Application of a convection-permitting ensemble prediction system to quantitative precipitation forecasts over southern China: Preliminary results during SCMREX

Xubin Zhang

CMA/Guangdong Provincial Key Laboratory of Regional Numerical Weather Prediction, Institute of Tropical and Marine Meteorology, CMA, Guangzhou, China

Correspondence
Xubin Zhang, Institute of Tropical and Marine Meteorology, Guangzhou 510640, China. Email: xbzhang@grmc.gov.cn

Funding information
Natural Science Foundation of Guangdong Province, 2017A030313225. Guangdong Province Science and Technology Project 2017B020244002, 2017B030314140, 2017B020218003. National Natural Science Foundation of China, 41405104.

As a preliminary attempt to cope with the low predictability of heavy rainfall over southern China in the pre-summer rainy season, an experimental convection-permitting ensemble prediction system (GM-CPEPS) based on the Global/Regional Assimilation and Prediction System (GRAPES) is developed. GM-CPEPS produces 12 h forecasts at 0.03° horizontal resolution based on 16 perturbed members. Perturbations from downscaling, ensemble of data assimilation, time-lagged scheme and topography are combined to generate the initial perturbations. Sea-surface temperature is perturbed and a combination of downscaling and balanced random perturbations is used to perturb the lateral boundary conditions. Stochastically perturbed parametrization tendencies, multi-physics, and perturbed parameters are all implemented.

In this study, GM-CPEPS was verified over a 15-day period during the Southern China Monsoon Rainfall Experiment (SCMREX) in May 2014. It was indicated that GM-CPEPS provided estimates of forecast uncertainty that are comparable to some international peers. Compared with the control forecasts (DET), some deterministic guidance, including the forecast distribution with 90th percentile, probability-matched mean, and linear combination of both (NPM), showed advantages in forecasting moderate and heavy rainfall; and the optimal-member technique was superior in reducing bias. Probabilistic guidance demonstrated the value over DET in detecting potential threats of severe weather, with both the optimal-probability and neighbourhood-probability technique leading to improvements in predicting lighter rainfall. Two cases were used to display the deterministic and probabilistic guidance intuitively and to illustrate corresponding advantages and drawbacks.

KEYWORDS
convection-permitting, ensemble prediction system, QPF, SCMREX

1 | INTRODUCTION

In the pre-summer rainy season (April–June) in southern China, heavy rainfall occurs frequently, accounting for roughly half the annual precipitation. The maximum seasonal rainfall accumulation is over 800 mm in several areas, often causing severe flooding (Luo et al., 2017; hereinafter L17). Mostly, precipitation during the pre-summer rainy season is of a convective nature with mesoscale organizational characteristics (Luo et al., 2013; Xia et al., 2015). Mesoscale convective systems (MCSs), and their interactions with larger-scale atmospheric systems and the complex underlying surface, are both closely related to the formation and evolution of heavy rainfall (Wang et al., 2014; Wu and Luo, 2016; L17). Unfortunately, numerical weather prediction (NWP) models are still very poor
at producing quantitative precipitation forecasts (QPFs) of heavy rainfall, especially of the warm-sector extreme rainfall over southern China during its pre-summer rainy season (L17).

Compared with NWP models with parametrized convection, convection-permitting (CP) NWP models, whose grid spacing (∼4 km or less) is fine enough to remove convective parametrization, have been shown to be beneficial for QPFs, especially for convective processes (Done et al., 2004; Weisman et al., 2008). For this reason, both operational and experimental CP NWP models for warm-season QPFs have been developed over the past decade. A CP regional model based on the Global/Regional Assimilation and Prediction System (GRAPES: Xue and Liu, 2007; Chen et al., 2008), named GRAPES-MARS3KM (Zhang et al., 2016; hereinafter Z16), has been developed recently for QPFs over southern China and has been in operational use since July 2014.

However, convective-scale forecast uncertainties due to highly nonlinear flows and rapid error growth (Lorenz, 1969; Hohenegger and Schär, 2007) lead inevitably to the low intrinsic predictability of atmospheric flows at CP scales (Zhang et al., 2003). The practical predictability is also seriously limited by deficiencies in the current data assimilation (DA) techniques and observation systems (Liu and Xue, 2008; Dong et al., 2011), and by imperfections in the present dynamics and physics configuration of NWP models (Morrison et al., 2009; Schwartz et al., 2010). In particular, the large sensitivities of CP-scale QPFs over southern China to initial conditions (ICs) and physics parametrization schemes are very pronounced in the pre-summer rainy season (Wu et al., 2013; Z16; Bao et al., 2017; L17) and thus reduce the predictability.

To cope with the poor CP-scale predictability, CP ensemble prediction systems (EPSs) are currently being developed by several NWP centres (Raynaud and Bouttier, 2016), considering the advantage of EPSs in estimating forecast uncertainties (Leith, 1974; Wilks, 2006). Some of these EPSs are in operation, for example, COSMO-DE-EPS (Peralta et al., 2012), AROME-EPS (Bouttier et al., 2012), MOGREPS-UK (Hagelin et al., 2017), the Storm-Scale Ensemble Forecast system (Clark et al., 2011), and the 3 km NCAR ARW EPS (Schwartz et al., 2015). However, there is still no operational or experimental CP EPS that is well designed and maturely applied to QPFs over southern China during the pre-summer rainy season. To fill this void, an experimental CP EPS based on GRAPES-MARS3KM (hereinafter GM-CPEPS) was developed recently by the Institute of Tropical and Marine Meteorology (ITMM) of the China Meteorological Administration.

The Southern China Monsoon Rainfall Experiment (SCMREX; http://exps.camscma.cn/scmrex) is a research and development project of the World Weather Research Programme of the World Meteorological Organization. One of its objectives is to improve QPFs over southern China in the pre-summer rainy season by conducting CP ensemble-forecasting experiments (L17). The field experiments of SCMREX were carried out over southern China in May and June of 2013–2015, when a number of heavy rainfall events were observed (fig. 3 of L17). Especially, the distribution of total rainfall accumulated during
8–23 May 2014 in southern China (Figure 1c) is similar to that of the seasonal-mean rainfall accumulated from April to June of 1981–2012 (fig. 1a of L17).

In support of the SCMREX programme, the present study attempts (for the first time in the literature) to evaluate the QPF performance of GM-CPEPS during SCMREX. The aim is to examine the current ability of GM-CPEPS in predicting the pre-summer precipitation and thereby to prompt the development of an operational CP EPS in southern China. Moreover, the ways of generating both deterministic and probabilistic guidance are investigated to explore how to provide better guidance to forecasters. Considering the limited computing resources and the good representation of precipitation events, the evaluation, based on a batch experiment and two case-studies, was implemented only for 8–23 May 2014. The GM-CPEPS configuration is described in section 2. Sensitivity tests to illustrate the rationale behind the design of GM-CPEPS are shown in section 3. Sections 4 and 5 present the results of the batch experiment and case-studies, respectively. Section 6 concludes the article and provides further discussion.

2 | METHODOLOGY

GM-CPEPS comprises one deterministic or control member (DET) and 16 perturbed members (PERs), all nested within a mesoscale EPS, namely GM-MSEPS. A summary of the acronyms for the components of GM-CPEPS can be found in Table 1. Aimed at improving the short-term (0–12 h) QPF skill, both GM-CPEPS and GM-MSEPS issue 12 h perturbed forecasts twice a day at 0000/1200 UTC.

2.1 | The GM-MSEPS ensemble

GM-MSEPS is based on GRAPES-MARS, which is a regional NWP model (Figure 1a) with a horizontal resolution of 0.09°×0.09° and 55 vertical levels. GM-MSEPS also has 1 DET and 16 PERs. For DET, the 6 h forecasts of the 0.5°×0.5° NCEP global forecast system (GFS) are used as background fields for the cold start of DA. The system used for DA and the observations assimilated in GM-MSEPS are the same as those in GM-CPEPS and will be introduced below. After a 6 h DA cycling which is started at 1800/0600 UTC per day, 24 h DET forecasts are issued with IC from DA analyses and lateral boundary conditions (LBCs) from GFS forecasts. Balanced random (BR: Barker, 2005; Meng and Zhang, 2008) perturbations are added to the GFS forecasts to generate both background fields and LBCs for the 16 PERs. The model uncertainty is represented by a combination of multi-physics (MP: Houtekamer et al., 1996; Stensrud et al., 2000) and stochastically perturbed parametrization tendency (SPPT: Palmer et al., 2009; Bouttier et al., 2012) schemes. MP and SPPT are used in both the 6 h DA cycling and 12 h forecasts. In MP, different parametrization schemes for microphysics and planetary boundary layer (PBL) processes are combined. During SCMREX, the ensemble mean (EM) of GM-MSEPS generally predicted both the pattern and evolution of rain bands, which were mainly influenced by the large-scale forcing, similar to the observed ones, and the ensemble spread can represent most of the forecast uncertainties of these rain bands (not shown).

2.2 | The GM-CPEPS ensemble

2.2.1 | Deterministic forecasts

The NWP model used in GM-CPEPS DET is GRAPES-MARS3KM, which is a nonhydrostatic model and adopts a semi-implicit, semi-Lagrangian scheme for temporal integration. Its horizontal grid is set on a longitude–latitude mesh with the Arakawa C-grid staggering. The GRAPES-MARS3KM domain covers southern China and the northern South China Sea (SCS) (Figure 1b),

| TABLE 1 | Acronyms for the configuration and products of GM-CPEPS |
|---------|------------------------------------------------------|
| DET     | Deterministic or control member                      |
| PER     | Perturbed member                                     |
| GRAPES-MARS | A mesoscale NWP model based on GRAPES                |
| GRAPES-MARS3KM | A convection-permitting NWP model based on GRAPES    |
| GM-MSEPS | A mesoscale EPS based on GRAPES-MARS                 |
| GM-CPEPS | A convection-permitting EPS based on GRAPES-MARS3KM  |
| GRAPES-CHAF3KM | A DA system based on 3D-Var analysis of GRAPES     |
| IDSC    | Downscaling for initial condition                    |
| EDA     | Ensemble of data assimilation                        |
| TLA     | Time-lagged scheme                                  |
| TO      | Topography (terrain height)                          |
| TS      | Sea-surface temperature                             |
| LDSC    | Downscaling for lateral boundary conditions          |
| BR      | Balanced random                                     |
| MP      | Multi-physics                                       |
| SPPT    | Stochastically perturbed parametrization tendency   |
| PP      | Perturbed parameters                                |
| EM      | Ensemble mean                                        |
| MED     | Ensemble median                                      |
| Q90     | Ensemble 90th percentiles                            |
| MAX     | Ensemble maximum                                     |
| PM      | Probability-matched mean                             |
| NPM     | Combination of PM and Q90                            |
| OPT     | Optimal or best member                               |
| PRO     | Original probability                                 |
| OPTP    | Optimal probability                                  |
| NEP     | Neighbourhood ensemble probability                  |

Blue acronyms indicate the components of GM-CPEPS; red acronyms indicate the perturbation methods used to design GM-CPEPS; green acronyms indicate the deterministic guidance of GM-CPEPS; and black acronyms indicate the probabilistic guidance of GM-CPEPS.
FIGURE 2  (a) Schematic configuration of GM-MSEPS (upper part) and GM-CPEPS (lower part). Only the ensemble forecasts issued at 0000 UTC and one of the 16 perturbed members are shown. A detailed description is given in the text. (b) Schematic showing the construction of the TLA ensemble using forecasts with different lead times. The red slanted numbers indicate the valid times at which forecasts are selected in the TLA ensemble. The dashed lines represent the GM-CPEPS DET forecasts used to construct the TLA ensemble, and the black circles represent the corresponding initial times. The forecasts initialized at 2100 UTC with grey come from the DA cycling of DET. Only the TLA ensemble used to issue GM-CPEPS forecasts at 0000 UTC is shown comprising 634×434 horizontal grid points with horizontal resolutions of 0.03°×0.03° and 55 vertical layers up to 28 km. Its vertical coordinate is terrain-following with the Charney–Phillips vertical layer skipping. The main forecast variables are zonal and meridional velocities (U, V), Exner pressure (Π), potential temperature (θ), and specific humidity (Q). The WRF Single-Moment 6-class (WSM6) microphysics scheme (Hong et al., 2006), the five-layer thermal diffusion (SLAB) land surface model, the medium-range forecast (MRF) PBL scheme (Hong and Pan, 1996), the rapid radiative transfer model (RRTM) long-wave radiation scheme, and the Dudhia short-wave radiation scheme are used in the physics parametrization.

The DA system used is GRAPES-CHAF3KM (Z16), which is based on the three-dimensional variational (3D-Var) analysis system of GRAPES (Xue et al., 2008) along with the multigrid technique (Xie et al., 2011) and the partial cycling strategy (Hsiao et al., 2012). Data from six types of observation (radiosonde, surface station, ship, Doppler radar, wind-profiling radar and aircraft) is assimilated in GRAPES-CHAF3KM every 3 h.

As shown in Figure 2a, the flow of GM-CPEPS DET is the same as that of GM-MSEPS DET except that the LBCs of the former come from forecasts of the latter rather than from GFS. Z16 can be referred to for more details of DET.

2.2.2 Perturbed forecasts
The uncertainties in IC, surface, LBCs, and model physics are all considered in generating the PERs of GM-CPEPS. However, there is still no generally accepted method for perturbing these four components of a CP EPS. Recently, blending or combining approaches have been developed to consider different benefits of various perturbation methods, not only for IC (Wang et al., 2011; Lang et al., 2012; Caron, 2013) but also for model physics (Hacker et al., 2011; Duda et al., 2014; Berner et al., 2015). Given the positive impacts of these approaches on forecasts as illustrated in the references cited
above, combining perturbations from multiple uncertainty sources using some leading perturbation methods (Table 1) was used to construct GM-CPEPS to consider various uncertainties from various sources as much as possible. Each PER is defined by adding perturbations to the IC, surface, LBCs, and model physics of the unperturbed member, i.e. DET.

2.2.3 IC perturbations

IC perturbations are generated by a linear combination of perturbations from the downscaling (IDSC), ensemble of DA (EDA), time-lagged scheme (TLA), and topography (TO), with \( \alpha_1, \alpha_2, \alpha_3 \) and \( \alpha_4 \) as the corresponding weighting factors, respectively. \( \alpha_4 \) is set to 1; whereas \( \alpha_1, \alpha_2, \) and \( \alpha_3 \) need to be determined to avoid excessive perturbations causing numerical instability, because IDSC, EDA and TLA perturbations all show significantly larger magnitude than TO perturbations (not shown).

For IC perturbations of a CP EPS, downscaling the initial perturbations from a coarser EPS is a common and standard technique (Kühnlein et al., 2014; Tennant, 2015). Here, IDSC perturbations are calculated by subtracting EM of the 16 initial PERs of GM-MSEPS from each initial PER and then interpolated to the GRAPES-MARS3KM domain. EDA is designed to estimate the uncertainty of analyses (Belo-Pereira and Berre, 2006; Vié et al., 2011) and is also used in GM-CPEPS. In the 6 h DA cycling of the GM-MSEPS PERs, the 3 h forecasts valid at 2100/0900 UTC are used as the perturbed background fields to create the first perturbed analyses with GRAPES-CHAF3KM. The 3 h integration of GRAPES-MARS3KM with MP and SPPT, which is described in section 2.2.6, will then provide the next perturbed background fields valid at 0000/1200 UTC to create the final perturbed analyses. The EM of these 16 perturbed analyses is subtracted from each perturbed analysis to calculate EDA perturbations.

Based on TLA, a set of successive forecasts valid at different ranges but valid at the same time can be used to reflect the IC uncertainty with time-evolving (flow-dependent) information (Mittermaier, 2007; Wang et al., 2016). The ready-made forecasts of GM-CPEPS DET are used to construct a TLA ensemble, which is also an integral part of GM-CPEPS. For TLA members with an initial time of 0000 UTC (Figure 2b), the forecasts used are the 21, 22, 23 and 24 h ones initialized at 0000 UTC on the previous day, the 9, 10, 11, 12, 13, 14 and 15 h ones initialized at 1200 UTC on the previous day, and the 1, 2 and 3 h ones initialized at 0000 UTC; the forecasts valid at 2300 UTC on the previous day and at 0000 UTC, which are initialized at 2100 UTC based on the analyses derived from the 6 h DA cycling, are also used. The order of these 16 forecasts used in constructing the 16 TLA members is random. To compute TLA perturbations, the EM of the 16 TLA members is subtracted from each TLA member.

The initialization and evolution of heavy rainfall over southern China are often influenced by the terrain (Wang et al., 2014). Due to weaknesses of NWP in describing real topography, the uncertainty of QPFs is associated with the uncertainty of representing topography in NWP. However, there are still few references proposing methods to perturb the topography. Here, the terrain height on land is perturbed by adding Gaussian random numbers with zero mean and a standard deviation of 300 m to the original value (Figure 1b).

The perturbation amplitude is restricted to be less than 450 m to avoid unrealistic perturbations, and the terrain height is left unperturbed if the perturbed height is negative. The value of parameters determining the perturbation amplitudes is set based on the analysis of differences in terrain height between current topography data and those with higher resolution (not shown). The 16 perturbed ICs are then generated by interpolating the IC of GM-CPEPS DET valid at 0000/1200 UTC based on the perturbed topographies, and TO perturbations are computed by subtracting the EM of these 16 ICs from each of them.

Perturbations of \( U, V, \Pi, \theta \) and \( Q \) are calculated in the above four types of perturbation. \( \theta \) and \( Q \) are left unperturbed if the perturbed \( Q \) of the IC reaches an unphysical value or exceeds a critical supersaturation value. Several further steps are conducted to determine the weighting factors for IDSC, EDA and TLA perturbations as follows.

1. Multivariate Empirical Orthogonal Function (MV-EOF) analysis is used in the combination of various perturbations (i.e. \( U, V, \Pi, \theta \) and \( Q \)). For the combined data at a certain vertical level, the temporal dimension used in traditional MV-EOF is replaced by the dimension of the ensemble member.

2. Eigenmodes corresponding to the leading eigenvalues, whose contributions to the total variance are above 90%, are calculated at every vertical level.

3. The eigenvalues at level \( k \) for IDSC, EDA and TLA perturbations are defined as \( E_{1}^{k}, E_{2}^{k} \) and \( E_{3}^{k} (k = 1, 2, \cdots, 55) \), respectively. The first estimation \( \beta^{k} \) for \( \alpha_1, \alpha_2 \) and \( \alpha_3 \) at level \( k \) is defined as

\[
\beta^{k} = (E_{\text{max}}^{k} + E_{\text{min}}^{k})/E_{\text{tot}}^{k},
\]

where \( E_{\text{max}}^{k}, E_{\text{min}}^{k} \) and \( E_{\text{tot}}^{k} \) represent the maximum, minimum and sum, respectively, of the three eigenvalues (i.e. \( E_{1}^{k}, E_{2}^{k} \) and \( E_{3}^{k} \)).

4. The mean eigenvalues for IDSC, EDA and TLA perturbations are defined as \( E_{1}^{m}, E_{2}^{m} \) and \( E_{3}^{m} \), respectively, by averaging \( E_{1}^{k}, E_{2}^{k} \) and \( E_{3}^{k} \) over all the vertical levels. \( \alpha_1, \alpha_2 \) and \( \alpha_3 \) at level \( k \) are then defined as

\[
\alpha_{i}^{k} = \beta^{i} \min(\beta, E_{\text{max}}^{k}/E_{1}^{m}, 1.0), (i = 1, 2, 3),
\]

where \( E_{\text{max}}^{m} \) represents the maximum of the three mean eigenvalues (i.e. \( E_{1}^{m}, E_{2}^{m} \) and \( E_{3}^{m} \)) and \( \beta_{i} \) denotes the tuning factor used to maintain the three perturbations’ mutual balance in contribution to the total IC perturbations. Specifically, to increase the contribution of IDSC perturbations, \( \beta_{1} \) is set to 1.2 and both \( \beta_{2} \) and \( \beta_{3} \) are set to 0.3.
2.2.4 | Surface perturbations
As a crucial aspect of monsoon precipitation, sea-surface temperature (SST) shows some influence on precipitation (Mo and Juang, 2003; Roxy, 2014), especially for the pre-summer precipitation over southern China due to the significant warm-moist air coming from the SCS (Huang and Mao, 2015; Wu and Luo, 2016). Thus, the uncertainty of SST is considered here by adding random numbers drawn from a Gaussian distribution with zero mean and a standard deviation 2.0 K to SST at the grid scale to generate SST perturbations. SST is kept unchanged during the model integration, resulting in constant forcing from the surface boundary to the atmosphere. SST perturbations are bounded within the range of ±2 (±1) standard deviations for PERs initialized at 0000 (1200) UTC, i.e. 0800 (2000) local solar time, to limit excessive perturbations. SST perturbations are allowed to be larger in daytime than at night-time, given the larger variability in the former (Gentemann et al., 2003).

2.2.5 | LBCs perturbations
Downscaling LBCs (LDSC) perturbations are blended with BR perturbations to generate LBCs perturbations for GM-CPEPS.

Introducing LDSC perturbations from a mesoscale EPS by nesting is a common way for CP EPS (Gebhardt et al., 2011; Schwartz et al., 2015), hence the forecasts of the GM-MSEPS PERs are interpolated to form the first LBCs of the GM-CPEPS PERs.

BR perturbations, which are produced by taking Gaussian random draws with zero mean and covariances from the regional background error covariances provided by the 3D-Var analysis system of GRAPES, are inflated multiplicatively as the forecast progresses (Torn, 2010; Schwartz et al., 2015). Specifically, the averaged inflation factor is roughly 1.04 (1.2) within (beyond) the 4 h forecasts. BR perturbations to generate LBCs perturbations. SST is kept unchanged during the model integration, resulting in constant forcing from the surface boundary to the atmosphere. SST perturbations are bounded within the range of ±2 (±1) standard deviations for PERs initialized at 0000 (1200) UTC, i.e. 0800 (2000) local solar time, to limit excessive perturbations. SST perturbations are allowed to be larger in daytime than at night-time, given the larger variability in the former (Gentemann et al., 2003).

2.2.6 | Model physics perturbations
A combination of SPPT, MP and Perturbed Parameters (PP: Gebhardt et al., 2011; Duda et al., 2014) schemes is used to generate model physics perturbations.

In the SPPT scheme, random perturbations are added to the total parametrized tendency of physical processes, which includes the radiation, microphysics and PBL processes, for the variables $U$, $V$, $\theta$ and $Q$. In this implementation, a random field $r$ drawn from a Gaussian distribution with zero mean, a standard deviation of 0.5, a spatial decorrelation scale of 50 km, and a temporal decorrelation scale of 1 h is produced during model integration. Here, $r$ is bounded within the range of ±2 standard deviations to avoid unrealistic perturbations causing numerical instability. Next, the physical tendency is multiplied at each time step by $f = 1 + r$ to form a perturbed tendency. No tendency perturbations are applied near the surface (below about 500 m above ground) or near the model top (<50 hPa), and the perturbations in the transition layers (500–1,500 m and 100–50 hPa) are ramped up smoothly to full amplitude. Additionally, perturbations of $Q$ are checked in a manner which is similar to the initial perturbations, to determine whether to implement tendency perturbations to both $\theta$ and $Q$.

WSM6 and the WRF Single-Moment 5-class (WSM5) microphysics scheme (Hong et al., 2004), as well as MRF and the Yonsei University (YSU: Hong et al., 2006) scheme, are used to construct four combinations of physics packages for parametrizing the microphysics and PBL processes in the MP scheme. Among the 16 perturbed ensemble members, members 1–4 use WSM6/MRF schemes, as is the case in DET, members 5–8 use WSM6/YSU schemes, members 9–12 use WSM5/MRF schemes, and the remaining members use WSM5/YSU schemes.

Both the rain intercept parameter ($N_{0r}$) in the microphysics scheme and the critical Richardson number ($Ri_c$) in the PBL scheme are perturbed in the PP scheme, considering the significant sensitivities of the parametrization schemes to these parameters (Hacker et al., 2011; Baker et al., 2014). In the WSM6/MRF, WSM6/YSU, WSM5/MRF and WSM5/YSU schemes, the default setting of $N_{0r}/Ri_c$ is set to 8 x 10^6/0.5, which is used in members 1, 5, 9 and 13; for the remaining three members of each combined scheme, $N_{0r}/Ri_c$ is set to 8 x 10^7/0.5, 8 x 10^8/0.5 and 8 x 10^9/1.0, respectively.

2.3 | Experimental design
To assess the performance of GM-CPEPS, retrospective GM-CPEPS forecasts were conducted during 8–23 May 2014, when 32 12 h forecasts were collected. The experimental period was dominated by several heavy rainfall events, with significant extreme daily rainfall (>100 mm/day) observed. As described in L17, two significant extremely heavy rainfall events, which were respectively characterized by local and large-scale MCSs and representative during the pre-summer rainy season, occurred on 8 and 22–23 May (hereinafter Cases 1 and 2, respectively). Among these events, those forecasts with initial times of 0000 UTC 8 May and 1200 UTC 10 May were selected to test the ratio nale behind the design of GM-CPEPS and to demonstrate the GM-CPEPS performance intuitively. Detailed descriptions of these events can be found in L17 and Huang and Luo (2017).

This study focuses on the performance of both deterministic and probabilistic guidance of GM-CPEPS for QPFs (Table 1). Specifically, the deterministic guidance provided by GM-CPEPS includes rainfall from the EM, the probability-matched mean (PM: Ebert, 2001; Schwartz et al., 2014), the ensemble median (MED), the ensemble 10th, 25th, 75th and 90th percentiles (Q_{10}, Q_{25}, Q_{75} and Q_{90}, respectively), the ensemble maximum (MAX) and minimum (MIN),
and the rainfall from each PER. However, the guidance based on ensemble percentiles whose thresholds are below 50th are generally not so often used in operational warning for severe weather, especially in southern China in the pre-summer rainy season. Therefore, the performances of MIN, Q10 and Q25 are not illustrated here. For rainfall with ensemble percentile threshold of q-th, i.e. Qq, it is calculated as follows:

\[ Q_q = b V_{a+1} + (1 - b) V_a, \]  

where \( a \) and \( b \) represent the integer and fractional portions of \((M + 1)q/100\) respectively and \( M \) represent the ensemble size, i.e. 16. \( V_{a+1} \) and \( V_a \) denote rainfall from the members ranked \((a + 1)\)th and \( a\)th, respectively. The probabilistic guidance includes the probability of rainfall at different thresholds.

To investigate the advantages of GM-CPEPS in QPFs, its deterministic guidance is directly compared with DET, and its probabilistic guidance is compared with the binary rainfall fields of DET following Zhang (2018) (hereinafter Z18). Specifically, DET was transformed into binary fields by setting grid points exceeding a certain threshold to a value of 1, while all other points were given a value of 0. All comparisons were based on the skill scores illustrated below, which were averaged over all 32 forecasts.

### 2.4 Forecast verification

Verification for the forecasts of variables at the upper levels was performed using 12-hourly radiosonde observations over the model domain (fig. 1a of Z16). Here, the forecasts were bilinearly interpolated onto the locations of radiosondes. Forecasts of precipitation and variables at the surface were verified against the surface meteorological stations and automatic weather stations (AWSs) over the verification domain (Figure 1b,c). These surface observations were interpolated onto the NWP model grids using Cressman interpolation to calculate the skill scores.

The root-mean-square error (RMSE) of the forecasts was calculated for variables at the upper levels and surface. To verify the rainfall for GM-CPEPS deterministic guidance as well as DET, the fraction skill score (FSS; Roberts and Lean, 2008) was used. FSS was computed for the 1 h accumulated rainfall in consideration of the more significant “double penalties” due to the displacement error in the verification of high-resolution forecasts (Ebert, 2008). In calculating FSS, “neighbourhood length” was defined as the length of area over which the fractions were computed. Probabilistic guidance verification was achieved by computing the continuous ranked probability score (CRPS: Hersbach, 2000), the Brier score (BS) and its reliability aspect (Wilks, 2006), and the area under the relative operating characteristic curve (AROC: Mason and Graham, 2002).

For the comparison of skill scores (including RMSE, FSS, CRPS, BS, reliability and AROC), the statistical significances of the differences between different forecast sources were assessed using bootstrap resampling. Specifically, random samples of skill scores were generated with replacement, and the differences of skill scores were then calculated. This procedure was repeated 1,000 times, and bootstrapping was performed on the score differences between the paired samples. The rank at which the resampled score differences crossed zero was used as the significance level to represent the probability that two skill scores were distinct. In section 4, only situations above the 85% or 90% significance level (indicating an 85 or 90% probability that two skill scores differed) are labelled.

### 3 Rational for GM-CPEPS Design

So far, the best practice in designing a CP EPS remains unclear (Romine et al., 2014). Insufficient ensemble spread (or dispersion), which often causes inaccurate representation of forecast uncertainties in the ensemble forecasts, is one of the most critical problems that needs to be improved in current CP EPSs (Hohenegger et al., 2008; Gebhardt et al., 2011; Vié et al., 2011; Schwartz et al., 2014). However, methods used to improve the ensemble spread do not necessarily improve the deterministic guidance (e.g. EM) and may degrade the individual-member forecasts, thereby limiting the operational use of CP EPSs especially in forecasting high-impact events (Romine et al., 2014).

Given the two aforementioned issues, the rationale for GM-CPEPS design is not only to increase the ensemble spread but also to increase both deterministic and probabilistic skills. For this purpose, GM-CPEPS is firstly constructed by simply incorporating perturbations from all uncertainty sources using some leading perturbation methods. The sensitivities of GM-CPEPS performance to the different perturbation methods are then tested to determine which perturbations should be included. This idea is also used in Z18 to construct a mesoscale EPS and seems to work.

In addition to the retrospective forecasts (hereinafter ALL) described in section 2.3, 10 additional comparison experiments – called noIDSC, noEDA, noTLA, noTO, noTS, noLDSC, noBR, noSPPT, noMP and noPP with IDSC, EDA, TLA, TO, TS, LDSC, BR, SPPT, MP and PP perturbations, respectively, removed from ALL – were carried out. Moreover, an experiment named CTL with the simplest perturbations (i.e. IDSC and LDSC perturbations) was designed as a benchmark in the comparison. The above ensemble forecast experiments and their respective identifiers are all listed in Table 2. These experiments were performed for Cases 1 and 2, but only the results of Case 1 are shown here because the conclusions are similar in either case.

During the whole 12 h forecasts, the EM RMSE of the 1 h accumulated rainfall in ALL was the smallest (Figure 3). The differences of RMSE among the experiments became more evident for the forecasts beyond 8 h when most heavy rainfall occurred (Huang and Luo, 2017) than for the whole 12 h forecasts. Thus, including all the perturbations mentioned...
TABLE 2  Set-up of the 11 ensemble forecast experiments in section 3

| Exp. | Perturbations                          |
|------|----------------------------------------|
| ALL  | Including all perturbations           |
| noIDSC | Removing IDSC perturbations from ALL   |
| noEDA | Removing EDA perturbations from ALL    |
| noTLA | Removing TLA perturbations from ALL    |
| noTO  | Removing TO perturbations from ALL     |
| noTS  | Removing TS perturbations from ALL     |
| noLDSC | Removing LDSC perturbations from ALL  |
| noBR  | Removing BR perturbations from ALL     |
| noSPPT | Removing SPPT perturbations from ALL  |
| noMP  | Removing MP perturbations from ALL     |
| noPP  | Removing PP perturbations from ALL     |
| CTL  | Including only IDSC and LDSC perturbations |

FIGURE 3  EM RMSE and CRPS for forecasts of 1 h accumulated precipitation initialized at 0000 UTC 8 May 2014 averaged over different lead times. The white (grey) bars represent RMSE averaged over forecasts valid from 1 h (8 h) to 12 h. The dots (circles) represent CRPS averaged over forecasts valid from 1 h (8 h) to 12 h.

in section 2.2 led to the highest deterministic QPF skill, especially for heavy rainfall. The benchmark experiment, i.e. CTL, which included only IDSC and LDSC perturbations, showed the largest RMSE among all the experiments during the whole 12 h forecasts, illustrating the importance of including all the perturbations in the design of GM-CPEPS. However, there was little difference in RMSE between ALL and noLDSC or noBR. Additionally, removing IDSC and MP perturbations from ALL resulted in more evident degradation of QPF skill, compared with other perturbations. The results of CRPS were very similar to those of RMSE, illustrating the highest probabilistic QPF skill from ALL. Therefore, the contributions of the various individual perturbations to the perturbations of ALL, which included all types of perturbations, the largest were IDSC and MP perturbations while the smallest were from LBCs.

For a reliable EPS, the ensemble spread should be consistent with the EM RMSE. At 1200 UTC when the quasi-linear MCS reached maturity (fig. 9a3 of Huang and Luo, 2017), large EM RMSEs (≥10 mm) were present in the regions with rainfall above 10 mm/h for all the experiments (not shown). In these regions, both the area and magnitude of large ensemble spreads (≥10 mm) in ALL were most consistent with those of large EM RMSEs, among all the experiments (Figure 4c), whereas the opposite was true for noIDSC, noEDA, noTLA and CTL, all of which had small spreads (Figure 4a,b). In the regions where large RMSEs were absent, all the experiments except ALL produced large spreads, especially noMP and noPP (Figure 4d). Therefore, CTL and the experiments removing IC or model physics perturbations from ALL produced obvious biases in ensemble dispersion for QPFs. However, the above biases were reduced by introducing all the perturbations. Overall, the advantage of ALL over noLDSC or noBR in ensemble dispersion was small, albeit indeed present (Figure 4c).

To summarize, the current design of GM-CPEPS was proved to be effective in representative case-studies and was thus used during SCMREX, since the experiment with all the perturbations evidently outperformed the benchmark experiment and most of the perturbations, i.e. the IC, surface, and model physics perturbations seemed to be indispensable to GM-CPEPS in increasing its performance. But for LBCs perturbations, they contributed marginally to the performance of GM-CPEPS, not only for Cases 1 and 2 but also for cases during the entire experimental period (not shown). As shown in some previous studies (e.g. Vié et al., 2011), LBCs perturbations often showed non-negligible impacts on CP EPSs performance, especially in the later period of forecasts. However, this conclusion was apparently contradicted here, probably because the forecast range of GM-CPEPS is short (i.e. 12 h) and its domain is also relatively large. Besides LBCs perturbations, not all the other perturbations led to markedly improved forecasts. Thus, the current design of GM-CPEPS can be optimized by excluding or improving some perturbation components, which will be essential in future work.

4  | RESULTS OF A BATCH EXPERIMENT

4.1  | Deterministic guidance for non-precipitation forecasts

Figure 5 shows the vertical profiles of RMSE for the 12 h forecasts. By and large, EM provided higher skill for the non-precipitation variables, especially horizontal wind (Figure 5a), compared with DET. For temperature and humidity, the improvements of EM over DET were mainly at lower levels (Figure 5b,c). When comparing ensemble spread with RMSE, GM-CPEPS is no exception to the common deficiency of CP EPSs in underdispersion. Specifically, the underdispersion was more serious for temperature and humidity (Figure 5b,c) than for wind (Figure 5a). However, observation error should be included in the verification as a non-negligible error source (Saetra et al., 2004). Thus, the total spread, which includes both ensemble spread and observation error, was also
FIGURE 4  (a–d) Ensemble spread (dashed lines for values equal to 10 mm) for forecasts of 1 h accumulated rainfall valid at 1200 UTC 8 May 2014. Contours and numbers in the top right corner of each panel with different colours represent the different experiments. The numbers represent the mean ensemble spread and EM RMSE (unit: mm), which are averaged over the region (excluding grid points with missing observations) shown in each panel, separated by “/” to indicate that the former (latter) are related to ensemble spread (EM RMSE). The grey shaded area denotes the region with observed rainfall above 10 mm, which generally corresponds to the region with EM RMSE above 10 mm calculated. The estimations of observation error reported by the National Centers for Environmental Prediction (NCEP) (fig. 2 of Wang et al., 2008) and the 3D-Var analysis system of GRAPES (e.g. about 1.7 g/kg for specific humidity at lower levels) were used here. The ratio of EM RMSE to total spread was close to 1, especially for variables at lower levels (1,000–850 hPa), indicating good consistency between spread and RMSE. However, GM-CPEPS remained underdispersive at middle levels (700–300 hPa) for temperature and humidity (Figure 5b,c) but became overdispersive at those levels for wind (Figure 5a). The ratio generally varied between 0.85 and 1.1 at lower levels, where dynamic and thermodynamic variables are very important for convective processes; moreover, the ratio was comparable to that for some mesoscale EPSs in the world (Hacker et al., 2011; Schwartz et al., 2014).

For both 10 m wind speed and 2 m specific humidity, both RMSE and ensemble spread decreased with lead time (Figure 6a,c), which is caused by the poor spin-up behaviour due to the imperfect coupling between the diagnosed and prognostic variables at the initial time in GRAPES-MARS3KM (Z16). EM showed significant improvements over DET in predicting the surface variables (Figure 6), although some degradations appeared in the first 2 forecasts of 10 m wind speed (Figure 6a). The estimations of observation error from NCEP for 10 m wind speed (1.4 m/s), 2 m temperature (1.2 K), and 2 m specific humidity (1.7 g/kg) were used here to calculate the total spread. After considering the impact of observation error on the verification, the consistency of spread and RMSE was significantly improved. Even so, underdispersion was still present for all the surface variables, especially for 10 m wind speed (Figure 6a). The ratio of EM RMSE to total spread was closest to 1 for 2 m specific humidity (Figure 6c), followed by 2 m temperature (Figure 6b). Overall, the ratios for surface variables, e.g. 10 m wind speed, 2 m temperature and 2 m specific humidity were roughly 1.5, 1.2 and 1.0, respectively, which are very comparable to those for some operational CP EPSs (Beck et al., 2016).

Thus, GM-CPEPS exhibited reasonable performances of EM and some capabilities for estimating forecast uncertainty, especially for variables at lower levels and the surface.

4.2 Deterministic guidance for precipitation forecasts

Compared with DET, EM improved the forecasts of precipitation by significantly reducing the RMSE (Figure 6d). In
FIGURE 5  Ensemble spreads of GM-CPEPS, EM RMSE, and DET RMSE for 12 h forecasts of (a) horizontal wind, (b) temperature, and (c) specific humidity during 8–23 May 2014. The dots indicate the vertical levels for which the significance level of the RMSE differences is larger than 90% for comparing EM with DET. “oberr” (“noberr”) represents the total (ensemble) spread including (not including) observation error. Numbers on the right-hand axes represent the ratio of EM RMSE to ensemble spread, separated by “/” to indicate that the former (latter) are related to total (ensemble) spread including (not including) observation error.

FIGURE 6  Ensemble spreads of GM-CPEPS, EM RMSE, and DET RMSE for forecasts of (a) 10 m wind speed, (b) 2 m temperature, (c) 2 m specific humidity, and (d) 1 h accumulated rainfall valid at different lead times during 8–23 May 2014. The dots indicate the lead times for which the significance level of the RMSE differences is larger than 90% for comparing EM with DET. “oberr” (“noberr”) represents the total (ensemble) spread including (not including) observation error. Numbers on the top axes represent the ratio of EM RMSE to ensemble spread, where the grey italic (black normal) ones indicate that the former (latter) are related to total (ensemble) spread including (not including) observation error.
FIGURE 7  FSS for forecasts of 1 h accumulated precipitation with thresholds of 0.1, 10, 20 and 40 mm and a neighbourhood length of 50 km averaged over lead times from (a,c) 1 to 6 h and (b,d) 7 to 12 h for various forecasts during 8–23 May 2014. Black (grey) crosses in (a) and (b) indicate the thresholds for which the significance level of the FSS differences between Q₉₀ and other forecast distributions is larger than 90% (85%), while black (grey) crosses in (c) and (d) indicate the thresholds for which the significance level of the FSS differences between DET and other forecasts is larger than 90% (85%).

calculating the total spread, the observation error was estimated as in Bouttier et al. (2012). Generally, the ratio of EM RMSE to total spread was around 1.2, indicating that GM-CPEPS’s underdispersion of precipitation was not very serious.

Forecast distributions of EPS (such as MED, EM, Q₇₅, Q₉₀ and MAX) can be directly used to generate deterministic guidance. Therefore, the performance of these products for different types of rainfall was assessed here by comparing the 50 km length FSS for the 1 h accumulated rainfall between various forecast distributions (Figure 7). The comparison of FSS with other neighbourhood lengths is not shown here because the conclusions are generally similar. FSS ranges from zero to one, with higher values corresponding to higher skills. In this study, rainfall with precipitation rates greater than 0.1, 10, 20 and 40 mm/h is defined as light, moderate, heavy and extremely heavy rainfall, respectively.

Overall, QPF skill increased with the percentile of forecast distribution, and MED showed the poorest performance. EM, Q₉₀ and MAX showed the best performance for light rainfall (Figure 7a,b). For moderate and heavy rainfall, the highest FSS was produced by Q₉₀, followed by Q₇₅ and MAX, especially for forecasts beyond 6 h (Figure 7b). Compared with other forecast distributions of EPS, both Q₉₀ and MAX showed significant advantage in predicting extremely heavy rainfall, especially the former. Actually, it was often very difficult for forecasts with a smaller percentile of forecast distribution to produce heavier rainfall. In particular, the magnitude of extremely heavy rainfall was overestimated (underestimated) for forecasts above (below) ensemble 90th percentiles (not shown). Thus, among the products from forecast distributions, Q₉₀ performed the best for almost all types of rainfall. Compared with DET, Q₉₀ was significantly more skilful in predicting light rainfall, and was slightly better or generally comparable in predicting heavier rainfall (Figure 7c,d). As shown in Figure 8a, there was underestimation (overestimation) for the area of light rainfall in DET (Q₉₀). Thus, the missing rate of DET in predicting light rainfall can be corrected by using Q₉₀, which partially contributed to higher FSS in Q₉₀ than in DET. The areal coverages of moderate and heavy rainfall were overestimated more evidently for Q₉₀ than for DET (Figure 8b,c), which explains the insignificant improvement in FSS of Q₉₀ over DET. Q₉₀ also yielded less but more accurate areal coverage of extremely heavy rainfall.
than DET, especially during the later experimental period
(Figure 8d).

Aimed at improving the rainfall estimation from the EM, the PM blends the spatial pattern of the EM with the frequency distribution of the whole EPS (Ebert, 2001). The PM field at a given time is produced in three steps. First, the frequency distribution is formed by pooling the rainfall amounts from all 16 PERs at each grid point in the verification domain, ranking them from largest to smallest and keeping every 16th value. Second, the EM rainfall amounts in the verification domain are also ranked from largest to smallest. Third, the grid point corresponding to the largest EM rainfall amount is assigned the largest value in the frequency distribution, and so on. Compared with DET, PM yielded significant degradation in predicting light rainfall during the whole forecasts (Figure 7c,d), which mainly resulted from more serious underestimation of areal coverage in PM than in DET (Figure 8a). For moderate and heavy rainfall, PM significantly outperformed DET for forecasts beyond 6 h (Figure 7d); especially, there was smaller overestimation of areal coverage in PM than in DET (Figure 8b,c). Because there were only a few observations or forecast samples for extremely heavy rainfall, the forecast differences between PM and DET were not very significant. For areal coverage of extremely heavy rainfall, PM showed evident underestimation, but the bias was much smaller than that of DET (Figure 8d).

In summary, the significant superiorities of PM (Q90) over DET were present in the forecasts of moderate and heavy (light and extremely heavy) rainfall. This gave us a hint that combining PM with Q90 may produce new PM forecasts (hereinafter NPM) with better performance than either individually. The superiority of Q90 over PM was mainly for light and extremely heavy rainfall; therefore, more weight is given to Q90 than to PM in constructing NPM for these types of rainfall. Note that, although bias was evidently present in Q90 for most of the rainfall thresholds, the characteristics of bias were not identical among rainfall with different thresholds (Figure 8). Actually, high bias of Q90 was chiefly present in predicting rainfall of moderate amount (e.g. above 5 mm and below 35 mm), but not evident in predicting rainfall of smaller (e.g. below 5 mm) or larger (e.g. above 35 mm) amount (not shown). In order to avoid serious bias caused by Q90, NPM should be constructed based on rainfall of different amounts. Specifically, NPM is generated by a linear combination of PM (with weighting factor $\varepsilon$) and Q90 (with weighting factor $\mu$). $\varepsilon$ ($\mu$) is set to 0.45 (0.55) when the predicted rainfall amount from PM (Q90) is below 5 mm or above 35 mm; while $\varepsilon$ ($\mu$) is set to 0.65 (0.35) when the predicted rainfall amount from PM (Q90) is above 5 mm and below 35 mm. According to the definition of NPM, there may be discontinuous fields which are not physical in NPM, where the corresponding fields in PM or Q90 are around 5 or 35 mm and characterized by small spatial variations. However, this discontinuity is generally not
In this study, the optimal or best member (hereinafter OPT) was selected based on its closeness to the NPM field, since the NPM forecasts generally behaved better than the PM, DET, and Q90 ones. Following Schwartz et al. (2014), the FSS with a 50 km neighbourhood length was calculated over the verification domain for precipitation rates greater than 0.1 and 20 mm/h for each PER using the corresponding NPM field as the truth. Following Z18, OPT was determined by maximizing the cost function

$$J_i = \frac{1}{9} \sum_{j=4}^{12} FSS_j(20) + \frac{1}{12} \sum_{j=1}^{12} FSS_j(0.1),$$

where $FSS_j(20)$ and $FSS_j(0.1)$ denote the FSS, which is calculated in the verification domain, for member $i$ at lead time $j$ for precipitation rates above 20 and 0.1 mm/h, respectively. Here, only those forecasts beyond 3 h were measured for heavy rainfall because NPM was less skilful than DET during the first 3 h of forecasts (not shown). Obviously, the closest member with the highest cost function was selected as OPT. Similar to Z18, if more than one member had a maximum cost function, the OPT forecasts were produced by averaging over these members.

Although OPT produced higher or comparable FSS for forecasts of light rainfall compared with PM and DET, it performed worse than Q90 and NPM (Figure 7c,d). There was no significant superiority for OPT in comparison with the other forecasts in predicting rainfall with larger amounts, although OPT seemed to perform best in the 7–12 h forecasts of extremely heavy rainfall. The conclusion that OPT did not significantly outperform DET in predicting any type of rainfall except light rainfall, as well as the reason behind the variation of OPT’s advantages over PM in the present study, were both similar to those in Z18. However, the bias of OPT was clearly smaller than that of DET in predicting all types of rainfall (Figure 8). On average, OPT yielded the smallest bias among all the deterministic guidance.

4.3 | Probabilistic guidance

It is now becoming increasingly accepted that a probabilistic forecasting approach is absolutely necessary for CP NWP, considering the chaotic behaviour and large uncertainties of the high-impact weather at convective scales (Stensrud et al., 2009; Snook et al., 2012).

The original probabilistic fields (hereinafter PRO) discussed in this study were simply the number of ensemble members with a certain threshold divided by the ensemble size, namely 16. The optimal-probability technique used in Z18 was also investigated here. Specifically, only the ensemble member whose cost function in Equation (4) accounted for 5% or more of the total cost function for all the PERs was included in the calculation of optimal probability (OPTP).

Recently, “neighbourhood approaches” were used to post-process high-resolution ensemble output to produce the probabilistic guidance (Theis et al., 2005). Introduced
by Schwartz et al. (2010), the “neighbourhood ensemble probability” (NEP) applied a neighbourhood approach to generate grid-point probabilities. Here, NEP fields with a neighbourhood length of 50 km were also verified.

The skills for various probabilistic guidance (PRO, OPT and NEP) were assessed and compared with those for the binary rainfall fields of DET. Several skill scores for the forecasts during SCMRex were calculated here (Figure 9). Considering that the comparison of these scores among different forecasts was generally similar at different lead times, only the scores averaged over the whole 12 h forecasts were shown in Figure 10. BS measures the accuracy of EPS for predicting probability. Reliability is one of the three components of BS (Candille and Talagrand, 2005) and measures the agreement between predicted probability and mean observed frequency. AROC measures the ability of a forecast to discriminate between two alternative outcomes and thus measures the resolution, which is the other component of BS. A better (smaller) BS is associated with better (smaller) reliability and better (larger) resolution.

Compared with DET, all the probabilistic guidance produced significantly smaller BS for all thresholds (Figure 9a), indicating that the latter performed more accurately. For light rainfall, both OPTP and NEP outperformed PRO, with smaller BS for OPTP than for NEP. In terms of moderate rainfall, BS for NEP was significantly smaller than that for the other probabilistic guidance, while OPTP did not show significant superiority over PRO (Figure 9a). Regarding heavy and extremely heavy rainfall, there was no significant difference in accuracy among the probabilistic guidance.

The results of comparing reliability between DET and the probabilistic guidance were similar to those for BS (Figure 9b), except that the improvement in reliability of the probabilistic guidance over DET was remarkably larger than that in accuracy (Figure 9a). In short, compared with PRO, OPTP was significantly advantageous only for light rainfall, whereas NEP was more reliable for almost all thresholds except for heavy and extremely heavy rainfall (Figure 9b). For DET, some areal biases were found across all lead times and thresholds (figure 5c,d in Z16; Figure 8 herein). Meanwhile, for the probabilistic guidance, underforecasting (overforecasting) for light (heavy) rainfall is present (not shown). As a result, the poorer reliabilities for both DET and PERs were largely attributed to their higher biases.

For AROC, the superiority of probabilistic guidance over DET was not as significant as that for reliability. Specifically, improvements of the probabilistic guidance in AROC over DET were statistically significant for moderate and heavy rainfall (Figure 9c). There were no significant differences in discrimination between PRO and OPTP for all types of rainfall. Overall, evident improvements (degradations) of NEP over PRO in discrimination were found in the forecasts of moderate (heavy) rainfall.

OPTP seemed to be more accurate than PRO in forecasting lighter rainfall, mainly due to the more reliable performance in the former. Additionally, the superiority of OPTP in discrimination was not significant, which resulted from the optimal-probability technique excluding fewer members in general. Since most of the PERs showed similar cost functions in Equation (4), especially the last few ones, the principle
used to select PERs in OPTP must be relaxed to include enough ensemble members. NEP significantly outperformed PRO in accuracy, reliability and discrimination for rainfall with smaller threshold, which was consistent with the conclusion of Schwartz and Sobash (2017) that NEP typically does not possess good reliability or resolution for rare events because of sharpness loss.

5 | RESULTS OF CASE-STUDIES

To further illustrate the performance of both deterministic and probabilistic guidance of GM-CPEPS, as well as to improve our understanding of the underlying physical causes of the performance, this section presents both forecasts initiated at 0000 UTC 8 May and 1200 UTC 10 May.

5.1 | Deterministic guidance

From 0900 to 1200 UTC 8 May, a local MCS, which formed in southwestern Guangdong (GD) Province, moved eastward and merged with some convective clusters on the southwest coast of GD, resulting in a quasi-linear MCS. This MCS caused severe rainfall (hereafter R1) at 1200 UTC, with 1 h accumulated rainfall over 60 mm (Figure 10a). DET yielded evident area overestimations in Guangxi (GX) Province during the whole forecasts as well as obvious westward displacement of R1 (Figure 10b). Although \( Q_{90} \) obviously overestimated the area of light rainfall, it did predict R1 well with a similar location to the observation (Figure 10c). Compared with \( Q_{90} \), PM evidently reduced the area overestimation of light rainfall (Figure 10d) but at the expense of more underestimations, which caused the lower FSS (Figure 7d). Additionally, PM predicted R1 more accurately than DET. Compared with both \( Q_{90} \) and PM, NPM showed better performance for R1, not only in location but also in size, despite still underestimating the intensity (Figure 10e). Generally, OPT performed the best for this local MCS (Figure 10f). In particular, OPT captured the observed characteristics of R1 well at 1200 UTC, with the predicted locations of 1 h accumulated rainfall exceeding 60 mm close to the observed ones (Figure 10a). OPT was actually member 14, which was one of the 16 PERs. Thus, the excessive smoothing of averaging, which was illustrated in Z18 and occurs when there are too many members with the same maximum cost function, was absent in the present case.

During the period from 1800 to 2100 UTC 10 May, a linear-shaped MCS formed in GX and moved southeastward, accompanied by a severe rain band R2 (Figure 10g). DET exhibited poor performance for this case, namely spurious precipitation at the border between GX and GD, and a northward displacement error for R2 (Figure 10h). \( Q_{90} \) reduced the intensity of spurious rainfall at the border between GX and GD in DET, but it also yielded evident area overestimations over western GD (Figure 10i). R2 was captured by \( Q_{90} \) despite the displacement error and intensity underestimation. PM underestimated the intensity of R2 seriously (Figure 10j). Overall, \( Q_{90} \) outperformed PM in forecasting R2 but underperformed for more spurious rainfall, and NPM performed between \( Q_{90} \) and PM (Figure 10k). Although OPT indeed showed some superiorities over DET in predicting R2, especially for location, it still fell behind \( Q_{90} \) in predicting R2 intensity (Figure 10i,l). For this large-scale MCS, both \( Q_{90} \) and NPM intuitively generated better guidance than DET in the forecasts of heavy rainfall. The evident displacement error for larger-scale MCSs present in single forecasts (i.e. DET and OPT) was partially corrected by including all the outcomes of PERs, which also inevitably resulted in false alarms.

5.2 | Probabilistic guidance

For the forecasts initialized at 0000 UTC 8 May, no PER was excluded in calculating OPTP, according to the principle used in the optimal-probability technique (Figure 11a). For R1 at 1200 UTC, whose intensity was seriously underestimated in DET (Figure 10b), PRO highlighted the occurrence of moderate (extremely heavy) rainfall with predicted probabilities of roughly 20% (5%) (Figure 11a), indicating the ability of GM-CPEPS to capture potential threats of severe weather. In addition, the predicted probabilities of PRO were below 20% in GX (Figure 11a), these being obviously smaller than those where R1 was observed. Therefore, the risk of excessive warning due to DET (Figure 10b) can be diminished to some extent with the help of GM-CPEPS. Obviously, the areas with non-zero probability in NEP were larger than those in PRO, well covering the areas where moderate rainfall of R1 was observed (Figure 11b). This aspect was absent in PRO (Figure 11a), indicating the better capacity for detection in NEP than in PRO. However, the probability of NEP was generally smaller than that of PRO, resulting directly from smoothing in the neighbourhood technique. Thus, severe weather with higher threshold (e.g. extremely heavy rainfall) cannot be identified well by NEP (Figure 11b).

In the second case-study, 12 members of the 16 PERs were used to calculate OPTP. Both PRO and OPTP identified the potential threat of R2, with the probability of moderate rainfall reaching 10–40% (Figure 11c) and 20–50% (Figure 11d), respectively. Additionally, partial maxima for extremely heavy rainfall were also captured by PRO and OPTP. Obviously, the probabilistic guidance provided us with additional information about potential severe weather, which was absent in DET. Similar to Z18, OPTP in this study also enhanced the probability of detection effectively but produced misleading guidance in the case of spurious rainfall. The absences of R2 in some areas in both PRO and OPTP were corrected evidently by NEP (Figure 11e). NEP also exhibited
FIGURE 10  One-hour accumulated rainfall (unit: mm) of AWS observation (a, g), DET (b, h) and various forecasts from deterministic guidance (c–f, i–l) valid at different lead times. R1 and R2 indicate the maximum observed 1 h accumulated rainfall valid at 1200 UTC 8 May and 2100 UTC 10 May respectively, and their outlines are contoured at the 10 mm threshold in orange lines. The regions of (a)–(f) are the same as the thick dashed range shown in Figure 1b, while the rectangles of (g)–(l) are the same as the thin dashed rectangle shown in Figure 1b.
FIGURE 11  Probability (%) of the 1 h accumulated rainfall exceeding 10 mm (shaded) and exceeding 40 mm (white contour for values greater than 5) for various forecasts from probabilistic guidance valid at different lead times. The outlines of the maximum observed 1 h accumulated rainfall marked as R1 and R2 are contoured at the 10 mm threshold in cyan lines. The regions of (a)–(b) are the same as the thick dashed range shown in Figure 1b, while the rectangles of (c)–(e) are the same as the thin dashed rectangle shown in Figure 1b

the advantage over PRO and OPTP in reducing the probability of spurious rainfall.

6 | CONCLUSIONS AND DISCUSSION

Poor CP-scale predictability of heavy rainfall over southern China in the pre-summer rainy season has severely limited QPF skill, thereby strongly calling for the development of a CP EPS. As part of this effort, an experimental GRAPES-based CP EPS, namely GM-CPEPS, has been established recently to provide forecasters with high-resolution deterministic and probabilistic products that are extremely scarce in current operational use.

GM-CPEPS was described fully herein, and its performance for QPF during SCMREX was also examined.

GM-CPEPS covers southern China and the northern SCS with a horizontal resolution of 0.03°. It consists of one unperturbed and 16 perturbed members, and it issues 12 h forecasts twice a day at 0000/1200 UTC. Several approaches are combined to generate perturbations of IC, LBCs and model physics, aiming at considering different benefits of various perturbation methods. Perturbations from downscaling, EDA, a time-lagged scheme and topography are combined with different weights to generate IC perturbations. Sea-surface temperature is also perturbed. LBCs perturbations are generated by blending downscaling perturbations with balanced random perturbations. Model physics
perturbations are generated based on the combination of MP, SPPT and PP schemes.

The sensitivities of GM-CPEPS performance to the different perturbation methods were tested in two cases, to validate the rationale of GM-CPEPS design. The current construction of GM-CPEPS, which is based on simply incorporating perturbations from all uncertainty sources using some leading perturbation methods, was proved to be effective in providing the best deterministic and probabilistic QPFs as well as ensemble dispersion for QPFs. Moreover, all the perturbation methods contributed to improved QPF performance but in different ways. Initial downscaling perturbations made the largest contributions followed by MP, whereas perturbations from LBCs made the smallest contributions.

The half-month performance of GM-CPEPS in the period of 8–23 May 2014 was evaluated by implementing objective verifications for both deterministic and probabilistic guidance. Two heavy rainfall cases, which were initialized at 0000 UTC 8 May and 1200 UTC 10 May, respectively, were used to investigate the GM-CPEPS performance for rainfall forecasting more intuitively and to illustrate the factors affecting its performance. The assessment was focused on the advantages of GM-CPEPS over DET and on comparisons among various post-process techniques.

Verification of the EM for both vertical and surface variables revealed the advantage of EM over DET in deterministic skill as well as the comparable performance in estimating forecast uncertainty compared with some existing operational or experimental regional EPSs. However, there was still underdispersion for all the surface variables, which requires further improvements in designing the perturbations.

For precipitation, various forecast distributions of GM-CPEPS with different percentiles were verified and compared with EM and DET. Generally, Q₉₀ (i.e. the forecast distribution with the 90th percentile) showed the highest deterministic skill, with significant improvements in light (moderate and heavy) rainfall over DET (EM). Although PM underperformed DET for light rainfall, the former improved QPFs over the latter for both moderate and heavy rainfall. NPM, which was a linear combination of Q₉₀ and PM based on their respective superiority in QPFs, was confirmed to be effective for showing better or comparable performance compared with Q₉₀ and PM. NPM was superior to DET for almost all types of rainfall, with most significant improvements in the forecasts of light rainfall at all lead times and the forecasts of moderate and heavy rainfall beyond the lead time of 6 h. Note that the parameters, e.g. the rainfall thresholds used to classify rainfall and the weighting factors for Q₉₀ and PM, which are used in constructing NPM, were currently selected without tuning. Thus, the performance of NPM could probably be improved by tuning these parameters, which will be explored in following work. An optimal-member technique, namely selecting the member closest to the NPM fields, was proposed here to improve the deterministic guidance for rainfall. The superiority of this technique over DET was evident in reducing the area overestimation of heavier rainfall.

Both the superiority and inferiority of various deterministic guidances were demonstrated intuitively in the case-studies. GM-CPEPS showed evident advantages over DET in predicting precipitation related to local or large-scale MCSs. The inferiority of PM in forecasting light rainfall was present, because the areal coverage of this type of rainfall was underestimated. Q₉₀ was apt to show spurious rainfall that caused serious area overestimation for most types of rainfall, whereas NPM could effectively reduce the area underestimation (over-estimation) of PM (Q₉₀) for light and extremely heavy rainfall (moderate and heavy rainfall). The superiorities over DET for the forecasts of heavy rainfall seemed to be different between NPM and OPT, especially for different MCS cases. Thus, exploring a new post-process technique including the advantages of both NPM and OPT merits further investigation. It cannot be ignored that both DET and PERs showed some biases in the forecasts, which caused GM-CPEPS’s unsatisfactory performance for reliability. Although NPM can reduce biases for some types of rainfall, it also added some biases compared with PM, especially for moderate and heavy rainfall. Thus, calibration based on some traditional methods (Hopson and Webster, 2010; Zhu and Luo, 2015) is required in following work to address this issue. Using calibrated NPM is expected to improve the performance of OPT and will also be investigated.

From the perspective of probability, the original probabilistic guidance performed significantly better than DET in not only accuracy but also reliability for all types of rainfall and in discrimination for moderate and heavy rainfall. Besides this guidance including all the 16 PERs, the probabilistic guidance based on the optimal-probability technique proposed in Z18 was also verified herein. Although this technique improved the probability of detection effectively, it caused misleading guidance in the case of spurious rainfall. However, this technique did not bring too much improvement except for light rainfall because there was too little variation between most PERs in the cost function used to determine the optimal member. The neighbourhood probability (i.e. NEP) with a neighbourhood length of 50 km was found to be effective in improving the accuracy, reliability and discrimination of probabilistic guidance for rainfall with smaller threshold. Because NEP includes additional information about neighbourhood locations, it can reduce the missing rates in the forecasts. Owing to the effect of smoothing caused by averaging the probability over neighbourhood grids, NEP can also reduce the probability of spurious rainfall and thereby improve the reliability. However, the averaging also causes sharpness loss, which was responsible for the limited superiority of NEP in heavier rainfall.

The case-studies demonstrated the significant advantage of GM-CPEPS probabilistic guidance over DET in detecting potential threats of severe weather and lessening excessive warning. The benefits of NEP were more remarkable for
reliability than for discrimination. It was noteworthy that the post-process techniques investigated in this study, i.e. OPTP and NEP, did lead to some improvements in the forecasts of cases with localized or large-scale forcing, but these improvements were still limited. Consequently, it is necessary to develop some new post-process techniques to further improve the probabilistic guidance for heavy rainfall forecasting. Even so, the discrimination, especially the probability of detection, may not necessarily be improved through post-process techniques.

Developing perturbation methods for IC, surface, and model physics that are more proper than those used in designing the current version of GM-CPEPS is probably the most effective way to improve the discrimination and will be explored in future work. Especially, TS perturbations can be potentially improved by implementing random perturbations with a spatial scale, and modifying the perturbations in the convectively unstable regions with convergent wind in SPPT also shows some positive impacts on the forecasts of heavy rainfall, based on the results of recent sensitivity tests in case-studies.

Although both deterministic and probabilistic guidance of GM-CPEPS showed some advantages over DET during SCMRREX, the corresponding verification was only conducted over a 15-day period, which is still not long enough to validate the generality of these results in the pre-summer rainy season. Consequently, more experiments covering more cases are needed to draw general conclusions about the performance of GM-CPEPS in QPF during the pre-season rainy season, which is very necessary for prompting the application of GM-CPEPS in operations.

ACKNOWLEDGEMENTS

This research was sponsored by the Natural Science Foundation of Guangdong Province through Grant 2017A030313225, Guangdong Province Science and Technology Project through Grant 2017B020244002, 2017B030314140, 2017B020218003, and National Natural Science Foundation of China through Grant 41405104. I would like to acknowledge the support with high-performance computing from Tianhe-2 provided by the National Supercomputing Center in Guangzhou. The author thanks two anonymous reviewers for their helpful comments, which evidently improved the quality of this paper; and also thanks Yali Luo, Huiling Yuan, Yuejian Zhu, Qilin Wan, and Quan Dong for helpful discussions.

REFERENCES

Ancell, B.C. (2013) Nonlinear characteristics of ensemble perturbation evolution and their application to forecasting high-impact events. Weather and Forecasting, 28, 1353–1365.

Baker, L., Rudd, A., Migliorini, S. and Bannister, R. (2014) Representation of model error in a convective-scale ensemble prediction system. Nonlinear Processes in Geophysics, 21, 19–39. https://doi.org/10.5194/npg-21-19-2014.

Bouttier, F., Vié, B., Nuisier, O. and Raynaud, L. (2012) Impact of stochastic physics in a convection-permitting ensemble. Monthly Weather Review, 140, 3706–3721.

Caron, J.F. (2013) Mismatching perturbations at the lateral boundaries in limited-area ensemble forecasting: a case study. Monthly Weather Review, 141, 356–374.

Chen, D.H., Xue, J.S., Yang, X.S., Zhang, H.L., Shen, X.S., Hu, J.L., Wang, Y., Ji, L.R. and Chen, J.B. (2008) New generation of multiscale NWP system (GRAPES): general scientific design. Chinese Science Bulletin, 53, 3433–3445.

Clark, A.J., Kain, J.S., Stensrud, D.J., Xue, M., Kong, F., Coniglio, M.C., Thomas, K.W., Wang, Y., Brewster, K., Gao, J., Wang, X., Weiss, S.J. and Du, J. (2011) Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. Monthly Weather Review, 139, 1410–1418.

Done, J., Davis, C.A. and Weisman, M.L. (2004) The next generation of NWP: explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. Atmospheric Science Letters, 5, 110–117.

Dong, J., Xue, M. and Droegemeier, K. (2011) The analysis and impact of simulated high resolution surface observations in addition to radar data for convective storms with an ensemble Kalman filter. Meteorology and Atmospheric Physics, 112, 41–61.

Duda, J.D., Wang, X., Kong, F. and Xue, M. (2014) Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. Monthly Weather Review, 142, 2196–2219.

Ebert, E.E. (2001) Ability of a poor man’s ensemble to predict the probability and distribution of precipitation. Monthly Weather Review, 129, 2461–2480. https://doi.org/10.1175/1520-0493(2001)129,2,461:AOAPMS.2.0.CO;2.

Ebert, E.E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. Meteorological Applications, 15, 51–64.

Gebhardt, C., Theis, S.E., Paulat, M. and Ben-Bouallègue, Z. (2011) Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmospheric Research, 100, 168–177. https://doi.org/10.1016/j.atmosres.2010.12.008.

Gentemann, C.L., Donlon, C.J., Stuart-Menteth, A. and Wentz, F.J. (2003) Diurnal signals in satellite sea surface temperature measurements. Journal of Geophysical Research, 108(C3), 803. https://doi.org/10.1029/2002JC001629.

Hacker, J.P., Ha, S.Y., Snyder, C., Berner, J., Eckel, F.A., Kuchera, E., Pocernich, M., Rugg, S., Schramm, J. and Wang, X. (2011) The U.S. Air Force Weather Agency’s mesoscale ensemble: scientific description and performance results. Tellus, 63A, 625–641.

Hagelin, S., Son, J., Svinhun, R., McCabe, A., Roberts, N. and Tennant, W. (2017) The Met Office convection-scale ensemble, MOGREPS-UK. Quarterly Journal of the Royal Meteorological Society, 143, 2846–2861. https://doi.org/10.1002/qj.3135.

Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting, 15, 559–570.
Mo, K.C. and Jiang, H.M.H. (2003) Influence of sea surface temperature anomalies in the Gulf of California on North American monsoon rainfall. *Journal of Geophysical Research*, 108, 4112. https://doi.org/10.1029/2002JD002043.

Morrison, H., Thompson, G. and Tatarskii, V. (2009) Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one- and two-moment schemes. *Monthly Weather Review*, 137, 991–1007.

Palmer TN, Buizza R, Doblas-Reyes F, Jung T, Leutbecher M, Shutt SJ, Steinheimer M, Weisheimer A. 2009. Stochastic parametrization and model uncertainty. Tech. Rep. ECMWF RD Tech. Memo. 598, 42 pp. Reading: ECMWF. Available at: http://www.ecmwf.int/publications/ (accessed October 2009).

Peralta, C., Ben Bouallégué, Z., Theis, S.E., Gebhardt, C. and Buchhold, M. (2012) Accounting for initial condition uncertainties in COSMO-DE-EPS. *Journal of Geophysical Research*, 117, D07108, 13pp. https://doi.org/10.1029/2011JD016381.

Raynaud, L. and Bouttier, F. (2016) Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quarterly Journal of the Royal Meteorological Society*, 142, 854–866.

Roberts, N.M. and Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136, 78–97.

Romine, G.S., Schwartz, C.S., Berner, J., Smith, K.R., Snyder, C., Anderson, J.L. and Weisman, M.L. (2014) Representing forecast error in a convection-permitting ensemble system. *Monthly Weather Review*, 142, 4519–4541. https://doi.org/10.1175/MWR-D-14-00100.1.

Roxy, M. (2014) Sensitivity of precipitation to sea surface temperature over the tropical summer monsoon region – and its quantification. *Climate Dynamics*, 43, 1159–1169. https://doi.org/10.1007/s00382-013-1881-y.

Saetra, O., Hersbach, H., Bidlot, J.R. and Richardson, D. (2004) Effects of observation errors on the statistics for ensemble spread and reliability. *Monthly Weather Review*, 132, 1478–1501. https://doi.org/10.1175/1520-0493(2004)132<1478:EOEOTR>2.0.CO;2.

Schwartz, C.S., Kain, J.S., Weiss, S.J., Xue, M., Bright, D.R., Kong, F., Thomas, K.W., Levitt, J.J., Coniglio, M.C. and Wandishin, M.S. (2010) Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, 25, 263–280.

Schwartz, C.S., Romine, G.S., Smith, K.R. and Weisman, M.L. (2014) Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Weather and Forecasting*, 29, 1295–1318.

Schwartz, C.S., Romine, G.S., Sobash, R.A., Fossell, K.R. and Weisman, M.L. (2015) NCAR’s experimental real-time convection-allowing ensemble prediction system. *Weather and Forecasting*, 30, 1645–1654. https://doi.org/10.1175/WAF-D-15-0103.1.

Schwartz, C.S. and Sobash, R.A. (2017) Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: a review and recommendations. *Monthly Weather Review*, 145, 3397–3418.

Snook, N., Xue, M. and Jung, Y. (2012) Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Monthly Weather Review*, 140, 2126–2146.

Stensrud, D.J., Bao, J. and Warner, T.T. (2000) Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, 128, 2077–2107.

Stensrud, D.J., Xue, M., Wicker, J.L., Kelleher, E.K., Foster, M.P., Schaefer, J.T., Schneider, R.S., Benjamin, S.G., Wegnandt, S.S., Ferree, J.T. and Tuell, J.P. (2009) Convective-scale warn-on-forecast system: a vision for 2020. *Bulletin of the American Meteorological Society*, 90, 1487–1499.

Tennant, W. (2015) Improving initial condition perturbations for MOGREPS-UK. *Quarterly Journal of the Royal Meteorological Society*, 141, 2324–2336. https://doi.org/10.1002/qj.2524.

Theis, S.E., Hense, A. and Damrath, U. (2005) Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteorological Applications*, 12, 257–269.

Torn, R.D. (2010) Performance of a mesoscale Ensemble Kalman Filter (EnKF) during the NOAA high-resolution hurricane test. *Monthly Weather Review*, 134, 4375–4392.

Vié, B., Nuisier, O. and Ducrocq, V. (2011) Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on initial conditions and lateral boundary conditions. *Monthly Weather Review*, 139, 403–423.
Wang, C., Huang, S., Chen, S., Chang, C. and Tsuboki, K. (2016) Cloud-resolving typhoon rainfall ensemble forecasts for Taiwan with large domain and extended range through time-lagged approach. Weather and Forecasting, 31(1), 151–172.

Wang, H., Lao, Y. and Jou, B.J. (2014) Initiation, maintenance, and properties of convection in an extreme rainfall event during SCMREX: observational analysis. Journal of Geophysical Research: Atmospheres, 119, 13206–13232. https://doi.org/10.1002/2014JD022339.

Wang, X., Barker, D.M., Snyder, C. and Hamill, T.M. (2008) A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. Monthly Weather Review, 136, 5116–5131. https://doi.org/10.1175/2008MWR2444.1.

Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Sahdan, S., Tian, W., Ma, X., Tascu, S. and Bazil, E. (2011) The central European limited-area ensemble forecasting system: ALADIN-LAEF. Quarterly Journal of the Royal Meteorological Society, 137, 483–502. https://doi.org/10.1002/qj.751.

Weisman, M.L., Davis, C., Wang, W., Manning, K.W. and Klemp, J.B. (2008) Experiences with 0–36-h explicit convective forecasts with the WRF-ARW Model. Weather and Forecasting, 23, 407–437.

Wilks, D.S. (2006) Statistical Methods in the Atmospheric Sciences. New York: Academic Press, pp. 284–292.

Wu, D., Meng, Z. and Yan, D. (2013) The predictability of a squall line in south China on 23 April 2007. Advances in Atmospheric Sciences, 30(2), 485–502.

Wu, M. and Lao, Y. (2016) Mesoscale observational analysis of lifting mechanism of a warm-sector convective system producing the maximal daily precipitation in China mainland during pre-summer rainy season of 2015. Journal of Meteorological Research, 30, 719–736. doi:10.1007/s13351-016-6,089-8

Xia, R., Zhang, D.L. and Wang, B. (2015) A 6-yr cloud-to-ground lightning climatology and its relationship to rainfall over central and eastern China. Journal of Applied Meteorology and Climatology, 54, 2443–2460. https://doi.org/10.1175/JAMC-D-15-0029.1.

Xie, Y., Koch, S., McGinley, J., Albers, S., Bieringer, P.E., Wolfson, M. and Chan, M. (2011) A space–time multiscale analysis system: a sequential variational analysis approach. Monthly Weather Review, 139, 1224–1240.

Xue, J.S. and Liu, Y. (2007) Numerical weather prediction in China in the new century: progress, problems and prospects. Advances in Atmospheric Sciences, 24(6), 1099–1108. doi:10.1007/s00376-007-1,099-1

Xue, J.S., Zhuang, S.Y., Zhu, G.F., Zhang, H., Liu, Z.Q., Liu, Y. and Zhaan, Z.R. (2008) Scientific design and preliminary results of three-dimensional variational data assimilation system of GRAPES. Chinese Science Bulletin, 53, 3446–3457.

Zhang, F., Snyder, C. and Rotunno, R. (2003) Effects of moist convection on mesoscale predictability. Journal of the Atmospheric Sciences, 60, 1173–1185. https://doi.org/10.1175/1520-0469(2003)060<1173:EMCOM>2.0.CO;2.

Zhang, X. (2018) A GRAPES-based mesoscale ensemble prediction system for tropical cyclone forecasting: configuration and performance. Quarterly Journal of the Royal Meteorological Society, 144, 478–498. https://doi.org/10.1002/qj.3220.

Zhang, X., Luo, Y., Wang, Q., Ding, W. and Sun, J. (2016) Impact of assimilating wind profiling radar observations on convection-permitting quantitative precipitation forecasts during SCMREX. Weather and Forecasting, 31, 1271–1292.

Zhu, Y. and Luo, Y. (2015) Precipitation calibration based on the frequency-matching method. Weather and Forecasting, 30, 1109–1124.

How to cite this article: Zhang X. Application of a convection-permitting ensemble prediction system to quantitative precipitation forecasts over southern China: Preliminary results during SCMREX. Q J R Meteorol Soc 2018;144:2842–2862. https://doi.org/10.1002/qj.3411