# Testing Stochastic and Perturbed Parameter Methods in an Experimental 1-km Warn-on-Forecast System Using NSSL's Phased-Array Radar Observations

DEREK R. STRATMAN,[a,b] NUSRAT YUSSOUF,[a,b,c] CHRISTOPHER A. KERR,[a,b] BRIAN C. MATILLA,[a,b] JOHN R. LAWSON,[d] AND YAPING WANG[e]

[a] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[b] *NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma*
[c] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[d] *Bingham Research Center, Utah State University, Vernal, Utah*
[e] *SAIC at NOAA/NWS/NCEP/EMC, College Park, Maryland*

(Manuscript received 24 April 2023, in final form 9 November 2023, accepted 14 November 2023)

ABSTRACT: The success of the National Severe Storms Laboratory's (NSSL) experimental Warn-on-Forecast System (WoFS) to provide useful probabilistic guidance of severe and hazardous weather is mostly due to the frequent assimilation of observations, especially radar observations. Phased-array radar (PAR) technology, which is a potential candidate to replace the current U.S. operational radar network, would allow for even more rapid assimilation of radar observations by providing full-volumetric scans of the atmosphere every ~1 min. Based on previous studies, more frequent PAR data assimilation can lead to improved forecasts, but it can also lead to ensemble underdispersion and suboptimal observation assimilation. The use of stochastic and perturbed parameter methods to increase ensemble spread is a potential solution to this problem. In this study, four stochastic and perturbed parameter methods are assessed using a 1-km-scale version of the WoFS and include the stochastic kinetic energy backscatter (SKEB) scheme, the physically based stochastic perturbation (PSP) scheme, a fixed perturbed parameters (FPP) method, and a novel surface-model scheme blending (SMSB) method. Using NSSL PAR observations from the 9 May 2016 tornado outbreak, experiments are conducted to assess the impact of the methods individually, in different combinations, and with different cycling intervals. The results from these experiments reveal the potential benefits of stochastic and perturbed parameter methods for future versions of the WoFS. Stochastic and perturbed parameter methods can lead to more skillful forecasts during periods of storm development. Moreover, a combination of multiple methods can result in more skillful forecasts than using a single method.

SIGNIFICANCE STATEMENT: Phased-array radar technology allows for more frequent assimilation of radar observations into ensemble forecast systems like the experimental Warn-on-Forecast System. However, more frequent radar data assimilation can eventually cause issues for prediction systems due to the lack of ensemble spread. Thus, the purpose of this study is to explore the use of four stochastic and perturbed parameter methods in a next-generation Warn-on-Forecast System to generate ensemble spread and help prevent the issues from frequent radar data assimilation. Results from this study indicate the stochastic and perturbed parameter methods can improve forecasts of storms, especially during storm development.

KEYWORDS: Severe storms; Numerical weather prediction/forecasting; Short-range prediction; Data assimilation; Ensembles; Stochastic models

## 1. Introduction

NOAA's experimental Warn-on-Forecast System (WoFS; Stensrud et al. 2009, 2013; Wheatley et al. 2015; Skinner et al. 2018; Jones et al. 2020), which is developed and tested at the National Severe Storms Laboratory (NSSL), is a regional, on-demand, rapidly cycled, multi-physics ensemble data assimilation and prediction system designed to provide probabilistic forecast guidance of severe convective weather on watch-to-warning scales to, e.g., National Weather Service forecasters, emergency managers, and broadcast meteorologists. While the current version of the experimental WoFS is almost ready for transitioning into NWS operations after several successful years of producing useful probabilistic forecasts (e.g., Gallo et al. 2022; Burke et al. 2022), work is already underway at NSSL in developing the next-generation WoFS, which includes moving from 3- to 1-km horizontal grid spacing (Wang et al. 2022; Kerr et al. 2023). In addition to the increase in grid resolution, an increase in both the spatial and temporal density of assimilated radar observations is also being explored for the next-generation WoFS.

Radar observations are crucial to the success of WoFS due to the intra-storm information they provide and their ability to quickly spin up storms in the model analyses and forecasts through data assimilation (DA; Yussouf et al. 2013; Wheatley et al. 2015; Jones et al. 2015). The current experimental WoFS assimilates gridded radar observations every 15 min, but Stratman et al. (2020) determined that more frequent assimilation of radar data could benefit the WoFS by more quickly spinning up storms in the model, leading to better, more

*Corresponding author*: Derek R. Stratman, derek.stratman@noaa.gov

DOI: 10.1175/MWR-D-23-0095.1

accurate forecasts at longer lead times. For that study, they assimilated data from NSSL's National Weather Radar Testbed (NWRT) phased-array radar (PAR; Forsyth et al. 2005; Weber et al. 2007; Weber et al. 2021), which is a potential candidate to replace the current aging WSR-88D radars due to its ability to provide full volumetric scans of the surrounding atmosphere every ~1 min instead of every ~5 min with the WSR-88D radars. A common existing problem for storm-scale ensemble forecast systems, including the current experimental WoFS, is ensemble underdispersion (e.g., Romine et al. 2014), which can result in filter divergence and model imbalances. Thus, while a PAR system would enable the ability to more frequently assimilate full volumes of radar observations, even more frequent DA cycling could further exacerbate the underdispersion problem and eventually lead to poorer analyses and forecasts of severe storms.

Numerous techniques have been developed over the last couple of decades to help mitigate ensemble underdispersion by accounting for different sources of error within the data assimilation and model systems. Some of the techniques, such as additive noise (Dowell and Wicker 2009; Sobash and Wicker 2015) and covariance inflation (Zhang et al. 2004; Houtekamer and Mitchell 2005; Anderson 2009; Whitaker and Hamill 2012), were developed to address ensemble underdispersion due to frequent DA. Another way to increase ensemble spread while accounting for various systematic errors is by adding perturbations to the initial (Leutbecher and Palmer 2008; Johnson and Wang 2016; Lang et al. 2019; Schwartz et al. 2020, 2022) and lateral-boundary conditions (Vié et al. 2011; Romine et al. 2014; Zhang 2019), which can improve ensemble forecasts of severe and hazardous weather. Some common methods for operational ensemble forecast systems to account for model errors and improve ensemble spread include multi-model (e.g., Candille 2009; Roberts et al. 2019), multi-physics (e.g., Stensrud et al. 2000; Fujita et al. 2007; Meng and Zhang 2007; Charron et al. 2010; Wheatley et al. 2015; Jones et al. 2020), and multi-parameter (e.g., Murphy et al. 2004; Hacker et al. 2011; Thompson et al. 2021) approaches. The current experimental WoFS (Jones et al. 2020) uses a combination of these methods, including mixed physics, perturbations to initial and lateral-boundary conditions, additive noise, and prior adaptive covariance inflation (Anderson 2009). However, even though the WoFS is a multi-physics ensemble, it currently does not use any techniques—e.g., stochastic or perturbed parameters—that account for uncertainty within each of those model physics (see e.g., Leutbecher et al. 2017).

There are numerous ways to introduce stochasticity and uncertainty into the model physics. For this study, four stochastic and perturbed parameter methods are explored using the PAR-observed 9 May 2016 tornado outbreak with an experimental 1-km-scale version of the WoFS. These methods include the stochastic kinetic energy backscatter (SKEB) scheme (Berner et al. 2009, 2011), the physically based stochastic perturbations (PSP) scheme using the planetary boundary layer schemes (Kober and Craig 2016; Rasp et al. 2018; Hirt et al. 2019), a fixed perturbed parameters (FPP) method using the microphysics scheme (Hacker et al. 2011; Christensen et al. 2015), and the surface-model scheme blending (SMSB) method, which is a novel technique introduced here that blends physical schemes in the land surface model. These four stochastic and perturbed parameter methods will be detailed in the following section. In section 3, an overview of the 9 May 2016 PAR observations will be described along with the data assimilation and forecast system, experiments, and evaluation methods. Results from the stochastic and perturbed parameter methods' experiments will be presented in section 4 followed by a summary and discussion of this study in the final section.

TABLE 1. SKEB scheme parameters and their values used in the experiments.

| SKEB parameter | Value |
|---|---|
| Backscatter rate for streamfunction | $1 \times 10^{-5}$ |
| Backscatter rate for potential temperature | $5 \times 10^{-6}$ |
| Decorrelation time | 1800 s |
| Spectral slope for perturbations | −1.83 |

## 2. Stochastic and perturbed parameter methods

### a. SKEB scheme

One of the common methods to introduce model uncertainty with stochastic perturbations is with the SKEB scheme (Berner et al. 2009, 2011). It is currently used in several operational ensemble prediction systems, including the National Centers for Environmental Prediction (NCEP)'s Global Ensemble Forecast System (GEFS; Zhou et al. 2022). The SKEB scheme used here was advanced and simplified by Berner et al. (2011) for the Advanced Research version of the Weather Research and Forecasting (WRF-ARW; Skamarock et al. 2008) Model following previous work by Shutts (2005) and Berner et al. (2009). In short, the SKEB scheme accounts for the variability in subgrid-scale processes by using a first-order autoregressive process to introduce spatially and temporally correlated perturbations in spectral space to the tendency terms of potential temperature and the rotational component of the horizontal wind via the streamfunction at each time step. As in Berner et al. (2011), the same perturbation pattern and values are used at all model levels.

The SKEB scheme was extensively tested and tuned for the 1-km-scale experiments by varying the backscatter rates, decorrelation times, and spectral slopes for perturbations for both streamfunction and potential temperature (not shown). The final parameter settings used in the study are detailed in Table 1 and are similar to previous studies that span a spectrum of spatial and temporal scales (e.g., Berner et al. 2011, 2012; Romine et al. 2014; Duda et al. 2016; Gasperoni et al. 2020), which suggests the generality of the SKEB parameter settings. Even so, the decorrelation time is smaller than those previous studies given that the forecasts for this study are shorter and for the next few hours as opposed to the next day or longer. An example of the spatial pattern of the perturbations for both components of the wind and

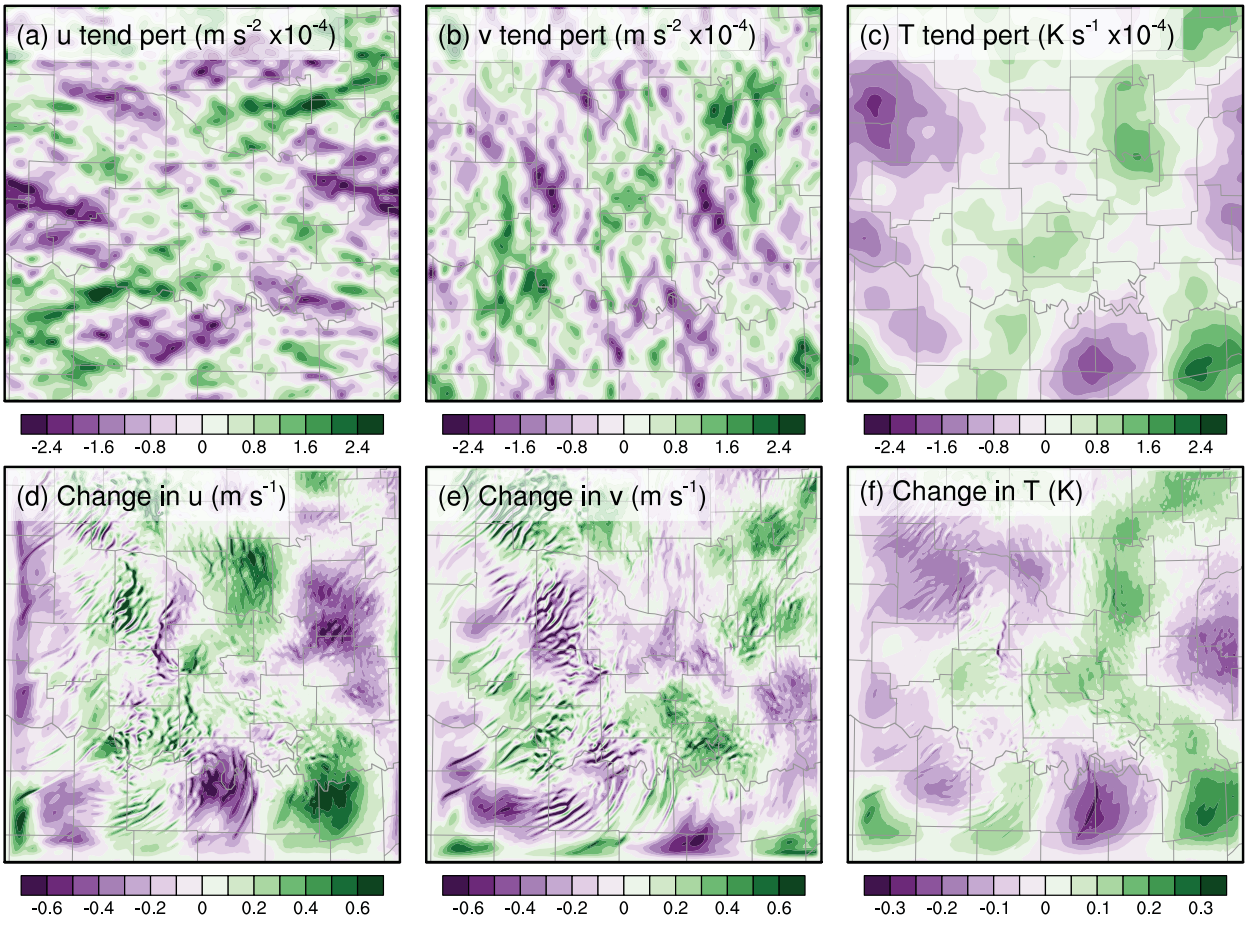


FIG. 1. (top) Example of the tendency perturbations for a single ensemble member for the (a) $u$ component of the wind (m s$^{-2}$), (b) $v$ component of the wind (m s$^{-2}$), and (c) potential temperature (K s$^{-1}$) from the SKEB scheme. (bottom) Example of the change in the (d) $u$ component of the wind (m s$^{-1}$), (e) $v$ component of the wind (m s$^{-1}$), and (f) potential temperature (K) fields for the same ensemble member after 1800 s, which is the decorrelation time of the perturbations.

potential temperature for a single ensemble member halfway through the decorrelation period (i.e., 900 s) are shown in Figs. 1a–c. By the end of the decorrelation time of 1800 s, these tendency perturbations lead to changes in the wind and potential temperature fields of around ±0.6 m s$^{-1}$ and ±0.3 K, respectively (Figs. 1d–f).

### b. *PSP scheme*

To help promote convection initiation in convection-allowing models, Kober and Craig (2016) proposed the PSP scheme as a way to reintroduce missing variability onto the grid scale within the boundary layer. The PSP scheme adds stochastic perturbations on the model's smallest resolvable scale with amplitudes proportional to the subgrid standard deviations of variables provided by the model's turbulence scheme. For this study, the following mathematical formulation from Rasp et al. (2018), which is a modification of the formula in Kober and Craig (2016), is used to introduce perturbations to the temperature $T$, water vapor mixing ratio qv, and vertical velocity $w$ tendencies:

$$\left(\frac{\partial \Phi}{\partial t}\right)_{\mathrm{PSP}} = \alpha_{\Phi}\eta\frac{1}{\tau_{\mathrm{eddy}}}\frac{l_{\mathrm{eddy}}}{\Delta x_{\mathrm{eff}}}\sqrt{\overline{\Phi'^2}}, \tag{1}$$

where $\Phi$ is the variable being perturbed, $\alpha_{\Phi}$ is the tuning parameter for each variable, $\eta$ is the 2D stochastic perturbation field, $\tau_{\mathrm{eddy}}$ is the representative eddy lifetime, $l_{\mathrm{eddy}}$ is the typical size of the largest eddies in a daytime convective boundary layer, $\Delta x_{\mathrm{eff}}$ is the effective model resolution, and $\sqrt{\overline{\Phi'^2}}$ is the subgrid standard deviation of $T$, qv, and $w$ provided by the turbulence schemes. The values for the PSP parameters used in the study are listed in Table 2 and are discussed next.

Ideally, the tuning parameter $\alpha_{\Phi}$ would be exactly 1 (Kober and Craig 2016), but previous studies that tested the PSP scheme found values greater than 1 worked best (Kober and Craig 2016; Rasp et al. 2018; Hirt et al. 2019). For this study, numerous sensitivity tests were completed, and the best results were achieved with $\alpha_{\Phi} < 1$ for all three variables. Additionally, it was found that using smaller tuning parameter values for $T$ and qv than for $w$ gave the best results for this

TABLE 2. PSP scheme parameters and their values.

| PSP parameter | Values |
|---|---|
| Tuning parameter $\alpha_\Phi$ | 0.5 ($w$), 0.05 ($T$, qv) |
| Representative eddy lifetime $\tau_{eddy}$ | 600 s |
| Largest typical eddy size $l_{eddy}$ | 1000 m |
| Effective model resolution $\Delta x_{eff}$ | 5000 m |
| Length scale | 5000 m |
| Decorrelation time | 600 s |
| Gridpoint standard deviation | 1.0 |
| Standard deviation cutoff | 3.0 |

case. These values are different from the previous studies, likely due to this study being the first to implement the PSP scheme into the WRF-ARW Model for *three* different boundary layer schemes and to use it in a mixed-physics ensemble prediction system. Following Rasp et al. (2018), $l_{eddy}$ and $\tau_{eddy}$ are set to 1000 m and 600 s, respectively, to represent the largest eddy sizes of a typical daytime convective boundary layer and their lifetime. Following the previous studies using the PSP scheme (Kober and Craig 2016; Rasp et al. 2018; Hirt et al. 2019), $\Delta x_{eff}$ is set to $5\Delta x$ or 5000 m.

The stochastic perturbation field $\eta$ has a mean of 0 with a gridpoint standard deviation of 1. The stochastic perturbations have a correlation length scale of 5000 m to approximate the smallest model-resolvable scale. The stochastic perturbation fields have a decorrelation time of 600 s to approximate the typical eddy lifetime and are evolved using a first-order autoregressive process similar to one used for the SKEB scheme in WRF-ARW. Hirt et al. (2019) showed that evolving the perturbation fields rather than holding their values constant, as in Kober and Craig (2016) and Rasp et al. (2018), resulted in comparable results. However, they concluded the evolving perturbation fields are still preferred since they are physically more realistic. Perturbations greater than three standard deviations are set to ±3. The variances $\Phi'^2$ for $T$, qv, and $w$ are calculated for the three boundary layer schemes used in this study following the level 2.5 model of Mellor and Yamada (1982) and Eqs. (4)–(6) in Kober and Craig (2016). Examples of the PSP scheme's tendency

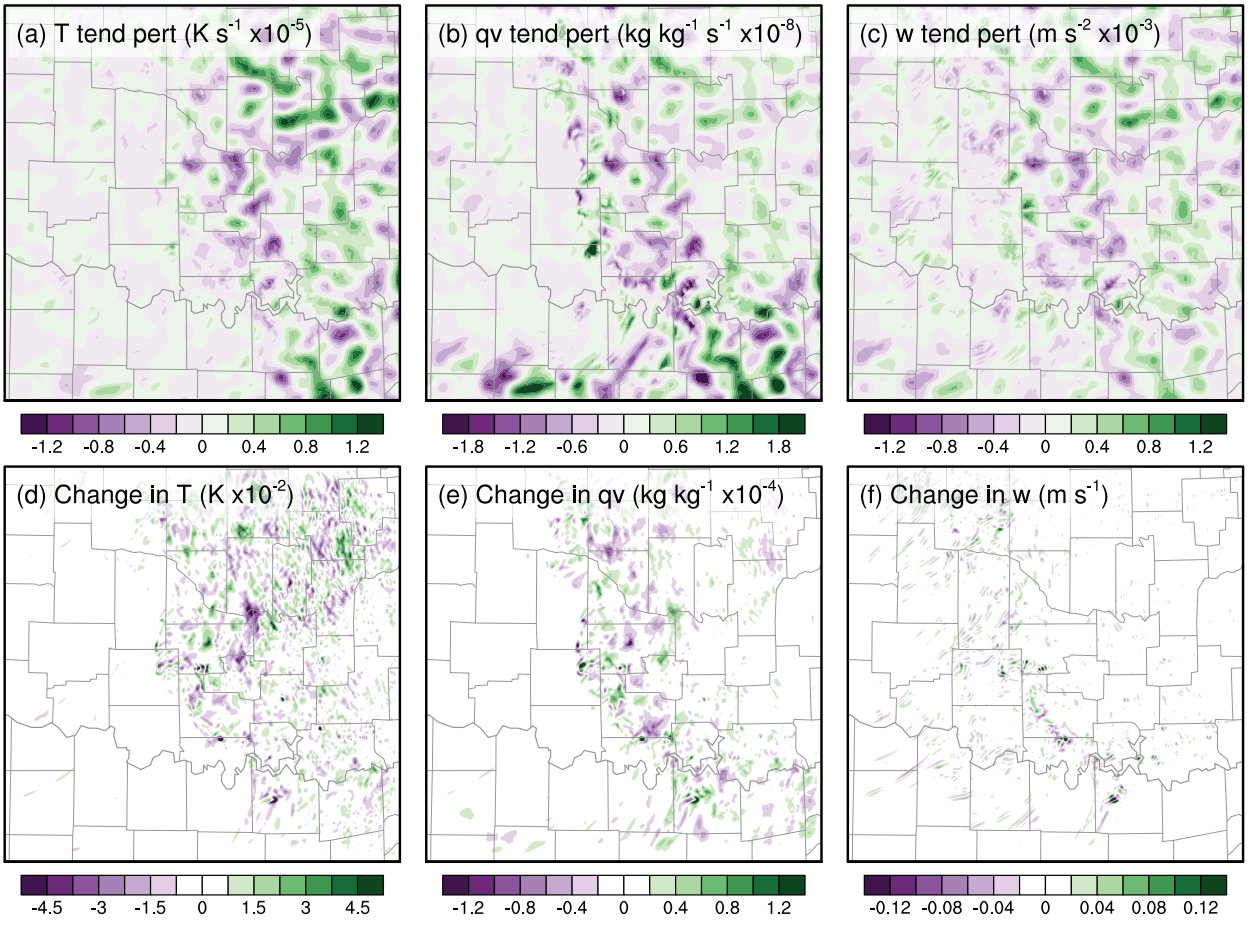


FIG. 2. (top) Example of the tendency perturbations for a single ensemble member for (a) potential temperature (K $s^{-1}$), (b) water vapor mixing ratio (kg $kg^{-1}$ $s^{-1}$), and (c) vertical velocity (m $s^{-2}$) produced by the PSP scheme. (bottom) Example of the change in the (d) potential temperature (K), (e) water vapor mixing ratio (kg $kg^{-1}$), and (f) vertical velocity (m $s^{-1}$) fields for the same ensemble member after 1800 s of model integration.

TABLE 3. NSSL two-moment microphysics parameters and their ranges of values.

| Microphysics parameter | Parameter ranges |
|---|---|
| ccn ($m^{-3}$) | [$0.3 \times 10^9$, $1.3 \times 10^9$] |
| alphah | [0.0, 3.0] |
| alphar | [0.0, 2.5] |
| ehw0 | [0.4, 1.0] |
| ehlw0 | [ehw0, 1.0] |

perturbations for $T$, qv, and $w$ and the changes to these fields after 30 min are shown in Fig. 2 and highlight how the scheme adds larger perturbations to areas with more variance in $T$, qv, and $w$ to help promote convection initiation. Also, the same perturbation pattern is used at all model levels just like the SKEB scheme, but unlike the SKEB scheme, the values can vary with model level with the largest perturbations often residing in the boundary layer (not shown).

### c. FPP method

We account for uncertainty in the microphysics scheme with a simple fixed perturbed parameter method similar to the one used in Christensen et al. (2015). In the FPP method, parameters are randomly determined within a specified range and held constant through the forecast model integration. As in the current experimental WoFS (Jones et al. 2020), the NSSL two-moment microphysics scheme (Mansell et al. 2010) is used for this study's ensemble prediction experiments. While several microphysics parameters can be perturbed, we settled on five parameters that in combination generally give optimal spread and skill based on preliminary testing: the base cloud condensation nuclei concentration (ccn), the shape parameters for graupel and rain size distributions (alphah and alphar, respectively), and the graupel–droplet and hail–droplet collection efficiencies (ehw0 and ehlw0, respectively). The parameter ranges are specified in Table 3 and are reasonably set based on previous studies (e.g., Mansell and Ziegler 2013 and T. Mansell, NOAA 2022, personal communication). Latin hypercube sampling with multidimensional uniformity (Deutsch and Deutsch 2012) is used to determine the fixed perturbed parameter values. This method ensures the full parameter phase space is adequately sampled and avoids clustering of randomly determined parameter values. While the FPP method results in more ensemble spread of microphysics-related variables in areas where model storms exist, it can also increase the ensemble spread of other model state variables, such as potential temperature. An example of the impact of the FPP method on the ensemble spread of composite reflectivity and near-surface potential temperature after an hour of model integration is shown in Fig. 3.

### d. SMSB method

A novel approach is applied to represent uncertainty within the land surface model, implemented via member-specific blending of different physical schemes within the model. The surface-model scheme blending (SMSB) method is used with the Noah land surface model with multiple-physics (Noah-MP; Niu et al. 2011; Yang et al. 2011; He et al. 2023) because of the numerous available physical scheme options. In this study, we blend the two available physical scheme options for both canopy stomatal resistance and lower boundary condition of soil temperature. We similarly blend the three available physical scheme options for both radiation transfer and soil moisture factor for stomatal resistance. Those schemes are blended together by first computing the associated parameters[1] for each of the physical scheme options. Next, these parameters are blended together by computing weighting coefficients for each physical scheme option, $n$, using the inverse distance weighting function:

$$\mathrm{wgt}_n = \frac{\frac{1}{d_n}}{\frac{1}{d_1} + \cdots + \frac{1}{d_N}}, \tag{2}$$

where $N$ is the total number of available schemes for a particular physical process (here $N$ is 2 or 3). The distances $d_n$ used in the weighting of each physical scheme option are computed using the equation:

$$d_n = \sqrt{(1 + r^2) - 2r\cos\left(\theta - \frac{2\pi}{n}\right)}, \tag{3}$$

where $r$ and $\theta$ are the radial and angular coordinates, respectively. These polar coordinates are stochastically determined for each ensemble member and physical scheme using Latin hypercube sampling with multidimensional uniformity (Deutsch and Deutsch 2012). To ensure uniform sampling, $r$ uses the square roots of the random values and has a range of [0, 1], while $\theta$ has a range of [0, $2\pi$]. These member-specific radial and angular coordinates are held fixed for all of the experiments using SMSB in this study. The resulting weights are also fixed across the domain for a given ensemble member and physical scheme. However, the blended parameters vary by grid point due to the spatial heterogeneity of the fields used to compute the parameters.

An example of the impacts of the SMSB method on four near-surface model state variables for a single ensemble member and on the ensemble spread for those same variables after 30 min of model integration are shown in Fig. 4. This example highlights that the impact of the SMSB method on the variables and their spread varies by location depending on the surface characteristics, such as land use. For example, the ensemble spread of the near-surface wind is generally smaller in the western part of the domain, which is mostly grassland, when using the SMSB method, while the eastern part of the domain, which is mostly woody savanna, has larger ensemble spread.

[1] The parameters for canopy stomatal resistance include sunlit and shaded leaf photosynthesis and sunlit and shaded leaf stomatal resistance. The parameter for lower boundary condition of soil temperature is energy influx from soil bottom. The parameters for radiation transfer include gap fraction for diffuse light and total gap fraction for direct beam. The parameter for lower boundary condition of soil temperature is the soil water transpiration factor.

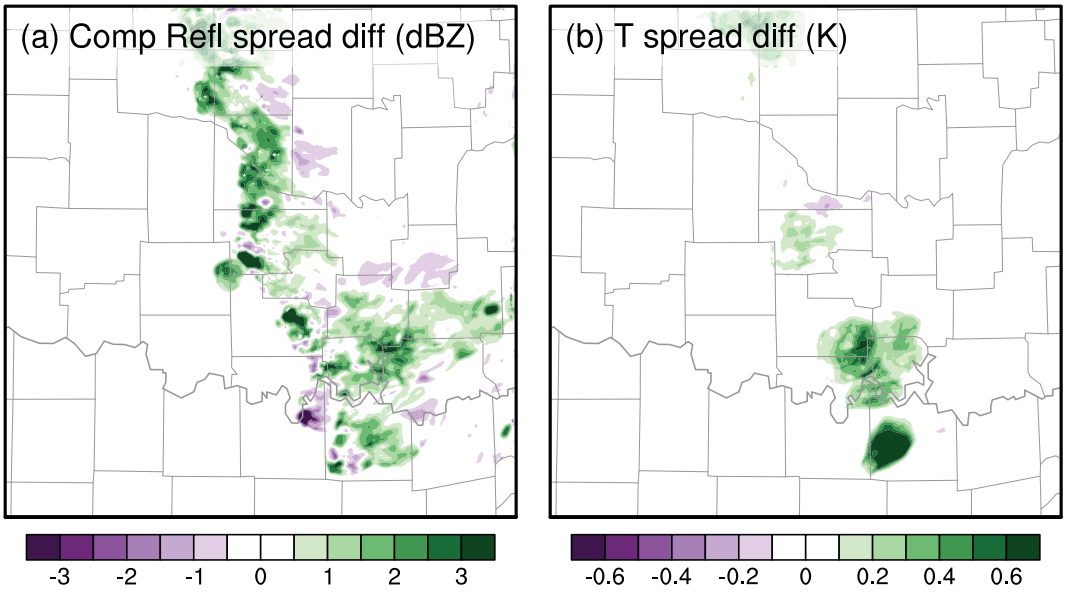


FIG. 3. Example of the difference in ensemble spread of (a) composite reflectivity (dBZ) and (b) near-surface potential temperature (K) with and without using the FPP method after 1 h of model integration. Green colors indicate where the FPP method added ensemble spread.

## 3. Experiment design

### a. PAR observations

During the 9 May 2016 Oklahoma tornado outbreak, 12 tornadoes, including one rated EF4 and three rated EF3, were produced by multiple supercell storms over the course of a few hours beginning around 2100 UTC (NOAA/NWS 2016). Unfortunately, two people were killed by the tornadoes along with a few injuries. For this study, we will focus on the first two tornadic supercells given their proximity to the

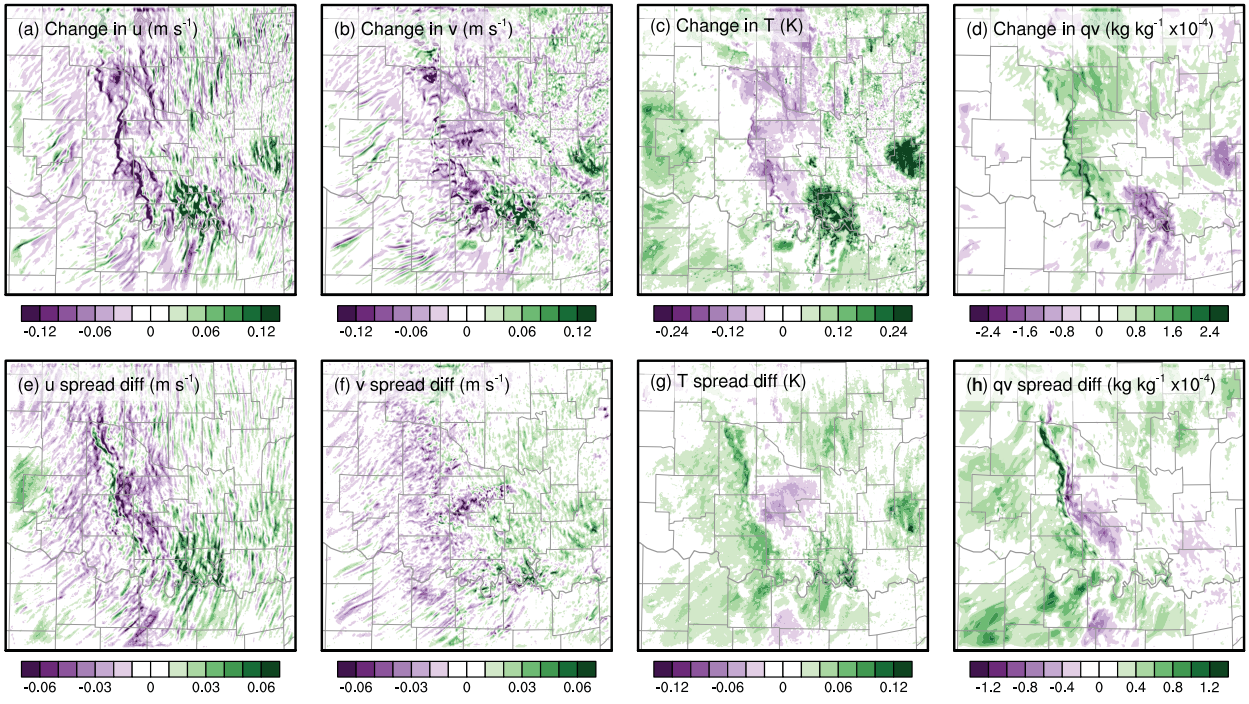


FIG. 4. (top) Example of the change in the near-surface (a) $u$ component of the wind (m s$^{-1}$), (b) $v$ component of the wind (m s$^{-1}$), (c) potential temperature (K), and (d) water vapor mixing ratio (kg kg$^{-1}$) fields for the same ensemble member after 1800 s of model integration. (bottom) Example of the difference in ensemble spread of near-surface (e) $u$ component of the wind (m s$^{-1}$), (f) $v$ component of the wind (m s$^{-1}$), (g) potential temperature (K), and (h) water vapor mixing ratio (kg kg$^{-1}$) fields after 1800 s of model integration.

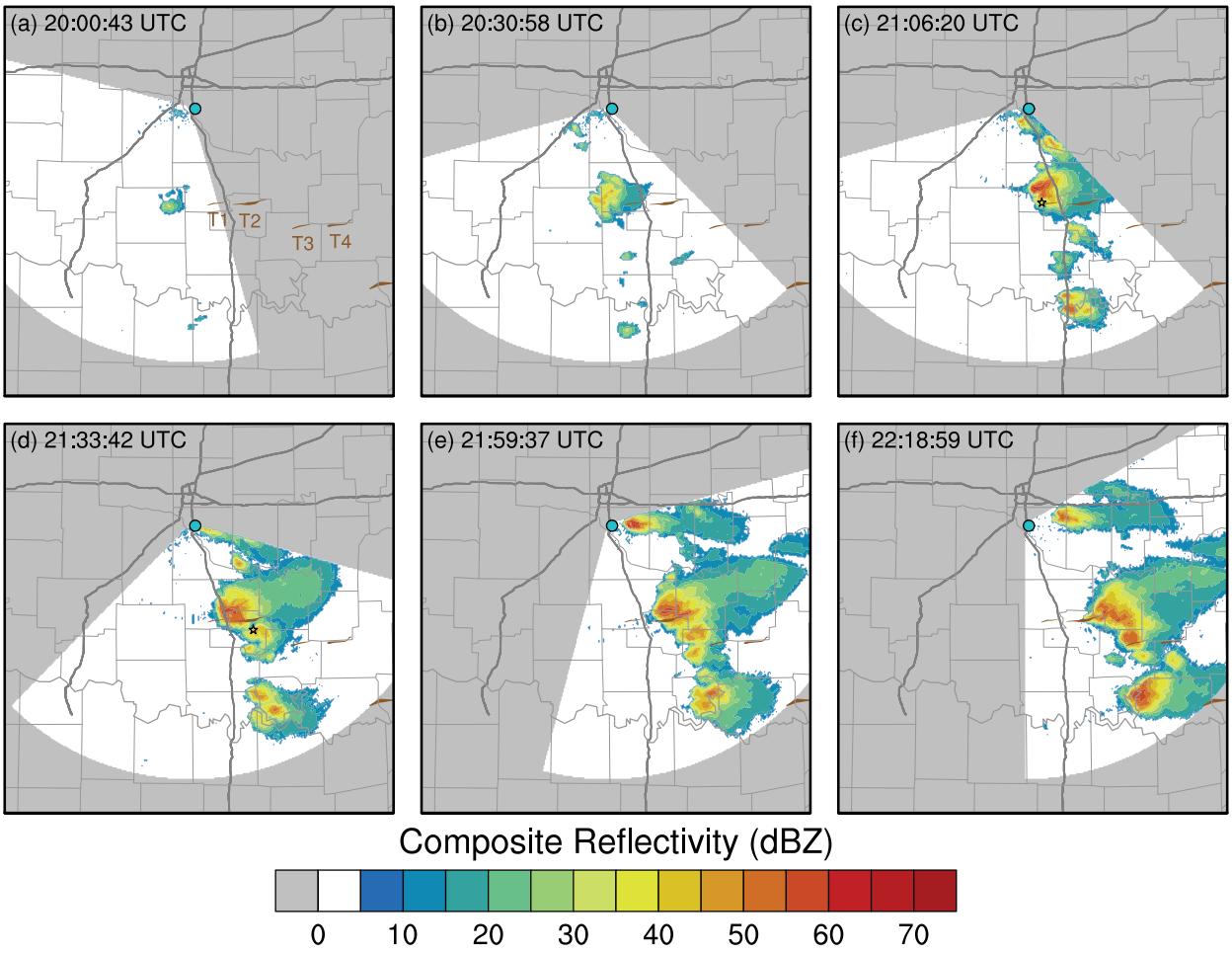


FIG. 5. (a)–(f) NWRT PAR composite reflectivity observations from six times during the 9 May 2016 tornadic supercell event in Oklahoma. The paths of the first four EF1+ tornadoes are chronologically annotated with T1, T2, T3, and T4 in (a). Black stars in (c) and (d) highlight the locations of Katie, OK, and Sulphur, OK, respectively.

NWRT PAR, which began scanning around 1800 UTC and concluded operations shortly before 2300 UTC. The first tornadic supercell of the event (S1) began developing around 2000 UTC (Fig. 5a) and first produced an EF4 tornado near Katie, Oklahoma (T1; 2106–2127 UTC; Fig. 5c). After a brief respite, S1 then produced an EF3 tornado (T2; 2134–2217 UTC; Fig. 5d), which passed to the north of Sulphur, Oklahoma. As S1 began to weaken, a new tornadic supercell (S2) developed to its southeast and first produced an EF3 (T3; 2218–2236 UTC; Fig. 5f). The second tornado by S2 was rated an EF1 (T4; 2246–2319 UTC). In addition to the tornadoes, these supercells also produced damaging hail up to the size of baseballs (2.75 in. or 7 cm in diameter).

PAR technology allows for faster sampling of weather since the radar beams are electronically steered via a set of radiating elements rather than mechanically aimed using a dish, which is required to rotate and tilt to fully sample the atmosphere (Weber et al. 2021). Because of the electronically steered beams, a PAR typically samples a single 90° volumetric sector in plan position indicator mode every ~1 min. However, for this event, the NWRT PAR was operated in range–height indicator mode while sampling a 120° sector. To sample a larger sector, two adjacent 60° sectors were sampled by rotating the PAR back and forth 60° after each subsector scan. Thus, full volumetric scans with 14 tilts ranging from 0.51° to 19.50° were produced every 90–100 min instead of every ~1 min. Additionally, the scanning sector was shifted during the event to follow the storms as they developed south of the radar and moved eastward (Fig. 5).

### b. *Data assimilation and forecast system*

The experiments in this study are completed using an ensemble data assimilation and forecast system similar to current experimental versions of the WoFS (e.g., Jones et al. 2020; Kerr et al. 2023). As with those systems, the Advanced Weather Research and Forecasting (WRF-ARW) Model (Skamarock et al. 2008) and the Development Testbed Center's ensemble square root filter (EnSRF; Whitaker and Hamill 2002) within the Gridpoint Statistical Interpolation (GSI) system (Kleist et al. 2009; Hu et al. 2016) are used to

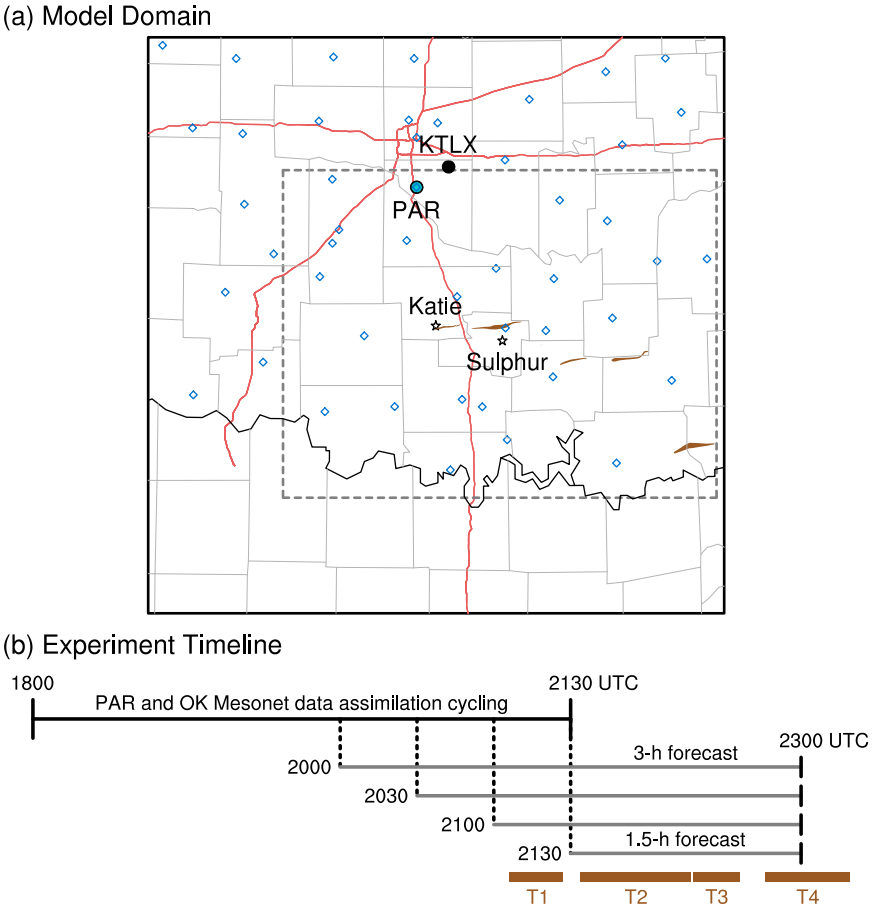


FIG. 6. (a) Model domain for experiments with PAR and KTLX locations (light blue and black circles, respectively), Oklahoma Mesonet locations (blue diamonds), and tornado paths (brown shading) overlaid. Dashed gray box outlines the area where the ensemble fraction skill score is computed. (b) Schematic of experiment time line along with an associated schematic of the tornado occurrence times.

generate ensemble analyses and forecasts of storms. The WoFS-like system has 36 ensemble members for data assimilation cycling and the free forecasts and is initialized at 1800 UTC using 1-h forecasts initialized from the experimental 36-member 1700 UTC High-Resolution Rapid Refresh ensemble's (HRRRE; Benjamin et al. 2016; Kalina et al. 2021; Dowell et al. 2022; James et al. 2022), which has 3-km horizontal grid spacing. The lateral-boundary conditions (LBCs) are provided by nine members of the 1500 UTC HRRRE, as in Skinner et al. (2018) for the 2017 experimental system. The 9 HRRRE members rotated through four times each in increasing sequential order (i.e., ensemble members groups 1–9, 10–18, 19–27, and 28–36 each use the nine HRRRE members).

The WoFS-like ensemble experiments are conducted on a 301 × 301 gridpoint domain (Fig. 6a) with 1-km horizontal grid spacing and 51 vertical levels. Data assimilation cycling, which is the process of repeatedly assimilating new observations and advancing the ensemble forecasts, is completed from 1800 to 2130 UTC (Fig. 6b). Ensemble forecasts are then initialized every 30 min from 2000 to 2130 UTC and integrated out until 2300 UTC. The physical parameterization schemes used in the ensemble forecasts for each member are the same as those used in the experimental 1-km-scale WoFS (Kerr et al. 2023), except that the Shin–Hong scale-aware boundary layer (SH; Shin and Hong 2015) scheme is used in place of the Yonsei University scheme to allow for the use of the PSP scheme. All members use the NSSL two-moment microphysics (Mansell et al. 2010) and Noah-MP (Niu et al. 2011; Yang et al. 2011) schemes. Physics diversity among the ensemble members (Table 4) is created by using the SH, Mellor–Yamada–Janjić (MYJ; Janjić 2001), and Mellor–Yamada–Nakanishi–Niino (MYNN; Nakanishi and Niino 2009) schemes for the planetary boundary layer, the Dudhia (Dudhia 1989) and Rapid Radiative Transfer Model–Global (RRTMG; Iacono et al. 2008) schemes for shortwave radiation, and the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) and the RRTMG schemes for longwave radiation.

To focus on the impact of stochastic and perturbed parameter methods on the rapid assimilation of radar data, only PAR and Oklahoma Mesonet observations are assimilated. The

TABLE 4. Physics scheme configuration for each ensemble member.

| Ensemble member | PBL scheme | SW radiation scheme | LW radiation scheme |
|---|---|---|---|
| 1, 7, 13, 19, 25, 31 | SH | Dudhia | RRTM |
| 2, 8, 14, 20, 26, 32 | SH | RRTMG | RRTMG |
| 3, 9, 15, 21, 27, 33 | MYJ | Dudhia | RRTM |
| 4, 10, 16, 22, 28, 34 | MYJ | RRTMG | RRTMG |
| 5, 11, 17, 23, 29, 35 | MYNN | Dudhia | RRTM |
| 6, 12, 18, 24, 30, 36 | MYNN | RRTMG | RRTMG |

Mesonet observations that are assimilated include 1.5 m AGL temperature, 1.5 m AGL dewpoint temperature, surface pressure, and 10 m AGL $u$- and $v$-wind components. Before they are able to be assimilated, the raw PAR reflectivity and radial velocity observations are processed using the Warning Decision Support System–Integrated Information (WDSS-II; Lakshmanan et al. 2007) program suite to quality control the radar data, dealias the radial velocity fields, and interpolate the observations onto a 1-km grid with 13 vertical levels, which consists of levels every 500 m for the 500–3000 m MSL layer and every 1000 m for the 3000–10 000 m MSL layer. Reflectivity intensities less than 15 dB$Z$ are set to 0 dB$Z$ and are thinned to a 4-km grid. As in Stratman et al. (2020), clear-air reflectivity from the Twin Lakes, Oklahoma, WSR-88D radar (KTLX) is used in the domain area outside of the PAR sector. Radial velocity observations are only assimilated in areas where reflectivity exceeds 15 dB$Z$. To avoid any potential issues with assimilating radar observations near the domain edges, radar observations within 30 km of the domain edges are not assimilated. The assumed observation errors for all Mesonet and PAR observations are listed in Table 5 along with the observations' horizontal and vertical localization radii, which are used in the Gaspari and Cohn (1999) localization function.

### c. Experiments

The experiments in this study are designed to explore different ways of using stochastic and perturbed parameter methods with the goal of helping guide development and implementation of these methods into future generations of the WoFS. In the first set of experiments, the stochastic and perturbed parameter methods are individually assessed by running experiments with 5-min cycling intervals for each method and comparing them to a Control experiment, which does not use stochastic or perturbed parameter methods (Table 6). Motivated by previous studies which found combining multiple methods generally led to better results than any single method (e.g., Jankov et al. 2017, 2019), another set of experiments is designed using either a combination of all of the stochastic and perturbed parameter methods (AllStoch) or a combination of only three of the methods (NoSKEB, NoPSP, NoFPP, NoSMSB). To see how the impact of the stochastic and perturbed parameter methods changes with DA cycling interval, experiments similar to the Control and one of the better performing combination experiments (i.e., NoPSP) are conducted using 2.5- and 15-min cycling intervals (Control2.5, NoPSP2.5, NoPSP15, Control15). For all of the stochastic and perturbed parameter experiments, the stochastic and perturbed methods are used during both the data assimilation cycling and the free forecasts.

TABLE 5. Observation types and their assumed errors and horizontal and vertical localization radii.

| Observation type | Error | $H$ local (km) | $V$ local [ln($p$/pref)] |
|---|---|---|---|
| Temperature | 1.0 K | 60.0 | 0.85 |
| Dewpoint temperature | 1.0 K | 60.0 | 0.85 |
| $u$ wind | 1.0 m s$^{-1}$ | 60.0 | 0.85 |
| $v$ wind | 1.0 m s$^{-1}$ | 60.0 | 0.85 |
| Surface pressure | 0.75 hPa | 60.0 | 0.85 |
| Reflectivity | 5.0 dB$Z$ | 6.0 | 0.80 |
| Clear-air reflectivity | 5.0 dB$Z$ | 6.0 | 0.80 |
| Radial velocity | 3.0 m s$^{-1}$ | 6.0 | 0.80 |

### d. Evaluation methods

Several subjective and objective approaches are taken to assess the impacts of the stochastic and perturbed parameter methods and whether any of those impacts could be beneficial for systems, such as the current and future generations of WoFS. To assess any changes in ensemble error and spread, observation-space diagnostics are computed for different observation types during the data assimilation cycling process and free forecasts. For the data assimilation cycling, the mean

TABLE 6. Experiment names and their descriptions.

| Expt name | Expt description |
|---|---|
| Control | 5-min cycling with no stochastic or perturbed parameter methods |
| SKEB | SKEB scheme only |
| PSP | PSP scheme only |
| FPP | FPP method only |
| SMSB | SMSB method only |
| AllStoch | All stochastic and perturbed parameter methods used |
| NoSKEB | Only PSP, FPP, and SMSB used |
| NoPSP | Only SKEB, FPP, and SMSB used |
| NoFPP | Only SKEB, PSP, and SMSB used |
| NoSMSB | Only SKEB, PSP, and FPP used |
| Control2.5 | As in Control, but with 2.5-min cycling |
| NoPSP2.5 | As in NoPSP, but with 2.5-min cycling |
| Control15 | As in Control, but with 15-min cycling |
| NoPSP15 | As in NoPSP, but with 15-min cycling |

innovation [Eq. (5) in Aksoy et al. 2009], root mean square innovation [RMSI; Eq. (4.1) in Dowell et al. 2011], total ensemble spread, which includes the observation errors [Eq. (3.3) in Dowell and Wicker 2009], and consistency ratio, which is the total ensemble spread divided by RMSI [Eq. (3.4) in Dowell and Wicker 2009], are computed at each data assimilation time for the prior and posterior model states. For consistency ratio, values near 1 represent an optimally dispersive ensemble. The model variables used in the data assimilation diagnostics include reflectivity, radial velocity, 1.5 m AGL temperature, and 10 m AGL horizontal wind, while the associated observations include PAR reflectivity and radial velocity observations and Oklahoma Mesonet 1.5 m AGL temperature and 10 m AGL horizontal wind observations. For the free forecasts, only RMSI and ensemble spread (i.e., without observation errors) are computed using forecast composite reflectivity, 2 m AGL temperature, 2 m AGL dewpoint temperature, and 10 m AGL horizontal wind components from the model forecasts along with Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) composite reflectivity and Oklahoma Mesonet 1.5 m AGL temperature, 1.5 m AGL dewpoint temperature, and 10 m AGL horizontal wind components.

The WoFS is designed to provide probabilistic guidance of individual storms, so this study will assess the impacts of the stochastic and perturbed parameter methods using two storm-related fields—reflectivity and updraft helicity (UH; Kain et al. 2008). For UH, the probability of 2–5-km UH (UH25) greater than 400 $m^2 s^{-2}$ is subjectively compared to the 2–5-km rotation fields (Miller et al. 2013) from the MRMS system. For UH probabilities, the maximum UH values are aggregated together through the forecast period for each ensemble member to form UH swaths. Next, the neighborhood ensemble maximum probability (Schwartz and Sobash 2017) fields are computed for UH values greater than 400 $m^2 s^{-2}$ using a $3 \times 3$ gridpoint window. To better highlight the differences between the stochastic and Control experiments, the difference in the UH probability fields are shown for the stochastic experiments. For MRMS rotation fields, the maximum rotation values are aggregated together through the forecast periods and then smoothed using a maximum-filter over a $3 \times 3$ gridpoint window.

For reflectivity, forecast composite reflectivity is objectively compared to MRMS composite reflectivity by computing the ensemble fractions skill score (eFSS; Duc et al. 2013):

$$\mathrm{eFSS} = 1 - \frac{\frac{1}{N \times M}\sum_{n=1}^{N}\sum_{m=1}^{M}\left[P^{o}_{n,m} - P^{f}_{n,m}\right]^2}{\frac{1}{N \times M}\left[\sum_{n=1}^{N}\sum_{m=1}^{M}\left(P^{o}_{n,m}\right)^2 + \sum_{n=1}^{N}\sum_{m=1}^{M}\left(P^{f}_{n,m}\right)^2\right]}, \quad (4)$$

where $N$ is the number of neighborhood windows, $M$ is the number of ensemble members, and $P^o$ and $P^f$ are the observed and forecast fractions, respectively, within each spatial window that exceed a threshold. The eFSS is an extension of the fractions skill score from Roberts and Lean (2008), but with the inclusion of the ensemble member dimension. For this study, the eFSS is computed using observed and forecast composite reflectivity values greater than 35 dB$Z$ and a 16-km neighborhood window width. An eFSS value of 1 indicates perfect skill, while a value of 0 indicates no skill. eFSS values less than 1 can be due to a combination of displacement errors and biases associated with the number and size of storms, so to understand the role of the biases on the eFSS values, the average asymptotic eFSS (AeFSS; Roberts and Lean 2008) is also computed for each experiment's ensemble forecast using the equation:

$$\mathrm{AeFSS} = \frac{2F^{o}F^{f}}{\left(F^{o}\right)^2 + \left(F^{f}\right)^2}, \quad (5)$$

where $F^o$ is the observed fraction for the entire domain and $F^f$ is the forecast fraction for the entire domain and ensemble. AeFSS values less than 1 indicate a bias, which can be due to an overprediction or underprediction of reflectivity.

## 4. Results

### a. *Individual method experiments*

#### 1) OBSERVATION-SPACE DIAGNOSTICS

To assess the impact of the stochastic and perturbed parameter methods on ensemble errors and spread, observation-space diagnostics are computed for the background forecasts and analyses during the data assimilation cycling for some of the assimilated observation types. Starting with reflectivity, the FPP method results in analyses with the smallest mean innovations until 2015 UTC and smallest RSMIs through 2100 UTC, while leading to background forecasts with the largest total ensemble spread through 2100 UTC (Figs. 7a–c). The other methods generally result in similar ensemble errors and spread as the Control experiment. The consistency ratios for background forecasts of reflectivity are highly variable between each data assimilation time and are similar among the experiments with values gradually approaching 1 and becoming less underdispersive (Fig. 7d). For radial velocity, all of the error and spread diagnostics are mostly similar among the experiments (Figs. 7e–g). Even so, the stochastic and perturbed parameter method experiments generally have higher consistency ratios than the Control experiment after 2000 UTC, indicating the various methods are contributing to higher spread-to-error ratios (Fig. 7h). For 1.5 m AGL temperature, the SMSB method consistently results in the smallest mean innovations in the background forecasts and analyses through the data assimilation period (Fig. 7i). After 1930 UTC, the SMSB method also largely results in smaller RMSIs (Fig. 7j). The total ensemble spread and consistency ratio are mostly similar across the experiments for 1.5 m AGL temperature (Figs. 7k,l). While not shown, these results are also consistent for 1.5 m AGL dewpoint temperature. For $v$ wind, the SKEB scheme results in the smallest mean innovations and RMSIs in the background forecasts and analyses for most data assimilation times (Figs. 7m,n). The SKEB scheme also results in the largest spread at most times, which in combination with lower RMSIs results in higher consistency ratios

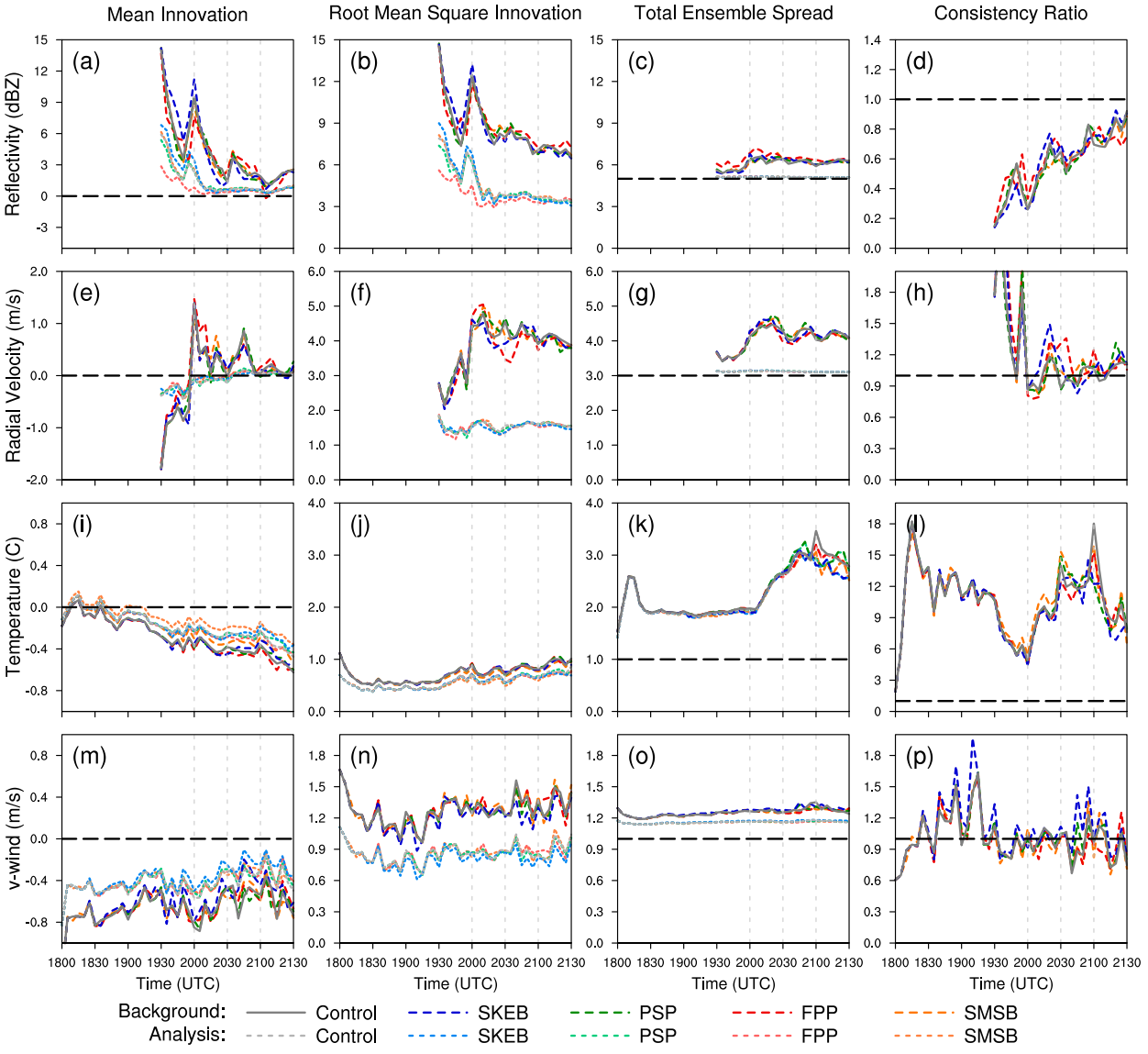


FIG. 7. Mean innovation, RMSI, total ensemble spread, and consistency ratio of (a)–(d) reflectivity (dBZ), (e)–(h) radial velocity (m $s^{-1}$), (i)–(l) 1.5 m AGL temperature (°C), and (m)–(p) 10 m AGL $v$-wind (m $s^{-1}$) for background forecasts (bolder lines) and analyses (fainter lines) from the Control and individual stochastic and perturbed parameter method experiments. Vertical gray dashed lines highlight forecast initialization times. Horizontal black dashed lines in the total ensemble spread plots in (c), (g), (k), and (o) depict the observation errors for each observation type.

than the other experiments for most data assimilation times (Figs. 7o,p).

To assess the impact of the stochastic perturbation methods on ensemble forecast errors and spread, RMSI and ensemble spread are computed for various variables during the four forecast periods. Beginning with composite reflectivity, the SKEB scheme results in the smallest errors for most forecast times in the forecasts initialized at 2000 and 2030 UTC (Figs. 8a,b). The FPP method results in the largest ensemble spread during the earlier times of each forecast, while the SKEB scheme results in the largest ensemble spread during the later forecast times (Figs. 8a–d). Conversely, the SMSB method consistently results in the smallest spread for all forecasts. For near-surface temperature and dewpoint temperature, the SMSB method results in the smallest errors for most forecast times, while generally having similar or slightly smaller ensemble spread than the other experiments (Figs. 8e–l). The differences in RMSI and spread for the 10 m AGL horizontal wind components are mostly similar among the different experiments through all forecast periods (Figs. 8m–t). Except for composite reflectivity before 2100 UTC in the 2000 UTC initialized forecast, all ensemble experiments are underdispersive for all five variables. How these ensemble error and spread results from the observation-space diagnostics translate to forecasts of storm-related fields will be explored next.

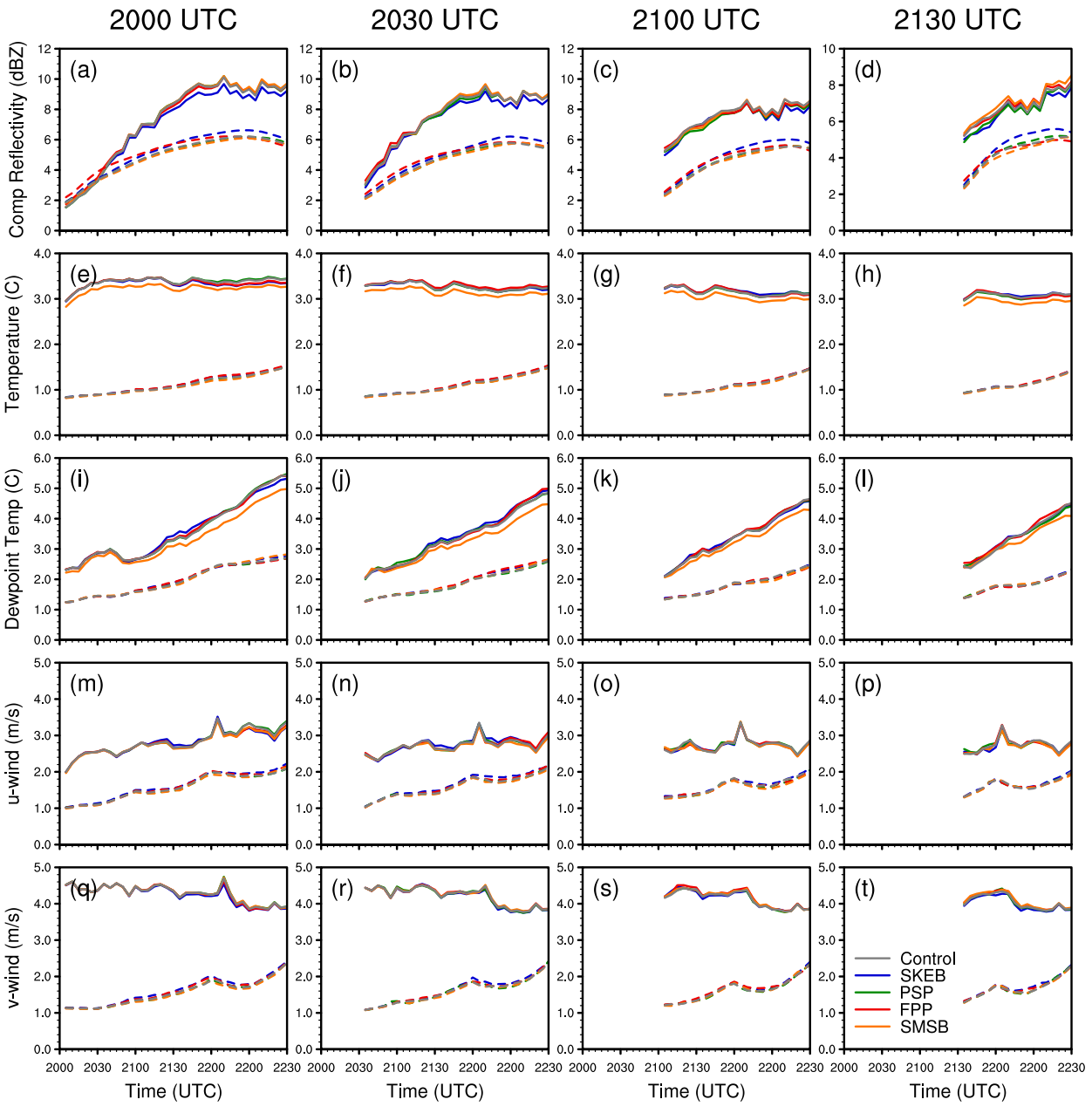


FIG. 8. RMSI (solid lines) and ensemble spread (dashed lines) of (a)–(d) composite reflectivity (dBZ), (e)–(h) near-surface temperature (°C), (i)–(l) near-surface dewpoint temperature (°C), (m)–(p) 10 m AGL $u$ wind (m s$^{-1}$), and (q)–(t) 10 m AGL $v$ wind (m s$^{-1}$) for the Control and individual stochastic and perturbed parameter method forecasts initialized at 2000, 2030, 2100, and 2130 UTC.

### 2) ENSEMBLE FORECAST RESULTS

To assess the impact of the stochastic and perturbed parameter methods on the skill and accuracy of the ensemble forecasts, both subjective and objective evaluation methods are used with the UH25 and composite reflectivity fields, respectively. Starting with UH25, the Control experiment produces probabilities over 40% near the start of S1's rotation track in the ensemble forecast initialized at 2000 UTC, but has lower probabilities for the remainder of the track (Fig. 9a). The SKEB and FPP experiments produce larger probabilities than the Control experiment for most of S1's rotation track (Figs. 9b,d), especially the FPP experiment. The SMSB experiment also has larger probabilities near the start of S1's rotation track, but it mostly has smaller probabilities for the remainder of the track (Fig. 9e). The PSP experiment is the most similar to the Control experiment for the forecasts initialized at 2000 UTC (Fig. 9c). For the forecasts initialized at 2030 UTC, the SKEB, PSP, and FPP experiments produce similarly good forecasts of S1's rotation track as the Control experiment with the PSP experiment producing the highest probabilities

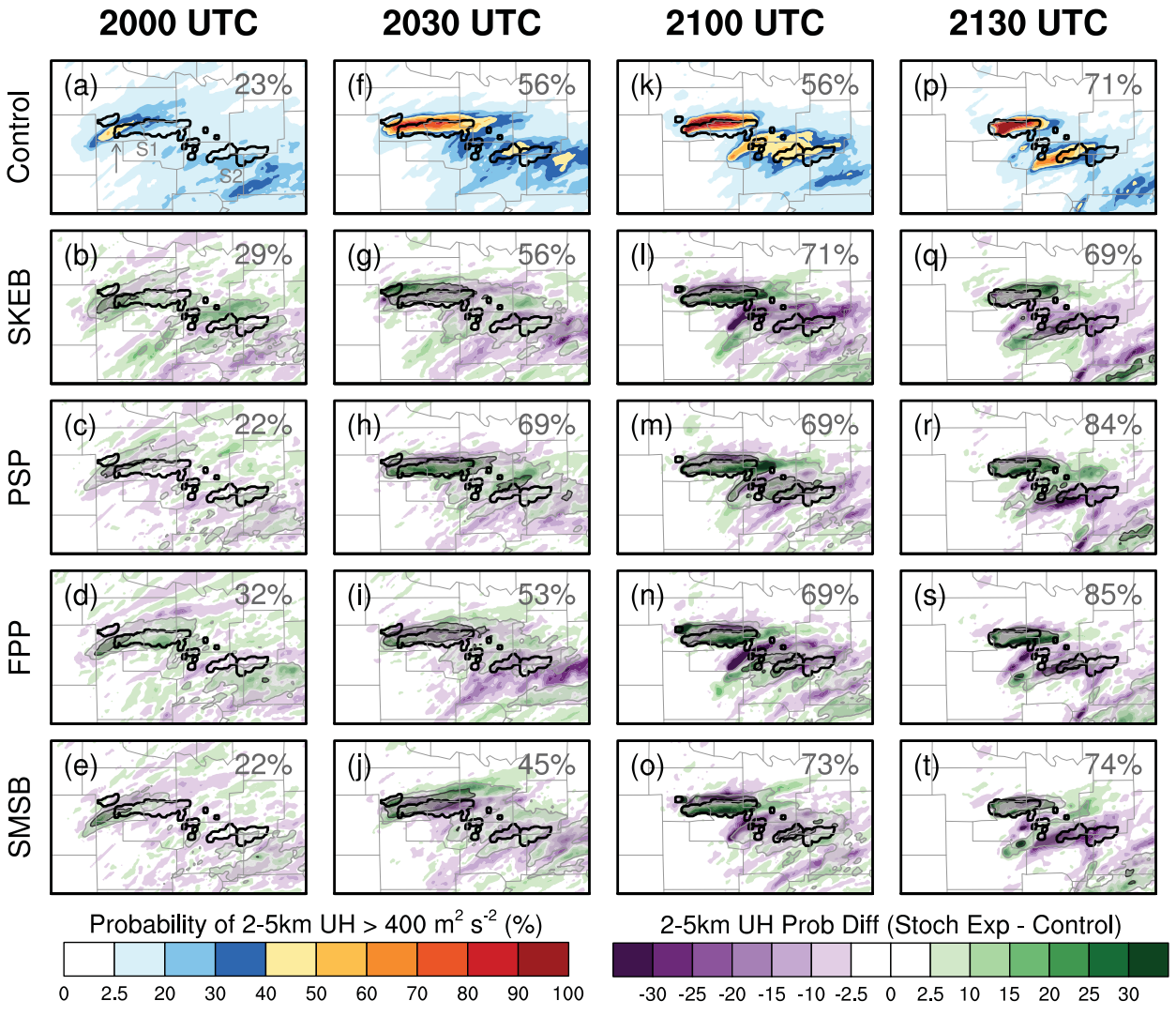


FIG. 9. For the Control experiment, probabilities of 2–5-km UH > 400 $m^2 s^{-2}$ for forecasts with initializations at 2000, 2030, 2100, and 2130 UTC are shown. For the stochastic and perturbed parameter experiments, the difference in probabilities of 2–5-km UH > 400 $m^2 s^{-2}$ between the stochastic experiments and the Control experiment are shown along with light and dark gray transparent shadings and contours for areas with probabilities greater than 20% and 40%, respectively. The nine-point maximum-filtered MRMS 2–5-km rotation field is contoured at 0.008 $s^{-1}$ (black contours). The average UH probability within the bounds of S1's unfiltered rotation track, which is indicated in (a) with the gray arrow, is annotated in the upper-right corner of each plot.

along the track (Figs. 9f–i). The SMSB experiment performs the worst with the swath of higher probabilities deviating to the north toward the end of S1's track (Fig. 9j). While all of the experiments' forecasts initialized at 2100 UTC produce probabilities over 80% along S1's rotation track, the stochastic and perturbed parameter experiments' probability swaths are slightly more accurate than the Control experiment's probability swath, which has a small northward bias (Figs. 9k–o). For the final set of forecasts initialized at 2130 UTC, all experiments produce similarly accurate, high-confidence UH25 probability swaths for S1 (Figs. 9p–t), but for S2's rotation tracks, the stochastic and perturbed parameter experiments, especially the SMSB experiment, generally result in lower probabilities than the Control experiment.

Next, the quantitative skill of the forecasts is evaluated for composite reflectivity greater than 35 dB$Z$ using the eFSS. Since these comparisons involve single forecasts from the different experiments, no significance testing is done, but differences in eFSS greater than 0.05 between the stochastic and perturbed parameter experiments and the Control experiment are used as an indication of a substantial difference in skill. Starting with the forecasts initialized at 2000 UTC, both the SKEB and FPP experiments produce composite reflectivity forecasts that are more skillful than the other experiments, including the Control experiment (Fig. 10a). The SKEB experiment's forecast is substantially and consistently more skillful than the Control experiment's forecast for almost the entire forecast period. For the FPP experiment, the largest

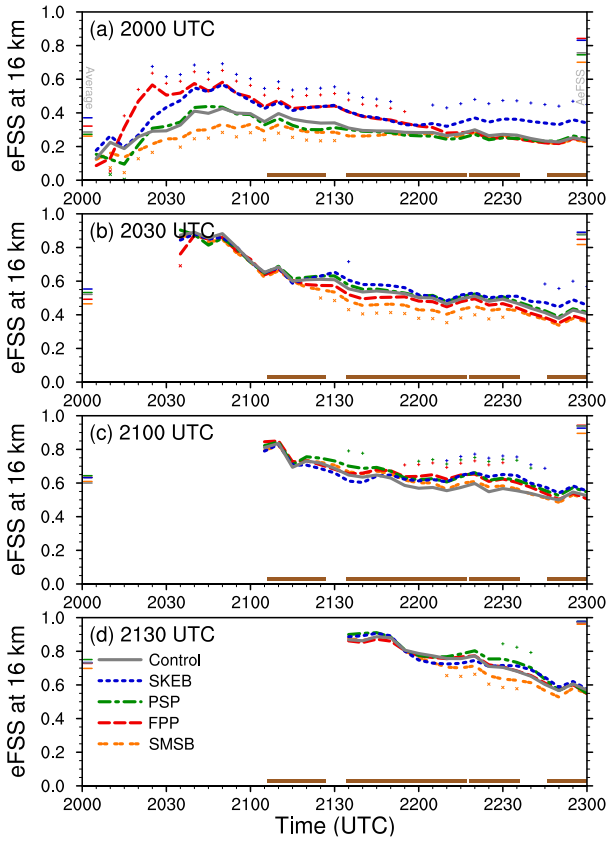


FIG. 10. Time series of eFSS for the Control and individual stochastic and perturbed parameter experiments' composite reflectivity forecasts. The average eFSS and AeFSS for each experiment's forecast is indicated by short lines on the left-hand and right-hand sides, respectively, of each subplot. A + marker indicates when a stochastic or perturbed parameter experiment has an eFSS more than 0.05 higher than the Control experiment, and a × marker indicates when a stochastic or perturbed parameter experiment has an eFSS more than 0.05 lower than the Control experiment. The brown lines represent the time periods of S1's T1 and T2 and S2's T3 and T4.

difference in eFSS from the Control experiment occurs within the first hour of the forecast, which is when S1 is developing and maturing. Similar to its spread in composite reflectivity, the FPP experiment's skill gradually decreases and becomes more similar to the Control's experiment skill after 2200 UTC. As alluded to by the UH25 results, the PSP experiment's forecast initialized at 2000 UTC is similarly skillful as the Control experiment's forecast, while the SMSB experiment's forecast is less skillful than the Control experiment's forecast through 2145 UTC. For the last hour of the forecast, all of the experiments are similarly skillful except for the SKEB experiment, which is likely more skillful due to producing a better forecast for S2 as is evident in the UH25 probability forecast (Fig. 9b). All experiments' forecasts have AeFSS values less than 1 indicating bias (Fig. 10a). The bias in these forecasts—and all other experiments' forecasts in this study—are due to an underforecasting bias (not shown), so the SKEB and FPP experiments have the smallest underforecasting bias in this forecast.

Differences in forecast skill among the experiments are smaller for the other three forecasts than for the first forecast. For the forecasts initialized at 2030 UTC, differences among the experiments do not develop until after 2115 UTC (Fig. 10b). After this time, the SKEB experiment's forecast is substantially more skillful than the Control experiment's forecast for only a few forecast times, while the SMSB's experiment's forecast is again the least skillful likely due to the deviation in forecast storm track as shown in the UH25 results (Fig. 9j). The FPP experiment's forecast is slightly less skillful than the Control experiment's forecast for several forecast times. The PSP experiment's forecast is again similarly skillful to the Control experiment's forecast. For the forecasts initialized at 2100 UTC, the stochastic and perturbed parameter experiments' forecasts are generally more skillful than the Control experiment's forecast, especially from 2200 to 2230 UTC for the SKEB, PSP, and FPP experiments (Fig. 10c). This improvement in skill is at least partly due to the slight northward displacement error of S1 as evident by the Control experiment's UH25 forecast (Fig. 9k). For the forecasts initialized at 2130 UTC, the experiments' forecasts are similarly skillful, except for the SMSB experiment, which is substantially less skillful than the Control experiment for several forecast times, and the PSP experiment, which is substantially more skillful for a few forecast times (Fig. 10d). The SMSB experiment's lower skill is at least partially due to having a worse forecast for S2. Overall, the biggest impacts of the stochastic and perturbed parameter methods occur in the forecasts initialized at 2000 UTC with the SKEB scheme and FPP method providing the most benefit, especially early on when S1 is developing in the forecasts.

### b. *Combination experiments*

Ideally, multiple stochastic and perturbed parameter methods would be used in a WoFS-like system to account for different sources of model uncertainty, so this next set of experiments explores *some* of the possible stochastic and perturbed parameter method combinations. For forecasts initialized at 2000 UTC, the various combination experiments' UH25 probabilities for S1 are generally similar to the Control experiment's probabilities (Figs. 11a–f). Even so, the NoPSP and NoSMSB experiments, which have the SKEB scheme and FPP method in common, mostly have higher probabilities along S1's rotation track than the Control experiment. The axis of higher probabilities for the forecasts initialized at 2030 UTC are generally too far north for all experiments (Figs. 11g–l), especially the AllStoch experiment likely due to the ensemble analysis and subsequent forecast latching onto S1's initial mesocyclone, which is depicted as the short rotation track to the northwest of S1's main rotation track (Fig. 11h). While the Control experiment's forecast initialized at 2100 UTC continues to have a northward bias, all of the combination experiments' UH25 probability swaths are more accurate resulting in higher probabilities along S1's rotation track (Figs. 11m–r). Also, all of the combination experiments' forecasts produce higher UH25 probabilities for S1's initial mesocyclone than the Control experiment's forecast indicating their analyses of S1 are better. In the forecasts initialized at 2130 UTC, all of the combination

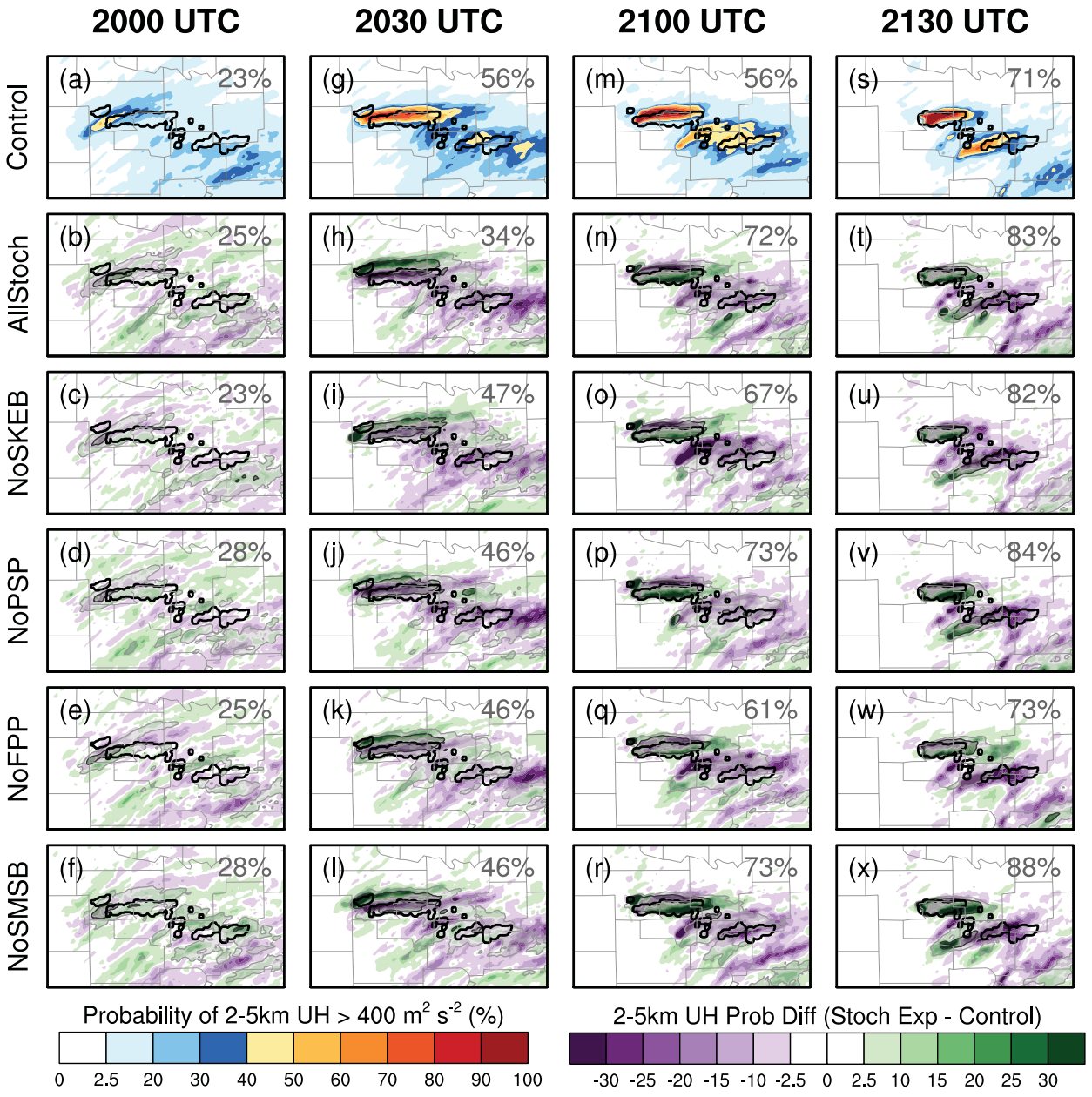


FIG. 11. As in Fig. 9, but for the combination experiments.

experiments predict S1's rotation track similarly well, while the Control experiment's forecast still exhibits a small northward bias (Figs. 11s–x). The largest differences among the experiments for the final forecast stem from their predictions of S2's rotation tracks with the Control experiment having the highest probabilities overlapping those areas.

The combination experiments' composite reflectivity forecasts initialized at 2000 UTC are substantially more skillful than the Control experiment's forecast for most forecast times leading up to and during T1 and the start of T2, except for the NoFPP experiment (Fig. 12a). This result highlights the beneficial impact the FPP method provides during S1's development. Later in the forecast period, the NoFPP and NoSMSB experiments are substantially more skillful than the Control experiment for at least few forecast times. For the forecasts initialized at 2030 UTC, the combination experiments are similarly skillful for almost all forecast times with average eFSS differences less than 0.08 (Fig. 12b). Even so, the AllStoch and NoSKEB experiments are substantially less skillful for several forecast times. The AllStoch, NoPSP, and NoSMSB experiments produce similarly skillful forecasts for the forecasts initialized at 2100 UTC and are substantially more skillful than the Control experiment for several forecast times (Fig. 12c). For the forecasts initialized at 2130 UTC, the differences in skill between the combination experiments and the Control experiment

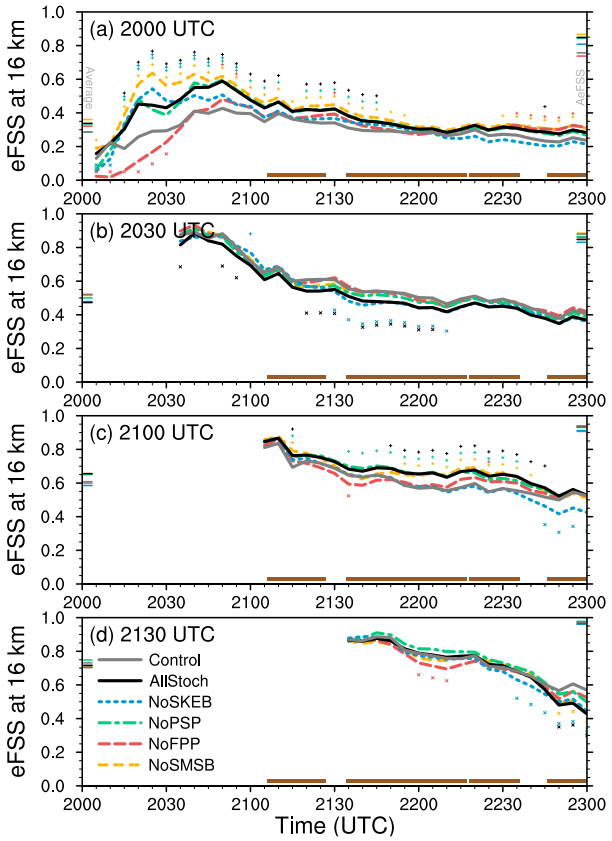


FIG. 12. As in Fig. 10, but for the combination experiments.

are again generally small (Fig. 12d). Except for the NoPSP experiment, the other combination experiments have at least a couple of forecast times with substantially less skill. Overall, for both the UH25 and composite reflectivity results, the combination experiments show that the SKEB scheme and FPP method provide the most benefit when combined, especially during the first forecast when S1 is developing and maturing.

### c. *Cycling interval experiments*

Since PAR technology allows for more frequent full-volumetric scans of the atmosphere than current operational radar technology, the ability to assimilate more frequent radar observations is possible for prediction systems like WoFS. This next set of experiments explores the impact of stochastic and perturbed parameter methods at three different DA cycling intervals, including the two previously shown 5-min DA cycling frequency experiments, Control and NoPSP. For the forecasts initialized at 2000 UTC, the NoPSP2.5 experiment's forecast mostly has higher probabilities in S1's rotation track than the Control2.5 experiment (Figs. 13a,b). This result is similar to the previous 5-min cycling experiments' results shown here again (Figs. 13c,d). Both of the 15-min cycling experiments perform similarly poorly with probabilities less than 10% within S1's rotation track (Figs. 13e,f). Unlike the 5-min cycling experiments where both experiments have the highest probabilities over similar areas (Figs. 13i,j), the NoPSP2.5 experiment's higher-probability swath is more accurate than the Control2.5 experiment's swath for the forecasts initialized at 2030 UTC (Figs. 13g,h). Both of the 15-min cycling experiments' UH25 probability swaths originate from S1's initial mesocyclone location and are thus too far north (Figs. 13k,l). Also, the NoPSP15 experiment's probabilities are smaller in S1's rotation track than the Control15 experiment's probabilities, as with the two 5-min cycling experiments.

For the forecasts initialized at 2100 UTC, the Control2.5 and NoPSP2.5 experiments produce similar forecasts (Figs. 13m,n), while larger differences exist for the two 5-min cycling experiments, where the Control experiment's probability swath is too far north, as previously mentioned (Figs. 13o,p). Both of the 15-min cycling experiments' UH25 probability swaths are similar and still too far north (Figs. 13q,r). The probability swaths are similarly accurate for the 2.5-min cycling experiments' forecasts initialized at 2130 UTC (Figs. 13s,t), but NoPSP2.5 undesirably has lower probabilities for the eastern half of S1's rotation track. Both experiments have poor forecasts for S2's rotation tracks with no probabilities over 40%. The 5- and 15-min cycling stochasticity experiments have higher probabilities than their no-stochasticity counterparts for S1's rotation track while the opposite is true for S2's rotation tracks (Figs. 13u–x).

For eFSS, the NoPSP2.5 and NoPSP experiments are similarly skillful than their respective counterparts, Control2.5 and Control, for the forecasts initialized at 2000 UTC (Fig. 14a). Both Control15 and NoPSP15 perform poorly, but their relatively low skill is mostly due to inaccurate forecast storm locations rather than a bias in spatial coverage since their AeFSS values are more similar to the NoPSP2.5 and NoPSP experiments than the Control2.5 and Control experiments. For the forecasts initialized at 2030 UTC, the 2.5- and 5-min cycling experiments all have similar skill for the entire forecast period, while the NoPSP experiment is substantially less skillful than the Control15 experiment for several forecast times (Fig. 14b). Both 15-min cycling experiments again have lower skill than the 2.5- and 5-min cycling experiments. While the NoPSP experiment is substantially more skillful than the Control experiment for several times throughout the forecasts initialized at 2100 UTC, the NoPSP2.5 experiment is substantially more skillful during the first hour of the forecast period and substantially less skillful during the last 45 min of the forecast period (Fig. 14c). The 15-min cycling experiments are more similar in skill to the other cycling experiments for this forecast, but still overall less skillful, while NoPSP15 is still mostly less skillful than Control15. For the final forecast initialized at 2130 UTC, the 2.5- and 5-min DA cycling experiments perform more similar to their cycling frequency counterparts than their stochastic counterparts (Fig. 14d). The 15-min cycling experiments continue to be more similar to each other, but the NoPSP15 experiment is now somewhat more skillful than the Control15 experiment during the first hour of the forecast period.

In Stratman et al. (2020), the more frequent cycling experiments were more skillful in forecasts initialized at earlier times. In this study, that holds true when comparing the 5- and 15-min cycling experiments, but not for the 2.5- and

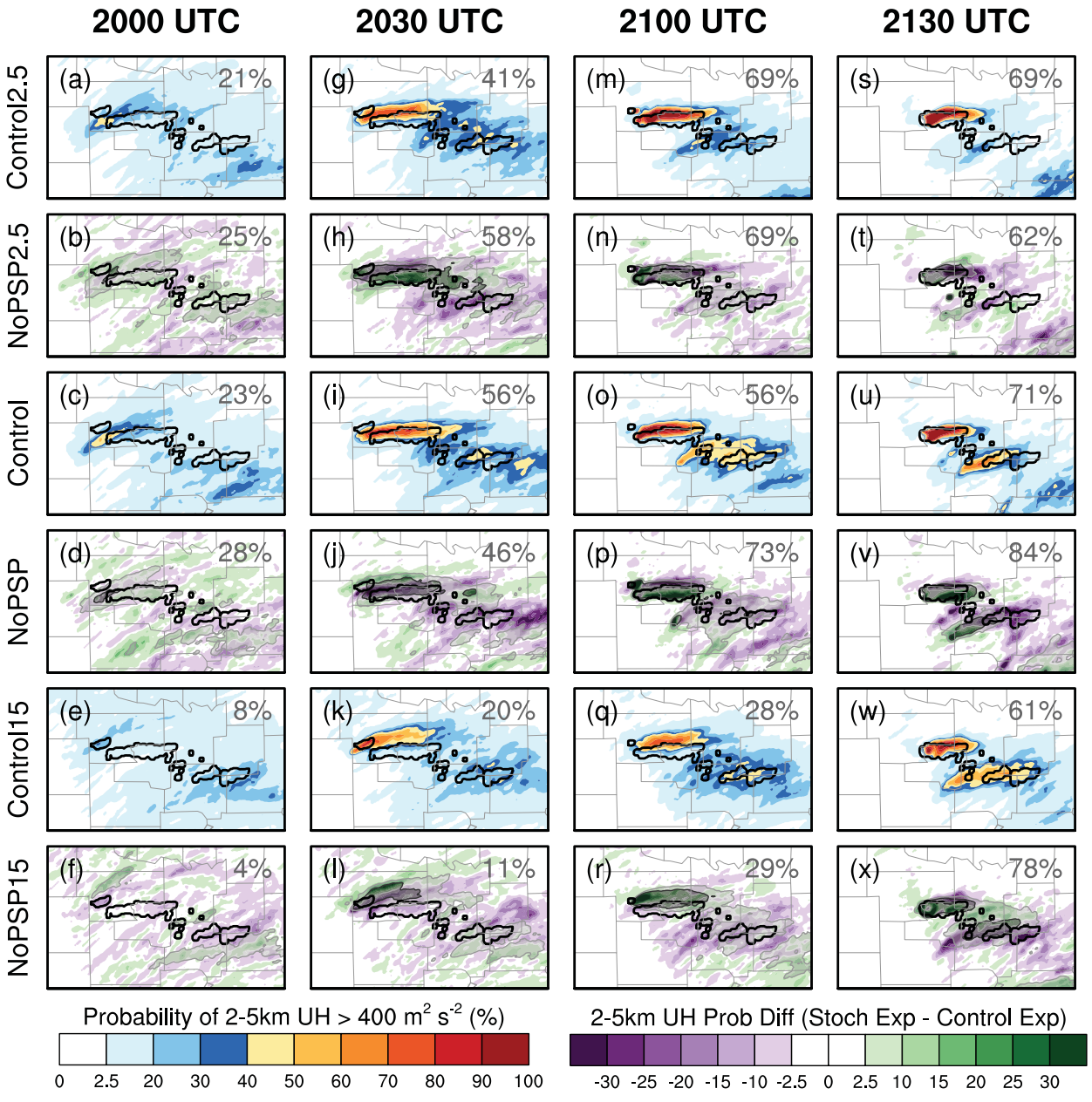


FIG. 13. As in Fig. 9, but for the cycling frequency experiments.

5-min cycling experiments. This result is perhaps due to not having forecasts initialized at earlier times than 2000 UTC. Another result that agrees with the findings in Stratman et al. (2020) is that skill eventually starts to decrease at later initialized forecasts for the shorter cycling intervals. Ideally, the stochastic and perturbed parameter methods would help combat this decrease in skill by generating more ensemble spread and helping prevent filter divergence, but that is not the result here likely due to the model imbalances that develop from cycling too frequently (Yang and Wang 2023). Even so, the beneficial impacts of the stochastic and perturbed parameter methods generally occur in earlier initialized forecasts for shorter cycling intervals.

## 5. Summary and discussion

In this study, four stochastic and perturbed parameter methods, including a new method, were tested within a 1-km-scale version of the WoFS as a way to increase ensemble spread during frequent PAR data assimilation cycling. PAR technology allows for more frequent full-volumetric scans of the atmosphere than the current dish-based operational weather radars and is thus a candidate to replace those systems. The more frequent radar observations could benefit storm-scale forecast systems like the experimental WoFS by allowing for more frequent updates of the model analyses. Using the temporally and spatially dense PAR observations from the 9 May 2016 Oklahoma tornado outbreak, various

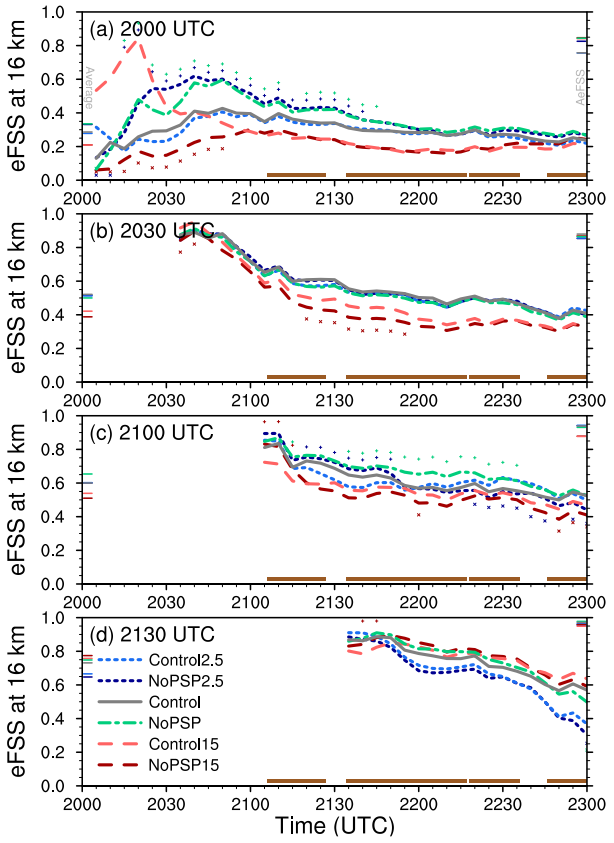


FIG. 14. As in Fig. 10, but for the cycling frequency experiments.

experiments were conducted to explore the impacts of stochastic and perturbed parameter methods individually, in different combinations, and with different cycling intervals. While the results are preliminary since only a single case was used for the experiments, some potentially important results for future versions of WoFS emerged from these initial experiments.

Perhaps the primary finding from this initial study is that stochastic and perturbed parameter methods can improve WoFS forecasts of storms when storms are developing and maturing, which means more skillful forecasts of storms at longer lead times. The beneficial impact of the stochastic and perturbed parameter methods in a multi-physics setup generally decreased in later initialized forecasts when the storms were more mature and not quickly developing or evolving. Even so, some benefit remained in later initialized forecasts as storms evolved or new storms developed. In the individual method experiments, the SKEB and FPP methods performed the best in the first initialized forecast when the first tornadic supercell was developing (Fig. 15). The PSP scheme had the smallest impact on the first forecast, but in later forecasts, the PSP experiment generally performed better than the Control experiment. While the novel SMSB method resulted in smaller near-surface temperature and dewpoint temperature errors, the SMSB experiment generally performed similar or worse than the Control experiment.

Another important result from the experiments is that combining multiple stochastic and perturbed parameter methods can sometimes produce better, more skillful forecasts than any individual method. This result mostly agrees with previous studies that found that adding uncertainty to the model physics in multiple ways can result in more skillful forecasts than any single method (e.g., Jankov et al. 2017). The two best performing combination experiments were NoPSP and NoSMSB (Fig. 15), which further highlights the success of the SKEB scheme and FPP method in this study. The cycling interval experiments revealed the stochastic and perturbed parameter methods can be beneficial at all cycling intervals, but the beneficial impact was evident in earlier forecasts for shorter cycling intervals (Fig. 15). In the

Average UH Probabilities

| 2000 | 2030 | 2100 | 2130 | | 2000 | 2030 | 2100 | 2130 |
|---|---|---|---|---|---|---|---|---|
| 23.4 | 56.2 | 55.7 | 71.2 | Control | 0.29 | 0.52 | 0.60 | 0.73 |
| 29.4 | 56.5 | 70.5 | 69.4 | SKEB | 0.37 | 0.55 | 0.63 | 0.73 |
| 21.9 | 68.6 | 69.1 | 83.8 | PSP | 0.27 | 0.53 | 0.64 | 0.75 |
| 32.0 | 53.4 | 68.8 | 85.1 | FPP | 0.32 | 0.49 | 0.63 | 0.73 |
| 21.6 | 45.4 | 73.0 | 74.2 | SMSB | 0.26 | 0.47 | 0.61 | 0.70 |
| 25.3 | 34.4 | 72.3 | 83.3 | AllStoch | 0.34 | 0.47 | 0.66 | 0.72 |
| 23.2 | 46.9 | 67.2 | 81.9 | NoSKEB | 0.29 | 0.48 | 0.59 | 0.70 |
| 28.1 | 46.0 | 73.5 | 84.1 | NoPSP | 0.33 | 0.50 | 0.65 | 0.75 |
| 24.6 | 46.0 | 61.4 | 73.2 | NoFPP | 0.32 | 0.52 | 0.61 | 0.70 |
| 28.2 | 46.1 | 72.5 | 87.9 | NoSMSB | 0.36 | 0.52 | 0.65 | 0.71 |
| 20.5 | 40.6 | 69.4 | 68.7 | Control2.5 | 0.28 | 0.51 | 0.60 | 0.67 |
| 24.9 | 57.9 | 69.5 | 62.1 | NoPSP2.5 | 0.33 | 0.52 | 0.60 | 0.65 |
| 8.3 | 19.6 | 28.5 | 60.8 | Control15 | 0.21 | 0.42 | 0.54 | 0.76 |
| 4.2 | 10.5 | 29.0 | 77.6 | NoPSP15 | 0.21 | 0.39 | 0.51 | 0.77 |

Average eFSS

FIG. 15. Summary of results showing (left) average UH25 probabilities in S1's rotation track and (right) average eFSS for each experiment and forecast. Color shading highlights where experiments have higher probabilities or more skill (green) and lower probabilities or less skill (purple).

later initialized forecasts, the detrimental effects of model imbalances along with possible filter divergence caused by cycling too frequently were evident, indicating the stochastic and perturbed parameter methods might not be enough to overcome these issues.

As a next step of this work, the stochastic and perturbed parameter methods will be tested and further tuned in the experimental WoFS-1km (Kerr et al. 2023; Wang et al. 2022) using multiple severe weather events from different geographical locations to produce more robust conclusions. In the current experimental WoFS, additive noise (Dowell and Wicker 2009; Sobash and Wicker 2015) is used to help spin up storms in the analyses and forecasts by generating spread in areas of convection, but additive noise was withheld in these experiments to focus on the impact of the stochastic and perturbed parameter methods. Therefore, future experiments with the WoFS-1km will compare the stochastic and perturbed parameter methods with different additive noise techniques.

While four stochastic and perturbed parameter methods were implemented and tested in this study, other methods exist and may be beneficial to prediction systems like the WoFS. For example, the stochastic perturbation of physics tendencies (SPPT) scheme (Buizza et al. 1999; Palmer et al. 2009), which perturbs the physics tendencies using multiplicative noise, is another potential way to represent uncertainty in physics parameterization schemes. Also, while the FPP method performed well in this study, another way to account for uncertainty in the microphysics scheme is to use a stochastically perturbed parameter (SPP) scheme (e.g., Ollinaho et al. 2017; Jankov et al. 2019; Lang et al. 2021; Thompson et al. 2021; McTaggart-Cowan et al. 2022), so a SPP scheme will likely be implemented into the NSSL two-moment microphysics scheme. The SMSB method had less benefit than expected, so additional work is needed to determine why it resulted in less skillful forecasts and if this outcome is specific to this case and geographical location or if it is a more general issue. If it is a recurring issue after many additional forecasts of severe thunderstorm events are evaluated, then work will need to be done to explore how to improve it—possibly through stochastically blending other variables in the Noah-MP scheme or stochastically blending only some of the variable options. Overall, though, the experiments in this study highlight the potential benefits of the stochastic and perturbed parameter methods for the next-generation WoFS, which will likely feature 1-km or less horizontal grid spacing and even more frequent radar DA than the current experimental WoFS permits.

*Acknowledgments.* Funding for this work was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA21OAR4320204, U.S. Department of Commerce and NOAA/Office of Oceanic and Atmospheric Research/Weather Portfolio Tornado Warning Improvement and Extension Program (TWIEP). The authors thank Anthony Reinhart for helping to process and quality control the NWRT PAR observations with WDSS-II; Edward Mansell for his latest NSSL two-moment microphysics code, helping determine which parameters to perturb, and for internally reviewing the manuscript prior to submission; Hailey Shin for her help with computing the variances in the Shin–Hong scale-aware PBL scheme for the PSP scheme; and Nicholas Gasperoni for his help with the SKEB scheme. The authors also thank Altug Aksoy and two other anonymous reviewers for their valuable reviews and Ron McTaggart-Cowan for his helpful suggestions that greatly improved this paper. Crucial local computing assistance was provided by Kent Knopfmeier and Brett Morrow.

*Data availability statement.* The model output and code used in this study are not currently available in a publicly accessible repository. The model output and code used to generate the results herein are available from the authors upon request.

## REFERENCES

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, https://doi.org/10.1175/2008MWR2691.1.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, https://doi.org/10.1111/j.1600-0870.2008.00361.x.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, https://doi.org/10.1175/2008JAS2677.1.

——, S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, https://doi.org/10.1175/2010MWR3595.1.

——, T. Jung, and T. N. Palmer, 2012: Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate*, **25**, 4946–4962, https://doi.org/10.1175/JCLI-D-11-00297.1.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

Burke, P. C., and Coauthors, 2022: Collaborating to increase warning lead time using the Warn-on-Forecast System. *30th Conf. on Severe Local Storms*, Santa Fe, NM, Amer. Meteor. Soc., 44, https://ams.confex.com/ams/30SLS/meetingapp.cgi/Paper/407184.

Candille, G., 2009: The multiensemble approach: The NAEFS example. *Mon. Wea. Rev.*, **137**, 1655–1665, https://doi.org/10.1175/2008MWR2682.1.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian Ensemble Prediction System. *Mon. Wea. Rev.*, **138**, 1877–1901, https://doi.org/10.1175/2009MWR3187.1.

Christensen, H. M., I. M. Moroz, and T. N. Palmer, 2015: Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *J. Atmos. Sci.*, **72**, 2525–2544, https://doi.org/10.1175/JAS-D-14-0250.1.

Deutsch, J. L., and C. V. Deutsch, 2012: Latin hypercube sampling with multidimensional uniformity. *J. Stat. Plann. Inference*, **142**, 763–772, https://doi.org/10.1016/j.jspi.2011.09.016.

Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927, https://doi.org/10.1175/2008JTECHA1156.1.

——, ——, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Wea. Rev.*, **139**, 272–294, https://doi.org/10.1175/2010MWR3438.1.

——, and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1.

Duc, L., K. Saito, and H. Seko, 2013: Spatial–temporal fractions verification for high-resolution ensemble forecasts. *Tellus*, **65A**, 18171, https://doi.org/10.3402/tellusa.v65i0.18171.

Duda, J. D., X. Wang, F. Kong, M. Xue, and J. Berner, 2016: Impact of a stochastic kinetic energy backscatter scheme on warm season convection-allowing ensemble forecasts. *Mon. Wea. Rev.*, **144**, 1887–1908, https://doi.org/10.1175/MWR-D-15-0092.1.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Forsyth, D. E., and Coauthors, 2005: The national weather radar testbed (phased array). *32nd Conf. on Radar Meteorology*, Albuquerque, NM, Amer. Meteor. Soc., 12R.3, https://ams.confex.com/ams/32Rad11Meso/techprogram/paper_96377.htm.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainty. *Mon. Wea. Rev.*, **135**, 1846–1868, https://doi.org/10.1175/MWR3391.1.

Gallo, B. T., and Coauthors, 2022: Exploring the watch-to-warning space: Experimental outlook performance during the 2019 spring forecasting experiment in NOAA's Hazardous Weather Testbed. *Wea. Forecasting*, **37**, 617–637, https://doi.org/10.1175/WAF-D-21-0171.1.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

Gasperoni, N. A., X. Wang, and Y. Wang, 2020: A comparison of methods to sample model errors for convection-allowing ensemble forecasts in the setting of multiscale initial conditions produced by the GSI-based EnVar assimilation system. *Mon. Wea. Rev.*, **148**, 1177–1203, https://doi.org/10.1175/MWR-D-19-0124.1.

Hacker, J. P., C. Snyder, S.-Y. Ha, and M. Pocernich, 2011: Linear and nonlinear response to parameter variations in a mesoscale model. *Tellus*, **63A**, 429–444, https://doi.org/10.1111/j.1600-0870.2010.00505.x.

He, C., and Coauthors, 2023: The community Noah-MP land surface modeling system technical description version 5.0. NCAR Tech. Note NCAR/TN-575+STR, 284 pp., https://doi.org/10.5065/ew8g-yr95.

Hirt, M., S. Rasp, U. Blahak, and G. C. Craig, 2019: Stochastic parameterization of processes leading to convective initiation in kilometer-scale models. *Mon. Wea. Rev.*, **147**, 3917–3934, https://doi.org/10.1175/MWR-D-19-0060.1.

Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289, https://doi.org/10.1256/qj.05.135.

Hu, M., H. Shao, D. Stark, K. Newman, C. Zhou, and X. Zhang, 2016: Gridpoint Statistical Interpolation (GSI) user's guide version 3.5. Developmental Testbed Center Rep., 148 pp., https://dtcenter.org/com-GSI/users/docs/users_guide/GSIUserGuide_v3.5.pdf.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, https://doi.org/10.1175/WAF-D-21-0130.1.

Janjić, Z. I., 2001: Nonsingular implementation of the Mellor–Yamada Level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp., http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf.

Jankov, I., and Coauthors, 2017: A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Mon. Wea. Rev.*, **145**, 1161–1179, https://doi.org/10.1175/MWR-D-16-0160.1.

——, J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, https://doi.org/10.1175/MWR-D-18-0092.1.

Johnson, A., and X. Wang, 2016: A study of multiscale initial condition perturbation methods for convection-permitting ensemble forecasts. *Mon. Wea. Rev.*, **144**, 2579–2604, https://doi.org/10.1175/MWR-D-16-0056.1.

Jones, T. A., D. Stensrud, L. Wicker, P. Minnis, and R. Palikonda, 2015: Simultaneous radar and satellite data storm-scale assimilation using an ensemble Kalman filter approach for 24 May 2011. *Mon. Wea. Rev.*, **143**, 165–194, https://doi.org/10.1175/MWR-D-14-00180.1.

——, and Coauthors, 2020: Assimilation of *GOES-16* radiances and retrievals into the Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 1829–1859, https://doi.org/10.1175/MWR-D-19-0379.1.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, https://doi.org/10.1175/WAF2007106.1.

Kalina, E. A., I. Jankov, T. Alcott, J. Olson, J. Beck, J. Berner, D. Dowell, and C. Alexander, 2021: A progress report on the development of the High-Resolution Rapid Refresh ensemble. *Wea. Forecasting*, **36**, 791–804, https://doi.org/10.1175/WAF-D-20-0098.1.

Kerr, C. A., B. C. Matilla, Y. Wang, D. R. Stratman, T. A. Jones, and N. Yussouf, 2023: Results from a pseudo-real-time next-generation 1-km Warn-on-Forecast System prototype. *Wea. Forecasting*, **38**, 307–319, https://doi.org/10.1175/WAF-D-22-0080.1.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP global data assimilation system. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Kober, K., and G. C. Craig, 2016: Physically based Stochastic Perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *J. Atmos. Sci.*, **73**, 2893–2911, https://doi.org/10.1175/JAS-D-15-0144.1.

Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The Warning Decision Support System—Integrated information. *Wea. Forecasting*, **22**, 596–612, https://doi.org/10.1175/WAF1009.1.

Lang, S., E. Hólm, M. Bonavita, and Y. Tremolet, 2019: A 50-member ensemble of data assimilations. *ECMWF Newsletter*, No. 158, ECMWF, Reading, United Kingdom, 27–29, https://doi.org/10.21957/nb251xc4sl.

——, S.-J. Lock, M. Leutbecher, P. Bechtold, and R. M. Forbes, 2021: Revision of the stochastically perturbed parametrisations model uncertainty scheme in the integrated forecasting system. *Quart. J. Roy. Meteor. Soc.*, **147**, 1364–1381, https://doi.org/10.1002/qj.3978.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, https://doi.org/10.1016/j.jcp.2007.02.014.

——, and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, https://doi.org/10.1002/qj.3094.

Mansell, E. R., and C. L. Ziegler, 2013: Aerosol effects on simulated storm electrification and precipitation in a two-moment bulk microphysics model. *J. Atmos. Sci.*, **70**, 2032–2050, https://doi.org/10.1175/JAS-D-12-0264.1.

——, ——, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, https://doi.org/10.1175/2009JAS2965.1.

McTaggart-Cowan, R., and Coauthors, 2022: Using stochastically perturbed parameterizations to represent model uncertainty. Part I: Implementation and parameter sensitivity. *Mon. Wea. Rev.*, **150**, 2829–2858, https://doi.org/10.1175/MWR-D-21-0315.1.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875, https://doi.org/10.1029/RG020i004p00851.

Meng, Z., and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Wea. Rev.*, **135**, 1403–1423, https://doi.org/10.1175/MWR3352.1.

Miller, M. L., V. Lakshmanan, and T. M. Smith, 2013: An automated method for depicting mesocyclone paths and intensities. *Wea. Forecasting*, **28**, 570–585, https://doi.org/10.1175/WAF-D-12-00065.1.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772, https://doi.org/10.1038/nature02771.

Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, https://doi.org/10.2151/jmsj.87.895.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multi-parameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

NOAA/NWS, 2016: The severe weather and tornado outbreak of May 9, 2016. NOAA/NWS, accessed 20 January 2023, https://www.weather.gov/oun/events-20160509.

Ollinaho, P., and Coauthors, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quart. J. Roy. Meteor. Soc.*, **143**, 408–422, https://doi.org/10.1002/qj.2931.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp., https://www.ecmwf.int/en/elibrary/75936-stochastic-parametrization-and-model-uncertainty.

Rasp, S., T. Selz, and G. C. Craig, 2018: Variability and clustering of midlatitude summertime convection: Testing the Craig and Cohen theory in a convection-permitting ensemble with stochastic boundary layer perturbations. *J. Atmos. Sci.*, **75**, 691–706, https://doi.org/10.1175/JAS-D-17-0258.1.

Roberts, B., I. L. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Postprocessing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, https://doi.org/10.1175/BAMS-D-18-0041.1.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

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

Schwartz, C. S., and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, https://doi.org/10.1175/MWR-D-16-0400.1.

——, M. Wong, G. S. Romine, R. A. Sobash, and K. R. Fossell, 2020: Initial conditions for convection-allowing ensembles over the conterminous United States. *Mon. Wea. Rev.*, **148**, 2645–2669, https://doi.org/10.1175/MWR-D-19-0401.1.

——, J. Poterjoy, G. S. Romine, D. C. Dowell, J. R. Carley, and J. Bresch, 2022: Short-term convection-allowing ensemble precipitation forecast sensitivity to resolution of initial condition perturbations and central initial states. *Wea. Forecasting*, **37**, 1259–1286, https://doi.org/10.1175/WAF-D-21-0165.1.

Shin, H. H., and S.-Y. Hong, 2015: Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. *Mon. Wea. Rev.*, **143**, 250–271, https://doi.org/10.1175/MWR-D-14-00116.1.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, https://doi.org/10.1256/qj.04.106.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast System. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Sobash, R. A., and L. J. Wicker, 2015: On the impact of additive noise in storm-scale EnKF experiments. *Mon. Wea. Rev.*, **143**, 3067–3086, https://doi.org/10.1175/MWR-D-14-00323.1.

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.

——, and Coauthors, 2009: Convective-scale Warn-on-Forecast System: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Stratman, D. R., N. Yussouf, Y. Jung, T. A. Supinie, M. Xue, P. S. Skinner, and B. J. Putnam, 2020: Optimal temporal frequency of NSSL phased-array radar observations for an experimental Warn-on-Forecast System. *Wea. Forecasting*, **35**, 193–214, https://doi.org/10.1175/WAF-D-19-0165.1.

Thompson, G., J. Berner, M. Frediani, J. A. Otkin, and S. M. Griffin, 2021: A stochastic parameter perturbation method to represent uncertainty in a microphysics scheme. *Mon. Wea. Rev.*, **149**, 1481–1497, https://doi.org/10.1175/MWR-D-20-0077.1.

Vié, B., O. Nuissier, and V. Ducrocq, 2011: Cloud-resolving ensemble simulations of Mediterranean heavy precipitation events: Uncertainty on initial conditions and lateral boundary conditions. *Mon. Wea. Rev.*, **139**, 403–423, https://doi.org/10.1175/2010MWR3487.1.

Wang, Y., N. Yussouf, C. A. Kerr, D. R. Stratman, and B. C. Matilla, 2022: An experimental 1-km Warn-on-Forecast System for hazardous weather events. *Mon. Wea. Rev.*, **150**, 3081–3102, https://doi.org/10.1175/MWR-D-22-0094.1.

Weber, M., J. Y. N. Cho, J. S. Herd, J. M. Flavin, W. E. Benner, and G. S. Torok, 2007: The next-generation multimission U.S. surveillance radar network. *Bull. Amer. Meteor. Soc.*, **88**, 1739–1752, https://doi.org/10.1175/BAMS-88-11-1739.

——, and Coauthors, 2021: Toward the next generation operational meteorological radar. *Bull. Amer. Meteor. Soc.*, **102**, E1357–E1383, https://doi.org/10.1175/BAMS-D-20-0067.1.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast System. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

——, and ——, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, https://doi.org/10.1175/MWR-D-11-00276.1.

Yang, Y., and X. Wang, 2023: Impact of radar reflectivity data assimilation frequency on convection-allowing forecasts of diverse cases over the continental United States. *Mon. Wea. Rev.*, **151**, 341–362, https://doi.org/10.1175/MWR-D-22-0095.1.

Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multi-parameterization options (Noah-MP): 2. Evaluation over global river basins. *J. Geophys. Res.*, **116**, D12110, https://doi.org/10.1029/2010JD015140.

Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, https://doi.org/10.1175/MWR-D-12-00237.1.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zhang, X., 2019: Multiscale characteristics of different-source perturbations and their interactions for convection-permitting ensemble forecasting during SCMREX. *Mon. Wea. Rev.*, **147**, 291–310, https://doi.org/10.1175/MWR-D-18-0218.1.

Zhou, X., and Coauthors, 2022: The development of the NCEP Global Ensemble Forecast System version 12. *Wea. Forecasting*, **37**, 1069–1084, https://doi.org/10.1175/WAF-D-21-0112.1.