The Impact of a Stochastically Perturbing Microphysics Scheme on an Idealized Supercell Storm

XIAOSH QIAO
Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science and Technology, and Key Laboratory of Transportation Meteorology, China Meteorological Administration, Nanjing, and Shenyang Central Meteorological Observatory, Shenyang, China

SHIZHANG WANG AND JINZHONG MIN
Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science and Technology, and Jiangsu Research Institute of Meteorological Science, Nanjing, China

(Manuscript received 12 March 2017, in final form 10 October 2017)

ABSTRACT

The concept of stochastic parameterization provides an opportunity to represent spatiotemporal errors caused by microphysics schemes that play important roles in supercell simulations. In this study, two stochastic methods, the stochastically perturbed temperature tendency from microphysics (SPTTM) method and the stochastically perturbed intercept parameters of microphysics (SPIPM) method, are implemented within the Lin scheme, which is based on the Advanced Regional Prediction System (ARPS) model, and are tested using an idealized supercell case. The SPTTM and SPIPM methods perturb the temperature tendency and the intercept parameters (IPs), respectively. Both methods use recursive filters to generate horizontally smooth perturbations and adopt the barotropic structure for the perturbation $r$, which is multiplied by tendencies or parameters from this parameterization. A double-moment microphysics scheme is used for the truth run. Compared to the multiparameter method, which uses randomly perturbed prescribed parameters, stochastic methods often produce larger ensemble spreads and better forecast the intensity of updraft helicity (UH). The SPTTM method better predicts the intensity by intensifying the midlevel heating with its positive perturbation $r$, whereas it performs worse in the presence of negative perturbation. In contrast, the SPIPM method can increase the intensity of UH by either positive or negative perturbation, which increases the likelihood for members to predict strong UH.

1. Introduction

Convection-permitting ensemble forecasts have been widely studied for many years (e.g., Stensrud et al. 2000; Tong and Xue 2008; Xue et al. 2009; Clark et al. 2010; Boultier et al. 2012; Yussouf and Stensrud 2012; Hally et al. 2014; Romine et al. 2014; Christensen et al. 2015; Snook et al. 2015; Watson et al. 2015; Schellander-Gorgas et al. 2017). Many of these studies (e.g., Xue et al. 2009; Snook et al. 2011; Boultier et al. 2012; Romine et al. 2014; Christensen et al. 2015) have accounted for the errors caused by parameterizations, including those of cumulus, radiation, planetary boundary layer (PBL), and microphysics, because these parameterizations are very important for storm simulations. Among these parameterizations, only the microphysics scheme directly simulates the phase transitions of water in storm clouds, such as the melting of hail or the evaporation of raindrops; thus, this scheme can have a great impact on storm simulations and has been the focus of many studies (e.g., Lin et al. 1983; Gilmore et al. 2004; Van Den Heever and Cotton 2004; Milbrandt and Yau 2005a; Morrison et al. 2005; Snook and Xue 2008; Dawson et al. 2010; Zheng and Chen 2014). Over the past several decades, numerous microphysics schemes have been developed, including the single-moment (SM) Lin scheme (Lin et al. 1983), the WRF single-moment 6-class microphysics scheme (WSM6; Hong et al. 2004), the 1.5-moment Thompson scheme (Thompson et al. 2004), the double-moment (DM) MY scheme (MY2M; Milbrandt and Yau 2005a), and the DM Morrison scheme (M2M; Morrison et al. 2005).
To date, none of these schemes have consistently outperformed all of the other schemes (Clark et al. 2012). For example, Gelpi (2016) studied several extreme precipitation events that occurred over Basque Country, which is located in northern Spain, and reported that the WSM6 scheme worked better in the western river basin and that the Lin scheme yielded smaller errors in the eastern river basin. Although DM schemes are usually better than SM schemes (Morison et al. 2009; Dawson et al. 2010; Rajeevan et al. 2010; Van Weverberg et al. 2014; Igel et al. 2015) in terms of producing more realistic stratiform precipitation regions in squall lines (Duda et al. 2014) and producing the smallest track errors for mesocyclones (Dawson and Palmer 2015; Dawson et al. 2015), the use of a DM scheme did not improve forecasts in several other cases (e.g., Wu and Petty 2010; Van Weverberg et al. 2013). Igel et al. (2015) noted that using a DM scheme to parameterize raindrop breakup processes is still poorly understood and implied that the forecasting of the mean diameters of raindrops using a DM scheme was not as accurate as expected. Because the uncertainty of the microphysics scheme is large and may play an important role in forecasts, it is reasonable to take this uncertainty into account when studying convective-scale ensemble forecast systems (Fresnay et al. 2012).

Several methods can be used to represent the errors caused by microphysics schemes, including (i) variation of microphysics scheme within the ensemble, (ii) varying the prescribed parameters (i.e., the input of microphysics) for each member, and (iii) stochastically perturbing the quantities involved in the microphysics parameterization at every time step and at every grid. These quantities can represent tendencies or any prescribed parameters. The first approach has been recognized as a multiphysics approach, which is easy to implement and has been widely used (e.g., Rao and Srinivas 2014; Berner et al. 2015; Ha et al. 2015; Dahl and Xue 2016; Federico 2016). However, this approach has some drawbacks, because the covariance of its ensemble is difficult to interpret (Hacker et al. 2011). The second approach is often named the multiparameter (MultiPara) method, which is used to estimate the uncertainties of parameters using a single microphysics scheme (e.g., Fresnay et al. 2012; Yussouf and Stensrud 2012). Yussouf and Stensrud (2012) determined that using MultiPara is helpful for producing accurate ensemble forecasts in idealized supercell cases. However, this method fixes the prescribed parameters for each member throughout the forecast while ignoring the fact that these parameters may vary substantially during forecasts because they represent complex microphysical processes (Khouider and Majda 2006). The third approach is known as the stochastic perturbation of parameterization. This approach is called the stochastically perturbed parameterization tendency (SPPT) method when it perturbs the tendency from the parameterization, whereas it is called the stochastically perturbed parameterization (SPP) when the prescribed parameters are perturbed. The former was first proposed by Buizza et al. (1999) and was defined as SPPT by Palmer et al. (2009). The earliest literature on the latter can be found in Bowler et al. (2008).

Romine et al. (2014) and Berner et al. (2015) examined the SPPT method in a convection-permitting ensemble prediction system and noted the benefit of this approach on the rapid growth of the ensemble spread. However, Berner et al. (2015) showed that the forecasting skill of the SPPT method is slightly poorer than that of MultiPara. This is partly due to the incorporation of several physics schemes in the SPPT implementation. The nonlinearity of the forecast model makes it difficult to determine the contribution of each perturbed physics scheme to the resulting predictions, which is similar to findings of Duda et al. (2014). Investigations of a purely stochastically perturbing microphysics scheme (e.g., Fresnay et al. 2012; Hally et al. 2014) showed that the performances of ensemble forecasts of convection activity are sensitive to microphysical perturbations in cases of weak predictability. Using a stochastic approach is thus supposed to improve the results of the ensemble forecasts of tornadic supercell cases, in which strong mesocyclone-scale rotations sometimes exhibit weak predictability, especially when large grid spacings (Clark et al. 2013) are used.

Thus, in light of the challenges associated with interpreting the output from a multiphysics ensemble and the weakness of the MultiPara ensemble in representing the spatiotemporal variations of errors, stochastic parameterization appears to be an ideal method. To date, many techniques of stochastically perturbing the cumulus (Bright and Mullen 2002; Teixeira and Reynolds 2008; Grell and Freitas 2014; Fowler et al. 2016) and PBL (Kober and Craig 2016) have been proposed, but to the best of the authors’ knowledge, no published studies have adequately investigated the impact of the stochastically perturbing microphysics scheme or the configuration of its parameters on supercell storms, especially on supercell-related mesocyclones. Therefore, this work is necessary.

SM schemes are still widely used in many convection-permitting studies (e.g., Rao and Srinivas 2014; Berner et al. 2015; Ha et al. 2015; Katragkou et al. 2015; Stegehuis et al. 2015; Dahl and Xue 2016; Supinie et al. 2016) and operation centers (e.g., Federico 2016; Nuissier et al. 2016; Schellander-Gorgas et al. 2017).
Therefore, it is also worth investigating whether the performances of SM schemes can be improved by stochastically perturbing the input of microphysics (e.g., IPs). Previous studies have indicated that IPs can vary significantly even within a single supercell storm (e.g., Dawson et al. 2010); thus, the errors associated with IPs are likely to be better represented through spatiotemporally stochastic perturbation than through the Multi-Para method. Although the observed variability in particle size distributions (PSDs) within supercell storms (Kumjian and Ryzhkov 2008) is not random, the stochastic approach still provides an opportunity to represent such variability to some extent.

Because this is a preliminary study that focuses on the application of stochastically perturbed microphysics to supercells, an idealized supercell storm case is used here. This approach has been used in many studies that have investigated the impacts of microphysics schemes on supercell storms (Van Den Heever and Cotton 2004; Snook and Xue 2008; Zheng and Chen 2014). To the best of the authors’ knowledge, stochastically perturbed microphysics schemes include one scheme that has been implemented based on the WRF Model and one scheme that has been implemented based on the Application of Research to Operations at the Mesoscale (AROME) model (Boutritt et al. 2012, 2016). Here, stochastic methods are implemented based on the Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001). This model has been successfully applied in many studies of supercells (e.g., Snook and Xue 2008; Schenkman et al. 2011, 2012; Zheng and Chen 2014). The primary goal of this study is to compare the performance of the stochastic approach and multiparameter approach on supercell ensemble forecasts. The secondary goal is to determine the sensitivity of supercell ensemble forecasts to stochastic parameterization. The remainder of this paper is organized as follows. In section 2, the general SPPT algorithm is reviewed and the SPPT designed for this study is given. The experimental designs are given in section 3, and the corresponding results are presented and discussed in section 4. This work is then summarized in section 5.

2. Methodology

a. The stochastic method

The design of the SPPT (Palmer et al. 2009) is based on the notion that errors of parameterization can be represented or sampled by a multiplicative term (Berner et al. 2015; Christensen et al. 2015). As previously described by many other authors (Boutritt et al. 2012; Romine et al. 2014; Berner et al. 2015; Christensen et al. 2015), this method is often applied to accumulated parameterization tendencies as follows:

$$\frac{\partial X}{\partial t} = D + (1 + r) \sum_{i=1}^{N} P_i,$$

where $\frac{\partial X}{\partial t}$ denotes the total tendency of the variable $X$, which consists of the horizontal wind components $u$ and $v$, the temperature $T$, and the water vapor mixing ratio $q_v$. The term $D$ is the tendency from the dynamic components, $P_i$ is the tendency from the $i$th parameterization, and $N$ is the total number of parameterization schemes. The random perturbation $r$ has a mean of zero. In this work, only the temperature tendency from the microphysics scheme is perturbed using Eq. (1) in order to easily investigate the impact of this perturbation on supercell simulations. This method is hereafter named the stochastically perturbed temperature tendency from microphysics (SPTTM) method. The tendencies of water-related variables are not perturbed in this work because it is difficult to interpret results when both types of variables are perturbed. Stochastically perturbing the tendencies of water-related variables will be examined in the future.

In the case of perturbing IPs, Eq. (1) becomes

$$\frac{\partial X}{\partial t} = D + \sum_{i=1}^{N-1} P_i + P_{\text{micro}}[(1 + r)N_{0t}],$$

where $P_{\text{micro}}$ is the tendency from the microphysics scheme and is a function of $N_{0t}$, which represents the IPs for rain $N_{0r}$, snow $N_{0s}$, and hail $N_{0h}$. This $N_{0r}$-related perturbation method is named the stochastically perturbed IPs of microphysics (SPIPM) method.

The effect of SPPT depends on the random perturbation $r$. In addition to the widely used spectral pattern generator (Boutritt et al. 2012; Romine et al. 2014; Berner et al. 2015), the random perturbation $r$ can also be produced using a recursive filter (Qiao et al. 2017). This recursive filter is identical to the one that was designed for the three-dimensional variational data assimilation system (Gao et al. 2004) in the ARPS package. This filter is used to yield the horizontal pattern of the perturbation $r$. The evolution of $r$ is driven by the following first-order autoregressive (AR1) process:

$$r(x, y, t + \Delta t) = (1 - \alpha)r(x, y, t) + \sqrt{\sigma}e(x, y, t),$$

where $\sigma$ is the noise amplitude, $\alpha$ is the autoregressive parameter that influences the decorrelation time of $\Delta t/\alpha$, $\Delta t$ is the time step of the model, and $e(x, y, t)$ is the noise at grid point $(x, y)$ that is obtained using the recursive filter. The initial value of $r(x, y, t)$ is set to zero. In the
vertical direction, the barotropic structure is adopted based on the consideration that the forecast error is often large in the boundary layer (Berner et al. 2011) and that the surface cold pool exerts a considerable impact on the evolution of small-scale convective systems.

To prevent the perturbation from being too large or too small and maintain numerical stability, minimum and maximum values are defined for $r$. For the SPTTM method, these values are defined as 0.1 and 2.0, respectively. For the SPIPM method, these values are defined as 0.1 and 10, respectively. Larger values are used for the latter method because IPs vary greatly in supercell storms (Dawson et al. 2010). Currently, the SPTTM and SPIPM methods are implemented based on the Lin scheme.

b. Metrics for forecast evaluation

Given that this work focuses on supercell storms and their related low-level spiral updrafts, the parameter of updraft helicity (UH) is used to evaluate the results of this experiment. This parameter is designed to track rotations in simulated storms and has been employed in the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) Spring Forecasting Experiments since 2008 (Clark et al. 2013). The UH is computed by integrating the vertical component of the helicity over a given layer. Following Qiao et al. (2017), the specified layer ranges from 0 to 3 km above ground level (AGL) to focus on low-level mesocyclones. The maximum UH (maxUH) of the entire domain is used as an indicator of the intensity of the low-level mesocyclone. The evolution of the UH is tracked using a modified method from Fiori et al. (2009). Additional details about this modified method can be found in Qiao et al. (2017).

As noted by Clark et al. (2013), UH is a very finescale feature and is associated with high amplitudes and sharp gradients. These small-scale features are sometimes unpredictable because they can be regarded as noise, which may increase the measured error (Roberts and Lean 2008). In this case, traditional gridpoint-by-gridpoint verification methods may not be desirable. Thus, the contiguous rain area (CRA; Ebert and McBride 2000) method is used to calculate the root-mean-square error (RMSE) of the prediction of the UH. The RMSE used here refers to the shifted error that has removed the displacement error from the total error (Ebert and McBride 2000). To determine the displacement error, the pattern matching is performed by overlaying the centers of UH from the truth run and the ensemble run. The CRA method is an object-oriented approach that is designed to isolate and verify the rain within a particular weather system; in this study, the CRA method is used for UH instead of rain.

Because of the finescale features of UH, the CRA approach is applied to each member instead of the ensemble mean to prevent the impact of smoothing. For brevity, only the average values of the RMSEs of the ensemble members are shown in section 4. However, some members will likely predict centers of UH that are located far from their true locations. Ebert and McBride (2000) classified these cases as “missed events.” In this study, predictions yielding UH centers that are more than 15 km away from their true locations are considered to be missed. This length is determined based on results demonstrating that the scale of UH greater than 50 m$^2$s$^{-2}$ is

---

**Table 1. Experimental designs for SPTTM.**

<table>
<thead>
<tr>
<th>Expt name</th>
<th>Horizontal scale (km)</th>
<th>$N_v$ ($m^4$)</th>
<th>$N_{vh}$ ($m^6$)</th>
<th>STD</th>
<th>Temporal scale (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP8e6_h100s0.5</td>
<td>100</td>
<td>$8 \times 10^6$</td>
<td>$4 \times 10^4$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP8e6_h50s0.5</td>
<td>50</td>
<td>$8 \times 10^6$</td>
<td>$4 \times 10^4$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP8e6_h20s0.5</td>
<td>20</td>
<td>$8 \times 10^6$</td>
<td>$4 \times 10^4$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP1e6_h100s0.5</td>
<td>100</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP1e6_h50s0.5</td>
<td>50</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP1e6_h20s0.5</td>
<td>20</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP8e5_h100s0.5</td>
<td>100</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP8e5_h50s0.5</td>
<td>50</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP8e5_h20s0.5</td>
<td>20</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>0.5</td>
<td>30</td>
</tr>
<tr>
<td>TP8e6_h100s0.25</td>
<td>100</td>
<td>$8 \times 10^6$</td>
<td>$4 \times 10^4$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP8e6_h50s0.25</td>
<td>50</td>
<td>$8 \times 10^6$</td>
<td>$4 \times 10^4$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP8e6_h20s0.25</td>
<td>20</td>
<td>$8 \times 10^6$</td>
<td>$4 \times 10^4$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP1e6_h100s0.25</td>
<td>100</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP1e6_h50s0.25</td>
<td>50</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP1e6_h20s0.25</td>
<td>20</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP8e5_h100s0.25</td>
<td>100</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP8e5_h50s0.25</td>
<td>50</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>0.25</td>
<td>30</td>
</tr>
<tr>
<td>TP8e5_h20s0.25</td>
<td>20</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>0.25</td>
<td>30</td>
</tr>
</tbody>
</table>
often less than 8 km and that only a few experiments have predicted UH locations more than 15 km away from the true location. Because the RMSE does not indicate the direction of the error, the evolution of maxUH is investigated for each member. To focus on the strongest stage of UH, maxUH values of greater than 300 m$^2$s$^{-2}$ are examined. This threshold is applied for evaluations of the RMSE, the evolution of maxUH, and the fractions skill score (FSS), which will be introduced later.

Although RMSE can be calculated member by member, this approach is not feasible for calculating the ensemble spread. Moreover, the CRA method used in this study can be applied only to a two-dimensional field and cannot be applied to three-dimensional variables. Therefore, the ensemble spread, which has been widely used in many other studies (e.g., Bouttier et al. 2012; Romine et al. 2014; Berner et al. 2015; Christensen et al. 2015) is calculated using a conventional method and

### Table 2. Experimental designs for SPIPM.

<table>
<thead>
<tr>
<th>Expt name</th>
<th>Horizontal scale (km)</th>
<th>$N_0$ ($m^{-1}$)</th>
<th>$N_0$ ($m^{-1}$)</th>
<th>STD</th>
<th>Temporal scale (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP8e6_h20S6.0</td>
<td>20</td>
<td>$8 \times 10^6$</td>
<td>$4 \times 10^4$</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>PP1e6_h20S6.0</td>
<td>20</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>PP8e5_h20S6.0</td>
<td>20</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>PP8e6_h50S6.0</td>
<td>50</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^4$</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>PP1e6_h50S6.0</td>
<td>50</td>
<td>$1 \times 10^6$</td>
<td>$9 \times 10^3$</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>PP8e5_h50S6.0</td>
<td>50</td>
<td>$8 \times 10^5$</td>
<td>$4 \times 10^3$</td>
<td>6</td>
<td>30</td>
</tr>
</tbody>
</table>

![Fig. 1](http://journals.ametsoc.org/doi/abs/10.1175/MWR-D-17-0064.1)
Based on the formula of Fortin et al. (2014). For brevity, the ensemble spread is calculated only for the parameters of $w, \theta, Z$ (reflectivity), and UH (i.e., values greater than 50 m$^2$ s$^{-2}$) at grid points with $Z$ values that are greater than 5 dBZ.

In addition to yielding the RMSE values, the CRA approach also provides spatial error information about the UH; however, it is not easy to interpret spatial errors using RMSE values. Therefore, only the RMSE and intensity error are evaluated using the CRA method. The FSS proposed by Roberts and Lean (2008) is applied to evaluate the accuracy of the spatial prediction of UH for each member. The percentile thresholds are used based on the difficulty of the FSS to focus on the spatial accuracy of the forecast in the presence of bias in the amount of rain (Roberts and Lean 2008). The 95th and 75th percentiles are selected as thresholds to focus on the forecast skill of strong UH. Meanwhile, given that UH is a localized quantity, only grid points with UH values of greater than 1 m$^2$ s$^{-2}$ are used in the FSS calculations.

Otherwise, the FSS will primarily reflect the forecast skill of zero UH.

In addition, the FSS is a scale-selective verification; thus, a spatial scale must be defined. Mittermaier and Roberts (2010) proposed that this scale should be large enough to encompass the length of a mesoscale system. Given the scale of UH and the displacement error mentioned above, the spatial scale of the FSS is empirically set to 15 km. Given that the FSS is calculated for each member and that showing the FSS values of all members will make figures difficult to understand, the average FSS values of each member (hereafter, mean FSS) are discussed below.

Notably, to comprehensively evaluate the performance of ensemble forecasts, other verification metrics, such as relative operating characteristic curves (ROC; Harvey et al. 1992), should be used. However, in order to perform ROC analysis, the sample size must be considered. In this study, the number of grids with UH values that are greater than 1 m$^2$ s$^{-2}$ is often less than 100 throughout the entire domain. For UH values

**FIG. 2.** The average values of the vertical stretching term of the vorticity equation (shaded) and perturbed potential temperature (blue contours) within 2 km of the center of UH for (a) the truth run, (b) Ctrl8e6, (c) Ctrl1e6, and (d) Ctrl8e5.
that are greater than \(100 \text{ m}^2\text{s}^{-2}\), the number of grids is reduced to 20 or less. Therefore, it is difficult to use ROC for UH diagnosis.

3. Experimental design

a. Settings of the truth run

The truth run is used to evaluate the ensemble performance. Since the performance of DM schemes is often superior to those of SM schemes, the MY2M scheme (Milbrandt and Yau 2005b) is selected for the truth run and an SM scheme (i.e., a Lin scheme) is used in the SPTTM and SPIPM ensembles. The remaining configurations of the truth run are briefly listed as follows and are identical to the settings of the 2-km runs used by Qiao et al. (2017). The domain comprises \(55 \times 55 \times 53\) grid points with a horizontal grid spacing of 2 km. A stretching function is applied to the vertical direction,
such that the resolution near the surface is 20 m, which gradually increases to 700 m near the top of the domain. The average of vertical grid spacing is approximately 400 m. Open conditions and rigid wall are selected for the lateral boundaries and the top boundary, respectively. No surface physics are employed for a flat-bottom boundary. A thermal bubble is used to initialize the supercell storm in a horizontally homogeneous environment generated by a sounding that was obtained on 20 May 1977 from Del City, Oklahoma. This thermal bubble is centered at $x = 60$, $y = 40$, and $z = 1.5$ km, with a maximum perturbed temperature of 6 K and horizontal and vertical radii of 10 and 1.5 km, respectively. A constant wind of $u = 3 \text{ m s}^{-1}$ and $v = 14 \text{ m s}^{-1}$ is subtracted from the sounding to prevent the storm from moving out of the domain (Tong and Xue 2005). Additional settings for the parameterizations are given in Qiao et al. (2017).

### b. Settings of control runs

Three deterministic forecasts using the Lin scheme with different IPs are designed to simulate the case in which errors due to the microphysics scheme are ignored. As in other studies (e.g., Gilmore et al. 2004; Snook and Xue 2008; Duda et al. 2014; Zheng and Chen 2014), the hail intercept ($N_{0h}$) and the rain intercept ($N_{0r}$) are defined separately for each control run. The default value of the snow intercept is not perturbed because it does not significantly influence the low-level spiral updraft (Snook and Xue 2008).

Early studies (e.g., Gilmore et al. 2004; Snook and Xue 2008) suggested that the IPs of the hail $N_{0h}$ and rain $N_{0r}$ can vary from $10^2$ to $10^6 \text{ m}^{-4}$ and from $10^5$ to $10^7 \text{ m}^{-4}$, respectively. However, Snook and Xue (2008) demonstrated that strong, long-lived vortices cannot be produced using very large or very small IPs. For this reason, the IPs for rain and hail used in this study are limited to the ranges of $10^5$–$10^6 \text{ m}^{-4}$ and $10^3$–$10^4 \text{ m}^{-4}$, respectively. Three control runs are named Ctrl8e6, Ctrl1e6, and Ctrl8e5, in which the word “Ctrl” in the experiment name represents a control run and the value after Ctrl indicates the value of $N_{0r}$ used in this experiment. The corresponding values of $N_{0h}$ in Ctrl8e6, Ctrl1e6, and Ctrl8e5 are $4 \times 10^4$, $9 \times 10^3$, and $4 \times 10^3 \text{ m}^{-4}$, respectively. The default values of $N_{0h}$ ($4 \times 10^4 \text{ m}^{-4}$) and $N_{0r}$ ($8 \times 10^6 \text{ m}^{-4}$) are set as the largest values because previous studies (e.g., Snook and Xue 2008; Dawson et al. 2010; Snook et al. 2015) have stated
that IPs that are greater than the default values are often unable to sustain a strong low-level vortex. The other model configurations of the control runs are identical to those of the truth run.

c. Settings of ensemble forecasts

Because the MultiPara ensemble has been successfully used in many studies (e.g., Fresnay et al. 2012; Yussouf and Stensrud 2012; Duda et al. 2014; Hally et al. 2014; Berner et al. 2015), a group of ensemble runs using the MultiPara method is designed here to serve as a series of reference ensemble runs for the evaluation of the SPTTM and SPIPM approaches. To focus on the impact of the model error, no initial perturbations are applied to the ensemble forecasts, as has been performed in other studies (e.g., Duda et al. 2016).

Three experiments, namely, MP8e6, MP1e6, and MP8e5, are designed based on the control runs, in which MP refers to the MultiPara method and the value after MP represents the mean $N_0h$ of its members. For example, MP8e6 uses a mean $N_0r$ value of $8 \times 10^6$ m$^{-4}$. The corresponding mean $N_0h$ values in MP8e6, MP1e6, and MP8e5 are $4 \times 10^4$, $9 \times 10^3$, and $4 \times 10^3$ m$^{-4}$, respectively, which are consistent with their control run counterparts. The perturbations of $N_0h$ and $N_0r$ for each member follow a normal distribution. The standard deviations (STD) of IPs are 30% of their mean values in this set of experiments. This configuration of the STD is designed to prevent zero or negative IPs. For example, the STD of $2.4 \times 10^6$ m$^{-4}$ for $N_0r$ works in MP8e6 but may result in a negative $N_0r$ value in MP8e5. As in Qiao et al. (2017), the ensemble size is 20. The other model configurations of these experiments are identical to those of the control run.

Here, three sets of experiments using SPTTM are designed to examine the performance of SPPT in a supercell simulation and to determine its optimal configurations. To accurately compare these experiments, the three combinations of $N_0h$ and $N_0r$ used in the control runs are separately assigned to the three sets of SPTTM experiments. In addition to IPs, the spatial decorrelation scale and the STD vary in...
different experiments. Three horizontal decorrelation scales of 100, 50, and 20 km are investigated. These values are defined based on the fact that the horizontal spatial scale of a supercell is approximately 50 km [e.g., Fig. 5 in Xue et al. (2001)]; thus, the impacts of SPTTM with perturbation scales that are larger than, comparable to, or smaller than the spatial scale of a supercell can be examined. The large and small STDs of the perturbation are set to 0.5 and 0.25, respectively. Examining different STDs helps determine the optimal perturbation amplitude because excessively large STDs may result in instability whereas excessively small STDs may be insufficient. A total of 18 experiments are defined and are listed in Table 1 and are named as follows: TP refers to the tendency perturbation; the value after TP indicates the value of \( N_0 \) used in the experiment; and the values after \( h \) and \( s \) denote the horizontal scale and the STD, respectively.

![Fig. 8. As in Fig. 3, but for experiments using SPTTM. The IPs in the plots decrease, from larger to smaller values, from top to bottom.](http://journals.ametsoc.org/doi/pdf/10.1175/MWR-D-17-0064.1)
For example, the experiment name TP8e6_h50s0.25 indicates that the temperature tendency from a microphysics scheme is perturbed with an STD of 0.25 and a spatial scale of 50 km. All experiments in Table 1 use a temporal decorrelation scale of 30 min. This temporal length is selected because our early experiments indicated that temporal scales greater than 60 min cause numerical instability and worse performance.

Additionally, the performance of SPIPM and its sensitivities to the horizontal scale and IPs are also investigated. The experimental settings for SPIPM are shown in Table 2. The PP in the experiment name indicates that the parameters of the physics scheme are perturbed. Our early investigation of IP perturbations indicated that perturbation patterns with a scale of 100 km are too smooth to represent the spatial variation of IPs in a storm, thus horizontal scales of 50 and 20 km are examined for the SPIPM experiments. Because IPs can vary significantly, an STD of 6.0 is used for the SPIPM experiments. The temporal decorrelation scale is set to 30 min because of the same reason discussed above.

4. Results

a. Results of control runs

First, the differences between the truth run and the three control runs are examined. No quantitative verification is performed for the control runs because the main purpose of performing these experiments is to determine the errors in the presence of an imperfect microphysics scheme.

Figure 1a shows that a clear hook echo with a maxUH of 350 m$^2$s$^{-2}$ appears in the truth run at 117 min, whereas all of the control runs produce maxUH values that are less than 260 m$^2$s$^{-2}$ at the same model time and are located to the east of their true locations (Figs. 1b–d). Moreover, the hook echoes of the control runs are not as clear as that of the truth run. Meanwhile, the coverage area of reflectivity values that are greater than 30 dBZ...
in the truth run is larger than those in the control runs. Overall, Fig. 1 indicates that the Lin scheme produces a smaller supercell and weaker UH values than the MY2 scheme. This result is consistent with those of Dawson et al. (2015), in which the DM scheme produced stronger vorticity than the SM scheme.

To further understand the differences in the UH structures between the truth run and the control runs, the vertical distributions of the stretching term of the vorticity equation ($z_s$, where the subscript $s$ means stretching) and the perturbed potential temperature ($\vartheta_0$) from the surface (level 3) to approximately 4.8 km (level 20) are investigated. The stretching term shown in Fig. 2 represents the average value of $z_s$ within a $3 \times 3 \times 3$ box centered at the location of the maxUH value. The other quantity is also averaged using the same method. Figure 2a demonstrates that there are several times when strong $z_s$ (>0.0065 s$^{-2}$) extends from the midlevel to lower levels. These extensions of the strong $z_s$ are followed by a strong maxUH (which will be shown in subsequent figures). The first time that this phenomenon appears is at approximately 90 min when it lasts for more than 10 min. The second, third, and fourth appearances occur between 120 and 140 min, and the fifth and sixth appearances occur after 150 min. For the sake of convenience, we define three stages of strong UH. The first stage lasts from 80 to 100 min, the second stage lasts from 120 to 140 min, and the third stage lasts from 150 to 170 min. Notably, a temporal discontinuity in $z_s$ can be observed in Fig. 2. This is because the location of maxUH does not always vary gradually. In addition to the distribution of $z_s$, other common features of these stages include perturbed potential temperatures that are often less than $-2$ K at level 6 before 150 min.

Compared with Fig. 1, the weaker maxUH at 117 min in the control runs is partly due to the smaller $z_s$ at the same time (Fig. 2). The potential temperatures that are much lower than that in the truth run ($-2$ vs $-1$ K at level 6) indicate that the cold pool is stronger in the control runs. This strong cold pool may segregate the midlevel cyclone from the low-level gust front (Snook and Xue 2008), which causes the weakening of the vertical stretching in Figs. 2b–d. The remainder of this section examines whether ensemble forecasts that allow
for errors due to the microphysical parameterization can capture the errors between the truth run and the control runs described above.

b. Results of the MultiPara ensemble

First, quantitative evaluations are performed for this set of experiments. To focus on the strongest stages of UH, RMSEs are calculated for model times with true maxUH values that are greater than 300 m$^2$s$^{-2}$. Figure 3a demonstrates that the RMSEs for all three experiments are small in the first stage but are much larger in the following two stages. MP8e5 produces the smallest RMSE in the second stage, which implies that using small IPs improves the performance of ensemble forecasts. The evolution of the intensity error shown in Fig. 3b also provides information similar to that using smaller IPs results in smaller error. The subsequent discussion of the FSS will reveal the reason that the intensity error in MP8e6 is the largest.

The evolution of maxUH from 90 to 180 min shows that the intensity forecast of UH appears to be better in MP1e6 and MP8e5 because more of their members predicted maxUH values greater than 300 m$^2$s$^{-2}$ during the second and third stages (Figs. 4b,c). However, no member predicted a maxUH evolution that is fully consistent with that of the truth run. Therefore, in this case, the main contribution of the MultiPara method to ensemble forecasts is that it increases the chances of maxUH being greater than 300 m$^2$s$^{-2}$ in each stage. Another feature of Fig. 4 is that all three ensemble forecasts tend to predict strong UH values during the third stage. This phenomenon will be further discussed below.

In addition to providing intensity information, Fig. 4 also demonstrates that the differences in the maxUH evolutions between different members are large in MP1e6 and MP8e5 but smaller in MP8e6. However, it is incorrect to deduce that the ensemble spread is larger in MP1e6 and MP8e5 than it is in MP8e6. To comprehensively evaluate the differences between members, the investigation of the ensemble spread of their model variables is required. Notably, the ensemble spread discussed in this section consists of displacement error, intensity error, and pattern error. It should thus not be used to make comparisons to the RMSEs shown in Fig. 3. Figure 5 shows that the ensemble spread in MP8e6 is larger than that in MP1e6 and MP8e5 between 60 and 120 min for all members except for Z. This result may be caused by the large STD used in MP8e6. However, after 120 min, the opposite situation occurs for UH. In comparison to Fig. 4, the smaller spread of UH in MP8e6 is partly due to the weaker intensity of the maxUH forecasts.

The FSS shown here is used to evaluate the spatial accuracy of these methods. It can be seen in Fig. 6 that the FSS values for MP1e6 and MP8e5 are greater than...
0.8 before 140 min for the 95th percentile, while they occur before 130 min for the 75th percentile. This result differs from those of other studies (Ebert and McBride 2000; Roberts and Lean 2008) in that the FSS tends to be higher for smaller percentiles. Further investigation (not shown) indicates that there are often fewer than 5 grids with percentile values that are greater than 95 for both the truth run and the ensemble run. In this case, the frequency of a hit event for the 95th percentile is often 100% when maxUH is located within 15 km (the scale used for the FSS) of its true location. For the 75th percentile, more grids are involved in the FSS calculation; thus, the FSS reflects not only the location accuracy of maxUH but also the accuracy of the coverage (or shape) of UH values that are greater than 1 m² s⁻² around the location of maxUH. Therefore, it is acceptable that the FSS for the 75th percentile is smaller than its corresponding value for the 95th percentile. MP8e6 yields the lowest FSS for the 95th and 75th percentiles during the second and third stages, which indicates that the maxUH in this experiment is located more than 15 km from its true location. This result indicates that the large displacement error causes the large intensity error in MP8e6 because the UH forecast in this experiment is classified as missed. Since MP1e6 and MP8e5 produce better results in terms of their RMSE and FSS values, two members that predict strong UH in the second stage are selected from both experiments to diagnose the vorticity structures. Figure 7 shows that strong 𝜖 values appear between 120 and 130 min in both members, which is consistent with the peaks of maxUH in Fig. 4b and Fig. 4c, respectively. Compared with
Fig. 2, the cold pools in both members, especially member 17 of MP8e5, are weaker than the corresponding control runs. Thus, perturbing the preset IPs improves the performance of the ensemble forecast on the UH intensity by producing a weaker cold pool in some members.

However, this strong $\zeta$, in the second stage lasts for less than 10 min in both members. As the $-2-K$ contour reaches level 6 at 130 min (Fig. 7b), the strong $\zeta$ disappears. This result indicates that the cold pool in the MultiPara ensemble is still too strong. Only a few members of MP1e6 and MP8e5 produce strong UH (Fig. 4), which implies that using the MultiPara ensemble cannot substantially increase the chance for members to produce a weak cold pool and strong UH.

In addition, Figs. 4 and 7 show that a strong $\zeta$ appears in the third stage in all control runs (except MP8e6) and selected members. This is the reason that maxUH is larger in Figs. 4b and 4c. However, the cold pool is still strong during this stage, which should not benefit the development of UH. Findings by Snook and Xue (2008) may partly explain the reason; they stated that the low-level spiral updraft requires a balance between the cold pool intensity and the environment inflow. The slow-moving storm in this case may influence the environmental conditions for approximately 150 min, which causes the formation of new balance and strong UH. Whether this result is valid for other cases, such as fast-moving storms, requires further study.

c. Results of the SPTTM ensemble

Compared to MP8e6, the RMSEs of most of the experiments using SPTTM with a $N_0r$ value of $8 \times 10^6$ m$^{-4}$ are larger in the first stage but smaller in the second and third stages (Fig. 8a). The average error in the second stage for TP8e6_h50s0.5 is substantially smaller than that of MP8e6. The SPTTM experiments with smaller IPs often produce smaller RMSEs than those with large IPs (Fig. 8c); this is consistent with the differences shown in Fig. 3. However, the parameters of SPTTM should be carefully selected. Using a horizontal scale of 100 km causes a much larger RMSE in TP1e6_h100s0.5 than that in TP1e6_h20s0.25 between 120 and 160 min. The intensity errors of most of the experiments (except for TP1e6_h100s0.5) that use an $N_0r$ value of $1 \times 10^6$ m$^{-4}$ are much less than those of the large IP experiments between 90 and 180 min, especially for TP1e6_h20s0.25 (Fig. 8d). Figure 8c shows that experiments with an STD of 0.25 often produce smaller RMSEs than those with an STD of 0.5. These results indicate that it is better to use a smaller horizontal scale and a smaller STD for SPTTM with an $N_0r$ value of $1 \times 10^6$ m$^{-4}$.

For experiments that use SPTTM with an $N_0r$ value of $8 \times 10^5$ m$^{-4}$, no RMSEs that are smaller than that of MP8e5 are found in the second stage. However, the RMSE in the third stage (Fig. 8d) is clearly smaller than that of MP8e5. Figure 8f indicates that the smaller RMSE in the third stage is partly due to the better forecast of the UH intensity because the intensity error is often less than 20 m$^2$s$^{-2}$ for all of the SPTTM experiments, especially TP8e5_h50s0.5. In addition, experiment TP8e5_h100s0.5, which uses a horizontal scale of 100 km and an STD of 0.5, still produces the worst results among the smaller-scale and STD experiments. Unlike experiments that use an $N_0r$ value of $1 \times 10^6$ m$^{-4}$, using an STD of 0.25 does not consistently yield a smaller RMSE than using an STD of 0.5.

The evolutions of maxUH (Figs. 9a,d,g and 10a,d,g) show that more members using SPTTM with an $N_0r$ value of $8 \times 10^6$ m$^{-4}$ predict maxUH values that are greater than 300 m$^2$s$^{-2}$ in the second and third stages when compared with the results shown in Fig. 4a. This difference indicates that the underprediction of the UH intensity in MP8e6 is alleviated by using SPTTM. For experiments that use an $N_0r$ value of $1 \times 10^6$ m$^{-4}$ (Figs. 9b,e,h and 10b,e,h), SPTTM does not always produce forecasts that are better than those obtained using the MultiPara method, implying that the horizontal decorrelation scale and STD should be carefully selected for SPTTM.

For horizontal decorrelation scale, most of the members in TP1e6_h100s0.5 produce weak UH intensity in the second stage (Fig. 9b). In contrast, more members in TP1e6_h20s0.5 yield maxUH exceeding the preset threshold (Fig. 9h) in this stage. For STD, the use of small STD does not certainly cause the intensity of maxUH being weaker than that using larger STD. In some cases, more members produce strong UH with smaller STD. This phenomenon can be observed in the
comparison of Figs. 9c and 10c. The UH intensities in the third stage are stronger in Figs. 10b,e,f,h, and 10i than their counterparts in Fig. 9. These results indicate that SPTTM performs better with small STD and horizontal scale in this idealized supercell case.

Unlike the MultiPara method, which gradually influences the ensemble spread, SPTTM appears to introduce a strong shock to the forecast because a rapid increase in the ensemble spread occurs for all investigated variables in almost all of the SPTTM experiments (Fig. 11). Figure 11 shows that the maximum values of this spread are greater than those in the MultiPara experiments. This result is consistent with the results of other studies (e.g., Romine et al. 2014; Berner et al. 2015), which indicated that stochastically perturbing the tendency from the parameterization benefits the growth of the ensemble spread. However, this result is not valid for UH, as the ensemble spread remains smaller than 40 m^2 s^{-2} after 60 min for SPTTM experiments (Figs. 11d,h,l), whereas this value is essentially smaller than the spread of UH for MP1e6 and MP8e5. Given that the RMSEs of the experiments using SPTTM with an \( N_0 \) value of \( 8 \times 10^6 \) are consistently smaller than that of MP8e6, the smaller ensemble spread for these experiments does not mean that SPTTM performs worse than the MultiPara method. This situation is also valid for SPTTM experiments with smaller IPs.

The evolution of the FSS during the SPTTM experiments (Fig. 12) indicates that the SPTTM substantially improves the spatial accuracy when using an \( N_0 \) value of \( 8 \times 10^6 \) m^{-4}. The FSS ranges from 0.2 to 0.8 during the second and third stages, while their corresponding values in MP8e6 are near zero. This is valid for both the 95th and 75th percentiles. However, no clear improvement is found for cases with smaller IPs. Moreover, some experiments, such as TP1e6_h100s0.5 (Figs. 12c,e) and TP8e5_h100s0.5 (Figs. 12d,f), perform substantially worse than MP1e6 and MP8e5, respectively. This result indicates that the use of a horizontal scale of 100 km is
not suitable for microphysics schemes in this supercell case, although using a scale of 150 km has worked reasonably well in other studies (e.g., Romine et al. 2014; Berner et al. 2015).

To further understand how stochastic perturbations impact the evolution of UH, the evolution of the mean perturbation $r$ within 8 km of the location of maxUH (Fig. 13) and its relation to the vorticity structure are qualitatively examined for each member. Two members are shown here. Both are selected because they yielded strong UH. The results shown in Fig. 14a indicate that continuously multiplying the temperature tendency by the positive $r$ (red curves, Fig. 13) results in strengthening of midlevel heating and low-level cooling. The strong $\zeta_s$ that extends from level 20 to nearly level 4 at approximately 120 min indicates that the positive $r$ may benefit the formation of strong UH. However, the strong $\zeta_s$ at the midlevels soon disappears as a result of the strong cold pool pushing the low-level vorticity away from below the midlevel vorticity. Therefore, the positive $r$ in SPTTM benefits the short-lived strong UH but is not valid for the long-lived UH.

The results in Fig. 14b indicate that a negative $r$ weakens both cold pool and midlevel latent heating;
thus, it cannot support the development of strong UH. When the perturbation \( r \) transitions from negative to positive, the strong \( \zeta \) appears and lasts for approximately 10 min (from 140 to 150 min). This result indicates that SPTTM can also produce strong UH by first weakening the cold pool through a negative \( r \) and then increasing both midlevel heating and low-level cooling to enhance the vertical stretching of vorticity.

Further examination (not shown) indicates that the relationship between UH and the perturbation \( r \) discussed above is not valid for all members, because it is not a linear relationship. The balance between the cold pool intensity and the environmental flow cannot be met by all of the members using SPTTM due to the stochastic nature of this method.

d. Results of the SPIPM ensemble

Figure 15a demonstrates that experiments using SPIPM with an \( N_{0r} \) value of \( 8 \times 10^4 \) m\(^{-4}\) produce substantially worse results than those that use smaller IPs. Compared to the results of TP8e6_h50s0.5, PP8e6_h50s6.0 does not produce significantly smaller RMSE. In contrast, the RMSEs in the experiments using SPIPM with smaller IPs are often smaller than those obtained using SPTTM. Figure 15b indicates that the large RMSEs in PP8e6_h50s6.0 and PP8e6_h20s6.0 are due to their large intensity errors. These results suggest that SPIPM performs better than SPTTM when using small IPs but performs comparably to the latter when large IPs are employed.

The results shown in Figs. 16a and 16b indicate that the large RMSEs in PP8e6_h50s6.0 and PP8e6_h20s6.0 are the result of the members in both experiments tending to either overpredict or underpredict the intensity of maxUH. For example, member 16 of PP8e6_h50s6.0 essentially captures three peaks of maxUH in the second stage, but all of these peaks are substantially higher than the magnitude of the truth simulation. In experiments using smaller IPs, more members produce maxUH intensities that are comparable to their true values, although overprediction still occurs in some members. The density of markers in Figs. 16c–f is higher than that in its SPTTM counterparts, which partly explains why the RMSEs of SPIPM experiments are smaller than those of SPTTM experiments. However, it is still difficult to determine a member in Figs. 16c–f that can accurately reproduce three peaks of maxUH, similar to that in the truth simulation. Therefore, applying SPIPM to the Lin scheme primarily improves the performance of UH intensity, but this approach is not valid.
for assessing temporal variations in the UH peaks during each stage.

Although the RMSEs of SPIPM experiments with small IPs are smaller than those of SPTTM experiments, the ensemble spread of the former is comparable to that of the latter in Fig. 17 after 60 min. Unlike SPTTM, which introduces a clear shock to the model, SPIPM gradually influences the ensemble spread (Fig. 17). Tuning parameters such as the horizontal scale and the STD does not substantially change the ensemble spread, especially for the reflectivity. However, these SPIPM experiments have a common feature: large IPs ($N_{or} = 8 \times 10^5$ m$^{-4}$) result in larger ensemble spreads for $w$ (Fig. 17a) and $\theta$ (Fig. 17b) from 60 to 140 min but a smaller spread for UH after 120 min (Fig. 17d). The production of large spreads of UH by large IPs in the early stage of the simulation is consistent with other studies (Snook and Xue 2008), which showed that large IPs can produce strong short-lived vorticity but do not benefit the maintenance of strong vorticity.

Figure 18a demonstrates that the SPIPM method accurately captures not only the intensity of UH but also its spatial accuracy. The FSS for SPIPM with an $N_{or}$ value of $8 \times 10^5$ m$^{-4}$ remains at nearly 1.0 before 160 min, which indicates that the predicted locations of maxUH in most of the members are within 15 km of the locations in the truth simulation. The FSS with a threshold of the 75th percentile remains near or greater than 0.8 throughout all three stages (Fig. 18b) for the SPIPM experiments that use small IPs, which indicates that the coverage area of the UH is accurately captured by most members in these experiments.

As in section 4c, the relationship between the perturbation $r$ and UH structure is investigated to further understand the impact of SPIPM on low-level spiral updraft. Two members (members 1 and 17) from PP1e6_h50s6.0 are selected for examination (Fig. 19). Both have negative perturbations after 80 min, whereas the perturbation in member 1 becomes positive after 120 min. The clearest feature in Fig. 20 is that strong stretching occurs at 120 min in both members when the perturbation is negative. The latent heating in the midlevels is also strong during the same period (the 3-K contour reaches level 14). This result differs from that of the SPTTM experiments, which have weak heating in the midlevels when the perturbation is negative. After 120 min, the cold pool in member 1 becomes stronger as the perturbation becomes positive. The positive perturbation increases the value of the IPs and thus results in the strengthening of the cold pool. As in the SPTTM experiment, this strong cold pool causes the rapid disappearance of strong stretching at midlevels. In contrast, strong stretching occurs again after 140 min in member 17 and lasts for approximately 20 min, which is accompanied by negative perturbations throughout this period. These results indicate that strong stretching may occur in the presence of either a positive or a negative perturbation when using SPIPM. This characteristic of SPIPM increases the chance of members to predict UH as strong as in the truth.
simulation and thus results in the much denser marks in Fig. 16 than that in Figs. 9 and 10.

To qualitatively evaluate the impact of a stochastically perturbing microphysics scheme on a storm, storm patterns at the first level above the ground (Fig. 21) are examined. Two members from TP8e5_h50s0.25 (member 8) and PP1e6_h50s6.0 (member 17) are selected. Both members are chosen because they predict strong maxUH. Compared to the truth run, in which reflectivity values that are greater than 30 dBZ cover nearly three-quarters of the domain, the coverage area of the reflectivity in member 8 of TP8e5_h50s0.25 (hereafter referred to as member 8) and member 17 of PP1e6_h50s6.0 (hereafter referred to as member 17) are smaller, especially that of member 17, in which the coverage area appears to represent less than half of the domain during all three stages. This small degree of coverage in member 17 is partly due to the increase in the diameters of raindrops caused by applying the negative perturbation r to IPs. This result occurred because smaller IPs result in larger raindrops that often form and fall near the updraft core; hence, the precipitation area becomes smaller. Although member 17 (Figs. 21c and 21f) has substantially fewer grids with large drops (>3 mm) than that observed in the truth run, its cold pool intensity is weaker than that in member 8 by 2 K; thus, it is more similar to the truth simulation.

With the strongest cold pool intensity, the shape of UH in member 8 becomes more elliptical because the outflow related to the rear-flank downdraft reaches farther southeast and results in a stretch of vorticity in the northwest–southeast direction. In contrast, with lower cold pool intensity, the shapes of UH in both the truth run and member 17 become approximately circular. This result partly explains why the FSS of TP8e5_h50s0.25 (Fig. 12f) is lower than 0.8 after 120 min while the corresponding FSS of PP1e6_h50s6.0 remains greater than approximately 0.8 before 140 min.

5. Summary and discussion

Here, two stochastic perturbation methods are designed to incorporate the errors associated with microphysics schemes into ensemble forecast systems. One perturbs the temperature tendency from the microphysics scheme, while the other one perturbs the IPs at every grid and at every time step. The former is named SPTTM and the latter is named SPIPM. The performances of both methods are compared to the performance of the MultiPara method using an idealized supercell case study.

Several conclusions can be drawn based on the results described in the previous section: (i) the use of an imperfect microphysics scheme may introduce forecast bias into the intensity of UH due to the strong cold pool; (ii) the use of the MultiPara method can improve the forecast of the intensity of UH for a few members, but the cold pools in most members are still strong; (iii) SPTTM performs better than MultiPara when forecasting the intensity by increasing the midlevel heating, but it performs worse with negative perturbations; (iv) SPIPM can improve forecasts of the intensity of UH with either positive or negative perturbations, increasing the chance for members to predict strong UH; and (v) the use of IPs smaller than the default value of the Lin scheme and a horizontal scale that is less than 100 km are helpful for improving the performance of the stochastic method used in this study.

This work represents a preliminary study of the stochastic perturbation of microphysics schemes in supercell simulations; the results obtained in this work demonstrate the potential benefits of this technique when it is applied to the ensemble forecasts of strong, long-lived, low-level spiral updraft. More work is required because the current
implementations of SPTTM and SPIPM are still imperfect. For example, the SPTTM scheme excludes the perturbation of water-related variables, which are also important for supercell simulations; therefore, it is necessary to determine whether perturbing the combination of tendencies of temperature and water-related variables can improve the performance of SPTTM. Moreover, the impact of using stochastic perturbation on a DM scheme has not yet been examined. Although some results have indicated that perturbing the IPs works better than perturbing the temperature tendency and that it is therefore preferable to perturb the PSD-related quantities such as the tendency of the number concentration from a DM scheme, the validity of this assumption requires further investigation.

**Acknowledgments.** This work is jointly sponsored by the Foundation of Beiji Ge (BJG201409), the National Natural Science Foundation of China (41505090), the National Natural Science Foundation of China (41430427), the National Natural Science Foundation of China (41505089), the Major State Basic Research Development Program of China (973 Program: 2013CB430102), the Startup Foundation for Introducing Talent

![FIG. 21. The mass-weighted mean diameters of raindrops (shaded), perturbed potential temperature (blue contours with an interval of 2 K), UH (dark contours with an interval of 50 m s⁻²), and reflectivity (red contours) at the first level above the ground for (left) the truth run, (middle) member 8 of TP8e5_h50s0.25, and (right) member 17 of PP1e6_h50s6.0 at (a)–(c) 112, (d)–(f) 132, and (g)–(i) 144 min.](image-url)
of NUIST 2014R007, and a project funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions.

REFERENCES


