An Assessment of Subseasonal Forecast Skill Using an Extended Global Ensemble Forecast System (GEFS)

Yuejian Zhu [*1], Xiaqiong Zhou [2], Wei Li [2], Dingchen Hou [1], Christopher Melhauser [2], Eric Sinsky [2], Malaquias Pena [2], Bing Fu [2], Hong Guan [3], Walter Kolczynski [2] Richard Wobus [2] and Vijay Tallapragada [1]

[1]EMC, NCEP, NWS, NOAA, College Park, Maryland

[2]IMSG at EMC, NCEP, NWS, NOAA, College Park, Maryland

[3]SRG at EMC, NCEP, NWS, NOAA, College Park, Maryland

Version 3.0

Update: 9/7/2017

In preparation to be submitted to Journal of Climate

*Corresponding Author: Yuejian Zhu, Email: Yuejian.Zhu@noaa.gov, Environmental Modeling Center/NCEP/NOAA, 5830 University Research Court, College Park, MD 20740

Abstract

In order to provide ensemble based subseasonal (weeks 3 & 4) forecasts to support NCEP CPC's operational mission, experiments have been designed through the SubX project to investigate potential predictability in both tropical and extratropical regions. The control experiment is the current operational GEFS version 11 extended from 16 days to 35 days. In addition to the control, parallel experiments have been designed to focus on three areas: 1) improving forecast uncertainty representation for the tropics through stochastic physical perturbations; 2) considering the impact of the ocean by using a 2-tiered SST approach; and 3) testing a new scale aware convection scheme to improve model physics for tropical convection and MJO forecasts. All experiments are initialized every 5 days at 0000 UTC during the period of May 2014 - May 2016 (25 months).

In the tropics, MJO forecast skill has been improved from an average of 12.5 days (control) to nearly 22 days by combining all three modifications to GEFS. For the experiment with the best overall score, a skill of 23 days could be reached for a strong MJO period. In the extratropics, anomaly correlation (AC) of 500 hPa geopotential height for the ensemble mean improved over weeks 3 & 4. In addition, CRPS of the Northern Hemisphere raw surface temperature (land only) improved as well. A similar result has been found for CONUS precipitation, although forecast skill is extremely low. Our results suggest that calibration may be important and necessary for surface temperature and precipitation forecast for the subseasonal time scale due to the large systematic model errors.

1. Introduction

The National Oceanic and Atmospheric Administration (NOAA) is working on a unified modeling system through the NWS Next Generation Global Prediction System (NGGPS) project (NOAA/NWS, 2014) to provide the best possible guidance to a wide customer base, including emergency managers, forecasters, and the aviation community, to protect life and property, and enhance the national economy. The dominant factors for short term (7 to 10 days) to extended range (multi-weeks) forecasts differ. A numerical weather prediction system (and/or ensemble forecast system) heavily relies on atmospheric initial conditions (including initial uncertainties) and model parameterizations for a short range forecast. As forecast lead time increases, the impact of the ocean and other external forcing in the Earth system become important, thus they cannot be neglected (Li et al. 2009, Liu et al. 2016). Various studies have shown that the oceanic variation at the subseasonal time scale is primarily related to thermodynamic forcing (Li et al., 2001; Ling et al., 2015). The impacts of large-scale ocean currents are of secondary importance (Takaya et al., 2010).

Subseasonal forecasts span the time period between weather and seasonal (climate) forecasts. Currently, there are no optimal configurations of numerical weather or climate models that can provide skillful forecast covering the subseasonal time scale. With the ultimate goal to improve forecast skill and deliver useful numerical guidance for subseasonal time scales, we explore the potential forecast skill of an extended Global Ensemble Forecasting System (GEFS) covering the subseasonal time scale. An early effort has already shown positive results when extending the GEFS and updating the underlying sea surface temperature (SST) boundary conditions (Zhu et al, 2017).

Based on the length of forecast Northern Hemisphere (NH; 20oN-80oN) 500 hPa geopotential height anomaly correlation (AC) exceeding 60%, the operational GEFS has provided more skillful numerical guidance compared to the operational high resolution deterministic Global Forecast System (GFS) (Fig. 1). The GEFS ensemble mean improves from 8.92 days of skillful forecast (AC >60%) in 2014 (GEFS v10) to 10.52 days in 2016 (GEFS v11) in contrast to the GFS that only improves from 7.9 days in 2014 to 8.45 days in 2016. Improving the skill of numerical guidance is achieved year by year through model upgrades that include enhancing data assimilation, model physics, initial perturbations, stochastic physics perturbations, and model resolution.

In contrast to current seasonal forecasting systems, there are several advantages in extending GEFS to cover the subseasonal time scale, including 1) improved initial perturbations using an ensemble Kalman filter (EnKF) data assimilation system (Zhou et al, 2017) which represents observation and analysis uncertainties; 2) increased horizontal resolution from weather into the subseasonal time scales allowing small scale process to be resolved and more realistic interactions between scales; 3) advanced model physics with various stochastic physics perturbation schemes to represent model uncertainties; 4) increased ensemble size (i, e, GEFS currently runs 80+4 members for one synoptic day) to provide more reliable probabilistic guidance; 5) suitable configuration (ensemble size and frequency) for real time reforecasts/hindcasts for calibration; and 6) seamless forecasts across weather and seasonal time scale.

Based on the performance of GEFS for the extended range forecast in our early investigation on SST (Zhu et al. 2017), this study builds on those results and provides a comprehensive investigation taking into account key factors that may contribute to improving forecast skill in the extended range for GEFS. In this investigation, the current National Centers of Environmental Prediction (NCEP) operational GEFS v11 (Zhou et. al, 2017) is used as the base configuration. The goal is to focus and improve our common understanding in three different physical processes (or scientific areas). The first area is to improve tropical forecast uncertainty. Unlike extratropical baroclinic systems, the tropics exhibit less potential error growth modes from optimal initial perturbations (Toth and Kalnay, 1993; 1997). The lack of error growth in the tropics is especially pronounced in the current GEFS due to its under dispersion (Zhou et al, 2017). To improve upon this deficiency, we use various stochastic physical perturbation schemes to represent tropical forecast uncertainty. The second area is to more realistically consider the day-to-day variability of the underlying SST using a two-tiered SST approach. This allows for a one way interaction between the ocean and atmosphere which may be an optimal strategy in representing the impact of ocean forcing (Melhauser et al, 2016; Zhu et al, 2017). The third area is to test a new convection scheme with a scale-aware parameterization that can more realistically represent the vertical motion, radiative impact, and cloud and precipitation processes associated with tropical convection (Han et al, 2017).

The experiment design and configurations will be described in Section 2. The forecast skill will be evaluated in Section 3 and an evaluation of ensemble size will be discussed in Section 4. Summary and conclusions along with future work will be presented in Section 5.

2. Experiment design and configurations

In this paper, we test three configurations to explore the forecast skill of GEFS on subseasonal prediction. In the design of each experiment configuration, we combined configuration changes based on early investigations on the effect of some of the configuration changes (Melhauser et al, 2016; Zhu et al, 2017, Han et al. 2017). Although it is useful to independently examine the impact of each configuration change for a full experiment period, running these permutations would be too computationally expensive with a high resolution GEFS and 21 ensemble members for the full experiment period.

2.1 Operational GEFS v11 and extended forecast - "CTL"

The operational GEFS v11 was implemented on 2 December 2015. It uses a reduced horizontal resolution version of the NCEP GFS Global Spectral Model (with Semi-Lagrangian dynamics) v12.0 (GSM). GEFS is initialized four times per day (0000, 0600, 1200 and 1800 UTC) and integrated to 16 days (Sela 1980; Han and Pan 2011; Han et al. 2016). The horizontal resolution is approximately 34 km for days 0-8 and 52 km for days 8-16 with 64 hybrid vertical levels. More details of GEFS v11 can be found in Zhou et al. (2017), Melhauser et al. (2016), and Zhu et al. (2017). For the extended GEFS forecast in this study, the horizontal resolution of forecast lead days 16-35 is the same as day 8-16 (about 52 km). In addition, the GEFS uses the same SST forcing as the GFS, which is initialized with the Real Time Global (RTG) SST analysis (Gemmill et al, 2008) and damped to analysis climatology (90-d e-folding, Melhauser et al, 2016; Zhu et al, 2017) during model integration. The sea ice concentration is initialized from the daily 0000 UTC sea ice analysis from the Interactive Multisensor Snow and Ice Mapping System (Ramsay 1998).

2.2 New stochastic perturbed physics schemes - "SPs"

The current operational GEFS uses the Stochastic Total Tendency Perturbation (STTP; Hou et al, 2006; 2008) to account for random model error. STTP has been in operations since GEFS v9 which was implemented in 2010. The impact of STTP is largely constrained to the extratropical region and the boreal winter; there are only minor impacts in the tropical region. A set of stochastic perturbed physical schemes have been implemented in the EnKF component of the NCEP hybrid data assimilation system to provide optimal estimates of the background error covariance. These include the 1) Stochastic Kinetic Energy Backscatter (SKEB; Shutts and Palmer, 2004; Shutts 2005; Berner et al., 2009) scheme; 2) Stochastically Perturbed Parametrization Tendencies (SPPT; Buizza et al 1999; Palmer et al, 2009) scheme; and 3) Stochastic Perturbed Humidity (SHUM; Tompkins and Berner, 2008) scheme.

For this study, the SPPT scheme uses spatial and temporal random patterns comprised of five horizontal spatial/temporal scale combinations to perturb the tendency of wind, temperature and water vapour; the SHUM scheme takes single random scale pattern to perturb lower level relative humidity only; the SKEB scheme only considers numerical dissipation (diffusion). An example of the individual independent scale random patterns and combined 5-scale random pattern is shown in Fig. 2 (values of sigma represent maximum amplitude). The impacts of the different individual stochastic schemes are shown in Fig. 3. The difference in ensemble spread between ensembles with the various stochastic perturbations and an ensemble with no stochastic perturbations are provided at a forecast lead time of 144 hours (6 days) for an average of 6 spring cases and 6 fall cases. In our application of SKEB, this scheme counteracts the dissipation of kinetic energy mainly due to horizontal diffusion. The contributions of the STTP and SKEB are similar (upper-middle and middle rows of Fig. 3) in the extratropics, both showing small differences relative to CTL. The same stochastic pattern as SPPT is used for SHUM, but only for first spatial/temporal scale (top-middle of Fig. 2) and the lower level relative humidity. In summary, these stochastic schemes incorporated into the GEFS GSM generally improve tropical forecast uncertainty, mainly from the inclusion of SPPT and SHUM, which may lead to improvements in tropical forecast skill.

2.3 Two tiered SST approach - "SPs+SST_bc"

In addition to the new stochastic perturbed physics schemes discussed in Section 2.2, this experiment uses a two-tiered SST approach for lower boundary conditions over ocean by considering an evolving ocean SST state with gradually increasing with lead time. The two-tiered SST approach relaxes the RTG SST analysis to a bias corrected operational Climate Forecast System (CFS) v2 (Saha et al, 2014) predicted SST from the latest CFS forecast rather than the climatological SST in the control GSM configuration and updated every 24 hours during model integration (Melhauser et al, 2016; appendix I of Zhu et al, 2017). There are two advantages from a two-tiered SST approach, 1) the lower boundary of atmosphere model receives evolving SSTs with smaller scale details which are bias corrected to remove systematic difference from CFS v2 and 2) the computational resources are significantly reduced since an ocean and sea ice model are not required.

However, it is not a true coupled approach, but the impact of using predicted ocean variation from the fully coupled system (CFS v2) can be determined through various evaluations, such as the Madden-Julian Oscillation (MJO) skill and other tropical low-level forecast elements of the extended GEFS forecast. Meanwhile, work is ongoing at NCEP to produce a fully coupled GEFS (atmosphere, ocean, sea ice, wave, and aerosol).

2.4 New scale aware convective scheme - "SPs+SST_bc+SA_CV"

In additional to the new stochastic physics experiment (SPs) and two-tiered SST approach (SPs+SST_bc), a final experiment tests an upgraded Simplified Arakawa-Schubert

(SAS) cumulus parameterization scheme, which is both scale- and aerosol-aware (Han et al, 2017). The upgraded scheme is used in GFS v14 implemented on 19 July 2017.

The main changes in the upgraded scheme include: 1) the rain conversion rate decreases with decreasing air temperature above the freezing level; 2) convective adjustment time in deep convection is proportional to the convective turnover time, with convective available potential energy (CAPE) approaching zero after the adjustment time; 3) cloud base mass flux in the shallow convection scheme is now a function of mean updraft velocity; 4) convective inhibition (CIN) in the sub-cloud layer is an additional trigger condition to suppress unrealistic spotty rainfall, especially over high terrain during the summer; and 5) convective cloudiness is enhanced by suspended cloud condensate in an updraft.

Retrospective runs of the NCEP GFS v14 implementation (approximately 13 km horizontal resolution deterministic forecast, Han et al., 2017) using this updated SAS scheme have indicated the improvement of short lead-time precipitation forecast for CONUS (Equitable Threat Score (ETS) score is higher and frequency bias is reduced; not shown here).

2.5 Experiment Configurations

There are four experiments (including CTL) in this investigation, described in Sections 2.1-2.4. The period of the experiments is from May 2014 to May 2016 (25 months), with each experiment initialized at 0000 UTC every 5 days. Each experimental configuration of the GEFS runs 20 perturbed forecasts and a control (unperturbed) forecast out to 35 days. The ensemble resolutions are T_L574 (approximately 32km) with 64 hybrid vertical levels for days 0-8 and T_L382 (approximately 52km) with 64 hybrid levels for days 8-35 (Zhou et al., 2017). The initial analyses and perturbations are identical for all experiments and taken from the NCEP GFS v13 (with hybrid analysis - 4DEnVar) data assimilation system, which was implemented in May 2016. Some of the initial analyses and perturbations (before GFS v13 went operational) are from GFS v13 retrospective runs.

3. Evaluation

The verification in this study will evaluate the performance (forecast skill) difference for the three designed experiments compared to CTL. The evaluation covers tropical and extratropical areas, including upper atmosphere and surface variables applying deterministic (ensemble mean) and probabilistic (ensemble distribution) verifications. The evaluation varies from single variables to phenomena that combine the effect of multiple variables such as MJO. For extended range forecasts, such as week 2 and weeks 3 & 4, an average over the period (or an accumulation within the period for precipitation) has been used to evaluate forecast skill. All verification are on 1 deg x 1 deg or 2.5 deg x 2.5 deg lat-lon grid points weighted by latitude. The proxy truths for verification use the NCEP Global Data Assimilation System (GDAS) analysis and the Climatological Calibrated Precipitation Analysis (CCPA; Hou et al 2014). The skills are defined with respect to NCEP/NCAR 40 year reanalysis (Kalnay et al, 1996) climatology.

3.1 Tropical evaluation - Forecast skill of MJO

As one of the sources of predictability for the subseasonal time scale (or extended range forecast), a skill of MJO prediction is always considered a key metric when evaluating the forecast capability of the operational models (Kim et al. 2014; Shelly et al. 2014; Ling et al. 2014; Xiang et al. 2015, Wang et al. 2014). In this study, we evaluated the Wheeler Hendon

(WH) MJO skill (Lin et al. 2008; Wheeler and Hendon 2004), which is defined as the bivariate anomaly correlation between the analysis and forecast of two principal component time series (Real-time Multivariate MJO - RMM1 and RMM2) from multivariate EOF of the MJO components using outgoing longwave radiation (OLR) and zonal wind at 200 hPa and 850 hPa respectively, i.e.

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

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

The MJO strength during the experimental period is shown in Fig. 4. Over the two-year period, there are several significant MJO events with the strongest occurring in April and July of 2015. The MJO skill of the four experiments during the period of 1 May 2014 - 26 May 2016 (Fig. 5) indicates that using SPPT combined with SHUM and SKEB (SPs hereafter) outperformed the STTP (CTL hereafter) on MJO forecast skill by about 4 days (AC>=50%), improving the MJO skill from 12.5 days to 16.8 days. Since this improvement is realized without changing the model or its external forcing, this improvement is largely attributed to the decrease of forecast error and an increase in spread of tropical zonal winds providing a better representation of tropical forecast uncertainties that impact the MJO circulation (Fig. 6) and an overall better performance of tropical convection (figure not shown).

In addition to the uncertainty associated with the model dynamics and physics, as an uncoupled forecast system, error in the underlying SST forcing is a significant factor limiting the forecast skill (Wang et al. 2015). As previously discussed, the SST forcing in the operational GEFS uses the RTG analysis damped to the climatology, therefore the day-to-day variability of the SST cannot be well represented. To more realistically represent the underlying SST forcing, an updated SST from the bias corrected CFS v2 forecast was used to replace the climatological SST. The updated SST forcing combined with SPs (i.e. SPs+SST_bc) further improves the MJO skill of the extended GEFS by 1.7 days (Fig. 5; Table 1).

The model physics has been found to impact the MJO skill (DeMott et al. 2014, Jiang et al. 2015; 2016). To examine the effect of the model physics, especially the convective parameterization, we updated the convection package with the updated scale-aware SAS convection scheme before applying the SPs in the GEFS 35-day experiment. This further improved the MJO skill for additional 3 days compared to the SPs + SST_bc (Fig. 5; Table 1).

To further explore the reason for the improvement in each experiment, in addition to examining the RMS error and spread in tropical zonal wind at higher and lower level for SPs scheme, we also investigated the variation of the profile of zonal wind spread as a function of latitude for all experiments (Fig. 7). Compared to the CTL (Fig. 7a), all SPs experiments showed an enhanced zonal wind spread over the tropics and most of NH midlatitude. The enhancement is even stronger in SPs+SST_bc scheme and strongest in SPs+SST_bc+SA_CV. Given the fact that the ensemble spread is comparable to the forecast error in SPs scheme (Fig. 6) and other SPs schemes (figure not shown), we hypothesize that the increase of the zonal wind spread over the tropics contributed to the improvement of the MJO forecast skill through an improvement of the zonal wind forecast skill. Of course, the pure effect of the two-tiered SST and new convection scheme can't be excluded in MJO forecast skill improvement. The update of these two physical processes could also lead to improvement of the large-scale circulation. Since in our experiment design, it is hard to separate the pure effect of the two-tiered SST or the new convection scheme from the effect of the stochastic schemes, we demonstrated the ultimate combined effect for each scheme to help explain the reason of the improvement for the MJO forecast skill.

The MJO forecast skill has dependence on the initial strength of MJO (Kim et al. 2014, Lin et al. 2008). In order to show the difference of the MJO forecast skill between strong and weak periods, we separated the 25-month experiment period into a strong MJO period that covers April 2015 - March 2016 and a weak period that covers May 2014 - March 2015 (Fig. 8). As expected, the MJO forecast is more skillful during a strong MJO period than a weak period, with a difference of ~5 days. This is consistent in all experiment (not shown; aggregate numbers provided in Table 1).

Table 1 summarizes the MJO forecast skill for the 2-yr average and the comparison between the strong and weak MJO period. Overall, the SPs+SST_bc+SA_CV is the best configuration that results in a MJO skill of 22 days. The second best is the SPs+SST_bc. The largest contribution of the improvement is due to the new stochastic scheme that largely reduced the error over tropical circulation. The convection scheme and external boundary forcing also contributed in some degree, but since we used stochastic physics in the experiment for the evaluation, the effect of the convection scheme and external forcing should be considered a combined effect with the stochastic physics.

3.2 Extra-tropical evaluation- Forecast skill of 500 hPa geopotential height

Although the MJO forecast skill is considered to be one of the key metrics to evaluate the performance of a forecast system on the subseasonal time scale, the ultimate goal of these experiments is to examine the impact on extratropical forecast skill. As such, it is important to determine if the improvements contributed by SPs over the tropics extend into the extratropics. SPs shows substantial improvement according to the Relative Operating Characteristics (ROC) area lead time series for 500 hPa geopotential height over the northern hemisphere (Fig. 9a). The forecast skill is enhanced further for SPs+SST_bc and the SPs+SST_bc-SA_CV, where SPs+SST_bc-SA_CV has a highest skill especially at later lead times.

The AC of 500 hPa geopotential height is a commonly used measure to estimate pattern similarities of the large scale circulation and is used in this study to quantify potential skill over the extratropics (Wilks 2011). The AC of 500 hPa geopotential height for SPs, SPs+SST_bc and SPs+SST_bc+SA_CV shows improvement compared to the CTL, where SPs+SST_bc+SA_CV has the highest AC (Fig. 9b). In addition to examining the northern hemisphere 500 mb geopotential height as a function of lead time, it is also helpful to examine as a function of verification date so that the variations can be detected with respect to date for a given lead time (Fig. 10). The 8-14 day lead time mean is used to evaluate week 2 and the 15-28 day lead time mean is used to evaluate weeks 3 & 4.

Over the northern hemisphere, there is a small 0.3% improvement in skill in the SPs+SST_bc+SA_CV experiment compared to the CTL experiment during week 2 over the 2-year period (Fig. 10). However, during weeks 3 & 4, there is considerable improvement when using SPs, CFS v2 SST, and the scale aware convection scheme. When SPs replaces the operational STTP, there is a 3.1% improvement. There is additional 0.2% improvement when CFS v2 SSTs are added and an additional 0.9% improvement when the scale aware convection scheme is added (Table 2). SPs+SST_bc and SPs+SST_bc+SA_CV show statistically significant improvement relative to CTL during weeks 3 & 4. In the Southern Hemisphere, the overall skill is slightly less than the Northern Hemisphere, especially for weeks 3 & 4. However, an improvement in skill over the Southern Hemisphere is much larger than for the Northern Hemisphere (Fig. 10 and Table 2). There is a 3-4% improvement for week 2 and a 10% for weeks 3 & 4 over the Southern Hemisphere.

There is a similar pattern of improvement over the southern hemisphere, but for both week 2 and weeks 3 & 4. SPs contributes the most to the forecast improvement (4.2%) during week 2 as well as the most (5.3%) during weeks 3 & 4 compared to the contributions of CFS v2 SSTs and scale-aware convection scheme (Fig. 10). Therefore, the results show that the use of SPs improves the forecast skill the most over the extratropical region, especially during weeks 3 & 4.

3.3 Evaluation of 2-m temperature and precipitation

Over the experiment period, the global land only 2-m temperature Ranked Probabilistic Skill Scores (RPSS, Melhauser et al., 2016) were calculated for the Northern Hemisphere (NH), Southern Hemisphere (SH), tropics (TR) and Northern America (NA) for two lead time averages (week 2 average (Fig. 11a) and weeks 3 & 4 average (Fig. 11b)). The TR and SH have the highest RPSS for both week 2 and weeks 3 & 4 average lead times with NA having the lowest. Comparing week 2 with weeks 3 & 4, the RPSS remains similar for the TR and SH for both shorter and longer lead times, but the NH and NA show a ~0.05-0.1 reduction in skill. Except for the SH, stochastic physics improves the results compared with CTL, although the only statistically significant (95% confidence level) improvement is found for week 2 in the tropics for SPs+SST_bc and SPs+SST_bc+SA_CV and weeks 3 & 4 for all experiments. This is not a surprise since all three configuration changes, in some way; improve tropical forecasts either by increasing ensemble spread, introducing additional SST information, or improving convection parameterization, a dominant process in the tropics. Although NA shows improvement for all experiments, none of these are statistically significant at the 95% confidence level for both lead time averages, but SPs is improved for weeks 3 & 4 at the 90% confidence level.

The Contiguous United States (CONUS) precipitation forecasts for week 2 and weeks 3 & 4 averages have been verified against the CCPA (Hou et al, 2014). There are positive RPSS for the week 2 average (Fig. 11c; left) and negative RPSS for the weeks 3 & 4 average (Fig. 11c; right). The lower (or negative) skill for accumulated precipitation compared with 2-m temperatures is no surprise, due to lower predictability for extended range precipitation forecasts and imperfections in the model physical parameterizations causing the model forecast to be biased.

4. Impact of ensemble size on extended range forecasts

Evaluations have been made for the 21-member GEFS based on the three different configurations (experiments) and a control experiment. It has been found that the experiment SPs+SST_bc+SA_CV generally shows the highest skill for many forecast variables and measures. Therefore, this experiment has been selected as the configuration for the National Multi-Model Ensemble (NMME) and Subseasonal Experiment (SubX) projects in real time (once per week on Wednesday 0000 UTC). The configuration of this final experiment is also used for an 18-year GEFS reforecast (1999-2016), which will be used for forecast calibration.

Due to limited computing resources, an optimum GEFS configuration (including the reforecasts) in terms of resolution and ensemble size (Ma et al., 2012), years of reforecast, and reforecast frequency need to be determined. In this section, an impact of ensemble size will be discussed for week 2 and weeks 3 & 4 average forecast skill in order to provide valuable information for reforecast configuration.

During both week 2 and weeks 3 & 4, the 21-member experiment has the greatest AC compared to experiments with smaller ensemble sizes for the northern hemisphere and southern hemisphere 500hPa geopotential height (Fig. 12), and the tropical zonal winds of 850hPa and 250hPa (Fig. 13). From Fig. 12 and Fig 13, we find that the 11-member and 21-member ensembles agreed comparatively well, thus indicating that an 11-member ensemble is satisfactory in reproducing expected features. Unlike deterministic forecasts, an ensemble mean will eliminate many "drop out" cases (e.g. highlighted features in Fig. 12 and Fig. 13). Overall, an ensemble system remains more skillful than a deterministic model for the SPs+SST_bc+SA_CV configuration. The significance of the degradation due to decreased ensemble size was also determined (Table 3). In the tropics during week 2, there is significant degradation in AC for the 5-member and control only experiment relative to the 21-member experiment for both the upper- and lower-level zonal wind (Table 3). During weeks 3 & 4, however, the degradation is not significant for the 5-member experiment for the upper level zonal wind. In the extratropics during week 2 and weeks 3 & 4, the AC score of the 5-member experiment is significantly degraded for NH. All AC scores for 11 members are lower than for 21 members, but the difference is not significant.

5. Summary and discussion

The three experiments, which emphasize three different physical processes (i.e. tropical forecast uncertainties, the ocean forcing on the atmosphere, and tropical deep convection) to improve forecast skill on the subseasonal timescale, have been evaluated and compared with the control version forecast. The purpose of this experiment design was to determine the benefit to the weeks 3 & 4 forecast skill through the combination of all three areas based on some preliminary investigation. Due to the limitation of computational resources that are required for a comprehensive investigation, we didn't perform a comprehensive experiment to focus on each physical process independently.

Based on the analysis of MJO forecast skill, the experiment SPs+SST_bc+SA_CV shows ~22 days skillful forecast (defined by AC scores is excess of 50%), thus is considered to be the best configuration. Comparing to CTL, a large gain in MJO forecast skill is found after introducing new stochastic physics perturbation schemes. Updating the underlying SST and new convection scheme further added positive impact on the MJO forecast skill. As such, we speculate that the improvement of the representation of the forecast uncertainties over the tropics through introducing new stochastic physics perturbation schemes led to the biggest gain in MJO skill. The improvement of the MJO skill in SPs+SST_bc (SPs+SST_bc+CV) should be considered as a combined effect of the SST and SPs (SST, new convection scheme, and SPs). We cannot make any conclusions in terms of the relative importance of the SPs versus SST and convection scheme due to a lack of independent experiments.

In the extratropics, large-scale AC of 500 hPa geopotential heights of SPs+SST_bc+SA_CV indicated that AC scores for NH weeks 3 & 4 forecast improved significantly (5.4%) from CTL (35.5%) with no significant difference for week 2 forecast.

In terms of 2-meter temperature and accumulated precipitation, we have seen some improvement for NA and CONUS for week 2 and weeks 3 & 4 in all experiments from CTL, except SPs precipitation in week 2, but the improvement is marginal compared to the improvement for upper atmosphere variables. A future study will consider a analyzing each component independent to assess the relative impacts. Over the subseasonal time scale, systematic model errors in both 2-m temperature and accumulated precipitation can dominate the forecast error. The results may indicate that forecast skill could be improved if we have sufficient model hindcast (or reforecast) to calibrate subseasonal forecasts.

In additional to the real time run, another big task for the extended GEFS is the long term reforecast for the calibration of the real time forecast. For the reforecast, the ensemble size is an important part of configurations that should be carefully considered. We performed some tests of the dependence of the forecast skill score on the ensemble size and found that the skills degraded significantly for both week 2 and weeks 3 & 4 forecast when decreasing the ensemble size from 21 members to 5 members; the difference from 21 members to 11 members is relatively smaller. This indicates that 11 members are sufficient for the GEFS 18 years reforecast. Our intention is to minimize the use of computational resources, but still provide a high quality and reliable reforecast.

We investigated the capability of extended GEFS to potentially improve subseasonal prediction through advancing three physical processes for the operational system. Because we focus on the subseasonal time scale, the impact of the initial condition uncertainty and the associated initial perturbations are gradually reduced (Zhu, 2005). Having a better representation of the forecast uncertainty, mainly over the tropics, improves the forecast skill of tropical variables including the MJO. Updating the underlying SST and convection scheme lead to further improvement of the tropical forecast skill. While we focus on the subseasonal forecast skill using an uncoupled forecast system in this study, a fully coupled system may further improve forecast skills.The purpose of this study is to help shed light on some key components providing subseasonal forecast skill. As an intermediate solution before a fully coupled system is available, a suitable forecast system that can provide a significant improvement in forecast skill compared to the current operational configuration should be considered. As seen in this study, by updating the three physical processes in the forecast system, we obtain larger improvement in the upper atmosphere of the tropics and extratropics compared to extratropical surface temperature and precipitation. More investigation in the future is required to explore the impact of sea ice on high latitude forecast skill, the teleconnection of tropical and extra-tropical pattern to improve NA's hazard (and/or extreme) forecasts, and in tropics, examine the potential predictability of tropical cyclones at this timescale, including genesis.

Acknowledgement:

The authors would like to thank Drs. Xingren Wu, Wanqiu Wang, Jongil Han, Xu Li, Ruiyu Sun at EMC (and CPC) for valuable discussion pertaining to the design our experiments. Additionally, all of the help from EMC ensemble team members. This study is partially supported through NWS OSTI and NOAA's Climate Program Office (CPO)'s Modeling, Analysis, Predictions, and Projections (MMAP) program.

Reference:

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 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., P. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. Mon. Wea. Rev., 133, 1076–1097, doi: 10.1175/MWR2905.1.

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

DeMott, C. A., C. Stan, D. A. Randall, and M. D. Branson (2014), Intraseasonal variability in coupled GCMs: The roles of ocean feedbacks and model physics, J. Clim., 27, 4970–4995.

Fu, X., 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 Dynamics, 41, 1067, doi:10.1007/s00382-013-1859-9.

Gemmill, W., B. Katz and X. Li, 2007: Daily Real-Time, Gloabl Sea Surface Temperature - High-Resolution Analysis: RTG_SST_HR. NCEP Office Note, #260 1-39. http://polar.ncep.noaa.gov/mmab/papers/tn260/MMAB260.pdf

Han, J., W. Wang, Y. C. Kwon, S.-Y. Hong, V. Tallapragada, and F. Yang, 2017: Updates in the NCEP GFS Cumulus Convection Schemes with Scale and Aerosol Awareness, submitted to Wea. and Forecasting, in process.

Hou, D., M. charles, Y. Luo, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D. Seo, M. Pena, and B. Cui, 2014: Climatology-Calibrated Precipitation Analysis at Fine Scales: Statistical Adjustment of Stage IV toward CPC Gauge-Based Analysis. J. Hydrometeor., 15, 2542–2557

Jiang, X., and Coauthors, 2015: Vertical structure and physical processes of the Madden-Julian oscillation: Exploring key model physics in climate simulations. *J. Geophys. Res. Atmos.*, **120**, 4718–4748, doi:https://doi.org/10.1002/2014JD022375

Jiang, X., M. Zhao, E. D. Maloney, and D. E. Waliser (2016), Convective moisture adjustment time scale as a key factor in regulating model amplitude of the Madden-Julian Oscillation, Geophys. Res. Lett., 43, 10,412–410,419, doi:10.1002/2016GL070898.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Wollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, D. Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, Vol. 77, 437-471

Kim, H.-M., P. J. Webster, V. E. Toma, and D. Kim, 2014: Predictability and prediction skill of the MJO in two operational forecasting systems. *J. Climate*, **27**, 5364–5378, doi:https://doi.org/10.1175/JCLI-D-13-00480.1.

Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. J. Geophys. Res., 114, D04114, doi:10.1029/2008JD010969.

Li, W., R. Yu, H. Liu, and Y. Yu, 2001: Impacts of diurnal cycle of SST on the intraseasonal variation of surface heat flux over the western Pacific warm pool, Adv. Atmos. Sci., 18(5), 793–806.

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.

Ling, T., M. Xu, X.-Z. Liang, J. X. L. Wang, and Y. Noh, 2015: A multilevel ocean mixed layer model resolving the diurnal cycle: Development and validation, J. Adv. Model. Earth Syst., 07, doi:10.1002/2015MS000476

Liu X, Wu T, Yang S, Li T, Jie W, Zhang L, Wang Z, Liang X, Li Q, Cheng Y, Ren H, Fang Y, Nie S, 2016: MJO prediction using the sub-seasonal to seasonal forecast model of Beijing Climate Center. Clim Dyn. doi:10.1007/s00382-016-3264-7

Ma, J., Y. Zhu, D. Wobus and P. Wang, 2012: An Effective Configuration of Ensemble Size and Horizontal Resolution for the NCEP GEFS. Advance in Atmospheric Sciences, Vol. 29, No. 4, 782-794

Melhauser. C. W. Li, Y. Zhu, X. Zhou, M. Pena and D. Hou, 2016: Exploring the Impact of SST on the Extended Range NCEP Global Ensemble Forecast System, STI Climate Bulletine: http://www.nws.noaa.gov/ost/climate/STIP/41cdpw_digest.htm p 30-34

National Academies of Sciences, Engineering, and Medicine, 2016: Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts. Washington, DC. The National Academies Press. doi: 10.17226/21873.

NOAA National Weather Service, 2014: Next Generation of Global Prediction System (NGGPS). The documentations are available through online: https://www.weather.gov/sti/stimodeling_nggps

Palmer, T., N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. Tech. Rep. ECMWF RD Tech. Memo. 598, 42 pp. [Available online at http://www.ecmwf.int/publications/.]

Ramsay, B. H., 1998. The interactive multisensor snow and ice mapping system. Hydrological Processes, 12, 1537-1546.

Saha, S. and Coauthors, 2014: The NCEP Climate Forecast System Version 2. J. Climate, 27, 2185–2208, doi: 10.1175/JCLI-D-12-00823.1.

Sela, J., 1980: Spectral modeling at the National Meteorological Center, Mon. Wea. Rev., 108, 1279-1292.

Shutts, G., and T. N. Palmer, 2004: The use of high-resolution numerical simulations of tropical circulation to calibrate stochastic physics schemes. 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, 83 – 102.

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

Takaya, Y., F. Vitart, G. Balsamo, M. Balmaseda, M. Leutbecher, F. Molteni, 2010: Implementation of an ocean mixed layer model in IFS, Proc. ECMWF Technical Memorandum No. 622, pp. 34

Tompkins, A. M., and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability, J. Geophys. Res., 113, D18101.

Toth, Z. and E. Kalnay, 1993: Ensemble Forecasting at NMC: The Generation of Perturbations. Bull. Amer. Meteor. Soc., 74, 2317–2330

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. Mon. Wea. Rev., 127, 3297-3318.

Wheeler, M., 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.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 676 pp.

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

Zhou, X. Y. Zhu, D. Hou, and D. Kleist 2016: Comparison of the Ensemble Transform and the Ensemble Kalman Filter in the NCEP Global Ensemble Forecast System. Weather and Forecasting, Vol. 31, 2058-2074

Zhou, X. Y. Zhu, D. Hou, Y. Luo, J. Peng and D. Wobus, 2017: The NCEP Global Ensemble Forecast System with the EnKF Initialization. Weather and Forecasting (Accepted).

Zhu, Y. 2005: Ensemble Forecast: A New Approach to Uncertainty and Predictability. Advance in Atmospheric Sciences, Vol. 22, No. 6, 781-788

Zhu, Y. X. Zhou, M. Pena, W. Li, C. Melhauser, and D. Hou, 2017: Design and Evaluation of Sea Surface Temperature Forcing on Weeks 3 & 4 Forecast Skill using the NCEP Global Ensemble Forecasting System, Submit to Weather and Forecasting (in review).

Figure caption list:

Fig. 1. Lead days at which forecast loses useful skill (AC=0.6) for Northern Hemisphere 500 hPa geopotential height of GFS forecast (blue) and GEFS ensemble mean forecast (red) during years 2014 - 2016.

Fig. 2: 5-scale random patterns used in Stochastic Perturbed Physics Tendencies (SPPT). On the top of each plot, the numbers (except for upper left) represent the scales of spatial and temporal perturbations with the maximum amplitude and contour intervals in the bracket.

Fig. 3. Global meridional cross section showing the impact of stochastic perturbations for the atmosphere (cross section) for 144 hour forecasts from six spring initializations (left) and six fall initializations (right). Paneled are the differences of zonal wind spread from CTL for (top) no stochastic physical perturbations, and the difference of STTP (upper middle); SKEB (middle); SPPT; (lower middle) and SHUM (bottom).

Fig. 4. Amplitude of MJO during May 2014- May 2016 from GDAS analysis data. The resolution of the time-series is 5 days.

Fig. 5. MJO RMM1+RMM2 forecast skill for CTL (black), SPs (red), SPs+SST_bc (blue), and SPs+SST_bc+SA_CV (green) during the period of May 2014 to May 2016.

Fig. 6. RMS error plotted every 12 hours out to 35 days of the ensemble mean (solid) and the ensemble spread (dash) for 850 hPa (top) and 250 hPa (bottom) tropical zonal wind (20oN – 20oS) during the May 2014 to May 2016 period comparing CTL (black) and the new SPs (red).

Fig. 7. Global meridional cross section of the zonal wind spread [m s^{-1}] at 360 forecast hours (15 days) for a) CTL, b) SPs minus CTL, c) SPs+SST_bc minus CTL; and d) SPs+SST_bc+SA_CV minus CTL. The result is calculated using 6 cases starting the 1st of March 2016 every 5-day.

Fig. 8. MJO RMM1+RMM2 forecast skill calculated for a weak MJO period (May 2014 – March 2015) and a strong MJO period (April 2015 – March 2016) for experiment SPs+SST_bc+SA_CV.

Fig. 9. Daily Relative Operating Characteristics (ROC; top) scores and Anomaly Correlation (AC; bottom) scores out to 35 days for Northern Hemisphere (20^{o}N-80^{o}N) 500 hPa geopotential height for CTL (black), SPs (red), SPs+SST_bc (green), and SPs+SST_bc+SA_CV (blue).

Fig. 10. Ensemble mean Anomaly Correlation time series for Northern Hemisphere 500 hPa geopotential height from May 2014 - May 2016 for CTL (black) and SPs (red) for a) days 8-14 and b) days 15-28 (weeks 3 & 4). Panel c) and d) are the same as a) and b) except for the Southern Hemisphere. Average scores are shown by straight dashed lines

matching the color of CTL and SPs.

Fig. 11. Ranked Probability Skill Scores for CTL (black), SPs (red), SPs+SST_bc (green), and SPs+SST_bc+SA_CV (blue) during the period of May 2014 - May 2016 for a) land only 2-m temperature week 2 (day 8-14) average; b) land only 2-m temperature weeks 3 & 4 (day 15-28) average; and c) CONUS only accumulated precipitation week 2 (day 8-14; left) and weeks 3 & 4 (day 15-28; right) average.

Fig. 12. Ensemble mean Anomaly Correlation time series for Northern Hemisphere 500 hPa geopotential height from May 2014 - May 2016 for 1 member (black), 5 members (red), 11 members (green), and 21 members (blue) for a) days 8-14 and b) days 15-28 (weeks 3 & 4). Panel c) and d) are the same as a) and b) except for the Southern Hemisphere. Average scores are shown by straight dashed lines matching the color of different member sizes

Fig. 13. Ensemble mean Anomaly Correlation time series for Tropical 850 hPa zonal wind from May 2014 - May 2016 for 1 member (black), 5 members (red), 11 members (green), and 21 members (blue) for a) days 8-14 and b) days 15-28 (weeks 3 & 4). Panel c) and d) are the same as a) and b) except for Tropical 250 hPa zonal wind. Average scores are shown by straight dashed lines matching the color of different member sizes.

Table caption list:

Table 1. Days of useful MJO forecast skill (50% of amplitude correlation) for the four experiments over the full period (May 1 2014 - May 26 2016) and a weak (May 2014 – March 2015) and strong period (April 2015 – March 2016).

Table 2. Anomaly Correlation (AC) averaged over 25 months for lead day 8-14 (week 2) and lead day 15-28 (weeks 3 & 4). The bolded blue values represent results that significantly improved from the CTL at the 95% confidence level.

Table 3. Anomaly Correlation (AC) for different ensemble sizes from SPs+SST_bc+SA_CV averaged over 25 months for lead days 8-14 (week 2) and lead days 15-28 (weeks 3 & 4). The bolded red values represent results that are significantly degraded from the 21-member ensemble experiment at the 95% confidence level.

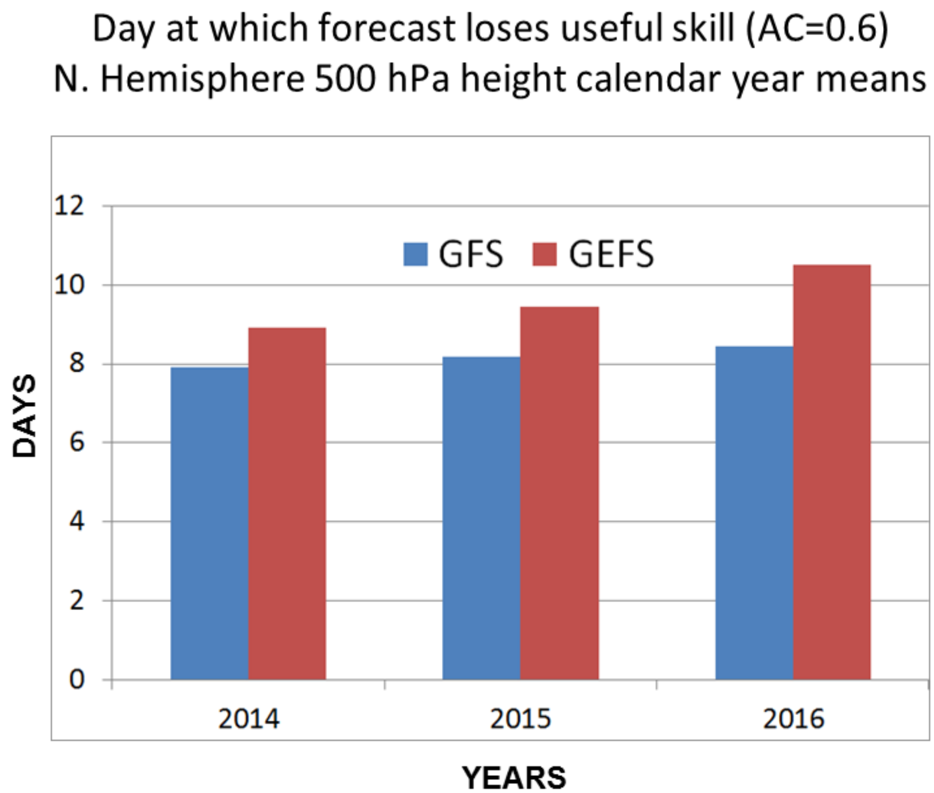

Fig. 1. Lead days at which forecast loses useful skill (AC=0.6) for Northern Hemisphere 500 hPa geopotential height of GFS forecast (blue) and GEFS ensemble mean forecast (red) during years 2014 - 2016.

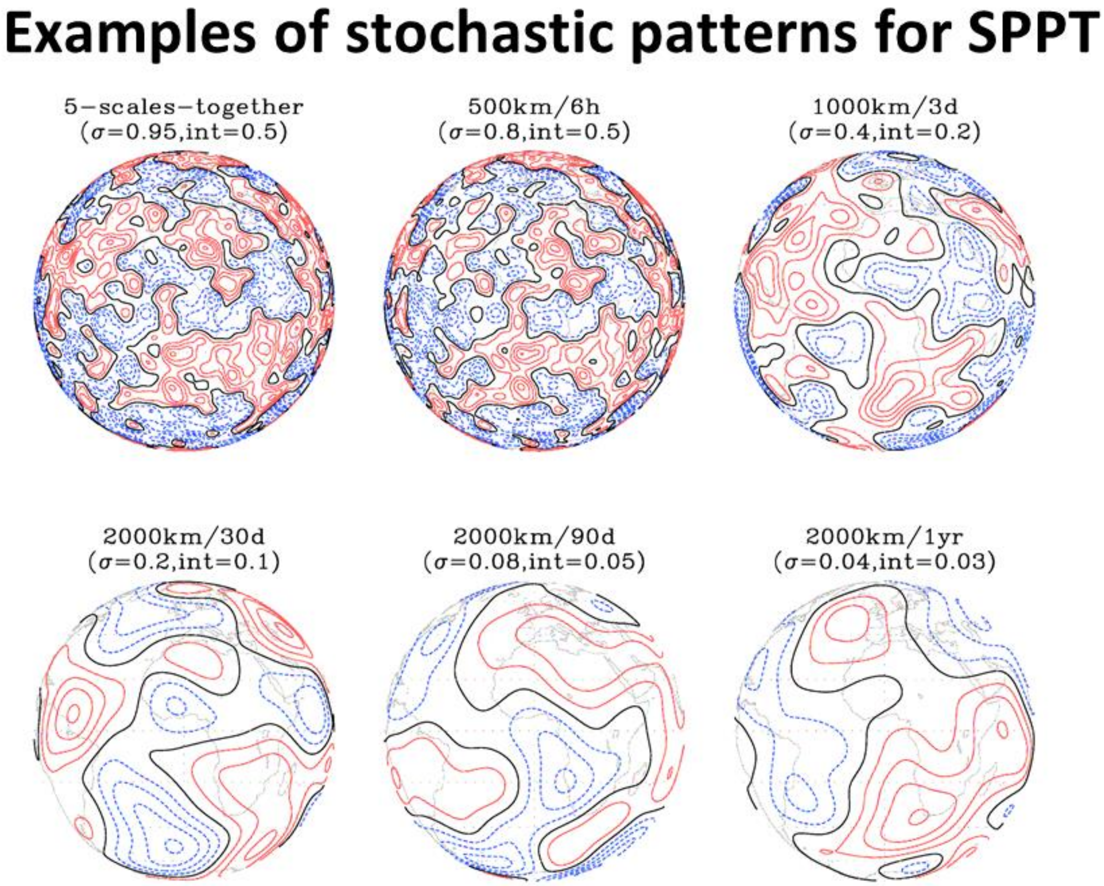

Fig. 2: 5-scale random patterns used in Stochastic Perturbed Physics Tendencies (SPPT). On the top of each plot, the numbers (except for upper left) represent the scales of spatial and temporal perturbations with the maximum amplitude and contour intervals in the bracket.

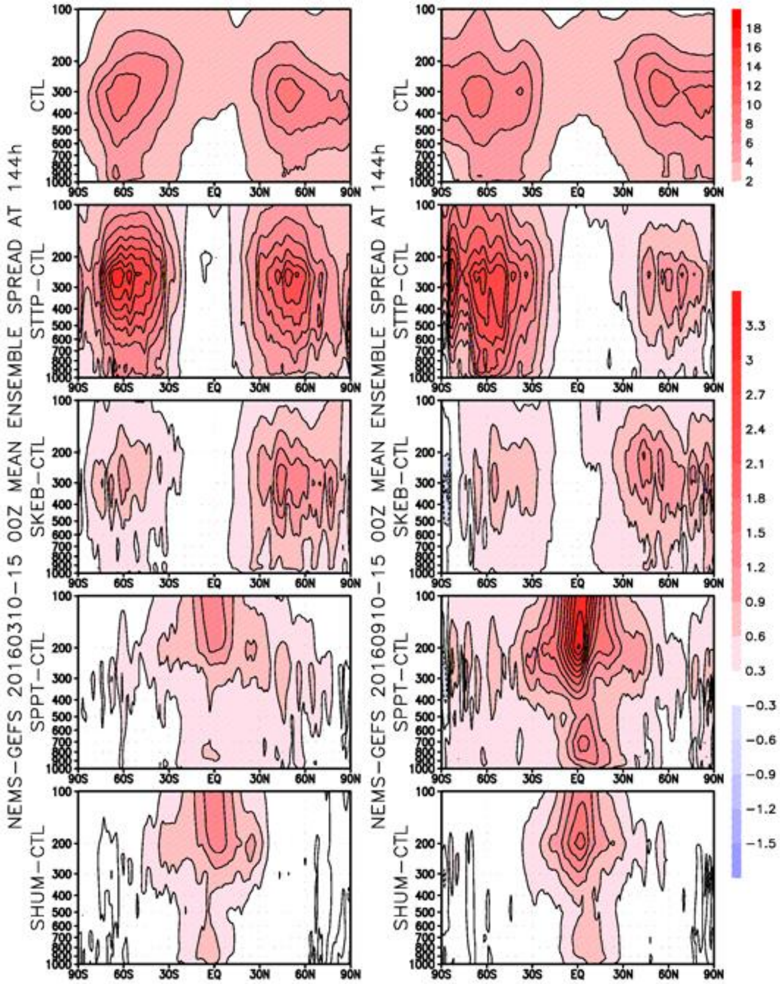

Fig. 3. Global meridional cross section showing the impact of stochastic perturbations for the atmosphere (cross section) for 144 hour forecasts from six spring initializations (left) and six fall initializations (right). Paneled are the differences of zonal wind spread from CTL for (top) no stochastic physical perturbations, and the difference of STTP (upper middle); SKEB (middle); SPPT; (lower middle) and SHUM (bottom).

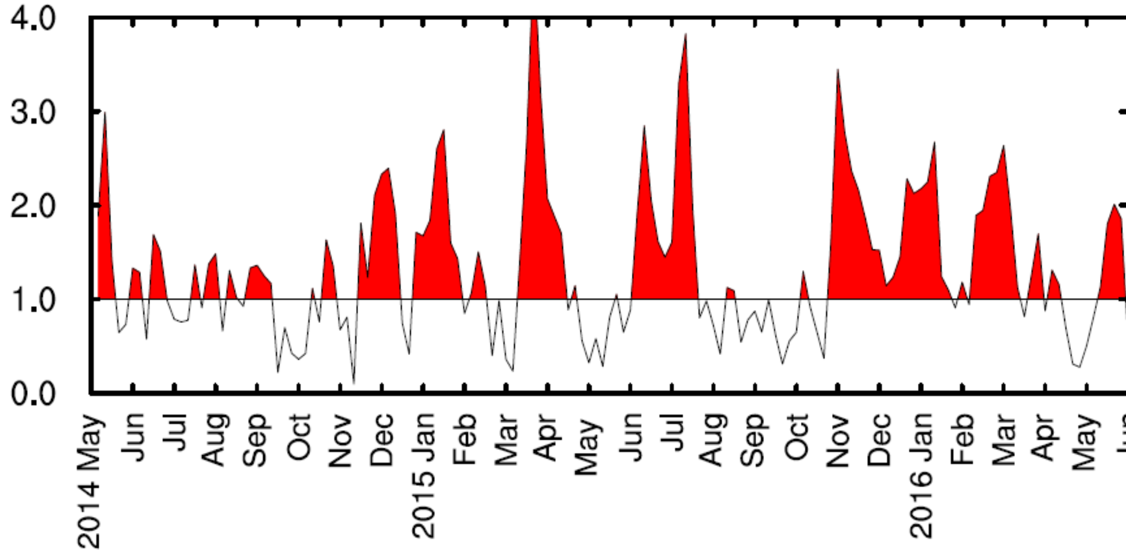

Fig. 4. Amplitude of MJO during May 2014- May 2016 from GDAS analysis data. The resolution of the time-series is 5 days.

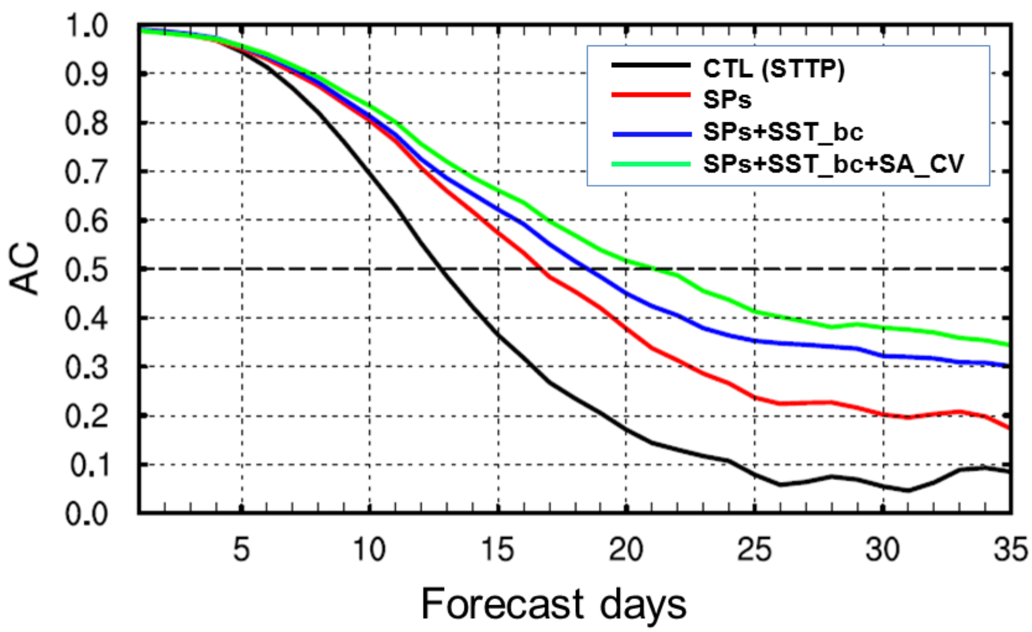

Fig. 5. MJO RMM1+RMM2 forecast skill for CTL (black), SPs (red), SPs+SST_bc (blue), and SPs+SST_bc+SA_CV (green) during the period of May 2014 to May 2016.

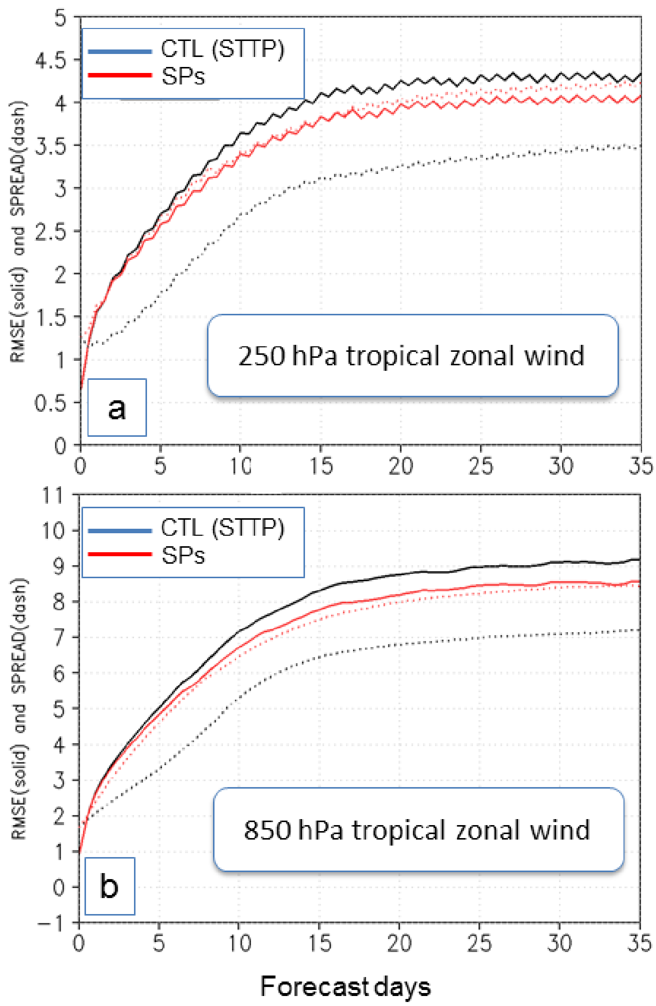

Fig. 6. RMS error plotted every 12 hours out to 35 days of the ensemble mean (solid) and the ensemble spread (dash) for 850 hPa (top) and 250 hPa (bottom) tropical zonal wind (20oN – 20oS) during the May 2014 to May 2016 period comparing CTL (black) and the new SPs (red).

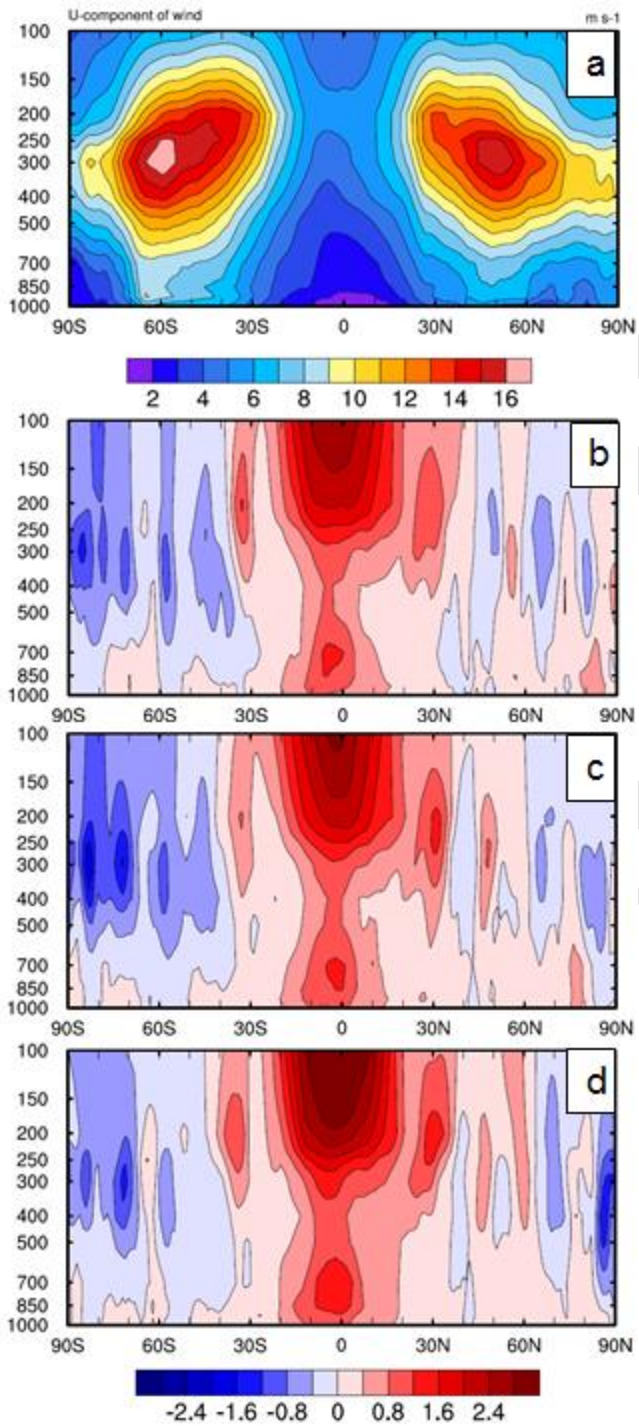

Fig. 7. Global meridional cross section of the zonal wind spread [m s-1] at 360 forecast hours (15 days) for a) CTL, b) SPs minus CTL, c) SPs+SST_bc minus CTL; and d) SPs+SST_bc+SA_CV minus CTL. The result is calculated using 6 cases starting the 1st of March 2016 every 5-days.

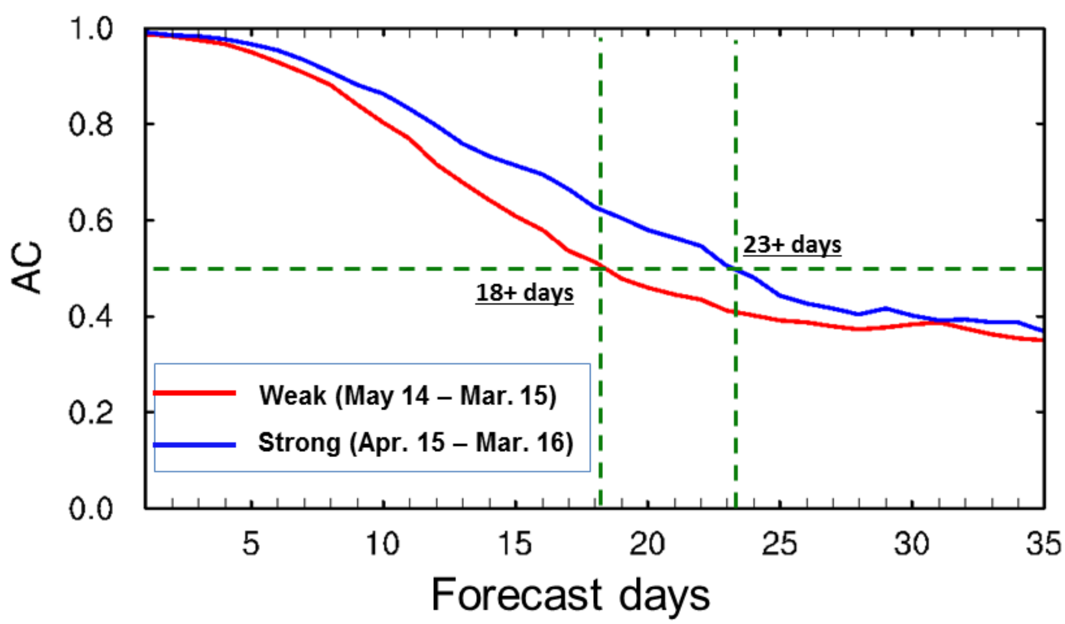

Fig. 8. MJO RMM1+RMM2 forecast skill calculated for a weak MJO period (May 2014 – March 2015) and a strong MJO period (April 2015 – March 2016) for experiment SPs+SST_bc+SA_CV.

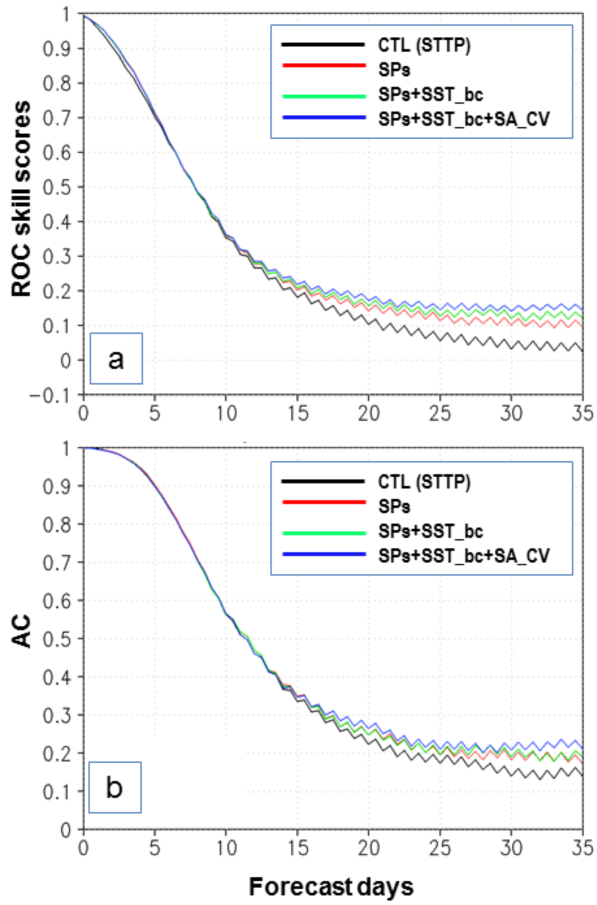

Fig. 9. Daily Relative Operating Characteristics (ROC; top) and Anomaly Correlation (AC; bottom) scores out to 35 days for Northern Hemisphere (20oN-80oN) 500 hPa geopotential height for CTL (black), SPs (red), SPs+SST_bc (green), and SPs+SST_bc+SA_CV (blue).

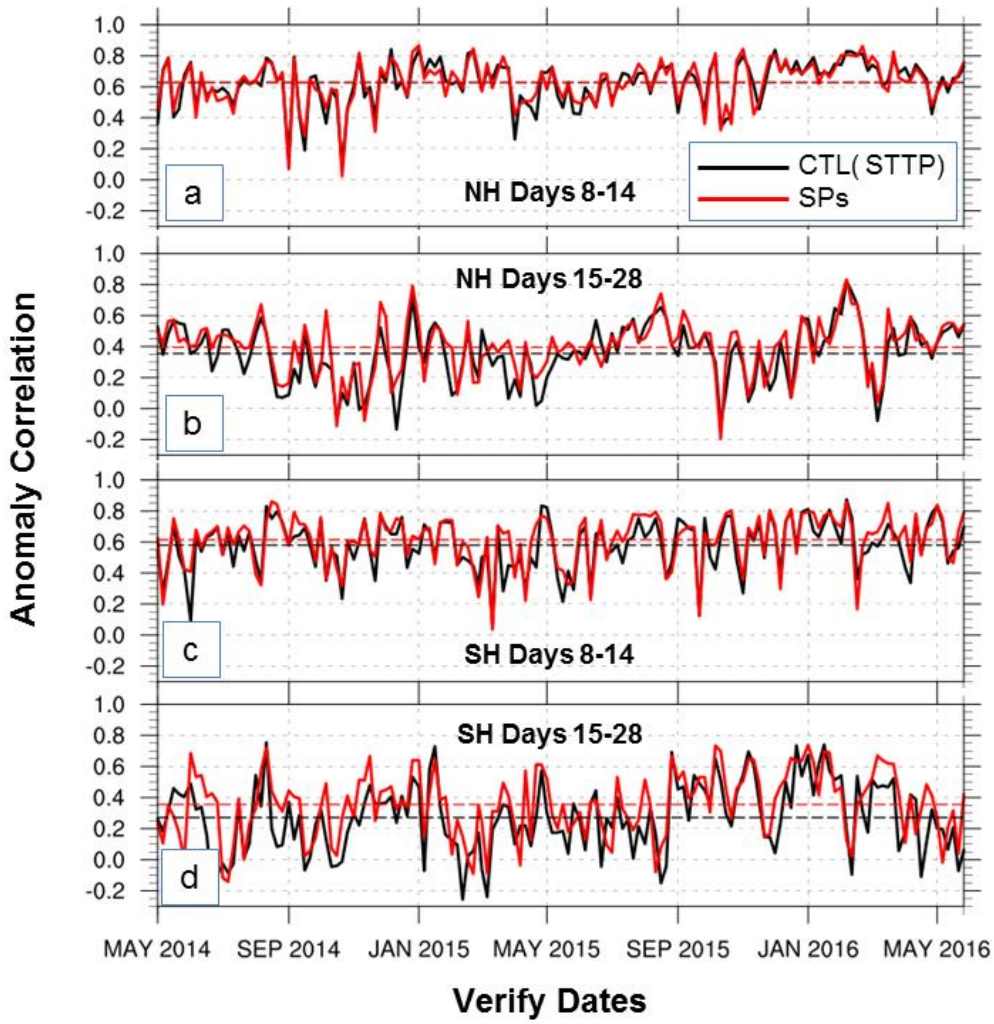

Fig. 10. Ensemble mean Anomaly Correlation time series for Northern Hemisphere 500 hPa geopotential height from May 2014 - May 2016 for CTL (black) and SPs (red) for a) days 8-14 and b) days 15-28 (weeks 3 & 4). Panel c) and d) are the same as a) and b) except for the Southern Hemisphere. Average scores are shown by straight dashed lines matching the color of CTL and SPs

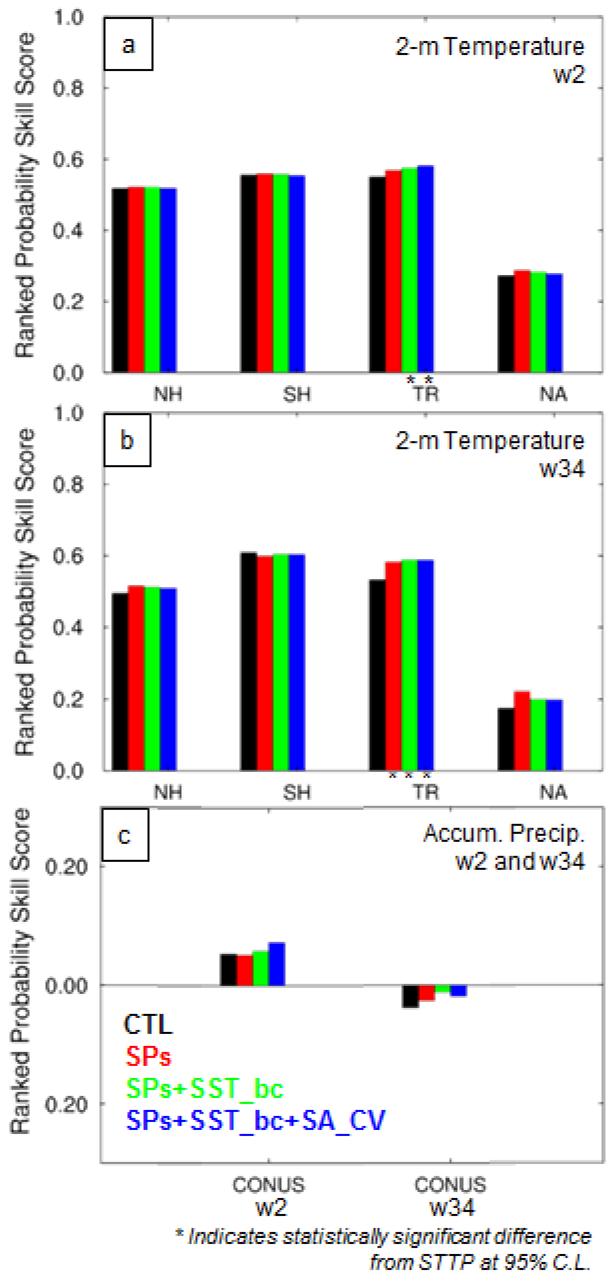

Fig. 11. Ranked Probability Skill Scores for CTL (black), SPs (red), SPs+SST_bc (green), and SPs+SST_bc+SA_CV (blue) during the period of May 2014 - May 2016 for a) land only 2-m temperature week 2 (day 8-14) average; b) land only 2-m temperature weeks 3 & 4 (day 15-28) average; and c) CONUS only accumulated precipitation week 2 (day 8-14; left) and weeks 3 & 4 (day 15-28; right) average.

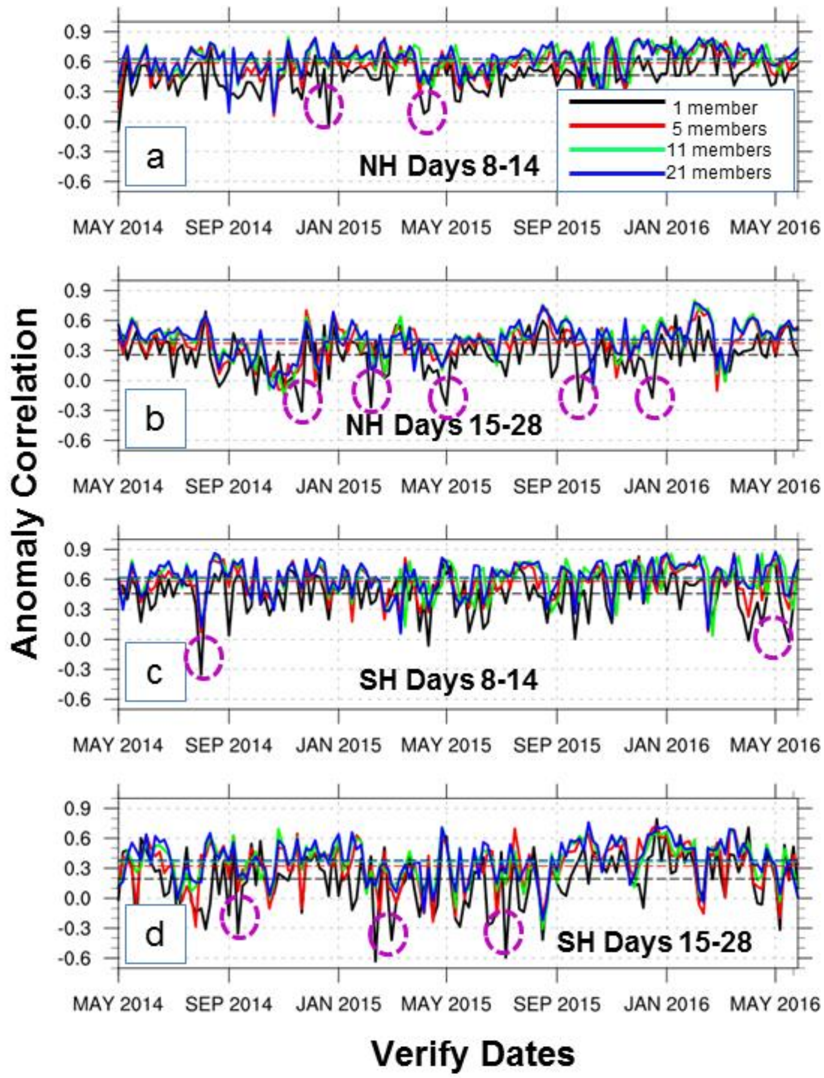

Fig. 12. Ensemble mean Anomaly Correlation time series for Northern Hemisphere 500 hPa geopotential height from May 2014 - May 2016 for 1 member (black), 5 members (red), 11 members (green), and 21 members (blue) for a) days 8-14 and b) days 15-28 (weeks 3 & 4). Panel c) and d) are the same as a) and b) except for the Southern Hemisphere. Average scores are shown by straight dashed lines matching the color of different member sizes.

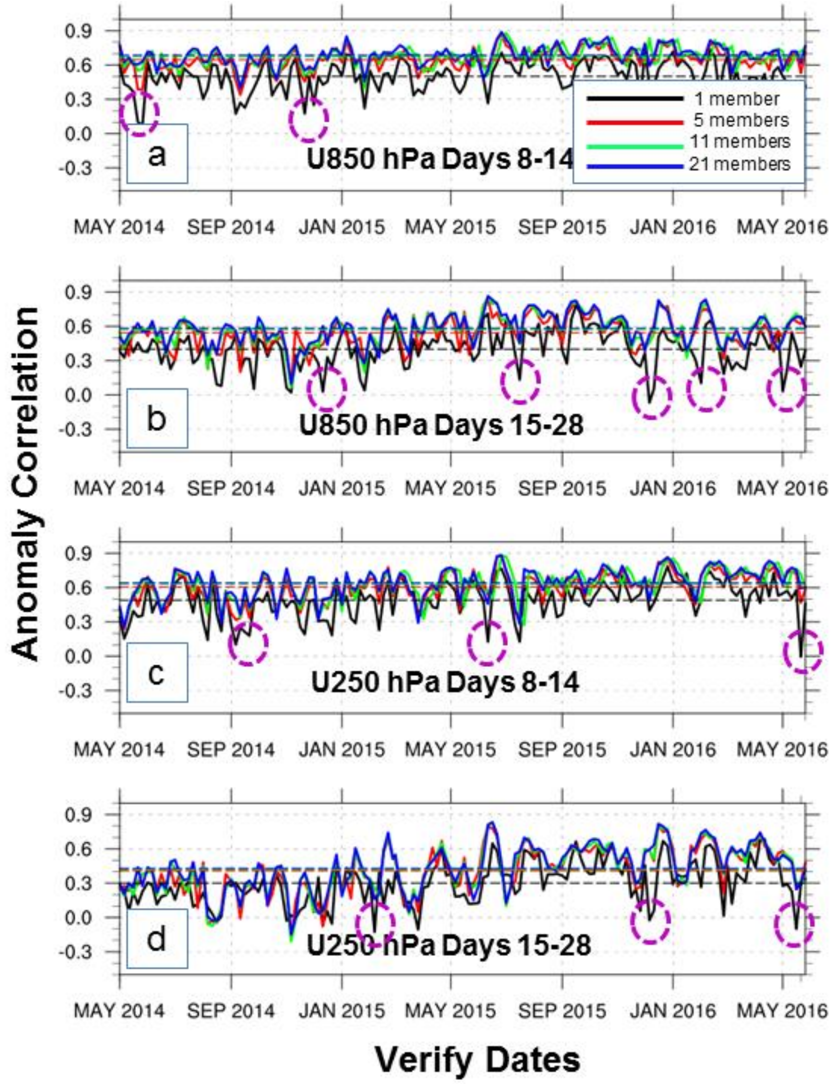

Fig. 13. Ensemble mean Anomaly Correlation time series for Tropical 850 hPa zonal wind from May 2014 - May 2016 for 1 member (black), 5 members (red), 11 members (green), and 21 members (blue) for a) days 8-14 and b) days 15-28 (weeks 3 & 4). Panel c) and d) are the same as a) and b) except for Tropical 250 hPa zonal wind. Average scores are shown by straight dashed lines matching the color of different member sizes.

Configurations	Weak	Strong	25 months
CTL (STTP)	12.2	12.8	**12.5**
SPs	15.8	18	**16.8**
SPs+SST_bc	17	19.5	**18.5**
SPs+SST_bc+SA_CV	18+	23+	**22.0**

Table 1. Days of useful MJO forecast skill (50% of amplitude correlation) for the four experiments over the full period (20140501-20160526) and a weak (May 2014 – March 2015) and strong period (April 2015 – March 2016).

AC scores	CTL	SPs	SPs+SST_bc	SPs+SST_bc+SA_CV
NH day 8-14	0.627	0.630	0.632	0.629
NH day 15-28	0.355	0.396	**0.398**	**0.409**
SH day 8-14	0.580	0.615	**0.620**	**0.618**
SH day 15-28	0.271	**0.366**	**0.367**	**0.379**

Table 2. Pattern Anomaly Correlation averaged over 25 months for lead day 8-14 (week 2) and lead day 15-28 (weeks 3 & 4). The bolded blue values represent results that significantly improved from the CTL at the 95% confidence level.

AC Scores	Domains	Variables	21 Members	11 Members	5 Members	1 Member
Day 8-14	NH	z500	0.628	0.619	**0.586**	**0.463**
	SH	z500	0.620	0.609	0.582	**0.458**
	TR	u850	0.686	0.673	**0.646**	**0.501**
		u250	0.641	0.630	**0.605**	**0.490**
Day 15-28	NH	z500	0.410	0.405	**0.372**	**0.257**
	SH	z500	0.380	0.363	0.323	**0.194**
	TR	u850	0.583	0.571	**0.544**	**0.400**
		u250	0.430	0.420	0.409	**0.300**

Table 3. Pattern Anomaly Correlation for different ensemble sizes from SPs+SST_bc+SA_CV averaged over 25 months for lead days 8-14 (week 2) and lead days 15-28 (weeks 3 & 4). The bolded red values represent results that are significantly degraded from the 21-member ensemble experiment at the 95% confidence level.