A Comparison of 3DEnVar and 4DEnVar for Convective-Scale Direct Radar Reflectivity Data Assimilation in the Context of a Filter and a Smoother

YUE YANG[a] AND XUGUANG WANG[a]

[a] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 11 April 2023, in final form 16 October 2023, accepted 18 October 2023)

ABSTRACT: The Gridpoint Statistical Interpolation (GSI)-based four- and three-dimensional ensemble–variational (4DEnVar and 3DEnVar) methods are compared as a smoother and a filter, respectively, for rapidly changing storms using the convective-scale direct radar reflectivity data assimilation (DA) framework. Two sets of experiments with varying DA window lengths (WLs; 20, 40, 100, and 160 min) and radar observation intervals (RIs; 20 and 5 min) are conducted for the 5–6 May 2019 case. The RI determines the temporal resolution of ensemble perturbations for the smoother and the DA interval for the filter spanning the WL. For experiments with a 20-min RI, evaluations suggest that the filter and the smoother have comparable performance with a 20-min WL; however, extending the WL results in the outperformance of the filter over the smoother. Diagnostics reveal that the degradation of the smoother is attributed to the increased degree of nonlinearity and the issue of time-independent localization as the WL extends. Evaluations for experiments with different RIs under the same WL indicate that the outperformance of the filter over the smoother diminishes for most forecast hours at thresholds of 30 dBZ and above when shortening the RI. Diagnostics show that more frequent interruptions of the model introduce model imbalance for the filter, and the increased temporal resolution of ensemble perturbations enhances the degree of nonlinearity for the smoother. The impact of model imbalance on the filter overwhelms the enhanced nonlinearity on the smoother as the RI reduces.

SIGNIFICANCE STATEMENT: The background uncertainties of rapidly changing storms suffer from fast error growth and high degrees of nonlinearities during the data assimilation (DA) period. Two variants of the ensemble-based DA method can account for such temporal evolution. The smoother uses background ensemble from multiple observation times over an assimilation period to estimate the propagation of statistics. The filter frequently calculates the statistics at multiple observation times over the same period. Current comparisons of the smoother and the filter were mostly performed using simple models; however, unknowns remain for convection-allowing forecasts with additional complexities. This study compares the filter and the smoother for the convective-scale analysis and prediction using a real-data study and finds that the comparison varies with the assimilation period and the observation interval.

KEYWORDS: Convective storms; Radars/Radar observations; Data assimilation; Numerical weather prediction/forecasting

1. Introduction

Initialization of convective-scale numerical weather predictions (NWPs) for rapidly changing storms requires a data assimilation (DA) method to properly portray the temporal evolution of flow-dependent background-error covariances (e.g., Lu et al. 2017; Davis et al. 2021). Such evolution can be depicted by two variants of the ensemble DA method. The four-dimensional (4D) ensemble DA method uses ensemble perturbations from multiple times within a given DA window length (WL) to estimate the propagation of background-error covariances. Here, the temporal resolution of ensemble perturbations is identical to the observation interval within the DA window. The 4D ensemble DA approach follows the spirit of a smoother (e.g., van Leeuwen and Evensen 1996). In contrast, the three-dimensional (3D) ensemble DA method accounts for the temporal evolution of background-error covariances through frequent assimilation cycling, where the same given WL is divided into short intervals. The 3D ensemble DA approach follows the spirit of a filter (e.g., van Leeuwen and Evensen 1996). By design, the DA frequency in the filter is inversely proportional to the observation interval. In studies that compare the filter and the smoother, the analyses produced at the end of the WL are compared. For more details on the comparison in the context of a smoother and a filter, please refer to Bennett and Budgell (1989) and van Leeuwen and Evensen (1996).

Previous studies, using simple dynamic models, examined the theoretical differences between the filter and the smoother in the linear and nonlinear regimes. It was found that the filter and the smoother have identical solutions at the final analysis time in linear models (Bennett and Budgell 1989; Evensen 2004). In comparison, the filter and the smoother provide different solutions in nonlinear models (van Leeuwen and Evensen 1996; Evensen 1997; Evensen and van Leeuwen 2000). Such differences in final analyses were attributed to different degrees of non-Gaussianity from prior ensembles (Evensen 1997; Evensen and van Leeuwen 2000). Compared to the filter, the prior ensemble in the smoother was obtained through a longer full nonlinear model integration and thus had more non-Gaussian features.

While recent studies have compared the 3D ensemble DA method and its 4D counterpart in real NWP models, such as for global forecasts (e.g., Wang and Lei 2014; Kleist and Ide 2015a,b;

Corresponding author: Yue Yang, yue.yang@ou.edu

DOI: 10.1175/MWR-D-23-0082.1

Lorenc et al. 2015; Lorenc and Jardak 2018), hurricanes (e.g., Lu et al. 2017; Zhang and Pu 2019; Shen et al. 2020; Davis et al. 2021), and the convective storm (S. Wang et al. 2013), none of these comparisons were made in the context of comparing a filter and a smoother. Specifically, in these studies, either the 3D ensemble DA only performs a single update at the center of the WL rather than frequent updates within the WL or the analysis from the 4D ensemble DA is output at the center of the WL while its 3D counterpart is frequently cycled until the end of the WL, making it impossible to compare both analyses at the end of the WL. Therefore, this current study examines 3D ensemble–variational (EnVar; X. Wang et al. 2013) and 4DEnVar (Wang and Lei 2014) in the context of the filter and the smoother for the real convective-scale analysis and prediction.

The comparison of the filter and the smoother in real NWP models adds additional complexities compared to simple models as follows. First, subgrid-scale physical processes in convective-scale NWP models such as turbulence, radiation, and microphysics are parameterized via a series of physics schemes, introducing model errors and additional nonlinearities. Both the source of the nonlinearity and the measure that reflects the degree of nonlinearity can be more complex in real NWP models than those in ideal models. Since small errors grow linearly before subsequent large errors enter the nonlinear phase with steadily decreasing growth rates (Lorenz 1965; Krishnamurthy 1993; van Leeuwen and Evensen 1996), the WL is treated as a proxy for the degree of nonlinearity in this study. Moreover, for a smoother in a nonlinear regime, increasing the temporal resolution of ensemble perturbations may change the degree of nonlinearity resolved by the prior ensemble. Therefore, in addition to extending the WL, this study also varies the temporal resolution of ensemble perturbations (the same as the observation interval) to change the degree of nonlinearity.

Second, localization is required to cut off spurious correlations at long distances when using a finite ensemble size (Anderson 2007). Performing localization in both space and time for a smoother, however, is not straightforward. Most 4D ensemble DA approaches still use a time-independent localization as in a 3D ensemble DA approach. This implementation is likely to lose useful correlations for a smoother with a long WL (e.g., Anderson 2007, 2012; Cosme et al. 2012; Bocquet 2016), especially for rapidly changing storms. Therefore, in addition to nonlinearity, the impact of WL on the comparison between the filter and the smoother may be affected by the use of time-independent localization.

Finally, more frequent interruptions of the NWP model, like in a filter, can introduce model imbalance (e.g., Lynch and Huang 1992; S. Wang et al. 2013; Pan and Wang 2019; Yang and Wang 2023). Therefore, this study also includes the impact of observation interval on the comparison between the filter and the smoother.

In this study, to compare 4DEnVar and 3DEnVar in the context of the smoother and the filter for the real convective-scale analysis and prediction, the Gridpoint Statistical Interpolation (GSI) is adopted to directly assimilate radar reflectivity (Wang and Wang 2017). As a first step, this study uses the Advanced Research Weather Research and Forecasting (WRF-ARW) model (Skamarock et al. 2008) for the case study of a rapidly evolving storm that occurred on 5–6 May 2019.

The current paper proceeds as follows. Section 2 introduces the experimental settings, including the GSI-based EnVar system in the context of the filter and the smoother, model and DA configurations for the 5–6 May 2019 case, and verification methods. In section 3, the filter and the smoother are compared by varying the WL and the radar observation interval (RI) to consider the aforementioned three aspects. Section 4 provides the summary and conclusions.

2. Experiment setup

a. *GSI-based EnVar system in the context of the filter and the smoother*

Given the effective combination of variational and ensemble-based DA frameworks, the GSI-based EnVar system (Wang 2010) shows benefits in NWP using operational global models (e.g., Hamill et al. 2011; X. Wang et al. 2013; Wang and Lei 2014) and regional models (e.g., Benjamin et al. 2016; Lu et al. 2017; Wang and Wang 2021a, 2023). The capability of directly assimilating radar reflectivity was incorporated into this system by Johnson et al. (2015) for the ensemble Kalman filter (EnKF) and Wang and Wang (2017) for the EnVar. This convective-scale GSI-based EnVar system has been widely adopted in convection-allowing prediction systems, such as the High-Resolution Rapid Refresh (HRRR)- and Rapid Refresh Forecast System (RRFS)-like contexts (e.g., Duda et al. 2019; Roberts et al. 2020; Wang and Wang 2021b; Gasperoni et al. 2022, 2023; Yang and Wang 2023). As mentioned in Wang and Lei (2014), 4DEnVar is an extension of 3DEnVar in time. Different from 3DEnVar that incorporates ensemble perturbations at a single time, 4DEnVar uses ensemble perturbations at multiple times in the variational minimization.

The flowchart of the filter in Fig. 1a mirrors the procedure of the one-way coupled GSI-based 3DEnVar system from X. Wang et al. (2013). Starting from the initial conditions (ICs) valid at t_{n-1}, the ensemble members and the control member are integrated forward to the analysis time at $t_n - m \times dt$ to produce the first-guess fields ($m + 1$ denotes the total number of DA cycles within the WL and dt represents the DA interval). The control member forecast is updated via the EnVar using the background-error covariances estimated by the short-term forecast ensemble without static covariances, and the ensemble members are updated via the EnKF. The resultant control analysis and analysis ensemble are then propagated to the next DA cycle. Through $m + 1$ DA cycles with a DA interval of dt spanning the WL (i.e., $m \times dt$), the final control analysis at t_n is generated. Therefore, the temporal evolution of background-error covariances within the WL is incorporated within the filter by frequent DA cycling.

Initialized from the same ICs at t_{n-1}, the 4DEnVar smoother procedure shown in Fig. 1b is obtained by expanding the time dimension of the filter. As an alternative approach to account for the temporal evolution of background-error covariances within the WL, the ensemble perturbations with a temporal resolution of dt spanning the WL are generated at

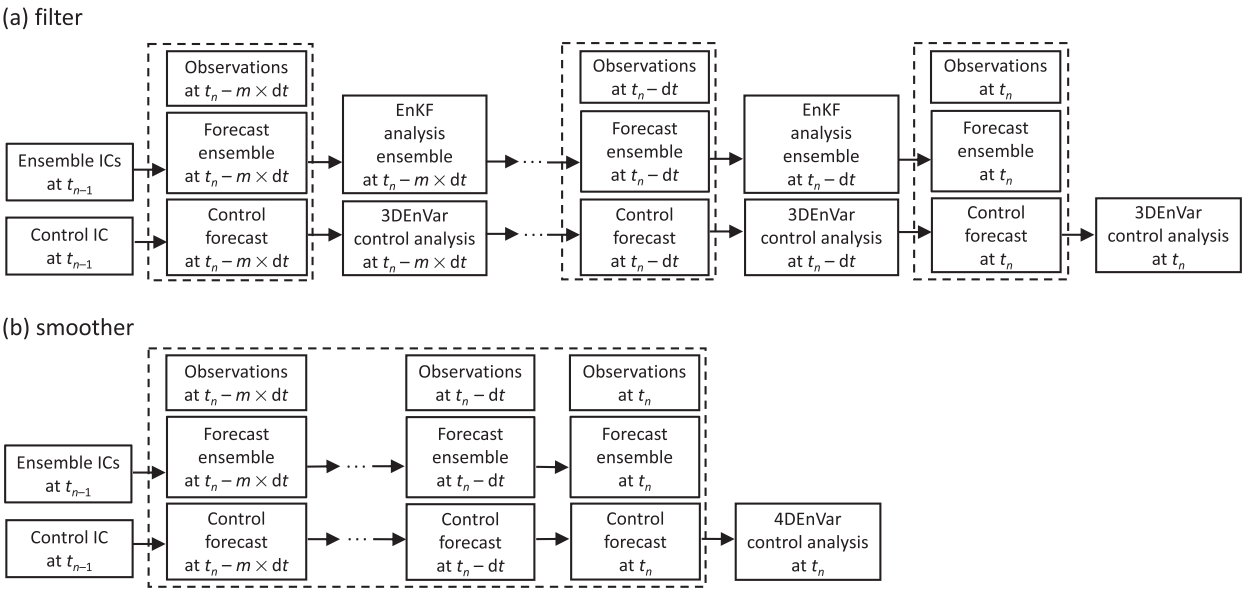

FIG. 1. Simplified schematics of (a) the filter and (b) the smoother within the GSI-based EnVar data assimilation system, where t_{n-1} indicates the time of the common initial conditions, t_n denotes the final analysis time, dt is the time interval for observations, and $m + 1$ is the number of observation times during the data assimilation period (i.e., $m \times dt$). For the filter in (a), the top row denotes the ensemble members updated every dt by the EnKF and the bottom row denotes the control member updated every dt by the EnVar. For the smoother in (b), the temporal resolution of ensemble perturbations is dt.

$t_n - m \times dt, \ldots, t_n - dt$, and t_n. Then, the 4D background-error covariances estimated by the ensemble perturbations are applied to update the final control forecast valid at t_n. As a result, the final analyses of the filter and the smoother can be compared at t_n.

b. *Case overview*

The filter and the smoother are compared for a rapidly evolving storm from 5 to 6 May 2019 to account for the temporal evolution of background-error covariances. As shown in Figs. 2a–d, widely scattered supercells developed and persisted along the dryline across a broad north–south expanse of High Plains from eastern Nebraska (NE) to western Texas (TX). Moist advection east of the dryline, together with diurnal heating and steep midlevel lapse rates, promoted rapid upscale growth of convection, with expected rapid error growth corresponding to convective and microphysical processes. A cluster of supercells in the southeast NE and northern Kansas (KS) initiated between 2100 and 2300 UTC 5 May (Fig. 2a) and drifted southeastward along a slow-moving surface cold front, resulting in an east–west-oriented line of storm around 0040 UTC 6 May (Figs. 2c,d). Thunderstorms that developed along convergence bands located at the southwest KS (Fig. 2a) moved southeastward into the southeast KS and northern Oklahoma (OK) (Fig. 2d). The isolated storms that formed along the dryline over the TX Panhandle (Fig. 2a) grew upscale and matured as they propagated toward the east (Figs. 2b–d).

c. *Configuration of data assimilation and forecast*

The WRF-ARW Model version 3.9 is coupled with the GSI-based EnVar DA system for the forecasts. The model domain over the continental United States (CONUS) consists of 51 full-η levels and 1621 × 1121 grid points at a 3-km horizontal grid spacing. The physics configuration from the HRRR system (Alexander et al. 2020; Dowell et al. 2022) is adopted as follows: the Rapid Update Cycle (RUC) land surface model (Benjamin et al. 2004), the aerosol-aware Thompson microphysics scheme (Thompson and Eidhammer 2014), the Mellor–Yamada–Nakanishi–Niino (MYNN) Level-2.5 planetary boundary layer parameterization (Nakanishi and Niino 2006, 2009; Olson et al. 2019), and the Rapid Radiative Transfer Model for general circulation models (RRTMG) shortwave and longwave radiation schemes (Iacono et al. 2008).

The DA cycling setup used to produce the common ICs for the filter and the smoother (Fig. 1) is designed in Fig. 3. At 1800 UTC 5 May 2019, perturbations for a 40-member ensemble are generated by recentering the 20-member Short-Range Ensemble Forecast (SREF) ensemble (Du et al. 2015) and the 20-member Global Ensemble Forecast System (GEFS) ensemble (Zhou et al. 2017) around the Global Forecast System (GFS) control analysis. The mean of the 40-member ensemble serves as the control member. Then, the 3DEnVar and 3DEnKF are performed to update the control member and ensemble members, respectively, via hourly assimilation of conventional data (i.e., routine surface observations, mesonet observations, and data from radiosondes, aircraft, ships, and buoys) from 1900 to 2200 UTC, emulating the hourly DA of conventional data for the HRRR system (Dowell et al. 2022). Next, assimilation of Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) radar reflectivity is performed every 20 min from 2200 to 2240 UTC. Note that only the 2-min reflectivity data closest to the analysis time are incorporated for the variational

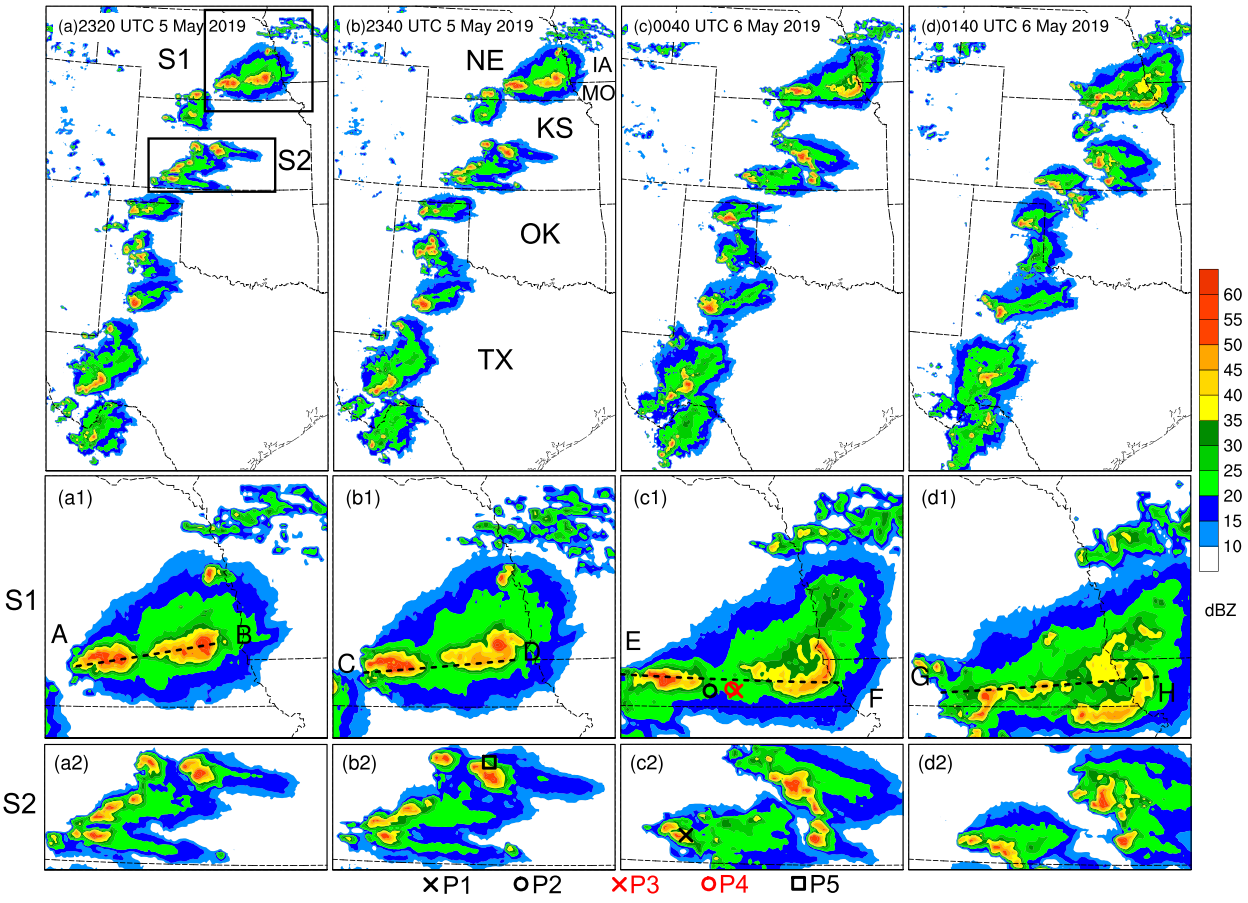

FIG. 2. (a)–(d) Observed Multi-Radar Multi-Sensor (MRMS) composite reflectivity (color fill; dBZ) for the 5–6 May 2019 case valid at the time indicated in each panel. Representative storms S1 and S2 are framed by black rectangles in (a). The associated state abbreviations are added in (b). The zoomed-in (a1)–(d1) S1 and (a2)–(d2) S2 with markers P1 in (c2), P2–P4 in (c1), and P5 in (b2) will be discussed later in section 3. The black dashed lines in (a1)–(d1) indicate the locations of cross sections in Fig. 6.

minimization at each cycle of 3DEnVar. The 20-min interval for the reflectivity assimilation is selected due to the most reliable performance discussed in Yang and Wang (2023).

Starting from 2240 UTC 5 May, the time series of root-mean-square error (RMSE) is calculated for locations with reflectivity above 10 dBZ from the control member (Fig. 4). The prior_run_RI20 describes the forecast error growth every 20 min from the common control analysis ICs at 2240 UTC (20 min before the 2300 UTC analysis cycle) out to 160 min after the 2300 UTC analysis cycle. The high-quality common ICs at 2240 UTC guarantee remarkable and rapidly growing errors during the subsequent forecasts. It is shown that the RMSE increases fast during the early period of the model integration, followed by gentle growth after 60 min into the 2300 UTC analysis cycle. Therefore, we can infer that various WLs from 2300 UTC can represent the different degrees of nonlinearity. In this study, the WLs of 20, 40, 100, and 160 min are chosen to design experiments in Table 1.

As listed in Table 1, we conducted six pairs of experiments to compare the filter and the smoother for direct reflectivity DA. The first four pairs are used to examine the impact of WL on the comparison between the filter and the smoother via fixing a 20-min RI and comparing different WLs from 20 to 160 min (i.e., WL20_RI20, WL40_RI20, WL100_RI20, and WL160_RI20). The last two pairs further shorten the RI from 20 to 5 min (i.e., WL20_RI5 and WL40_RI5) aiming to explore the impact of RI on the comparison between the filter and the smoother. The prefixes of 3D and 4D correspond to the filter (e.g., 3D_WL40_RI20) and the smoother (e.g., 4D_WL40_RI20), respectively.

As an example, Fig. 3 briefly demonstrates the procedures of DA cycling and free forecasts for 3D_WL40_RI20 and 4D_WL40_RI20 following the flowcharts of the filter and the smoother (Fig. 1). Initialized from the common ICs valid at 2240 UTC, 3D_WL40_RI20 assimilates reflectivity every 20 min starting at 2300 UTC until 2340 UTC to generate the final control analysis, whereas 4D_WL40_RI20 outputs the control analysis only at 2340 UTC for the comparison. Although the final analysis time is different among experiments with various WLs, free deterministic forecasts are launched from the final control analyses out to 0500 UTC 6 May.

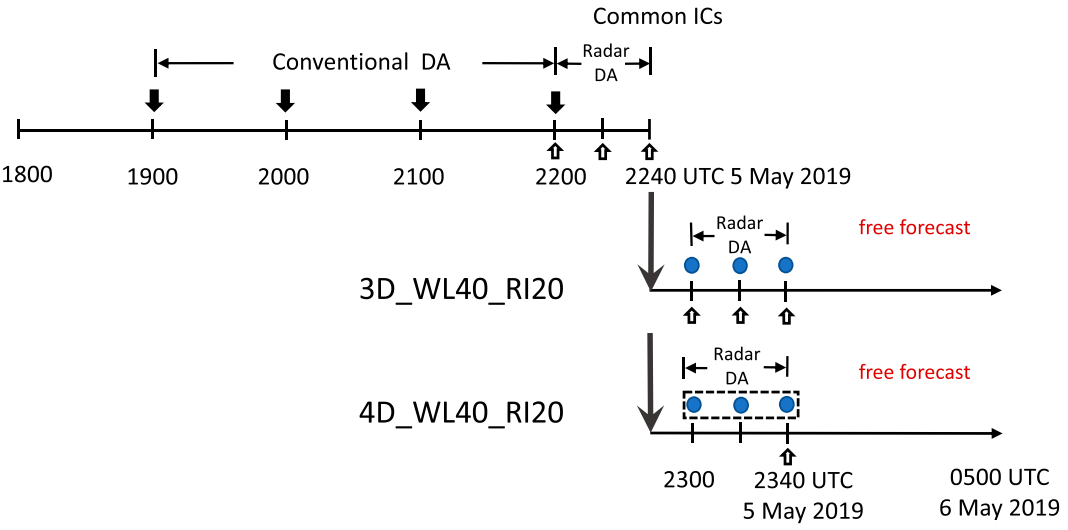

FIG. 3. Flowchart to generate common initial conditions valid at 2240 UTC 5 May 2019. Schematics of the DA setups and the free forecasts of the filter and the smoother experiments, where a 40-min DA window length and a 20-min radar observation interval (i.e., 3D_WL40_RI20 and 4D_WL40_RI20 in Table 1) are chosen as examples.

d. *Verification methods*

Figures 2a–d show the verification domain that fully covers the evolution of dominant storms. Following the common usage in severe weather forecasting (e.g., Roberts et al. 2012; Degelia et al. 2020), several statistical metrics as a function of threshold are chosen to assess forecast skills of the filter and the smoother for the composite reflectivity.

Contingency table–based verification indices (Doswell et al. 1990) include the probability of detection (POD), false alarm ratio (FAR) or its opposite, success ratio (SR = 1 − FAR), bias, and critical success index (CSI; Schaefer 1990). The geometric relationships among these measures are derived by Roebber (2009) as follows:

$$\mathrm{CSI} = \frac{1}{1/\mathrm{SR} + 1/\mathrm{POD} - 1}, \quad \text{and} \tag{1}$$

$$\mathrm{bias} = \frac{\mathrm{POD}}{\mathrm{SR}} = \tan\theta, \tag{2}$$

where θ is the angle from the x axis. Hence, the relative differences in POD, SR, bias, and CSI can be visualized in a performance diagram, where a perfect forecast is located at the upper right.

As a neighborhood-based metric, the fractions skill score (FSS; Roberts and Lean 2008) allows for spatial displacements of meso- and storm-scale features. Based on the fractions Brier score (FBS), FSS is calculated as

$$\mathrm{FSS} = 1 - \frac{\mathrm{FBS}}{\mathrm{FBS}_{\mathrm{worst}}} = 1 - \frac{\frac{1}{N_v}\sum_{i=1}^{N_v}[\mathrm{NP}_{F(i)} - \mathrm{NP}_{O(i)}]^2}{\frac{1}{N_v}\left[\sum_{i=1}^{N_v}\mathrm{NP}_{F(i)}^2 + \sum_{i=1}^{N_v}\mathrm{NP}_{O(i)}^2\right]}, \tag{3}$$

where N_v denotes the total number of grid points at a given threshold, $\mathrm{NP}_{F(i)}$ is the neighborhood forecast probability at the ith grid point for a given threshold, and $\mathrm{NP}_{O(i)}$ is the corresponding observed neighborhood probability. The FBS measures the squared difference between forecast and observed fractions and the $\mathrm{FBS}_{\mathrm{worst}}$ represents the maximum possible value of FBS when no collocation exists between nonzero fractions. The value of FSS starts from 0 (no skill) and goes to 1 (perfect forecast). Given that overlarge neighborhood radii mitigate the risk of double penalty at the expense of losing detail (e.g., Roberts 2008; Roberts and Lean 2008), a 48-km neighborhood radius (16 grid points) is applied based on previous work of convective-allowing forecasts (e.g., Johnson and Wang 2012; Duda et al. 2014; Gasperoni et al. 2020; Yang and Wang 2023). Without bias correction, the differences in FSS account for both random and systematic errors.

e. *Non-Gaussian measure*

To measure the degree of non-Gaussianity for error distributions, Kullback–Leibler divergence (D_{KL}; Kullback and Leibler 1951; Ruiz et al. 2021) is utilized, given by

$$D_{KL} = \int p(x)\log\frac{p(x)}{q(x)}dx, \tag{4}$$

where $p(x)$ is the probability density function (PDF) for the state variable from the histogram of the ensemble and $q(x)$ is the PDF of the fitted Gaussian distribution with the mean and the standard deviation from the ensemble. The term D_{KL} equals 0 when $p(x)$ and $q(x)$ are identical; otherwise, it is positive by the Gibbs inequality. Therefore, a larger D_{KL} suggests that $p(x)$ has a greater departure from the Gaussian distribution. In the present study, x refers to the perturbation of a variable, such as reflectivity and vertical velocity. As suggested by Kondo and Miyoshi (2019), an ensemble size of 1000 is sufficient to describe a non-Gaussian PDF that avoids sampling error contaminations. To increase the sample size in this

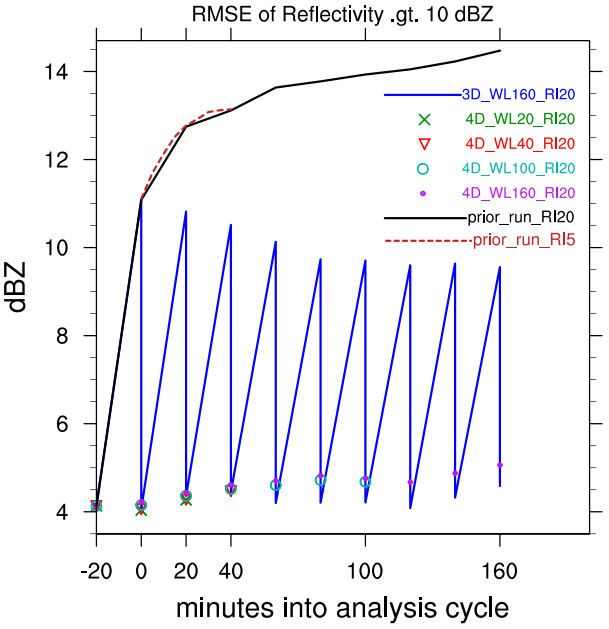

FIG. 4. Evolution of root-mean-square error (RMSE) of reflectivity above 10 dBZ for the control member of the experiments. The "0" represents the 2300 UTC analysis cycle. The RMSE statistics for the control first guess and the control analysis of the filter with various data assimilation window lengths (WLs) are shown in the upper and lower points of the sawtooth pattern, respectively (blue curve; named with the 160-min WL). The black curve represents the RMSE for the control first guess of the smoother with a 160-min WL and a 20-min radar observation interval (RI), and the brown dashed curve denotes the smoother with a 40-min WL and a 5-min RI. The markers show the RMSE for the control analysis of the smoother with a 20-min RI under indicated WLs.

study, the ensemble perturbations from a square around a given grid point of the 40-member ensemble are selected to calculate D_{KL}. Tests for the sensitivity of D_{KL} to the sample size via increasing the size of the square suggest that using a square of 5×5 grid points (i.e., 1000 samples) can alleviate sampling errors. Reducing subsamples from the 5×5 square suggests that the correlations associated with the neighboring grid points do not change the relative performance between the filter and the smoother.

3. Results

a. *Impact of data assimilation window length on the comparison between the filter and the smoother*

1) COMPARISON OF FINAL CONTROL ANALYSES

The RMSE for reflectivity is used to compare the closeness of final control analyses to the observations between the filter and the smoother under various WLs (Fig. 4). Note that we only compare the filter and the smoother that share the same settings for the WL and RI, and the posterior RMSE statistics are not used to measure the accuracy of analyses. For experiments with 20- and 40-min WLs, the filter and the smoother have a similar degree of closeness to the observations at the final analysis time. As the WL extends to 100 and 160 min, however, the posterior RMSE of the smoother gradually drifts away from the observations. It turns out that the final analysis of the filter fits closer to the observations than that of the smoother. Thus, the comparison between the filter and the smoother in the fitness of the final control analysis to the observations varies with the WL.

TABLE 1. List of experiments for the comparison between the filter and the smoother, where the settings for the data assimilation window length (WL) and the radar observation interval (RI) are changed.

Expt	WL (min)	RI (min)
WL20_RI20	20	20
WL40_RI20	40	20
WL100_RI20	100	20
WL160_RI20	160	20
WL20_RI5	20	5
WL40_RI5	40	5

The comparisons of final control analyses between the filter and the smoother are consistent for most isolated storms over the whole domain (Figs. 2a–d). As an example, Fig. 5 shows the final control analyses of representative storms S1 and S2 from the filter and the smoother with different WLs. For composite reflectivity, the substantial differences between the filter and the smoother [denoted by black rectangles in Figs. 5f(2)–h(2)] begin to appear when the WL exceeds 20 min. The pattern and intensity of composite reflectivity in 3D_WL20_RI20 [Figs. 5a(1),a(2)] are comparable to those in 4D_WL20_RI20 [Figs. 5e(1),e(2)], both of which resemble the observations valid at 2320 UTC 5 May [Figs. 2a(1),a(2)]. As the WL extends to 40 min, some reflectivity cores located at KS in 4D_WL40_RI20 become slightly weaker than 3D_WL40_RI20 [Figs. 5b(2),f(2)]. The observed composite reflectivity valid at 2340 UTC [Figs. 2b(1),b(2)] is well captured by 3D_WL40_RI20 [Figs. 5b(1),b(2)]. Compared to 4D_WL100_RI20 [Figs. 5g(1),g(2)], the analyzed composite reflectivity of 3D_WL100_RI20 [Figs. 5c(1),c(2)] agrees better with the observations valid at 0040 UTC 6 May [Figs. 2c(1),c(2)]. In 4D_WL100_RI20, several reflectivity peaks in southern KS are underestimated, such as location P1 in Fig. 5g(2). In addition, 4D_WL100_RI20 has spurious scattered composite reflectivity in the northeast NE and northwest IA and overestimated radar echoes in the vicinity of location P2 [Fig. 5g(1)]. Similar underestimation and overestimation occur in 4D_WL160_RI20 [Figs. 5h(1),h(2)], and the analysis of 3D_WL160_RI20 [Figs. 5d(1),d(2)] is closer to the observations valid at 0140 UTC [Figs. 2d(1),d(2)].

In terms of the maximum vertical velocity (Wmax), the smoother deviates from the filter when the WL exceeds 40 min [denoted by red rectangles in Figs. 5g(1),h(1)]. In all filter experiments [Figs. 5a(1)–d(1),a(2)–d(2)] and smoother experiments with 20- and 40-min WLs [Figs. 5e(1),f(1),e(2),f(2)], the coverage outlined by Wmax of 4 m s^{-1} mainly overlaps the composite reflectivity above 35 dBZ at the final analysis time. However, 4D_WL100_RI20 and 4D_WL160_RI20 have updraft zones that do not coincide with intense reflectivity echoes for the storm S1 [Figs. 5g(1),h(1)], for instance, two locations P3 and P4 in Fig. 5g(1).

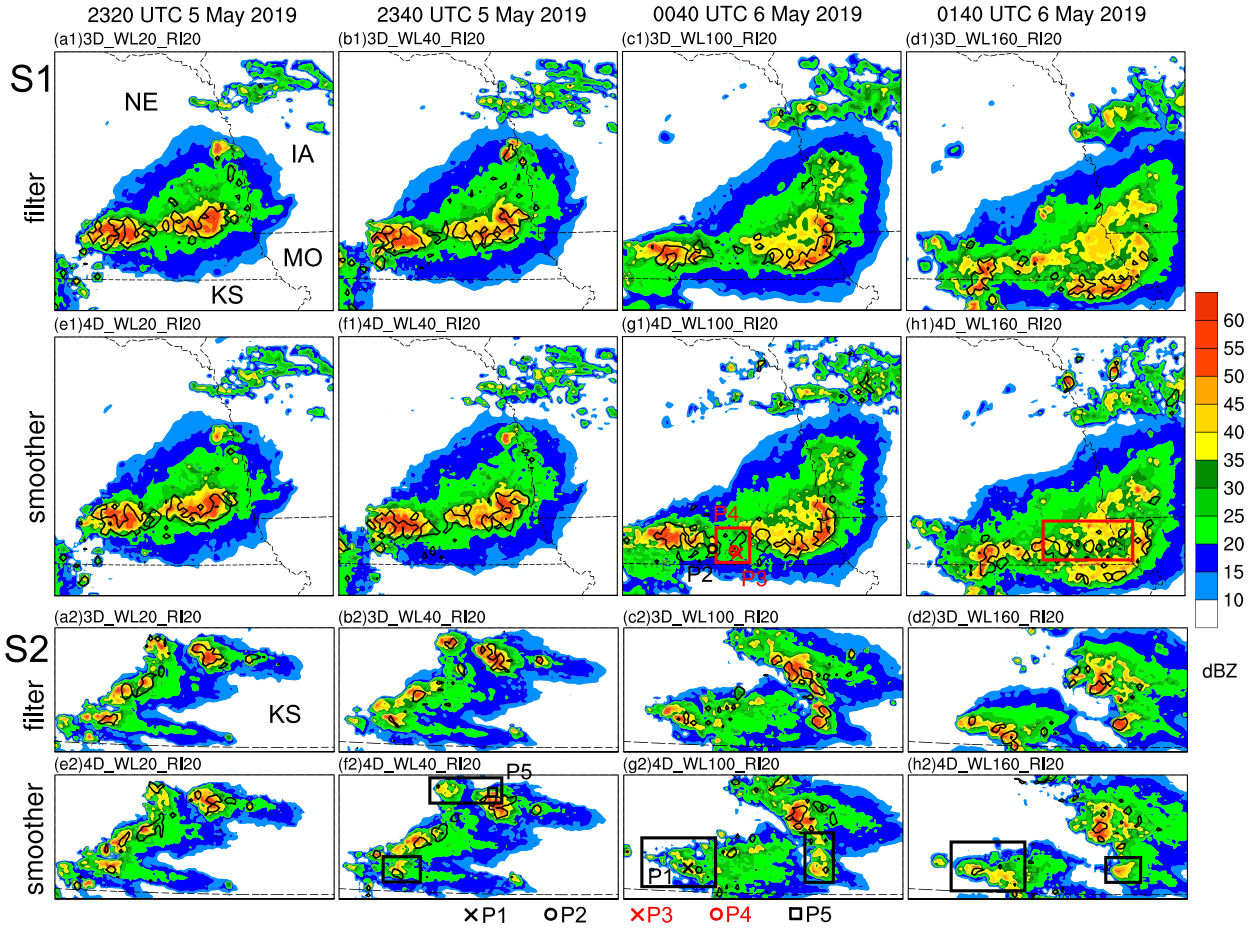

FIG. 5. Final control analyses of composite reflectivity (color fill; dBZ) and maximum vertical velocity (black contours; 4 m s^{-1}) for representative storms S1 and S2 from (a1)–(d1),(a2)–(d2) filter and (e1)–(h1),(e2)–(h2) smoother experiments with various DA window lengths (WLs) valid at the final analysis time indicated. Experiments in each column share the same settings for the WL and the radar observation interval (RI). The red rectangles in (g1) and (h1) denote the regions with remarkable differences in the updraft between the filter and the smoother, and the black rectangles in (f2)–(h2) denote differences in the composite reflectivity.

To explore the relationship between the mid- to upper-level radar echoes and vertical velocities for storm S1 at the final analysis time, observed and analyzed cross sections above 3 km along the black dashed lines in Figs. 2a(1)–d(1) are plotted accordingly (Figs. 6a–l). It is found that the filter and the smoother under the same WL have no substantial difference in the reflectivity pattern (Figs. 6e–h vs Figs. 6i–l). For vertical velocity in the filter and smoother experiments with 20- and 40-min WLs, the updraft cores usually collocate with the high-reflectivity regions above 35 dBZ (Figs. 6e,f,i,j). However, when the WL is larger than 40 min, the smoother always has an extra updraft zone with a maximum exceeding 4 m s^{-1} ranging from 6 to 12 km compared to the filter (red rectangles in Figs. 6k,l vs Figs. 6g,h). Scatterplots of the mid- to upper-level reflectivity and vertical velocities within red rectangles of Figs. 5g(1) and 5h(1) show that the smoother has much more instances that strong vertical motions unrealistically correspond to weak echoes compared to the filter (Figs. 6m–p).

Taking storm S1 as an example, Fig. 7 provides the distributions of the final analyzed cold pools represented by potential temperature perturbations at the lowest model level from the filter and the smoother with various WLs. The coverage and strength of the cold pool in 3D_WL20_RI20 are comparable to those in 4D_WL20_RI20 at 2320 UTC 5 May (Figs. 7a,e). When the WL exceeds 20 min, the differences in the cold pool between the filter and the smoother become distinct (denoted by white rectangles in Fig. 7). In the region between two high-reflectivity cores, the potential temperature perturbation in 3D_WL40_RI20 is up to 10 K higher than that in 4D_WL40_RI20 at 2340 UTC (Figs. 7b,f). Moreover, 3D_WL40_RI20 has a ~3 K weaker cold pool than 4D_WL40_RI20 underneath the eastern reflectivity core. These differences are also seen with growing spatial extent as the WL increases to 100 min (Figs. 7c,g). For the final analysis valid at 0140 UTC 6 May, broader and stronger cold pools exist in 4D_WL160_RI20 compared to 3D_WL160_RI20 (Figs. 7d,h). Relative to the cold pool in 3D_WL160_RI20, the cold pool underneath the eastern reflectivity core in 4D_WL160_RI20 is shifted to the northeast. It should be noted that the large updraft zones in 4D_WL100_RI20

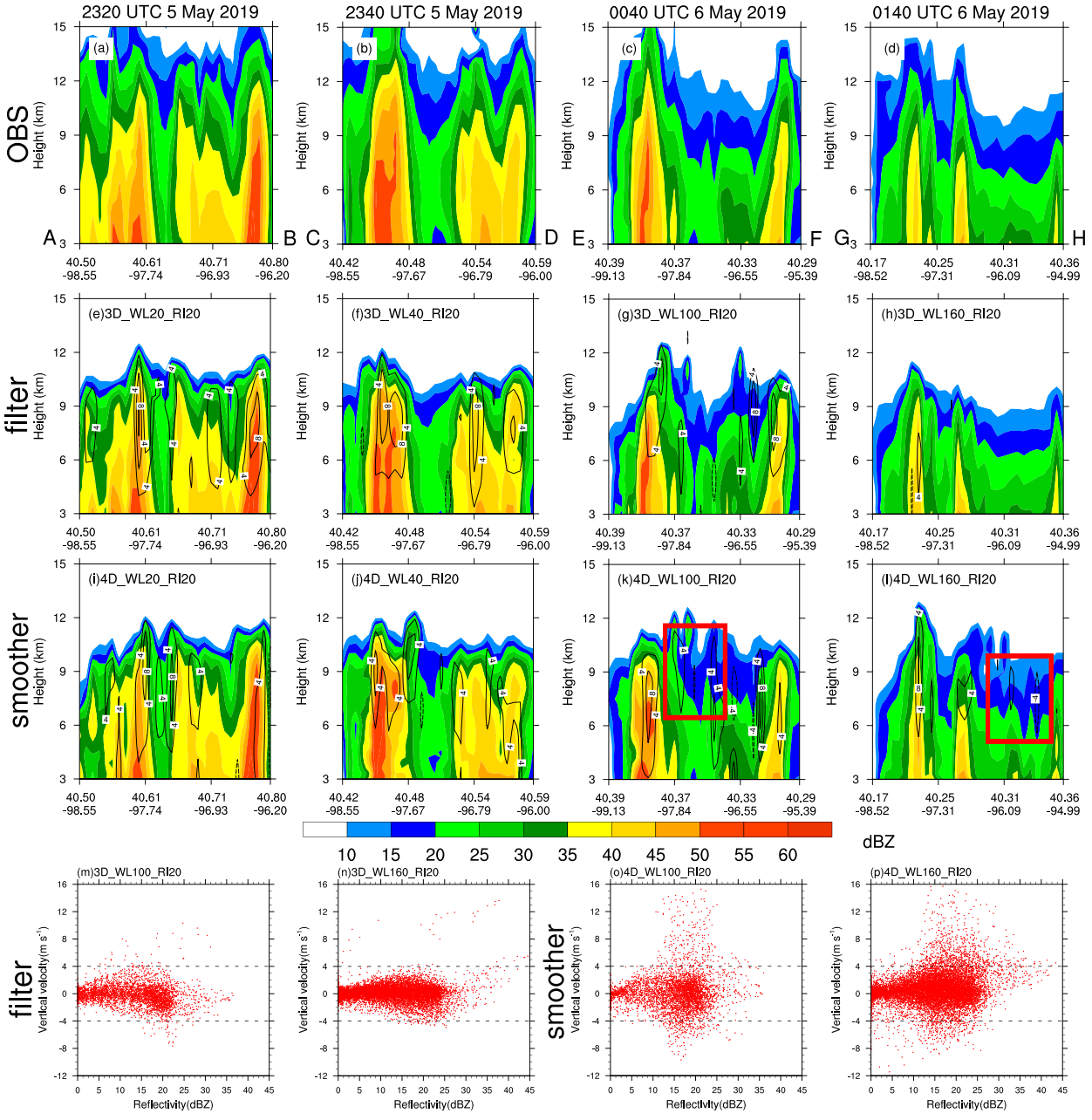

FIG. 6. Cross sections of (a)–(d) observed reflectivity (color fill; dBZ), analyzed reflectivity (color fill; dBZ), and analyzed vertical motions exceeding 4 m s^{-1} (black contours) above 3 km for (e)–(h) filter and (i)–(l) smoother experiments with various DA window lengths (WLs) valid at the final analysis time indicated. Experiments in each column share the same settings for the WL and the radar observation interval (RI). The red rectangles in (k) and (l) denote the extra updrafts in the smoother compared to the filter. Scatterplots of the reflectivity and vertical motions between 6 and 12 km for (m),(n) filter and (o),(p) smoother experiments with WLs of 100 and 160 min within red rectangles of Figs. 5g(1) and 5h(1), respectively.

and 4D_WL160_RI20 are collocated with strong cold pools featuring potential temperature perturbations from −3 to −5 K (Figs. 7g,h).

2) COMPARISON OF CONTROL FORECASTS

As pointed out by Sun et al. (2016), the quality of the analysis should be assessed using the subsequent forecasts rather than the closeness of the analysis to the observations. Therefore, the control forecasts from the filter and the smoother are compared under various WLs, including the temperature near the surface and the composite reflectivity.

The forecasts of temperature at the lowest model level (T) and composite reflectivity at the first forecast hour (0000 UTC for experiments with 20- and 40-min WLs, 0100 UTC for

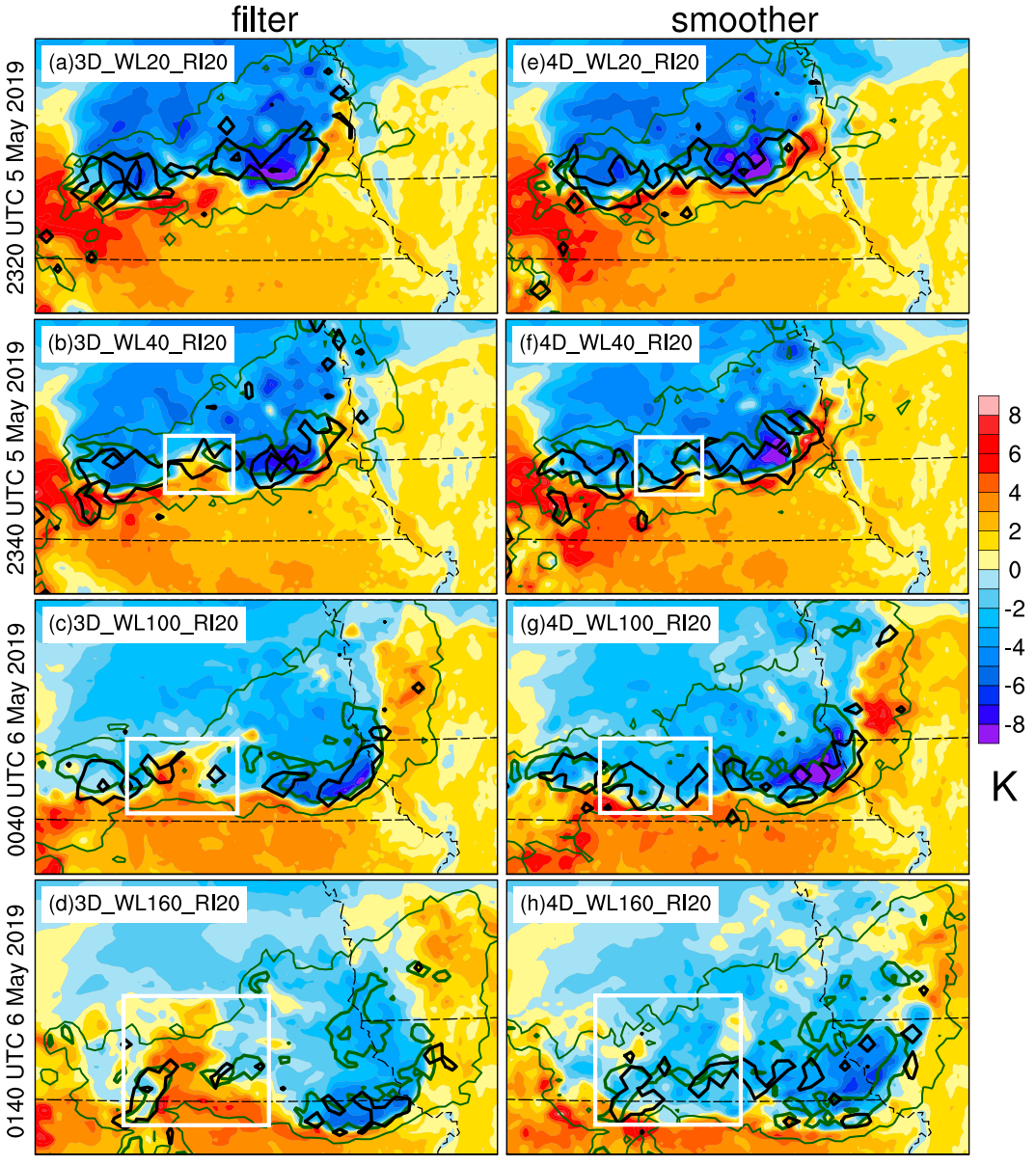

FIG. 7. Cold pool identified by potential temperature perturbations at the lowest model level (color fill; K), composite reflectivity (green contours; thin contours for 20 dBZ and thick contours for 40 dBZ), and maximum vertical velocity (black contours; 4 m s^{-1}) for (a)–(d) filter and (e)–(h) smoother experiments with various DA window lengths (WLs) valid at the final analysis time indicated. Experiments in each row share the same settings for the WL and the radar observation interval (RI). The white rectangles denote the regions with remarkable differences in the cold pool between the filter and the smoother.

experiments with a 100-min WL, and 0200 UTC for experiments with a 160-min WL) are shown in Fig. 8 overlaid with observed 2-m temperature and MRMS composite reflectivity. In terms of low-level temperature, remarkable differences between the filter and the smoother exist in cold pools near storm cores. The smoother tends to produce lower T than the filter counterpart (denoted by black arrows in Figs. 8e–h vs Figs. 8a–d). Such differences in T become broader and more remarkable as the WL increases. For instance, the cold pool related to storm S1 in 4D_WL20_RI20 is stronger by about 7 K than its 3D counterpart (Figs. 8a,e). Equipped with a WL of 40 min, such differences in T increase to 9 K (Figs. 8b,f). When the storm moves southeast approaching the boundary between NE and KS an hour later at 0100 UTC, the observation marked by the black arrow verifies that 3D_WL100_RI20 outperforms 4D_WL100_RI20 in forecasting T (Figs. 8c,g). The better forecast T in 3D_WL160_RI20 against this observation is further maintained at 0200 UTC, whereas T in 4D_WL160_RI20

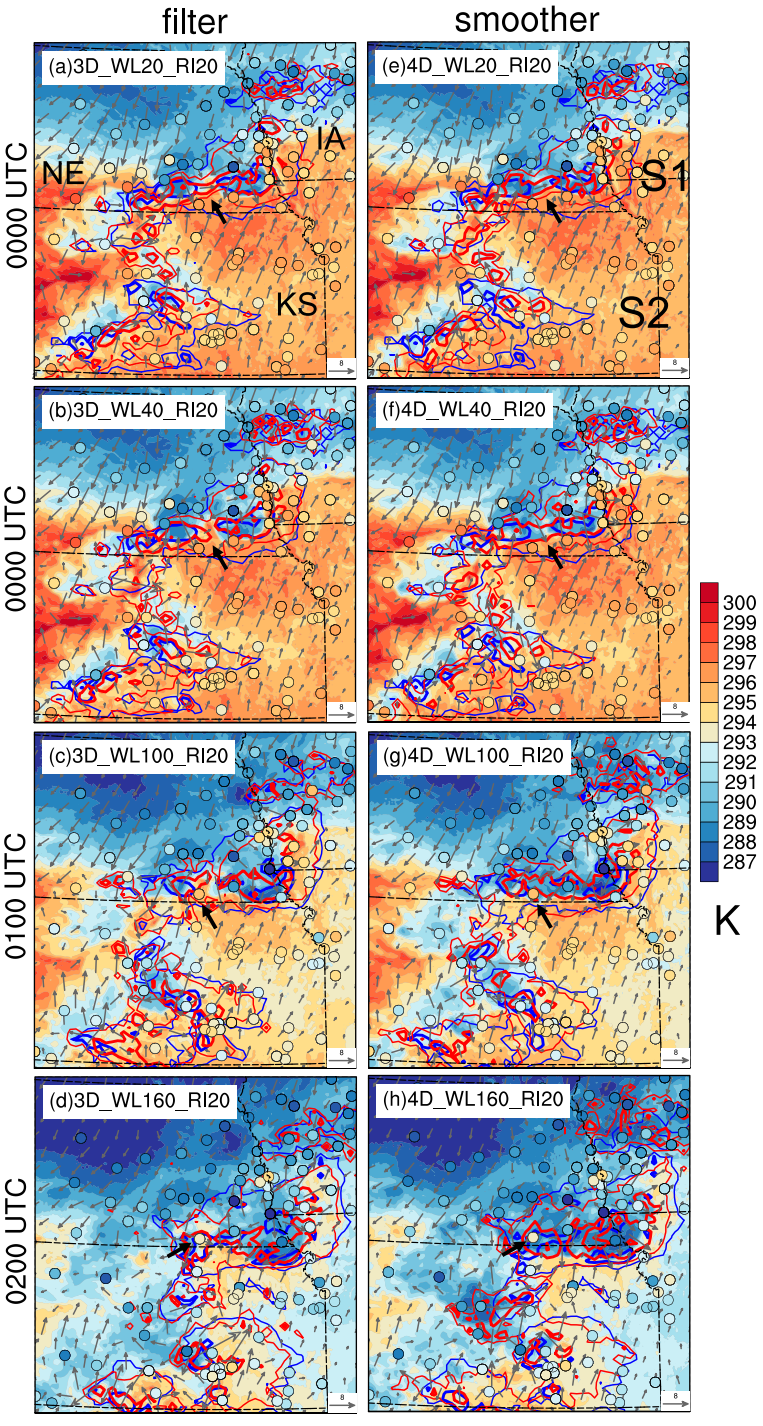

FIG. 8. Forecasts of temperature at the lowest model level (color fill; K) and composite reflectivity (red contours; thin contours for 20 dBZ and thick contours for 40 dBZ) overlaid with the observed composite reflectivity from the MRMS system (blue contours; thin contours for 20 dBZ and thick contours for 40 dBZ) for (a)–(d) filter and (e)–(h) smoother experiments with various DA window lengths (WLs) valid at the forecast hour indicated. Experiments in each row share the same settings for the WL and the radar observation interval (RI). Observations of 2-m temperature are overlaid using shaded circles. The black arrows denote the regions/locations with remarkable differences in the temperature between the filter and the smoother.

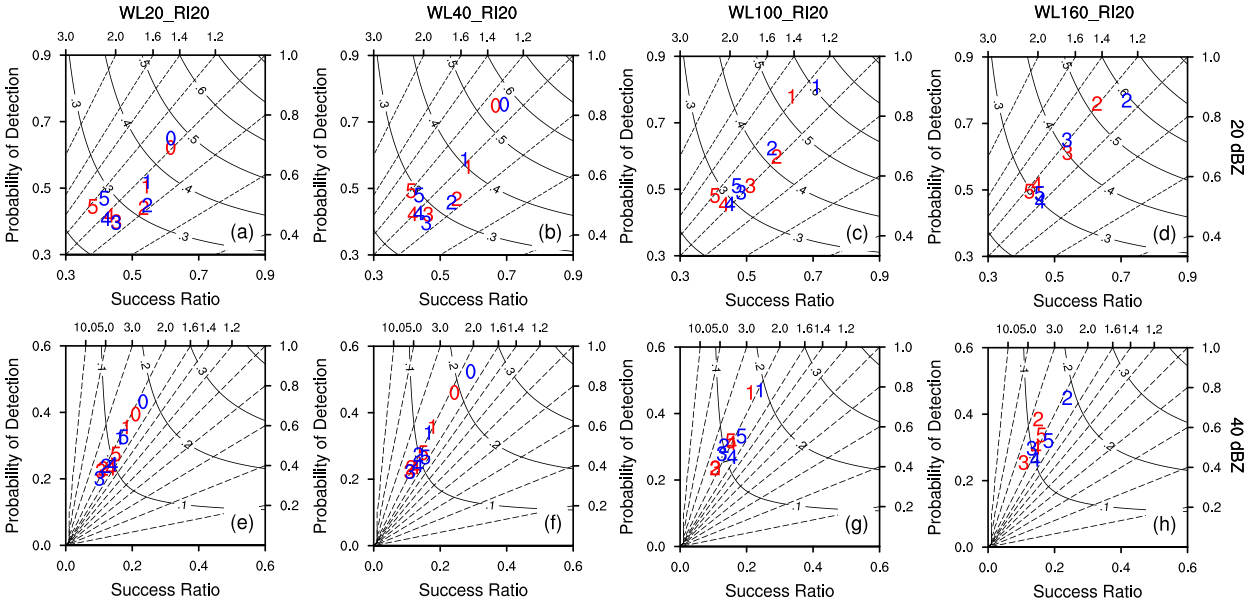

FIG. 9. Performance diagrams of composite reflectivity at (a)–(d) 20- and (e)–(h) 40-dBZ thresholds over the whole domain for the filter (blue markers) and smoother (red markers) experiments with various DA window lengths (WLs). Experiments in each column share the same settings for the WL and the radar observation interval (RI) as indicated. The colored numbers denote the valid leading time (UTC) on 6 May 2019.

has dropped to 289 K, about 4 K less than the observation (Figs. 8d,h). The forecasts of the too-strong cold pools in 4D_WL100_RI20 and 4D_WL160_RI20 are likely attributed to the large-area analyzed updrafts, consistent with the analyzed cold pools (Figs. 7g,h). Therefore, the analyzed updrafts within the red rectangles of Figs. 5 and 6 are spurious. During the forecast period, the spurious updrafts are able to lift extra moisture upward to the lifting condensation level, and the increased falling precipitation results in the enhanced cold pool due to evaporative cooling.

For composite reflectivity forecasts, 3D_WL20_RI20 and 4D_WL20_RI20 generate similar storm patterns at 0000 UTC 6 May, both of which match well with the observations (Figs. 8a,e). However, as the WL extends beyond 20 min, the smoother performs worse than the corresponding filter in storm coverage (Figs. 8f–h vs Figs. 8b–d). For example, 3D_WL40_RI20 has the capability of reproducing two isolated 40-dBZ composite reflectivity cores within storm S1 in southeast NE similar to observations at 0000 UTC (Fig. 8b). In contrast, 4D_WL40_RI20 overpredicts the 40-dBZ composite reflectivity core in an east–west-oriented line (Fig. 8f). Such overestimated core regions in the smoother (Figs. 8f–h) can be explained by the intensive cold pools that serve as continued initiation zones for additional convection. Compared to the weaker cold pools in the filter, the increased outflow surges from the stronger cold pools in the smoother easily converge with the ambient southerly flow along the leading edges of the cold pools. The newly developed updrafts along the convergence regions facilitate the formation of convective lines. Moreover, the smoother experiments with WL longer than 20 min still underestimate the coverage and intensity of storm S2 in south-central KS following the trends of the final analyses [Figs. 5f(2)–h(2)]. Likewise, the spurious scattered composite reflectivity in northeast NE and northwest IA from the final analyses of 4D_WL100_RI20 and 4D_WL160_RI20 [Figs. 5g(1),h(1)] remain during the early forecast period (Figs. 8g,h). Therefore, the differences in the final control analyses between the filter and the smoother under various WLs are transferred as similar differences in the control forecasts.

The performance diagrams for composite reflectivity over the whole domain in Fig. 2 provide an objective evaluation of forecast skill for the filter and the smoother under different WLs (Fig. 9). At the threshold of 20 dBZ, the CSIs of 3D_WL20_RI20 and 4D_WL20_RI20 are close in the early stage of forecasting until 0040 UTC 6 May (Fig. 9a). The deviations of forecasts from observations between the filter and the smoother gradually grow as the WL extends, especially during the early forecast period (Figs. 9b–d). Taking the first forecast hour as an example, the difference in CSI between 3D_WL40_RI20 and 4D_WL40_RI20 is less than 0.05 at 0000 UTC (Fig. 9b), the difference between 3D_WL100_RI20 and 4D_WL100_RI20 becomes larger than 0.05 at 0100 UTC (Fig. 9c), and the difference between 3D_WL160_RI20 and 4D_WL160_RI20 approaches 0.1 at 0200 UTC (Fig. 9d). The smoother with a WL above 20 min usually performs worse than its filter counterpart for most valid times. Such comparisons of CSI at the 40-dBZ threshold (Figs. 9e–h) show a similar trend to that at the 20-dBZ threshold but with larger deviations of forecasts from observations between the filter and the smoother for the early forecast period. Compared to the filter, the inferior performance of the smoother is mainly attributed to the larger bias, the smaller SR, and the smaller

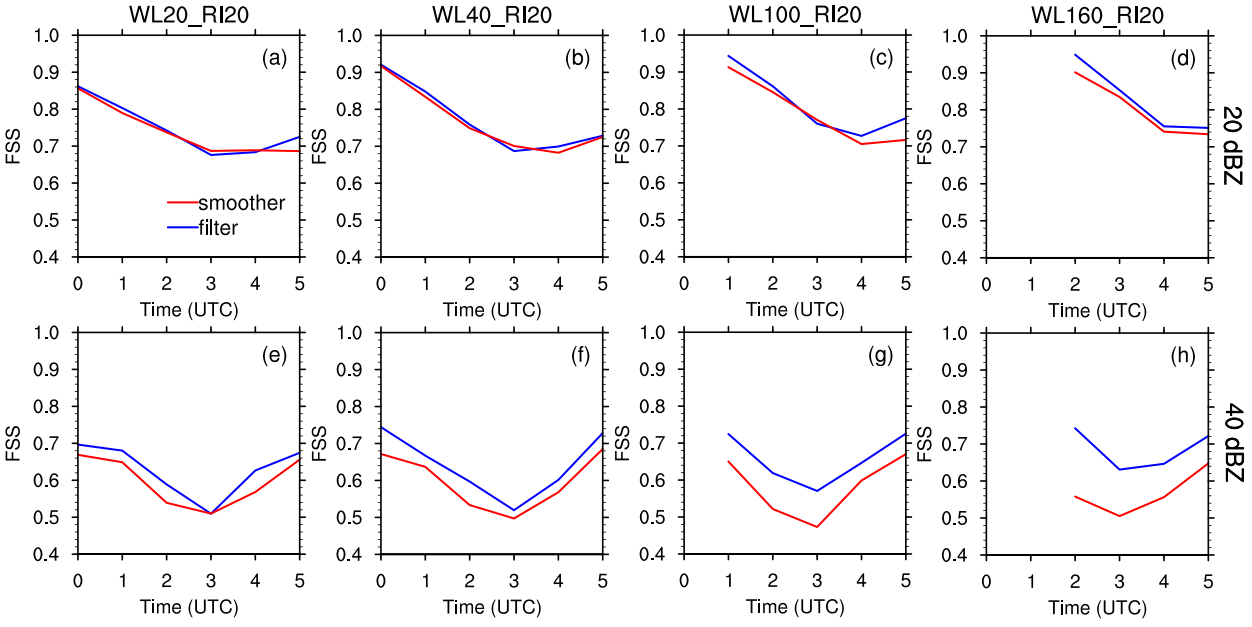

FIG. 10. FSSs of composite reflectivity at (a)–(d) 20- and (e)–(h) 40-dBZ thresholds over the whole domain for the filter (blue curves) and smoother (red curves) experiments with various DA window lengths (WLs) valid at times on 6 May 2019. Experiments in each column share the same settings for the WL and the radar observation interval (RI) as indicated.

POD, which are consistent with the subjective comparisons in Fig. 8.

The neighborhood-based FSS for the composite reflectivity over the whole domain in Fig. 10 draws a similar conclusion to the CSI in evaluating the forecast skills of the filter and the smoother with various WLs. At the threshold of 20 dBZ, the FSSs of the filter and the smoother are approximately equal when the WL is no more than 40 min (Figs. 10a,b). As the WL further extends, the smoother has lower FSSs than the filter, especially in the early forecasting stage (Figs. 10c,d). Compared to the 20-dBZ threshold, the gap between the smoother and the filter further widens at the threshold of 40 dBZ throughout almost the entire forecast period under each WL setting, showing worse performance of the smoother than the filter (Figs. 10e–h).

3) POTENTIAL EXPLANATIONS FOR THE DIFFERENCES IN THE FINAL CONTROL ANALYSES

The discrepancies in the control forecasts between the filter and the smoother under various WLs show close relationships with the discrepancies in the final control analyses. Compared to the filter, the lower forecast skills of the smoother result from poorer analyses. As discussed in section 1, the extension of the WL may increase the degree of nonlinearity and trigger issues of time-independent localization for the smoother. Therefore, the explanations for the differences in final control analyses between the smoother and the filter are investigated from the DA algorithm perspective using points P1–P5 in Fig. 5. We focus on the inferior reflectivity analyses of the smoother than the filter for WLs above 20 min and the inferior updrafts analyses when the WL exceeds 40 min.

For convective scales, Kondo and Miyoshi (2019) and Ruiz et al. (2021) indicated that non-Gaussianity is largely associated with highly nonlinear processes. Therefore, D_{KL} in this study is an indicator of the degree of nonlinearity. Figures 11a–d provide the distributions of D_{KL} at 0040 UTC 6 May for the reflectivity at the 13th model level (~3 km) in storms S1 and S2 of 3D_WL100_RI20 and 4D_WL100_RI20. It can be seen that the areas with substantial differences in final reflectivity analyses between 3D_WL100_RI20 and 4D_WL100_RI20 [Figs. 5c(1),c(2) vs Figs. 5g(1),g(2)] usually correspond to the areas where 4D_WL100_RI20 has larger values of D_{KL} than 3D_WL100_RI20. At P1, D_{KL} over 4.5 from 4D_WL100_RI20 (Fig. 11h) is much larger than D_{KL} of 0.48 from 3D_WL100_RI20 (Fig. 11g). Therefore, for P1 in 4D_WL100_RI20, we can infer that the higher level of nonlinearity is one reason for its worse reflectivity analysis because of the linear assumption. On the other hand, the autocorrelations of reflectivity at the 13th model level between the ensemble perturbation of P1 and the ensemble perturbation field at the final analysis time are calculated at different observation times within the DA window for 4D_WL100_RI20 (Figs. 12a–c). The positive correlations at P1 and its surroundings are reasonable because the reflectivity is missing in the first-guess field. However, due to the time-independent localization, the positive correlation peak gradually moves southeastward until out of the localization range from 0400 UTC 6 May back to 2300 UTC 5 May. The moving direction is determined by the location of P1 at different observation times relative to the background storm at the final analysis time. In Figs. 12a and 12b, the weak correlations within the localization radius cause the underestimation of reflectivity at P1 [Fig. 5g(2)]. Therefore, the time-independent localization at P1 of 4D_WL100_RI20 also leads to worse reflectivity analysis

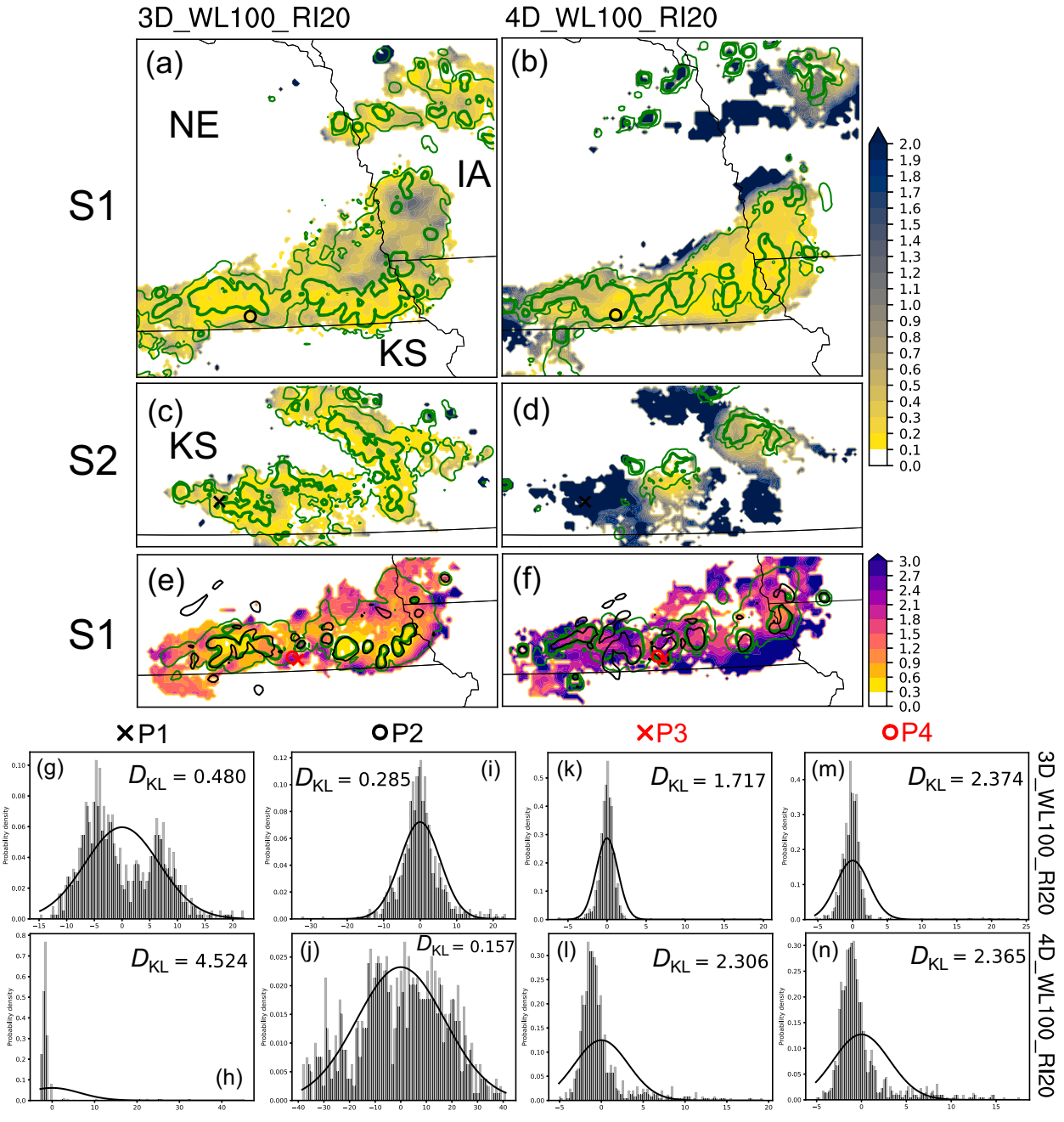

FIG. 11. (a)–(f) Spatial distributions of Kullback–Leibler divergence (D_{KL}; color fill) for (a)–(d) reflectivity of S1 and S2 at the 13th model level (~3 km) and for (e),(f) vertical velocity of S1 at the 19th model level (~6 km) valid at the final analysis time (i.e., 0040 UTC 6 May 2019) of 3D_WL100_RI20 in (a), (c), and (e) and 4D_WL100_RI20 in (b), (d), and (f). In (a)–(d), the corresponding control first-guess reflectivity of 20 dBZ (thin green contours) and 40 dBZ (thick green contours) are overlaid. In (e) and (f), the corresponding control first-guess fields of 2 m s^{-1} vertical velocity (black contours) and 20-dBZ (thin green contours) and 40-dBZ (thick green contours) reflectivity are overlaid. The D_{KL} is calculated when the control analysis of reflectivity exceeds 10 dBZ for reflectivity in (a)–(d) and 20 dBZ for vertical velocity in (e) and (f). (g)–(n) Ensemble-based histograms at P1–P4 with values of D_{KL}.

than 3D_WL100_RI20. Although 3D_WL100_RI20 has a slightly larger D_{KL} than 4D_WL100_RI20 at P2 [0.285 vs 0.157; Figs. 11i,j], the time-independent localization (Figs. 12d–f) plays a dominant role in degrading the final control analysis of 4D_WL100_RI20 at P2 [Fig. 5g(1)]. When the WL extends to 160 min, similar reasons discussed above lead to the worse reflectivity analysis of 4D_WL160_RI20 than 3D_WL160_RI20 (not shown). For P5 where 4D_WL40_RI20 has overestimated reflectivity [Fig. 5f(2)], D_{KL} from 4D_WL40_RI20 is 1.2 larger than that from 3D_WL40_RI20 (not shown). The higher degree of

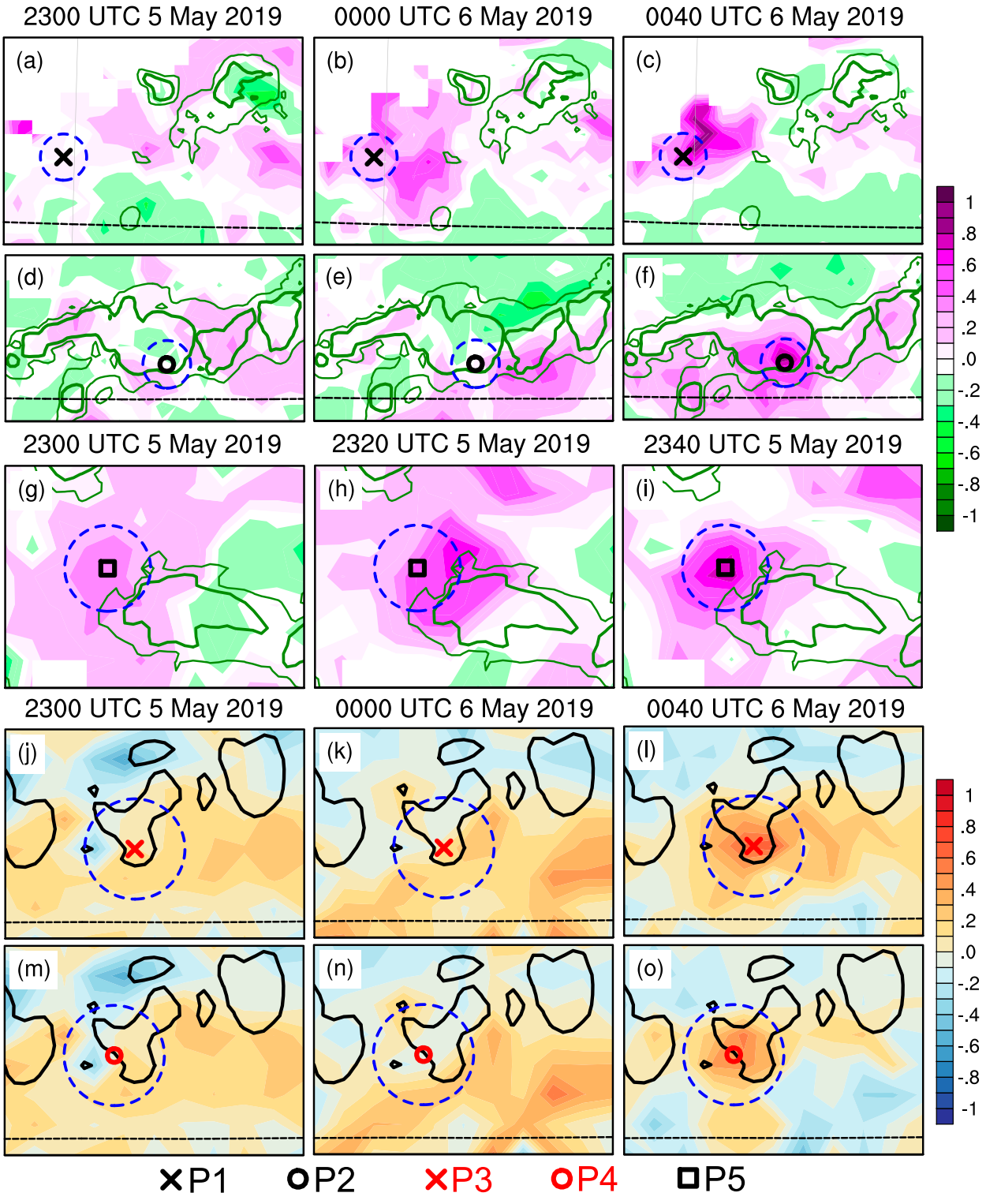

FIG. 12. (a)–(i) Autocorrelation of reflectivity at the 13th model level (~3 km) and (j)–(o) cross correlation of vertical velocity with reflectivity at the 19th model level (~6 km) between the single-point ensemble perturbation for reflectivity (markers; P1–P5) at the time indicated and the ensemble perturbation field at the final analysis time (indicated in the right column) for smoother experiments. Green contours in (a)–(i) denote the control first-guess reflectivity (thin contours for 20 dBZ and thick for 40 dBZ) at the 13th model level. Black contours in (j)–(o) denote the control first-guess 2 m s^{-1} vertical velocity at the 19th model level. Dashed navy circles represent the 15-km horizontal localization radii.

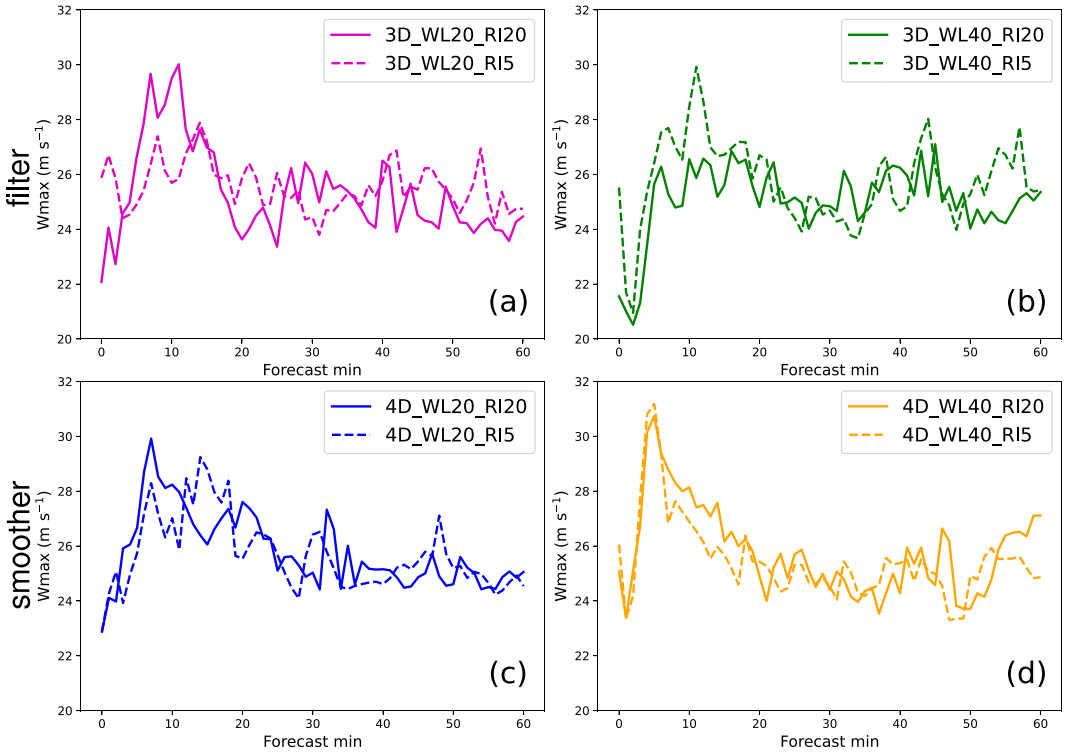

FIG. 13. The 1-h evolution of the maximum vertical velocity (Wmax; m s^{-1}) per min over the CONUS in forecasts starting at 2320 UTC 5 May 2019 from the (a) filter (magenta lines) and (c) smoother (blue lines) with a 20-min DA window length (WL) and at 2340 UTC 5 May 2019 from the (b) filter (green lines) and (d) smoother (orange lines) with a 40-min WL. The solid lines indicate the experiments using a 20-min radar observation interval (RI), and the dashed lines are used for experiments with a 5-min RI.

nonlinearity mainly explains the worse reflectivity analysis of P5 in 4D_WL40_RI20, since the location of the positive correlation peak changes slightly during the DA period (Figs. 12g–i).

To explain the spurious updrafts in the final control analysis of 4D_WL100_RI20 [Fig. 5g(1)], distributions of D_{KL} at 0040 UTC 6 May for the vertical velocity of storm S1 (Figs. 11e,f) and cross correlations of vertical velocity field with single-point reflectivity at P3 and P4 (Figs. 12j–o) are plotted at the 19th model level (~6 km). 4D_WL100_RI20 generally has larger values of D_{KL} than 3D_WL100_RI20 (Figs. 11e,f). The D_{KL} at P3 of 4D_WL100_RI20 is ~0.6 greater than that in 3D_WL100_RI20 (Figs. 11k,l), and the cross-correlation peak at P3 of 4D_WL100_RI20 moves eastward until out of the localization range from 0400 UTC 6 May back to 2300 UTC 5 May (Figs. 12j–l). Therefore, the combined effects of nonlinearity and time-independent localization hinder the elimination of the spurious first-guess updraft at P3 of 4D_WL100_RI20 through the reflectivity DA. Due to a nearly identical D_{KL} at P4 of 3D_WL100_RI20 compared to 4D_WL100_RI20 (2.374 vs 2.365; Figs. 11m,n), the spurious analyzed updraft at P4 of 4D_WL100_RI20 can be attributed to the limitation of time-independent localization (Figs. 12m–o). Similar reasoning of nonlinearity and time-independent localization also explains the spurious analyzed updrafts in 4D_WL160_RI20 compared to 3D_WL160_RI20 (not shown).

b. *Impact of shortening observation interval on the comparison between the filter and the smoother*

The sensitivity of the comparison between the filter and the smoother to the observation interval is examined using experiments with 20- and 40-min WLs. The time series of Wmax over the CONUS initialized from final control analyses is used to compare the adjustment of model imbalance before and after shortening the RI (Fig. 13). For filter experiments with the same WL, the Wmax curves from experiments with a 5-min RI show striking differences relative to experiments with a 20-min RI during the first 20 min (Figs. 13a,b). Given that both 3D_WL20_RI20 and 3D_WL40_RI20 take about 20 min to obtain a stable Wmax value around 25 m s^{-1}, 20 min is enough for them to reach a new model balance before the next reflectivity DA cycle without introducing noises. However, 3D_WL20_RI5 and 3D_WL40_RI5 cannot attain a stable Wmax within 5 min. Therefore, too frequent interruptions of the model in 3D_WL20_RI5 and 3D_WL40_RI5 could accumulate noise through successive cycles, affecting the subsequent forecasts from the final control analyses.

For smoother experiments with the same WL, shortening the RI from 20 to 5 min slightly changes the adjustment of model imbalance (Figs. 13c,d). All the smoother experiments can build a stable 26–27 m s^{-1} Wmax around 20 min. Nevertheless, applying a shorter RI may change the degree

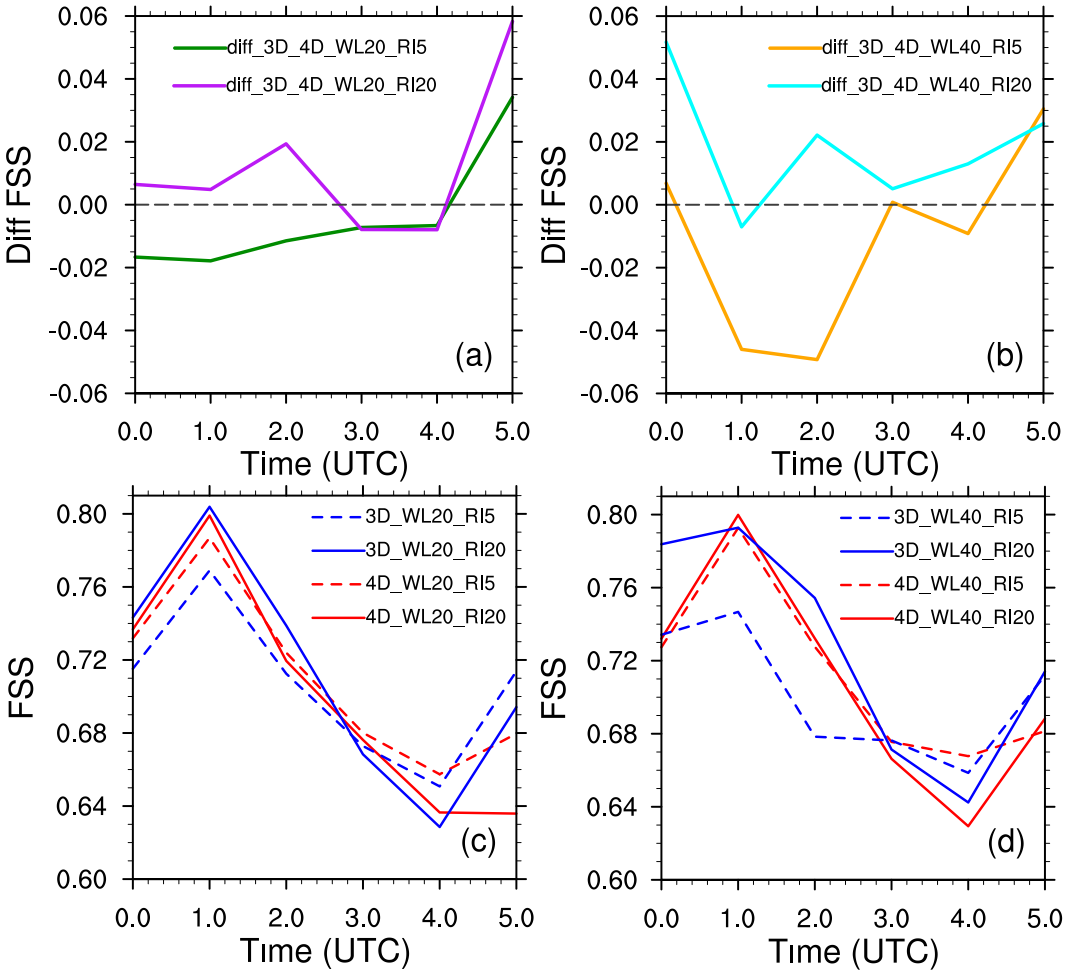

FIG. 14. Differences in FSS (filter minus smoother) with (a) 20- and (b) 40-min DA window lengths (WLs) for composite reflectivity at the 30-dBZ threshold valid at times on 6 May 2019 before and after shortening the radar observation interval (RI) from 20 to 5 min. (c),(d) The detailed FSS results of the filter and the smoother corresponding to (a) and (b), respectively.

of nonlinearity for the smoother. For example, compared to prior_run_RI20, prior_run_RI5 has slower error growth rates near 20 min and between 30 and 40 min into the analysis cycle (Fig. 4), which represent more prominent nonlinear features. As a result, shortening the RI probably affects the performance of the smoother due to the increased degree of nonlinearity, which more seriously violates the linear assumption for the evolution of background-error covariances.

To explore the impact of RI on the comparison between the filter and the smoother, their differences in FSS are calculated before (i.e., diff_3D_4D_WL20_RI20 and diff_3D_4D_WL40_RI20) and after shortening the RI (i.e., diff_3D_4D_WL20_RI5 and diff_3D_4D_WL40_RI5). It is found that the FSS differences before and after shortening the RI are similar at the thresholds below 30 dBZ (not shown), but the differences are substantial for thresholds of 30 dBZ and above. Thus, we use the results at the 30-dBZ threshold in Fig. 14 to discuss the trade-off between the model imbalance for the filter and the enhanced nonlinearity for the smoother. In Figs. 14a and 14b, positive values indicate that the filter is superior to the smoother and negative values denote inferior results of the filter than the smoother. Given the smaller values of diff_3D_4D_WL20_RI5 than diff_3D_4D_WL20_RI20 (Fig. 14a) and the smaller values of diff_3D_4D_WL40_RI5 than diff_3D_4D_WL40_RI20 (Fig. 14b) at most valid times, the advantages of the filter over the smoother reduce as the RI decreases. In the early forecast, such as 0200 UTC 6 May, the smoother with a 5-min RI performs better than its filter counterpart. Figures 14c and 14d show the detailed FSSs of the smoother and the filter before and after shortening the RI. It can be seen that reducing the RI makes the filter more degraded compared to the corresponding smoother during the earlier stage of forecasts, e.g., differences between 3D_WL20_RI20 and 3D_WL20_RI5 versus differences between 4D_WL20_RI20 and 4D_WL20_RI5 from 0000 to 0020 UTC (Fig. 14c). Using a shortened RI makes the filter less improved compared to the corresponding smoother during the later stage of forecasts, e.g., the difference between

3D_WL40_RI5 and 3D_WL40_R20 versus the difference between 4D_WL40_RI5 and 4D_WL40_RI20 at 0040 UTC (Fig. 14d). The changes in the comparison between the filter and the smoother as the RI reduces likely reflect the negative impact of model imbalance on the filter surpassing the negative impact of enhanced nonlinearity on the smoother.

4. Summary and conclusions

Both the four-dimensional (4D) and three-dimensional (3D) ensemble data assimilation (DA) methods in the context of a smoother and a filter have the capability of accounting for the temporal evolution of flow-dependent background-error covariances, which is necessary to initialize convective-scale numerical weather predictions (NWPs) for rapidly changing storms. Using the Weather Research and Forecasting model, this study compares the 4D ensemble–variational (EnVar) method and its 3D counterpart in the context of the smoother and the filter for direct radar reflectivity DA. Consistent with the radar observation interval (RI) spanning the DA window length (WL), the temporal resolution of ensemble perturbations for the smoother is identical to the DA interval for the filter. As a first step, a case study is performed for the rapidly evolving event that occurred on 5–6 May 2019. Given that the comparison of the smoother and the filter in a real NWP model introduces more complexities (i.e., nonlinearity, issues of time-independent localization for the smoother, and model imbalance for the filter) than a simple model, two sets of experiments with the varied WL and RI are conducted to consider the impacts from these complexities. Under a 20-min RI when model imbalance has not kicked in, experiments with WLs of 20, 40, 100, and 160 min (e.g., 3D_WL100_RI20 and 4D_WL100_RI20) are designed to explore the impact of WL on the comparison between the filter and the smoother. To investigate the impact of RI on the comparison, a 5-min RI is adopted under the WLs of 20 and 40 min.

For the impact of WL, the filter and the smoother are compared as the WL increases. Detailed comparisons for the final control analysis show that the performance of the smoother and the filter converges when both adopt a 20-min RI and a 20-min WL. As the WL exceeds 20 min, the final control analysis of the filter becomes closer to observations than the smoother. For example, compared to 3D_WL100_RI20, 4D_WL100_RI20 has spurious and weaker analyzed reflectivity, larger upper-level analyzed updraft zones without matching strong radar echoes, and stronger analyzed cold pools collocated with updrafts. Such differences become larger and more pronounced with the increasing WL. For the forecasts of experiments with a WL over 20 min, the spurious reflectivity remains in the smoother. The warmer near-surface temperature in the control forecast of the filter agrees better with nearby surface observations than the smoother. The enhanced cold pool resulted from spurious updrafts in the smoother promotes storm maintenance and overpredicted reflectivity. Objective evaluations including performance diagrams and FSSs of composite reflectivity confirm the inferior forecast skill of the smoother relative to the filter when the WL exceeds 20 min. The gap of the forecast skill between the smoother and the filter gradually widens as the WL increases.

Given the differences in control forecasts between the filter and the smoother with various WLs are directly tied to differences in the final control analyses, diagnostics were performed to investigate potential DA-related explanations for the worse analyses of the smoother compared to the filter. Diagnostics included Kullback–Leibler divergence as a statistical measure of non-Gaussianity tied to areas of nonlinearity and correlations evolving within the localization radius. These diagnostics revealed that the extension of the WL not only increases the degree of nonlinearity that violates the smoother's linear assumption for the evolution of background-error covariances but also triggers degradations related to using a time-independent localization. Such issues become increasingly severe with longer WLs.

When shortening the RI from 20 to 5 min under 20- and 40-min WLs, the advantages of the filter over the smoother reduce for most forecast hours at thresholds of 30 dBZ and above, leading to the outperformance of the smoother in the early stage of the forecast. Diagnostics of maximum updraft velocity as an indicator of model imbalance suggest that a shorter RI triggers model imbalance for the filter from frequent model interruptions. The time series of RMSE show that a shorter RI further enhances the nonlinearity of the smoother via depicting the error growth in detail. By forecast skill comparison, the negative impact of model imbalance on the filter surpasses the enhanced nonlinearity on the smoother when the RI reduces.

As a first step, this work was performed to understand the impacts of WL and RI on the comparison of a filter and a smoother for analyses and forecasts of rapidly evolving convection using the EnVar method in a case study. Such a filter versus smoother comparison can also be done within other types of DA algorithms, such as an EnKF approach without the variational component. This study found that a smoother with a high temporal resolution of ensemble perturbations shows advantages over a filter. Systematic experiments with more cases should be warranted in the future to identify optimal configurations of WL and RI for the filter and the smoother. Results of this study further motivate the investigation of potential improvements to the smoother by incorporating time-dependent localization, especially for longer WLs.

Acknowledgments. This research was supported by NA19OAR4590231. Computing resources for data assimilation and forecasts were provided by the Extreme Science and Engineering Discovery Environment (XSEDE). Computational verification and post plotting were performed at the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma (OU). Dr. Yongming Wang is acknowledged for providing the GSI-based EnVar DA system with the direct radar reflectivity assimilation capability. We appreciate Dr. Nicholas A. Gasperoni for proofreading the manuscript. The authors also thank the editor and three anonymous peer reviewers for their helpful comments that substantially improved the quality of this manuscript.

Data availability statement. The initial and lateral boundary conditions in this study are provided by NOAA/NCEI (2003, 2017) and NOAA/NCEP (2014). The radar reflectivity data used for assimilation and verification can be accessed from the Multi-Radar/Multi-Sensor System (NOAA/NCEP 2015). The GSI-EnVar/EnKF (Shao et al. 2016) developed by the Developmental Testbed Center is used to perform data assimilation and available at https://dtcenter.org/community-code/gridpoint-statistical-interpolation-gsi. Version 3.9 of the WRF-ARW Model (Skamarock et al. 2008) used for forecasts is available under a Creative Commons Attribution-Noncommercial 4.0 International License at https://www2.mmm.ucar.edu/wrf/users/download/get_sources.html.

REFERENCES

Alexander, C., and Coauthors, 2020: Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) model development. *26th Conf. on Numerical Weather Prediction (NWP)/30th Conf. on Weather Analysis and Forecasting (WAF)*, Boston, MA, Amer. Meteor. Soc., 8A.1, https://ams.confex.com/ams/2020Annual/webprogram/Paper370205.html.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99–111, https://doi.org/10.1016/j.physd.2006.02.011.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371, https://doi.org/10.1175/MWR-D-11-00013.1.

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic–terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494, https://doi.org/10.1175/1520-0493(2004)132<0473:MWPWTR>2.0.CO;2.

——, 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.

Bennett, A. F., and W. P. Budgell, 1989: The Kalman smoother for a linear quasi-geostrophic model of ocean circulation. *Dyn. Atmos. Oceans*, **13**, 219–267, https://doi.org/10.1016/0377-0265(89)90041-9.

Bocquet, M., 2016: Localization and the iterative ensemble Kalman smoother. *Quart. J. Roy. Meteor. Soc.*, **142**, 1075–1089, https://doi.org/10.1002/qj.2711.

Cosme, E., J. Verron, P. Brasseur, J. Blum, and D. Auroux, 2012: Smoothing problems in a Bayesian framework and their linear Gaussian solutions. *Mon. Wea. Rev.*, **140**, 683–695, https://doi.org/10.1175/MWR-D-10-05025.1.

Davis, B., X. Wang, and X. Lu, 2021: A comparison of HWRF six-hourly 4DEnVar and hourly 3DEnVar assimilation of inner core tail Doppler radar observations for the prediction of Hurricane Edouard (2014). *Atmosphere*, **12**, 942, https://doi.org/10.3390/atmos12080942.

Degelia, S. K., X. Wang, D. J. Stensrud, and D. D. Turner, 2020: Systematic evaluation of the impact of assimilating a network of ground-based remote sensing profilers for forecasts of nocturnal convection initiation during PECAN. *Mon. Wea. Rev.*, **148**, 4703–4728, https://doi.org/10.1175/MWR-D-20-0118.1.

Doswell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585, https://doi.org/10.1175/1520-0434(1990)005<0576:OSMOSI>2.0.CO;2.

Dowell, D. C., 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.

Du, J., G. DiMego, D. Jovic, B. Ferrier, B. Yang, and B. Zhou, 2015: Short Range Ensemble Forecast (SREF) system at NCEP: Recent development and future transition. *27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., 2A.5, https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273421.html.

Duda, J. D., X. Wang, F. Kong, and M. Xue, 2014: Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. *Mon. Wea. Rev.*, **142**, 2198–2219, https://doi.org/10.1175/MWR-D-13-00297.1.

——, ——, Y. Wang, and J. R. Carley, 2019: Comparing the assimilation of radar reflectivity using the direct GSI-based Ensemble-Variational (EnVar) and indirect cloud analysis methods in convection-allowing forecasts over the continental United States. *Mon. Wea. Rev.*, **147**, 1655–1678, https://doi.org/10.1175/MWR-D-18-0171.1.

Evensen, G., 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1342–1354, https://doi.org/10.1175/1520-0493(1997)125<1342:ADAFSN>2.0.CO;2.

——, 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.*, **54**, 539–560, https://doi.org/10.1007/s10236-004-0099-2.

——, and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867, https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

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.

——, ——, and ——, 2022: Using a cost-effective approach to increase background ensemble member size within the GSI-Based EnVar system for improved radar analyses and forecasts of convective systems. *Mon. Wea. Rev.*, **150**, 667–689, https://doi.org/10.1175/MWR-D-21-0148.1.

——, ——, and ——, 2023: Valid time shifting for an experimental RRFS convection-allowing EnVar data assimilation and forecast system: Description and systematic evaluation in real-time. *Mon. Wea. Rev.*, **151**, 1229–1245, https://doi.org/10.1175/MWR-D-22-0089.1.

Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. J. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 668–688, https://doi.org/10.1175/2010MWR3456.1.

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.

Johnson, A., and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble.

Mon. Wea. Rev., **140**, 3054–3077, https://doi.org/10.1175/MWR-D-11-00356.1.

——, ——, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, https://doi.org/10.1175/MWR-D-14-00345.1.

Kleist, D. T., and K. Ide, 2015a: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, https://doi.org/10.1175/MWR-D-13-00351.1.

——, and ——, 2015b: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, https://doi.org/10.1175/MWR-D-13-00350.1.

Kondo, K., and T. Miyoshi, 2019: Non-Gaussian statistics in global atmospheric dynamics: A study with a 10 240-member ensemble Kalman filter using an intermediate atmospheric general circulation model. *Nonlinear Processes Geophys.*, **26**, 211–225, https://doi.org/10.5194/npg-26-211-2019.

Krishnamurthy, V., 1993: A predictability study of Lorenz's 28-variable model as a dynamical system. *J. Atmos. Sci.*, **50**, 2215–2229, https://doi.org/10.1175/1520-0469(1993)050<2215:APSOLV>2.0.CO;2.

Kullback, S., and R. A. Leibler, 1951: On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86, https://doi.org/10.1214/aoms/1177729694.

Lorenc, A. C., and M. Jardak, 2018: A comparison of hybrid variational data assimilation methods for global NWP. *Quart. J. Roy. Meteor. Soc.*, **144**, 2748–2760, https://doi.org/10.1002/qj.3401.

——, N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea. Rev.*, **143**, 212–229, https://doi.org/10.1175/MWR-D-14-00195.1.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17A**, 321–333, https://doi.org/10.3402/tellusa.v17i3.9076.

Lu, X., X. Wang, M. Tong, and V. Tallapragada, 2017: GSI-based, continuously cycled, dual-resolution hybrid ensemble–variational data assimilation system for HWRF: System description and experiments with Edouard (2014). *Mon. Wea. Rev.*, **145**, 4877–4898, https://doi.org/10.1175/MWR-D-17-0068.1.

Lynch, P., and X. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120**, 1019–1034, https://doi.org/10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level 3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, https://doi.org/10.1007/s10546-005-9030-8.

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

NOAA/NCEI, 2003: Global Forecast System (GFS). NOAA/National Centers for Environmental Information, accessed 11 April 2023, https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast.

——, 2017: Global Ensemble Forecast System (GEFS). NOAA/National Centers for Environmental Information, accessed 11 April 2023, https://www.ncei.noaa.gov/.

NOAA/NCEP, 2014: Short Range Ensemble Forecast (SREF). NOAA/National Centers for Environmental Prediction, accessed 11 April 2023, https://www.nco.ncep.noaa.gov/pmb/products/sref/.

——, 2015: NOAA/NSSL's national mosaic and QPE system, Multi-Radar Multi-Sensor (MRMS) system. NOAA/National Centers for Environmental Prediction, accessed 11 April 2023, https://mrms.ncep.noaa.gov/data.

Olson, J. B., J. S. Kenyon, W. M. Angevine, J. M. Brown, M. Pagowski, and K. Sušelj, 2019: A description of the MYNN-EDMF scheme and the coupling to other components in WRF–ARW. NOAA Tech. Memo. OAR GSD-61, 37 pp., https://doi.org/10.25923/n9wm-be49.

Pan, Y., and M. Wang, 2019: Impact of the assimilation frequency of radar data with the ARPS 3DVar and cloud analysis system on forecasts of a squall line in southern China. *Adv. Atmos. Sci.*, **36**, 160–172, https://doi.org/10.1007/s00376-018-8087-5.

Roberts, B., B. T. Gallo, I. L. Jirak, A. J. Clark, D. C. Dowell, X. Wang, and Y. Wang, 2020: What makes a convection allowing ensemble of opportunity buy us in forecasting thunderstorms? *Wea. Forecasting*, **35**, 2293–2316, https://doi.org/10.1175/WAF-D-20-0069.1.

Roberts, N. M., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169, https://doi.org/10.1002/met.57.

——, 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.

Roberts, R. D., A. R. S. Anderson, E. Nelson, B. G. Brown, J. W. Wilson, M. Pocernich, and T. Saxen, 2012: Impacts of forecaster involvement on convective storm initiation and evolution nowcasting. *Wea. Forecasting*, **27**, 1061–1089, https://doi.org/10.1175/WAF-D-11-00087.1.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Ruiz, J., G.-Y. Lien, K. Kondo, S. Otsuka, and T. Miyoshi, 2021: Reduced non-Gaussianity by 30 s rapid update in convective-scale numerical weather prediction. *Nonlinear Processes Geophys.*, **28**, 615–626, https://doi.org/10.5194/npg-28-615-2021.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.

Shao, H., and Coauthors, 2016: Bridging research to operations transitions: Status and plans of community GSI. *Bull. Amer. Meteor. Soc.*, **97**, 1427–1440, https://doi.org/10.1175/BAMS-D-13-00245.1.

Shen, F., D. Xu, J. Min, Z. Chu, and X. Li, 2020: Assimilation of radar radial velocity data with the WRF Hybrid 4DEnVar system for the prediction of Hurricane Ike (2008). *Atmos. Res.*, **234**, 104771, https://doi.org/10.1016/j.atmosres.2019.104771.

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.

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.

Sun, J., H. Wang, W. Tong, Y. Zhang, C.-Y. Lin, and D. Xu, 2016: Comparison of the impacts of momentum control variables on high-resolution variational data assimilation and

precipitation forecasting. *Mon. Wea. Rev.*, **144**, 149–169, https://doi.org/10.1175/MWR-D-14-00205.1.

Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1.

van Leeuwen, P. J., and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.*, **124**, 2898–2913, https://doi.org/10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.

Wang, S., M. Xue, and J. Min, 2013: A four-Dimensional asynchronous Ensemble Square-Root Filter (4DEnSRF) algorithm and tests with simulated radar data. *Quart. J. Roy. Meteor. Soc.*, **139**, 805–819, https://doi.org/10.1002/qj.1987.

Wang, X., 2010: Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, https://doi.org/10.1175/2010MWR3245.1.

——, and T. Lei, 2014: GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Mon. Wea. Rev.*, **142**, 3303–3325, https://doi.org/10.1175/MWR-D-13-00303.1.

——, D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar based ensemble–variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, https://doi.org/10.1175/MWR-D-12-00141.1.

Wang, Y., and X. Wang, 2017: Direct assimilation of radar reflectivity without tangent linear and adjoint of the nonlinear observation operator in the GSI-based EnVar system: Methodology and experiment with the 8 May 2003 Oklahoma City tornadic supercell. *Mon. Wea. Rev.*, **145**, 1447–1471, https://doi.org/10.1175/MWR-D-16-0231.1.

——, and ——, 2021a: Development of convective-scale static background error covariance within GSI-based hybrid EnVar system for direct radar reflectivity data assimilation. *Mon. Wea. Rev.*, **149**, 2713–2736, https://doi.org/10.1175/MWR-D-20-0215.1.

——, and ——, 2021b: Rapid update with EnVar direct radar reflectivity data assimilation for the NOAA regional convection-allowing NMMB model over the CONUS: System description and initial experiment results. *Atmosphere*, **12**, 1286, https://doi.org/10.3390/atmos12101286.

——, and ——, 2023: A multivariate additive inflation approach to improve storm-scale ensemble-based data assimilation and forecasts: Methodology and experiment with a tornadic supercell. *J. Adv. Model. Earth Syst.*, **15**, e2022MS003307, https://doi.org/10.1029/2022MS003307.

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.

Zhang, S., and Z. Pu, 2019: Numerical simulation of rapid weakening of Hurricane Joaquin with assimilation of high-definition sounding system dropsondes during the tropical cyclone intensity experiment: Comparison of three- and four-dimensional ensemble–variational data assimilation. *Wea. Forecasting*, **34**, 521–538, https://doi.org/10.1175/WAF-D-18-0151.1.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP global ensemble forecast system in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, https://doi.org/10.1175/WAF-D-17-0023.1.