

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

JONATHAN LABRIOLA

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

NATHAN SNOOK

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

KEVIN W. THOMAS

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

(Manuscript received 6 September 2018, in final form 31 May 2019)

ABSTRACT

Day-ahead (20–22 h) 3-km grid spacing convection-allowing model forecasts are performed for a severe hail event that occurred in Denver, Colorado, on 8 May 2017 using six different multimoment microphysics (MP) schemes including: the Milbrandt–Yau double-moment (MY2), Thompson (THO), NSSL double-moment (NSSL), Morrison double-moment graupel (MOR-G) and hail (MOR-H), and Predicted Particle Properties (P3) schemes. Hail size forecasts diagnosed using the Thompson hail algorithm and storm surrogates predict hail coverage. For this case hail forecasts predict the coverage of hail with a high level of skill but underpredict hail size. The storm surrogate updraft helicity predicts the coverage of severe hail with the most skill for this case. Model data are analyzed to assess the effects of microphysical treatments related to rimed ice. THO uses diagnostic equations to increase the size of graupel within the hail core. MOR-G and MOR-H predict small rimed ice aloft; excessive size sorting and increased fall speeds cause MOR-H to predict more and larger surface hail than MOR-G. The MY2 and NSSL schemes predict large, dense rimed ice particles because both schemes predict separate hail and graupel categories. The NSSL scheme predicts relatively little hail for this case; however, the hail size forecast qualitatively improves when the maximum size of both hail and graupel is considered. The single ice category P3 scheme only predicts dense hail near the surface while above the melting layer large concentrations of low-density ice dominate.

1. Introduction

Hailstorms cause substantial property damage in the United States; between 2017 and 2018 seven hail events caused more than \$1 billion (U.S. dollars) in damage (NCEI 2017). The high cost of recent severe hail events can be attributed in part to the growth and expansion of cities; urban populations have increased by more than 56% and urban areas have grown by 154% since

1960 (Joyce et al. 2008). As cities in hail-prone regions continue to expand, damage from hail is expected to further increase (Rosencrants and Ashley 2015). In an effort to mitigate the impacts and damage of hail, meteorologists are increasingly relying on numerical weather prediction (NWP) models to improve hail forecast skill and extend hail warning lead time. This study assesses next-day hail forecasts for the 8 May 2017 Colorado hail event and analyzes how the treatments of rimed ice in the multimoment microphysics effect surface hail size forecasts.

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

DOI: 10.1175/MWR-D-18-0319.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

Convection-allowing models (CAMs) are run at horizontal grid spacings capable of representing dominant circulations in midlatitude convective storms (Weisman et al. 1997), but insufficient to resolve finescale convective features (Kain et al. 2008)—such grid spacings are typically around 3–4 km. CAMs are able to represent deep, moist convection without too large of a computational cost so that forecasts over large regions [such as the contiguous United States (CONUS)] with relatively long lead times (up to a couple of days) can be produced (Clark et al. 2012a).

CAMs are increasingly used operationally (e.g., Benjamin et al. 2016; Jirak et al. 2018), and multiple research organizations run CAM forecasts routinely (e.g., Coniglio et al. 2010; Clark et al. 2012a; Schwartz et al. 2015; Gallo et al. 2017; Sobash et al. 2016). The Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE), performed each year during the climatological maximum for severe weather over the United States (Clark et al. 2012a), has analyzed next-day CAM ensemble forecast output since 2007 (Xue et al. 2007; Kong et al. 2007). Additionally, the National Weather Service is running deterministic CAM forecasts called High-Resolution Rapid Refresh (HRRR; Smith et al. 2008; Benjamin et al. 2016) and produces loosely integrated ensemble-of-opportunity-type ensemble forecasts called High-Resolution Ensemble Forecast (HREF; Jirak et al. 2018) as outgrowth of the HWT SFE activities.

The horizontal grid spacing used by CAMs is too coarse to resolve many severe weather phenomena including large hail, severe winds, and tornadoes—for this reason, storm surrogates are often used instead to diagnose severe weather. Updraft helicity (UH):

$$UH = \int_{z_0}^{z_1} \zeta w \, dz, \quad (1)$$

is the vertical integration of the product of vertical vorticity ζ and updraft speed w between two atmospheric layers, z_0 and z_1 (typically 2 and 5 km above the surface, respectively), and is used to identify intense rotating updrafts found in supercells (Kain et al. 2008). Because supercell thunderstorms produce a relatively large amount of reported severe hail, wind, and tornadoes, UH can serve as a surrogate predictor for severe weather including hail (e.g., Sobash et al. 2011, 2016; Clark et al. 2013). Other storm surrogate fields include simulated radar reflectivity Z at the -10°C level and column integrated total graupel (CTG). Maximum storm surrogate values calculated during the forecast every time step (i.e., forecast maximum) are often used to capture the rapid evolution of storms (Kain et al.

2010; Clark et al. 2012b, 2013), this is done because model data at typical model output intervals are unable to capture the dynamic and microphysical evolution of storms that occurs at subminute time scales.

Despite the high impact of severe hail events, explicit hail prediction remains relatively understudied. Forecasting hail is difficult; hail growth and melting are dependent upon interactions of growing hailstones with the surrounding environment and involves complex internal microphysical and kinematical processes. Nelson (1983) and Foote (1984) documented that hail growth is dependent upon embryo trajectories through the storm updraft; these trajectories are in turn largely governed by storm dynamics, including updraft orientation, volume, and intensity. A number of studies have documented the intricacies of hail microphysical processes, such as the transition between wet and dry growth (Lesins and List 1986; Garcia-Garcia and List 1992), variation in density (Heymsfield 1978; Knight and Heymsfield 1983; Ziegler et al. 1983; Gilmore et al. 2004; Knight et al. 2008), and the interaction of water on the surface of hail (Chong and Chen 1974; Rasmussen and Heymsfield 1987; Miller et al. 1988; Garcia-Garcia and List 1992; Phillips et al. 2014). Microphysical parameterization (MP) schemes are used to simulate the complex microphysical processes observed in convective storms for CAMs.

In real-time NWP (operational or experimental) all CAMs use bulk MP schemes (hereafter MP schemes) to predict the evolution of precipitation. MP schemes assume a particle size distribution (PSD) function for each hydrometeor species (x), and determine parameters of the function from predicted quantities such as the hydrometeor mass mixing ratio (q_x), number concentration (N_{tx}), reflectivity (Z_x), and volume (v_x) or density (ρ_x). The most commonly assumed distribution for hydrometeor PSDs is a three-parameter gamma-distribution (Ulbrich 1983):

$$N(D) = N_{0x} D^{\alpha_x} e^{(-\lambda_x D)}, \quad (2)$$

where N_{0x} is the intercept parameter, α_x is the shape parameter, and λ_x is the slope parameter of the PSD. MP schemes vary in complexity based upon the number of PSD parameters that are predicted. The number of degrees of freedom or parameters predicted has impacts on how the scheme represents microphysical processes, including melting and sedimentation.

MP schemes usually predict PSDs for distinct hydrometeor categories with predefined characteristics (e.g., rain, cloud water, snow). Certain hydrometeor types such as rain can more easily be represented via this approach; however, rimed ice typically includes a spectrum of particle characteristics such as: density, size, shape,

and fall velocity within convective storms that make it difficult to represent with static parameters. Many MP schemes predict a single category for rimed ice (e.g., Lin et al. 1983; Rutledge and Hobbs 1983; Meyers et al. 1992; Ferrier 1994; Thompson et al. 2008; Morrison et al. 2005; Morrison and Grabowski 2008); user defined parameters determine if the rimed ice behaves more similar to graupel (low-density rimed ice) or hail (high-density rimed ice). MP schemes such as Milbrandt and Yau (2005b) predict two separate rimed ice categories that represent both low- and high-density rimed ice. Prognostic equations for rimed ice volume (and thus density) have also been successfully implemented within MP schemes (e.g., Mansell et al. 2010; Milbrandt and Morrison 2013; Morrison and Milbrandt 2015) and allow NWP models to represent a spectrum of ice and rimed ice characteristics.

Hail size information can be extracted from the CAM predicted microphysical state variables. The Thompson hail size algorithm (Thompson et al. 2018; Gagne et al. 2019) uses hail PSDs predicted by MP schemes to approximate the maximum observable hail size at each grid point; variants of this method have been used to verify forecasts and understand process-level hail growth and decay processes (Milbrandt and Yau 2006a; Snook et al. 2016; Labriola et al. 2017, 2019; Luo et al. 2018). Surveys conducted during the SFE indicate that forecasters find the additional information provided by explicit hail size forecasts useful relative to surrogate fields (e.g., UH) (Gallo et al. 2017).

Objective hail forecast verification is a substantial challenge. Surface-based hail reports exhibit considerable population biases (Wyatt and Witt 1997; Davis and LaDue 2004); hail in rural regions tends to be underreported. Surface-based reports also exhibit considerable size biases; the general public often report hail size in terms of familiar circular or spherical objects (e.g., dimes, golf balls), causing the overrepresentation of hail sizes corresponding to such objects (Jewell and Brimelow 2009). Further, data gaps in surface observations are typically too large to capture the rapid evolution of severe hail producing storms.

The U.S. Next Generation Weather Radar (NEXRAD) system (Crum et al. 1993) is an observational platform capable of capturing the evolution of severe hail within thunderstorms in three dimensions. Radar-derived hail products have been used to verify hail forecasts in several recent studies (e.g., Gagne et al. 2015, 2017; Snook et al. 2016; Labriola et al. 2017, 2019; Luo et al. 2017, 2018) and are often used operationally to diagnose maximum hail size (Cintineo et al. 2012). Radar-derived hail proxies, such as hydrometeor classification algorithm (HCA; Park et al. 2009; Putnam et al. 2017) output, are preferable to surface

reports because they produce high-resolution surface hail size estimates that are not subject to population biases or gaps in data over rural areas (Cintineo et al. 2012). HCAs use single- and dual-polarization radar data (i.e., Z , differential reflectivity Z_{dr} , correlation coefficient ρ_{hv} , and differential phase φ_{dp}) to diagnose the dominant hydrometeor species (e.g., rain, snow, hail) within a given radar observation volume. The HCA of Ryzhkov et al. (2013) and Ortega et al. (2016) further classifies hail into one of three size categories: non-severe, severe, or significant severe (corresponding to 5, 25, or 50 mm in hailstone diameter, respectively).

It is noted that the HCA provides hail size information only at the location of radar observations (i.e., along a radar beam) and not explicitly the surface, which is often the location of greatest interest. Additionally, HCA output is highly sensitive to biases in differential reflectivity measurements (Ortega et al. 2016). For calibrated radar observations, HCA membership functions have been found to classify hail size within a 120-km range of a radar with greater skill than other radar-derived hail products such as the maximum estimated size of hail (MESH) (Ortega et al. 2016). Following Labriola et al. (2019), HCA output is used to verify forecasts in this study both subjectively and objectively.

In this study, we present results from six CAM forecasts using different MP schemes for a multiple hailstorms that occurred over the Denver, Colorado, metropolitan area on 8 May 2017. Rimed ice treatments are compared between MP schemes to understand how microphysical assumptions impact surface hail size forecast skill. The remainder of this paper is organized as follows: In section 2, we discuss the ensemble configuration and the evaluated hail forecast fields, and provide a brief description of the verification methodologies used in this study. Hail size forecasts are evaluated in section 3, together with a microphysical analysis of the model output. Finally, results are summarized and further discussed in section 4.

2. Methods

a. Case overview

On 8 May 2017, positive vorticity advection from an upper-level trough located over Baja California and low-level forcing along a weak frontal boundary located over the Palmer divide (an east–west area of raised terrain to the south of Denver) initiated multicellular thunderstorms south and east of Denver around 1930 UTC. This event serves as an example of a prominent Denver cyclone (Szoke et al. 1984; Blanchard and Howard 1986). Surface wind convergence along the elevated terrain of the Palmer divide positively contributed to thunderstorm

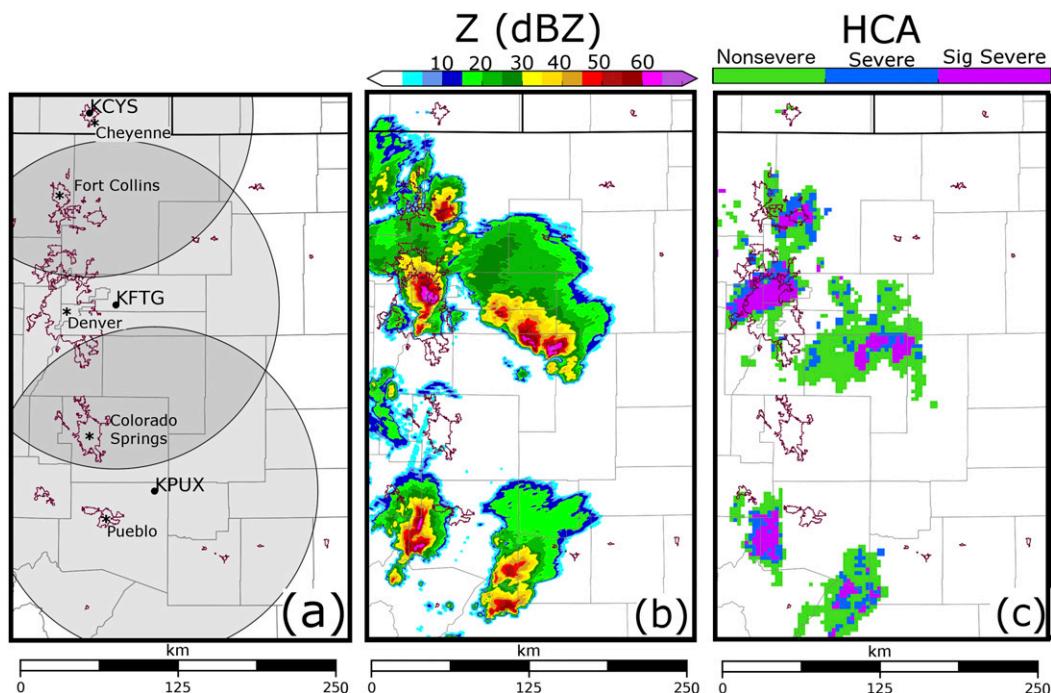


FIG. 1. (a) The verification domain during this study. S-Band radars KCYS, KFTG, and KPUX are plotted, gray shading indicates the 130-km range of the radars (the HCA is applied within this radius). Radar locations (\bullet) and cities (\ast) are included on the map. (b) Maximum observed Z the lowest radar tilt (0.5°) for radars plotted in (a) at 2130 UTC. (c) Merged HCA output between 2000 and 2200 UTC using the same radars as in (a). Merged HCA output is calculated on a 500-m grid and scaled to the native 3-km grid for this image. Output categorizes hail sizes as nonsevere (diameter > 5 mm), severe (diameter > 25 mm), and significant severe (diameter > 50 mm). Purple contours represent urban boundaries.

initiation and intensification. Upslope flow to the north of the frontal boundary also initiated discrete thunderstorms along the Front Range of the Rocky Mountains near Denver and Fort Collins, Colorado, around 2005 UTC. The storms (Fig. 1b) produced large hail between 2000 and 2200 UTC; HCA output (Fig. 1c) and surface-based reports indicate surface maximum hail sizes from these storms ranged between 1 and 2.75 in. (25–70 mm) in diameter. Damage from these storms was particularly extensive because one of the storms produced hail up to 2.75 in. (70 mm) in diameter over the Denver metropolitan area during the evening rush hour. Estimated insured losses from this storm are approximately \$2.3 billion (U.S. dollars) (NCEI 2017), making it the most costly insured catastrophe in Colorado state history (Fritz 2017).

This severe hail event was unique in that the observed surface dewpoint temperature was approximately 8°C , this was an abnormally dry environment to support the development severe hailstorms in Colorado (Modahl 1979). Although the environment was relatively dry, vertical wind shear contributed to the development of prolific hailstorms. The observed 0–6-km bulk shear for this event was 30–40 kt ($1 \text{ kt} \approx 0.5144 \text{ m s}^{-1}$) (Marsh 2017). Idealized simulations run by Dennis and Kumjian

(2017) suggest that increased deep layer shear elongates storm updrafts and increases hail residence time in a favorable growth region. The vertical wind shear also supported the development of supercell thunderstorms for this event, many of the large hail producing thunderstorms exhibited rotation (not pictured). Of the different storm morphologies, supercell thunderstorms most frequently produce large hail in the United States (Blair et al. 2017) and are often associated with hail events that cause more than \$1 billion (U.S. dollars) in damage (e.g., Changnon and Burroughs 2003; Changnon 2009).

b. Model configuration

This study uses an ensemble of six WRF-ARW forecasts produced using configurations closely based upon the Center for Analysis and Prediction of Storms (CAPS) storm-scale ensemble forecast (SSEF) that was run operationally during the 2017 HWT SFE (Jung et al. 2018). During the HWT SFE, the CAPS SSEF was initialized at 0000 UTC every day. CAPS SSEF experiments have been included in multiple studies that analyze next-day severe weather predictions (e.g., Clark et al. 2012a,b; Loken et al. 2017; Gagne et al. 2017), a relatively long lead time for convective-scale forecasts. The 0000 UTC

TABLE 1. Ice categories and prognostic variables for the six analyzed MP schemes during this study. N/A indicates that the category is not predicted by the MP scheme. See main text (section 2c) for more details.

	Graupel	Hail	Ice (general)	Snow	Cloud ice
THO	q_g , diagnostic (N_{ig}), $\rho_g = 500 \text{ kg m}^{-3}$	N/A	N/A	q_s	q_i, N_{ii}
MOR-G	$q_g, N_{ig}, \rho_g = 400 \text{ kg m}^{-3}$	N/A	N/A	q_s, N_{is}	q_i, N_{ii}
MOR-H	N/A	$q_h, N_{ih}, \rho_h = 900 \text{ kg m}^{-3}$	N/A	q_s, N_{is}	q_i, N_{ii}
MY2	$q_g, N_{ig}, \rho_g = 400 \text{ kg m}^{-3}$	$q_h, N_{ih}, \rho_h = 900 \text{ kg m}^{-3}$	N/A	q_s, N_{is}	q_i, N_{ii}
NSSL	q_g, N_{ig}, v_g	q_h, N_{ih}, v_h	N/A	q_s, N_{is}	q_i, N_{ii}
P3 (1 category)	N/A	N/A	$q_{i_tot}, N_{ii_tot}, q_{i_rim}, v_{i_rim}$	N/A	N/A

8 May 2017 operational North American Mesoscale Forecast System (NAM) analysis is used as the analysis background while the 12-km NAM forecasts at 3-hourly intervals are used to provide lateral boundary conditions. Weather Surveillance Radar-1988 Doppler (WSR-88D) data, along with available soundings and upper air observations, are analyzed using the Advanced Regional Prediction System (ARPS) 3DVAR/Cloud-analysis system (Hu et al. 2006a,b) to generate initial conditions on a 3-km CAM grid.

The 3-km grid has $1621 \times 1121 \times 51$ grid points and covers the CONUS. Six CAM forecasts initialized from the 3DVAR analysis differ only in the MP scheme used. They include the Thompson et al. (2008) (THO), Morrison et al. (2005, 2009) and Morrison and Milbrandt (2011) double-moment graupel (MOR-G) and double-moment hail (MOR-H), Milbrandt and Yau (2005b) double-moment (MY2), Mansell et al. (2010) National Severe Storms Laboratory (NSSL) double-moment schemes, and the Predicted Particle Properties (Morrison et al. 2015; Morrison and Milbrandt 2015) single ice category (P3) scheme. The forecasts are run to 24 h, using the WRF-ARW Model version 3.9.1.1 (Skamarock et al. 2008). Other relevant model settings and parameterizations follow those of the 2017 CAPS 3DVAR SSEF control member (Kong 2017). This includes the use of the MYJ planetary boundary layer scheme, the Noah land surface model, and the longwave and shortwave radiation parameterization using the Rapid Radiative Transfer Model for general circulation models (RRTMG).

c. Overview of microphysics schemes

Microphysics schemes selected for this study are partially or fully double-moment schemes that are frequently selected in the CAPS HWT SFE ensemble (e.g., Kong 2017). Multimoment schemes were chosen because single-moment schemes do not explicitly predict the sedimentation of number concentrations and are generally unable to represent size sorting processes (e.g., Wacker and Seifert 2001; Milbrandt and Yau 2005b; Dawson et al. 2010; Jung et al. 2010) unless special treatment is made such as that in the Thompson

et al. (2004) scheme for snow. While there are many differences between the MP schemes used in this study, this paper will primarily focus on differences in the treatment of rimed ice categories (Table 1). Hail and graupel categories are frequently referred to as “rimed ice categories” in this study; however, many of the rimed ice categories also include unrimed frozen raindrops. Unrimed ice is often included in the graupel or hail categories because this is the hydrometeor type they most closely resemble.

Unlike the other MP schemes examined in this study, THO is single-moment for a low-density rimed ice category, for which it predicts only mixing ratio (q_g). In a compromise to represent both hail and graupel within the single rimed ice category, rather than modify density, the THO scheme diagnoses N_{ig} and thus modifies the size of rimed ice. This is done because rimed ice surface accumulations are shown to be more sensitive to changes in the assumed intercept parameter than density (Gilmore et al. 2004). The THO scheme assumes N_{ig} to be a function of the mean mass diameter of supercooled rain and q_g (Khain et al. 2015) (Table 1). Supercooled liquid water is considered to account for wet-growth processes in convective storms (G. Thompson 2019, personal communication). Near storm updrafts, where raindrop diameters and q_g are large, THO increases the size of graupel by decreasing N_{ig} ; outside the updraft region graupel particles are relatively small in diameter.

Both the MOR-G and MOR-H schemes predict a single low-density rimed ice category. The MOR-H scheme implements the same prognostic equations as the MOR-G scheme; however, rimed ice characteristics are more similar to hail (i.e., high-density, and larger terminal velocities) (Table 1). Running both versions of the MOR scheme allow forecasts to be run with different rimed ice particle characteristics that are shown to have a large impact of storm structure (Morrison and Milbrandt 2011).

The MY2 and NSSL schemes have two rimed ice categories, that separately represent high-density (hail) and low-density (graupel) rimed ice particles. They predict both mixing ratio and total number concentration

for both hail and graupel. The NSSL scheme additionally includes prognostic equations for hail and graupel volume mixing ratios, allowing the scheme to derive spatially varying particle densities. The variable density rimed ice categories are used by the NSSL scheme to represent a spectrum of rimed ice characteristics and to improve representation of hailstone production (Labriola et al. 2019).

Unlike the other four MP schemes, the P3 scheme has a user specified number of “free” ice-phase categories, each of which can represent a wide range of ice particle types (e.g., cloud ice, snow, graupel, and hail). The P3 scheme includes four prognostic variables (Table 1) that describe the evolution of a general ice category; this includes: the total ice mass mixing ratio (q_{i_tot}), the total ice number concentration (N_{i_tot}), the accreted rime mass mixing ratio (q_{i_rim}) and the accreted rime volume mixing ratio (v_{i_rim}). The predicted moments define a single PSD that includes up to four size regimes: small circular dense ice, large nonspherical unrimed ice (e.g., snow), partially rimed ice, and fully rimed ice (e.g., graupel, hail); each of these species uses different mass–diameter and fall speed–diameter relationships. The P3 scheme also diagnoses α_x as a function of λ_x . Though multiple ice phase categories prevent the dilution of ice particle characteristics (Milbrandt and Morrison 2016), this study uses the single ice category version of the P3 scheme used by the CAPS SSEF during the 2017 HWT SFE (Kong 2017).

d. Forecast evaluation and verification

This study verifies and analyzes forecasts over a subdomain based in Colorado (Fig. 1a) between 2000 and 2200 UTC, this is when the largest hail was observed. Both storm surrogate methods and the Thompson hail method are used to make hail size forecasts. Storm surrogate fields used include UH, CTG, and Z at the -10°C level; Gagne et al. (2015, 2017) used these fields to diagnose severe and significant severe hail. Thresholds used to identify severe and significant severe hail for storm surrogate fields are consistent with Gagne et al. (2017) as listed in Table 2. The Thompson hail size algorithm (hereafter THAIL) is used to infer the maximum “observable” hail size from the predicted model PSD category most similar to hail, defining the maximum observable hail diameter as the diameter at which the integrated N_{th} of larger hailstones is 10^{-3} m^{-3} , or one hailstone within a $100\text{ m} \times 100\text{ m}$ patch with a depth of 1 m. The choice of this threshold is subjective and open to interpretation; however, variants of THAIL using a similar threshold have been used in other recent hail forecasting studies (e.g., Milbrandt and Yau 2006a; Snook et al. 2016; Labriola et al. 2017, 2019; Luo et al. 2018).

TABLE 2. The thresholds for each field to diagnose nonsevere, severe, and significant severe hail sizes. Thresholds are defined in Gagne et al. (2017). N/A indicates a threshold does not exist for the specific hail size bin.

	Nonsevere	Severe	Significant severe
Updraft helicity	N/A	$>75\text{ m}^2\text{ s}^{-2}$	$>150\text{ m}^2\text{ s}^{-2}$
Column total graupel	N/A	$>25\text{ kg m}^{-2}$	$>50\text{ kg m}^{-2}$
Z at -10°C level	N/A	$>60\text{ dBZ}$	$>60\text{ dBZ}$
Thompson hail	$>5\text{ mm}$	$>25\text{ mm}$	$>50\text{ mm}$

THO, MOR-G, MOR-H, and MY2 assume the hydrometeor category most similar to hail follows an inverse-exponential distribution (i.e., the hail PSD shape parameter is assumed to be zero). This distribution has an extended tail for large hail sizes and can cause the scheme to predict larger ice particles than schemes that assume nonzero shape parameters (e.g., NSSL and P3). For both explicit hail size fields and observations, nonsevere, severe, and significant-severe hail are defined as hail exceeding 5, 25, and 50 mm in diameter, respectively.

Hydrometeor classification algorithm (HCA) output from polarimetric WSR-88D observations is used in this study to verify hail forecasts; this dataset was chosen because it is unaffected by population biases and is generally in agreement with ground-based reports. The HCA used and its implementation are identical to those of Putnam et al. (2017). All hail falls within 130 km of at least one WSR-88D during this event; at approximately this range the HCA algorithm is capable of quite accurately estimating surface hail size (Ortega et al. 2016). A two-dimensional HCA output field is generated by interpolating data from the lowest observed radar tilt for WSR-88D sites Cheyenne (KCYS), Denver (KFTG), and Pueblo (KPUX) to a grid with 500-m grid spacing (Fig. 1a). Environmental information (i.e., air temperature, moisture) required by the HCA is obtained from the 3-km WRF-ARW Model forecasts. To resolve small-scale features (i.e., hail cores) while avoiding excessive noise, radar data are first interpolated to the 500-m grid, and a 9-point smoothing filter is then applied to Z_{dr} and ρ_{hv} . The HCA algorithm is performed for all data within 130 km of a WSR-88D radar. Although the HCA is not applied to the entire subdomain, only to regions shaded gray in Fig. 1a, it covers all areas impacted by severe hailstorms. HCA output from the three radar sites are then merged together; where multiple radars observe the same location, the largest indicated hail size is selected. To reduce noise and avoid spurious detections, a smoothing filter is applied to the merged HCA output. The smoothing filter increases (decreases) the hail size detection when the four closest grid points are larger (smaller); the updated detection is set to

match the four surrounding grid points. This technique follows that of (Labriola et al. 2019). Finally, the radar-HCA 500-m grid is then remapped to the WRF 3-km grid by taking the largest value within a 3-km pixel (Fig. 1c).

Due to the potentially large error present in day-ahead (18–24 h) convective-scale forecasts, CAMs are oftentimes unable to predict the initiation and development of individual observed storms accurately. The model is capable; however, of skillfully predicting an environment favorable for convective development, though the model environment is often modified by modeled thunderstorms preceding the time of interest. Due to the limits of intrinsic predictability (Lorenz 1969), verification of 18–24-h severe and significant severe hail predictions using a point-by-point approach provides little insight; this study will instead verify the spatial coverage of hail. Fractions skill score (FSS; Roberts and Lean 2008) is a metric that compares the fractional coverage of predicted and observed events at varying neighborhood lengths. The fractional coverage approach taken by FSS circumvents the “double penalty” issue associated with displacement errors suffered in point-by-point verifications. FSS is defined with the mean square error (MSE) and the reference mean square error (MSE_{ref}):

$$MSE_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{(n)ij} - M_{(n)ij}]^2, \quad (3)$$

$$MSE_{(n)\text{ref}} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{(n)ij}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M_{(n)ij}^2 \right], \quad (4)$$

$$FSS_{(n)} = 1 - \frac{MSE_{(n)}}{MSE_{(n)\text{ref}}}, \quad (5)$$

where O is the binary observational field; M is the binary forecast field; N_x and N_y are the x and y dimensions of the domain, respectively; and n is the neighborhood length. FSS can be calculated over an increasingly larger neighborhood length to determine the smallest neighborhood scale where a forecast predicts the spatial coverage of hail with skill. A forecast is considered to have acceptable (or useful) skill when the FSS meets the criterion:

$$FSS > 0.5 + \frac{f_0}{2}, \quad (6)$$

where f_0 is the random forecast skill. Verification is performed for occurrence of hail exceeding a given size within a 42-km radius of a grid point. This distance is

consistent with that used in current Storm Prediction Center (SPC) convective outlooks (i.e., “within 25 miles of a point”) and previous CAM studies (e.g., Gagne et al. 2017). FSS is used along with subjective comparisons to identify strengths and weaknesses of the different hail forecast fields and compare forecasts performed using different MP schemes. Forecasts are compared for a single hail event, additional case studies are required to draw meaningful conclusions as microphysical processes vary substantially due to changes in the storm morphology and environment (e.g., Nelson 1983; Heymsfield 1983; Dennis and Kumjian 2017).

3. Results

Storm surrogate (UH, CTG, Z at the -10°C level) and surface maximum hail size (diagnosed via THAIL) forecast fields produced using the THO, MOR-G, MOR-H, MY2, NSSL, and P3 schemes are first evaluated using HCA output as a proxy for hail size observations (Fig. 1c). The forecast verification domain (see Fig. 1) extends along the front range of the Rocky Mountains in Colorado between 2000 and 2200 UTC, when HCA output indicates hail exceeding 50 mm in diameter in several storms (Fig. 1c).

a. Verification

Hail size forecasts (Fig. 2) diagnosed via THAIL are the primary focus of subjective verification because they can be directly compared against HCA output (Fig. 1c). The P3 hail size forecast is not included within this study because, at the time of the study, the THAIL algorithm was not coded for the P3 scheme. MOR-G, MOR-H, and MY2 hail size forecasts (Figs. 2b–d) predict the total area coverage of hail >5 mm to be similar to observations (Fig. 1c); however, the THO and NSSL forecasts predict spatially small swaths of hail (Figs. 2a,e). Although the NSSL forecast predicts storms to produce updrafts that are similar in intensity to the other forecasts (Fig. 3), the scheme predicts storms to produce mostly graupel at the surface and relatively little hail. Since in the HWT SFE the THAIL algorithm does not consider the NSSL graupel category, the forecast underestimates to coverage of rimed ice at the surface (Fig. 1c). The NSSL hail and graupel categories are further analyzed later, in section 3b.

THAIL uses hail PSDs diagnosed from model output at the lowest model level above the surface to estimate surface maximum hail size. This method is subject to the various microphysical characteristics and assumptions made by each MP scheme, differences among MP schemes impact the surface diagnosed hail size. None of the forecasts (Fig. 2) predict significant severe hail

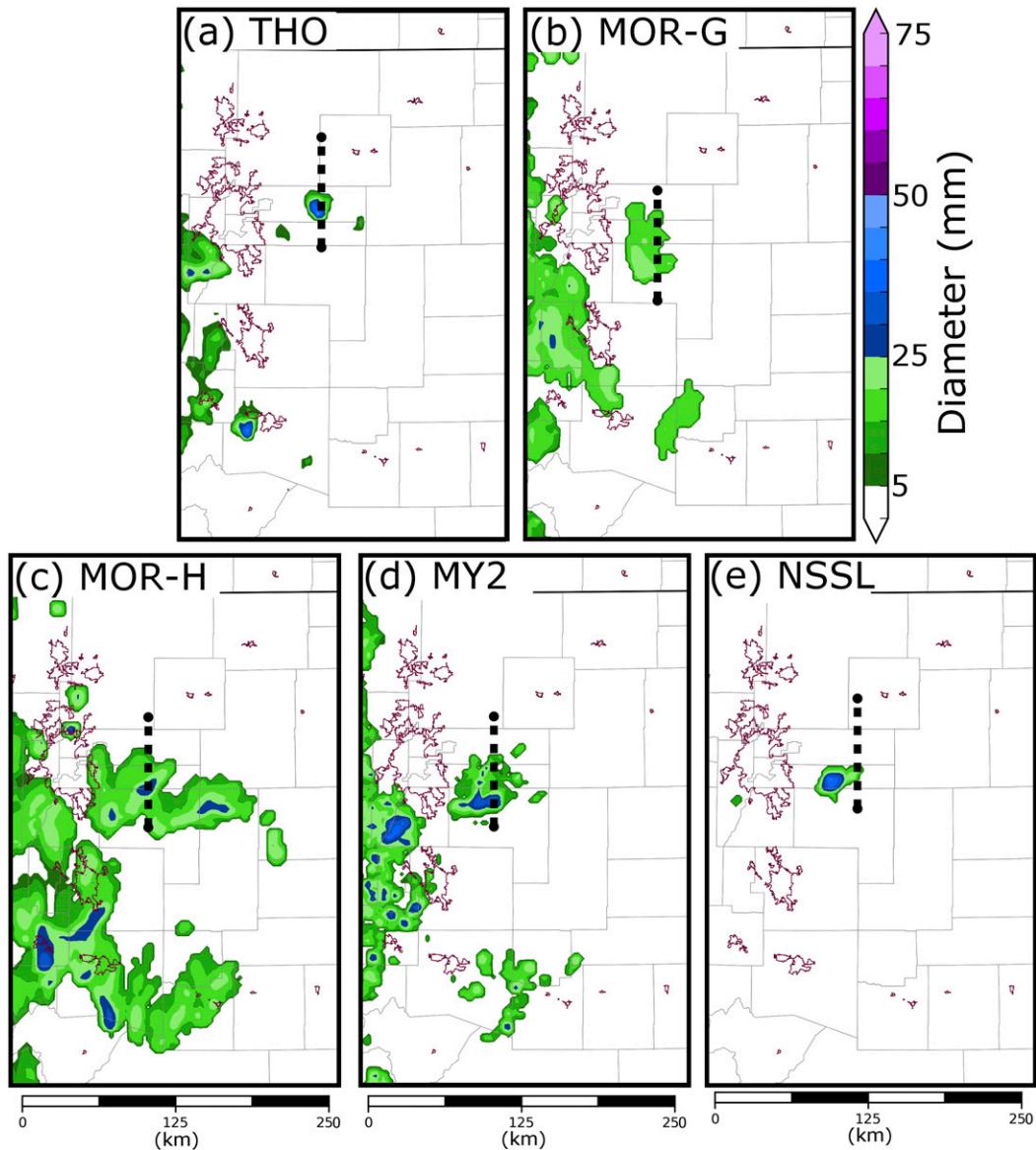


FIG. 2. Forecast maximum hail size between 2000 and 2200 UTC. Surface hail size forecasts are diagnosed by the THAIL algorithm; forecasts are run using the (a) THO, (b) MOR-G, (c) MOR-H, (d) MY2, and (e) NSSL schemes. The dashed black line marks the location where vertical cross sections are analyzed in Figs. 5, 8, and 12–14. The background is the same as Fig. 1.

coverage; it is suspected this hail size underprediction is due in part to predicted storm updrafts that are relatively weak ($<28 \text{ m s}^{-1}$) (Fig. 3). MOR-G (Fig. 2b), MOR-H (Fig. 2c), and MY2 (Fig. 2d) surface hail size forecasts predict widespread nonsevere hail. The THO, MOR-H, MY2, and NSSL forecasts (Figs. 2a,c–e) also predict localized cores of severe hail that better capture large surface hail sizes observed in HCA output (Fig. 1c) than the MOR-G forecast (Fig. 2b). A more in-depth analysis of the microphysical treatment of rimed ice for each MP scheme is presented in section 3b.

FSS is used in this study to examine at which scales forecasts skillfully predict the fractional coverage of hail. It is important to note that the FSS is calculated over a small domain that includes multiple hail producing thunderstorms. The fractional observed hail coverage within this domain is thus much larger than what is normally observed over the full CONUS, and the FSS required for a skillful forecast is thus unusually high (Table 3). Additionally, the fractional coverage of observed hail (Table 3) is larger than what HCA output suggests because the verification method considers the largest observed hail

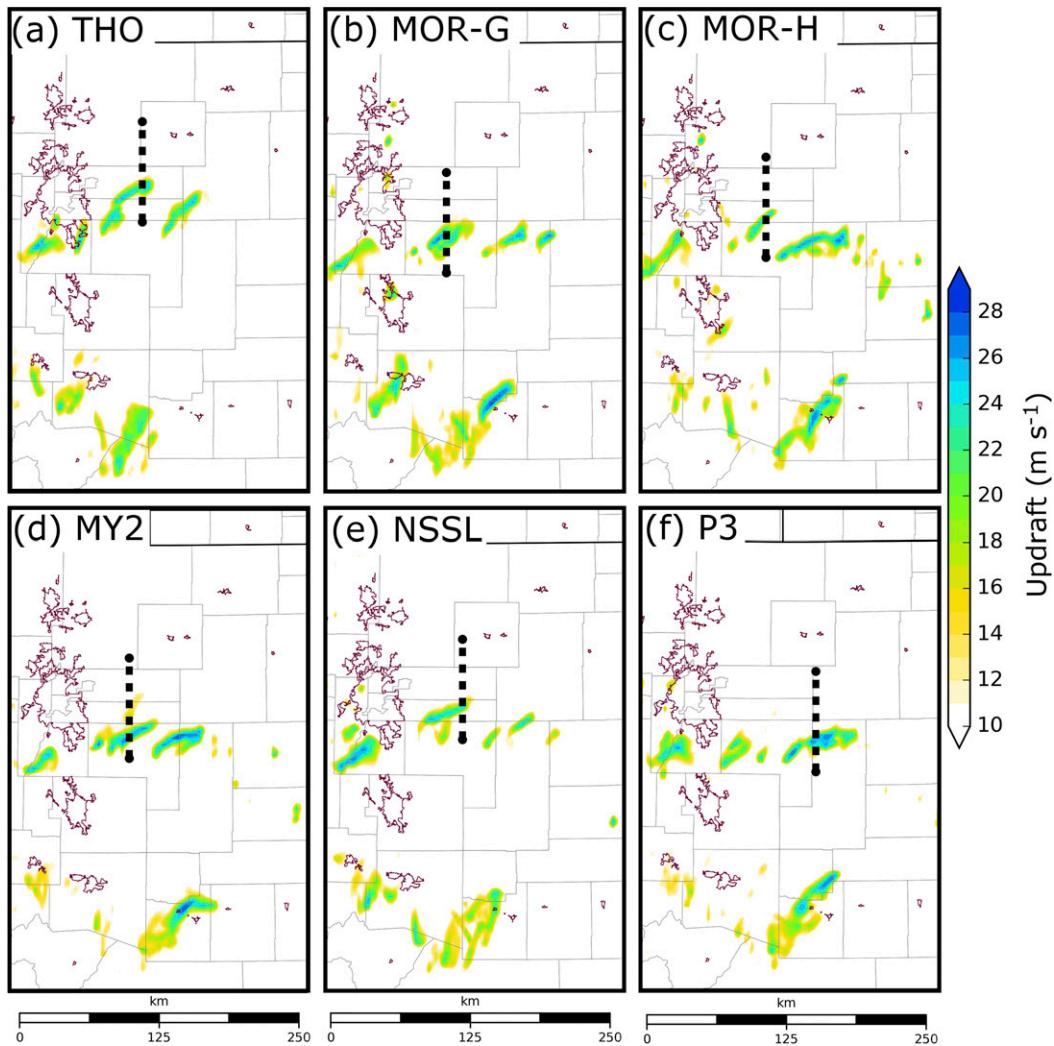


FIG. 3. Forecast maximum updraft velocity between 2000 and 2200 UTC. Forecasts are run using the (a) THO, (b) MOR-G, (c) MOR-H, (d) MY2, (e) NSSL, and (f) P3 schemes. The dashed black line marks the location where vertical cross sections are analyzed in Figs. 5, 8, and 12–14. The background is the same as Fig. 1.

diameter within a 42-km radius of a grid point. Although the skill threshold is large for this study, severe hail is considered to be a rare event climatologically throughout the CONUS; for these types of rare hazards, 0.5 is considered to be the lower limit for skill (Mittermaier and Roberts 2010). The FSS of significant severe hail is not plotted because forecasts exhibit little skill at scales less than 80 km. The FSS is instead calculated for hail exceeding 5 mm in diameter, which is roughly considered the coverage of “all hail” (including both severe and non-severe hail). Storm surrogate methods have generally not been calibrated to diagnose regions of nonsevere hail and therefore are omitted from the “all hail” category.

Storm surrogates that are defined using storm dynamic fields (i.e., UH) predict severe hail with more skill than surrogates defined using microphysical variables

(i.e., CTG, Z at the -10°C level). Maximum hail size diagnosed via THAIL exhibits the most skill at predicting severe hail coverage for the MOR-H (Fig. 4c) and MY2 (Fig. 4d) forecasts; however, UH is the only parameter that has a FSS that exceeds 0.5 for all forecasts (Fig. 4). Although UH predicts the coverage of severe hail with some skill, there are variations in skill

TABLE 3. Observed fractional coverage of hail considered by FSS and the criterion for acceptable skill for each hail size threshold.

Size category	Fractional coverage	Skill
All hail	0.588	0.794
Severe	0.503	0.751
Significant severe	0.437	0.719

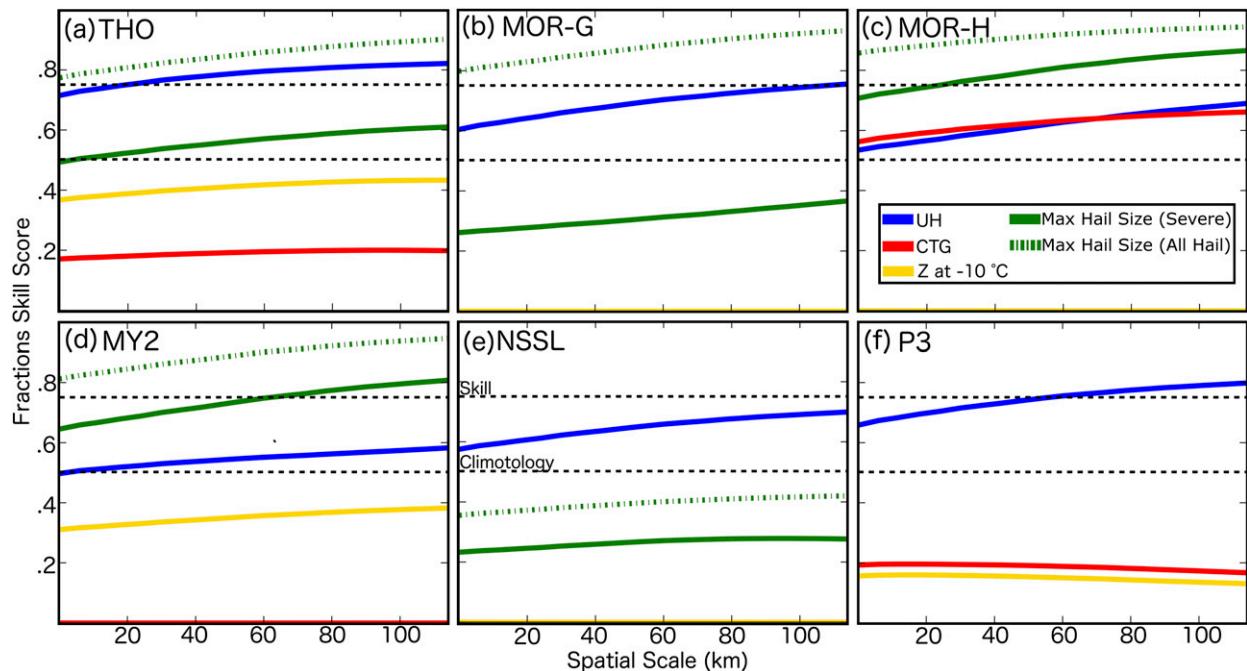


FIG. 4. The FSS of predicting all hail and severe hail coverage using storm surrogate methods (UH, CTG, Z at the -10°C level) and maximum surface hail size diagnosed via THAIL (maximum hail size). Forecasts are created using the (a) THO, (b) MOR-G, (c) MOR-H, (d) MY2, (e) NSSL, and (f) P3 schemes. In each figure the bottom horizontal black dashed line represents the climatology of severe hail over the verification domain between 2000 and 2200 UTC (Fig. 1), the top horizontal black dashed line represents the line of skill for severe hail. Maximum hail size is the only field plotted for the all hail size category (dashed green line) because storm surrogate methods do not identify nonsevere hail coverage. Some curves [e.g., Z at the -10°C line in (b)], are not found in the plots because they fall below the lower bound of plotted domain. The P3 scheme forecast does not include maximum surface hail size.

between forecasts for this case. Variations in the predictive skill of UH are attributed to differences in storm track length and position (both of which vary substantially between members) during the 2-h verification window. Results are similar to Gagne et al. (2017) that found UH to be a skilled predictor of large hail because the parameter identifies storms with strong rotating updrafts. Storms with rotating updrafts tend to be most capable of producing severe hail. Because UH is strictly based upon storm dynamics, the field is not directly affected by the large differences in microphysical variables between MP schemes.

The skill of storm surrogates defined by microphysical variables is strongly dependent upon choice of MP scheme, in particular assumptions made within and the predicted PSD details. The THO and MY2 schemes predict large hail more than 10 km above mean sea level (MSL), leading to increased performance of the storm surrogate Z at the -10°C level (Figs. 4a,d). Z at the -10°C level underestimates the coverage of hail for all other forecasts (Figs. 4b,c,e,f), because the other MP schemes predict either smaller rimed ice particles aloft or large rimed ice is not lofted as high above the ground. With the exception of MOR-H (Fig. 4c) that

predicts larger concentrations of rimed ice aloft than the other forecasts, CTG shows no skill at predicting the coverage of severe hail in terms of the FSS (Fig. 4).

Although hail diagnostic fields (e.g., UH, CTG) exhibit some skill when predicting the coverage of severe hail, verifications are limited to the scope of a single hail event. Previous studies that analyzed multiple severe hail events (e.g., Gagne et al. 2017), documented that storm surrogates can produce extreme false alarm rates when the NWP model forecasts storms that are never observed in nature. Further, UH is tuned to detect hail produced by rotating thunderstorms and is unable to detect hail for other storm modes. Results of verifications provide information on the skill of the hail forecasts for this event, but additional case studies are necessary to understand how hail forecasts vary between cases.

b. Microphysical analysis

Model output is used in this study to understand the modeled microphysical processes that impact the overall quality of explicit hail size and storm surrogate forecasts. To this end, vertical cross sections are presented for various predicted variables, including: hydrometeor mass content (m_x), number concentration (N_x), density

(ρ_x), mass-weighted mean diameter (D_{mx}), and rime fraction. The mass content of hydrometeor species x (m_x) is the product of the hydrometeor mass mixing ratio and air density. Vertical cross sections are plotted at the location of the maximum surface hail size according to hail size forecasts, this is typically collocated with the storm updraft. Comparison of vertical cross sections are performed at 2130 UTC; at this time each scheme predicts a rotating thunderstorm ($UH > 60 \text{ m}^2 \text{ s}^{-2}$) and rimed ice ($q > 1 \times 10^{-4} \text{ g kg}^{-1}$) at the surface.

Contoured frequency by altitude diagrams (CFADs; Yuter and Houze 1995) will also be used to gain a better understanding of microphysical trends for: N_{0x} , λ_x , ρ_x , and hail size diagnosed from model output via THAIL. CFADs, which are calculated over the verification domain (see Fig. 1) at 5 min intervals during the forecast evaluation period (2000–2200 UTC), provide the number of event occurrences per bin at each vertical level with frequencies normalized by the total number of occurrences. Information gained from this analysis can be used to understand spatial trends in the microphysical variables predicted by the MP schemes.

1) SINGLE RIMED ICE CATEGORY SCHEMES

Characteristics of rimed ice particles vary significantly between MP schemes, and these differences are largely governed by the scheme-imposed properties of the rimed ice category in question. The THO, MOR-G, and MOR-H schemes predict a single hydrometeor category that represents rimed ice. Because all rimed ice particles are represented within a single category, all three schemes predict large mass contents of ice aloft (Figs. 5a,d,g). Despite differences in assumed rimed ice density (Table 1), MOR-G (Fig. 5f) and MOR-H (Fig. 5i) predict rimed ice to remain mostly small ($D_{mx} < 2 \text{ mm}$) above the 0°C isotherm, larger D_{mx} values are found primarily beneath the 0°C isotherm. The THO scheme predicts N_{ig} (Fig. 5b) to decrease within the updraft, this region corresponds with a relative increase in D_{mg} (Fig. 5c).

CFADs of the rimed ice category PSD parameters indicate the THO, MOR-G, and MOR-H schemes predict large intercept (Figs. 6a,d,g) and slope parameters (Figs. 6b,e,h), causing the rimed ice particles to be smaller than 20 mm in diameter (Figs. 6c,f,i) in general. MOR-G (Fig. 6d) and MOR-H (Fig. 6g) predict the intercept parameter of rimed ice to occupy a larger range of values than THO (Fig. 6a), this is in part because the THO scheme constrains N_{0g} to remain between 10^4 and $3 \times 10^6 \text{ m}^{-4}$. All three MP schemes predict the intercept parameter to be within the observed range of N_{0g} values (10^4 – 10^{10} m^{-4}) (Knight et al. 1982). Although MOR-H predicts rimed ice most similar to hail and THO predicts a graupel/hail hybrid

category, both schemes predict the intercept parameter of rimed ice to be several orders of magnitude larger than the MY2 or NSSL hail categories. MY2 and NSSL hail PSD parameters are discussed later in the section.

MOR-G (Figs. 6d,e) and MOR-H (Figs. 6g,h) rimed ice PSD parameters behave similarly, modifications made to the bulk density and terminal velocity of rimed ice does not appear to impact rimed ice PSDs. Because MOR-H does not predict substantially larger hailstones than MOR-G in a bulk sense it is speculated that the increased fall speed of hail compared to graupel causes MOR-H to predict more rimed ice to reach the surface before melting resulting in the increased surface coverage of hail (Fig. 2c). CFADs indicate that at the lowest model grid heights (approximately 2 km above MSL), MOR-H (Fig. 6i) predicts an increased frequency of rimed ice reaching the surface than MOR-G (Fig. 6f).

The diagnostic graupel intercept parameter in THO causes the size of graupel to increase. At 2–10 km above MSL the THO scheme predicts a greater frequency (approximately 2.5×10^{-3}) (Fig. 6a) of low N_{0g} values (approximately 10^4 m^{-4}) causing the maximum size of hail to exceed 70 mm in diameter (Fig. 6c). At 8–10 km above MSL m_g is relatively large (Fig. 5a) and supercooled liquid water is present, these conditions cause the THO scheme graupel PSD to predict relatively low N_{0g} values (10^4 m^{-4}) aloft. As part of the diagnostic procedure the THO scheme requires all grid points within a vertical column to assume the lowest N_{0g} value above. Many of lowest N_{0g} values are located 8–10 km above MSL, this subsequently enhances the size of rimed ice toward the surface in the storm updraft region (Fig. 5c). More than 10 km above MSL there is no supercooled liquid water and thus N_{0g} is diagnosed to increase to 10^6 m^{-4} (Fig. 6a). Although the THO scheme predicts rimed ice to grow farther aloft than MOR_H or MOR_G, the scheme is in closer agreement with the observed hail growth zone, which is located well above the 0°C isotherm (e.g., Ziegler et al. 1983; Foote 1984).

The diagnostic intercept parameter in the THO scheme causes sharp local gradients in hail size because variables used to diagnose N_{ig} are determined locally (G. Thompson 2019, personal communication). To illustrate how modifying N_{ig} impacts rimed ice particle size, two graupel PSDs from adjacent grid points are sampled at the “×” in Fig. 5c. Although q_g is relatively similar between the two grid points, the THO scheme diagnoses N_{ig} to be much smaller in location 1 than location 2 because supercooled liquid is present at the former grid point (Table 4). At location 1 m_g is distributed among fewer hailstones causing the graupel PSD to

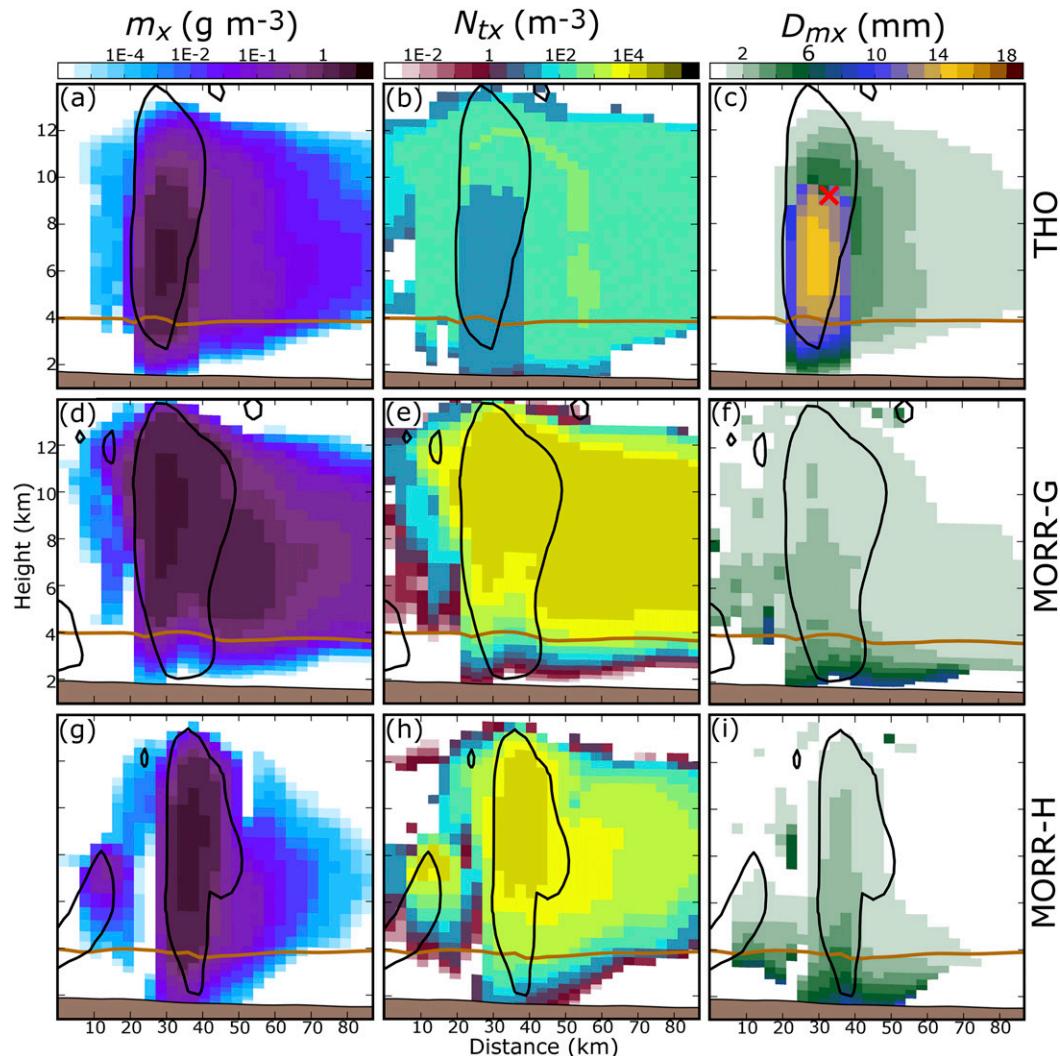


FIG. 5. A vertical cross section of a hail core at 2130 UTC, the location of the cross section is marked in Fig. 2. Vertical cross sections are taken of the rimed ice category predicted by the (a)–(c) THO, (d)–(f) MORR-G, and (g)–(i) MORR-H schemes. Plotted variables include rimed ice: (a),(d),(g) m_x , (b),(e),(h) N_{tx} , and (c),(f),(i) D_{mx} . 1 m s^{-1} updrafts are contoured in black, the brown horizontal line denotes the 0°C isotherm. The red “x” in (c) marks the location where graupel particle size distributions are sampled.

favor fewer but larger graupel particles than location 2 (Fig. 7).

Neither the MORR-G (Fig. 6f) nor the MORR-H (Fig. 6i) schemes replicate the growth of rimed ice aloft, instead both schemes predict the diameter of rimed ice to increase toward the surface. It is speculated that large mean sizes near the surface are due to excessive size sorting. Excessive size sorting is frequently documented in two-moment MP schemes (e.g., Milbrandt and Yau 2005a; Dawson et al. 2014; Morrison et al. 2015; Johnson et al. 2016, 2019, manuscript submitted to *Mon. Wea. Rev.*), and typically occurs when the mass fall speed is larger than number fall speed, causing more mass to reach the surface and increase the particle mass-weighted mean

diameter. Because the THO scheme only predicts graupel mass mixing ratio, this scheme cannot explicitly predict gravitational size sorting of rimed ice. Instead, the scheme requires all grid points in a vertical column to assume the lowest N_{0g} value aloft; this treatment increases the size of the rimed ice particles in the hail core and simulates the effects of size sorting to some extent.

The effects of excessive size sorting are exacerbated for MP schemes that assume inverse exponential distributions (e.g., MORR-G and MORR-H). Many double-moment MP schemes make additional efforts to improve the simulation of size sorting. For example, a version of the two-moment Milbrandt and Yau (2006b) scheme diagnoses hydrometeor PSD shape parameters instead of

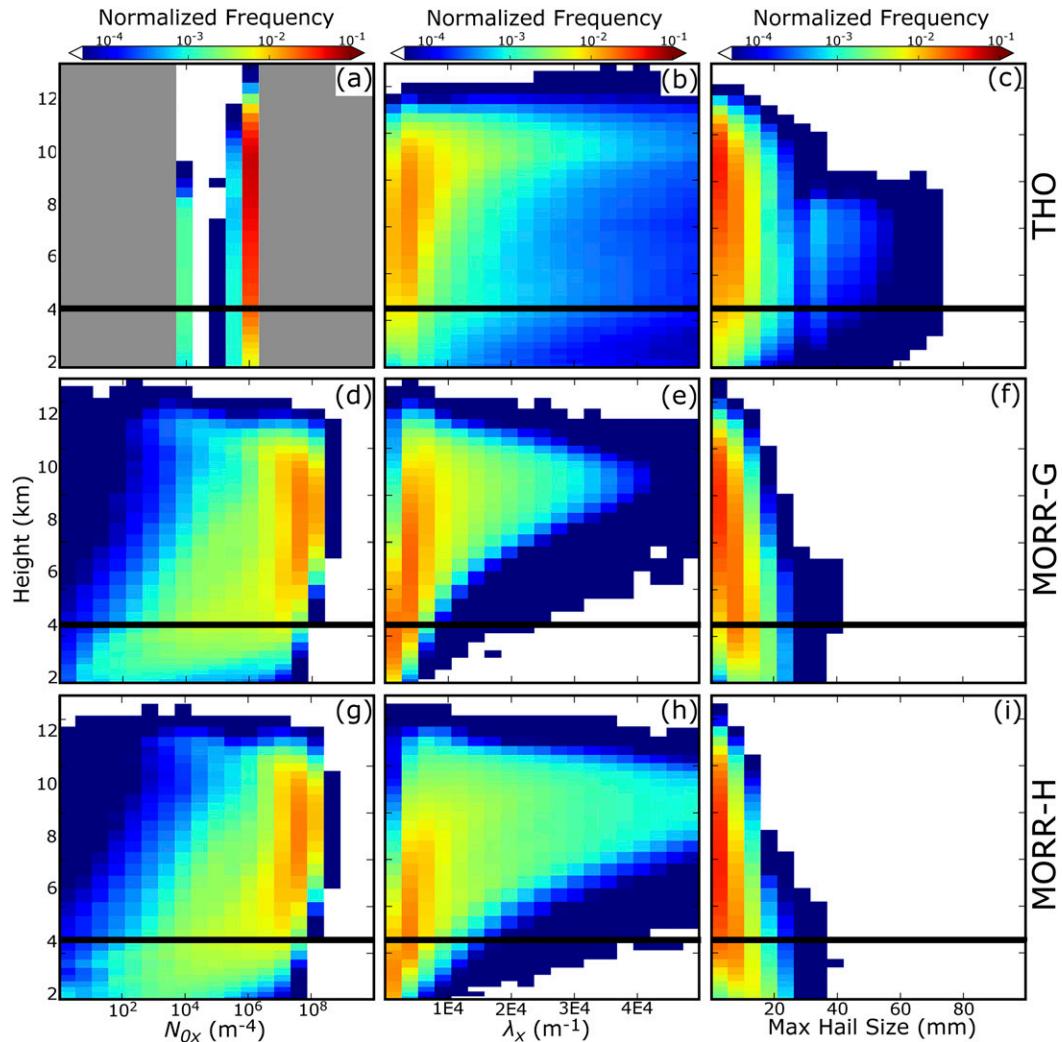


FIG. 6. CFADs for (a),(d),(g) N_{ox} , (b),(e),(h) λ_x , and maximum hail size diagnosed via (c),(f),(i) THAIL for the rimed ice category predicted by the (a)–(c) THO, (d)–(f) MORR-G, and (g)–(i) MORR-H schemes during the forecast evaluation period (2000–2200 UTC). Data are considered over the verification domain (Fig. 1). N_{ox} is plotted using exponentially growing bin intervals. Each CFAD has 20 horizontal bins evenly distributed over range of LOG (10^0)–LOG (10^{10}) m^{-4} for N_{ox} , 0–50 000 m^{-1} for λ_x , and 0–100 mm for maximum hail size. The approximate $0^\circ C$ isotherm height is plotted with a horizontal black line. Gray shading in (a) marks the allowable range of N_{og} in the THO scheme.

using fixed values, the THO scheme (Thompson et al. 2008) increases the rain number terminal velocity to resemble more closely mass terminal velocity, the NSSL scheme adjusts hydrometeor number concentration to prevent artificial growth in reflectivity (Mansell 2010), and the P3 scheme uses drop breakup to limit unreasonably large raindrops at the lower levels (Morrison et al. 2015).

2) TWO RIMED ICE CATEGORY SCHEMES

The MY2 (Figs. 8a–c) and NSSL (Figs. 8d–f) schemes predict separate hail and graupel categories, with the hail category generally producing mass and number

concentration values that are orders of magnitude lower than corresponding values for graupel. The ability to predict two separate categories allows the hail category to represent generally larger rimed ice particles, while smaller, more numerous rimed ice particles remain confined to the graupel category. PSD parameter CFADs (Fig. 9) demonstrate that the MY2 and NSSL schemes favor fewer but larger ice particles in the hail category during the forecast evaluation period. Both the MY2 and NSSL schemes predict N_{oh} (Figs. 9a,d) to be approximately 2 orders of magnitude smaller than the single rimed ice category schemes intercept parameters; neither scheme predicts λ_h (Figs. 9b,e) to exceed

TABLE 4. THO predicted (q_g) and diagnosed (N_{ig}) graupel moments sampled at adjacent grid points marked by a red “×” in Fig. 5c. Variables diagnosed particle size distributions in Fig. 7.

	q_g	N_{ig}	q_r
Location 1 ($x = 33$ km, 2.67 g kg $^{-1}$, $y = 9.2$ km above MSL)	2.67 g kg $^{-1}$	29.51 m $^{-3}$	4.8×10^{-3} g kg $^{-1}$
Location 2 ($x = 33$ km, 2.31 g kg $^{-1}$, $y = 9.7$ km above MSL)	2.31 g kg $^{-1}$	328.79 m $^{-3}$	0.0 g kg $^{-1}$

7.5×10^3 m $^{-1}$. This combination of PSD parameters causes the MY2 and NSSL forecasts to predict larger hail (Figs. 9c,f) than MOR_G (Fig. 6f) or MOR_H (Fig. 6i).

In previous studies such as Johnson et al. (2016), the MY2 scheme was found to produce many small hailstones due to the three-component freezing of rainwater (i.e., frozen raindrops). Vertical cross sections taken through the MY2 forecast indicate that N_{th} (Fig. 8b) increases and D_{mh} (Fig. 8c) decreases toward the 0°C isotherm in the hail core; this is due to the production of many small hailstones. While this process is not obvious in the forecast maximum hail size CFAD diagram (Fig. 9c), in part because the CFAD is calculated over the entire verification domain, it is noted that the MY2 scheme predicts a high-frequency occurrence ($>10^{-2}$) of hailstones with a maximum diameter of approximately 5 mm near the 0°C isotherm. Although the MY2 scheme produces small hailstones, the CFAD diagram of forecast hail size indicates the MY2 forecast (Fig. 9c) predicts large hailstones (>50 mm) to be approximately 4–9 km above MSL. Maximum hail size CFADs (Figs. 9c,f) suggest that both the NSSL and MY2 schemes predict hail to grow primarily above the 0°C isotherm, this is in agreement with many previous observational studies such as Ziegler et al. (1983).

The MY2 scheme assumes that graupel is created via the three-component freezing of rain, and as a result both the hail and graupel categories exhibit similar behaviors in terms of mass (Fig. 10a) and number concentration growth (not pictured). The NSSL scheme creates hail through a series of microphysical interactions: graupel increases in density during wet growth, and is subsequently converted into hail (Mansell et al. 2010). Since hail is only produced from dense graupel (as opposed to from freezing of raindrops), the NSSL scheme generally predicts relatively few ($N_{th} < 10$ m $^{-3}$), but large ($D_{mh} > 14$ mm) hailstones (Figs. 8e,f) within the storm updraft region. These results are in agreement with Johnson et al. (2016) and (Labriola et al. 2019).

Due to a multistep hail production process, the NSSL forecast produces less rimed ice in the hail category (Fig. 10b) than the MY2 scheme (Fig. 10a). The MY2

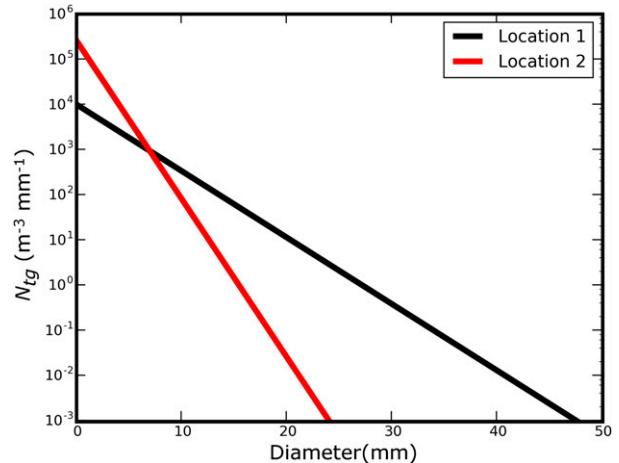


FIG. 7. Particle size distributions collected at two adjacent points in the hail core of a thunderstorm predicted by the THO scheme. The location where the PSDs are sampled is marked with an “×” in Fig. 5c. The predicted (q_g) and diagnosed (N_{ig}) graupel moments that diagnose the graupel PSDs are provided in Table 4.

forecast predicts more hail than the NSSL forecast because the scheme creates hail quickly from frozen rain drops. Although the NSSL scheme populates the hail category more slowly than the MY2 scheme, both schemes predict storms to produce large quantities of graupel during the forecast evaluation period (Fig. 10). NSSL and MY2 predicted graupel is not considered in current operational surface hail size forecasts (Fig. 2) because the THAIL algorithm only diagnoses the maximum size of ice in the hail category for these schemes.

To determine the impact of including graupel within the THAIL algorithm, the forecast maximum size of hail (FCST-H) (Figs. 11b,d) is compared to the forecast maximum size of both hail and graupel (FCST-HG) (Figs. 11c,e) at 2130 UTC. FCST-H is equivalent to the THAIL algorithm implemented in the HWT SFE (Fig. 2). FCST-HG predicts larger swaths of surface hail that qualitatively resemble HCA output (Fig. 11a) more closely than FCST-H for the NSSL scheme (Figs. 11d,e). MY2 FCST-H and FCST-HG (Figs. 11b,c) are relatively similar, in part because the scheme predicts multiple swaths of hail at the surface. Although NSSL FCST-HG (Fig. 11e) exhibits qualitative improvement in surface hail size forecast skill, additional case studies are needed to evaluate the benefit of including both rimed ice categories.

Vertical cross sections of m_g (Figs. 12a,d) and N_{ig} (Figs. 12b,e) suggest that the NSSL scheme predicts storms to produce more graupel at the surface than the MY2 scheme. Although the size of graupel is relatively similar between the two forecasts (Figs. 12c,f), the NSSL scheme includes prognostic volume mixing ratio equations that increase the bulk density (and thus fall speed)

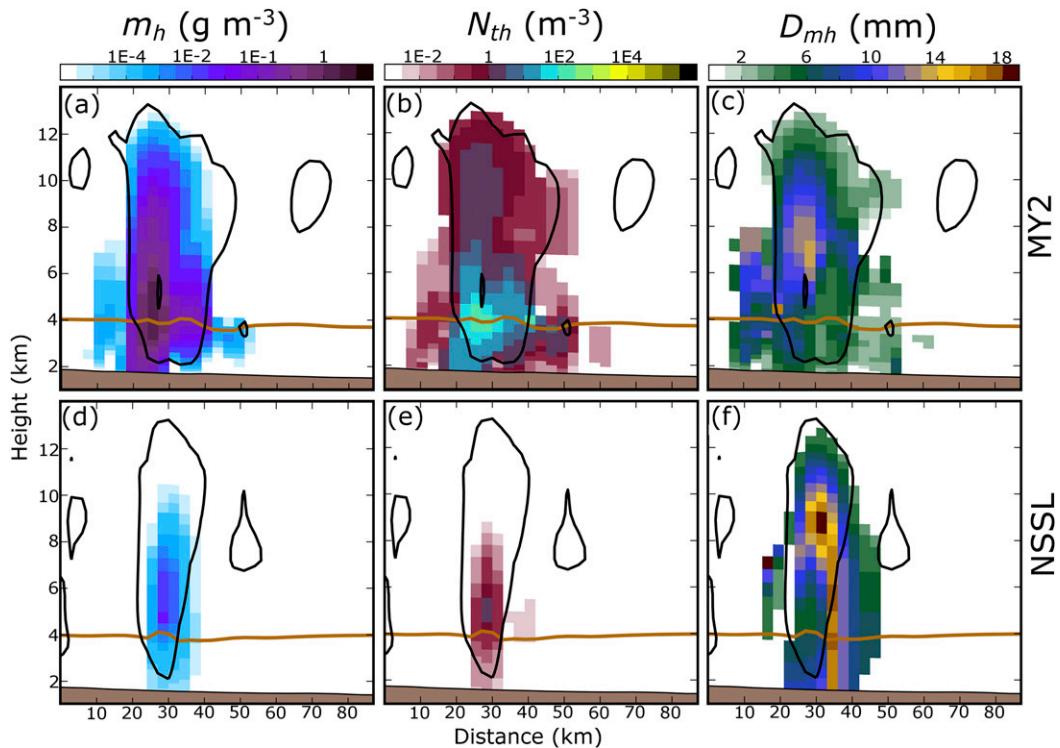


FIG. 8. A vertical cross section of a hail core at 2130 UTC, the location of the cross section is marked in Fig. 2. Vertical cross sections are taken of the predicted rimed ice category most similar to hail in the (a)–(c) MY2 and (d)–(f) NSSL schemes. Plotted variables include (a),(d) m_h , (b),(e) N_{th} , and (c),(f) D_{mh} . 1 m s^{-1} updrafts are contoured in black, the brown horizontal line denotes the 0°C isotherm.

of graupel to resemble high-density rimed ice (i.e., hail) in the melting layer. Graupel particles with larger fall speeds fall more quickly through the melting layer and are less likely to melt before reaching the surface (Figs. 12d,e). The MY2 scheme assumes low-density graupel throughout the storm (Table 1), we suspect reduced graupel fall speeds limit the amount of graupel that reaches the surface (Figs. 12a,b). Prognostic rimed ice density MP schemes (i.e., NSSL and P3) are further discussed in the following subsection.

3) PROGNOSTIC RIMED ICE PROPERTY SCHEMES

Variable density rimed ice categories allow MP schemes (e.g., Mansell et al. 2010; Milbrandt and Morrison 2013; Morrison and Milbrandt 2015) to improve representation of particle fall speeds and microphysical processes. The P3 and NSSL schemes, and some other schemes (e.g., Milbrandt and Morrison 2013), predict the volume mixing ratio of rime to diagnose ice particle density (Table 1). The NSSL scheme diagnoses the density of separate hail and graupel categories. In contrast, the P3 scheme's single general ice category can represent any dominant type of ice, including fully rimed ice (graupel/hail), in its size distribution depending on its diagnostic

values (e.g., rime fraction, bulk density). In the P3 scheme forecast, ice particles are most similar to hail near the surface: ice mass (Fig. 13a) and number concentration (Fig. 13b) are relatively small ($m_{i_tot} < 0.1 \text{ g m}^{-3}$; $N_{i_tot} < 10 \text{ m}^{-3}$), but the mass-weighted mean diameter (Fig. 13c) of ice is relatively large ($D_{mi_tot} > 6 \text{ mm}$). Although D_{mi_tot} exceeds 6 mm near the surface, P3 predicts ice to be smaller than rimed ice in the MY2 or NSSL scheme hail categories (Figs. 8c,f). Small ice particle sizes are in part a consequence of the strict mean size upper limit imposed within the P3 scheme that prevents the formation of larger ice; when the limiter is relaxed the predicted surface ice coverage and size increases (Johnson et al. 2019, manuscript submitted to *Mon. Wea. Rev.*). Fully rimed ice is found within the central storm updraft (Fig. 13d), many observational studies (e.g., Nelson 1983; Heymsfield 1983; Foote 1984; Ziegler et al. 1983) have shown this is primarily where most accretional growth occurs.

Vertical cross sections of hydrometeor density show the NSSL scheme predicts more widespread coverage of high-density rimed ice ($>600 \text{ kg m}^{-3}$) (Figs. 14b,c) than the P3 scheme (Fig. 14a). The P3 scheme predicts ice densities to be mostly less than 300 kg m^{-3} above the 0°C

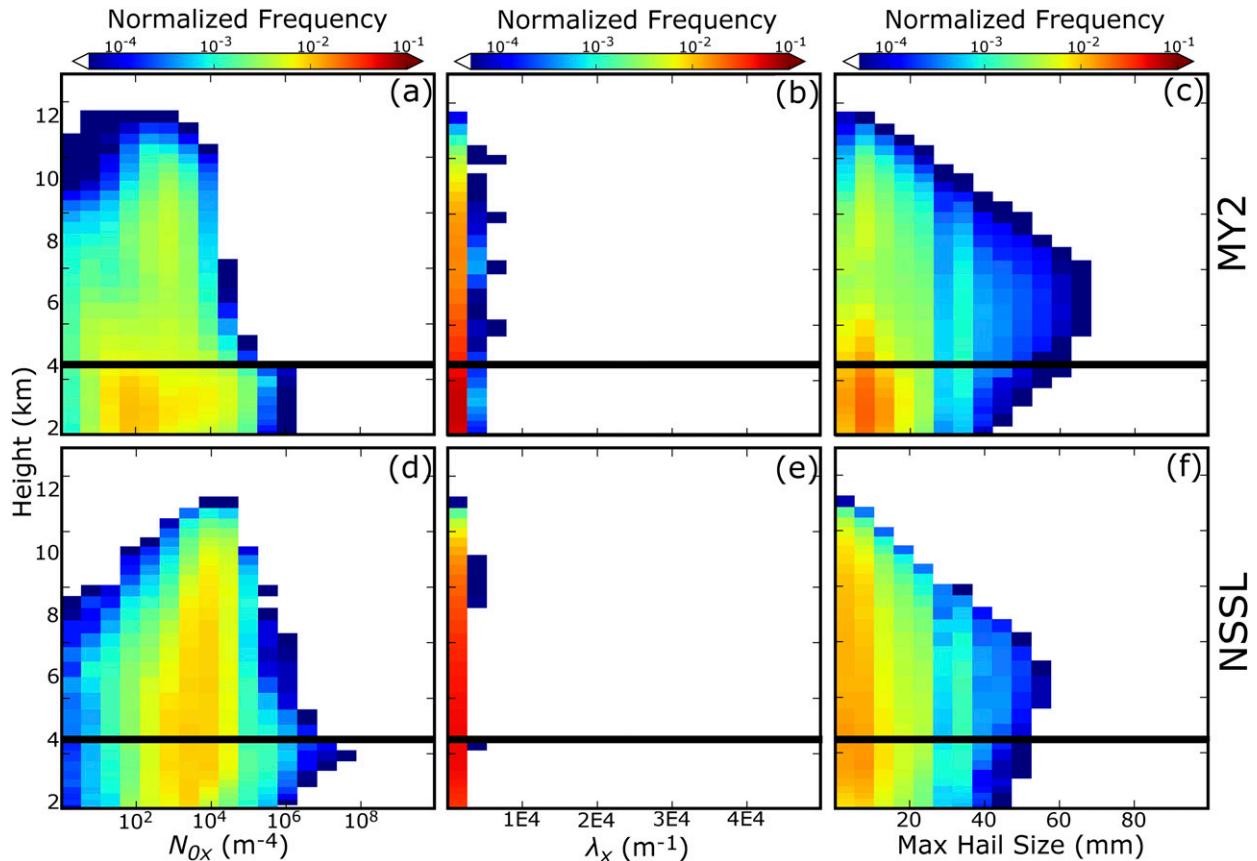


FIG. 9. As in Fig. 6, but for the (a)–(c) MY2 and (d)–(f) NSSL scheme hail categories.

isotherm. In this region large concentrations ($N_{i_{tot}} > 1 \times 10^4 \text{ m}^{-3}$) (Fig. 13b) of low-density particles dilute the properties of high-density hail, closer to the storm updraft (where riming occurs) $\rho_{i_{tot}}$ increases to approximately 500 kg m^{-3} (Fig. 14a). In the NSSL scheme, distinct graupel (Fig. 14c) and hail (Fig. 14b) categories allow the prediction of moderate to high-density ($>400 \text{ kg m}^{-3}$)

rimed ice throughout the entire storm. Due to the presence of rimed ice throughout the storm, NSSL predicts the storm to produce a swath of hail and graupel at the surface exceeding 50 km in width (Figs. 14b,c). Higher-density ice in the P3 scheme is primarily confined to the storm updraft region. Subsequently, P3 predicts the storm to produce a hail swath that is approximately 30 km

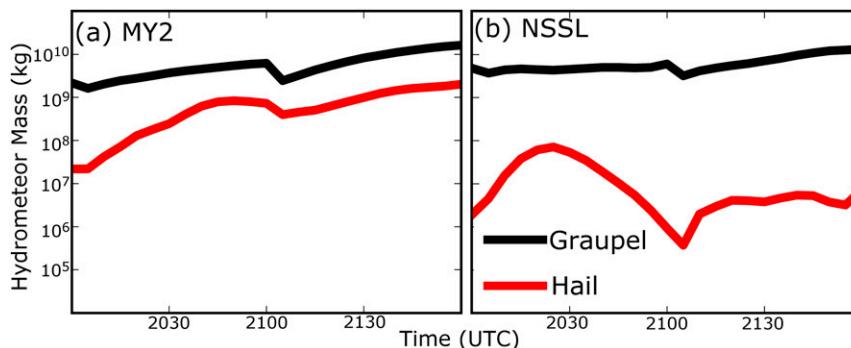


FIG. 10. A time series of the mass of hail and graupel predicted by the (a) MY2 and (b) NSSL schemes from 2000 to 2200 UTC. Values are the sum of all hail and graupel where $UH > 10 \text{ m}^2 \text{ s}^{-2}$ within the three-dimensional domain that is used for verification (Fig. 1).

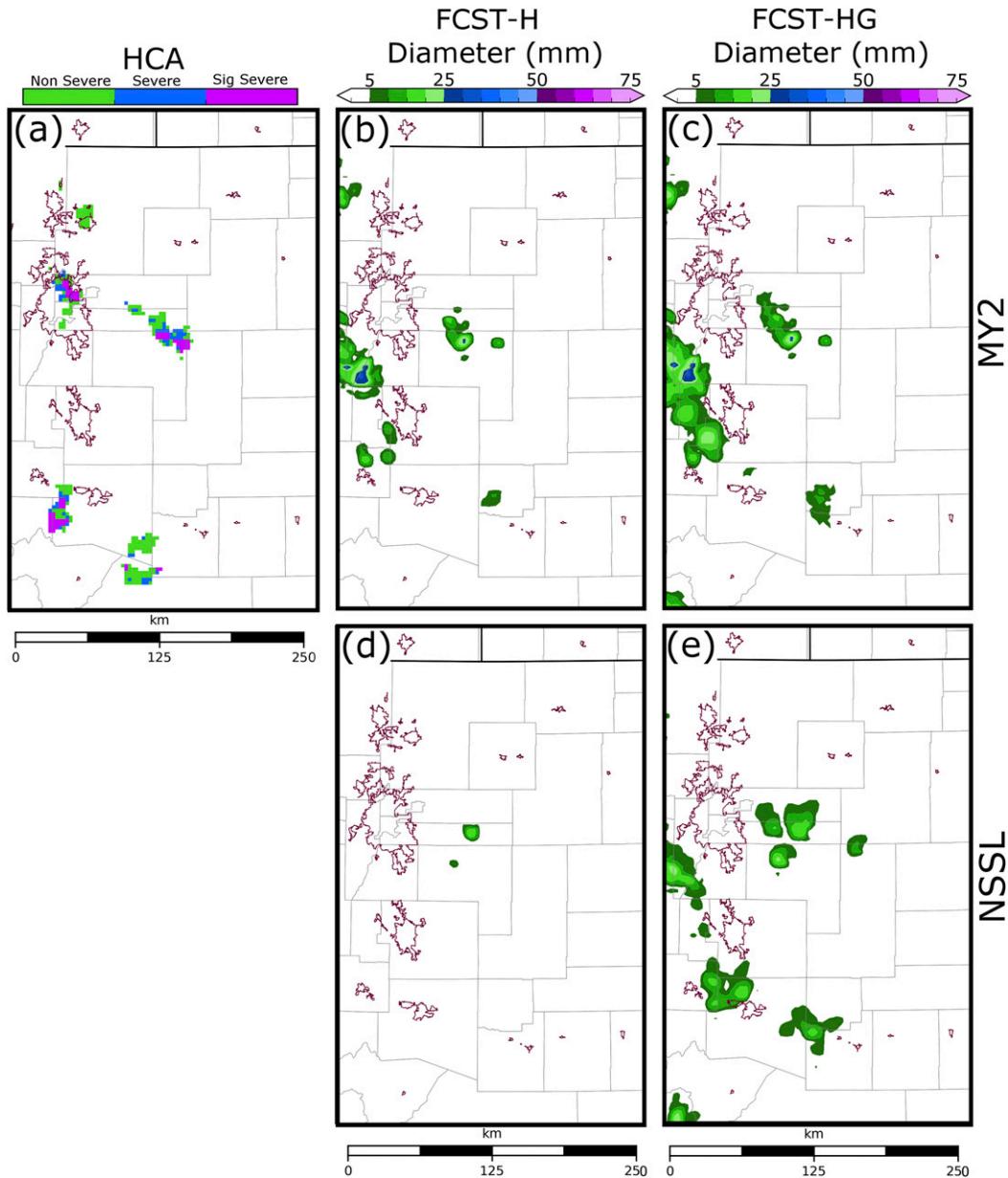


FIG. 11. (a) HCA output at 2130 UTC. (b),(c) MY2 and (d),(e) NSSL forecast maximum size of hail [(b),(d); FCST-H] and hail and graupel [(c),(e); FCST-HG] are provided at the same time. Background is the same as Fig. 1.

in width near the surface (Fig. 14a). The 2130 UTC HCA output (Fig. 11a) indicates that storms in the general vicinity of the cross sections produced surface hail cores that were approximately 30 km in diameter.

CFAD diagrams of rimed ice density (Fig. 15) indicate that P3 and NSSL schemes predict different ice particle densities aloft. With the exception of what is expected to be high-density small ice spheres located more than 10 km above MSL, moderate and high-density ice ($\rho_{i_tot} > 500 \text{ kg m}^{-3}$) is rarely predicted by the P3 forecast (Fig. 15a) because the single ice category is

dominated by large concentrations (Fig. 13b) of low-density (Fig. 14a) ice. The underrepresentation of dense rimed ice may be improved in the P3 scheme through the use of two or more ice categories that would result in reduced mixing/dilution of ice particle types, as shown in Milbrandt and Morrison (2016). In general, the NSSL scheme hail category predicts rimed ice density (Fig. 15b) to be larger than 500 kg m^{-3} , the minimum hail density threshold within the scheme (Mansell et al. 2010). Beneath the 0°C isotherm in the NSSL forecast, the frequency occurrence of high-density hail and

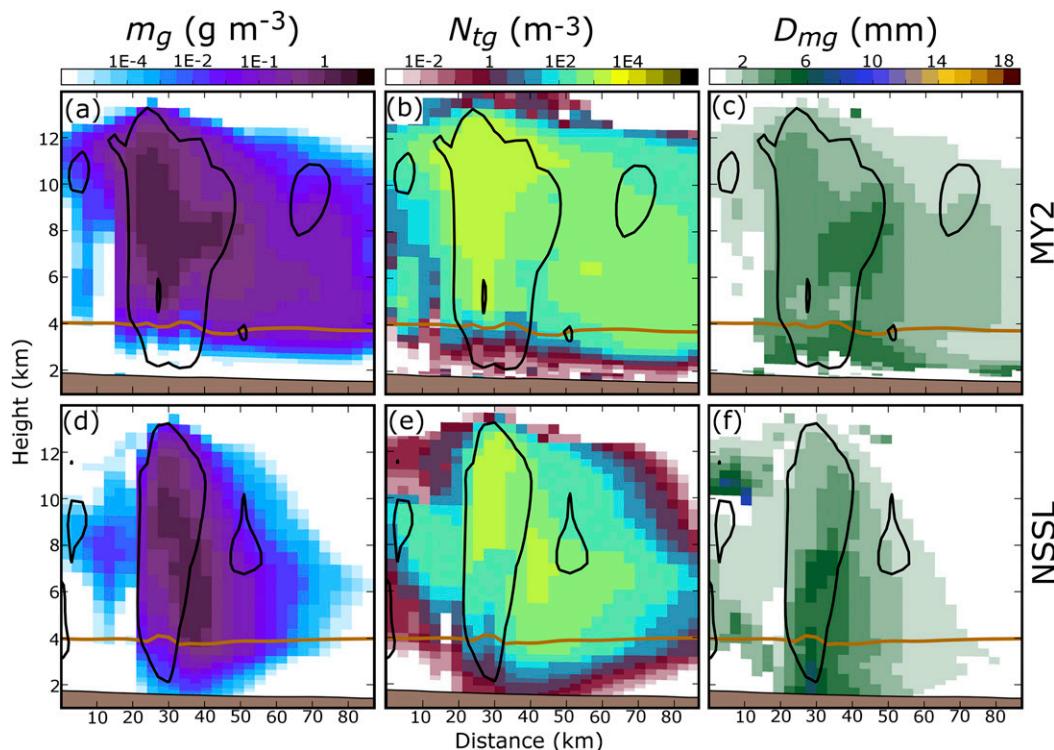


FIG. 12. The same vertical cross sections as Fig. 8. Plotted variables include: (a),(d) m_g , (b),(e) N_{tg} , and (c),(f) D_{mg} .

graupel ($\rho_{h,g} > 900 \text{ kg m}^{-3}$) increases due to particle collapse during melting (Figs. 15b,c).

4. Summary and further discussion

In this study, we analyze 20–22-h hail forecast skill for a severe hail event in Colorado on 8 May 2017, based on 3-km grid spacing model forecasts using six different MP schemes including the Milbrandt–Yau double-moment (MY2), Thompson (THO), NSSL double-moment (NSSL), Morrison double-moment graupel (MOR-G), Morrison double-moment hail (MOR-H), and Predicted Particle Properties (P3) schemes. Subjective comparisons, along with objective verification using fractions skill score (FSS) are used to evaluate both severe and significant severe hail forecasts. Surface hail size forecasts are diagnosed from model output via the Thompson algorithm (THAIL) and storm surrogate fields [i.e., updraft helicity (UH), column integrated total graupel (CTG), and reflectivity Z at the -10°C level]. MP scheme treatment of hail is further analyzed using three-dimensional model output to identify microphysical properties that impact the skill of surface hail size forecasts and how hail-related processes are represented in each scheme.

Objective verification metrics indicate that UH predicts hail greater than 25 mm in diameter (i.e., severe hail)

with a moderate level of skill for this case, confirming the findings of Gagne et al. (2017). While maximum hail size forecasts exhibit more skill at predicting the coverage of severe hail than UH for some forecasts (i.e., MOR-H and MY2); UH is the only parameter to exhibit at least minimally useful predictive skill for all forecasts. This is in part because UH is a storm surrogate field developed for convective-allowing models (CAMs) run at coarse resolutions (3–4-km grid spacing), and identifies storms with strong rotating updrafts that are generally capable of producing severe and significant severe hail. Other storm surrogates that are microphysically based (i.e., CTG, Z at -10°C level) produce less optimal results; forecast skill for these parameters is highly variable and strongly dependent upon choice of MP scheme and the treatment of microphysical processes within.

Surface hail size forecasts are diagnosed from model output via THAIL, and the level of forecast skill is strongly dependent upon choice of MP scheme. All forecasts underpredict the maximum size of hail for this case study; most forecasts predict limited severe hail and none of the forecasts predict significant severe hail despite observations indicating otherwise. Although forecasts underestimate surface hail size, THO, MOR-H, MOR-G, and MY2 schemes predict the coverage of hail greater than 5 mm in diameter, or “all hail sizes” with a high level of skill.

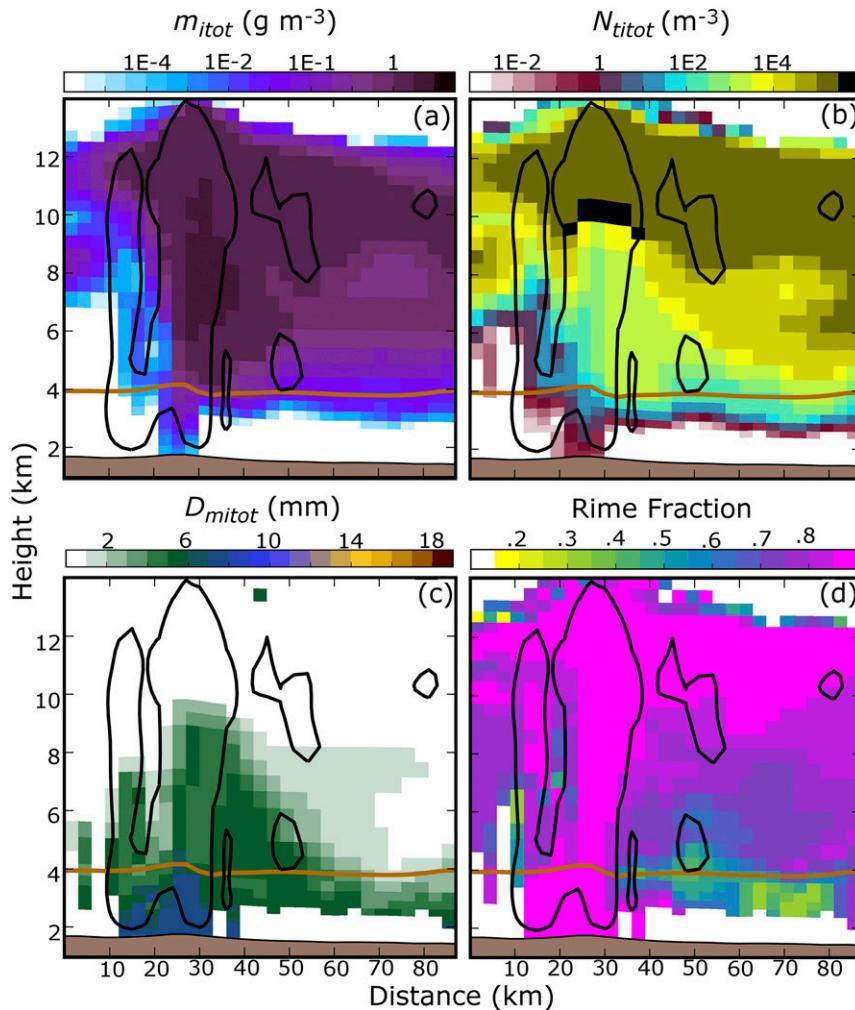


FIG. 13. A vertical cross section of a hail core at 2130 UTC, the location of the cross section is marked in Fig. 3. Vertical cross sections are taken of the P3 predicted bulk ice category. Plotted variables include: (a) $m_{i,tot}$, (b) $N_{i,tot}$, (c) $D_{mi,tot}$, and (d) rime fraction. 1 m s^{-1} updrafts are contoured in black, the brown horizontal line denotes the 0°C isotherm.

A microphysical analysis is conducted in this study to understand how the treatment of rimed ice in MP schemes affects the skill of surface maximum size hail forecasts. The THO scheme predicts a single-moment graupel category and employs diagnostic equations to determine the number concentration. Modifying the graupel number concentration allows the scheme to predict storms that produce large “hail-like” rimed ice. MOR-G and MOR-H also predict a single rimed ice category. Both schemes include the same prognostic equations but assume different static properties for rimed ice (particle density, fall speed). MOR-G and MOR-H predict storms to produce large quantities of relatively small rimed ice aloft; however, excessive size sorting and an increased fall speed causes the latter scheme to produce larger, severe hail at the surface.

MP schemes that predict two rimed ice categories similar to graupel and hail (MY2 and NSSL) are able to represent large rimed ice in the hail category, generally smaller rimed ice remains confined to the graupel category.

The NSSL scheme uses prognostic hail and graupel volume mixing ratios to represent graupel and its product more realistically (i.e., hail is created from dense graupel); however, this multistep hail growth process causes the scheme to produce relatively little rimed ice in the hail category for this case study. Although the scheme predicts relatively little rimed ice in the hail category, the NSSL graupel category includes dense rimed ice. Applying the THAIL algorithm to both the NSSL hail and graupel categories qualitatively improves surface hail size forecast skill by increasing the coverage of surface hail. Although forecast improvements

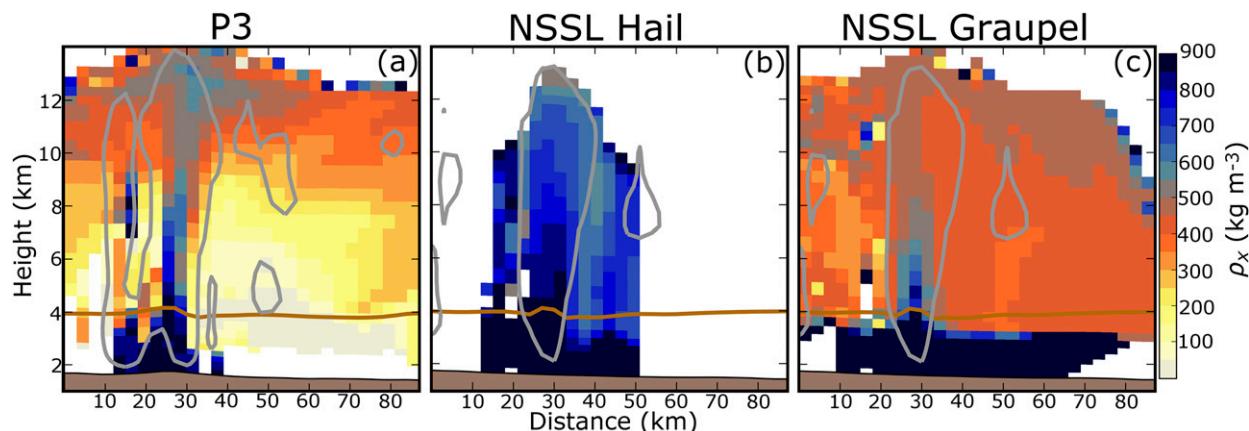


FIG. 14. (a) Vertical cross sections of the bulk ice density predicted by the P3 scheme and (b) hail and (c) graupel predict by the NSSL scheme. The location of the cross sections is provided in Fig. 3. 1 m s^{-1} updrafts are contoured in gray, the brown horizontal line denotes the 0°C isotherm.

are promising, further analysis is required to determine if including both the graupel and hail categories improves surface hail size forecast skill for multimimed ice category schemes (e.g., MY2 and NSSL).

To avoid converting between different ice species the P3 scheme used in this study predicts the evolution of a single ice category. This single ice category is predominately composed of large concentrations of low-density ice particles above the melting layer, and consequently the single ice category configuration of the P3 scheme is unable to represent dense hail aloft. Future studies should analyze the P3 scheme with multiple ice categories (Milbrandt and Morrison 2016) to determine if multiple categories are effective in capturing the full spectrum of ice phase particles found to coexist within convective storms.

Postprocessing calibration of model output has been shown to improve forecast accuracy and reliability. Current surface hail size forecasts skillfully predict the spatial coverage of hail for this case, but the algorithm requires further optimization to improve surface hail size estimates. THAIL should be updated to include the support of newer MP schemes, such as the P3 scheme, and be optimized and/or calibrated to better diagnose maximum surface hail size. The THAIL algorithm is dependent upon an arbitrary minimum observable number concentration threshold that can be adjusted to increase or decrease predicted surface hail sizes. Thresholds can be more realistically determined using observational studies, for example Smith and Waldvogel (1989) evaluated the relationship between observed maximum surface hail size and sampling volume. Although calibrations

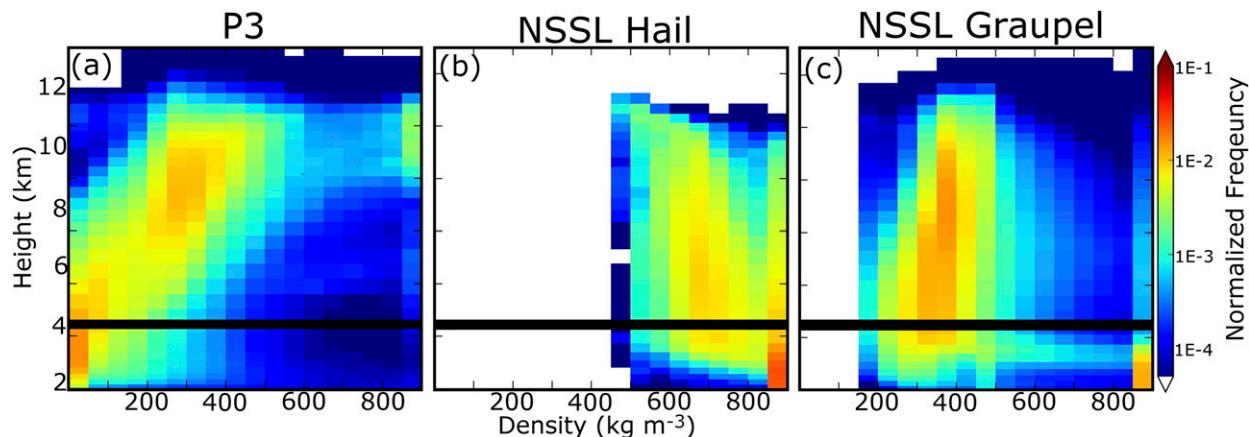


FIG. 15. (a) CFADs for the density of bulk ice predicted by the P3 scheme and (b) hail and (c) graupel by the NSSL scheme during the forecast evaluation period (2000–2200 UTC). Data are considered over the verification domain (Fig. 1). Bins are distributed evenly between 0 and 900 kg m^{-3} , over increments of 50 kg m^{-3} for each diagram. The approximate 0°C isotherm height is plotted with a horizontal black line.

and additional model information could potentially improve surface hail size forecast skill, it is important to note that the skill remains constrained by the accuracy of surface observations.

Additional CAM surface hail size forecasting methods should be analyzed. Model-derived hail proxies such as the maximum estimated size of hail (MESH) have shown skill at predicting surface hail size (e.g., Snook et al. 2016; Labriola et al. 2017; Luo et al. 2017, 2018). Milbrandt and Yau (2006a) noted the surface flux of hail exceeding a critical diameter is important to consider at the surface and provides information about the number of large hailstones to reach the surface, and Luo et al. (2018) examined the prediction of surface accumulated hail number concentration for a real case. Machine learning algorithms applied to CAM forecast output represent another way of using environmental, dynamic, and microphysical information to calibrate surface hail size forecasts. A storm-based machine learning approach developed by Gagne et al. (2017) improves the skill of surface hail size forecasts by reducing the number of false alarm events and increasing forecast reliability. Additional research using machine learning algorithms to calibrate CAM forecasts has the potential to further improve day-ahead surface hail size predictions beyond the skill of current storm surrogate and explicit hail forecasting methods, and is being pursued by this current research group.

Conclusions drawn from this study are limited because this is a single case study and hail forecast skill is expected to vary substantially from case to case. While only one case study, results provide insight into hail diagnostic parameters and the impact of microphysical assumptions on the representation of rimed ice within several multimoment MP schemes. Analyzing forecasts run for a diverse set of well observed hail events is necessary to draw more significant results.

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. Forecast background and initial conditions were generated under the support of NOAA Grants NA15OAR4590186 and NA16NWS4680002. The SHARP team includes David Gagne, Amy McGovern, Youngsun Jung, Amanda Burke, and coauthors Jonathan Labriola, Nathan Snook, and Ming Xue. Computing was performed primarily on the XSEDE Stampede2 supercomputer at the University of Texas Advanced Computing Center (TACC). Gregory Thompson and two anonymous scientists reviewed and substantially improved the quality of this manuscript. The authors

thank dissertation committee members Youngsun Jung, Cameron Homeyer, Amy McGovern, and Phillip Scott Harvey, along with Ted Mansell who provided useful feedback on the project analysis.

REFERENCES

- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Blair, S. F., and Coauthors, 2017: High-Resolution hail observations: Implications for NWS warning operations. *Wea Forecasting*, **32**, 1101–1119, <https://doi.org/10.1175/WAF-D-16-0203.1>.
- Blanchard, D. O., and K. W. Howard, 1986: The Denver Hailstorm of 13 June 1984. *Bull. Amer. Meteor. Soc.*, **67**, 1123–1131, [https://doi.org/10.1175/1520-0477\(1986\)067<1123:TDHOJ>2.0.CO;2](https://doi.org/10.1175/1520-0477(1986)067<1123:TDHOJ>2.0.CO;2).
- Changnon, S. A., 2009: Increasing major hail losses in the U.S. *Climatic Change*, **9**, 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>.
- Chong, S.-L., and C. S. Chen, 1974: Water shells on ice pellets and hailstones. *J. Atmos. Sci.*, **31**, 1384–1391, [https://doi.org/10.1175/1520-0469\(1974\)031<1384:WSOIPA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<1384:WSOIPA>2.0.CO;2).
- Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail climatology of the contiguous United States. *Wea. Forecasting*, **27**, 1235–1248, <https://doi.org/10.1175/WAF-D-11-00151.1>.
- Clark, A. J., and Coauthors, 2012a: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, <https://doi.org/10.1175/BAMS-D-11-00040.1>.
- , J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012b: 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>.
- , J. Gao, P. T. Marsh, T. M. Smith, J. S. Kain, J. Correia Jr., M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387–407, <https://doi.org/10.1175/WAF-D-12-00038.1>.
- Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF Model output for severe weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*, **25**, 408–427, <https://doi.org/10.1175/2009WAF2222258.1>.
- 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).
- Davis, S. M., and J. G. LaDue, 2004: Nonmeteorological factors in warning verification. *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., P2.7, https://ams.confex.com/ams/11aram22sls/techprogram/paper_81766.htm.
- Dawson, D. T., II, M. Xue, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms.

- Mon. Wea. Rev.*, **138**, 1152–1171, <https://doi.org/10.1175/2009MWR2956.1>.
- , 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>.
- 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>.
- 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).
- Footo, 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*, accessed 25 May 2017, https://www.washingtonpost.com/news/capital-weather-gang/wp/2017/05/25/remember-that-crazy-hail-storm-in-denver-its-going-to-be-colorados-costliest-catastrophe/?utm_term=.1d3f823248f4.
- Gagne, D. J., II, A. McGovern, J. Brotzge, M. Coniglio, J. Correia Jr., and M. Xue, 2015: Day-ahead hail prediction integrating machine learning with storm-scale numerical weather models. *27th Conf. on Innovative Applications of Artificial Intelligence*, Austin, TX, Amer. Meteor. Soc., 3954–3960.
- , —, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, <https://doi.org/10.1175/WAF-D-17-0010.1>.
- , 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., and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting*, **32**, 1541–1568, <https://doi.org/10.1175/WAF-D-16-0178.1>.
- Garcia-Garcia, F., and R. List, 1992: Laboratory measurements and parameterizations of supercooled water skin temperatures and bulk properties of gyrating hailstones. *J. Atmos. Sci.*, **49**, 2058–2073, [https://doi.org/10.1175/1520-0469\(1992\)049<2058:LMAPOS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<2058:LMAPOS>2.0.CO;2).
- Gilmore, M. S., J. M. Straka, and E. N. Rasmussen, 2004: Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. *Mon. Wea. Rev.*, **132**, 2610–2627, <https://doi.org/10.1175/MWR2810.1>.
- Heymsfield, A. J., 1978: The characteristics of graupel particles in northeastern Colorado cumulus congestus clouds. *J. Atmos. Sci.*, **35**, 284–295, [https://doi.org/10.1175/1520-0469\(1978\)035<1689:OOMAAI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<1689:OOMAAI>2.0.CO;2).
- , 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).
- Hu, M., M. Xue, and K. Brewster, 2006a: 3DVAR and cloud analysis with WSR-88D Level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675–698, <https://doi.org/10.1175/MWR3092.1>.
- , —, J. Gao, and K. Brewster, 2006b: 3DVAR and cloud analysis with WSR-88D Level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134**, 699–721, <https://doi.org/10.1175/MWR3093.1>.
- Jewell, R., and J. Brimelow, 2009: Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States. *Wea. Forecasting*, **24**, 1592–1609, <https://doi.org/10.1175/2009WAF2222230.1>.
- Jirak, I. L., A. J. Clark, B. Roberts, B. Gallo, and S. J. Weiss, 2018: Exploring the optimal configuration of the High Resolution Ensemble Forecast System. *25th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 14B.6, <https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345640.html>.
- Johnson, M., Y. Jung, D. T. I. Dawson, and M. Xue, 2016: Comparison of simulated polarimetric signatures in idealized supercell storms using two-moment bulk microphysics schemes in WRF. *Mon. Wea. Rev.*, **144**, 971–996, <https://doi.org/10.1175/MWR-D-15-0233.1>.
- Joyce, C. A., and Coeditors, 2008: *The World Almanac and Book of Facts: U.S. Cities, States, and Population*. World Almanac Books, 1008 pp.
- Jung, Y., M. Xue, and G. Zhang, 2010: Simulations of polarimetric radar signatures of a supercell storm using a two-moment bulk microphysics scheme. *J. Appl. Meteor. Climatol.*, **49**, 146–163, <https://doi.org/10.1175/2009JAMC2178.1>.
- , —, G. Zhao, F. Kong, K. Thomas, T. Supinie, K. Brewster, and N. Snook, 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, Québec, 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>.
- Khain, A. P., and Coauthors, 2015: Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization. *Rev. Geophys.*, **53**, 247–322, <https://doi.org/10.1002/2014RG000468>.
- Knight, N. C., and A. J. Heymsfield, 1983: Measurement and interpretation of hailstone density and terminal velocity. *J. Atmos. Sci.*, **40**, 1510–1516, [https://doi.org/10.1175/1520-0469\(1983\)040<1510:MAIOHD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<1510:MAIOHD>2.0.CO;2).
- Knight, C. A., W. A. Cooper, D. W. Breed, I. R. Paluch, P. L. Smith, and G. Valu, 1982: Microphysics. *Hailstorms of the Central High Plains*, C. Knight and P. Squires, Eds., Colorado Associated University Press, 151–193.
- , P. T. Schlatter, and T. W. Schlatter, 2008: An unusual hailstorm on 24 June 2006 in Boulder, Colorado. Part II: Low-density growth of hail. *Mon. Wea. Rev.*, **136**, 2833–2848, <https://doi.org/10.1175/2008MWR2338.1>.
- Kong, F., 2017: 2017 CAPS Spring Forecast Experiment Program Plan. Center for the Analysis and Prediction of Storms, 28 pp., forecast.caps.ou.edu/SpringProgram2017_Plan-CAPS.pdf.

- , and Coauthors, 2007: Preliminary analysis on the real-time storm-scale ensemble forecasts produced as a part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 3B.2, https://ams.confex.com/ams/22WAF18NWP/techprogram/paper_124667.htm.
- 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, 2019: 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>.
- Lesins, G. B., and R. List, 1986: Sponginess and drop shedding of gyrating hailstones in a pressure-controlled icing wind tunnel. *J. Atmos. Sci.*, **43**, 2813–2825, [https://doi.org/10.1175/1520-0469\(1986\)043<2813:SADSOG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1986)043<2813:SADSOG>2.0.CO;2).
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2).
- Loken, E. D., A. J. Clark, M. Xue, and F. Kong, 2017: Comparison of next-day probabilistic severe weather forecasts from coarse- and fine-resolution CAMs and a convection-allowing ensemble. *Wea. Forecasting*, **32**, 1403–1421, <https://doi.org/10.1175/WAF-D-16-0200.1>.
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21A**, 289–307, <https://doi.org/10.1111/j.2153-3490.1969.tb00444.x>.
- Luo, L., M. Xue, K. Zhu, and B. Zhou, 2017: Explicit prediction of hail using multimoment microphysics schemes for a hailstorm of 19 March 2014 in eastern China. *J. Geophys. Res. Atmos.*, **122**, 7560–7581, <https://doi.org/10.1002/2017JD026747>.
- , —, —, and —, 2018: Explicit prediction of hail in a long-lasting multi-cellular convective system in eastern China using multimoment microphysics schemes. *J. Atmos. Sci.*, **75**, 3115–3137, <https://doi.org/10.1175/JAS-D-17-0302.1>.
- Mansell, E. R., 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>.
- Meyers, M. P., P. J. DeMott, and W. R. Cotton, 1992: New primary ice-nucleation parameterizations in an explicit cloud model. *J. Appl. Meteor.*, **31**, 708–721, [https://doi.org/10.1175/1520-0450\(1992\)031<0708:NPINPI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1992)031<0708:NPINPI>2.0.CO;2).
- Milbrandt, J. A., 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>.
- , and M. K. Yau, 2005a: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064, <https://doi.org/10.1175/JAS3534.1>.
- , and —, 2005b: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62**, 3065–3081, <https://doi.org/10.1175/JAS3535.1>.
- , and —, 2006a: A multimoment bulk microphysics parameterization. Part III: Control simulation. *J. Atmos. Sci.*, **63**, 3114–3136, <https://doi.org/10.1175/JAS3816.1>.
- , and —, 2006b: A multimoment bulk microphysics parameterization. Part IV: Sensitivity experiments. *J. Atmos. Sci.*, **63**, 3137–3159, <https://doi.org/10.1175/JAS3817.1>.
- Miller, L. J., J. D. Tuttle, and C. A. Knight, 1988: Airflow and hail growth in a severe northern High Plains supercell. *J. Atmos. Sci.*, **45**, 736–762, [https://doi.org/10.1175/1520-0469\(1988\)045<0736:AAHGIA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<0736:AAHGIA>2.0.CO;2).
- Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Weather*, **25**, 343–354, <https://doi.org/10.1175/2009WAF2222260.1>.
- Modahl, A. C., 1979: Synoptic parameters as discriminators between hailfall and less significant convective activity in northeast Colorado. *J. Appl. Meteor.*, **18**, 671–681, [https://doi.org/10.1175/1520-0450\(1979\)018<0671:SPADBH>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<0671:SPADBH>2.0.CO;2).
- 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. Milbrandt, 2011: Comparison of two-moment bulk microphysics schemes in idealized supercell thunderstorm simulations. *Mon. Wea. Rev.*, **139**, 1103–1130, <https://doi.org/10.1175/2010MWR3433.1>.
- , and —, 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. A. 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. Tessoroff, 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>.
- NCEI, 2017: U.S. billion-dollar weather and climate disasters: Overview. National Centers for Environmental Information, accessed 2 August 2018, <https://www.ncdc.noaa.gov/billions/>.
- 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).
- 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. Zrníc, 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>.
- Phillips, V. T., A. Khain, N. Benmoshe, and E. Ilotoviz, 2014: Theory of time-dependent freezing. Part I: Description of scheme for wet growth of hail. *J. Atmos. Sci.*, **71**, 4527–4557, <https://doi.org/10.1175/JAS-D-13-0375.1>.
- Putnam, B. J., M. Xue, Y. Jung, G. Zhang, and F. Kong, 2017: Simulation of polarimetric radar variables from 2013 CAPS Spring Experiment storm-scale ensemble forecasts and evaluation of microphysics schemes. *Mon. Wea. Rev.*, **145**, 49–73, <https://doi.org/10.1175/MWR-D-15-0415.1>.
- Rasmussen, R. M., and A. J. Heymsfield, 1987: Melting and shedding of graupel and hail. Part I: Model physics. *J. Atmos. Sci.*, **44**, 2754–2763, [https://doi.org/10.1175/1520-0469\(1987\)044<2754:MASOGA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1987)044<2754:MASOGA>2.0.CO;2).
- Roberts, N., and H. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- 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>.
- Rutledge, S. A., and P. Hobbs, 1983: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. VIII: A model for the “seeder-feeder” process in warm-frontal rainbands. *J. Atmos. Sci.*, **40**, 1185–1206, [https://doi.org/10.1175/1520-0469\(1983\)040<1185:TMAMSA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<1185:TMAMSA>2.0.CO;2).
- Ryzhkov, A. V., M. R. Kumjian, S. M. Ganson, and P. Zhang, 2013: Polarimetric radar characteristics of melting hail. Part II: Practical implications. *J. Appl. Meteor. Climatol.*, **52**, 2871–2886, <https://doi.org/10.1175/JAMC-D-13-074.1>.
- 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>.
- Smith, P. L., and A. Waldvogel, 1989: On determinations of maximum hailstone sizes from hailpad observations. *J. Appl. Meteor.*, **28**, 71–76, [https://doi.org/10.1175/1520-0450\(1989\)028<0071:ODOMHS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1989)028<0071:ODOMHS>2.0.CO;2).
- Smith, T. L., S. G. Benjamin, J. M. Brown, S. Weygandt, T. Smirnova, and B. Schwartz, 2008: Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh (HRRR) model. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 11.1, https://ams.confex.com/ams/24SLS/techprogram/paper_142055.htm.
- Snook, N., 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*, **4**, 811–825, <https://doi.org/10.1175/WAF-D-15-0152.1>.
- Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728, <https://doi.org/10.1175/WAF-D-10-05046.1>.
- , C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, <https://doi.org/10.1175/WAF-D-15-0138.1>.
- Szoke, E. J., M. L. Weisman, J. M. Brown, F. Caracena, and T. W. Schlatter, 1984: A subsynoptic analysis of the Denver tornadoes of 3 June 1981. *Mon. Wea. Rev.*, **112**, 790–808, [https://doi.org/10.1175/1520-0493\(1984\)112<0790:ASAOTD>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<0790:ASAOTD>2.0.CO;2).
- Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542, [https://doi.org/10.1175/1520-0493\(2004\)132<0519:EFOWPU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2).
- , 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>.
- , S. Tessendorf, A. Heymsfield, and I. Kyoko, 2018: Improving graupel and hail treatment in the Thompson microphysics parameterization. *North American Hail Workshop*, Boulder, CO, NCAR, <https://www.mmm.ucar.edu/north-american-hail-workshop>.
- 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).
- Wacker, U., and A. Seifert, 2001: Evolution of rain water profiles resulting from pure sedimentation: Spectral vs. parameterized description. *Atmos. Res.*, **58**, 19–39, [https://doi.org/10.1016/S0169-8095\(01\)00081-3](https://doi.org/10.1016/S0169-8095(01)00081-3).
- Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527–548, [https://doi.org/10.1175/1520-0493\(1997\)125<0527:TRDOEM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2).
- Wyatt, A., and A. Witt, 1997: The effect of population density on ground-truth verification of reports used to score a hail detection algorithm. *28th Conf. on Radar Meteorology*, Austin, TX, Amer. Meteor. Soc., P3.17.
- Xue, M., and Coauthors, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 3B.1, https://ams.confex.com/ams/22WAF18NWP/techprogram/paper_124587.htm.
- Yuter, S. E., and R. A. Houze Jr., 1995: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part II: Frequency distributions of vertical velocity, reflectivity, and differential reflectivity. *Mon. Wea. Rev.*, **123**, 1941–1963, [https://doi.org/10.1175/1520-0493\(1995\)123<1941:TDKAME>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<1941:TDKAME>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:HGIAOM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1983)040<1768:HGIAOM>2.0.CO;2).