

Evaluating Ensemble Kalman Filter Analyses of Severe Hailstorms on 8 May 2017 in Colorado: Effects of State Variable Updating and Multimoment Microphysics Schemes on State Variable Cross Covariances

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 6 September 2019, in final form 23 March 2020)

ABSTRACT

Ensemble Kalman filter (EnKF) analyses of the storms associated with the 8 May 2017 Colorado severe hail event using either the Milbrandt and Yau (MY) or the NSSL double-moment bulk microphysics scheme in the forecast model are evaluated. With each scheme, two experiments are conducted in which the reflectivity (Z) observations update in addition to dynamic and thermodynamic variables: 1) only the hydrometeor mixing ratios or 2) all microphysical variables. With fewer microphysical variables directly constrained by the Z observations, only updating hydrometeor mixing ratios causes the forecast error covariance structure to become unreliable, and results in larger errors in the analysis. Experiments that update all microphysical variables produce analyses with the lowest Z root-mean-square innovations; however, comparing the estimated hail size against hydrometeor classification algorithm output suggests that further constraint from observations is needed to more accurately estimate surface hail size. Ensemble correlation analyses are performed to determine the impact of hail growth assumptions in the MY and NSSL schemes on the forecast error covariance between microphysical and thermodynamic variables. In the MY scheme, Z is negatively correlated with updraft intensity because the strong updrafts produce abundant small hail aloft. The NSSL scheme predicts the growth of large hail aloft; consequently, Z is positively correlated with storm updraft intensity and hail state variables. Hail production processes are also shown to alter the background error covariance for liquid and frozen hydrometeor species. Results in this study suggest that EnKF analyses are sensitive to the choice of MP scheme (e.g., the treatment of hail growth processes).

1. Introduction

Hail causes significant damage to both crops and personal property in the United States—often exceeding \$1 billion (U.S. dollars) in damage annually (Changnon 2009). Damage from hail is particularly costly when urban areas are affected. Just since the late 1990s, multiple major metropolitan areas have sustained more than \$1 billion in damage from a single hail event, including Dallas/Fort Worth, Texas; Denver, Colorado; Kansas City, and St. Louis, Missouri; and greater Chicago (Edwards and Thompson 1998; Changnon

and Burroughs 2003; Changnon et al. 2009; NOAA National Centers for Environmental Information 2018). The continued expansion of these metropolitan areas is expected to further increase hail damage in the future (Rosencrants and Ashley 2015).

Increasing the warning lead time of severe hail events can potentially mitigate damage by allowing more time for preparation. Current warning lead times for severe hail are limited because the National Weather Service (NWS) issues warnings based upon the detection of severe hail (diameter > 1.0 in.). Generally, hail detections are based upon either surface-based reports or radar signatures that are indicative of severe hail. To extend severe hail warning lead times the NWS is

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

DOI: 10.1175/MWR-D-19-0300.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by UNIVERSITY OF OKLAHOMA LIBRARY | Unauthenticated | Downloaded 12/30/20 10:35 PM UTC

investigating a warn-on-forecast framework (Stensrud et al. 2009, 2013), where instead of issuing warnings based upon the detection of severe hail, the NWS will issue warnings based upon high-resolution, frequently updated numerical weather prediction (NWP) model guidance. While convection-allowing forecasts can skillfully predict convective hazards (e.g., Kain et al. 2008, 2010; Clark et al. 2012; Dawson et al. 2012; Snook et al. 2016, 2019; Yussouf et al. 2015; Yussouf and Knopfmeier 2019; Johnson et al. 2015; Johnson and Wang 2017; Schwartz et al. 2015; Skinner et al. 2018; Dawson et al. 2017; Labriola et al. 2017, 2019a; Jones et al. 2016, 2019; Supinie et al. 2016, 2017; Gallo et al. 2019; Stratman et al. 2020), forecast skill is highly dependent upon errors introduced by the initial conditions and model physics.

Previous studies have shown, either through observing system simulation experiments (OSSEs) or real-data experiments (e.g., Snyder and Zhang 2003; Zhang et al. 2004, 2006; Dowell et al. 2004; Tong and Xue 2005; Caya et al. 2005; Xue et al. 2006; Snook et al. 2011; Dawson et al. 2012; Romine et al. 2013; Schwartz et al. 2015; Johnson and Wang 2017; Lawson et al. 2018; Stratman et al. 2020) that an ensemble Kalman filter (EnKF; Evensen 1994, 2003) can be successfully applied to convection-resolving forecasts. EnKF methods employ an ensemble of forecasts to sample error covariance. Flow-dependent error covariances derived from ensemble forecasts are used to correct errors in unobserved variables during data assimilation (DA) that strongly influence convective-scale dynamics such as updraft speed, in-cloud temperature, and microphysical properties (Tong and Xue 2005, hereafter TX05; Tong and Xue 2008a). This feature of EnKF is particularly useful when assimilating radar data (e.g., Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Tong and Xue 2005; Caya et al. 2005; Dowell and Wicker 2009; Aksoy et al. 2009), which provide indirect but high spatial and temporal resolution observations.

Even when forecasts are initialized from relatively accurate initial conditions, forecast error will continue to increase with time due to model errors including those associated with subgrid-scale processes (Zhu and Navon 1999; Houtekamer et al. 2005; Zhang et al. 2006; Hawblitzel et al. 2007; Melhauser and Zhang 2012; Zhang et al. 2015). For convective-scale simulations, one of the largest sources of model error are microphysical parameterizations. These parameterizations represent fine, subgrid-scale processes that undergo rapid, nonlinear transformations at convective scales and often rely upon ad hoc procedures due to the limited understanding of microphysical processes.

Bulk microphysics schemes (referred to as MP schemes) are the most commonly used type of microphysical parameterization in convective modeling studies. MP schemes predict the evolution of an assumed hydrometeor particle size distribution (PSD) that typically follows a gamma distribution (Ulbrich 1983):

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

where D is the particle diameter, N_{0x} is the intercept parameter, λ_x is the slope parameter, and α_x is the shape parameter for hydrometeor species x . To diagnose the PSD, MP schemes predict certain p moments of the distribution:

$$M_x(p) = \int_0^{\infty} D^p N_x(D) dD = \frac{N_{0x} \Gamma(\alpha_x + p + 1)}{\lambda_x^{\alpha_x + p + 1}}, \quad (2)$$

where $\Gamma(n)$ is a gamma function. The zeroth, third, and sixth moments of a PSD are most commonly predicted by MP schemes and are proportional to hydrometeor number concentration (N_{tx}), mixing ratio (q_x), and reflectivity (Z_x), respectively. The number of predicted moments corresponds to the number of diagnosed PSD parameters. All undiagnosed parameters are typically assumed to be constant.

While single-moment MP schemes (e.g., Lin et al. 1983) are generally computationally efficient because they predict only hydrometeor mixing ratio, they are unable to accurately represent microphysical processes such as selective melting/evaporation and size sorting. Multimoment MP schemes (e.g., Ferrier 1994; Milbrandt and Yau 2005a; Morrison et al. 2005; Morrison and Grabowski 2008; Thompson et al. 2008; Mansell et al. 2010; Milbrandt and Morrison 2013; Morrison and Milbrandt 2015; Milbrandt and Morrison 2016; Lim and Hong 2010) can replicate these microphysical processes and improve the representation of storm structure. Double-moment schemes that predict both q_x and N_{tx} (diagnose N_{0x} , λ_x), can reproduce observed polarimetric radar signatures (Jung et al. 2012; Johnson et al. 2016; Putnam et al. 2014, 2017a,b); however, multiple studies (e.g., Milbrandt and Yau 2005b; Dawson et al. 2014; Johnson et al. 2016; Morrison et al. 2015) note these schemes suffer from excessive size sorting. Excessive size sorting can be mitigated using a correction mechanism (Thompson et al. 2008; Mansell 2010), a diagnostic shape parameter, or by using a triple-moment scheme (Milbrandt and Yau 2005b).

In some newer MP schemes, instead of predicting many static hydrometeor categories (e.g., Straka and Mansell 2005), the scheme predicts the evolution of particle characteristics such as density (Mansell et al. 2010; Milbrandt and Morrison 2013; Morrison and

Milbrandt 2015; Morrison et al. 2015; Milbrandt and Morrison 2016). MP schemes generally predict the density of rimed ice; this is because ice particle density can vary substantially, and because ice particles can undergo large fluctuations in density during the riming process. Prognostic density equations allow the MP scheme to update rimed ice fall speeds and improve the representation of hail production processes (Labriola et al. 2019a).

Predicted microphysical variables can be used to diagnose surface hail size in a NWP forecast, either via simulated radar-derived hail products (e.g., Snook et al. 2016; Luo et al. 2017, 2018; Labriola et al. 2017) such as the maximum estimated size of hail (MESH; Witt et al. 1998a), or by using hail PSDs to diagnose the largest observable hail size (e.g., Snook et al. 2016; Labriola et al. 2017, 2019a,b; Luo et al. 2018; Gagne et al. 2019). Hail size forecasts not only provide useful information for forecasters but are also useful in evaluating MP scheme treatment of hail growth and decay processes. Due to the limited coverage and biased nature of surface hail reports (e.g., Sammler 1993; Witt et al. 1998b; Doswell et al. 2005), most hail size forecasts are evaluated against radar-derived hail products that serve as a proxy for hail size such as MESH or output from a hydrometeor classification algorithm (HCA; Heinselman and Ryzhkov 2006; Park et al. 2009; Ryzhkov et al. 2013; Ortega et al. 2016).

Over the continental United States, the Next Generation Weather Radar (NEXRAD; Crum et al. 1993) network provides full volumetric scans at a temporal resolution sufficient to observe the rapid evolution of convective storms. Reflectivity (Z) is proportional to the sixth moment of a hydrometeor PSD and is strongly influenced by larger particles within a radar volume. After a recent polarimetric upgrade to the NEXRAD system, radars now emit both horizontally and vertically polarized signals. Dual-polarization variables such as differential reflectivity (Z_{dr}), copolar cross-correlation coefficient (ρ_{hv}), and specific differential phase (K_{dp}), which in addition to single-polarization products, Z and radial velocity (V_r), provide information about hydrometeor size, orientation, phase, and shape (Doviak et al. 2000; Kumjian and Ryzhkov 2008). Due to the large amount of hydrometeor information that can be inferred from polarimetric radar products, several recent studies have used simulated or observed polarimetric variables (e.g., Jung et al. 2010a, 2012; Ryzhkov et al. 2011; Kumjian et al. 2014; Dawson et al. 2014; Putnam et al. 2014, 2017a,b, 2019; Johnson et al. 2016, 2018, 2019; Snook et al. 2016; Snyder et al. 2017a,b) to evaluate microphysical parameterizations.

Generally, Z is the primary observation of microphysical relevance assimilated during DA. Despite a

limited number of observations, OSSEs have shown that an EnKF system can reasonably update single-moment MP scheme variables when assimilating Z and V_r (TX05). Multimoment schemes, which predict approximately twice as many microphysical variables as a single-moment scheme, are even more underconstrained by observations. Although Xue et al. (2010) (hereafter XJZ10) obtained a reasonably good analysis using a double-moment scheme in an OSSE, it was noted that Z alone may be insufficient to constrain the increased number of variables in multimoment schemes because many different microphysical configurations can correspond to the same Z . Additionally, the large number of predicted variables introduces error into the analysis because increasing the number of error-containing parameters decreases the accuracy of the estimation (Aksoy et al. 2006; Tong and Xue 2008b; Jung et al. 2010b).

Relatively few studies have considered developing an optimal EnKF-based framework for updating multimoment MP scheme variables during DA. Previously, to improve constraint of microphysical variables, XJZ10 assimilated both radar Z and V_r , and Labriola et al. (2019a) used an EnKF system to update only hydrometeor mixing ratio; however, both studies conclude that there is a need to further analyze the ability of an EnKF system to update multimoment scheme variables. This study evaluates EnKF analysis sensitivity to selective state variable updating and the microphysical assumptions made within two double-moment MP schemes used in the prediction model. Knowledge gained through this study provides information that can be used to optimize EnKF configurations when using such multimoment MP schemes. To this end, EnKF analyses are evaluated for the 8 May 2017 Colorado severe hail event. The rest of the paper is organized as follows: in section 2a brief overview of the 8 May 2017 Colorado severe hail event is provided along with the experimental design, including the EnKF and prediction model settings. Verification statistics and ensemble correlation analyses are provided in section 3. Finally, section 4 contains a summary and discussion of the results.

2. Methods

a. Case overview

On 8 May 2017, multiple hail producing thunderstorms occurred over north-central Colorado. According to surface-based hail reports received by the Storm Prediction Center (SPC), these storms produced severe (diameter > 25 mm) and significant severe (diameter > 50 mm) hail starting at approximately 2000 UTC. One storm that impacted downtown Denver (Fig. 1) caused a particularly large amount of damage; due to heavy traffic volume

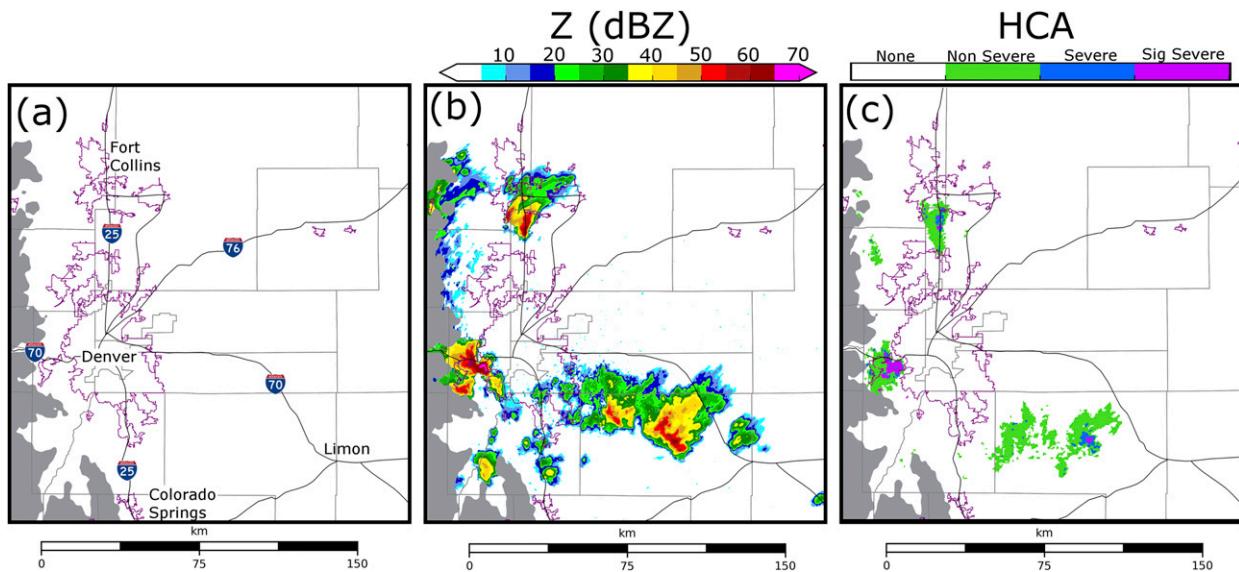


FIG. 1. (a) An areal map of the Denver region. (b) Observed Z from the lowest radar tilt (0.5°) of KFTG (Denver) at 2040 UTC. (c) Merged HCA output between 2000 and 2040 UTC. The HCA is applied to the lowest radar tilts (0.5°) of the two closest radars: KFTG (Denver) and KCYS (Cheyenne). Purple contours represent urban boundaries, thin black lines represent major highways. All regions of the domain that are more than 2.5 km above mean sea level are shaded gray.

during evening rush hour, a large number of cars sustained major hail damage during the storm (Fritz 2017). As of the date of this writing, an estimated \$2.3 billion in insurance claims (NOAA National Centers for Environmental Information 2018) have been filed in Colorado for hail damage sustained during this severe weather event.

Forcing for large-scale ascent was moderate during this severe weather event. Atmospheric conditions were conducive for the development of supercell thunderstorms with moderate shear [0–6 km shear: 30–40 kt ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$)] and instability (mixed-layer CAPE: 1000–1500 J kg^{-1}) (Marsh 2017). On 8 May 2017, Colorado was located downstream of a trough associated with an upper-level low pressure system stationed over Baja California. Cyclonic vorticity advection from the trough paired with low-level convergence along a weak frontal boundary initiated multicellular storms to the south and east of Denver at approximately 1930 UTC. Additional, more isolated storms initiated along the front range of the Rocky Mountains via upslope flow, one of these storms later produced significant severe hail over Denver (Fig. 1c). For additional information regarding this severe weather event we refer the reader to Labriola et al. (2019b).

b. Prediction model settings

All experiments use the Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001) NWP model on a domain of $483 \times 443 \times 53$ grid points with a horizontal grid spacing of 500 m. The model grid is stretched in the vertical, with a minimum grid spacing of 50 m at the

surface and an average vertical grid spacing of 425 m. Model physics include NASA Goddard Space Flight Center short- and longwave radiation parameterization; surface fluxes calculated from surface drag coefficients, surface temperature, and volumetric water; a two-layer soil model; and a 1.5-order turbulent mixing and planetary boundary layer parameterization (Xue et al. 2000, 2001).

Ensemble forecasts are run using either the Milbrandt and Yau (2005a) double-moment (MY) MP scheme or the NSSL double-moment variable density rimed ice MP scheme (Mansell et al. 2010). These MP schemes were selected for this study because they are commonly used multimoment MP schemes that have been used in several previous experiments to produce explicit hail forecasts (Milbrandt and Yau 2006; Snook et al. 2016; Luo et al. 2017, 2018; Labriola et al. 2017, 2019a,b). Following Labriola et al. (2019a), the minimum hydrometeor number concentration threshold is set to 10^{-8} m^{-3} for both schemes to avoid excessive removal of hail near the surface, which can result when using the MY default setting (10^{-3} m^{-3}).

Initial and boundary conditions are interpolated from the Center for Analysis and Prediction of Storms (CAPS) EnKF storm-scale ensemble forecast (SSEF; Jung et al. 2018)—a 40-member ensemble of forecasts run using the Advanced Research version of the Weather Research and Forecasting Model (WRF-ARW; Skamarock et al. 2008). The SSEF uses 3-km horizontal grid spacing, and its model domain spans the contiguous United States. The SSEF was initialized via EnKF at 1800 UTC 8 May 2017 from the North American Mesoscale Forecast System

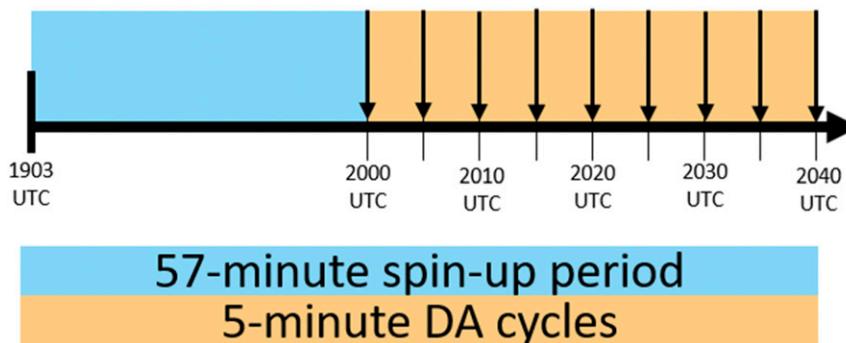


FIG. 2. The experiment configuration for ensembles MY-Q, MY-ALL, NSSL-Q, and NSSL-ALL. The blue region represents the approximate 1-h spinup period and the orange region represents the 40-min period when 5-min DA cycling occurs. Downward pointing arrows indicate where assimilation occurs.

(NAM; [Environmental Modeling Center 2017](#)) analysis, with initial perturbations introduced by the Short-Range Ensemble Forecast (SREF; [Du et al. 2015](#)). The SSEF is run until 0000 UTC 9 May 2017. At 1903 UTC the 500-m domain is interpolated from the SSEF, with boundary conditions obtained from the SSEF every 9 min; these times were selected because the SSEF outputs model data every 9 min. No microphysical information is provided along domain boundaries to avoid development of spurious convection. For additional information regarding the 2017 CAPS SSEF, we refer the reader to [CAPS \(2017\)](#).

c. DA and observation operator settings

After initialization, the 500-m ensemble forecasts undergo a ~ 60 -min spinup period until 2000 UTC, when DA cycling begins ([Fig. 2](#)). 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, assimilates all available observations every 5 min between 2000 and 2040 UTC ([Fig. 2](#)).

Surface (ASOS and AWOS) observations and data from the Denver NEXRAD radar (KFTG) are assimilated during the experiment. Observation error settings follow [Snook et al. \(2016\)](#). For surface observations, errors are assumed to be 1.5 m s^{-1} for horizontal wind components u and v , 2.0 K for potential temperature θ , and 2.0 K for dewpoint temperature T_d . The covariance localization radius is set to 300 km in the horizontal and 6 km in the vertical. Surface observations update only select model state variables (i.e., u , v , θ , q_v), as was done in [Labriola et al. \(2019a\)](#) because unreliable covariances led to the intensification of spurious updrafts during DA. KFTG observation errors are assumed to be 4.0 m s^{-1} for V_r and 6.0 dBZ for Z . These values are larger than typical instrument errors associated with NEXRAD radars (e.g.,

[Doviak and Zrnić 1993](#); [Ryzhkov et al. 2005](#)), but for data assimilation purposes, the errors also include representative errors and other sources of uncertainties. Also, specifying relatively large observation errors can help alleviate underdispersion problems of the EnKF ensemble and past sensitivity studies have shown better performance of the ensemble analyses and subsequent ensemble forecasts (e.g., [Dowell et al. 2004](#); [Snook et al. 2013](#)). The [Gaspari and Cohn \(1999\)](#) covariance localization cutoff radius is set to 3 km in both the horizontal and vertical. KFTG data are interpolated horizontally to the model domain, but the height of the radar beam is preserved in the vertical ([Xue et al. 2006](#)); Z and V_r observations are preprocessed in the horizontal directions to 1 km grid spacing in regions of precipitation ($Z > 5 \text{ dBZ}$) and to 2 km spacings in regions of clear air ($Z < 5 \text{ dBZ}$). Given that the radar observations have average resolution of about 1 km, such observation density preserves most useful information in the observations. The relaxation-to-prior spread ([Whitaker and Hamill 2012](#)) algorithm is used to inflate the posterior ensemble spread to 95% of the prior spread for all thermodynamic variables (i.e., u , v , w , θ , p) and hydrometeor mixing ratio q_x . Inflating all hydrometeor state variables (i.e., q_x , N_{rx} , and v_x) can potentially cause PSDs to become unbalanced, and cause hydrometeor properties to become unrealistic (e.g., hail is too dense). Following [Labriola et al. \(2019a\), \$N_{rx}\$ and \$v_x\$ are updated during inflation to preserve EnKF estimated mean mass diameter and density, respectively.](#)

The radar observation operator used during this experiment, which is a variant of the [Jung et al. \(2010a\)](#) T-matrix method, uses lookup tables based upon PSD parameters to calculate the hydrometeor scattering amplitude; an in-depth discussion of this observation operator is provided in [Putnam et al. \(2019\)](#). Generally, MP

schemes do not track meltwater, instead the water fraction is diagnosed in the observation operator. The Jung et al. (2008a) melting model is used to create an ice–water hybrid mixture for both hail and graupel in the observation operator. The mixing ratio of the ice–water hybrid category is determined via a ratio between ice and rainwater mixing ratios, the number concentration is updated to preserve particle mean mass diameter and prevent the excessive accumulation of rainwater on the surface of ice particles (Labriola et al. 2019a).

A total of four experiments are conducted during this study; ensembles run using either the MY or NSSL schemes will assimilate observations where either Z updates: 1) hydrometeor mixing ratios only or 2) all microphysical state variables (q_x, N_{ix}, v_x); in addition to dynamic and thermodynamic variables (u, v, w, θ, q_v). To determine the impact of assimilating Z observations, no other observations update microphysical state variables. Experiments are named after the MP scheme used and the microphysical variables that are updated (i.e., MY-Q, MY-ALL, NSSL-Q, NSSL-ALL). In addition to assimilating Z , V_r is used in all experiments to update u, v , and w , along with water vapor mixing ratio (q_v). The V_r forward observation operator does not consider hydrometeor terminal velocities due to the large uncertainties in hydrometeor size distribution and fall speed assumptions and small elevation angles of radar beams.

Analyses and forecasts during the DA period (2000–2040 UTC) are objectively and subjectively evaluated to determine which experiment produces the most accurate initial state estimate. This time period was selected for evaluation because radar derived hail products (i.e., HCA output) indicate that multiple thunderstorms were producing significant severe surface hail, the largest HCA hail size bin. While storms caused more substantial damage after 2040 UTC when they impacted downtown Denver, they produced similarly large hailstones both during and after the DA period. Forecast evaluation after 2040 UTC is beyond the scope of this study, instead we refer the reader to previous studies (e.g., Snook et al. 2016; Luo et al. 2018) that evaluate surface hail size forecasts.

d. Analysis evaluation procedure

Due to the nature of this severe weather event, estimated surface hail size is verified against radar-based hail size estimates. Surface-based hail reports are generally too sparse and unreliable to be used for verification; severe hail occurrences are underreported in rural areas and away from highways, as well as when a more severe weather event, such as a tornado, occurs nearby (Doswell et al. 2005; Witt et al. 1998b; Allen and Tippett 2015). For example, there are only 4 hail reports recorded by the SPC during the DA period (2000–2040 UTC)

and all reports are in Denver. Radar observations during the same period (Fig. 1c) indicate multiple hail producing storms were present in the experiment domain. Further, hail sizes that correspond with familiar circular objects (e.g., quarters, golfballs) are overreported by the public (Sammler 1993), skewing reported hail sizes. Radar derived hail products such as MESH (Witt et al. 1998a) and HCA output (Heinselman and Ryzhkov 2006; Park et al. 2009; Ryzhkov et al. 2013; Ortega et al. 2016; Putnam et al. 2017b) have been shown to be superior to surface-based reports for hail size verification because they have a high spatial resolution and fewer systematic biases (Cintineo et al. 2012).

When working with unbiased observations, HCA output is shown to better discriminate hail size than the MESH algorithm (Ortega et al. 2016). Subjective comparisons also demonstrate the HCA output closely resembles surface-based hail reports (when available) both in terms of hail size and the location of the report. The Ortega et al. (2016) HCA applies a fuzzy logic algorithm to both single- (Z) and dual-polarization (Z, Z_{dr}, ρ_{hv}) radar observations along with environmental information (i.e., temperature) to classify the dominant hail size for a grid point as either: nonsevere (≥ 5 mm), severe (≥ 25 mm), or significant severe (≥ 50 mm). Output from the Ortega et al. (2016) HCA (Fig. 1c) along with radar observed Z (Fig. 1b) is used to evaluate ensemble analyses and forecasts during the DA period.

In this study, the HCA is applied to NEXRAD radar observations from both Denver (KFTG) and Cheyenne (KCYS); these are the two closest radars to the hail producing storms. Radar observations are interpolated to the 500-m model grid for verification, with a 9-point smoother applied to Z_{dr} and ρ_{hv} to reduce noise. Environmental information is obtained from the ensemble forecasts. HCA output is considered within a 120 km radius of a given radar, this is the maximum distance Ortega et al. (2016) uses to diagnose surface hail size. HCA output from KFTG and KCYS is merged by taking the maximum detected hail size at each grid point and applying a smoothing filter that decreases hail size detections by one size bin (e.g., significant severe to severe) when the majority of grid points within a 1 km radius are smaller than the given HCA detection (Labriola et al. 2019a).

NWP models do not explicitly predict surface hail size, instead hail PSDs are frequently used to diagnose the maximum observable hail size at a grid point (e.g., Snook et al. 2016; Labriola et al. 2017, 2019a,b; Luo et al. 2018; Gagne et al. 2019). This study uses the Snook et al. (2016) maximum hail size algorithm, which defines the maximum hail size (D_{max}) as the largest diameter for

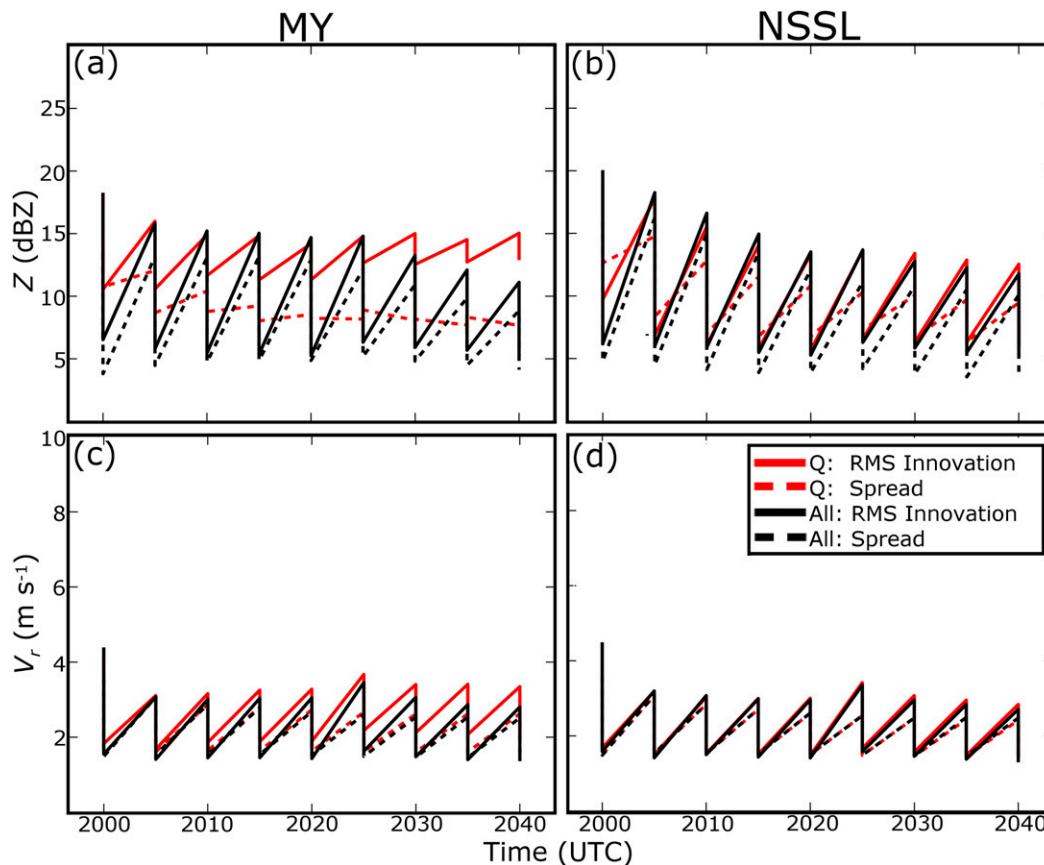


FIG. 3. The ensemble RMSI (solid lines) and spread (dotted lines) for (a),(b) Z and (c),(d) V_r . The performance of ensembles (a),(c) MY-Q and (b),(d) NSSL-Q are marked with red lines, and (a),(c) MY-ALL and (b),(d) NSSL-ALL are marked with black lines. Statistics are calculated over the experiment domain from 2000 to 2040 UTC, calculations are limited to locations where the observed and/or model (ensemble mean) Z exceeds 15 dBZ.

which the PSD predicts at least 1 hailstone in a $100\text{ m} \times 100\text{ m}$ box located in the lowest meter of the atmosphere ($N_{\min} = 10^{-4}\text{ m}^{-4}$). This criterion is similar to minimum number concentration thresholds defined for hail in previous studies (e.g., Milbrandt and Yau 2006; Gagne et al. 2019). Although the minimum number concentration threshold can be modified, the average D_{\max} value does not substantially change (approximately 1–2 mm) when the threshold is increased or decreased by an order of magnitude. The spatial coverage of analysis estimated D_{\max} at the lowest model height ($\sim 25\text{ m}$ AGL) is compared subjectively and objectively to the HCA output (Fig. 1c).

3. Results

a. Evaluating ensemble forecast and analysis innovations

The root-mean-square innovation (RMSI) and ensemble spread of Z and V_r (Fig. 3) quantitatively

evaluate EnKF analyses and forecasts. The RMSI is defined as

$$\text{RMSI} = \sqrt{\langle d^2 \rangle}, \tag{3}$$

where $\langle d^2 \rangle$ is the mean squared innovation or difference between the observation and model mapped to observation space; d is defined as

$$d = y^o - \overline{H(x)}, \tag{4}$$

where y^o is the observation, H is the forward operator that maps the model state vector to observation space, and x is either the model state forecast or analysis vector, respectively. Innovations are averaged over the KFTG volume in regions where either the observed or ensemble mean simulated Z exceeds 15 dBZ. This criterion includes both observed precipitation and spurious echoes, but eliminates potentially large regions of clear air and light precipitation from the statistics (Snook et al. 2011; Jung et al. 2012).

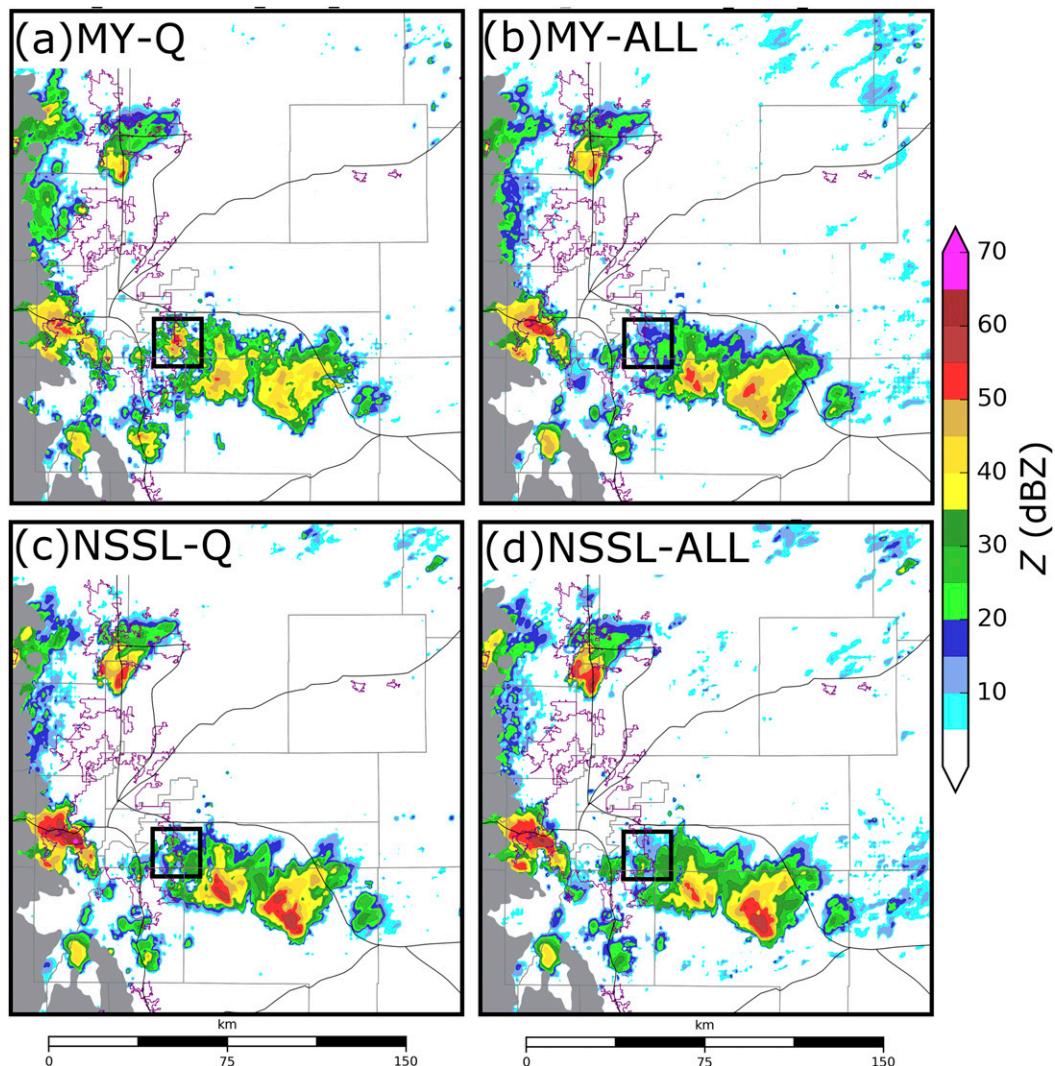


FIG. 4. Ensemble mean simulated Z interpolated to the lowest tilt (0.5°) of KFTG at the time of the final analysis (2040 UTC) for ensembles (a) MY-Q, (b) MY-ALL, (c) NSSL-Q, and (d) NSSL-ALL. A black square in each figure marks the subdomain analyzed in Fig. 6, the background map is the same as Fig. 1.

Although innovations are relatively similar between NSSL-Q and NSSL-ALL (Figs. 3b,d), RMSIs differ substantially between MY-Q and MY-ALL for Z (Fig. 3a). The MY-Q and MY-ALL RMSIs for V_r are relatively similar (Fig. 3c); this is partly because only winds in the precipitation regions are included in the statistics while differences in the outflow regions are not included. MY-Q RMSIs for Z exhibit only a small reduction during DA cycling, suggesting updating only one PSD parameter for multimoment schemes may lead to an unreasonable PSD in MY-Q (Fig. 3a). RMSIs for MY-Q and MY-ALL V_r are more similar than RMSIs for Z . Wind speed is indirectly influenced by microphysical state variables through thermodynamic feedbacks, and therefore unlike Z , is less affected

by updates to total number concentrations during assimilation. When only parts of PSD variables are updated, the RMSI for MY-Q V_r does not change much with cycles (Fig. 3c). This behavior is not observed in NSSL-Q for reasons that will be discussed later in this section.

Analysis mean simulated Z (Fig. 4) and KFTG observations (Fig. 1b) are subjectively compared at the time of the final DA cycle (2040 UTC) to identify why the RMSI for MY-Q Z (Fig. 3a) is larger than any other ensemble (Figs. 3a,b). Although there are slight variations in simulated Z intensity, MY-ALL (Fig. 4b), NSSL-Q (Fig. 4c), and NSSL-ALL (Fig. 4d) analyses closely match observed Z (Fig. 1b). MY-Q (Fig. 4a) Z is somewhat different from observations (Fig. 1b); storm

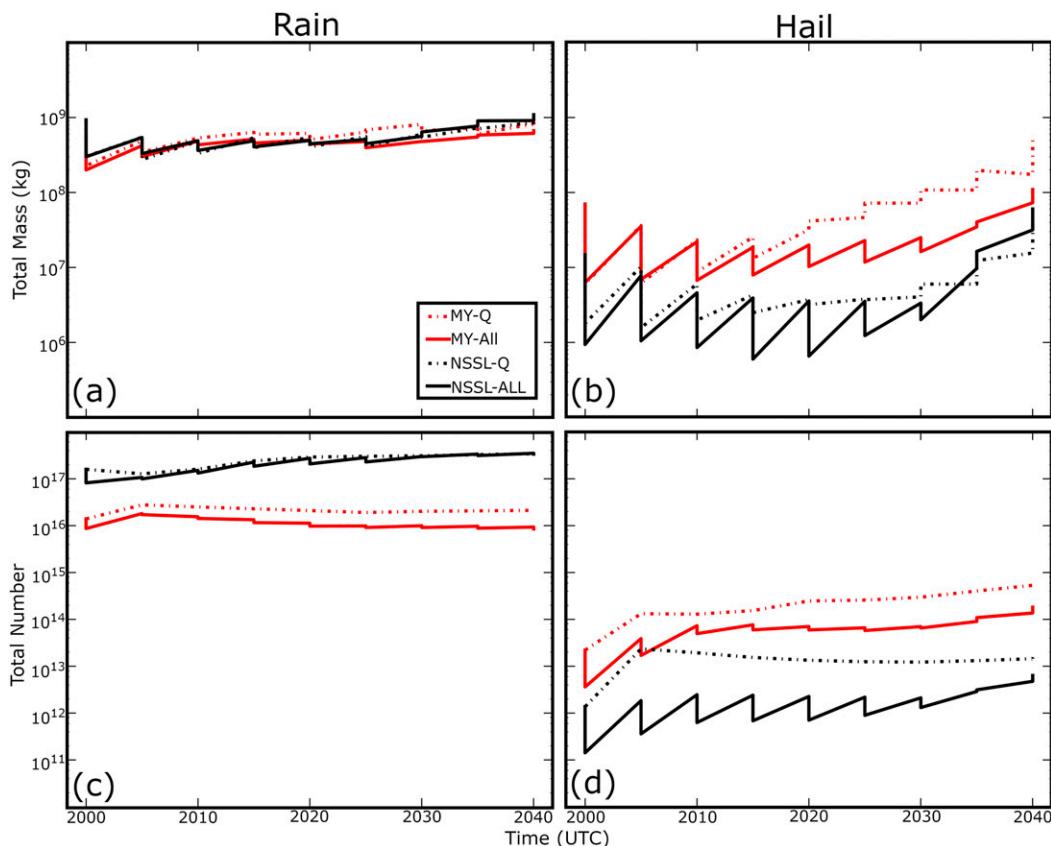


FIG. 5. Ensemble mean (forecast and analysis) (a),(b) total mass and (c),(d) number of (a),(c) rain and (b),(d) hail between 2000 and 2040 UTC. Hydrometeor concentrations are summed over a volume that spans the experiment subdomain.

structure is less organized, and the MY-Q analysis frequently underestimates Z . Additionally, MY-Q predicts reflectivity to be spuriously intense in regions of relatively little observed precipitation, such as to the east of Denver (Fig. 4a). Poor MY-Q Z estimates that increase the RMSI for MY-Q Z (Fig. 3a), are due in part to unreliable error covariances that develop as a consequence of unbalanced updates of hydrometeor mixing ratios and number concentrations.

Despite predicting weaker Z values in the convective core of the observed storms, MY-Q predicts the total mass and number of rain (Figs. 5a,c) and hail (Figs. 5b,d) to be larger than MY-ALL. This is partly because MY-Q predicts more spurious Z throughout the domain during DA (Fig. 4a); however, the ensemble also predicts observed storms to produce more precipitation. NSSL-Q (Fig. 4c) and NSSL-ALL (Fig. 4d) predict more intense Z than either MY ensemble (Figs. 4a,b) despite both experiments predicting substantially less hail in terms of both mass and number (Figs. 5b,d). This is because although the NSSL scheme predicts storms to produce relatively large hail aloft that contributes to high

Z values. Results agree with previous studies (e.g., Johnson et al. 2016, 2019), which note the MY scheme predicts storms to produce a large number of small hailstones aloft while the NSSL scheme predicts storms to produce fewer but larger hailstones. Differences in simulated Z between the experiments demonstrate that Z is not a monotonic function of hydrometeor mixing ratio but is sensitive to changes in the hydrometeor particle size distribution.

For an EnKF system to accurately estimate model state variables that are not directly observed, the filter must develop reliable multivariate covariances during the assimilation period. The cross covariance between state variables and observation priors is used to retrieve unobserved variables and modify storm structure. Previous studies, such as TX05 and XJZ10, analyzed forecast cross correlations between model state variables and an observation prior at an assumed observation location. Such correlation analyses demonstrate the expected impact of assimilating that observation on analyzed model state variables; a positive (negative) correlation between state variables and

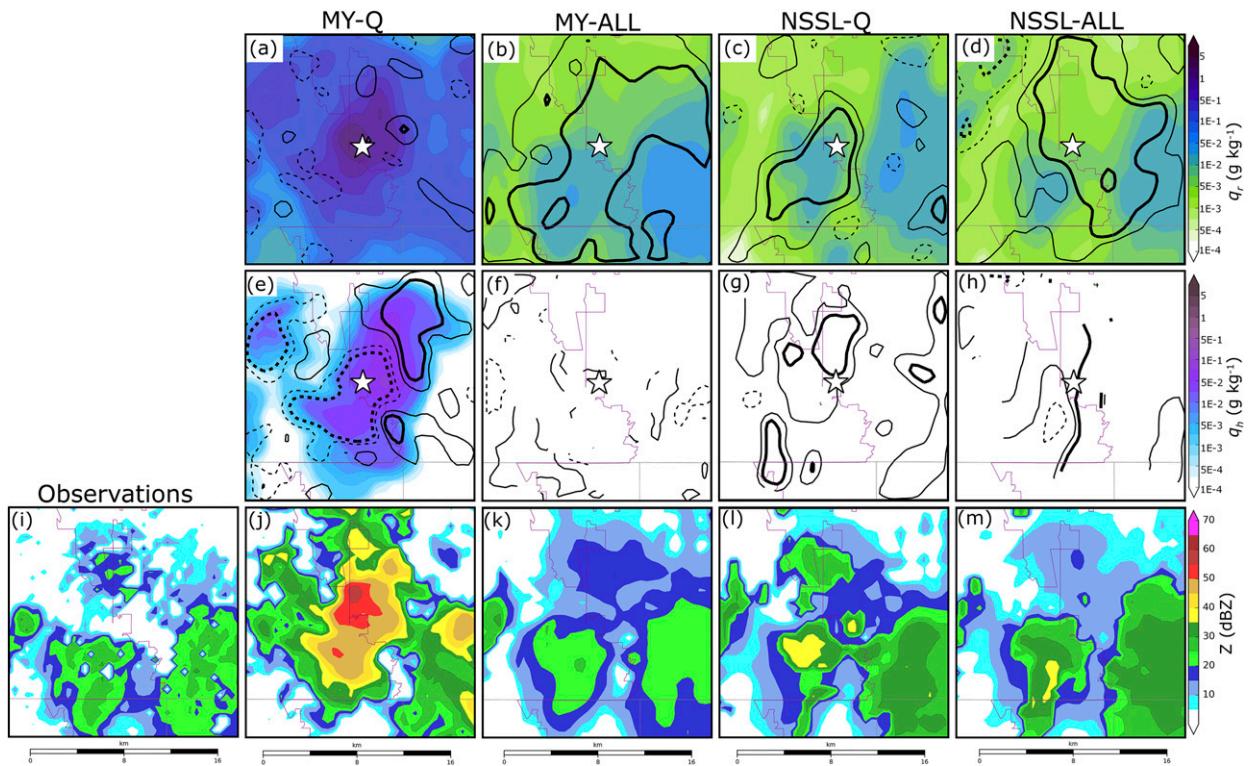


FIG. 6. Forecast error correlations for ensembles (a),(e) MY-Q, (b),(f) MY-ALL, (c),(g) NSSL-Q, and (d),(h) NSSL-ALL. The location of the subdomain is shown in Fig. 4. Correlations are calculated between the (a)–(d) mass of rain q_r , or (e)–(h) mass of hail q_h and an assumed Z observation (white star) interpolated to the lowest tilt of (0.5°) KFTG prior to the final assimilation cycle at 2040 UTC. Positive correlations (solid lines) and negative correlations (dashed lines) are plotted in increments of 0.15 between -0.3 and 0.3 . Stronger correlations (0.3 and -0.3) are contoured with a thick black line, weaker correlations (0.15 , -0.15) are contoured with a thin black line. To reduce noise, the correlation field is smoothed using a 9-point filter. Color shading represents ensemble forecast mean (a)–(d) q_r , or (e)–(h) q_h . Observed (i) Z and ensemble analysis mean Z for ensembles (j) MY-Q, (k) MY-ALL, (l) NSSL-Q, and (m) NSSL-ALL are provided at the same time and location.

observation prior suggests the EnKF will adjust both variables in the same (opposite) direction. For instance, if the Z observation prior and vertical velocity are positively correlated the EnKF will strengthen the storm updraft (i.e., intensify the storm) when the Z prior is adjusted upward toward the observed Z value by the filter. Figure 1 of Snyder and Zhang (2003) illustrates this idea for a radial velocity observation and vertical velocity.

Examples of an ensemble correlation analysis are provided in Figs. 6a–h. The ensemble correlation analysis is performed over a subdomain containing spuriously strong Z values in MY-Q, highlighted in Fig. 4. The assumed Z observation prior (sampled at the location of the white star) and hydrometeor mixing ratios are interpolated to the lowest radar tilt of KFTG (0.5°) from the closest model grid points both above and below the radar beam prior to calculating correlation; this allows for comparison between simulated (Figs. 6j–m) and observed Z (Fig. 6i). In previous OSSEs (e.g., Zhang

et al. 2004; Caya et al. 2005; TX05), analyzed state variable estimates are often poor during the first several DA cycles because the error covariance between model state variables and observation priors is unreliable. Ensemble covariances are typically examined when the state estimation and ensemble covariance become reasonably reliable in later cycles (e.g., TX05). In this study, correlation analyses are performed on the ensemble forecasts prior to the final DA at 2040 UTC when multivariate ensemble covariances can be reasonably good (Fig. 2).

MY-Q predicts high Z values where no organized convection is observed (Fig. 4a); such spurious precipitation develops within the model during the DA cycles near 2030 UTC and is not effectively suppressed by the filter in this case. A correlation analysis is conducted to understand why the MY-Q EnKF intensifies Z in a spurious storm to the east of Denver. Although MY-Q analyses predict the storm to produce large Z values, the updraft is weak and disorganized, which remains

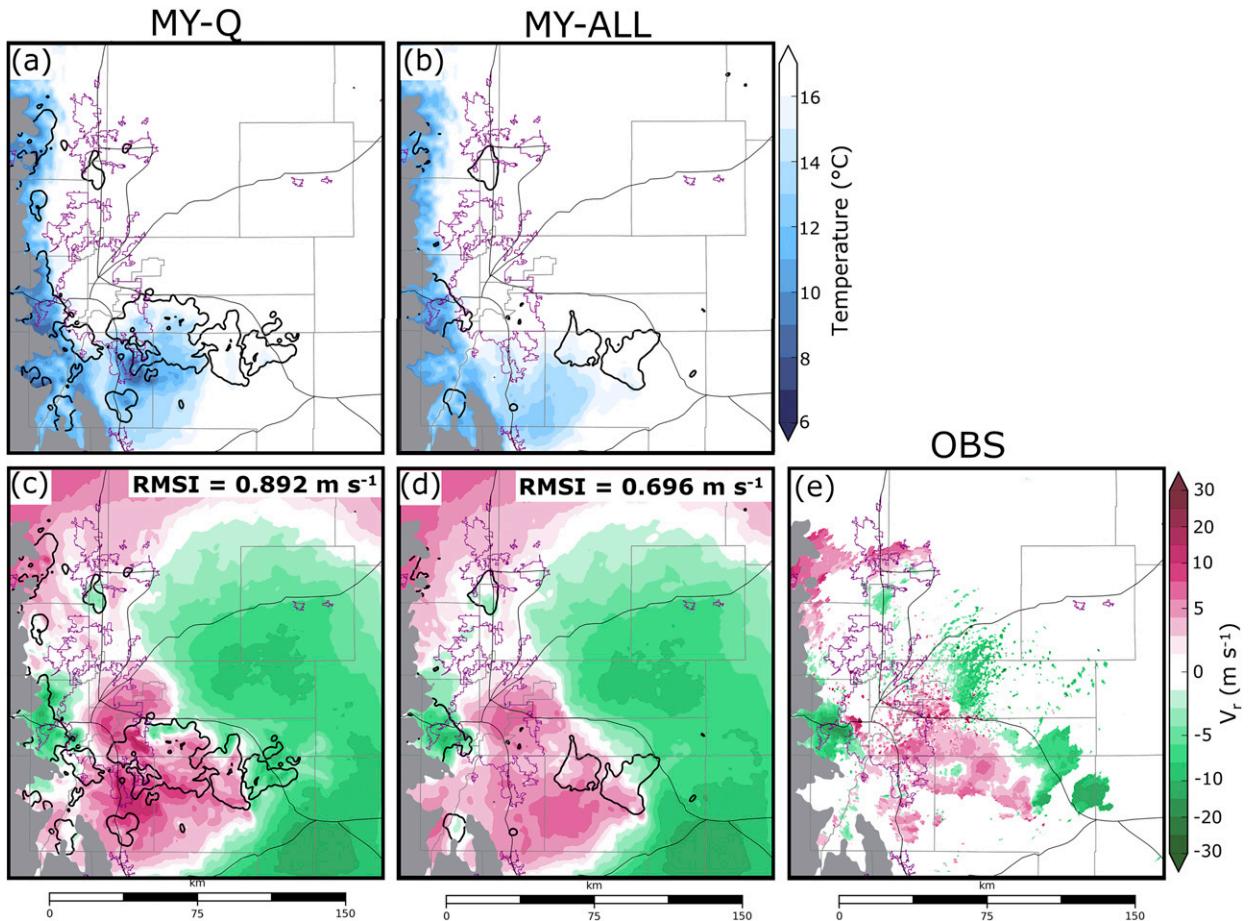


FIG. 7. Ensemble mean forecast (a),(b) temperature at the lowest model grid level above the surface and (c),(d) radial velocity at the lowest KFTG grid tilt for (a),(c) MY-Q and (b),(d) MY-ALL prior to the final assimilation cycle at 2040 UTC. (e) Observed KFTG radial velocity at the same time is provided. Thick black contours represent the 30 dBZ ensemble forecast mean Z smoothed using a 9-point filter. Background maps are the same as Fig. 1.

unchanged during DA. In the vicinity of the spurious storm, Z and q_r in MY-Q are relatively uncorrelated (Fig. 6a), consequently the EnKF is unable to decrease q_r in the spurious storm when assimilating Z observations. In the same location MY-Q Z and q_h are negatively correlated (Fig. 6e), this relationship intensifies reflectivity in the spurious storm during DA (Fig. 6j) because the EnKF increases q_h when decreasing spuriously large model diagnosed Z . While experiment MY-Q exhibits the weakest correlations between Z and q_r (Fig. 6a), the other ensembles exhibit mostly strong positive correlations throughout much of the subdomain (>0.3) (Figs. 6b–d). Differences in the forecast error covariance structure between MY-Q (Fig. 6a) and MY-ALL (Fig. 6b) suggest the importance of correctly updating all model state variables. MY-ALL, NSSL-Q, and NSSL-ALL analyses predict weaker Z values that better fit observations in the subdomain (Figs. 6k–m). This is in part because the background better fits observations

(i.e., no spurious storm prior to assimilation) and the background error covariance produces a more optimal analysis.

Updates to microphysical state variables, such as q_r , also impact thermodynamic state variables. For example, enhanced evaporational cooling attributed to an increase in precipitation during data assimilation (Figs. 5a,b) may cause MY-Q to predict more intense cold pools (Figs. 7a,b) and stronger (Figs. 7c,d) winds than MY-ALL. Winds modified by spurious storms also increase the RMSI for MY-Q V_r (Fig. 7c) more than for MY-ALL V_r (Fig. 7d).

b. Evaluating surface hail size

For an analysis to skillfully estimate surface hail size, microphysical variables must be properly updated so that the analyzed hail PSD produces reasonable hail size estimates. Box-and-whisker plots (Fig. 8) compare the ensemble member analysis estimated areal surface

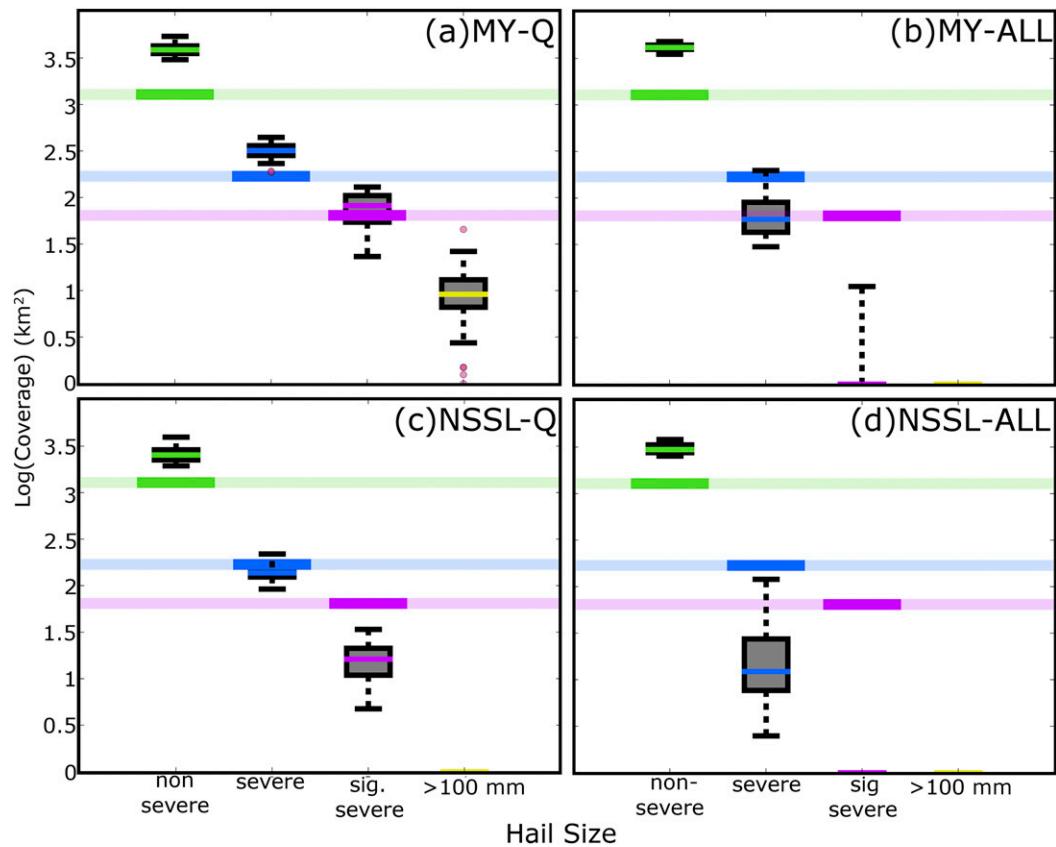


FIG. 8. Analysis estimated coverage of nonsevere (green), severe (blue), and significant severe (purple) hail according to D_{\max} between 2000 and 2040 UTC for ensembles (a) MY-Q, (b) MY-ALL, (c) NSSL-Q, and (d) NSSL-ALL. Hail coverage is only considered within the domain shown in Fig. 1 and where land elevation is less than 2.5 km above mean sea level. Actual coverage of nonsevere, severe, and significant severe hail based upon HCA output (Fig. 1c) is marked with a thick horizontal line of the corresponding color. Although the largest observed hailstone on this day was approximately 70 mm, the coverage of $D_{\max} > 100$ mm (yellow) is included to identify hail size overestimation biases.

coverage of nonsevere, severe, and significant severe hail to HCA output (Fig. 1c). Additional subjective comparisons are performed by comparing the probability-matched mean (Ebert 2001) of the ensemble analysis D_{\max} (Fig. 9) to HCA output (Fig. 1c). Unlike a simple ensemble mean, which typically smooths out extreme values (i.e., large hail sizes), the probability-matched mean maintains the frequency distribution of the ensemble to preserve extreme values.

All ensembles overestimate the spatial coverage of nonsevere hail (Fig. 8); in addition, MY-Q overestimates the coverage of severe and significant severe hail (Fig. 8a). In the probability-matched mean of the ensemble analyses, D_{\max} in MY-Q (Fig. 9a) predicts the largest hail ($D_{\max} > 200$ mm) in spurious convection located to the south and east of Denver. HCA output is unable to identify the maximum hail size because all hail greater than 50 mm in diameter is classified as “significant severe”; however, the largest hail report

recorded by the SPC for this event is 70 mm, suggesting that MY-Q substantially overestimates surface hail size. Further, almost all MY-Q estimated severe and significant severe hail coverage occurs in spurious convection away from where HCA output indicates large hail to occur (Fig. 1c). Although NSSL-Q underestimates the coverage of significant severe hail (Fig. 8c), the analyses (Fig. 9c) exhibit some qualitative skill and much of the largest hail occurs in regions where significant severe hail is detected in HCA output (Fig. 1c). MY-ALL (Fig. 8b) and NSSL-ALL (Fig. 8d) underestimate the spatial coverage of severe and significant severe hail, and their probability-matched means (Figs. 9b,d) suggest the analyses rarely indicate hail exceeding 25 mm in diameter.

Hail size changes as the DA system modifies q_h and N_{th} . The MY-Q filter, which does not modify N_{th} , frequently estimates large hailstones (Fig. 9a), primarily because the EnKF makes large adjustments to q_h in

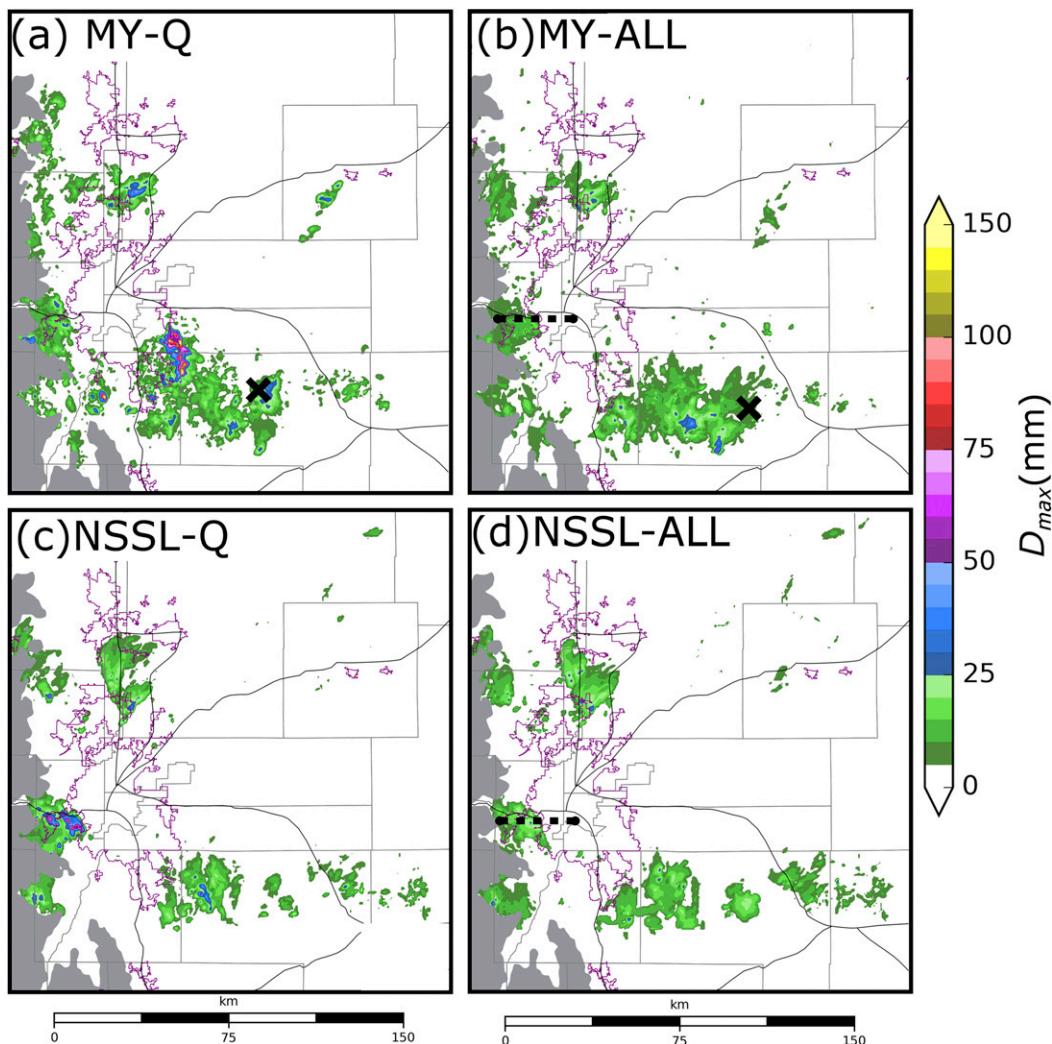


FIG. 9. The probability match mean of D_{max} for ensemble analyses (a) MY-Q, (b) MY-ALL, (c) NSSL-Q, and (d) NSSL-ALL between 2000 and 2040 UTC. A black “x” in (a) and (b) marks the location where hail PSDs are sampled in Fig. 10. Horizontal dashed lines in (b) and (d) mark the locations where vertical cross sections are taken in Figs. 11–15.

regions where relatively little hail is present. When large quantities of added ice mass are shared between a relatively small number of hailstones, the average hailstone diameter greatly increases. To demonstrate this behavior a PSD is sampled from a hailstorm southeast of Denver (“x” in Fig. 9a) using the MY-Q

ensemble mean both before and after assimilation at 2140 UTC. At the MY-Q sampled grid point q_h increases by $5.3 \times 10^{-2} \text{ g kg}^{-1}$ (Table 1) during DA. Because q_h increases but N_{th} remains constant, the slope of the analysis hail PSD becomes more shallow than the forecast (Fig. 10a) and D_{max} increases by

TABLE 1. MY-Q and MY-ALL forecast and analysis mean q_h , N_{th} , and D_{max} at 2040 UTC. The locations where the MY-Q and MY-ALL hail variables are sampled is marked with an “x” in Figs. 9a and 9b, respectively. Hail state variables diagnose the hail PSDs shown in Fig. 10.

	MY-Q forecast	MY-Q analysis	MY-ALL forecast	MY-ALL analysis
$q_h \text{ (g kg}^{-1}\text{)}$	1.64×10^{-2}	6.94×10^{-2}	6.02×10^{-2}	2.68×10^{-1}
$N_{th} \text{ (m}^{-3}\text{)}$	4.55	4.55	24.19	92.32
$D_{max} \text{ (mm)}$	17	28	17	19

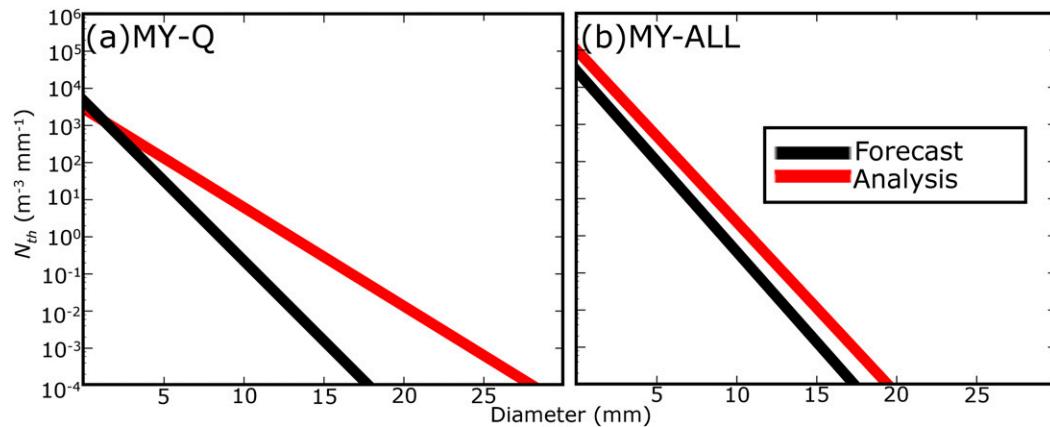


FIG. 10. Hail PSDs diagnosed from model level 15 (1.96 km above ground level) of (a) MY-Q and (b) MY-ALL ensemble mean forecasts and analyses at 2040 UTC. The grid point where the MY-Q and MY-ALL hail PSDs are diagnosed is marked with an “ \times ” in Figs. 9a and 9b, respectively. The hail variables used to diagnose the PSDs are provided in Table 1. PSDs were selected to highlight examples of how the EnKF updates a hail PSD during assimilation.

approximately 11 mm (Table 1). In extreme instances DA causes D_{\max} to exceed 200 mm in MY-Q (Fig. 9a). NSSL-Q estimates hail sizes (Fig. 9c) that more closely resemble HCA output (Fig. 1c). A more reliable multivariate covariance, due in part to fewer spurious storms in the NSSL-Q background forecasts compared to the MY-Q background forecasts (not pictured), prevents large updates to q_h and limits the most extreme hail sizes.

Hail PSDs sampled from the MY-ALL ensemble forecast and analysis mean at 2140 UTC in a hailstorm to southeast of Denver (“ \times ” in Fig. 9b) demonstrate how the EnKF adjusts hail size through changes to both q_h and N_{th} (Fig. 10b). During DA, the MY-ALL EnKF increases q_h by approximately $2.08 \times 10^{-1} \text{ g kg}^{-1}$ and nearly triples N_{th} (Table 1). Although q_h in MY-ALL increases more than in MY-Q (Table 1), D_{\max} remains relatively unchanged because of the increase in N_{th} . The slope of the MY-ALL analysis hail PSD is similar to that of the forecast PSD (Fig. 10b) and subsequently D_{\max} increases by only 2 mm (Table 1). For this analyzed grid point (Fig. 10b) D_{\max} is classified as nonsevere, despite the HCA output (Fig. 1c) detecting significant severe hail in the surrounding region. Microphysical variables, such as N_{th} , need additional constraint to accurately estimate hail size; previous work by XJZ10 suggests assimilating additional data improves initial condition estimates by further constraining microphysical variables.

Although MY-ALL and NSSL-ALL analyses estimate microphysical properties, such as Z with a moderate to high level of skill (Figs. 3a,b), there are an insufficient number of observations to constrain all microphysical variables associated with multimoment

microphysics schemes and, as a result, the performance of hail size forecast is limited. To the west of Denver where the HCA detects significant severe hail (Fig. 1c), MY-ALL (Fig. 9b) and NSSL-ALL (Fig. 9d) analyses underestimate hail size in part because the EnKF modifies moments of the hail PSD so that the slope of the PSD remains relatively constant (e.g., Fig. 10b). This causes hail mass to be split among more hailstones, limiting the ability to increase the mean hailstone diameter.

c. Ensemble correlation analysis

Differences in the microphysical parameterizations cause hydrometeor properties (e.g., size, density, fall speed) to differ substantially between MP schemes. For example, Johnson et al. (2016) note that hail behavior differs between the MY and NSSL schemes primarily due to differences in hail production processes. Typically, the MY scheme produces many small hail particles aloft because the MY hail category is primarily composed of small, frozen raindrops that have been converted into hail particles (Johnson et al. 2016). The NSSL scheme produces fewer and larger hailstones because hail is produced from dense graupel that has undergone wet growth (Mansell et al. 2010). Simulated Z is intrinsically related to such microphysical assumptions (e.g., Jung et al. 2008b,a, 2012; Dawson et al. 2014; Putnam et al. 2014, 2017b; Johnson et al. 2016, 2018); these assumptions also determine the forecast error covariance (TX05; XJZ10).

The ensemble correlation between observation prior Z and model state variables (i.e., w , θ , q_h , N_{th} , q_r , N_{tr}) illustrates how different hail treatments in MP schemes impact the analysis increments. Ensemble correlation

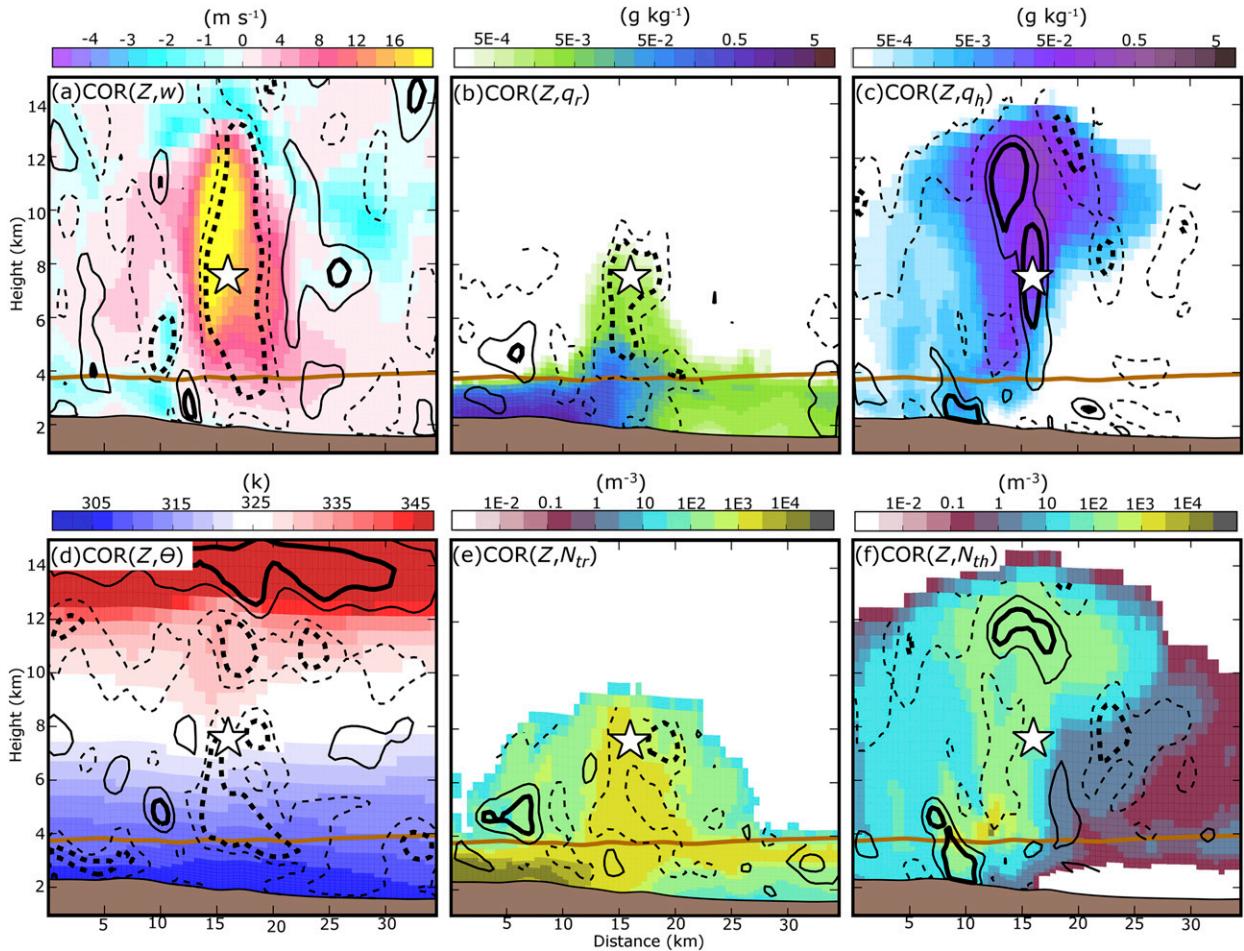


FIG. 11. Vertical cross sections of MY-ALL ensemble forecast mean (a) w , (b) q_r , (c) q_h , (d) θ , (e) N_{tr} , and (f) N_{th} prior to the final assimilation cycle at 2040 UTC. Cross sections are taken through the Denver hailstorm, denoted as east–west line in Fig. 9b. Forecast error correlations between an assumed Z observation taken from within the hail growth zone (white star) and state variables are plotted. The same plotting convention is used to contour correlation as in Fig. 6. The horizontal brown line represents the 0°C isotherm.

analyses are performed on the hailstorm located to the west of Denver over Interstate-70 (Fig. 1) prior to the final DA cycle at 2040 UTC. Vertical cross sections are taken through the updraft of the hailstorm (Figs. 9b,d) and coincide with where most hail growth processes occur. The correlation analysis is performed in the vertical to capture the variation of hail particle behavior both above and below the 0°C isotherm.

When an assumed observation is taken from the ice-phase dominant region of the MY-ALL updraft, w and Z become negatively correlated (Fig. 11a). This correlation pattern was first observed in XJZ10 and was attributed to microphysical assumptions made in the MY scheme. Stronger updrafts loft more rain and cloud water aloft, leading to a positive correlation between w and rainwater throughout much of the updraft above the 0°C isotherm (i.e., supercooled water) (Figs. 12b,d).

Supercooled raindrops are converted into hail within the MY scheme, hence updraft intensity is also positively correlated with N_{th} (Fig. 12e) throughout much of the storm. Because raindrops are relatively small, the mean mass diameter of hail decreases in strong updrafts (e.g., Johnson et al. 2016; XJZ10), and thus Z is negatively correlated with w in MY-ALL forecasts. Weaker updrafts advect less supercooled liquid above the 0°C isotherm and lower the production rate of small hailstones; additionally larger hailstones aloft fall through the weaker updrafts to enhance Z .

MY-ALL correlations determine how model state variables are modified during reflectivity assimilation. Similar to XJZ10, Z at the observation location is negatively correlated with θ in the updraft region (Fig. 11d) because w and θ are positively correlated in the updraft

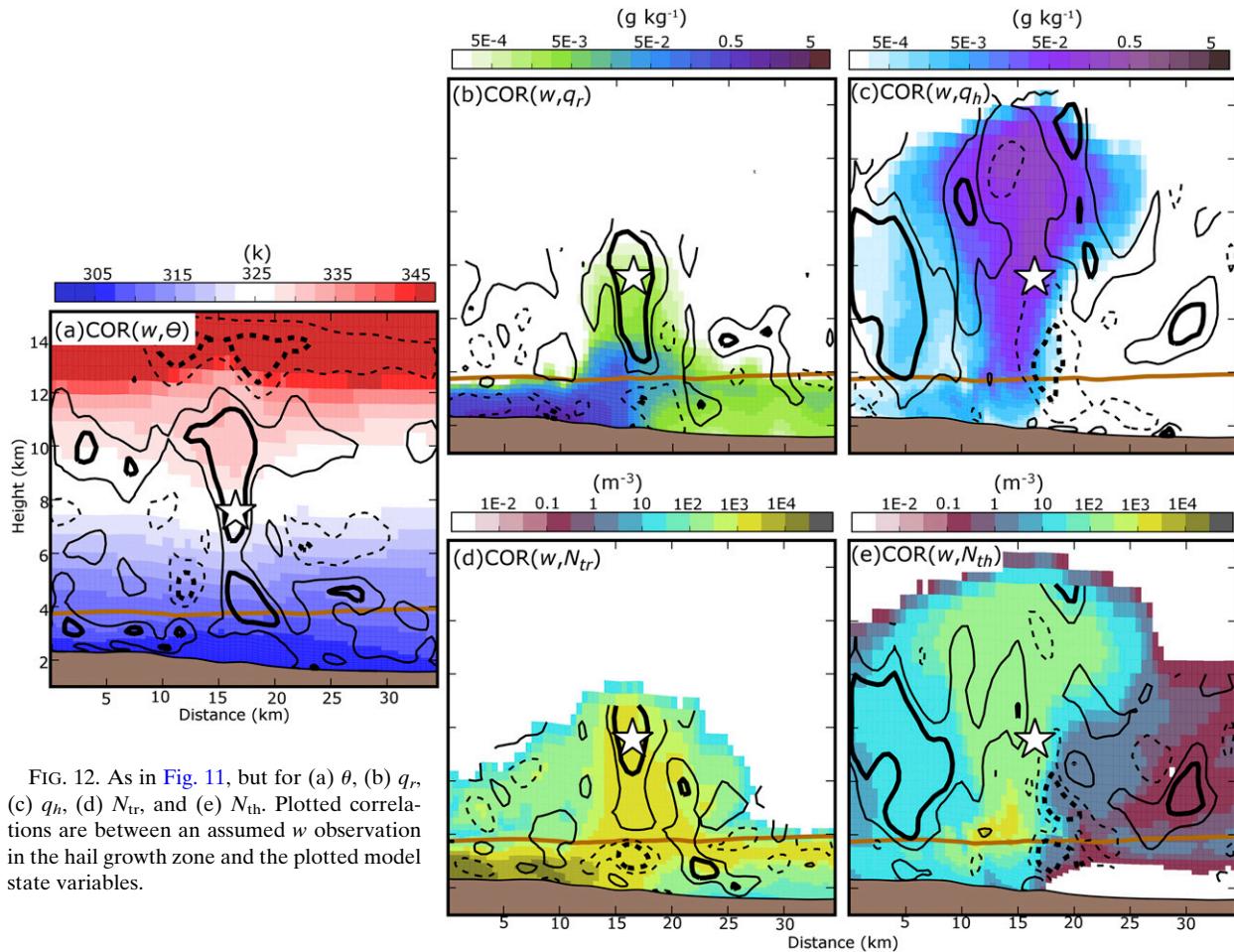


FIG. 12. As in Fig. 11, but for (a) θ , (b) q_r , (c) q_h , (d) N_{tr} , and (e) N_{th} . Plotted correlations are between an assumed w observation in the hail growth zone and the plotted model state variables.

region (Fig. 12a). Hydrometeor state variables: q_r (Fig. 11b), N_{tr} (Fig. 11e), and N_{th} (Fig. 11f) are negatively correlated with Z between 4 and 8 km above MSL; this is where frozen raindrops are converted to hail. Unlike with N_{th} , Z exhibits mostly positive correlations with q_h (Fig. 11c) above the 0°C isotherm, suggesting the complex relationship between Z and the hail size spectrum.

Unlike MY-ALL, Z at the observation location is positively correlated with w (Fig. 13a) throughout much of the NSSL-ALL storm updraft; however, similar to MY-ALL, w is positively correlated with both rain and hail state variables (Figs. 14b–e). Strong updrafts in the NSSL-ALL storm loft liquid water above the 0°C isotherm. Instead of being converted into small hail particles, the supercooled liquid is accreted by graupel and hail, and causes rimed ice particles to increase in size (Labriola et al. 2019a). It is also noted that Z and θ are positively correlated throughout the updraft in the NSSL-ALL storm (Fig. 13d) due to the same processes in MY-ALL.

Prognostic graupel and hail volume (v_g and v_h , respectively) equations allow the NSSL scheme to vary rimed ice density during the forecast and modify both particle sedimentation and hail production processes. In the storm updraft region, Z and v_h are positively correlated (Fig. 15a). Hail accreting more liquid water in the stronger updraft increases hail volume and Z in the updraft (Figs. 15a,b). It is noted that hail density is a nonunique solution, and that an infinite combination of q_h , v_h pairs can produce the same density value; correlations are therefore not indicative of modifications to hail density.

While this study primarily evaluates the ensemble background error correlation fields within strong thunderstorms at 2040 UTC, it is important to note the background error covariance evolves with time through the DA cycles. Correlations between observation prior Z and w taken from within the updraft of a storm to the east of Denver are evaluated throughout the DA window (Fig. 16). This storm is selected because it is present in observations throughout the DA window. Prior to the

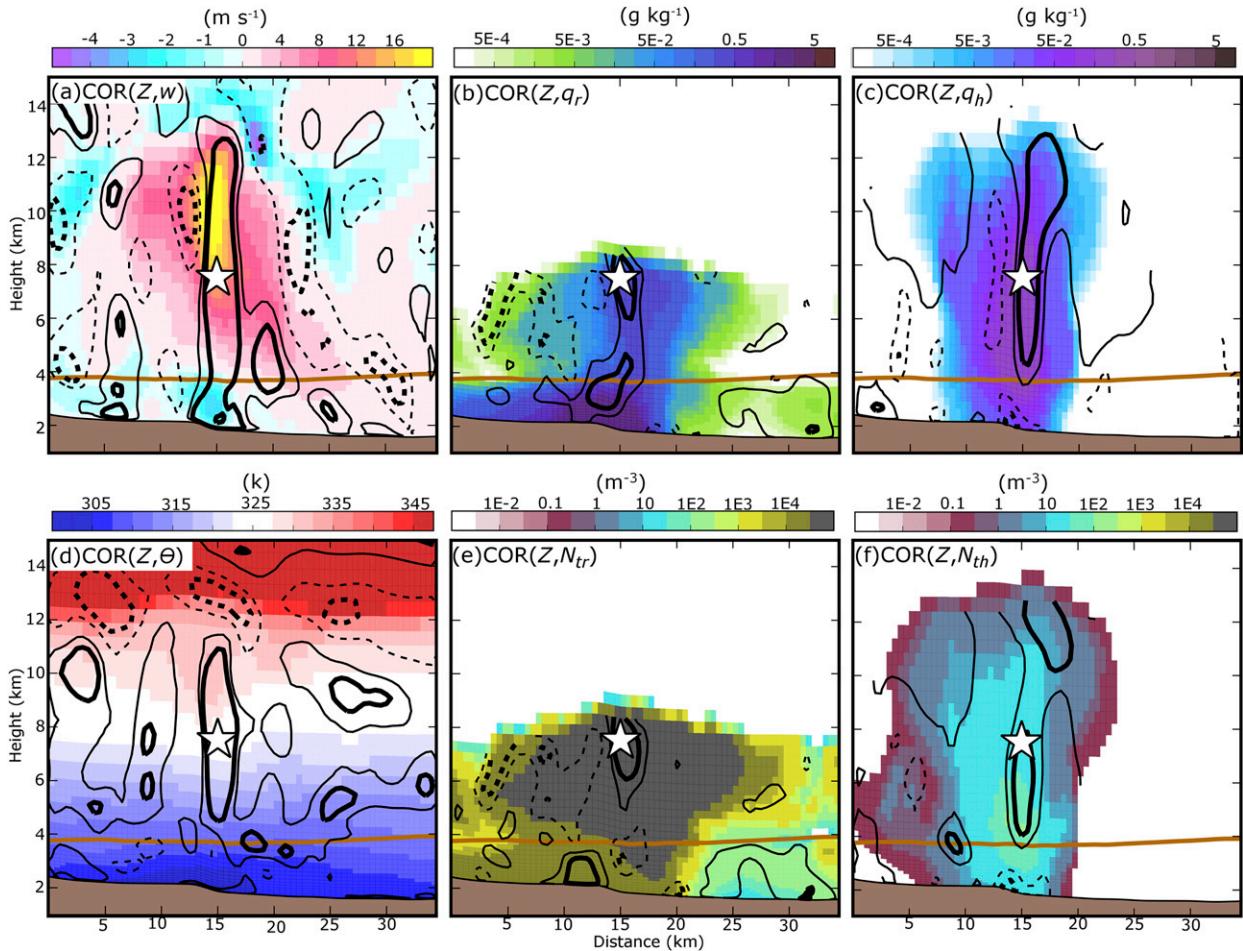


FIG. 13. As in Fig. 11, but for NSSL-ALL (a) w , (b) q_r , (c) q_h , (d) θ , (e) N_{tr} , and (f) N_{th} . Plotted correlations are between an assumed Z observation in the hail growth zone and the plotted model state variables. Cross sections are taken through the Denver hailstorm, denoted as east–west line in Fig. 9d.

first DA cycle both MY-ALL and NSSL-ALL are not confident in the location of organized convection, which causes correlations between w and Z to be noisy (Figs. 17a,d). Both ensembles predict storm updrafts to be weak ($w < 15 \text{ m s}^{-1}$) at 2020 UTC (Figs. 17b,e). Because of that, the MY scheme predicts the storm to produce little hail. Instead, the weak updrafts loft precipitation above the 0°C isotherm and enhance Z ; w and Z become negatively correlated with time as the storm intensifies and strong updrafts increase the number of small hail stones aloft. These correlation fields are similar to those observed in the Denver hail storm (Fig. 11a). Unlike MY-ALL, w and Z are positively correlated in NSSL-ALL between 2020 and 2040 UTC (Figs. 17e,f) because the updrafts advect more precipitation above the 0°C isotherm and enhance the production of large hailstones. The evolving correlation fields demonstrate the sensitivity of the background error covariance to the microphysical assumptions made

within MP schemes as well as other factors including the quality of the state estimation. In this paper, we focus more on the later time (2040 UTC) when the state estimation becomes better.

Correlations between model state variables (i.e., w and θ) and Z have important consequences on the representation of analyzed storms. The correlation patterns observed in this study were also observed in other strong storms at 2040 UTC (e.g., Figs. 17c,f) and during earlier DA cycles (not shown). When MY-ALL underestimates (overestimates) Z , the filter will decrease (increase) w (Fig. 11a) and make the updraft air temperature cooler (warmer) (Fig. 11d). In effect the EnKF in MY-ALL weakens the analyzed storm. Both w (Fig. 13a) and θ (Fig. 13d) are positively correlated with Z in NSSL-ALL and cause the filter to intensify the storm under similar circumstances. Z is generally a function of the diameter of a hailstone to the sixth power although it becomes more complex for a large or wet hailstone due to the Mie

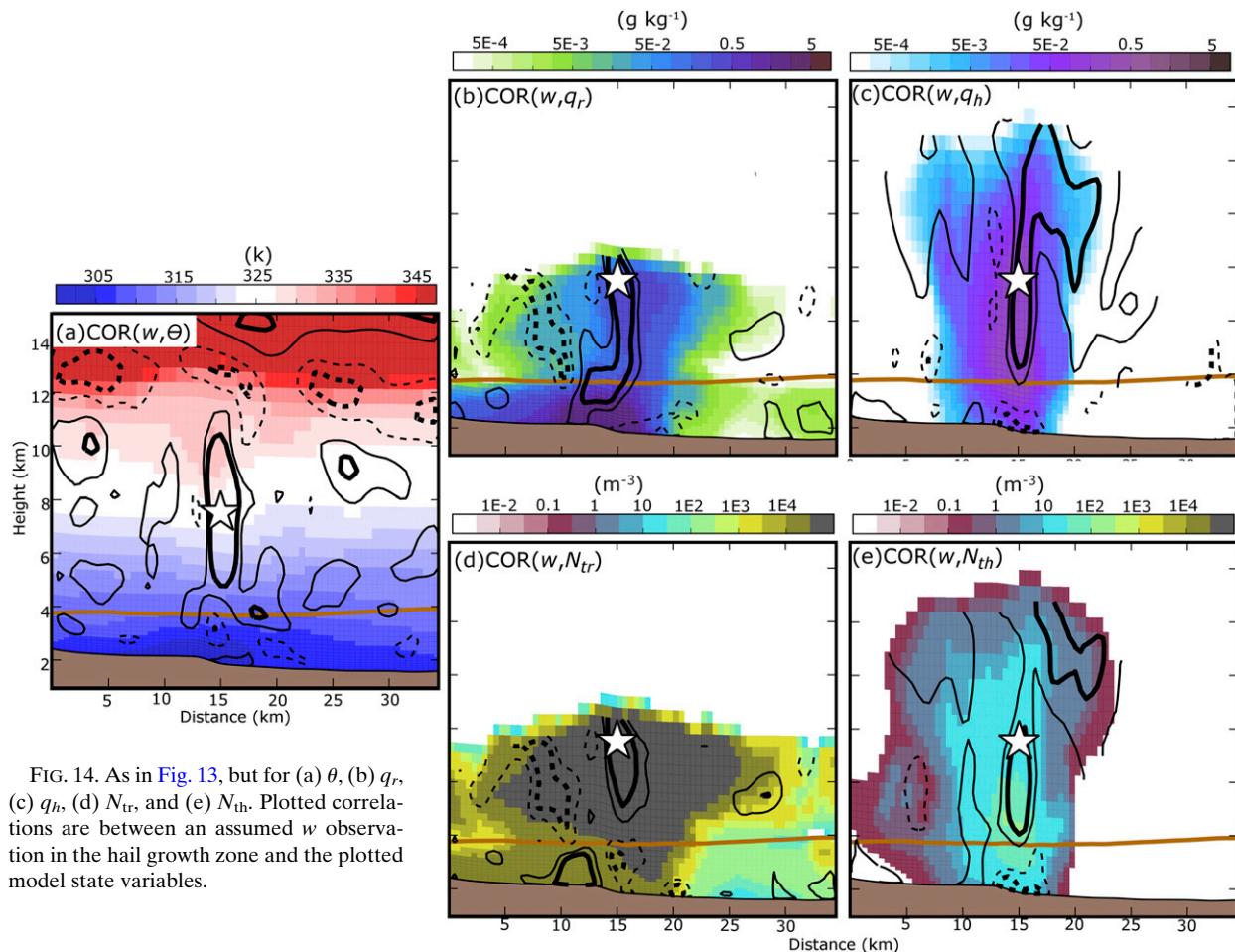


FIG. 14. As in Fig. 13, but for (a) θ , (b) q_r , (c) q_h , (d) N_{tr} , and (e) N_{th} . Plotted correlations are between an assumed w observation in the hail growth zone and the plotted model state variables.

scattering effect. In addition, numerous observational studies (e.g., Heymfield 1983; Nelson 1983; Ziegler et al. 1983; Foote 1984; Dennis and Kumjian 2017) have suggested that hail size is not a monotonic function of updraft strength, but is also influenced by vertical wind shear, environmental moisture, and updraft volume. More observations are needed to determine which microphysics scheme produces an analysis increment that more closely reflects reality.

4. Summary and further discussion

Newer multimoment bulk microphysics schemes (MP schemes) are increasing in complexity and predicting more state variables in order to improve the representation of microphysical processes (e.g., riming, sedimentation). As the number of degrees of freedom within a MP scheme increases, initial state estimation using an EnKF becomes more challenging, in part because the large number of microphysical state variables predicted by a MP scheme are insufficiently

constrained by the limited number observations that can infer microphysical properties [e.g., radar reflectivity (Z)].

In this study a cycled EnKF framework is used to update the microphysical properties of hail-producing storms for the 8 May 2017 Colorado severe hail event. Four ensemble forecast experiments are conducted using either the Milbrandt and Yau (2005a) double-moment (MY) scheme or the NSSL double-moment variable density rimed ice scheme (Mansell et al. 2010). An EnKF is used to update either only hydrometeor mixing ratios (MY-Q, NSSL-Q) or all microphysical state variables (MY-ALL, NSSL-ALL); in addition to dynamic and thermodynamic information. While most previous studies use observed Z to update all microphysical variables, this study examines whether updating a limited number of particle size distribution moments (i.e., mass mixing ratio) provides sufficient constraint for an ensemble system run using a multimoment scheme. The model forecasts and analyses are evaluated against observed Z in addition to output from a hydrometeor

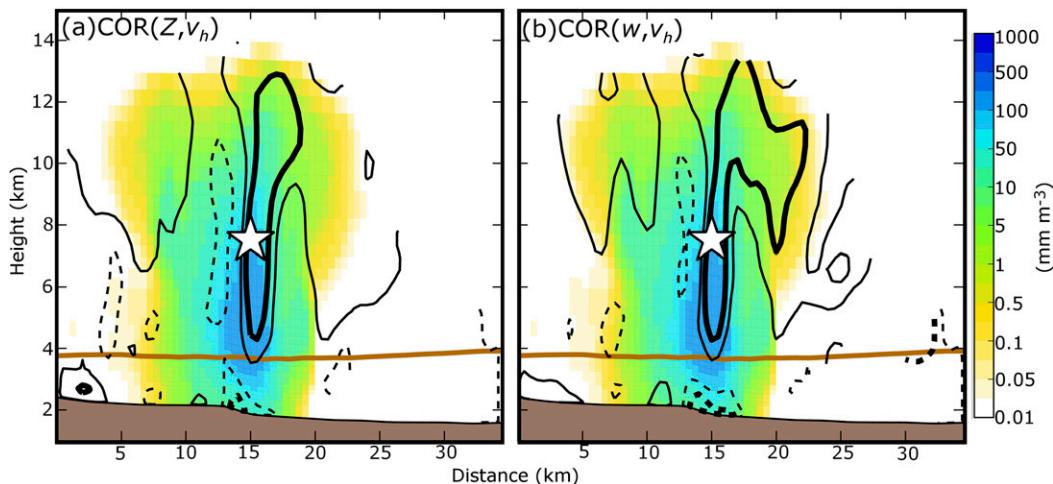


FIG. 15. As in Fig. 13, but for v_h . Plotted correlations are between v_h and an assumed (a) Z observation or (b) w observation in the hail growth zone.

classification algorithm (HCA) to determine which configuration produces the more realistic state variable estimates related to hail.

For ensembles that update only hydrometeor mixing ratio, in particular MY-Q, the forecast error covariance is often unreliable and limits the accuracy of state variable estimates. We suspect that the forecast error covariance is unreliable because updating a limited number of microphysical variables introduces large imbalances into the ensemble prediction system during assimilation. For example, negative correlations between Z and hydrometeor mixing ratios in experiment MY-Q cause spurious radar echoes to be enhanced during assimilation, even if clear-air reflectivity data are assimilated. Further, due to the poor multivariate error covariance structure,

the EnKF in MY-Q was unable to replicate Z intensity or structure within mature storms. Generally, the forecast error covariance of MY-ALL and the NSSL ensembles (NSSL-Q and NSSL-ALL) is more reliable than that of MY-Q, allowing the ensemble analyses to estimate Z and radial velocity (V_r) with less error.

Comparison (and verification) of analyzed surface hail size against HCA output provides insight into how the different EnKF configurations update hail particle size distributions. Generally, ensembles that update only mixing ratio (MY-Q, NSSL-Q) estimate large surface hail sizes because large quantities of hail mass can be distributed among a relatively small number of hailstones. Ensembles that update all microphysical

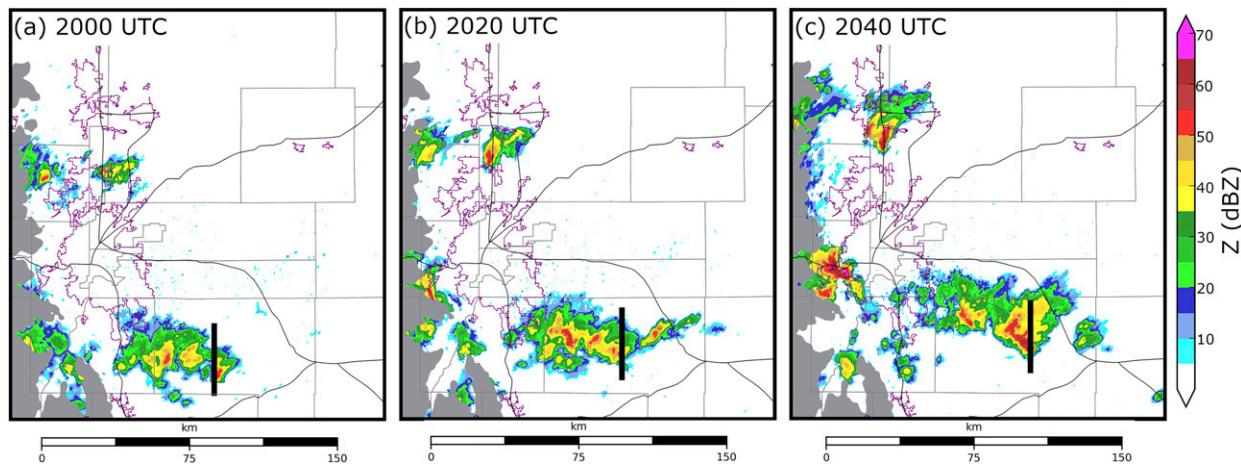


FIG. 16. KFTG observed Z at the lowest radar tilt (0.5°). Thick black lines denote the approximate location of where vertical cross sections are taken in Fig. 17.

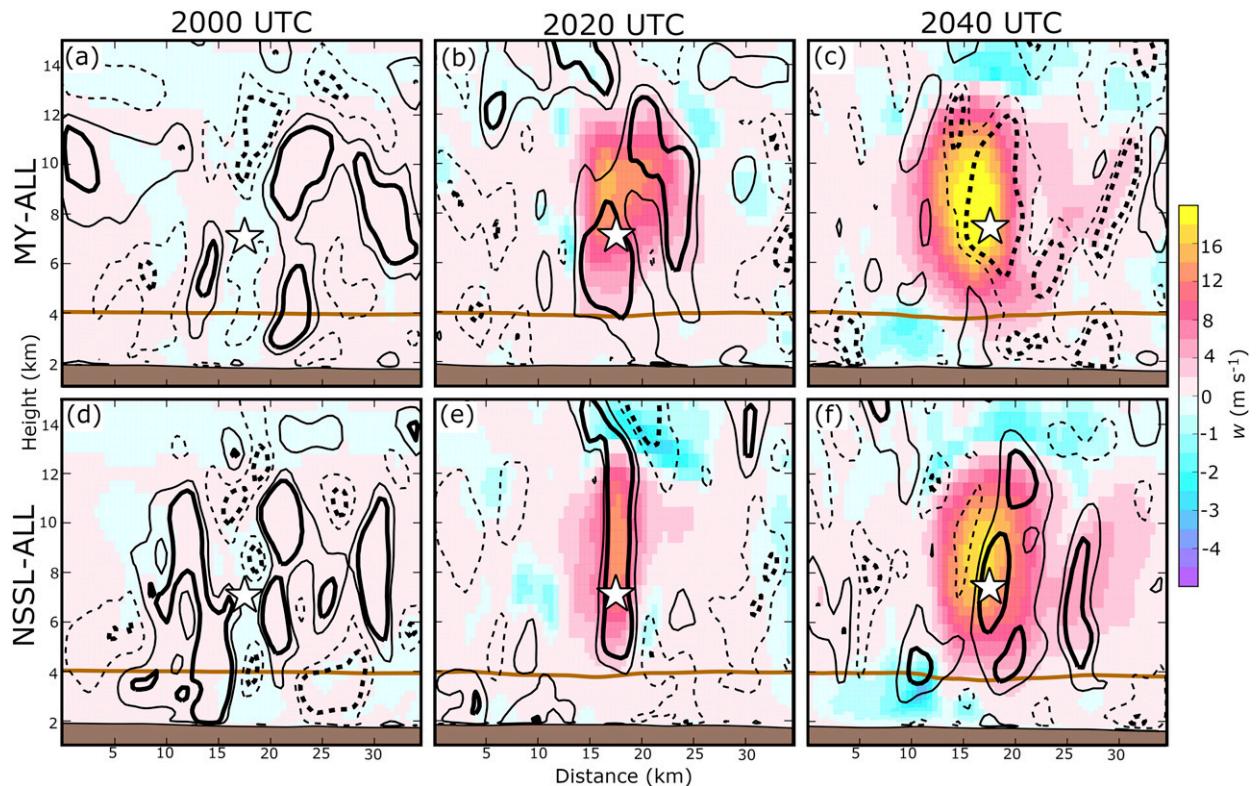


FIG. 17. Vertical cross section of the (a)–(c) MY-ALL and (d)–(f) NSSL-ALL ensemble mean w at (a),(d) 2000, (b),(e) 2020, and (c),(f) 2040 UTC. Cross sections are taken from within a storm to the east of Denver. Forecast error correlations between an assumed Z observation taken from within the hail growth zone (white star) and w are plotted. The same plotting convention is used to contour correlation as in Fig. 6. The horizontal brown line represents the 0°C isotherm.

variables (MY-ALL and NSSL-ALL) tend to underestimate surface hail size because these ensembles typically predict larger hail number concentrations and cause hail mass to be split between many small hailstones. Results suggest the need to assimilate additional observations of microphysical relevance in order to better constrain the increased number of state variables (e.g., number concentration).

Hail production and growth assumptions made by MP schemes substantially influence the forecast error covariance. The MY scheme generates hail primarily from small frozen raindrops; this process increases the number of small hailstones above the 0°C isotherm and causes Z to be negatively correlated with air temperature and updraft strength. Due to differences in hail production processes, the opposite correlation patterns are observed in the NSSL scheme forecasts for intense hailstorms. Hail growth assumptions also influence correlation patterns for hydrometeor state variables. For this study the NSSL-ALL experiment favors more positive correlations between Z and hail variables (mass and number) than the MY-ALL experiment. While this study provides insight into the complexities of updating microphysical variables via an

EnKF, it is noted that only a limited number of the available schemes are evaluated. A large number of multimoment schemes (e.g., Thompson et al. 2008; Morrison et al. 2005, 2009; Lim and Hong 2010; Morrison and Milbrandt 2015; Morrison et al. 2015; Milbrandt and Morrison 2016) are used in weather prediction systems; however, relatively few studies have analyzed how underlying microphysical assumptions made within these schemes impact multivariate ensemble background error covariances and state variable updates within ensemble DA. We have also shown that the multivariate ensemble covariances can be sensitive to the quality of storm analysis and possibly also the storm intensity and morphology.

Due to the many possible combinations of variables that are able to produce a given Z value, microphysical state variables will remain insufficiently constrained by observed reflectivity. The assimilation of polarimetric observations is shown to provide additional constraint in observation system simulation experiments (OSSEs) conducted by Jung et al. (2008b, 2010b) and a real case study conducted by Putnam et al. (2019); however, assimilating polarimetric variables remains nontrivial. Although MP schemes can replicate basic polarimetric

signatures (e.g., Johnson et al. 2016, 2019; Putnam et al. 2017a,b, 2019), the NWP output from these schemes is often biased in intensity and coverage. To benefit from the assimilation of polarimetric observations, more effort is needed to improve the representation of microphysical processes that generate polarimetric signatures and to find the optimal configurations that can maximize the impact of polarimetric data. Improving the representation of microphysical processes has the potential to not only improve EnKF estimates, but also mitigate misrepresentations of subgrid-scale processes and reduce model errors.

Acknowledgments. This work was primarily supported by NSF Grant AGS-1261776 as part of the Severe Hail, Analysis, Representation, and Prediction (SHARP) project. Supplemental support was provided by NOAA Grant NA16OAR4590239. The third and fourth authors are also supported by the NOAA Warn-on-Forecast (WoF) and Spectrum Efficient National Surveillance Radar (SENSR) Grant NA16OAR4320115. The SHARP team includes David Gagne, Amy McGovern, Amanda Burke, in addition to the coauthors. Computing was performed primarily on the XSEDE Stampede2 supercomputer at the University of Texas Advanced Computing Center (TACC). The authors thank Kevin Thomas for support in obtaining data along with Marcus Johnson who provided useful feedback on the project.

REFERENCES

- Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophys. Res. Lett.*, **33**, L12801, <https://doi.org/10.1029/2006GL026186>.
- , D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, <https://doi.org/10.1175/2008MWR2691.1>.
- Allen, J. T., and M. K. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteor.*, **10** (3), <https://ejssm.org/ojs/index.php/ejssm/article/viewArticle/149>.
- CAPS, 2017: CAPS spring forecast experiment program plan. Center for The Analysis and Prediction of Storms, accessed 14 July 2018, http://www.caps.ou.edu/~fkong/sub_atm/spring17.html.
- Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094, <https://doi.org/10.1175/MWR3021.1>.
- 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>.
- , and J. Burroughs, 2003: The tristate hailstorm: The most costly on record. *Mon. Wea. Rev.*, **131**, 1734–1739, <https://doi.org/10.1175/2549.1>.
- , D. Changnon, and S. D. Hilberg, 2009: Hailstorms across the nation: An atlas about hail and its damages. Illinois State Water Survey Contract Rep. 2009-12, 92 pp.
- 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>.
- Clark, A. J., J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012: Forecasting tornado pathlengths using a three-dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113, <https://doi.org/10.1175/WAF-D-11-00147.1>.
- 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](https://doi.org/10.1175/1520-0477(1993)074<0645:RAAUWD>2.0.CO;2).
- Dawson, D. T., II, 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>.
- Dawson, L. C., G. S. Romine, R. J. Trapp, and M. E. Baldwin, 2017: Verifying supercellular rotation in a convection-permitting ensemble forecasting system with radar-derived rotation track data. *Wea. Forecasting*, **32**, 781–795, <https://doi.org/10.1175/WAF-D-16-0121.1>.
- 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>.
- Doviak, R. J., and D. S. Zrnić, 1993: *Doppler Radar and Weather Observations*. 2nd ed. Academic Press, 562 pp.
- , V. Bringi, A. Ryzhkov, A. Zahrai, and D. Zrnić, 2000: Considerations for polarimetric upgrades to operational WSR-88D radars. *J. Atmos. Oceanic Technol.*, **17**, 257–278, [https://doi.org/10.1175/1520-0426\(2000\)017<0257:CFPUTO>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<0257:CFPUTO>2.0.CO;2).
- Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927, <https://doi.org/10.1175/2008JTECHA1156.1>.
- , 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](https://doi.org/10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2).
- Du, J., G. Dimego, B. Zhou, D. Jovic, B. Ferrier, and B. Yang, 2015: Regional ensemble forecast systems at NCEP. *27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., 2A.5, https://ams.confex.com/ams/27WAF23NWP/webprogram/Manuscript/Paper273421/NWP2015_NCEP_RegionalEnsembles_paper.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, [https://doi.org/10.1175/1520-0493\(2001\)129<2461:AOAPMS>2.0.CO;2](https://doi.org/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, [https://doi.org/10.1175/1520-0434\(1998\)013<0277:NCOHSW>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0277:NCOHSW>2.0.CO;2).
- Environmental Modeling Center, 2017: North American Mesoscale (NAM) analysis and forecast system characteristics. Environmental Modeling Center, 9 pp.
- 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:ADMMPP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<0249:ADMMPP>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](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, <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/>.
- Gagne, D. J., II, S. E. Haupt, D. W. Nychka, and G. Thompson, 2019: Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Wea. Rev.*, **147**, 2827–2845, <https://doi.org/10.1175/MWR-D-18-0316.1>.
- Gallo, B. T., A. J. Clark, B. T. Smith, R. L. Thompson, I. Jirak, and S. R. Dembek, 2019: Incorporating UH occurrence time to ensemble-derived tornado probabilities. *Wea. Forecasting*, **34**, 151–164, <https://doi.org/10.1175/WAF-D-18-0108.1>.
- 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>.
- Hawblitzel, D. P., F. Zhang, Z. Meng, and C. A. Davis, 2007: Probabilistic evaluation of the dynamics and predictability of the mesoscale convective vortex of 10–13 June 2003. *Mon. Wea. Rev.*, **135**, 1544–1563, <https://doi.org/10.1175/MWR3346.1>.
- Heinselman, P. L., and A. V. Ryzhkov, 2006: Validation of polarimetric hail detection. *Wea. Forecasting*, **21**, 839–850, <https://doi.org/10.1175/WAF956.1>.
- 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:CSOAH1>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<1482:CSOAH1>2.0.CO;2).
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, <https://doi.org/10.1175/MWR-2864.1>.
- Johnson, A., and X. Wang, 2017: Design and implementation of a GSI-based convection-allowing ensemble data assimilation and forecast system for the PECAN field experiment. Part I: Optimal configurations for nocturnal convection prediction using retrospective cases. *Wea. Forecasting*, **32**, 289–315, <https://doi.org/10.1175/WAF-D-16-0102.1>.
- , —, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, <https://doi.org/10.1175/MWR-D-14-00345.1>.
- Johnson, M., Y. Jung, D. 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>.
- , —, —, T. Supinie, M. Xue, J. Park, and Y.-H. Lee, 2018: Evaluation of unified model microphysics in high-resolution NWP simulations using polarimetric radar observations. *Adv. Atmos. Sci.*, **35**, 771–784, <https://doi.org/10.1007/s00376-017-7177-0>.
- , —, J. A. Milbrandt, H. Morrison, and M. Xue, 2019: Effects of the representation of rimed ice in bulk microphysics schemes on polarimetric signatures. *Mon. Wea. Rev.*, **147**, 3785–3810, <https://doi.org/10.1175/MWR-D-18-0398.1>.
- Jones, T., P. Skinner, N. Yussouf, K. Knopfmeier, A. Reinhart, and D. Dowell, 2019: Forecasting high-impact weather in land-falling tropical cyclones using a warn-on-forecast system. *Bull. Amer. Meteor. Soc.*, **100**, 1405–1417, <https://doi.org/10.1175/BAMS-D-18-0203.1>.
- , K. H. Knopfmeier, D. M. Wheatley, G. J. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental warn-on-forecast system. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, <https://doi.org/10.1175/WAF-D-15-0107.1>.
- 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 —, 2010a: 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 —, 2010b: Simultaneous estimation of microphysical parameters and the atmospheric state using simulated polarimetric radar data and an ensemble Kalman filter in the presence of an observation operator error. *Mon. Wea. Rev.*, **138**, 539–562, <https://doi.org/10.1175/2009MWR2748.1>.
- , —, 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>.
- , R. Kong, C. Liu, J. Luo, M. Xue, and G. Zhao, 2018: CAPS real-time storm-scale EnKF data assimilation and forecasts for the NOAA hazardous weather testbed spring forecasting experiments: Towards the goal of operational ensemble-variational data assimilation for convection permitting models. *Eighth EnKF Data Assimilation Workshop*, Sainte-Adèle, QC, Canada, McGill University, https://web.meteo.mcgill.ca/enkf/abstracts_html/5966e38b03.php.
- Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, <https://doi.org/10.1175/WAF2007106.1>.

- , S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542, <https://doi.org/10.1175/2010WAF2222430.1>.
- 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>.
- , A. P. Khain, N. Benmoshe, E. Ilotoviz, A. V. Ryzhkov, and V. T. J. Phillips, 2014: The anatomy and physics of ZDR columns: Investigating a polarimetric radar signature with a spectral bin microphysical model. *J. Appl. Meteor. Climatol.*, **53**, 1820–1843, <https://doi.org/10.1175/JAMC-D-13-0354.1>.
- Labriola, J., N. 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>.
- , —, —, and M. Xue, 2019a: Explicit ensemble prediction of hail in 19 May 2013 Oklahoma City thunderstorms and analysis of hail growth processes with several multimoment microphysics schemes. *Mon. Wea. Rev.*, **147**, 1193–1213, <https://doi.org/10.1175/MWR-D-18-0266.1>.
- , —, M. Xue, and K. W. Thomas, 2019b: Forecasting the 8 May 2017 severe hail storm in Denver, Colorado, at a convection-allowing resolution: Understanding rimed ice treatments in multimoment microphysics schemes and their effects on hail size forecasts. *Mon. Wea. Rev.*, **147**, 3045–3068, <https://doi.org/10.1175/MWR-D-18-0319.1>.
- Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to warn-on-forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, <https://doi.org/10.1175/WAF-D-17-0145.1>.
- Lim, K.-S. S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*, **138**, 1587–1612, <https://doi.org/10.1175/2009MWR2968.1>.
- 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](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 multicellular 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., 2010: On sedimentation and advection in multimoment bulk microphysics. *J. Atmos. Sci.*, **67**, 3084–3094, <https://doi.org/10.1175/2010JAS3341.1>.
- , 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>.
- Marsh, P. T., 2017: Mesoscale discussion 664. Storm Prediction Center, accessed 5 August 2017, <http://www.spc.noaa.gov/products/md/2017/md0664.html>.
- Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, <https://doi.org/10.1175/JAS-D-11-0315.1>.
- Milbrandt, J. A., and M. K. Yau, 2005a: 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 —, 2005b: 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 —, 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, 2013: Prediction of graupel density in a bulk microphysics scheme. *J. Atmos. Sci.*, **70**, 410–429, <https://doi.org/10.1175/JAS-D-12-0204.1>.
- , and —, 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>.
- Morrison, H., and W. W. Grabowski, 2008: A novel approach for representing ice microphysics in models: Description and tests using a kinematic framework. *J. Atmos. Sci.*, **65**, 1528–1548, <https://doi.org/10.1175/2007JAS2491.1>.
- , 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>.
- , J. Curry, M. D. Shupe, and P. Zuidema, 2005: A new double-moment microphysics parameterization for application in cloud and climate models. Part II: Single-column modeling of Arctic clouds. *J. Atmos. Sci.*, **62**, 1678–1693, <https://doi.org/10.1175/JAS3447.1>.
- , G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, <https://doi.org/10.1175/2008MWR2556.1>.
- , J. A. Milbrandt, G. H. Bryan, I. Kyoko, S. A. Tsendorf, 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](https://doi.org/10.1175/1520-0469(1983)040<1965:TIOSFS>2.0.CO;2).
- NOAA National Centers for Environmental Information, 2018: United States Billion-Dollar Weather and Climate Disasters. NOAA, accessed 2 August 2018, <https://www.ncdc.noaa.gov/billions/>.
- Ortega, K. L., 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, multinet radar data,

- and the ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 141–162, <https://doi.org/10.1175/MWR-D-13-00042.1>.
- , —, —, —, and —, 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>.
- , —, —, N. Snook, and G. Zhang, 2019: Ensemble Kalman filter assimilation of polarimetric radar observations for the 20 May 2013 Oklahoma tornadic supercell case. *Mon. Wea. Rev.*, **147**, 2511–2533, <https://doi.org/10.1175/MWR-D-18-0251.1>.
- Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284, <https://doi.org/10.1175/MWR-D-12-00112.1>.
- Rosencrants, T. D., and W. S. Ashley, 2015: Spatiotemporal analysis of tornado exposure in five US metropolitan areas. *Nat. Hazards*, **78**, 121–140, <https://doi.org/10.1007/s11069-015-1704-z>.
- Ryzhkov, A. V., S. E. Giangrande, V. M. Melnikov, and T. J. Schuur, 2005: Calibration issues of dual-polarization radar measurements. *J. Atmos. Oceanic Technol.*, **22**, 1138–1155, <https://doi.org/10.1175/JTECH1772.1>.
- , M. Pinsky, A. Pokrovsky, and A. Khain, 2011: Polarimetric radar observation operator for a cloud model with spectral microphysics. *J. Appl. Meteor. Climatol.*, **50**, 873–894, <https://doi.org/10.1175/2010JAMC2363.1>.
- , 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, <https://doi.org/10.1175/JAMC-D-13-074.1>.
- 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., 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 —, 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>.
- , 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>.
- , M. Xue, and Y. Jung, 2019: Tornado-resolving ensemble and probabilistic predictions of the 20 May 2013 Newcastle–Moore EF5 tornado. *Mon. Wea. Rev.*, **147**, 1215–1235, <https://doi.org/10.1175/MWR-D-18-0236.1>.
- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677, <https://doi.org/10.1175/2555.1>.
- Snyder, J. C., H. B. Bluestein, D. T. Dawson, and Y. Jung, 2017a: Simulations of polarimetric, X-band radar signatures in supercells. Part I: Description of experiment and simulated ρ_{hv} rings. *J. Appl. Meteor. Climatol.*, **56**, 1977–1999, <https://doi.org/10.1175/JAMC-D-16-0138.1>.
- , —, —, and —, 2017b: Simulations of polarimetric, X-band radar signatures in supercells. Part II: ZDR columns and rings and KDP columns. *J. Appl. Meteor. Climatol.*, **56**, 2001–2026, <https://doi.org/10.1175/JAMC-D-16-0139.1>.
- 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>.
- Straka, J. M., and E. R. Mansell, 2005: A bulk microphysics parameterization with multiple ice precipitation categories. *J. Appl. Meteor.*, **44**, 445–466, <https://doi.org/10.1175/JAM2211.1>.
- Stratman, D., N. Yussouf, Y. Jung, T. Supinie, M. Xue, P. S. Skinner, and B. J. Putnam, 2020: Optimal temporal frequency of NSSL phased-array radar observations for an experimental warn-on-forecast system. *Wea. Forecasting*, **35**, 193–214, <https://doi.org/10.1175/WAF-D-19-0165.1>.
- Supinie, T. A., Y. Jung, M. Xue, D. J. Stensrud, M. M. French, and H. B. Bluestein, 2016: Impact of VORTEX2 observations on analyses and forecasts of the 5 June 2009 Goshen County, Wyoming, supercell. *Mon. Wea. Rev.*, **144**, 429–449, <https://doi.org/10.1175/MWR-D-15-0171.1>.
- , N. Yussouf, Y. Jung, M. Xue, J. Cheng, and S. Wang, 2017: Comparison of the analyses and forecasts of a tornadic supercell storm from assimilating phased-array radar and WSR-88D observations. *Wea. Forecasting*, **32**, 1379–1401, <https://doi.org/10.1175/WAF-D-16-0159.1>.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- 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**, 1649–1668, <https://doi.org/10.1175/2007MWR2071.1>.
- 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](https://doi.org/10.1175/1520-0450(1983)022<1764:NVITAF>2.0.CO;2).
- 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](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>.
- 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](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](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., and K. H. Knopfmeier, 2019: Application of the warn-on-forecast system for flash-flood-producing heavy convective rainfall events. *Quart. J. Roy. Meteor. Soc.*, **145**, 2385–2403, <https://doi.org/10.1002/qj.3568>.
- , D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, <https://doi.org/10.1175/MWR-D-14-00268.1>.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, [https://doi.org/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).
- , A. M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149–166, <https://doi.org/10.1175/WAF909.1>.
- Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2015: Practical predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma: Sensitivity to synoptic timing and topographical influence. *Mon. Wea. Rev.*, **143**, 2973–2997, <https://doi.org/10.1175/MWR-D-14-00394.1>.
- Zhu, Y., and J. M. Navon, 1999: Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint. *Mon. Wea. Rev.*, **127**, 1497–1517, [https://doi.org/10.1175/1520-0493\(1999\)127<1497:IOPEOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<1497:IOPEOT>2.0.CO;2).
- Ziegler, C. L., P. S. Ray, and N. C. Knight, 1983: Hail growth in an Oklahoma multicell storm. *J. Atmos. Sci.*, **40**, 1768–1791, [https://doi.org/10.1175/1520-0469\(1983\)040<1768:HGAOM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<1768:HGAOM>2.0.CO;2).