Simulation of Polarimetric Radar Variables from 2013 CAPS Spring Experiment Storm Scale Ensemble Forecasts and Evaluation of Microphysics Schemes

Bryan J. Putnam[1, 2, 3], Ming Xue[1, 2], Youngsun Jung[1], Guifu Zhang[2, 3], and Fanyou Kong[1]

[1] Center for Analysis and Prediction of Storms, Norman, Oklahoma
[2] Advanced Radar Research Center, Norman, Oklahoma
[3] School of Meteorology, University of Oklahoma, Norman, Oklahoma 73072

Submitted to Monthly Weather Review
November 2015
Accepted September 2016

Corresponding author address:
Ming Xue
Center for Analysis and Prediction of Storms
University of Oklahoma,
120 David Boren Blvd., Norman, OK 73072
mxue@ou.edu

Abstract

Polarimetric radar variables are simulated from members of the 2013 CAPS Storm Scale Ensemble Forecasts (SSEF) with varying microphysics (MP) schemes and compared with observations. The polarimetric variables provide information on hydrometeor types and particle size distributions (PSDs), neither of which can be obtained through reflectivity (Z) alone. The polarimetric radar simulator pays close attention to how each MP scheme (including single- (SM) and double-moment (DM) schemes) treats hydrometeor types and PSDs. The recent dual-polarization upgrade to the entire WSR-88D network provides nation-wide polarimetric observations, allowing for direct evaluation of the simulated polarimetric variables.

Simulations for a mesoscale convective system (MCS) and supercell cases are examined. Five different microphysics schemes: Thompson, DM Milbrandt and Yau (MY), DM Morrison, WRF DM 6 category (WDM6) and WRF SM 6 category (WSM6) are used in the ensemble forecasts. Forecasts using the partially DM Thompson, and fully DM MY and Morrison schemes better replicate the MCS structure and stratiform precipitation coverage as well as supercell structure compared to WDM6 and WSM6. Forecasts using the MY and Morrison schemes better replicate observed polarimetric signatures associated with size sorting than those using the Thompson, WDM6, and WSM6 schemes, in which such signatures are either absent or occur at abnormal locations. Several biases are suggested in these schemes, including too much wet graupel in MY, Morrison, and WDM6; a small raindrop bias in WDM6 and WSM6; and the under-forecast of liquid water content in regions of pure rain for all schemes.

1. Introduction

The national WSR-88D S-band weather radar network has completed its polarimetric upgrade, providing unprecedented polarimetric radar variable measurements over the CONUS (ROC 2013). The polarimetric radar variables provide additional information about the cloud hydrometeor types and their particle size distributions (PSDs) compared to reflectivity (Z), in particular information on hydrometeor size and diversity. They include: 1) differential reflectivity (Z_{DR}) that is sensitive to hydrometeor shape, orientation, and phase; 2) specific differential phase (K_{DP}) that is sensitive to rainwater content/rain rate; and 3) cross-correlation coefficient (ρ_{hv}) that is sensitive to diverse and mixed-phase hydrometeors (Bringi and Chandrasekar 2001). Common dynamical and microphysical processes lead to patterns in these variables which occur at specific locations and in specific circumstances within convective storms, referred to as polarimetric signatures (Kumjian and Ryzhkov 2008). For example, there is a relative Z_{DR} maxima along the right forward flank of supercells due to hydrometeor size sorting, known as the Z_{DR} arc. In mesoscale convective systems (MCS), high Z_{DR} is observed on the leading edge of the convective line due to the size sorting of larger drops that fall ahead of the system (Park et al. 2009).

The hydrometeor variables in microphysics (MP) schemes of numerical weather prediction (NWP) models, such as mixing ratio (q), are typically not directly observed. One way to evaluate the model prediction of hydrometeor fields and the MP parameterization schemes is to simulate polarimetric variables from the model output and compare them with observations. The model state variables, including MP variables, are connected to observed polarimetric fields by the so-called polarimetric radar data simulator (PRDS) (Jung et al. 2008a; Jung et al. 2010, hereafter JXZ10), or the observation operators in data assimilation terminology. These operators are derived from scattering calculations of polarized radar radio waves by hydrometeor particles within each radar sampling volume.

Most MP schemes represent hydrometeor PSDs in bulk form using the simplified gamma distribution

$$N(D)_x = N_{0x} D_x^{\alpha_x} e^{(-\Lambda_x D)} \quad , \tag{1}$$

which defines the number of particles of hydrometeor x with diameter D in a unit volume (Ulbrich 1983; Milbrandt and Yau 2005b). Three free parameters govern the distribution: 1) the slope parameter Λ_x; 2) the intercept parameter N_{0x}; and 3) the shape parameter α_x. MP schemes can be broadly categorized by the number of these free parameters that they derive from predicted microphysical variables for each species. For example, q is proportional to the third PSD moment (mass) and is used to solve for Λ_x. Single-moment (SM), double-moment (DM), and triple-moment (TM) schemes predict one, two, and three moments of the PSD and can therefore determine one, two or three of the PSD parameters, respectively. Parameters that are not derived from predicted variables are either diagnosed or set as constant. Another significant feature of a given MP scheme is the number of hydrometeor species included. Five categories are most commonly considered in ice MP schemes: cloud water (c), cloud ice (i), rain water (r), snow (s), and graupel (g) or hail (h), and some but relatively few schemes (e.g., the Milbrandt and Yau (MY) scheme (Milbrandt and Yau 2005a)) include graupel and hail as separate species.

It is important that the observation operators developed for a PRDS are consistent with the MP scheme so that the simulated variables reflect the model microphysical state and dynamical processes. Increasing the number of model variables predicted (e.g., moving from a SM to a DM scheme) increases the amount of predicted microphysical information that can and should be used in the operators. Some schemes, including Thompson (Thompson et al. 2008) and WDM6 (Lim and

Hong 2010) schemes are partially double moment, predicting a second moment for rain (number concentration, N_{tr}) but only one moment for other hydrometeor species. Though most SM and DM schemes set $\alpha_x = 0$ by default, resulting in an exponential distribution, WDM6 uses $\alpha_r = 1$ for rain and the Thompson scheme uses a combined exponential and gamma distribution for snow.

Limitations of MP schemes may preclude the model from replicating certain polarimetric signatures and highlight microphysical state differences. Some current schemes, including the Thompson, WSM6 (Hong and Lim 2006), WDM6, and Morrison (Morrison et al. 2005; Morrison et al. 2009) schemes, contain a graupel category but not hail[1]. In a supercell simulation experiment using these schemes by Johnson et al. (2016), the hail signature in the forward flank downdraft, a decrease in Z_{DR} associated with large, dry hail (Kumjian and Ryzhkov 2008), was not replicated by those schemes that only include a graupel category due to the small size of and limited amount of graupel present near the surface and associated high rainwater content. Additionally, Wacker and Seifert (Wacker and Seifert 2001) and Milbrandt and Yau (2005b) have shown that SM MP schemes cannot represent sedimentation, or size sorting, and thus a DM or higher order scheme is required to produce polarimetric signatures associated with size sorting (JXZ10; Kumjian and Ryzhkov 2012). Jung et al. (2012) demonstrated that the Z_{DR} arc signature could be replicated with the DM MY scheme but not a SM Lin (Lin et al. 1983) scheme when the states of a supercell are estimated using a cycled ensemble Kalman filter. Putnam et al. (2014) showed for an MCS case that the size sorting of large drops and subsequent increase in Z_{DR} in the convective line compared to the stratiform region could be replicated by the DM MY scheme but not the SM Lin scheme.

Since the spring of 2007, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma has been producing storm-scale ensemble forecasts (SSEF) for the CONUS (Kong et al. 2007; Xue et al. 2007) as part of the NOAA Hazardous Weather Testbed (HWT) Spring Experiment (Weiss et al. 2007; Clark et al. 2012). In the spring of 2013, the SSEF system had a 4 km convection-permitting grid spacing, allowing explicit representation of convective storms (Kong 2013). The system used a variety of MP schemes among its ensemble members, including several that predicted two moments of some of the hydrometeor species within the schemes. One set of special products produced from the ensemble forecast output were simulated polarimetric radar variables using the PRDS developed at CAPS. The availability of polarimetric observations from the upgraded WSR-88D network and the PRDS with the ability to simulate polarimetric variables from a variety of microphysics schemes provided an unprecedented opportunity to compare and contrast the ability of the various microphysics schemes commonly used in convective-scale forecasts in reproducing known polarimetric signatures.

The purpose of this paper is to document the real-time implementation of the PRDS, evaluate the simulated polarimetric variables and polarimetric variable forecasts within the SSEF system against WSR-88D polarimetric observations, and to infer strengths and weaknesses of MP schemes as implemented in the 2013 ensemble. Biases identified with the microphysics schemes can help the scheme developers to improve their schemes, and help the scheme users to interpret their simulation results in their research. As observed quantities are often more intuitive to forecasters, simulated polarimetric variables can be used by forecasters to monitor and nowcast severe weather when the association of polarimetric signatures with weather events/features is well recognized. Towards that

[1] The rimed ice category in the Morrison scheme can be switched to represent either graupel or hail. In this study it was represented as graupel.

end, knowledge gained on the behavior of the MP schemes and the PRDS can help forecasters better understand the dual-pol forecast products.

Up to the time of this exercise, CAPS's PRDS had mainly been used with the MY DM scheme (Jung et al. 2012; Putnam et al. 2014); this effort represents the first time that multiple DM schemes are evaluated in a common framework in terms of their ability to produce polarimetric radar signatures for real cases. The recent study of Johnson et al. (2016) had a similar goal but it was based on a set of idealized supercell simulations therefore no real radar observations could be used for comparison. Since significant polarimetric radar signatures are relatively local and isolated within convective systems, simple grid-point-based evaluation scores typically applied to precipitation forecasts, such as the equitable threat score, are not very revealing, especially when different types of convective systems are mixed together (different types of convective systems tend to produce different kinds of polarimetric signatures in different parts of the systems). Because of the many challenges facing objective evaluations of the forecast of polarimetric signatures, which tend to be highly localized and in the current forecasts contain significant biases, we choose to focus on two cases from the CAPS 2013 Spring Experiment only in this study: one case with MCSs and one case with supercells. Focusing on two cases allows us to perform more detailed subjective evaluations and at the same time access the objective evaluation methods and procedures. Such a study would provide the groundwork for future studies evaluating the MP and PRDS performances over the entire Spring Experiment period.

The rest of this paper is organized as follows. The methodologies, including the general design of the CAPS SSEF, the PRDS and MP schemes used, and the quality control of observations, are given in section 2. Section 3 presents evaluation results for the MCS and supercell cases. A summary and conclusion reviewing notable trends in polarimetric size sorting signatures and biases in graupel and water content is given in section 4. Some challenges faced in evaluation are also noted.

2. Methodology

a. Overview of the 2013 CAPS SSEF

The 2013 CAPS SSEF forecasts were run as part of the NOAA HWT Spring Experiment (Kong 2013). Official forecasts began on 6 May 2013 and continued through 7 June 2013. Daily 48-hour forecasts initialized at 0000 UTC were run on a CONUS domain using a horizontal grid spacing of 4 km with 51 vertical levels (Fig. 1). Twenty nine ensemble members were run using three mesoscale NWP models: the WRF-ARW model (version 3.4.1, 26 members, Skamarock et al. 2008), the U.S. Navy COAMPS model (2 members, Hodur 1997), and the CAPS Advanced Regional Prediction System (ARPS, version 5.3, 1 member, Xue et al. 2003). This paper focuses on the WRF-ARW members since the COAMPS and ARPS members used single-moment microphysics schemes only.

The 26 WRF-ARW members varied in terms of their initial conditions (IC), boundary conditions (BC), and physics packages. The control member IC was obtained by assimilating surface, upper-air, and WSR-88D radar observations using the ARPS 3DVAR and complex cloud analysis system (Xue et al. 2003; Gao et al. 2004; Hu et al. 2006a,b), with the NCEP 12 km NAM (Rogers et al. 2009) 0000 UTC analysis used as the background, and used a BC that was obtained from the 0000 UTC NAM forecast. An additional 11 members used this IC and BC while 13 members used this IC and BC with added perturbations derived from the 2100 UTC NCEP Short-Range Ensemble Forecasts (SREF, Du et al. 2006). One member was initialized from the NAM analysis directly. For the purpose

of investigating the performance of various physics packages in the WRF, the subset of members which used the same IC and BC as the control member differed in their use of land surface, boundary layer, radiation, and microphysics schemes. Since MP scheme differences are the focus of this study, polarimetric variable simulations were performed for the control and for those members that differed from the control only in their choice of MP scheme (the arw_cn, arw_m20, arw_m21, arw_m22, and arw_m26 members) (Kong 2013). These members used the NOAH land surface model (Chen and Dudhia 2001) and the Mellor–Yamada–Janjic boundary layer scheme (MYJ, Mellor and Yamada 1982; Janjic 2002). More details on the MP schemes used are provided in section 2c.

b. Polarimetric simulation and general experiment settings

The PRDS originally developed for the ARPS output (Jung et al. 2008a; JXZ10) was adapted and applied to the WRF output with several different MP schemes. The PRDS calculations include only the rain, snow, graupel, and hail categories, when applicable. Despite the important role that cloud water and cloud ice play in precipitation processes, the radar returns from these hydrometeors are minimal. Important details of the PRDS, including the axis-ratio relation, canting angle of particles, the melting model, and radar scattering amplitudes, are briefly summarized here.

The PRDS operators include complex scattering amplitudes calculated using the T-matrix method (Vivekanandan et al. 1991; Bringi and Chandrasekar 2001) for both rain and ice species via numerical integration over the PSDs. The raindrop axis ratio decreases with diameter based on the relation in Brandes et al. (2002); this ratio is set to 0.75 for hail, graupel, and snow. The mean canting angle for all hydrometeor types is 0° with a standard deviation of 0° for rain, 20° for snow, and ranging from 0° to 60° for hail and graupel depending on the water fraction. Since most MP schemes do not predict mixed-phase hydrometeors, a mixing ratio fraction of wet (melting) snow, wet hail, or wet graupel is considered present when rain (q_r) coexists at a particular model grid point with snow (q_s), hail (q_h), or graupel (q_g), creating mixed-phase mixing ratios denoted q_{rs}, q_{rh} and q_{rg}. The water fraction model used there for the mixed phases is described in detail in Jung et al. (2008a), and the water fraction model used during the 2013 CAPS Spring Experiment does not vary across the size spectrum. The density (ρ) of each mixed-phase species increases as the fractional amount of rain increases and the dielectric constant is calculated using the Maxwell-Garnett mixing formula (Maxwell-Garnett 1904). These variables are used in separate calculations of Z_{rs}, Z_{rh}, and Z_{rg} for mixtures, in addition to Z_r, Z_s, Z_g, and Z_h, with the log of the sum giving the final simulated Z. A radar wavelength of 107 mm is used to match the WSR-88D S-band network. For reference, from JXZ10, Z is calculated using equation (3), Z_{DR} from the quotient of (3) and (4), and K_{DP} from (6).

c. Spring experiment microphysics schemes

The 2013 SSEF WRF-ARW members used 6 different MP schemes: the MY, Morrison, Thompson, WDM6, NSSL (Mansell 2010), and WSM6 schemes. The NSSL scheme has not yet been added to the PRDS because its representation of hydrometeor PSDs is a lot more complex than the other schemes. The original PRDS operators were already compatible with the WSM6 and MY schemes. The Morison scheme follows the same PSD and has the same predicted moments as MY (having either graupel or hail) so it was easily implemented. Modifications were required for the other schemes. The Thompson and WDM6 schemes predict N_t and q for rain but only predict q for the remaining categories as used in the PRDS (prediction of N_{tr} was added to the Thompson scheme subsequent to Thompson et al. (2008). WDM6 diagnoses N_{0s} using temperature and uses a fixed value for N_{0g}. WDM6 also uses a fixed shape parameter of 1.0 for α_r. The Thompson scheme has been further updated since Thompson et al. (2008) to use temperature and the mean volume diameter of rain to diagnose N_{0g}. The Thompson scheme also deviates from the typical representation of the

bulk PSD for snow, using a combined exponential and gamma distribution, but the simulation of polarimetric variables at and above the freezing level will be the focus of future studies. Table 1 summarizes whether a fixed value or predicted model variables are used to calculate N_0 and α for each hydrometeor category for each MP scheme.

d. Polarimetric radar observations

The upgraded WSR-88D radars provide domain-wide polarimetric observations which are used for comparison to the simulated variables. Z and Z_{DR} observations are filtered using a 5-point along-the-radial median filter, while K_{DP} is calculated from similarly filtered differential phase (Φ_{DP}) observations using the least squares fit method of Ryzhkov and Zrnic (1996). Nine range gates are used when $Z > 40$ dBZ and 25 range gates are used for $Z < 40$ dBZ.

The availability of polarimetric observations allows for extensive quality control of the data using fuzzy logic (Park et al. 2009). The fuzzy logic method uses ranges of polarimetric radar data values and weights to determine the most likely hydrometeor type of the observation. The Z, Z_{DR}, ρ_{hv}, standard deviation (SD) of Z (1 km running average), and SD(Φ_{DP}) (2 km running average) membership functions are used along with their respective weights. The confidence vectors are not included. Additionally, the temperature profile and the presence of frozen hydrometeors from the forecast model are used to help further narrow down potential hydrometeor types before classification. For example, frozen categories are not considered at heights where full melting has occurred in the forecast and rain-associated categories are not considered above the freezing level. The MY forecast member was chosen since this scheme included the most number of hydrometeor categories and produced storms with reasonable structure and intensity based on a qualitative comparison of the results. Those observations that are determined to be ground clutter, anomalous propagation, or biological scatterers are removed. This is important since SSEF forecasts begin at 0000 UTC and short term forecasts in the late spring/early summer months will be at a time when observed radar blooms due to birds and insects are prominent (Lakshmanan et al. 2007). An example of Z, Z_{DR}, and K_{DP} observations at an elevation angle of 0.5° before and after the removal of non-meteorological echoes for one of the cases evaluated in this study, at 0400 UTC 20 May 2013, is given in Fig. 2. The locations of the WSR-88D radars used are included as black dots in Fig. 2a. Obvious clutter from the late evening radar bloom (10 pm Central Standard Time) is almost completely removed. Data points that are determined to be a three body scatter spike are also removed (Mahale et al. 2014).

3. Evaluation of simulation results

In this section, results for two example cases chosen from the 2013 Spring Experiment are evaluated. The first is a four-hour forecast initialized at 0000 UTC 20 May 2013 for a series of MCSs. The second is a 21-hour forecast of several supercell thunderstorms also initialized at 0000 UTC 20 May 2013. These cases provide two different convective modes that contain distinct and different polarimetric signatures for evaluation. Additionally, the first case is a short-term forecast with sufficient lead time to allow for microphysical processes such as size sorting to develop while not too long for the storm systems originally initialized from radar data to dissipate. The divergence between the ensemble of model solutions and the observations is also relatively small at this point. Because of the early evening initialization time, a similar situation is difficult to find for supercell thunderstorms, which typically dissipate or grow upscale at night. For these reasons, the 21-hour long forecasts valid in the afternoon of 20 May were chosen. It was also a point of emphasis to choose cases that had storm systems that were well placed so more focus could be on the differences in the polarimetric variable values and not storm structure and placement. For convenience, the chosen ensemble

members (section 2a) are referred to by their respective microphysics schemes: TOM (Thompson), MY (Milbrandt and Yau), MOR (Morrison), WDM (WDM6), and WSM (WSM6).

In this study, the forecast results are compared to the observations by creating a 0.5° elevation mosaic of observed and simulated radar data from all WSR-88D sites within the domain from data in a 'gridtilt' format for direct comparison. The simulated variables are left on the model grid in the horizontal but mapped (via weighted average) in the vertical to the elevations or tilts of each radar using the beam pattern weighting function given in Xue et al. (2006). Conversely, the radar observations are interpolated to the model grid points in the horizontal but left on the radar elevation levels in the vertical. As a result, both model and radar data are transferred to a common 'gridtilt' space with respect to individual radars. A 0.5° elevation mosaic is created by combining the lowest available elevation angles (0.5°) from the WSR-88D radars located within the domain, using the observation closest to the surface where two or more radars overlap. The 0.5° elevation angle is chosen and the closest-to-the-surface value is used because polarimetric signatures associated with hydrometeor size sorting are most prevalent near the surface. Our method is similar to how the 'reflectivity at lowest altitude' or 'RALA' product is produced in the Multi-Radar Multi-Sensor (MRMS) system (Smith et al. 2016). A near-surface constant height mosaic will have large areas of missing observations because the radar beam height increases with distance. Additionally, vertical interpolation of radar elevation level data to a constant model level can have large error as the radar beam width increases with distance from the radar (Sun and Crook 2001).

a. The 20 May 2013 mesoscale convective system case

Height falls associated with an upper level trough moving into the central plains and ample low level moisture led to the development of multiple areas of severe thunderstorms during the mid-afternoon of 19 May 2013. Over time, these clusters grew upscale to form several MCSs that stretched from the upper Mississippi valley south into Oklahoma. The most intense of these system resulted from storms initially forming over central Kansas that continued into eastern Iowa. At 0400 UTC 20 May this system exhibited the elements of a classic MCS including leading convection and trailing stratiform precipitation (Fritsch and Forbes 2001). Additional linear convective storms formed along an outflow-reinforced cold front that stretched southwestward into northern Oklahoma. Widespread damaging wind and hail was reported across the Midwest, and several tornadoes were reported in southwestern Missouri (SPC 2014a).

1) QUALATATIVE EVALUATION OF FORECASTS

Mosaics of observed and simulated Z, Z_{DR}, and K_{DP} in the gridtilt format described at the beginning of section 3 are plotted in Fig. 3, 4, and 5. Locations of WSR-88D radar sites used for both the observed and simulated variable plots are included in Fig. 3a, 4a, and 5a. A 20 dBZ Z contour is included for reference in Fig. 4 and 5. Overall, the observed features are well placed in the five forecasts. However, the intensity and structure of the forecast precipitation differ from the observations, and differ among the forecast members. The MY and MOR members produce more widespread, high Z in the convective areas (Fig. 3c,d) in eastern Iowa compared to the observations but show a decrease in intensity in the stratiform precipitation region further west in central Iowa. Z in WDM and WSM is lower than in MY and MOR in the convective areas and more closely matches the observations (Fig. 3e,f). However, the stratiform precipitation over Iowa is almost non-existent in WDM, and is significantly under-forecast in WSM. For the WDM case, Lim and Hong (2010) similarly found that WDM had low rain rates in the stratiform region of a 2D simulated MCS and attributed these low rates to higher rain number concentrations and increased evaporation. The WSM and WDM forecasts are overall very similar as only warm rain processes are DM in WDM. The

importance of predicting a second moment for ice processes was noted in Putnam et al. (2014), who found that a DM MY forecast, which predicts a second moment for snow and cloud ice, better maintained separate convective and stratiform precipitation regions in an MCS compared to a SM Lin forecast due in part to improved transport of frozen hydrometeors between the convective towers and stratiform precipitation region. The TOM Z appears most reasonable in terms of both intensity and coverage (Fig. 3b). The placement and structure of convection in western Missouri and northern Oklahoma, and the associated Z intensity, match the observations well. However, the convective region in eastern Iowa is disorganized with no discrete linear convective line ahead of the trailing precipitation.

The range of Z may be due to different hydrometeors sizes, types, and water contents; simulated Z_{DR} and K_{DP} provide further insight to better differentiate the microphysical states of the members. Additionally, observed Z_{DR} and K_{DP} can be used to diagnose the hydrometeor categories using the fuzzy logic method described in section 2d (Fig. 6). The observed HCA is computed for each radar and plotted as a 0.5° elevation mosaic as described at the beginning of section 3. For the forecasts, the dominant category is determined based on which hydrometeor type provides the majority contribution to the linear simulated Z, including the diagnosed mixed phase species described in section 2b. If no category contributes at least 50% to Z, the category is considered a 'mix'. Since the model MP scheme categories do not match the Park et al. (2009) HCA categories, hail, wet (melting) graupel, and 'mix' are added to the classification list for identification when present in the PRDS results. Additionally, the forecast rain category does not differentiate between the 'big drop' and 'heavy rain' categories so all rain is combined into one category and wet (melting) hail from the PRDS results is considered 'rain and hail'. Non-meteorological categories are not included, as they were removed during quality control.

Observed Z_{DR} is generally greater than 2.0 dB in the convective regions with a maximum of around 3.0 dB and is less than 2.0 dB in the stratiform regions (Fig. 4a). This matches the typical observed Z_{DR} pattern in an MCS caused by differing rain PSDs; areas of convective precipitation have high Z_{DR} due to the presence of large raindrops, with the maxima occurring along the leading edge of the convective line where the size sorting of smaller raindrops, which have a low terminal velocity and are carried further rearward in the convective line, isolates larger raindrops and leads to high Z_{DR} values. The stratiform region, which is not supported by an intense updraft, contains moderate precipitation with small to medium-sized raindrops and lower Z_{DR} values, respectively (Zhang et al. 2008). MY and MOR have the highest simulated Z_{DR} in convective regions, comparable to the observations, but the coverage of these high values is more widespread (Fig. 4c,d). The widespread high Z_{DR} matches the areas of over-forecast high Z where the presence of larger raindrops in intense convection would be expected. The high Z_{DR}, along with very high Z, is also indicative of large oblate wet graupel, the dominant hydrometeor category in these convective regions (Fig. 6c-d). MY predicts hail but shows a similar high bias in wet graupel as MOR. The MY scheme implemented in the WRF as used in the CAPS SSEF was modified from Milbrandt and Yau (2005a) to include a strict minimum size threshold for hail. Hail below this threshold is converted back to graupel and this threshold has been shown to result in forecasts that produce little, if any, hail (Van Weverberg et al. 2012). This may explain the presence of too much graupel compared to the observations in this case. Z_{DR} tends to increase towards the leading edge of the convective lines in southeastern Kansas and northeastern Oklahoma, as well as in central Missouri and eastern Iowa for MOR (indicated by arrows in Fig. 4c,d); high Z_{DR} at the leading edge of convection is a commonly found polarimetric signature associated with size sorting. However, there are many convective areas where large drops are embedded within the convection and a size sorting signature is not evident. MOR has a strong size

sorting signature in southwestern Missouri (indicated by 'A' in Fig. 4d), but this corresponds with low Z and weak precipitation. It is not unusual to see high Z_{DR} in developing convection as size-sorting begins to occur, but the spike in Z_{DR} may be due to the fact that the drop breakup rate in MOR is dependent on q_r and the rate will be low where q_r is small. As a result, in low precipitation regions with small q_r, the low rate of drop breakup may lead to a locally high number of large drops. These spikes in Z_{DR} were also noted in Johnson et al. (2016).

There is a more significant difference in the microphysical state of MOR and MY in the stratiform precipitation region in central Iowa. There is widespread moderate Z_{DR} in MOR, only 0.5 dB less than the Z_{DR} in the convective line to the east, while Z_{DR} decreases away from the leading convective lines in MY, similar to the observations. The distinguishing Z_{DR} that differentiates the PSDs of the convective and stratiform regions is more prominent in MY; this matches the findings of Putnam et al. (2014). The difference in the similar DM MOR and MY forecast results highlights MP scheme challenges that extend beyond simply adding a second moment.

Z_{DR} in TOM and WDM shows no clear organization based on the structure of the convective systems (Fig. 4b,e). There is little difference between Z_{DR} in the convective and stratiform regions, and no clear size sorting signatures that match the observations. In fact, the highest Z_{DR} in TOM is in an area of light precipitation on the rear side of the convective line in Western Missouri, to the north and south of the convection in Iowa, and in isolated light showers in Missouri (indicated by arrows in Fig. 4b). Relatively high Z_{DR} values are often seen with developing storms like those in Missouri, but the remaining noted Z_{DR} patterns do not match the observations. Both TOM and WDM are DM for rain but only SM for graupel, and thus there is no size-sorting of graupel, which has a greater impact on the development of low-level Z_{DR} signatures (along with hail) than the size-sorting of rain (Dawson et al. (2014)). Despite the apparent lack of size sorting, which is one of the features tied to maintaining a stratiform precipitation region in Putnam et al. (2014), TOM still represents the coverage of the convective and stratiform precipitation regions relatively well. Wheatley et al. (2014) found in their real case EnKF study of an MCS that the Thompson scheme replicated the convective and stratiform regions well due to broad and intense development of snow aloft. Thompson is not DM for snow, but it does use a unique snow PSD and diagnostic N_{0s}, and TOM contains stratiform coverage similar to that produced by the fully DM MOR and MY results. WDM also has areas where high Z_{DR} is located on the rear edge of convection in northern Missouri and southern Iowa (indicated by arrows in Fig. 4e). In fact, Z_{DR} is relatively low in the most intense areas of convection (Z > 40 dBZ), indicating a large number of small to moderate-sized drops. Given that wet graupel is the dominant category present (indicated by arrows in Fig. 6e), it is likely that the wet graupel that exists is small and shedding small raindrops. This differs from MOR and MY, which also show a significant contribution to Z from wet graupel but have much higher Z_{DR} values, indicating larger wet graupel and a significant difference in the graupel PSDs between the schemes.

Z_{DR} differs more substantially between WDM and WSM compared to Z (Fig. 3e,f and 4e,f). The Z_{DR} maxima from the two schemes are about the same, but in WSM Z and Z_{DR} have a monotonic relationship, with the highest Z_{DR} co-located with the highest Z; this is because the WSM scheme with a fixed N_0 is incapable of size sorting. WDM does not improve upon WSM, however, compared to the observations, since Z and Z_{DR} do not differ between convective and stratiform precipitation regions and the size sorting signatures appear in the wrong locations. WDM is DM for rain and cloud droplets, but SM for ice species as in WSM, further emphasizing the impact that the size sorting of graupel has on low-level rain PSDs and the associated Z_{DR} signatures.

Simulated K_{DP} is generally lower in all members compared to observations with values mostly less than 2.0 $^{\circ}$ km^{-1}. A few more intense convective areas (Z > 50 dBZ) in MY and WSM have K_{DP} higher than 2.0 $^{\circ}$ km^{-1} (Fig. 5c,f) which agrees more closely with the observations (Fig. 5a). MY and MOR have similar Z_{DR} maxima but generally lower K_{DP} compared to the observations so the rain PSDs must contain a lower concentration of small to moderate-sized drops (Fig. 5a,c,d). The convective regions in MY, MOR, and WDM contain significant wet graupel (Fig. 6c-e). Low Z_{DR} and K_{DP} are indicative of small wet graupel in WDM (Fig. 4e and 5e) while significantly higher Z_{DR} in MY and MOR (Fig. 4c,d) suggests larger wet graupel but with a low water ratio. Low Z_{DR} is also indicative of the lack of large raindrops in WDM due to the gamma distribution with a short tail in the large drop's end. WSM has lower Z_{DR} but similar K_{DP} (Fig. 5f) compared to MY and similar Z_{DR} but higher K_{DP} compared to TOM, indicating a higher concentration of small to moderate-sized drops due to a large fixed N_{0r}. A high bias in graupel compared to observations (Fig. 6a) could potentially explain low K_{DP} (as compared to an all-rain scenario), but WSM, as well as TOM, have similarly low K_{DP} values in areas of pure rain.

Another significant difference in the forecasts compared to the observations is the lack of significant K_{DP} in the stratiform precipitation region over central Iowa. Those members that show this trailing precipitation (TOM, MY, MOR) only have noteworthy K_{DP} where Z exceeds 35 dBZ (Fig. 5b,c,d); the observations exhibit low (but consistent) K_{DP} throughout the stratiform region (Fig. 5a). Although the melting layer differs between the forecasts, which could contaminate the K_{DP} results, K_{DP} is also lower in areas of pure rain. In MOR, Z_{DR} is higher than observed in the stratiform region, indicating PSDs with a few larger raindrops and lower water content overall (Fig. 4d). Z_{DR} in TOM and MY is closer to the observations but similarly low K_{DP} values suggest an overall low bias among the DM schemes in the concentration of moderately-sized raindrops and their water content (Fig. 5b,c).

2) QUANTITATIVE EVALUATION OF FORECASTS

Although forecast errors at all scales continuously grow as forecast length increases, the error growth rate is much larger for smaller scale phenomenon (e.g., convective cells). Additionally, diagnostic metrics are highly sensitive to displacement errors for storm scale forecasts, especially for fine-scale structures, making one-to-one quantitative verification even more difficult. A few different methods of quantitative comparison that help account for these errors are considered.

Percentile histograms of simulated Z, Z_{DR}, and K_{DP} are calculated over the domain used for Fig. 3, 4, and 5 to gauge the overall distribution of values (Fig. 7). The histograms are created by first ranking all observed and simulated values individually and then distributing the simulated values within 10 bins representing the observed percentiles between 0.0 and 1.0. Percentiles relative to the observations are used to account for potential biases and outlier values in the model results. The range of the observed values corresponding to the 0.0 to 0.2, 0.2 to 0.4, etc., percentile bins is indicated by coordinating the colors between the observed values used as the bounds for their respective percentile bins. For reference, the observed histograms for each variable are provided in Fig. 7a-c, with a black line indicating the number of observed values per bin included for all simulated results and a red line indicating the ideal distribution of the simulated model values, if these values were distributed evenly. The total number of simulated values is also included in each subplot.

There is a high bias in the simulated Z results relative to the largest observed Z values, indicative of strong convective precipitation, especially for MY and MOR (Fig. 7g,j). Qualitative comparisons suggested that the convective regions contained more intense precipitation over a larger area and that wet graupel contributed to high Z. The observed distribution peaks at around 35 dBZ

(Fig. 7a), which corresponds to Z from the large region of stratiform precipitation in central Iowa and contributes to the number of values in the middle percentiles. The MOR and TOM distributions are most similar to the observations in this mid-range, though all forecasts have at least some low bias. MOR has a consistent high bias for low and high Z due to an over-forecast of precipitation coverage (Fig. 7j). Conversely, WDM has a consistent low bias at nearly all percentiles, lacking the stratiform precipitation and significantly under-forecasting precipitation coverage overall (Fig. 7m).

The observed distribution of Z_{DR} peaks around 1 dB, and extends up to 4 dB (Fig. 7b). TOM, and particularly MY and MOR, have a high bias in the higher percentiles due to the over-forecast east-west extent of the convective lines as well as PSDs that contain too many large drops. MOR has a significant bias above the 90th percentile (Fig. 7k) related to the large drops in the stratiform region, oblate wet graupel in the convective region, and over-forecast in precipitation coverage overall. On the other hand, WDM and WSM have a similar low bias above the 50th percentile. The low bias in WSM may be related to the relatively high fixed N_{0r} of 8 x 10^6 m^{-4} (Fig. 7q). Although WDM is DM for rain, it follows a similar behavior as the SM WSM, which was also noted in the qualitative evaluation.

The K_{DP} histograms are limited to values above 0.5 °km^{-1} because values below this threshold can be indistinguishable from noise in the observations (Jung et al. 2008b). This limits the K_{DP} assessment to the convective regions, where the precipitation coverage for all members matches the results better as well. The TOM, WDM, and WSM results are relatively similar to the observations for the low to mid percentiles while MY and MOR overestimate these values due to the broader width of the convective lines in these forecasts (Fig. 7f,i,l,o,r). However, for the highest percentiles, which represent the peak of the observed K_{DP} values in the intense precipitation convective regions, there is a low bias in all members. In general, all members have simulated K_{DP} values lower than observed in the convection regions due to apparent lower liquid water contents resulting from contamination by graupel (MY, MOR, and WDM (Fig. 7i,l,o)), a bias towards larger drops as indicated by Z_{DR} (TOM, MY, and MOR (Fig. 7f,i,l)), and a high concentration of small drops due to the fixed intercept parameter in WSM (Fig. 7r).

Another quantitative measure to evaluate the PSDs is scatterplots of Z versus Z_{DR} at a given location for the observations and forecasts (Fig. 8). Data points where $Z > 5$ dBZ from the Fig. 3, 4 and 5 domain are considered and are categorized by dominant hydrometeor type using the same process as for Fig. 6. TOM, MY, and MOR show a broad overall distribution of Z and Z_{DR} value combinations similar to the observations. The additional free PSD parameter in the DM scheme allows for greater flexibilities in the range of possible PSDs in the forecast. The high density of data points for rain in the observations results from the broad region of stratiform precipitation (Fig. 8a). TOM, MOR, and MY, which performed well in terms of stratiform precipitation in the qualitative evaluation, have a similar concentration of data points for rain (Fig. 8b-d). The distribution in MOR is shifted toward slightly higher Z and Z_{DR} due to the widespread coverage of more moderate to large raindrops in the stratiform region. The large amount of melting ice species (snow, graupel, or a mix of both (Fig. 8d)) leads to the over-forecast convective intensity in MOR based on a comparison of the hydrometeor types associated with these values in Fig. 3d and 6d; the same is true to a lesser extent with MY (Fig. 8c). The Z_{DR} maxima displaced from the leading edge of the convection in WDM results from wet snow and graupel (Fig. 8e, indicated by arrows in Fig. 4e and 6e), while that with TOM is mainly associated with rain (Fig. 8b, indicated by arrows in Fig. 4b and 6b).

WDM and WSM exhibit very similar distributions (Fig. 8e-f). There is little spread in the data points in WSM given the one-to-one relationship between Z and Z_{DR} in a SM scheme. Again,

WDM, being only DM for rain, and with a diagnostic N_{0s}, exhibits the least variation compared to the more complex TOM and fully DM MOR and MY. Most of the variation in both WSM and WDM is associated with the presence of mixed-phase precipitation where changing liquid and frozen water contents will lead to various Z and Z_{DR} combinations.

Since traditional numerical measures like root mean square error (RMSE) will indicate poor results when spatial errors are present, neighborhood methods have been developed to account for placement errors when the overall storm structure is otherwise good (Ebert 2008). One of these techniques, the fractions skill score (FSS, Roberts 2008; Roberts and Lean 2008), has been considered in past studies involving the CAPS SSEF and is used again here (Schwartz et al. 2009; Cintineo et al. 2014). The FSS is calculated by finding the fraction of forecast grid points in a neighborhood with a given radius that exceeds a threshold value compared to the observations. The FSS is designed so that as the radius for the neighborhood increases to the size of the domain the score will asymptote towards an ideal finite value of 1. If there is bias present in the forecast, then the score will be less than 1, except for relatively small scale neighborhoods (Mittermaier and Roberts 2010). A forecast can be considered to have measurable skill when:

$$FSS > 0.5 + \frac{O_{domain}}{2}, \tag{3}$$

where O_{domain} is the domain-wide fraction of grid points where observations exceed the given threshold (Roberts and Lean 2008).

The FSS is calculated over the Fig. 3, 4, and 5 domain for several thresholds for radii ranging from 0 to 200 km to account for the regional nature of MCS coverage (Fig. 9). Overall precipitation coverage, including both the convective and stratiform regions, is assessed using a threshold of Z > 15 dBZ (Fig. 9a). All forecasts perform well at this threshold. TOM, MY, and MOR show skill with initial scores around 0.65 that increase to around 0.9 at the 100 km radius. WDM has the worst score overall, averaging around 0.2 less than other members, likely due to the low precipitation coverage bias. When the Z threshold is increased to 40 dBZ to assess the prediction of intense convective precipitation (Fig. 9b), there is more spread between the members, and all members exhibit very poor scores for small radii when large spatial errors are present. Interestingly, WDM has the highest score for the 100 km radius due to less extensive east-west coverage bias of the convective lines compared to the other members. WDM under-forecasts precipitation overall but matches the observations better in the convective regions. TOM and MY show some skill using a 100 km radius, but do not improve much with increased radii. MOR, which exhibits substantial high Z bias due to the over-forecast of intense convective precipitation, has no measurable skill.

A Z_{DR} threshold of 2.5 dB is chosen to assess convective regions in the observations where the largest drops are present, specifically the maxima seen with the polarimetric signatures associated with size sorting on the leading edges of convective lines (Fig. 9c). Scores are generally poor; the only skillful forecasts are TOM, which shows skill at a 100 km radius, and MY, which shows skill at a 150 km radius. This result indicates the coverage and intensity of significant Z_{DR} in TOM in convective regions is closest to the observations without over-forecasting large drops overall within the stratiform region. However, the qualitative evaluation in section 3a(1) showed Z_{DR} in TOM is displaced and large radii neighborhoods miss these fine-scale details. MOR is likely negatively impacted by the high Z_{DR} coverage bias in the stratiform region and greater east-west extent of convective regions. WDM and WSM are biased towards small hydrometeor sizes.

Similarly to Z_{DR}, a threshold of 0.6 $^{\circ}$km^{-1} for K_{DP} is chosen to highlight the convective cores where higher liquid water content is present (Fig. 9d). The K_{DP} maximum is generally lower in all

members than in the observations and skill scores at higher thresholds will be very poor. All members have skill for radii greater than 50 km (Fig. 9d). Thus, convective regions with high K_{DP} are relatively well placed, with the caveat that graupel contamination may affect the upper range of these values. The better FSS scores for K_{DP} compared to Z_{DR} are likely due to the more direct linkage between high liquid water content and intense convection while Z_{DR} patterns associated with size sorting are not co-located with Z maxima. Z_{DR} provides a more stringent assessment of microphysical processes and states.

Since the range of simulated polarimetric variable values in each forecast may not match the overall range of the observations, the FSS scores are calculated again for the same thresholds using percentiles (Fig. 10). The percentile value in the observations consistent with each numeric threshold is used as the threshold to assess the forecast percentile values. This method normalizes the scores for those forecasts which do not produce values as high as the observations, effectively removing biases and providing a fairer assessment of feature placement, as in the Z_{DR} values for WDM and WSM. The scores for $Z > 15$ dBZ and $K_{DP} > 0.6\,^{\circ}\text{km}^{-1}$ are very similar to the previous results. However, the new scores are improved significantly for the $Z > 40$ dBZ convective assessment and $Z_{DR} > 2.5$ dB which are more affected by maximum value biases. More specifically, those forecast members that showed no skill without using percentiles (MOR for $Z > 40$ dB; TOM, WDM, and WSM for $Z_{DR} > 2.5$ dB) now show skill at higher radii when using percentiles. The most significant improvement is in WDM and WSM; these members have a low bias compared to the other members for high Z_{DR} values. The percentile calculations show the highest Z_{DR} values from WDM and WSM forecasts are well placed compared to the observations but underestimated in value. Future quantitative assessment methods of polarimetric variables may need to take into account maximum and minimum value biases.

b. The 20 May 2013 supercell case

Several supercell thunderstorms developed along a stationary front across the southern plains early afternoon 20 May 2013, the most intense of which occurred over central and southern Oklahoma. Dewpoints in the low 70s (F) and 5000 J kg^{-1} of CAPE combined with winds in excess of 50 kts at 500 hPa associated with an upper level trough to create a volatile severe weather environment. The most intense storm, which produced an EF-5 tornado, formed along the stationary front southwest of the Oklahoma City area shortly before 2000 UTC. The tornado killed 24 people and caused over 1 billion dollars in damage across the southern Oklahoma City metropolitan area (NWS 2014). Additional tornadoes were reported across Oklahoma as well as widespread large hail reports over southern Oklahoma (SPC 2014b).

1) QUALATATIVE EVALUATION OF FORECASTS

Mosaics of 0.5° tilt observed and simulated Z, Z_{DR}, and K_{DP} for all members at 2100 UTC (Fig. 11, 12, and 13) as well as hydrometeor classifications using the same process in section 3a(1) (Fig. 14) are evaluated. Locations of WSR-88D radar sites within the domain used to create the observed and simulated variable mosaics are included in Fig. 11a, 12a, 13a, and 14a. A 20 dBZ Z contour is included for reference in Fig. 12 and 13. The placement and coverage of forecast convection are worse than in the MCS case due to the longer forecast lead time, the isolated nature of discrete supercell storms, and that storm development is not directly influenced by assimilated radar data at the initial condition time. The Z patterns for the southern Oklahoma storms in MY and MOR (indicated by 'A' in Fig. 11c,d) both exhibit classic supercell structure with a hook-echo/rear flank downdraft, indicative of the presence of a mesocyclone. It should be noted that the structures are rather large compared to the observations; this is often seen in forecasts using a 4 km grid spacing (Lean et al. 2008, Johnson et al. 2013). TOM, WDM, and WSM have a line of cells that are smaller

and low-precipitation in comparison but have supercell characteristics (Fig. 11b,e,f). For reference, a plot of the Spring Experiment hourly max updraft helicity product (Kain et al. 2008; Kong 2013) for 2100 UTC is included (Fig. 15); the high values greater than 150 $m^2 s^{-2}$ for the southern Oklahoma storm in all members are indicative of a mesocyclone (indicated by arrows in Fig. 15). As in the MCS case, WDM under-forecasts the precipitation coverage, while MOR forecasts the most widespread, high *Z*. *Z* in the observations (Fig. 11a) generally peaks at a higher value ($Z > 50$ dBZ) than in the model forecasts; MOR is most similar to the observations. However, the high *Z* in MOR is due to the presence of wet graupel (Fig. 14d), while high *Z* observations are mostly due to rain and some hail. There is spurious convection in the northwest corner of the domain in all members.

The observed Z_{DR} is generally higher on the right (southeastern) edge of the forward flanks of the observed cells (Fig. 12a), exhibiting a distinctive Z_{DR} arc. Z_{DR} is lower in the center of the forward flanks of the central Oklahoma storms, possibly due to a hail-induced Z_{DR} hole, but the hydrometeor classification algorithm does not identify widespread, consistent areas of hail (Fig. 14a). The lower Z_{DR} is associated with higher K_{DP} (Fig. 13a), indicating moderately-sized drops and a high rain rate. Z_{DR} increases along the right forward flank of the dominant southern Oklahoma storm in MOR, MY, and to a lesser extent in WDM (indicated by arrows in Fig. 12c,d,e), while other convective cells are less organized and don't show this signature. The HCA identifies wet graupel in MOR, MY, and WDM along the right forward flank of this storm (indicated by arrows in Fig. 14c-e), and the size sorting of melting graupel has been shown by Dawson et al. (2014) to have a substantial impact on the model representation of the Z_{DR} arc. MOR shows the highest Z_{DR} farther downwind the forward flank than the extent of the wet graupel, consistent with Dawson et al. (2014; see their Fig. 17). Wet graupel is present along the entire right forward flank in MY and WDM, and the maximum Z_{DR} is not located along the immediate edge. Dawson et al. (2014) also found hail better replicated the observed coverage and intensity of the Z_{DR} arc compared to graupel which can lead to an over-extensive forward flank, as seen in this case.

TOM and WSM did not exhibit the same Z_{DR} pattern as the observations (Fig. 12b,f). There is a one-to-one relationship in WSM where high Z_{DR} occurs with high *Z* in the center of the cells. Similar to the MCS case, a relative Z_{DR} maximum occurs along the edges of smaller cells with less intense precipitation in TOM, MOR, and WDM. This occurs in the southwest portion of the domain for MOR and TOM, in northwest Oklahoma for MOR, and in central Oklahoma for TOM and WDM (indicated by 1, 2, and 3 in Fig. 12b,d,e, respectively). As noted for the MCS case, high Z_{DR} is sometimes associated with aggressive size sorting in developing convection. The drop breakup scheme in MOR may have also contributed to the spikes in Z_{DR} values associated with weak precipitation in that case. The Z_{DR} in TOM appears particularly misplaced; there is a large area of Z_{DR} exceeding 2.5 dB that occurs between the more intense precipitation associated with the storm cells. The PSDs in this region are heavily weighted toward a few large drops given the low *Z*. Compared to MY and MOR, TOM, WSM, and WDM are not DM for graupel, which was noted to have an impact on size sorting in supercells.

K_{DP} is under-forecast in all members compared to the observations. The observations peak above 3.0 $^\circ km^{-1}$ (Fig. 13a), while only TOM has a maximum above 1.75 $^\circ km^{-1}$ (Fig. 13b). Of note, the highest K_{DP} values are not co-located with the highest Z_{DR} values in the southern Oklahoma storm in MY and MOR (Fig. 13c,d). K_{DP} has a relative maximum in the center of the storm but decreases in the right forward flank where Z_{DR} is higher, another indication that size sorting has resulted in a few large raindrops in the Z_{DR} arc compared to elsewhere in the forward flank. Although there is some

graupel present in the forecast that may contaminate the K_{DP} results, particularly in MOR, most of the forecast convection is classified as pure rain, suggesting liquid water content is under-forecast overall.

2) QUANTITATIVE EVALUATION OF FORECASTS

As in Section 3a(2), percentile histograms (Fig. 16) and FSSs (Fig. 17) are considered for quantitative evaluation. Both are calculated over a sub-domain that focuses on the line of supercells that extends from southeast Kansas to northwest Texas. The MY and MOR histograms show an over-forecast of precipitation coverage (Fig. 16g,j); this over-forecast is present but confined to lower percentiles in TOM and WSM (Fig. 16d,p). Given this circumstance, the mostly even distribution of intensities in the observations (Fig. 16a) is matched relatively well by these members. The WDM precipitation coverage is under-forecast so substantially that the Z distribution is lower than the observations for all percentiles (Fig. 16m).

There is a high bias in the amount of Z_{DR} values due to the greater precipitation coverage in TOM, MY, and MOR. Z_{DR} in MY has a relatively even distribution compared to the observations, with the caveat that the over-forecast coverage of precipitation leads to a consistently higher number of values at most percentiles overall (Fig. 16h). After considering the grid scale, simulated Z_{DR} in MY represents the varying degree of maximum raindrop size in the rain PSD well. MOR has a significant peak at the 70th percentile (Fig. 16k), corresponding to the widespread Z_{DR} around 2.5 dB and greater in the central forward flank regions where wet graupel is present (Fig. 14d). Z_{DR} in WSM, for all but the lowest percentiles, is lower than the observations compared to TOM, MOR, and MY (Fig. 16q); this is expected given the fixed intercept parameters used in this SM scheme. Z_{DR} in WDM has a significant low bias (Fig. 16n) due to small raindrops and small, wet graupel in precipitation under-forecast in coverage and intensity.

In the forecast K_{DP} distributions (Fig. 16f,i,l,o,r), all members but WDM have a similar number of values, with a generally decreasing trend overall towards the higher percentiles. WDM has the lowest number of values compared to the observations while WSM has the highest number of values in comparison, particularly for the lowest percentiles. All forecast members appear to have generally lower liquid water contents than the observations in pure rain areas.

The FSS is calculated for radii only up to 100 km due to the more localized nature of the supercell case. All members show some skill for radii of 20-40 km or more for a Z threshold of 15 dBZ (Fig. 17a). WSM and TOM have noticeably higher scores than the other members because the precipitation coverage is more similar to the observations; precipitation coverage is over-forecast in MY and MOR and under-forecast in WDM. The qualitative evaluation showed more realistic Z and Z_{DR} patterns in MY and MOR compared to WSM and TOM but the size of the supercells in the former cases was notably larger. Members generally show no skill for Z_{DR} (Fig. 17b). The K_{DP} threshold (Fig. 17c) is decreased slightly compared to the MCS case since the simulated values are lower overall (.4 $^\circ$ km^{-1} instead of .6 $^\circ$ km^{-1}). MY and WSM, as in the MCS case, have the best skill for K_{DP} for large radii (> 60 km), but these scores are not high (< .7). Issues of grid-scale and storm placement leave many quantitative challenges for the simulated polarimetric variables, especially in terms of Z_{DR} patterns for a supercell case.

Fig. 18 replicates the FSS calculations for Fig. 17 using percentile values. As in the MCS case, those forecasts with very poor scores for Z_{DR} > 2.5 dB due to lower maximum values than the observations (WDM and WSM) show significant improvement. However, all forecasts still show little to no skill overall. The spatial extent and coverage of Z_{DR} signatures for the supercell case appear more difficult to match than the MCS case. Unlike the MCS case, the K_{DP} scores are also

improved for all members. The histograms for each case show that the distribution of K_{DP} values is a better match in the MCS case than for the supercell case where there is a greater low bias in K_{DP} values. The use of percentiles helps better match the observed and forecast distributions for comparison so that all but WDM show at least some skill at the larger radii values. It is clear the poor K_{DP} scores when percentiles aren't used are due to the low bias of values in the forecast.

4. Summary and conclusions

Polarimetric variables are simulated from the CAPS spring experiment storm scale ensemble forecasts (SSEF) for evaluation of both single-moment (SM) and double-moment (DM) model microphysics (MP) schemes. An existing polarimetric radar data simulator (PRDS, Jung et al. 2008a; Jung et al. 2010) is modified to add several new MP schemes including Thompson (TOM), Morrison (MOR), and WDM6 (WDM); Milbrandt and Yau (MY) and WSM6 (WSM) were already included. Careful attention is paid in the simulation to the hydrometer types and particle size distributions (PSDs) of each scheme to properly represent the forecast microphysical state. Two cases are considered: a 4-hour forecast for a series of mesoscale convective systems (MCS) from 20 May 2013 and a 21-hour forecast of supercell thunderstorms from the 20 May 2013 Oklahoma tornado outbreak. Simulated reflectivity (Z), differential reflectivity (Z_{DR}), and specific differential phase (K_{DP}) from a single ensemble member forecast using each scheme with otherwise similar model settings are compared to observations from the recently upgraded WSR-88D radar network.

Z_{DR} in MOR and MY in the supercell case, as well as classification of the hydrometeors present, produce results consistent with Dawson et al. (2014), which demonstrated the role that the size sorting of graupel plays in the formation of the Z_{DR} arc. The other schemes examined are not DM for graupel and do not show this pattern. In addition, the two schemes that best represent polarimetric size sorting signatures (MY and MOR) also show better coverage of stratiform precipitation compared to the SM WSM scheme. TOM, only DM for rain with a unique snow PSD and diagnostic N_{0s}, shows incorrect size sorting signatures but still represents the stratiform precipitation region well. Qualitative and quantitative evaluation shows that WDM, despite being DM for rain, has a similar one-to-one relationship between Z and Z_{DR} as WSM and no stratiform precipitation development. The other DM schemes include more complex diagnostic equations (TOM) or are fully DM (MY and MOR), demonstrating that size sorting of hydrometeor categories in addition to rain is as important in improving the forecast microphysical state. TOM, MOR, and WDM all have incorrect Z_{DR} maxima associated with isolated, weak convection on the back side of convective lines where isolated large drops are not expected.

Notable biases are present in each scheme. Z and Z_{DR} in the stratiform precipitation region of the MCS are too high in TOM, MY, and particularly MOR, indicating that the forecast rain PSDs contain too many large drops for stratiform rain. The MY, MOR, and WDM forecasts contain a large amount of wet (melting) graupel in convective areas, as determined by the coexistence of rain and graupel in the model, while a hydrometeor classification algorithm (HCA) used indicates a little hail but mostly rain in similar locations in the observations. Although wet graupel is not included as a category in the classification scheme, significant graupel would not be expected near the surface for these warm season cases. These areas of wet graupel contribute to more extensive intense Z compared to the observations. MY includes a hail category but contains a similarly significant amount of graupel, likely due to a strict minimum hail size threshold in the scheme. Finally, simulated K_{DP} values are lower in all members for both cases, particularly in intense convective precipitation regions. K_{DP} increases with large amounts of moderate sized drops and higher liquid water contents,

but large raindrops and graupel with a low water ratio are apparent in TOM, MOR, and MY, while WDM and WSM have a bias toward small raindrops and graupel. The use of a triple-moment (TM) microphysics scheme with an effectively variable shape parameter would provide greater flexibility to represent a wider range of possible PSDs, including those that have a positive shape parameter (α_x) with a brace-like shape. This leads to a maximum of moderately-sized drops and higher liquid water contents/rain rates compared to an exponential distribution, which tends to underestimate liquid water content. Additionally, K_{DP} is a measurement related to mass in a volume. With a 4 km grid the volumes are quite large for calculating K_{DP}, which may vary greatly over a few km distance, and more localized maxima that may be present could be missed, particularly in intense convective precipitation areas.

There are several challenges inherent in large-domain convective-scale forecasts that can hamper our ability to gain information about the different microphysics schemes from the simulated variables. A poor forecast of storm structure for a given supercell or MCS will be missing notable polarimetric value patterns. For example, TOM, WDM, and WSM have poor supercell structures that make Z_{DR} arc comparisons more difficult. Previous studies have shown that forecasts performed using 4 km horizontal grid spacing may miss some fine-scale details in convection (Bryan et al. 2003), result in larger scale structures (Lean et al. 2008, Johnson et al. 2013), and impede processes such as development of trailing stratiform precipitation (Bryan and Morrison 2012; Xue et al. 2013). Other studies that have considered simulated polarimetric variables use a smaller grid scale than 4 km: 2 km in Putnam et al. (2014), 1 km in Jung et al. (2012), and 1 km in Li and Mecikalski (2012). These patterns may also be displaced compared to observations, making quantitative comparisons difficult.

Forecast members show some skill in terms of the fractions skill score (FSS) for Z and K_{DP} in the MCS case but higher scores require larger radii, and all forecasts exhibit very poor skill for Z_{DR} in both cases. Although qualitative comparisons indicate that MY and MOR represent Z_{DR} patterns relatively well, substantial spatial error leads to FSS scores with no skill. Normalizing the FSS using percentile values results in a significant improvement in skill for foreasts that do not contain simulated values as high as the observations. Future studies should continue to adapt these methods as forecasts are refined and improved before general statistics can be produced for all forecasts over the Spring Experiment period. Such information could be used in the future to provide additional forecast products as well as serve research purposes like determining which MP scheme may best represent polarimetric signatures in supercells for use in dual-pol data assimilation experiments.

Finally, we point out that there are also many uncertainties with the polarimetric radar simulator. There are various assumptions made on the water drop aspect ratio, canting angle of snow, hail, and graupel, and water fraction for mixed phase species. These are some of the aspects that still need refinement and tuning, and they can affect the microphysics evaluation. Dawson et al. (2014) developed an alternative water fraction model for the mixed phases that depends on the size spectrum. The relative performance of this model should be evaluated in the future.

Acknowledgements: This research was primarily supported by NSF grant AGS-1046171. The first and fourth authors were also supported by NOAA grant NA11OAR4320072. The second and third authors were also supported by NSF grant AGS-1261776. The CAPS Spring Experiment ensemble forecasts and fifth author were supported by a NOAA Collaborative Science and Technology Applied Research (CSTAR) Grant NA13NWS4680001, and were produced on a supercomputer operated by NICS at the University of Tennessee as part of the NSF XSEDE resources. A research grant from the Korea Meteorological Administration supported improvements to the polarimetric radar data

simulator. Detailed comments from Dr. Edward Mansell and two anonymous reviewers significantly improved the original manuscript.

References

Brandes, E. A., G. Zhang, and J. Vivekanandan, 2002: Experiments in rainfall estimation with a polarimetric radar in a subtropical environment. *J. Appl. Meteor.*, **41,** 674-685.

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

Bryan, G. H. and H. Morrison, 2012: Sensitivity of a Simulated Squall Line to Horizontal Resolution and Parameterization of Microphysics. *Mon. Wea. Rev.*, **140,** 202–225.

Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131,** 2394-2416.

Chen, F. and J. Dudhia, 2001: Coupling an Advanced Land Surface-Hydrology Model with the Penn State-NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. *Mon. Wea. Rev.*, **129,** 569.

Cintineo, R., J. A. Otkin, M. Xue, and F. Kong, 2014: Evaluating the Performance of Planetary Boundary Layer and Cloud Microphysical Parameterization Schemes in Convection-Permitting Ensemble Forecasts Using Synthetic GOES-13 Satellite Observations. *Mon. Wea. Rev.*, **142,** 163–182.

Clark, A. J., S. J. Weiss, J. S. Kain, I. L. Jirak, M. Coniglio, C. J. Melick, C. Siewert, R. A. Sobash, P. T. Marsh, A. R. Dean, M. Xue, F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, J. Du, D. R. Novak, F. E. Barthold, M. J. Bodner, J. J. Levit, C. B. Entwistle, T. L. Jensen, and J. Correia, 2012: An Overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bulletin of the American Meteorological Society***,** 10.1175/bams-d-11-00040.1.

Dawson, D. T., II. , E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2014: Low-level ZDR signatures in supercell forward flanks: The role of size sorting and melting of hail. *J. Atmos. Sci.*, **71,** 276-299.

Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New dimension of NCEP Short-Range Ensemble Forecasting (SREF) system: Inclusion of WRF members. *Preprint, WMO Expert Team Meeting on Ensemble Prediction System*, Exeter, UK, 5pp.

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteorological Applications*, **15,** 51-64.

Fritsch, J. M. and G. S. Forbes, 2001: Mesoscale convective systems. *Meteor. Monogr.*, **50,** 323-358.

Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Ocean. Technol.*, **21,** 457-469.

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.*, **125,** 1414-1430.

Hong, S.-Y. and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42,** 129-151.

Hu, M., M. Xue, and K. Brewster, 2006a: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134,** 675-698.

Hu, M., M. Xue, J. Gao, and K. Brewster, 2006b: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134,** 699-721.

Janjic, Z., 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Mesomodel. NCEP Office Note 437, NOAA/NWS, 61.

Johnson, A., X. Wang, F. Kong, and M. Xue, 2013: Object-Based Evaluation of the Impact of Horizontal Grid Spacing on Convection-Allowing Forecasts. *Mon. Wea. Rev.*, **141,** 3413–3425.

Johnson, M., Y. Jung, D. Dawson, and M. Xue, 2016: Comparison of simulated polarimetric signatures in idealized supercell storms using two-moment bulk microphysics schemes in WRF. *Mon. Wea. Rev.*, **144,** 971-996.

Jung, Y., G. Zhang, and M. Xue, 2008a: Assimilation of simulated polarimetric radar data for a convective storm using ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Wea. Rev.*, **136,** 2228-2245.

Jung, Y., M. Xue, G. Zhang, and J. Straka, 2008b: Assimilation of simulated polarimetric radar data for a convective storm using ensemble Kalman filter. Part II: Impact of polarimetric data on storm analysis. Mon. Wea. Rev., **136**, 2246–2260.

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.

Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 29-30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric data. *Mon. Wea. Rev.*, **140,** 1457-1475.

Kong, F., 2013: 2013 CAPS Spring Forecast Experiment Program Plan, 24 pp.

Kong, F., M. Xue, D. Bright, M. C. Coniglio, K. W. Thomas, Y. Wang, D. Weber, J. S. Kain, S. J. Weiss, and J. Du, 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. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 3B.2.

Kumjian, M. R. and A. V. Ryzhkov, 2008: Polarimetric signatures in supercell thunderstorms. *J. Appl. Meteor. Climatol.*, **47,** 1940-1961.

Kumjian, M. R. and A. V. Ryzhkov, 2012: The Impact of Size Sorting on the Polarimetric Radar Variables. *J. Atmos. Sci.*, **69,** 2042–2060.

Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. Stumpf, 2007: An Automated Technique to Quality Control Radar Reflectivity Data. *J. Appl. Meteor. Climatol.*, **46,** 288-305.

Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of High-Resolution Versions of the Met Office Unified Model for Forecasting Convection over the United Kingdom. *Mon. Wea. Rev.*, **136,** 3408–3424.

Li, X. and J. R. Mecikalski, 2012: Impact of the Dual-Polarization Doppler Radar Data on Two Convective Storms with a Warm-Rain Radar Forward Operator. *Mon. Wea. Rev.*, **140,** 2147–2167.

Lim, K.-S. S. and S.-Y. Hong, 2010: Development of an Effective Double-Moment Cloud Microphysics Scheme with Prognostic Cloud Condensation Nuclei (CCN) for Weather and Climate Models. *Mon. Wea. Rev.*, **138,** 1587-1612.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climat. Appl. Meteor.*, **22,** 1065-1092.

Mahale, V., G. Zhang, and M. Xue, 2014: Fuzzy logic classification of S-band polarimetric radar echoes to identify three-body scattering and improve data quality. *J. Appl. Meteor. Clim.*, **53,** 2017-2033.

Mansell, E. R., 2010: On Sedimentation and Advection in Multimoment Bulk Microphysics. *J. Atmos. Sci.*, **67,** 3084–3094.
Maxwell-Garnett, J. C., 1904: Colors in metal glasses and in metallic films. *Philos. Trans. Roy. Soc. London*, **Ser. A,** 203.
Mellor, G. L. and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.* , **20,** 851-875.
Milbrandt, J. A. and M. K. Yau, 2005a: A multi-moment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62,** 3065-3081.
Milbrandt, J. A. and M. K. Yau, 2005b: A multi-moment bulk microphysics parameterization. Part I: Aanlysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62,** 3051-3064.
Mittermaier, M. and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Weather and Forecasting*, **25,** 343-354.
Morrison, H., J. A. Curry, and V. I. Khvorostyanov, 2005: A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. *J. Atmos. Sci.*, **62,** 1665-1677.
Morrison, H., 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.
NWS, cited 2014: The Tornado Outbreak of May 20, 2013. [Available online from http://www.srh.noaa.gov/oun/?n=events-20130520.]
Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. *Wea. Forecasting*, **24,** 730–748.
Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2014: The Analysis and Prediction of Microphysical States and Polarimetric Radar Variables in a Mesoscale Convective System Using Double-Moment Microphysics, Multinetwork Radar Data, and the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **142,** 141–162.
Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15,** 163-169.
Roberts, N. M. and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev*, **136,** 78-97.
ROC, N.: WSR-88D Dual Polarization Deployment Progress. [Available online from http://www.roc.noaa.gov/WSR88D/PublicDocs/DualPol/DPstatus.pdf.]
Rogers, E., G. DiMego, T. Black, M. Ek, B. Ferrier, G. Gayno, Z. Janjic, Y. Lin, M. Pyle, V. Wong, W. S. Wu, and J. Carley, 2009: The NCEP North American Mesoscale Modeling System : Recent changes and future plans. *23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*, Omaha, NE, 2A.4.
Ryzhkov, A. and D. S. Zrnic, 1996: Assessment of rainfall measurement that uses specific differential phase. *J. Appl. Meteor.*, **35,** 2080–2090.
Schwartz, C., J. Kain, S. Weiss, M. Xue, D. Bright, F. Kong, K. Thomas, J. Levit, and M. Coniglio, 2009: Next-day convection-allowing WRF model guidance: A second look at 2 vs. 4 km grid spacing. *Mon. Wea. Rev.*, **137,** 3351-3372.
Skamarock, W. and Coauthors: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-4751STR. [Available online from www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]
Smith, T., V. Lakshmanan, G. Stumpf, K. Ortega, K. Hondl, K. Cooper, K. Calhoun, D. Kingfield, K. Manross, R. Toomey, and J. Brogden, 2016: Multi-radar Multi-sensor (MRMS) Severe

Weather and Aviation Products: Initial Operating Capabilities. *Bull. Amer. Meteor. Soc.* **,** in press.

SPC, cited 2014: SPC Filtered Storm Reports for 5/19/2013. [Available online from http://www.spc.noaa.gov/climo/reports/130519_rpts.html.]

SPC, cited 2014: SPC Filtered Storm Reports for 5/20/2013. [Available online from http://www.spc.noaa.gov/climo/reports/130520_rpts.html.]

Sun, J. and N. A. Crook, 2001: Real-time low-level wind and temperature analysis using single WSR-88D data. *Wea. Forecasting*, **16,** 117-132.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136,** 5095-5115.

Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distributions. *J. Appl. Meteor.*, **22,** 1764-1775.

Van Weverberg, K., A. M. Vogelmann, H. Morrison, and J. A. Milbrandt, 2012: Sensitivity of idealized squall-line simulations to the level of complexity used in two-moment bulk microphysics schemes. *Mon. Wea. Rev.*, **140,** 1883-1907.

Vivekanandan, J., W. M. Adams, and V. N. Bringi, 1991: Rigorous Approach to Polarimetric Radar Modeling of Hydrometeor Orientation Distributions. *J. Appl. Meteor.*, **30,** 1053–1063.

Wacker, U. and A. Seifert, 2001: Evolution of rain water profiles resulting from pure sedimentation: spectral vs. parameterized description. *Atmos. Res.*, **58,** 19-39.

Weiss, S. J., J. S. Kain, D. R. Bright, J. J. Levit, G. W. Carbin, M. E. Pyle, Z. I. Janjic, B. S. Ferrier, J. Du, M. L. Weisman, and M. Xue, 2007: The NOAA Hazardous Weather Testbed: Collaborative testing of ensemble and convection-allowing WRF models and subsequent transfer to operations at the Storm Prediction Center. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 6B.4.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman Filter Analyses and Forecasts of a Severe Mesoscale Convective System Using Different Choices of Microphysics Schemes. *Monthly Weather Review*, **142,** 3243-3263.

Xue, M., M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square-root Kalman filter for evaluating impact of data from radar networks on thunderstorm analysis and forecast. *J. Atmos. Ocean Tech.*, **23,** 46–66.

Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82,** 139-170.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, and K. K. Droegemeier, 2013: Prediction of Convective Storms at Convection-Resolving 1 km Resolution over Continental United States with Radar Data Assimilation: An Example Case of 26 May 2008 and Precipitation Forecasts from Spring 2009. *Advances in Meteorology*, **2013,** 9.

Xue, M., F. Kong, D. Weber, K. W. Thomas, Y. Wang, K. Brewster, K. K. Droegemeier, J. S. K. S. J. Weiss, D. R. Bright, M. S. Wandishin, M. C. Coniglio, and J. Du, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Amer. Meteor. Soc., CDROM 3B.1.

Zhang, G., M. Xue, Q. Cao, and D. Dawson, 2008: Diagnosing the intercept parameter for exponential raindrop size distribution based on video disdrometer observations. *J. Appl. Meteor. Climatol.*, **47,** 2983-2992.

Scheme	$\mathbf{N_{0r}}$	$\mathbf{N_{0s}}$	$\mathbf{N_{0g}}$	$\mathbf{N_{0h}}$	$\boldsymbol{\alpha_r}$	$\boldsymbol{\alpha_s}$	$\boldsymbol{\alpha_g}$	$\boldsymbol{\alpha_h}$
Thompson	F(q_r, N_{tr})	F(q_s,T)	F(q_g,T)	NA	0	0.6357	0	NA
MY	F(q_r, N_{tr})	F(q_s, N_{ts})	F(q_g, N_{tg})	F(q_h, N_{th})	0	0	0	0
Morrison	F(q_r, N_{tr})	F(q_s, N_{ts})	F(q_g, N_{tg})	NA	0	0	0	NA
WDM6	F(q_r, N_{tr})	F(T)	4×10^6 m^{-4}	NA	1	0	0	NA
WSM6	8×10^6 m^{-4}	F(T)	4×10^6 m^{-4}	NA	0	0	0	NA

Table 1: Reference table for which predicted model variables are used to calculate the intercept (N_0) and shape (α) parameters for rain (r), snow (s), graupel (g), and hail (h) from each of the microphysics scheme considered. The values are either fixed and listed, or calculated as a function of mixing ratio (q), number concentration (N_t), and/or air temperature (T). Schemes which do not contain a given hydrometeor category are listed as not applicable (NA).

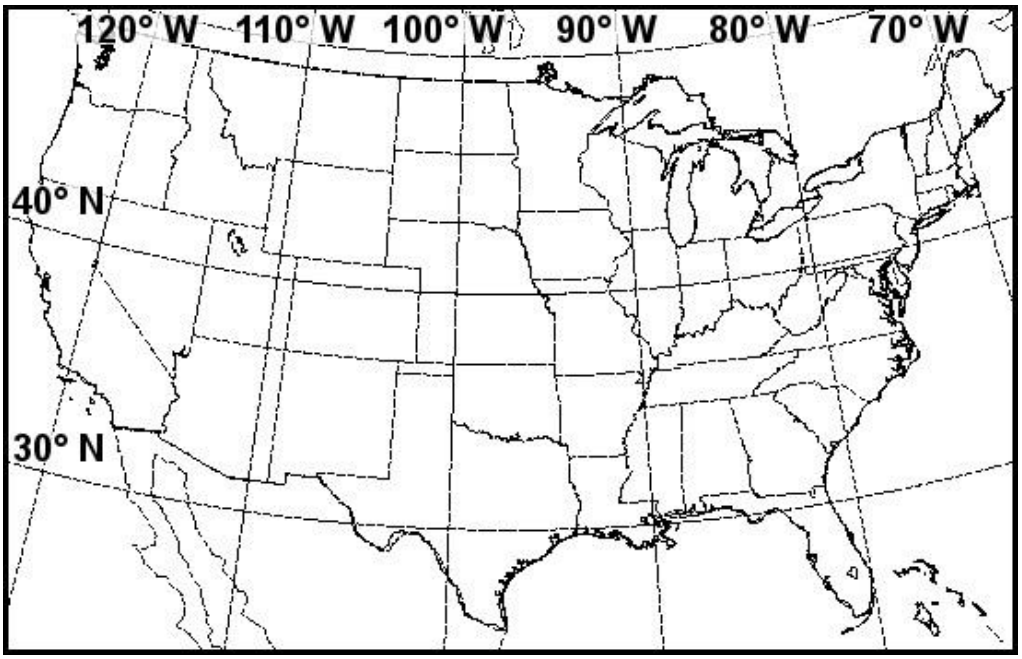

Fig. 1. Model domain for 2013 CAPS Spring Experiment storm scale ensemble forecasts.

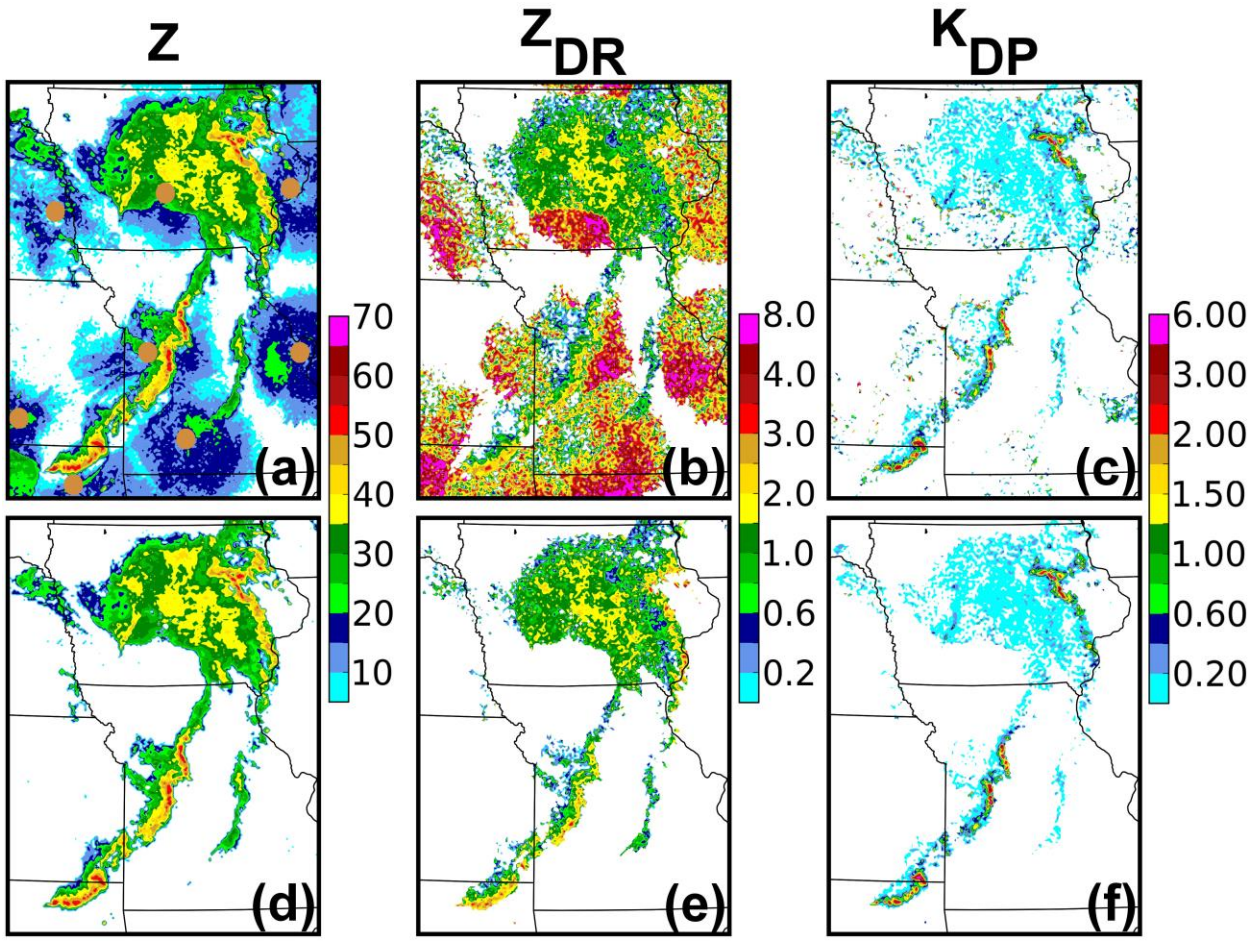

Fig. 2. Mosaics of observed (a) reflectivity (dBZ), (b) differential reflectivity (dB), and (c) specific differential phase (° km^{-1}) at a .5° tilt before ground clutter/biological scatterer removal and (d-f) after. The locations of the WSR-88D radars used are included as tan dots in (a).

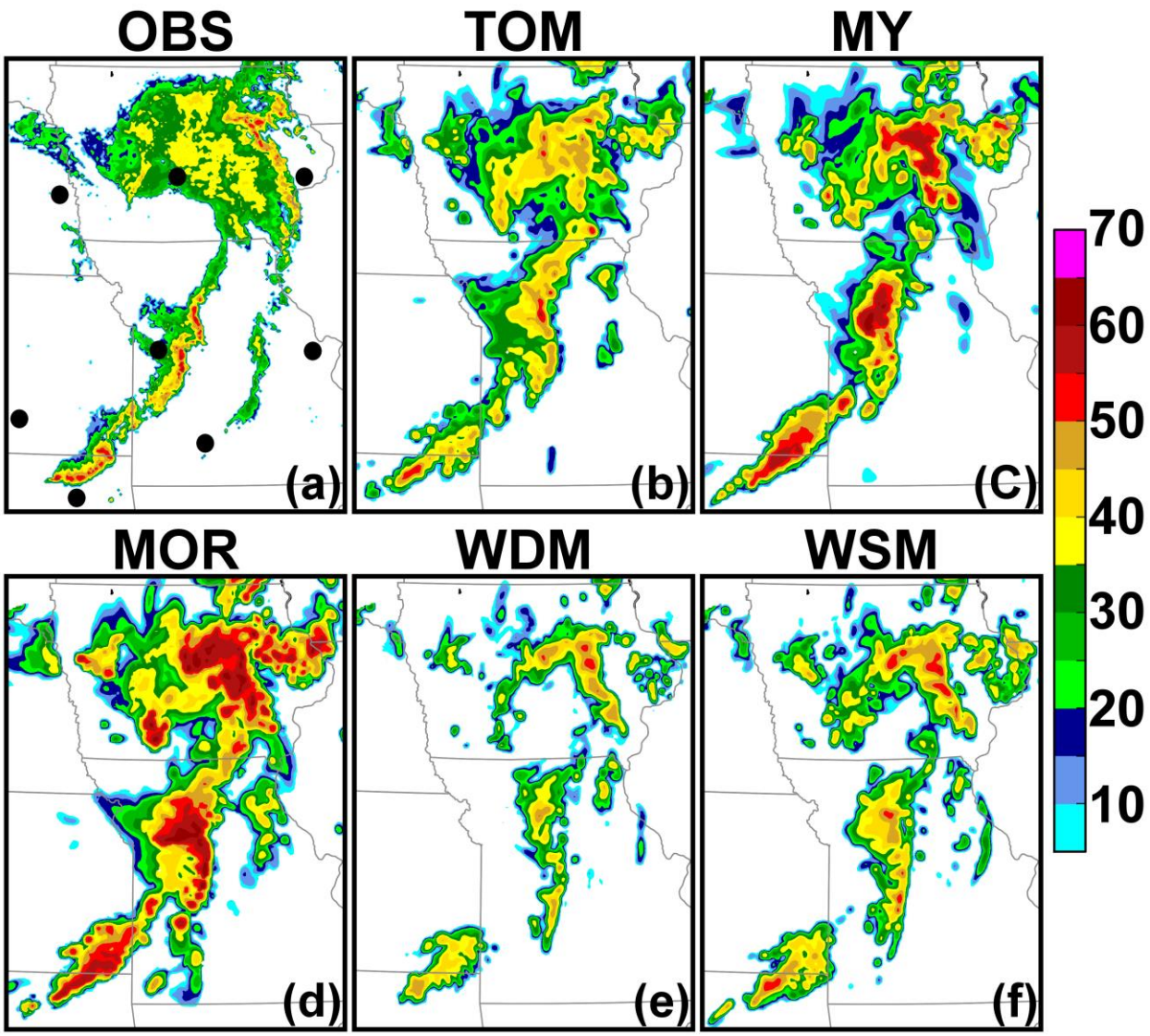

Fig. 3. Mosaics of observed (a) reflectivity (dBZ) at a .5° tilt 0400 UTC 20 May 2013 and simulated values at the same tilt locations from the (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

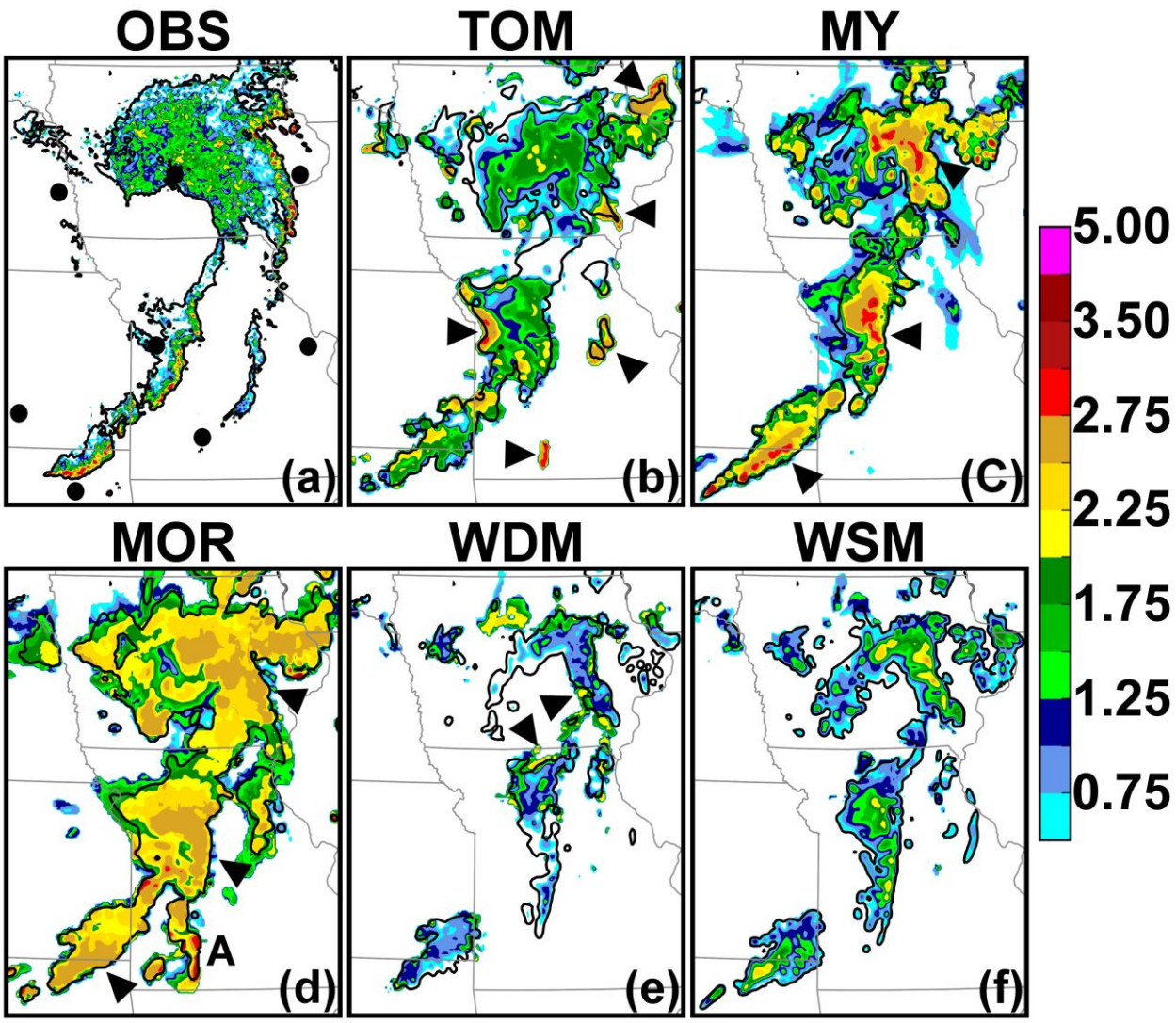

Fig. 4. Mosaics of observed (a) differential reflectivity (dB) at a .5° tilt 0400 UTC 20 May 2013 and simulated values at the same tilt locations from the (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Features of interest referenced in the text are noted by arrows. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

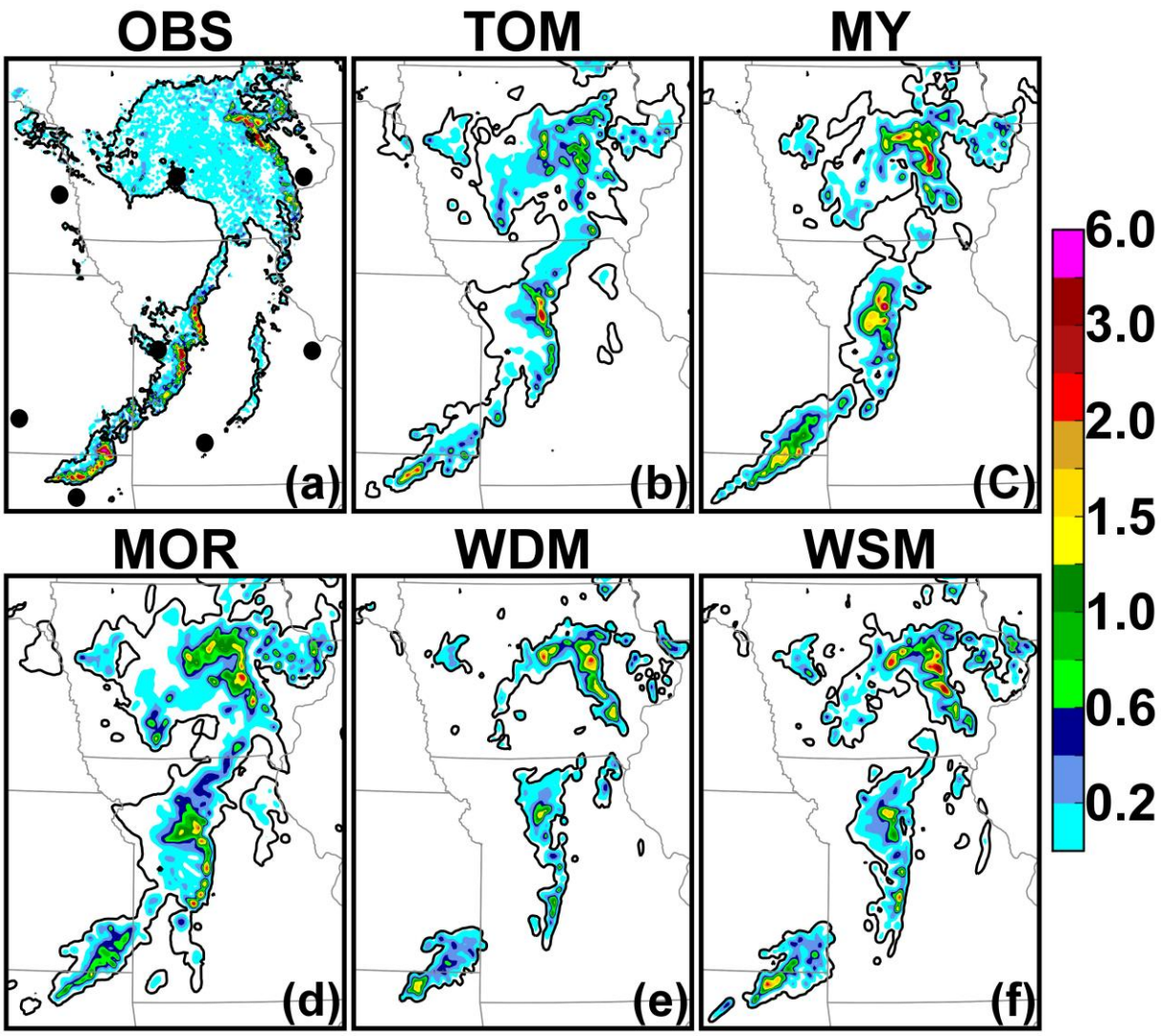

Fig. 5. Mosaics of observed (a) specific differential phase (° km^{-1}) at a .5° tilt 0400 UTC 20 May 2013 and simulated values at the same tilt locations from the (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

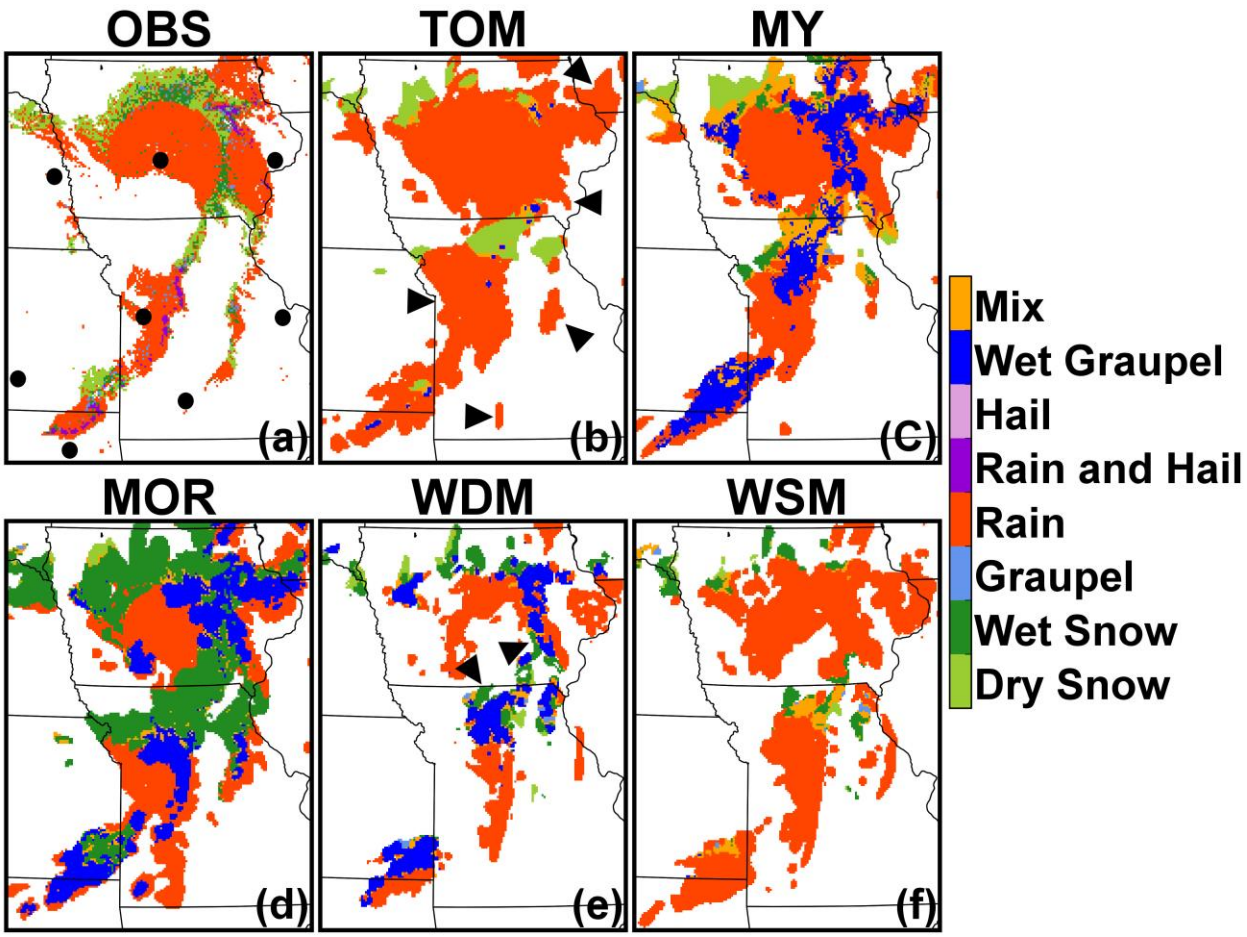

Fig. 6. Mosaic of hydrometeor classification using fuzzy logic for (a) observations at a .5° tilt 0400 UTC 20 May 2013 as well as classification of highest simulated linear reflectivity at the same tilt locations for (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Features of interest referenced in the text are noted by arrows. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

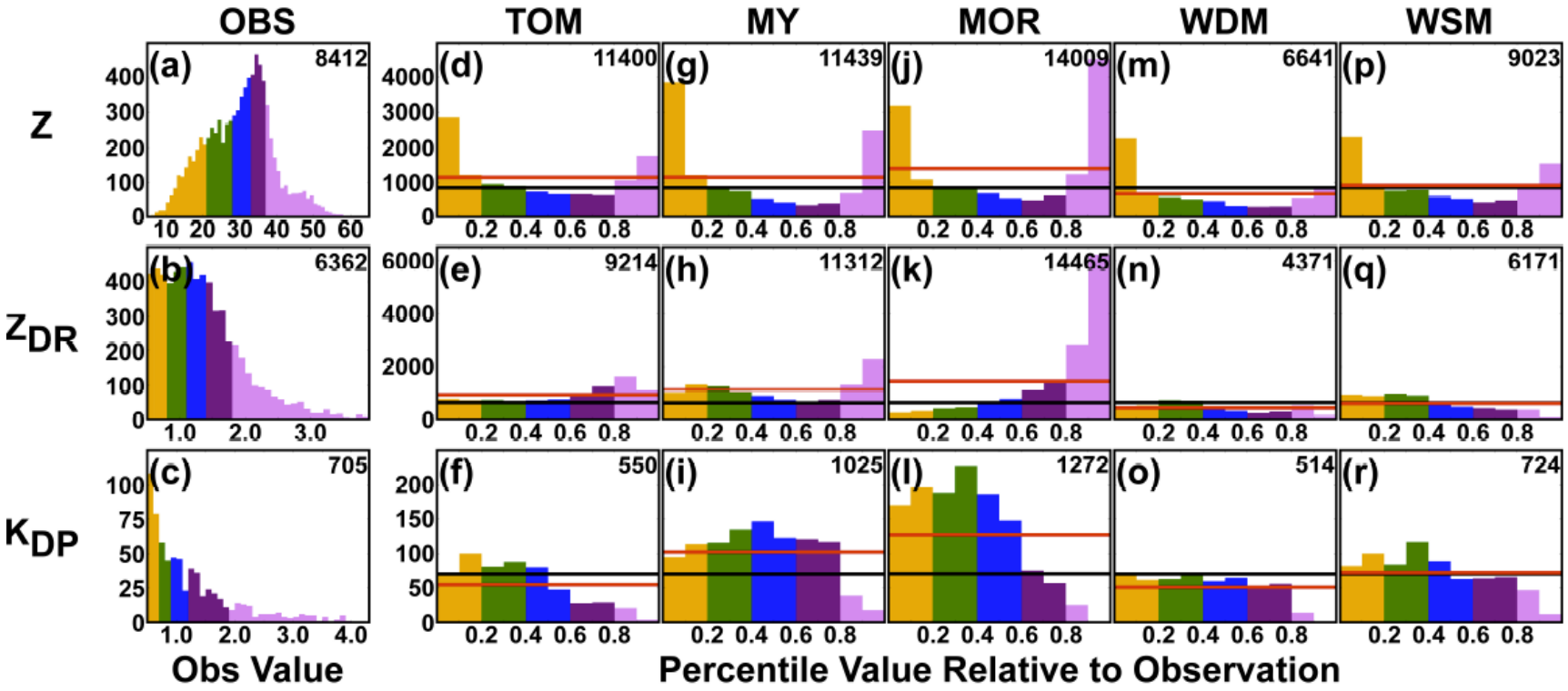

Fig. 7. Histograms of observed (a) reflectivity (dBZ), (b) differential reflectivity (dB), and (c) specific differential phase (° km^{-1}) mosaic values at a .5° tilt from Figs. 3, 4, and 5 as well as percentile histograms of simulated values at the same tilt locations from the (d-f) TOM, (g-i) MY, (j-l) MOR, (m-o) WDM, and (p-r) WSM forecasts distributed into bins based on the observed percentiles (noted by the solid black line). The idealized distribution given the number of values in each of the simulated results is given by the red line. The total number of values is noted in the upper right hand corner of each subplot. The observed values used as the bounds for the percentile bins are color coded between the observed and simulated value plots.

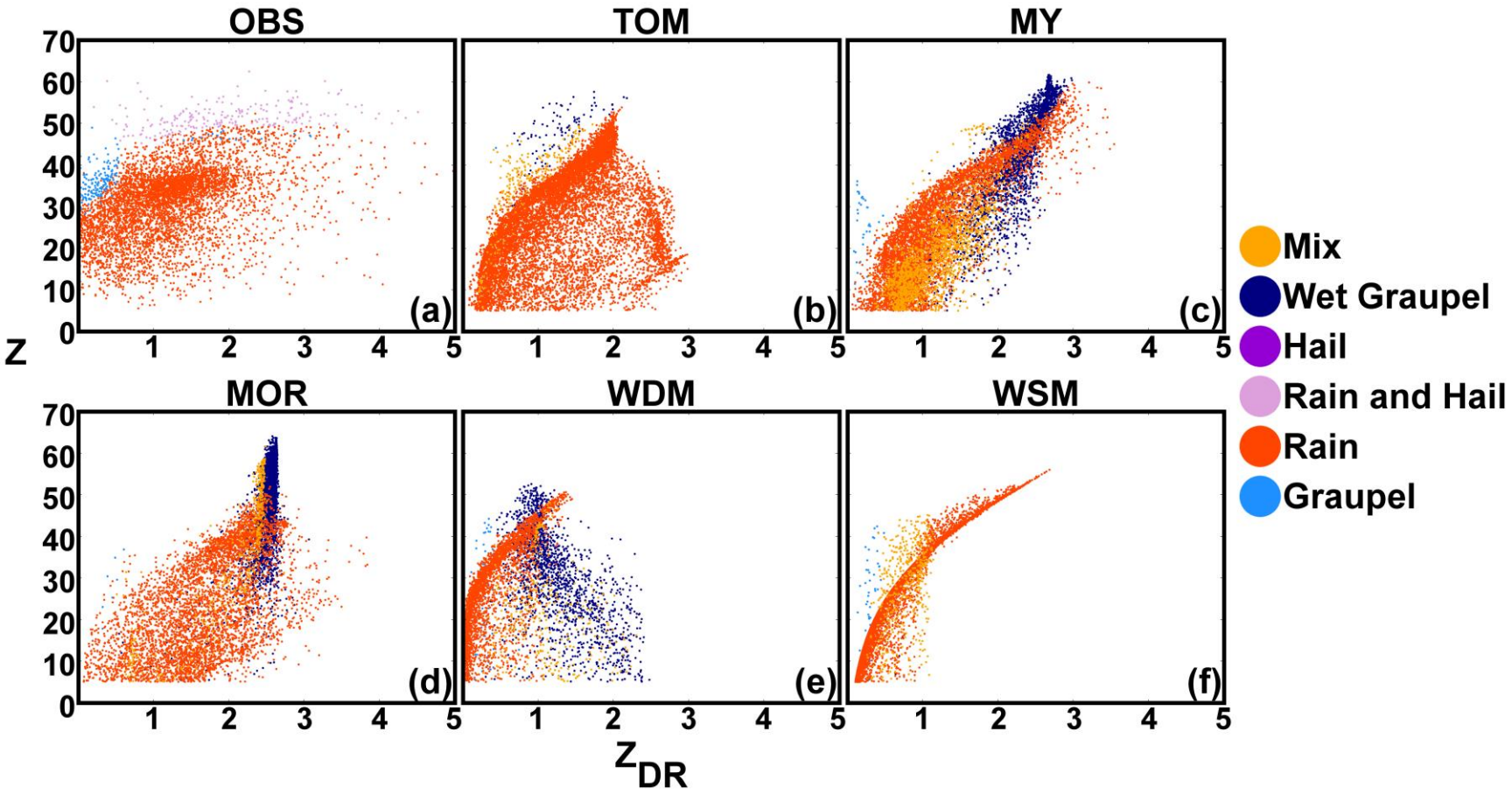

Fig. 8. Scatter plot of (a) observed reflectivity (dBZ) and differential reflectivity (dB) mosaic values at a .5° tilt from Figs. 3, 4, and 5 as well as scatter plots of simulated values at the same tilt locations from the (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts.

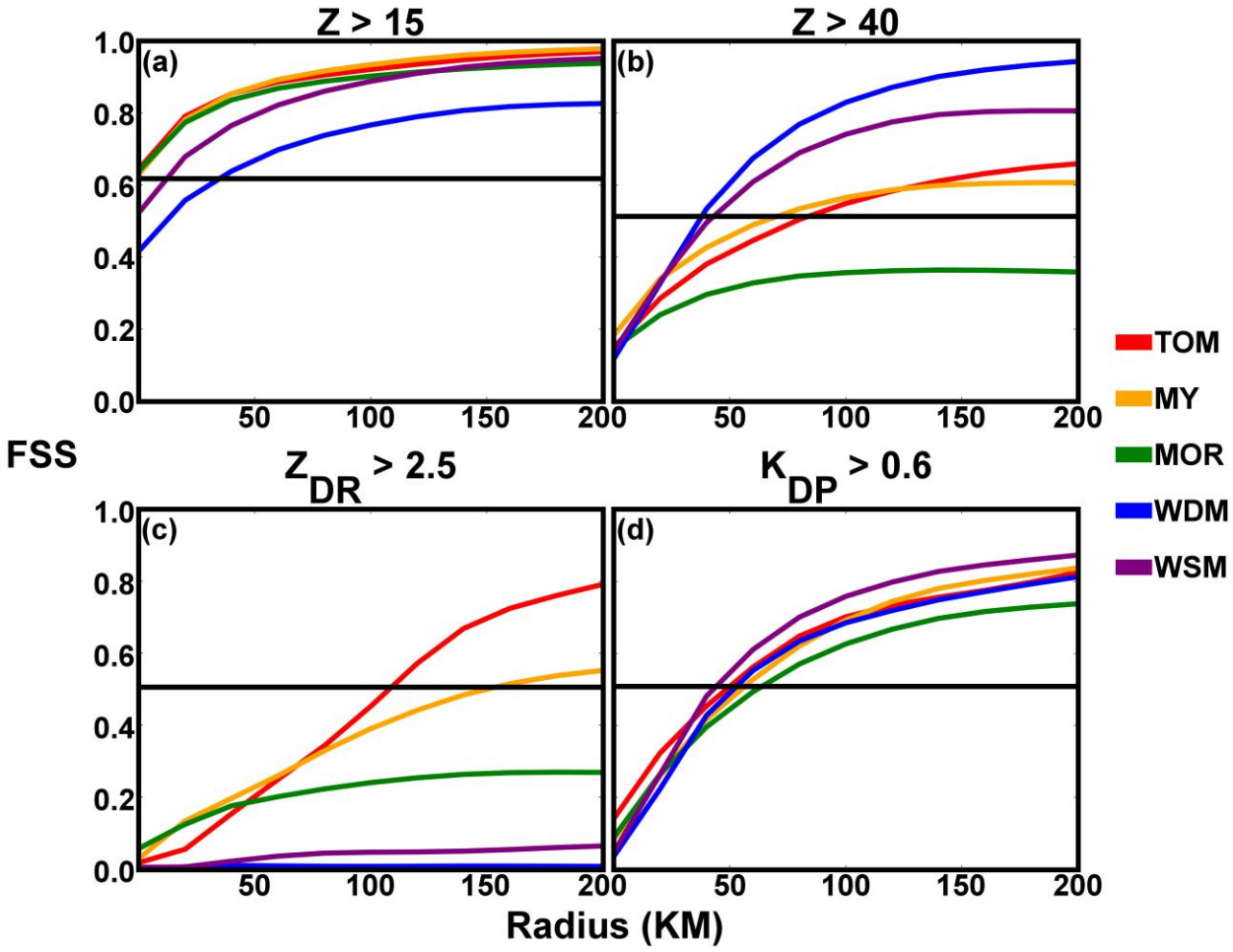

Fig. 9. Fractions skill scores for the TOM, MY, MOR, WDM, and WSM forecast results at increasing neighborhood radii for (a) reflectivity values exceeding 15 dBZ, (b) reflectivity values exceeding 40 dBZ, (c) differential reflectivity values exceeding 2.5 dB, and (d) specific differential phase values exceeding 0.6 ° km^{-1} for the mosaics in Figs. 3, 4, and 5. The black line indicates skill greater than a random forecast.

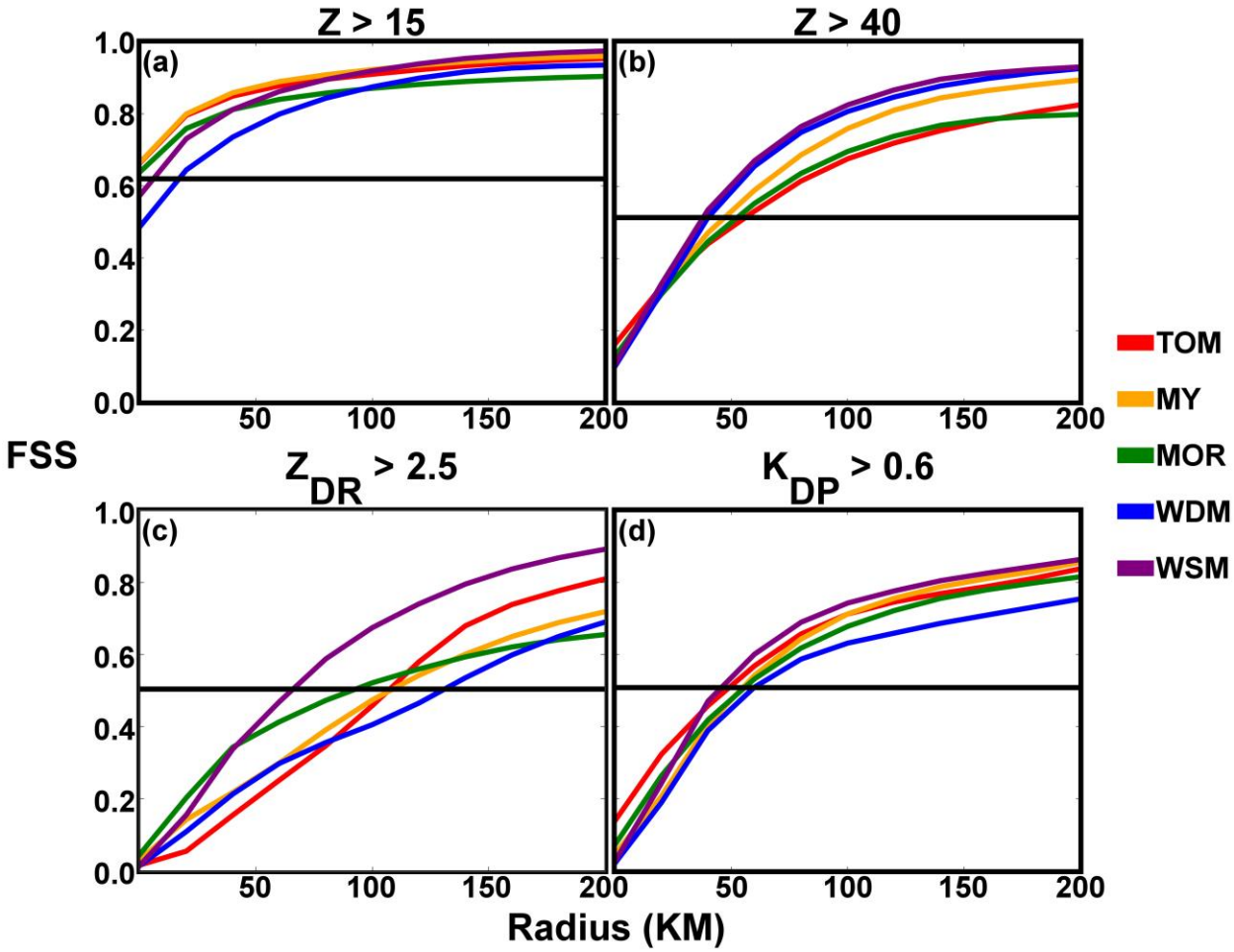

Fig. 10. As in Fig. 9, but with the FSS scores calculated based on percentile values relative to the observations.

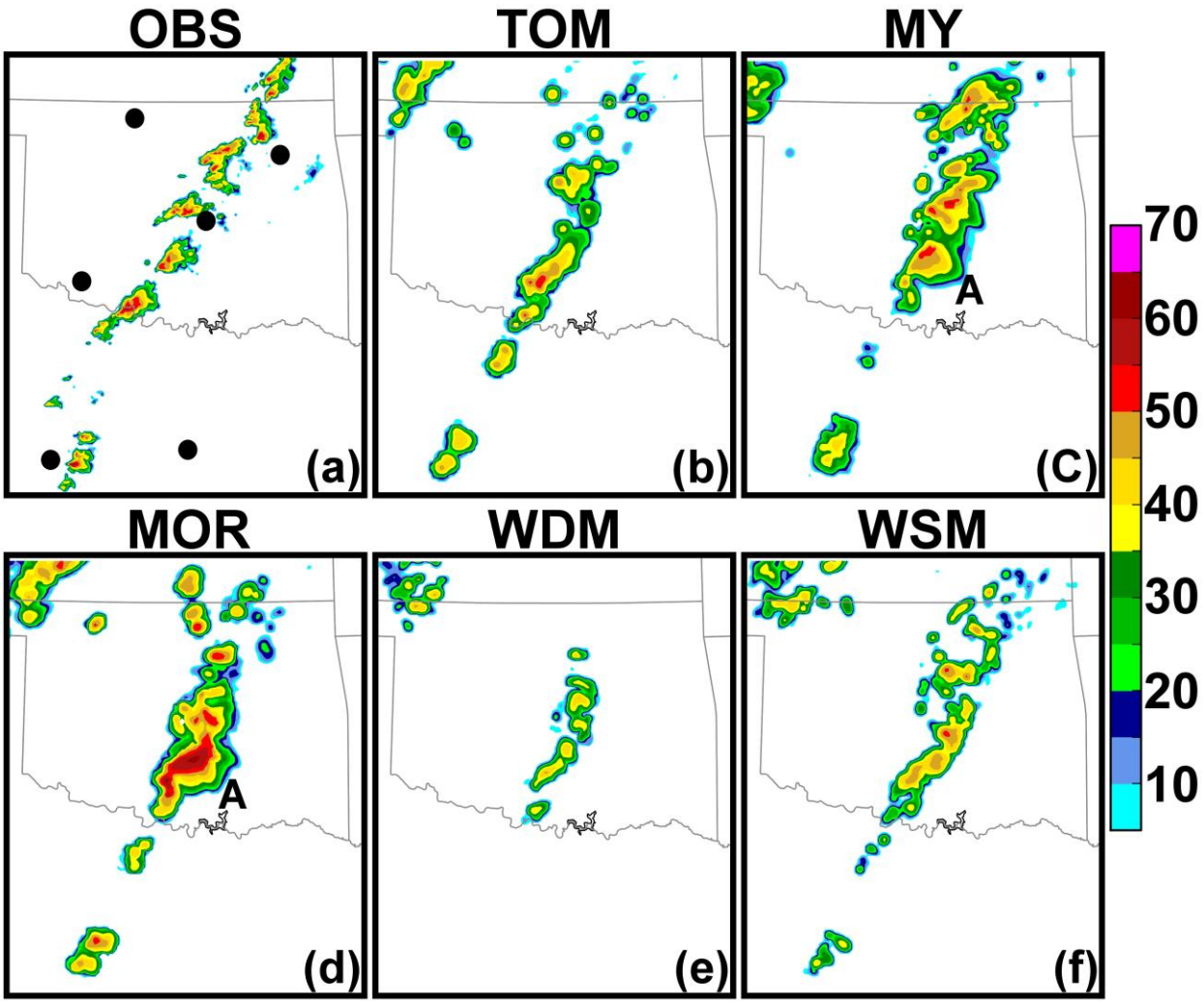

Fig. 11. Mosaics of observed (a) reflectivity (dBZ) at a .5° tilt 2100 UTC 20 May 2013 and simulated values at the same tilt locations from the (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Features of interest referenced in the text are noted by letters. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

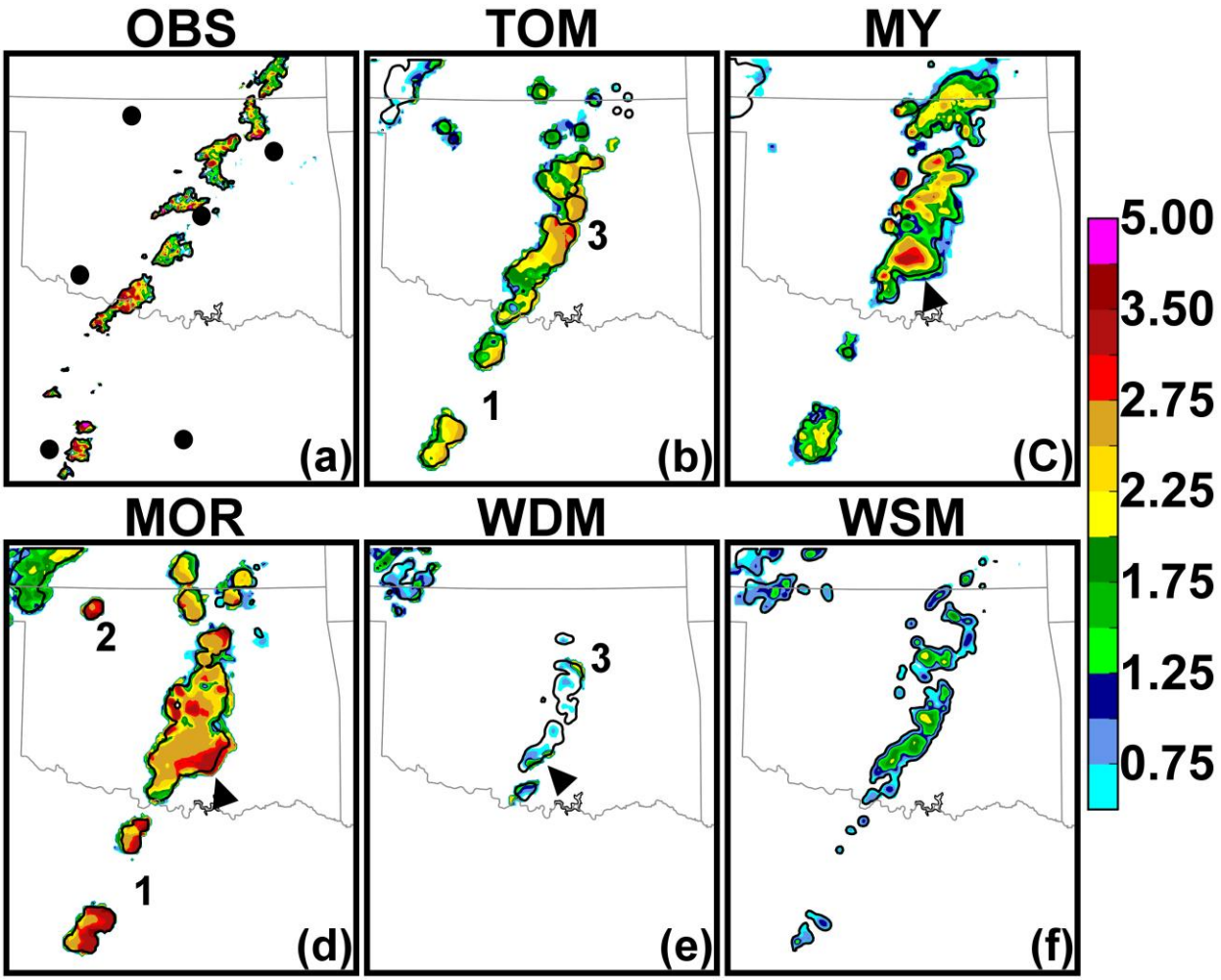

Fig. 12. Mosaics of observed (a) differential reflectivity (dB) at a .5° tilt 2100 UTC 20 May 2013 and simulated values at the same tilt locations from the (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Features of interest referenced in the text are noted by arrows and numbers. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

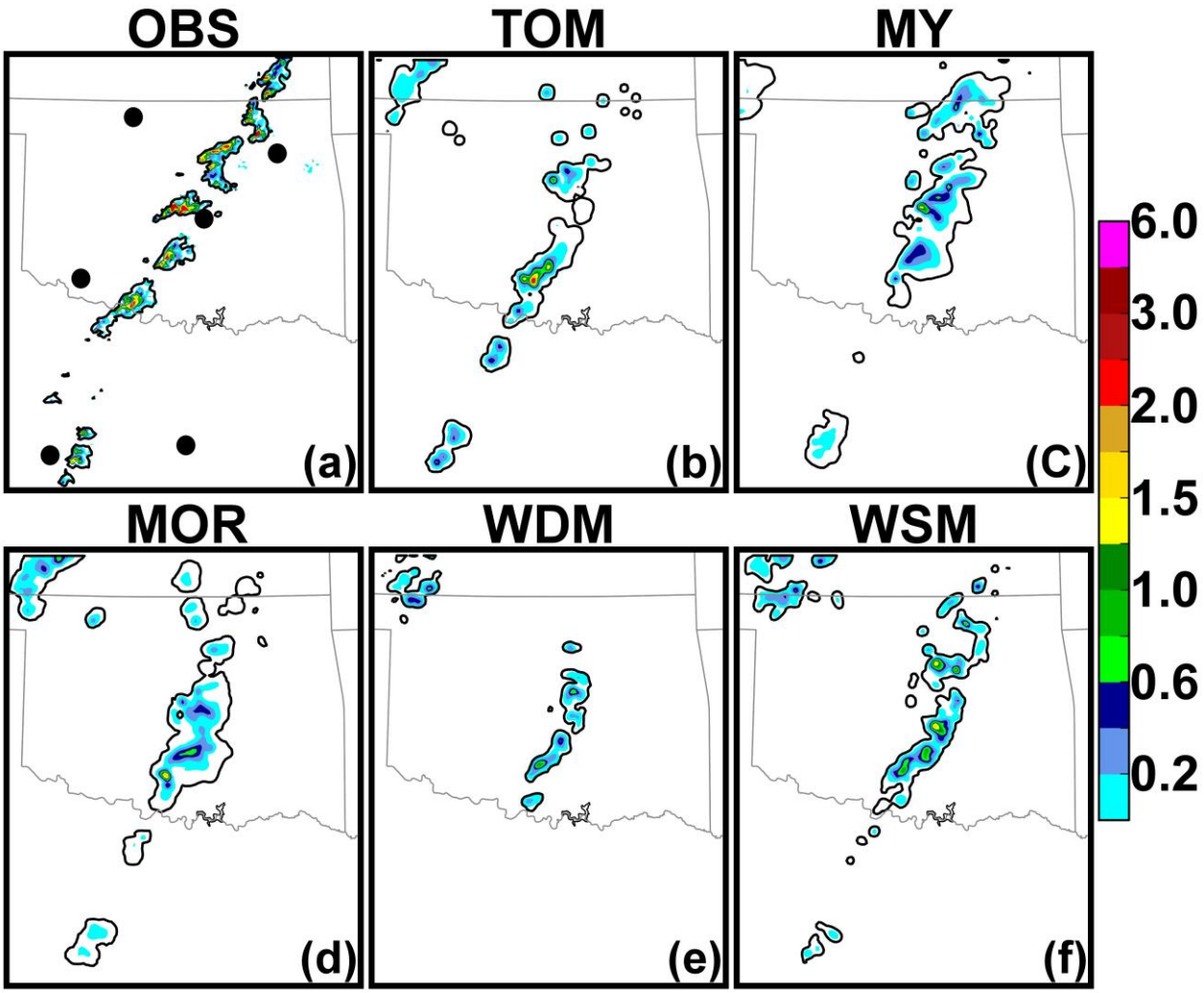

Fig. 13. Mosaics of observed (a) specific differential phase (° km^{-1}) at a .5° tilt 2100 UTC 20 May 2013 and simulated values at the same tilt locations from the (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

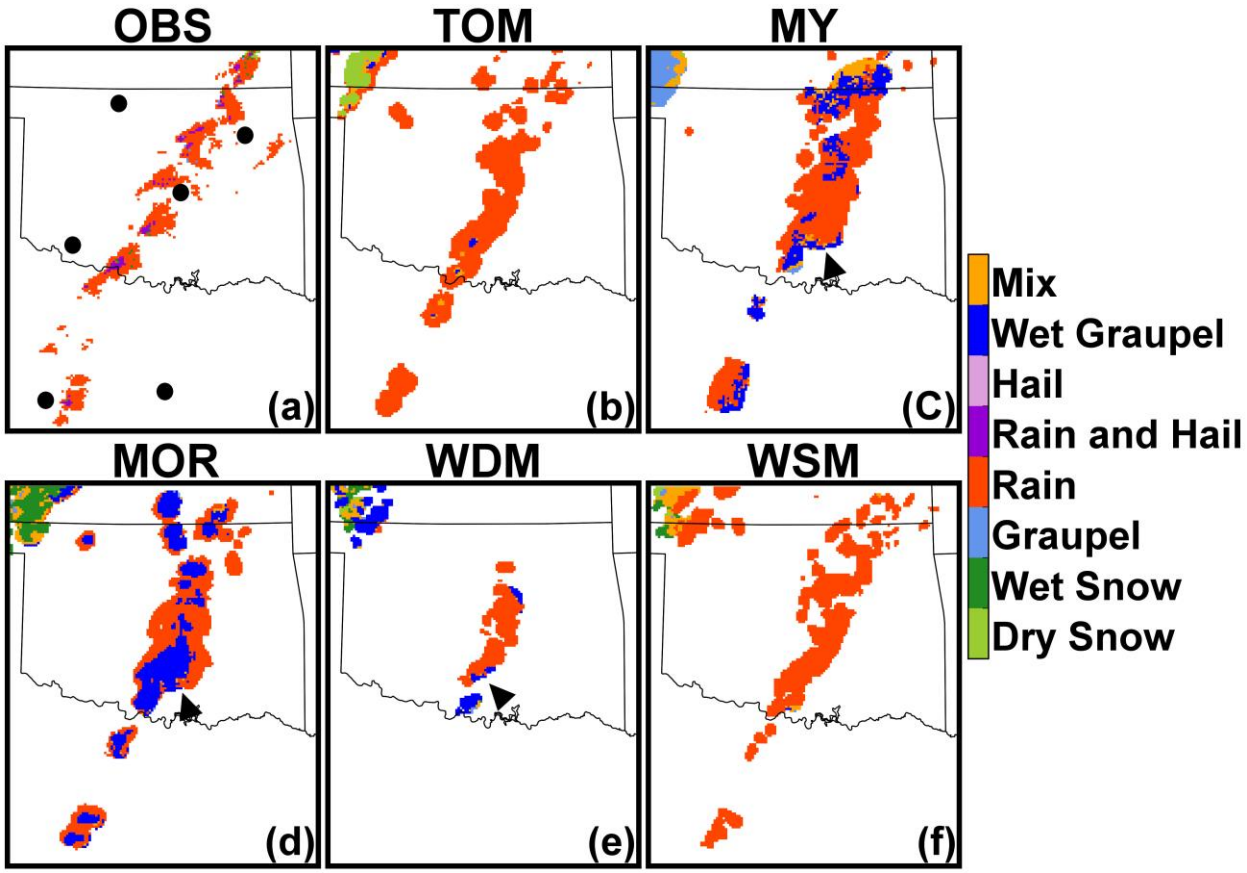

Fig. 14. Mosaic of hydrometeor classification using fuzzy logic for (a) observations at a .5° tilt 2100 UTC 20 May 2013 as well as classification of highest simulated linear reflectivity at the same tilt locations for (b) TOM, (c) MY, (d) MOR, (e) WDM, and (f) WSM forecasts. Features of interest referenced in the text are noted by arrows. Locations of WSR-88D sites used for both the observed and simulated variable plots are noted with black dots in (a).

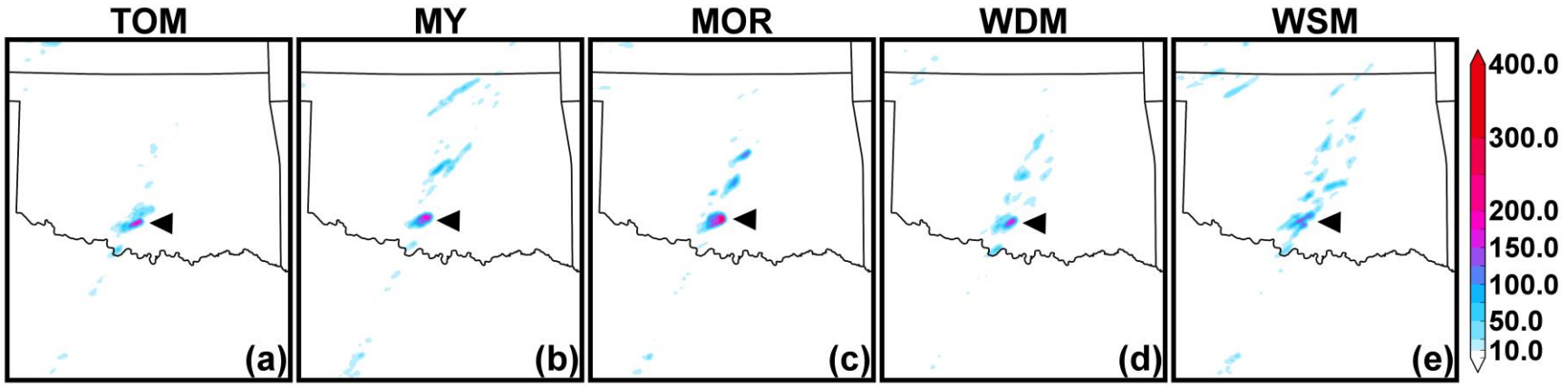

Fig. 15. 2D plots of maximum updraft helicity ($m^2 s^{-2}$) at 2100 UTC 20 May 2013 for (a) TOM, (b) MY, (c) MOR, (d) WDM, and (e) WSM. Features of interest referenced in the text are noted by arrows.

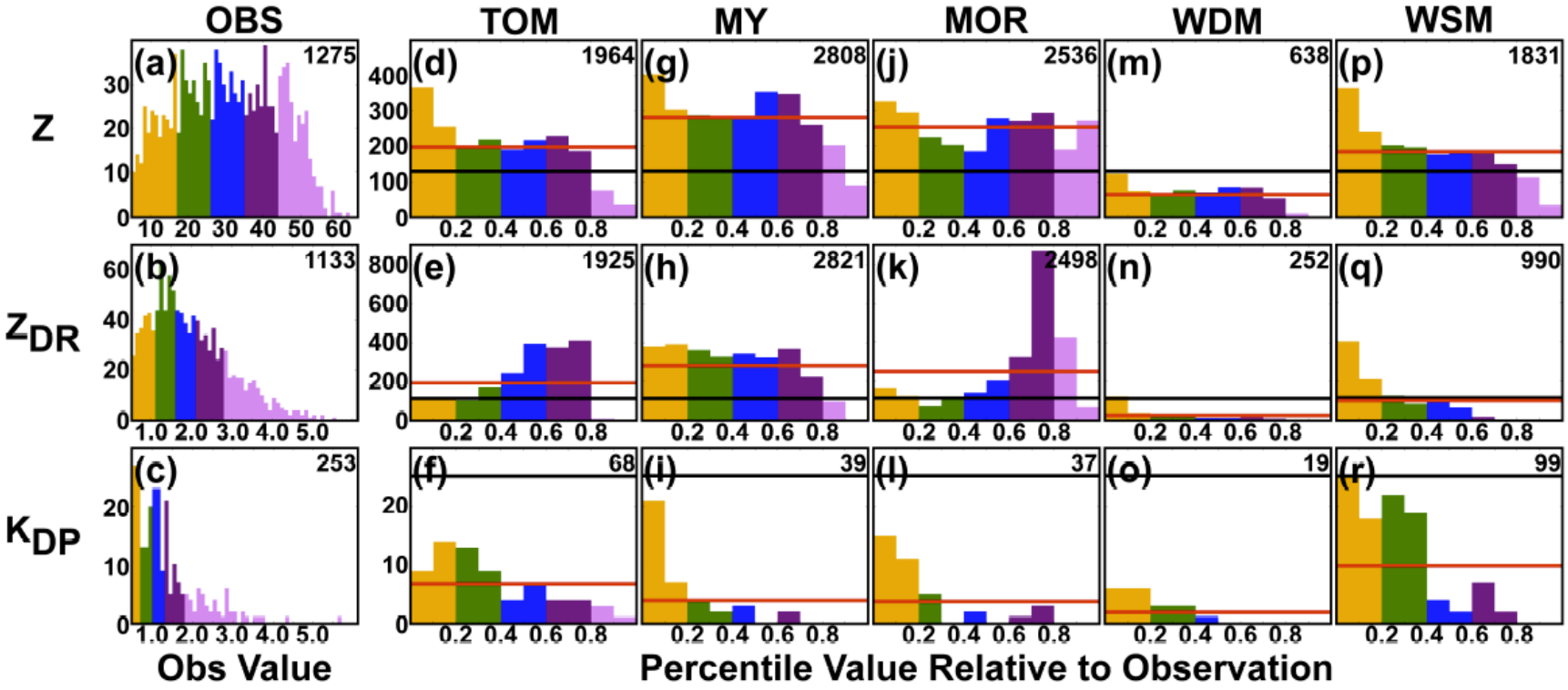

Fig. 16. Histograms of observed (a) reflectivity (dBZ), (b) differential reflectivity (dB), and (c) specific differential phase (° km^{-1}) mosaic values at a .5° tilt from Figs. 11, 12, and 13 as well as percentile histograms of simulated values at the same tilt locations from the (d-f) TOM, (g-i) MY, (j-l) MOR, (m-o) WDM, and (p-r) WSM forecasts distributed into bins based on the observed percentiles (noted by the solid black line). The idealized distribution given the number of values in each of the simulated results is given by the red line. The total number of values is noted in the upper right hand corner of each subplot. The observed values used as the bounds for the percentile bins are color coded between the observed and simulated value plots.

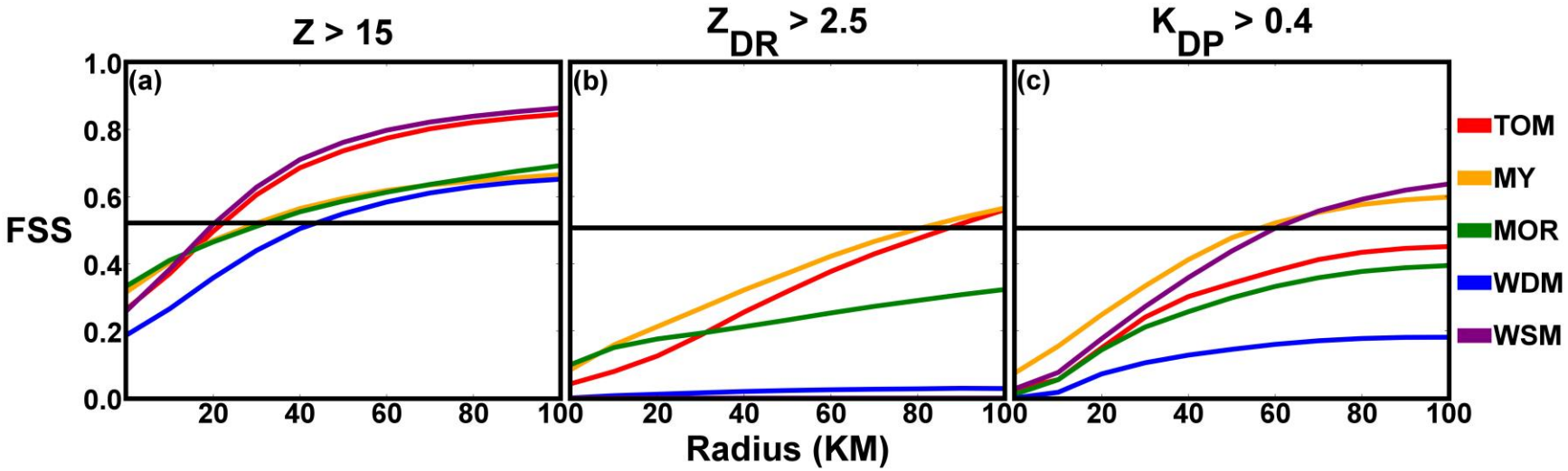

Fig. 17. Fractions skill scores for the TOM, MY, MOR, WDM, and WSM forecast results at increasing neighborhood radii for (a) reflectivity values exceeding 15 dBZ, (b) differential reflectivity values exceeding 2.5 dB, and (c) specific differential phase values exceeding 0.4 ° km^{-1} for the mosaics in Figs. 11, 12, and 13. The black line indicates skill greater than a random forecast.

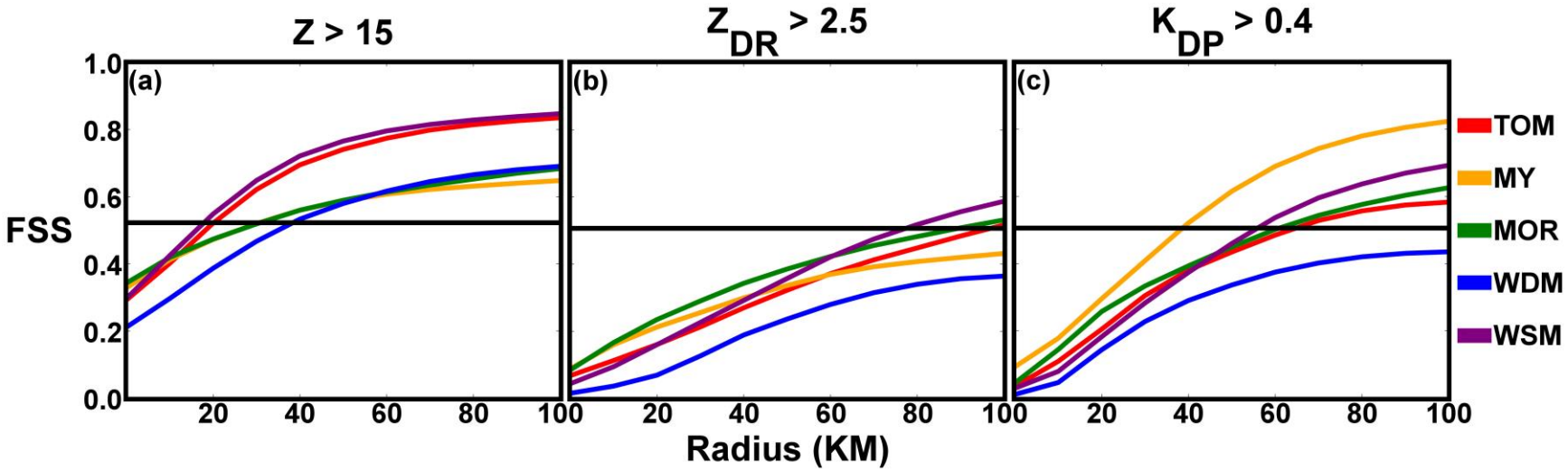

Fig. 18. As in Fig. 17, but with the FSS scores calculated based on percentile values relative to the observations.