

# Prediction and Ensemble Forecast Verification of Hail in the Supercell Storms of 20 May 2013

NATHAN SNOOK AND YOUNGSUN JUNG

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

JERALD BROTZGE

*New York State Mesonet, and University at Albany, State University of New York, Albany, New York*

BRYAN PUTNAM AND MING XUE

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

(Manuscript received 2 November 2015, in final form 18 March 2016)

## ABSTRACT

Despite recent advances in storm-scale ensemble NWP, short-term (0–90 min) explicit forecasts of severe hail remain a major challenge as a result of the fast evolution and short time scales of hail-producing convective storms and the substantial uncertainty associated with the microphysical representation of hail. In this study, 0–90-min ensemble hail forecasts for the supercell storms of 20 May 2013 over central Oklahoma are examined and verified, with the goals of 1) evaluating ensemble forecast performance, 2) comparing the advantages and limitations of different forecast fields potentially suitable for the prediction of hail and severe hail in a Warn-on-Forecast setting, and 3) evaluating the use of dual-polarization radar observations for hail forecast validation. To address the challenges of hail prediction and to produce skillful forecasts, the ensemble uses a two-moment microphysics scheme that explicitly predicts a hail-like rimed-ice category and is run with a grid spacing of 500 m. Radar reflectivity factor and radial velocity, along with surface observations, are assimilated every 5 min for 1 h as the storms were developing to maturity, followed by a 90-min ensemble forecast. Several methods of hail prediction and hail forecast verification are then examined, including the prediction of the maximum hail size compared to Storm Prediction Center (SPC) and Meteorological Phenomena Identification Near the Ground (mPING) hail observations, and verification of model data against single- and dual-polarization radar-derived fields including hydrometeor classification algorithm (HCA) output and the maximum estimated size of hail (MESH). The 0–90-min ensemble hail predictions are found to be marginally to moderately skillful depending on the verification method used.

## 1. Introduction

Severe hail is a major weather hazard that can cause injury and property damage. When severe hail strikes an urban area, the potential for injury is increased, and damage from a single event can exceed \$1 billion. For example, the “Mayfest” hailstorm that struck Fort Worth, Texas, on 5 May 1995 resulted in 109 hail-related injuries and caused over \$2 billion in damage (Edwards and Thompson 1998). It is estimated that hail causes over \$1 billion in

crop damage and \$1 billion in property damage annually (Jewell and Brimelow 2009). Despite these impacts, direct prediction of hail using NWP ensembles has received relatively little attention compared to the prediction of other thunderstorm hazards (e.g., tornadoes, severe winds, and heavy rainfall).

Forecasting severe hail is a challenging problem for numerical weather prediction (NWP). Explicit prediction of hail requires the NWP model to predict the development, motion, and intensity of convective storms accurately; furthermore, hail in the model is closely tied to microphysical parameterization, which is a major source of model uncertainty and error at the convective scale. The predictability of hail is also limited by the rapid development and evolution of the convective storms that

---

*Corresponding author address:* Nathan Snook, Center for Analysis and Prediction of Storms, University of Oklahoma, Rm. 5425, 120 David Boren Blvd., Norman, OK 73072.  
E-mail: nsnook@ou.edu

produce it; forecast errors grow quickly at the convective scale, necessitating an accurate estimation of the atmospheric state and information on the level of uncertainty in this estimate. Ensemble forecasting is needed to provide forecast uncertainty information.

Ensemble Kalman filter (EnKF; Evensen 1994, 2003) data assimilation (DA) techniques have been widely applied in recent years to produce skillful NWP analyses and forecasts ranging from global to convective scales (Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Anderson 2001; Whitaker and Hamill 2002; Snyder and Zhang 2003; Dowell et al. 2004; Zhang et al. 2004; Tong and Xue 2005; Dirren et al. 2007; Tong and Xue 2008a; Snook et al. 2011; Dawson et al. 2012). EnKF methods employ an ensemble of forecasts to sample the error covariances of the background state; the resulting flow-dependent covariances and cross covariances allow state variables not directly observed to be retrieved during DA (e.g., Tong and Xue 2008a)—a feature that is particularly important at the storm scale, where radar data are vital, but observe only a very limited number of variables. Retrieval of wind, temperature, and microphysical fields can be achieved through EnKF DA of radar data (e.g., Dowell et al. 2004; Tong and Xue 2005; Tong 2006; Snook et al. 2011), and the quality of the resulting analyses can be further improved when conventional observations are also assimilated (Snook et al. 2015). Maximizing the accuracy of the initial state in this way is vital for subkilometer-scale NWP applications, such as hail prediction. EnKF-based ensembles have been shown to produce more skillful probabilistic forecasts at the global scale in operational and quasi-operational settings (e.g., Houtekamer et al. 2005; Hamill and Whitaker 2011). Quasi-operational forecast ensembles incorporating cycled EnKF DA are also being applied experimentally for real-time storm-scale NWP as part of the annual Center for Analysis and Prediction of Storms (CAPS) spring experiment (e.g., Xue et al. 2008; CAPS 2015) toward the goals of the “Warn on Forecast” vision (Stensrud et al. 2009).

Forecast verification is another challenging aspect of hail prediction, in large part because of the lack of high quality, high-resolution hail observation datasets. Most currently available in situ hail observation datasets, including storm reports collected and archived by the Storm Prediction Center (SPC) and crowdsourced reports submitted and archived via the National Severe Storms Laboratory (NSSL) Meteorological Phenomena Identification Near the Ground (mPING) smartphone application (Elmore et al. 2014), rely on reports from members of the public and, thus, may suffer from substantial biases. These biases include substantial underreporting of hail in sparsely populated regions (Cintineo

et al. 2012), as well as a bias in reported hail size toward the sizes of common circular or spherical objects (e.g., dime-, quarter-, and baseball-sized hail), because such objects are often used as references in hail reports submitted by the public (Schaefer et al. 2004).

A cost-efficient method of addressing the limitations of observational hail data is the use of radar observations, particularly from dual-polarization radars, as a supplementary source of hail information. Radar data are already used operationally to estimate hail size and coverage via the maximum estimated size of hail (MESH) algorithm (Witt et al. 1998). The MESH algorithm uses a weighted integration of the reflectivity factor exceeding 40 dBZ above the melting layer to estimate the maximum size of hail occurring at the surface. While the application of MESH can provide high-resolution gridded estimates of hail coverage, biases and errors are possible as a result of the fact that MESH relies entirely upon observations in the upper levels of the atmosphere to estimate hail size at the surface. Less indirect observations of near-surface hail can be obtained using dual-polarization radar observations.

Between 2011 and 2013, the WSR-88D S-band radar network (Crum et al. 1993) used operationally by the NWS was upgraded with the capability to collect dual-polarization radar observations. Dual-polarization radars observe the atmosphere using both horizontally and vertically polarized radar beams. Observed fields obtained from these data, such as differential reflectivity  $Z_{dr}$ , specific differential phase  $K_{dp}$ , and copolar correlation coefficient  $\rho_{hv}$ , allow us to infer storm properties such as the types of hydrometeors present, their relative sizes, and the presence of nonmeteorological scatterers such as insects or debris through the identification of polarimetric signatures (Kumjian and Ryzhkov 2008). Using dual-polarization observations, regions of hail can be identified either by manually examining these polarimetric signatures, or by applying an automated hydrometeor classification algorithm (HCA). In this way, dual-polarization radar observations can be used as a high quality supplementary data source to provide estimates of the geographic distribution of hail. HCAs (e.g., Vivekanandan et al. 1999; Heinselman and Ryzhkov 2006; Park et al. 2009; Lim et al. 2013; Bechini and Chandrasekar 2015) have been successfully applied in research and operational settings for the detection of hail, and have provided a basis for hail size discrimination using polarimetric radar observations (e.g., Ryzhkov et al. 2013).

In this study, we explore the ability of a storm-scale ensemble using subkilometer horizontal grid spacing to produce explicit 0–90-min ensemble hail forecasts directed toward model-based Warn-on-Forecast prediction

of hail and severe hail. The goals of this study include 1) evaluating the performance of the ensemble in terms of hail forecast probability; 2) comparing the relative advantages, limitations, and biases of different hail-related model forecast fields suitable for the prediction of hail and severe hail in a Warn-on-Forecast setting; and 3) evaluating the use of dual-polarization radar observations for hail forecast validation. To accomplish these goals, the Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001) atmospheric model is used together with its ensemble Kalman filter DA system (Tong and Xue 2005; Xue et al. 2006; Tong and Xue 2008b) to assimilate WSR-88D radar reflectivity factor and radial velocity data and surface observations for the supercell storms that occurred over central Oklahoma during the afternoon of 20 May 2013. The ensemble analyses produced are used to initialize ensemble forecasts of the supercell storms. These forecasts are verified against WSR-88D observations, and their skill in predicting hail is assessed using several different methods.

Probabilistic forecasts of maximum hail size are produced explicitly from the number concentration and mass mixing ratio predicted by the model microphysics, as well as by applying the MESH algorithm used operationally by the National Weather Service (Witt et al. 1998) to simulated radar observations from the ensemble forecasts; these forecasts are verified against surface hail observations and MESH calculated from WSR-88D observations. The hail mass field within the ensemble is also verified against observed dual-polarization data via the application of a fuzzy-logic-based HCA (Park et al. 2009) to identify hail swaths in the observed data. Where possible, forecast skill is objectively assessed using metrics such as equitable threat score (ETS) (Schaefer 1990), the area under the relative operating characteristic curve (AUC), and reliability diagrams.

The remainder of this paper is organized as follows. In section 2, we provide a summary of the 20 May 2013 supercell case, introduce the ensemble DA and forecast systems used, and describe the experiments performed and the verification techniques used. Section 3 discusses the results of the ensemble forecasts, as well as the results and limitations of the hail forecast verification techniques examined. In section 4, overall hail prediction results are summarized, and the relative merit of the hail prediction methods and implications for future hail prediction and Warn-on-Forecast efforts are discussed.

## 2. Data and methods

On the afternoon of 20 May 2013, multiple supercell storms formed and rapidly intensified along a dryline over central and southwestern Oklahoma between 1830

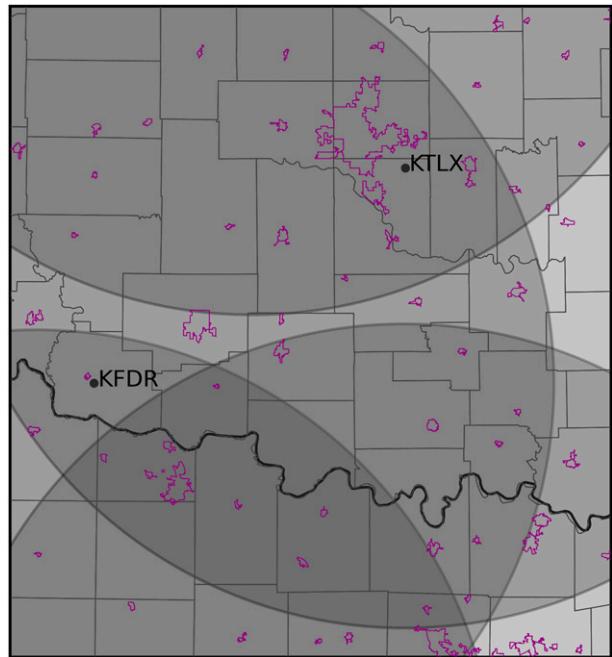


FIG. 1. Geographic extent of the 500-m experiment domain. Assimilated radar sites are marked, along with 230-km range rings shaded in gray (KVNK, KDYX, and KFWS are located outside the domain boundary, but still observe portions of the model domain). Darker shades of gray indicate regions of overlapping radar coverage. Urban boundaries are shown in purple.

and 1930 UTC. These storms moved to the northeast, passing through central Oklahoma during the subsequent hours, producing numerous reports of severe hail and several tornadoes, including the [enhanced Fujita (EF) scale] EF5 Newcastle–Moore tornado (Burgess et al. 2014). For more a detailed discussion of the storms of 20 May 2013, we refer the reader to Zhang et al. (2015). Ensemble analyses and forecasts are produced using the ARPS and its EnKF DA system (Xue et al. 2006; Tong and Xue 2008a).

The ensemble forecasts are performed using a horizontal grid spacing of 500 m on a grid consisting of  $603 \times 653 \times 63$  grid points that covers much of Oklahoma and portions of far northern Texas (Fig. 1). The model grid is stretched in the vertical, with a minimum vertical spacing of 50 m at the surface and an average vertical spacing of 425 m. Our EnKF DA and ensemble forecasting uses 40 members and is initialized at 1800 UTC on 20 May 2013 from the 40-member CAPS real-time storm-scale ensemble forecasts (SSEFs; Kong et al. 2014) based on version 3.4.1 of the ARW (Skamarock et al. 2008) model at a horizontal grid spacing of 4 km. The SSEFs also provide lateral boundary conditions for the ensemble on our 500-m grid. The ARPS model settings largely follow Snook et al. (2015): radiation is parameterized via

the NASA Goddard Space Flight Center long- and shortwave radiation parameterization, a two-layer soil model with surface fluxes parameterized using predicted surface temperature and water content is applied, subgrid turbulence is parameterized using a 1.5-order turbulent kinetic energy (TKE) based scheme, and terrain data are interpolated from a USGS dataset with a resolution of 30 s of arc. References on these schemes can be found in Xue et al. (2001) and Snook et al. (2015). Microphysical processes are parameterized using the two-moment microphysics scheme of Milbrandt and Yau (2005), which we will henceforth refer to as MY2.

From the interpolated 1800 UTC SSEF forecasts, 30-min spinup ensemble forecasts are first performed. Storm-scale perturbations are added to the ensemble at 1800 UTC in order to enhance small-scale perturbations absent in the parent (4-km grid spacing) ensemble and improve the ensemble spread in the 500-m forecast experiment; they are smoothed Gaussian perturbations generated using a recursive filter with horizontal and vertical decorrelation scales of 6 km. The smoothed perturbations are applied to the horizontal wind ( $u$ ,  $v$ ) and potential temperature  $\theta$  fields with mean standard deviations of  $0.5 \text{ m s}^{-1}$  and  $0.5 \text{ K}$ , respectively. From 1830 to 1930 UTC, available radar and surface observations are assimilated at 5-min intervals using the ARPS EnKF system (Xue et al. 2006; Tong and Xue 2008b), which is based on the ensemble square root filter (EnSRF) algorithm of Whitaker and Hamill (2002). At the end of the assimilation period, 90-min ensemble forecasts are performed from 1930 to 2100 UTC.

During EnKF DA, observation errors for radar data are assumed to be  $4.0 \text{ m s}^{-1}$  for the radial velocity and  $6.0 \text{ dBZ}$  for the radar reflectivity factor (Snook et al. 2013); observation errors for surface observations are assumed to be  $1.5 \text{ m s}^{-1}$  for  $u$  and  $v$ ,  $1.5 \text{ K}$  for air temperature,  $2.0 \text{ K}$  for dewpoint, and  $2.0 \text{ hPa}$  for air pressure. The radar observation operators used follow Jung et al. (2008), and a Gaussian power-gain function (Wood and Brown 1997) is used in the vertical to provide beam pattern weighting in the operators (Xue et al. 2006). Covariance localization with a radius of 3 km in the vertical and horizontal is applied for radar data; for surface observations, localization radii of 6 km in the vertical and 300 km in the horizontal are used. To maintain ensemble spread during DA, relaxation-to-prior-spread covariance inflation (Whitaker and Hamill 2012) is applied, with an inflation coefficient of 0.95. These localization and covariance inflation settings were set based on a set of earlier sensitivity experiments and drew upon experience from our earlier studies. Radar data from five WSR-88D radar sites are assimilated: Oklahoma City, Oklahoma (KTLX); Vance Air Force Base,

Oklahoma (KVNK); Frederick, Oklahoma (KFDR); Fort Worth, Texas (KFWS); and Dyess Air Force Base, Texas (KDYX). Together, these WSR-88D sites give radar coverage over the full domain, with the full line of storms observed by at least two radars during the assimilation period (Fig. 1). Surface observations from Oklahoma Mesonet sites within the domain are assimilated.

Hail within the model is closely linked to the microphysics parameterization; for this reason, the ability of the microphysics scheme to accurately capture processes related to hail formation and growth strongly impacts the quality of the resulting hail forecast. All experiments use the MY2 microphysical parameterization scheme, in which the hydrometeor particle size distribution (PSD) is represented by a gamma distribution function:

$$N(D) = N_0 D^\alpha e^{-\lambda D}, \quad (1)$$

where  $N(D)$  is the number of drops/particles of a given diameter,  $N_0$  is the intercept parameter,  $\alpha$  is the shape parameter, and  $\lambda$  is the slope parameter. In MY2, hydrometeor mixing ratios and number concentrations are predicted, allowing two unknowns,  $N_0$  and  $\lambda$ , to be retrieved, while  $\alpha$  is held constant at 0. The variation of hydrometeor PSDs permitted by this two-moment microphysical scheme allows a better representation of microphysical processes, including size sorting and melting, than is possible using a single-moment scheme. This improved representation is important because size sorting and melting are the primary contributors to predicted dual-polarization radar signatures within the model (Dawson et al. 2014). In contrast, single-moment microphysics schemes are generally inadequate for realistically representing dual-polarization signatures (Dawson et al. 2010; Jung et al. 2010; Jung et al. 2012; Putnam et al. 2014). The results, analyses, and forecasts are presented in the next section.

### 3. Results

In this section, the ensemble analyses and 0–90-min ensemble forecasts produced by the model are evaluated using several qualitative and quantitative verification methods, with a focus on model representation and prediction of hail. The ensemble analyses are qualitatively evaluated in terms of the predicted storm structure and radar reflectivity factor. Quantitative verification of predicted hail is performed using the methods summarized in the previous section. Maximum hail size predicted explicitly from the hail PSD within the model (using methods described below in section 3a), and via application of the MESH algorithm

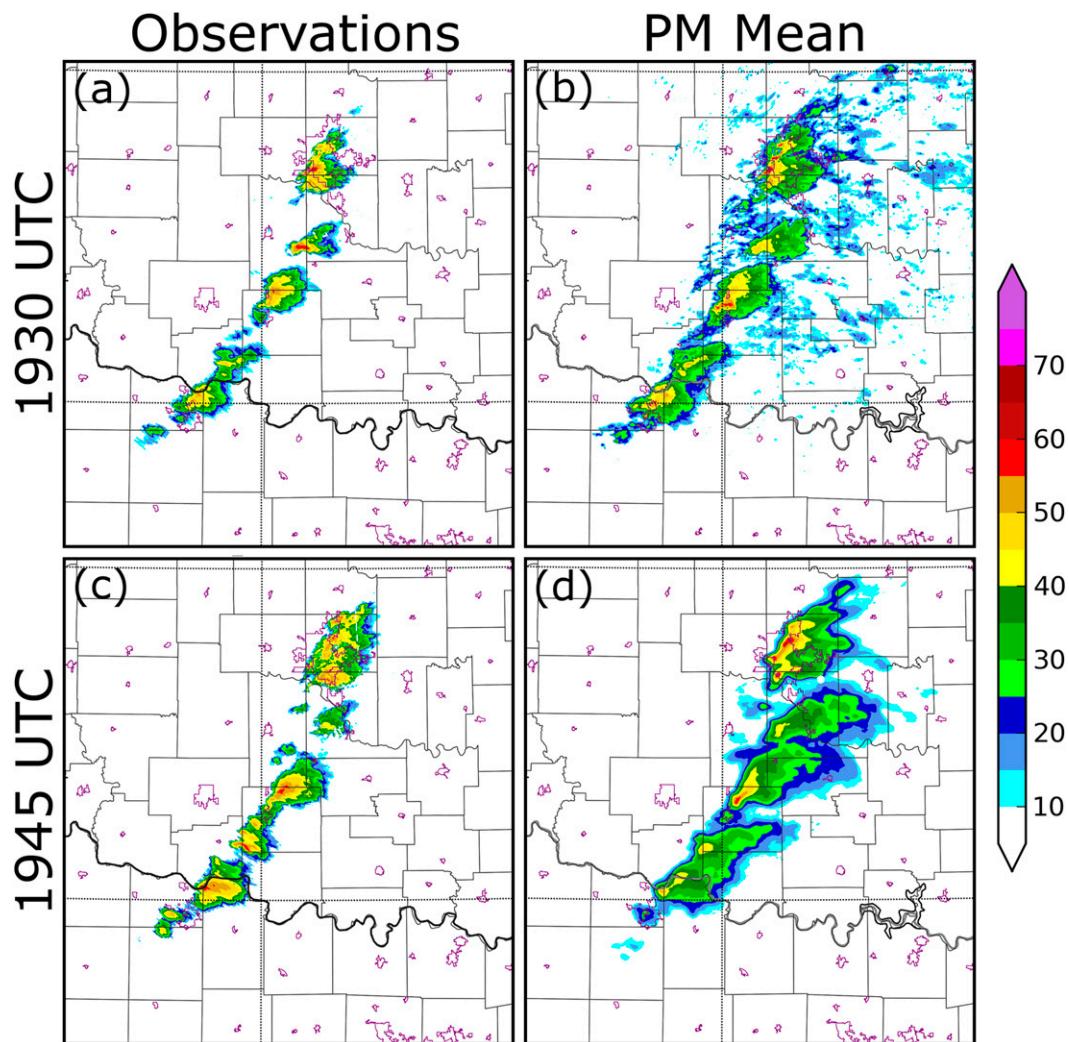


FIG. 2. Observed KTLX 0.5° reflectivity factor and ensemble PM mean reflectivity factor interpolated to the KTLX 0.5° tilt at (a),(b) 1930 and (c),(d) 1945 UTC.

to radar reflectivity factor and temperature fields predicted by the model, are verified against MESH observations calculated from WSR-88D radar observations. Dual-polarization radar observations are also used for verification: hail swaths predicted by the ensemble, in terms of the hail mass field, are verified qualitatively against dual-polarization radar observations via application of a fuzzy-logic-based HCA (Park et al. 2009).

The final EnKF analysis at the end of the DA period (1930 UTC) is compared to radar observations in terms of the radar reflectivity factor in Figs. 2a and 2b. By this time, the line of supercell thunderstorms was relatively mature, with several cells possessing robust mesocyclones and radar presentation typical of supercell storms. The EnKF analysis accurately captures the developing line of supercell storms; the probability-matched (PM; Ebert 2001) ensemble mean analysis

(Fig. 2b) closely matches the observed reflectivity (Fig. 2a). The skillful nature of the EnKF analysis makes it a suitable initial condition for the subsequent ensemble hail prediction forecasts. Some spurious light precipitation is present in the northeastern portion of the forecast domain at 1930 UTC (Fig. 2b), but this spurious precipitation quickly dissipates, and is almost entirely absent from the 15-min ensemble PM mean forecast at 1945 UTC (Fig. 2d), which still shows good agreement with observations (Fig. 2c) at that time.

a. Verification of forecasts of PSD-estimated maximum hail size

The forecast ensemble contains direct information about hail within the model from the output of the microphysical scheme. Using the gamma distribution assumed by the model [defined by (1)], we can directly

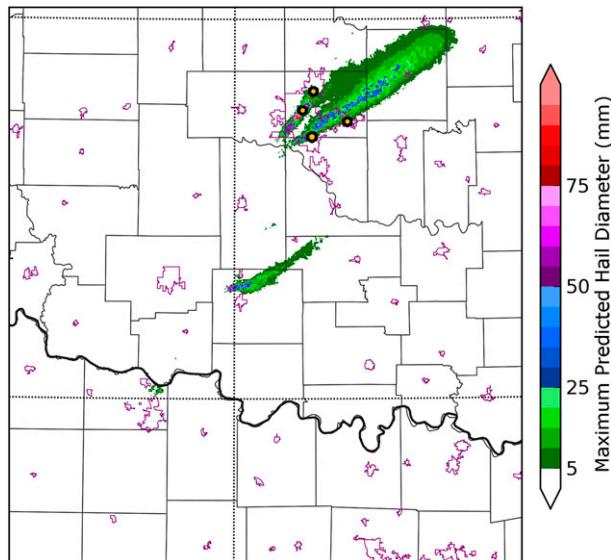


FIG. 3. Swath of maximum predicted hail diameter at the surface within the forecast ensemble between 1930 and 2100 UTC (0–90 min of forecast time), shown in terms of the PM mean of the forecast ensemble. Hail reports from SPC and mPING data received between 1930 and 2100 UTC are indicated by orange dots.

obtain the hail PSD at each model grid point given the hail mixing ratio  $q_h$  and the hail total number concentration  $n_h$ . To translate the model PSD into a meaningful estimate of maximum hail size, we must consider the nature of the hail reports—typically, the size of the largest hailstones is reported. The PSD defined by (1) does not have a corresponding maximum diameter value; instead, it extends to arbitrarily large values of hail diameter (with extremely low probabilities). We address this by defining the maximum observable hail diameter  $D_{\max}$ , estimated from the hail PSD, as the diameter value for which  $n_h(D_{\max}) = 10^{-4} \text{ m}^{-3}$ . This definition, which is similar to the definition proposed by Milbrandt and Yau (2006a), is chosen subjectively for delineating between PSD-defined hail concentrations that would be physically observable and those that would not.

A swath of maximum  $D_{\max}$  between 1930 and 2100 UTC, calculated using the PM ensemble mean at 5-min intervals, is presented in Fig. 3. SPC storm data and mPING hail reports received during the forecast period are indicated in Fig. 3 by orange dots. Though radar observations indicated several substantial hail swaths (which will be discussed in greater detail below), SPC and mPING hail reports were rather sparse. Only four reports of hail were received between 1930 and 2100 UTC, all of which were associated with the northernmost supercell, and all within the Oklahoma City metropolitan area (OKC metro). The forecast ensemble predicted two hail

swaths in the vicinity of these reports (Fig. 3); the two most northwesterly reports fall directly along and within the track of a small but intense hail swath extending through the northwestern portion of the OKC metro area, while the other two reports fall along the southern edge of a larger swath of small to marginally severe hail crossing the southeastern portion of the OKC metro area and extending several tens of kilometers to the northeast. The forecast PM ensemble mean predicted another swath of hail to the south of the OKC metro area, though no SPC or mPING hail reports were received in the vicinity of this predicted hail swath.

From the ensemble, we can also derive the probability of  $D_{\max}$  exceeding a specified threshold. During severe weather operations, NWS forecasters issue severe thunderstorm warnings for storms likely to be producing hail 1 in. (25.4 mm) or greater in diameter at the surface. While there is no commonly accepted definition on the minimum diameter of hail, the smallest hail size commonly accepted by mPING or the NWS is 0.25 in. (6.35 mm), corresponding to pea-sized hail. With these guidelines in mind, we consider two thresholds: one to identify severe hail within the ensemble (25 mm) and one to identify areas where any hail is present (5 mm). The probability of  $D_{\max}$  exceeding (Fig. 4a) 5.0 mm and (Fig. 4b) 25.0 mm using a neighborhood ensemble probability (NEP; Schwartz et al. 2010) method are presented in Fig. 4. For NEP calculations, a neighborhood radius of 2.5 km is used. Severe reports exceeding the specified diameter are indicated by orange dots in Fig. 4.

The region of  $P(D_{\max} > 5 \text{ mm}) > 0.4$  (Fig. 4a) roughly corresponds to the region of  $D_{\max} > 5 \text{ mm}$  in the PM ensemble mean (Fig. 3). Areas of lower probability extend primarily to the north and east, reflecting variation within the ensemble in the speed and direction of motion of the storms and in the length of time that hail was produced. Similar as for the PM mean (Fig. 3), the observed SPC and mPING hail reports fall near or within areas of high predicted probability of hail (Fig. 4a), though reports are absent in the region of high probability in the rural areas northeast of the OKC metro area. Two of the received hail reports were severe (reported hail diameter  $> 1$  in.); these reports, which occurred in the southeastern part of the OKC metro area, are on the southern edge of a relatively low-probability region for severe hail probability [ $P(D_{\max} > 25 \text{ mm})$ ; Fig. 4b] within the ensemble. Overall, the probability of severe hail for this case was relatively modest, with the highest probabilities of around 0.3–0.6 occurring in two swaths over the OKC metro area. The proximity of the reported severe hail to these swaths, however, is an encouraging indication of model skill for explicit short-term severe hail prediction.

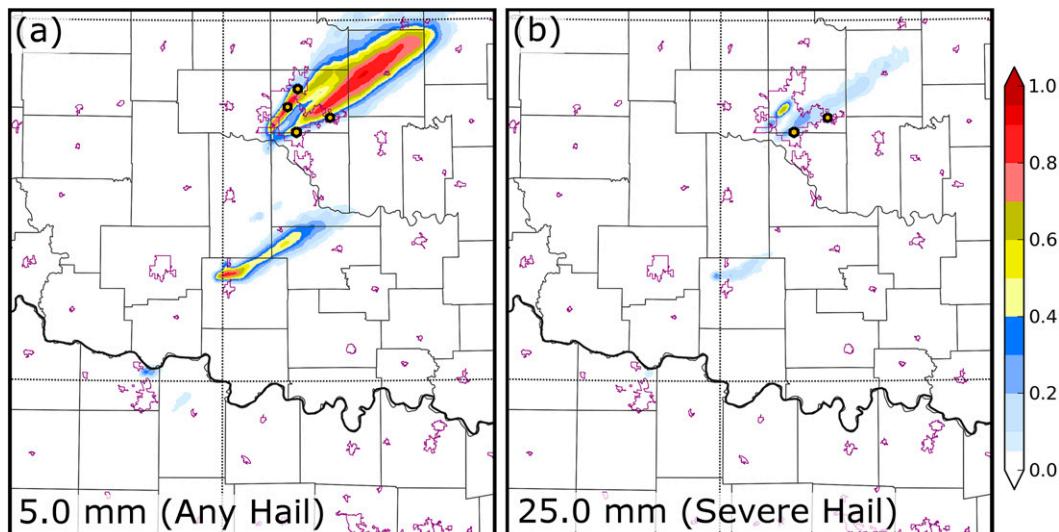


FIG. 4. Swaths of the neighborhood ensemble probability of  $D_{\max}$  exceeding (a) 5.0 and (b) 25.0 mm between 1930 and 2100 UTC (0–90 min of forecast time). Hail reports from SPC and mPING data received between 1930 and 2100 UTC that meet or exceed the corresponding size are indicated by dots.

#### b. Radar-based ensemble hail forecast verification using MESH

As noted in our discussion of methods, a major limitation of hail reports is a substantial population bias (e.g., Wyatt and Witt 1997; Davis and LaDue 2004). Most hail reports are received in high-population regions, with substantial underreporting in rural areas. For this reason, it is unclear whether the large areas of hail predicted within the model to the north and east of the OKC metro area (Figs. 3 and 4a) are false detections, or whether hail occurred in those areas but was not reported. WSR-88D radar observations, which offer full volumetric coverage of the storms that occurred on 20 May 2013, can be used as an indirect source of observations to address this ambiguity. In this section, we consider two methods of radar-based hail verification: prediction and verification of MESH and verification of ensemble-predicted regions of hail against areas of hail identified by dual-polarization WSR-88D observations using an implementation of the fuzzy-logic-based HCA of Park et al. (2009).

MESH (Witt et al. 1998) is calculated using a weighted integration of the radar reflectivity factor exceeding 40 dBZ above the melting level to estimate the maximum size of hail occurring at the surface. Temperature and simulated reflectivity factor, calculated using the radar observation operator of Jung et al. (2008), are readily available from our forecast ensemble, making it straightforward to directly apply the MESH algorithm to produce gridded forecast MESH fields. For verification,

we interpolate hourly MESH observations from the WSR-88D network to our model grid. A comparison between the WSR-88D observed hourly MESH swath from 1900 to 2100 UTC (Fig. 5a), the available verification period most closely corresponding to the 1930–2100 UTC forecast period, and the ensemble PM mean mesh swath during the forecast period (Fig. 5b) shows a relatively favorable comparison between the forecast ensemble and the observations. In the observed MESH field (Fig. 5a), three primary MESH swaths are present, corresponding to the Newcastle–Moore supercell over the OKC metro area and two other supercells to the south. These three primary swaths are also present in the PM mean forecast (Fig. 5b), in locations that match well with the observed swaths. The MESH forecasts show skill ( $ETS > 0$ ) at the 5- and 25-mm thresholds, with ETS values of 0.28 and 0.12 at these thresholds, respectively.

Compared to the observations, the forecast PM mean MESH swaths have narrower cores of high MESH values and higher maximum values for the northern two storms (Fig. 5). Because storms in many ensemble members moved to the northeast faster than the observed storms, the forecast swaths extend farther northeast than the observed swaths; this behavior is most prominent in the northernmost storm. In contrast, the ensemble underpredicted the extent of MESH in the southernmost storm, located near the Oklahoma–Texas border (Fig. 5). While hail was present at upper levels in the southern storm in the ensemble, as we will discuss below, this storm, in many ensemble members, exhibited a relatively low radar reflectivity factor ( $<45$  dBZ) within the updraft core in

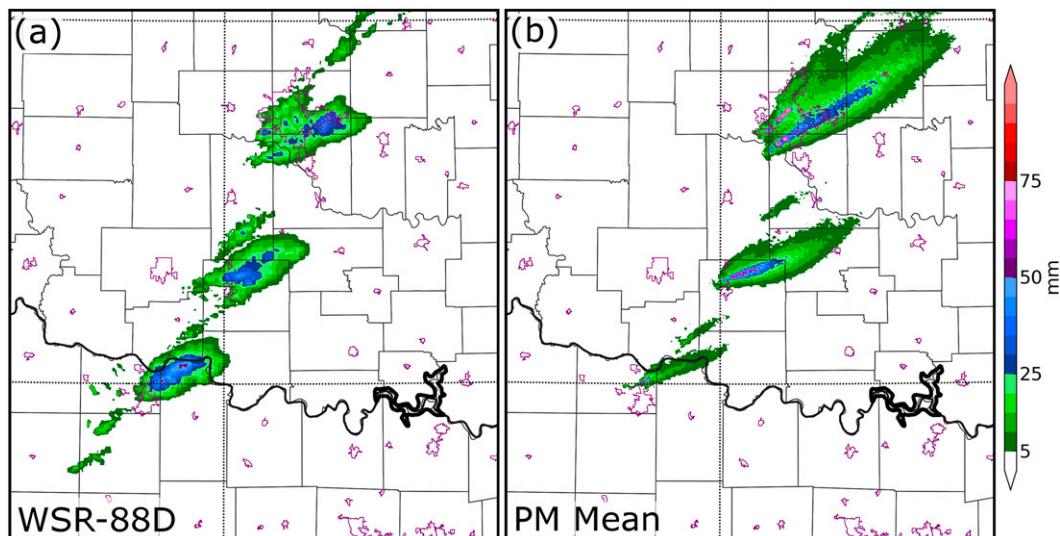


FIG. 5. Swaths of MESH shown as (a) hourly WSR-88D gridded MESH data between 1900 and 2100 UTC interpolated to the model grid, and (b) PM mean gridded MESH predicted by the forecast ensemble between 1930 and 2100 UTC (0–90 min of forecast time).

the upper portion of the storm (not shown). Because MESH is calculated using a weighted integration of the radar reflectivity factor exceeding 40 dBZ above the 0°C level, this led to low MESH values in the southern storm within the ensemble.

MESH, which uses the reflectivity factor aloft as a proxy for estimating hail size at the surface, is employed operationally to produce estimates of maximum hail size that may or may not agree with the microphysically predicted  $D_{\max}$  at the surface. When ensemble-predicted MESH (Fig. 5b) is compared to predicted  $D_{\max}$  at the surface (Fig. 3), however, the MESH estimate of maximum hail diameter at the surface exceeds the actual  $D_{\max}$  at the surface in most areas, and the hail swaths predicted by MESH cover a larger region for each of the hail-producing storms in this case. At least for this case, MESH, when applied to the ensemble, appears to have a high bias in predicted hail size, including prediction of nonsevere hail in areas near hail-producing storms where no actual hail was present in the microphysical fields at the surface (Figs. 3 and 5b). Unfortunately, because high-resolution observed hail size data are not available at the surface, it is not feasible to determine whether similar biases in MESH are also present in the observations, though prior studies have noted high biases in MESH calculated from WSR-88D observations (e.g., Ortega et al. 2009; Cintineo et al. 2012).

### c. Verification of the ensemble hail forecast using dual-polarization radar observations

Dual-polarization radar observations provide a method for obtaining dense, high quality estimates of the spatial

distribution of hail in the area observed by the radar; such observations are well suited for hail forecast verification. To estimate the hydrometeor type from observed dual-polarization radar data, we apply a variant of the HCA of Park et al. (2009), which uses fuzzy-logic-based classification to determine the dominant hydrometeor type from observed radar data (horizontally polarized radar reflectivity factor  $Z$ , differential reflectivity  $Z_{\text{dr}}$ , and copolar correlation coefficient  $\rho_{\text{hv}}$ ) and the model vertical temperature profile. The variant of the Park et al. (2009) HCA used in this study applies the membership functions and weights for  $Z$ ,  $Z_{\text{dr}}$ , and  $\rho_{\text{hv}}$ , as well as for standard deviations of  $Z$  and differential phase along radials. Confidence vectors are not used; despite their omission, however, HCA output using this configuration did not appear to suffer from any data quality issues. The radar data used in the HCA underwent preliminary quality control, including the rejection of radar observations, where  $\rho_{\text{hv}}$  was less than 0.85. As the model directly predicts hail, we compare observed HCA-generated swaths directly to ensemble-predicted swaths of hail defined using the hail mass field  $m_h$ ,

$$m_h = q_h \times \rho_{\text{air}}, \quad (2)$$

where  $q_h$  is the hail mixing ratio and  $\rho_{\text{air}}$  is the air density.

HCA-indicated hail swaths from the lowest-elevation scans of KTLX and KFDR (the two nearest WSR-88D radars to the storms in this case study) between 1930 and 2100 UTC are shown in Fig. 6, with the elevation of

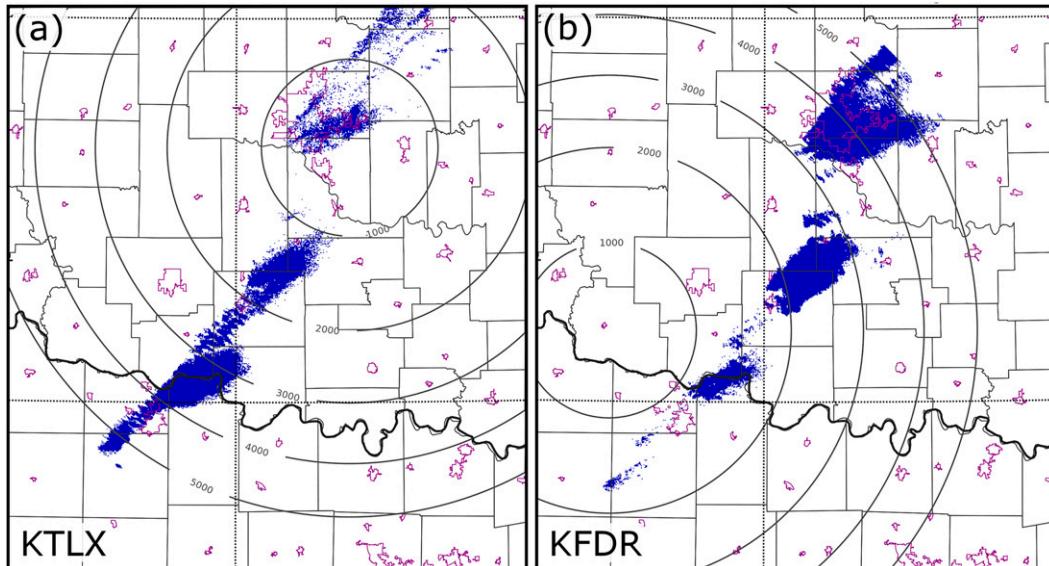


FIG. 6. HCA-indicated hail swaths (blue-shaded regions) from (a) KTLX and (b) KFDR observations at the  $0.5^\circ$ -elevation angle. Height of the radar beam above mean sea level (m) is indicated by the gray contours.

the beam above mean sea level indicated. These swaths are used to verify ensemble forecasts of  $P(m_h > 0.03 \text{ kg m}^{-3})$  in Fig. 7, with HCA-indicated hail swaths marked by the contour and the probabilistic ensemble forecast shaded. In Fig. 7, the model  $m_h$  field, calculated using (2), is interpolated to the same vertical level as the HCA-indicated hail swaths plotted in Fig. 6, resulting in  $m_h$  on the model grid in the horizontal, but located at the elevation of the radar beam in the vertical. The HCA-indicated hail swaths, smoothed for visual clarity, are also plotted in Fig. 7 for comparison and verification. The threshold of  $0.03 \text{ kg m}^{-3}$  for  $m_h$  used in Fig. 7 was chosen based on preliminary experiments using a wide range of  $m_h$  thresholds; lower  $m_h$  thresholds produced forecasts with a higher probability of detection, but resulted in substantial overprediction of the geographic extent of hail swaths, while higher thresholds improved forecast reliability but reduced the probability of detection (not shown). The threshold of  $0.03 \text{ kg m}^{-3}$  was found subjectively to produce the best balance between forecast reliability and the high probability of detection, at least for this case.

The ensemble forecast NEP swath for  $P(m_h > 0.03 \text{ kg m}^{-3})$  generally captures the three main hail-producing cells, producing relatively skillful forecasts over the portion of the domain observed by the radars, generating forecasts with AUC values of 0.904 for KTLX (Fig. 7a) and 0.899 for KFDR (Fig. 7b). For reference, a perfect forecast produces an AUC of 1.0, while a forecast with no skill produces an AUC of 0.5. The ensemble-predicted hail swaths also show fair to

good forecast reliability (Fig. 8), although hail swath coverage is substantially overforecast when verified against the KFDR observations (Fig. 8b). In both the model and the observations, the geographic extent of the hail swaths is greater at higher altitudes (i.e., farther from the radar site). This pattern is more prominent in the model swaths than in the observed data; the model shows large hail swaths in the northern two storms several kilometers above the surface for the KFDR radar surface (Fig. 7b), but relatively limited hail from these storms reaching the near-surface region observed by KTLX (Fig. 7a). Likewise, the model predicts a hail swath close to the HCA-indicated hail swath for the southern storm along the Texas–Oklahoma border for KTLX (Fig. 7a), but almost no hail for this storm over the near-surface region indicated by the HCA in the KFDR observations (Fig. 7b).

When the ensemble-predicted MESH field (Fig. 5b) is compared to the predicted  $D_{\text{max}}$  at the surface (Fig. 3), it is apparent that the MESH field consistently has greater geographic coverage than the  $D_{\text{max}}$  field. MESH, which relies on the model state above the  $0^\circ\text{C}$  level, indicates the presence of abundant hail and graupel in the three hail-producing storms, with geographic coverage comparable to MESH calculated from WSR-88D observations (Fig. 5). However, the predicted PM mean  $D_{\text{max}}$  field shows less extensive hail swaths in each of the three storms, with almost no hail predicted for the southernmost storm. While the intensity of the southernmost storm is underpredicted in all forecast fields we have considered thus far, the systematic difference between

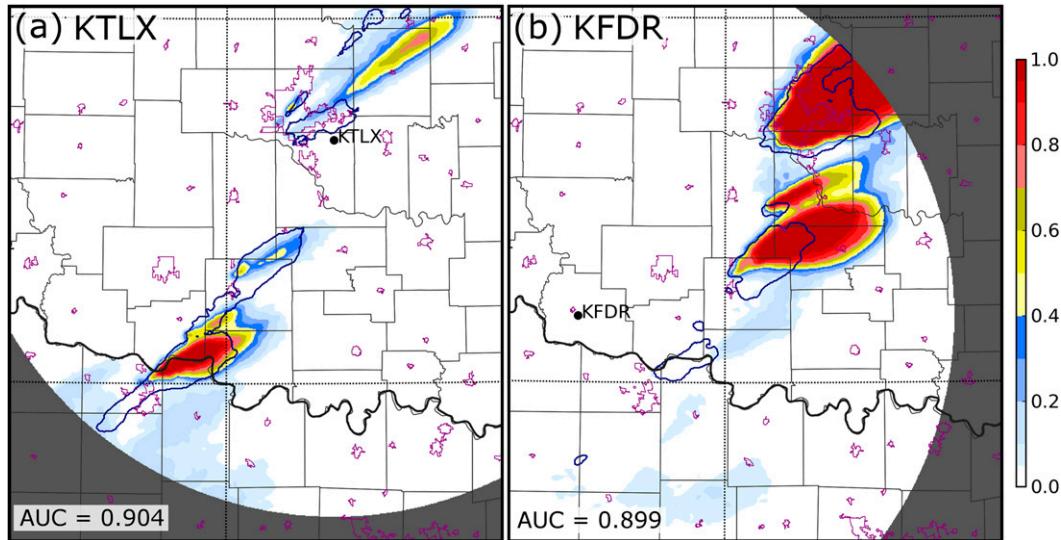


FIG. 7. Neighborhood ensemble probability for  $P(m_h > 0.03 \text{ kg m}^{-3})$  on the  $0.5^\circ$ -elevation level of (a) KTLX and (b) KFDR during the 90-min forecast period (1930–2100 UTC). Observed HCA-indicated hail swaths during this period, smoothed for clarity, are indicated by the dark blue contours.

MESH and  $D_{\max}$  at the surface is also apparent in the northern storms where forecast MESH agreed well with MESH calculated from WSR-88D observations.

To help illuminate the reasons for the noted differences in the MESH and  $D_{\max}$  fields, we examine vertical cross sections of the hail and graupel mass fields taken through the core of the northern storm in an individual ensemble member (Fig. 9). While the cross sections in Fig. 9 are taken from a single member, they are typical of such cross sections within the ensemble. Both hail and graupel are present in the storm: hail is tightly concentrated, while graupel is more widespread within the upper levels of the storm. These results suggest that contributions to reflectivity factor from graupel are in part responsible for the larger geographic extent of MESH swaths compared to swaths of  $D_{\max}$  at the surface. We also note that while substantial hail mass is present in the mid- and upper levels of the storm, exceeding  $4.0 \text{ kg m}^{-3}$ , the mass of hail reaching the surface is much lower (approximately  $0.4\text{--}0.6 \text{ kg m}^{-3}$ ), consistent with the pattern of low  $m_h$  at lower altitudes noted above in our discussion of Fig. 7. These results are consistent with the findings of Johnson et al. (2016), who compared several two-moment microphysics schemes for simulations of an idealized supercell storm produced using the WRF Model and found that the size of the hail simulated with MY2 was relatively small, making hail in the MY2 scheme prone to melting quickly.

While gridded hail observations are not available at the surface, we can see a similar systematic difference between the geographic coverage of MESH (Fig. 5a) and

hail swaths identified in the KTLX and KFDR radar data via the HCA (Fig. 6). The extent of the observed MESH swaths is greater than the extent of the HCA-indicated hail swaths near the surface (e.g., in areas less than 2 km above the surface in Fig. 6), but of similar geographic extent for HCA-based observed hail swaths farther aloft (e.g., in areas 3 km or more above the surface in Fig. 6). This suggests that the discrepancy between MESH and hail swath extent is present to some extent in the observations for this case and is not entirely due to model error (though model error, particularly resulting from the microphysical scheme and faster storm motion, is likely still a sizeable contributor). There are several possible sources of bias in the HCA classifications. For example, it is possible for the HCA to misclassify areas of very heavy rain as containing hail. It is also possible that the HCA-indicated swaths might miss some regions of small hail when they are mixed with intense rainfall, because the HCA identifies only the dominant hydrometeor category; in such a case (low concentrations of small hail mixed with heavy rainfall), the HCA might indicate only rain instead of rain mixed with hail.

#### 4. Summary and discussion

In this study, we evaluated the ability of an EnKF-based storm-scale ensemble forecast (at 500-m horizontal grid spacing) to explicitly predict hail on a 0–90-min time scale for the supercell thunderstorms that occurred over Oklahoma on 20 May 2013. Radar radial velocity and reflectivity data, along with surface observations, were

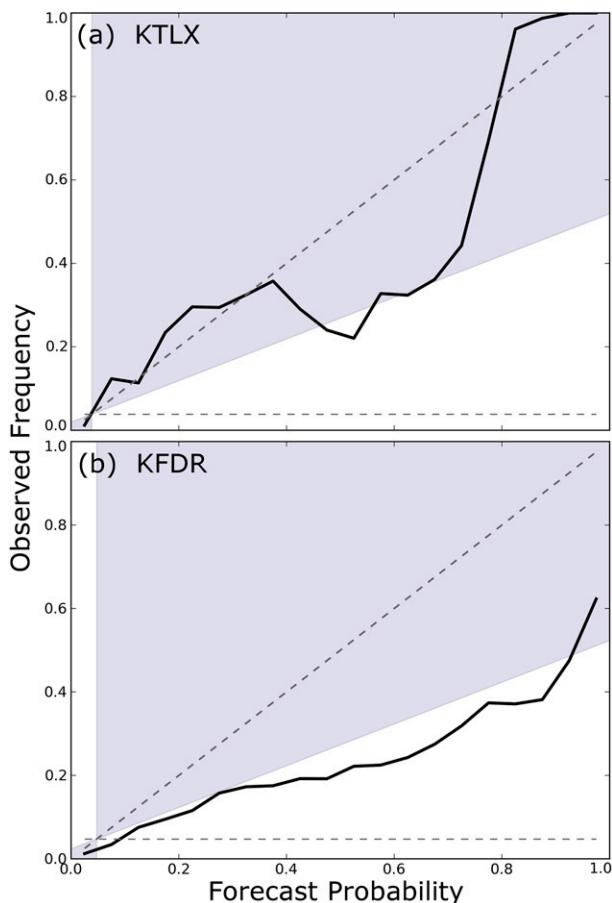


FIG. 8. Reliability diagrams for the probabilistic hail forecasts shown in Fig. 7 over the portion of the forecast domain observed by (a) KTLX and (b) KFDR. Statistics were calculated on the 0.5° radar tilt where it fell within the model domain. The shaded area indicates the region of the reliability diagram corresponding to positive forecast skill.

assimilated into a 40-member ensemble using the ARPS EnKF system; at the end of the data assimilation period, an ensemble of 90-min forecasts was launched. The ensemble analyses and forecasts generally captured the three major observed hail-producing supercells in terms of location, size, and evolution. Some modest errors were noted in the forecast; the predicted storms in the ensemble moved to the northeast somewhat more quickly than the observed storms, and the southernmost of the three main hail-producing storms weakened somewhat in the forecast ensemble, contrary to observations, late in the forecast period.

Ensemble-based hail forecasts were produced and verified using several methods, including the prediction of maximum hail diameter at the surface from the number concentration and mass mixing ratio predicted by the model microphysical scheme, prediction of maximum estimated size of hail (MESH) obtained by applying the

operational MESH algorithm to the ensemble output, and direct prediction of hail mass within the ensemble. These predictions were verified against several different data sources, including observed hail reports at the surface in data collected by the SPC and mPING, MESH calculated from WSR-88D radar observations, and hail swaths obtained from WSR-88D dual-polarization observations using a fuzzy-logic-based hydrometeor classification algorithm.

The ensemble produced forecasts ranging from marginally skillful to highly skillful depending on the forecast and verification method chosen. Surface hail reports were relatively sparse from this case (only four reports were recorded in SPC and mPING data), but all occurred very near or within hail swaths predicted by the ensemble. Population bias in the human reports was likely a major factor in this case; the only surface hail reports received were located within the Oklahoma City metropolitan area, while radar-based hail observations showed large hail swaths in rural areas as well as hail over the metropolitan area. Of the verification methods examined, ensemble prediction of MESH showed the closest agreement with the observations (in this case, MESH calculated from WSR-88D observations interpolated to the model grid). The predicted MESH field correctly identified hail swaths in the three main hail-producing storms, and showed relatively good agreement with observations in terms of maximum hail size and swath extent, except for some underprediction of MESH in the southernmost storm; objective verification of the MESH forecast using the equitable threat score (ETS) indicated positive skill ( $ETS > 0$ ) both for the prediction of all hail ( $>5$  mm) and for the prediction of severe hail ( $>25$  mm). Forecasts of hail mass, when verified against hail swaths identified in dual-polarization WSR-88D observations by the hydrometeor classification algorithm, similarly showed good skill, but revealed a systematic pattern of abundant hail present aloft, but not much hail actually reaching the surface. While the geographic extent of hail swaths observed at upper levels was generally larger than that near the surface, this pattern appeared overly strong in the model, with hail in some swaths (such as for the southernmost storm) melting almost entirely before reaching the surface. Microphysical representation of hail, particularly in multimoment microphysical schemes, is an ongoing area of research; it is hoped that model performance can be improved by future work focusing on the characterization and mitigation of model errors related to the microphysical treatment of hail.

Using dual-polarization WSR-88D observations as a supplemental source of information about the location and extent of hail within the storms using a hydrometeor

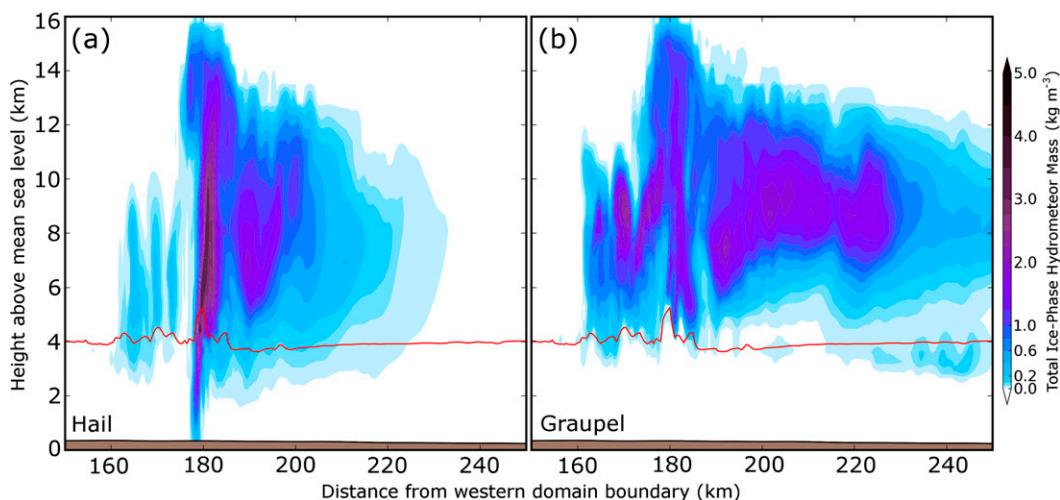


FIG. 9. Vertical west–east cross sections of particle mass fields ( $\text{kg m}^{-3}$ ) for (a) hail and (b) graupel through the region of densest hail in the northernmost storm of ensemble member 23 at 2000 UTC. The height of the  $0^{\circ}\text{C}$  isotherm is indicated by the red line in each panel.

classification algorithm (HCA) provided a valuable means of validating forecasts of the hail field within the ensemble. In particular, the full 3D coverage of the hail region estimates obtained from the HCA allowed for validation of the ensemble hail forecast even in regions of sparse surface hail observations, as well as aloft. We note, however, that such validation is subject to possible error in the HCA; this is a possible topic for future study. Further, since the HCA does not tell us the exact size or the amount of hail present, it should be used in combination with other sources of observations whenever possible.

One way that hail prediction may be improved is through better explicit prediction of the properties of rimed ice. For example, prediction of the maximum hail size at the surface may be improved by using a three-moment scheme (Milbrandt and Yau 2006b). Mansell et al. (2010) and Milbrandt and Morrison (2013) predict bulk graupel density to account for the wide variation in the density of rimed particles. More recently, the predicted particle properties (P3) scheme, which predicts the total ice mass, ice number, ice mass from rime growth, and bulk rime volume, was shown to simulate wide ranges of rimed-ice particle size distributions reasonably well for squall-line and orographic precipitation cases (Morrison and Milbrandt 2015; Morrison et al. 2015). Future studies using these more advanced microphysics schemes could offer new insights and advances in hail prediction.

Prediction of maximum hail diameter at the surface from the number concentration and mass mixing ratio predicted by the model microphysics produced hail swaths in close proximity to observed surface hail reports, though surface reports were relatively sparse and

limited to the Oklahoma City metropolitan area. Because hail surface observations are often relatively sparse, as was true in this case, we note that it is difficult to quantify the advantage of ensemble hail predictions over “nowcasting” methods (e.g., extrapolating based upon the presence of a supercell and estimated storm motion), although we note that nowcasting does not provide quantitative prediction of the size and geographic distribution of hail. To better quantify the benefit of ensemble hail predictions, additional study will be needed, ideally including a variety of storms ranging from those not producing hail at the surface to those producing very large hail at the surface. It may also be beneficial in future studies to apply a hail size discrimination algorithm, such as that of Ryzhkov et al. (2013), to polarimetric radar observations as a supplemental source of information on hail size.

Overall, the results of this study demonstrate the promise of high-resolution, storm-resolving ensembles using state-of-the-art data assimilation schemes to skillfully predict hail on time scales of up to 90 min, as well as the utility of radar observations (especially dual-polarization radar observations) as a supplemental source of information for hail verification. We note, though, that additional studies (including storms of varying convective modes, in varying locations, and at varying times of the day and the year) will be necessary to evaluate the robustness of the results obtained for this case. Furthermore, while it is our hope that storm-scale ensembles of this type will someday become feasible using operational computing resources, the computational resources needed to perform the analyses and forecasts in this study were quite substantial,

necessitating the use of significant supercomputing resources and precluding, at least for now, the use of such a method in real time. As such, we encourage future work aimed at improving the computational efficiency of storm-scale ensembles similar to that used in this study.

**Acknowledgments.** This work was primarily supported by NSF Grant AGS-1261776 as part of the Severe Hail Analysis, Representation, and Prediction (SHARP) project. Supplementary support was also supported by NSF Grants AGS-0802888 and AGS-1046171. 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, Keith Brewster and David Gagne for assistance with radar data processing, and Tim Supinie for helpful discussions regarding postprocessing methods.

#### REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- Bechini, R., and V. Chandrasekar, 2015: A semisupervised robust hydrometeor classification method for dual-polarization radar applications. *J. Atmos. Oceanic Technol.*, **32**, 22–47, doi:10.1175/JTECH-D-14-00097.1.
- Burgess, D. W., and Coauthors, 2014: 20 May 2013 Moore, Oklahoma, tornado: Damage survey and analysis. *Wea. Forecasting*, **29**, 1229–1237, doi:10.1175/WAF-D-14-00039.1.
- CAPS, 2015: CAPS 2015 HWT Spring Experiment EnKF-based ensemble forecast products. Center for Analysis and Prediction of Storms, accessed 7 August 2015. [Available online at [http://www.caps.ou.edu/~fkong/sub\\_atm/spring15-enkf.html](http://www.caps.ou.edu/~fkong/sub_atm/spring15-enkf.html).]
- 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, doi: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, doi:10.1175/1520-0477(1993)074<0645:RAAUWD>2.0.CO;2.
- Davis, S. M., and J. G. LaDue, 2004: Nonmeteorological factors in warning verification. Preprints, *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., P2.7. [Available online at <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, doi: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.
- , E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2014: Low-level ZDR signatures in supercell forward flanks: The role of size sorting and melting of hail. *J. Atmos. Sci.*, **71**, 276–299, doi:10.1175/JAS-D-13-0118.1.
- Dirren, S., R. D. Torn, and G. J. Hakim, 2007: A data assimilation case study using a limited-area ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 1455–1473, doi:10.1175/MWR3358.1.
- Dowell, D. C., L. J. Wicker, and D. J. Stensrud, 2004: High-resolution analyses of the 8 May 2003 Oklahoma City storm. Part II: EnKF data assimilation and forecast experiments. Preprints, *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 12.5. [Available online at <https://ams.confex.com/ams/pdfpapers/81393.pdf>.]
- 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.
- 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, doi: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. Rothfus, 2014: MPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342, doi: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, 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.
- Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHKEFV>2.0.CO;2.
- , and J. S. Whitaker, 2011: What constrains spread growth in forecasts initialized from ensemble Kalman filters? *Mon. Wea. Rev.*, **139**, 117–131, doi:10.1175/2010MWR3246.1.
- Heinselman, P. L., and A. V. Ryzhkov, 2006: Validation of polarimetric hail detection. *Wea. Forecasting*, **21**, 839–850, doi:10.1175/WAF956.1.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- , —, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, doi:10.1175/MWR-2864.1.
- 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. Dawson II, 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., G. Zhang, and M. Xue, 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.
- , M. Xue, and G. Zhang, 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, 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, doi:10.1175/MWR-D-11-00032.1.
- Kong, F., and Coauthors, 2014: CAPS storm-scale ensemble forecasting system: Impact of IC and LBC perturbations. *Proc. 26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction*, Atlanta, GA, Amer. Meteor. Soc., 119. [Available online at <https://ams.confex.com/ams/94Annual/webprogram/Paper234762.html>.]
- 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.
- Lim, S., R. Cifelli, V. Chandrasekar, and S. Y. Matrosov, 2013: Precipitation classification and quantification using X-band dual-polarization weather radar: Application in the Hydro-meteorology Testbed. *J. Atmos. Oceanic Technol.*, **30**, 2108–2120, doi:10.1175/JTECH-D-12-00123.1.
- Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, doi:10.1175/2009JAS2965.1.
- Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment 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 of a hailstorm. *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, K. Ikeda, 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, A. G. Kolodziej, K. A. Scharfenberg, A. Witt, and J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530, doi:10.1175/2009BAMS2815.1.
- Park, H. S., A. 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. A. Snook, and G. Zhang, 2014: The analysis and prediction of microphysical states and polarimetric variables in a mesoscale convective system using double-moment microphysics, multinet radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 141–162, doi:10.1175/MWR-D-13-00042.1.
- 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.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, doi:10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.
- , J. J. Levitt, S. J. Weiss, and D. W. McCarthy, 2004: The frequency of large hail over the contiguous United States. *14th Conf. on Applied Meteorology*, Seattle, WA, Amer. Meteor. Soc., 3.3. [Available online at <https://ams.confex.com/ams/pdfpapers/69834.pdf>.]
- 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.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.
- 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 —, 2013: Impacts of assumed observation errors in EnKF analyses and ensemble forecasts of a tornadic mesoscale convective system. *Proc. 17th Conf. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Austin, TX, Amer. Meteor. Soc., 13.2. [Available online at <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, doi:10.1175/MWR-D-13-00262.1.
- 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 on Forecast System: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, doi:10.1175/2009BAMS2795.1.
- Tong, M., 2006: Ensemble Kalman filter assimilation of Doppler radar data for the initialization and prediction of convective storms. Ph.D. dissertation, University of Oklahoma, 243 pp. [Available online at <https://shareok.org/handle/11244/1001>.]
- , 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 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 radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, **136**, 1649–1668, doi:10.1175/2007MWR2071.1.

- Vivekanandan, J., D. S. Zrnić, S. M. Ellis, R. Oye, A. Ryzhkov, and J. M. Straka, 1999: Cloud microphysics retrieval using S-band dual-polarization radar measurements. *Bull. Amer. Meteor. Soc.*, **80**, 381–388, doi:[10.1175/1520-0477\(1999\)080<0381:CMRUSB>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0381:CMRUSB>2.0.CO;2).
- 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](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, doi:[10.1175/MWR-D-11-00276.1](https://doi.org/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](https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2).
- Wood, V. T., and R. A. Brown, 1997: Effects of radar sampling on single-Doppler velocity signatures of mesocyclones and tornadoes. *Wea. Forecasting*, **12**, 928–938, doi:[10.1175/1520-0434\(1997\)012<0928:EORSOS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0928:EORSOS>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 non-hydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161–193, doi:[10.1007/s007030070003](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, doi:[10.1007/s007030170027](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 forecast. *J. Atmos. Oceanic Technol.*, **23**, 46–66, doi:[10.1175/JTECH1835.1](https://doi.org/10.1175/JTECH1835.1).
- , and Coauthors, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. Preprints, *24th Conf. Several Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.2. [Available online at [https://ams.confex.com/ams/24SLS/techprogram/paper\\_142036.htm](https://ams.confex.com/ams/24SLS/techprogram/paper_142036.htm).]
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observations on the 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](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2).
- Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2015: Practical predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma: Sensitivity to synoptic timing and topographical influence. *Mon. Wea. Rev.*, **143**, 2973–2997, doi:[10.1175/MWR-D-14-00394.1](https://doi.org/10.1175/MWR-D-14-00394.1).