Explicit Ensemble Prediction of Hail in 19 May 2013 Oklahoma City Thunderstorms and Analysis of Hail Growth Processes with Several Multimoment Microphysics Schemes

JONATHAN LABRIOLA

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

NATHAN SNOOK AND YOUNGSUN JUNG

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

MING XUE

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

(Manuscript received 31 July 2018, in final form 23 January 2019)

ABSTRACT

Hail forecast evaluations provide important insight into microphysical treatment of rimed ice. In this study we evaluate explicit 0–90-min EnKF-based storm-scale (500-m horizontal grid spacing) hail forecasts for a severe weather event that occurred in Oklahoma on 19 May 2013. Forecast ensembles are run using three different bulk microphysics (MP) schemes: the Milbrandt–Yau double-moment scheme (MY2), the Milbrandt–Yau triple-moment scheme (MY3), and the NSSL variable density-rimed ice double-moment scheme (NSSL). Output from a hydrometeor classification algorithm is used to verify surface hail size forecasts. All three schemes produce forecasts that predict the coverage of severe surface hail with moderate to high skill, but exhibit less skill at predicting significant severe hail coverage. A microphysical budget analysis is conducted to better understand hail growth processes in all three schemes. The NSSL scheme uses two-variable density-rimed ice categories to create large hailstones from dense, wet growth graupel particles; however, it is noted the scheme underestimates the coverage of significant severe hail. Both the MY2 and MY3 schemes produce many small hailstones aloft from unrimed, frozen raindrops; in the melting layer, hailstones become much larger than observations because of the excessive accretion of water. The results of this work highlight the importance of using a MP scheme that realistically models microphysical processes.

1. Introduction

Each year, hail causes on average more than $1.4 billion (U.S. dollars) in property and crop damage in the United States (Changnon et al. 2009). Densely populated regions in particular are susceptible to costly hail damage; for example, the 1995 ''Mayfest'' hailstorm near Fort Worth, Texas, caused more than $2 billion in damage and injured 109 people at an outdoor festival (Edwards and Thompson 1998). The cost of hail-related damage is expected to increase as cities expand; most hail events that have caused more than $1 billion in insured losses have occurred since 2001 and the frequency of occurrence of such events is increasing (Changnon 2009).

Extending the warning lead time for severe hail has the potential to mitigate hail-related damage, though this is challenging, as severe hail events often develop quickly. To extend severe weather warning lead time, the National Weather Service (NWS) is shifting from issuing severe weather warnings based upon detection by, for example, radar (warn-on-detection), to a paradigm where warnings will be issued based upon high-resolution numerical weather prediction (NWP) model guidance (warn-on-forecast; Stensrud et al. 2009, 2013) Many studies analyze the skill of convective-scale forecasts; however, only a handful of studies (e.g., Milbrandt and Yau 2006; Snook et al. 2016, hereafter

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

DOI: 10.1175/MWR-D-18-0266.1

S16; Luo et al. 2017; Gagne et al. 2017; Labriola et al. 2017, hereafter L17; Adams-Selin and Ziegler 2016; Luo et al. 2018) assess the skill of explicit hail prediction (hail size and coverage). Hail prediction using storm-scale NWP forecasts remains understudied because of the difficult and complex nature of hail storm prediction, hail growth processes, and challenges associated with hail forecast verification.

The skill of explicit hail prediction is dependent upon the NWP model's ability to predict the complex processes and the storm environment that support the growth of hail; this requires an accurate representation of the hailstorm structure, dynamic, thermodynamic, and microphysical processes, as well as the surrounding environment. Hail grows where supercooled liquid accretes to the surface of rimed ice; however, growth is limited if a large number of hail embryos depletes the supercooled liquid (e.g., Heymsfield 1983). Hail particle growth is strongly correlated with storm updraft strength (e.g., Rasmussen and Heymsfield 1987). Strong updrafts suspend large rimed ice particles in regions of supercooled liquid; however, if an updraft is too strong the particles are more rapidly advected outside of the optimal growth zone. Strong vertical wind shear also contributes to hail growth (e.g., Nelson 1983; Foote 1984; Dennis and Kumjian 2017). Increased deep layer shear extends the storm updraft horizontally and increases the residence time of hail within the growth zone. Because of the complex nature of riming, a large number of microphysical parameterizations that have different treatments for rimed ice were developed with the goal of improving the representation of rimed ice.

Microphysical processes in convective-scale NWP models are mostly parameterized using bulk microphysics (MP) schemes (e.g., Lin et al. 1983; Milbrandt and Yau 2005a). Bulk MP schemes predict the bulk characteristics for hydrometeor species x such as mixing ratio q_x, total number concentration N_{tx}, or reflectivity Z_x to define a commonly used three-parameter gamma particle size distribution (PSD):

$$N(D) = N_{0x} D^{\alpha_x} e^{-\lambda_x D}, \tag{1}$$

where λ_x is the slope parameter, α_x is the shape parameter, and N_{0x} is the intercept parameter (Ulbrich 1983). A few MP schemes (Mansell et al. 2010; Morrison and Milbrandt 2015; Morrison et al. 2015; Milbrandt and Morrison 2016) also have the capacity to predict hydrometeor volume (v_x) and thus explicitly derive density (ρ_x); this is primarily done for rimed ice hydrometeor categories, which undergo large fluctuations in density during growth and melting. Variable density MP schemes (e.g., Mansell et al. 2010) have the ability to update particle fall speeds and rimed ice production.

To save computational time, one or more PSD parameters are often assumed to be constant, though such assumptions can limit the model's ability to realistically represent microphysical processes. For example, most double-moment MP schemes predict the third (q_x) and zeroth (N_{tx}) moments of a PSD to diagnose λ_x and N_{0x}, but assume α_x is constant. Double-moment schemes are able to simulate qualitatively polarimetric signatures (Jung et al. 2012; Johnson et al. 2016; Putnam et al. 2017b), but the schemes often suffer from excessive size sorting (e.g., Milbrandt and Yau 2005b; Dawson et al. 2014; Johnson et al. 2016; Morrison et al. 2015). Triple-moment MP schemes, which additionally predict the sixth moment of the PSD (Z_x) can diagnose α_x. The shape parameter narrows the hydrometeor size spectra and limits size sorting by having the weighted fall speeds of a hydrometeor type converge toward a singular value (Milbrandt and Yau 2005a; Dawson et al. 2014).

Convective-scale forecasts have been shown to predict spatial extent and size of hail at the surface with some skill. Multiple studies (e.g., S16; L17; Luo et al. 2017, 2018) have used predicted hail quantities (mass, number concentration) to identify regions where hail reaches the surface. Simulated radar products such as the maximum estimated size of hail (MESH; Witt et al. 1998a) have also been used to derive surface hail size from model output (S16; L17; Luo et al. 2017, 2018). Another method, the Thompson hail method (Thompson et al. 2018), uses model diagnosed hail PSDs to approximate the largest ''observable'' hailstone. Variants of this method have been used to diagnose the maximum surface hail size and distinguish regions of hail growth (Milbrandt and Yau 2006; S16; L17; Luo et al. 2018). The skill of hail forecasts derived directly from model output is strongly dependent upon choice of MP scheme; to account for model biases, machine learning algorithms can be applied to CAM output to create storm-based hail forecasts (Gagne et al. 2015, 2017; McGovern et al. 2017).

Convective-scale explicit hail forecasts are subject to significant error associated with the NWP model (e.g., initial conditions, microphysics). Many recent convective-scale studies use cycled ensemble Kalman filter (EnKF; Evensen 1994, 2003) data assimilation (DA) to generate initial conditions (e.g., Dowell et al. 2004; Snook et al. 2011, 2012, 2015; Dawson et al. 2012; Jung et al. 2012; Putnam et al. 2014, 2017a; Yussouf et al. 2013, 2016; Wheatley et al. 2014; Schwartz et al. 2015; Skinner et al. 2018). This technique is preferred because flow-dependent error covariances derived from the forecast ensemble allow the DA system to update unobserved

variables such as temperature, pressure, and microphysical variables from available observations (e.g., Tong and Xue 2005, 2008a). An EnKF DA system can thus use radar observations, which indirectly observe hydrometeor information, to improve the microphysical state variables of hail producing storms (S16; L17).

Hail forecast verification is a substantial challenge. Hail events are underreported in rural areas, away from major highways, as well as in cases when a more extreme severe weather event (e.g., one or more tornadoes) occurs in addition to hail (Doswell et al. 2005; Witt et al. 1998b; Allen and Tippett 2015). Additionally, hail sizes that correspond to familiar circular or spherical objects (e.g., dimes, softballs) are overreported (Sammler 1993), as such objects are commonly used as size references in reports from the public. New observational databases attempt to mitigate observed hail size biases. The meteorological phenomena identification near the ground (mPING; Elmore et al. 2014) phone application requires users to report hail in size increments of 0.25 in., and the severe hail analysis and verification experiment (SHAVE; Ortega et al. 2009) directly interviews the public to create a hail observation database with high spatial resolution. Nevertheless, surface-based reports remain inadequate for objective hail forecast verification.

The next-generation radar (NEXRAD; Crum et al. 1993) system is the only observational platform that performs full-volume scans of the atmosphere at a temporal frequency (approximately 5 min) sufficient to capture the rapid evolution of convective storms. Radar-derived hail products, such as MESH, serve as a proxy for hail size and can be used for forecast verification (e.g., S16). Although MESH varies in skill throughout the United States, it is not subject to population bias and thus superior to ground-based reports (Cintineo et al. 2012).

Polarimetric radars provide observations such as differential reflectivity Z_{dr}, copolar correlation coefficient ρ_{hv}, and differential phase Φ_{dp}, in addition to Z and radial velocity V_r. With these additional observations, one is able to infer hydrometeor properties such as shape, size, orientation, and phase (Kumjian and Ryzhkov 2008). Hydrometeor classification algorithms (HCAs; Park et al. 2009) apply a fuzzy logic algorithm to both single-and dual-polarization radar data to classify the dominant hydrometeor type. This study implements the Putnam et al. (2017b) HCA, along with hail size discrimination algorithm membership functions from Ortega et al. (2016) to classify the dominant hail size into one of three bins: nonsevere (5 mm), severe (25 mm), and significant severe (50 mm). HCAs often classify hail size with more skill than MESH (Ortega et al. 2016), although biased observations degrade the skill of the classifications.

Previous studies (i.e., S16; L17; Luo et al. 2017, 2018) evaluated hail forecasts produced using single-moment, double-moment, and triple-moment MP schemes; however, no studies have evaluated hail forecasts produced using a variable density MP scheme, which has the ability to represent a spectrum of rimed ice particle characteristics. In this study double-moment, triple-moment, and variable-density double-moment MP schemes are used to produce explicit ensemble hail forecasts for a severe hailstorm event that occurred in the Oklahoma City, Oklahoma, metropolitan area. Surface hail size forecasts are subjectively and objectively verified against HCA output (Ortega et al. 2016) using the Brier skill score (Brier 1950), reliability diagrams, and relative operating characteristic curves. In addition to forecast evaluations, a microphysical budget analysis is conducted. This study analyzes hail growth and decay processes predicted by each MP scheme to identify sources of forecast error, and determine the impact of microphysical processes on the prediction of hail.

The rest of the paper is organized as follows. In section 2 we provide a brief overview of the 19 May 2013 Oklahoma City severe hail event and describe the experimental design, including the DA and forecast model settings. Hail forecast verification statistics and an in-depth analysis of hail growth and decay processes are presented and discussed in section 3. A summary and some further discussion on the results are given in section 4.

2. Case, experiment configuration, and verification procedures

a. *Case overview*

On 19 May 2013 an upper-level low pressure system developed in the northern plains of the United States as a negatively tilted trough moved over the southern plains. An 1800 UTC (1300 LST) sounding launched from the Norman Weather Forecast Office (WFO) indicated the environment was both highly unstable (CAPE $\sim$4878 J kg^{-1}) and strongly sheared [0–6-km shear $\sim$49 kt ($\sim$25 m s^{-1})], ideal for the development of supercell thunderstorms. Thunderstorms initiated along a dryline boundary west of Oklahoma City starting at approximately 1945 UTC. Three supercell thunderstorms (Fig. 1a) produced large regions of severe hail, and localized regions of significant-severe hail, in the southwestern Oklahoma City metropolitan area (Figs. 1b,c). Aside from producing large hail throughout

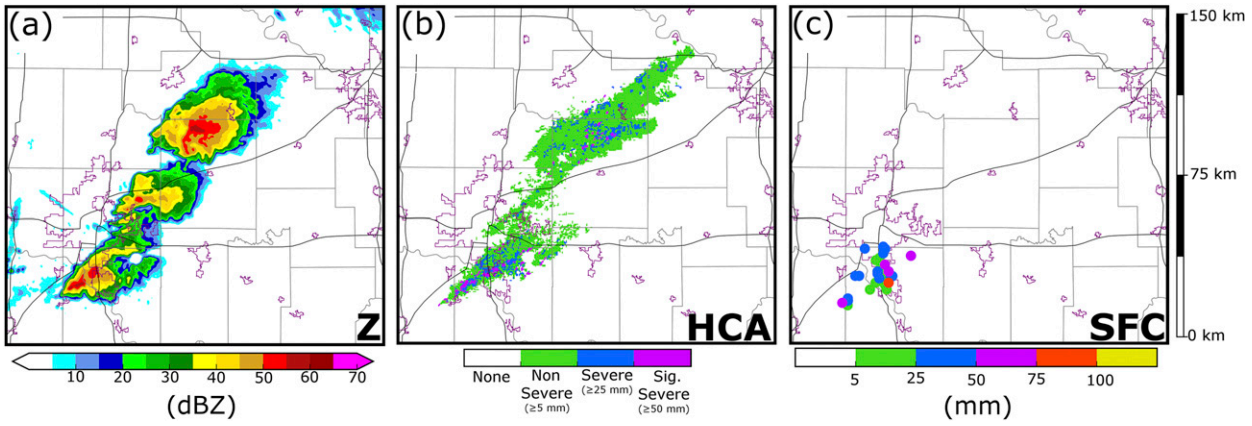

FIG. 1. (a) KTLX observed Z at the lowest radar tilt (0.5°), valid for 2220 UTC. (b) A swath of the largest observed hail size between 2155 and 2320 UTC. Hail size is determined by applying an HCA to the lowest radar tilt of KTLX and KINX (0.5°) and merging the subsequent output. (c) Surface-based hail size reports (SPC and mPING) between 2155 and 2320 UTC. On the background maps thin black lines are highways and purple lines are urban boundaries.

the region, these storms also produced two long-track tornadoes that produced EF-3 damage [on the enhanced Fujita (EF) scale] in Carney, Oklahoma, and EF-4 damage in Shawnee, Oklahoma. For further discussion on this event we refer the reader to Carlin et al. (2017) and Wienhoff et al. (2018).

b. *Prediction model configurations*

All experiments use the Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001) as the prediction model. The model domain consists of 573 × 515 × 53 grid points with a 500-m horizontal grid spacing. The model grid is stretched in the vertical, with a minimum vertical grid spacing of 50 m at the surface and an average vertical grid spacing of 425 m. The ARPS model settings follow those of S16 and L17. Model physics (Xue et al. 2000, 2001) include NASA Goddard Flight Center shortwave and longwave radiation, surface fluxes calculated from surface drag coefficients, surface temperature and volumetric water, a two-layer soil model, and a 1.5-order turbulent kinetic energy-based subgrid-scale turbulence parameterization (Deardorff 1980; Klemp and Wilhelmson 1978; Moeng 1984).

Hail forecasts are produced using three different MP schemes: the Milbrandt and Yau (2005a) double-moment MP scheme (MY2), the Milbrandt and Yau (2005a) triple-moment MP scheme (MY3), and the National Severe Storms Laboratory (NSSL) double-moment variable-density-rimed ice MP scheme (Mansell et al. 2010). Forecast maximum surface hail size (D_{max}) is calculated using a variant of the Thompson hail algorithm introduced in S16. D_{max} is calculated at the first model level above the surface (~25 m AGL) and is defined as the largest diameter for which the hail PSD has at least 1 stone per 10 000 m^4 (Milbrandt and Yau 2006).

The MY2 scheme is selected for this experiment because it has been used in multiple recent hail prediction studies (e.g., S16, L17, Luo et al. 2017, 2018). The MY2 and MY3 schemes use the same prognostic equations for hydrometeor mass mixing ratio and number concentration; however, MY3 also predicts hydrometeor Z; Z is used to diagnose the PSD shape parameter. This parameter narrows the size distribution and limits further size sorting by causing weighted fall speeds to converge toward a singular value (Dawson et al. 2014). The NSSL scheme includes mass mixing ratio and number concentration, as well as prognostic equations for both hail and graupel volume (and thus density). Variable density is used to update particle fall speeds and conversion rates between hydrometeor species (e.g., dense wet-growth graupel is converted to hail; Mansell et al. 2010). The minimum number concentration threshold is set to $10^{-8}\,m^{-3}$ for the MY2 and MY3 schemes; the default setting ($10^{-3}\,m^{-3}$) led to excessive removal of hail, including some large hail near the surface. Unlike the MY2 and MY3 schemes, which threshold via number concentration and mass, the NSSL cheme only thresholds via mixing ratio, and thus does not remove as many hail grid points near the surface.

Vertical cross sections taken through the hail core of a thunderstorm, after a 60-min forecast, are used to demonstrate the impact of reducing the minimum number threshold. The default MY2 scheme (Figs. 2a,c,e) removes near-surface hail that is relatively sparse in both mass ($<5 \times 10^{-3}\,g\,m^{-3}$) and number ($<5 \times 10^{-5}\,m^{-3}$); however, the updated scheme (Figs. 2b,d,f)

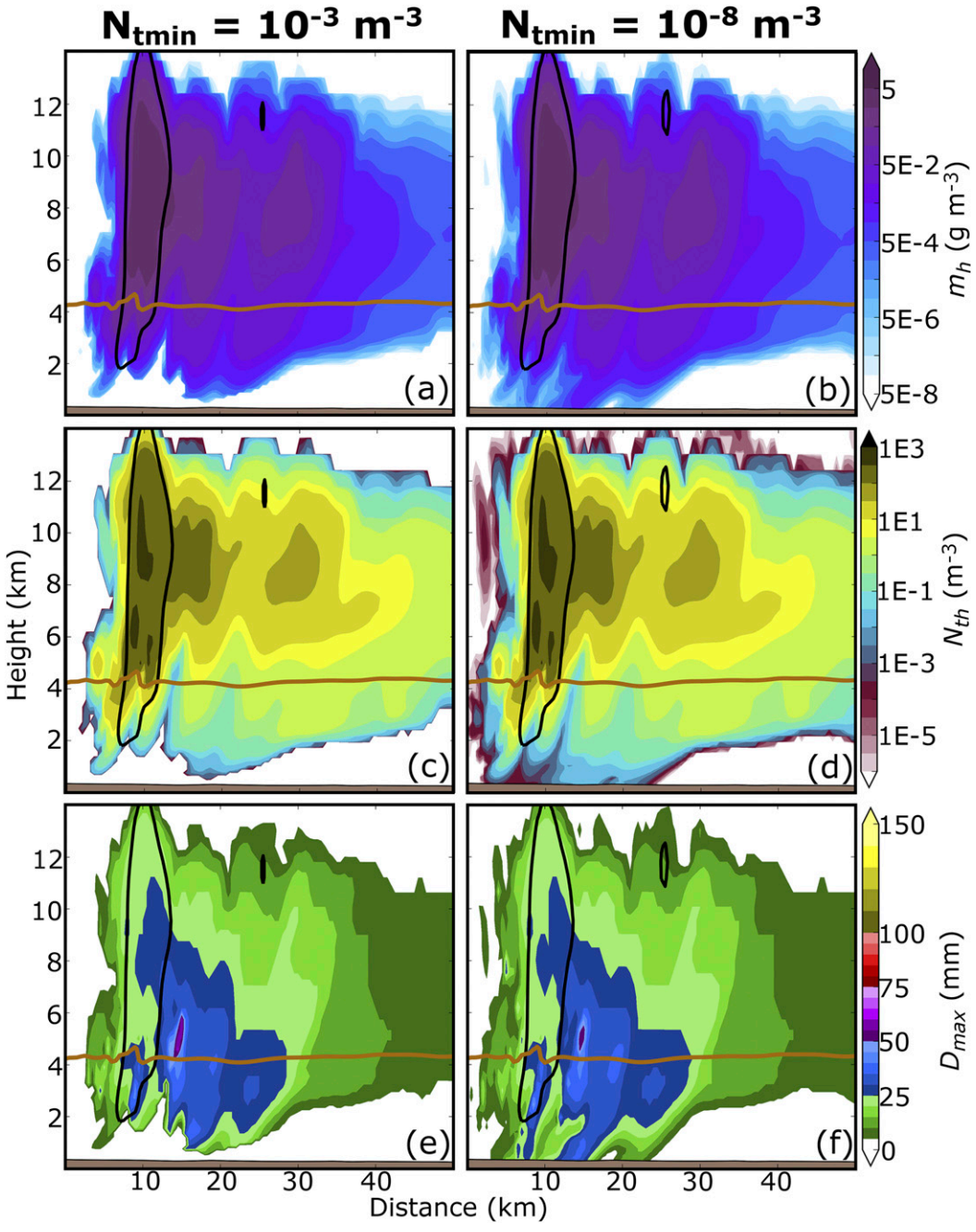

FIG. 2. Vertical cross sections taken through the hail core of a supercell thunderstorm after a 60-min forecast. Forecasts are run using the MY2 scheme where the minimum number concentration is set to (a),(c),(e) 10^{-3} m^{-3} and (b),(d),(f) 10^{-8} m^{-3}. Cross sections are of (a),(b) m_h, (c),(d) N_{th}, and (e),(f) D_{max}. In each plot a brown horizontal line represents the 0°C isotherm and black contours indicate regions where updraft velocity exceeds 5 m s^{-1}.

predicts more hail to reach the surface that is classified as observable by the D_{max} algorithm (Fig. 2f). Decreasing the minimum number threshold in the MY2 and MY3 schemes substantially increases surface hail size and coverage (Figs. 3c,d) compared to the default scheme (Figs. 3a,b) because the lower threshold limits the removal of near-surface hail. Altering the minimum number threshold affects mostly near-surface hail size forecasts (Fig. 3), while model state variables (e.g., w, q_h, N_{th}) remain relatively similar between the forecasts (Figs. 2a–d) in the rest of the storm. Reducing the minimum number concentration

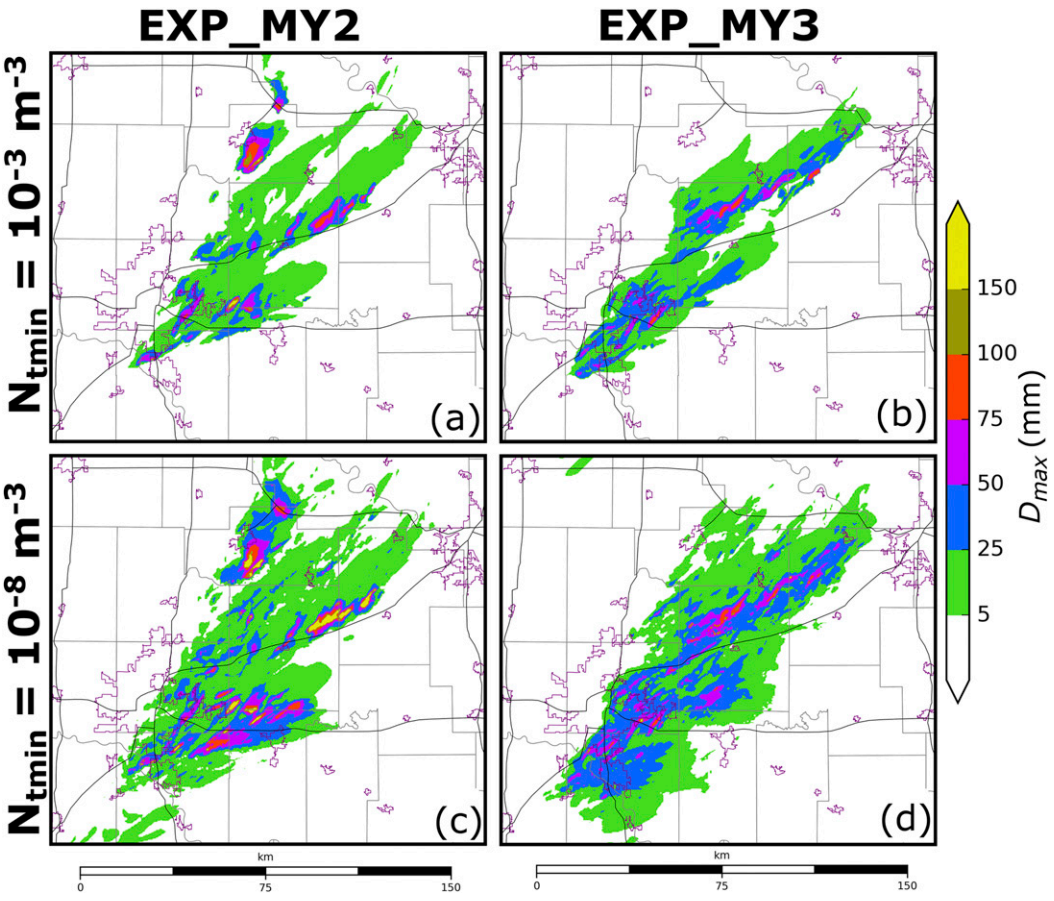

FIG. 3. Swaths of forecast maximum surface D_{max} calculated at every model time step between 2155 and 2320 UTC. Forecasts are run using the (a),(c) MY2 and (b),(d) MY3 schemes, where the minimum number concentration threshold is set to (a),(b) 10^{-3} m^{-3} or (c),(d) 10^{-8} m^{-3}. Background maps are the same as in Fig. 1.

threshold does not increase the number of spurious storms for this study.

c. *Radar observation operators and EnKF DA configurations*

The EnKF DA requires accurate and efficient radar reflectivity observational operators consistent with the MP schemes used. The Rayleigh scattering approximation (Jung et al. 2008a,b) is typically used to simulate the Z of ice hydrometeors in our prior EnKF studies (Dawson et al. 2010; Snook et al. 2011, 2015, S16; Jung et al. 2012; Putnam et al. 2014; L17). While this observation operator is computationally efficient, it frequently overestimates the Z of large ice particles that cause Mie scattering (e.g., hail). The T-matrix radar simulator/observation operator (Jung et al. 2010) accounts for Mie scattering; it uses the T-matrix calculations (Vivekanandan et al. 1991; Bringi and Chandrasekar 2001) to derive the scattering amplitude of hydrometeors for different size bins and Z is then calculated by integrating the scattering amplitude over a given PSD. Recent updates have optimized the computational efficiency of the T-matrix method; for example, Putnam et al. (2018, manuscript submitted to *Mon. Wea. Rev.*) calculates scattering amplitude for select PSD parameters and uses lookup tables and interpolation to derive Z. This study uses the Putnam et al. (2018, manuscript submitted to *Mon. Wea. Rev.*) T-matrix-based operator, which accounts for Mie scattering at computational speeds comparable to the Rayleigh scattering approximation.

Adjustments are made to the radar forward observation operator to accommodate the different MP schemes. To reduce the size of scattering amplitude lookup tables, the Rayleigh scattering approximation is assumed for the NSSL snow category, since snow particles are small enough that we can assume Rayleigh scattering with little to no error. To simulate the melting layer, a variant of the Jung et al. (2008a) melting model is used to create ice–water mixtures for both graupel and

hail. The ice–water hybrid category (which is generated offline) is determined via a ratio between ice and rainwater mixing ratios. For example, the hail–water mixing ratio (f_{wh}) is:

$$f_{wh} = \frac{q_r}{q_r + q_h}. \quad (2)$$

While computationally efficient, this calculation does not account for shedding and thus allows too much liquid to accumulate on the surface of an ice particle. In regions where q_r is large, this can cause the diameter of melting ice to overinflate. To maintain computational efficiency but to produce more realistic Z we modify the ice–water mixture number concentration so that the mean mass diameter of the melting ice species is preserved during melting.

A 40-member ensemble is used for both the EnKF DA and forecast periods during this experiment, and one ensemble experiment is run using each of the microphysical schemes listed above (MY2, MY3, NSSL). Hereafter, the ensembles run using the MY2, MY3, and NSSL schemes are referred to as EXP_MY2, EXP_MY3, and EXP_NSSL, respectively. Initial and boundary conditions are provided by the Center for Analysis and Prediction of Storms (CAPS) EnKF-initialized storm-scale ensemble forecast (SSEF; Jung et al. 2014). The EnKF-initialized SSEF is a 40-member ensemble of convection allowing (4-km horizontal grid spacing) forecasts that were run using the Advanced Research version of the Weather Research and Forecasting (WRF) Model (ARW; Skamarock et al. 2008) for the NOAA Hazardous Weather Testbed Spring Forecasting Experiment. The SSEF model domain spans the contiguous United States, forecasts are initialized at 1800 UTC 19 May 2013 and are run until 0000 UTC 20 May 2013. Additional information about the SSEF is provided in Kong (2013).

At 1900 UTC, our 500-m domain ensemble is initialized by interpolating from the SSEF. To introduce additional storm-scale perturbations, a 2-dimensional recursive filter is used to create smoothed Gaussian perturbations with horizontal and vertical correlation scales of 6 km and 3 km, respectively, which are added to the interpolated 1900 UTC SSEF ensemble. The storm-scale perturbations have standard deviations of 2 m s^{-1}, 2 K, and 0.5 g kg^{-1} for horizontal wind components (u, v), potential temperature θ, and water vapor mixing ratio q_v, respectively. Sixty-minute spin-up ensemble forecasts are launched from the perturbed initial conditions and run until 2000 UTC when EnKF DA is first performed. The CAPS EnKF system (Xue et al. 2006; Tong and Xue 2008b), which is based upon the Whitaker and Hamill (2002) ensemble square root filter (EnSRF) algorithm, is used to assimilate observations every 10 min from 2000 to 2150 UTC, the time of final EnKF analysis. This configuration resulted in the improved suppression of spurious storms.

TABLE 1. A list of the assimilated observations along with the corresponding model state variables that are updated by each observation.

Observation	Updated state variables
Surface	u, v, θ, q_v
Profiler	u, v
Sounding	u, v, θ, q_v
KTLX Z	u, v, w, θ, q_v, q_x
KTLX V_r	u, v, w, q_v

Assimilated conventional observations include surface (Oklahoma Mesonet, ASOS, and AWOS) and profiler observations. Conventional observations are not used to update vertical velocity (w) because in preliminary tests, unreliable covariances led to the development of spurious updrafts. A list of the model state variables updated by each observation type is provided in Table 1. Observation error settings largely follow S16: surface observation errors are assumed to be 1.5 m s^{-1} for u and v, 2.0 K for θ, and 2.0 K for dewpoint temperature (T_d). The surface observation covariance localization radius is set to 300 km in the horizontal and 6 km in the vertical. Wind profiler observations are used to update background u and v winds and assume an error of 2.5 m s^{-1} for both variables and a covariance localization radius of 80 km in the horizontal and 6 km in the vertical.

Radar observations can infer important microphysical information and are assimilated. Data from the nearest Oklahoma City NEXRAD radar (KTLX) are interpolated horizontally to the model grid column locations but are preserved in the vertical at the height of the radar beam (Xue et al. 2006); Z and V_r observations are also thinned in the horizontal to one observation every 1 km in regions of precipitation (observed $Z > 5$ dBZ) and to one observation every 2 km in regions of clean air (observed $Z < 5$ dBZ). Radar data are thinned horizontally to control the ensemble spread reduction due to assimilation of very dense observations. KTLX observation errors are assumed to be 4.0 m s^{-1} for V_r and 6.0 dBZ for Z (Snook et al. 2013), and the covariance localization function (Gaspari and Cohn 1999) cutoff radius is set to 3 km in both the horizontal and vertical.

As multicategory microphysical variables are underconstrained by observed Z, large uncertainties may exist within analyzed variables. Uncertainties may increase for multimoment schemes because of the larger number

of predicted variables (Xue et al. 2010); for example, the MY2, MY3, and NSSL schemes predict 12, 17, and 15 state variables, respectively. Multimoment schemes better represent the nonmonotonic relationship between q_x and Z, which improves the representation of simulated radar variables (e.g., Putnam et al. 2014), and motivates the use of multimoment schemes in an EnKF during radar DA (Jung et al. 2012). To limit the influence of unreliable covariances we update only hydrometeor mixing ratio during DA (Table 1). Updating only hydrometeor mixing ratios may lead to the slower filter convergence rate but we found the system produces fewer spurious storms than updating more variables.

The relaxation-to-prior-spread (Whitaker and Hamill 2012) method is used to inflate the posterior ensemble spread to 95% of the prior ensemble spread for all updated variables. Although N_{tx}, υ_x, and Z_x are not updated (and thus never inflated), the variables are modified to avoid unrealistic PSD behavior during inflation. During inflation, N_{tx} is updated to preserve hydrometeor mean mass diameter, υ_x is updated to preserve particle density, and Z_x is updated to preserve PSD shape parameter. Subjective comparisons between MY3 predicted Z after inflation and radar-observed Z (not shown) indicate that the diagnostic update can produce the posterior Z that compares well with observations.

d. *Verification procedure*

HCA output from the two nearest NEXRAD radars (KTLX and KINX) is used to verify hail size forecasts. Observations from KTLX and KINX are interpolated to the 500-m model grid and a 9-point smoother is applied to the Z_{dr} and ρ_{hv} fields to reduce noise. The HCA is performed for radar data on the lowest elevation angle within 120 km of a given radar—120 km is the maximum distance Ortega et al. (2016) used to evaluate hail size. Although the HCA is applied to the lowest radar tilt, the membership functions derived by Ortega et al. (2016) are tuned to detect surface hail size. Next, the data are merged (the largest hail size is selected where radar volumes overlap). Because of small-scale noise, a final smoothing filter is applied to the merged HCA output. The smoothing filter decreases (increases) hail size detections when the four closest grid points are smaller (larger); when a detection is updated, it is set to match the four surrounding grid points. Merged HCA output (Fig. 1b) is used for both subjective and objective forecast verification in this study.

Hail forecast verification is challenging in part because large surface hail is a highly localized phenomenon. When using a traditional neighborhood ensemble probability (NEP; Schwartz et al. 2010) approach, small position errors between ensemble members can lead to mostly nonzeros probabilities that are difficult to verify. Neighborhood maximum ensemble probability (NMEP; Ben Bouallègue and Theis 2014; Schwartz and Sobash 2017) considers the probability of an event occurrence within distance i; this expands both the number of probabilistic and observed events and increases the sharpness of the forecast. For objective verification the NMEP method defined in Schwartz and Sobash (2017) is used to verify the probability of an event occurrence within 10 km of a grid point with a smoothing Gaussian filter of 10 km. This distance was selected because it accounts for small forecast storm displacement errors, but maintains the high spatial resolution necessary for warn-on-forecast applications.

Probabilistic hail size forecasts are evaluated using the Brier skill score (Brier 1950), reliability diagrams, and relative operating characteristic (ROC; Mason 1982) curves. The Brier skill score is defined by three separate components: forecast reliability, resolution, and uncertainty. Forecast reliability compares predicted probability of occurrence to observed frequency, which is used to infer model bias and is most commonly displayed in terms of a reliability diagram. In an unbiased system, observed frequency and predicted probability are similar; if there is a mismatch, the model suffers an overprediction (underprediction) bias when the predicted probability of occurrence is larger (smaller) than the observed frequency. ROC curves determine the model's ability to discern between events and nonevents. ROC curves are generated by plotting probability of detection (POD) against probability of false detection (POFD) for increasing probability thresholds. Forecast skill is usually determined by the area under the ROC curve (AUC), if the AUC exceeds 0.7 then the forecast is skilled at predicting the event (Buizza et al. 1999).

3. Results

a. *Forecast evaluation and verification*

The 90-min-long hail size forecasts starting from the final EnKF analyses at 2150 UTC are verified between 2155 and 2320 UTC, a period during which multiple supercell thunderstorms produce severe and significant severe hail over the Oklahoma City metropolitan area (Figs. 1b,c). Accumulated swaths of the forecast maximum D_{max} (calculated at every model time step) are verified against accumulated swaths of observed maximum hail size as indicated by HCA output (calculated every radar volume scan). We note that radars observe the atmosphere far less frequently (~5 min) than

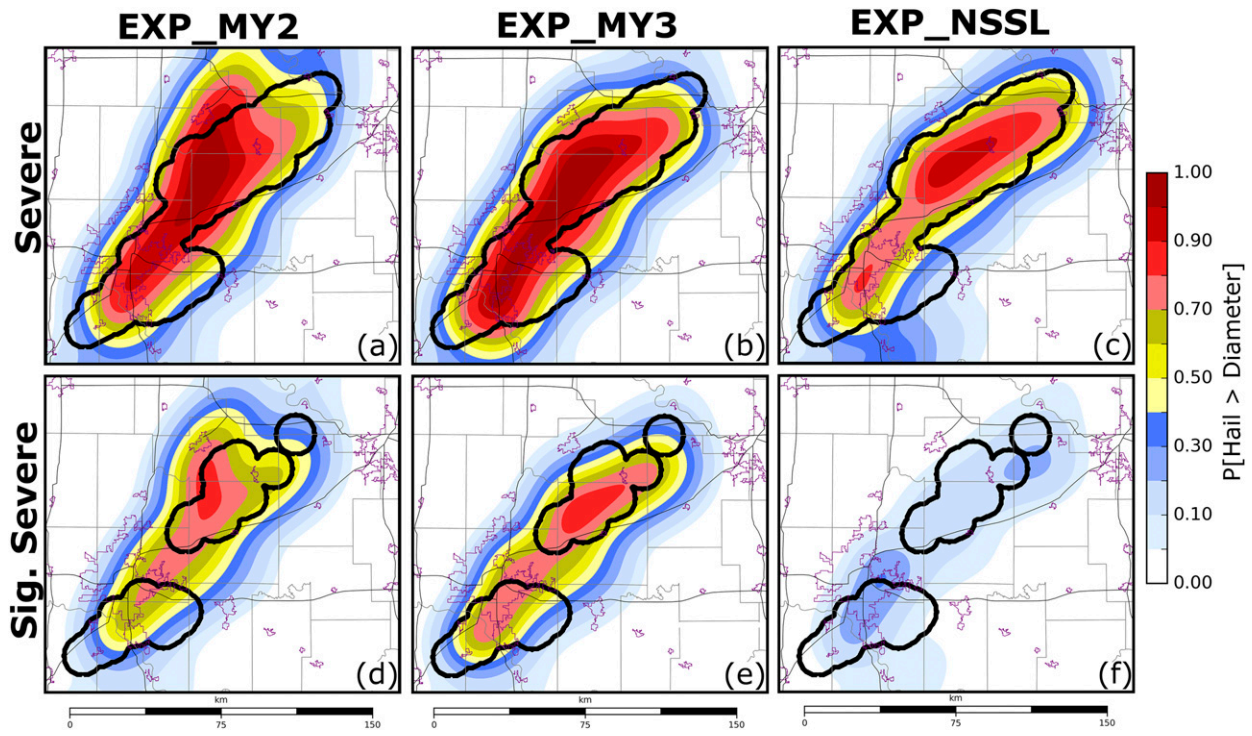

FIG. 4. The NMEP of (a)–(c) severe hail and (d)–(f) significant severe hail predicted using the (a),(d) MY2, (b),(e) MY3, and (c),(f) NSSL MP schemes between 2155 and 2320 UTC. Thick black contours represent the locations of observed severe or significant severe hail according to HCA output. Background maps are the same as in Fig. 1.

a typical storm-scale model time step (~1 s); however, we believe that the higher temporal resolution of the model output better captures the extent and evolution of hail in the model storms, and thus choose to perform verification with the highest temporal frequency data available. Hail size forecasts are verified for two separate diameter thresholds including severe hail (diameter $>$ 25 mm) and significant severe hail (diameter $>$ 50 mm).

For this study, we first subjectively evaluate the NMEP of severe and significant severe hail and note forecast biases. The predicted probability of severe hail (P[severe hail]) (Figs. 4a–c) is generally high ($>$ 0.7) in regions where severe hail is observed. Both EXP_MY3 (Fig. 4b) and EXP_NSSL (Fig. 4c) predict the northernmost hail swath to be oriented toward the northeast, suggesting these ensembles predict the rightward deviation in storm motion commonly observed in supercell thunderstorms (Bunkers et al. 2000). Unlike the above two ensembles, EXP_MY2 (Fig. 4a) predicts a bifurcation in maximum P[severe hail] values for the northernmost storm. This bifurcation is due to the production of spurious storms in EXP_MY2 forecasts. EXP_MY2 forecasts produce many spurious storms, this increases the overall coverage of severe hail and the total area where P[severe hail] $>$ 0.4 (Fig. 4a). Both EXP_MY2 (Fig. 4d) and EXP_MY3 (Fig. 4e) predict P[significant severe hail] to exceed 0.5 where HCA output (Fig. 1b) indicates the occurrence of significant severe hail.

Reliability diagrams indicate that ensembles predict the spatial coverage of severe hail (Fig. 5a) with relatively little bias. Although Brier skill scores (Table 2) vary slightly (0.532–0.633), all three ensembles produce skillful severe hail forecasts. For significant severe hail forecasts, the Brier skill score (Table 2) decreases for all three ensembles (0.132–0.463), indicating the ensembles predict more extreme events with less skill. EXP_NSSL predicts significant severe hail with the least skill (Table 2), in part because the ensemble produces under confident significant severe hail size forecasts (Fig. 5b) and rarely predicts P[significant severe] $>$ 0.2 (Fig. 4f). Despite lower Brier skill scores (Table 2), EXP_MY2 and EXP_MY3 produce reliable significant severe hail forecasts (Fig. 5b) and predict significant severe hail with higher confidence than EXP_NSSL.

Contrary to the other verification metrics, ROC curves (Fig. 6) indicate the ensembles predict both severe and significant severe hail with a high level of skill (AUC $>$ 0.92). Large variations in P[significant severe] between the ensembles does not appear to have a large impact on the AUC score. It is noted that verifications

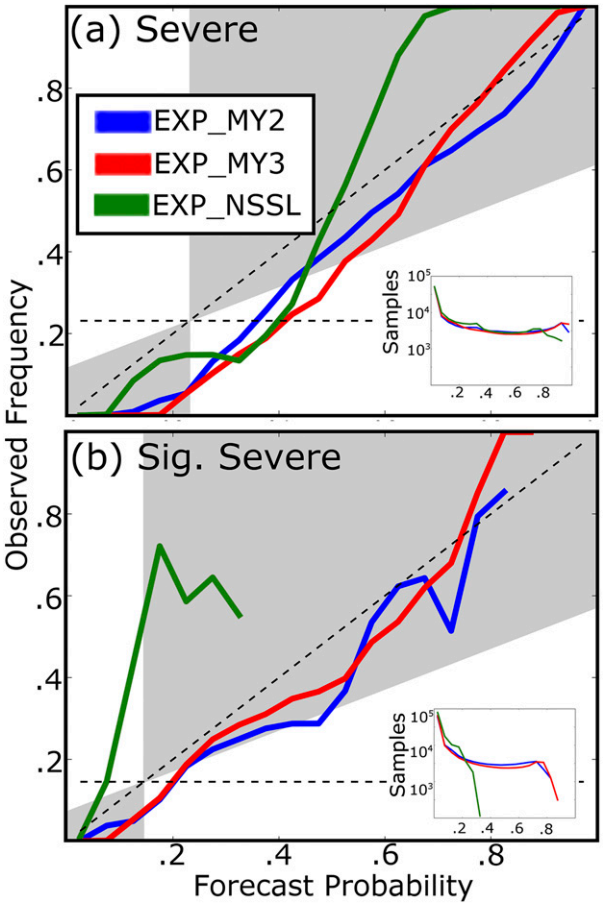

FIG. 5. Reliability diagrams for the probabilistic (a) severe hail forecasts and (b) significant severe hail forecasts shown in Fig. 4. In the diagrams blue lines correspond to EXP_MY2, red lines to EXP_MY3, and green lines to EXP_NSSL. Forecast exhibit skill when the forecast reliability falls within the gray shading. The number of samples per forecast probability bin is provided for both (a) severe hail and (b) significant severe hail.

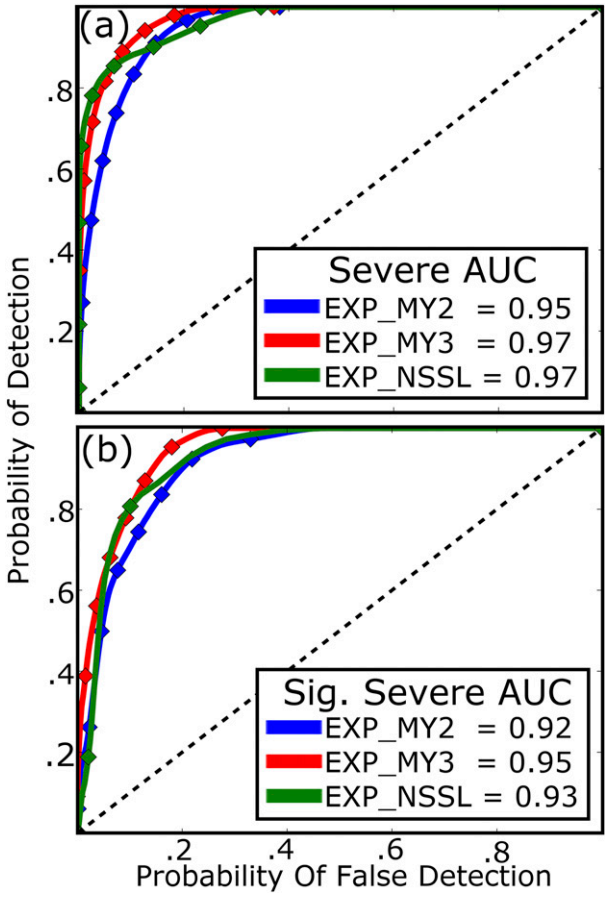

FIG. 6. ROC curves for the probabilistic (a) severe and (b) significant severe hail forecasts shown in Fig. 4. In the diagrams blue lines correspond to EXP_MY2, red lines to EXP_MY3, and green lines to EXP_NSSL. Forecast AUC scores are provided, scores that are greater than 0.7 predict severe or significant severe hail with skill.

are performed for a single event and additional case studies are needed; however, findings from this study are believed to be meaningful.

Swaths of D_{max} produced by single ensemble members (Fig. 7) provide insight into hail predictions. EXP_MY2 overpredicts the size of hail at the surface, with both members 20 (Fig. 7a) and 40 (Fig. 7d) predicting hail to exceed 150 mm (~6 in.) in diameter. The HCA output (Fig. 1b) is unable to determine the upper limit on observed hail size for hail exceeding 50 mm because the algorithm classifies all hail 50 mm and larger as significant severe; however, mPING reports suggest that hail, which occurred during this event, did not exceed 3 in. (~75 mm) (Fig. 1c) in diameter. The MY2 scheme surface hail size overprediction bias is noted in several previous studies; both L17 and Luo et al. (2018) note the scheme produces excessively large surface hail. The MY3 scheme employs the same mass and number concentration tendency equations as the MY2 scheme; as such, EXP_MY3 also overpredicts hail size (Figs. 7b,e). Despite an overprediction bias, EXP_MY3 members (Figs. 7b,e) predict smaller D_{max} values than EXP_MY2 members (Figs. 7a,d). Forecast differences are in part because microphysical processes (e.g., sedimentation)

TABLE 2. The Brier skill score for EXP_MY2, EXP_MY3, and EXP_NSSL. Skill scores are calculated for severe hail and significant severe hail forecasts between 2155 and 2320 UTC. The corresponding NMEP used to derive the Brier skill score is provided in Fig. 4.

	Severe	Significant severe
EXP_MY2	0.532	0.352
EXP_MY3	0.633	0.463
EXP_NSSL	0.629	0.132

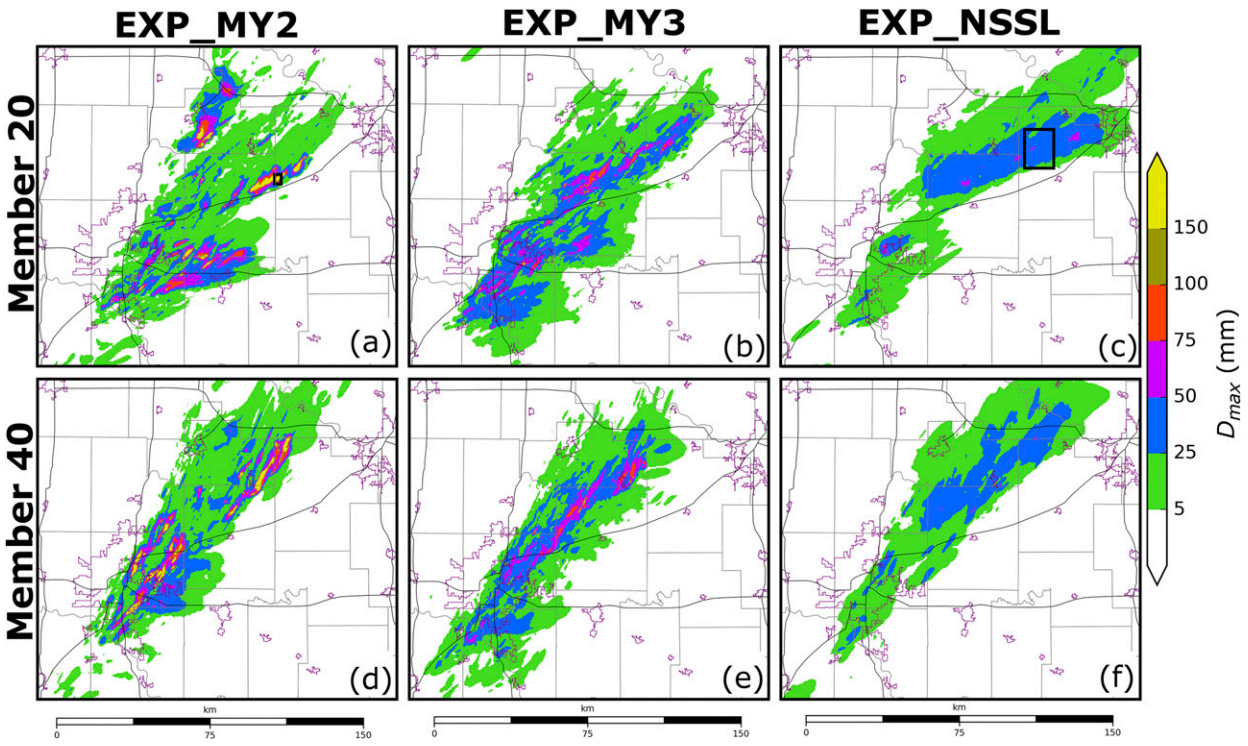

FIG. 7. Swaths of forecast maximum surface D_{max} calculated at every model time step between 2155 and 2320 UTC. Swaths are created for ensemble members (a)–(c) 20 and (d)–(f) 40, which are run using the (a),(d) MY2, (b),(e) MY3, and (c),(f) NSSL MP schemes. These members are representative of the other ensemble member forecasts. Black polygons in (a) and (c) highlight the locations where hail processes are analyzed later in the study. Background maps are the same as in Fig. 1.

are sensitive to variations in the shape parameter (Milbrandt and Yau 2005a).

EXP_NSSL forecasts (Figs. 7c,f) qualitatively most closely match HCA output in terms of maximum hail size. EXP_NSSL members predict the northernmost storm to produce more severe hail than what is observed in HCA output (Fig. 1b), though this is partly because we compare low temporal frequency (~5 min) radar observations to high temporal frequency (~1 s) NWP output. Additionally, the NSSL scheme predicts the storms to produce hail cores that are wider than observations.

b. Hail production processes

Similar to the microphysical budget analysis conducted in L17, this study analyzes hail growth and decay process predicted by each MP scheme to identify sources of forecast error. Hail microphysical tendency terms (e.g., accretion, sublimation, melting) from a single time step are integrated along a horizontal plane that spans the verification domain (Fig. 4). Terms are plotted as a function of average model level height in order to illustrate the model's treatment of hail both above and below the 0°C isotherm. Vertical cross sections are taken through the hail core of a supercell thunderstorm during the microphysical budget analysis to illustrate the impact of microphysical processes on the representation of hail. We conduct the microphysical analysis at 2250 UTC; this is one hour into the forecast period when multiple supercell thunderstorms are producing hail.

Hail growth assumptions in MP schemes impact the representation of hail within the model. In EXP_MY2 (Fig. 8b) and EXP_MY3 (Fig. 8d), hailstones are almost exclusively created by the three-component accretion of rain; this was first reported in Johnson et al. (2016). Three-component accretion of rain occurs when rainwater freezes to the surface of a frozen hydrometeor species to produce a subsequent third ice hydrometeor type (e.g., rain freezing onto snow to create hail). A static threshold is used to determine the resulting hydrometeor category. The NSSL scheme hail category does not contain frozen raindrops (Mansell et al. 2010), instead hail is created from dense graupel ($\rho_g > 800\,\mathrm{kg\,m^{-3}}$) in wet growth conditions (Fig. 8f). EXP_NSSL N_{th} tendencies (Fig. 8f) are approximately one to two orders of magnitude smaller than EXP_MY2 (Fig. 8b) or EXP_MY3 (Fig. 8d) N_{th}

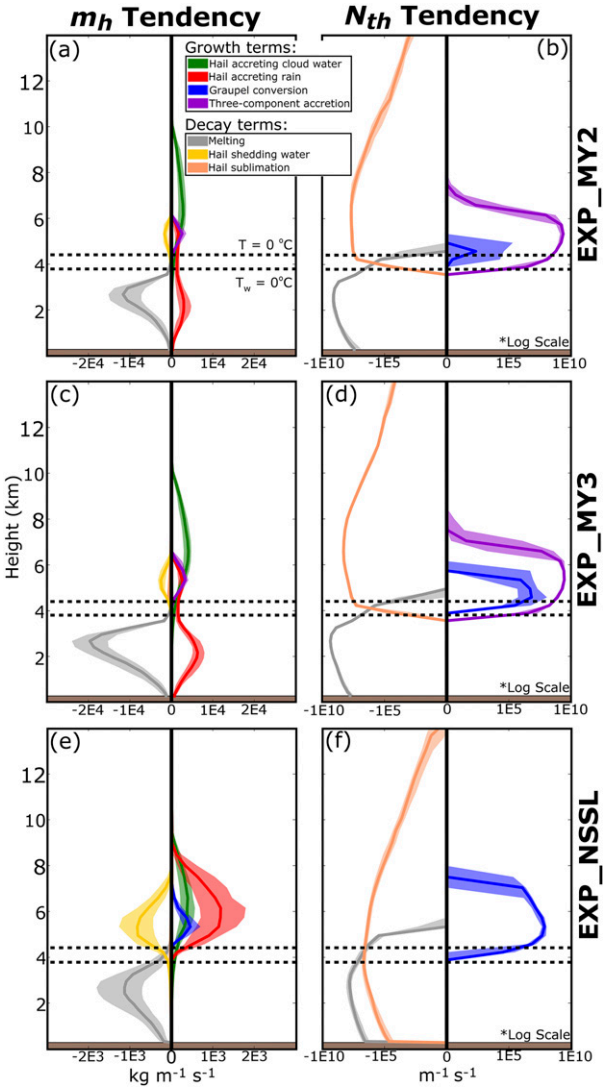

FIG. 8. Microphysical source and sink terms for (a),(c),(e) m_h and (b),(d),(f) N_{th} predicted by (a),(b) EXP_MY2, (c),(d) EXP_MY3, and (e),(f) EXP_NSSL at 2250 UTC. Microphysical terms are integrated over each model vertical level spanning the verification domain shown in Fig. 4. The terms are plotted as a function of average model height per vertical level. Terms are summed for each ensemble member, the interquartile range of the ensemble is shaded and median value is denoted as a dark line. Only terms that contribute greater than 10 kg m^{-1} for m_h and 10 m^{-1} for N_{th} are plotted. In each plot the upper horizontal dashed black line is the average 0°C isotherm height and the lower horizontal dashed line is the average wet-bulb temperature 0°C isotherm height.

tendencies because dense, wet-growth graupel particles are much less numerous than frozen raindrops.

The three-component accretion of rain increases the number of hailstones above the 0°C isotherm in the EXP_MY2 (Fig. 8b) and EXP_MY3 (Fig. 8d) forecasts; however this process contributes relatively little to the total mass of hail aloft because frozen raindrops are relatively small (Figs. 8a,c). The two largest m_h growth processes for the MY2 (Fig. 8a) and MY3 (Fig. 8c) schemes are the accretion of rain and cloud water. Similar to L17, the MY2 (Fig. 8a) and MY3 (Fig. 8c) schemes both predict hail to accrete rainwater near the surface despite the wet-bulb temperature exceeding 0°C. The consequences of this process are discussed later in section c of this chapter. Unlike MY2 and MY3, the NSSL scheme assumes all accreted water is shed in and beneath the melting layer for hail, eliminating the possibility of m_h growth. EXP_NSSL m_h tendencies (Fig. 8e) are smaller in magnitude than either EXP_MY2 (Fig. 8a) or EXP_MY3 (Fig. 8c) because the rate of accretion and melting is proportional to hail number concentration. The NSSL scheme predicts fewer hailstones than the MY2 and MY3 schemes, and thus predicts less rain and cloud water droplets are accreted.

Vertical cross sections of hail-related microphysical variables (Fig. 9) are evaluated to understand the impact of microphysical processes on the behavior of hail. Although the EXP_MY2 (Figs. 7a,d) and EXP_MY3 (Figs. 7b,e) predict storms to produce $D_{max} > 100$ mm at the surface, cross sections indicate hail is much smaller above the 0°C isotherm (Figs. 9g,h). The NSSL scheme predicts storms to produce smaller D_{max} values at the surface ($\leq$75 mm) (Figs. 7c,f); however, D_{m_h} remains large throughout the vertical depth of the updraft (Fig. 9i). Differences in microphysical assumptions cause EXP_MY2 and EXP_MY3 to predict larger hail than EXP_NSSL beneath the 0°C isotherm, but smaller hail aloft.

Both the MY2 (Fig. 9d) and MY3 (Fig. 9e) schemes predict N_{th} to be large above the 0°C isotherm at 5 km above mean sea level; this is the approximate level where the three-component accretion of rain is maximized (Figs. 8b,d). Since frozen raindrops are relatively small in diameter, the three-component accretion of rain decreases the D_{m_h} above the 0°C isotherm for both EXP_MY2 (Fig. 9g) and EXP_MY3 (Fig. 9h) members. The small hail particles aloft are advected downwind of the updraft, creating a plume of hail that extends more than 40 km. Although the MY2 and MY3 schemes produce large plumes of hail aloft, the small hail particles in the plumes mostly melt before reaching the surface. S16 previously noted the MY2 scheme melts hail too quickly and underpredicts the spatial coverage of hail at the surface. The underprediction bias noted in S16 is mitigated when a less stringent minimum N_{th} threshold of 10^{-8} m^{-3} is used, this allows hailstones that are more sparse in number concentration to reach the surface.

Above the 0°C isotherm, the NSSL scheme produces fewer but larger hailstones than either the MY2 or MY3

schemes because hail is created from wet-growth graupel. Large hail behaves differently from small rimed ice particles: first, upper-level winds do not advect the NSSL scheme hail as far downstream, and most large hail ($D_{m_h} > 4.5$ mm) thus remains confined to the boundaries of the storm updraft region (Fig. 9i). Second, the larger hailstones have an increased terminal velocity, and thus higher fall speeds. An increased fall speed causes a higher percentage of hailstones to reach the surface before melting and produces relatively wide swaths of severe hail (Figs. 7c,f) that closely match observations (Figs. 1b,c).

c. *Hail melting processes*

In this study, EXP_MY2 forecasts (Figs. 7a,d) predict $D_{\max}$ to frequently exceed 150 mm (approximately 6 in.), despite HCA output (Fig. 1b) indicating that surface hail is mostly nonsevere or severe (<50 mm in diameter). Much of the hail growth in EXP_MY2 forecasts occurs in the melting layer; EXP_MY2 member 20, which is representative of most EXP_MY2 members, predicts $D_{\max}$ to increase from 25 mm to more than 150 mm while traversing the 0°C isotherm (Fig. 10a). L17 suggests much of this growth is because hail in the MY2 scheme accretes rain and cloud water in and beneath the melting layer.

To better understand MY2 scheme hail growth, we integrate the total microphysical tendency of m_h and N_{th} over a subdomain that encompasses only the hail core of EXP_MY2 member 20 (Fig. 7a). Microphysical tendencies vary substantially throughout a single storm, integrating the tendencies over the hail core (Figs. 10b,c) provides more insight into hail growth processes than integrating over the verification domain (Figs. 8a,b). Above the 0°C isotherm N_{th} increases rapidly due to the three-component accretion of rain (Fig. 10c). The introduction of many small, frozen raindrops decreases the mean size of hail aloft. Beneath the 0°C isotherm, m_h increases (Fig. 10b) because hail accretes more rain and cloud water than it melts and sheds. At approximately the same layer below the freezing level, N_{th} slowly decreases due to melting (Fig. 10c), this is because accretional growth does not increase N_{th}. When m_h increases but N_{th} decreases the slope parameter of the hail PSD λ_h becomes shallow (Fig. 11b) and causes $D_{\max}$ to exceed 150 mm (Fig. 10a). Accretional growth maintains or expands hail size, allowing more hail to reach the surface.

EXP_MY3 predicts $D_{\max}$ (Figs. 7b,e) to exceed 150 mm less frequently than EXP_MY2 (Figs. 7a,d), even though both MP schemes use the same m_h and N_{th} tendency equations. Hail PSDs selected from three grid points within a hail core of EXP_MY3 member 5 (Fig. 11a) indicate that when $\alpha_h > 0$ the PSD follows a gamma distribution which narrows the PSD and moderates $D_{\max}$ (Fig. 11b). Multiple studies (e.g., Milbrandt and Yau 2005b; Kumjian and Ryzhkov 2012; Dawson et al. 2014) have documented that α_h increases due to sedimentation. Although the MY3 scheme moderates $D_{\max}$, hail continues to accrete liquid in and beneath the melting layer, which causes EXP_MY3 to predict hail greater than 100 mm in diameter (Figs. 7b,e).

EXP_MY2 and EXP_MY3 member 20 forecasts are rerun without the accretion of rain and cloud water beneath the 0°C isotherm (EXP_MY2_NOAC and EXP_MY3_NOAC, respectively) to demonstrate the sensitivity of surface hail size to the accretion of water beneath the 0°C isotherm. EXP_MY2_NOAC (Fig. 12a) and EXP_MY3_NOAC (Fig. 12b) predict the maximum surface hail size to be smaller in diameter than EXP_MY2 (Fig. 7a) and EXP_MY3 (Fig. 7b) forecasts, and more closely resemble observations (Figs. 1b,c). It is noted that EXP_MY2_NOAC (Fig. 12a) predicts hail to exceed 100 mm; however, most of this hail is attributed to initialization from the EXP_MY2 member 20 analysis, and within 10 min of forecast initialization most of the storms cease to produce hail greater than 100 mm in diameter. Forecast experiments EXP_MY2_NOAC (Fig. 12a) and EXP_MY3_NOAC (Fig. 12b) are in agreement with the conclusions of this study and suggest the accretion of liquid water beneath the 0°C isotherm is important to the production of large hailstones ($D_{\max} > 100$ mm).

Coverage of severe hail increases at the surface for EXP_MY2_NOAC (Fig. 12a) and EXP_MY3_NOAC (Fig. 12b) when compared to EXP_MY2 (Fig. 7a) and EXP_MY3 (Fig. 7b) forecasts. In all three MP schemes the N_{th} melting tendency term is used to preserve D_{m_h} during melting; however, both the MY2 and MY3 schemes reduce the N_{th} melting term by 90% when accretional growth occurs. Except in updrafts where large quantities of liquid water are accreted, when the N_{th} melting term is reduced D_{m_h} decreases because m_h decreases more quickly than N_{th}. When accretion of liquid water is eliminated beneath the 0°C isotherm the N_{th} melting tendency is not reduced, and thus EXP_MY2_NOAC and EXP_MY3_NOAC predict hail to more frequently exceed 25 mm in diameter (Fig. 12).

EXP_NSSL exhibits limited hail growth across the 0°C isotherm. EXP_NSSL member 20, for example, predicts a change of only 0–5 mm across the melting layer (Fig. 13a), compared to a change of 150 mm across the melting layer in EXP_MY2 member 20 (Fig. 10a). Microphysical tendency terms in the hail core of EXP_NSSL member 20 (Fig. 7c) suggest m_h (Fig. 13b) and N_{th} (Fig. 13c) decrease at approximately the same

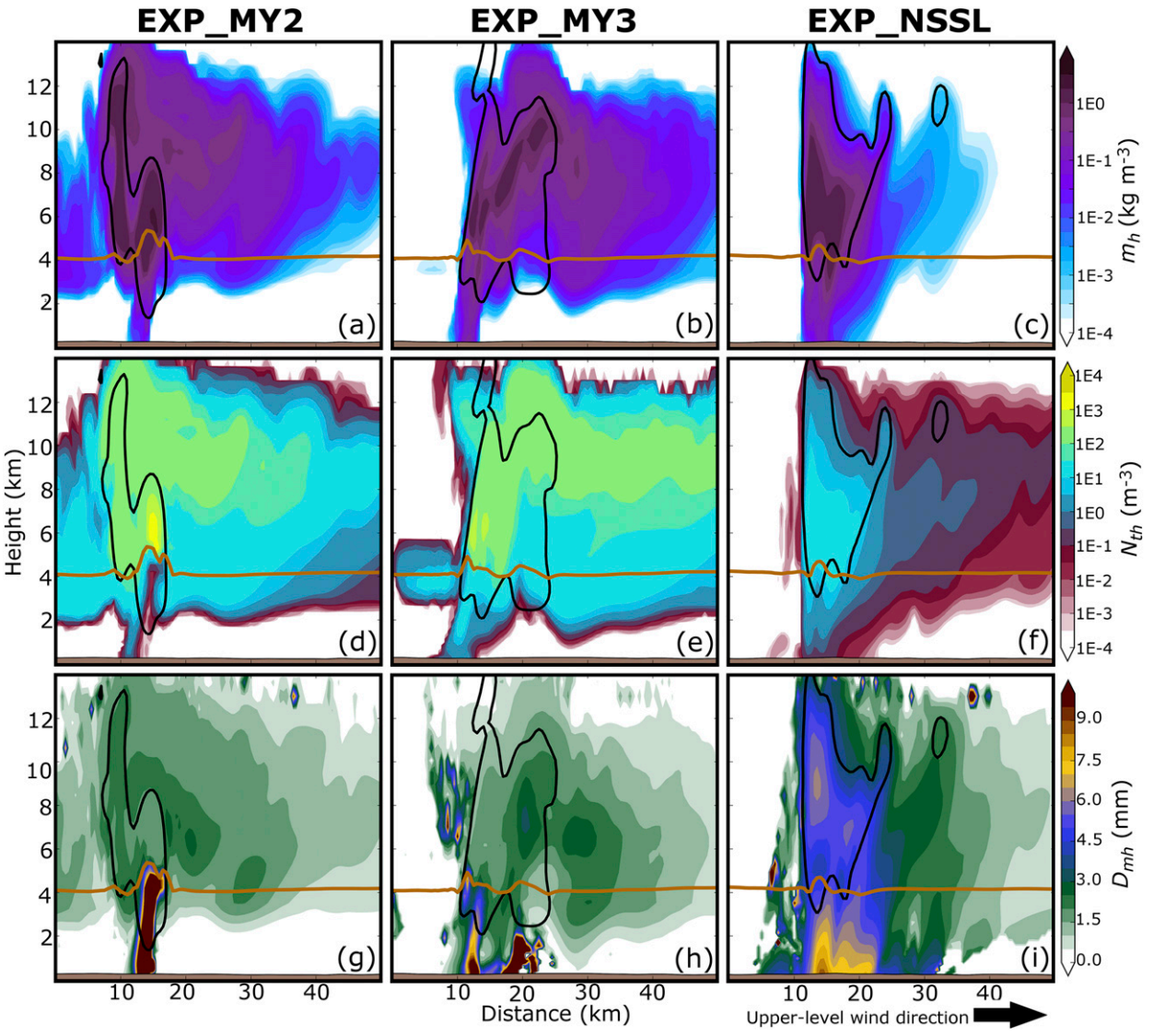

FIG. 9. A vertical cross section taken through the hail core of the northernmost supercell thunderstorm predicted by member 20 of (a),(d),(g) EXP_MY2, (b),(e),(h) EXP_MY3, and (c),(f),(i) EXP_NSSL at 2250 UTC. Cross sections are of (a)–(c) m_h, (d)–(f) N_{th}, and (g)–(i) D_{m_h}. In each plot a brown horizontal line represents the 0°C isotherm and black contours indicate regions where updraft velocity exceeds 5 m s^{-1}.

rate beneath the 0°C and isotherm. Both terms decrease at the same rate because hail sheds all accreted liquid when the air temperature is warmer than 0°C. Although the NSSL scheme limits hail growth within the melting layer, EXP_NSSL forecasts severe and significant severe hail (Figs. 4c,f) because large hailstones are created from wet-growth graupel above the 0°C isotherm.

Most MP schemes assume ice is dry; this assumption has important consequences on the treatment of hail beneath the 0°C isotherm. Collection of water beneath the 0°C isotherm will increase the ice water fraction and potentially hail size (Lesins and List 1986); however, if too much liquid water is collected than the hailstone will shed and evaporate excess liquid (Rasmussen and Heymsfield 1987; Chong and Chen 1974). Instead of representing the collection of water, the MY2 and MY3 schemes freeze rain and cloud water onto the surface of hail (i.e., accretion). Since the accreted water is frozen, excess liquid cannot be removed via shedding, and instead m_h increases when the air temperature is well above freezing (Fig. 10b). In contrast, the NSSL scheme does not allow any liquid water collection beneath the 0°C isotherm (Fig. 13b); this limits a potential hail growth process within the scheme and potentially limits the maximum surface hail size. Without the ability to represent ice–water hybrid categories these

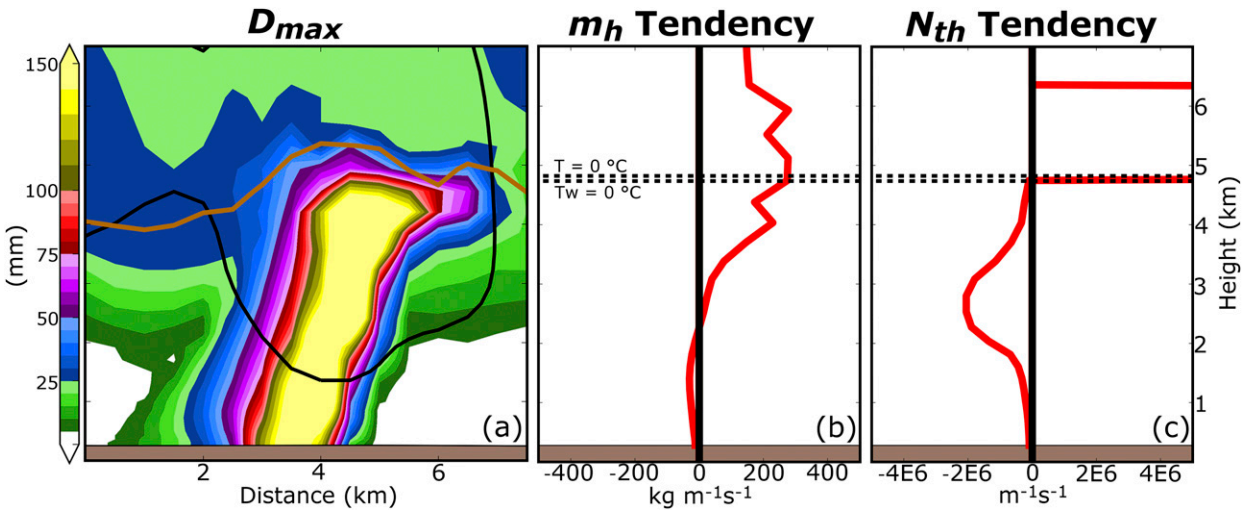

FIG. 10. (a) A vertical cross section of D_{max} taken through the hail core of the northernmost storm predicted by EXP_MY2 member 20 at 2250 UTC. (b),(c) The total microphysical tendency for m_h and N_{th}, respectively, are also shown. The total microphysical tendency is integrated over each model vertical level in the storm hail core, which is highlighted by the black polygon in Fig. 7a. The tendencies are plotted as a function of average model level height. In the vertical cross section in (a) the horizontal brown line represents 0°C isotherm and the black contour represents where updraft velocity exceeds 5 m s^{-1}. In (b) and (c) the upper horizontal dashed black line is the 0°C isotherm and the bottom dashed line is the wet-bulb 0°C isotherm.

MP schemes must continue to make assumptions on the treatment of hail during melting.

4. Summary and discussion

In this study we evaluate 90-min-long EnKF-initialized explicit hail ensemble forecasts at 500-m grid spacing for the 19 May 2013 Oklahoma City, Oklahoma, supercell hailstorm event. For the initial conditions, we assimilate all available surface, profiler, and radar (reflectivity and radial velocity) observations every 10 min, over a 110-min period ending at the final analysis time of 2150 UTC. Three ensembles, consisting of 40 members each, are run for both the data assimilation and forecast periods. Each ensemble is similarly configured but uses one of three microphysics (MP) schemes: the Milbrandt and Yau (2005a) double-moment MP scheme (MY2), the Milbrandt and Yau (2005a)

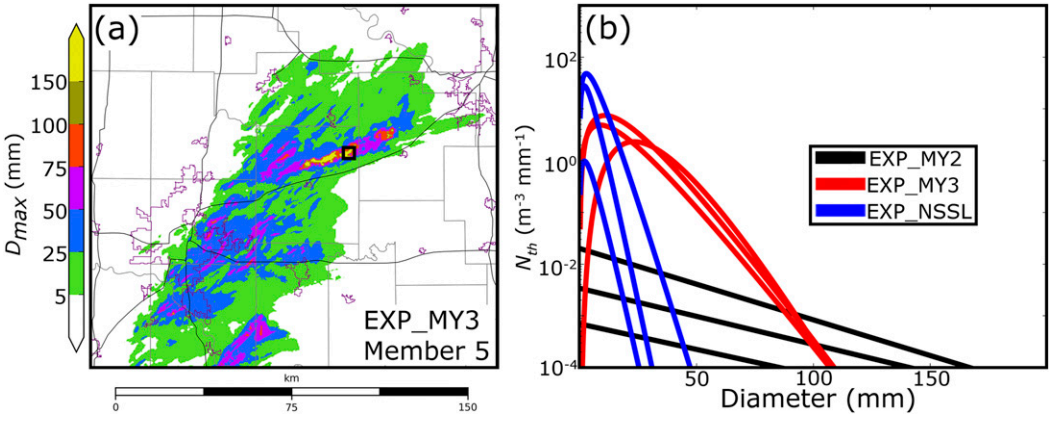

FIG. 11. (a) A Swath of forecast maximum surface D_{max} calculated at every model time step between 2155 and 2320 UTC for EXP_MY3 member 5. A black polygon indicates where hail PSDs are sampled for the MY3 scheme. Background maps are the same as in Fig. 1. (b) A comparison of hail core PSDs for EXP_MY2 member 20 (Fig. 7a), EXP_MY3 member 5 (Fig. 11a), and EXP_NSSL member 20 (Fig. 7c) at 2250 UTC. PSDs are sampled from within the black polygons in Fig. 7a, Fig. 11a, and Fig. 7c, respectively. Because the minimum y-axis value is 10^{-4} m^{-3} mm^{-1} the x intercept is considered to be the maximum observable hail size D_{max}. For EXP_MY3, a rightward shift in the PSD peak is caused by an increased shape parameter.

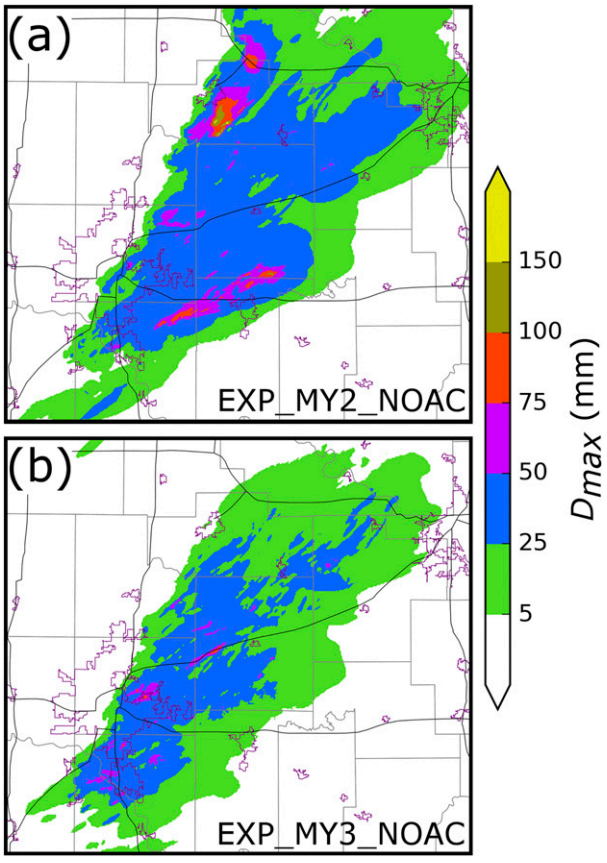

FIG. 12. Swaths of forecast maximum surface D_{max} calculated at every model time step between 2155 and 2320 UTC. Forecasts are initialized from member 20 analyses of (a) EXP_MY2 and (b) EXP_MY3. Forecasts (a) EXP_MY2_NOAC and (b) EXP_MY3_NOAC are run with modified MY2 and MY3 schemes, respectively, where the accretion of rainwater and cloud water does not occur beneath the 0°C isotherm. Background maps are the same as in Fig. 1.

triple-moment MP scheme (MY3), or the NSSL double-moment variable-density-rimed ice MP scheme (Mansell et al. 2010). Ensembles run using these schemes are hereafter referred to as EXP_MY2, EXP_MY3, and EXP_NSSL, respectively.

Surface hail size forecasts are verified against output from a hydrometeor classification algorithm (HCA). We determine the forecast maximum hail size (D_{max}) via hail particle size distributions (PSDs), which approximates the largest observable hailstone at each model grid point. Area under the relative operating characteristic (AUC) scores indicate all three MP schemes predict the coverage of severe (diameter >25 mm) and significant severe (diameter >50 mm) hail with a high level of skill (≥0.92); albeit Brier skill scores suggest the schemes predict significant severe hail coverage with less skill than severe hail coverage. The ensembles produce reliable hail size forecasts, with the exception of EXP_NSSL, which produces under confident significant severe hail forecasts. Although EXP_MY2 and EXP_MY3 produce skillful severe and significant severe hail forecasts, both ensembles overestimate the maximum surface hail size. Qualitatively, EXP_NSSL most closely resembles HCA output but the scheme predicts limited significant severe hail coverage.

Hail production tendencies are evaluated during the forecast period to understand hail size forecast differences and biases. Storms predicted by either EXP_MY2 or EXP_MY3 generally produce a large number of small hail particles above the 0°C isotherm because the MP schemes create hail via the three-component accretion of rain (i.e., frozen raindrops). The small hail particles are advected downstream of the updraft to produce large plumes of hail aloft that extend more than 40 km downstream of the updraft. EXP_NSSL predicts storms to produce fewer but relatively larger hail particles above the 0°C isotherm because the NSSL scheme hail category consists of only dense graupel that has undergone wet growth.

Hail size forecast skill is also determined by how MP schemes model hail in and beneath the melting layer. Because water fraction is not predicted by the MY2, MY3, or NSSL schemes, neither scheme is able to represent the collection of liquid water in the melting layer. Both the MY2 and MY3 MP schemes assume all water collected by hail is converted to ice both above and beneath the melting layer; this process causes model predicted hail to exceed 150 mm (~6 in.) in diameter beneath the 0°C isotherm. In regions where hail is exceptionally large the MY3 scheme often increases the shape parameter (α_h) of hail. Increasing α_h narrows the hail PSD and moderates the most extreme hail sizes (diameter >150 mm); however, EXP_MY3 continues to overpredict hail size. The NSSL scheme sheds all accreted liquid water when the air temperature is warmer than 0°C. While this assumption prevents the rapid growth of hail in the melting layer, it is not realistic when compared to observations (e.g., Rasmussen et al. 1984) and neglects potential hail growth via the collection of water (Lesins and List 1986).

To further improve the hail forecast skill, we must use MP schemes that accurately model hail growth processes. MP schemes such as the predicted particle properties (P3; Morrison and Milbrandt 2015; Morrison et al. 2015; Milbrandt and Morrison 2016) scheme and the Ice-Spheroids Habit Model with Aspect-ratio Evolution (ISHMAEL; Jensen et al. 2017) scheme explicitly predict the mass and volume of accreted rime to avoid converting between predefined hydrometeor species. Hydrometeor conversion between unrimed and rimed

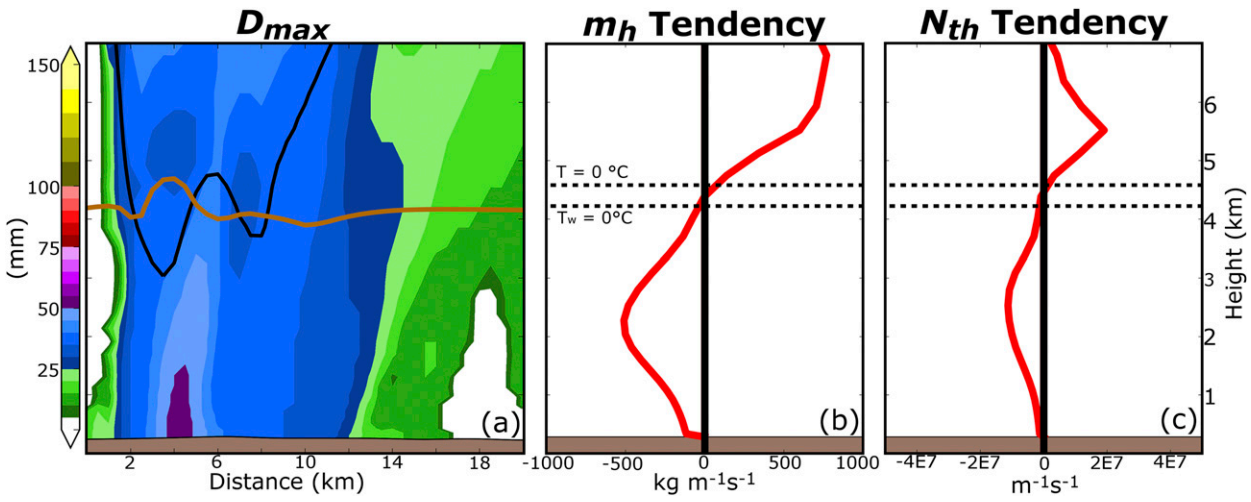

FIG. 13. (a) A vertical cross section of D_{max} taken through the hail core of the northernmost supercell predicted by EXP_NSSL member 20 at 2250 UTC. (b),(c) The total microphysical tendency for m_h and N_{th}, respectively, are also shown. The total microphysical tendency is integrated over each model vertical level in the storm hail core, which is highlighted by a polygon in Fig. 7c. The total tendency is plotted as a function of average model level height. In the vertical cross section in (a) the horizontal brown line represents 0°C isotherm and the black contour represents where updraft velocity exceeds 5 m s^{-1}. In (b) and (c) the upper horizontal dashed black line is the 0°C isotherm and the bottom dashed line is the wet-bulb 0°C isotherm.

ice categories is a large source of model error, this is most evident in EXP_MY2 and EXP_MY3 forecasts when frozen rain is converted to hail. ISHMAEL also predicts particle shape, which is shown to impact the spatial distribution of ice due to size sorting (Jensen et al. 2018). Further investigation into the inclusion of prognostic water fraction equations such as those in Ferrier (1994) could also improve representation of ice within the melting layer. Integrating these novel MP schemes into a hail prediction system could be part of an investigation of the benefit of including additional microphysical variables and processes.

Data assimilation can also be used to improve the microphysical state of hailstorms; however, we must determine if microphysical variables are being realistically updated by the DA system. Performing data assimilation when using a multimoment MP scheme is nontrivial. Because of the large number of predicted hydrometeor properties, microphysical variables are typically under constrained by observations (i.e., reflectivity). Microphysical variables also have very nonlinear relations with observations; Xue et al. (2010) notes this introduces large errors into the initial conditions. For this study, we update only hydrometeor mixing ratios to limit the degrees of freedom of the model that are updated. Although this technique produces a stable configuration, the resulting analysis is suboptimal because D_{max} is often too large in early cycles when forecast errors are large. An optimal data assimilation procedure for multimoment MP schemes should be developed to improve initial condition estimates and produce a more skilled hail size forecast.

A limited number of studies have analyzed hail forecasts for storm modes other than supercell thunderstorms. Many nonrotating storms produce large hail and cause extensive damage, one such hail event caused more than $1.4 billion in hail damage in Colorado (Fritz 2017). Studies such as Luo et al. (2017) have evaluated the skill of next-day CAM forecasts for a severe-hail-producing pulse thunderstorm, but to our knowledge no studies have evaluated hail forecasts for nonrotating storms using a frequently updated, high-resolution ensemble forecast framework. In the future we plan to expand the number of case studies to include multiple hail events.

Acknowledgments. This work 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 Stampede2 supercomputer at the University of Texas Advanced Computing Center (TACC). Supplemental computations were carried out at the Oklahoma Supercomputing Center for Education and Research (OSCER). Ensemble background and initial conditions were generated under the support of NOAA CSTAR Grant NA10NWS4680001. The authors thank Kevin Thomas for support in obtaining data. Ted Mansell and Daniel T. Dawson provided important insight into microphysics schemes. The authors also thank Bryan Putnam, Marcus Johnson, Amy McGovern, David Gagne, and

Amanda Burke, who offered code for and provided useful feedback on the project. Comments from three anonymous reviewers improved the quality of this manuscript.

REFERENCES

Adams-Selin, R. D., and C. L. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Wea. Rev.*, **144**, 4919–4939, https://doi.org/10.1175/MWR-D-16-0027.1.

Allen, J. T., and M. K. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteor.*, **10** (3), http://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/149.

Ben Bouallègue, Z., and S. E. Theis, 2014: Spatial techniques applied to precipitation ensemble forecasts: From verification results to probabilistic products. *Meteor. Appl.*, **21**, 922–929, https://doi.org/10.1002/met.1435.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Bringi, V. N., and V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar: Principles and Applications*. Cambridge University Press, 636 pp.

Buizza, R., A. Hollingsworth, F. Lalaurette, and A. Ghelli, 1999: Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. *Wea. Forecasting*, **14**, 168–189, https://doi.org/10.1175/1520-0434(1999)014<0168:PPOPUT>2.0.CO;2.

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

Carlin, J. T., J. Gao, J. C. Snyder, and A. V. Ryzhkov, 2017: Assimilation of Z_{DR} columns for improving the spinup and forecast of convective storms in storm-scale models: Proof-of-concept experiments. *Mon. Wea. Rev.*, **145**, 5033–5057, https://doi.org/10.1175/MWR-D-17-0103.1.

Changnon, S. A., 2009: Increasing major hail losses in the U.S. *Climatic Change*, **96**, 161–166, https://doi.org/10.1007/s10584-009-9597-z.

——, D. Changnon, and S. D. Hilberg, 2009: Hailstorms across the nation: An atlas about hail and its damages. ISWS Contract Rep. CR 2009-12, Illinois State Water Survey, Champaign, IL, 92 pp., https://www.ideals.illinois.edu/handle/2142/15156.

Chong, S.-L., and C. S. Chen, 1974: Water shells on ice pellets and hailstones. *J. Atmos. Sci.*, **31**, 1384–1391, https://doi.org/10.1175/1520-0469(1974)031<1384:WSOIPA>2.0.CO;2.

Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail climatology of the contiguous United States. *Wea. Forecasting*, **27**, 1235–1248, https://doi.org/10.1175/WAF-D-11-00151.1.

Crum, T. D., R. L. Alberty, and D. W. Burgess, 1993: Recording, archiving, and using WSR-88D data. *Bull. Amer. Meteor. Soc.*, **74**, 645–653, https://doi.org/10.1175/1520-0477(1993)074<0645:RAAUWD>2.0.CO;2.

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://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, https://doi.org/10.1175/MWR-D-11-00008.1.

——, E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2014: Low-level Z_{DR} signatures in supercell forward flanks: The role of size sorting and melting of hail. *J. Atmos. Sci.*, **71**, 276–299, https://doi.org/10.1175/JAS-D-13-0118.1.

Deardorff, J. W., 1980: Stratocumulus-capped mixed layers derived from a three-dimensional model. *Bound.-Layer Meteor.*, **18**, 495–527, https://doi.org/10.1007/BF00119502.

Dennis, E. J., and M. R. Kumjian, 2017: The impact of vertical wind shear on hail growth in simulated supercells. *J. Atmos. Sci.*, **74**, 641–663, https://doi.org/10.1175/JAS-D-16-0066.1.

Doswell, C. A., H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595, https://doi.org/10.1175/WAF866.1.

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

Edwards, R., and R. L. Thompson, 1998: Nationwide comparisons of hail size with WSR-88D vertically integrated liquid water and derived thermodynamic sounding data. *Wea. Forecasting*, **13**, 277–285, https://doi.org/10.1175/1520-0434(1998)013<0277:NCOHSW>2.0.CO;2.

Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfusz, 2014: MPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342, https://doi.org/10.1175/BAMS-D-13-00014.1.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249–280, https://doi.org/10.1175/1520-0469(1994)051<0249:ADMMPF>2.0.CO;2.

Foote, G. B., 1984: A study of hail growth utilizing observed storm conditions. *J. Climate Appl. Meteor.*, **23**, 84–101, https://doi.org/10.1175/1520-0450(1984)023<0084:ASOHGU>2.0.CO;2.

Fritz, A., 2017: Remember that crazy hailstorm in Denver? It's going to be Colorado's costliest catastrophe. *Washington Post*, 25 May 2017, https://www.washingtonpost.com/news/capital-weather-gang/wp/2017/05/25/remember-that-crazy-hail-storm-in-denver-its-going-to-be-colorados-costliest-catastrophe/?utm_term=.1d3f823248f4.

Gagne, D. J., II, A. McGovern, J. Brotzge, M. Coniglio, J. Correia Jr., and M. Xue, 2015: Day-ahead hail prediction integrating machine learning with storm-scale numerical weather models. *AAAI'15 Proc. 29th AAAI Con. on Artificial Intelligence*, Austin, TX, AAAI, 3954–3960.

——, ——, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting

with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, https://doi.org/10.1175/WAF-D-17-0010.1.

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

Heymsfield, A. J., 1983: Case study of a hailstorm in Colorado. Part IV: Graupel and hail growth mechanisms deduced through particle trajectory calculations. *J. Atmos. Sci.*, **40**, 1482–1509, https://doi.org/10.1175/1520-0469(1983)040<1482:CSOAHI>2.0.CO;2.

Jensen, A. A., J. Y. Harrington, H. Morrison, and J. A. Milbrandt, 2017: Predicting ice shape evolution in a bulk microphysics model. *J. Atmos. Sci.*, **74**, 2081–2104, https://doi.org/10.1175/JAS-D-16-0350.1.

——, ——, and ——, 2018: Microphysical characteristics of squall-line stratiform precipitation and transition zones simulated using an ice particle property-evolving model. *Mon. Wea. Rev.*, **146**, 723–743, https://doi.org/10.1175/MWR-D-17-0215.1.

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, https://doi.org/10.1175/MWR-D-15-0233.1.

Jung, Y., M. Xue, G. Zhang, and J. M. Straka, 2008a: 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, https://doi.org/10.1175/2007MWR2083.1.

——, ——, ——, and ——, 2008b: Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part II: Impact of polarimetric data on storm analysis. *Mon. Wea. Rev.*, **136**, 2246–2260, https://doi.org/10.1175/2007MWR2288.1.

——, ——, 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, https://doi.org/10.1175/2009JAMC2178.1.

——, ——, and M. Tong, 2012: 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, https://doi.org/10.1175/MWR-D-11-00032.1.

——, ——, F. Kong, Y. Wang, K. W. Thomas, and F. Shen, 2014: CAPS real-time storm-scale EnKF data assimilation and forecast system for the hazardous weather testbed spring experiment. *Sixth EnKF Workshop*, Center for Analysis and Prediction of Storms, University of Oklahoma, http://hfip.psu.edu/fuz4/EnKF2014/EnKF-day2/EnKF_2014_yjung.pdf.

Klemp, J. B., and R. B. Wilhelmson, 1978: The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci.*, **35**, 1070–1096, https://doi.org/10.1175/1520-0469(1978)035<1070:TSOTDC>2.0.CO;2.

Kong, F., 2013: 2013 CAPS Spring Forecast Experiment Program Plan. NOAA/NSSL, Storm Prediction Center, Norman, OK, 24 pp., http://forecast.caps.ou.edu/SpringProgram2013_Plan-CAPS.pdf.

Kumjian, M. R., and A. V. Ryzhkov, 2008: Polarimetric signatures in supercell thunderstorms. *J. Appl. Meteor. Climatol.*, **47**, 1940–1961, https://doi.org/10.1175/2007JAMC1874.1.

——, and ——, 2012: The impact of size sorting on the polarimetric radar variables. *J. Atmos. Sci.*, **69**, 2042–2060, https://doi.org/10.1175/JAS-D-11-0125.1.

Labriola, J., N. A. Snook, Y. Jung, B. J. Putnam, and M. Xue, 2017: Ensemble hail prediction for the storms of 10 May 2010 in south-central Oklahoma using single- and double-moment microphysical schemes. *Mon. Wea. Rev.*, **145**, 4911–4936, https://doi.org/10.1175/MWR-D-17-0039.1.

Lesins, G. B., and R. List, 1986: Sponginess and drop shedding of gyrating hailstones in a pressure-controlled icing wind tunnel. *J. Atmos. Sci.*, **43**, 2813–2825, https://doi.org/10.1175/1520-0469(1986)043<2813:SADSOG>2.0.CO;2.

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, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2.

Luo, L., M. Xue, K. Zhu, and B. Zhou, 2017: Explicit prediction of hail using multimoment microphysics schemes for a hailstorm of 19 March 2014 in eastern China. *J. Geophys. Res. Atmos.*, **122**, 7560–7581, https://doi.org/10.1002/2017JD026747.

——, ——, ——, and ——, 2018: Explicit prediction of hail in a long-lasting multi-cellular convective system in eastern China using multimoment microphysics schemes. *J. Atmos. Sci.*, **75**, 3115–3137, https://doi.org/10.1175/JAS-D-17-0302.1.

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, https://doi.org/10.1175/2009JAS2965.1.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.

McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1.

Milbrandt, J. A., and M. K. Yau, 2005a: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064, https://doi.org/10.1175/JAS3534.1.

——, and ——, 2005b: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62**, 3065–3081, https://doi.org/10.1175/JAS3535.1.

——, and ——, 2006: A multimoment bulk microphysics parameterization. Part III: Control simulation. *J. Atmos. Sci.*, **63**, 3114–3136, https://doi.org/10.1175/JAS3816.1.

——, and H. Morrison, 2016: Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. Part III: Introduction of multiple free categories. *J. Atmos. Sci.*, **73**, 975–995, https://doi.org/10.1175/JAS-D-15-0204.1.

Moeng, C.-H., 1984: A large-eddy-simulation model for the study of planetary boundary-layer turbulence. *J. Atmos. Sci.*, **41**, 2052–2062, https://doi.org/10.1175/1520-0469(1984)041<2052:ALESMF>2.0.CO;2.

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, https://doi.org/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, https://doi.org/10.1175/JAS-D-14-0066.1.

Nelson, S. P., 1983: The influence of storm flow structure on hail growth. *J. Atmos. Sci.*, **40**, 1965–1983, https://doi.org/10.1175/1520-0469(1983)040<1965:TIOSFS>2.0.CO;2.

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, https://doi.org/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://doi.org/10.1175/JAMC-D-15-0203.1.

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, https://doi.org/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, https://doi.org/10.1175/MWR-D-13-00042.1.

——, ——, ——, N. A . Snook, and G. Zhang, 2017a: Ensemble probabilistic prediction of a mesoscale convective system and associated polarimetric radar variables using single-moment and double-moment microphysics schemes and EnKF radar data assimilation. *Mon. Wea. Rev.*, **145**, 2257–2279, https://doi.org/10.1175/MWR-D-16-0162.1.

——, ——, ——, G. Zhang, and F. Kong, 2017b: Simulation of polarimetric radar variables from 2013 CAPS spring experiment storm-scale ensemble forecasts and evaluation of microphysics schemes. *Mon. Wea. Rev.*, **145**, 49–73, https://doi.org/10.1175/MWR-D-15-0415.1.

Rasmussen, R. M., and A. J. Heymsfield, 1987: Melting and shedding of graupel and hail. Part I: Model physics. *J. Atmos. Sci.*, **44**, 2754–2763, https://doi.org/10.1175/1520-0469(1987)044<2754:MASOGA>2.0.CO;2.

——, V. Levizzani, and H. R. Pruppacher, 1984: A wind tunnel and theoretical study of the melting behavior of atmospheric ice particles. III: Experiment and theory for spherical ice particles of radius >500 um. *J. Atmos. Sci.*, **41**, 381–388, https://doi.org/10.1175/1520-0469(1984)041<0381:AWTATS>2.0.CO;2.

Sammler, W., 1993: An updated climatology of large hail based on 1970-1990 data. Preprints, *17th Conf. on Severe Local Storms*, St. Louis, MO, Amer. Meteor. Soc., 32–35.

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

——, and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, https://doi.org/10.1175/2009WAF2222267.1.

——, G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, https://doi.org/10.1175/WAF-D-15-0103.1.

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

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

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, https://doi.org/10.1175/MWR-D-10-05053.1.

——, ——, and ——, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140**, 2126–2146, https://doi.org/10.1175/MWR-D-11-00117.1.

——, ——, and ——, 2013: Impacts of assumed observation errors in EnKF analyses and ensemble forecasts of a tornadic mesoscale convective system. *17th Conf. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Austin, TX, Amer. Meteor. Soc., 13.2, https://ams.confex.com/ams/93Annual/webprogram/Paper219933.html.

——, ——, and ——, 2015: Multiscale EnKF assimilation of radar and conventional observations and ensemble forecasting for a tornadic mesoscale convective system. *Mon. Wea. Rev.*, **143**, 1035–1057, https://doi.org/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, https://doi.org/10.1175/WAF-D-15-0152.1.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

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

Thompson, G., S. Tessendorf, A. Heymsfield, and I. Kyoko, 2018: Improving graupel and hail treatment in the thompson microphysics parameterization. *North American Hail Workshop*, Boulder, CO, NCAR.

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, https://doi.org/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, https://doi.org/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**, 1630–1648, https://doi.org/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, https://doi.org/10.1175/1520-0450(1983)022<1764:NVITAF>2.0.CO;2.

Vivekanandan, J., W. M. Adams, and V. N. Bringi, 1991: Rigorous approach to polarimetric radar modeling of hydrometeor orientation distributions. *J. Appl. Meteor.*, **30**, 1053–1063, https://doi.org/10.1175/1520-0450(1991)030<1053:RATPRM>2.0.CO;2.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman filter analyses and forecasts of a severe mesoscale convective system using different choices of microphysics schemes. *Mon. Wea. Rev.*, **142**, 3243–3263, https://doi.org/10.1175/MWR-D-13-00260.1.

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

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

Wienhoff, Z. B., H. B. Bluestein, L. J. Wicker, J. C. Snyder, A. Shapiro, C. K. Potvin, J. B. Houser, and D. W. Reif, 2018: Applications of a spatially variable advection correction technique for temporal correction of dual-Doppler analyses of tornadic supercells. *Mon. Wea. Rev.*, **146**, 2949–2971, https://doi.org/10.1175/MWR-D-17-0360.1.

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. W. Mitchell, and K. W. Thomas, 1998a: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.

——, ——, ——, E. D. W. Mitchell, J. T. Johnson, and K. W. Thomas, 1998b: Evaluating the performance of WSR-88D severe storm detection algorithms. *Wea. Forecasting*, **13**, 513–518, https://doi.org/10.1175/1520-0434(1998)013<0513:ETPOWS>2.0.CO;2.

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, https://doi.org/10.1007/s007030070003.

——, and Coauthors, 2001: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143–165, https://doi.org/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, https://doi.org/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: Experiments with simulated radar data. *Quart. J. Roy. Meteor. Soc.*, **136**, 685–700, https://doi.org/10.1002/qj.593.

Yussouf, N., 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, https://doi.org/10.1175/MWR-D-12-00237.1.

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