Morris, M. T., K. A. Brewster, and F. H. Carr, 2021: Assessing the impact of non-conventional radar and surface observations on high-resolution analyses and forecasts of a severe hailstorm. *Electronic J. Severe Storms Meteor.*, **16** (1), 1–39.

Electronic Journal of
SEVERE STORMS METEOROLOGY

Assessing the Impact of Non-Conventional Radar and Surface Observations on High-Resolution Analyses and Forecasts of a Severe Hailstorm

MATTHEW T. MORRIS*, KEITH A. BREWSTER, AND FREDERICK H. CARR

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

(Submitted 2 May 2020; in final form 26 March 2021)

ABSTRACT

A 2009 National Research Council study recommended that new mesoscale observing networks be integrated with existing networks to form a nationwide "network of networks". The report also recommended that research testbeds be established, such as the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) DFW Testbed, to ascertain the potential benefit of proposed observing systems. In this work, we use various conventional and non-conventional observing systems from the DFW Testbed in a series of observing system experiments (OSEs). Of special interest are radar data from Terminal Doppler Weather Radars and CASA X-band radars, as well as novel surface observations. The Advanced Regional Prediction System (ARPS) model is used to perform OSEs that are designed to assess the impact of these observing systems. A three-dimensional variational analysis system and companion complex cloud analysis are used to produce analysis increments, which are assimilated in ARPS using Incremental Analysis Updating. Experiments are performed on a supercell thunderstorm case from 11 April 2016 that produced large, damaging hail. The analysis includes quantitative comparisons of model-derived hail with radar-observed hail, along with verification of surface fields. The CASA radial velocity data benefited the forecasted storm structure, as it positively affected subsequent storm morphology and model-derived hail forecasts. Of note in surface observation impacts, the dewpoint measurements from the non-conventional Earth Networks and CWOP networks slightly degrade the forecasted dewpoint field compared to independent standard observations, but did not prevent the successful prediction of hail.

1. Introduction

The primary recommendation from a 2003 United States Weather Research Program (USWRP) workshop on surface observing systems was to establish nationwide three-dimensional mesoscale observing networks that collect observations at a higher spatiotemporal resolution than existing networks (Dabberdt et al. 2005). A subsequent report by the National Research Council (2009) expanded these findings by recommending that existing and new mesoscale networks be integrated to form a nationwide "network of networks" to maximize the observational benefit of disparate networks.

The current US operational S-band (10-cm wavelength) radar network (WSR-88D; Crum and Alberty 1993) is unable to observe roughly 70% of the troposphere below 1 km AGL due to its spacing and the Earth's curvature. This deficiency could be remedied by including additional radars in the network, such as the Federal Aviation Administration (FAA) Terminal Doppler Weather Radars (TDWRs) and low-power, short-range radars, strategically placed to fill gaps in the network (Dabberdt et al.

**Current affiliation:* Systems Research Group at NOAA/NWS/NCEP/EMC, 5830 University Research Ct., College Park, MD 20740.

Corresponding author address: Matthew T. Morris, E-mail: matthew.t.morris@noaa.gov

2005). The gap-filling concept has been demonstrated by the Collaborative Adaptive Sensing of the Atmosphere (CASA) National Science Foundation Engineering Research Center, which developed and deployed a network of four X-band (3-cm wavelength) radars in southwest Oklahoma in 2006 (McLaughlin et al. 2009).

Observational testbeds have been recommended as a means of collaboration among federal, private, and academic partners (Dabberdt et al. 2005 and National Research Council 2009), especially in regions that present operational challenges, such as urban, coastal, and mountainous regions, with a goal to objectively assess the future benefit of proposed observing systems. One such testbed has been established in the Dallas-Fort Worth (DFW) Metroplex, known as the CASA DFW Urban Demonstration Network, or, for brevity, the DFW Testbed (National Research Council 2012). The DFW Testbed is a joint endeavor among academic institutions (University of Massachusetts, University of Oklahoma, and Colorado State University), private companies, local governments, and the National Weather Service (NWS; Brewster et al. 2017). A network of seven closely spaced X-band radars supplements the existing WSR-88D radar (KFWS) in Fort Worth by providing increased low-level coverage in the area (Bajaj and Philips 2012). These radars are further supplemented by the TDWRs serving two major airports in the DFW area. Seven WSR-88D radars outside the region also contribute to coverage aloft in North Texas. Other observing systems in the testbed include satellites, radiosondes, aircraft data, Sonic Detection and Ranging (SODARs), microwave radiometers, and various conventional and non-conventional surface observation networks.

Measuring the impact of individual observing systems in such testbeds is accomplished through observing system experiments (OSEs). In the standard OSE, an analysis and subsequent forecast are generated for a control experiment in which all available observations are assimilated. The control forecast is then compared to data-denial experiments, in which observations from a particular system are withheld, to determine the changes in the forecast accuracy attributable to the withheld data. Recent OSE studies have considered the impact of sounding and profiler data (e.g., Graham et al. 2000; Agustí-Panareda et al. 2010; Benjamin et al. 2010), GPS-derived precipitable water (e.g., Smith et al. 2007; Benjamin et al. 2010), aircraft data (e.g., Benjamin et al. 2010), satellite radiances and satellite derived winds (e.g., Bouttier and Kelly 2001; Zapotocny et al. 2002, 2005, 2007; Kazumori et al. 2008; Bi et al. 2011), and radar radial winds and reflectivity (e.g., Schenkman et al. 2011a,b). OSEs have also been used to assess the impact of data from field projects such as the Mesoscale Predictability Experiment (MPEX; e.g., Coniglio et al. 2016; Keclik et al. 2017). OSEs also were used to assess the value of observational data in the DFW Testbed for a severe-weather case on 4 April 2014 (Gasperoni et al. 2018). Note that OSEs differ from observing system simulation experiments (OSSEs) in that they use actual, rather than simulated, observations.

Assimilation of data from Doppler radars is crucial for modeling ongoing thunderstorms, as they are one of the few remote-sensing platforms capable of observing convective storms with the requisite spatiotemporal resolution. The dense network of X-band CASA radars in the DFW Testbed is expected to better observe the lowest levels of the atmosphere, filling low-level gaps in the WSR-88D radar network. Data from gap-filling radars may improve forecasts of the Warn on Forecast (WoF; Stensrud et al. 2009, 2013) system, owing to their increased spatiotemporal resolution and additional low-level coverage in areas not well observed by the WSR-88Ds.

Several studies have looked at the impact of X-band radar data from the CASA Integrated Project One (IP-1) deployment in southwestern Oklahoma. Schenkman et al. (2011a,b) found the addition of CASA radial velocity data to WSR-88D data resulted in improved simulations of a tornadic mesoscale convective system (MCS) and its associated line-end vortex, as well as more accurate forecasts of tornadic mesovortices, when using a three-dimensional variational analysis (3DVAR) with complex cloud analysis. Snook et al. (2012) found that assimilating CASA and WSR-88D radar data into a forecast ensemble using an ensemble Kalman filter (EnKF) also resulted in improved probabilistic forecasts of mesovortices. Stratman and Brewster (2015) examined the influence of assimilating CASA radar data on tornado-track forecasts for a cluster of supercells on 24 May 2011. The forecasts used 3DVAR with complex cloud analysis, implemented with multiple microphysics parameterization schemes. Although successful simulations were made

using all data, the impact of CASA data in this case was limited, likely due to complex interactions among multiple pre-tornadic supercells downstream of the overlapping coverage area.

Hu et al. (2006a) reproduced a tornadic thunderstorm in the Fort Worth area, including reductions in timing and location errors. They included radar reflectivity data via the complex cloud analysis procedure, which includes a latent-heat adjustment. Hu et al. (2006b) also found additional forecast improvements with the assimilation of radial velocity data using the Advanced Regional Prediction System (ARPS) 3DVAR analysis system, although the impact was smaller than that from reflectivity. Similarly, Dawson and Xue (2006) demonstrated that forecasts of a strong, bow-shaped MCS most closely matched the observed system when the same Hu et al. (2006a) complex cloud analysis package was used in conjunction with the ARPS Data Analysis scheme (Brewster et al. 2005a), thus resulting in the elimination of the 2- to 3-h model "spin-up" time.

Also using ARPS 3DVAR with complex cloud analysis, Zhao and Xue (2009) showed that the assimilation of radial-velocity data from coastal WSR-88Ds was most impactful for improving the track forecast of Hurricane Ike (2008), while reflectivity data were most useful for improving the intensity forecast. Further, Xue et al. (2013) found that assimilating radial velocity and reflectivity data from the entire WSR-88D network aided forecasts of convection throughout most of the continental United States for a period of at least 24 h. While the methods employed in these studies varied (3DVAR, EnKF, etc.), the objective of each analysis was to generate an optimum analysis in the least-squares minimization.

With increasing numbers of surface observing systems outside the traditional government-owned systems becoming available, it is important to assess the benefit of such surface observations to objective analyses and numerical weather prediction (NWP). Tyndall and Horel (2013) found the impact of surface observations was less for those located in metropolitan areas than for similar observations in more remote locations. Hilliker et al. (2010) found improvements in forecasts of temperature and dewpoint using surface observations from the network commonly known as "WeatherBug" from Earth Networks. Improvements for wind-speed forecasts were more limited, perhaps as a result of biases in wind speed measurements due to suboptimal siting of instruments in terms of height above ground and exposure.

Carlaw et al. (2015) examined the impact of several non-conventional data sources on high-resolution forecasts of a tornadic supercell that impacted Cleburne, TX on 15 May 2013. Measurements from the WeatherBug stations indicated increased levels of moisture compared to the model background field, particularly near Cleburne, an area not well-covered by conventional observations. Enhanced buoyancy due to the increased analyzed humidity led to increases in both updraft velocity and vertical vorticity in the resultant storm, improving comparisons to the observed storms.

The purpose of this research is to extend the work of Carlaw et al. (2015) by using OSEs to determine the value of several observing systems in the DFW Testbed, focusing on impacts to a major hailstorm that occurred within the testbed. Section 2 chronicles the test case, a high-impact hail event of 11 April 2016. Section 3 describes the observational datasets used in this study, along with pre-processing and quality control procedures applied to these datasets. Section 4 details the numerical model and its associated analysis system that are used for simulations presented in this work, along with the experiment design. Section 5 includes the results of the OSEs. Concluding remarks, including implications of this work to the "network of networks" vision, are presented in Section 6.

2. 11 April 2016 hailstorm

During the afternoon and early evening hours of 11 April 2016, a prolifically hail-producing supercell affected north-central Texas, including the northern portion of the DFW Metroplex. The supercell formed around 1900 UTC just southwest of Wichita Falls, and quickly became severe as it tracked to the east-southeast. Severe storm reports from the NCEI Storm Events Database, along with reports from Meteorological Phenomena Identification Near the Ground (mPING; Elmore et al. 2014) are shown in Figure 1, with numerous significant severe hail reports (diameter >5 cm) occurring along the track of this storm. Grapefruit-sized hail (10-cm diameter) was reported in Denton

and Rockwall Counties. The largest hail associated with the storm (13.3 cm) was reported in Wylie (southern Collin County), where damage occurred to 80% of homes, at a cost of >$300 million (NWS 2016).

Examining the environment of this storm, a shortwave trough at 500 hPa was present over the southern Plains, extending from southeast Colorado through eastern New Mexico at 1200 UTC 11 April 2016 (Figure 2). This trough deepened and moved to the east during the day, aiding development of a surface low pressure system centered over northwest Texas. By 1800 UTC, the surface low was just south of Wichita Falls (Figure 3), with a dryline extending southward through the Texas Hill Country. A cold front extended westward from the surface low into New Mexico and eastward across southeastern Oklahoma.

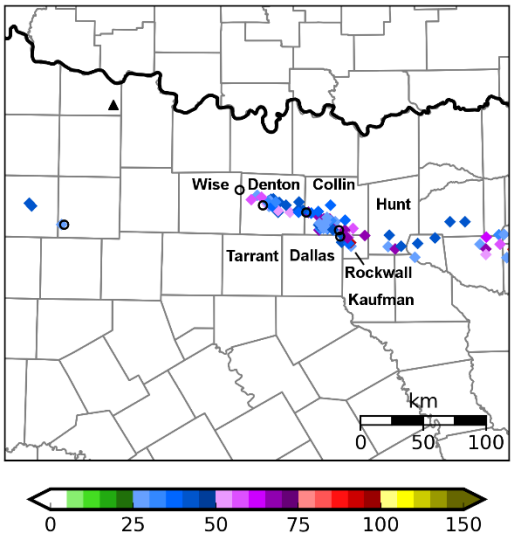


Figure 1: Severe hail reports (shaded diamonds; mm) from the NCEI Storm Events Database and mPING, along with severe wind reports from the Storm Events Database (open circles). Counties named in the text are labeled, and the black triangle denotes the location of Wichita Falls, TX.

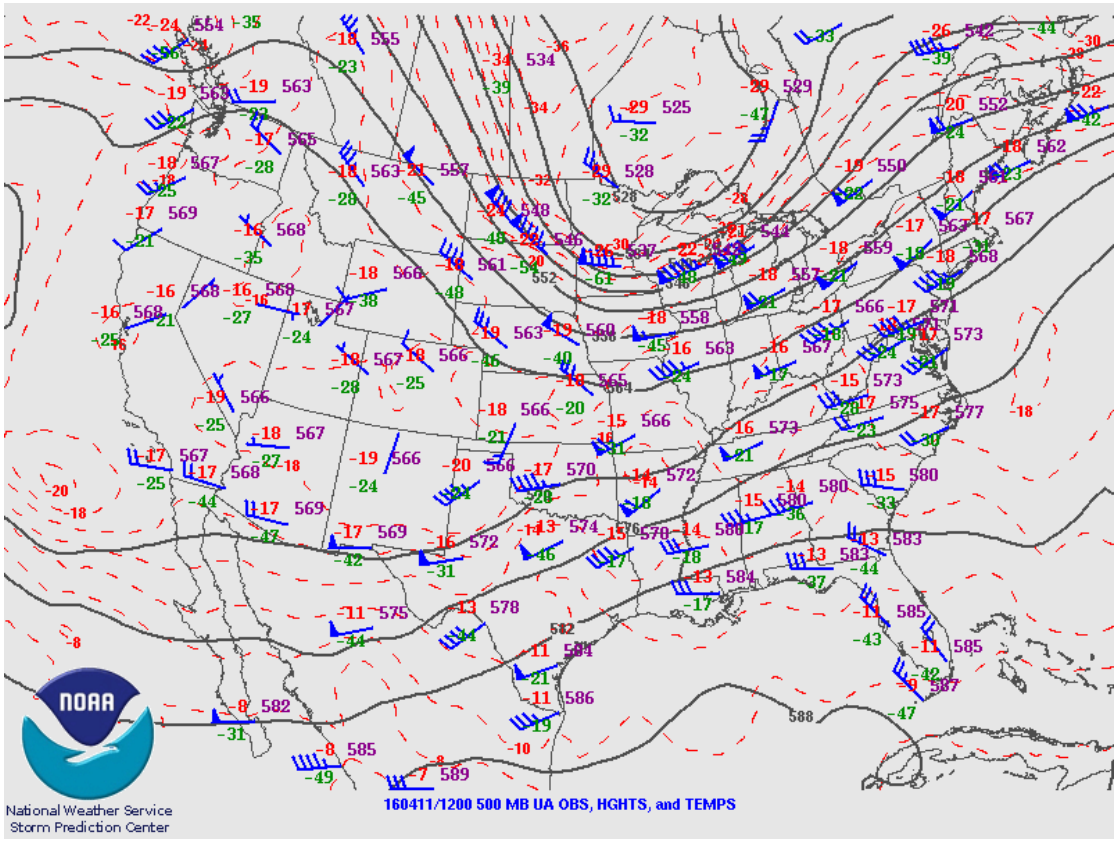


Figure 2: 500-hPa upper-air analysis valid 1200 UTC 11 April 2016 from the NOAA Storm Prediction Center. Solid black lines represent geopotential height contours (isohypses every 60 m), while dashed red lines represent isotherms (every 2°C). Wind barbs are in kt and use standard conventions.

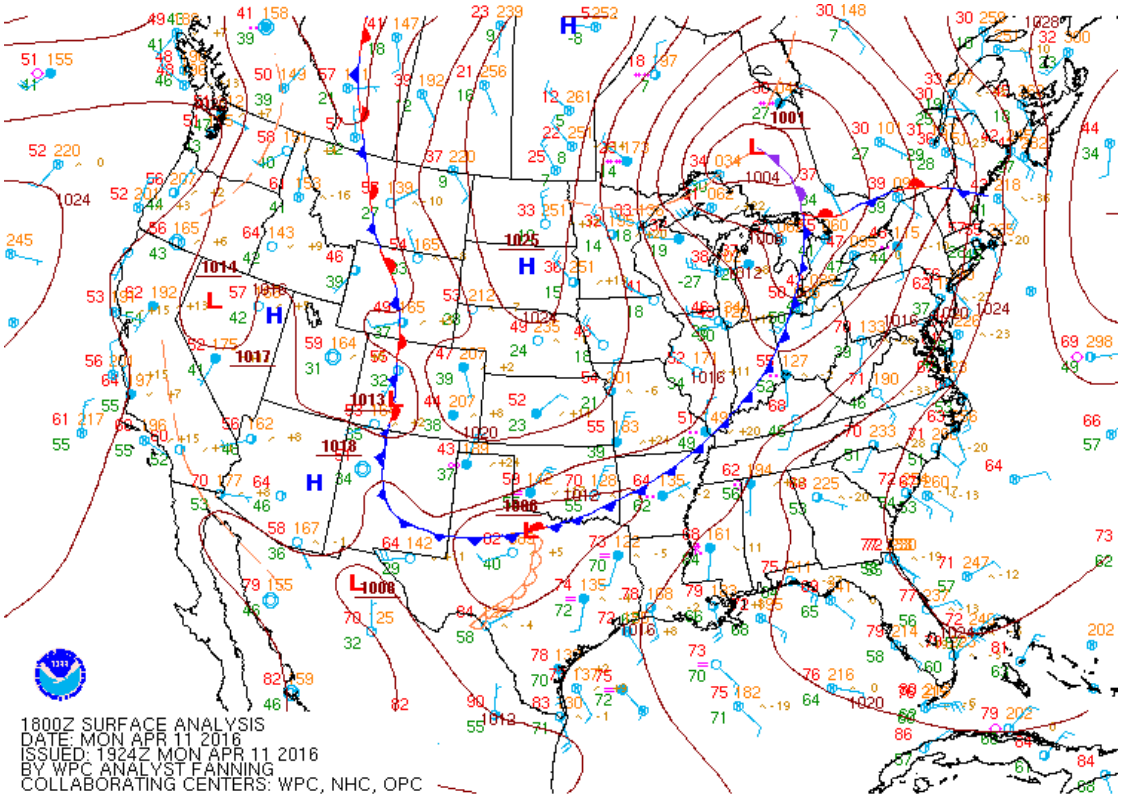


Figure 3: Surface analysis from the Weather Prediction Center (WPC) valid 1800 UTC 11 April 2016. MSL pressure is contoured every 4 hPa; the standard station model is used, including temperature (red, °F), dewpoint (green, °F), MSL pressure (orange, tenths of hPa above 1000), and wind barbs (kt).

The special 1800 UTC sounding from Fort Worth (Figure 4) shows that midlevel lapse rates >7.5°C km^{-1}, together with surface heating and adequate low-level moisture, resulted in surface-based (SB) CAPE in excess of 4000 J kg^{-1} and no SB convective inhibition (CINH). Hail was the dominant indicated severe hazard, owing to deep-layer shear >20 m s^{-1} and ample CAPE in the –10 to –30°C layer. This strong supercell did not produce a tornado in the DFW Metroplex, as the storm was quickly undercut by cool air from the cold front depicted in Fig. 3 and cool outflow from the storm itself. The presence of this shallow stable layer separating the updraft from the ground, in conjunction with unfavorable wind shear within the surging, cool outflow, significantly reduced the potential for tornadogenesis.

3. Observations

a. Conventional observations

In this work, conventional surface data sources refer to those that are available in the US federal observing networks and assimilated into operational forecast models. These sources include Automated Surface Observing System (ASOS), Automated Weather Observing System (AWOS), and established surface mesoscale networks such as the Oklahoma Mesonet (McPherson et al. 2007) and the West Texas Mesonet (Schroeder et al. 2005). Aircraft data are obtained from the Meteorological Data Collection and Reporting System (MDCRS), which provides observations of flight-level temperature, dewpoint, and wind (Moninger et al. 2003). NWS radiosonde data are not available for the midafternoon time period of this study. Data from eight radars in the WSR-88D network lie within the study domain, with the KFWS radar (Fort Worth, TX) being closest to the storm. Visible (0.63 μm) and infrared (10.5 μm) data from the Geostationary Operational Environmental Satellite (GOES) are also incorporated through the complex cloud analysis.

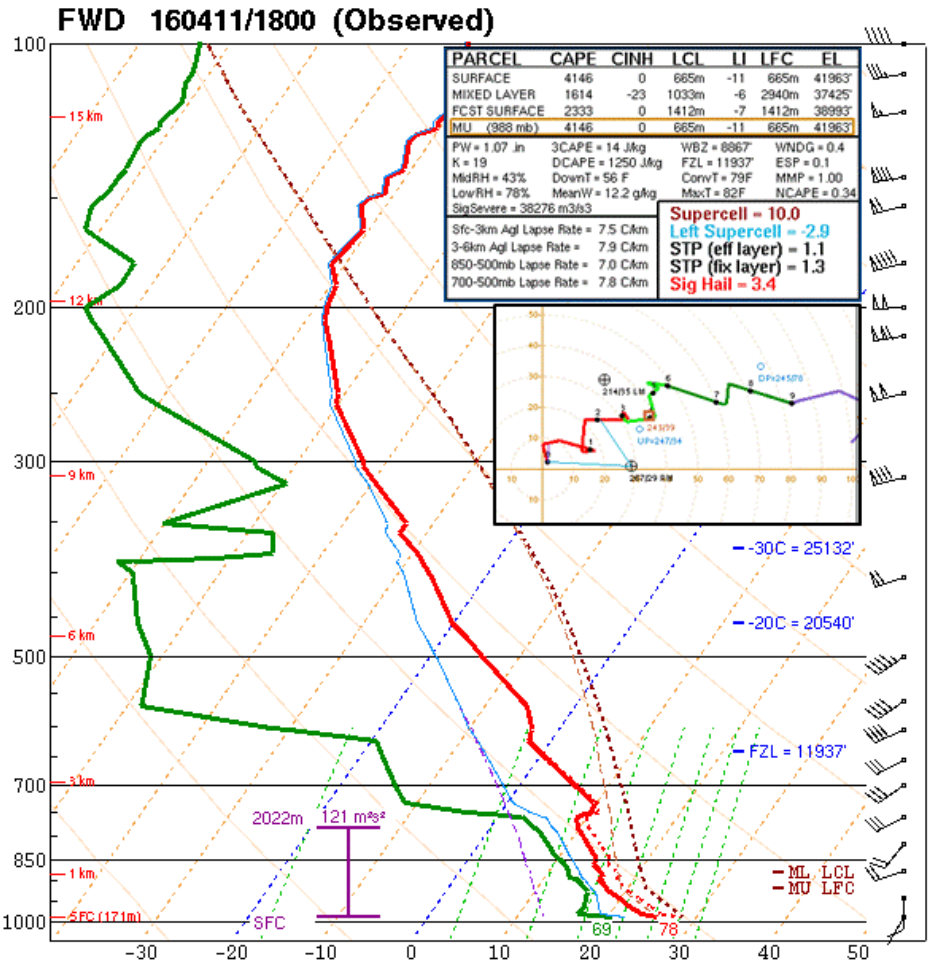


Figure 4: Skew *T*–log*p* diagram of observed sounding from Fort Worth (FWD) at 1800 UTC 11 April 2016. The temperature and dewpoint profiles are shown in red and green, respectively, with wind barbs in kt. Analyzed variables are produced by the NSHARP program (Hart and Korotky, 1991). Hodograph inset, units of kt. *Click image to enlarge*.

b. Non-conventional surface and boundary-layer observations

Global Science and Technology (GST) has developed the Mobile Platform Environmental Data (MoPED) mobile observation system, which collects observations of temperature, humidity, pressure, and precipitation from sensors mounted on fleets of trucks and other vehicles (Dahlia 2013). Since active vehicles collect data approximately every ten seconds, data thinning is applied here by averaging data from each individual vehicle over a 5-min time window or as soon as the distance traveled reaches 1 km.

The Citizen Weather Observer Program (CWOP) network is comprised of Automatic Position Reporting System as a Weather Network (APRSWXNET) stations (CWOP 2014), in which surface temperature, dewpoint, pressure, and wind observations from instruments owned by amateur radio operators and other volunteers are accumulated in the Meteorological Assimilation Data Ingest System (MADIS). These observations are subjected to the MADIS Quality Control and Monitoring System (QCMS), which performs a variety of quality control checks (MADIS 2018).

A third non-conventional data source is the Earth Networks GroundTruth WeatherBug network, herein labelled Earth Networks. This expansive network of surface observations reports temperature, humidity, and wind at sites commonly supported by schools and television stations.

A recently deployed network of weather stations from Understory Weather represents the fourth non-conventional data source (Willmot et al. 2017). Ten available stations on 11 April 2016 mostly were within Dallas County. This was early in their buildout of 140 stations across the DFW Testbed. Temperature, pressure, and humidity variables are measured using standard sensors, while wind, rain, and hail impacts are calculated based upon the forces acting on a metallic ball. While these data are mainly intended to verify weather-related insurance losses, they also may improve NWP.

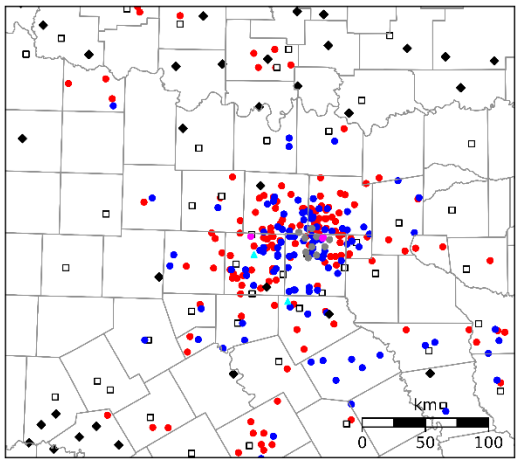


Figure 5: Spatial distribution within the analysis domain of the conventional and non-conventional surface data assimilated at the first analysis time (2150 UTC 11 April 2016). Observations include CWOP (red dots—148), ASOS/AWOS (open squares—44), Earth Networks (blue dots—105), Understory (gray dots—10), mesonet (black diamonds—32), GST MoPED (magenta dots—3), and SODAR (teal triangles—2). Not shown are 98 MDCRS observations.

Two SecondWind (now part of Vaisala) SODARs have been installed by WeatherFlow in the DFW Testbed, one at the Fort Worth NWS forecast office and the other in Midlothian, to cover temporal gaps in the NWS radiosonde network. These ground-based remote sensing instruments measure wind speed and direction in the boundary layer, up to about 200 m (Lang and McKeogh 2011). Figure 5 displays a typical distribution of both conventional and non-conventional surface data sources, along with the locations of two SODARs, in the DFW Testbed.

c. Non-conventional radar data

X-band Doppler radars in the DFW Testbed are susceptible to attenuation in regions of heavy rainfall and hail, and are thus deployed in networks with overlapping radar coverage whenever possible (Brewster et al. 2005b,c). In the DFW Testbed, X-band radars operated at Addison (ADD), Arlington (UTA), Cleburne (JCO), Denton (UNT), Fort Worth (FTW), and Midlothian (MDL) by 11 April 2016. Two additional radars are planned to complete the network (Brewster et al. 2017). This network is the result of a multisector partnership among CASA, the private sector, the NWS, and members of the North Central Texas Council of Governments (NCTCOG; Bajaj and Philips 2012). The X-band radars in the DFW Testbed, as operated in 2016, scan only at 2–3 elevations ≤4.5°. Thus, these radars provide improved coverage below 500 m and faster update times than the existing WSR-88D network.

In addition, two TDWR radars (Istok et al. 2008) surveil the two major passenger airports in the DFW Testbed. These C-band (5-cm wavelength) radars are operated by the FAA, primarily for the detection of hazards affecting aircraft operations. Figure 6 displays the spatial coverage of the eight WSR-88Ds, two TDWRs, and six X-band radars used in this study, along with two future X-band radar locations.

d. Quality-control procedures

Observations assimilated in this work are subject to several quality-control (QC) procedures. Radar data QC includes velocity unfolding and removal of nonmeteorological echoes (Brewster et al. 2005a). QC checks for MADIS observations are detailed in the NWS Techniques Specification Package (NWS 1994). In addition, the ARPS 3DVAR program employs several QC checks to remove spurious observations, including temporal and spatial consistency checks, as well as threshold checks against the background field (Brewster 1996).

4. Methods

a. Advanced Regional Prediction System (ARPS)

The ARPS model is used to perform the OSEs presented in this research. The Center for Analysis and Prediction of Storms (CAPS) at the

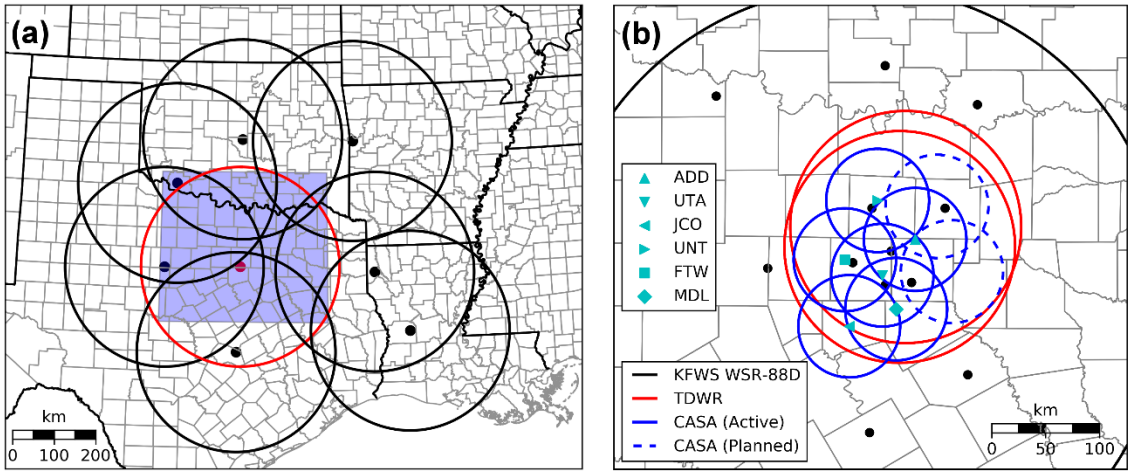


Figure 6: a) Locations of the 8 WSR-88D radars whose data are used in this work, where the location of KFWS is indicated in red and the blue shaded region represents the model domain used, and b) locations of the radars used in this study, as well as the ten ASOS and two Oklahoma Mesonet stations used for verification (black dots). Range rings for the CASA, TDWR, and KFWS WSR-88D radars are shown with radii of 30 km, 90 km, and 240 km, respectively.

University of Oklahoma developed the first version of ARPS during the 1990s (Xue et al. 2000, 2001). ARPS is a compressible, non-hydrostatic model with a terrain-following vertical coordinate on an Arakawa C-grid. The vertical coordinate is stretched using a hyperbolic tangent function to permit higher resolution at the lower levels. Simulations of tropical cyclones (e.g., Zhao and Xue 2009), MCSs (e.g., Dawson and Xue 2006) and tornadoes (e.g., Xue et al. 2014) have been performed using ARPS.

In this work, double-moment microphysics (Milbrandt and Yau 2005), NASA atmospheric radiation transfer and 1.5-order turbulent kinetic energy (TKE) planetary boundary layer (PBL; Deardorff 1980) parameterization schemes are used. Advection is fourth-order in the vertical and horizontal, while convection is explicitly resolved. For the land surface, a two-layer diffusive soil model is used (Noilhan and Planton 1989).

b. ARPS Three-Dimensional Variational (3DVAR) analysis system

The ARPS 3DVAR (Gao et al. 2004) produces an analysis by minimizing a scalar cost function, $J(x)$, which is given by:

$$J(x) = \frac{1}{2}(x - x_b)^{\mathrm{T}}\mathbf{B}^{-1}(x - x_b) + \frac{1}{2}(H(x) - y_o)^{\mathrm{T}}\mathbf{R}^{-1}(H(x) - y_o) + J_c \qquad (1)$$

The first term on the right-hand side measures the distance between the analysis of the state variable, x, and the background field, x_b, and is weighted by the inverse of the background error covariance matrix, **B**. The second term represents the distance between the analysis, x, brought to observation locations by the forward operator, H, and the observed variables, y_o, and is weighted by the inverse of the observation error covariance matrix, **R**. Cross-correlations between model variables are not included in the **B** matrix, and a first-order recursive filter (Hayden and Purser 1995) is used to generate the isotropic Gaussian spatial error correlations. Furthermore, observational errors are assumed to be uncorrelated, resulting in a diagonal observation error covariance matrix.

The final term in Eq. (1) is a penalty term, and represents a weak anelastic mass continuity constraint:

$$J_c = \frac{1}{2}\lambda_c D^2 \qquad (2)$$

where D is given by:

$$D = \alpha\left(\frac{\partial\bar{\rho}u}{\partial x} + \frac{\partial\bar{\rho}v}{\partial y}\right) + \beta\left(\frac{\partial\bar{\rho}w}{\partial z}\right) \qquad (3)$$

Here, λ_c represents a weighting coefficient for the mass continuity constraint, α and β correspond to weighting terms for the horizontal and vertical divergence terms, respectively, and $\bar{\rho}$ is the mean air density at a given height. The

anelastic mass continuity is a weak constraint that acts to derive non-radial wind information from the observed radial velocities (Gao et al. 2004; Hu et al. 2006b). When the grid aspect ratio is near unity (i.e., the horizontal and vertical grid spacing are nearly the same), the anelastic mass divergence constraint produces accurate analyses of vertical and horizontal velocity (Hu et al. 2006b). However, when the horizontal grid spacing is much larger than the vertical grid spacing (i.e., the aspect ratio is over 100), which is often true in the lowest levels of the model, adjustments to the vertical velocity dominate adjustments to the horizontal component of the wind. This work follows that of Carlaw et al. (2015), which uses a horizontal weighting coefficient ($\alpha\lambda_c = 1000$) that is an order of magnitude larger than the vertical weighting coefficient ($\beta\lambda_c = 100$).

c. Complex cloud analysis

The ARPS complex cloud analysis package is used to analyze hydrometeors (Brewster et al. 2005a; Hu et al. 2006a). The cloud analysis procedure uses satellite, radar, and surface observations of cloud layers to modify hydrometeor fields by using equations that relate hydrometeor mixing ratio values and observed radar reflectivity (e.g., Ferrier 1994; Rogers and Yau 1989) remapped to the 3D model grid and mosaicked using the maximum reflectivity value where there is overlapping radar coverage (Brewster et al. 2005a). This procedure was updated to allow inversion of the hydrometeor-to-reflectivity equations for all the microphysics schemes used in ARPS and WRF (Brewster and Stratman 2015). The complex cloud analysis is performed after the 3DVAR minimization is completed for all other fields (reflectivity is not included in the 3DVAR).

The background hydrometeor mixing ratio values are replaced by reflectivity-derived values where radar reflectivity is above a user-defined threshold (20 dBZ everywhere, except 25 dBZ within the lowest 2 km to account for non-precipitation echoes). In regions where radar reflectivity is below the prescribed threshold (0 dBZ), precipitation in the model background field is removed and relative humidity set to a maximum of 95%, thus removing spurious convection from the model field. Finally, the cloud analysis procedure uses a 1-D cloud model with entrainment, to adjust the temperature profile in regions where clouds and updrafts are present. This accounts for the latent heat released during condensation.

d. Incremental Analysis Updating

When numerical models are forced to adjust to large volumes of information, all applied at the initial time, adjustment processes manifested as high-frequency, short-wavelength oscillations often occur (e.g., Bloom et al. 1996; Brewster 2003). To combat this issue, we use Incremental Analysis Updating (IAU), which applies analysis increments computed at the initial time gradually over an assimilation window (Bloom et al. 1996). The analysis increments are applied using a triangular distribution in time, ramping from zero at the initial time to a maximum at mid-window, then ramping to zero at the end. Increments are not applied to the pressure and vertical velocity fields during IAU at storm scales, as these fields are not well-observed at this scale, and rapidly respond to changes in other model fields. The distribution can be made different in time for each variable (known as IAU with Variable-Dependent Timing, IAU-VDT; Brewster et al. 2015). Specifically, one can apply a larger portion of the wind and latent-heat increments early in the assimilation window, while applying a greater portion of the hydrometeor increments later in the window. This allows sufficient vertical velocity to evolve in the model to support the added hydrometeors, which otherwise might fall out immediately.

e. Experiment design

The experiments presented here use the ARPS model on a single grid with a horizontal spacing of 1 km. The 400 × 360-km domain is centered at 33.0 °N, 97.25 °W (Fig. 6). There are 50 vertical levels. Vertical grid spacing averages 400 m, and increases with height from a minimum of 20 m at the lowest model level. Land-surface features are specified using Global Ecosystems Database (GED) surface data files (Kineman 1992), while terrain elevation information is interpolated from the 30-s U.S. Geological Survey (USGS) terrain dataset (Danielson and Gesch 2011).

Two consecutive cycles are used to assimilate observations into the model forecast, the first beginning at 2150 UTC. Analysis increments are determined by the 3DVAR system, using observations in the window 2150–2200 UTC.

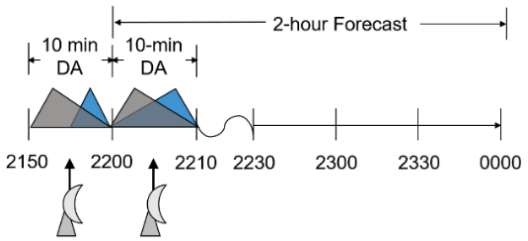


Figure 7: Assimilation procedure for the experiments presented. Data assimilation cycles begin at 2150 UTC, with a 2-h forecast beginning at 2200 UTC. Gray (blue) triangles represent the weighting of fractions of the computed wind and latent heat (hydrometeor) analysis increment introduced during each assimilation window.

The increments are applied gradually during that 10-min window using IAU-VDT (Figure 7). The second cycle, from 2200–0000 UTC, assimilates data during the first 10 min, which is followed by a 110-min free forecast. Initial conditions are obtained for the model grid using the 2100 and 2200 UTC operational Rapid Refresh (RAP) analyses, which have a 13-km grid. These model analyses are interpolated in space and time to produce the ARPS initial conditions valid at 2150 UTC. Lateral boundary conditions are also derived from the RAP analyses valid at 2100, 2200, 2300, and 0000 UTC.

The ARPS 3DVAR analysis system used here includes four analysis passes to account for the diverse spacing of observation types assimilated. The horizontal correlation scale distance for the passes is 100 km, 50 km, 10 km, and 0.8 km, respectively. Profiler data (here from the two SODARs) are incorporated on the first and second passes, which allows this information to be spread across the model domain. Conventional surface observations (ASOS/AWOS) and MDCRS flight data are incorporated in the second and third passes. Mesonet and non-conventional surface data are assimilated in the third analysis pass, and radar velocity data are incorporated only during the final pass. In this way, smaller-scale details are added to the analysis as the correlation scale is decreased on subsequent passes, in concert with the data spacing. The vertical correlation scale is defined to be four grid points for all four analysis passes. Satellite and radar reflectivity data are analyzed via the complex cloud scheme after being remapped to the model grid and mosaicked.

The OSEs are listed in Table 1. Note that in CONTROL, all available data are used, including WSR-88D, CASA, and TDWR reflectivity and radial velocity data within the domain, while other experiments are designed to withhold information from one or more observing systems.

Table 1: Observing system experiments (OSEs) performed.

Experiment	Conventional surface data	Non-conventional surface data	88D data	CASA data	TDWR data	Upper-air profiles
CONTROL	All	All	All	All	All	All
NOTESTBED	All	None	All	None	None	Deny SODARs
NONEWSFC	All	None	All	All	All	All
NOGST	All	Deny MoPED	All	All	All	All
NOERNET	All	Deny Earth Networks	All	All	All	All
NOCWOP	All	Deny CWOP	All	All	All	All
NOUNDERSTORY	All	Deny Understory	All	All	All	All
NOCWOP_ERNET	All	Deny CWOP and Earth Networks	All	All	All	All
NOCASA	All	All	All	None	All	All
NOCASAVR	All	All	All	Deny V_r	All	All
NOTDWR	All	All	All	All	None	All
88DONLY	All	All	All	None	None	All
NOKFWS	All	All	Deny KFWS	All	None	All
NOKFWS_CASA	All	All	Deny KFWS	None	None	All
CASAONLY	All	All	None	All	None	All
NORADAR	All	All	None	None	None	All

5. Results

The experiments are evaluated, qualitatively, in terms of observed impacts on forecasted reflectivity and winds, including location and storm structure, and, quantitatively, on forecasts of hail and surface weather.

a. Qualitative storm-morphology comparison

Figure 8 shows how the CONTROL experiment's simulated reflectivity at model level 21, ≈2 km AGL, compares with 0.5° tilt reflectivity observations from the KFWS WSR-88D. The reflectivity pattern in the CONTROL experiment at the end of the second assimilation window (2210 UTC) generally agrees well with the observations at this time, as it captures the most intense precipitation occurring in northern Denton County (Fig. 1). The simulated echo is a little broader and stronger, due to the mosaicking technique that uses the maximum reflectivity from all radars during the data assimilation period. By 2300 UTC, 50 min into the free forecast, the CONTROL simulated storm was centered in southern Collin County, in general agreement with the radar observations. By 2330 UTC, a few differences emerged, as the rotation center and hook echo of the simulated storm are displaced about 20 km to the south, compared to the observed reflectivity.

By 0000 UTC, the low-level rotation of the simulated storm is in extreme eastern Kaufman County, while the observed storm had progressed further east and appears to have weakened. The CONTROL experiment exhibits a high bias in reflectivity, with reflectivity predicted over a larger area than what was observed. This bias could be a result of choices made in the microphysics scheme (e.g., Wainwright et al. 2014). A fine line also is evident in the radar data south of the supercell, indicating the cold front and outflow boundary that undercut the storm early in its lifespan. This feature is suggested in the 2 km wind fields and is more apparent at lower model levels (not shown).

Figure 9 shows the 2-km AGL simulated reflectivity and wind vectors at 2210 UTC, the end of the final assimilation window, for the radar denial experiments, compared to the 0.5° KFWS radar reflectivity (Fig. 9j). All experiments but NORADAR and CASAONLY (Fig. 9b,f) capture the main supercell of interest, including reflectivity values >65 dBZ. NORADAR has no reflectivity at this time and level, owing to the denial of all radar data during the assimilation period. The CASAONLY experiment contains only a small area of reflectivity, since CASA radar scans near the storm were at or below 2 km AGL. The differences between CASAONLY and NOKFWS (Fig. 9f,h) are attributed solely to the assimilation of data from the seven surrounding WSR-88Ds.

Figure 10 shows the reflectivity one hour into the free forecast, when the observed storm was dropping its largest hail in Wylie. The CONTROL experiment (Fig. 10a) has a well-defined hook echo at this time, centered in northeastern Dallas County. Although convective initiation has occurred in NORADAR (Fig. 10b), the echoes are weaker and displaced to the northwest relative to CONTROL. The 88DONLY experiment (Fig. 10c) has weaker reflectivity values than CONTROL, demonstrating some positive contribution from the non-conventional radar data. NOCASA (Fig. 10d) has a supercell in roughly the same location as CONTROL, although the low-level circulation is not as well organized. NOCASAVR (Fig. 10e) has a weaker supercell than NOCASA, suggesting the CASA radial velocity data were more impactful for this case than the reflectivity data.

Given that the CASA radars, which were taking only two or three elevation scans, were incapable of observing the full depth of the storm, it took some time for the echoes to grow upward in CASAONLY (Fig. 10f). This simulation allowed a stronger cell to develop northwest of the main cell, as the cold pool and associated southward gust front propagation was delayed. Nonetheless, while NORADAR and CASAONLY (Fig. 10b,f) are both weaker relative to CONTROL, CASAONLY is superior to NORADAR as the assimilation of low-level CASA radial velocity data, with IAU-VDT, allowed for earlier convective initiation. NOTDWR (Fig. 10g) shows an evolution similar to that in CONTROL. NOKFWS (Fig. 10h) has a slightly weaker supercell structure than CONTROL, but this feature is better defined than in NOKFWS_CASA (Fig. 10i), suggesting that the CASA data were beneficial in substituting for KFWS data.

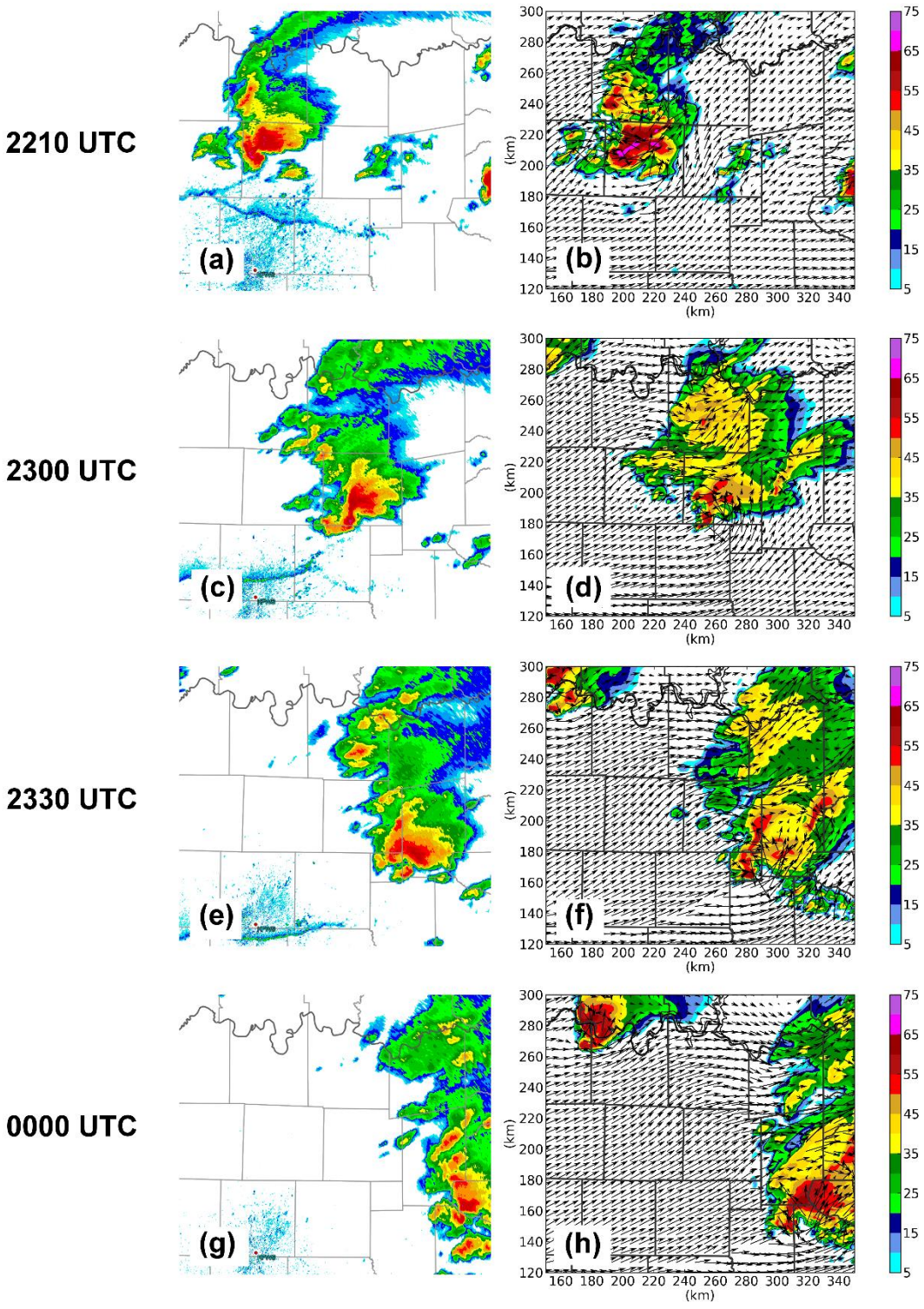


Figure 8: Observed reflectivity from the KFWS 0.5° scan (left column) and simulated reflectivity and wind vectors at 2 km AGL for the CONTROL experiment (right column), at the four times labeled. *Click image to enlarge.*

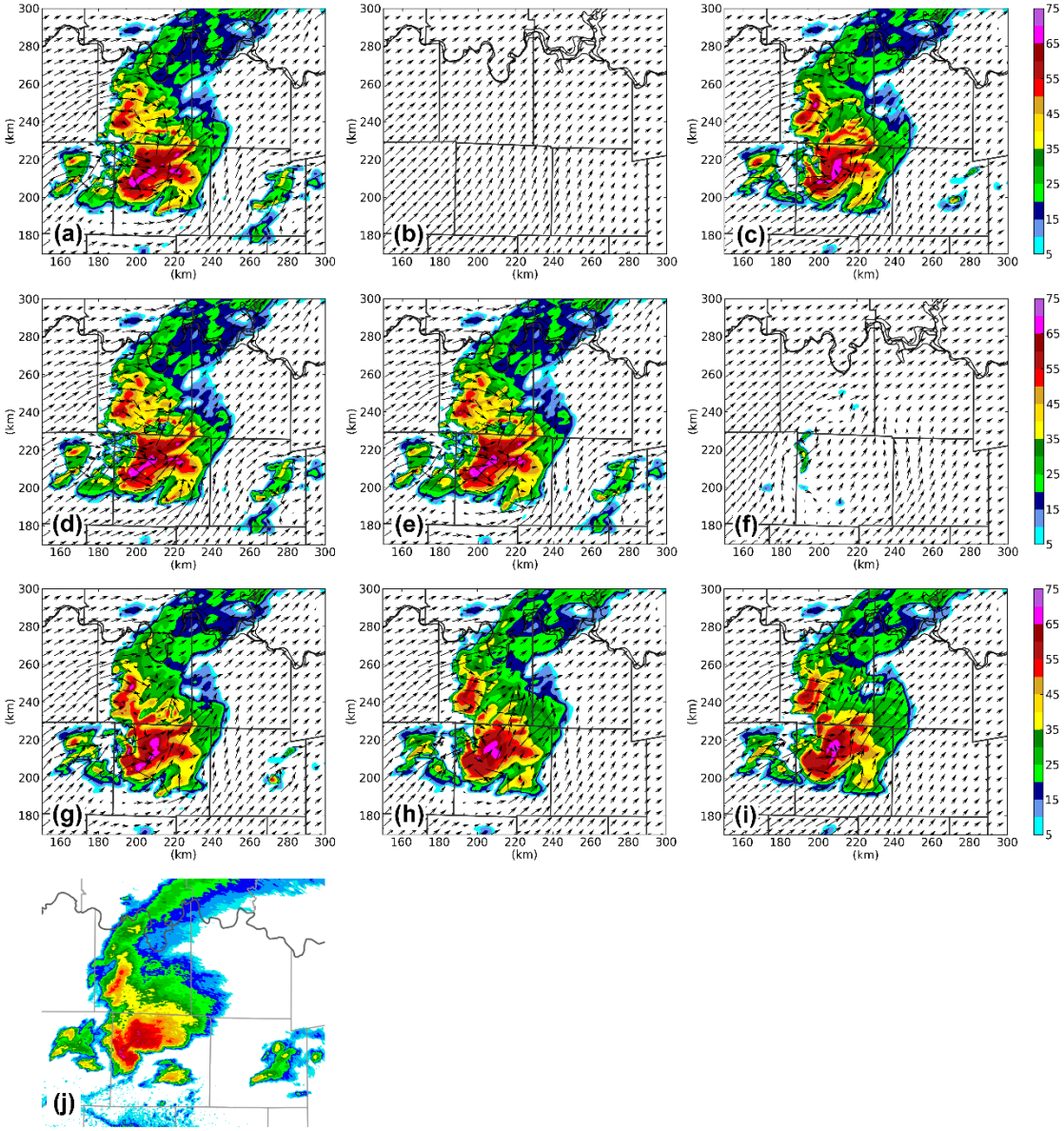


Figure 9: Simulated reflectivity and wind vectors at 2 km AGL at 2210 UTC for: a) CONTROL, b) NORADAR, c) 88DONLY, d) NOCASA, e) NOCASAVR, f) CASAONLY, g) NOTDWR, h) NOKFWS, and i) NOKFWS_CASA, as well as the j) KFWS reflectivity 0.5° scan at 2209 UTC. *Click image to enlarge.*

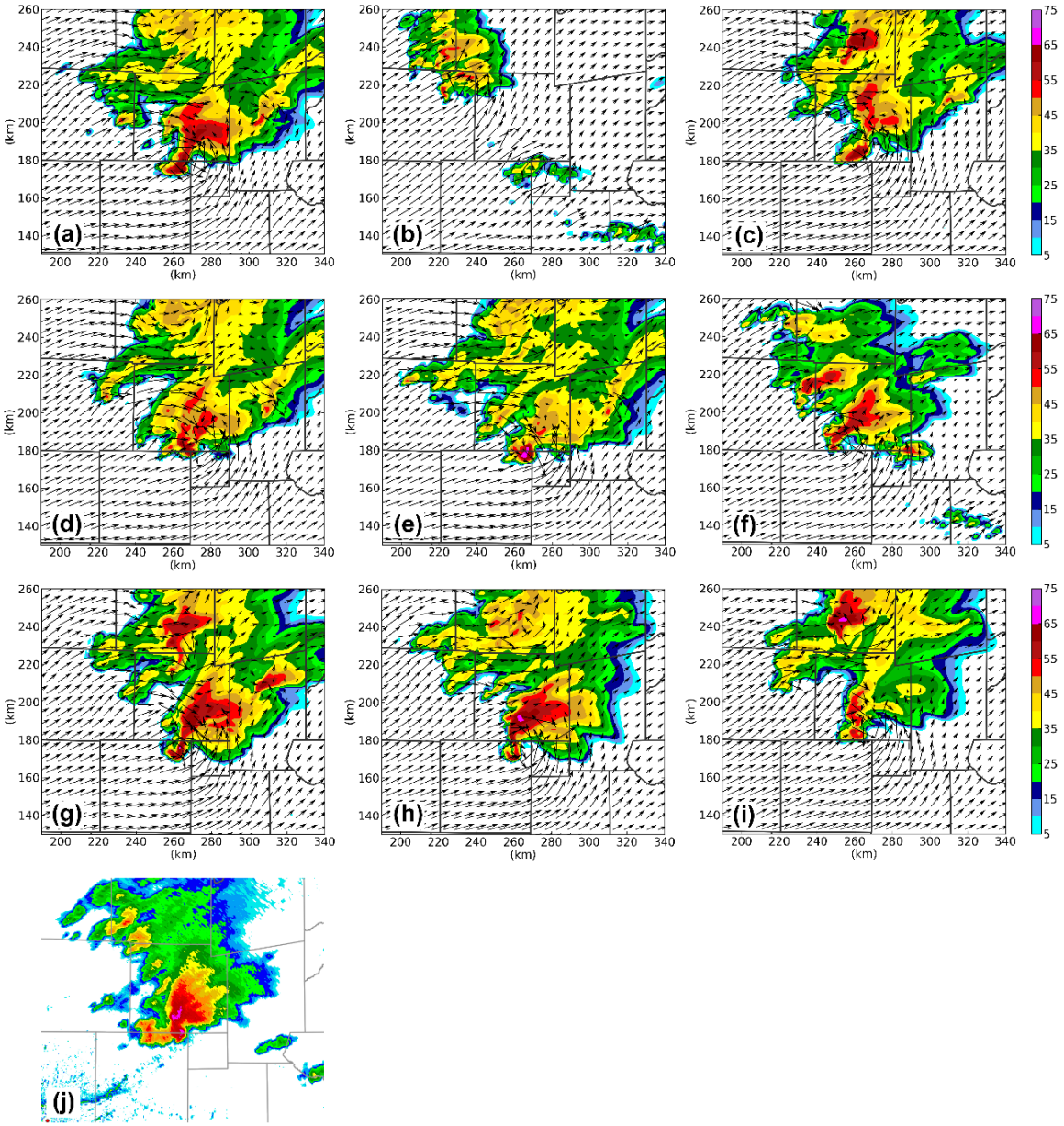


Figure 10: As in Fig. 9, but model output for 2310 UTC, and KFWS reflectivity at 2309 UTC. *Click image to enlarge.*

Figure 11 shows the 1- to 3-km updraft helicity (UH) valid at 2210 UTC for the radar data-denial experiments. The CONTROL experiment (Fig. 11a) has a UH center in central Denton County at this time. No UH center was present in NORADAR (Fig. 11b) due to the denial of all radar data. The storm structure is weaker in 88DONLY (Fig. 11c) suggesting some benefit of the non-conventional radar data. UH values were slightly higher than CONTROL in NOCASA and NOCASAVR (Fig. 11d,e), with little, if any, discernible differences seen among the simulations. No UH center was present in CASAONLY (Fig. 11f), although the wind field suggests some convergence is present at 2 km AGL, which would aid in convective initiation. The UH magnitude is considerably weaker in NOTDWR (Fig. 11g), suggesting that the reduction seen in 88DONLY was due to TDWR, rather than CASA, radar data. The magnitude is also lower in NOKFWS (Fig. 11h) relative to CONTROL, but NOKFWS is clearly superior to

NOKFWS_CASA (Fig. 11i), which shows that the CASA data were able to recover some of the low-level information lost due to the denial of KFWS data.

One impact of the CASA data can be seen in a patch of stronger northerly (v) component of the wind along the western boundary of Denton County at 1 km AGL, at the end of the second assimilation window in CONTROL (2210 UTC, Figure 12a) compared to NOCASA at the same time (Fig. 12c). Also, the southerlies on the east side of the center of rotation were slightly weaker in CONTROL. The patch of stronger northerly winds propagated to the southeast, affecting the subsequent cycle of mesocyclone formation, resulting in the center of the new updraft circulation in CONTROL forming 3–5 km west of the NOCASA position at 2235 UTC (Fig. 12b,d). The position of the new mesocyclone in CONTROL, west of the Denton-Collin County border, agrees better with the observed

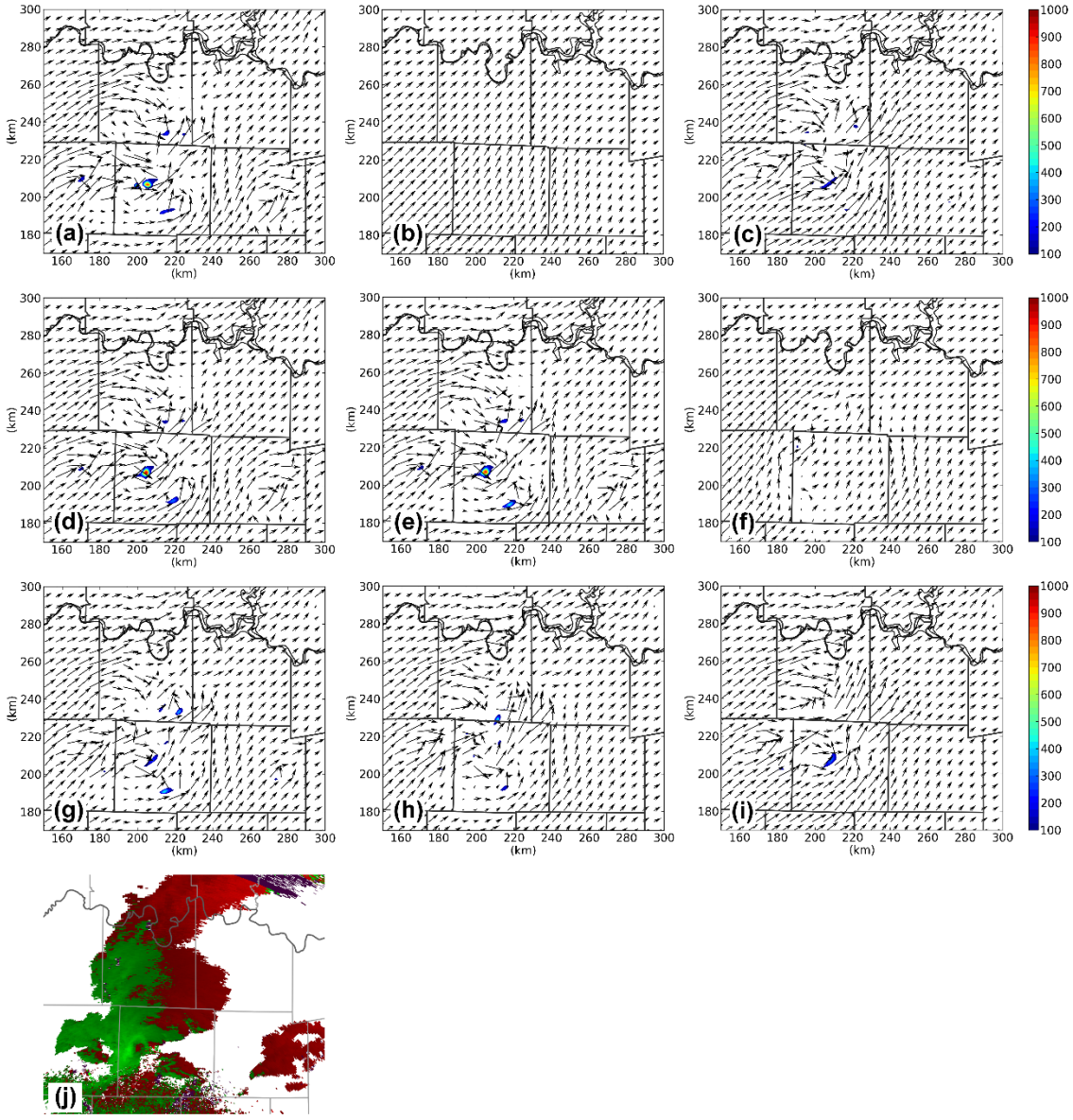


Figure 11: 1- to 3-km updraft helicity (m^2 s^{-2}) and wind vectors at 2 km AGL valid at 2210 UTC for: a) CONTROL, b) NORADAR, c) 88DONLY, d) NOCASA, e) NOCASAVR, f) CASAONLY, g) NOTDWR, h) NOKFWS, and i) NOKFWS_CASA, as well as the j) KFWS radial velocity 0.5° scan at 2209 UTC. *Click image to enlarge.*

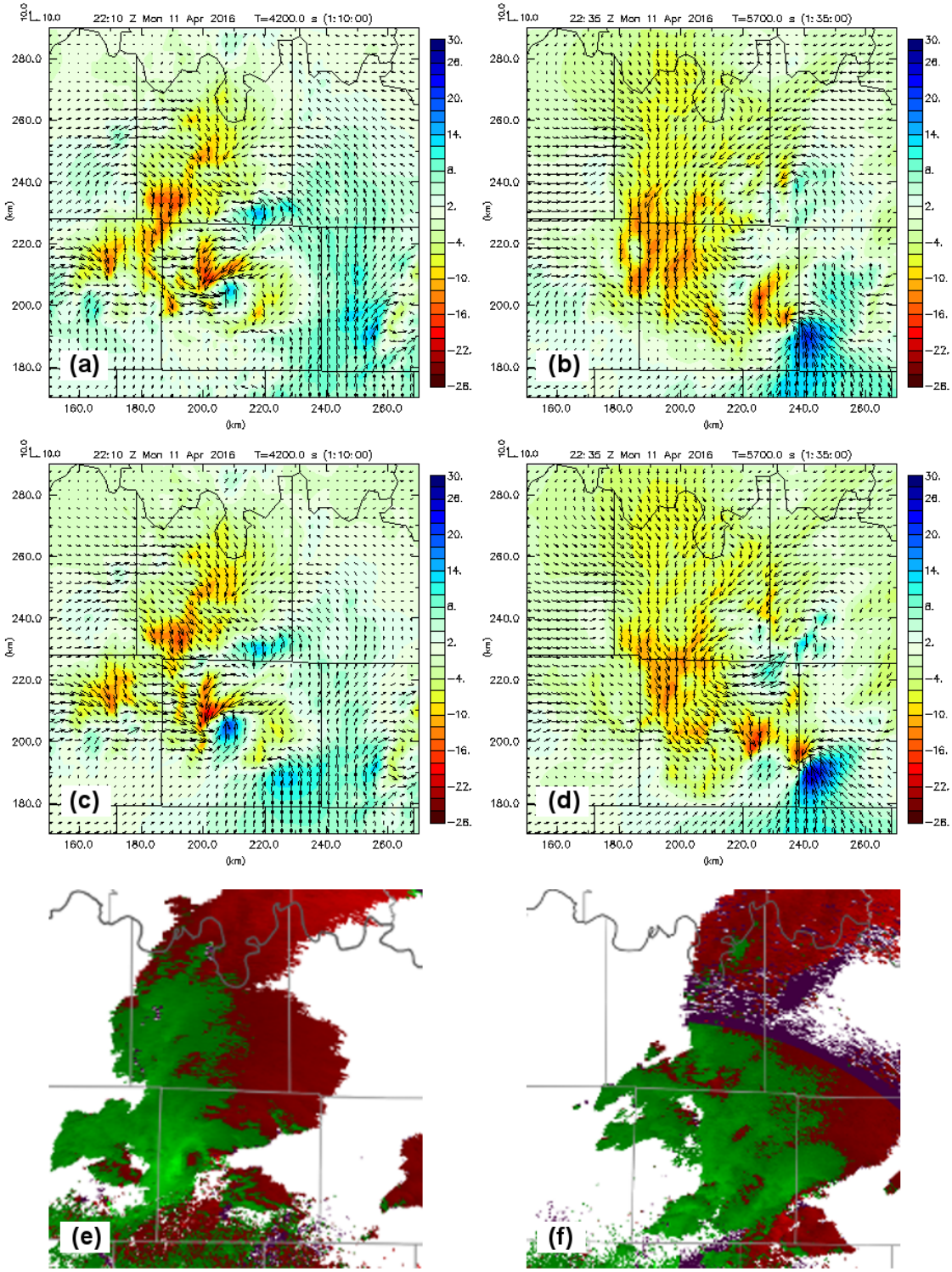


Figure 12: Wind vectors and northward (v-component) wind (m s^{-1}) at 1 km AGL for the CONTROL simulation at a) 2210 UTC and b) 2235 UTC, NOCASA experiment at c) 2210 UTC and d) 2235 UTC, and KFWS radial velocity 0.5° scan at e) 2209 UTC and f) 2234 UTC. *Click image to enlarge.*

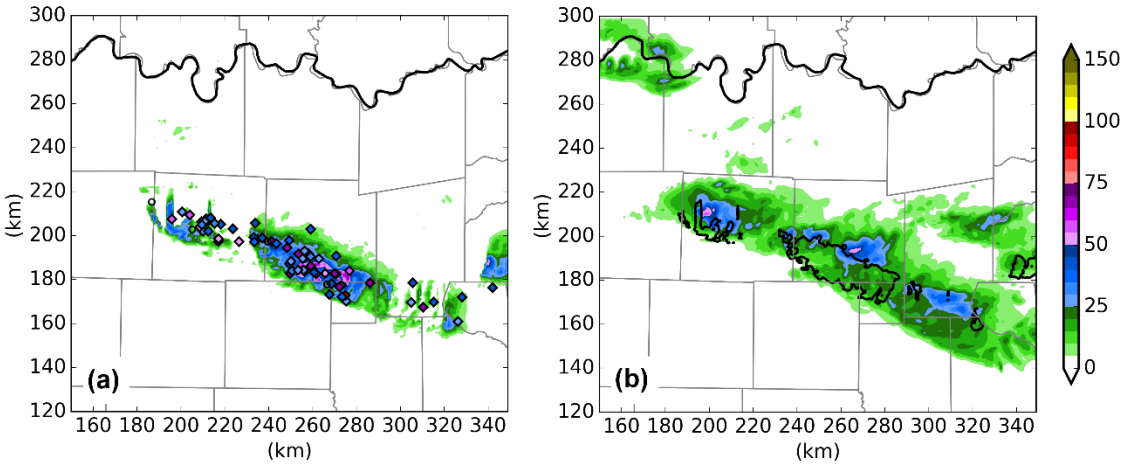


Figure 13: a) Observed MESH and b) CONTROL forecasted MESH (in mm) for the forecast period, 2200 to 0000 UTC. Diamonds are reports from mPING (Elmore et al. 2014) and Storm Events Database. The black contour in (b) corresponds to observed regions of MESH ≥25 mm.

radar radial velocity signature of the circulation at that time (Fig. 12f). The westward displacement also provides a better hail verification (section 5b).

When comparing the surface data-denial experiments (not shown), little, if any, differences are seen in the simulated wind and reflectivity fields at 2210 UTC; this is because each experiment assimilated all available radar data. Furthermore, by 2310 UTC, 1 h into the free forecast, the storm placement was similar to CONTROL in all surface denial experiments, suggesting that the surface data had little impact on short-term storm motion.

b. Hail verification

The maximum estimated size of hail (MESH) can be determined in the WSR-88D network using a hail detection algorithm (HDA; Witt et al. 1998). MESH is empirically estimated from a weighted vertical integration of the horizontal reflectivity factor (Z) >40 dBZ above the melting level. Snook et al. (2016) considered several hail-verification metrics for ensemble forecasts of the supercell thunderstorms of 20 May 2013, and found that forecasts utilizing MESH agreed most closely with the observed hail (WSR-88D derived MESH swaths). In this case, KFWS radar data remapped to the model grid are used to compute observed MESH. Similarly, forecast MESH is computed by the same method using model simulated reflectivity.

The CONTROL experiment (Figure 13) generally has a good forecast of the hail swath location and orientation. CONTROL showed hail >25 mm, although the forecast swath is a bit longer and wider than the Radar MESH and hail reports. CONTROL under-predicted the maximum hail size in some places, and displaced about 20 km northeastward the maximum axis in the central and eastern part of the swath.

Figure 14 shows the forecast MESH swaths for the radar denial experiments, with the black contour showing where Radar MESH values exceeded 25 mm. Hail was sparsely forecast in the 2-h period during the NORADAR experiment (Fig. 14b), owing to the 40-min "spin-up" period the model needed for precipitation to form when no radar data are assimilated. The 88DONLY experiment (Fig. 14c) has a somewhat smaller swath of MESH-indicated severe hail than the CONTROL experiment (Fig. 14a), as well as under-predicted hail size in the westernmost portion of the observed severe hail swath. This indicates that the TDWR and CASA radars were beneficial in initializing the hail-producing elements of the storm. This is also indicated by the NOCASA experiment (Fig. 14d), which has similar coverage as the CONTROL albeit smaller hail sizes. The NOCASAVR experiment (Fig. 14e) has less areal coverage of severe hail than the CONTROL and NOCASA experiments, which indicates that retaining CASA reflectivity data while denying radial velocity data resulted in a

more substantial degradation in the forecast than denying all available CASA data, showing the primary benefit coming from the CASA radial velocity data. The NOTDWR (Fig. 14g) result was mixed, as it has a more continuous severe hail swath than the CONTROL experiment with less northeastward displacement, but somewhat smaller hail sizes than observed at the beginning of the forecast.

The NOKFWS experiment (Fig. 14h) is clearly superior to NOKFWS_CASA (Fig. 14i), indicating that CASA reflectivity and radial velocity data are able to recover some of the low-level information lost, in the event the KFWS radar had been unavailable or if the configuration of WSR-88D radars was such that the closest radar was more distant from the storm as in more-sparsely covered areas of the country. Like NOTDWR, there is less northeastward displacement of the main hail swath in NOKFWS when compared to CONTROL, albeit with smaller than observed hail sizes at the beginning of the forecast.

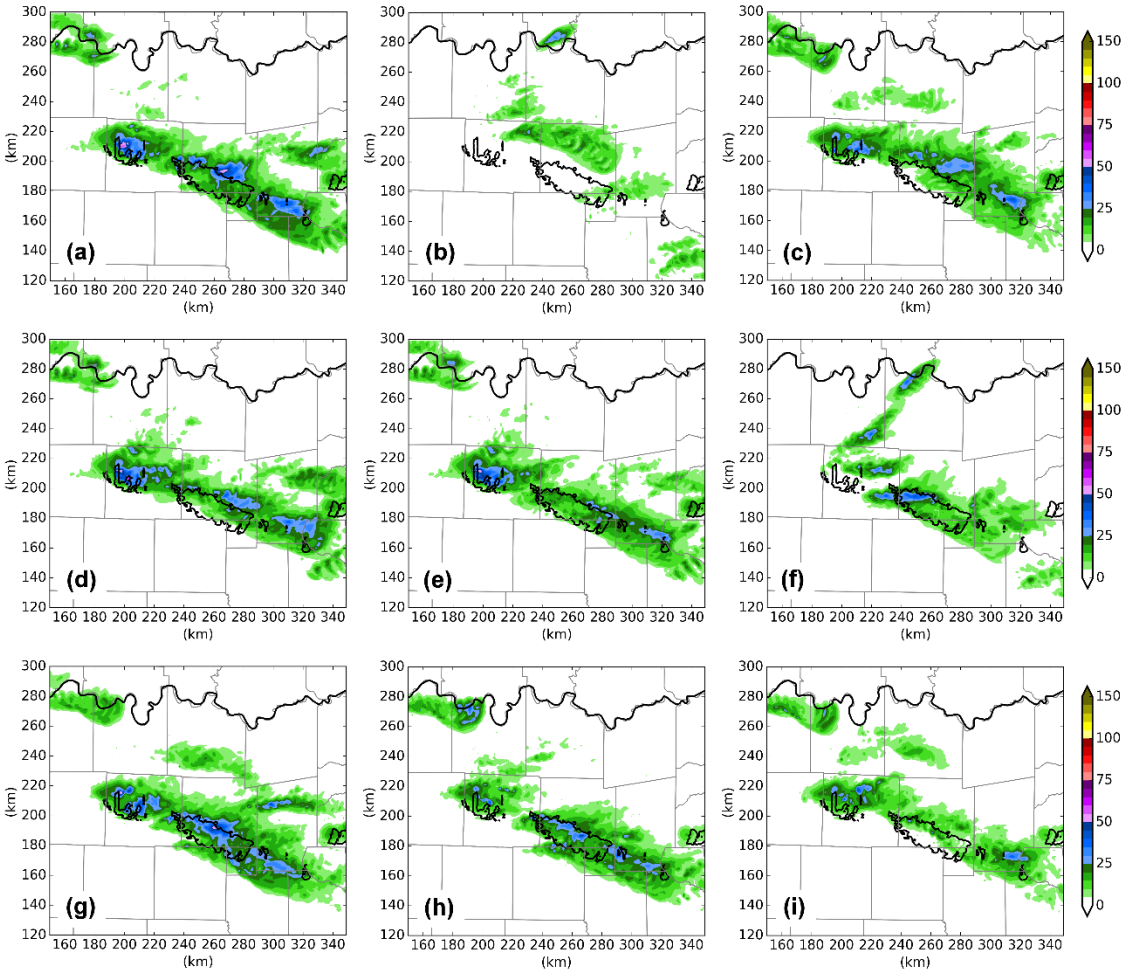


Figure 14: Simulated MESH (in mm) for the a) CONTROL, b) NORADAR, c) 88DONLY, d) NOCASA, e) NOCASAVR, f) CASAONLY, g) NOTDWR, h) NOKFWS, and i) NOKFWS_CASA experiments for the 2-h forecast period 2200–0000 UTC. *Click image to enlarge.*

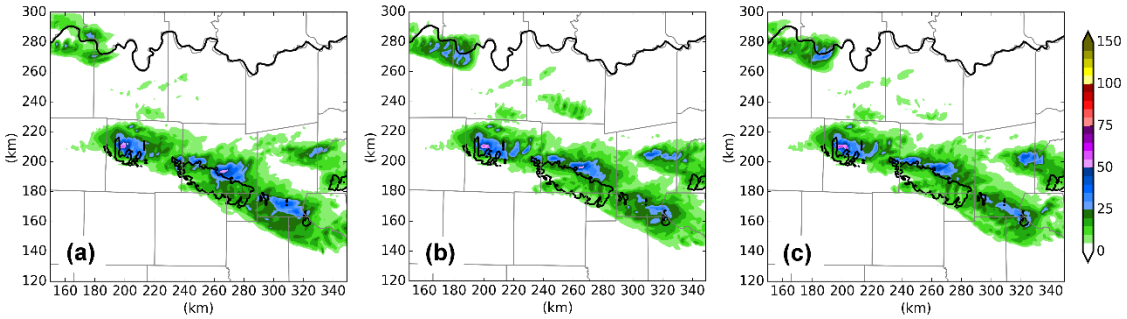


Figure 15: As in Fig. 14, but for: a) CONTROL, b) NONEWSFC, and c) NOCWOP_ERNET. *Click image to enlarge.*

Figure 15 shows the forecast MESH swaths for a subset of the surface data-denial experiments. NONEWSFC (Fig. 15b) produced comparable hail to CONTROL throughout most of the period, with subtle differences emerging at the end of the forecast period, as the region of severe hail outside of what is observed in Radar MESH was reduced in NONEWSFC along the main supercell track. The NOCWOP_ERNET (Fig. 15c) swath is comparable to the prediction from NONEWSFC, albeit with slightly smaller maximum hail sizes in Collin County, which suggests that the primary source of the differences between CONTROL and NONEWSFC was the CWOP and Earth Networks data.

Forecasts of MESH are verified quantitatively using performance diagrams (Roebber 2009) that are based on the forecast contingency table. In performance diagrams, the y-axis represents the probability of detection (POD), while the x-axis corresponds to the frequency of hits (FOH) or success ratio (SR). SR is defined as $1 - FAR$, where FAR is the false alarm rate.

$$POD = \frac{Hits}{Hits+Misses} \quad (4)$$

$$FAR = \frac{False\ Alarms}{Hits+False\ Alarms} \quad (5)$$

$$FOH = SR = \frac{Hits}{Hits+False\ Alarms} \quad (6)$$

Dashed lines represent bias, with values <1 corresponding to an under-prediction bias and values >1 representing over-prediction. The hyperbolic lines correspond to the critical success index (CSI):

$$CSI = \frac{Hits}{Hits+Misses+False\ Alarms} \quad (7)$$

Since a perfect forecast has no misses or false alarms, it would be located in the upper-right corner of the diagram where $POD = 1$, $SR = 1$, and $CSI = 1$.

Because severe hail is such a localized phenomenon, the Radar MESH and forecast MESH swaths are modified using a neighborhood threshold, so the forecast is not penalized for small position errors. For each grid point, if hail is observed (forecasted) within the specified radius, then observations (forecasts) of hail are expanded to include that grid point.

Figure 16 presents the performance diagram for 25 mm hail, using a neighborhood threshold of 20 km, for the radar-denial experiments. The CONTROL experiment has the highest POD, indicating that including all available data results in the best severe hail forecast. While the NOKFWS experiment has a somewhat lower POD than the CONTROL experiment, the NOKFWS_CASA experiment has a substantially lower POD than the NOKFWS experiment, indicating that low-level data from the CASA network are able to positively supplement upper-level information from the surrounding WSR-88D radars. The 88DONLY experiment has a lower POD than CONTROL, indicating that the addition of CASA and TDWR data improves the prediction of this hail event. There is a minor decrease in POD for NOTDWR. The NOCASA experiment shows a reduced POD, as well, when compared to CONTROL, indicating that the degradation seen in 88DONLY is likely due to a combination of absent TDWR and CASA data. A lower POD for NOCASAVR, when compared to NOCASA, indicates that including CASA reflectivity data, without the support of the radial velocity data, results in a slight degradation

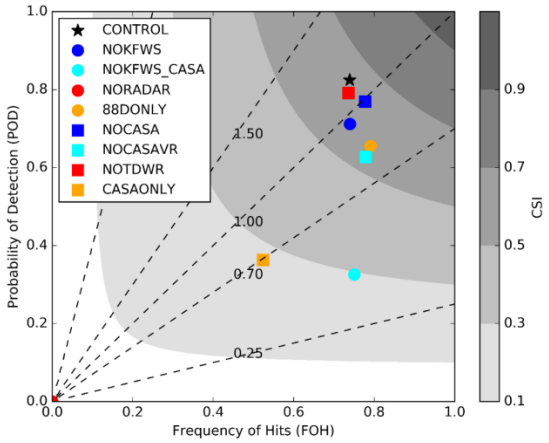


Figure 16: Performance diagram for the radar data-denial experiments using a hail size of 25 mm and neighborhood threshold of 20 km. *Click image to enlarge.*

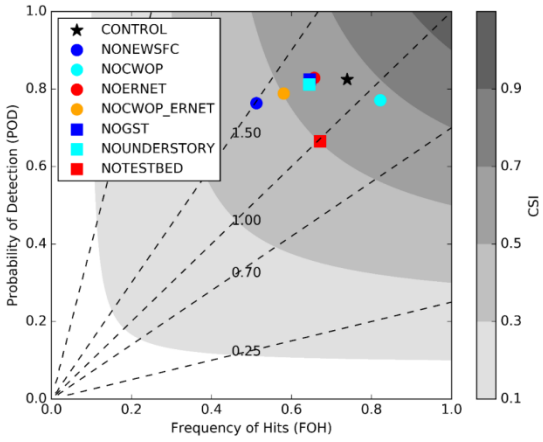


Figure 17: As in Fig. 16, but for the surface data-denial experiments. *Click image to enlarge.*

of the forecast. The NOKFWS_CASA, NORADAR, and CASAONLY experiments suffer from an under-prediction bias, consistent with the loss of the major KFWS data source, and the resulting smaller severe hail in Fig. 14.

Figure 17 shows the performance diagram for the surface data-denial experiments, using a hail size of 25 mm and neighborhood threshold of 20 km. The NOTESTBED experiment is an extension of 88DONLY, in that it denies non-conventional surface data in addition to the non-conventional radar data (CASA and TDWR). Thus, this experiment has a similar performance to 88DONLY. While all experiments shown have a similar POD, there is some variation in the FOH. The NONEWSFC experiment has a lower FOH than CONTROL, as does NOCWOP_ERNET. However, denying individual data sets (NOERNET, NOCWOP, NOUNDERSTORY, and NOGST) does not result in significant differences from CONTROL, indicating that the degradation in NONEWSFC is largely in response to the combined denial of CWOP and Earth Networks data.

c. Verification of surface fields

Forecast performance is also evaluated by considering the forecasts of surface variables; the focus here is on 2-m temperature and dewpoint, which had the greatest variation among the experiments. The root mean square difference (RMSD) and bias compared to twelve independent observing stations (i.e., not assimilated in the 3DVAR analysis system) are used in the RMSD and bias calculations. These reference stations include ten ASOS stations and two Oklahoma Mesonet stations, as shown in Fig. 6b; these reference stations were not included in the data counts in Fig. 5 (44 and 32 for ASOS and mesonet, respectively). These stations were chosen because stations within these networks are carefully sited, well maintained, and regularly calibrated (Brock et al. 1995). Analyses and forecasts are linearly interpolated to the observation location to compute the RMSD and bias values.

Figure 18a shows the 2-m dewpoint RMSDs for the suite of surface data-denial experiments. The RMSD at the beginning (0 min) represents the difference between the observations and the RAP model background. The vertical line at 20 min corresponds to the end of the last data-assimilation cycle. There is some spread evident during the assimilation window of up to 0.5°C. NONEWSFC has the lowest RMSD throughout the majority of the forecast period. NOCWOP_ERNET exhibits a similar pattern to NONEWSFC, indicating that the majority of differences seen in NONEWSFC are attributed to the CWOP and Earth Networks observations. Since both NOCWOP and NOERNET exhibit an RMSD at the end of the forecast period that is similar to CONTROL, the loss of these observation types is likely resulting in the higher skill in NONEWSFC. Thus, non-conventional observations of dewpoint do not contribute to forecast skill for this case study. Bias (model

minus observations) for the 2-m dewpoint field (Fig. 18b) is lower for NONEWSFC throughout the forecast period, again indicating the lack of benefit of non-conventional surface data. The majority of the bias in the dewpoint field is introduced by the CWOP observations.

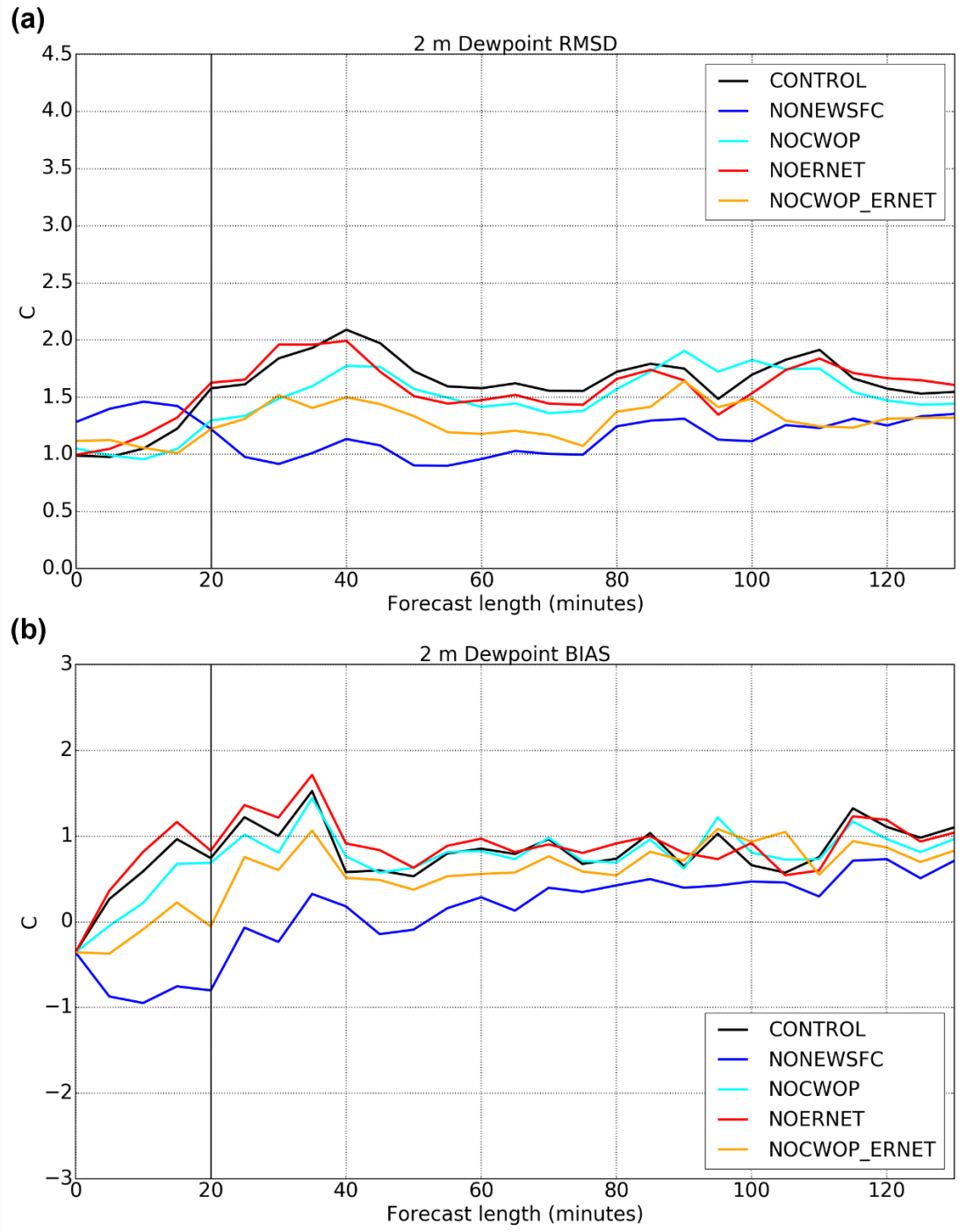


Figure 18: a) Root mean square difference (RMSD), and b) bias for the 2-m dewpoint field in the suite of surface data-denial experiments (legend). The vertical line at 20 min represents the end of the second assimilation window.

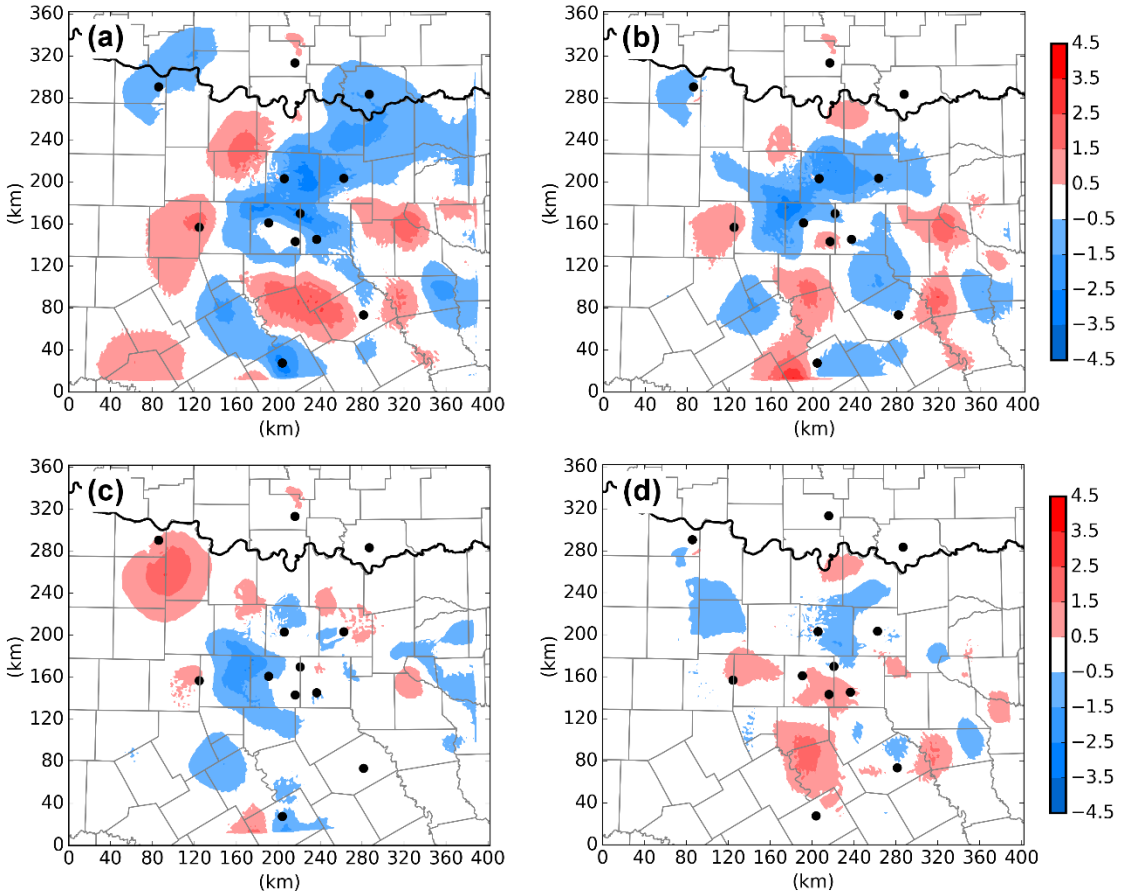


Figure 19: Specific humidity of vapor (q_v) difference fields (g kg^{-1}) at the surface at 2210 UTC, which are determined by subtracting the q_v values for CONTROL from the values for a) NONEWSFC, b) NOCWOP_ERNET, c) NOCWOP, and d) NOERNET, respectively, along with the verification stations shown in Fig. 6b.

Figure 19 shows surface specific humidity (q_v) difference fields at the end of the data assimilation period (2210 UTC), determined by subtracting the q_v values for CONTROL from those of four data denials. NONEWSFC, NOCWOP_ERNET, and NOERNET all exhibit a region of positive q_v difference fields west of the DFW area at 2210 UTC, while NOCWOP does not, indicating that the Earth Networks observations are introducing a dry region to the model background field in this region. Conversely, NONEWSFC, NOCWOP_ERNET, and NOCWOP exhibit a negative q_v difference field just to the east of the aforementioned region, indicating that CWOP observations introduce a moist region to the model background field. Since more verification stations are collocated with the moist area caused by CWOP observations, the moist bias seen in Fig. 18b is largely driven by CWOP observations.

When considering 2-m temperature RMSD for the suite of surface denial experiments (Figure 20a), the NOCWOP_ERNET and NONEWSFC experiments have slightly lower RMSD values than NOCWOP, CONTROL, and NOERNET at the end of the forecast period, although the RMSDs vary by ≈0.25°C throughout the forecast period. The RMSD value for all experiments shown is roughly 2.5°C at the start of the forecast period owing to a position error in the frontal placement in the background fields (i.e., RAP) at 2150 UTC. The RMSD values decreased throughout the forecast period as the front passed the verification stations. As for RMSD, the bias (Fig. 20b) is similar for all experiments throughout the forecast period, although there is a notable cold bias of up to –1.75°C in NONEWSFC from 20–50 min into the forecast. The results are mixed, however, as NONEWSFC has the lowest bias

value at the end of the forecast period. A cold bias of slightly more than 2°C at the beginning of the forecast is due to the aforementioned error in frontal placement.

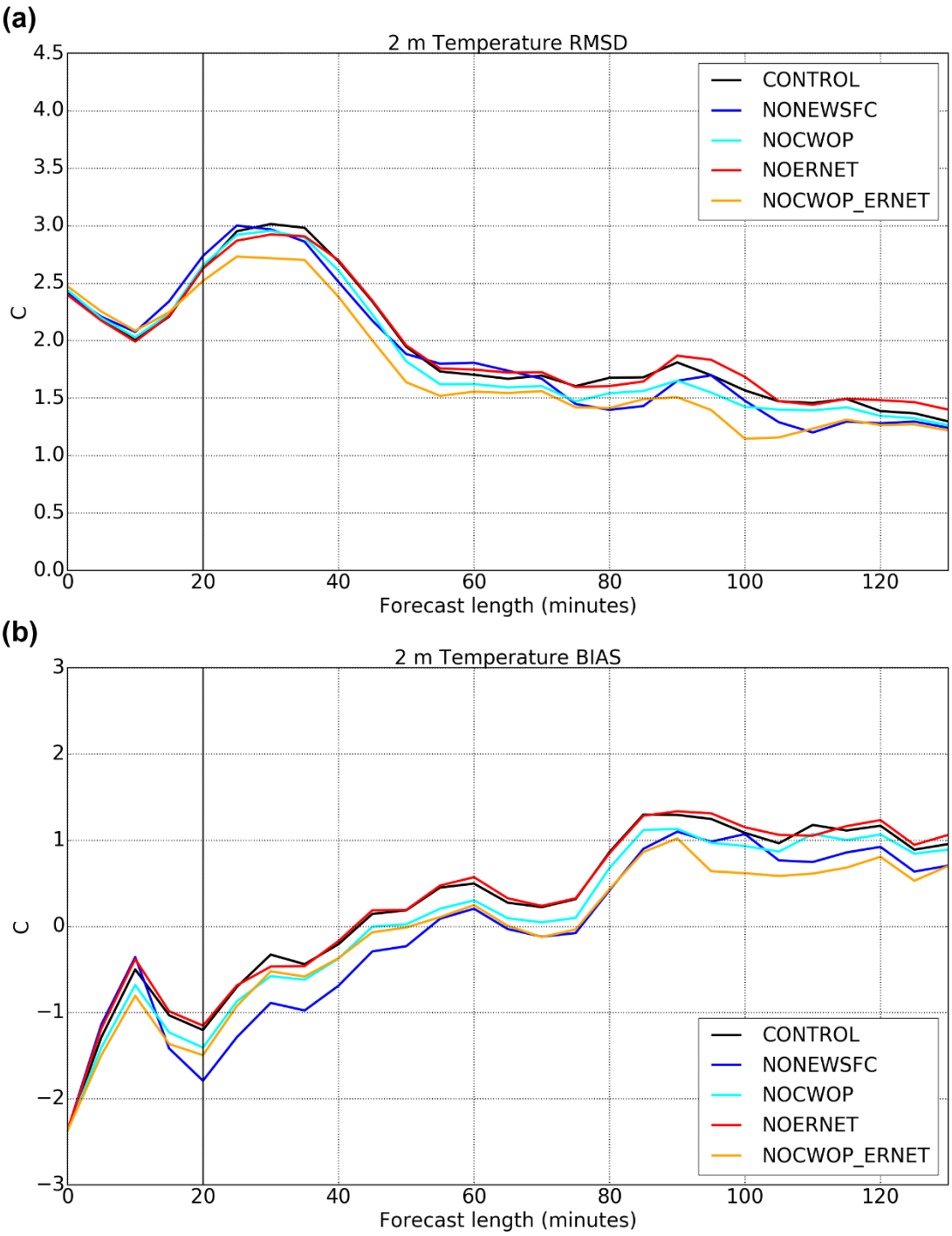


Figure 20: As in Fig. 18, but for 2-m temperature.

6. Summary and discussion

The DFW Testbed was established recently in the DFW Metroplex, as a proof-of-concept for the proposed nationwide "network of networks" (National Research Council 2009), providing the opportunity for OSEs to assess the value of a large variety of non-conventional radar and surface data sources. In this work, OSEs are performed on a prolific hail-producing supercell that impacted the northern portion of the DFW Testbed on 11 April 2016. ARPS and its associated 3DVAR with complex cloud analysis and IAU-VDT data assimilation are used for the experiments.

The CONTROL experiment captures the behavior of the observed supercell as it: (i) correctly depicts the most intense precipitation occurring in Denton County at the end of the data assimilation window (2210 UTC), (ii) handles well the first cycling of the mesocyclone (at 2235 UTC), and (iii) provides a good match between the storm placement at 2300 UTC and the 0.5°-tilt reflectivity observations from the KFWS WSR-88D radar. These qualities make it credible as a control experiment for OSEs. The 88DONLY and NOCASA simulations have an eastward displacement in the updraft and corresponding hail track, and a somewhat weaker circulation than CONTROL (e.g., at 2310 UTC). The hail verification demonstrates that including CASA and TDWR data in the CONTROL experiment aided in the production of a more accurate supercell hailstorm simulation. However, based upon the CASAONLY experiment, the CASA radars with the currently employed low-level-only scanning strategies clearly are insufficient as a standalone data source, as they do not cover the entire depth of the storm.

When comparing the forecast MESH swath from the CONTROL experiment against the Radar MESH swath, the forecast MESH swath: is similar in shape and orientation to the Radar MESH swath, underpredicts the maximum hail size, shows a somewhat larger areal extent of hail compared to Radar MESH, and displaces the swath 20 km northeastward. For the case study considered here, including CASA data improved the forecasted hail, while including TDWR data degraded it somewhat. Low-level CASA data were able to recover some, but not all, of the low-level information that would have been lost had only the KFWS WSR-88D radar been inactive for this case, indicating that CASA radars potentially could alleviate WSR-88D gaps where radar density is less than in the Southern Plains.

The positive impact of the CASA radar data on the forecast was due to changes from the X-band radial velocity in the lowest levels of the atmosphere that are not as well observed by the existing network of WSR-88Ds. Differentiation of impact of CASA radial velocity versus reflectivity follows from slight degradations in the reflectivity and forecasted hail observed in NOCASAVR versus NOCASA. In other words, including CASA reflectivity data without the CASA radial velocity data results in a worse simulation than denying the entire CASA dataset.

While the focus of this study was not on sensitivity to microphysics parameterization schemes, the choice and tunable configuration of a parameterization scheme still can influence the characteristics of simulated hail fields. Loftus et al. (2014) found that a triple-moment microphysics scheme produced large hail that agreed well with observed hail reports for a supercell on 29 June 2000, whereas a double-moment scheme only produced small hail aloft. Furthermore, Dawson et al. (2014) demonstrated that size sorting in the microphysics scheme can impact the simulated differential reflectivity (Z_{DR}) signature. However, Labriola et al. (2019) compared the Milbrandt–Yau double- and triple-moment schemes with the NSSL variable density-rimed ice double-moment scheme and found that the schemes all produced skillful forecasts of the severe hail coverage.

Unlike for Carlaw et al. (2015), the non-conventional dewpoint observations do not contribute to forecast skill for this case study, as RMSD and bias are somewhat lower when these observations are withheld from the analysis in NONEWSFC.

One major limitation of this work is that it only considers one convective mode (i.e., supercell) for one case. While the supercell considered here resulted in extensive damage, it was elevated in the cold sector, which potentially could have limited the forecast benefit of the non-conventional surface data and low-level radial velocity data from the CASA X-band radars. Other studies, though, have shown that CASA radar data are useful for simulating a

quasi-linear convective system (e.g., Schenkman et al. 2011a) and tornadic supercell (e.g., Schenkman et al. 2011b; Stratman and Brewster 2015). More complete results could also be obtained by considering data-denial experiments over a longer period or a series of active days, as the aggregated results would serve as more substantial evidence of the potential value of these observing systems.

Additionally, the seventh CASA X-band radar has been deployed in the eastern portion of the DFW Testbed, with the final radar planned for the northeast corner of the network (see dashed blue circles in Fig. 6b). Once the entire radar network is in place, a longer-term study should be used to ascertain the value of the completed network. The Understory Weather observations used in this study represent a small subset (10) of the observations available beginning in the spring of 2017 (140). A longer-term study also could gauge the potential forecast improvements of this recently completed network.

ACKNOWLEDGMENTS

This work was largely supported by the NOAA/NWS National Mesonet Program via Global Science and Technology (GST; SA15-OU001 and ORA4-47) and NOAA grant NA16OAR4320115. The authors thank Jonathan Labriola and the work done under SHARP project NSF Grant AGS-1261776, for assistance with forecast verification and hail prediction, and Kevin Thomas for data management. The authors also thank Lee Carlaw, Brice Coffer, and Daniel Dawson for their thorough reviews of this manuscript. Observational data were provided by the National Oceanic and Atmospheric Administration (NOAA) via the Meteorological Assimilation Data Ingest System (MADIS). Oklahoma Mesonet data are provided courtesy of the Oklahoma Mesonet, a cooperative venture between Oklahoma State University and the University of Oklahoma, supported by the taxpayers of Oklahoma. Additional surface data were provided by GST and Understory Weather, and crowd-sourced weather event reports were provided via mPING. Computing was performed at the OU Supercomputing Center for Education & Research (OSCER).

REFERENCES

Agustí-Panareda, A., A. Beljaars, C. Cardinali, I. Genkova, and C. Thorncroft, 2010: Impacts of assimilating AMMA soundings on ECMWF analyses and forecasts. *Wea. Forecasting*, **25**, 1142–1160.

Bajaj, A. N. and B. J. Philips, 2012: Casting the net—A revolutionary business model for deploying weather radar networks. *Meteor. Tech. Int.*, 106–108.

Benjamin, S. G., B. D. Jamison, W. R. Moninger, S. R. Sahm, B. E. Schwartz, and T. W. Schlatter, 2010: Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and Mesonet observations via the RUC hourly assimilation cycle. *Mon. Wea. Rev*., **138**, 1319–1343.

Bi, L., J. A. Jung, M. C. Morgan, and J. F. Le Marshall, 2011: Assessment of assimilating ASCAT surface wind retrievals in the NCEP global data assimilation system. *Mon. Wea. Rev.*, **139**, 3405–3421.

Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271.

Bouttier, F., and G. Kelly, 2001: Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **127**, 1469–1488.

Brewster, K., 1996: Application of a Bratseth analysis scheme including Doppler radar data. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 92–95.

——, 2003: ADAS-ARPS data assimilation using incremental analysis updating, 5 pp. [Available online at http://www.caps.ou.edu/ARPS/ARPS5DOC/ADASNudging.pdf.]

——, and D. R. Stratman, 2015: An updated high resolution hydrometeor analysis system using radar and other data. Proc., *23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., 31.

——, M. Hu, M. Xue, and J. Gao, 2005a: Efficient assimilation of radar data at high resolution for short-range numerical weather prediction. Preprints, *World Weather Research Program Int. Symp. on Nowcasting and Very Short Range Forecasting*, Toulouse, France, World Meteor. Org., 3.06.

——, E. C. Fay and F. Junyent, 2005b: How will X-band attenuation affect tornado detection in the CASA IP1 radar network? Preprints, *32nd Conf. on Radar Meteorology*, Albuquerque, NM, Amer. Meteor. Soc., 14.4.

——, L. White, B. Johnson, and J. Brotzge, 2005c: Selecting the sites for CASA NetRad, a collaborative radar network. Preprints, *Ninth Symp. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans and Land Surface (IOAS-AOLS)*, San Diego, CA, Amer. Meteor. Soc., P3.4.

——, F. H. Carr, K. W. Thomas, and D. R. Stratman, 2015: Utilizing heterogeneous radar systems in a real-time high resolution analysis and short-term forecast system in the Dallas/Fort Worth testbed. Proc., *37th Conf. on Radar Meteorology*, Norman, OK, Amer. Meteor. Soc., 14A.3.

——, A. Bajaj, B. J. Philips, D. L. Pepyne, E. Lyons, and F. H. Carr, 2017: CASA Dallas-Fort Worth Urban Testbed observations: Network of Networks at work. Proc., *Special Symp. on Meteorological Observations and Instrumentation*, Seattle, WA, Amer. Meteor. Soc., 1027.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The Oklahoma Mesonet: A technical overview. *J. Atmos. Oceanic Technol.*, **12**, 5–19.

Carlaw, L. B., J. A. Brotzge, and F. H. Carr, 2015: Investigating the impacts of assimilating surface observations on high-resolution forecasts of the 15 May 2013 tornado event. *Electronic J. Severe Storms Meteor.*, **10** (2), 1–34.

Coniglio M. C., S. M. Hitchcock, and K. H. Knopfmeier, 2016: Impact of assimilating preconvective upsonde observations on short-term forecasts of convection observed during MPEX. *Mon. Wea.Rev.*,**144**, 4301–4325.

Crum, T. D., and R. L. Alberty, 1993: The WSR-88D and the WSR-88D operational support facility. *Bull. Amer. Meteor. Soc.*, **74**, 1669–1687.

CWOP, 2014: Hourly number of weather stations on APRS-IS. [Available online at http://www.wxqa.com/checkservers2.html.]

Dabberdt, W. F., and Coauthors, 2005: Multifunctional mesoscale observing networks. *Bull. Amer. Meteor. Soc.*, **86**, 961–982.

Dahlia, J., 2013: The National Mesonet Program: Filling in the gaps. *Weatherwise*, **66**, 26–33.

Danielson, J., and D. B. Gesch, 2011: Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey Tech. Rep. Open-File Rep. 2011-1073, 26 pp. [Available online at https://pubs.usgs.gov/of/2011/1073/pdf/of2011-1073.pdf.]

Dawson, D. T. II, and M. Xue, 2006: Numerical forecasts of the 15–16 June 2002 southern Plains mesoscale convective system: Impact of mesoscale data and cloud analysis. *Mon. Wea. Rev.*, **134**, 1607–1629.

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

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

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

Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249–280.

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

Gasperoni, N. A., X. Wang, K. A. Brewster, and F. H. Carr, 2018: Assessing impacts of the high-frequency assimilation of surface observations for the forecast of convection initiation on 3 April 2014 within the Dallas–Fort Worth test bed. *Mon. Wea. Rev.*, **146**, 3845–3872.

Graham, R. J., S. R. Anderson, and M. J. Bader, 2000: The relative utility of current observation systems to global-scale NWP forecasts. *Quart. J. Roy. Meteor. Soc*., **126**, 2435–2460.

Hart, J. A., and W. D. Korotky, 1991: The SHARP workstation v1.50 users guide. NOAA/National Weather Service, 30 pp.

Hayden, C. M., and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. *J. Appl. Meteor.*, **34**, 3–15.

Hilliker, J. L., G. Akasapu, and G. S. Young, 2010: Assessing the short-term forecast capability of nonstandardized surface observations using the national digital forecast database (NDFD). *J. Appl. Meteor. Climatol.*, **49**, 1397–1411.

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

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

Istok, M., and Coauthors, 2008: Terminal Doppler Weather Radar for NWS operations: Phase 3 update. Proc., *24th Conf. on Interactive Information and Processing Systems*, New Orleans, LA, Amer. Meteor. Soc., 6B.10.

Kazumori, M., Q. Liu, R. Treadon, and J. C. Derber, 2008: Impact study of AMSR-E radiances in the NCEP global data assimilation system. *Mon. Wea. Rev.*, **136**, 541–559.

Keclik, A. M., C. Evans, P. J. Roebber, and G. S. Romine, 2017: The influence of assimilated upstream, preconvective dropsonde observations on ensemble forecasts of convection initiation during the Mesoscale Predictability Experiment. *Mon. Wea. Rev.*, **145**, 4747–4770.

Kineman, J. J., 1992: Global Ecosystems Database version 1.0 users guide: Key to geophysical records documentation no. 26, USDOC/NOAA National Geophysical Data Center, Boulder, CO., 121 pp.

Labriola, J., N. Snook, Y. Jung, and M. Xue, 2019: Explicit ensemble prediction of hail in 19 May 2013 Oklahoma City thunderstorms and analysis of hail growth processes with several multi-moment microphysics schemes. *Mon. Wea. Rev.*, **147**, 1193–1213.

Lang, S., and E. McKeogh, 2011: LIDAR and SODAR measurements of wind speed and direction in upland terrain for wind energy purposes. *Remote Sens.*, **3**, 1871–1901.

Loftus, A. M., and W. R. Cotton, 2014: A triple-moment hail bulk microphysics scheme. Part II: Verification and comparison with two-moment bulk microphysics. *Atmos. Res.*, **150**, 97–128.

MADIS, 2018: MADIS quality control. Accessed 28 January 2021. [Available online at https://madis.ncep.noaa.gov/madis_qc.shtml.]

McLaughlin, D., and Coauthors, 2009: Short-wavelength technology and the potential for distributed networks of small radar systems. *Bull. Amer. Meteor. Soc.*, **90**, 1797–1817.

McPherson, R. A., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **24**, 301–321.

Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064.

Moninger, W. R., R. D. Mamrosh, and P. M. Pauley, 2003: Automated meteorological reports from commercial aircraft. *Bull. Amer. Meteor. Soc.*, **84**, 203–216.

National Research Council, 2009: *Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks*. The National Academies Press, 250 pp.

——, 2012: *Urban Meteorology: Forecasting, Monitoring, and Meeting Users' Needs*. The National Academies Press, 176 pp.

Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549.

NWS, 1994: Technique specification package 88-21-R2 for AWIPS-90 RFP Appendix G requirements numbers: Quality control incoming data. AWIPS Document Number TSP-032-1992R2. [Available online at ftp://ftp.library.noaa.gov/noaa_documents.lib/NWS/NWS_TSP_88-21-R2.pdf.]

——, 2016: April 11th 2016 hail storm. [Available online at http://www.weather.gov/fwd/20160411.]

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608.

Rogers, R. R., and M. K. Yau, 1989: *A Short Course in Cloud Physics*. Butterworth-Heinemann, 290 pp.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011a: The analysis and prediction of the 8–9 May 2007 Oklahoma tornadic mesoscale convective system by assimilating WSR-88D and CASA radar data using 3DVAR. *Mon. Wea. Rev.*, **139**, 224–246.

——, ——, ——, ——, and ——, 2011b: Impact of CASA radar and Oklahoma mesonet data assimilation on the analysis and prediction of tornadic mesovortices in an MCS. *Mon. Wea. Rev.*, **139**, 3422–3445.

Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, and J. W. Lipe, 2005: The West Texas Mesonet: A technical overview. *J. Atmos. Oceanic Technol.*, **22**, 211–222.

Smith, T. L., S. G. Benjamin, S. I. Gutman, and S. Sahm, 2007: Short-range forecast impact from assimilation of GPS-IPW observations into the Rapid Update Cycle. *Mon. Wea. Rev.*, **135**, 2914–2930.

Snook, N., M. Xue, and Y. Jung, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140**, 2126–2146.

——, Y. Jung, J. Brotzge, B. Putnam, and M. Xue, 2016: Prediction and ensemble forecast verification of hail in the supercell storms of 20 May 2013. *Wea. Forecasting.*, **31**, 811–825.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16.

Stratman, D. R., and K. A. Brewster, 2015: Impact of assimilating CASA X-Band radar data for 24 May 2011 tornadic storms using various microphysics schemes at 1-km grid resolution. Proc., *37th Conf. on Radar Meteorology*, Norman, OK, Amer. Meteor. Soc., P221.

Tyndall, D. P., and J. D. Horel, 2013: Impacts of mesonet observations on meteorological surface analyses. *Wea. Forecasting*, **28**, 254–269.

Wainwright, C. E., D. T. Dawson, M. Xue, and G. Zhang, 2014: Diagnosing the intercept parameters of the exponential drop size distributions in a single-moment microphysics scheme and impact on supercell storm simulations. *J. Appl. Meteor. Climatol.*, **53**, 2072–2090.

Willmot, E., S. Bussmann, N. Homeier, and A. Kubicek, 2017: Examples of ground-based hail, rain, and wind sensor networks operating in real-time. Proc., *IEEE International Geoscience and Remote Sensing Symp. (IGARSS)*, 5987–5989.

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303.

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161–193.

——, and Coauthors, 2001: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143–165.

——, 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. *Adv. Meteor.*, **2013**, 1–9.

——, M. Hu, and A. D. Schenkman, 2014: Numerical prediction of the 8 May 2003 Oklahoma City tornadic supercell and embedded tornado using ARPS with the assimilation of WSR-88D data. *Wea. Forecasting*, **29**, 39–62.

Zapotocny, T. H., W. P. Menzel, J. P. Nelson III, and J. A. Jung, 2002: An impact study of five remotely sensed and five in situ data types in the Eta data assimilation system. *Wea. Forecasting*, **17**, 263–285.

——, ——, J. A. Jung, and J. P. Nelson III, 2005: A four-season impact study of rawinsonde, GOES, and POES data in the Eta data assimilation system. Part II: Contribution of the components. *Wea. Forecasting*, **20**, 178–198.

——, J. A. Jung, J. F. Le Marshall, and R. E. Treadon, 2007: A two-season impact study of satellite and in situ data in the NCEP global data assimilation system. *Wea. Forecasting*, **22**, 887–909.

Zhao, K., and M. Xue, 2009: Assimilation of coastal Doppler radar data with the ARPS 3DVAR and cloud analysis for the prediction of Hurricane Ike (2008). *Geophys. Res. Lett.*, **36**, L12803.

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

REVIEWER A (Lee B. Carlaw):

Initial Review:

Recommendation: Accept with major revisions.

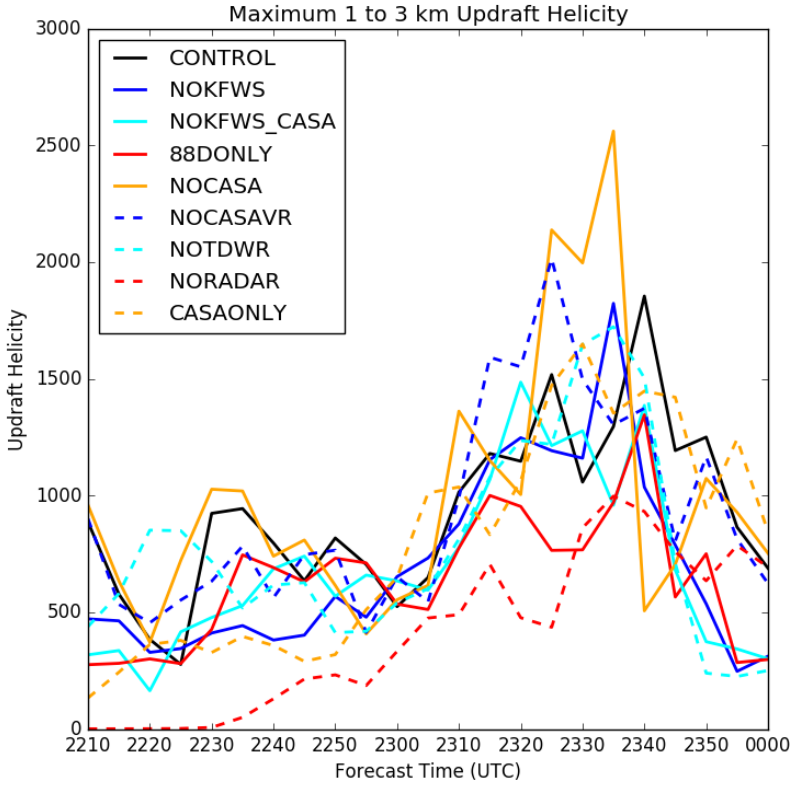
General Comments: This is a well-written manuscript describing the benefits of including several "non-conventional" observing systems during the data assimilation process in the simulation of a prolific hail-producing supercell near DFW. I believe this work has applications to the developing Warn-on-Forecast initiative, specifically as it relates to rapidly-cycled short range simulations of convection, with additional implications for the use of low-cost X-band radars to fill gaps in the nation's WSR-88D network. My comments are mostly of the minor-to-moderate type, but there are a few somewhat more notable concerns regarding the level of analysis performed for the radar denial experiments, as well as some consistency issues with some of the figures that I'd like to see addressed before formal publication, rendering the "major revisions" designation.

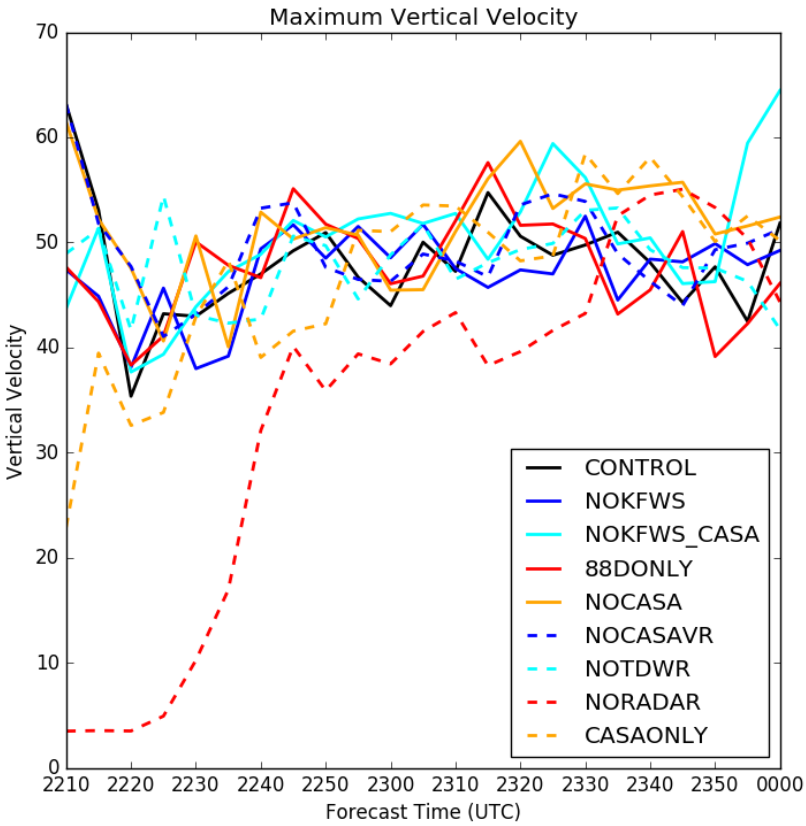
Major Comments: One of the primary findings of this paper is the demonstrable improvement in the simulation of the supercell storm (as it pertains to the forecasted hail swaths) when allowing both TDWR and CASA data during the assimilation window. Given the importance of this finding (at least for this particular event/simulation), I feel there is a lack of discussion on what information was added by the TDWR and CASA Z and V increments during the 20-min assimilation cycles which aided in the simulation of a clearly more robust mid-level mesocyclone. Did the inclusion of lower to mid-level scans of radial velocity allow the weak anelastic mass constraint in the 3DVAR step to gradually develop more robust mid-level vertical velocities through the assimilation window? You seem to nod at this possibility with the NOCASAVR experiment, as well as the note where you state "...including the low-level radial velocity data from the CASA X-band radars afforded a better initial analysis of the wind field, which resulted in a superior simulation in the CONTROL experiment", but this analysis seems incomplete.

Furthermore, as Dawson and and Xue (2006) showed, there can be notable impacts on how ARPS develops vertical velocities by adjusting to mid-level warming introduced during the cloud analysis. Could the inclusion of low- and mid-level reflectivities from the TDWR and CASA data have supported the development of a more robust updraft through the assimilation steps? A time-series analysis of vertical velocity or some similar diagnostic variable in a region bounding the updraft for the CONTROL, NOTDWR, NOCASA, NOCASAVR, and 88DONLY throughout the cycled 20-min data-assimilation window might provide some useful insights here, as an initial thought.

The NOCASA and NOCASAVR have very similar results, so we conclude that most of the impact is due to the CASA radial velocities. These velocities would be directly affected in the lowest 2 km. Careful examination of the wind fields at the end of the second assimilation window shows the CONTROL has some stronger north winds along the Wise-Denton County line, to the west of the mesocyclone at that time. This affected the evolution of the storm circulation in the ensuing 30 min as the initial rotation center and updraft were replaced by one to the south as the storm cycled. The effect was to form the new circulation center slightly further west from that of NOCASA and NOCASAVR at 2235 UTC, and this more closely matched the low-level circulation center as viewed from KFWS at that time. The MESH tracks are also somewhat further west in CONTROL, closer to the verification track. These details were added to Section 5a, along with a supporting figure.

Examination of updraft helicity values in the vicinity of the storm indicates that the CONTROL experiment has slightly lower UH values at the end of the assimilation, but both are quite strong values. Subsequent to this time there is a slight reduction in the UH as the mesocyclone cycles as described above, so stronger is not necessarily better. Plots of the max UH and w time-series are provided here per reviewer request but are omitted from the paper as there is not a clear indication of which is more "correct".


Maximum 1 to 3 km Updraft Helicity
Updraft Helicity
Forecast Time (UTC)
CONTROL
NOKFWS
NOKFWS_CASA
88DONLY
NOCASA
NOCASAVR
NOTDWR
NORADAR
CASAONLY
0
500
1000
1500
2000
2500
3000
2210
2220
2230
2240
2250
2300
2310
2320
2330
2340
2350
0000


Maximum Vertical Velocity
Vertical Velocity
Forecast Time (UTC)
CONTROL
NOKFWS
NOKFWS_CASA
88DONLY
NOCASA
NOCASAVR
NOTDWR
NORADAR
CASAONLY
0
10
20
30
40
50
60
70
2210
2220
2230
2240
2250
2300
2310
2320
2330
2340
2350
0000

I wonder if the 2-m dewpoint analysis might be unnecessary as a standalone subsection for this paper. While I can see the value of discussing the impacts of including the non-conventional observations on the near-surface moisture fields in a dissertation, I'm not sure it adds enough information to warrant inclusion in this manuscript. Additionally, it looks like you might be close to overfitting to the non-conventional observations based on the difference analyses in Fig. 19, so I'd be interested to see the observation error values if you decide to keep this section. Otherwise, my suggestion would be to simply reference the noted degradation in the moisture forecast in the text. The main findings here—which I think are more novel and interesting—relate to the inclusion of TDWR and CASA radar data, and I think this dewpoint analysis section detracts from this a bit.

The dewpoint discussion was included as a means of distinguishing this case study from that discussed in Carlaw et al. (2015), which is discussed in the introduction section. In the case study presented here, the non-conventional dewpoint observations did not improve the RMSD and bias scores at the selected verification stations. However, the non-conventional surface data did result in overall improvements in the hail forecast as seen in the hail swaths and performance diagram. One important caveat is that this is a single case study and conclusions should not be extrapolated to new cases without more complete studies, such as those discussed in the future work.

The verification of the 88DONLY's hail forecast on the performance diagram in Fig. 16 appears inconsistent with the plan view forecast when compared to the other experiments in Fig. 14, as well as the reflectivity output in Figs. 9 and 10, and the text discussion. There are no pixels of hail sizes >20 mm anywhere near the >25-mm MESH contours in Fig. 14c (88DONLY), which should result in a POD and FOH around 0. This leads me to believe there may have been some mislabeling, traced back most likely to Fig. 14, but this needs to be verified.

The MESH swath figures were mislabeled initially. Panel (b) should have been 88DONLY, while panel (c) should have been NORADAR. The figure has been updated to reverse the ordering of the two panels, which follows more closely with the discussion in the text than simply relabeling the figures. The labels for the remaining experiments were also verified.

[Minor comments omitted…]

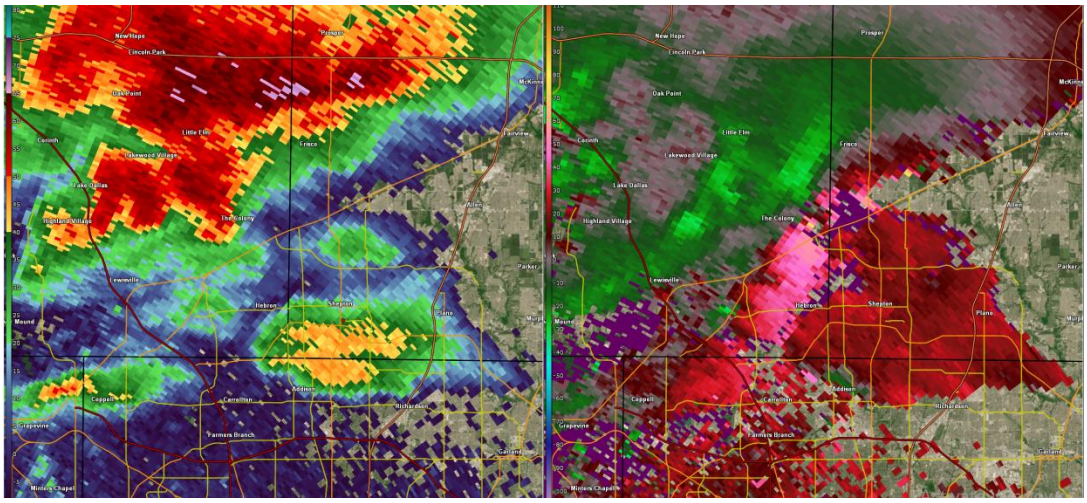
Second Review:

Recommendation: Accept with minor revisions.

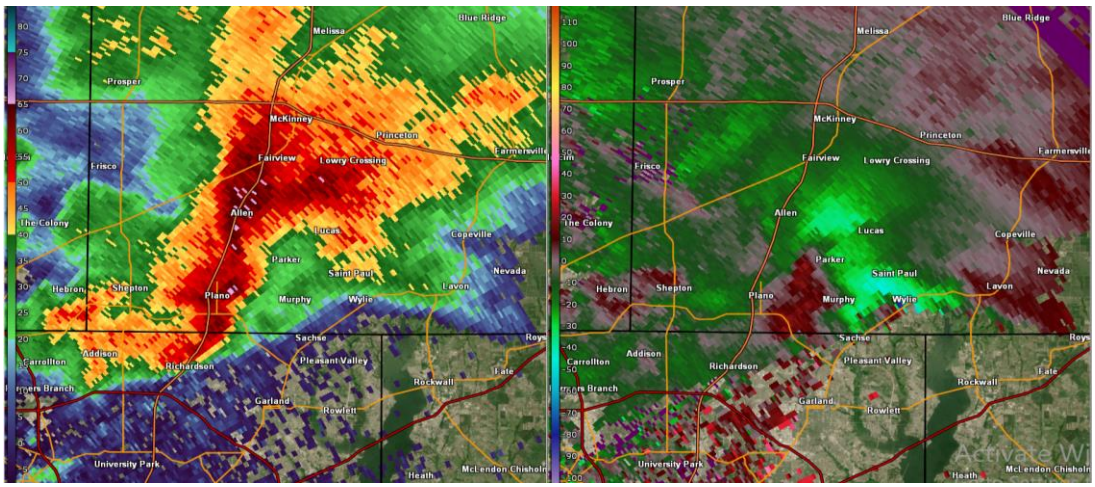
General Comments: This manuscript is a resubmission of a paper examining the impacts of various data denial experiments on the simulation of a hail-producing supercell near DFW. The authors have done a good job responding to my more substantive comments relating to the level of the analysis among the various denial experiments. Reading the manuscript with the added material did bring up an additional thought regarding analysis that attempts to drill further into the differences that manifest specifically between the CONTROL and NOCASAVR experiments, which ultimately is one of the more salient points of the manuscript. Since this discussion starts to veer away from the level of analysis required of a typical OSE, I don't think this necessitates another round of major revisions, but I'd nonetheless like to see the author's thoughts on a few additional comments.

Substantive Comments: The additional discussion relating to the various CASA denial experiments was very helpful, and the authors have clearly spent some time improving the manuscript in this regard. Fleshing out seemingly small, innocuous differences in a high-resolution simulation that can propagate downstream in quite a nonlinear manner can be extremely challenging! To that end, the more that I look at Figs. 12 (v-wind analysis) and 14 (MESH swaths), the more it appears the most notable impacts between CONTROL and NOCASA/NOCASAVR seem to occur deeper into the simulation (i.e. towards and past 2230 UTC and likely even around 2300-2310 UTC) when this supercell was producing its largest hail in southern Collin County. The differences between CONTROL and NOCASAVR are fairly striking at that point: NOCASAVR fails to produce much larger than 25-mm MESH tracks while the CONTROL simulation produces corridors of "significant" hail roughly in the proper location, with the addition of CASA radial velocities.

Looping through KFWS radar data from that event, there are no fewer than five reflectivity "tags" (see https://www.weather.gov/media/ilx/Research-Case_Studies/barker_sls06.pdf and https://ams.confex.com/ams/27SLS/webprogram/Manuscript/Paper254200/Shimon-2014AMS-SLS-AbstractFINALa.pdf for examples) that rotate northward into the main supercell's inflow. Your 1-km AGL v-wind analysis seems to capture one of the more predominant "tags" in far southwestern Collin County (see the included image below). The outbound velocity orientation is getting pretty oblique to the KFWS radar in this area, so the radial component is muted compared to the stronger southerlies that are likely occurring; this would align with the stout southerlies in Fig. 12. By about 2255 UTC, this feature is completely absorbed into the supercell, at which point extreme updraft accelerations can be inferred by the development of a notable BWER which eventually extends past 26 000 ft (7925 m) (see KFWS 5.1° scan at 23:04:28 for reference; not included here).



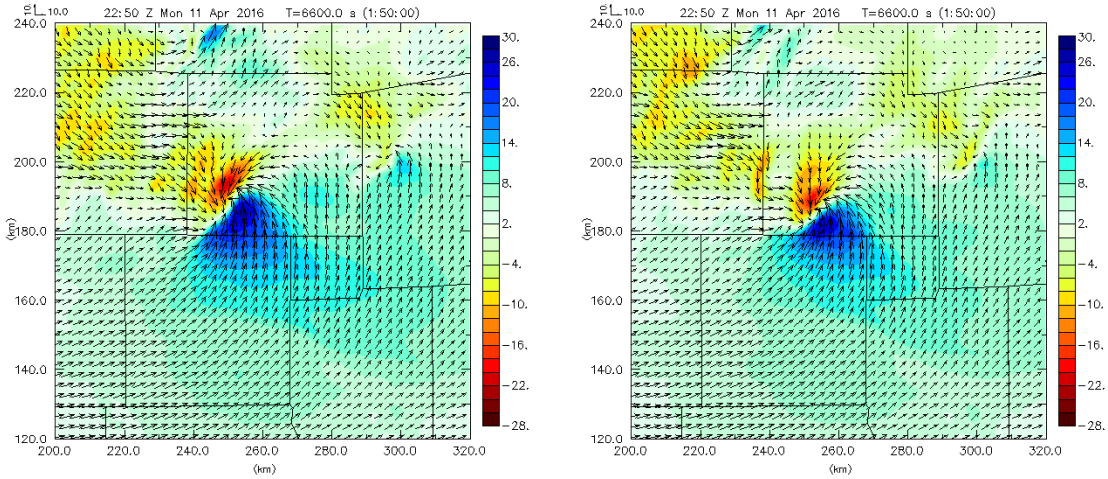
[Above] 0.9° scan from KFWS *Z* (left) and V (right) at 22:34:57, about the same time as Figs. 12b and 12d. While there is likely some degree of sidelobe contamination in southeastern Denton County, the region of southerlies in southwestern Collin County seems to match up well with Figs. 12b and 12d. For reference, the radar beam is ≈4000–4500 ft (1219–1372 m) ARL in the storm inflow.



[Above] 0.5° scan at 22:57:23 as the last reflectivity "tag" is ingested. This is the last volume scan before significant updraft accelerations occurred. Strong inbounds (~40 kt; north or northeasterly winds) at ≈4000 ft (1219 m) ARL.

Do the 1–2 km AGL V-wind and reflectivity plan views look any different, specifically between the CONTROL and NOCASAVR experiments during the (roughly) 2250 UTC time frame as the last reflectivity tag is ingested into the supercell inflow? Is this tag or associated strengthening northeasterly winds even present in the simulations at this point? I'm not sure you'd be able to track any of these differences back to the last assimilation step at 2210 UTC without something more robust like a backward Lagrangian trajectory analysis, but the inclusion of the CASA V_r data seems to result in some storm-scale effects that ultimately manifest in a more robust hail-producing core in southern Collin County.

We have examined carefully animations of the CONTROL and NOCASAVR at 1.5 km AGL in the period 2240–2340 UTC. The below figure shows the CONTROL (left) and NOCASAVR (right) v-wind forecast valid 2250 UTC. Although we cannot confidently identify specific reflectivity tags that match those in the radar images, we can see waves of stronger winds (downdraft surges) that enter the updraft/circulation area generally from the NW; we see this in both forecasts. The main improvement remains in the more westward location of the circulation center and in the hail forecast.



Is there a specific reason for selecting the 1–3-km AGL layer to use for the updraft helicity calculations in Fig. 11 and section 5? I'm generally used to seeing the 2–5-km AGL layer in operational forecast products, and this generally seems to be the preferred layer in research since it encompasses more of the region where theory tells us vertical dynamic perturbation pressure gradient forces are/should be maximized.

The 1–3-km AGL layer was chosen to capture the effects of the CASA radial velocity data in the vicinity of the supercell. With the exception of the Cleburne radar, the beam heights of the CASA radars do not extend above 2.5 km. While not shown explicitly in the manuscript, other variations of updraft helicity were assessed; these showed that the differences were greatest in the 1–3-km layer.

[Minor comments omitted…]

Third Review:

Recommendation: Accept with minor revisions.

General Comments: The authors have addressed my previous questions. There are just a few minor points that came up during my last read-through of the revised manuscript. Because of the minor nature of these comments (don't be alarmed by the length of the review!), I do not need to see the paper again, but would like to see the author's responses to the comments below. Otherwise, I believe this paper is in acceptable form for publication.

[Minor comments omitted…]

REVIEWER B (Brice E. Coffer):

Initial Review:

Recommendation: Accept with major revisions.

General comments: This paper presents data-denial experiments assimilating novel, non-conventional data from the CASA DFW Testbed into high-resolution convection forecasts for a high-impact severe hail event. The research presented is interesting, albeit relatively straightforward (i.e., more observations in the lower troposphere generally improves the forecast), and would be of interest to the readers of EJSSM. I am not an expert in data assimilation, so I approached this review more from the storm-scale processes viewpoint. Most of my comments revolve around more interpretability in how the supercell and other environmental fields evolve in response to more/less data assimilation in the lower troposphere. Along these lines, I would like to see more discussion on *why* the simulations are different, not just that there are differences. Considering the scope of the comments, I would like to see the manuscript again before full acceptance.

Major comments: Can the authors elaborate how the CASA radars impact the forecast? Does the higher spatial resolution associated with X-Band help the most or is it a function of increased temporal frequency of scans through the lowest few km?

More detail was added to Section 5a, see also reply to Reviewer A, Carlaw, Major Comment 1.

We know for sure that the positive impact of the CASA radar data on the forecast was largely due to the inclusion of radial velocity data; this manifests as degradations in the reflectivity and forecasted hail when comparing NOCASAVR against NOCASA, which shows that including CASA reflectivity data without the CASA radial velocity data results in a worse simulation than denying the entire CASA dataset. The CASA data were only assimilated at 2150 UTC and 2200 UTC, so the influence of the increased temporal frequency of scans was not assessed in this study.

In general, there is little discussion for why the simulations are different. The authors should go more in-depth than just cursory glances at the wind field and reflectivity at 2 km. How do fields like surface boundaries, CAPE/shear, updraft helicity, cold pool strength, etc., vary between the simulations?

A figure showing the 1–3-km updraft helicity valid at the end of the assimilation (2210 UTC) has been added to the manuscript. Some additional discussion was added per another reviewer's similar request. Many fields were examined but are not discussed owing to space constraints and only small differences among results.

In section 5a, there is a conclusion that including the low-level radial velocity data from the CASA X-band radars resulted in an improved analysis of the wind field and a superior simulation in the control experiment. It is unclear to me how the authors come to that conclusion, considering there is no "truth" to compare the wind field to. In general, I find it hard to compare the wind field between the simulations. Perhaps difference fields could help here.

Please refer to response to Carlaw's comment #1 for more detail on this.

Not enough data are presented for the surface data-denial experiments. It seems natural to also show reflectivity and MESH for the simulations discussed in Fig. 17, especially considering the success ratio almost doubles with the added surface data. What do these storms look like? How do the surface fields differ?

The reflectivity fields for the surface data denial experiments are similar to CONTROL at 2210 UTC and the storm placement is similar to CONTROL in all surface data denial experiments at 2310 UTC. This has been discussed briefly in section 5a without the addition of new reflectivity figures. The MESH swaths for

several surface data denial experiments, namely NONEWSFC and NOCWOP_ERNET have been added to section 5b.

Section 5c leaves the reader with more questions than answers. It is strange that only dewpoint is considered despite this apparently contributing no forecast skill to this case study. RMSD and bias are lower in NONEWSFC, however the false alarm rate is nearly double in NONEWSFC compared to the CONTROL run (Fig. 17). Surely the new surface observations are providing value to the forecast, however the authors do not elaborate on why. Similar to my previous comment, more analysis for the surface data denial experiments seems warranted.

The RMSD and bias time series for temperature have been added to section 5c, which further adds to the discussion on the surface data denial experiments.

The language in the conclusions is vague at times. For example, "Comparing the forecast MESH swath to radar MESH swath, the location of the hail swath is captured quite well". "Quite well" seems rather arbitrary without a direct meaning. In this location and others, I encourage the authors to use less-ambiguous terms.

The comparisons between the CONTROL-simulated reflectivity (MESH) and observed reflectivity (MESH) in the conclusions section were made more detailed.

[Minor comments omitted…]

Second Review:

Recommendation: Accept.

General Comment: I believe the authors have adequately addressed my comments, as well as the other two reviewers, and thus I would be comfortable recommending publication in EJSSM.

REVIEWER C (Daniel T. Dawson II):

Initial Review:

Recommendation: Accept with major revisions.

Synopsis: This study performs analyses and simulations of a severe hail-producing supercell that traversed the CASA DFW Testbed using the ARPS 3DVAR+cloud analysis+forward model system. A control experiment is performed in which conventional surface and radar observations are assimilated along with surface and radar observations from several additional special observational networks (most notably the CASA radars). Then, several data-denial experiments are performed and a combination of qualitative (for radar reflectivity) and quantitative (for hail observations) comparisons with observations are made to assess the value of each network in terms of its impact on forecast skill. The authors find that assimilating the CASA and TDWR radar data improves forecasts of the hail swath from the supercell, but that the assimilation of special surface networks does not show a clear benefit, and even degrades the forecast of the surface moisture field. Overall I think this study is a useful contribution, mainly by virtue of the demonstration of the utility of "gap-filling" radars to improve forecasts of severe convective storms. However, there are some fairly substantial issues with the manuscript in its present form. My biggest concerns have to do with a relatively weak connection of the study with past research (specifically the use of 3DVAR vs. EnKF and the pros/cons thereof), a somewhat underdeveloped comparison of the forecast and observed reflectivity and kinematic structure, and a lack of discussion of the importance of the microphysics scheme in regard to the simulated hail swath. These and other more minor issues are described in detail below and in various comments embedded in the annotated copy of the manuscript. My overall recommendation is acceptance pending major revisions.

Substantive Comments: The discussion of previous storm-scale radar data assimilation and forecast studies in the introduction could use some more context and connection with the current study. In particular, the studies cited differed in the data assimilation methods used. Some used 3DVAR+cloud analysis, like the present study, but others used EnKF. Of course, in both cases the end goal is to produce an optimum analysis (in the least-squares minimization sense), but the methods differ substantially. At a minimum, these differences should be briefly discussed, compared, and contrasted, and the reasons for choosing 3DVAR in the present study clearly outlined.

Within the introduction section, we have clarified which data assimilation system (3DVAR, EnKF, etc.) was used for each of the articles cited. In addition, the following text has been added to the end of the discussion on radar assimilation experiments: "While the methods employed in these studies varied (3DVAR, EnKF, etc.), the objective of each analysis was to generate an optimum analysis in the least-squares minimization." It is beyond the scope of this article to discuss detailed differences among various data assimilation methods. This work is a follow-on to Carlaw et al. (2015), which used 3DVAR for analysis increments; here IAU has been updated to the IAU with Variable-Dependent Timing for the assimilation of those increments.

Referring to Fig. 10c, the fact that the CASAONLY experiment still manages to produce an intense supercell in the forecast at 2310 UTC that is only moderately displaced from that of the other experiments, despite it being absent (at least in the hydrometeor fields) in the final analysis at 2210 UTC is somewhat surprising, and strikes me as an important result, but not much is said about this. Are the wind and/or thermodynamic fields still well-analyzed, such that the hydrometeor fields quickly "spin up" in the forecast? Or, is the mesoscale background forcing analyzed well enough such that a storm would form anyway, regardless of the assimilation of radar data? Looking at the forecast hail swath for the NORADAR experiment (Fig. 14b), it seems that this indeed the case, as the overall intensity of the swath is comparable to that of CONTROL (Fig. 14a). EDIT: it appears that Fig. 14b is mislabeled and in fact Fig. 14c should be the NORADAR experiment. But, there is still a hail swath, albeit broken and weak, so this may still be relevant.

While CASAONLY and NORADAR are both degraded from CONTROL, owing to delayed convective initiation, the CASAONLY simulation is superior to NORADAR, as evidenced by the MESH swath and performance diagram. At 2210 UTC, there are precipitation echoes present at 2 km AGL in CASAONLY that are not present in NORADAR. Convective initiation was delayed in NORADAR relative to CASAONLY, resulting in a weaker storm that was displaced well to the northwest of the observed cell, which implies the differences between CASAONLY and NORADAR are due to the CASA radial velocity data as both simulations used the same surface data. This can be attributed to the use of IAU-VDT and the latent heat from the cloud analysis in the data assimilation; the storm is largely forced from the bottom up, so a better depiction of the low-level convergence owing to additional radial velocity data would enable faster convective initiation, and the latent heat perturbations applied at low-levels in this high-CAPE environment would rapidly produce a storm. This has been briefly discussed in section 5a.

In section 1a, the authors qualitatively analyze the experiments regarding the reflectivity structure as compared with low-level observed radar scans. This is fine as a first step, but later in the section they also compare the mesocyclone circulation are also made between the simulations, but with no corresponding reference to the observations. Instead, they assume that a stronger/more coherent circulation is "superior". (However, the quantification of improved hail swath forecasts for these experiments discussed in the subsequent section does lend some *indirect* support to these statements). Without reference to the observations, such an inferential leap is dubious. Thus, these value judgments should be removed, or at the very least clarified as speculative. Even better would be to compare the simulated velocity structure directly with the observations (which could be done by computing the radial velocity field from the model kinematic fields and comparing with the remapped radial velocity field from KFWS or one or more of the CASA radars). Another possibility is to create additional 3DVAR analyses to cover the free forecast time and compare the forecasts to these.

See our reply to Carlaw, Major Comment 1.

Regarding the quantitative verification of hail swaths, numerous studies have shown that the characteristics of simulated hail fields (not to mention numerous other aspects of storms) are strongly sensitive to a chosen microphysics scheme and to tunable parameters within that scheme. Though I understand that the focus of this study is not on microphysics sensitivity, I think it is a weak point that the authors don't even mention this issue or discuss the characteristics of the MY scheme in regard to its performance for hail simulation/forecast verification. (A few studies that have looked at this in addition to the already cited Snook et al. 2016 are shown below). I think that at the very least some justification for the choice of the MY double-moment scheme needs to be made.

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

Labriola, J., N. A. Snook, Y. Jung, and M. Xue, 2019: Explicit ensemble prediction of hail in 19 May 2013 Oklahoma City thunderstorms and analysis of hail growth processes with several multi-moment microphysics schemes. *Mon. Wea. Rev.*, **147**, 1193–1213.

Loftus, A. M., and W. R. Cotton, 2014: A triple-moment hail bulk microphysics scheme. Part II: Verification and comparison with two-moment bulk microphysics. *Atmos. Res.*, **150**, 97–128.

Some discussion on microphysics sensitivity was added to the discussion section, along with a brief overview of the findings from the above three studies. Labriola et al. (2019) showed that the double-moment scheme produced skillful forecasts of severe hail coverage.

[Minor comments omitted…]

Second Review:

Recommendation: Accept with minor revisions.

Synopsis: This paper has improved substantially from the initial submission. Most of my concerns have been addressed, and the authors have done a great job synthesizing them with those of the other reviewers. I think the paper is closer to acceptance, but I still have one concern that I do not think was adequately addressed. This concern is regarding the discussion of the simulated velocity structure (particularly the mesocyclone) of the storm without any comparisons to the observations. Otherwise, I only have a few other minor comments and overall I recommend acceptance pending minor revisions.

Substantive Comment: This comment is a follow-up to my previous substantive comment #3 regarding the qualitative storm morphology comparison in section 5a. In their response to that comment, the authors pointed to their response to Lee Carlaw's (Reviewer 1) original major comment #1. While the additional discussion of the differences between the experiments in the context of the impact of the various special radar datasets is definitely welcome, this response did not actually address my concern (also echoed by Brice Coffer's (Reviewer 2) original major comment #3) which was instead focused on the lack of a comparison of the simulated velocity structure with observations. Granted, the authors did concede that it wasn't clear which forecast was superior in that response, but this was not reflected in the revised text at all, as far as I could see. Again, the only observations that are explicitly shown in this discussion and the associated figures are the low-level radar reflectivity scans in Fig. 8. Yet the text discusses the forecasts of the circulation relative to CONTROL and uses such wording as "degraded" or "recovered" as if it is known that CONTROL has the best one. There is even a discussion of the cycling of the mesocyclone and the associated superior performance of CONTROL. But, no observations (or analyses derived from them) during the forecast period are presented, in figure form or otherwise, to back up these claims. So, my original comment on this issue still stands. I reiterate my earlier recommendation to either walk back the claims regarding the forecasts of the mesocyclone or provide some explicit comparison of the relevant observations with the forecasts.

The text in Section 5a has been updated to address these concerns; in addition, observed reflectivity and radial velocity panels have been added to several figures in section 5a (Figs. 9–12). For instance, in the discussion on Fig. 10, the text has been updated to state that the 88DONLY experiment has weaker reflectivity values than CONTROL, rather than implying that the structure is degraded. Furthermore, in the discussion on Fig. 11 (1–3-km UH), the text has been updated to state that the storm structure is weaker in 88DONLY as the 1–3-km UH field shows the low-level circulation is weaker.

Third Review:

Recommendation: Accept with minor revisions.

General Comment: The authors have done a good job responding to my concerns and I have no further substantial comments. I only have one very minor suggestion included as a comment in the tracked changes version of the document. Otherwise, I think the paper is ready for publication.

[Minor comments omitted…]