Practical Predictability of Supercells: Exploring Ensemble Forecast Sensitivity to Initial Condition Spread

MONTGOMERY L. FLORA

School of Meteorology, and Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

COREY K. POTVIN

Cooperative Institute for Mesoscale Meteorological Studies, and School of Meteorology, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

LOUIS J. WICKER

NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

ABSTRACT

As convection-allowing ensembles are routinely used to forecast the evolution of severe thunderstorms, developing an understanding of storm-scale predictability is critical. Using a full-physics numerical weather prediction (NWP) framework, the sensitivity of ensemble forecasts of supercells to initial condition (IC) uncertainty is investigated using a perfect model assumption. Three cases are used from the real-time NSSL Experimental Warn-on-Forecast System for Ensembles (NEWS-e) from the 2016 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. The forecast sensitivity to IC uncertainty is assessed by repeating the simulations with the initial ensemble perturbations reduced to 50% and 25% of their original magnitudes. The object-oriented analysis focuses on significant supercell features, including the mid- and low-level mesocyclone, and rainfall. For a comprehensive analysis, supercell location and amplitude predictability of the aforementioned features are evaluated separately.

For all examined features and cases, forecast spread is greatly reduced by halving the IC spread. By reducing the IC spread from 50% to 25% of the original magnitude, forecast spread is still substantially reduced in two of the three cases. The practical predictability limit (PPL), or the lead time beyond which the forecast spread exceeds some prechosen threshold, is case and feature dependent. Comparing to past studies reveals that practical predictability of supercells is substantially improved by initializing once storms are well established in the ensemble analysis.

1. Introduction

With convection-allowing ensembles in operational use, it is critical to understand storm-scale predictability. Beyond assessing predictability limits, predictability studies are also important for evaluating and comparing the impacts of different forecast error sources [e.g., initial condition (IC) uncertainty, coarse IC resolution, and model error]. Understanding the impacts of various error sources can guide our priorities for storm-scale modeling system design. For example, Potvin et al. (2017) found modeled supercells are relatively insensitive to IC resolution, with missing scales <10 km regenerating in 10–20 min. These results promote using a dual-resolution ensemble, in which analyses are generated on a coarse grid and then downscaled onto a finer grid for forecast initialization. In this study, we choose to investigate the sensitivity of ensemble forecasts of supercells to IC spread.

Traditionally, atmospheric predictability is divided into two domains: *intrinsic* and *practical*. *Intrinsic predictability* is defined as the extent to which prediction is possible given an optimal procedure and infinitesimal IC errors, while *practical predictability* is the extent to which prediction is possible given the best-known procedures and contemporary IC errors or those expected in the foreseeable future (Lorenz 1969, 1996; Melhauser and Zhang 2012). Unlike intrinsic predictability,

Corresponding author: Montgomery L. Flora, monte.flora@noaa.gov

practical predictability is largely determined by current (or future) observing networks and modeling systems (which are highly imperfect). First proposed in Lorenz (1969), the practical predictability limit (PPL) is defined as "the time interval within which the errors in prediction do not exceed some prechosen magnitude."

Since storm-scale numerical weather prediction (NWP) was first proposed (e.g., Lilly 1990), several studies have explored the sensitivity of explicit forecasts of severe thunderstorms to IC uncertainty (McPherson and Droegemeier 1991; Droegemeier and Levit 1993; Wandishin et al. 2008, 2010; Potvin and Wicker 2013; Cintineo and Stensrud 2013 (hereafter CS13); Durran and Weyn 2016; Zhang et al. 2015, 2016; Miglietta et al. 2016, 2017; Weyn and Durran 2017). Early studies by McPherson and Droegemeier (1991) and Droegemeier and Levit (1993) evaluated the sensitivity of idealized simulations of supercells and supercells versus multicellular thunderstorms, respectively, to the characteristics of the thermal bubble used for storm initiation. More recently, Potvin and Wicker (2013) used an idealized observing system simulation experiment (OSSE) framework to perform ensemble Kalman filter (EnKF) radar data assimilation and prediction of supercells. They found that useful probabilistic guidance of low-level rotation, a proxy for tornado potential, is possible out to at least 30–60 min. CS13 compared Rapid Update Cycle (RUC) model forecast soundings to observed soundings to estimate mesoscale analysis error. These errors were then used to construct an ensemble of soundings that were used to initialize an ensemble of simulations. This ensemble allowed examination of the sensitivity of supercell forecasts to contemporary IC uncertainty. They found storm location based on the 40-dBZ contour is more predictable than midlevel mesocyclone location, while 5-min heavy rainfall location and cold pool area are virtually unpredictable. Zhang et al. (2015, 2016) explored the practical and intrinsic predictability of the 20 May 2013 tornadic supercells. Zhang et al. (2015) found the timing of convection initiation (CI) is strongly modulated by the planetary boundary layer (PBL) evolution, which is sensitive to the initialization time and local topography. Using an ensemble generated from small-magnitude perturbations, Zhang et al. (2016) found the intrinsic predictability limit of the event to be 3–6 h. Exploring the predictability of a supercell that formed over northern Italy, Miglietta et al. (2016) found that due to strong orographic effects, large-scale forcing and initialization time had a substantial impact on rainfall, while the results were fairly insensitive to choice of PBL parameterization. Miglietta et al. (2017) showed that in the Mediterranean, predictability can be

extended when topographic features are responsible for the triggering of supercells.

There are three general limitations of previous supercell predictability studies that can be improved upon to increase the scientific and operational relevance of the diagnosed PPLs. First, ensembles in past studies were generated by shifting the initialization time (e.g., Zhang et al. 2015; Miglietta et al. 2016) or introducing quasi-random perturbations (e.g., CS13) onto a control state. None of these studies, however, featured fully flow-dependent IC errors across all scales. Since storm predictability is governed by both upscale growth of intrastorm errors (e.g., Zhang et al. 2007, 2016) and downscale growth of larger-scale errors (e.g., Durran and Gingrich 2014; Durran and Weyn 2016), the IC perturbations used in this study are obtained from storm-scale ensemble analyses generated by an EnKF on a mesoscale domain. Second, past studies have primarily focused on initialization prior to CI. In this study, however, we are working within the Warn-on-Forecast (WoF) paradigm (e.g., Stensrud et al. 2009, 2013), where forecasts are only seriously considered once storms have existed long enough to have been well assimilated. Our study, therefore, assesses the predictability of storms already established in the EnKF analyses. Assimilating real radar, satellite, and surface data to initialize our ensemble forecasts also provides more realistic simulations than the idealized frameworks of some previous studies (e.g., CS13; Potvin and Wicker 2013). Third, investigations of storm predictability ought to account for the discrete, object-like nature of storm-scale phenomena. Traditional root-mean-square differences (RMSDs) calculated between two nearly identical forecasts can be heavily penalized by operationally insignificant phase errors. Focusing on continuous fields (e.g., of temperature or wind; Zhang et al. 2015, 2016) rather than on storm features (e.g., midlevel mesocyclone) also limits the applicability of the results to operational forecasting and to advancing conceptual understanding of storm predictability. This is especially true when errors are computed over domains dominated by storm-free regions, since the much slower error growth in the latter mutes the signal of the intrastorm error growth in the domain-wide calculations. Thus, using an approach similar to CS13, this study will evaluate the practical predictability of individual supercell features.

The rest of the paper is organized as follows. Section 2 discusses our model configuration, the process of reducing IC spread in a preexisting ensemble, quantification of the PPL, and the three cases used. The practical predictability of supercells given current and reduced magnitudes of IC uncertainty that will potentially be achieved in the future is explored in section 3. Section 4
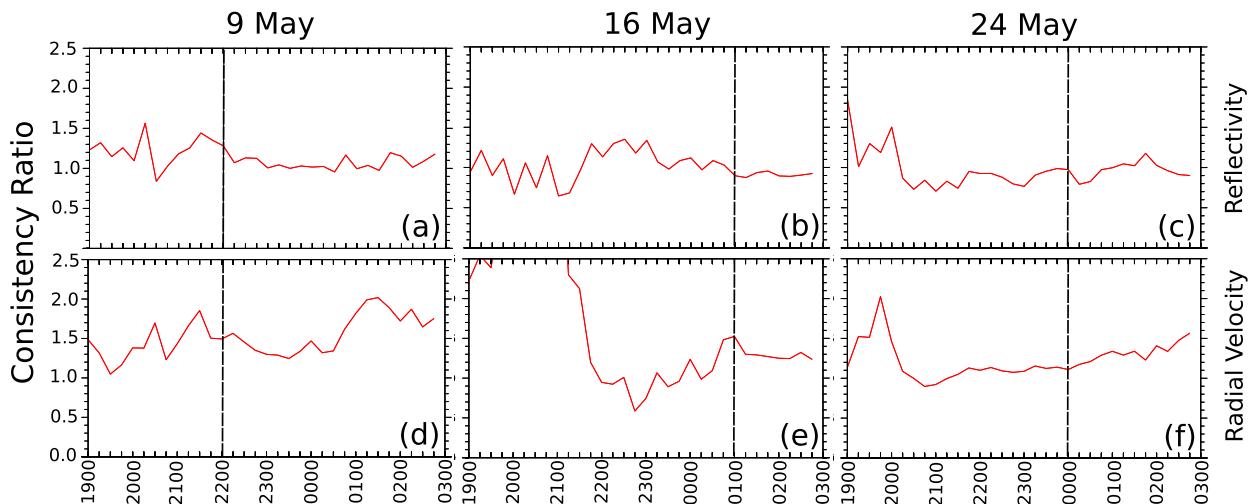
FIG. 1. Consistency ratios (calculated using ensemble priors) during the radar data assimilation period for (a),(d) 9; (b),(e) 16; and (c),(f) 24 May. The top and bottom rows are for radar reflectivity and radial velocity, respectively. Vertical dashed lines indicate the initialization time used in this study for all three cases.

presents a summary of the results, as well as limitations of the study and recommendations for future work.

2. Methods

a. NEWS-e

The ICs and lateral boundary conditions (LBCs) for our simulations are from the NSSL Experimental WoF System for ensembles (NEWS-e; Wheatley et al. 2015; Jones et al. 2016) analyses, generated in real time during the 2016 Hazardous Weather Testbed Spring Forecasting Experiment. The NEWS-e is a high-spatiotemporal-resolution ensemble data assimilation and prediction system nested within the experimental 3-km High-Resolution Rapid Refresh Ensemble (HRRRE; Dowell et al. 2016). NEWS-e consists of 36 WRF-ARW (Skamarock et al. 2008) ensemble members with physical parameterization diversity [see Wheatley et al.'s (2015) Table 2 for PBL and radiation schemes used] run over a 250×250 gridpoint domain with a 3-km horizontal grid spacing. The domain is daily recentered on the region of greatest severe weather potential. The NEWS-e is initialized daily at 1800 UTC with ICs and LBCs provided by the HRRRE. After initialization, radar, satellite, mesonet (when available), and other conventional observations are assimilated every 15 min using the ensemble adjustment Kalman filter (Anderson 2001) included in the Data Assimilation Research Testbed (DART) software.

To evaluate the performance of the radar data assimilation in the three cases used (see below), we used the consistency ratio (Dowell et al. 2004), which is the ratio of the sum of the prescribed observation error variance and ensemble forecast error variance to the ensemble forecast root-mean-square innovation (RMSI). Consistency ratios near unity suggest nearly optimal ensemble spread. In all three cases, the consistency ratio for radar reflectivity is around 1.0 ± 0.4 (Figs. 1a–c), similar to results in Wheatley et al. (2015). For radial velocity (Figs. 1d–f), the values are slightly higher, falling in a similar range as found in Wheatley et al. (2015). The radial velocity consistency ratio for 16 May is near 2.5 for the first several assimilation cycles, presumably due to the lack of radar observations in the NEWS-e domain. However, as soon as storms enter the domain around 2200 UTC and the number of observations increases, the consistency ratio is dramatically reduced. The values of the consistency ratio suggest reasonable ensemble spread is obtained by the initialization times used in our experiments.

A common problem for storm-scale data assimilation is ensemble underdispersion. Figure 2 shows vertical profiles of initial spread in the 3-km ensemble analyses for the three cases. The RUC analyses errors in CS13 (cf. 1-h RUC errors in their Fig. 1) are comparable for relative humidity, but horizontal wind spread is noticeably larger. This is probably due to the much greater influence of storms (and attendant much larger U and V) in the NEWS-e versus RUC. Temperature spread is smaller in the NEWS-e, but it is conceivable that the desired initial temperature spread has been reduced due to inclusion of mesonet, radar, and satellite observations in the NEWS-e system. Ultimately, we argue that the NEWS-e spread in these cases is reasonable, as indicated by the radar consistency ratios and loose consistency with the 1-h RUC errors.
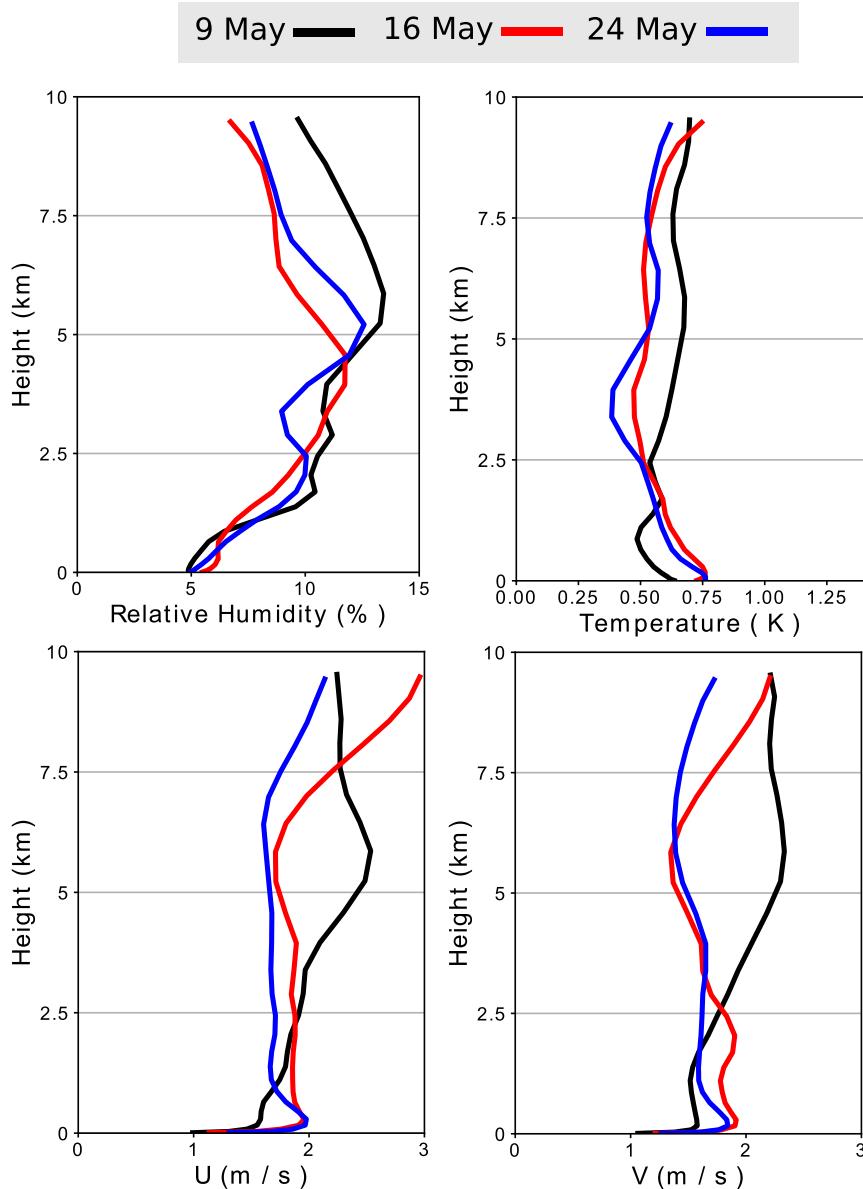
FIG. 2. Vertical profiles of level-average initial ensemble spread in relative humidity (%), temperature (K), and horizontal wind components (m s^{-1}) U and V for 9 (black), 16 (red), and 24 May (blue) on the 3-km grid.

b. Model configuration

The numerical model used is the WRF-ARW version 3.6.1. We use one-way nested domains of 3- and 1-km grid spacing with 51 vertical levels, 11 of which are in the lowest 2 km above ground level (AGL). The 3-km grid has 250 × 250 grid points, while the 1-km domain size is case dependent (Table 1). A 1-km grid spacing has been used in past supercell predictability studies (e.g., CS13; Zhang et al. 2015, 2016; Miglietta et al. 2016) and is a reasonable compromise between available computing resources and the ability of the model to represent essential physical processes for supercell prediction, such as the mid- and low-level mesocyclone (e.g., Potvin and Flora 2015). To isolate the sensitivity of ensemble forecast spread to IC uncertainty, a perfect model assumption is adopted, which requires a single set of physics be used in our simulations. We use the following parameterization schemes: Thompson microphysics (Thompson et al. 2008), MM5 similarity surface layer (Zhang and Anthes 1982), RUC land

TABLE 1. Number of grid points in the north–south (*NY*) and west–east (*NX*) directions for the 1-km grid.

	NY	*NX*
9 May	196	226
16 May	211	391
24 May	220	253

surface model (Smirnova et al. 1997, 2000), Mellor–Yamada–Nakanishi–Niino (MYNN) level-3.0 PBL (Nakanishi and Niino 2006), RRTM longwave radiation (Mlawer et al. 1997), and Dudhia shortwave radiation (Dudhia 1989). Our simulations are integrated for 3 h with the initialization time differing from case to case with model output every 5 min. The WRF Model

prognostic variables include the three wind components (*u*, *v*, and *w*), perturbation potential temperature, perturbation geopotential height, perturbation surface pressure of dry air, and (from the Thompson scheme) mixing ratios of water vapor, ice, rain, graupel, snow, and cloud water and number concentration of cloud ice and rain.

c. Reducing initial condition spread

Reducing IC spread in a preexisting ensemble requires collapsing each ensemble member toward a single deterministic state. The ensemble mean, which is supposed to be the best estimate of the true state, is an obvious option for this deterministic state. Unfortunately, the ensemble mean tends to be unrealistically smooth, especially within storms, due to phase
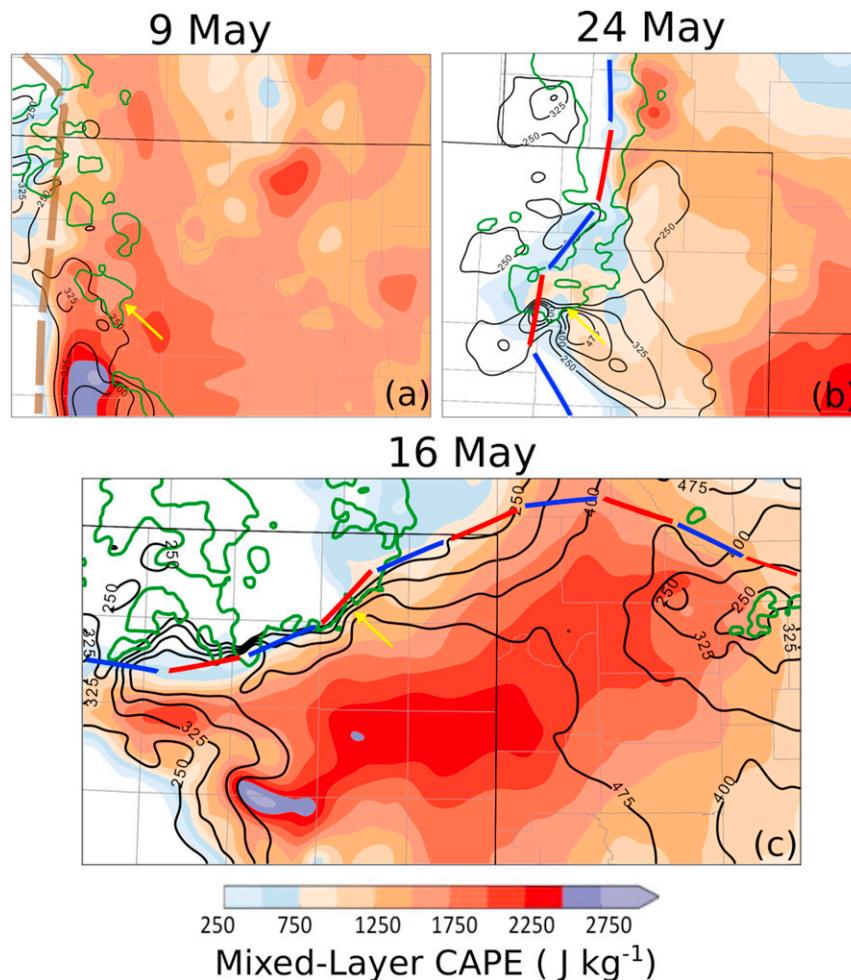

FIG. 3. Ensemble-mean mixed layer (lowest 75 hPa) CAPE (filled contour), ensemble-mean 0–3-km SRH (m² s⁻²; black contours), and the control member 30-dB*Z* reflectivity at 1.5 km AGL (green contours) are shown for (a) 9, (b) 24, and (c) 16 May. The yellow arrows denote the subjectively defined storm inflow region for the primary storm of interest for the three cases. Approximate dryline and stationary front position shown by dashed brown curve and red and blue dashed curves, respectively.

TABLE 2. Various bulk parameters qualitatively derived from a subjectively defined storm inflow region on the 3-km grid.

	9 May	16 May	24 May
CAPE (J kg^{-1})	1800	1500–1800	1000–1500
CIN (J kg^{-1})	−75	−200	−100
0–1-km shear (m s^{-1})	10	11–14	11–14
0–6-km shear (m s^{-1})	<20	36	36–39
0–1-km SRH (m^2 s^{-2})	120	200–250	200
SFC T_d (°F)	60–64	56–64	45–48

differences among the ensemble members. Rather than use the ensemble mean, we selected a control member from the 36-member ensemble that met the following criteria, listed from highest to lowest priority:

1) Small deviation of environment from initial ensemble-mean temperature, 3D wind, and water vapor mixing ratio.
2) Early storm evolution (e.g., first 30 min) close to ensemble-mean evolution.
3) Storm remains relatively isolated and survives through the end of the 3-h simulation.

The member that best matched the observations is not automatically chosen as the control simulation for two reasons. First, closely replicating the observed evolution of a particular event is not necessary for a general investigation of predictability. Second, the ensemble forecasts deviated from the observations in all three cases, an expected consequence of biases in the NEWS-e analyses and the model. Since the control member must remain close to the ensemble mean to ensure the remaining members do not become unduly offset from the control (which would impede interpretation of the results), minimizing the deviation of the control from the ensemble (not observations) was the higher priority. To evaluate the performance of the ensemble forecasts, the control member evolution is taken as truth.

To generate the ensemble perturbations, the control member state variables fields are subtracted from each ensemble member state over the entire 3D domain. The resulting perturbations are then reduced to 50% and 25% of their original magnitudes. The reduced perturbations are next added to the control member state to generate new ensembles with reduced IC spread. Finally, each reduced-spread ensemble analysis is downscaled to the 1-km grid. To clarify, we are simulating the effect of reducing current storm-scale IC uncertainty through improvement in the observation network and/or model, and not an artificial reduction in IC spread with no corresponding improvements in ensemble data assimilation. For example, the 50% experiments represent the forecast spread evolution that would occur if contemporary storm-scale IC spread were halved.
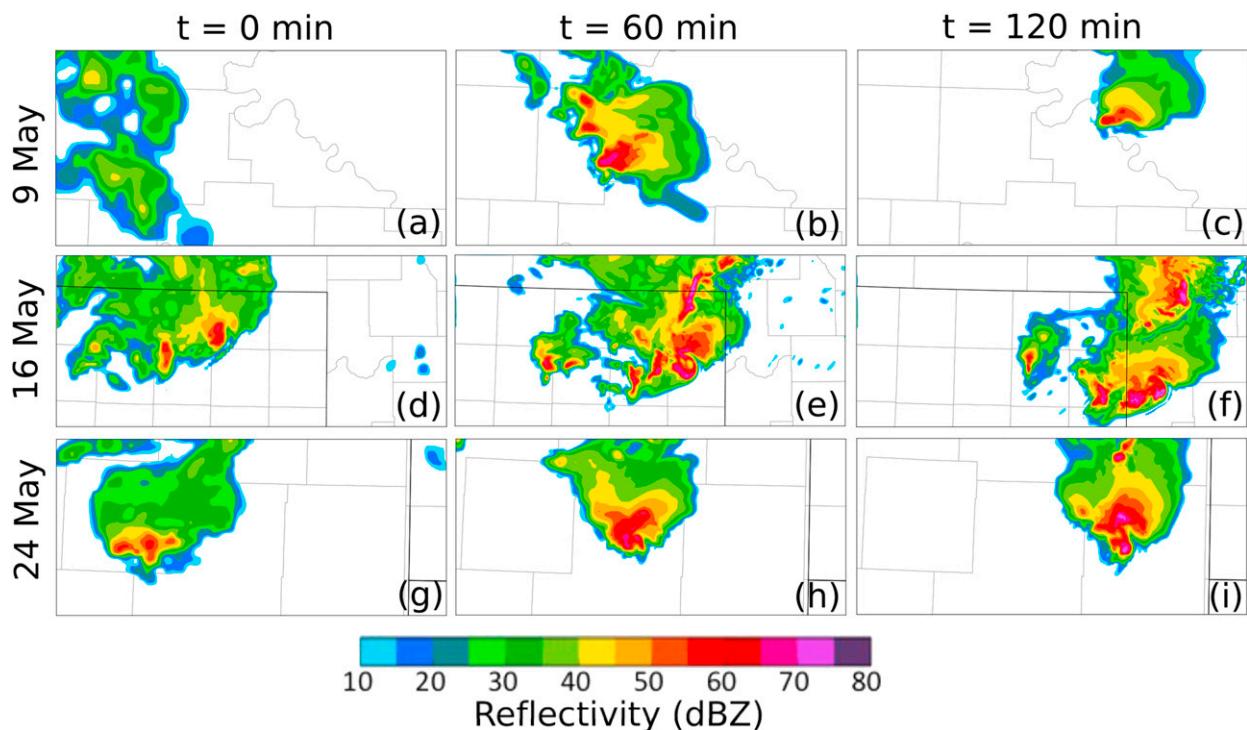

FIG. 4. Control member reflectivity (dBZ) at 1.5 km AGL at $t = 0$, 60, and 120 min for (a)–(c) 9, (d)–(f) 16, and (g)–(i) 24 May.
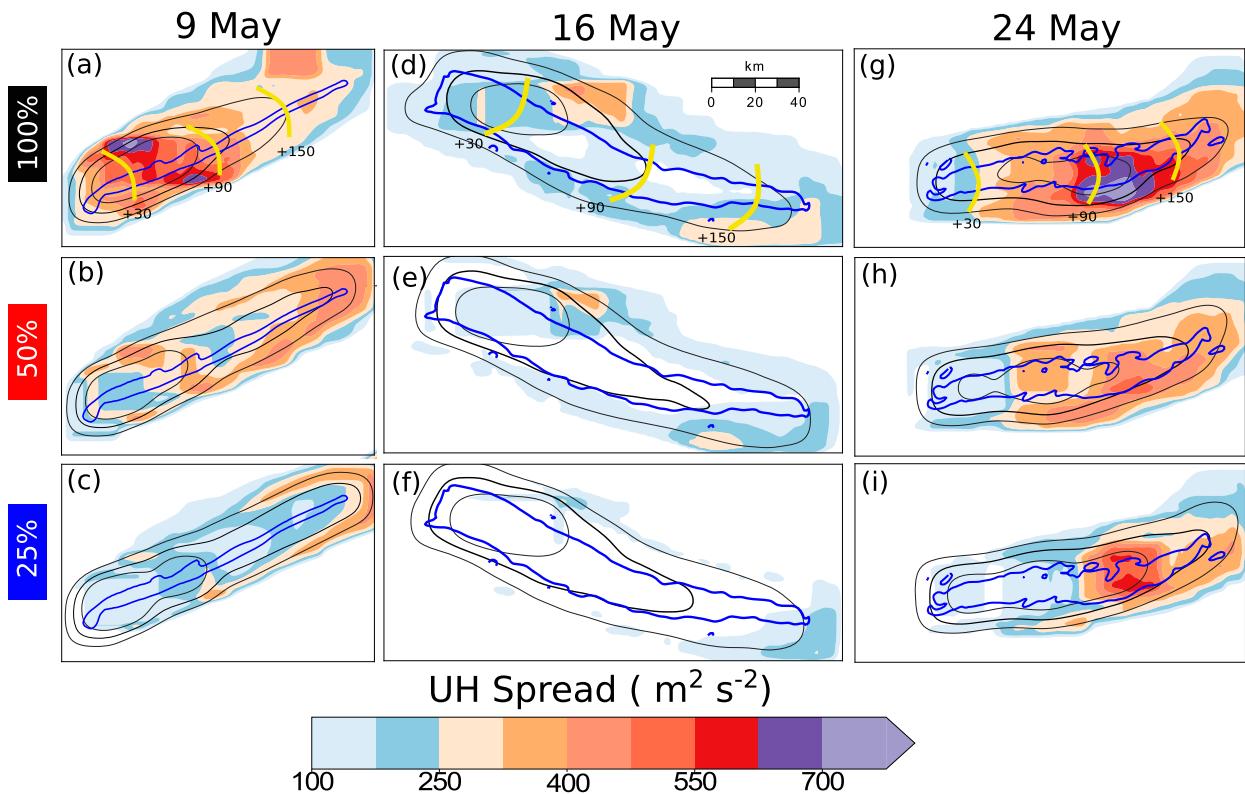
FIG. 5. The 2–5-km AGL UH spread swaths maximized in a 20-km-radius neighborhood (filled contours) with probability-matched mean contoured (black lines; thick line = $700 \, \text{m}^2 \, \text{s}^{-2}$ with a contour interval of $300 \, \text{m}^2 \, \text{s}^{-2}$) for (a)–(c) 9, (d)–(f) 16, and (g)–(i) 24 May for the (a),(d),(g) 100%, (b),(e),(h) 50%, and (c),(f),(i) 25% experiments. Yellow lines show approximate timing of ensemble every 60 min after initial 30 min. The $300 \, \text{m}^2 \, \text{s}^{-2}$ control member isoline is overlaid in blue for each case. Map scale shown in (d) is valid for all plots.

Therefore, the risk of forecast underdispersion does not increase as IC spread is decreased in our experiments.

d. Measuring predictability

Since Lorenz (1969) initially proposed a definition for the PPL (see section 1), no universal quantitative definition has been established, which is unsurprising, given that the optimal choice of error threshold is application dependent. Therefore, we have adopted PPL criteria from previous studies, as well as novel criteria that are particularly appropriate for severe thunderstorm forecasting. The two main PPL criteria are based on ensemble spread and probability, while an additional criterion, ensemble bias, is used primarily to evaluate causes of limited practical predictability.

Traditionally, in an ensemble framework, predictability is evaluated by measuring evolution of ensemble spread, where rapid growth of spread is associated with poor predictability. The two most common metrics, RMSD and standard deviation, both measure deviation from some defined state. In the case of standard deviation, the state is the ensemble mean. In the case of the RMSD, this state is more generally defined; in our application, it is the control member. Although the ensemble mean and control member are initially similar (as required by our control member selection process), they diverge substantially during their forecasts in some experiments, causing RMSD to be heavily influenced by ensemble bias (relative to the control member), which is undesirable since we wish to separately measure spread and bias. Given that storm-scale ensemble perturbations can be highly non-Gaussian, we considered that standard deviation may likewise not be a robust measure of spread. However, for all storm attributes examined in this study, the ensemble perturbations are sufficiently Gaussian for the standard deviation to generally well represent the ensemble spread. Therefore, we adopt standard deviation from the ensemble mean as our spread metric. Ensemble spread, however, was not evaluated in whole, but separated into two components: phase and amplitude spread. For this study, phase spread represents the degree of storm location uncertainty, defined by spread in location of maximum UH, while amplitude spread represents the degree of uncertainty in the maximum value of a storm variable. This allowed us to separately consider the predictability limits of the storm location and the amplitudes of selected features.
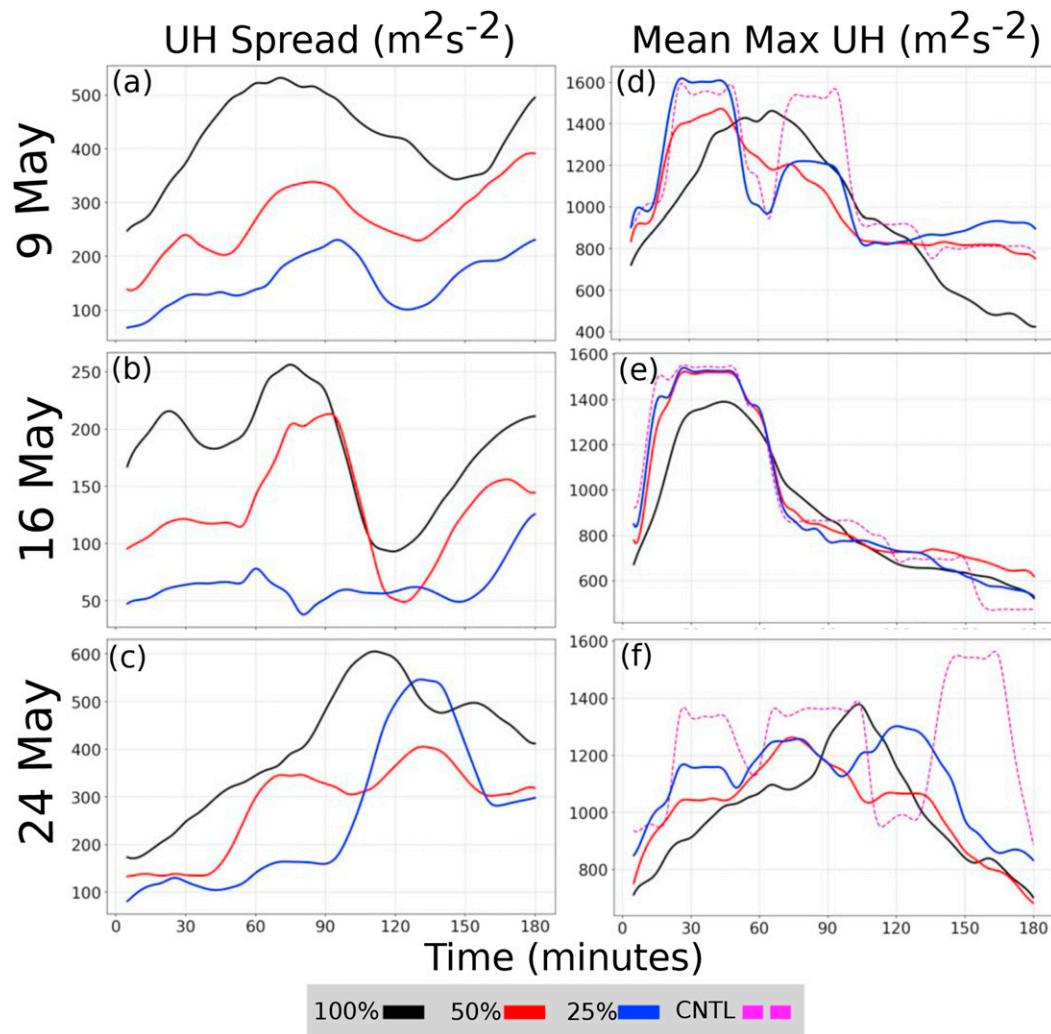
FIG. 6. (left) Time series of the ensemble standard deviation in maximum UH within a 20-min window for (a) 9, (b) 16, and (c) 24 May. (right) Ensemble-mean maximum UH within a 20-min window for (d) 9, (e) 16, and (f) 24 May (no additional time smoothing). The 100% ensembles, 50% ensembles, 25% ensembles, and control member time series are black, red, blue, and magenta, respectively.

Although ensemble spread is traditionally used to evaluate predictability in studies, an additional useful metric for forecast operations is the ensemble probability of exceedance for a storm attribute. In general, the probability of exceedance is the number of ensemble members exceeding some threshold divided by the total number of ensemble members. In CS13, loss of practical predictability is considered to occur once domain-maximum probability of exceedance falls below 60%. We adopt a similar definition for the PPL in this study, except the probability exceedance threshold is varied among 30%, 50%, and 70%. However, unlike CS13, the probability of exceedance in this study is calculated within a 3-km-radius neighborhood to mitigate the effects of operationally tolerable storm location errors.

This type of ensemble probability is referred to as the neighborhood maximum ensemble probability (NMEP; Schwartz and Sobash 2017).

Finally, a novel criterion for evaluating practical predictability in this study is bias. For this study, bias refers to the difference between the control member and ensemble mean for a given variable. If the magnitude of the forecast bias substantially increases as the IC perturbations are increased, then this may indicate poor practical predictability. As will be shown, contemporary analysis uncertainty can lead to premature storm demise in some experiments (a bifurcation in the ensemble forecast), which introduces bias and greatly limits the practical predictability.
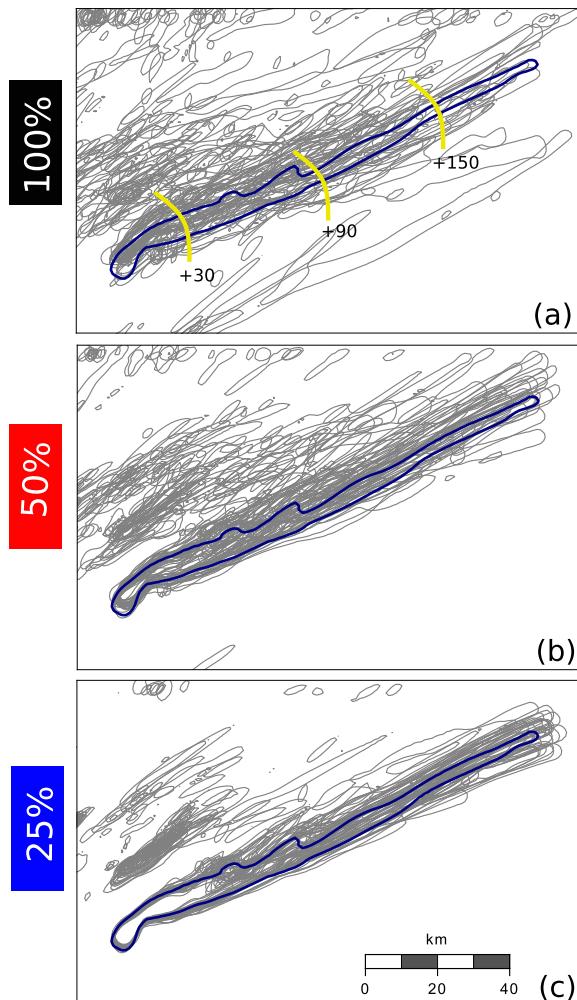
UH = 300 m²s⁻² isolines

FIG. 7. The $300\,m^2\,s^{-2}$ UH isolines (gray) for the 9 May (a) 100%, (b) 50%, and (c) 25% experiments with the control member overlaid in blue. Yellow lines show approximate timing of ensemble every 60 min after initial 30 min.

e. Three supercell cases

Anticipating that the predictability of supercell evolution is case dependent, we performed experiments for three different 2016 events using NEWS-e analyses valid at 2200 UTC 9 May, 0100 UTC 17 May, and 0000 UTC 25 May. Using the local time (CST) to identify the date, these cases are hereafter labeled 9 May, 16 May, and 24 May, respectively. The thermodynamic and kinematic characteristics vary substantially across the three cases (Fig. 3 and Table 2). The 9 May environment (Fig. 3a) is moderately unstable, with $1800\,J\,kg^{-1}$ of ensemble-mean mixed layer CAPE with the weakest low-level wind profile of the three cases. The 16 May environment (Fig. 3c) is also moderately unstable, but

the 0–3-km storm-relative helicity (SRH) is significantly stronger than in the 9 May case with substantial deep layer shear as well. Finally, the 24 May environment (Fig. 3b) is marginally to moderately unstable (e.g., $1000\,J\,kg^{-1}$ of CAPE), with a relatively dry boundary layer (e.g., surface dewpoint temperature between 45° and 48°F), substantial low- and deep-level wind shear, and strong 0–3-km SRH.

The evolution of 1.5 km AGL reflectivity for all three control simulations is presented in Fig. 4. In all three cases, supercells are present in the initial conditions. The 9 May simulation (Figs. 4a–c) features a right-moving supercell initially surrounded by secondary, weaker storms, but eventually becoming isolated. In the 16 May case (Figs. 4d–f), there is a large heavy precipitation (HP) supercell that maintains many supercell characteristics, although it is steadily transitioning upscale. For example, there is a persistent hook echo signature, which is masked at times by convection initiated off the gust front. Finally, in the 24 May case, there is a relatively isolated right-moving supercell with a strong hook echo signature (Figs. 4g–i).

3. Results

a. Updraft helicity

The feature that best distinguishes supercells from other convective modes is the deep, quasi-steady rotating updraft known as the midlevel mesocyclone. Midlevel updraft helicity (UH), the most common parameter for detecting midlevel mesocyclones and characterizing their intensity, is defined as

$$UH = \int_{z_b}^{z_t} w\zeta\,dz, \qquad (1)$$

where w is vertical velocity ($m\,s^{-1}$), ζ is vertical vorticity (s^{-1}), and z_b and z_t are heights AGL, typically set (including in this study) to 2 and 5 km, respectively. Figure 5 shows time-maximum UH amplitude spread (hereafter referred to as spread swaths) with probability-matched mean[1] (Ebert 2001) UH contours overlaid for all three cases. To isolate amplitude spread by eliminating phase errors, maximum UH is computed within a large 20-km-radius neighborhood for each member.

---

[1] To obtain the probability matched mean, forecast values for all n ensemble members for the entire domain are pooled together, ranked from greatest to smallest, then every nth value is extracted to produce an array of ranked ensemble member values. The values of the ensemble-mean forecast are also ranked from greatest to smallest, but with the spatial location of each value stored along with its rank. Finally, the grid point with the highest ensemble-mean value is assigned the highest value from the distribution of ensemble members, and so on.
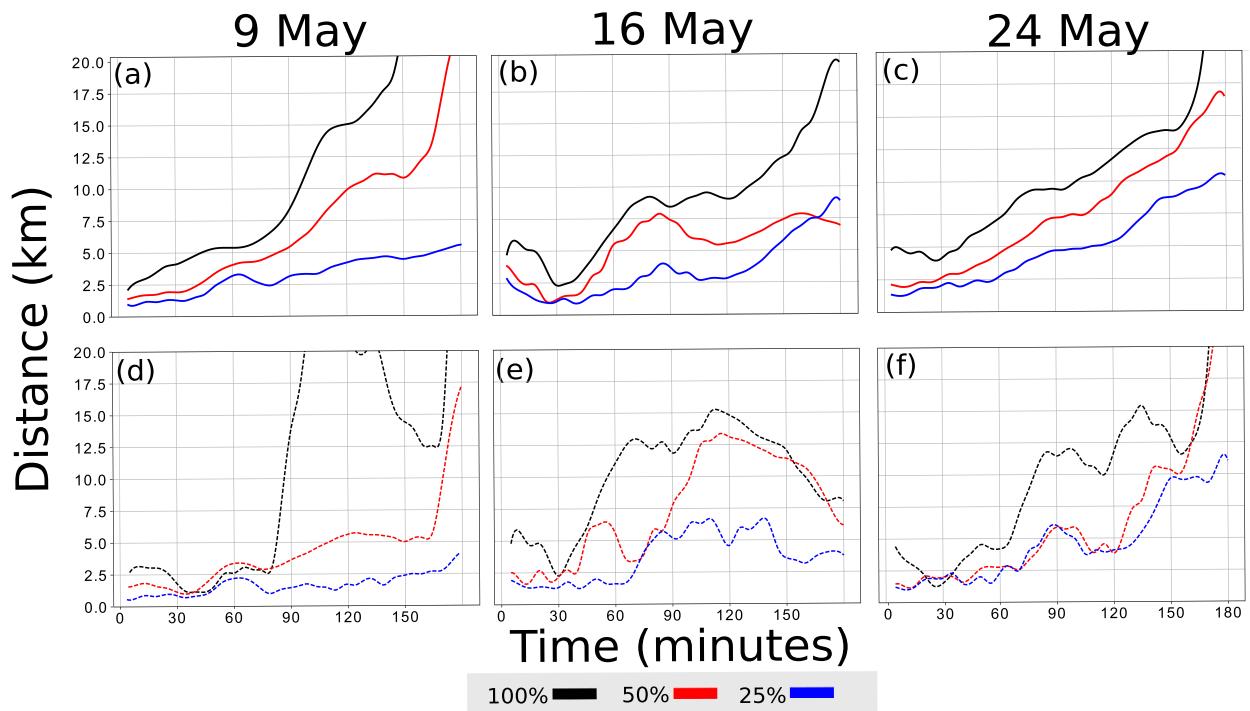
FIG. 8. (top) Time series of the ensemble-average distance from the ensemble-mean storm location (i.e., maximum UH) for (a) 9, (b) 16, and (c) 24 May. (bottom) Time series of the distance between the ensemble mean and control member storm location for (d) 9, (e) 16, and (f) 24 May.

This neighborhood is based on Fig. 8, which shows that 20 km was the maximum ensemble-average displacement from the ensemble-mean storm location for all three cases. The resulting spread swaths are slightly smoothed, and, for illustration purposes, the probability-matched mean contours are heavily smoothed to focus on the general evolution of the ensemble. As we can see in Fig. 5, reducing IC spread greatly reduces the UH amplitude spread in all three cases. In the 9 and 16 May cases (Figs. 5a–c, d–f), the UH amplitude spread continues to benefit from reductions in IC spread, with large decreases between both the 100% and 50% and 50% and 25% experiments. However, in the 24 May case (Figs. 5g–i), the 100%, 50%, and 25% experiments all have similar spread toward the end of the simulation. The diminishing returns in the 24 May case may indicate the intrinsic predictability limit is being approached prior to the end of the simulation. Finally, the UH amplitude spread in the 16 May case was less than half that in the other cases, suggesting it may be the most predictable case. Recall that the 9 and 24 May supercells are fairly discrete and isolated, while the 16 May supercell is steadily organizing on larger scales as it grows upscale. Therefore, the 16 May supercell is likely inheriting the greater predictability of the larger scales, as opposed to the other two cases, which results in a slower forecast spread growth.

Although forecast spread is generally reduced by decreasing the IC spread, forecast spread in some parts of the 9 and 24 May domains did increase. For example, after 150 min in the 9 May case, the spread in the 50% experiment is greater than that in the 100% experiment (cf. Figs. 5a and 5b). This is related to premature storm demise in the 100% experiment, which manifests as a substantially smaller probability-matched mean UH value in the 100% ensemble. Therefore, as the storm lifetime is lengthened in the 50% experiment, compared to the 100% experiment, so is the spread growth time. In the 24 May case, the spread in the 25% experiment (Fig. 5i) is greater than in the 50% experiment (Fig. 5h) after 90 min. In this case, the primary storm of interest is impacted by an upstream secondary storm that forms in its wake. This implies that the amplitude spread in the last 30 min is muddled by the influence of the secondary storm; this is verified upon closer inspection (not shown).

To further examine UH amplitude spread, Fig. 6 shows time series of spread in domain-maximum UH, with the mean domain-maximum UH shown for reference. To avoid the influence of secondary storms, maximum values are only extracted from within a subjectively drawn polygon based on the ensemble members' UH isolines for the primary storm (not shown).
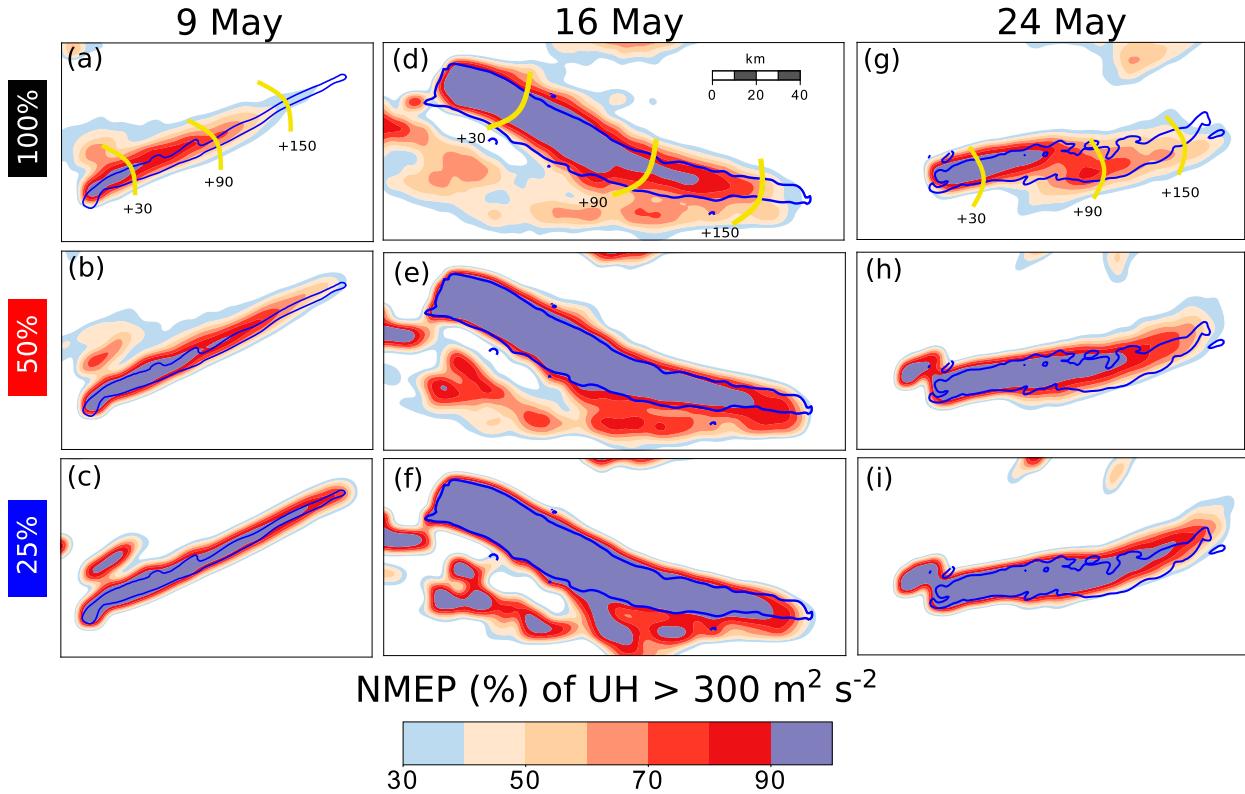
FIG. 9. As in Fig. 5, but for NMEP of UH > 300 m² s⁻² (filled contour) within a 3-km-radius neighborhood.

To limit the impact of relatively minor timing errors and focus on the longer-term spread evolution, values of domain-maximum UH were computed within a 20-min window, and the resulting curves were averaged using the same window. As we can see in Fig. 6, only the 24 May case (Fig. 6c) experiences diminishing returns, which is consistent with spread swaths above. To demonstrate the principle of diagnosing the amplitude PPL, we arbitrarily select an amplitude spread threshold of $UH = 500\,\mathrm{m^2\,s^{-2}}$ then determine the forecast lead time at which this threshold is exceeded in the 100% ensembles. The appropriateness of the $UH = 500\,\mathrm{m^2\,s^{-2}}$ will vary with event and application. However, we consider this a reasonable choice for the purpose of illustration. For example, given a fixed, mean vertical vorticity of $10^{-2}\,\mathrm{s^{-1}}$ and mean vertical velocity of $30\,\mathrm{m\,s^{-1}}$ with a spread of $20\,\mathrm{m\,s^{-1}}$ (which distinguishes between strong and weak storms), the corresponding spread in UH is $600\,\mathrm{m^2\,s^{-2}}$. In the 9 and 24 May cases (Figs. 6a,c), the amplitude PPL is approximately 50 and 90 min, respectively. While the amplitude spread in both cases drops below the $UH = 500\,\mathrm{m^2\,s^{-2}}$ threshold later in the forecast, this is related to the premature storm demise in 9 May and general storm demise in 24 May, which can be seen in the dramatic decrease of the 100%

experiment ensemble-mean domain-maximum UH (Figs. 6d,f). As for the 16 May case, the PPL threshold is never met, so the amplitude PPL limit is beyond 3 h, according to our criteria.
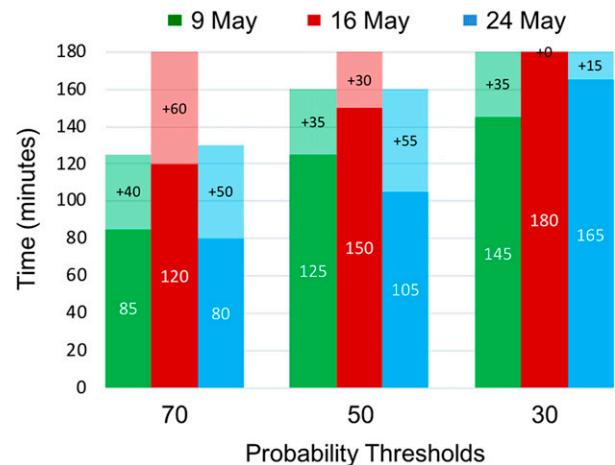

FIG. 10. Times when domain-maximum NMEP of UH > 300 m² s⁻² drops below 70%, 50%, and 30% for 9 (solid green), 16 (solid red), and 24 May (solid blue). Lead time gained with a 50% IC spread reduction shown in translucent colors.
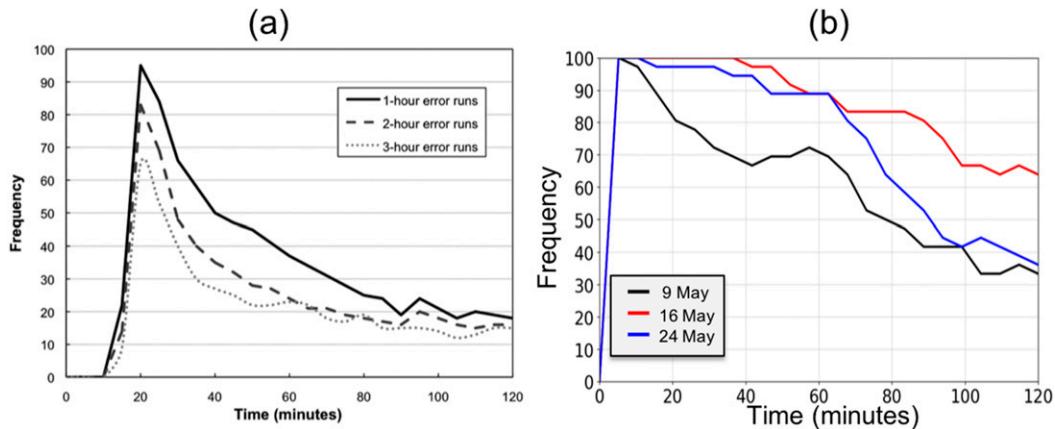
FIG. 11. (a) Figure 14 from CS13; time series of the domain-maximum frequency of UH > 50 m^{2}s^{-2} for all 100 runs perturbed with 1-, 2-, and 3-h environmental forecast errors. (b) Time series of domain-maximum frequency of UH > 50 m^{2}s^{-2} in our 100% experiments (with no neighborhood to be consistent with CS13).

To qualitatively examine UH location spread, Fig. 7 shows 300 m^{2}s^{-2} isolines[2] for the 9 May case. As the IC spread is decreased, the contours collapse toward the control member forecast, indicating location spread and bias are being substantially reduced. The 100% and 50% experiments are not substantially different from each other (cf. Figs. 7a and 7b), but the location spread is greatly reduced in the 25% experiment (Fig. 7c). For the other two cases, the location spread is greatly reduced with a 50% IC spread reduction (not shown).

To quantify the UH location uncertainty, we plotted time series of ensemble-average distance from the ensemble-mean maximum UH location for each case (Fig. 8). In the 100% experiment for all three cases, it takes approximately 90–100 min before the location uncertainty exceeds 10 km. With a 50% IC spread reduction, the time before the location uncertainty exceeds 10 km is extended by 30–40 min in the 9 and 24 May cases, while in the 16 May case, the location uncertainty never exceeds 10 km. Nowcasting techniques, such as extrapolation and Bunkers motion (Bunkers et al. 2000), were used as baselines for supercell location prediction, but due to the unusually large phase errors they produced, the results were not shown. In the case of Bunkers motion, the large phase errors were due to the sensitivity of the calculation near frontal boundaries, which was noted in Bunkers et al. (2000).

Both amplitude and location spread are useful statistics for describing forecast uncertainty, but NMEP can be most easily applied to operations. Figure 9 shows the NMEP of UH > 300 m^{2}s^{-2} in a 3-km-radius neighborhood for all three cases. The neighborhood is employed to limit the impact of tolerable phase errors from reducing NMEP values. In the 9 and 24 May cases (Figs. 9a–c and 9g–i), the NMEP is substantially improved by reducing IC spread. As for the 16 May case (Figs. 9d–f), the NMEP in the 100% ensemble is already very large, leaving little room for improvements in the 50% and 25% experiments. Using the CS13 definition presented in section 2d, but varying the domain-maximum NMEP threshold, Fig. 10 shows PPLs diagnosed from the 100% experiments, as well as the lead time increase with a 50% IC spread reduction. The PPLs for 9 and 24 May are quite similar for all probability thresholds, while the PPLs in 16 May are noticeably longer. With a 50% IC spread reduction, the PPL increased, on average, by 45 min across the three cases and probability thresholds of 70% and 50%, signifying that the PPL would be considerably lengthened if model and observational improvements reduced typical storm-scale analysis uncertainty by half.

To compare our results with CS13, we computed time series of domain-maximum probability of midlevel UH > 50 m^{2}s^{-2} with no neighborhood, allowing us to compare to their Fig. 14 (Fig. 11). The most comparable results are the 1-h error forecasts since our initial spread profiles are similar to their 1-h error profiles (cf. our Fig. 2 with their Fig. 1). As we can see in Fig. 11, the lead time when probabilities fall below 60% is case dependent and ranges between 65 and 135 min. In two of the three cases, the PPL is nearly triple the 1-h error

---

[2] Past studies have used 50 and 180 m^{2}s^{-2} (e.g., CS13; Zhang et al. 2015, 2016) for a 1-km grid. However, assuming vertical velocity and vorticity within the mesocyclone are $O(10)$ m s^{-1} and $O(10^{-2})$ s^{-1}, respectively, between 2 and 5 km AGL, then based on Eq. (1), 300 m^{2}s^{-2} is a fairly appropriate threshold for a mature supercell on a 1-km grid.
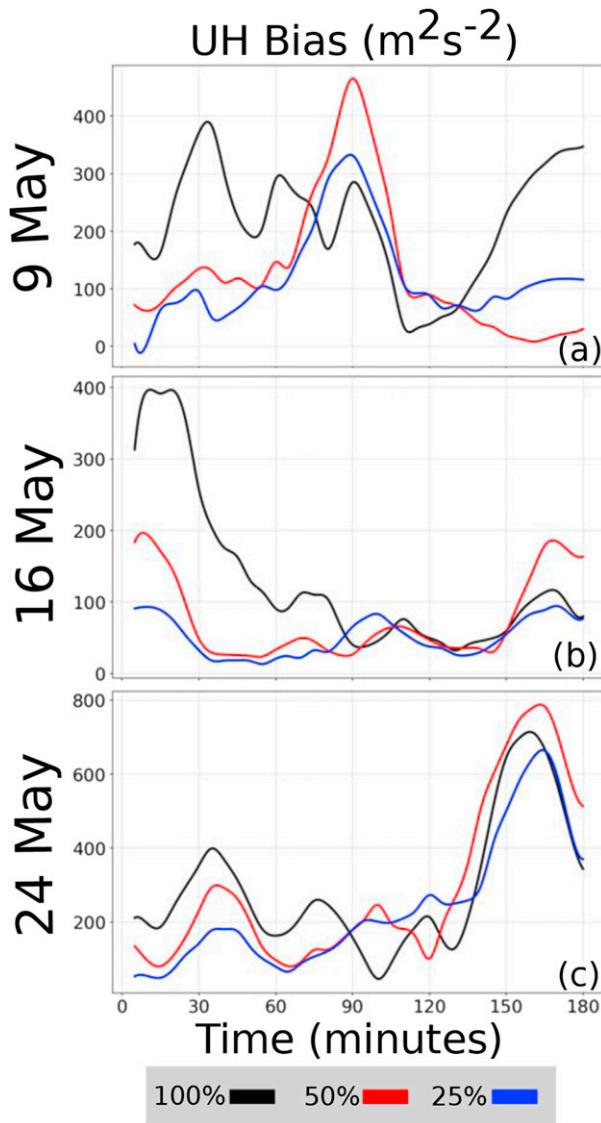
FIG. 12. Time series of the absolute difference between the maximum UH of the ensemble mean and control member for (a) 9, (b) 16, and (c) 24 May. The time series are averaged over a 20-min window to facilitate interpretation.

limit in CS13 (40 min). Given the similarity between the initial spread profiles, if the three case studies presented here are representative, then initializing forecasts post-CI rather than 1 h prior to CI appears to significantly improve supercell practical predictability.

A novel criterion for practical predictability in this study is ensemble bias. Figure 12 shows time series of UH ensemble bias for all three cases. Of the three cases, 9 May (Fig. 12a) has the largest difference in bias between the 100% and 25% experiments, especially in the last hour of the simulation. This is because some ensemble members did not sustain a supercell throughout

the 3-h period. In a sense, 9 May is the least predictable case, since IC perturbations consistent with current analysis uncertainty led to a critical bifurcation in the ensemble. However, with a 50% IC spread reduction, the bias in the 9 May case is substantially reduced after 140 min. The bias is also reduced in the 16 and 24 May cases (Figs. 12b,c), with reductions in IC spread reduction. Toward the end of the forecast period, the bias dramatically increased in some of the ensembles in all three cases, but this is due to interactions of secondary storms with the primary storm of interest. Overall, if current IC spread is reduced by 50%–75% in the future, then both forecast spread and bias will be substantially reduced.

b. Low-level vorticity

One primary role for probabilistic numerical guidance in the severe weather warning process is improving tornado forecasts (Stensrud et al. 2009, 2013). Since operational model resolutions are far from resolving tornadoes, model proxies such as maximum low-level (0–2 km AGL) vorticity (LLV) are used to assess tornado potential (e.g., Potvin and Wicker 2013; Wheatley et al. 2015; Jones et al. 2016; Yussouf et al. 2015, 2016). The utility of LLV as a tornado proxy, particularly at grid spacings fine enough to begin resolving low-level mesocyclones, stems largely from the fact that nearly half of observed low-level mesocyclones produce a tornado (Trapp et al. 2005). Given the importance of LLV forecasts to assessing tornado potential, it is surprising that the predictability of LLV has largely been neglected in past supercell predictability studies.

In the 16 and 24 May cases, the low-level mesocyclones are intense and long-lived (not shown). However, in the 9 May case, the low-level mesocyclone is slightly weaker than in those two cases with two distinct intensification periods. Thus, the 9 May ensemble provides insight into the predictability of supercells with marginal tornadic potential. Figures 13a–c show that LLV forecast spread is greatly reduced by IC spread reductions with diminishing returns occurring only in the last 20–30 min of the simulation. Although it is not shown, the forecast bias is also reduced by reducing the IC spread. Furthermore, by reducing the IC spread, the NMEP of LLV[3] exceeding $0.015\,s^{-1}$ was substantially increased where the control member LLV exceeds $0.015\,s^{-1}$ (Figs. 13d–f). Forecasts of the initial low-level

---

[3] Assuming solid-body vortex is an appropriate approximation for a low-level mesocyclone, then given an average tangential velocity between 5 and $10\,m\,s^{-1}$ (i.e., $7.5\,m\,s^{-1}$) and a radius of 1 km, then $\zeta = 2V_T/R = 0.015\,s^{-1}$.

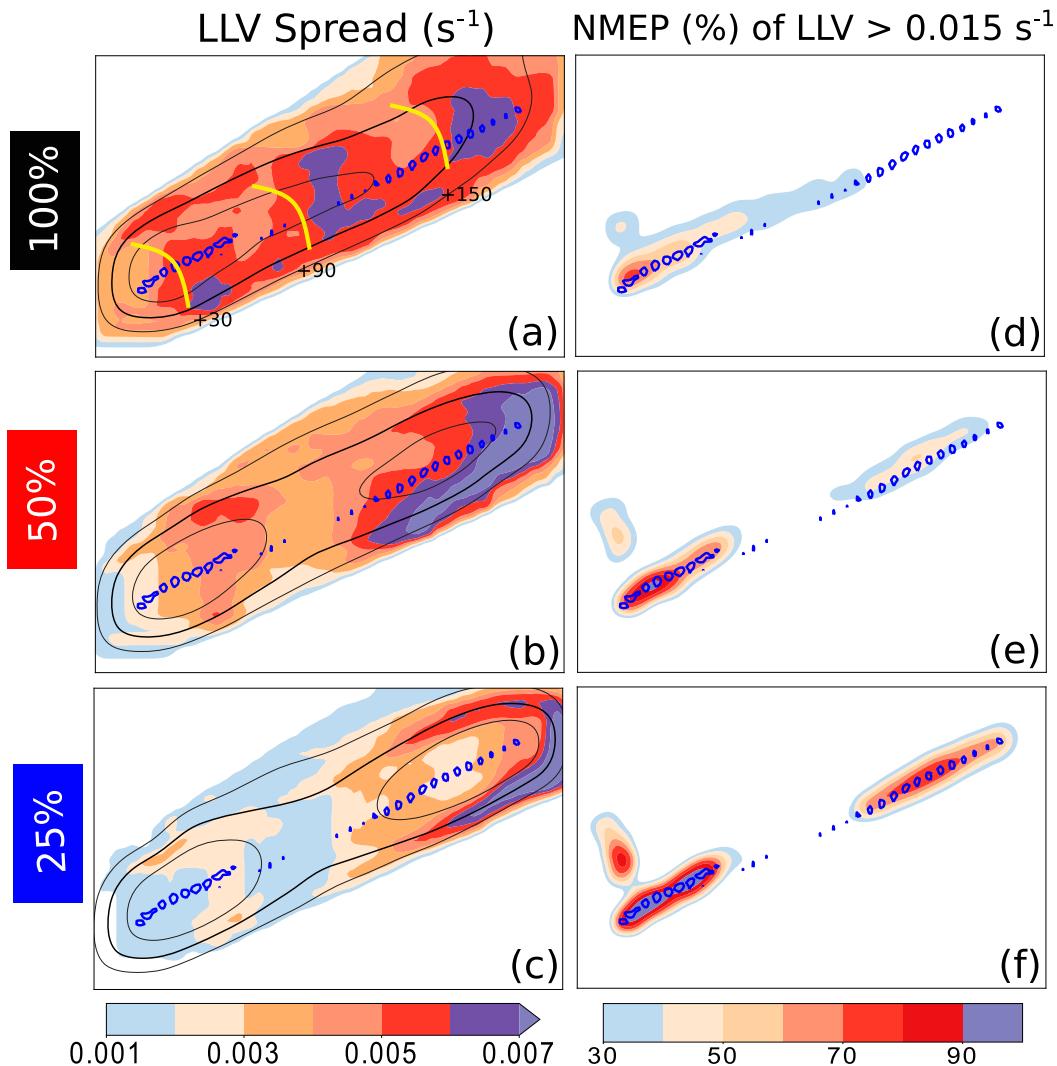## LLV Spread (s^{-1})　　　NMEP (%) of LLV > 0.015 s^{-1}

FIG. 13. Ensemble forecasts of 0–2-km AGL maximum LLV for the 9 May (a),(d) 100%; (b),(e) 50%; and (c),(f) 25% experiments. (left) LLV spread swaths maximized in a 20-km-radius neighborhood (filled contours) with probability-matched mean contoured (black lines; thick line = 0.012 s^{-1} with a contour interval of 0.003 s^{-1}). Yellow lines show approximate timing of ensemble every 60 min after initial 30 min. (right) NMEP of LLV > 0.015 s^{-1} (filled contour) within a 3-km-radius neighborhood. The 0.015 s^{-1} control member isoline is overlaid in blue for each column.

mesocyclone intensification are greatly improved by a 50% IC spread reduction, with modest improvements in forecasts of the second intensification (cf. Figs. 13d and 13e). In the 25% experiment, the probabilities are further increased relative to the 50% experiment in the forecasts of the initial intensification period, with a substantial improvement in the forecasts of the latter intensification period. Using different domain-maximum NMEP thresholds, Fig. 14 shows the LLV PPLs for each case, as well as the increase in the PPL with a 50% IC spread reduction. The PPLs are case dependent, with the 9 May PPL noticeably shorter than

in the other two cases. This is because the initial low-level mesocyclone intensification ends around 60 min, and a 50% reduction in IC spread is necessary to capture the latter intensification. Using a slightly lower LLV threshold for the low-level mesocyclone in the NMEP (i.e., 0.01 s^{-1}) for 9 May, the PPLs were extended on average by 50 min, making them more like the other two cases (not shown). The substantial forecast improvements gained by 50% IC spread reduction suggest that tornado prediction out to 3-h lead times could greatly benefit from a realizable reduction in current analysis uncertainty.
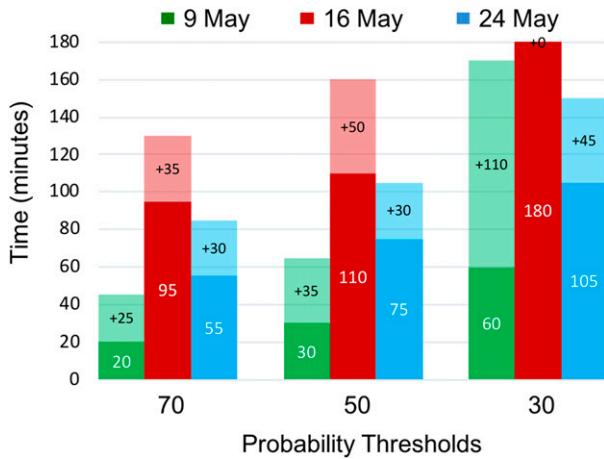
FIG. 14. As in Fig. 10, but for domain-maximum NMEP of maximum LLV $> 0.015\,\mathrm{s}^{-1}$.

c. Rainfall

The hourly rainfall for the three cases varied widely. The 9 May ensemble produced, on average, 0.5 in. (12.7 mm) after the first 90 min, while the 16 May ensemble exceeded well over 2 in. (50.8 mm) throughout the forecast period

(not shown). In the 24 May case, heavier hourly rainfall is produced, but there is poor agreement among the members in the 100% experiment. As we can see in Figs. 15a and 15d, the spread in hourly accumulated rainfall for the 24 May case is nearly 0.5 in., while the NMEP of hourly rainfall >0.75 in. (19.05 mm), a representative threshold for heavier rainfall in all three cases, is 40%–50%. Nevertheless, forecast spread is greatly reduced between both the 100% and 50% and the 50% and 25% experiments (Figs. 15a–c). These spread reductions arose in part from a substantial decrease in location uncertainty, which lead to a large increase in NMEP of hourly rainfall >0.75 in. (shown in Figs. 15d–f) and the collapse of >0.75-in. hourly rainfall isolines about the control member (not shown). Closer inspection (not shown) reveals that this strong sensitivity of rainfall forecast uncertainty to IC spread arises largely from differences in storm motion between ensemble members, which leads to differences in both the paths of heaviest rainfall and the duration of heavy rainfall at a given location. Moreover, rainfall predictability in the 24 May 100% experiment is limited not just by considerable location uncertainty, but also by large forecast bias (Figs. 15g–i), which is nearly 1 in. (25.4 mm) throughout the forecast period. Consistent with the results above,
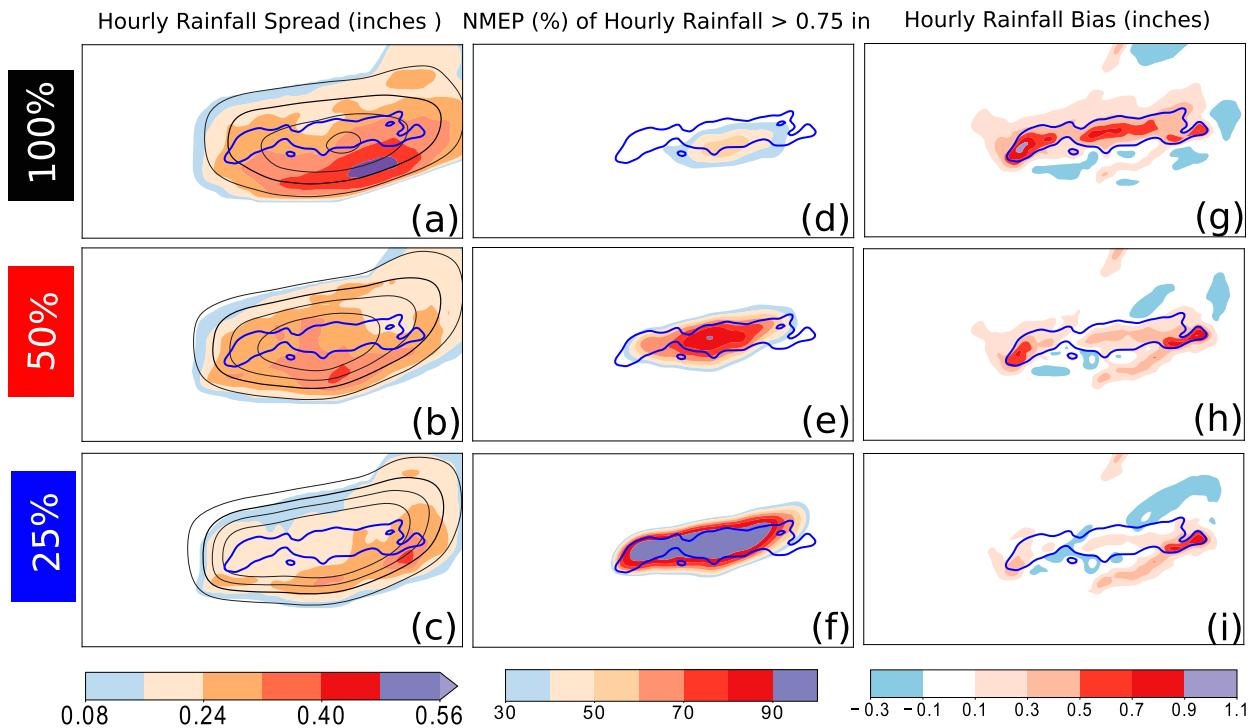

FIG. 15. Ensemble forecasts of hourly accumulated rainfall for the 24 May (a),(d),(g) 100%; (b),(e),(h) 50%; and (c),(f),(i) 25% experiments. (left) Hourly accumulated rainfall spread swaths maximized in a 20-km-radius neighborhood (filled contours) with probability-matched mean contoured (black lines; thick line = 0.5 in. with a contour interval of 0.25 in.). (middle) NMEP of hourly accumulated rainfall > 0.75 in. (filled contour) within a 3-km-radius neighborhood. (right) Swath of ensemble bias in hourly accumulated rainfall bias. The 0.75-in. control member isoline is overlaid in blue for each column.
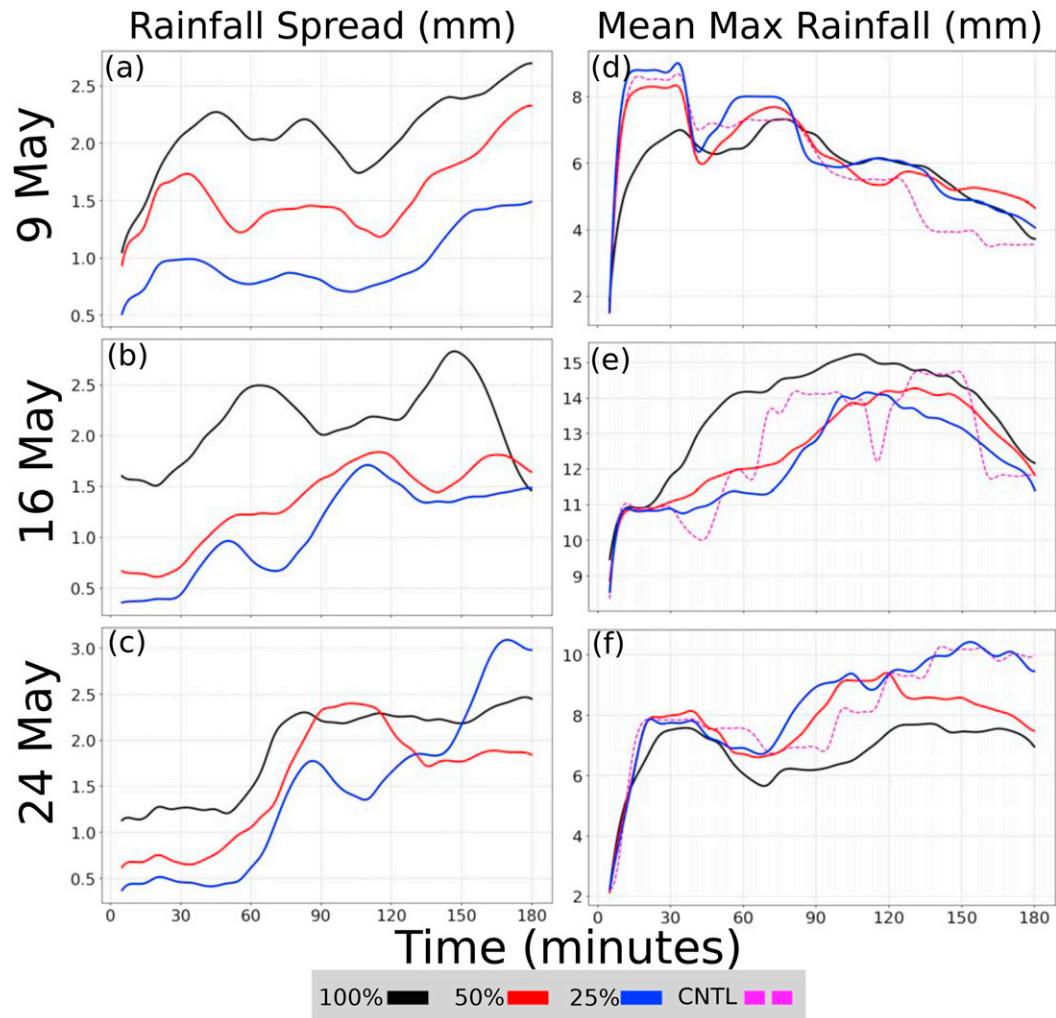
FIG. 16. As in Fig. 6, but for 5-min maximum rainfall.

however, the forecast bias in hourly rainfall in the 24 May case is substantially reduced by reducing the IC spread. Reducing IC uncertainty also substantially decreased forecast uncertainty of heavy rainfall location in the 16 May case (not shown). The 9 May control simulation produced relatively light rainfall; reducing IC spread correctly lowered the NMEP (not shown). For thresholds of 0.75 and 1.0 in., the maximum NMEP for all three cases never falls below 70% within the 3 h in the 100% experiments (not shown), making a 50% IC spread reduction practically unnecessary.

Figure 16 shows time series of spread in maximum 5-min rainfall, as well as mean maximum 5-min rainfall for all three cases. In the 24 May case, there is considerable overlap of the spread curves, consistent with the other supercell features for this case. Overall, the magnitude and growth rate of spread is similar among the three cases (cf. Figs. 16a–c). The small amplitude spread and large location spread in our experiments is consistent with past studies where location errors contributed more than amplitude errors to rainfall forecast uncertainty (e.g., Park 1999; Yussouf et al. 2016).

4. Conclusions

Developing a suitable full-physics NWP framework for studying storm-scale predictability is a necessary step for assessing the capabilities of storm-scale ensembles and for understanding the relative importance of different sources of forecast errors. The method developed herein enables systematic evaluation of the impact of IC uncertainty on ensuing forecasts of supercells, while leveraging the added realism (relative to idealized experiments) of storms, their environments, and analysis errors therein provided by contemporary convection-allowing ensembles. Even if the NEWS-e analyses used

in this study are underdispersed, despite evidence to the contrary (see Fig. 1), the experiments still serve to illustrate the effects of reducing IC spread. In addition, our focus on supercell features rather than traditional point-based metrics increases the interpretability and operational relevance of our results. Using this framework, we assess the practical predictability of the updraft helicity (primarily associated with the midlevel mesocyclone), low-level vorticity (a proxy for tornado potential), and rainfall over multiple cases. The IC spread is alternately set to 100%, 50%, and 25% of that in the real-world EnKF analyses from which our ensembles are initialized. This allows us to estimate contributions to forecast errors of contemporary and future analysis uncertainty.

Our major findings are as follows:

- Practical predictability of supercells is case and feature dependent.
- Practical predictability in the 100% experiments was substantially limited by forecast storm location uncertainty and ensemble bifurcations in forecast storm intensity.
- Reducing IC spread by 50% produced substantial reductions in forecast spread for all evaluated supercell features. Thus, severe thunderstorm, flash flood, and tornado warnings should be considerably improved if and when current storm-scale analysis uncertainty is halved.
- Further decreasing the IC spread to 25% of the original magnitude produced additional reductions in forecast spread in two of the three cases.
- Initializing supercell forecasts after the storms are well established in the ensemble analysis substantially improves their practical predictability.
- Rainfall predictability appears to be more heavily degraded by phase errors than are the predictabilities of the low- and midlevel mesocyclones.

To isolate the sensitivity of forecast spread to IC spread, we have adopted a perfect-model assumption in this study. In practice, however, model errors often produce large biases in ensemble forecasts. In addition, to account for model uncertainty in operational ensembles, physics diversity and/or stochastic physics schemes are often used, causing ensemble spread to increase faster than in our experiments. It follows that the PPLs diagnosed in this study are an upper-bound estimate of those for real ensemble systems. Our framework, therefore, can be used to set realistic limitations on expectations for the performance of real-world ensemble prediction systems. Furthermore, while it is unclear how literally the practical predictability increases gained by reducing IC spread in our experiments translate into real-world, imperfect-model ensembles, our results (and any future results leveraging the framework developed herein) provide valuable qualitative guidance for ensemble system design.

There are at least three major limitations of this study that should be addressed in future work. First, given that supercell predictability is case dependent, more cases must be investigated to produce general conclusions. All cases in this study are from the central Great Plains, which limits a direct application of the results to environments in other parts of the country, such as the low-CAPE environments that often occur in the southeastern United States during early spring. Second, all three supercell events were associated with fairly active large-scale patterns. We argued in section 3a that large-scale forcing enhanced 16 May's predictability, but it may be enhancing the predictability in the other two cases as well. Third, the predictability of other important supercell features, such as hail and the cold pool, are not investigated here. Hail is a frequently high-impact supercell hazard, and the cold pool is a critical factor in supercell longevity and tornado genesis, maintenance, and decay. Investigating the predictability of these and other supercell features would, therefore, provide additional valuable guidance for storm-scale ensemble design.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Bunkers, M. J., B. A. Klimowski, J. W. Zeitler, R. L. Thompson, and M. L. Weisman, 2000: Predicting supercell motion using a new hodograph technique. *Wea. Forecasting*, **15**, 61–79, https://doi.org/10.1175/1520-0434(2000)015<0061:PSMUAN>2.0.CO;2.

Cintineo, R. M., and D. J. Stensrud, 2013: On the predictability of supercell thunderstorm evolution. *J. Atmos. Sci.*, **70**, 1993–2011, https://doi.org/10.1175/JAS-D-12-0166.1.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005, https://doi.org/10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2.

——, and Coauthors, 2016: Development of a High-Resolution Rapid Refresh Ensemble (HRRRE) for severe weather forecasting. *28th Conf. on Severe Local Storms*, Portland, OR,

Amer. Meteor. Soc., 8B.2, https://ams.confex.com/ams/28SLS/webprogram/Paper301555.html.

Droegemeier, K. K., and J. Levit, 1993: The sensitivity of numerically-simulated storm evolution to initial conditions. Preprints, *17th Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., 431–435.

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.

Durran, D. R., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not of practical importance. *J. Atmos. Sci.*, **71**, 2476–2488, https://doi.org/10.1175/JAS-D-14-0007.1.

——, and J. A. Weyn, 2016: Thunderstorms do not get butterflies. *Bull. Amer. Meteor. Soc.*, **97**, 237–243, https://doi.org/10.1175/BAMS-D-15-00070.1.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

Lilly, D. K., 1990: Numerical prediction of thunderstorms—Has its time come? *Quart. J. Roy. Meteor. Soc.*, **116**, 779–798, https://doi.org/10.1002/qj.49711649402.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

——, 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.

McPherson, R. A., and K. K. Droegemeier, 1991: Numerical predictability experiments of the 20 May 1977 Del City, OK supercell storm. Preprints, *Ninth Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 734–738.

Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, https://doi.org/10.1175/JAS-D-11-0315.1.

Miglietta, M. M., A. Manzato, and R. Rotunno, 2016: Characteristics and predictability of a supercell during HyMeX SOP1. *Quart. J. Roy. Meteor. Soc.*, **142**, 2839–2853, https://doi.org/10.1002/qj.2872.

——, J. Mazon, and R. Rotunno, 2017: Numerical simulations of a tornadic supercell over the Mediterranean. *Wea. Forecasting*, **32**, 1209–1226, https://doi.org/10.1175/WAF-D-16-0223.1.

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.

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.

Park, S. K., 1999: Nonlinearity and predictability of convective rainfall associated with water vapor perturbations in a numerically simulated storm. *J. Geophys. Res.*, **104**, 31 575–31 587, https://doi.org/10.1029/1999JD900446.

Potvin, C. K., and L. J. Wicker, 2013: Assessing ensemble forecasts of low-level supercell rotation within an OSSE framework. *Wea. Forecasting*, **28**, 940–960, https://doi.org/10.1175/WAF-D-12-00122.1.

——, and M. L. Flora, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for Warn-on-Forecast. *Mon. Wea. Rev.*, **143**, 2998–3024, https://doi.org/10.1175/MWR-D-14-00416.1.

——, E. M. Murillo, M. L. Flora, and D. M. Wheatley, 2017: Sensitivity of supercell simulations to initial-condition resolution. *J. Atmos. Sci.*, **74**, 5–26, https://doi.org/10.1175/JAS-D-16-0098.1.

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.

Skamarock, W., 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.

Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Evolution of soil moisture and temperature in the MAPS/RUC assimilation. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 172–175.

——, S. G. Benjamin, J. M. Brown, B. Schwartz, and D. Kim, 2000: Validation of long-term precipitation and evolved soil moisture and temperature fields in MAPS. Preprints, *15th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 43–46.

Stensrud, D. J., 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.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Trapp, R. J., G. J. Stumpf, and K. L. Manross, 2005: A reassessment of the percentage of tornadic mesocyclones. *Wea. Forecasting*, **20**, 680–687, https://doi.org/10.1175/WAF864.1.

Wandishin, M. S., D. J. Stensrud, S. L. Mullen, and L. J. Wicker, 2008: On the predictability of mesoscale convective systems: Two-dimensional simulations. *Wea. Forecasting*, **23**, 773–785, https://doi.org/10.1175/2008WAF2007057.1.

——, ——, ——, and ——, 2010: On the predictability of mesoscale convective systems: Three-dimensional simulations. *Mon. Wea. Rev.*, **138**, 863–885, https://doi.org/10.1175/2009MWR2961.1.

Weyn, J. A., and D. R. Durran, 2017: The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *J. Atmos. Sci.*, **74**, 2191–2210, https://doi.org/10.1175/JAS-D-17-0006.1.

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.

Yussouf, N., D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, https://doi.org/10.1175/MWR-D-14-00268.1.

——, J. S. Kain, and A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Wea. Forecasting*, **31**, 957–983, https://doi.org/10.1175/WAF-D-15-0160.1.

Zhang, D., and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer—Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594–1609, https://doi.org/10.1175/1520-0450(1982)021<1594:AHRMOT>2.0.CO;2.

Zhang, F., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, https://doi.org/10.1175/JAS4028.1.

Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2015: Practical predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma: Sensitivity to synoptic timing and topographical influence. *Mon. Wea. Rev.*, **143**, 2973–2997, https://doi.org/10.1175/MWR-D-14-00394.1.

——, ——, ——, and ——, 2016: Intrinsic predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma at storm scales. *Mon. Wea. Rev.*, **144**, 1273–1298, https://doi.org/10.1175/MWR-D-15-0105.1.