Ensemble Hail Prediction for the Storms of 10 May 2010 in South-Central Oklahoma Using Single- and Double-Moment Microphysical Schemes

JONATHAN LABRIOLA

Center for Analysis and Prediction of Storms and School of Meteorology, University of Oklahoma, Norman, Oklahoma

NATHAN SNOOK AND YOUNGSUN JUNG

Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

BRYAN PUTNAM AND MING XUE

Center for Analysis and Prediction of Storms and School of Meteorology, University of Oklahoma, Norman, Oklahoma

(Manuscript received 23 February 2017, in final form 8 September 2017)

ABSTRACT

Explicit prediction of hail using numerical weather prediction models remains a significant challenge; microphysical uncertainties and errors are a significant contributor to this challenge. This study assesses the ability of storm-scale ensemble forecasts using single-moment Lin or double-moment Milbrandt and Yau microphysical schemes in predicting hail during a severe weather event over south-central Oklahoma on 10 May 2010. Radar and surface observations are assimilated using an ensemble Kalman filter (EnKF) at 5-min intervals. Three sets of ensemble forecasts, launched at 15-min intervals, are then produced from EnKF analyses at times ranging from 30 min prior to the first observed hail to the time of the first observed hail. Forty ensemble members are run at 500-m horizontal grid spacing in both EnKF assimilation cycles and subsequent forecasts. Hail forecasts are verified using radar-derived products including information from single- and dualpolarization radar data: maximum estimated size of hail (MESH), hydrometeor classification algorithm (HCA) output, and hail size discrimination algorithm (HSDA) output. Resulting hail forecasts show at most marginal skill, with the level of skill dependent on the forecast initialization time and microphysical scheme used. Forecasts using the double-moment scheme predict many small hailstones aloft, while the single-moment members predict larger hailstones. Near the surface, double-moment members predict larger hailstone sizes than their single-member counterparts. Hail in the forecasts is found to melt too quickly near the surface for members using either of the microphysics schemes examined. Analysis of microphysical budgets in both schemes indicates that both schemes suboptimally represent hail processes, adversely impacting the skill of surface hail forecasts.

1. Introduction

Severe hail is, on average, directly responsible for over a billion dollars of damage in the United States annually (Jewell and Brimelow 2009). Many businesses and industries are vulnerable to hail, including insurance, agriculture, aviation, transportation, and auto sales. There have been multiple instances in which a single hailstorm has caused over \$1.0 billion (U.S. dollars) in hail damage (Changnon and Burroughs 2003). Furthermore, the cost of damages resulting from severe hail is expected to increase in the future as a result of increasing population and rising property values.

Current methods of forecasting maximum hail size are limited in their ability to represent the surrounding environment. One commonly used hail size forecast model, HAILCAST (Jewell and Brimelow 2009), applies a one-dimensional model that uses proximity soundings to predict the growth of hailstones from initial hailstone embryos. Adams-Selin and Ziegler (2016) developed a version of HAILCAST that is implemented within convection allowing numerical weather

Corresponding author: Jonathan Labriola, j.labriola@ou.edu

DOI: 10.1175/MWR-D-17-0039.1

^{© 2017} American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

prediction (NWP) model forecasts to provide high spatial and temporal resolution hail forecasts. HAILCAST does not account for microphysical and dynamic features found within convective storms that have nonnegligible impacts on hail growth. Forecasting hail using NWP models remains a relatively understudied subject in the field of meteorology, particularly when compared to efforts focused on the prediction of other weather hazards such as flooding and tornadoes. Even as NWP forecast skill at convective scales continues to increase, skillful explicit prediction of hail remains difficult, requiring the model to accurately represent the storm intensity, mode, and motion, as well as microphysical and other physical processes important for hailstone formation. The short time scale of hailstorms and the limitations of the parameterizations used to represent microphysical processes within these storms contribute to rapid model error growth in hail-related forecast fields. In this study, we directly address the challenges of explicit hail prediction by investigating the ability of high-resolution ensemble forecasts using varying microphysics (MP) schemes to predict hail at varying lead times for a group of supercell thunderstorms that occurred over Oklahoma on 10 May 2010.

Ensemble prediction has become a vital tool for convective-scale forecasting because it is helpful in accounting for the fast error growth present at convective scales (Lorenz 1969; Clark et al. 2009, 2010). A well-designed ensemble will generate an "envelope" of possible outcomes that contains the true atmospheric state (Kalnay 2002; Kalnay et al. 2006). Furthermore, the ensemble can be used to generate probabilistic forecasts and provide uncertainty information. Two of the major factors affecting the ability of the ensemble to generate skillful short-term ensemble convective-scale forecasts are the data assimilation (DA; i.e., initial conditions) and MP parameterizations employed by the forecast system.

To maximize the accuracy of short-term storm-scale ensemble forecasts, it is necessary to reduce errors in the initial conditions by incorporating observational information via DA. To perform this task, modern stormscale NWP systems, both in operations and research, rely upon cycled DA systems that assimilate observations at regular intervals (e.g., Snyder and Zhang 2003; Dowell et al. 2004, 2011; Zhang et al. 2004; Tong and Xue 2005; Caya et al. 2005; Xue et al. 2006, 2010; Lei et al. 2008; Snook et al. 2011, 2015, 2016; Dawson et al. 2012; Jung et al. 2012a; Yussouf and Stensrud 2012; Yussouf et al. 2013; Putnam et al. 2014). While both variational and statistical techniques have been used, ensemble-based DA techniques, such as the ensemble Kalman filter (EnKF; Evensen 1994, 2003), are often chosen because they benefit from nonstatic spatial covariances carried by and derived from the ensemble. Flow-dependent covariances are particularly important at the storm scale, as they allow for the correction of errors in fields that, while not usually directly observed, strongly influence convective dynamics (e.g., vertical velocity, in-cloud temperature, pressure, and MP properties) (Tong and Xue 2005, 2008b). In this study, we use an EnKF DA system to assimilate radar and other observational data.

The parameterization of microphysical processes is complex, and is generally a large contributor to uncertainty within convective-scale models. In most research and operational applications, bulk MP schemes are used to parameterize cloud and precipitation processes because of their computational efficiency compared to spectral bin MP schemes. Most bulk MP schemes assume a threeparameter gamma distribution to describe the hydrometeor particle size distribution (PSD):

$$N(D) = N_{\alpha} D^{\alpha} e^{-\lambda D}, \qquad (1)$$

where N_o , λ , and α are the intercept, slope, and shape parameters, respectively (Ulbrich 1983). MP schemes can vary in complexity depending on the method used to specify these three parameters. The simplest schemes, single-moment (SM) schemes, predict only the mass mixing ratio (referred to in this paper as the mixing ratio) of hydrometeors. More complex double-moment (DM) schemes typically predict the total number concentration in addition to the mixing ratio. A number of recent studies (e.g., Jung et al. 2010; Dawson et al. 2010; Jung et al. 2012a; Putnam et al. 2014; Johnson et al. 2016) have found that the additional degrees of freedom within the predicted PSD when using a DM MP scheme allow for a more realistic representation of features observed in convective storms (e.g., cold pool structure, hydrometeor size sorting, polarimetric signatures). In this study, we will compare the relative performance of a DM scheme and an SM scheme within the EnKF DA and ensemble forecasts for hail.

Objective verification of hail forecasts is also a major challenge. Hail reports contain size and population biases, making verification of hail forecasts using these datasets difficult. In one of the first real-data convective-scale explicit hail prediction studies, Snook et al. (2016, hereafter S16) used products derived from Weather Surveillance Radar-1988 Doppler (WSR-88D) observations as supplementary sources of information for verification of hail forecasts to limit the impact of surface hail report biases and to facilitate verification in sparsely populated areas. More recently, Luo et al. (2017) used radar-derived products to verify hail simulations using one-, two-, and three-moment versions of the Milbrandt and Yau (2005) MP scheme. The Milbrandt and Yau (2005) threemoment version was found to perform the best. One radar product that is often used operationally as a proxy for hail observations is the maximum estimated size of hail (MESH). MESH is computed using a vertical integration of radar reflectivity Z multiplied by a temperature-based weighting function to estimate the maximum hail size at the surface (Witt et al. 1998). MESH was also used in the study to evaluate hail prediction. The recently upgraded WSR-88D network observes several polarimetric fields, including differential reflectivity Z_{dr} , copolar correlation coefficient $\rho_{\rm hv}$, and differential phase $\Phi_{\rm dp}$ in addition to radial velocity V_r and Z. These variables can be used to infer various properties of hydrometeors, including their size, shape, orientation, and phase (Kumjian and Ryzhkov 2008). When used as inputs for a fuzzy logic hydrometeor classification algorithm (HCA; Heinselman and Ryzhkov 2006; Park et al. 2009), polarimetric radar observations can be helpful in inferring the dominant hydrometeor species present, as well as in discriminating between regions of severe and nonsevere hail (Ryzhkov et al. 2013; Ortega et al. 2016).

In this paper, storm-scale ensemble forecasts with 500-m horizontal grid spacing are performed using single- and double-moment MP schemes for the case of a group of supercell thunderstorms occurring over southern Oklahoma on 10 May 2010. The goals of this study are 1) to evaluate the ability of ensemble forecasts with varying lead times to skillfully predict hail from these storms, 2) to evaluate the strengths and weaknesses of hail forecasts produced using different MP schemes, and 3) to investigate biases related to hail prediction in the MP schemes used in this study. These results can be used to further improve MP schemes and improve the viability of explicit hail prediction within a warn-on-forecast framework (Stensrud et al. 2009).

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the case, the ensemble configuration, and the techniques used in this study to verify ensemble hail forecasts. In section 3, we examine the sensitivity of hail prediction to the choice of MP scheme and forecast lead time, as well as giving consideration to the biases present in the MP schemes used. Finally, an overall summary of the results and discussion of potential areas for future investigation are presented in section 4.

2. Data and methods

a. Case overview

On the afternoon of 10 May 2010, a severe weather outbreak occurred in Oklahoma, in which a line of discrete supercell thunderstorms developed over the central and south-central parts of the state, producing multiple tornadoes and severe hail (Fig. 1). These thunderstorms developed ahead of a dryline at approximately 2200 UTC (Fig. 1a) and moved very quickly $(80-90 \text{ km h}^{-1})$. For ease of reference, we will refer to southern storms by a letter designation in alphabetical order from north to south (see Fig. 1). The northernmost, left-splitting storm is labeled storm A. Storms B-D are right-moving supercell thunderstorms. At 2315 UTC (Figs. 1c,d), storm C produced an EF3 tornado; Storm Prediction Center storm reports indicate that storm D simultaneously produced hail of up to 4-in. (101.6 mm) diameter. For an observationally oriented discussion of this case, we refer the reader to Palmer et al. (2011) and Bodine et al. (2014).

b. Experiment setup

1) PREDICTION MODEL AND DOMAIN

The Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001) is used as the prediction model in this study. All experiments use the same $723 \times 723 \times 53$ model grid, with a horizontal grid spacing of 500 m and a stretched vertical grid with a minimum vertical grid spacing of 50m near the surface that increases to a maximum vertical spacing of 800 m at the model top (approximately 21 km above the surface). The 500-m horizontal grid spacing used in this study is consistent with Snook et al. (2016) and is necessary to resolve finescale structures associated with hail processes. Full model physics are used (Xue et al. 2000), including the NASA Goddard Space Flight Center short- and longwave radiation scheme; surface fluxes calculated from surface drag coefficients, surface temperature, and surface volumetric water; a two-layer soil model; and a 1.5-order subgrid-scale turbulent mixing and planetary boundary layer parameterization. The initial ensemble forecasts for the DA cycles are initialized at 2000 UTC (approximately 75 min prior to convective initiation), from a 40-member ARPS parent ensemble run at 4-km horizontal grid spacing (Jung et al. 2012b 2013). The parent ensemble (Fig. 2a) is initialized at 1500 UTC and assimilates surface and radar observations hourly (1500-1700 UTC), the frequency of assimilation increases to every 10 min 1 h prior to the completion of DA at 1800 UTC. After 1800 UTC, a 6-h forecast is run until 0000 UTC 11 May 2010; the initial and external boundary conditions are interpolated from the parent ensemble forecast. External boundary conditions are obtained hourly from the parent ensemble (Fig. 2a), and linear time interpolation is used to interpolate between boundary conditions. Hydrometeor information is not



FIG. 1. KOUN 0.5° tilt observed reflectivity field on 10 May 2010 at (a) 2200, (b) 2230, (c) 2300, and (d) 2330 UTC. Letter identifiers for individual storms are included. Purple contours represent urban boundaries.

included in the boundary conditions ("dry boundary conditions") to minimize the occurrence of spurious convection initiating along the model boundaries.

2) MICROPHYSICS SCHEMES

The Milbrandt and Yau (2005) DM MP scheme (hereafter MY2) and the Lin et al. (1983) scheme as modified by Tao and Simpson (1993) (hereafter LIN) are used in two different sets of ensemble forecast experiments, which will be described in detail below. Hail forecasts produced using the Milbrandt and Yau (2005) SM MP scheme (not shown) were found to be inferior to both the LIN and MY2 scheme forecasts and, thus, will not be discussed. The MY2 scheme has been used in a number of previous ARPS studies (e.g., Dawson et al. 2010; Jung et al. 2012a; Putnam et al. 2014; Dawson et al. 2015), including the recent hail prediction study of S16. The LIN scheme, which is implemented in the parent ensemble (Jung et al. 2012b), is one of the most widely used single-moment MP schemes (Snook et al. 2011, 2015; Putnam et al. 2014), has previously been compared to the MY2 scheme (Jung et al. 2012a), and includes a rimed ice category that is most similar to hail. The MY2 scheme includes predictive equations for hydrometeor number concentrations N_{tx} and mixing ratios q_x for cloud water (N_{tc}, q_c) , cloud ice (N_{ti}, q_i) , rain (N_{tr}, q_r) , snow (N_{ts}, q_s) , graupel (N_{tg}, q_g) , and hail (N_{th}, q_h) , allowing the scheme to diagnose N_{0x} and λ_x while assuming a fixed-shape parameter (which, in this study, is set to 0). The LIN scheme predicts the mixing ratios of five hydrometeor species: cloud water (q_c) , cloud ice (q_i) , rain (q_r) , snow (q_s) , and hail (q_h) , allowing the scheme to diagnose λ_x for precipitating species while using preset constant values for N_{0x} . Unlike the MY2 scheme, the LIN scheme does not include a separate graupel category. The LIN hail category assumes a particle density of 913 kg m⁻³ and $N_{0h} = 4.4 \times 10^4$ m⁻⁴. To reduce cold pool intensity, the default N_{0r} (8.0 \times 10^6 m^{-4}) is reduced to $2.0 \times 10^6 \text{ m}^{-4}$ following Snook and Xue (2008).



FIG. 2. (a) Forecast diagram detailing the parent ensemble configuration for the case study. Red region represents 1-h DA cycling, orange is 10-min DA cycling, and the \times denotes when external boundary conditions for the case study ensemble are obtained from the parent ensemble. The time at which the case study ensemble is initialized is denoted above the timeline. (b) Forecast diagram detailing the case study ensemble configuration for both the MY2 and LIN ensembles. Blue region represents the spinup period, red is 15-min DA cycling, and orange is 5-min DA cycling. Significant weather events that are of importance to the case study are listed above the timeline. In both panels green regions represent forecast periods and downward arrows indicate when assimilation occurs.

c. DA settings

Starting from 2000 UTC, a 30-min spinup ensemble forecast is performed. After this spinup period, cycled DA is performed (Fig. 2b) using the CAPS EnKF DA system (Tong and Xue 2005; Xue et al. 2006; Tong and Xue 2008a). This system employs a variant of the ensemble square root filter (EnSRF; Whitaker and Hamill 2002). Available S-band WSR-88D and Engineering Research Center for Collaborative and Adaptive Sensing of the Atmosphere (CASA; McLaughlin et al. 2009) X-band radar observations, soundings, profiler data, Oklahoma and West Texas Mesonet observations, along with ASOS and AWOS observations, are assimilated at 15-min intervals from 2030 to 2115 UTC. Starting from 2115 UTC, the approximate time of convective initiation (Fig. 2b), the DA interval is reduced to 5 min for the remainder of the DA period. During DA, the radarindicated severe hail is first observed at 2150 UTC (there are too few surface-based reports to use direct hail size observations), and at 2155 UTC the storms produce multiple tornado reports (Fig. 2b). Storm D develops late within the DA period. The storm produces Z > 20 dBZ first at 2155 UTC, causing it to be undersampled by DA for ensembles launched at 2200 and 2215 UTC and completely unsampled by DA for ensembles launched at 2145 UTC.

In preliminary experiments, performing many frequent DA cycles caused time integration instability in some ensemble members because fast-moving storms on this day resulted in the repeated introduction of large, unbalanced adjustments to the model state variables. To alleviate this problem, the incremental analysis update (IAU) technique (Bloom et al. 1996) is used to incorporate the analysis increments into the forecast model. Incremental updates decrease with time during the 5-min DA cycles. IAU is particularly desirable for its filtering properties that reduce the impact of introducing large adjustments into a model. To maintain ensemble spread, a relaxation-to-prior-spread algorithm is applied following Whitaker and Hamill (2012) using a coefficient of 0.95, along with multiplicative covariance inflation in regions directly influenced by radar observations with an inflation factor of 1.10 (Tong and Xue 2005). From these continuous DA cycles, three separate ensemble forecast experiments are launched at 2145, 2200, and 2215 UTC. To refer to these forecast ensembles, we adopt a naming convention in which a forecast ensemble is denoted by the MP scheme used (MY2 or LIN) and the time T_0 at which the forecast was launched (e.g., MY2_2145T₀ or LIN_2215T₀).

Other configurations of the DA system largely follow S16. Radar observation errors of $6.0 \, \text{dBZ}$ and $4.0 \, \text{m s}^{-1}$ are assumed for Z and V_r , respectively. The cutoff radius of the covariance localization function (Gaspari and Cohn 1999) is set to 3 km in the horizontal and vertical directions to be consistent with Jung et al. (2012a) and Snook et al. (2016). Radar observations are interpolated horizontally onto the model grid but are not interpolated vertically (Xue et al. 2006). The forward observation operator for DA uses the T-matrix method for raindrops and the Rayleigh scattering approximation for ice particle scattering amplitude calculations (Jung et al. 2008). Radar data from the four S-band WSR-88Ds closest to the storms of interest are assimilated; these radars include Twin Lakes, Oklahoma (KTLX); Fredrick, Oklahoma (KFDR); Dyess Air Force Base, Texas (KDYX); and Dallas-Fort Worth, Texas (KFWS). Data from four X-band radars operated by CASA (McLaughlin et al. 2009) are also assimilated. The Chickasha (KSAO), Rush Springs (KRSP), Cyril (KCYR), and Lawton (KLWE) CASA radars were located between KTLX and KFDR in southwest Oklahoma, adding more near-surface radar coverage in this area. Reflectivity data from the CASA radars were processed using automated attenuation correction; for CASA radars, only Z observations exceeding $25 \, \text{dBZ}$ are assimilated to avoid assimilation of spurious, fully attenuated observations. The CASA data preprocessing procedure follows Snook et al. (2011).

Observation errors for surface measurements are assumed to be 1.5 m s^{-1} for horizontal wind components uand v, 1.5 K for air temperature T, 2.0 K for dewpoint T_d , and 2.0 hPa for air pressure p. The horizontal covariance localization radius for surface observations is 300 km, and the vertical localization radius is 6 km. Assumed errors for sounding data are 1.2 K for T, 2.0 K for T_d , and 0.6 hPa for p. For both soundings and wind profiler observations, an error of 2.5 m s^{-1} is assumed for uand v. A horizontal localization radius of 800 km and a vertical localization radius of 6 km are used for these upper-air observations. The above settings follow Snook et al. (2015).

d. Verification procedure

For verification, the model predictions of Z are calculated using a T-matrix method to obtain scattering amplitudes for both rain and frozen model hydrometeor fields (Jung et al. 2010). This method is capable of replicating the Mie scattering caused by large hailstones, resulting in realistic estimates for radar variables from the model data. The direct use of the T-matrix method as the forward operator within DA for ice hydrometeor categories is too expensive and is, therefore, only done for forecast verification; Z is highly relevant in evaluating the skill of hail forecasts because it provides valuable information about storm mode, structure, intensity, and microphysical state.

Because the storms in this study occurred largely over rural areas, there are relatively few surface hail reports available, and the available reports likely do not fully represent the region over which hail actually fell due to population biases typical of severe weather reports (e.g., Wyatt and Witt 1997; Davis and Ladue 2004). To supplement surface reports, radar-derived products such as MESH and hail swaths indicated by a HCA are used as supplementary sources of observed hail information, as was done in S16. MESH uses a weighted vertical integration of Z in regions of Z > 40 dBZ above the freezing level H_0 to estimate the maximum hail size:

severe hail index =
$$0.1 \times \int_{H_0}^{H} W(H) E \, dH$$
 and (2)

$$MESH = 2.54 (severe hail index)^{0.5}, \quad (3)$$

where E is the weighted Z and W is a temperature-based weighting function described in Witt et al. (1998). Ortega et al. (2009, 2016) document that MESH has a large false alarm ratio, and a wide range of maximum hail sizes can be found within a single MESH value. However, MESH is a commonly used operational product, and in many cases it is the only dataset capable of providing high spatiotemporal resolution estimates of surface hail size, making it an important supplementary data source for local storm reports in less populated regions (Melick et al. 2014). Though polarimetric observations were not available from operational WSR-88D observations in 2010, available polarimetric radar observations from the experimental dual-polarization WSR-88D in Norman, Oklahoma (KOUN), are accessed to determine likely regions of hail using a variant of the HCA of Park et al. (2009) with modifications following S16. KOUN is situated 100–150 km away from the storms, and the lowest radar tilt samples the storms at approximately 1.5-3.0 km AGL. This HCA uses observed Z, Z_{dr} , and ρ_{hv} as inputs for a fuzzy logic



FIG. 3. The NEP of the ensemble predicted reflectivity exceeding 25 dBZ (color shading) for (a) MY2_2145T₀, (b) MY2_2200T₀, (c) MY2_2215T₀, (d) LIN_2145T₀, (e) LIN_2200T₀, and (f) LIN_2215T₀ valid at 2245 UTC 10 May 2010. The observed 25-dBZ reflectivity contours are plotted in black, along with the corresponding observed storm letter identifiers. Purple contours represent urban boundaries.

algorithm to distinguish among dominant observed hydrometeor types. Observed V_r is used in addition to these fields to distinguish ground clutter.

Ensemble forecasts produced using neighborhood ensemble probability (NEP; Schwartz et al. 2010, 2017) and probability-matched ensemble means (Ebert 2001) are verified in this study. NEP analyzes the fractional coverage of event occurrence over a given neighborhood. This method is used because, at the convective scales, models often perform poorly in terms of point-to-point verification (e.g., threat scores) as a result of common displacement errors, particularly for highly localized phenomena such as hail. NEP mitigates this issue by allowing for small displacement errors, producing a probabilistic forecast by verifying each point against a nearby neighborhood of points, determined using a radius of influence (2.5 km in this study, following \$16). The probability-matched mean method produces a single best-guess forecast from the ensemble using the spatial distribution of the ensemble mean, but maintaining the frequency distribution of the full ensemble; this preserves extreme values that would otherwise be smoothed away when computing the ensemble mean.

Radar-derived MESH is used for objective verification in this study, as it is the only data source available with the spatial and temporal resolution necessary for surface hail size verification. The fractions skill score (FSS; Roberts and Lean 2008) is employed in this study to compare the fractional coverage of predicted surface hail to the fractional coverage of radar-derived MESH. This verification technique is particularly useful because it provides the scales at which the model predicts the spatial extent of the surface hail with acceptable skill. For this study a hail size threshold of 5 mm is used for verification; larger hail sizes (e.g., severe hail, significant severe hail) will not be verified because the sample size of these events is too small.

3. Results

a. Ensemble prediction of general reflectivity structure

The forecast NEP for P(Z > 25 dBZ) at 2245 UTC is shown in Fig. 3 for the ensemble forecasts started at 2145, 2200, and 2215 UTC using the MY2 and LIN MP schemes, respectively. In each experiment, the storms in the ensemble forecasts move eastward more quickly than the observed storms. Storm motion bias is frequently documented in convective modeling studies (e.g., S16; VandenBerg et al. 2014; Yussouf et al. 2016), and is likely attributed to model error. As expected, ensembles with the longest lead times [i.e., those launched the earliest; MY2_2145T₀ (Fig. 3a) and LIN_ 2145T₀ (Fig. 3d)] contain larger regions of moderate probability values and fewer areas with probability near 1.0, reflecting the greater variability in storm placement and intensity within these ensembles than in ensembles launched at later times, though there is still strong agreement among the members of these ensembles in predicting the northern part of the storms in the subdomain. The MY2_2145T₀, MY2_2200T₀, and LIN_ 2145T₀ results (Figs. 3a, 3b, and 3d, respectively) have lowered confidence in storm D; P(Z > 25 dBZ) is below 0.4 for these ensembles. Analysis of individual ensemble members indicates that the LIN scheme generally predicts smaller regions where Z exceeds 25 dBZ than the MY2 scheme. Both LIN_2145T₀ (Fig. 3d) and LIN_ $2200T_0$ (Fig. 3e) predict the development of a new storm behind the primary line, displaced from a secondary line of observed storms farther to the west. The MY2 ensembles (Figs. 3a-c) and LIN_2215T₀ (Fig. 3f) also predict this secondary line; however, $P(Z > 25 \, \text{dB}Z)$ is less than 0.2. We can see that LIN_{2215T_0} (Fig. 3f) captures the spatial extent of the storms better than $MY2_{2215T_0}$ (Fig. 3c). At longer lead times, the MY2 and LIN ensemble forecasts capture the spatial extent of the storms with similar levels of skill.

b. Ensemble prediction of MESH

Ensemble hail forecasts valid from 2215 to 2315 UTC are compared and verified during this study. During this time, the WSR-88D MESH swath (Fig. 4) for storm A remains nonsevere (MESH < 25 mm) and is oriented more to the northeast than the other swaths (which are oriented more to the east-northeast). Storm B, located just south of storm A, also produces a swath of nonsevere MESH before dissipating (Fig. 4). Both storms C and D produce long swaths of MESH that contain indications of severe hail (maximum predicted MESH exceeding 25 mm).

The MY2 and LIN ensembles predict MESH swaths (Fig. 5) near the observed MESH swaths (Fig. 4), but greatly overpredict the spatial extent of the MESH field. As noted earlier, the forecasted storms in all ensembles move more quickly than the observed storms (maximum observed storm motion $80-90 \text{ km h}^{-1}$), causing the MESH swaths to extend too far to the north and east. Ensembles using both the MY2 and LIN schemes perform poorly in predicting MESH for storm D; the storm intensifies to maturity only after the end of the DA period, reducing the ability of the models to accurately forecast this storm. Only LIN_2200T₀ predicts a MESH



FIG. 4. MESH field calculated from a mosaic of KOUN, KTLX, and KFDR radar observations mapped onto the model grid between 2215 and 2315 UTC. Letter identifiers for observed storms are included; purple contours represent urban boundaries.

swath associated with storm D (Fig. 5e). Contrary to the WSR-88D MESH (Fig. 4), the ensemble forecasts predict storm B to produce the highest-intensity MESH, until after approximately 2250 UTC, when the ensembles with shorter lead times [MY2_2200T₀ (Fig. 5b), MY2_2215T₀ (Fig. 5c), LIN_2200T₀ (Fig. 5e), and LIN_2215T₀ (Fig. 5f)] predict that storm C intensifies. Because of the initially large hail in terms of simulated MESH for storm B predicted by both the MY2 (Figs. 5a–c) and LIN (Figs. 5d–f) ensembles, it is believed that the predicted storm B is similar to the observed storm C (Fig. 4) except that it is geographically displaced.

While LIN_2200T₀ produces a more accurate forecast of MESH for storm D, the LIN ensembles consistently overpredict the intensity of MESH produced by all of the storms. In all three LIN ensembles (Figs. 5d-f), MESH exceeds 75 mm for both storms B and C, while the observed MESH swaths produced by these storms never exceed 50 mm. Maximum MESH values produced by the MY2 ensembles (Figs. 5a-c) are more similar to those in the observations. All three MY2 ensembles generally predict MESH values of less than 25 mm (nonsevere) for storms B and C, with only small regions of MESH exceeding 25 mm. The LIN ensembles also fail to capture the weakening of storm A later during the 2215–2315 UTC forecast period. In the observations, the left-splitting storm A stops producing MESH $> 5 \,\text{mm}$ by 2305 UTC (see Fig. 4), while the LIN ensembles predict a swath of MESH for storm A continuing through 2315 UTC (Figs. 5d-f). In LIN_2215T₀, storm A spuriously intensifies late in the forecast period,



FIG. 5. The probability-matched mean of the ensemble predicted MESH swath between 2215 and 2315 UTC. Letter identifiers for predicted storms are included; purple contours represent urban boundaries.

producing large MESH values near the eastern end of the 1-h forecast MESH swath (Fig. 5f). Though the MY2 ensembles (Figs. 5a–c) predict longer MESH swaths for storm A than the observations (Fig. 4), the intensity of the storm is better represented in these ensembles than in the LIN ensembles. Both MY2_2200T₀ (Fig. 5b) and MY2_2215T₀ (Fig. 5c) predict weakening in the MESH field of storm A, with maximum MESH values decreasing to only 5 mm by 2315 UTC.

Both MP schemes show limited skill in predicting the spatial coverage of MESH > 5 mm at scales beneficial to explicit short-term hail prediction. The ensemble average FSS (Roberts and Lean 2008) indicates that the LIN scheme (Figs. 6d,e) predicts the spatial coverage of MESH > 5 mm with more skill than the MY2 scheme (Figs. 6a-c). Both LIN_2215T₀ (Fig. 6f) and LIN_ $2200T_0$ (Fig. 6e) have several ensemble members that predict the spatial coverage of MESH $> 5 \,\text{mm}$ with skill at scales of less than 10 km. No ensemble members from MY2_2215T₀ (Fig. 6c) and MY2_2200T₀ (Fig. 6b) predict the spatial coverage of MESH > 5 mm with skill at scales less than 10 km. It is important to note that FSS remains unaffected by the LIN scheme's overprediction bias in maximum hail size (Figs. 5e,f) because a low threshold of nonsevere hail (MESH > 5 mm) is analyzed. MESH values associated with severe hail (MESH > 25 mm) are too localized to be objectively analyzed for this case study. Reduced MESH forecast skill is likely due to an overprediction bias in the spatial area of MESH (Fig. 5); this is because the predicted storms move too quickly in both the MY2 and LIN schemes.

To better understand the differences in the predicted MESH swaths from the LIN and MY2 ensembles, we will now consider explicit predictions of hail from the different experiments. The total mass of frozen hydrometeor species (hail and graupel) is presented as a function of height for ensemble members 10 and 30 of MY2_2215T₀ and LIN_2215T₀ in Fig. 7. These two ensemble members are selected because member 10 predicts large quantities of hail, particularly in MY2_2215T₀, while member 30 predicts more modest quantities of hail. The hydrometeor mass content for hydrometeor $x (m_x)$ is defined as

$$m_x = q_x \times \rho_{\rm air},\tag{4}$$

where q_x is the mixing ratio of hydrometeor x and ρ_{air} is the air density. The LIN scheme exhibits less variability between members in m_h and, generally, predicts larger m_h values above the freezing layer than the MY2 scheme for both members 10 and 30 (Figs. 7a,c). The MY2 scheme likely predicts smaller m_h values aloft due in part to the fact that the scheme has two separate rimed ice categories (graupel and hail), while LIN represents all rimed ice in the form of dense ice precipitation as hail. The MY2 scheme predicts a substantial fraction of dense frozen precipitation to be graupel, with



FIG. 6. FSSs of ensembles (a) MY2_2145T₀, (b) MY2_2200T₀, (c) MY2_2215T₀, (d) LIN_2145T₀, (e) LIN_2200T₀, and (f) LIN_2215T₀ for predicting the spatial coverage of MESH > 5 mm between 2215 and 2315 UTC. Radarderived MESH shown in Fig. 4 is used as the observations. The threshold for a skillful forecast is the horizontal top dashed line; climatology is the bottom dashed line in each panel. Each of the individual ensemble member FSSs is marked as a gray line, and the mean FSS of the ensemble is marked as a boldfaced colored line.

the mass of graupel m_g exceeding m_h in both members (Figs. 7a,c). The combined mass of frozen hydrometeors is overall much larger in the MY2 members than in the LIN members (Fig. 7), and in terms of vertical distribution, does not extend as far below the melting layer in MY2 as in the LIN members. Consistent with most other MY2 members, MY2 member 30 predicts almost no hail reaching the surface. This result is consistent with the findings of S16, who reported that very little hail reached the surface in their MY2 forecasts of supercell thunderstorms, even where hail was observed, leading to substantial underprediction of the near-surface hail mass content. In the MY2 members, graupel, which is smaller in diameter than hail in the simulations, contributes much less to the total Z than does hail, despite being the dominant hydrometeor species above the freezing layer in terms of mass (Figs. 7b,d). The component of Z from graupel (Z_g) is approximately $30 \, dBZ$ over much of the vertical extent of the storm, while Z_h exceeds 50 dBZ above the freezing layer. The MY2 scheme (Figs. 8a,b), which has been previously reported to predict many smaller hailstones (Johnson et al. 2016), predicts smaller hailstones aloft than does the LIN scheme (Figs. 8c,d). Small hailstones are a consequence of MY2 growth mechanisms, frozen raindrops significantly increase N_{th} in the freezing layer but do not proportionally modify m_h ; this decreases the mean mass

diameter of hail aloft. The LIN scheme assumes a constant intercept parameter, and the mean mass diameter increases aloft as m_h increases. A more detailed microphysical budget analysis of the two schemes will be presented in a later section. Reflectivity Z, which is proportional to the sixth moment of the PSD, is most sensitive to larger hailstones; consequently, LIN predicts larger Z_h aloft than MY2 (Figs. 7b,d). Having an additional low-density rimed ice category (graupel) may also further reduce Z in MY2. MESH is a function of Zabove the melting layer and is, thus, strongly sensitive to the distribution and intensity of the reflectivity field within the storm. Overall, the MY2 scheme predicts more realistic Z above the freezing layer than the LIN scheme (see Fig. 5). Differences between the hail PSDs predicted by the two schemes will be further analyzed in section 3c.

c. Explicit prediction of hail

1) EVALUATION OF HAIL MASS CONTENT

We will now compare the predicted hail mass content m_h against radar-indicated hail swaths defined using the HCA described earlier in section 2b, following the methods of S16. In both the LIN and MY2 ensembles, storms B and C are predicted to produce the largest mass of hail (Fig. 9). The LIN ensembles (Figs. 9d–f) predict



FIG. 7. Member 10 of (a),(b) MY2_2215T₀ and LIN_2215T₀ valid at 2245 UTC. Member 30 of (c),(d) MY2_2215T₀ and LIN_2215T₀ valid at 2245 UTC. Shown in (a) and (c) are the total hydrometeor species mass per model grid height within a subdomain that encompasses storms A–D. A threshold of 100 g m⁻¹ is plotted to delineate meaningful quantities of hail predicted by the model. Shown in (b) and (d) are the average reflectivities contributed by hydrometeor species per model height in a subdomain surrounding the storms of interest. The 5th–95th percentile freezing layer heights within the subdomain are shaded in blue. Only 2–16 km AGL is plotted to highlight features above the freezing layer.

 m_h exceeding 5.0 g m⁻³ in these storms; in contrast, MY2_2200T₀ (Fig. 9b) and MY2_2215T₀ (Fig. 9c) predict m_h of less than 2.0 g m⁻³ for storms B and C, while MY2_2145T₀ (Fig. 9a) predicts even smaller m_h values and produces hail only for storm B in many members. The MY2 ensemble members predict m_h of less than $0.01 \,\mathrm{g}\,\mathrm{m}^{-3}$ for storm A, which is located closest to the KOUN radar site. From the predicted MESH field (Figs. 5a-c), it is evident that the MY2 ensembles predict high Z associated with storm A above the freezing layer, but hail rapidly melts in the model below the freezing level (Figs. 7a,c), leading to very low m_h for storm A at the height of the 0.5° KOUN radar tilt (approximately 1 km above the surface near storm A) in the MY2 ensembles. Both LIN_2200T₀ (Fig. 9e) and LIN_2215T₀ (Fig. 9f) predict hail at the height of the 0.5° KOUN radar tilt for storm A. We find that LIN_2145T₀ (Fig. 9d), which generally poorly represents storm A, does not predict any hail associated with storm A. The swaths of hail predicted by the LIN ensembles (Figs. 9d–f) are larger in geographic extent than those predicted by the MY2 ensembles (Figs. 9a–c) and generally more intense. All LIN ensembles predict m_h exceeding 0.03 g m⁻³ (the threshold for verification of the m_h content used in S16) throughout much of the southern third of the subdomain; this region includes large areas where no surface hail was observed. In contrast, the MY2 scheme predicts more limited hail swaths with fewer false alarms.

An east-west vertical cross section of the m_h content through the hail core of storm B at 2245 UTC for members 10, 20, 30, and 40 of MY2_2215T₀ and LIN_ 2215T₀ (Fig. 10) reveals that the largest m_h values are collocated with the primary storm updraft. The LIN members (Figs. 10e-h) contain a large region of hail approximately 10 km wide within the updraft of storm B,



FIG. 8. Vertical cross sections of the hail mean mass diameter field at 2245 UTC for MY2 members (a) 10 and (b) 30 and LIN members (c) 10 and (d) 30. The cross section is taken through the m_h maximum of storm B at 1 km. The freezing layer is marked with a brown line. The *u* and *w* winds are denoted with arrows.

with m_h values exceeding 2.0 g m⁻³. The MY2 members mostly predict narrower, weaker updrafts for storm B, along with correspondingly lower amounts of hail, with m_h rarely exceeding $1.0 \,\mathrm{g \, m^{-3}}$ for storm B in MY2 members 20, 30, and 40 (Figs. 10b-d). MY2 member 10 (Fig. 10a), which is initialized with the strongest updraft for storm B of the selected members, is markedly different, exhibiting a robust updraft, with m_h exceeding $2.5 \,\mathrm{g}\,\mathrm{m}^{-3}$. For all selected members, the MY2 scheme predicts smaller, more localized hail cores, typically less than 5 km wide. Hail in the MY2 members melts more quickly below the freezing layer than the LIN members. Only MY2 member 10 predicts hail $(m_h > 0.03 \text{ gm}^{-3})$ that reaches the surface within the vertical cross section, while all LIN members predict hail that reaches the surface at this time. This is in agreement with the observations.

2) EVALUATION OF MAXIMUM HAIL SIZE

The PSD for hail can be used to estimate the maximum hail diameter D_{max} at any given point, as suggested by Milbrandt and Yau (2006a). The predicted PSD in the model is based upon the exponential distribution, with the shape parameter being set to 0 in Eq. (1) for this study. As hail diameter increases, the number concentration decreases. For very large diameters, the predicted hail becomes so sparse that it would not be physically observable (e.g., one hailstone per cubic kilometer). In S16, a cutoff value for number concentration of $10^{-4} \text{ m}^{-3} \text{ mm}^{-1}$ (equivalent of one hailstone in a cube with sides approximately 20 m in length) is used to denote D_{max} . This choice of cutoff value is similar (though not identical) to that used by Milbrandt and Yau (2006a).

Hail PSDs for LIN and MY2 members 10 and 30 at varying heights in the lower troposphere are compared in Fig. 11, highlighting differences in the predicted hail PSD, particularly near and within the melting layer. The PSDs presented in Fig. 11 are calculated at varying heights, placed horizontally at the location of the m_h maximum of storm B at 1000 m above mean sea level for each member. The lowest number concentration plotted in Fig. 11 is 10^{-4} m⁻³ mm⁻¹, thus the *x* intercept of each PSD in Fig. 11 is D_{max} .



FIG. 9. Probability matched mean of m_h presented as swaths valid for 2215–2315 UTC. Data are linearly interpolated to the height of the KOUN 0.5° tilt. Swaths of hail diagnosed by the HCA from KOUN 0.5° data during the same time span are contoured in red. Letter identifiers for observed storms are included; purple contours represent urban boundaries.

Within the melting layer, the MY2 scheme predicts PSDs that contain relatively few but large hailstones, as evidenced by the lower slope of the PSDs at 4000 m (Figs. 11a,b). At 5000 m, 500 m above the freezing layer,

the MY2 scheme predicts a large number of small hailstones in both member 10 (Fig. 11a) and member 30 (Fig. 11b). Within the melting layer at 4000 m, the sizes of the largest hailstones increase. At this height the MY2



FIG. 10. A vertical cross section of the m_h field at 2245 UTC. The cross section is taken through a 1-km maximum in m_h in storm B. Shown are members (a) 10, (b) 20, (c) 30, and (d) 40 of MY2_2215T_0 and LIN_2215T_0 members (e) 10, (f) 20, (g) 30, and (h) 40. The freezing layer is marked by a brown line; u and w winds are denoted by arrows.



FIG. 11. PSDs for hail within the hail core of storm B at 2245 UTC at varying heights MSL for MY2_2215T₀ members (a) 10 and (b) 30 and for LIN_2215T₀ members (c) 10 and (d) 30. PSDs are taken at the horizontal location of the m_h maximum at 1000 m MSL. The model surface is at 300 m MSL and the freezing layer is at 4500 m MSL.

scheme decreases N_{th} to preserve mean mass diameter due to melting, the scheme also predicts that hail accretes rain and cloud water, causing m_h to increase. Accretional growth while hail is simultaneously melting causes the MY2 scheme to predict that the hail mean mass diameter increases in the melting layer; Milbrandt and Yau (2006a) documented similar results. An analysis of the treatment of hail within the MP schemes will be performed later in section 3d. Hailstones in MY2 member 10 (Fig. 11a) grow much larger than in MY2 member 30 (Fig. 11b), consistent with the much more robust storm updraft predicted in member 10. Because N_0 is constant in the SM LIN scheme ($N_{0h} = 4 \times$ 10^{-4} m^{-4}), LIN members 10 (Fig. 11c) and 30 (Fig. 11d) exhibit PSDs with many small hailstones, even as hail melts below the freezing level. Because of its fixed N_0 value, the LIN scheme decreases in mass through melting larger hailstones and decreasing hail mean mass diameter; this is unrealistic and is a disadvantage of SM schemes. Despite this disadvantage, the LIN scheme does predict a larger D_{max} just below the freezing layer (Figs. 11c,d), with less variation between ensemble members than for the MY2 members (Figs. 11a,b).

Vertical cross sections of D_{max} through the hail core of storm B in selected MY2 and LIN members are shown in

Fig. 12. Because of the constraints on the SM scheme, the D_{max} field in the LIN members (Figs. 12e-h) is directly proportional to the m_h content (Figs. 10e-h). Large hail cores containing severe hail $(D_{\text{max}} > 25 \text{ mm})$ are predicted in the LIN members; these hail cores are collocated with the storm updraft. The largest D_{max} value in each of the LIN members is around 45-50 mm and is consistently located near the freezing level (Figs. 12e-h). Because the MY2 scheme allows the intercept parameter to vary, D_{max} is not directly proportional to m_h for the MY2 members (Figs. 12a–d). In MY2 members 20, 30, and 40 (Figs. 12b-d), the extent of large hail in the hail core of storm B is less than in the corresponding LIN members, and the maximum D_{max} is only slightly lower, at around 35-40 mm, despite the far lower m_h values in the hail cores of the MY2 members compared to the LIN members (Figs. 10b-d and 10f-h). Member 10 of MY2_2215T₀ (Fig. 12a) shows a marked difference in both the m_h and D_{max} fields compared to the other MY2_2215T₀ members. Values of m_h in the hail core of member 10 exceed $2.0 \,\mathrm{g}\,\mathrm{m}^{-3}$, which is comparable to LIN_2215T₀ member 10, and much higher values of D_{max} are present, up to 90 mm; the largest values of D_{max} are predicted below the freezing layer (Fig. 12a), which is consistent with Fig. 11a. Values of $D_{\rm max}$ indicate that hail remains largely nonsevere



FIG. 12. A vertical cross section of the D_{max} field at 2245 UTC from four ensemble members from MY2_2215T₀ and LIN_2215T₀, respectively, similar to those shown in Fig. 10. The cross section is taken through the m_h maximum of storm B at 1 km. The freezing layer is marked with a brown line.

above the freezing layer in MY2 member 10. This is because above the freezing level there is an increase in N_{th} , distributing m_h between more, but smaller, hailstones. In the other three selected MY2 ensemble members (Figs. 12b–d), the updrafts are weaker, and almost no hail with D_{max} greater than 10 mm reaches the surface. It is important to note that although $D_{max} >$ 5 mm near the surface in these cross sections, members predict $m_h < 0.03$ g m⁻³ at the surface (Fig. 10).

To produce a forecast estimate of maximum hail size at the surface, we consider swaths of maximum D_{max} , presented as probability-matched mean values for each ensemble. The correlation between near-surface D_{max} (SFC) and MESH is analyzed (Table 1) to better understand the relationship between these two fields. Model-simulated MESH is dependent upon simulated observations above the melting layer and implicitly accounts for melting within the algorithm (Witt et al. 1998); D_{max} explicitly accounts for melting calculated within the MP scheme. MESH and D_{max} should be highly correlated because both fields represent the maximum hail size at the surface; however, the correlation between the two fields is low (0.256-0.302) for MY2 and even lower (0.174-0.227) for LIN (Table 1). Low correlations between the D_{max} and MESH indicate discrepancies in melting treatment. Analyzing the maximum D_{max} within the lowest kilometer (1KM) of the atmosphere reduces the impact of near-surface melting, which can be excessive as noted earlier in our discussion of Figs. 10 and 12, as well as prior results (e.g., S16). Reducing the impacts of near-surface melting improves the overall correlation between the model-simulated MESH and D_{max} for both the MY2 (0.438–0.461) and LIN (0.481–0.519) schemes (Table 1). Because of the improved correlation between the two fields, this study will analyze swaths of maximum D_{max} within the lowest kilometer of the atmosphere (Fig. 13).

Both the MY2 and LIN ensembles produce very little near-surface hail for storms C and D. The MY2 (Figs. 13b,c) and LIN (Figs. 13e,f) ensemble forecasts were launched at 2200 UTC and later predict near-surface hail from storm C. The ensemble forecasts underestimate the spatial extent of hail when compared to the WSR-88D MESH (Fig. 4). Only LIN_2145T₀ (Fig. 13d) predicts that storm A does not produce near-surface hail; the other ensemble forecasts (Figs. 13a–c and 13e–f) more closely resemble the WSR-88D MESH field (Fig. 4). Storm B produces the most extensive and

TABLE 1. The average correlation coefficient between LIN- and MY2-derived MESH results with predicted D_{max} at the surface (SFC) and maximum D_{max} within the lowest kilometer of the atmosphere (1KM).

	2145T ₀	2200T ₀	2215T ₀
MY2 SFC	0.302	0.284	0.256
MY2 1KM	0.461	0.456	0.438
LIN SFC	0.227	0.174	0.212
LIN 1KM	0.519	0.481	0.492



FIG. 13. The probability matched mean of the maximum D_{max} field within the lowest kilometer of the atmosphere between 2215 and 2315 UTC for the (a)–(c) MY2 and (d)–(f) LIN ensembles. Individual storm letter identifiers are included for predicted storms; purple contours represent urban boundaries.

largest surface hail in the ensemble forecasts. Initially, $MY2_{2200T_0}$ (Fig. 13b) and $MY2_{2215T_0}$ (Fig. 13c) overestimate the maximum hail size in the storm B hail swath, but D_{max} quickly decreases to values close to those indicated by the observed MESH (Fig. 4). The initial overestimates of D_{max} in the vicinity of storm B can largely be attributed to the addition of large hailstones near and below the freezing layer during EnKF DA. Storm B weakens during the forecast in most ensemble members, consequently producing increasingly smaller hailstones beneath the freezing level. Member 10, which is initialized with a more robust updraft, maintains the intensity of storm B during the forecast and produces large hailstones below the freezing layer during the forecast (Fig. 12a). The corresponding LIN ensembles, LIN_{2200T_0} (Fig. 13e) and LIN_{2215T_0} (Fig. 13f), do not initially overpredict D_{max} for storm B, remaining relatively consistent with the WSR-88D MESH values for the full 2215-2315 UTC swath. The MY2 and LIN ensembles with the earliest initialization time, MY2_2145T $_0$ (Fig. 13a) and LIN_2145T $_0$ (Fig. 13d), produce similar hail-swath predictions that underestimate the spatial extent of hail near the surface compared to WSR-88D MESH, although the MY2 ensemble at this time better captures the hail swath associated with the left-splitting storm A (Fig. 13a).

While both MP schemes demonstrate the ability to explicitly predict the size of hail near the surface, the ensembles lack skill in representing the fractional coverage of $D_{\text{max}} > 5 \,\text{mm}$ at scales adequate for operational use (Fig. 14). The ensembles that were initialized the latest, MY2_2215T₀ (Fig. 14c) and LIN_2215T₀ (Fig. 14f), are capable of representing the fractional coverage of near-surface hail at scales of approximately 45 km. These ensembles assimilated more data than ensembles initialized at 2145 and 2200 UTC, leading to an overall improvement in the skill of the hail forecasts. For MY2_2145T₀ (Fig. 14a), MY2_2200T₀ (Fig. 14b), LIN_2145T₀ (Fig. 14d), and LIN_2200T₀ (Fig. 14e), we show limited skill in capturing the fractional coverage of near-surface hail at spatial scales of less than 50 km. It is important to note that radar-derived MESH (Fig. 4) is used to verify D_{max} forecasts (Fig. 14); weak correlation between the two fields likely suppresses FSS values. The FSS of the simulated MESH verified against radarderived MESH (Fig. 6) increases for all ensembles; this is because both fields contain similar biases in estimated surface hail size and spatial extent (Ortega et al. 2009, 2016). Additionally, when D_{max} is calculated in the melting layer, the algorithm inherits biases associated with the MP scheme treatment of melting hail that degrade forecast skill (e.g., extreme hail growth in the



FIG. 14. FSSs for ensembles (a) MY2_2145T₀, (b) MY2_2200T₀, (c) MY2_2215T₀, (d) LIN_2145T₀, (e) LIN_2200T₀, and (f) LIN_2215T₀ for predicting the spatial coverage of $D_{max} > 5$ mm between 2215 and 2315 UTC. Note that D_{max} is the maximum hail size using the D_{max} algorithm within the lowest kilometer of the atmosphere. Radar-derived MESH is used as the observations. The threshold for a skillful forecast is shown by the top horizontal dashed line; climatology is the bottom dashed line in each figure. Each of the individual ensemble member FSSs is marked as a gray line, and the mean FSS of the ensemble is marked as a boldface colored line.

MY2 scheme). An analysis of the microphysical treatment of hail is continued later in the section.

Though MESH can be used to provide an estimate of maximum hail size at the surface from Z and temperature information above the freezing level, the availability of dual-polarization radar observations provides additional information that can be utilized to verify surface hail size. To this end, we apply a hail size discrimination algorithm (HSDA; Ryzhkov et al. 2013) modified by Ortega et al. (2016) with confidence vectors equal to one (J. Krause 2016, personal communication) to verify model-predicted maximum hail size. The HSDA is run in addition to the HCA in regions where the HCA output identifies rain/hail. It includes additional membership functions for Z, Z_{dr} , and ρ_{hv} that classify hail into one of three size categories: nonsevere (0-25 mm), severe (25-50 mm), and significant severe (>50 mm). Compared to the WSR-88D MESH field (Fig. 4), the HSDA (Fig. 15) output produces localized hail swaths with a larger maximum hail size. The HSDA output indicates multiple regions with hailstones larger than 50mm in diameter, while the MESH field hardly contains values exceeding 50mm. These significant severe swaths tend to be far away from KOUN, where the radar observes high above the ground near the freezing layer. Hailstones near the freezing layer are expected to be larger than those near the surface as they have undergone less melting.

For direct comparison to the HSDA (Fig. 15), we interpolate the predicted D_{max} field to the KOUN 0.5° radar tilt (Fig. 16). The LIN_2200T₀ (Fig. 16e) and LIN_ $2215T_0$ (Fig. 16f) ensembles predict that storm A produces nonsevere hail near the surface, similar to WSR-88D MESH. Hail associated with the left-splitting storm A is underpredicted or absent in LIN_2145T₀ (Fig. 16d) and the MY2 ensembles (Figs. 16a-c). Ensembles MY2_2200T₀ (Fig. 16b) and MY2_2215T₀ (Fig. 16c) predict storm A to produce near-surface hail, but excessive melting limits the size and spatial extent of the hail. The MY2 ensembles (Figs. 16a-c) predict more localized swaths of severe hail than the LIN ensembles (Figs. 16d–f) for storm B at the elevation of the KOUN radar data. Both MY2_2200T₀ (Fig. 16b) and MY2_ $2215T_0$ (Fig. 16c) predict relatively narrow swaths of hail exceeding 95 mm in diameter, resembling a swath of significant severe hail produced by storm C in the HSDA output (Fig. 15). Localized pockets of significant severe hail predicted by the MY2 ensembles more



FIG. 15. Swath of HSDA output from 2215 to 2315 UTC for KOUN 0.5° tilt observations. To reduce noise in the output, differential reflectivity and copolar correlation coefficient are smoothed using a nine-point filter before use in the HCA. Observed storm letter identifiers are included; purple contours represent urban boundaries.

closely resemble severe hail regions in the HSDA output than the LIN ensembles, which predict wider swaths of severe hail no larger than 50 mm in diameter. The MY2 and LIN ensembles both predict spatially larger swaths of hail greater than 5 mm in diameter than the HSDA output.

At the elevation of the 0.5° KOUN tilt, the LIN ensembles (Figs. 16d-f) produce a larger region of hail in the southern portion of the subdomain than the MY2 ensembles (Figs. 16a-c) and much of this hail is associated with storm D (see Fig. 4). The MY2 ensembles produce limited nonsevere hail in this region, with D_{max} not exceeding 20 mm in diameter (Figs. 16a-c). The LIN ensembles predict hail as large as 35 mm in diameter in this region, a result more in agreement with the HSDA output (Fig. 15) and WSR-88D MESH (Fig. 4), although much of the hail predicted by the LIN ensembles to the south of the observed storm D is spurious. The D_{max} field interpolated to the elevation of the 0.5° KOUN tilt (Fig. 16) more closely resembles the WSR-88D MESH field (Fig. 4) in both spatial coverage and size of hail than the forecast MESH field (Fig. 5) for both the MY2 and LIN ensembles.

Some of the ensembles predict storms (particularly storm D) that do not produce hail, although WSR-88D-derived MESH (Fig. 4) indicates otherwise. It is important to note that each of the observed storms, including storm D, is present in most ensembles. A swath of NEP of P(Z > 35 dBZ) for the 2215–2315 UTC forecast period (Fig. 17) reveals that all ensembles, with



FIG. 16. The probability matched mean of the D_{max} field linearly interpolated to the KOUN 0.5° radar tilt between 2215 and 2315 UTC. The ensembles are plotted following the convention in Fig. 3. Predicted storm letter identifiers are included; purple contours represent urban boundaries.



FIG. 17. The NEP of $Z > 35 \,\text{dB}Z$ between 2215 and 2315 UTC. Model-predicted radar reflectivity is linearly interpolated onto the KTLX 0.5° radar tilt. Observed reflectivity exceeding 35 dBZ is contoured in black. Observed storm letter identifiers are included; purple contours represent urban boundaries.

the exception of LIN_2145T₀ (Fig. 17d), capture the storm-splitting event that produces storm A, and storms B, C, and D. Given that the storms are well captured by most of the ensembles in terms of Z but not in terms of near-surface hail (i.e., D_{max}), further investigation of the MP scheme's treatment of hail beneath the freezing layer is warranted.

d. Evaluation of microphysical budgets

To identify the cause of biases observed within D_{max} and the MESH forecasts, microphysical hail source and sink terms are examined using a microphysical budget analysis. These terms are shown (Fig. 18) in the form of accumulated totals within the updrafts ($w > 5.0 \text{ m s}^{-1}$) and downdrafts ($w < -2.5 \text{ m s}^{-1}$) of supercell thunderstorms in south-central Oklahoma at 2245 UTC. Terms are summed along a horizontal plane, and the resulting summed values are plotted as a function of model grid height to illustrate the model's treatment of hail both above and below the 0°C isotherm.

The magnitude of m_h within the storm is highly dependent upon the choice of MP scheme; the LIN scheme predicts larger m_h values than the MY2 scheme (e.g., Figs. 9 and 10). Most m_h in the LIN scheme is generated in the storm updrafts via accretion of cloud water, accretion of snow, and three-component accretion of rain (Fig. 18e); the accretion of cloud water and snow within the LIN scheme results in the generation of more m_h

than do all of the MY2 m_h source terms combined (Fig. 18a). The LIN scheme predicts hail that accretes relatively little rain; this is because the accretion of rain and shedding water are treated as a single term during the wet growth of hail. Because of the large accretion of snow and cloud water terms during wet growth, the LIN scheme predicts hailstones that shed more water than accrete rain. Both the MY2 (Fig. 18b) and LIN (Fig. 18f) schemes predict m_h growth is limited within storm downdrafts; these regions are typically unsaturated with respect to water and do not support the accretion of cloud water. In storm downdrafts modest m_h growth is attributed to hail colliding with hydrometeors such as rain in the MY2 scheme (Fig. 18b) and snow in the LIN scheme (Fig. 18f).

Large accretional m_h growth terms cause the LIN scheme to predict large m_h values aloft, confined to storm updrafts (Figs. 10e-h). Because the LIN scheme is an SM scheme, the hail mean mass diameter is largest aloft (Figs. 8c,d), where m_h is largest. MESH, which is dependent upon Z above the freezing layer, increases as a result of the large hail predicted to be aloft by the LIN scheme (Figs. 5d-f). Within the storm updraft the m_h source and sink terms for the MY2 scheme (Fig. 18a) are smaller in magnitude than those in the LIN scheme (Fig. 18e); the three most significant mass growth terms in the scheme are the accretion of cloud water, the accretion of rain, and three-component accretion of rain.



FIG. 18. The m_h tendency terms for (a),(b) MY2_2215T₀ and (e),(f) LIN_2215T₀, as well as the N_{th} tendency terms for (c),(d) MY2_2215T₀ at 2245 UTC. There are no N_{th} tendency terms predicted by the LIN scheme. Tendency terms are summed at each vertical height over a subdomain that encompasses storms A–D, either within storm (a),(c),(e) updraft or (b),(d),(f) downdraft regions, where $w > 5 \text{ m s}^{-1}$ or $w < -2.5 \text{ m s}^{-1}$, respectively. Terms are contoured between the 5th and 95th percentile values within the ensemble; the median value is marked with a boldfaced colored line. Only significant m_h tendency terms (terms in which the median exceeds 0.5 g m⁻¹ in magnitude) and N_{th} tendency terms in which the median exceeds 10 stones m⁻¹) are plotted. The average freezing layer within the subdomain is marked as the upper horizontal dashed black line, and the average wet-bulb freezing layer over the subdomain is marked as the lower dashed black line.

Of particular note is the continued accretion of rainwater by hail throughout the melting layer in both the storm updraft and downdraft regions; this is an unphysical behavior of the MY2 scheme (Figs. 18a,b). Above the melting layer in the storm updraft region, N_{th} quickly increases in the MY2 forecasts (Fig. 18c). This is due to three-component accretion of rain and the conversion of graupel to hail. A large increase in the N_{th} above the freezing layer causes the mean mass diameter of hail to decrease aloft within storm updrafts in the MY2 simulations (e.g., Fig. 8a); this pattern of behavior is the opposite of what occurs within the LIN scheme. The tendency of the MY2 scheme to predict small hail aloft is noted in Johnson et al. (2016); these small hailstones produce lower Z, resulting in lower predicted MESH values (Figs. 5a–c).

Vertical cross sections of hail mean mass diameter (Figs. 8a,b), D_{max} (Fig. 12), and hail PSDs (Fig. 11) indicate that the MY2 scheme predicts rapid growth in hail size beneath the 0°C isotherm. Analysis of the MY2 total

contribution to m_h within the storm updraft (Fig. 19a) indicates that the scheme predicts a net growth in m_h beneath the wet-bulb freezing level. Within this layer hail accretes cloud water and rain; however. almost no water is shed (Figs. 18a,b). The inability of hail to efficiently shed water in the melting layer significantly increases hail size in the MY2 simulations. Instead of being shed, accreted water freezes to hailstones and increases m_h , even when hailstones undergo melting and N_{th} decreases (Fig. 19b). Increasing m_h and decreasing N_{th} causes the mean mass diameter of hail to inflate in the upper melting layer (Figs. 8a,b). In extreme examples, N_{th} decreases quickly and m_h increases, resulting in predicted $D_{\text{max}} > 95 \,\text{mm}$ (Fig. 12a). Within thunderstorm downdrafts hail does not increase in size as rapidly, and the total m_h and N_{th} tendencies are much smaller and remain neutral or decreasing throughout the vertical extent of the downdraft (Fig. 19b).

PSDs throughout the hail core predicted by the LIN scheme (Figs. 11c,d) indicate that hail can grow



FIG. 19. The total m_h tendency for (a),(b) MY2_2215T₀ and (e),(f) LIN_2215T₀ and the total N_{th} tendency for (c),(d) MY2_2215T₀ at 2245 UTC. There are no N_{th} tendency terms predicted by the LIN scheme. Tendency terms are summed at each vertical height over a subdomain that encompasses storms A–D either within storm (a),(c),(e) updraft or (b),(d),(f) downdraft regions, where $w > 5 \text{ m s}^{-1}$ or $w < -2.5 \text{ m s}^{-1}$, respectively. Terms are contoured between the 5th and 95th percentiles values within the ensemble, and the median value is marked with a boldface colored line. The average freezing layer within the subdomain is marked as the upper horizontal black dashed line, and the average wet-bulb freezing layer over the subdomain is marked as the lower black dashed line.

modestly in size near the upper melting layer, primarily as a result of three-component accretion of rain, accretion of cloud water, and accretion of snow near the 0°C isotherm (Fig. 18e). Similar to the MY2 scheme, the LIN scheme predicts the hail to shed relatively little water beneath the 0°C isotherm; however, hail growth in the melting layer is limited. This is because the scheme does not allow accretional growth of hail beneath the wet-bulb freezing level (Figs. 18e,f), causing the melting term to be dominant beneath the 0°C isotherm (Figs. 19e,f). The LIN scheme also assumes a constant intercept parameter; thus, hail diameter is a monotonic function of m_h .

4. Summary and discussion

In this study, we evaluate the ability of SM and DM MP schemes in an EnKF-initialized storm-scale ensemble forecasting system to explicitly predict hail in 0–90-min forecasts for both left- and right-moving supercells in south-central Oklahoma on 10 May 2010. The experiments assimilate Z and V_r from both operational WSR-88D and experimental CASA X-band radars, as well as incorporating surface observations using EnKF and 40 ensemble members. Ensemble forecasts are then launched at 15-min intervals starting approximately 30 min prior to the first indication of hail by the WSR-88D data in terms of the MESH. The DA and forecast experiments are performed using one of two MP schemes: the SM LIN and DM MY2 schemes. To supplement the sparse surface hail reports recorded on 10 May 2010, hail products derived from WSR-88D observations are used to verify hail forecasts (as in S16); these products include MESH and the output of a HCA. Ensemble forecasts are also compared to the output of a HSDA run on WSR-88D observations.

The forecast ensembles exhibit marginal to moderate skill in predicting Z produced by the supercell thunderstorms over southern and central Oklahoma on 10 May 2010. Forecasts were poorer for the ensembles that were launched earliest (the MY2 and LIN ensembles launched at 2145 UTC); these ensembles only

predict the single strongest storm. Ensembles launched later (at 2200 and 2215 UTC) predict most of the hailproducing storms, though they still exhibit limited ability to predict storms that matured after the end of the DA period. The storms in all ensemble forecasts move to the north and east more quickly than do the observed storms. Similar biases in storm motion have been noted in prior studies, including S16. The previous literature suggests storm motion biases are likely attributable to model error (Yussouf et al. 2016). This study provides no additional information for why simulated storms move too quickly.

Both MP schemes predict the geographic extent and size of hail with marginal skill for 0-75-min time scales for forecasts launched when most of the storms of interest are mature or close to mature (i.e., at 2200 or 2215 UTC). The geographic extent of hail in the ensemble forecasts resembles that of the observations, but predicted hail size often differs from radar-derived hail sizes. Ensembles launched earlier, when the storms of interest were still developing (LIN_2145T₀ and MY2_ $2145T_0$), exhibit low skill in predicting hail; these ensembles underestimate the intensity of the storms throughout the duration of the experiment. Because the forecast storms consistently moved faster than the observed storms, all experiments exhibit hail swaths extending farther to the east than in surface observations or radar-indicated hail fields.

The LIN scheme predicts spatially larger swaths of hail than the MY2 scheme as a result of the microphysical treatment of hail. The LIN ensembles often predict m_h values nearly an order of magnitude larger than those predicted in corresponding MY2 ensembles because of the accretion of significant quantities of cloud water and snow by hail within the LIN scheme. Because SM MP schemes assume a constant intercept parameter, both D_{max} and the hail mean mass diameter are proportional to m_h in the LIN scheme. Since m_h is largest above the melting layer in the LIN forecasts, the largest hail sizes are similarly located above the melting layer. We note that unlike the MY2 scheme, the LIN scheme does not have a separate graupel category. When m_g and m_h are combined in the MY2 forecasts, the combined total exceeds m_h predicted in the LIN forecasts. Because the MY2 scheme predicts N_{th} to modify the intercept parameter, both D_{max} and the hail mean mass diameter can be increased without modifying m_h in the MY2 forecasts. This allows the MY2 scheme to predict large quantities of small hailstones aloft and fewer, larger hailstones near the surface. Occasionally, the MY2 scheme predicts unrealistically large near-surface hail sizes. This behavior can be attributed to the scheme allowing hail mass to increase via the accretion of cloud

and rainwater in the melting level (where N_{th} simultaneously decreases because of melting and no accreted water is shed). Overall, in the ensemble forecasts produced in this study, the LIN scheme predicts wide swaths of severe hail ($25 < D_{max} < 50 \text{ mm}$), while the MY2 scheme predicts localized swaths of severe hail, including some significant severe hail ($D_{max} > 50 \text{ mm}$).

The WSR-88D-derived MESH is used operationally to assess the potential size of hail at the surface based on the observed Z and temperature aloft. When the MESH algorithm is applied to the ensemble forecasts, the LIN ensembles generally overpredict MESH both in terms of intensity and spatial extent. The LIN ensembles predict significant severe hail (MESH > 50 mm) in several storms, while MESH derived from WSR-88D observations does not exceed 50 mm. The widespread intense MESH in the LIN ensembles can be attributed in part to the larger hailstones predicted within the storms, producing high values of simulated $Z (Z > 50 \, \text{dB}Z)$ and, thus, very large values of MESH. In contrast, the MY2 scheme favors smaller hail, which lowers the simulated Z. Including separate high- and low-density rimed ice categories in the MY2 scheme may additionally lower Z. Overall, the MY2-predicted MESH more closely resembles WSR-88D MESH than the LIN-predicted MESH. Accurate MP representation of different hydrometeor categories is an ongoing area of research; future work focusing on the characterization of hail and graupel has the potential to mitigate errors introduced into the ensemble by the MP scheme, allowing for more accurate prediction of the size and distribution of hydrometeors within storms.

Compared to proxy surface observations, including MESH derived from WSR-88D observations, and the output of a HSDA in close proximity to the radar performed on the lowest radar tilt, all ensembles generally underpredict the explicit (i.e., D_{max}) size and spatial coverage of hail near the surface. In the LIN ensembles, the default value of hail N_0 of the LIN scheme results in incorrect preferential melting of the largest hailstones, limiting the size of a hailstones near the surface. In contrast, in the MY2 forecasts, the scheme decreases the N_{th} to preserve the mean mass diameter during melting. This is a more physical approach. In some instances accretional growth within the melting layer significantly increases the maximum hail size within the MY2 simulations. Additionally the MY2 scheme predicts that hail melts too quickly, resulting in almost no hail reaching the surface. Also, even when temperatures exceed 0°C, the MY2 scheme predicts hail to grow in diameter beneath the freezing layer as a result of the accretion of rain and cloud water (wet growth). This process is most notable in storms with intense updrafts. Because there

are few in situ observations of hail near the freezing level, verification of such processes is difficult; surfacebased hail reports and radar observations offer only limited information on the melting of hail aloft. Currently, WSR-88D MESH is the best method for verifying predicted surface hail despite it being mainly dependent on reflectivity observations above the freezing layer. With the polarimetric upgrade of the operational WSR-88Ds, HCA and HSDA algorithms based on low-level polarimetric data have the potential to become preferred over MESH for hail verification. For this case study, surface hail reports are limited because most of the storms occurred in a sparsely populated rural portion of southern Oklahoma; future work will investigate hail-producing storms occurring over urban areas for which more widespread surface observations are available.

One potential way of further improving hail prediction is through the use of a three-moment MP scheme, such as that of Milbrandt and Yau (2006b). Introduction of a third moment allows the scheme to predict the shape parameter of the PSD. This may help to better represent hydrometeor sedimentation and mitigate the melting bias noted in this study, potentially allowing more realistic quantities of hail to reach the surface. Other recently developed schemes, such as those of Mansell et al. (2010) and Milbrandt and Morrison (2013), attempt to better represent the variety of rimed ice particles present in a storm. These MP parameterizations show promise in improving hail prediction. By explicitly predicting bulk density of graupel and/or hail, these schemes allow a greater range of possible particle characteristics. In another recently developed method, the predicted particle properties (P3) scheme of Morrison and Milbrandt (2015) and Morrison et al. (2015), the rimed ice category predicts the rimed ice number concentration, volume, mass, and mixing ratio, allowing for four degrees of freedom for describing rimed ice particles. The use of these new MP schemes may further improve short-term hail prediction.

This study finds that 0–90-min hail predictions are unreliable near the surface, regardless of the complexity of the MP scheme. The skill levels of surface hail size forecasts are intrinsically tied to the model's ability to correctly represent microphysical processes of hail formation, growth, and decay; such forecasts will thus likely continue to show limited skill unless the treatment of hail, particularly during melting processes, is improved within the model. The forecast fields that perform best in this study are those not impacted by the effects of nearsurface hail melting: m_h and D_{max} at the vertical location of the 0.5° radar tilt, which is relatively far from the radar site, as well as MESH. Thus, this study highlights the importance of improving the performance of MP schemes with regard to the treatment of the melting of hail. A more rigorous examination of the impact of using different rimed ice categorizations on hail prediction in future studies could help clarify the benefits and impacts of more recently developed MP schemes that allow for a spectrum of rimed ice characteristics, such as the P3 scheme. Further research on these problems is strongly desirable as this study and other recent studies have shown the potential of explicit hail prediction.

Acknowledgments. This works was primarily supported by NSF Grant AGS-1261776 as part of the Severe Hail, Analysis, Representation, and Prediction (SHARP) project. Computing was performed primarily on the XSEDE Stampede supercomputer at the University of Texas Advanced Computing Center (TACC). The authors thank Kevin Thomas for support in obtaining data; David Gagne, Cameron Homeyer, and Amy McGovern, who provided useful feedback on the project analysis. We also thank Ted Mansell and two anonymous reviewers, whose constructive comments helped to improve the manuscript.

REFERENCES

- Adams-Selin, R. D., and C. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Wea. Rev.*, 144, 4919–4939, doi:10.1175/MWR-D-16-0027.1.
- Bloom, S. C., L. L. Tackacs, A. M. Da Silva, and D. Ledvina, 1996:
 Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271, doi:10.1175/1520-0493(1996)124<1256:
 DAUIAU>2.0.CO;2.
- Bodine, D. J., R. D. Palmer, and G. Zhang, 2014: Dual-wavelength polarimetric radar analyses of tornadic debris signatures. *J. Appl. Meteor. Climatol.*, 53, 242–261, doi:10.1175/ JAMC-D-13-0189.1.
- Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094, doi:10.1175/ MWR3021.1.
- Changnon, S. A., and J. Burroughs, 2003: The tristate hailstorm: The most costly on record. *Mon. Wea. Rev.*, **131**, 1734–1739, doi:10.1175//2549.1.
- Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, 24, 1121–1140, doi:10.1175/ 2009WAF2222222.1.
- —, —, and —, 2010: Growth of spread in convectionallowing and convection-parameterizing ensembles. *Wea. Forecasting*, **25**, 594–612, doi:10.1175/2009WAF2222318.1.
- Davis, S. M., and J. G. Ladue, 2004: Nonmeteorological factors in warning verification. 22nd Conf. on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., P2.7, https://ams.confex. com/ams/pdfpapers/81766.pdf.

- Dawson, D. T., II, M. Xue, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms. *Mon. Wea. Rev.*, **138**, 1152–1171, https://dx.doi.org/10.1175/ 2009MWR2956.1.
- —, L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2012: Impact of the environmental low-level wind profile on ensemble forecasts of the 4 May 2007 Greensburg, Kansas, tornadic storm and associated mesocyclones. *Mon. Wea. Rev.*, 140, 696–716, doi:10.1175/MWR-D-11-00008.1.
- —, M. Xue, J. A. Milbrandt, and A. Shapiro, 2015: Sensitivity of real-data simulations of the 3 May 1999 Oklahoma City tornadic supercell and associated tornadoes to multimoment microphysics. Part I: Storm- and tornado-scale numerical forecasts. *Mon. Wea. Rev.*, **143**, 2241–2265, doi:10.1175/ MWR-D-14-00279.1.
- Dowell, D., L. J. Wicker, and D. Stensrud, 2004: High resolution analyses of the 8 May 2003 Oklahoma City storm. Part II: EnKFData assimilation and forecast experiments. 22nd Conf. on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., 12.5, https://ams.confex.com/ams/11aram22sls/techprogram/ paper_81393.htm.
 - —, —, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on stormscale analyses. *Mon. Wea. Rev.*, **139**, 272–294, doi:10.1175/ 2010MWR3438.1.
- 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, doi:10.1175/1520-0493(2001)129<2461: AOAPMS>2.0.CO;2.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res., 99, 10143–10162, doi:10.1029/94JC00572.
- —, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor.* Soc., 125, 723–757, doi:10.1002/qj.49712555417.
- Heinselman, P. L., and A. V. Ryzhkov, 2006: Validation of polarimetric hail detection. *Wea. Forecasting*, **21**, 839–850, doi:10.1175/WAF956.1.
- Jewell, R., and J. Brimelow, 2009: Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States. *Wea. Forecasting*, 24, 1592–1609, doi:10.1175/ 2009WAF2222230.1.
- Johnson, M., Y. Jung, D. T. I. Dawson, and M. Xue, 2016: Comparison of simulated polarimetric signatures in idealized supercell storms using two-moment bulk microphysics schemes in WRF. *Mon. Wea. Rev.*, **144**, 971–996, doi:10.1175/ MWR-D-15-0233.1.
- Jung, Y., M. Xue, G. Zhang, and J. M. Straka, 2008: Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Wea. Rev.*, 136, 2228–2245, doi:10.1175/2007MWR2083.1.
 - —, —, and —, 2010: Simulations of polarimetric radar signatures of a supercell storm using a two-moment bulk microphysics scheme. J. Appl. Meteor. Climatol., 49, 146–163, doi:10.1175/2009JAMC2178.1.

- —, —, and M. Tong, 2012a: Ensemble Kalman filter analyses of the 29–30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric radar data. *Mon. Wea. Rev.*, **140**, 1457–1475, doi:10.1175/MWR-D-11-00032.1.
- —, —, Y. Wang, Y. Pan, and K. Zhu, 2012b: Tests of a cycled EnKF data assimilation and forecasts for the 10 May 2010 tornado outbreak in the central US domain. 25th Conf. on Severe and Local Storms, Nashville, TN, Amer. Meteor. Soc., 8A.4, https://ams.confex.com/ams/26SLS/webprogram/ Paper211285.html.
- —, —, —, and —, 2013: Multi-scale ensemble Kalman filter data assimilation and forecasts of the 10 May 2010 tornado outbreak in central United States. *Sixth Conf. on Data Assimilation*, College Park, MD, WMO, http://das6.umd. edu/program/Posters/uploads/Bp09-Jung_Youngsun.pdf.
- Kalnay, E., 2002: Atmospheric Modeling, Data Assimilation, and Predictability. Cambridge University Press, 341 pp.
- —, B. Hunt, E. Ott, and I. Szunyogh, 2006: Ensemble forecasting and data assimilation: Two problems with the same solution? *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University of Press, 157–180.
- Kumjian, M. R., and A. V. Ryzhkov, 2008: Polarimetric signatures in supercell thunderstorms. J. Appl. Meteor. Climatol., 47, 1940–1961, doi:10.1175/2007JAMC1874.1.
- Lei, T., M. Xue, and T. Yu, 2008: Multi-scale analysis and prediction of the 8 May 2003 Oklahoma City tornadic supercell storm assimilating radar and surface network data using EnKF. 13th Conf. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface (IOAS-AOLS), Phoenix, AZ, Amer. Meteor. Soc., 6.4, https://ams. confex.com/ams/89annual/techprogram/paper_150404.htm.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, 22, 1065–1092, doi:10.1175/1520-0450(1983)022<1065: BPOTSF>2.0.CO;2.
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, 21, 289–307.
- Luo, L., M. Xue, K. Zhu, and B. Zhou, 2017: Explicit prediction of hail using multi-moment microphysics schemes for a hailstorm of 19 March 2014 in eastern China. J. Geophys. Res. Atmos., 122, 7560–7581, doi:10.1002/2017JD026747.
- Mansell, E. R., C. Ziegler, and E. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. J. Atmos. Sci., 67, 171–194, doi:10.1175/ 2009JAS2965.1.
- McLaughlin, D., and Coauthors, 2009: Short-wavelength technology and the potential for distributed networks of small radar systems. *Bull. Amer. Meteor. Soc.*, **90**, 1797–1817, doi:10.1175/ 2009BAMS2507.1.
- Melick, C., I. L. Jirak, J. Correia, A. R. Dean, and S. J. Weiss, 2014: Exploration of the NSSL maximum expected size of hail (MESH) product for verifying experimental hail forecasts in the 2014 Spring Forecasting Experiment. 27th Conf. on Severe Local Storms, Madison, WI, Amer. Meteor. Soc., 76, https:// ams.confex.com/ams/27SLS/webprogram/Paper254292.html.
- Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part II: A proposed threemoment closure and scheme description. J. Atmos. Sci., 62, 3065–3081, doi:10.1175/JAS3535.1.
- —, and —, 2006a: A multimoment bulk microphysics parameterization. Part III: Control simulation. J. Atmos. Sci., 63, 3114–3136, doi:10.1175/JAS3816.1.

- —, and —, 2006b: A multimoment bulk microphysics parameterization. Part IV: Sensitivity experiments. J. Atmos. Sci., 63, 3137–3159, doi:10.1175/JAS3817.1.
- —, and H. Morrison, 2013: Prediction of graupel density in a bulk microphysics scheme. J. Atmos. Sci., 70, 410–429, doi:10.1175/ JAS-D-12-0204.1.
- Morrison, H., and J. A. Milbrandt, 2015: Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part I: Scheme description and idealized tests. *J. Atmos. Sci.*, **72**, 287–311, doi:10.1175/JAS-D-14-0065.1.
- —, —, G. H. Bryan, I. Kyoko, S. A. Tessendorf, and G. Thompson, 2015: Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part II: Case study comparisons with observations and other schemes. J. Atmos. Sci., 72, 312–339, doi:10.1175/JAS-D-14-0066.1.
- Ortega, K. L., T. M. Smith, K. L. Manross, K. A. Scharfenberg, W. Arthur, A. G. Kolodziej, and J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530, doi:10.1175/2009BAMS2815.1.
- —, J. M. Krause, and A. V Ryzhkov, 2016: Polarimetric radar characteristics of melting hail. Part III: Validation of the algorithm for hail size discrimination. J. Appl. Meteor. Climatol., 55, 829–848, https://dx.doi.org/10.1175/JAMC-D-15-0203.1.
- Palmer, R. D., and Coauthors, 2011: Observations of the 10 May 2010 tornado outbreak using OU-PRIME: Potential for new science with high-resolution polarimetric radar. *Bull. Amer. Meteor. Soc.*, **92**, 871–891, doi:10.1175/2011BAMS3125.1.
- Park, H. S., V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS. *Wea. Forecasting*, 24, 730–748, doi:10.1175/2008WAF2222205.1.
- Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2014: The analysis and prediction of microphysical states and polarimetric radar variables in a mesoscale convective system using double-moment microphysics, multinetwork radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 141–162, doi:10.1175/MWR-D-13-00042.1.
- Roberts, N., and H. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:10.1175/ 2007MWR2123.1.
- Ryzhkov, A. V., M. R. Kumjian, S. M. Ganson, and P. Zhang, 2013: Polarimetric radar characteristics of melting hail. Part II: Practical implications. J. Appl. Meteor. Climatol., 52, 2871– 2886, doi:10.1175/JAMC-D-13-074.1.
- 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, doi:10.1175/MWR-D-16-0400.1.
- Snook, N., and M. Xue, 2008: Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophys. Res. Lett.*, 35, L24803, doi:10.1029/2008GL035866.
- Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, doi:10.1175/ 2009WAF2222267.1.
- Snook, N., M. Xue, and Y. Jung, 2011: Analysis of a tornadic mesoscale convective vortex based on ensemble Kalman filter assimilation of CASA X-band and WSR-88D radar data. *Mon. Wea. Rev.*, 139, 3446–3468, doi:10.1175/MWR-D-10-05053.1.
- —, —, and —, 2015: Multiscale EnKF assimilation of radar and conventional observations and ensemble forecasting for a

tornadic mesoscale convective system. *Mon. Wea. Rev.*, **143**, 1035–1058, doi:10.1175/MWR-D-13-00262.1.

- —, Y. Jung, J. Brotzge, B. J. Putnam, and M. Xue, 2016: Prediction and ensemble forecast verification of hail in the supercell storms of 20 May 2013. *Wea. Forecasting*, **31**, 811–825, doi:10.1175/WAF-D-15-0152.1.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, 131, 1663–1677, doi:10.1175//2555.1.
- Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-onforecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, 90, 1487–1499, doi:10.1175/2009BAMS2795.1.
- Tao, W., and J. Simpson, 1993: Goddard cumulus ensemble model. Part 1: Model description. *Terr. Atmos. Oceanic Sci.*, 4, 35–72, doi:10.3319/TAO.1993.4.1.35(A).
- Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807, doi:10.1175/MWR2898.1.
- —, and —, 2008a: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part I: Sensitivity analysis and parameter identifiability. *Mon. Wea. Rev.*, **136**, 1630–1648, doi:10.1175/2007MWR2070.1.
- —, and —, 2008b: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, **136**, 1649–1668, doi:10.1175/2007MWR2070.1.
- Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distribution. J. Climate Appl. Meteor., 22, 1764–1775, doi:10.1175/1520-0450(1983)022<1764: NVITAF>2.0.CO:2.
- VandenBerg, M. A., M. C. Coniglio, A. J. Clark, M. A. VandenBerg, M. C. Coniglio, and A. J. Clark, 2014: Comparison of next-day convection-allowing forecasts of storm motion on 1- and 4-km grids. *Wea. Forecasting*, 29, 878–893, doi:10.1175/WAF-D-14-00011.1.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, 130, 1913–1924, doi:10.1175/1520-0493(2002)130<1913: EDAWPO>2.0.CO;2.
- —, and —, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.
- Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. W. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, doi:10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.
- Wyatt, A., and A. Witt, 1997: The effect of population density on ground-truth verification of reports used to score a hail detection algorithm. Preprints, 28th Conf. on Radar Meteorology, Austin, TX, Amer. Meteor. Soc., 368–369.
- Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161–193, doi:10.1007/s007030070003.
- —, and Coauthors, 2001: The Advanced Regional Predicition System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143–165, doi:10.1007/ s007030170027.

- —, M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square root Kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting. J. Atmos. Oceanic Technol., 23, 46– 66, doi:10.1175/JTECH1835.1.
- —, Y. Jung, and G. Zhang, 2010: State estimation of convective storms with a two-moment microphysics scheme and an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **136**, 685– 700, doi:10.1002/gj.593.
- Yussouf, N., and D. Stensrud, 2012: Comparison of singleparameter and multiparameter ensembles for assimilation of radar observations using the ensemble Kalman filter. *Mon. Wea. Rev.*, 140, 562–586, doi:10.1175/ MWR-D-10-05074.1.
- —, E. R. Mansell, L. J. Wicker, D. Wheatley, and D. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, doi:10.1175/MWR-D-12-00237.1.
- —, J. S. Kain, A. J. Clark, N. Yussouf, 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 continuousupdate-cycle storm-scale ensemble system. *Wea. Forecasting*, **31**, 957–983, doi:10.1175/WAF-D-15-0160.1.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.