Assessment of the Forecast Skill of Multiphysics and Multistochastic Methods within the GRAPES Regional Ensemble Prediction System in the East Asian Monsoon Region

ZHIZHEN XU

Department of Atmospheric and Oceanic Sciences, and Institute of Atmospheric Sciences, Fudan University, Shanghai, and Chinese Academy of Meteorological Sciences, China Meteorological Administration, Beijing, China

JING CHEN

Numerical Weather Prediction Center, China Meteorological Administration, Beijing, China

ZHENG JIN

Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

HONGQI LI AND FAJING CHEN

Numerical Weather Prediction Center, China Meteorological Administration, Beijing, China

(Manuscript received 31 January 2019, in final form 22 March 2020)

ABSTRACT

To more comprehensively and accurately address model uncertainties in the East Asia monsoon region, a single-physics suite, where each ensemble member uses the same set of physics parameterizations as the control member in combination with multiple stochastic schemes, is developed to investigate if the multistochastic schemes that combine different stochastic schemes together can be an alternative to a multiphysics suite, where each ensemble member uses a different set of physics parameterizations (e.g., cumulus convection, boundary layer, surface layer, microphysics, and shortwave and longwave radiation). For this purpose, two experiments are performed for a summer monsoon month over China: one with a multiphysics suite and the other with a single-physics suite combined with multistochastic schemes. Three stochastic schemes are applied: the stochastically perturbed parameterizations (SPP) scheme, consisting of temporally and spatially varying perturbations of 18 parameters in the microphysics, convection, boundary layer, and surface layer parameterization schemes; the stochastically perturbed parameterization tendencies (SPPT) scheme; and the stochastic kinetic energy backscatter (SKEB) scheme. The combination of the three stochastic schemes is compared with the multiphysics suite in the Global and Regional Assimilation and Prediction Enhanced System–Regional Ensemble Prediction System with a horizontal grid spacing of 15 km. Verification results show that, overall, a single-physics suite that combines SPP, SPPT, and SKEB outperforms the multiphysics suite in precipitation verification and verification for upper-air weather variables, 10-m zonal wind, and 2-m temperature in the East Asian monsoon region. The indication is that a single-physics suite combining SPP, SPPT, and SKEB may be an appropriate alternative to a multiphysics suite. This finding lays a foundation for the development and design of future regional and global ensembles.

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-19-0021.s1>.

Corresponding author: Jing Chen, chenj@cma.gov.cn

1. Introduction

Ensemble prediction systems (EPSs) have been widely applied in various numerical weather prediction (NWP) centers to represent the forecast uncertainties that arise from uncertainties in the initial conditions and the model (Palmer et al. 1993; Molteni et al. 1996; Toth and Kalnay 1993; Houtekamer et al. 1996; Buizza et al. 1999; Palmer 2001; Chen et al. 2002; Leutbecher and Palmer 2008;

DOI: 10.1175/WAF-D-19-0021.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

Berner et al. 2011; Bauer et al. 2015). In the early years, some NWP centers addressed forecast uncertainty in part through initial perturbations; however, it is known that an EPS with only initial uncertainties is likely to be under-dispersive and insufficient in entirely and accurately explaining the forecast uncertainty (Buizza et al. 2005; Romine et al. 2014; Beck et al. 2016). As such, the need to account for model uncertainties that might derive from misrepresentation of unresolved physical parameterization processes is increasingly recognized as an essential part of producing a reliable probabilistic forecast (Palmer et al. 2009; Berner et al. 2011; Christensen et al. 2015; Leutbecher et al. 2017).

A variety of model-error schemes have been developed to represent model uncertainties. These include multimodel (e.g., Krishnamurti et al. 1999, 2016), multiphysics (e.g., Houtekamer et al. 1996; Stensrud et al. 2000), multiparameter (e.g., Stainforth et al. 2005), and stochastic schemes (e.g., Buizza et al. 1999). One widely employed option is the multiphysics ensemble configuration method, which introduces intramodel diversity by varying the suite of parameterizations (e.g., cumulus convection, boundary layer, surface layer, microphysics, and shortwave and longwave radiation) between ensemble members (Stensrud et al. 2000; Berner et al. 2011). The use of different parameterizations in multiphysics allows model deficiencies to be taken into account within the EPS (Beck et al. 2016). As part of the forecast error is due to model deficiencies, perturbing the model by using a multiphysics suite will likely lead to an improved ensemble forecast (e.g., Houtekamer et al. 1996; Berner et al. 2011; Hacker et al. 2011b; Wu et al. 2013). Additionally, for the multiphysics suite, each ensemble member has a different physics combination and hence has a different climatology and bias, which may introduce large diversity among the ensemble members, and that is one of the reasons why the use of a multiphysics suite typically increases the spread (Houtekamer et al. 1996; Eckel and Mass 2005; Berner et al. 2011; Hacker et al. 2011b; Arnold et al. 2013; Berner et al. 2017). However, although multiphysics suites perform well, there are several theoretical and practical disadvantages involved. First, ensemble members are systematically different from each other, and this does not fit with the aim of representations of forecast uncertainty to simulate the effect of the random component of the forecast errors but not the systematic component. Furthermore, ensemble members from multiphysics tend to cluster by model (Alhamed et al. 2002), such that the form of a forecast ensemble probability density function (PDF) may be dictated more by the distribution of models or schemes than by the synoptically dependent forecast

uncertainty (Bowler et al. 2008). Moreover, it is hard to find different sets of physics parameterizations that are physically consistent with each other (Berner et al. 2011). Finally, multiphysics requires additional development and maintenance costs, and choosing an optimal suite of physics parameterizations that work well together is a time-consuming task (Berner et al. 2011; Hacker et al. 2011b). Therefore, an alternative to the multiphysics suite is needed to better represent model uncertainties.

While a deterministic parameterization only represents the average subgrid-scale effect, a stochastic parameterization can better represent unsolved subgrid variability and regime transitions via inherent nonlinear processes, and thus can be viewed as a possible realization of the subgrid-scale motion (Arnold et al. 2013; Lierwalqui et al. 2018). In recent years, several stochastic schemes for representing model uncertainties have been developed and implemented in various EPSs (Buizza et al. 1999; Shutts 2005; Bowler et al. 2008; Berner et al. 2009; Palmer 2012; Yuan et al. 2016). Those schemes have similar model bias and climatology for all ensemble members, resulting in more statistically consistent (i.e., ensemble variance matches the mean square error of the ensemble mean) ensemble distributions over the multiphysics suite (Eckel and Mass 2005; Berner et al. 2011; Jankov et al. 2017). One stochastic scheme is the stochastically perturbed parameterization tendencies (SPPT) scheme (Buizza et al. 1999; Palmer et al. 2009), which has been found to have a beneficial impact on improving the ensemble spread, as well as the probabilistic skill of medium-range, seasonal, and climate forecasts (Berner et al. 2017; Leutbecher et al. 2017). Another stochastic scheme is the stochastic kinetic energy backscatter (SKEB) scheme (Mason and Thomson 1992; Shutts 2005; Berner et al. 2009, 2011), which can increase the ensemble spread and better represent certain aspects of tropical climate variability (Shutts 2015). A third kind of stochastic scheme is the stochastically perturbed parameterizations scheme (SPP; Li et al. 2008; Hacker et al. 2011b; Christensen et al. 2015), which has already shown certain advantages and beneficial impacts (e.g., the SPP scheme does not increase model biases, and improves the representation of light to heavy rain in the short range) in implementations of some NWP centers (e.g., Ollinaho et al. 2017; Jankov et al. 2017).

It is not possible to characterize the errors entirely and accurately with one single stochastic perturbation scheme, as there are many different sources of model error (Leutbecher et al. 2017). Thus, a combination of different stochastic schemes (hereinafter referred to as multistochastic schemes) has the potential to more

comprehensively represent model uncertainties from different sources and at different stages of the model integration process. A few studies have shown that multistochastic schemes outperform any single scheme and can best capture the model error and improve the forecast skill (Palmer et al. 2009; Charron et al. 2010; Berner et al. 2011, 2017; Hacker et al. 2011b; Leutbecher et al. 2017; Wolff et al. 2016; Jankov et al. 2017). Such a combined approach has been applied at some operational NWP centers—for example, Environment Canada (Charron et al. 2010) and European Centre for Medium-Range Weather Forecasts (ECMWF) (Palmer et al. 2009) combine SKEB and SPPT, and the Met Office EPS (MOGREPS) uses the random parameter (RP) scheme together with the SKEB scheme (McCabe et al. 2016). This suggests potential for multistochastic schemes to better deal with model uncertainties.

In this study, a single-physics suite combined with multistochastic schemes (SPP, SPPT, and SKEB) is applied in the Global and Regional Assimilation and Prediction Enhanced System–Regional Ensemble Prediction System (GRAPES-REPS) to consider possible sources of uncertainties in the East Asian monsoon region. A similar multistochastic experiment was conducted over the North American domain by Jankov et al. (2017). In their SPP implementation, they perturbed four uncertain parameters in convective and boundary layer parameterization schemes, and applied the same temporal and spatial decorrelations of stochastic patterns to perturb the parameters in SPP as those in SPPT [also see Christensen et al. (2015)]. A comprehensive SPP is implemented in this study. It perturbs 18 parameters to account for the uncertainties in four physics parameterization schemes: the convection, boundary layer, surface layer, and microphysics parameterizations. In addition, the temporal and spatial correlations of the stochastic pattern used to perturb the parameters in SPP are set differently from those in SPPT, and we believe these should be set differently in the two methods. The SPP scheme should be developed to measure the temporal and spatial correlation scales of the perturbations and to improve the forecast skill.

East Asia is a well-documented monsoon region with a unique monsoon climate and complex topography and landscape (e.g., the Tibetan Plateau) compared to other regions of the world (Flohn 1957; Tao and Chen 1987; Ding 2004). From June to August, driven by diverse factors (e.g., the thermal difference between the Asian mainland and the adjacent oceans, and annual variation of solar radiation), the East Asian summer monsoon (EASM) propagates northward

and northeastward, and severe monsoon rainfall is brought to the Yangtze River basin, the southwestern Philippine Sea, North China, and tropical western Pacific (Ding and Chan 2005). The EASM and related seasonal rain belts assume great variability at intra-seasonal, interannual, and interdecadal time scales, and the monsoon-related rainfall is usually associated with nonlinear interactions among a complex mix of planetary, synoptic, and mesoscale (i.e., meso- α - and meso- β -scale) rainfall systems (e.g., Ninomiya and Akiyama 1992). Additionally, the EASM system is affected by the Tibetan Plateau and complex land–sea distributions in this region (Flohn 1957; Hsu and Liu 2003; Wu et al. 2007; Chen and Bordonì 2014). Therefore, the complexity of the monsoon climate and topography of the EASM region present great challenges and uncertainties in numerical simulations and physics parameterizations (e.g., cumulus convection, boundary layer, surface layer, microphysics) for this region (Kang et al. 2002; Wang and Li 2004). In addition, under the influence of the EASM, the mechanism and characteristics of error growth, which is closely related to the rapid growing moist convective instabilities in the EASM region, are generally complicated, and thus, how to represent the forecast error of the EASM region and construct a reasonable model perturbation technique is a problem worthy of further study.

In this study, we compare the performance of a single-physics suite combined with multistochastic schemes (SPP, SPPT, and SKEB) and a multiphysics suite in the East Asian monsoon region. The aim of this study is to investigate whether the multistochastic schemes can outperform the multiphysics suite and further be an alternative to it. The paper is organized as follows: The model configurations, experimental design, and data are described in section 2. A detailed description of the three stochastic schemes is provided in section 3. The results of both the precipitation verification and verification for upper-air weather variables, 10-m zonal wind, and 2-m temperature are presented in section 4. Finally, a summary and discussion are presented in section 5.

2. Methodology

a. Model configurations

The GRAPES-REPS, applied in this study, is a regional EPS based on GRAPES Mesoscale (GRAPES-Meso) forecast system, which has been developed since 2010 (Long et al. 2011) and was operationally implemented in 2014 in the China Meteorological Administration

FIG. 1. Domain and topography for the model simulation and the distribution of ground-based meteorological observation stations.

(CMA) (Zhang et al. 2014). It has 15 members (1 control and 14 perturbed members) and covers the domain (15° – 55° N, 70° – 140° E). The domain and topography for the model simulation, along with the distribution of 2412 ground-based meteorological observation stations, are shown in Fig. 1. A geographical map of East Asia (Fig. S1 in the online supplemental material) has also been provided in the supporting information. This system adopts terrain-following coordinates with a 15-km horizontal grid spacing and has 51 vertical levels that are not spaced equally (the model top is at 10 hPa). It runs twice a day (initialized at 0000 and 1200 UTC) out to a forecast length of 72 h with 6-h output (model integration time step is 60 s). The boundary and initial conditions are provided by the GRAPES global EPS, which is also running operationally at the CMA. Model uncertainty is addressed by applying a multiphysics suite along with the SPPT scheme, and GRAPES-REPS applies its multiphysics suite via the combination of two boundary parameterization schemes and four convective cumulus parameterization schemes. The configurations of the multiphysics suite of GRAPES-REPS are given in Table 1.

b. Experimental design

In this study, two experiments are conducted for a summer monsoon month (1–30 June 2015) based on GRAPES-REPS over China (see Table 2), and forecasts are initialized at 0000 UTC and integrated for 48 h. The MP experiment employs the multiphysics suite in GRAPES-REPS (see Table 1). A single-physics suite is employed and combined with the SPPT, SPP,

and SKEB schemes in the SPPT_SPP_SKEB experiment. The single-physics suite includes the new Kain–Fritsch convective parameterization scheme (Kain 2004), the MRF boundary layer parameterization scheme (Hong and Pan 1996), the WSM6 microphysics parameterization scheme (Hong and Lim 2006), the Monin–Obukhov surface layer parameterization scheme (Beljaars 1995), the RRTM longwave and Dudhia shortwave radiation parameterization schemes (Mlawer et al. 1997; Dudhia 1989), and the Noah land surface parameterization scheme (Mahrt and Ek 1984).

c. Data

The background data for GRAPES-REPS are obtained by dynamical downscaling of the T639 [T639 global medium-term numerical forecast system (Guan and Chen 2008)] ensemble forecast background data. The GRAPES 15-km gridded analysis from the National Meteorological Information Center in CMA is used to verify surface and upper-air weather variables. Synoptic rain gauge observations from 2412 ground-based stations in China are used for precipitation verification (see Fig. 1 for their distribution). Note that the number of stations utilized for the construction of this dataset is somewhat limited and therefore may underestimate the precipitation, and no observation uncertainties were taken into account in the current study.

3. Descriptions of the three stochastic schemes

The spatially and temporally correlated random field plays an important role in the implementation of SPP,

TABLE 1. Multiphysics suite in GRAPES-REPS.

Ensemble member	GRAPES-REPS	
	Boundary layer physical parameterization process BL_PBL_PHYSICS	Cumulus convective physical parameterization process CU_PHYSICS
Control forecast	MRF	New Kain–Fritsch
1	MRF	Old Kain–Fritsch
2	MRF	Betts–Miller–Janjić
3	MRF	New Kain–Fritsch
4	MRF	Old Kain–Fritsch
5	MRF	Betts–Miller–Janjić
6	MRF	New Kain–Fritsch
7	MRF	Old Kain–Fritsch
8	YSU	Simplified Arakawa–Schubert
9	YSU	Betts–Miller–Janjić
10	YSU	New Kain–Fritsch
11	YSU	Simplified Arakawa–Schubert
12	YSU	Betts–Miller–Janjić
13	YSU	Old Kain–Fritsch
14	YSU	Simplified Arakawa–Schubert

SPPT, and SKEB. In GRAPES-REPS, the three stochastic schemes (SPP, SPPT, SKEB) use the same method, based on first-order Markov processes, to generate the random field (but with different settings of namelist parameters) (Li et al. 2008; Yuan et al. 2016). Descriptions of the random field and all three stochastic schemes are as follows.

a. Description of the random field

The random field $\varphi(\lambda, \phi, t)$, which is described by the first-order Markov chain with spherical harmonics expansion, and has a time–space correlated continuous horizontal structure as in Li et al. (2008), is defined as

$$\varphi(\lambda, \phi, t) = \mu + \sum_{l=1}^L \sum_{m=-l}^l \alpha_{l,m}(t) Y_{l,m}(\lambda, \phi), \quad (1)$$

where the variables λ , ϕ , and t are longitude, latitude, and time, respectively; μ is the mean of the random field $\varphi(\lambda, \phi, t)$; L is the spatial correlation scale (i.e., wave-number) of the random field; $Y_{l,m}(\lambda, \phi)$ are the spherical harmonics; and $\alpha_{l,m}(t)$ are the spectral coefficients of time-related random fields, with l and m being the total horizontal and zonal wavenumber, respectively. The evolution of $\alpha_{l,m}(t)$ is obtained by the first-order Markov chain:

$$\alpha_{l,m}(t + \Delta t) = e^{-\Delta t/\tau} \alpha_{l,m}(t) + \sqrt{\frac{4\pi\sigma^2(1 - e^{-2\Delta t/\tau})}{L(L+2)}} R_{l,m}(t), \quad (2)$$

where t and Δt are the time step of the model and the model integration time step (60 s in this case), respectively; τ denotes the temporal correlation scales of the random field and the model integration time step (60 s in this case); $R_{l,m}(t)$ follows a Gaussian distribution with zero mean and a variance of one, and are independent for each l , m , t and each ensemble member; and σ is the specified standard deviation of the random field $\varphi(\lambda, \phi, t)$. As described in Li et al. (2008), one often needs to keep a perturbed parameter within some specified bounds. Thus, a stretching function $S(\varphi, \mu)$ as in Li et al. (2008) is applied to $\varphi(\lambda, \phi, t)$ in Eq. (1) to obtain a new random field $\psi(\lambda, \phi, t)$ in Eq. (3):

$$\psi(\lambda, \phi, t) = \mu + S(\varphi, \mu)[\varphi(\lambda, \phi, t) - \mu], \quad (3)$$

which can ensure the new random field lies within specified bounds and has the ability to modify its PDF. The PDF distribution of the random field $\varphi(\lambda, \phi, t)$ is wider after stretching. The stretching function $S(\varphi, \mu)$ is given by

TABLE 2. Experiments conducted in this study.

Experiment	Initial condition	Model physical perturbation	Description
MP	Downscaling	Multiphysics	Control EPS
SPPT_SPP_SKEB	Downscaling	SPPT, SPP, and SKEB (single-physics)	Combination of SPPT, SPP, and SKEB

$$S(\varphi, \mu) = 2 - \frac{1 - \exp\left[\beta\left(\frac{\varphi - \mu}{\psi_{\max} - \mu}\right)^2\right]}{1 - \exp(\beta)}, \quad (4)$$

where $\mu = (\psi_{\max} + \psi_{\min})$ in which $\psi_{\max}$ and $\psi_{\min}$ are the upper and lower boundaries of $\psi(\lambda, \phi, t)$, and β is an empirical constant equal to -1.27 [the same as Li et al. (2008)].

The random fields for different members are independent, which is achieved by using different random seeds for the pseudo-number generator that supplies the $R_{l,m}(t)$ in Eq. (2) for different members to initiate the Markov process. Note that the perturbations are applied to latitude–longitude cross sections at each vertical level and the perturbations for each vertical level are identical.

Finally, the random field varies with several namelist parameters (e.g., σ , L , and τ). In the following, we discuss the SPP, SPPT, and SKEB schemes, which use the above equations to generate the random field but with different settings of namelist parameters. The stochastic perturbation parameter options for SPP, SPPT, and SKEB are presented in Table 3.

b. The SPP scheme

1) PARAMETER SELECTION

In this study, the 18 key parameters that may have an important impact on precipitation are selected from the new Kain–Fritsch convection (Kain and Fritsch 1990, 1993; Kain 2004), MRF planetary boundary layer (PBL) (Hong and Pan 1996), WRF single-moment 6-class (WSM6) microphysics (Hong and Lim 2006), and Monin–Obukhov (Beljaars 1995) surface layer parameterization schemes. Descriptions and the ranges of the parameters selected are presented in Table 4.

In the following, we offer a brief explanation of the motivation behind selecting the above parameters (Xu et al. 2019). The parameters and their ranges were determined based on the literature [e.g., ECMWF 2018; Reynolds et al. 2011; Baker et al. 2014; Johannesson et al. 2014; Di et al. 2015; McCabe et al. 2016] and consultations with GRAPES physics parameterization experts (J. Chen, G. Xu, and Q. Liu 2017, personal communication). The original symbols of the parameters in the references are presented as Table S1.

In the boundary layer parameterization scheme, boundary layer height is defined as the level where the bulk Richardson number reaches its critical value (BRCR) (ECMWF 2018), and Hong and Pan (1996)

TABLE 3. Stochastic perturbation parameter options for SPP, SPPT, and SKEB.

	SPP	SPPT	SKEB
Temporal correlation scale τ	12 h	6 h	6 h
Spatial correlation scale L	20	24	[50, 100]
Standard deviation σ	0.8	0.27	0.27
Specified bounds	Varying with parameters	[0.2, 1.8]	[−0.8, 0.8]
Mean of the random field μ	0.0	1.0	0.0
Constant β	−1.27	−1.27	−1.27

found that convective precipitation is particularly sensitive to the BRCR, thus the BRCR was selected. In addition, the profile shape exponent for calculating the momentum diffusivity coefficient (PFAC) is highly sensitive in the simulation of precipitation, because it directly affects the mixing intensity of turbulent eddies and development of convection in the boundary layer (Di et al. 2015). The Von Kármán constant (KARMAN), which is a constant of the logarithmic wind profile in the surface layer, and the CFAC, which is a coefficient for the Prandtl number, have also been proven to be both sensitive and important (Reynolds et al. 2011; Di et al. 2015). These parameters were therefore selected for the boundary layer parameterization scheme.

In the surface layer parameterization scheme, Zhang and Anthes (1982) found that the structure of the PBL is highly sensitive to the roughness length, and Baker et al. (2014) found that the roughness length can be indirectly perturbed through the Charnock parameter (CZO), which is the multiplier for the roughness length, and determines the magnitude of the wind speed-dependent roughness length over the oceans. The multiplier for the heat/moisture exchange coefficient (XKA) has also been proven to be sensitive and important, as the XKA value predominantly reveals the strength of the flux exchange (Di et al. 2015). The parameters CZO and XKA were therefore selected in the surface layer parameterization scheme.

In the convection parameterization scheme, the most important and uncertain parameters are the downdraft and entrainment mass flux rates, which represent mixing of the cloud with the environment (Kain 2004). The multipliers for the downdraft and entrainment mass flux rates, PD and PE, which are closely related to the downdraft and entrainment mass flux rates and physically affect the convective process, are two of the most sensitive parameters in the convective process (Yang et al. 2012; Di et al. 2015). The starting height of the

TABLE 4. Selected parameters. The identifiers of parameters and the schemes to which they apply are presented in the first and the second columns. The default values of the parameters are given in the third column, and the fourth column indicates the empirically realistic ranges of the parameters. Finally, definitions of the parameters are provided in the last column.

Parameter	Scheme	Default	Range	Description
BRCR	MRF	0.5	[0.25, 0.75]	Critical Richardson number
PFAC	MRF	2.0	[1, 3]	Profile shape exponent for calculating the momentum diffusivity coefficient
KARMAN	MRF	0.4	[0.38, 0.42]	Von Kármán constant
CFAC	MRF	7.8	[3.9, 15.6]	Coefficient for Prandtl number
XKA	Monin–Obukhov	2.4×10^{-5}	$[1.2 \times 10^{-5}, 5 \times 10^{-5}]$	Multiplier for heat/moisture exchange coefficient
CZO	Monin–Obukhov	0.0156	[0.01, 0.026]	Charnock parameter
EACRC	WSM6	1.0	[0.6, 1]	Snow/cloud water collection efficiency
DENG	WSM6	500	[300, 700]	Density of graupel (g m^{-3})
XNCR	WSM6	3×10^8	$[1 \times 10^7, 1 \times 10^9]$	Maritime cloud concentration (m^{-3})
N0R	WSM6	8×10^6	$[5 \times 10^6, 1.2 \times 10^7]$	Intercept parameter of rain (m^{-4})
PEAUT	WSM6	0.55	[0.35, 0.85]	Collection efficiency for cloud water converting to rain [kg (kg s)^{-1}]
DIMAX	WSM6	5×10^{-4}	$[3 \times 10^{-4}, 8 \times 10^{-4}]$	Limited maximum value for the cloud-ice diameter (m)
PD	New Kain–Fritsch	1.0	[0.5, 2]	Multiplier for downdraft mass flux rate
PE	New Kain–Fritsch	1.0	[0.5, 2]	Multiplier for entrainment mass flux rate
PH	New Kain–Fritsch	150	[50, 350]	Starting height of downdraft above updraft source layer (hPa)
TIMEC	New Kain–Fritsch	2400	[1800, 3600]	Average consumption time of convective available potential energy (s)
W0	New Kain–Fritsch	0.75	[0.04, 1]	Threshold vertical velocity in the trigger function (m s^{-1})
TKEMAX	New Kain–Fritsch	5.0	[3, 12]	Maximum turbulent kinetic energy (TKE) value in subcloud layer ($\text{m}^2 \text{s}^{-2}$)

downdraft above the updraft source layer (PH), which controls the structure of the downdraft, has a marked effect on the convection process (Yang et al. 2012; Di et al. 2015). In addition, the average consumption time (TIMEC) of the convective available potential energy (CAPE) efficiently controls the development of convection, and has a considerable impact on convective precipitation (Yang et al. 2012; Johannesson et al. 2014; Di et al. 2015). The average consumption time of the CAPE was also selected in the “RP” scheme (Bowler et al. 2008). The threshold vertical velocity (W0) in the trigger function is highly sensitive (Kain 2004; Li et al. 2008), and can be stochastically perturbed for ensemble forecasts (Bright and Mullen 2002). The intensity of updraft mass flux at the updraft source layer is assumed to be a function of turbulent kinetic energy (TKE) for shallow convection, and the maximum turbulent kinetic energy (TKEMAX) has been proven to be important and sensitive (Yang et al. 2012; Di et al. 2015). The parameters PD, PE, PH, TIMEC, W0, and TKEMAX were therefore selected in the convection parameterization scheme.

In the microphysics parameterization scheme, the properties of the scheme are sensitive to the size distribution of ice particles, and therefore the intercept

parameter (N0R), which directly influences the distribution of the entire range of drop sizes in the exponential distribution of rain drop size, was selected (Hacker et al. 2011a; Di et al. 2015). Additionally, Baker et al. (2014) and Di et al. (2015) confirmed the significance for precipitation of the collection efficiency for the conversion of cloud water to rain (PEAUT), and the limited maximum value for the diameter of cloud ice (DIMAX), because these parameters affect the conversion of cloud ice to rainwater. Based on the parameter sensitivity analysis work of Johannesson et al. (2014), the snow/cloud water collection efficiency (EACRC), which represents the ratio of cloud coagulation, was also selected. The coagulation between large and small cloud droplets can convert cloud droplets into precipitation, so there is a direct impact on precipitation. Finally, following the GRAPES physics parameterization experts’ suggestions, two other uncertain parameters in GRAPES-REPS were also selected: first, the density of graupel (DENG), which greatly influences the precipitation efficiency, was selected; second, the maritime cloud concentration (XNCR), which is a multiplier for the automatic conversion rate, is the direct factor of influence in the transformation of cloud water to rainwater, and thereby has an important effect on

FIG. 2. Structures of the perturbation patterns for (a) SPP, (b) SPPT, and (c) SKEB from a randomly chosen ensemble member and model time step.

precipitation. The parameters N0R, PEAUT, DIMAX, EACRC, DENG, and XNCR were therefore selected in the microphysics scheme.

2) DESIGN OF THE SPP SCHEME

The SPP scheme represents the uncertainties in parameters within physics parameterizations. In the SPP scheme, the selected 18 key parameters are stochastically perturbed at each time step, and vary temporally and spatially within a reasonable range set within the realistic empirically tuned value of each parameter. Temporal and spatial correlations are obtained through a first-order Markov process (Li et al. 2008; Berner et al. 2009). A uniform distribution (Ollinaho et al. 2017) is applied here to describe the distribution of the perturbed parameters as follows:

$$\xi_j = \exp(\psi_j) \hat{\xi}_j, \quad \psi_j \sim N(\mu, \sigma^2), \quad (5)$$

where the perturbed and unperturbed parameters are referred to as ξ_j and $\hat{\xi}_j$, respectively; different perturbed parameters are distinguished by an integer index j ($j = 1, 18$); μ is the mean of the random field and set to $\mu = 0$ for all parameters; and σ is the specified standard deviation of the random field and set to $\sigma = 0.8$ for all parameters based on the sensitivity experiments (Xu et al. 2019). The random fields for different parameters are independent, which is achieved by using different random seeds for the pseudo-number generator that supplies the $R_{l,m}(t)$ in Eq. (2) for different parameters to initiate the Markov process, and setting different realistic empirically tuned bounds for different parameters. Figure 2a illustrates the structure of the perturbation field $\exp(\psi_j)$ of the typical parameter BRCR, which is the critical bulk Richardson number from a randomly chosen ensemble member and model time step.

Note that SPP uses the same random field generator as SPPT and SKEB but with different correlation

patterns: the spatial correlation scale is set to $L = 20$ and the temporal correlation scale is set to $\tau = 12$ h for all parameters in the SPP implementation, based on sensitivity experiments on temporal and spatial decorrelations conducted in GRAPES-REPS (Xu et al. 2019).

Finally, we keep the perturbed parameters within strictly specified bounds (between $\xi_{\max}$ and $\xi_{\min}$) to prevent them from attaining physically unrealistic values. We make the following adjustments: if ξ_j is outside of the prescribed range, it is adjusted to a new parameter value:

$$\begin{aligned} \text{If } \xi_j < \xi_{\min}, \quad & \text{then } \xi_j = \xi_{\min}; \\ \text{if } \xi_j > \xi_{\max}, \quad & \text{then } \xi_j = \xi_{\max}, \end{aligned} \quad (6)$$

We adopt this enforced clipping method (Li et al. 2008) to avoid physically unrealistic values, since the tails of the lognormal distribution may extend to unphysically extreme values. However, the potential disadvantage of the clipping method, in that it may cluster the PDF at the tails, needs further justification.

c. The SPPT scheme

The SPPT scheme represents structural uncertainties associated with the physics parameterizations by perturbing the net parameterization tendencies with noise correlated in space and time (Buizza et al. 1999; Palmer et al. 2009). The net tendency term is referred to as X , and $\hat{X}$ denotes the perturbed net tendency:

$$X = \psi(\lambda, \phi, t) \hat{X}. \quad (7)$$

The random field $\psi(\lambda, \phi, t)$ is presented in Eq. (1). All namelist parameter settings of SPPT are the same as their operational configuration. The value of the standard deviation σ of the random field $\varphi(\lambda, \phi, t)$ is set to

0.27 [adopted from Charron et al. (2010)], and β in Eq. (5) is -1.27 [adopted from Li et al. (2008)]. The temporal correlation scale τ of the random field is set to 6 h, and the spatial correlation scale L is set to 24, as in the operational configuration of GRAPES-REPS. The perturbation is set within the range $[0.2\text{--}1.8]$, with a mean value of 1.0. The structure of the random field $\psi(\lambda, \phi, t)$ of SPPT is different from that of SPP and SKEB, and an example from a randomly chosen ensemble member and model time step is shown in Fig. 2b.

d. The SKEB scheme

The SKEB algorithm was originally developed and designed by Mason and Thomson (1992) and Shutts (2005) to represent model uncertainty arising from energy dissipation by stochastic perturbation of the streamfunction (the temperature may also be perturbed). In this scheme, a fraction of the dissipated energy, which acts as a streamfunction forcing for the resolved-scale flow, is backscattered upscale in the physical parameterization process.

In the implementation of SKEB in GRAPES-REPS, the temperature has not been perturbed yet, and the horizontal wind has been stochastically perturbed according to

$$\left(\frac{\partial u}{\partial t}\right)_{\text{SKEB}} = -\frac{1}{a} \frac{\partial F_{\psi}}{\partial \phi}, \quad (8)$$

$$\left(\frac{\partial v}{\partial t}\right)_{\text{SKEB}} = -\frac{1}{a \cos \phi} \frac{\partial F_{\psi}}{\partial \lambda}. \quad (9)$$

Following Shutts (2005), the stream forcing function F_{ψ} is given by

$$F_{\psi} = \frac{a\Delta x}{\Delta t} \psi(\lambda, \phi, t) \sqrt{\Delta t \hat{D}(\lambda, \phi, \eta, t)}, \quad (10)$$

where $\Delta x = 15$ km and $\Delta t = 60$ s are a typical model grid length and time step, respectively. The dissipation rate $\hat{D}(\lambda, \phi, \eta, t)$ is set to 0.03 and the adjustment coefficient a is set to 1.5, as in the operational configuration of GRAPES-REPS. All realizations of the random field $\psi(\lambda, \phi, t)$ lie within the bounds $[-0.8, 0.8]$, with a mean of 0.0 and a standard deviation of 0.27 [adopted from Charron et al. (2010)]. Since the forcing is mainly exerted at small scales, the spatial correlation scales of the random field are set within the range $[50, 100]$. The temporal correlation scale τ of the random field is set to 6 h, as in the operational configuration. The quantity β in Eq. (5) is -1.27 [adopted from Li et al. (2008)]. The structure of the random field $\psi(\lambda, \phi, t)$ of SKEB for a randomly chosen

ensemble member and model time step is shown in Fig. 2c.

4. Results

To compare the performances of the two experiments, we assessed the precipitation, zonal wind at three levels (250 hPa, 850 hPa, and 10 m), and dry-bulb temperature at three levels (500 hPa, 850 hPa, and 2 m). A set of verification measures was employed: the area under the relative operating characteristic curve (AROC) score (Mason 1982), the Brier skill score (BSS; Brier 1950; Murphy 1973; Weigel et al. 2007), and probability distribution of precipitation exceeding specific thresholds for probabilistic verification of precipitation; the fractions skill score (FSS; Roberts and Lean 2008; Roberts 2008), the spatial mean correlation coefficient and root-mean-square error (RMSE), as well as the ensemble mean frequency bias (Schaefer 1990) for deterministic verification of precipitation, and ensemble spread, RMSE, consistency (defined as the ratio of the spread to the RMSE), relative operating characteristic (ROC) curves (Mason 1982; Harvey et al. 1992), the decomposed continuous ranked probability score (CRPS) with its reliability and resolution components (Hersbach 2000), rank histograms (Talagrand et al. 1997; Hamill 2001), and outlier scores (i.e., the sum of the two end bins of the rank histograms) for verification of upper-air weather variables, 10-m zonal wind, and 2-m temperature. For a detailed discussion of these metrics, readers are referred to Jolliffe and Stephenson (2012).

Additionally, to test the statistical significance of the results, an unpaired Student's t test, in which we reject a null hypothesis at the 0.05 level of significance, was performed—the null hypothesis being that the difference between the time-averaged reference MP experiment and the time-averaged SPPT_SPP_SKEB experiment is zero. Statistically significant differences at the 95% confidence level are denoted by square points along lines. All the verification results are the monthly average for the 0000 UTC cycle between 1 and 30 June 2015.

a. Precipitation verification

1) PROBABILISTIC VERIFICATION OF PRECIPITATION

(i) AROC

The AROC score, which represents the area between the ROC curve and the no-discrimination line and measures the statistical discrimination (i.e., the ability of the forecast to discriminate among observations) capability of an EPS, is a commonly used metric for verification of probabilistic precipitation

FIG. 3. Domain-averaged AROC scores of 24-h accumulated precipitation for four thresholds (a) 0.1, (b) 10, (c) 25, and (d) 50 mm for the two experiments, varying with forecast hour. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% confidence level are denoted by filled square points along lines.

forecasts (Mason 1982). It has a range of 0–1, where a score of 1 is attained for a perfect forecast and a score of 0 indicates no skill. Figure 3 shows the AROC score (Figs. 3a–d) of 24-h accumulated precipitation for 0.1-, 10-, 25- and 50-mm thresholds. In comparison with the MP experiment, the SPPT_SPP_SKEB experiment is characterized by quite similar AROC scores for the 0.1-mm threshold (Fig. 3a), and achieves higher AROC scores for the 10- (Fig. 3b), 25- (Fig. 3c), and 50-mm (Fig. 3d) thresholds and for all forecast lead times, indicating that the SPPT_SPP_SKEB experiment is characterized by better statistical discrimination capability and generally produces more accurate forecasts, especially for heavier precipitation (above the 25-mm threshold). The improvements of the discrimination in the SPPT_SPP_SKEB experiment over the MP experiment are statistically significant at the 95% level for most of the forecast lead times and for most of the thresholds, except for the 0.1-mm threshold, where there is a

moderately statistically significant (in the 60%–80% level) improvement.

(ii) BSS

The BSS measures the improvement in skill of a probabilistic forecast relative to a reference forecast (usually the climatology) (Weigel et al. 2007), which is positively oriented and has a perfect value of 1. It is calculated following the equation:

$$\text{BSS} = 1 - \frac{\langle (f - o)^2 \rangle}{\langle (\langle o \rangle - o)^2 \rangle}, \quad (11)$$

where f and o are the forecast and observation probability, respectively, and $\langle o \rangle$ presents the “climatological” probability. The angle brackets represent an average over the entire observation set. Figure 4 shows the BSSs for the 24-h accumulated precipitation of the SPPT_SPP_SKEB and MP experiments using thresholds of 0.1, 10, 25, and 50 mm.

FIG. 4. BSS for 24-h accumulated precipitation in the SPPT_SPP_SKEB and MP experiments using precipitation thresholds of (a) 0.1, (b) 10, (c) 25, and (d) 50 mm. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% confidence level are denoted by filled square points along lines.

For all precipitation thresholds and all forecast lead times (Figs. 4a–d), both experiments achieve positive values of BSS, which indicate that the forecasts are more skillful when compared to the climatology. In addition, the SPPT_SPP_SKEB experiment is characterized by overall higher BSS values compared to the MP experiment for all thresholds at all lead times, and the improvements are statistically significant at the 95% level (t test) for most of the forecast lead times, indicating that the SPPT_SPP_SKEB experiment generally outperforms the MP experiment for all precipitation thresholds studied.

(iii) Probability distribution of precipitation

In addition to statistical analysis, probabilities of precipitation exceeding the 10- (Figs. 5a,b), 25- (Figs. 5c,d),

and 50-mm (Figs. 5e,f) thresholds for monthly mean 24-h accumulated precipitation, are also evaluated. Figure 5g shows the corresponding observation field for monthly mean 24-h accumulated precipitation. It shows that the concentrated region of precipitation occurs over the middle and lower reaches of the Yangtze River, East China Sea, and the Bay of Bengal. These areas are the typical areas of EASM precipitation in June, and their locations can be referred to in Fig. 1. A comparison of the probabilities of precipitation exceeding the 10- (Figs. 5a,b) and 25-mm (Figs. 5c,d) thresholds for the MP (Figs. 5a,c) and SPPT_SPP_SKEB (Figs. 5b,d) experiments shows higher probabilities for the SPPT_SPP_SKEB experiment (Figs. 5b,d) for the majority of the precipitation region in the southeastern region of the Qinghai–Tibet

FIG. 5. Probability distribution of precipitation exceeding (a),(b) 10-, (c),(d) 25-, and (e),(f) 50-mm thresholds for monthly mean 24-h accumulated precipitation for MP in (a), (c), and (e) and SPPT_SPP_SKEB in (b), (d), and (f); and (g) the corresponding observation field for monthly mean 24-h accumulated precipitation. The results are the monthly averages for the 0000 UTC cycle during June 2015.

Plateau, the Bay of Bengal, middle and lower reaches of the Yangtze River, and East China Sea, characterized by precipitation exceeding the specified thresholds, which indicates that the SPPT_SPP_SKEB experiment can better simulate the probability distribution of precipitation, and provides better guidance in precipitation

forecasts compared to the MP experiment. The probabilities of precipitation exceeding the 50-mm threshold (Figs. 5e,f) are also evaluated, which also indicates that the SPPT_SPP_SKEB experiment generally performs better, especially for the southeastern region of the Qinghai-Tibet Plateau and the Bay of Bengal.

2) DETERMINISTIC VERIFICATION OF PRECIPITATION

From a certain point of view, the ensembles can be considered as finite sets of deterministic forecast realizations, and the ensemble mean of the forecasts can be considered as a single deterministic forecast (Jolliffe and Stephenson 2012). Thus, more insight into the performance of the SPPT_SPPT_SKEB and MP experiments in predicting the accumulated precipitation can be seen in the deterministic verification of precipitation by employing deterministic metrics as follows:

(i) FSS

A neighborhood (or fuzzy) approach can provide a different insight into the quality of a forecast with regard to spatial information compared to traditional verification approaches (Gilleland et al. 2009, 2010). The FSS (Roberts and Lean 2008; Roberts 2008), which compares the forecast and observed fractional coverage of gridbox events in spatial windows of increasing neighborhood size, is also applied for precipitation verification. It is positively oriented and has a range of 0 (no skill) to 1 (perfect skill). According to the definition in Roberts and Lean (2008), the FSS is given by

$$\text{FSS} = 1 - \frac{\frac{1}{N} \sum_i \sum_j [O(i,j) - M(i,j)]^2}{\frac{1}{N} \sum_i \sum_j O(i,j)^2 + \frac{1}{N} \sum_i \sum_j M(i,j)^2}, \quad (12)$$

where $O(i, j)$ and $M(i, j)$ are the ensemble mean forecast and observation fractions at location (i, j) (the horizontal and vertical direction are indicated by i and j , respectively), and the fraction is defined as the proportion of the neighborhood covered by precipitation exceeding a given threshold compared to the full neighborhood. The variable N is the total number of grid points in the domain (Roberts and Lean 2008; Skok and Roberts 2016). FSS curves against neighborhood length for 24-h accumulated precipitation and for thresholds of 0.1, 10, 25, and 50 mm are displayed in Fig. 6. The FSS for the 0.1-mm threshold (Fig. 6a) indicates that the two experiments are of similar value, which implies a comparable forecast skill over all scales. For the higher thresholds of 10 mm (Fig. 6b), 25 mm (Fig. 6c), and 50 mm (Fig. 6d), the SPPT_SPP_SKEB experiment is generally more skillful over all scales than the MP experiment, indicating an improvement of forecast skill, especially for the medium- to heavy-rainfall events. Additionally, for both experiments, higher accumulation thresholds (e.g., 25, 50 mm) generally lead to a lower FSS at all scales

(Figs. 6c,d), and the FSS typically varies with the horizontal scale, with a larger horizontal scale leading to a higher FSS for all verified thresholds.

(ii) Spatial mean correlation coefficient and RMSE between the ensemble mean simulation and observation

The correlation coefficient may be a good measure of the linear association or phase error between forecasts and observations, and the RMSE measures the “average” forecast error. As shown in Table 5, the spatial mean correlation coefficient and RMSE of 24-h accumulated precipitation between the simulation and observation in June 2015 averaged over the simulated domain and over the verification period indicate that the SPPT_SPP_SKEB experiment is characterized by higher correlation coefficients and lower RMSE than the MP experiment. Specifically, the correlation coefficient of the MP experiment is 0.639 and that of the SPPT_SPP_SKEB experiment is 0.714. Moreover, the RMSE of the SPPT_SPP_SKEB experiment is 3.675, which is lower than that (3.712) of the MP experiment. The higher spatial mean correlation coefficient and lower spatial mean RMSE of the SPPT_SPP_SKEB experiment implies an improvement in the skill of simulating the spatial pattern of precipitation compared to the MP experiment.

(iii) Frequency bias

Frequency bias is calculated as the ratio of the frequency of forecast events to the frequency of observed events (Schaefer 1990), whose values higher (lower) than one implies the forecast system overpredicts (underpredicts) the exceedance of a given threshold. It has a range of zero to infinity, with a perfect value of 1.0. As shown in Fig. 7, when the ensemble mean frequency bias is evaluated for the 0.1-mm precipitation threshold (Fig. 7a), the frequency bias values are greater than one for both experiments, and the SPPT_SPP_SKEB experiment has a value closer to one. For the 10-mm threshold (Fig. 7b), the frequency bias values are higher and lower than one for the SPPT_SPP_SKEB and MP experiments, respectively, and the SPPT_SPP_SKEB experiment has a value closer to one for most of the forecast lead times. For higher precipitation thresholds (25 and 50 mm), frequency bias values are less than one for both experiments and the SPPT_SPP_SKEB experiment has higher values (Figs. 7c,d). In general, the SPPT_SPP_SKEB experiment exhibits a better representation for rainfall amounts (the value of frequency bias is closer to 1.0) than the MP experiment for all the thresholds and for most of the forecast lead times, especially for heavier precipitation (threshold above 10 mm).

FIG. 6. Graphs of FSS against neighborhood length for 24-h accumulated precipitation in the SPPT_SPP_SKEB and MP experiments using precipitation thresholds of (a) 0.1, (b) 10, (c) 25, and (d) 50 mm. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% confidence level are denoted by filled square points along lines.

Overall, all the precipitation verification results shown above indicate that the SPPT_SPP_SKEB experiment is characterized by a generally better performance compared with the MP experiment in both the probabilistic and the deterministic verification of precipitation for most of the lead times and for most of the thresholds.

b. Verification for upper-air weather variables, 10-m zonal wind, and 2-m temperature

1) ENSEMBLE SPREAD, RMSE, AND THE CORRESPONDING CONSISTENCY

First, the domain-averaged ensemble spread and RMSE, as well as the corresponding consistency—defined as the ratio of the ensemble spread to the RMSE (Leutbecher and Palmer 2008)—are calculated for the two experiments.

Figure 8 illustrates the domain-averaged RMSE and ensemble spread values (Figs. 8a,c,e,g,i,k) and consistency (Figs. 8b,d,f,h,j,l) for 250-hPa zonal wind (Figs. 8a,b), 500-hPa temperature (Figs. 8c,d), 850-hPa zonal wind (Figs. 8e,f), 850-hPa temperature (Figs. 8g,h), 10-m zonal wind (Figs. 8i,j), and 2-m temperature (Figs. 8k,l). As indicated in Berner et al. (2011), there should be approximate

TABLE 5. Spatial mean correlation coefficient and RMSE of 24-h accumulated precipitation between the simulation and observation in June 2015 averaged over the simulated domain and over the verification period.

Experiment	Correlation coefficient	RMSE
MP	0.639	3.712
SPPT_SPP_SKEB	0.714	3.675

FIG. 7. Ensemble mean frequency bias of 24-h accumulated precipitation in the SPPT_SPP_SKEB and MP experiments for four thresholds (a) 0.1, (b) 10, (c) 25, and (d) 50 mm for the two experiments, varying with forecast hour. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% confidence level are denoted by filled square points along lines.

agreement between the root-mean-square (RMS) ensemble mean error and “total spread,” which includes both the ensemble spread and observation/analysis error for a reliable ensemble. Thus, similarity between the ensemble mean error and ensemble spread is desirable, and a perfect spread–error consistency has a value of 1.0, if observation/analysis error is not taken into account. The consistency is presented in Fig. 8 (right).

As shown in Fig. 8, the SPPT_SPP_SKEB experiment is characterized by larger spread than the MP experiment, and the consistency increases from severe underdispersion (0.45–0.8) in the MP experiment to a higher value closer to one, especially for 250-hPa zonal wind (Figs. 8a,b), 850-hPa zonal wind (Figs. 8e,f), and 10-m zonal wind (Figs. 8i,j). The improvement was also seen for 500-hPa temperature (Figs. 8c,d) and 850-hPa temperature (Figs. 8g,h), but it is relatively slight compared to that for wind—the SPPT_SPP_SKEB experiment exhibits slightly higher or similar spread and consistency compared with the MP experiment for 500-hPa temperature and 850-hPa temperature. The

domain-averaged RMSE is overall similar for both experiments, which implies that the SPPT_SPP_SKEB experiment generally does not cause an increase in RMSE compared with the MP experiment. There is an exception for 2-m temperature (Figs. 8k,l), where the SPPT_SPP_SKEB experiment exhibits slightly higher RMSE and slightly lower spread compared with the MP experiment, and thus results in a slightly lower consistency compared to the MP experiment (Fig. 8l). As argued by Berner et al. (2009), Tennant et al. (2011), and Jankov et al. (2017), the SKEB algorithm can increase the spread of the ensemble and, in particular, has a larger impact on the spread for upper-level wind. Furthermore, the SKEB scheme in GRAPES-REPS only acts on the wind field and does not affect the temperature field. The above factors may explain why the SPPT_SPP_SKEB experiment has a greater impact on the ensemble spread for wind than that for temperature. In general, the improvement in spread and consistency between the SPPT_SPP_SKEB and MP experiments is statistically significant at the 99.99% level (t test) for

FIG. 8. The (left) domain-averaged RMSE of the ensemble mean for the SPPT+SPPT_SKEB (red line) and MP (blue line) experiments, control forecast RMSE (gray line) and ensemble spread for the SPPT+SPPT_SKEB (red column) and MP (blue column) experiments, and (right) consistency for (a),(b) 250-hPa zonal wind, (c),(d) 500-hPa temperature, (e),(f) 850-hPa zonal wind, (g),(h) 850-hPa temperature, (i),(j) 10-m zonal wind,

FIG. 9. Horizontal distributions of (a),(c) ensemble spread and (b),(d) RMSE of the ensemble mean for the (top) MP and (bottom) SPPT_SPP_SKEB experiments, for 850-hPa zonal wind at the 48-h forecast lead time. The results are the monthly average for the 0000 UTC cycle during June 2015.

250-, 850-hPa, and 10-m zonal wind for all lead times, and at the 95% level (t test) for 500- and 850-hPa temperature. The differences in RMSE for all variables and the differences in spread and consistency for 2-m temperature are not statistically significant.

In summary, the SPPT_SPP_SKEB experiment increases the ensemble spread without causing an increase in the forecast mean error (with the exception of a slight increase in RMSE of 2-m temperature). Therefore, the overall spread–error consistency is improved, indicating that the model is better at capturing error growth and the flow-dependent forecast uncertainty is better represented by the ensemble spread, which results in a more reliable EPS. However, the underdispersion has not been entirely fixed, since both experiments are characterized by spread values well below their corresponding RMSE for almost every variable and lead time. This

may be because either the multiphysics or the multi-stochastic scheme is unable to address all sources of uncertainty, and observation error has not been taken into account; plus, model systematic errors that primarily derive from inherent deficiencies in the model [e.g., grid resolution, finite differences, representation of physics, dynamics (Harr et al. 1983; Krishnamurti et al. 2016)] may also contribute to the underdispersion in terms of verification metrics (Wang et al. 2018).

In addition to the domain-averaged curves of spread and RMSE, the horizontal distributions (latitude–longitude cross sections) of the ensemble spread and RMSE of the ensemble mean for the MP and SPPT_SPP_SKEB experiments at the 48-h forecast lead time are also evaluated. As shown in Fig. 9, for 850-hPa zonal wind (Figs. 9a–d), the SPPT_SPP_SKEB experiment is characterized by a larger spread (Fig. 9c) and a similar or

←

and (k),(l) 2-m temperature, varying with forecast hour. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% confidence level are denoted by filled square points along lines.

FIG. 10. As in Fig. 9, but for 500-hPa temperature.

slightly higher RMSE of the ensemble mean (Fig. 9d) compared to that of the MP experiment (Figs. 9a,b) all over the domain. Besides, for 500-hPa temperature (Figs. 10a–d), the two experiments exhibit generally similar performance over the domain. Specifically, the SPPT_SPP_SKEB experiment produces a slightly higher ensemble spread and slightly higher RMSE of the ensemble mean compared with the MP experiment in the Qinghai–Tibet Plateau and northeastern China. The slightly inferior performance for the RMSE of temperature may indicate that the SPPT_SPP_SKEB experiment offers limited improvements to the ensemble mean error relative to the MP experiment. In general, for both the 850-hPa zonal wind (Figs. 9a–d) and the 500-hPa temperature (Figs. 10a–d), the magnitude of ensemble spread (Figs. 9a,c and 10a,c) is generally small compared to the RMSE of the ensemble mean (Figs. 9b,d and 10b,d) over the domain for both experiments at the 48-h forecast lead time. This indicates that both experiments are underdispersive, as seen in the corresponding domain-averaged consistency shown in Figs. 8f and 8d, where we can also see that this underdispersion is true not only at the 48-h forecast lead time but for all forecast lead times.

2) THE CRPS TOGETHER WITH ITS RELIABILITY AND RESOLUTION COMPONENTS

As an additional measure of ensemble performance, the CRPS is also calculated, for measuring the mean absolute error between the forecast probability and observations (Hersbach 2000). The CRPS is defined as

$$\text{CRPS} = \frac{1}{N} \sum_{i=1}^N \int_{x=-\infty}^{x=\infty} [F_i^f(x) - F_i^o(x)]^2 dx, \quad (13)$$

where $F_i^f(x)$ and $F_i^o(x)$ are the forecast and observation probability cumulative distribution function (CDF), respectively, for the i th case; and N is the total number of forecast–observation pairs. Note that lower values of CRPS denote a better forecast skill. The CRPS can be decomposed into three parts: a reliability part, a resolution part, and an uncertainty part (Hersbach 2000). The reliability part (negatively oriented) tests whether for each bin i on average the verifying analysis is found to be with a fraction i/N below this bin, and therefore it tests here whether the ensemble can generate cumulative distributions that have, on average, this desired statistical property. The resolution part (positively oriented) expresses the improvement gained by issuing probability forecasts that are case dependent, which

FIG. 11. The CRPS, together with its reliability and resolution components, for the (a) 250-hPa zonal wind, (b) 500-hPa temperature, (c) 850-hPa zonal wind, (d) 850-hPa temperature, (e) 10-m zonal wind, and (f) 2-m temperature. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% confidence level are denoted by filled square points along lines.

would have positive values if the EPS performs better than the climatological probabilistic forecast. The uncertainty part is the best achievable value of the CRPS, in cases where only a PDF-based climatology is available (Hersbach 2000; Candille and Talagrand 2005). Figure 11 shows the CRPS together with its reliability and resolution components. For most of the variables and most forecast lead times, the SPP_SKEB_SPPT experiment is characterized by a slightly lower (higher skill) or similar CRPS. Specifically, it generally yields a 1%–4% improvement in the CRPS compared with the MP experiment and there is a moderately statistically significant (in the 70%–90% level) improvement by the SPPT_SPP_SKEB experiment over the MP experiment, which implies that the SPPT_SPP_SKEB experiment provides limited additional improvements to the CRPS relative to the MP experiment. The reliability component of the CRPS for the SPPT_SPP_SKEB experiment is marginally, but statistically significantly, lower than that of the MP experiment for 850-hPa zonal wind

(Fig. 11c) and 10-m zonal wind (Fig. 11e), and there is a moderately statistically significant (in the 80% level) improvement for the 250-hPa zonal wind (Fig. 11a), which implies that the SPPT_SPP_SKEB experiment is capable of improving the reliability for wind, especially for low-level wind. Additionally, there is a moderately statistically significant (in the 60%–80% level) improvement of the reliability by the SPPT_SPP_SKEB experiment over the MP experiment for 850-hPa temperature (Fig. 11d) and 2-m temperature (Fig. 11f), indicating that the improvement of reliability for temperature is generally slight compared with that for wind. Note that the reliability component of the CRPS for 500-hPa temperature (Fig. 11b) is very close to zero for both experiments. In terms of the resolution component of the CRPS, the SPPT_SPP_SKEB experiment exhibits generally higher resolution than that of the MP experiment for the 250-hPa zonal wind (Fig. 11a) and 850-hPa zonal wind (Fig. 11c), indicating a better skill. The resolution component for 10-m zonal wind (Fig. 11e) is

statistically significantly improved in SPPT_SPP_SKEB over MP at the forecast lead time of 24 h (and before) but degraded at longer forecast lead times. Besides, the SPPT_SPP_SKEB and MP experiments have quite similar resolution components of CRPS for 500-hPa, 850-hPa, and 2-m temperature, and there is a moderately statistically significant (in the 60%–80% level) improvement of the resolution by the SPPT_SPP_SKEB experiment over the MP experiment, indicating a relatively slight improvement for temperature. In general, the SPPT_SPP_SKEB performs from neutral to slightly better than the MP experiment, and both the reliability and resolution components of the CRPS are improved compared to the MP experiment, especially for the wind. The CRPS results along with the findings from the RMSE may suggest that the SPPT_SPP_SKEB experiment provides limited additional improvements to the mean error relative to the MP experiment, based on the fact that the multiphysics has been proven to have merit in removing the mean error, as suggested by [Berner et al. \(2011, 2017\)](#). For the SPPT_SPP_SKEB experiment, since perturbing the parameters independently in the SPP scheme may bring in systematic errors into the system, taking the interactions between the perturbed parameters into consideration may further improve the skill.

3) ROC CURVE

The ROC is a curve of the hit rate against the false alarm rate over a range of different thresholds, which can be used to diagnose forecast accuracy, as proposed by [Harvey et al. \(1992\)](#). Perfect performance is represented on ROC axes by the upper-left-hand corner (e.g., [Mason and Mimmack 2002](#)). [Figure 12](#) shows ROC curves for the forecast lead time at 48 h for the SPPT_SPP_SKEB and MP experiments. Note that the ROC curves at other forecast lead times (e.g., 12, 24, 36 h) show similar conclusions as those for 48 h and are thus not shown here. As apparent in the ROC curves for 250-hPa zonal wind ([Fig. 12a](#)), 850-hPa zonal wind ([Fig. 12c](#)), 850-hPa temperature ([Fig. 12d](#)), and 10-m zonal wind ([Fig. 12e](#)), the SPPT_SPP_SKEB experiment shows an advantage over the MP experiment, indicating lower false alarm rates for specific hit rates in the SPPT_SPP_SKEB experiment. However, for the 500-hPa temperature ([Fig. 12b](#)) and 2-m temperature ([Fig. 12f](#)), the SPPT_SPP_SKEB experiment exhibits similar or slightly worse performance compared with the MP experiment, while the differences are not statistically significant.

4) RANK HISTOGRAMS AND THE “OUTLIERS”

Another common verification metric is the rank histogram ([Talagrand et al. 1997](#); [Hamill and Colucci 1997](#)),

which is used to check where the verifying observation usually falls with respect to the ensemble forecast data, ranked in increasing order at each grid point. If an EPS is reliable, the ensemble members and observations should be statistically indistinguishable. Therefore, for an ideal EPS that represents forecast uncertainty well, a flat distribution should be expected ([Talagrand et al. 1997](#)). For the 250-hPa zonal wind ([Fig. 13a](#)), 850-hPa zonal wind ([Fig. 13e](#)), and 10-m zonal wind ([Fig. 13i](#)), the distributions of the SPPT_SPP_SKEB experiment are more uniform compared to those of the MP experiment, which indicate ensemble spread is more reliable and better reflect true error distribution in the SPPT_SPP_SKEB experiment. Moreover, the diagram for 250-hPa zonal wind ([Fig. 13a](#)) is generally flat for both experiments, which implies that GRAPES-REPS captures the observed data well for 250-hPa zonal wind. However, the “U”-shaped histogram observed for 10-m zonal wind ([Fig. 13i](#)) is a characteristic of underdispersion (i.e., spread under estimates uncertainty) for both experiments: since the spread for 10-m zonal wind ([Fig. 8i](#)) is low on average, there is a probability that an observation may not be captured by the ensemble. Additionally, there is a notable tendency for the distributions to be most populated at the extreme ranks for the diagrams of the 500-hPa temperature ([Fig. 13c](#)), 850-hPa temperature ([Fig. 13g](#)), and 2-m temperature ([Fig. 13k](#)), indicating that ensemble forecasts are overdispersive (i.e., spread overestimates uncertainty) and capture the observation too often, which may be attributable to a systematic bias in the forecast or insufficient variability among the members.

The sum of the two end bins of the rank histograms (the “outliers”) refers to the frequency of when the observations fall outside the ensemble envelope. Lower values indicate a more reliable ensemble. For a perfect EPS with n members, the outlier should be $2/(n + 1)$. [Figure 13](#) (right) shows outlier scores of the two experiments. The SPPT_SPP_SKEB experiment is characterized by notably lower outliers compared with the MP experiment for 250-hPa zonal wind ([Fig. 13b](#)), 850-hPa zonal wind ([Fig. 13f](#)), and 10-m zonal wind ([Fig. 13j](#)), which indicates a notable improvement for wind. However, the improvement for 500-hPa temperature ([Fig. 13d](#)) and 850-hPa temperature ([Fig. 13h](#)) is not so dramatic, and for 2-m temperature ([Fig. 13l](#)) the SPPT_SPP_SKEB experiment is characterized by slightly higher outlier scores than the MP experiment. In general, outlier scores are improved by the SPPT_SPP_SKEB experiment compared with the MP experiment, and the improvement is statistically significant at the 99.99% level (t test) for wind and at the 95% level for 500-hPa and 850-hPa temperature at all lead times, but not

FIG. 12. ROC diagrams for the SPPT_SPP_SKEB experiment (red) and MP experiment (blue) for (a) 250-hPa zonal wind, (b) 500-hPa temperature, (c) 850-hPa zonal wind, (d) 850-hPa temperature, (e) 10-m zonal wind, and (f) 2-m temperature at the 48-h forecast lead time. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% confidence level are denoted by filled square points along lines.

statistically significant for most of the lead times for 2-m temperature.

5. Summary and discussion

In this study, a single-physics suite combined with multi-stochastic schemes (SPP, SPPT, and SKEB) is applied in

GRAPES-REPS with a horizontal grid spacing of 15km, to investigate whether it can be an alternative to the multi-physics suite. Two experiments—SPPT_SPP_SKEB and MP—are performed for a summer monsoon month (1–30 June 2015) over China. Forecasts are initialized at 0000 UTC and integrated for 48-h. The SPP, SPPT, and SKEB schemes use the same method, based on

FIG. 13. (left) Rank histograms and (right) outlier scores at the 48-h forecast lead time for (a),(b) 250-hPa zonal wind, (c),(d) 500-hPa temperature, (e),(f) 850-hPa zonal wind, (g),(h) 850-hPa temperature, (i),(j) 10-m zonal wind, and (k),(l) 2-m temperature, varying with forecast hour. The results are the monthly averages for the 0000 UTC cycle during June 2015. Statistically significant improvements at the 95% - confidence level are denoted by filled square points along lines.

first-order Markov processes, to generate the temporally and spatially varying random field (but with different settings of namelist parameters). In the SPP scheme, a total of 18 key parameters selected from the Kain–Fritsch convection, WSM6 microphysics, MRF PBL, and Monin–Obukhov surface layer parameterization schemes are temporally and spatially perturbed. In the SPPT scheme, net parameterization tendencies are perturbed with noise correlated in space and time. In the implementation of SKEB in GRAPES-REPS, the horizontal wind has been stochastically perturbed. To evaluate and compare the general performance of the two experiments, various verification metrics are employed. The six verified variables (250-hPa zonal wind, 500-hPa temperature, 850-hPa zonal wind, 850-hPa temperature, 10-m zonal wind, and 2-m temperature) are clustered into “notably improved” (250-hPa zonal wind, 850-hPa zonal wind, and 10-m zonal wind) and “slightly improved” (500-hPa temperature, 850-hPa temperature, and 2-m temperature) situations for discussing the results. The main conclusions are as follows:

For precipitation, the SPPT_SPP_SKEB experiment produces generally more skillful precipitation forecasts than the MP experiment, especially for heavier precipitation thresholds (above 10 mm). The AROC, BSS, FSS, and ensemble mean frequency bias of the SPPT_SPP_SKEB experiment are improved by about 6%, 10%, 5%, and 7%, respectively, at most of the forecast lead times, and the improvements are statistically significant at the 95% level for most of the forecast lead times and thresholds, except for the 0.1-mm threshold for the AROC score and FSS. Furthermore, the SPPT_SPP_SKEB experiment can better simulate the probability distribution of precipitation compared to the MP experiment. Additionally, the SPPT_SPP_SKEB experiment is characterized by a higher spatial mean correlation coefficient and lower RMSE of the observations and simulations, which implies an improvement in the skill of simulating the spatial pattern of the precipitation compared to the MP experiment. Overall, the SPPT_SPP_SKEB experiment is characterized by a generally better performance compared with the MP experiment in the precipitation verification.

For upper-air zonal winds and 10-m zonal wind, the SPPT_SPP_SKEB experiment yields an overall notable improvement compared to the MP experiment. The ensemble spread and consistency increase by about 26% and 28%, respectively, compared to that of the MP experiment at most of the lead times. The domain-averaged RMSE is overall similar for both experiments, which implies that the SPPT_SPP_SKEB experiment does not cause an increase in RMSE compared with the

MP experiment for wind. The outliers reduce by about 32% for wind at most of the lead times. The improvements for ensemble spread, consistency, and outliers are statistically significant at the 95% level for most of the forecast lead times. In addition, the SPPT_SPP_SKEB experiment generally yields a 2%–4% improvement in the CRPS over the MP experiment and has a statistically significant lower reliability component and a statistically significant higher-resolution component of the CRPS than that of the MP experiment, especially for low-level wind, indicating a generally better skill. Similar results are also shown in the ROC curves. Overall, the SPPT_SPP_SKEB experiment shows an advantage over the MP experiment for wind.

For upper-air temperature and 2-m temperature, the SPPT_SPP_SKEB experiment generally yields a slight improvement compared to the MP experiment. The spread and consistency of the SPPT_SPP_SKEB experiment increase by about 5% and 4%, respectively, compared to that of the MP experiment at most of the forecast lead times, and the improvements are statistically significant at the 95% level for most of the lead times. Moreover, the SPPT_SPP_SKEB experiment produces a similar or slightly higher RMSE (e.g., for 2-m temperature) compared to the MP experiment, while the differences in RMSE are not statistically significant. The outliers reduce by about 7% for 500- and 850-hPa temperature at most of the lead times; however, they increase by about 1% for 2-m temperature. In addition, the SPPT_SPP_SKEB experiment produces a quite similar CRPS (and RMSE, as mentioned above) with the MP experiment for temperature, which may suggest that the SPPT_SPP_SKEB experiment provides limited additional improvements to the mean error relative to the MP experiment for temperature. Moreover, there is a moderately statistically significant (in the 60%–80% level) improvement of the reliability and resolution component by the SPPT_SPP_SKEB experiment over the MP experiment for temperature. Afterward, for the ROC curves, the SPPT_SPP_SKEB experiment shows a similar or slightly worse performance (e.g., for 2-m temperature) compared with the MP experiment for temperature. In general, the SPPT_SPP_SKEB experiment provides limited added value to the overall performance for temperature relative to the MP experiment, and the improvements for temperature are not so dramatic as for wind, being only slight. This might be because the SKEB scheme, which possibly contributes to improving the overall performance (as suggested in [Bernier et al. 2009](#)), only acts on the wind field and does not affect the temperature field in GRAPES-REPS. On the other hand, even if SKEB does not act on the temperature field in

GRAPES-REPS, the SPPT_SPP_SKEB and MP experiments perform comparably for the temperature. If SKEB is further improved in the future (i.e., simultaneously perturbing the wind and temperature fields in our implementation), better results and improvements may be obtained.

In summary, the above verification results indicate that a single-physics suite combining SPP, SPPT, and SKEB is characterized by an overall better performance compared with the multiphysics suite in precipitation verification and verification for upper-air variables, 10-m zonal wind, and 2-m temperature in the East Asian monsoon region. Our results are consistent with the previous findings of Berner et al. (2011, 2017) and Hacker et al. (2011a,b) in that model error can be better represented by a combination of model-error schemes than a single scheme alone. In addition, while Charron et al. (2010) stressed that using a single set of subgrid-scale parameterizations with and without stochastic perturbations is not as skillful as the multiphysics suite, our results illustrate that a single-physics suite combining multistochastic schemes (SPP, SPPT, and SKEB) is characterized by generally better performance compared with the multiphysics suite, and may be considered as an appropriate and attractive alternative to a multiphysics suite in the design of future regional and global ensembles.

Acknowledgments. We are grateful to Dr. Jiong Chen, Dr. Qijun Liu, and Dr. Guoqiang Xu for their kind instruction and fruitful suggestions on choosing the parameters within the SPP scheme. Constructive comments from three anonymous reviewers have significantly improved this revised manuscript. This work is sponsored by The National Key Research and Development (R&D) Program of the Ministry of Science and Technology of China Grants 2018YFC1507405 and 2015BAC03B01.

REFERENCES

- Alhamed, A., S. Lakshmivarahan, and D. J. Stensrud, 2002: Cluster analysis of multimodel ensemble data from SAMEX. *Mon. Wea. Rev.*, **130**, 226–256, [https://doi.org/10.1175/1520-0493\(2002\)130<0226:CAOMED>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0226:CAOMED>2.0.CO;2).
- Arnold, H. M., I. M. Moroz, and T. N. Palmer, 2013: Stochastic parametrizations and model uncertainty in the Lorenz '96 system. *Philos. Trans. Roy. Soc.*, **371A**, 20110479, <https://doi.org/10.1098/rsta.2011.0479>.
- Baker, L. H., A. C. Rudd, and S. Migliorini, 2014: Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes Geophys.*, **21**, 19–39, <https://doi.org/10.5194/npg-21-19-2014>.
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, <https://doi.org/10.1038/nature14956>.
- Beck, J., F. Bouttier, and L. Wiegand, 2016: Development and verification of two convection-allowing multi-model ensembles over Western Europe. *Quart. J. Roy. Meteor. Soc.*, **142**, 2808–2826, <https://doi.org/10.1002/qj.2870>.
- Beljaars, A. C. M., 1995: The parametrization of surface fluxes in large-scale models under free convection. *Quart. J. Roy. Meteor. Soc.*, **121**, 255–270, <https://doi.org/10.1002/qj.49712152203>.
- 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**, 603–626, <https://doi.org/10.1175/2008JAS2677.1>.
- , S. Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, <https://doi.org/10.1175/2010MWR3595.1>.
- , and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, <https://doi.org/10.1175/BAMS-D-15-00268.1>.
- Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, <https://doi.org/10.1002/qj.234>.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Bright, D. R., and S. L. Mullen, 2002: Short-range ensemble forecasts of precipitation during the Southwest monsoon. *Wea. Forecasting*, **17**, 1080–1100, [https://doi.org/10.1175/1520-0434\(2002\)017<1080:SREFOP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<1080:SREFOP>2.0.CO;2).
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, <https://doi.org/10.1002/qj.49712556006>.
- , P. L. 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, <https://doi.org/10.1175/MWR2905.1>.
- Candille, G., and O. Talagrand, 2005: Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**, 2131–2150, <https://doi.org/10.1256/qj.04.71>.
- Charron, M., G. Pellerin, and L. Spacek, 2010: Toward random sampling of model error in the Canadian Ensemble Prediction System. *Mon. Wea. Rev.*, **138**, 1877–1901, <https://doi.org/10.1175/2009MWR3187.1>.
- Chen, J., and S. Bordon, 2014: Orographic effects of the Tibetan Plateau on the East Asian summer monsoon: An energetic perspective. *J. Climate*, **27**, 3052–3072, <https://doi.org/10.1175/JCLI-D-13-00479.1>.
- , D. Chen, and H. Yuan, 2002: A brief review on the development of ensemble prediction system. (in Chinese) *Yingyong Qixiang Xuebao*, **13**, 497–507.
- Christensen, H. M., I. M. Moroz, and T. N. Palmer, 2015: Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *J. Atmos. Sci.*, **72**, 2525–2544, <https://doi.org/10.1175/JAS-D-14-0250.1>.
- Di, Z., Q. Duan, and W. Gong, 2015: Assessing WRF model parameter sensitivity: A case study with 5 days summer precipitation forecasting in the Greater Beijing area. *Geophys. Res. Lett.*, **42**, 579–587, <https://doi.org/10.1002/2014GL061623>.
- Ding, Y., 2004: Seasonal march of the East Asian summer monsoon. *The East Asian Monsoon*, C.-P. Chang, Ed., World Scientific, 3–53, https://doi.org/10.1142/9789812701411_0001.

- , and J. C. L. Chan, 2005: The East Asian summer monsoon: An overview. *Meteor. Atmos. Phys.*, **89**, 117–142, <https://doi.org/10.1007/s00703-005-0125-z>.
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, [https://doi.org/10.1175/1520-0469\(1989\)046<3077:NSOCOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2).
- Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350, <https://doi.org/10.1175/WAF843.1>.
- ECMWF, 2018: IFS documentation—cycle CY45r1—Part IV: Physical processes. ECMWF, 223 pp., <https://www.ecmwf.int/en/publications/ifs-documentation>.
- Flohn, H., 1957: Large-scale aspects of the “summer monsoon” in South and East Asia. *J. Meteor. Soc. Japan*, **35A**, 180–186, https://doi.org/10.2151/jmsj1923.35a.0_180.
- Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430, <https://doi.org/10.1175/2009WAF2222269.1>.
- , —, —, and E. E. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.*, **91**, 1365–1376, <https://doi.org/10.1175/2010BAMS2819.1>.
- Guan, C., and Q. Chen, 2008: Experiments and evaluations of global medium range forecast system of T639L60 (in Chinese). *Meteor. Monogr.*, **34**, 11–16.
- Hacker, J. P., C. Snyder, S.-Y. Ha, and M. Pocerich, 2011a: Linear and non-linear response to parameter variations in a mesoscale model. *Tellus*, **63A**, 429–444, <https://doi.org/10.1111/j.1600-0870.2010.00505.x>.
- , and Coauthors, 2011b: The U.S. Air Force Weather Agency’s mesoscale ensemble: Scientific description and performance results. *Tellus*, **63A**, 625–641, <https://doi.org/10.1111/j.1600-0870.2010.00497.x>.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, [https://doi.org/10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2).
- , and S. J. Colucci, 1997: Verification of Eta–RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, [https://doi.org/10.1175/1520-0493\(1997\)125<1312:VOERSR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2).
- Harr, P. A., T. L. Tsui, and L. R. Brody, 1983: Identification of systematic errors in a numerical weather forecast. *Mon. Wea. Rev.*, **111**, 1219–1227, [https://doi.org/10.1175/1520-0493\(1983\)111<1219:IOSEIA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111<1219:IOSEIA>2.0.CO;2).
- Harvey, L. O., Jr., K. R. Hammond, C. M. Lusk, and E. F. Mross, 1992: The application of signal detection theory to weather forecasting behavior. *Mon. Wea. Rev.*, **120**, 863–883, [https://doi.org/10.1175/1520-0493\(1992\)120<0863:TAOSDT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0863:TAOSDT>2.0.CO;2).
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Hong, S. Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339, [https://doi.org/10.1175/1520-0493\(1996\)124<2322:NBLVDI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2).
- , and J. O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- Houtekamer, P., L. Lefavre, J. Derome, H. Richie, and H. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242, [https://doi.org/10.1175/1520-0493\(1996\)124<1225:ASSATE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2).
- Hsu, H. H., and X. Liu, 2003: Relationship between the Tibetan Plateau heating and East Asian summer monsoon rainfall. *Geophys. Res. Lett.*, **30**, 2066, <https://doi.org/10.1029/2003GL017909>.
- Jankov, I., J. Berner, and J. Beck, 2017: A performance comparison between multi-physics and stochastic approaches within a North American RAP ensemble. *Mon. Wea. Rev.*, **145**, 1161–1179, <https://doi.org/10.1175/MWR-D-16-0160.1>.
- Johannesson, G., D. Lucas, Y. Qian, L. P. Swile, and T. M. Wildey, 2014: Sensitivity of precipitation to parameter values in the Community Atmosphere Model version 5. Sandia Tech Rep. SAND2014-0829, Sandia National Laboratories, Albuquerque, NM, <https://doi.org/10.2172/1204103>.
- Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*. 2nd ed. John Wiley and Sons, 292 pp.
- Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181, [https://doi.org/10.1175/1520-0450\(2004\)043<0170:TKCPAU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2).
- , and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802, [https://doi.org/10.1175/1520-0469\(1990\)047<2784:AODEPM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2).
- , and —, 1993: Convective parameterization for mesoscale models: the Fritsch–Chappell scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr., No. 24, Amer. Meteor. Soc., 165–170.
- Kang, I.-S., and Coauthors, 2002: Intercomparison of the climatological variations of Asian summer monsoon precipitation simulated by 10 GCMs. *Climate Dyn.*, **19**, 383–395, <https://doi.org/10.1007/s00382-002-0245-9>.
- Krishnamurti, T. N., C. M. Kishtawal, T. E. Larow, D. R. Bachiochi, Z. Zhang, E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multi-model superensemble. *Science*, **285**, 1548–1550, <https://doi.org/10.1126/science.285.5433.1548>.
- , V. Kumar, A. Simon, A. Bhardwaj, T. Ghosh, and R. Ross, 2016: A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes. *Rev. Geophys.*, **54**, 336–377, <https://doi.org/10.1002/2015RG000513>.
- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, <https://doi.org/10.1016/j.jcp.2007.02.014>.
- , S. Lock, and P. Ollinaho, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, <https://doi.org/10.1002/qj.3094>.
- Li, X., M. Charron, and L. Spacek, 2008: A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.*, **136**, 443–462, <https://doi.org/10.1175/2007MWR2109.1>.
- Lierwalqui, M. V., H. Morrison, M. R. Kumjian, O. P. Prat, and K. J. Reimel, 2018: How best to add structural complexity to cloud microphysics parameterization schemes? *2018 Fall Meeting*, Washington, DC, Amer. Geophys. Union, Abstract NG33B-0961.
- Long, K. J., J. Chen, X. Ma, and Y. Ji, 2011: Preliminary study on regional set forecasting based on set Kalman transform (in Chinese). *J. Chengdu Univ. Info. Technol.*, **26** (1), 40–49.
- Mahrt, L., and M. Ek, 1984: The influence of atmospheric stability on potential evaporation. *J. Climate Appl. Meteor.*, **23**, 222–234, [https://doi.org/10.1175/1520-0450\(1984\)023<0222:TIOASO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1984)023<0222:TIOASO>2.0.CO;2).

- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Mason, P. J., and D. J. Thomson, 1992: Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.*, **242**, 51–78, <https://doi.org/10.1017/S0022112092002271>.
- Mason, S. J., and G. M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Climate*, **15**, 8–29, [https://doi.org/10.1175/1520-0442\(2002\)015<0008:COSSMO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0008:COSSMO>2.0.CO;2).
- McCabe, A., R. Swinbank, W. Tennant, and A. Lock, 2016: Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting. *Quart. J. Roy. Meteor. Soc.*, **142**, 2897–2910, <https://doi.org/10.1002/qj.2876>.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Lacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, <https://doi.org/10.1029/97JD00237>.
- Molteni, F., R. Buizza, T. Palmer, and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, <https://doi.org/10.1002/qj.49712252905>.
- Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600, [https://doi.org/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2).
- Ninomiya, K., and T. Akiyama, 1992: Multi-scale features of Baiu, the summer monsoon over Japan and the East Asia. *J. Meteor. Soc. Japan*, **70**, 467–495, https://doi.org/10.2151/jmsj1965.70.1B_467.
- Ollinaho, P., S. J. Lock, M. Leutbecher, P. Bechtold, A. Beljaars, and A. Bozzo, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quart. J. Roy. Meteor. Soc.*, **143**, 408–422, <https://doi.org/10.1002/qj.2931>.
- Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, <https://doi.org/10.1002/qj.49712757202>.
- , 2012: Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.*, **138**, 841–861, <https://doi.org/10.1002/qj.1923>.
- , F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *Proc. Seminar on Validation of Models over Europe*, Vol. 1, Reading, United Kingdom, ECMWF, 21–66.
- , R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parameterization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp., <https://www.ecmwf.int/en/elibrary/11577-stochastic-parametrization-and-model-uncertainty>.
- Reynolds, C. A., J. G. McLay, J. S. Goerss, E. A. Serra, D. Hodyss, and C. R. Sampson, 2011: Impact of resolution and design on the U.S. Navy global ensemble performance in the tropics. *Mon. Wea. Rev.*, **139**, 2145–2155, <https://doi.org/10.1175/2011MWR3546.1>.
- Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169, <https://doi.org/10.1002/met.57>.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, [10.1175/2007MWR2123.1](https://doi.org/10.1175/2007MWR2123.1).
- Romine, G. S., C. S. Schwartz, and J. Berner, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541, <https://doi.org/10.1175/MWR-D-14-00100.1>.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, [https://doi.org/10.1175/1520-0434\(1990\)005<0570:TCSIAA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2).
- Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, <https://doi.org/10.1256/qj.04.106>.
- , 2015: A stochastic convective backscatter scheme for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **141**, 2602–2616, <https://doi.org/10.1002/qj.2547>.
- Skok, G., and N. Roberts, 2016: Analysis of Fractions Skill Score properties for random precipitation fields and ECMWF forecasts. *Quart. J. Roy. Meteor. Soc.*, **142**, 2599–2610, <https://doi.org/10.1002/qj.2849>.
- Stainforth, D. A., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406, <https://doi.org/10.1038/nature03301>.
- Stensrud, D. J., J. W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, [https://doi.org/10.1175/1520-0493\(2000\)128<2077:UICAMP>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2).
- Tao, S., and L. Chen, 1987: A review of recent research on the East Asian summer monsoon in China. *Monsoon Meteorology*, C.-P. Chang and T. N. Krishnamurti, Eds., Oxford University Press, 60–92.
- Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability*, Reading, United Kingdom, ECMWF, 1–25.
- Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190–1206, <https://doi.org/10.1175/2010MWR3430.1>.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, [https://doi.org/10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2).
- Wang, B., and T. Li, 2004: East Asian monsoon and ENSO interaction. *East Asian Monsoon*, C.-P. Chang, Ed., World Scientific, 172–212, https://doi.org/10.1142/9789812701411_0005.
- Wang, J., J. Chen, J. Du, Y. Zhang, Y. Xia, and G. Deng, 2018: Sensitivity of ensemble forecast verification to model bias. *Mon. Wea. Rev.*, **146**, 781–796, <https://doi.org/10.1175/MWR-D-17-0223.1>.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: Generalization of the discrete Brier and ranked probability skill scores for weighted multimodel ensemble forecasts. *Mon. Wea. Rev.*, **135**, 2778–2785, <https://doi.org/10.1175/MWR3428.1>.
- Wolff, J., I. Jankov, J. Beck, L. Carson, J. Frimel, M. Harrold, and H. Jiang, 2016: Addressing model uncertainty through stochastic parameter perturbations within the High Resolution Rapid Refresh (HRRR) ensemble. *2016 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract IN33A-1801.
- Wu, D., Z. Meng, and D. Yan, 2013: The predictability of a squall line in South China on 23 April 2007. *Adv. Atmos. Sci.*, **30**, 485–502, <https://doi.org/10.1007/S00376-012-2076-X>.

- Wu, G., and Coauthors, 2007: The influence of mechanical and thermal forcing by the Tibetan Plateau on Asian climate. *J. Hydrometeor.*, **8**, 770–789, <https://doi.org/10.1175/JHM609.1>.
- Xu, Z., J. Chen, Y. Wang, H. Li, F. Chen, and Y. Fan, 2019: Sensitivity test of Stochastically Perturbed Parameterizations (SPP) scheme for mesoscale precipitation ensemble prediction (in Chinese). *Acta Meteor. Sin.*, **77**, 849–868.
- Yang, B., Y. Qian, G. Lin, R. Leung, and Y. Zhang, 2012: Some issues in uncertainty quantification and parameter tuning: A case study of convective parameterization scheme in the WRF regional climate model. *Atmos. Chem. Phys.*, **12**, 2409–2427, <https://doi.org/10.5194/acp-12-2409-2012>.
- Yuan, Y., X. Li, J. Chen, and Y. Xia, 2016: Stochastic parameterization toward model uncertainty for the GRAPES meso-scale ensemble prediction system (in Chinese). *Meteor. Mon.*, **42**, 1161–1175, <https://doi.org/10.7519/j.issn.1000-0526.2016.10.001>.
- Zhang, D., and A. R. Anthes, 1982: A high-resolution model of the planetary boundary layer-sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594–1609, [https://doi.org/10.1175/1520-0450\(1982\)021<1594:AHMOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1982)021<1594:AHMOT>2.0.CO;2).
- Zhang, H. B., J. Chen, X. Zhi, Y. Li, and Y. Sun, 2014: Study on the application of GRAPES regional ensemble prediction system (in Chinese). *Meteor. Mon.*, **40**, 1076–1087.