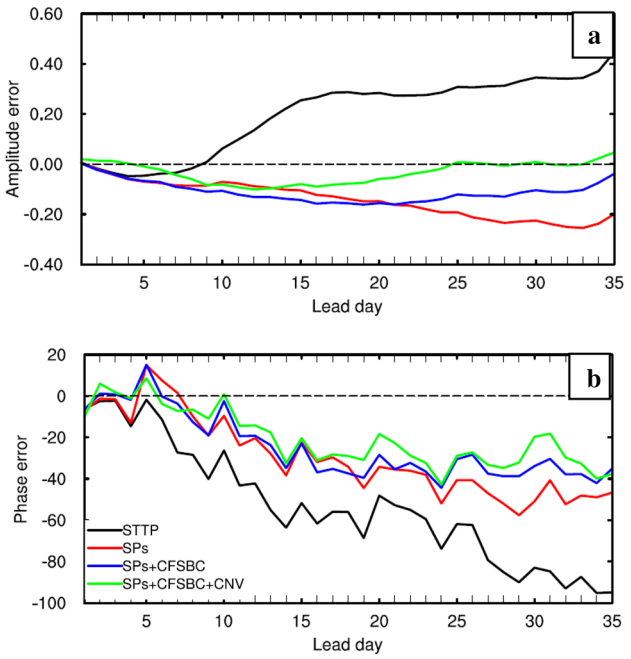

Fig. 3 MJO amplitude error (**a**) and phase error (**b**) as a function of lead days for the four experiments. The MJO amplitude error is defined as the mean bias of the MJO index with respect to the analysis data over the period of May 1st 2014-May 26 2016 (5 day-inteval). The MJO phase error is defined as the mean bias of the MJO phase angle against same analysis data over the same period. MJO phase is defined as $= tan^{-1}\left(\frac{RMM2}{RMM1}\right), \theta = [-180, 180]$

predicted MJO amplitude and phase. As in Vitart (2017), Fig. 3a, b show the evolution of the MJO amplitude and phase error as a function of lead time averaged over the experimental period. Over the 35 lead days, The MJO amplitude predicted by STTP is too strong and the MJO propagated too slowly compared to the analysis data. Opposite to the STTP, the MJO amplitude in SPs, SPs + CFSBC and SPs + CFSBC + CNV are predicted weak and the MJO still propagated slowly, which is common in many models (Wang et al. 2014; Rashid et al. 2011) but better than the STTP. For both MJO amplitude and phase speed, SPs + CFSBC + CNV has the least bias among all experiments. This is consistent with the MJO prediction skill in the same configuration showed in Fig. 1. Comparing the SPs + CFSBC to SPs, the SST forcing combined with the stochastic physics perturbation helped reduce the MJO amplitude and propagation bias for the lead time beyond 3 weeks and updating the convection scheme further reduced the MJO amplitude and phase bias.

Regarding the MJO prediction skill, the anomaly correlation coefficient (ACC) describes the forecast system's performance on the deterministic forecast. For an ensemble forecast system, to evaluate performance of the ensemble members in multiple categories, we used ranked probability skill score (RPSS) to demonstrate the performance of GEFS on the probabilistic forecast of the MJO. RPSS describes the squared error of the forecast system with respect to the analysis data in each category. Thus it is very sensitive to the distance between the forecast and the truth. To calculate the RPSS, we first calculated the RMM1 and RMM2 for each member for each experiment. Then we applied below RPSS formula to RMM1 and RMM2 to obtain the probability skill score for each of them.

For a certain initial date and forecast lead time, let Y_n denotes the cumulative probability in forecast RMM1 or RMM2 and A_n denotes the cumulative probability in the analysis. Based on Wilks (2011),

$$Y_n = \sum_j^n y_j\,, \quad \mathrm{n} = 1, \mathrm{N}, \tag{3}$$

and

$$A_n = \sum_j^n a_j, \quad \mathrm{n} = 1, \mathrm{N}. \tag{4}$$

In our case, since 20-member forecast were counted in the calculation of the probability score, $y_j$is thus the probability in each bin from the 20 member. We used N = 6 with bin range of [− 6, 6] for both RMM1 and RMM2. For example, if $y_1 = 0$, $y_2 = 0.1$, $y_3 = 0.4$, $y_4 = 0.2$, $y_5 = 0.2$, $y_6 = 0.1$, then $Y_1 = y_1 = 0$, $Y_2 = y_1 + y_2 = 0.1$, $Y_3 = y_1 + y_2 + y_3 = 0.5$, $y_4 = y_1 + y_2 + y_3 + y_4 = 0.7$, $Y_5 = y_1 + y_2 + y_3 + y_4 + y_5 = 0.9$ and $Y_6 = y_1 + \cdots + y_6 = 1$. If the analysis drops in the 3rd bin, then $a_3 = 1$ and for other bins $a = 0$. Thus, $A_1 = a_1 = 0$, $A_2 = a_1 + a_2 = 0$, $A_3 = a_1 + a_2 + a_3 = 1$, $A_4 = a_1 + a_2 + a_3 + a_4 = 1$, $A_5 = a_1 + \cdots + a_5 = 1$, and $A_6 = a_1 + \ldots + a_6 = 1$.

The ranked probability score (i.e. the sum of the squared difference between the cumulative forecast and analysis vectors) is then defined as:

$$RPS = \frac{1}{N}\sum_j^N \left(Y_n - A_n\right)^2. \tag{5}$$

To get the ranked probability skill score, we used the *RPS* of the climatology as the reference. Let

$$C_n = \sum_j^n c_j, \quad \mathrm{n} = 1,\ \mathrm{N}. \tag{6}$$

The RPS of the climatology data is defined as:

$$RPS_c = \frac{1}{N}\sum_j^N \left(C_n - A_n\right). \tag{7}$$

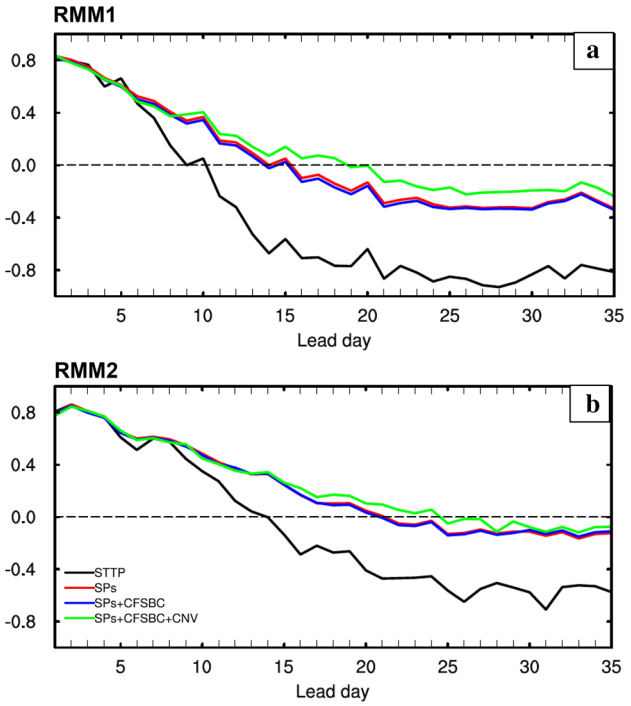

Fig. 4 Mean ranked probability skill score (RPSS) of RMM1 (**a**) and RMM2 (**b**) as a function of lead days averaged over the period of May 1 st, 2014–May 26, 2016. 6 bins ranged from −6 to 6 are used to calculate the score. For each lead day and initial date, 20 members were used to calculate the RPSS

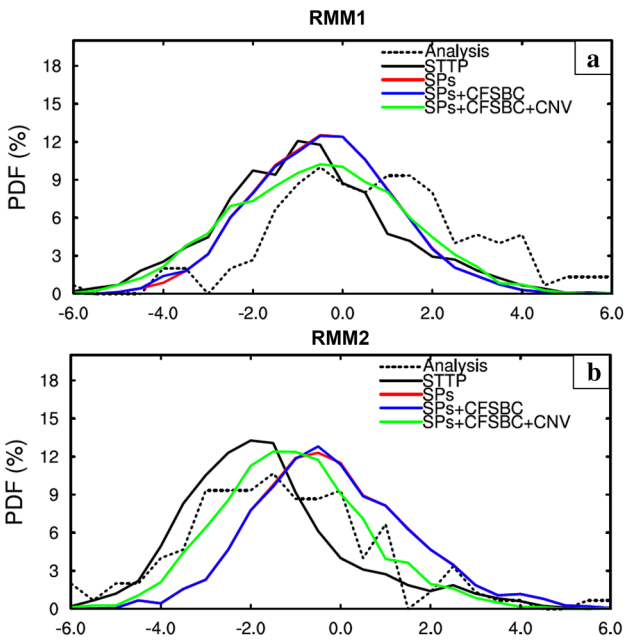

Fig. 5 Probability density distribution (PDF) of MJO RMM1 (**a**) and RMM2 (**b**) for the four experiments and analysis data for lead day 15. 25 bins are used to calculate the PDF. 150 samples are used to calculate the PDF of the analysis and 3000 samples (i.e. 150 initial days × 20 members) are used to calculate the PDF of the forecast data

The ranked probability skill score is then can be obtained using:

$$RPSS = 1 - \frac{RPS}{RPS_c}. \tag{8}$$

Different from the calculation of RPS for forecast, which is based on 20-member ensemble, RPS_c, is obtained from the 30-year (1988–2017) RMM indices from Bureau of Meteorology in Australia (http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt).

RPSS of the RMM1 and RMM2 as a function of lead day for each experiment averaged over the experimental period is shown in Fig. 4. Overall, probability forecast on RMM2 is more skillful than RMM1, which is consistent with the ensemble mean deterministic forecast. For

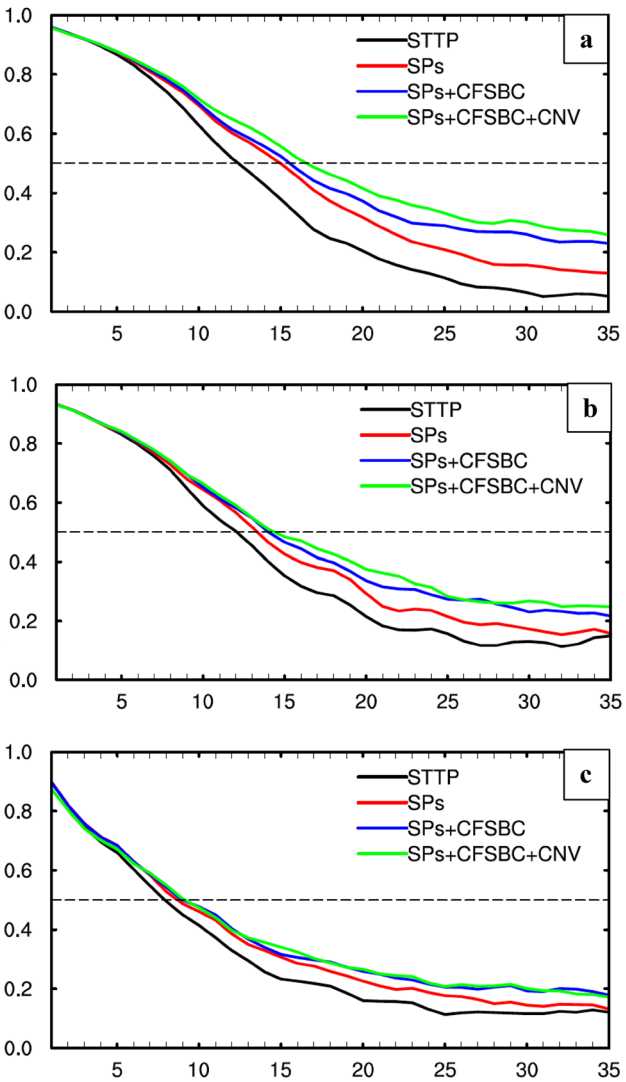

Fig. 6 Correlation coefficient of **a** U200, **b** U850 and **c** OLR anomalies over 15°S–15°N between the analysis and forecast data over the period of May 1st 2014–May 26 2016 for the four experiments

RPSS of both RMM1 and RMM2, experiments with new stochastic physics parameters (SPs, SPs + CFSBC and SPs + CFSBC + CNV) are all more skillful than STTP, with the SPs + CFSBC + CNV showing the best skill. The large difference between the STTP and new stochastic experiments is likely due to the RMMs bias in STTP. Figure 5 shows the probability distribution of the RMM1 and RMM2 in four experiments and analysis data. 20-ensemble in 150 initial dates are counted in the PDF for forecast data and 150 initial dates are counted in analysis data. For lead day 15, RMM1 and RMM2 of STTP from all ensemble members shift towards left in the PDF curve, indicating GEFS STTP predicted too much strong negative RMM1 and RMM2 (phase1 and 8 and phase 2 and 3) than other experiments and the analysis data. The PDF curve of SPPT + CFSBC + CNV matches the analysis data better than the other experiments, indicating that the SPPT + CFSBC + CNV predicted less weak RMM1 and RMM2.

3.2 Evaluation of the large-scale circulation and convection

Since the MJO skill using WH indices uses the upper and lower level zonal wind and the OLR anomaly to represent MJO associated circulation and convection, the prediction skill of the large-scale circulation and the convection are examined to help better understand the improvement in MJO skill. Figure 6 shows the correlation of the tropical 15°S–15°N mean zonal wind and OLR anomaly during the 2-year period (150 initial days) for all lead days and four experiments. In all experiments, the correlation score for upper level zonal wind is higher than the lower level zonal wind which is higher than the OLR. This is reasonable because lower level winds are more dependent on the model uncertainty in surface process than are the upper level winds. The OLR forecast shows the lowest score, likely because it is a function of more complicated physical and dynamical processes that involves considerable uncertainty thus limiting the prediction skill. For all three components, it is obvious that all the new stochastic experiments (including SPs, SPs + CFSBC and SPs + CFSBC + CNV) performed better than the production version of the GEFS (STTP), indicating the positive impact of the updated stochastic scheme on the performance of the forecast system over the tropical region. Among the SPs experiments, SPs + CFSBC + CNV outperformed SPs + CFSBC which outperformed SPs in wind components, especially in 200 hPa zonal wind. Although there are some differences between the SPs and the other two experiments (i.e. SPs + CFSBC and SPs + CFSBC + CNV) in OLR for longer lead time, the difference between the SPs + CFSBC and SPs + CFSBC + CNV is not as evident as in the zonal wind fields.

To further explore where in the tropics does the substantial improvement occur, we examined correlation of the wind and OLR anomaly on each model grid over the

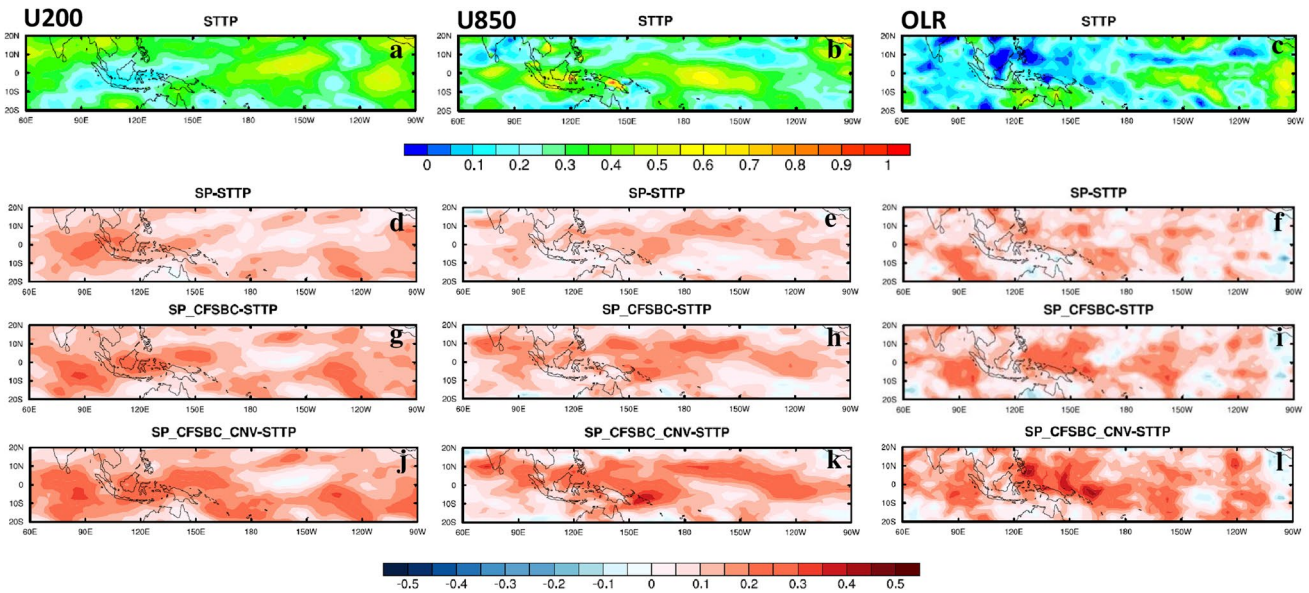

Fig. 7 Spatial distribution of the correlation coefficient of U200, U850 and OLR anomaly between the analysis and the forecast data during the period of May 1st 2014–May 26 2016 at lead day 15. **a–c** Correlation coefficient for STTP, **d–l** the difference of the correlation between each experiment and the STTP. The sample size to calculate the correlation coefficient is 150 for each plot

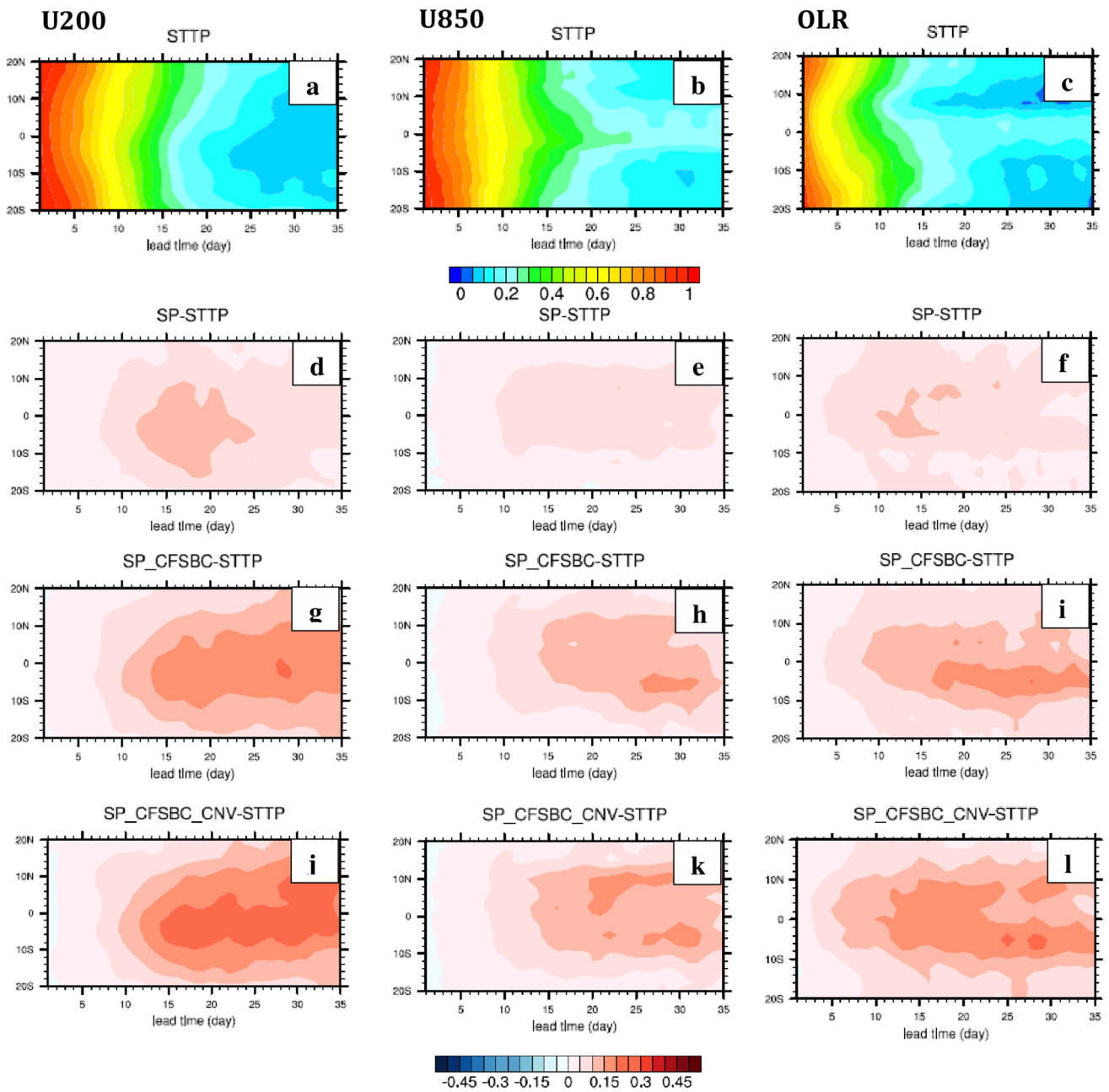

Fig. 8 Correlation coefficient as a function of latitude and lead time for U200, U850 and OLR anomaly for STTP (**a**–**c**) and the difference of the correlation coefficient of each experiment from the STTP (**d**–**l**) for each variable

tropics for lead day 15 (Fig. 7). The operational version of GEFS shows better performance on 850 hPa zonal wind anomaly forecast than the other two variables over the region from tropical Indian Ocean to west Pacific (Fig. 7a–c). The updated stochastic schemes help improved the 200 hPa zonal wind anomaly over the tropical Indian Ocean to most of the Maritime Continent, as well as the 850 hPa zonal wind anomaly over the tropical west Pacific and Indian Ocean (Fig. 7d–f). In the WH RMM calculation, the zonal wind is weighted more than the OLR anomaly. Although the updated stochastic schemes also resulted in an improvement of the OLR anomaly forecast

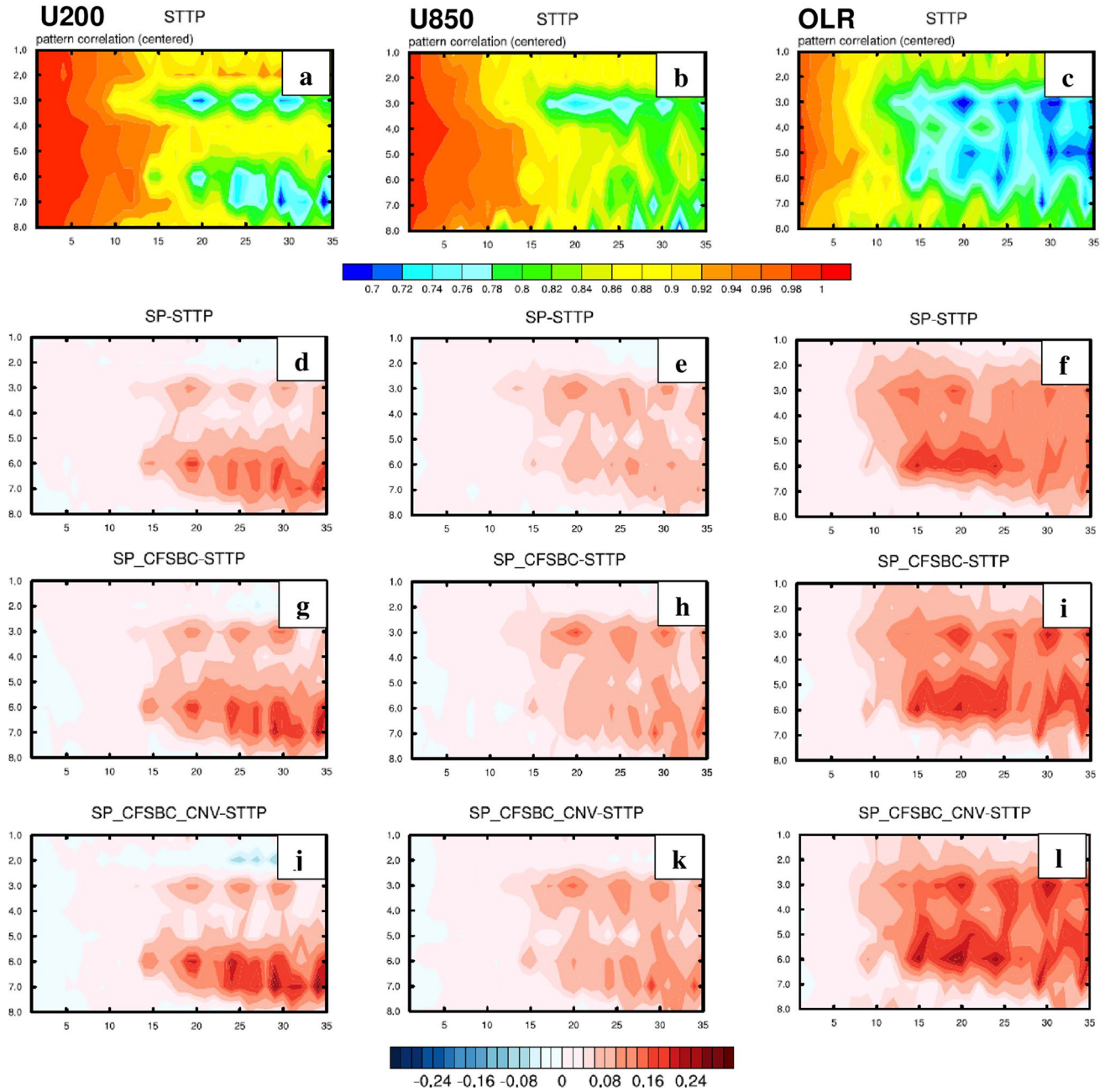

Fig. 9 Pattern correlation as a function of MJO phase and lead time for the composite U200, U850 and OLR over each MJO phase between the analysis and forecast data. **a–c** Pattern correlation for STTP experiment for the U200, U850 and OLR respectively. **d–l** the difference of the correlation coefficient of each experiment from the STTP (**d–l**) for each variable

over the tropical region, the improvement in the wind component appears to be the main reason that related to the improved MJO forecast in the updated stochastic schemes. A further update on underlying SST, combined with the updated stochastic schemes enhanced the improvement of the three variables. Updated the stochastic scheme, the underlying SST, and the convection scheme provided additional skill to the wind and OLR prediction (Fig. 7i–l). The improvement of the forecast anomaly of the three variables can also be demonstrated in a latitude and lead time cross-section of the three variables (Fig. 8). The SPs + CFSBC + CNV is the configuration that leads

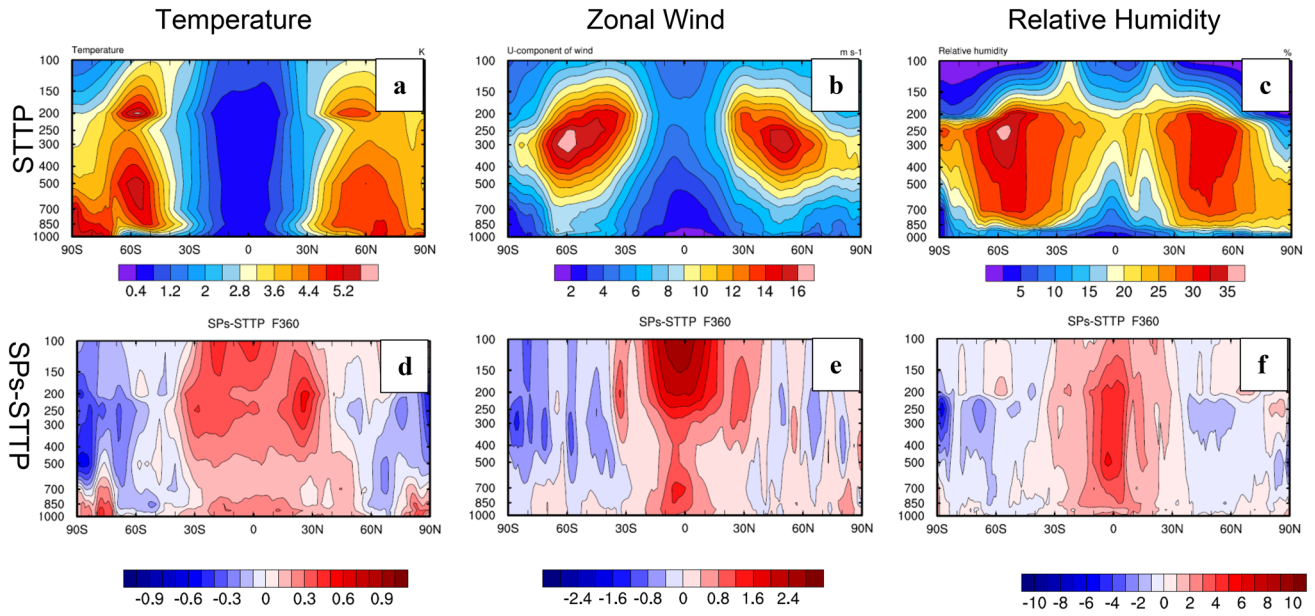

Fig. 10 Averaged ensemble spread of the perturbed members in GEFS for temperature, zonal wind and relative humidity at 360 forecast hour (**a**–**c**) and the difference between SPs and STTP for the corresponding variables (**d**–**f**). For each plot, 6 sample during March 2016 (March 1, 6, 11, 16, 21 and 26) was used to calculate the averaged ensemble spread. **b** and **e** are from Fig.7a and b in Zhu et al. 2018

to the largest improvement. Among the three variables, the largest improvement occurs in 200 hPa zonal wind, especially beyond two weeks of the lead time. The conclusion that of these tested here, the SPs + CFSBC + CNV is the best configuration for MJO forecasts can also be substantiated by a MJO phase and lead time cross-section of the pattern correlation of the composite 200 hPa, 850 hPa zonal wind and OLR fields (Fig. 9). For all three variables, the large improvement occurs mostly on phase 3 and phase 6–7 after 2 weeks of the lead time. This is consistent with the improvement of the forecast over tropical west Pacific and Indian Ocean shown in Fig. 7 and with other works (Wang et al. 2014).

3.3 Impact of the stochastic physics scheme

Based on the evaluation of the prediction skill of the MJO and its associated variables (Figs. 1, 4), the prediction skill exhibits a jump after using an updated stochastic scheme. As such, we examined the spread of the perturbed variables in the SPs experiment to further check the contribution of the stochastic scheme on the performance of the GEFS. The SPs is a combined stochastic physics scheme of three components, which include SPPT, SHUM and SKEB. The individual impact of the SPs has been discussed in Zhu et al. 2017. The combined effect of the SPs, compared to the operational STTP version, increased the spread mainly over the tropical region (Fig. 10) and the increase of the spread lead to an improved temperature, zonal wind and relative humidity profile in the same region (Fig. 11). For the weeks 3 and 4 forecast of the tropical mean zonal wind at 250 hPa and 850 hPa, the prediction skill improved by 43% and 19% respectively (Fig. 12). The weeks 3 and 4 prediction skill of the tropical zonal wind is strongly dependent on the initial dates. The skill range can be as large as 1 for the zonal wind at both levels. The prediction skill of the weeks 3 and 4 tropical mean zonal wind at 250 hPa and 850 hPa during a strong MJO period is 63% and 18% larger than that of the weak MJO period in SPs experiment and 50% and 12% larger in STTP experiment.

4 Conclusion and discussion

Improving prediction skill for the source of the subseasonal predictability (i.e, the MJO) is critical to the improvement of tropical and extratropical forecasts of other phenomena, especially for extreme events. In this work, experiments were performed using different

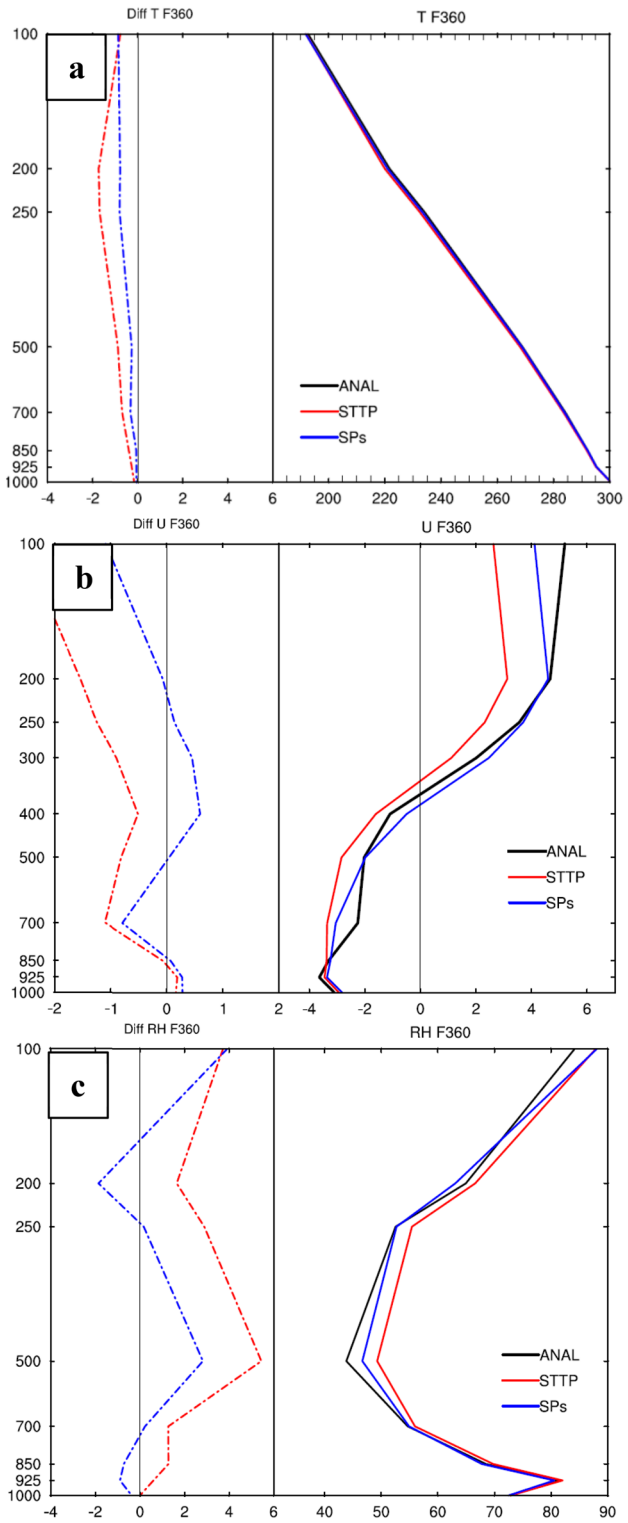

Fig. 11 The 360-h forecast of the tropical mean (15°S–15°N, 0°–360°E) temperature, zonal wind and relative humidity in Analysis, STTP and SPs (right panel of each plot) and the bias of each experiment (left panel of each plot). The samples used for this plot is same as in Fig. 10, i.e. March 2016 (March 1, 6, 11, 16, 21 and 26)

configurations of the NCEP GEFS to evaluate the MJO prediction skill and its key components. Based on the four experiments, it was determined that the improved-MJO forecast benefited largely from the use of a stochastic physics scheme that resulted in the improvement of the forecast of key MJO component variables. The improvement was mostly related to the increase of the spread of the temperature, wind and humidity profiles over the tropics. The addition of an updated underlying SST from the bias corrected CFSv2 forecast and a new convection scheme simulates tropical convective systems more realistically further enhanced the improvement. The improvement of the MJO prediction skill is mainly related to better representation of the circulation and convection over the tropical West Pacific and Indian Ocean, especially in the configuration of SPs + CFSBC + CNV. A combination of the updated stochastic physics perturbation, more realistic SST, and convection scheme in the forecast model led to about a 9-day extension of given skill threshold for MJO forecasts.

Following this work, our next step will be the evaluation for the MJO-skill impacts after bias-correction. Application of this configuration is enabled by the recent completion and post processing of an 18-year reforecast. In addition to the evaluation of tropical forecasts, an analysis of the teleconnection relation between tropical and extratropical prediction skill is necessary and critical. The current investigation suggests that there is a statistically significant lag correlation between the North Atlantic Oscillation (NAO) and MJO indices, which encourages further analysis on this direction. The extreme events associated with the MJO and other sources of predictability for sub-seasonal time scale are also included in plans for future studies.

Linking the prediction skill to the configuration of the forecast system, model developers and researchers may also be interested in isolating the impact of the stochastic scheme, the underlying SST, and the convection scheme. Related to the computational cost, the more granular sensitivity tests were not conducted as part of this work. Rather, the design of each configuration used herein was based on the positive impact of each factor (SPs, SST and convection) derived from small sample tests and early investigation (Zhu et al. 2018). A more comprehensive investigation can be left to other researchers who have interest.

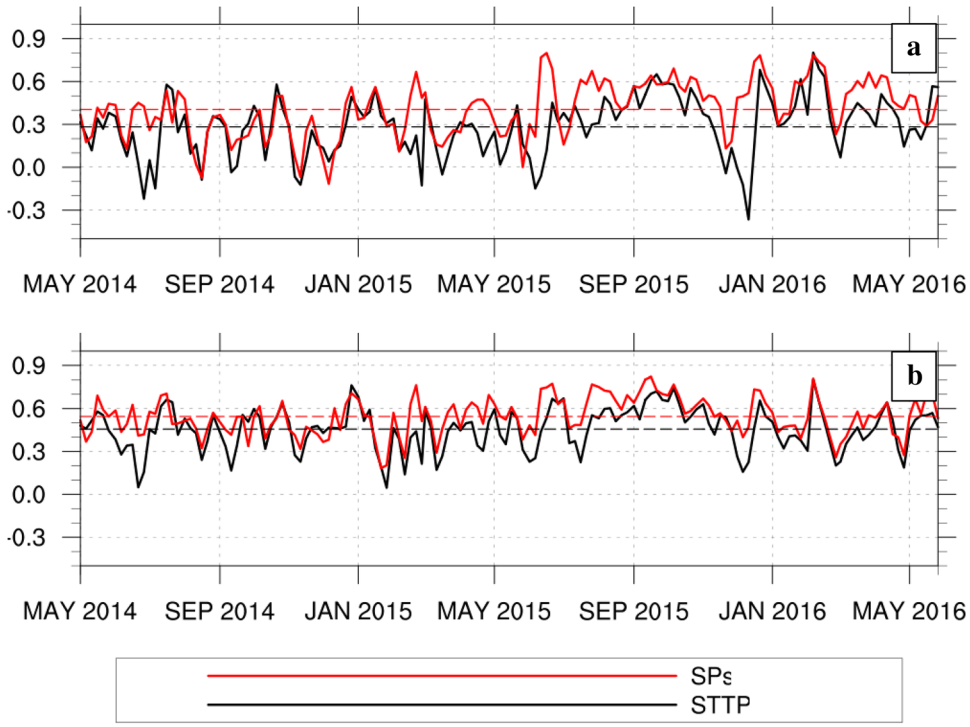

Fig. 12 The ensemble mean anomaly correlation of the week 3 and 4 averaged 250 hPa (**a**) and 850 hPa (**b**) zonal winds as a function of initial time for the STTP and SPs. The average score for 250 hPa zonal wind (**a**) is 0.404 and 0.283 for SPs and STTP. The average score for 850 hPa zonal wind (**b**) is 0.545 and 0.457 for SPs and STTP respectively

Acknowledgements The authors would like to thank Drs. Walter Kolczynski and Bing Fu for their help on stochastic physics perturbation settings. We really appreciate Dr. Shuguang Wang for the valuable discussion and constructive suggestion on the figures. We thank Drs. Qin Zhang, Wanqiu Wang and Ping Liu for providing valuable discussion on MJO. The authors are grateful to Dr. Xingren Wu on SST experiment discussion. We also thank Drs. Partha Bhattacharjee and Jack Kain for the EMC internal review and the two anonymous reviewers for their insightful comments that helped to improve this manuscript. This study is partially supported through NWS OSTI and NOAA's Climate Program Office (CPO)'s Modeling, Analysis, Predictions, and Projections (MAPP) program.

References

Arakawa A, Schubert WH (1974) Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. J Atmos Sci 31(3):674–701

Berner J, Shutts GJ, Leutbecher M, Palmer TN (2009) A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. J Atmos Sci 66(3):603–626

Buizza R, Milleer M, Palmer T (1999) Stochastic representation of model uncertainties in 504 the ECMWF ensemble prediction system. Q J R Meteorol Soc 125(560):2887–2908

Ferranti L, Palmer TN, Molteni F, Klinker K (1990) Tropical-extratropical interaction associated with the 30–60 day oscillation and its impact on medium and extended range prediction. J Atmos Sci 47:2177–2199

Fu X, Lee J-Y, Hsu P-C, Taniguchi H, Wang B, Wang W, Weaver S (2013) Multi-model MJO forecasting during DYNAMO/CINDY period. Clim Dyn 41:1067–1081. https://doi.org/10.1007/s00382-013-1859-9

Gemmill W, Katz B, Li X (2007) Daily real-time, global sea surface temperature-high-resolution analysis: RTG_SST_HR. NCEP Office Note, #260, pp 1–39. http://polar.ncep.noaa.gov/mmab/papers/tn260/MMAB260.pdf

Gottschalck J et al (2010) A framework for assessing operational Madden–Julian oscillation forecasts: a CLIVAR MJO working group project. Bull Am Meteor Soc 91:1247–1258. https://doi.org/10.1175/2010BAMS2816.1

Grell G (1993) Prognostic evaluation of assumptions used by cumulus parameterizations. Mon Weather Rev 121(3):764–787 (**2016/03/25**)

Han J, Pan H-L (2011) Revision of convection and vertical diffusion schemes in the ncep global forecast system. Weather Forecast 26(4):520–533 (**2016/03/25**)

Han J, Wang W, Kwon Y, Hong S, Tallapragada V, Yang F (2017) Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. Weather Forecast 32:2005–2017

Hou D, Toth Z, Zhu Y (2006) A stochastic parameterization scheme within NCEP Global Ensemble Forecast System. In: Preprints, 18th Conf. on Probability and Statistics, Atlanta, GA, Amer. Meteor. Soc., p 4.5

Hou D, Toth Z, Zhu Y, Yang W (2008) Impact of a stochastic perturbation scheme on NCEP Global Ensemble Forecast System. In: Preprints, 19th Conf. on Probability and Statistics, New Orleans, LA, Amer. Meteor. Soc., p 1.1

Hudson D, Marshall AG, Yin Y, Alves O, Hendon H (2013) Improving intraseasonal prediction with a new ensemble generation strategy. Mon Weather Rev 141:4429–4449. https://doi.org/10.1175/MWR-D-13-00059.1

Johnson NC, Collins D, Feldstein S, L'Heureux M, Riddle E (2014) Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. Weather Forecast 29:23–38. https://doi.org/10.1175/WAF-D-13-00102.1

Leutbecher M, Lock SJ, Ollinaho P, Lang STK, Balsamo G, Bechtold P, Bonavita M, Christensen HM, Diamantakis M, Dutra E, English S, Fisher M, Forbes RM, Goddard J, Haiden T, Hogan RJ, Juricke

S, Lawrence H, MacLeod D, Magnusson L, Malardel S, Massart S, Sandu I, Smolarkiewicz PK, Subramanian A, Vitart F, Wedi N, Weisheimer A (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. Q J R Meteorol Soc 143(707):2315–2339. https://doi.org/10.1002/qj.3094

Li C, Jia X, Ling J, Zhou W, Zhang C (2009) Sensitivity of MJO simulations to convective heating profiles. Clim Dyn 32:167–187

Lin H, Brunet G (2009) The influence of the Madden–Julian oscillation on Canadian wintertime surface air temperature. Mon Weather Rev 137:2250–2262

Lin JL, Lee MI, Kim D, Kang IS, Frierson DMW (2008) The impacts of convective parameterization and moisture triggering on AGCM-simulated convectively coupled equatorial waves. J Clim 21:883–909

Liu P, Wang B, Sperber KR, Li T, Meehl GA (2005) MJO in the NCAR CAM2 with the Tiedtke convective scheme. J Clim 18:3007–3020. https://doi.org/10.1175/JCLI3458.1

Liu X et al (2016) MJO prediction using the sub-seasonal to seasonal forecast model of Beijing Climate Center. Clim Dyn. https://doi.org/10.1007/s00382-016-3264-7

Liu et al (2017) MJO prediction using the sub-seasonal to seasonal forecast model of Beijing Climate Center. Clim Dyn 48:3283–3307

Ma J, Zhu Y, Wobus D, Wang P (2012) An effective configuration of ensemble size and horizontal resolution for the NCEP GEFS. Adv Atmos Sci 29(4):782–794

Maloney ED, Hartmann DL (2001) The sensitivity of intraseasonal variability in the NCAR CCM3 to changes in convective parameterization. J Clim 14:2015–2034

Neena JM, Lee JY, Waliser D, Wang B, Jiang X (2014) Predictability of the Madden–Julian oscillation in the intraseasonal variability hindcast experiment (ISVHE). J Clim 27(1):4531–4543. https://doi.org/10.1175/JCLI-D-13-00624

Ollinaho P, Lock S-J, Leutbecher M, Bechtold P, Beljaars A, Bozzo A, Forbes RM, Haiden T, Hogan RJ, Sandu I (2017) Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. Q J R Meteorol Soc 143:408–422. https://doi.org/10.1002/qj.2931

Palmer T, Buizza NR, Doblas-Reyes F, Jung T, Leutbecher M, Shutts G, Steinheimer M, Weisheimer A (2009) Stochastic parametrization and model uncertainty. Tech RepECMWF RD Tech Memo, vol 598, p 42. http://www.ecmwf.int/publications/ (**Available online**)

Pan HL, Wu W-S (1995) Implementing a mass flux convection parameterization package for the nmc medium-range forecast model. NMC Office Note 409:40

Pegion K, Sardeshmukh PD (2011) Prospects for improving subseasonal predictions. Mon Weather Rev 139:3648–3666

Ramsay BH (1998) The interactive multisensor snow and ice mapping system. Hydrol Process 12:1537–1546

Rashid HA, Hendon HH, Wheeler MC, Alves O (2011) Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. Clim Dyn 36:649–661. https://doi.org/10.1007/s00382-010-0754-x

Saha S et al (2014) The NCEP climate forecast system version 2. J Clim 27:2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Shutts G (2005) A kinetic energy backscatter algorithm for use in ensemble prediction systems. Q J R Meteorol Soc 131:3079–3102

Shutts G, Palmer TN (2004) The use of high-resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. In: Proc. ECMWF/CLIVAR Simulation and Prediction of Intra-Seasonal Variability with Emphasis on the MJO, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, pp 83–102

Tian DW, Eric, Yuan X (2017) CFSv2-based sub-seasonal precipitation and temperature prediction skill over the contiguous United States. Hydrol Earth Syst Sci 21:1477–1490

Tompkins AM, Berner J (2008) A stochastic convective approach to account for model uncertainty related to unresolved humidity variability. J Geophys Res 113:D18101

Tracton MS, Kalnay E (1993) Operational ensemble prediction at the National Meteorological Center: practical aspects. Weather Forecast 8:379–398

Troccoli A (2010) Seasonal climate forecasting. Meteorol Appl 17:251–268. https://doi.org/10.1002/met.184

Vitart F (2009) Impact of the Madden Julian oscillation on tropical storms and risk of landfall in the ECMWF forecast system. Geophys Res Lett 36:L15802. https://doi.org/10.1029/2009GL039089

Vitart F (2014) Evolution of ECMWF sub-seasonal prediction skill scores. Q J R Meteor Soc 140:1889–1899. https://doi.org/10.1002/qj.2256

Vitart F (2017) Madden–Julian oscillation prediction and teleconnections in the S2S database. Q J R Meteor Soc 143:2210–2220. https://doi.org/10.1002/qj.3079

Vitart F, Molteni F (2010) Simulation of the Madden–Julian oscillation and its teleconnections in the ECMWF forecast system. Q J R Meteor Soc 136:842–855

Waliser DE, Lau KM, Stern W, Jones C (2003) Potential predictability of the Madden–Julian oscillation. Bull Am Meteorol Soc 84:33–50

Wang W, Schlesinger ME (1999) The dependence on convection parameterization of the tropical intraseasonal oscillation simulated by the UIUC 11-layer atmospheric GCM. J Clim 12:1423–1457

Wang W, Hung M-P, Weaver SJ, Kumar A, Fu X (2014) MJO prediction in the NCEP climate forecast system version 2. Clim Dyn 42:2509–2520. https://doi.org/10.1007/s00382-013-1806-9

Wang W, Kumar A, Fu JX, Hung M-P (2015) What is the role of the sea surface temperature uncertainty in the prediction of tropical convection associated with the MJO? Mon Weather Rev 143:3156–3175. https://doi.org/10.1175/MWR-D-14-00385.1

Wheeler MC, Hendon HH (2004) An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. Mon Weather Rev 132:1917–1932

Wilks DS (2011) Statistical methods in the atmospheric sciences, 3rd edn. Academic Press, Oxford, Waltham

Xiang B, Zhao M, Jiang X, Lin S-J, Li T, Fu X, Vecchi G (2015) 3–4 week MJO prediction skill in a GFDL Coupled model. J Clim. https://doi.org/10.1175/JCLI-D-15-0102.1

Zhang GJ, Song X (2009) Interaction of deep and shallow convection is key to Madden–Julian oscillation simulation. Geophys Res Lett 36:L09708. https://doi.org/10.1029/2009GL037340

Zhou L, Neale R, Jochum M, Murtugudde R (2012) Improved Madden–Julian oscillations with improved physics: the impact of modified convection parameterizations. J Clim 25:1116–1136

Zhou X, Zhu Y, Hou D, Luo Y, Peng J, Wobus R (2017) Performance of the new NCEP global ensemble forecast system in a parallel experiment. Weather Forecast. https://doi.org/10.1175/WAF-D-17-0023.1

Zhu Y, Zhou X, Peña M, Li W, Melhauser C, Hou D (2017) Impact of sea surface temperature forcing on weeks 3 & 4 prediction skill in the NCEP global ensemble forecasting system. Weather Forecast 32:2159–2174

Zhu Y, Zhou X, Li W, Hou D, Melhauser C, Sinsky E, Peña M, Fu B, Guan H, Kolczynski W, Tallapragada V (2018) Toward the improvement of subseasonal prediction in the National Centers for environmental prediction global ensemble forecast system, J Geo Res. https://doi.org/10.1029/2018JD028506