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.

Electronic Journal of
SEVERE STORMS METEOROLOGY

Investigating the Impacts of Assimilating Surface Observations on High-Resolution Forecasts of the 15 May 2013 Tornado Event

LEE B. CARLAW
NOAA/NWS, Weather Forecast Office, Tucson, Arizona

JERALD A. BROTZGE
New York State Mesonet, University at Albany, State University of New York, Albany, New York

FREDERICK H. CARR
School of Meteorology, University of Oklahoma, and Center for Analysis and Prediction of Storms, Norman, Oklahoma

(Submitted 17 March 2015; in final form 20 October 2015)

ABSTRACT

In this study, the Advanced Regional Prediction System (ARPS), and its associated three-dimensional variational analysis (3DVAR) package are used to simulate a tornadic supercell at 400-m grid spacing. This storm produced an EF3 tornado in Johnson County, TX during the evening of 15 March 2013. Data from Doppler radar, satellite, aircraft, radiosondes, profilers, and surface observations are assimilated in this work. We show that the assimilation of non-conventional surface observations from three networks: the Citizen Weather Observer Program (CWOP), Global Science and Technology (GST) and Automated Weather Stations (AWS) operated by EarthNetworks, in and around the storm inflow, were fundamentally important to the development of an intense low-level mesocyclone. Simulations that did not incorporate this non-conventional data either developed a weak mesocyclone that was displaced to the east of the actual tornado track, or were unable to develop a defined mesocyclone in the first place. In particular, the assimilation of thermodynamic variables (temperature, pressure and moisture) from a subset of AWS observations near the storm's inflow environment leads to the most accurate simulation of the low-level mesocyclone.

1. Introduction

Evaluating the impacts of observations on atmospheric analyses and numerical forecasts has been an active and expanding topic of research since the 1960s. An understanding of the relative influence of individual observing platforms is important in guiding cost-benefit decisions on the optimal mix of future observations. In two recent reports, the National Research Council (NRC) notes that the current state of observation networks in the United States is inadequate to satisfy the needs of producing high-resolution analyses and forecasts of high-impact weather events (NRC 2009, 2012).

In these reports, the NRC recommends the development of a distributed adaptive "network of networks", in which observations from a multitude of providers are integrated into a coordinated, national observing infrastructure. In this way, the present chaotic nature of the mesoscale observing networks—brought on by the varying sensor qualities and siting specifications—is reduced as comprehensive metadata describing the sensor characteristics of each data source are centralized and improved. An additional critical recommendation by the NRC is for the development of research testbeds,

Corresponding author address: Lee Carlaw, NOAA/National Weather Service, 520 North Park Avenue Suite 304, Tucson, AZ 85719 E-mail: lee.carlaw@noaa.gov

in which the relative impact of observing networks are determined through extensive research and testing.

Historically, observation impact has been quantified either through the use of observing system simulation experiments (OSSEs) or observing system experiments (OSEs). In OSSEs, impacts from potential future observation systems are anticipated by running simulations using synthetic observations, which are then compared to a model "truth" run (e.g., Lord et al. 1997; Atlas 1997). In OSEs, real observations from a particular observing system are denied from a control simulation in which all available observation types are assimilated and then compared to the control run. The OSE design has the benefit of using real data at the cost of 1) being more computationally expensive since a new experiment must be run for each observation type, and 2) having errors in the control run.

As computing power has increased, OSE-related research likewise has expanded. Initially, OSEs tested the impact of incorporating satellite and upper-air data on a model's portrayal of synoptic-scale weather systems (e.g. Gilchrist 1982; Graham et al. 2000; Zapotocny et al. 2002, 2005, 2007). Subsequent studies determined the impacts of assimilating GPS-derived precipitable water observations, mesonet observations, and profiler data on meso-α-scale features (Benjamin et al. 2004, 2010). More recently, research has tested the effects of including radar radial winds and reflectivity on high-resolution forecasts of deep moist convection (Schenkman et al. 2011a, 2011b). In this study, we attempt to quantify the impacts of assimilating surface observations on high-resolution simulations of a supercell on 15–16 May 2013, which spawned an EF3 tornado southwest of the Dallas-Fort Worth Metroplex.

The inclusion of reflectivity and radial velocity data from Doppler radars has been shown to be very important to high-resolution numerical prediction of thunderstorms because radars provide data at the spatiotemporal scale needed to sample convective storms (Hu et al. 2006a, 2006b; Dawson and Xue 2006; Schenkman et al. 2011a). Schenkman et al. (2011b) found that the addition of radial winds from four low-scanning Collaborative Adaptive Sensing of the Atmosphere (CASA; McLaughlin et al. 2009) X-band radars was critical to accurately predict the evolution of the low-level wind field and gust fronts associated with a large mesoscale convective system. Those authors also found modest positive forecast impact with the inclusion of 5-min Oklahoma Mesonet data owing to improved sampling of both the low-level wind field and thermodynamic variables, but these improvements were not as important as the assimilation of the CASA radial winds.

The positive results of radar-assimilation research have substantial implications for the new Warn-on-Forecast initiative (Stensrud et al. 2009, 2013), in which warnings of severe convective storms (and in particular, tornadoes) moves from a warn-on-detection basis to one based increasingly on high-resolution numerical simulations. Obviously, this vision cannot be realized until explicit and accurate forecasts of tornadic thunderstorms become commonplace. Accurate initial conditions, which depend on the optimal assimilation of available observations, are a crucial part of realizing this vision.

The present study is unique in that we assimilate surface observations from several non-conventional surface observations such as the EarthNetworks WeatherBug network, the Citizen Weather Observer Program (CWOP 2014), and Global Science and Technology (GST) (see section 4a), in addition to standard data from METARs, upper-air rawinsonde observations (RAOBs), satellites, aircraft, and operational radars. To the best of our knowledge, this is the first study to assess the impacts of assimilating these non-conventional observations on high-resolution simulations of a supercell. We use the Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001, 2003) as the numerical prediction model and its associated three-dimensional variational analysis (3DVAR; Gao et al. 2004) package for data assimilation. Our goal is to quantify the impact of these surface-observing systems for the accurate analysis and prediction of high-impact weather.

Section 2 provides a brief synoptic and mesoscale overview of the 15 May 2013 north-central Texas tornado outbreak. Section 3 provides details on the ARPS forecast model, data assimilation system, and complex cloud-analysis package. Section 4 outlines the numerical simulation methodology and data-denial experiments. Section 5 provides a discussion of analysis and experiment results. Finally, summary and concluding remarks are presented in section 6.

2. Case overview

a. Event details

A localized tornado and severe weather event occurred during the late afternoon and evening of 15 May 2013 across portions of north-central Texas. A total of twenty tornadoes were reported. The first significant tornado—an EF4—impacted Granbury, TX in Hood County. The second significant tornado to develop, which is the focus of this study, caused EF3 damage in Johnson County near Cleburne, Texas. This tornado caused $124 million in property damage, seven injuries, and was notable for its erratic movement and large size. In particular, after an initial motion to the east, the tornado moved to the southeast before finally making a hard left turn towards Lake Pat Cleburne before dissipating just outside the city limits. At its largest, the Cleburne tornado approached a mile (1.6 km) in diameter. Figure 1 shows the tornado path and EF-rating from the Dallas-Fort Worth National Weather Service (NWS) damage survey.

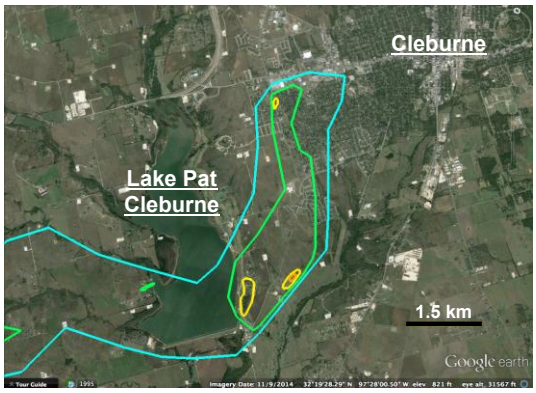

Figure 1: The Cleburne EF3 tornado track as surveyed by the National Weather Service in Dallas-Fort Worth, TX. The cyan, green, yellow, and orange outlines correspond to EF0, EF1, EF2 and EF3 damage respectively (National Weather Service 2015).

b. Meteorological setup and discussion

During the overnight hours on 14 May 2013, a low-amplitude 500-hPa shortwave trough moved northeast out of the Texas Big Bend region. By the morning of 15 May, an upper-level low was located near Wichita Falls, TX with a deepening surface low in the Texas Panhandle (Fig. 2a). Throughout the morning and early afternoon, clouds and precipitation associated with vorticity maxima progressing around the base of the upper-level trough moved east towards the Dallas-Fort Worth metropolitan region. By mid-afternoon, skies had cleared across much of western and central Texas leading to the development of an axis of increased buoyancy under cooler midtropospheric temperatures, with surface-based (SB) CAPE values increasing to near 3000 J kg^{-1} (Fig. 2b).

Between 2000 and 2100 UTC, visible satellite imagery (not shown) revealed a developing cumulus field in the vicinity of strong surface confluence and a diabatic-heating axis between Wichita Falls and San Angelo, TX. Around 2300 UTC, thunderstorms initiated from this cumulus field following the erosion of the morning capping inversion. Increasingly backed surface flow to the east of this cumulus field, along with 20–25 m s^{-1} of south-southwesterly 850-hPa flow (Fig. 2c), produced an environment favorable for supercells, with supercell composite parameter (SCP; Thompson et al. 2003) values increasing to near 20 (Fig. 2d) across the Dallas-Fort Worth Metroplex during the evening hours.

The supercell that spawned the EF3 tornado in Johnson County, TX, developed and intensified slowly between 2230 and 2330 UTC on 15 May, initially near Erath County. By 0132 UTC, gate-to-gate shear in excess of 40 m s^{-1} was visible on KFWS WSR-88D radar (Fig. 3a). At this point, the low-level circulation was moving generally to the southeast. At 0212 UTC, EF0 damage was reported just to the southwest of Cleburne, at which point the tornado began to move to the northeast. The tornado then passed over Lake Pat Cleburne and reached its maximum EF3 rating with an estimated width of one mile. Around 0218 UTC, the tornado made a final hard left turn towards Cleburne as it began dissipating at 0223 UTC (Fig. 3b).

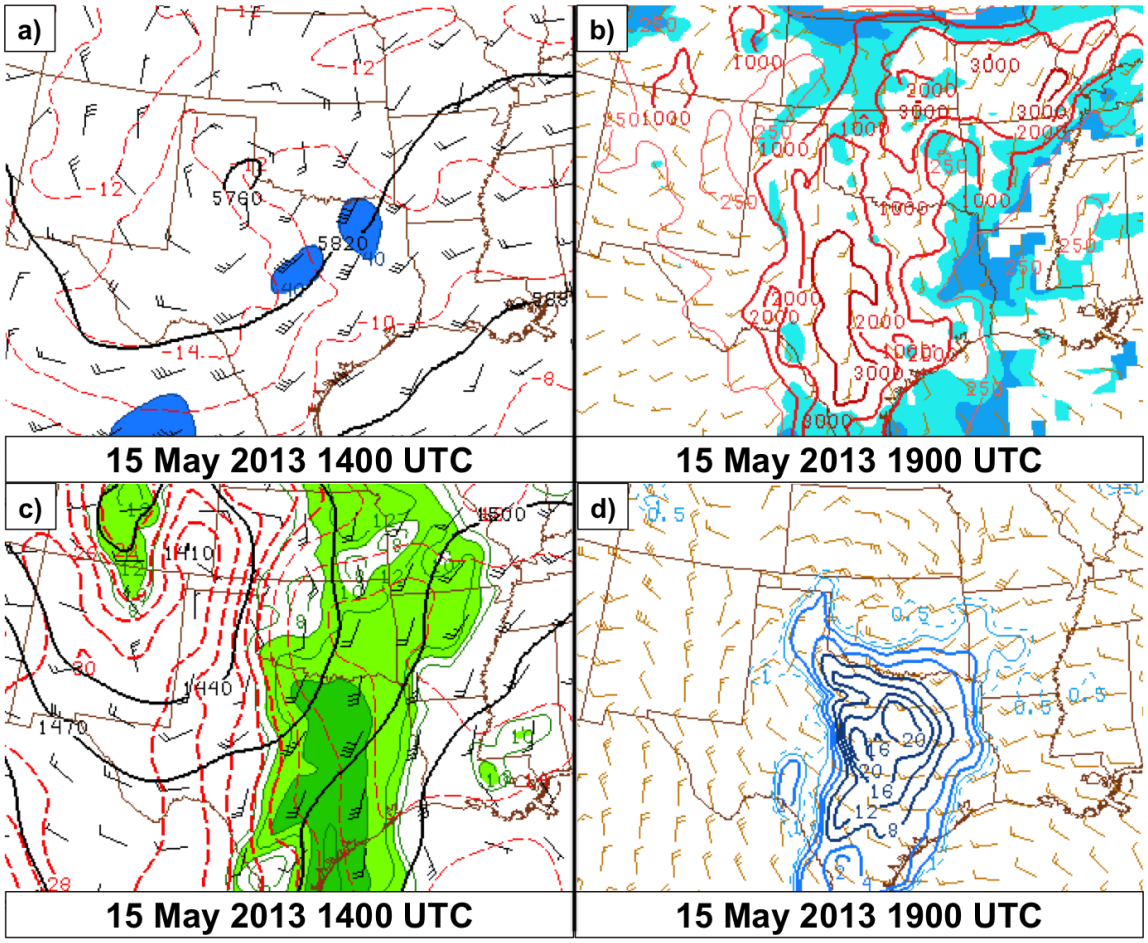

Figure 2: a) 500-hPa geopotential height (black contours, m), temperature (dashed red contours, °C), and wind speeds >40 kt (20 m s^{-1} ,color fill) at 1400 UTC 15 May 2013; b) surface-based CAPE (red contours, J kg^{-1}) and convective inhibition (color fill, J kg^{-1}) at 1900 UTC; c) 850-hPa geopotential heights (solid black contours, m), isotherms (dashed, °C), and dewpoint (color fill, °C) at 0000 UTC 16 May 2013; and d) supercell composite parameter (SCP, Thompson et al. 2003) and storm-motion vectors valid at the same time. These graphics are from the Surface Objective Analysis dataset (Bothwell et al. 2002).

3. Forecast model packages and computational domains

a. Forecast model and parameters

For this study, we use the ARPS, a compressible, non-hydrostatic general-purpose model that has been used previously to simulate events from tropical cyclones (Zhao and Xue 2009), mesoscale convective systems (Dawson and Xue 2006), dryline convection initiation (Liu and Xue 2008), and tornadoes (Schenkman et al. 2014), for example. Here, we employ full model physics, a 1.5-order turbulent kinetic energy-based subgrid-scale closure scheme, fourth-order advection in both the horizontal and vertical, NASA Goddard Space Flight Center longwave radiation parameterization (Chou and Suarez 1994), and Lin et al. (1983) single-moment, three-ice-species microphysics.

Following Snook and Xue (2008) and Schenkman et al. (2011a), we use a reduced value of the rain-intercept parameter, which defines the maximum concentration of raindrops of zero diameter in a Marshall-Palmer distribution (Dawson et al. 2010). Based on visual inspection of simulations, this change results in generally improved forecasts of reflectivity and low-level vertical vorticity fields. Snook and Xue (2008) find that reducing the rain-intercept parameter lead to the production of more large raindrops and fewer small drops.

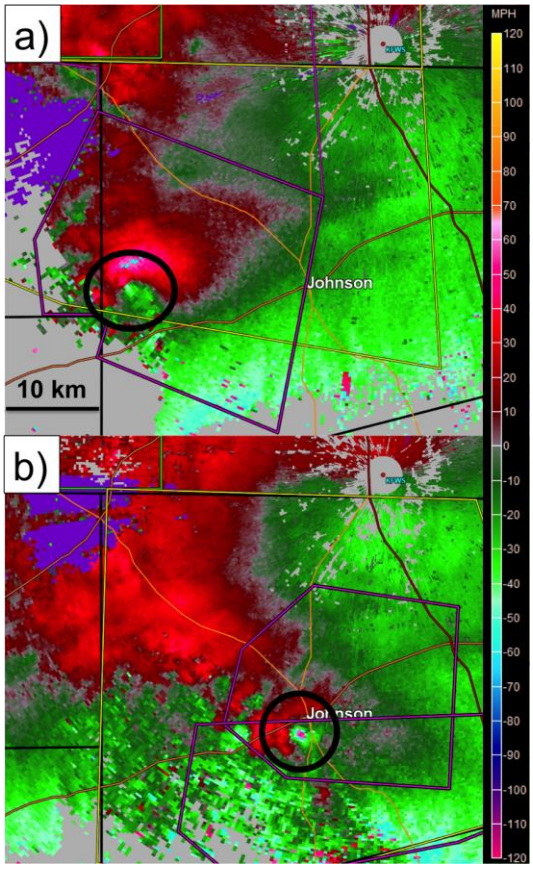

Figure 3: Observed 0.5° radial velocity from the Forth Worth, TX Doppler radar (KFWS, located in the upper right of each panel) valid at a) 0132 UTC 16 May 2013 and b) 0223 UTC. The velocity couplet is circled in black. Severe thunderstorm, tornado and flash flood warnings valid at the indicated times are shown in yellow, magenta, and green outlines, respectively.

This, in turn, results in less evaporational cooling owing to: 1) generally faster terminal fall speeds, and 2) reduced total surface area compared to small droplets containing the same amount of rainwater. The reduction in evaporational cooling leads to the development of weaker low-level cold pools and stronger and more sustained updrafts within supercells. In our experiments, the rain-intercept parameter is set to the standard value of 8.0×10^6 m^{-4} (Lin et al. 1983) on the coarse-resolution nest and reduced to 8.0×10^5 m^{-4} on the high-resolution computational domain (Fig. 4) when reflectivity data from Doppler radars are assimilated (see section 4).

b. Model grids and terrain

We use a set of two one-way nested grids with horizontal grid spacings of 3 km and 400 m, respectively. Both domains use 53 vertical levels, with a vertical grid spacing that stretches from 20 m near the surface to about 400 m at model top (21.1 km AGL) following a hyperbolic tangent function (Xue et al. 1995). The 3-km domain is centered on 32.7 °N, 97.55 °W and is roughly 900 km × 900 km. The 400-m domain is centered on 32.65 °N, 97.2 °W and is roughly 193 km × 161 km. This grid covers the immediate Dallas-Fort Worth metropolitan region, including its southwestern suburbs in Hood and Johnson counties. Terrain elevation for both the 3-km and 400-m grids is supplied by the U.S. Geological Survey 30-s terrain-elevation dataset.

c. The ARPS 3DVAR

The ARPS 3DVAR (Gao et al. 2004) is used as the data assimilation tool in this study. The ARPS 3DVAR minimizes a cost function, written as:

$$J(x) = \frac{1}{2}(x - x_B)^T \mathbf{B}^{-1}(x - x_B) + (H[x] - y_O)^T \mathbf{O}^{-1}(H[x] - y_O) + J_C \quad (1)$$

where the first term on the right-hand side represents the difference between the analysis state vector, x and the background field, x_B weighted by the inverse of the background error covariance matrix, **B**. The second term measures the difference between the model state projected to observation locations and observed values, y_O by the linearized forward operator, H weighted by the inverse of the observation error covariance matrix **O**. Currently, the cross-correlations between variables are not included in the background error matrix and observation errors are assumed to be uncorrelated, resulting in a diagonal **O** matrix. Observation errors used in this study are shown in Table 1. These values were chosen based on several assimilation tests (not shown) that varied the observation to

Table 1: Observation errors assigned to various sensors.

Data Source	u (m s^{-1})	v (m s^{-1})	Pressure (hPa)	Temperature (K)	RH (%)
RAOBs	2–3.5	2–3.5	0.4–0.6	1–2	5–15
Profilers	2–3.5	2–3.5	—	—	—
ASOS	1.5	1.5	0.75	0.8	5
AWOS	1.5	1.5	1.22	0.8	10
ACARS	1.5	1.5	1.22	1.11	10
Mesonets	1.5	1.5	0.75	1.11	5
Non-Conventional	2	2	1	3	10

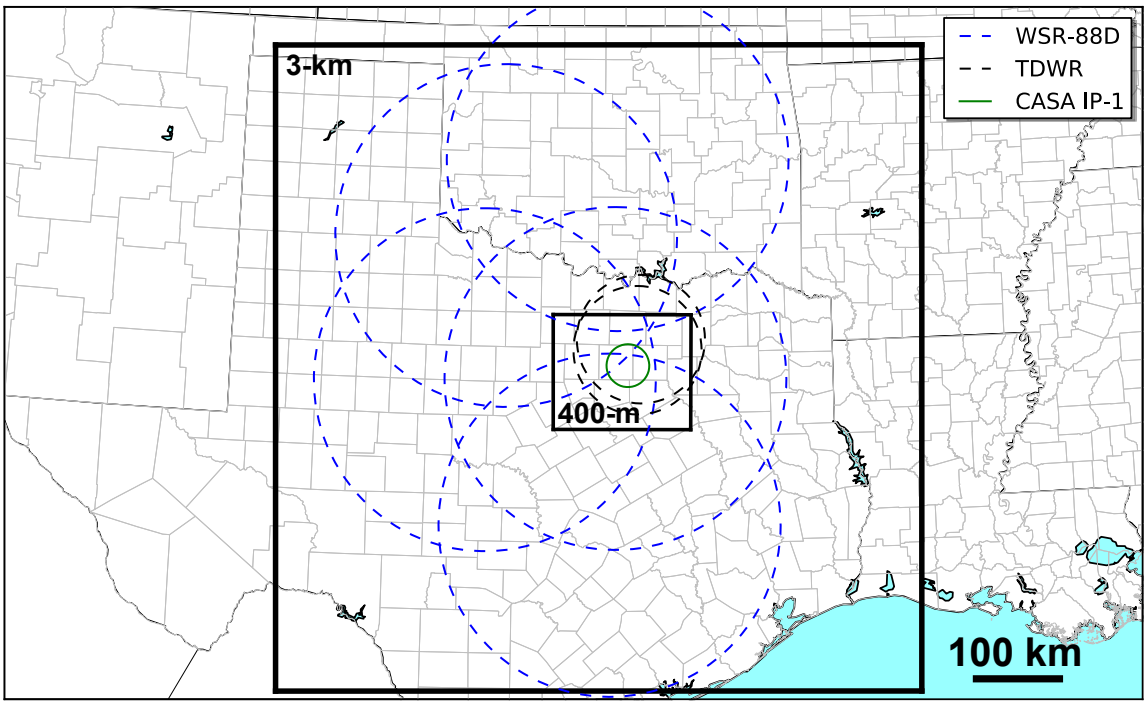

Figure 4: Overview of the coarse 3-km domain and high-resolution, 400-m nest. WSR-88D range rings (dashed blue lines, 240 km), Terminal Doppler Weather Radar (TDWR, dashed black, 90 km), and CASA (solid green, 30 km) are shown as well.

background error ratios. We then qualitatively assessed the smoothness of the analyses to avoid overfitting to the observations and creating analysis bullseyes. Spatial correlations of background error are modeled by repeated applications of a first order recursive filter (Hayden and Purser 1995) in all directions. Multiple passes of this recursive filter result in Gaussian and isotropic background error correlations. The ARPS 3DVAR uses an incremental form of the cost function with preconditioning using the square root of the **B** matrix (Gao et al. 2004).

The final term in Eq. (1), J_C, is a penalty term, which in this case consists of a weak anelastic mass divergence constraint. This term is written as:

$$J_C = \frac{1}{2}\lambda_C^2 D^2 \qquad (2)$$

The variable D is:

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

where ρ is the mean air density at individual model levels, α and β are horizontal and vertical

weighting coefficients, and λ_C determines the overall weight of the J_C penalty term in (2). As discussed by Gao (2004) and Hu et al. (2006b), the effect of this additional term is to apply a weak anelastic mass constraint to the analyzed wind field, which acts to disperse observed radar radial velocities into the cross-beam and vertical directions. In Hu et al. (2006b), the authors note that the true 3D divergence constraint (that is, when $\alpha = \beta = 1$) only works for grids with a vertical-to-horizontal grid aspect ratio near unity. In cases with large aspect ratios near the surface (as with this work), the effect of the full 3D constraint is to force most of the wind-field adjustment into vertical velocity with very little adjustment in the horizontal direction. For this reason, following Hu et al. (2006b), we use a horizontal weighting factor of $\alpha \lambda_C = 1000$ and a reduced vertical weighting factor of $\beta \lambda_C = 100$ in our experiments. Because this additional cost term is a weak constraint, mass divergence is not forced to be zero and, therefore, the analysis is able to develop physically justifiable convergence-divergence dipoles near thunderstorm updrafts, for example.

Some substantial changes to this release of the ARPS 3DVAR over previous versions include: the ability to specify decorrelation lengths rather than grid points when observations are spread in the vertical, general improvements to the processing and analysis of radar reflectivity data, and the addition of a quality-control check to ensure unconventional observations with incorrect surface elevations are screened out of the analysis process. Gao at al. (2004) offer more information about the 3DVAR formulation and the mass divergence constraint.

d. Complex cloud analysis

The ARPS complex cloud analysis incorporates cloud and moisture information from surface observations, satellite (both visible and infrared channels), and reflectivity data from Doppler radars to update model fields of hydrometeors and potential temperature. Because no direct links between reflectivity and temperature exist, this procedure is not done variationally, but as an additional step after the minimization of the 3DVAR cost function.

In our case, reflectivity data are converted into hydrometeor species using the Ferrier equations (Ferrier 1994) for rainwater and Rogers and Yau (1989) for snow and hail. Radar-observed hydrometeor fields are used to replace those present in the background whenever they are available. This is done based on the assumption that the radar-observed precipitation fields are, in general, superior to those present in the model background field. A noted issue with the ARPS 3DVAR procedure—and one that we experience in this work—arises when spurious hydrometeors exist in the background field while radial velocity data do not. In this situation, it is difficult to remove precipitation from the model grid, especially when the environment is unstable (Xue et al. 2013). To an extent, assimilating clear-air velocities from Doppler radar can mitigate this problem, but these data often are not available in the necessary locations.

The final step in the ARPS 3DVAR complex cloud analysis is an in-cloud thermal adjustment, which attempts to account for the effects of latent heat release during the condensation process. In this step, a moist adiabatic ascent profile, subject to dilution by mixing, is calculated from the analyzed cloud base. Where this temperature exceeds that of the background, the temperature is increased. The complex cloud analysis also increases the environmental temperature in regions that have analyzed clouds along with upward vertical velocities. Hu et al. (2006a) and Brewster et al. (2005) further describe the ARPS complex cloud analysis procedure.

e. Observation quality control

An important step in the data-assimilation procedure is the quality control of observations. Observations are quality controlled in a multi-step process during data assimilation. First, differences between observations and background values are computed and measurements that exceed user-defined limits and climatological thresholds are discarded. Next, observations are checked for temporal consistency during the previous hour. Additionally, observations are "buddy-checked" against surrounding values via a Barnes interpolation procedure (Barnes 1964). Owing to the inclusion of data from several unconventional sources, another quality-control step removes surface stations with very

low wind speeds resulting from transmission errors or poor siting (this issue will be discussed at more length in the next section). Stations reporting wind speeds <1 m s^{-1} concurrent with background values >5 m s^{-1} are removed from the assimilation dataset, similar to Horel and Dong (2010). Observations obtained via the Meteorological Assimilation Data Ingest System (MADIS) are also subject to internal quality-control checks, based on procedures outlined in the NWS Techniques Specification Package (NWS 1994).

4. Experiment design and methodology

a. Observational datasets

In this study, we assimilate observations from conventional sources such as upper air soundings (RAOBs), wind profilers, the Aircraft Communications Addressing and Reporting System (ACARS, Moninger et al. 2003), Automated Surface Observing System (ASOS) and Automated Weather Observing System (AWOS), and surface mesonet sites whenever available. In addition to these traditional datasets, we also assimilate observations from several non-conventional sources, which are outlined below.

1) CWOP OBSERVATIONS

Initially started by a group of landowners with access to electronic weather stations, the CWOP since has expanded to include over 10,000 members who voluntarily share their weather information via the Automatic Position Reporting System as a Weather Network (APRSWXNET, from CWOP 2014). These observations, near homes and business, are transmitted to the Meteorological Assimilation Data Ingest System (MADIS) and are subject to the same quality-control checks as other datasets. Information about the quality control can be found in NWS (1994). In this study, we refer to the observations obtained through the APRSWXNET as CWOP stations.

2) AUTOMATED WEATHER STATION OBSERVATIONS

EarthNetworks, which operates the WeatherBug network, has installed weather sensors at over 8000 locations across the country. Initially, stations were installed at public schools to provide real-time weather data for use by broadcast studios. This live weather information also was integrated into school curricula in order for science, technology, engineering, and math, or STEM, instruction. Over the past few years, coverage of the Automated Weather Station (AWS; EarthNetworks 2015) network has increased greatly. The provider has partnered with broadcast studios, private companies and local governmental agencies to provide a dense network of surface observations to research institutions, the NWS and the public.

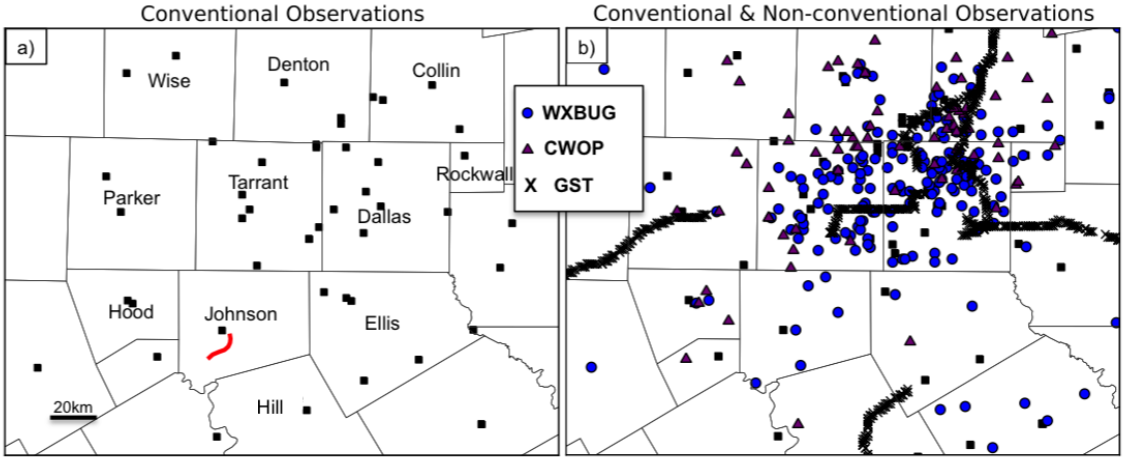

Figure 5: Plot of the locations of a) conventional (black squares) and b) non-conventional surface stations assimilated in our experiments. Conventional, WeatherBug (blue dot), and CWOP (black triangle) locations plotted are from 0100 UTC 16 May 2013 and are generally representative of stations available through the assimilation procedure. Locations of all GST observations assimilated in the 0100–0145 UTC window are shown as X. Selected county names and the Cleburne tornado track (red) appear in (a).

3) GST OBSERVATIONS

In 2009, GST, NOAA and the National Mesonet Program started a Mobile Platform Environmental Data observation network (MoPED; Heppner 2013). This network consists of observing sensors mounted on buses and transportation fleets around the U. S. The GST system does not observe wind data because of contamination issues while vehicles are traveling at highway speeds. In this work, roughly 400–600 GST observations are available during each 5-min assimilation window (see Section 4c). Observations during each 5-min assimilation period within a distance of 2 Δx (800 m) are combined into a single synthetic observation via a Barnes weighting factor (Barnes 1964).

Figure 5 provides a comparison of the conventional and non-conventional surface observations that are assimilated in our experiments. As mentioned previously, because of the siting of non-conventional stations in backyards (CWOP) and on school rooftops (AWS), the windspeed observations from these sources are found to be consistently low-biased. We attempt to alleviate this issue, to an extent, via the previously mentioned low-windspeed quality-control check, as well as inflating non-conventional wind observation errors during the analysis step (Table 1).

b. 3-km forecasts

A single forecast is run on the 3-km grid for the purpose of providing background fields and lateral boundary conditions for the 400-m runs. The initial background field for the coarse-resolution 3-km domain is supplied by the 13-km Rapid Refresh (RAP; Brown et al. 2012) analysis valid at 2100 UTC 15 May 2013. This grid is forced at the boundaries by subsequent RAP forecasts at hourly intervals until 0400 UTC 16 May. Hourly data assimilation cycles are performed using conventional observations including wind profilers, aircraft observations (ACARS), upper-air soundings (RAOBs), and surface sites.

Data from wind profilers and upper-air radiosondes (when available) are analyzed in the first 3DVAR pass using a horizontal decorrelation length (defined as the radius at which the observational weight from the recursive filter is *e*-folded, or reduced by roughly 67%) of 150 km. Surface observations from ASOS, AWOS, mesonet, and aircraft are assimilated in a second pass with a reduced decorrelation length of 50 km, which acts to fill in additional information about the kinematic and thermodynamic fields within the lowest model layers.

Table 2: Details of the 400-m assimilation-forecast experiments.

Experiment	Conventional surface data	Non-conventional surface data	CASA data	WSR-88D data	TDWR data
CTL	All	All	All	All	All
NOSFC	None	None	All	All	All
NONEWSFC	All	None	All	All	All
WXBUGADD	All	WXBUG in Johnson County, TX	All	All	All
WXBUGWIND	All	WXBUG wind in Johnson County, TX	All	All	All
WXBUGTHERMO	All	WXBUG T, T_d, P variables in Johnson County, TX	All	All	All
AMATREDUCE	All	All, but reduced observation error	All	All	All

Table 3: 3DVAR analysis-pass configuration.

Analysis Pass	Avg. station spacing (d, km)	Horizontal decorrelation length (km)	Vertical decorrelation length (km)	Incorporated Data
1	50 +	100	0.8	RAOBS, wind profilers
2	30	50	0.5	ASOS, AWOS, Mesonet, ACARS
3	10	10	0.5	AWS, GST, CWOP
4	0.4	0.8	0.3	Radar radial velocity

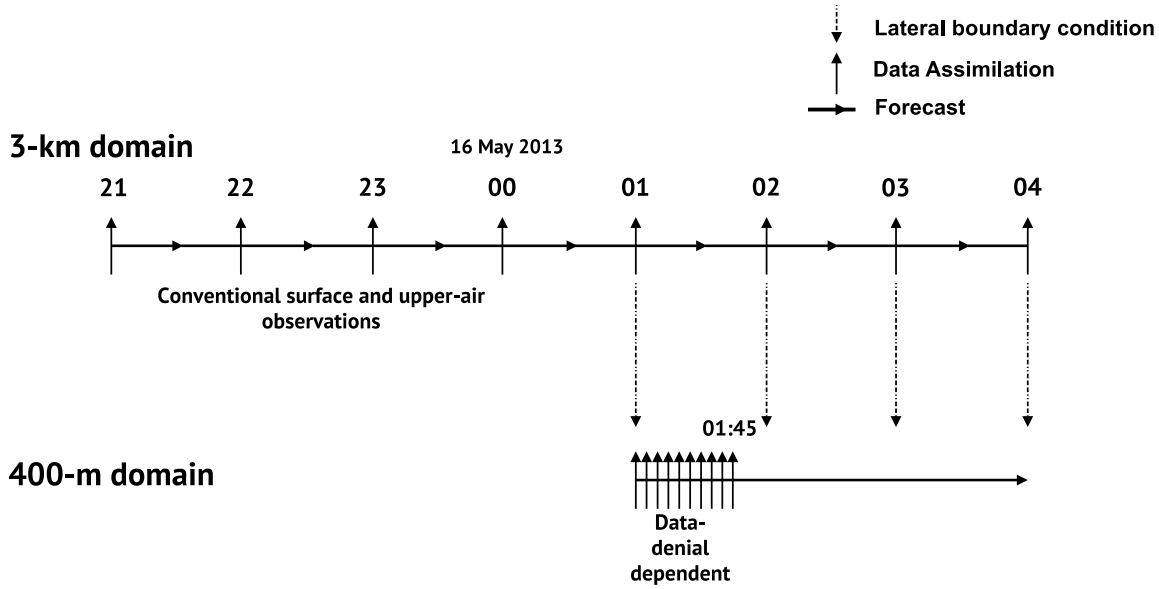

Figure 6: Schematic representation of assimilation cycles for the 3-km and 400-m domains. For clarity, only boundary-condition interpolations at the top of each hour are shown. In reality, the 3-km domain supplies interpolated boundary conditions to the 400-m grid at 5-min intervals.

c. 400-m forecasts

The 3-km forecast valid at 0100 UTC 16 May 2013 is interpolated to the 400-m domain to provide background fields for our high-resolution data-denial experiments. Each experiment differs by the set of observations that are incorporated during a 45-min intermittent assimilation period using 5-min cycling intervals between 0100 and 0145 UTC. Forecasts begin at 0145 UTC and end at 0400 UTC, and are forced every 5 min by interpolated boundary conditions from the 3-km nest. Figure 6 shows a schematic of the general assimilation procedures for both the 3-km and 400-m grids. An example namelist file used to run the 400-m simulations, which has been provided as supplemental data, can be found here.

Several additional 400-m simulations were also launched after a 0000 to 0045 UTC assimilation period. This was an attempt to capture the evolution of an earlier EF4 tornado in Granbury, TX, spawned by the same thunderstorm that produced the Cleburne EF3 tornado. However, these simulations were not able to produce a persistent supercell, possibly because of the lack of surface observations west of Hood County, near the thunderstorm and during this assimilation window (see Fig. 4).

Table 2 provides an overview of the experiments run in this study. In the control experiment (CTL), all available conventional and non-conventional observations are assimilated along with radial velocity and reflectivity data from eight Doppler radars (Fig. 3). Because we assimilate observations from different sources, we use multiple passes of the ARPS 3DVAR with decreasing decorrelation scales to account for variations in observation spacing. Conventional upper-air and surface observations are assimilated in the first two 3DVAR passes with decorrelation scales of 100 km and 50 km, respectively, while non-conventional surface stations are incorporated in the third pass with a reduced decorrelation scale of 10 km to account for increased station density. Doppler-radar radial velocities (V_r) are analyzed in the fourth pass with the smallest decorrelation length of 0.8 km. Table 3 summarizes the 3DVAR analysis strategy employed on the 400-m grid.

Experiment AMATREDUCE is the same as CTL, except the observation errors in the **O**-matrix in (1) of the non-conventional surface stations are artificially reduced to match those of the ASOS stations. Experiments NONEWSFC and NOSFC deny non-conventional surface stations and all surface stations, respectively, from the analysis cycling procedure. Experiment

WXBUGADD is identical to NONEWSFC, except observations from the WeatherBugnetwork within Johnson County, TX (near the Cleburne supercell) are included in the analysis procedure. Results from the WXBUGWIND and WXBUGTHERMO are identical to WXBUGADD except that they

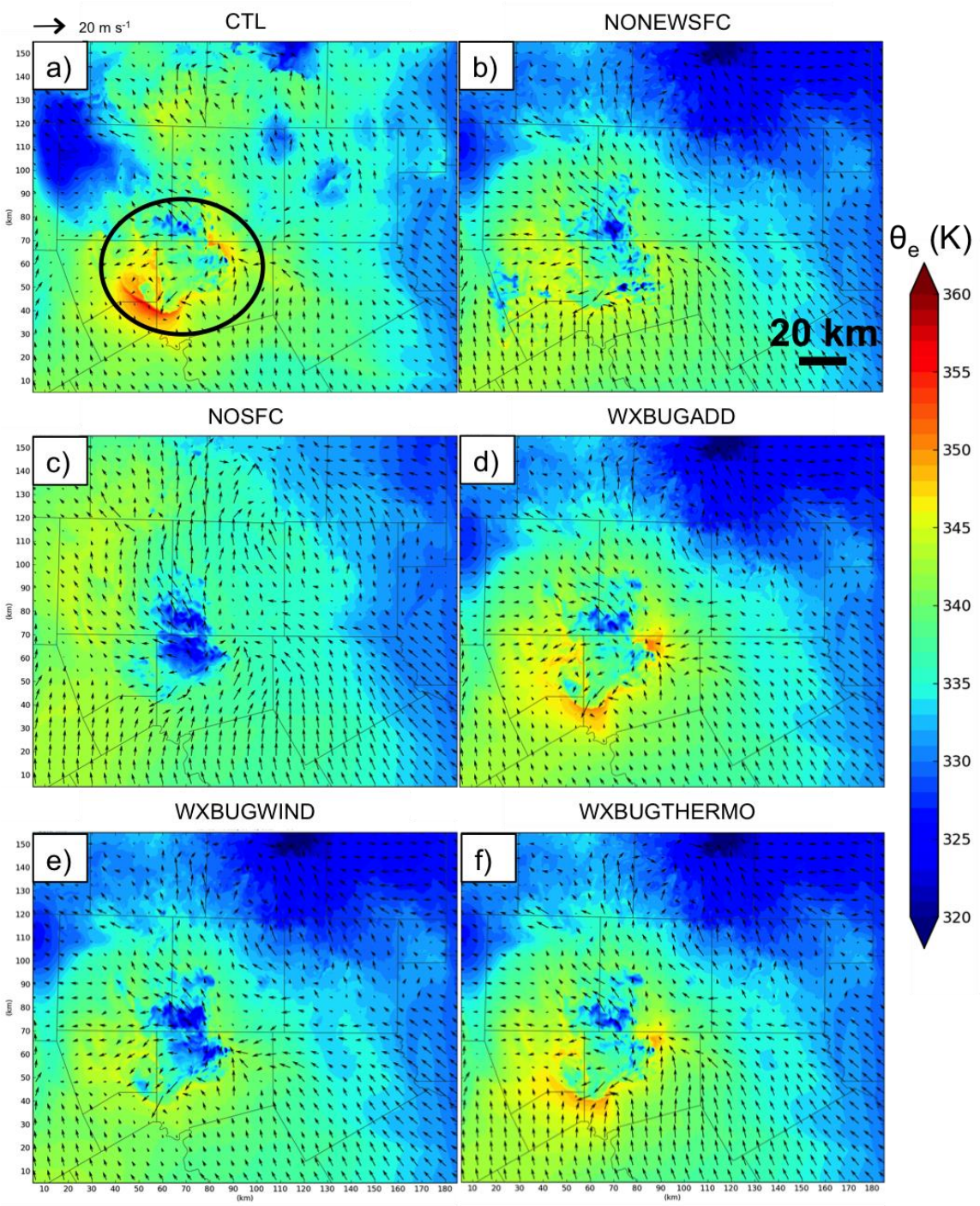

Figure 7: Plots of 0145 UTC 15 May 2013 10-m θ_e (shaded) and wind (vectors) analyses for experiments a) CTL, b) NONEWSFC, c) NOSFC, d) WXBUGADD, e) WXBUGWIND, and f) WXBUGTHERMO. For reference, the supercell is circled in (a).

WXBUGADD experiment encouraged two additional tests to determine the impact of the WeatherBug kinematic and thermodynamic variables on storm evolution. Experiments incorporate only kinematic and thermodynamic observations, respectively, from the Johnson County WeatherBug stations.

An additional data-denial experiment assimilating the single CASA radar located in Arlington, TX was run, but because its unambiguous range covered none of the Cleburne storm, little impact was noted on the subsequent forecast evolution. Furthermore, this paper primarily attempts to address the forecast impact from various surface observation systems, and therefore radar-data denials will be investigated in future work.

5. Results

In this section, we present results of the data denial and sensitivity experiments with particular emphasis on the Cleburne supercell. The first section focuses on differences in model fields at the end of the assimilation window (valid at 0145 UTC), followed by a discussion of the effects of these differences on the evolution of features in the 400-m forecasts. A forecast-verification comparison is discussed at the end of this section.

a. Analysis comparisons and verification

Qualitatively, the greatest differences among the data-denial experiments can be seen in the low-level moisture fields. A comparison of 10-m equivalent potential temperature (θ_e) and winds valid at 0145 UTC is shown in Fig. 7. In CTL, a region of relatively higher θ_e (with values around 360 K) is noted across portions of eastern Hood, northeastern Somervell, and extreme southwestern Johnson counties (Fig. 7a). NONEWSFC fails to depict a similar θ_e pattern, while NOSFC similarly fails to depict values above 340 K in the same vicinity (Figs. 7b,c). The inclusion of WXBUG stations in Johnson County in WXBUGADD and WXBUGTHERMO (Figs. 7d,f) results in increased 10-m θ_e values, which are more consistent with those seen in CTL.

This result is consistent with the surface observation distribution in the southwest portion of our domain. Figure 5 shows a low density of conventional observing systems in the Johnson-Hood county area, with a mean station separation of roughly 40–50 km. Moreover, the only conventional surface stations that exist in this vicinity are the generally less-accurate AWOS sites, meaning these observations have lower weights in the ARPS 3DVAR scheme. The inclusion of non-conventional stations in CTL and WXBUGADD greatly improves the density and spatial distribution of surface observations, which permits greater detail and accuracy in the third 3DVAR analysis pass.

Considerable differences in the 10-m θ_e field are also noted in northwestern Parker County, where CTL analyzes a local minimum around 320 K associated with earlier thunderstorm outflow. The denial of all surface stations in NOSFC results in values that are roughly 20–25 K warmer. NONEWSFC is also warmer than CTL, but not to the level of NOSFC. Experiments WXBUGADD, WXBUGTHERMO and WXBUGWIND closely resemble NONEWSFC as expected, since no additional data are assimilated in this region.

Figure 8 displays θ_e biases (observed value – analysis value) valid at the final analysis time of 0145 UTC 16 May, using data from ten ASOS sites. The data from these stations are not strictly independent due to the fact that we assimilate these observations in the first cycle at 0100 UTC. These ten sites are included in the first analysis pass and used as verification points because we consider them to be the most accurate sensors in the domain, as well as the most representative of the local environment given their siting characteristics.

Consistent with the results described above, the θ_e bias values in Parker County at KMWL are considerably reduced between CTL (2.3 °C) and both NONEWSFC (–3.5 °C) and NOSFC (–14.1 °C). These results indicate that the decrease in θ_e values in northwestern Parker County in the CTL experiment more closely resemble observed values. Similarly, biases at KDTO, located in Denton County, are improved from 11.6 °C in NONEWSFC to –0.9 °C in the CTL simulation. In general, the greatest improvements in equivalent potential temperature bias values in CTL occur in locations that are data-sparse with respect to conventional surface stations (i.e. at KMWL, KDTO, KTKI, KTRL, and KCRS), which implies a benefit of assimilating non-conventional observations.

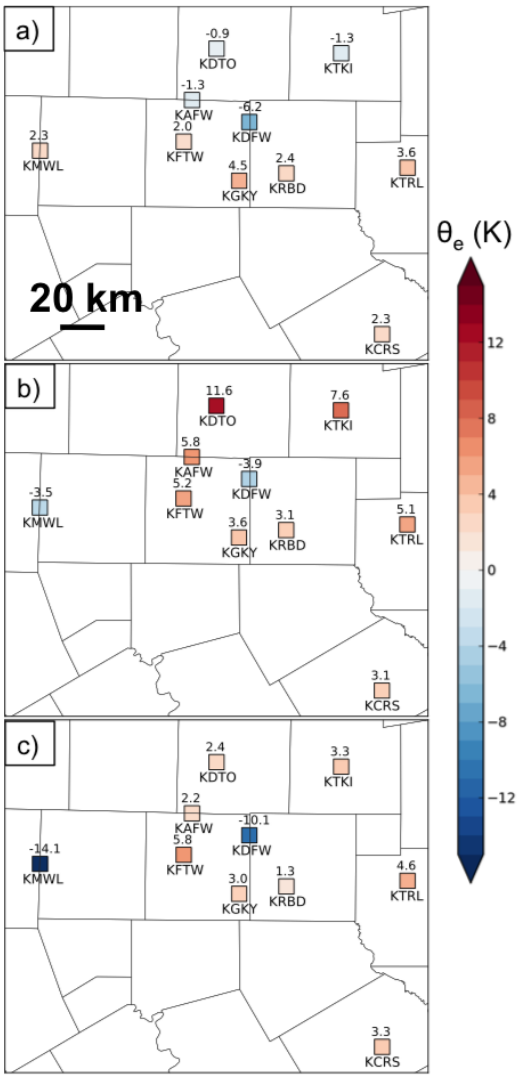

Figure 8: Bias values (observed – analysis, K) of surface level equivalent potential temperature for a) CTL, b) NONEWSFC, and c) NOSFC at final analysis time.

Table 4 provides root-mean-square-difference (*rmsd*), *bias*, and mean absolute errors (*mae*), of 10-m θ_e computed at all 10 verifying ASOS stations. In terms of *rmsd*, NOSFC verifies the worst of all seven experiments, with a similarly large *mae*. The NOSFC *bias* is close to zero owing to the presence of large offsetting negative and positive θ_e biases (Fig. 8). The addition of conventional surface stations in NONEWSFC results in only slight improvements to the *rmsd*. However, the assimilation of non-conventional surface observations in CTL substantially improves the *rmsd*, *bias*, and *mae* values, with the *rmsd* more than halved over NOSFC. Experiment AMATREDUCE results in a marginally worse final θ_e analysis verification with respect to CTL, even though the same observations were used, owing to the specification of higher accuracy than warranted to the non-conventional stations.

Table 4: θ_e root mean square error, bias, and mean absolute error calculated at ten ASOS stations at 0145 UTC 16 May 2013 for various experiments.

Experiment	RMSD (°C)	BIAS (°C)	MAE (°C)
CTL	3.1	0.7	2.7
NONEWSFC	5.8	3.8	5.3
NOSFC	6.3	0.2	5.0
WXBUGWIND	5.9	3.8	5.3
WXBUGTHERMO	5.6	3.1	4.9
WXBUGADD	5.5	2.8	4.7
AMATREDUCE	3.6	0.9	3.3

Verification against five ASOS stations in Dallas and Tarrant Counties alone reveals that non-conventional observations do not have quite as large an effect compared to the domain-wide statistics discussed above (not shown). These results are consistent with the findings of Tyndall and Horel (2013) in which the authors noted the impact of mesonet observations in major metropolitan centers were reduced due to the large number of additional stations available to correct the background field. In contrast, mesonet observations were seen to have a higher impact in data-sparse regions, where the background provided most of the information about the current atmospheric state.

b. Storm-evolution forecast comparisons

1) TORNADIC CIRCULATION STRUCTURE AND EVOLUTION

In this subsection we discuss forecast differences among the experiments outlined in section 4c with respect to the forecast evolution of the Cleburne tornadic supercell.

The aforementioned differences in the low-level thermodynamic fields near Johnson County lead to differences in the evolution of the Cleburne supercell. Given the high-resolution of the model domain, these differences are most substantial during the first hour of the 400-m simulation (between 0145 and 0245 UTC)—the primary focus of our discussion.

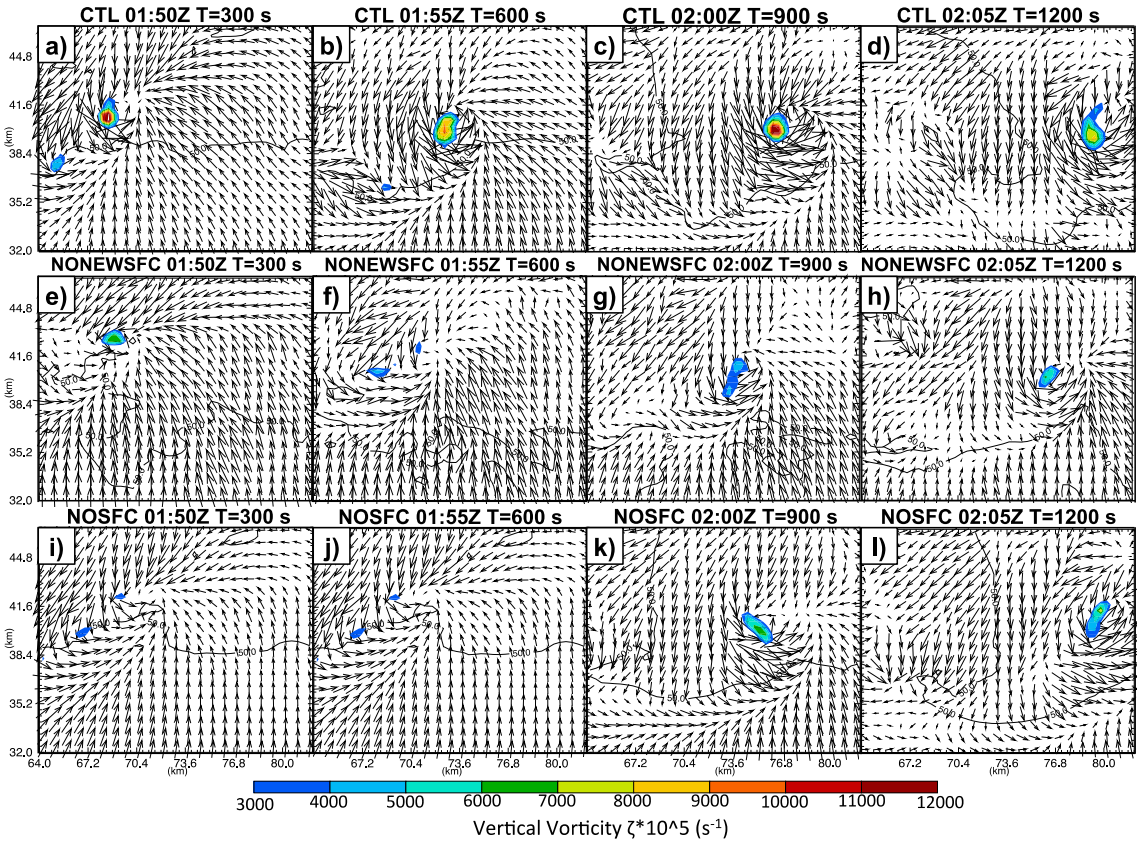

Figure 9: 10-m vertical vorticity (color fill, s^{-1}) and 50 dBZ reflectivity contour. Plots from left to right show forecasts at 300, 600, 900, and 1200 seconds. Experiment CTL is shown in a–d, NONEWSFC in e–h, and NOSFC in i–l. Click image for enlargement.

Before discussing the results of the 400-m forecasts, we briefly mention the reasoning behind calling the intense low-level circulations that develop in the various model runs *tornado-like vortices* (TLV)[1]. Because of the relatively course 400-m simulation resolution compared to the width of a typical tornado, we cannot resolve an actual tornadic circulation accurately.

Within the first 5 min of experiment CTL, a persistent and well-defined low-level TLV develops in southwestern Johnson County with peak 10-m vertical vorticity (ζ) exceeding 0.1 s^{-1} (Fig. 9a). Ten minutes later, at 0200 UTC, the low-level circulation reaches its maximum intensity with surface wind speeds approaching 55 m s^{-1} (Fig. 9c) (EF2-equivalent intensity). At this time, a secondary rear flank downdraft (RFD) surge forms and approaches the low-level circulation from the west. Within the next 5 min, this RFD surge overtakes the TLV, leading to rapid weakening of the low-level ζ field (Fig. 9d).

In the NONEWSFC experiment, the initial 10-m vortices that were carried through the assimilation cycle dissipate during the first 10 min (Fig. 9e,f). In contrast to the rapid intensification of the 10-m circulation noted in CTL, the NONEWSFC simulation does not develop any concentrated regions of 10-m ζ until after 0210 UTC (Fig. 9h), at which point a weak TLV develops in southeastern Johnson County before quickly dissipating roughly 5 min later (not shown).

[1]Following Weisman and Trapp (2003) and Schenkman et al. (2011a) we classify a TLV as an intense region of 10-m vertical vorticity (ζ) that exceeds 0.05 s^{-1} and persists for >15 min. We have doubled the value of ζ used in Schenkman et al. (2011a) since intense mesovortices not associated with the central circulation often exceed the 0.025 s^{-1} value used in that paper.

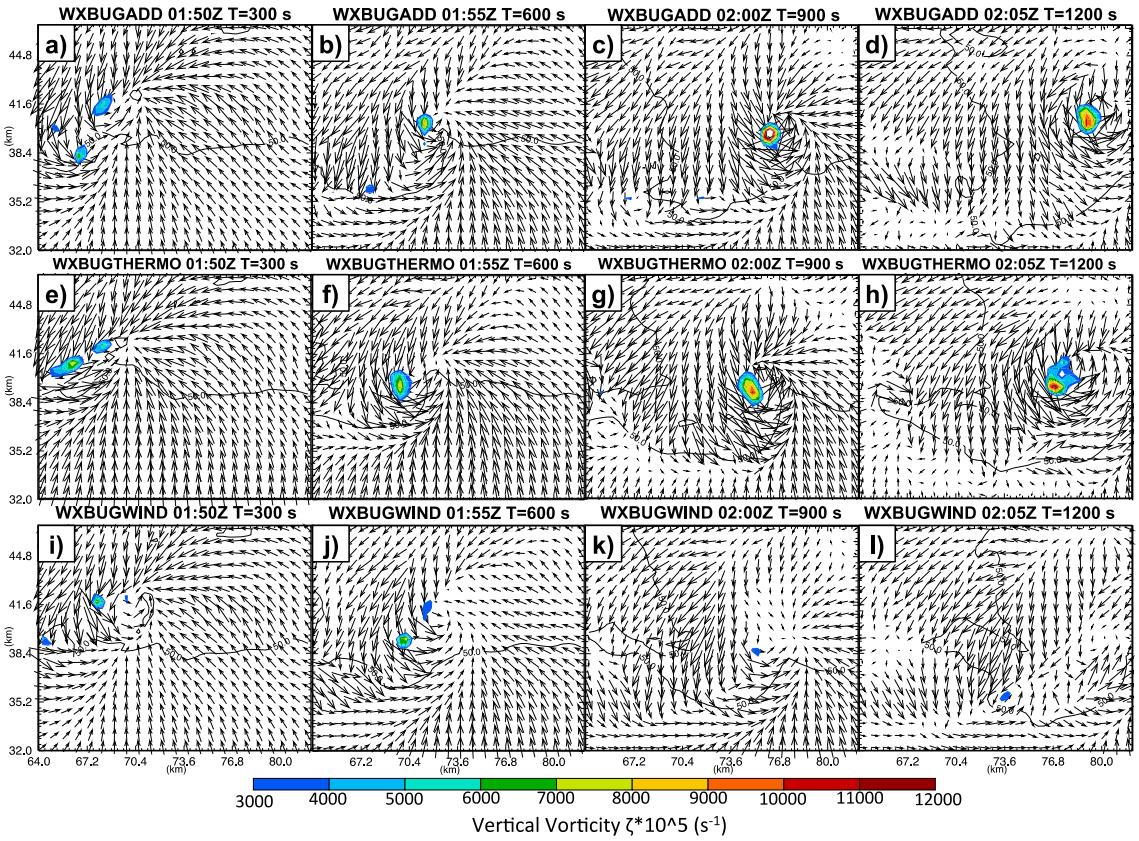

Figure 10: As in Fig. 9 but for (a–d) WXBUGADD, (e–h) WXBUGTHERMO, and (i–l) WXBUGWIND. Click image for enlargement..

Experiment NOSFC initially develops a weak 10-m circulation with ζ generally <0.05 s^{-1} by 0155 UTC (Fig. 9j), eventually wrapping northeastward into the center of the storm while overtaken by the RFD between 0205 and 0210 UTC (Fig. 9l). Qualitatively, the NOSFC simulation closely resembles that of NONEWSFC through the first 20 min. The 10-m circulation dissipates by 0215 UTC.

Experiments WXBUGADD (all conventional data and WeatherBug observations in Johnson County) and WXBUGTHERMO (same as WXBUGADD but only thermodynamic variables from WeatherBug observations) also develop an intense TLV with very similar timing to CTL. (Fig. 10b). Both forecasts quickly reorganize existing low-level vorticity into a well-defined circulation within the first 10 min as the storm develops and moves east. By 900–1200 seconds (0200–0205 UTC), ζ of the 10-m circulation exceeds 0.15 s^{-1} (Fig. 10c,d). Following the other experiments, the TLVs in both WXBUGADD and WXBUGTHERMO occlude around 0210 UTC and dissipate shortly thereafter (not shown).

WXBUGWIND initially develops a weak low-level circulation by 0150 UTC that appears to be embedded within the main RFD (Fig. 10i). Over the next 5 min or so, this circulation center moves quickly to the south-southeast with the RFD surge and dissipates soon thereafter (Fig. 10j,k). AMATREDUCE, like WXBUGWIND, is unable to produce a well-developed TLV (not shown).

Figure 11 shows 10-m isotach fields (color-filled above 29 m s^{-1}) for each of the experiments with radar-observed vorticity tracks from NSSL displayed in (a). These swaths show maximum 10-m wind speeds in each grid box at 30-s intervals. In order to reduce the effects of contamination from strong RFD winds, surface winds were plotted only if they were within three

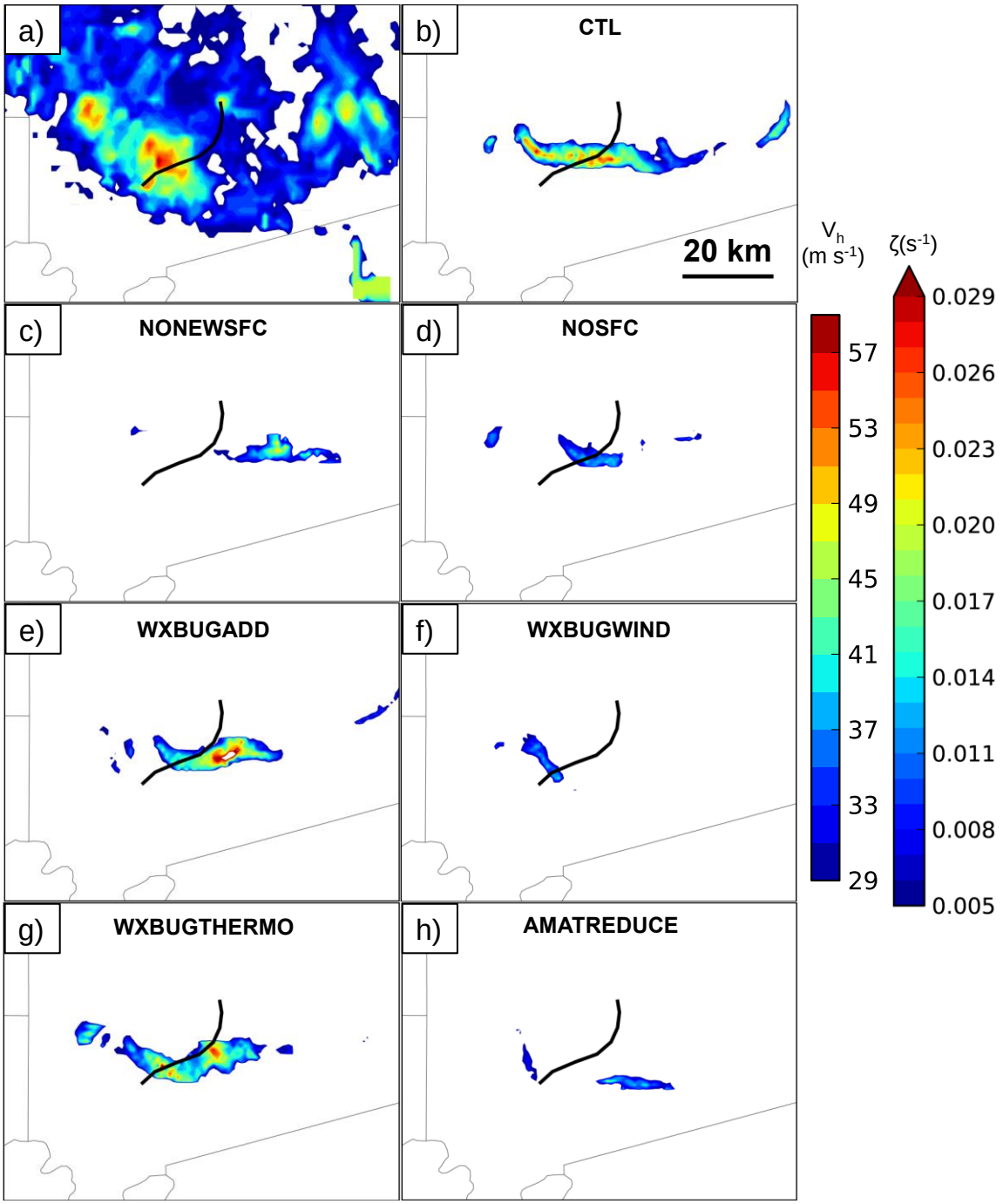

Figure 11: Maximum 10-m wind speeds (m s^{-1}) associated with 10-m AGL ζ >0.05 s^{-1} (b–h, V_h). Only winds >29 m s^{-1} are plotted. NSSL radar observed rotation tracks (s^{-1}) are plotted in (a, ζ). The black overlaid line is an approximation of the NWS-surveyed tornado track.

gridpoints of surface ζ >0.05 s^{-1}. An approximation of the Cleburne tornado track from radar imagery is overlaid on each plot as a black line. Consistent with the circulation strengths previously discussed, experiments CTL, WXBUGADD, and WXBUGTHERMO all produce weak to EF2-equivalent wind speeds (associated with high surface ζ values) across southwestern and south-central Johnson county, while NOSFC, WXBUGWIND, and AMATREDUCE forecast substantially reduced surface winds. NONEWSFC develops

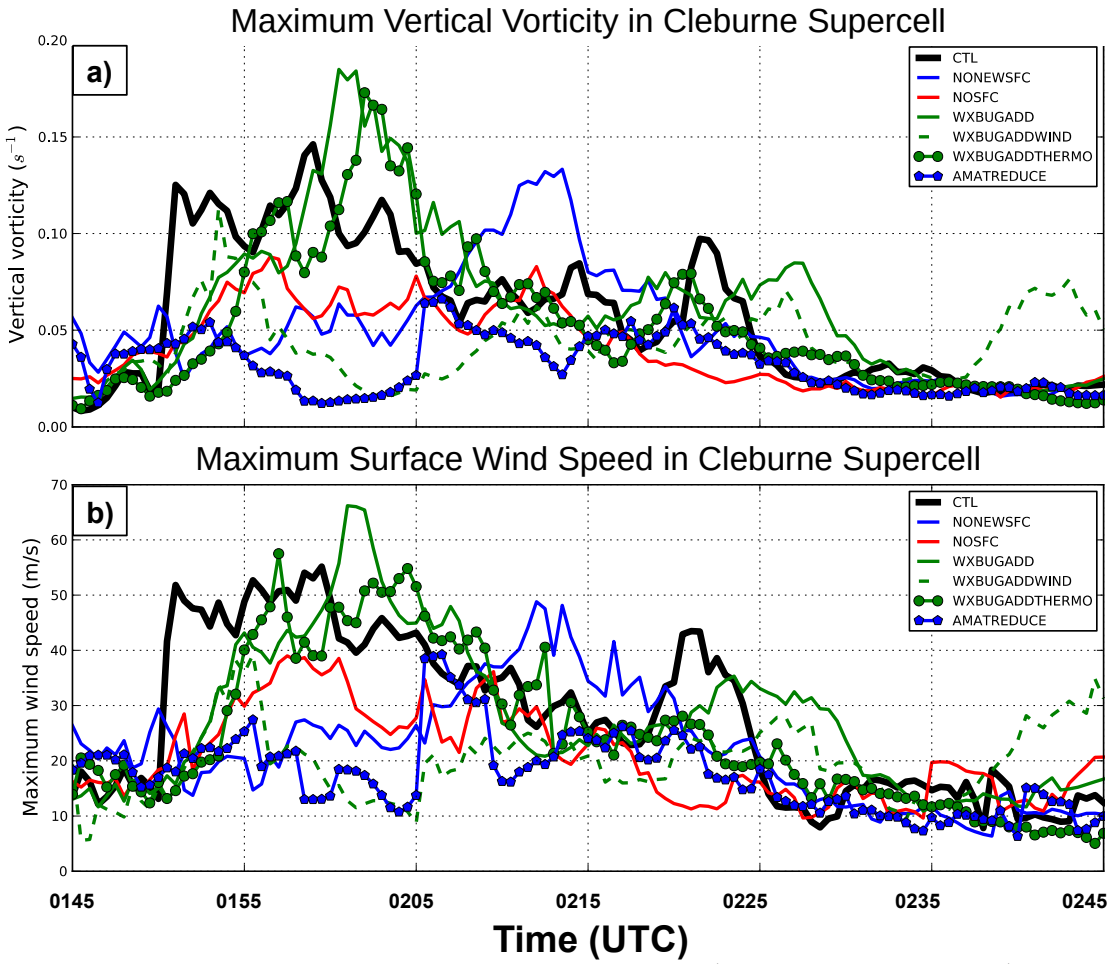

Figure 12: Time series plots of maximum 10-m values of: a) ζ (s^{-1}) and b) wind speeds (m s^{-1}).

weak EF1-strength wind speeds roughly 10 km east of the actual tornado track, indicating a delay in the development of the low-level circulation.

Figure 12 shows time series plots of maximum near-surface ζ and wind speeds computed in the same manner as in Fig. 11. CTL quickly develops high values of ζ and surface winds. Both WXBUGADD and WXBUGTHERMO produce the largest ζ in roughly the same time frame. WXBUGWIND, as mentioned above, initially develops a low-level circulation that quickly dissipates within the RFD. NONEWSFC develops the highest wind speeds 15–20 min later than CTL, which is actually closer in time to the observed tornado. However, this delay leads to a roughly 10-km eastward displacement in the NONEWSFC wind swath (cf. Fig. 11).

The TLVs and associated high wind speeds only last about half as long as the observed tornado. This is possibly a result of both the relatively coarse 400-m grid spacing compared to the width of the actual tornado, and the use of single-moment microphysics which has been shown to create large and intense cold pools (Dawson et al. 2010). Even with the reduced rain-intercept parameter discussed in 3a, our simulations likely are still creating cold pools that are unrealistically strong, ultimately shortening the TLV lifetime.

2) TRAJECTORY ANALYSIS

To evaluate the source region of high-vorticity air present in the updraft, backward parcel trajectories are calculated within the low-level central updraft at approximately 1 km AGL

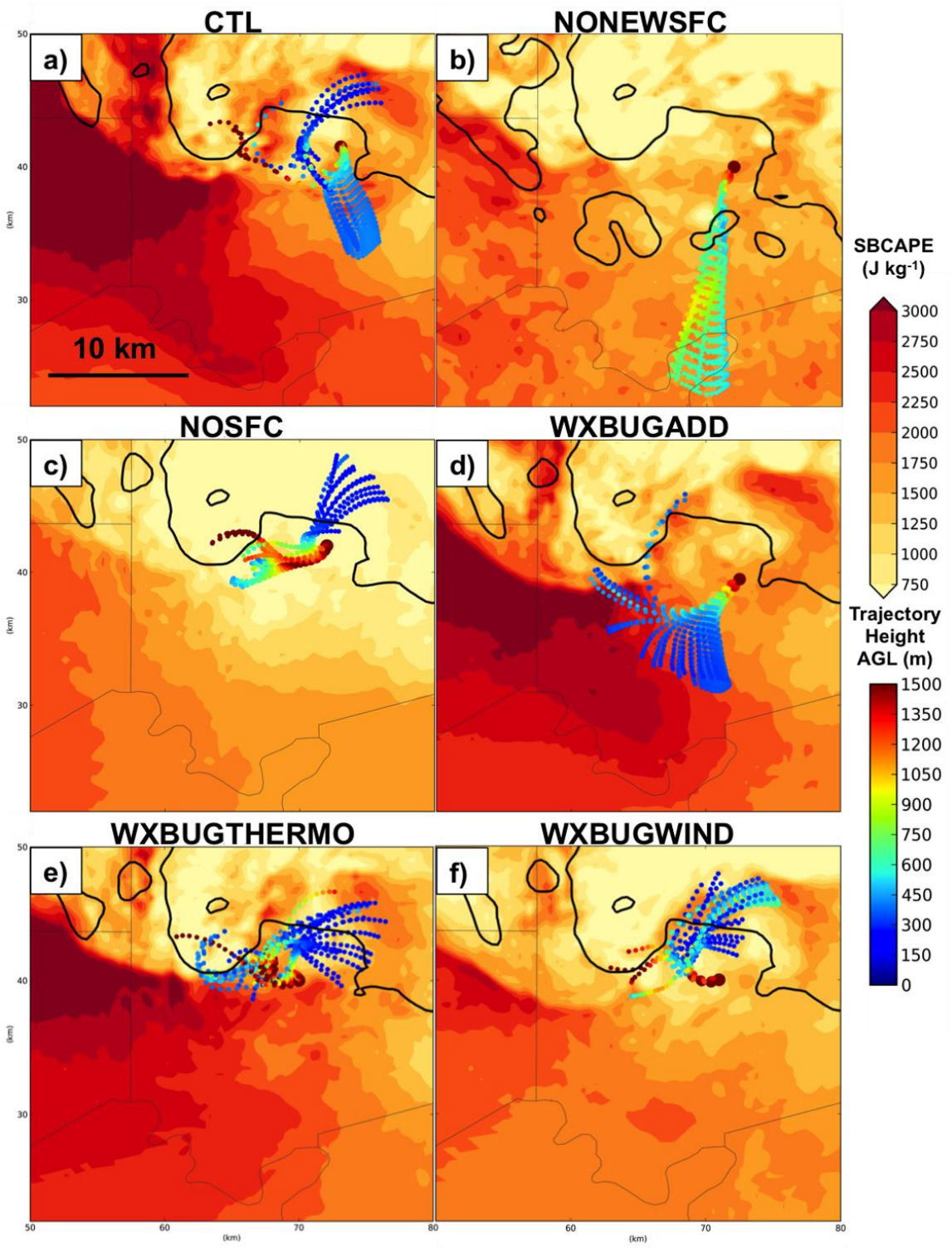

Figure 13: CAPE (color fill, J kg^{-1}) and trajectories for experiments labeled above each panel. CAPE is plotted at 0145 UTC along with the 25-dB*Z* reflectivity contour. Backward trajectories are shown in all panels from 0155 UTC to 0145 UTC.

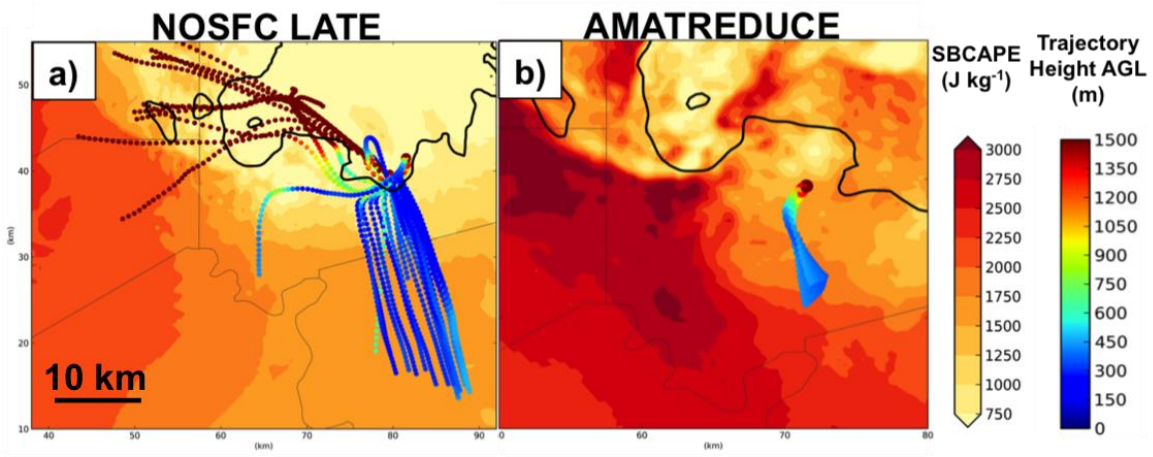

Figure 14: As in Fig. 13, but for a) NOSFC and b) AMATREDUCE. NOSFC backward trajectories taken ending at 0210 UTC.

leading up to 0155 UTC. In all cases, 72 parcels in a 300-m radius ring surrounding a central point are integrated backwards using a fourth-order Runge-Kutta scheme every 30 s for 10 min, or to the final analysis time at 0145 UTC. The trajectories are shown in Fig. 13 where heights AGL are indicated by colored dots (surface levels are cooler colors and higher levels are warmer). In the experiments that quickly develop persistent and deep low- to mid-level updrafts (i.e. CTL, WXBUGADD, and WXBUGTHERMO), parcels entering the main updraft originate from either the pocket of unstable air to the northeast of the forward-flank downdraft (FFD) or from the large reservoir of high-CAPE air to the southwest of the storm (Fig. 13a,d,e). Furthermore, inflow air intercepting the base of the forward-flank gust front (FFGF) emanates from the southwest, within an additional region of high-buoyancy air (not shown). Experiments WXBUGWIND and NOSFC, which were unable to produce persistent and strong updrafts, reveal parcels originating from less-buoyant air (Fig. 13f, c).

Experiments NOSFC and AMATREDUCE evolve in a somewhat different manner. In NOSFC, surface ζ maximizes later than in CTL, WXBUGADD, and WXBUGTHERMO (cf. Fig. 12), and the central updraft does not organize until after 0205 UTC. The trajectories at 0155 UTC show parcels originating from a region of somewhat reduced SBCAPE around 2000 J kg^{-1} in northern Bosque County (Fig. 13b). However, backward trajectories taken 15 min later show moderately to strongly unstable air (SBCAPE between 2250 to 2750 J kg^{-1}) originating from Hood and Somervell counties supporting a strong low-level updraft at 0210 UTC (Fig. 14a). Similarly, most of the parcels entering the central updraft in AMATREDUCE originate from a region of SBCAPE around 2500 J kg^{-1} to the southeast of the FFGF with no indication that parcels from the CAPE "pocket" to the northeast have entered the updraft at this time (Fig. 14b) as was observed in CTL.

Based on the noted differences in the updrafts of the Cleburne storm, the stretching of vertical vorticity should play a key role in determining the strength of the low-level circulations. Time-height profiles of maximum vertical velocity (w) and ζ are displayed in Figs. 15 and 16 respectively. The w plots are of the maximum in a square of 2.4 km around the largest value of ζ around the low-level TLV. A persistent and strong low-level updraft is present in CTL throughout the first 20 min of the simulation, coincident with the rapid increase of near-surface ζ. A similar updraft structure is noted in both WXBUGADD and WXBUGTHERMO along with the associated development of very high values of ζ (>0.1 s^{-1}), supporting the assumption that stretching plays a primary role in the development of the TLV seen in these experiments. NONEWSFC, which spawned a TLV roughly 20 min later than CTL, initially does not develop appreciably strong low- and

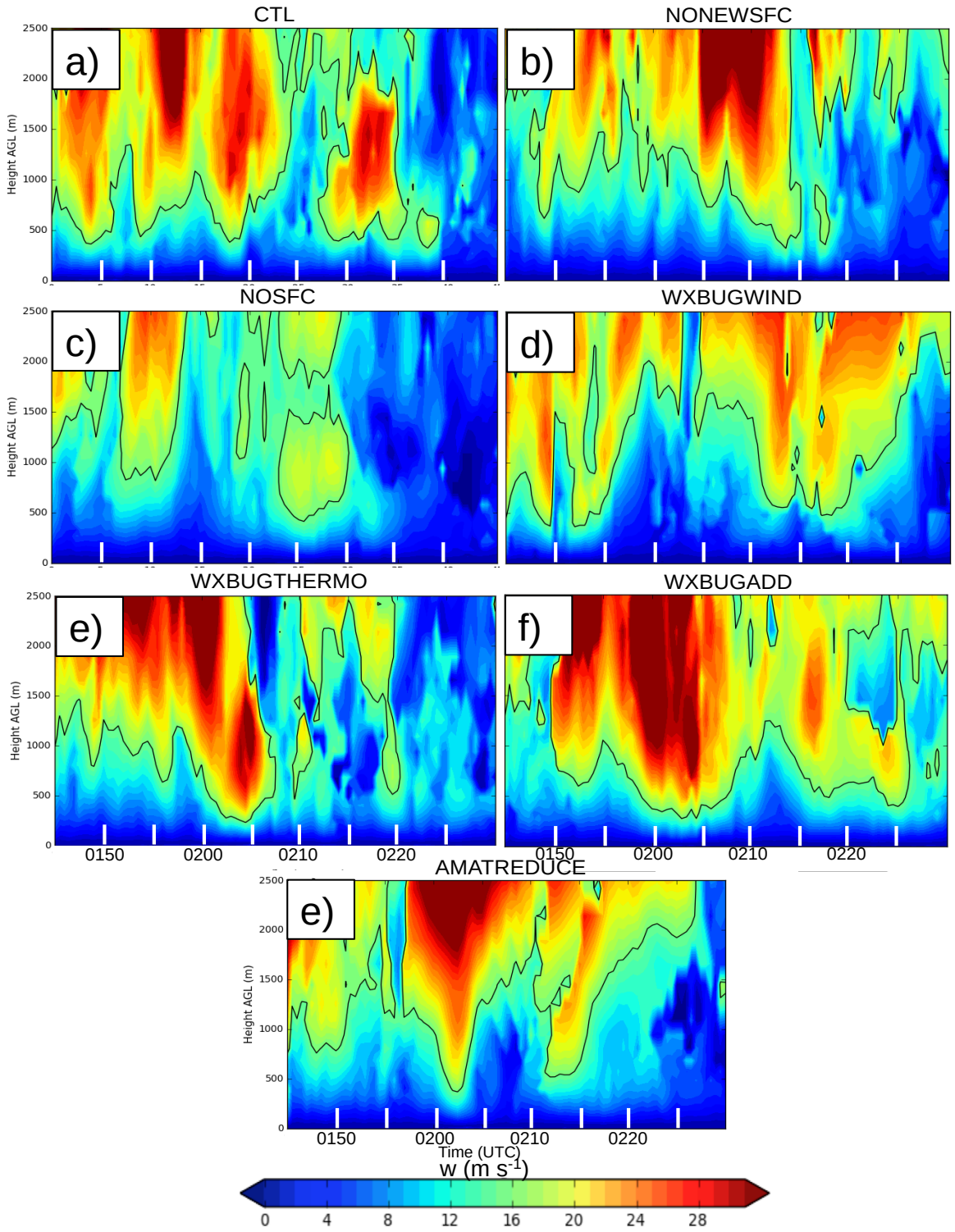

Figure 15: Time-height profiles of w (m s^{-1}) within a square of length 2.4 km around the maximum value of ζ between 0145 and 0230 UTC. Black contour drawn at w = 15 m s^{-1}.

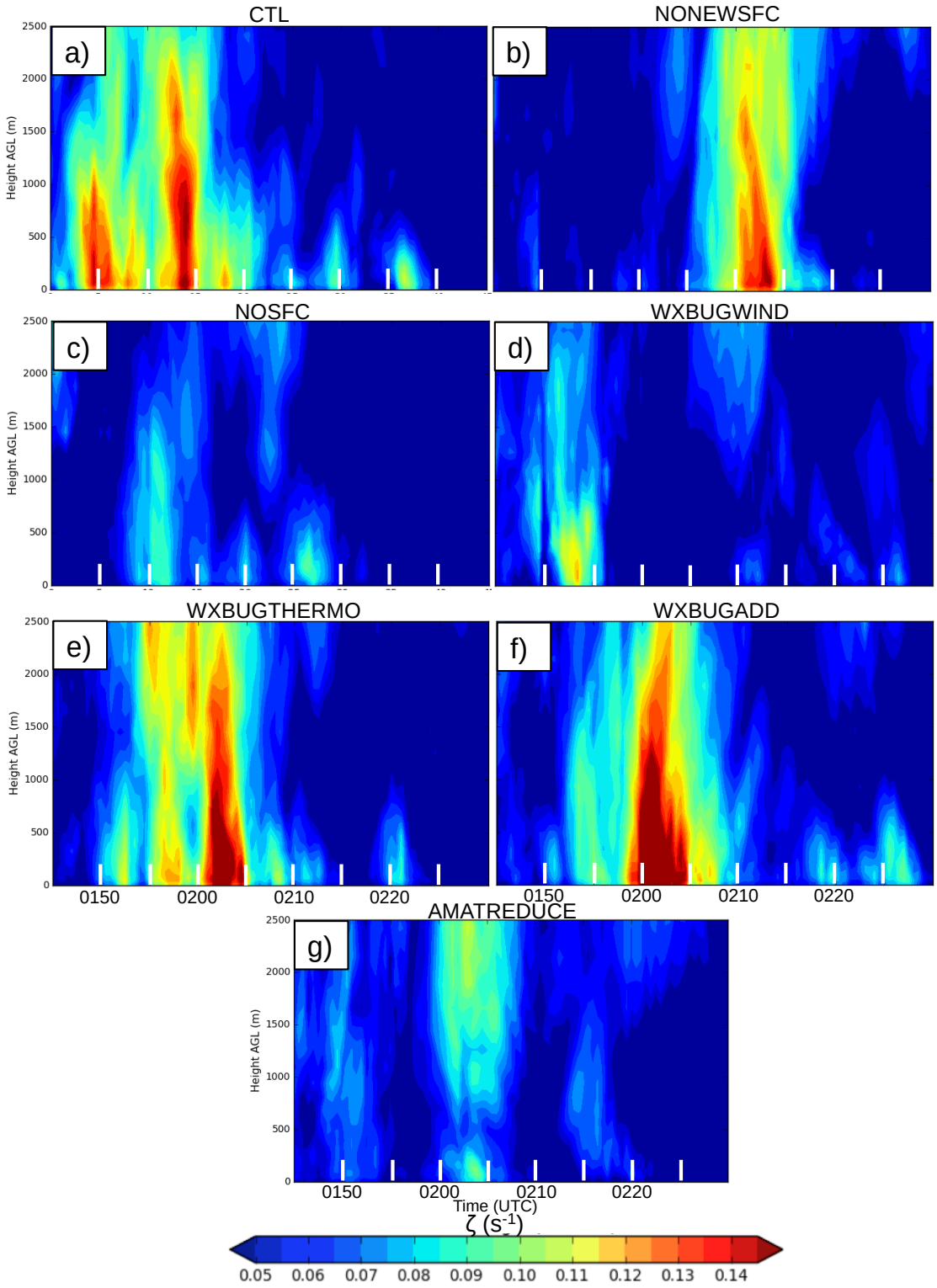

Figure 16: As in Fig. 15, but for vertical vorticity (s^{-1}). mid-level *w*. A sudden increase in updraft strength below 3.5 km occurs 1200–1500 s in the forecast, attendant with a short-lived, intense surface circulation. Much weaker *w* (generally <25 m s^{-1} through the first 3 km) is noted in NOSFC.

Also, the updraft heights in CTL, WXBUGTHERMO, and WXBUGADD are noticeably lower than in NOSFC, WXBUGWIND, AMATREDUCE, and in the first 20 min of NONEWSFC, with w >15 m s^{-1} routinely originating below 500 m as low-level ζ increases (Fig. 15). This consistent decrease in low-level updraft height is not seen in NOSFC, AMATREDUCE, or the early stages of NONEWSFC. While other processes are present here, the rapid increase in low-level updraft strength, likely from the ingestion of high-θ_e air, plays a large role in the development of intense mesovortices due to ζ stretching.

c. Surface-level forecast verification

To examine the impact of surface data on the forecast evolution of surface fields, *rmsd* values are computed for 60 min at the ten independent ASOS observation locations mentioned in section 5a. We verify forecasts using the ASOS rather than non-conventional observations for two primary reasons: 1) they are the most accurate and representative data available in the domain, and 2) are the only observations that are independent (except for their use in the very first analysis cycle). The *rmsd* values for 2-m temperature, 2-m dewpoint, and surface pressure are shown in Fig. 17. The *rms* differences in CTL for both temperature and dewpoint are the lowest out of all simulations for roughly the first 30 min, while the AMATREDUCE experiment shows slightly higher values than CTL throughout the 60-min verification window. This result is consistent with our expectations owing to the erroneously reduced errors ascribed to the non-conventional surface observations. Interestingly, experiments NONEWSFC, WXBUGTHERMO, and WXBUGWIND begin to verify better than CTL after 30–40 min in the temperature and dewpoint fields. Closer inspection of individual station time series reveals a rapid decrease in both temperature and dewpoint at KDTO (Denton Municipal Airport) 15–20 min into the simulation related to the passage of a gust front (Fig. 18). NONEWSFC, WXBUGTHERMO, and WXBUGWIND, which all initially start out with temperatures and dewpoints that are too low, quickly approach the observed values after the gust front passage, resulting in the (erroneous) improvement in the *rmsd* values over CTL in Fig. 17.

These verification results are consistent with the nature of convective-scale predictability. While previous research has shown that the predictability of the larger-scale flow is on the order of days to weeks (Lorenz 1969), it decreases with diminishing horizontal scale. Cintineo and Stensrud (2013) found that while storm location and the lightest areas of rainfall were predicable out to 1–2 hours, locations of cold pools or mesoscyclones were only predictable out to tens of minutes.

6. Discussion and conclusion

While numerous studies have tested the impacts of assimilating Doppler radar velocities and reflectivity on high-resolution forecasts of convective storms, few have attempted to quantify the impacts from observations from surface stations. This is likely due to the reduced spatial and temporal resolution of surface networks over traditional radar networks. Nevertheless, surface observations provide useful information about the low-level thermodynamic and kinematic structures that are typically not resolved by conventional Doppler radars. Additionally, surface observations are able to directly influence the distribution of heat and moisture in the lowest model levels, something that is not possible through the assimilation of radar data alone. To assess potential impacts of surface data, we performed data denials of conventional and non-conventional surface observations on 400-m-resolution forecasts of a supercell that developed on 15 May 2013 in north-central Texas.

Our results show that assimilating observations from non-conventional sites such as the EarthNetworks WeatherBug network or those from the Citizen Weather Observer Program (CWOP), along with properly specified observation errors, lead to improved surface analysis of the low-level thermodynamic field in this example test case. Additionally, *rmsd* values of forecast temperature, dewpoint, and surface pressure were improved over simulations that denied these stations from the assimilation window when compared to observations from several independent surface stations.

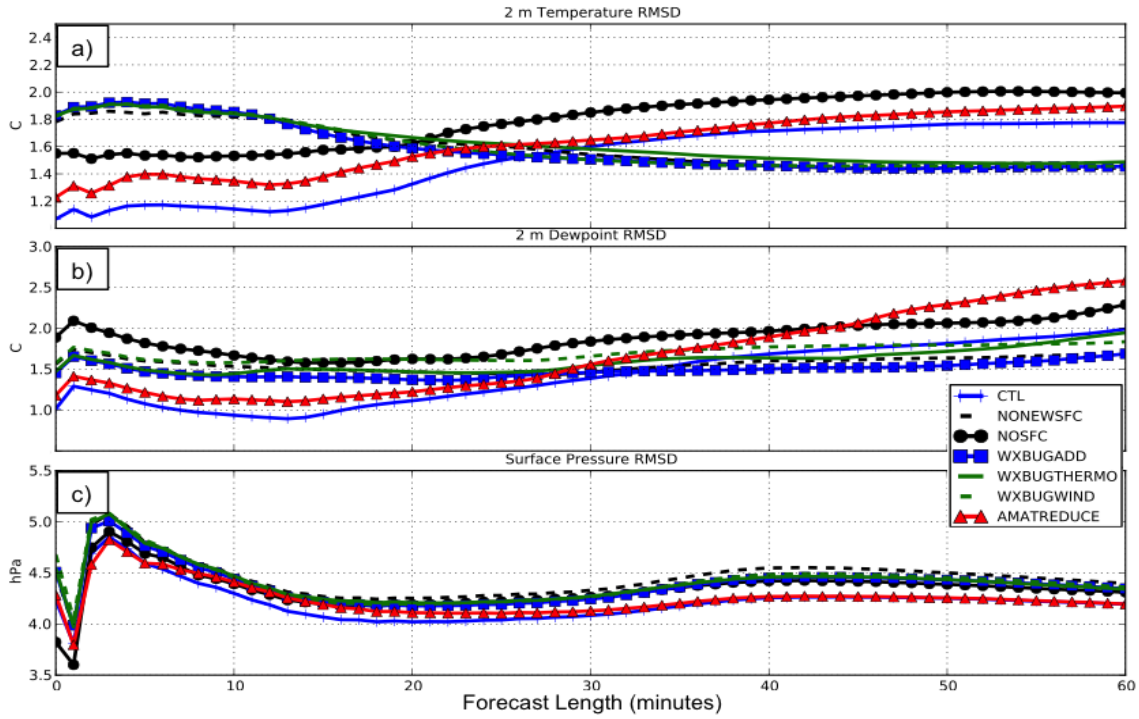

Figure 17: Root mean square difference (RMSD) time-series of (a) 2-m temperature (°C), b) 2-m dewpoint (°C), and c) surface pressure (hPa) computed at 10 independent surface stations between 0145 and 0245 UTC 16 May 2013, for experiments as labeled in the legend.

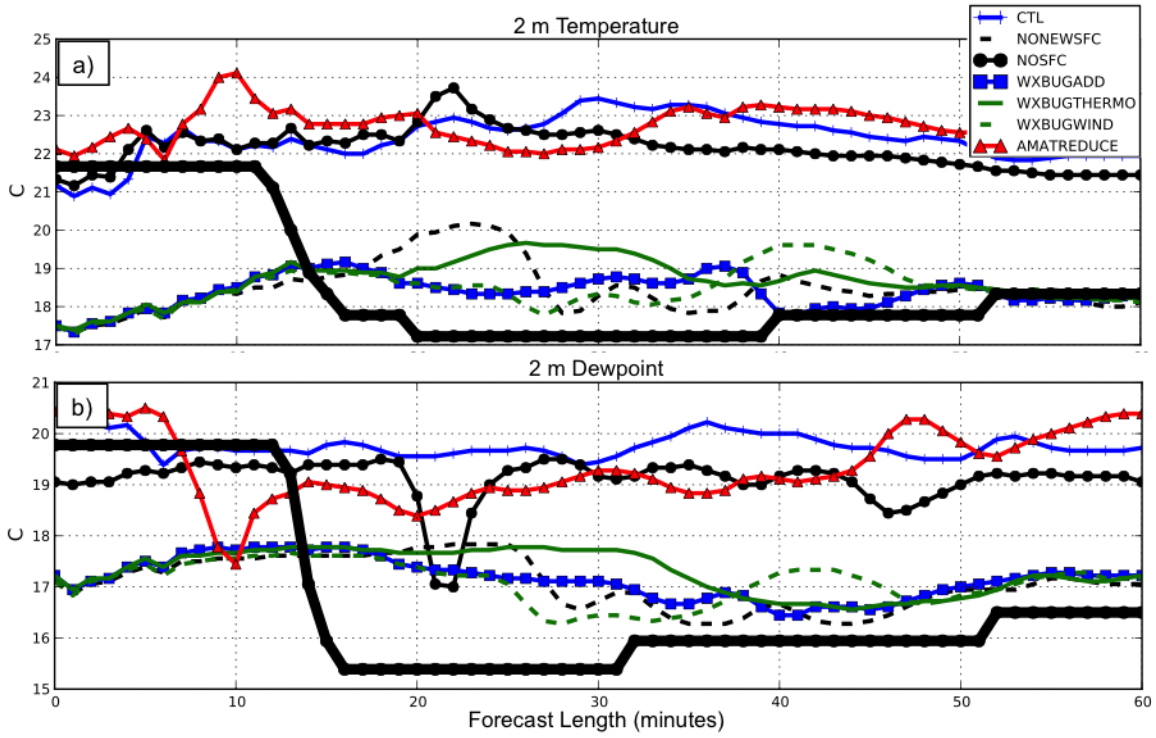

Figure 18: Traces of a) 2-m temperature (°C) and b) 2-m dewpoint (°C) at Denton Municipal Airport (KDTO) between 0145 and 0245 UTC 16 May 2013, for experiments labeled in the legend.

A substantial increase in the 10-m θe field were noted in the data sparse region of Johnson County, TX, where the addition of thermodynamic data from several WeatherBug sites ultimately proved to be critically important in forecasting the evolution of a tornadic supercell that moved through the town of Cleburne. The experiments that did not assimilate non-conventional surface data—in particular, data from the WeatherBug network—either delayed the development of an intense surface-level TLV, or were unable to develop one at all. In contrast, the control simulation, and those that made use of the additional thermodynamic information from several WeatherBug sites around the Cleburne supercell, were able to develop an intense surface-level circulation that roughly coincided with the observed tornado track.

These differences in storm evolution seem to stem from the ability of the non-conventional observation sites to provide information about the near-surface thermodynamic field and to "fill in the gaps" in data coverage in an otherwise data-sparse portion of the domain. Increases in the 10-m θe and SBCAPE within the inflow-region of the Cleburne supercell are ultimately associated with the development of vigorous low- and mid-level updrafts. While purely speculative, the rapid intensification of low-level ζ coincident with increases in updraft strength is possibly an indication that the low-level updraft plays a key role in the stretching of ζ leading up to the development of the Cleburne TLV. However, further testing is needed to show that this is the case, and is out of the scope of this current work.

While these results are encouraging, further investigation is necessary to determine if the positive analysis and forecast impact seen here applies to additional high-impact weather events. Rather than deny observations by sensor type, additional experiments testing the impacts of removing observations from particular regions of the domain may reveal more about the relative importance of surface data in relation to storm location. Two key results of this study are that 1) the incorporation of additional surface observations led to improvements in the analysis and forecast of a tornadic supercell in north-central Texas and 2) the implementation of an accurate Warn-on-Forecast procedure (Stensrud et al. 2009, 2013), at least in this case, appears to be highly dependent upon the accurate depiction of the low-level thermodynamic field.

ACKNOWLEDGMENTS

This project was funded through the Department of Commerce and NOAA through grant NA110AR4320072, "Prototyping and Evaluation of Key Network of Network Technologies." Additionally, archived observations and metadata were graciously provided by EarthNetworks and Global Science and Technology. This manuscript was greatly improved with helpful and insightful reviews from John Horel, Leigh Orf and Brice Coffer. The authors also wish to thank, in no particular order, Ariel Cohen, Kevin Thomas, Keith Brewster, Yunheng Wang, Nate Snook, Chris Cook, and Alex Schenkman for their considerable assistance during the course of this research project. The computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU).

REFERENCES

Atlas, R., 1997: Atmospheric observations and experiments to assess their usefulness in data assimilation. *J. Meteor. Soc. Japan,* **75**, 111–130.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396–409.

Benjamin, S. G., B. E. Schwartz, S. E. Koch, E. J. Szoke, 2004: The value of wind profiler data in U.S. weather forecasting. *Bull. Amer. Meteor. Soc.*, **85**, 1871–1886.

——, 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.

Bothwell, P. D., J. A. Hart, and R. L. Thompson, 2002: An integrated three-dimensional objective analysis scheme in use at the Storm Prediction Center. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, *Amer. Meteor. Soc.*, JP3.1.

Brewster, K., M. Hu, M. Xue, and J. Gao, 2005: Efficient assimilation of radar data at high resolution for short-range numerical weather prediction. Preprints, *WWRP Int. Symp. Nowcasting Very Short Range Forecasting*, Toulouse, France, Météo-France, 3.06. [Available online at http://twister.ou.edu/papers/BrewsterWWRP_Nowcasting.pdf.]

Brown, J., and Coauthors, 2012: Rapid Refresh replaces the Rapid Update Cycle at NCEP. Preprints, *21st Conf. on Numerical Weather Prediction/25th Conf. on Weather Analysis and Forecasting*, Montreal, QC, Canada, Amer. Meteor. Soc. 3B1.2.

Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo. 104606, Vol. 3. NASA Goddard Space Flight Center, 85 pp.

Cintineo, R. M., and D. J. Stensrud, 2013: On the predictability of supercell thunderstorm evolution. *J. Atmos. Sci.,* **70**, 1993–2011.

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

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

——, ——, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms. *Mon. Wea. Rev.,* **138**, 1152–1171.

EarthNetworks, 2015: Weather Station. [Available online at http://www.earthnetworks.com/Products/WeatherStation.aspx.]

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.

Gilchrist, A., 1982: JSC study conference on observing systems experiments, Exeter, 19–22 April 1982. *GARP/WCRP Numerical Experimentation Program. Rep. 4*, WMO Bull., 83 pp.

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

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

Heppner, P. O. G. Jr., 2013: Building a national network of mobile platforms for weather detection. Preprints, *29th Conf. on Environmental Information Processing Technologies*, Austin, TX, Amer. Meteor. Soc., 5.8.

Horel J. D., and X. Dong, 2010: An evaluation of the distribution of Remote Automated Weather Stations (RAWS). *J. Appl. Meteor. Climatol.,* **49**, 1563–1578.

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

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

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

Liu, H., and M. Xue, 2008: Prediction of convective initiation and storm evolution on 12 June 2002 during IHOP 2002. Part I: Control simulation and sensitivity experiments. *Mon. Wea. Rev.,* **136**, 2261–2282.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21**, 289–307.

Lord, S. J., E. Kalnay, R. Daley, G. D. Emmitt, and R. Atlas, 1997: Using OSSEs in the design of the future generation of integrated observing systems. *Preprints, 1st Symp. on Integrated Observation Systems,* Long Beach, CA, Amer. Meteor. Soc., 45–47.

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.

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.* National Academies Press, 250 pp.

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

NWS, 1994: Technique specification package 88-21-R for AWIPS-90 RFP. Appendix G requirements numbers: Quality control incoming data, AWIPS Doc. TDP-03201992R2, NOAA/National Weather Service Office of Systems Development, 39 pp. [Available online at http://docs.lib.noaa.gov/noaa_documents/NWS/NWS_TSP_88-21-R2.pdf.]

NWS, cited 2015: Tornadoes across North Texas, May 15th, 2013. [Available online at http://www.srh.noaa.gov/fwd/?n=tornadoes051513.]

Rogers, R. R., and M. K. Yau, 1989: *A Short Course in Cloud Physics.* 3d ed. Pergamon Press, 293 pp.

Schenkman, A., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011a: The analysis and prediction of the 8–9 May 2007 Oklahoma tornado 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.

——, ——, and M. Hu, 2014: Tornadogenesis in a high-resolution simulation of the 8 May 2003 Oklahoma City supercell. *J. Atmos. Sci.*, **71**, 130–154.

Snook, N. and M. Xue, 2008: Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophys. Res. Lett.*, **35**, L24803.

Stensrud, D. J., and Coauthors, 2009: Convective-scale Warn-on-Forecast system. *Bull. Amer. Meteor. Soc.,* **90**, 1487–1499.

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

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261.

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

Weisman, M. L., and R. J. Trapp, 2003: Low-level mesovortices within squall lines and bow echoes. Part I: Overview and dependence on environmental shear. *Mon. Wea. Rev.*, **131**, 2779–2803.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster 1995: *ARPS Version 4.0 710 User's Guide*, 380 pp. [Available online at http://www.caps.ou.edu/ARPS].

——, ——, ——, 2000: The Advanced Regional Prediction System (ARPS) — A multiscale non-hydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.,* **75**, 161–193.

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

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

——, M. Hu, and A. D. Schenkman, 2013: Numerical prediction of the 8 May 2003 Oklahoma City tornadic supercell and embedded tornado using ARPS with the assimilation of WSR-88D radar data. *Wea. Forecasting,* **29**, 39–62.

Zapotocny, T. H., W. P. Menzel, J. P. Nelson, 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, 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 (John D. Horel):

Initial Review:

Recommendation: Revisions required.

The manuscript by Carlaw et al. is well written and presents some interesting results. The figures are appropriate, although the labeling can get a bit confusing. My only concerns center around the interpretation of some of the validation results, particularly relying on comparisons to ASOS stations. In addition, there tends to be too much discussion of the accuracy as opposed to representativeness of the various surface observation types.

We thank you for your numerous insightful comments, which have added considerably to the readability and scientific content of this manuscript.

Major comments:. The ARPS-3DVAR system has been around quite a while. It wasn't clear to me why it was necessary to review the details of the system quite as much as is present in the manuscript. But, since the electronic journal format allows a bit more flex in terms of paper length, it was a helpful review for me. I'm OK with much of this description in its current form, but it would be helpful to highlight any significant changes from earlier versions.

The main reason for the rather detailed description of the ARPS 3DVAR system stemmed from our need to address changes we made to standard inputs, as well as notable pitfalls of some procedures. In particular, the discussions about altering the 3D divergence constraint coefficients, inability of the ARPS 3DVAR to remove spurious hydrometeors from analyses, and the addition of a low wind-speed check, were facilitated by a general discussion of the entire 3DVAR system.

We fully agree with your suggestion to highlight changes from previous versions and have added text within the discussion of the ARPS 3DVAR system.

I understand why the focus of the verification of the analyses up to 0145 UTC relied more on the "independent" ASOS stations (even though they are used in the first pass of the recursive analysis), but that is not clear for verifying the forecasts from 0145–0245 UTC, e.g., Fig. 17 for the ten "independent stations" (all the station observations are now independent for the forecast period) and KDTO specifically in Fig. 18. Why didn't you verify the forecasts versus the observations of the 6 WeatherBug stations in Johnson County? It would seem to be more relevant to know how well the various sensitivity runs did in that area, rather than in the other regions of the domain.

The decision to verify both analyses and the subsequent simulations using the ASOS sites came down to a matter of consistency and the desire to perform as unbiased [of] a verification as possible given our limited choices of highly accurate and representative surface observations. While the nonlinearity of the simulations should quickly remove much of the dependence on the observations used in the analyses, some degree of influence still exists in the forecasts. This could provide an "unfair advantage" to the non-conventional dataset when verifying forecasts that incorporated that data during each of the analysis cycles, especially given how many more non-conventional observations are assimilated compared to the number of ASOSs. We decided instead that using the 10 ASOS stations provided a good "big-picture" overview as to the general performance of each forecast, while letting the simulations of the supercell/TLV themselves speak to the importance of the non-conventional observations. We have elaborated upon this rationalization at the beginning part of 5c: Surface-level forecast verification.

Throughout the paper, ASOS stations are described as more "accurate". However, sensor accuracy in

terms of the characteristics of the physical hardware and related maintenance protocols are only part of the issue. More critical is the representativeness of the observations, which is never mentioned in the manuscript. For example, I believe all the WxBug stations in Johnson County are at schools. The wind sensors are likely mounted right above the roof line, which may be why the winds are not representative of conditions generally and the simulations using WxBug winds do poorly. Temperature and moisture observations at the schools are less sensitive to the siting issues and are likely more representative of conditions nearby.

We appreciate this insightful observation. This indeed is an issue that we noted (in related but separate work during related Master's thesis work) but failed to include in this manuscript. We mentioned the addition of a quality-control step that removes low wind speeds because of this issue, but were vague and incomplete about the reason. We have expanded the "Observation Dataset" section (4a) to address this issue of representativeness. Additionally, we added discussion about the motivation to use the 10 ASOS stations as verification points based on both their improved physical/sensing characteristics and their—in general—better representativeness of the local environment over the non-conventional sites.

The observation errors assigned in Table 1 should include assumptions about representativeness. But, these values seem low, and there is no information supplied on these values relative to those of the background errors. For example, Tyndall et al. (2010, WAF) guesstimated 2:1 METAR/background error variances with slightly higher ratios for the PUBLIC (CWOP and WeatherBug) observations. I don't see a lot of evidence for it in the figures, but there is a chance that you are overfitting the analyses to the observations that are assumed to be more accurate than they are. The Non-conventional WxBug temp observations (last row in Table 1) may actually affect the analysis more because you are not weighting them highly relative to the other observations and creating local analysis bullseyes. In those data-sparse regions they can then help correct a poor background.

The OB:BACKGROUND ratio varies somewhat owing to variations of values with height and observed field, but in general these ratios are near 0.8-1:1 for the conventional stations, and between 1-2:1 for the non-conventional observations. Several previous assimilation tests were performed using all available data to qualitatively determine a "best-guess" for the observation:background ratios by varying the observation errors and then qualitatively assessing how smooth the final product was (attempting to avoid the production of too many analysis bullseyes). Additionally, the use of multiple 3DVAR passes with incrementally decreasing de-correlation radii helps to create generally smoother analysis fields, necessitating the use of somewhat smaller error ratios. Finally, the AMATREDUCE experiment, which artificially reduced the non-conventional observation variances to match those of the ASOS stations did show a decrease in verification scores compared to the Control run, indicating that our choice of ob:background error for the non-conventional sites—at least in this case—is likely close to optimal for this situation. We have added a summary of these points in the ARPS3DVAR section.

[Minor comments omitted...]

Second review:

Recommendation: Accept.

General comment: The authors have done a good job responding to the reviewers' comments. My specific concerns have been addressed adequately and recommend that the paper be accepted for final publication.

REVIEWER B (Leigh G. Orf):

Initial Review:

Reviewer recommendation: Revise and resubmit.

General comments: In this paper, the authors describe the effect of assimilating surface data from nonconventional sources. They find that by assimilating surface data in otherwise data-sparse regions, ARPS simulations show improved results as compared to simulations that exclude these data.

Major comments (issues I have with paper, in order of most to least important): I do not find the results of this research to be very novel it is not really a surprise that having (even somewhat imprecise) surface data in data sparse regions such as in key locations for supercell morphology is concerned would result in a more accurate model simulation of a supercell.

While it certainly stands to reason that the assimilation of observations in data sparse regions should—given proper error values—improve an analysis, it was not clear to us that the inclusion of surface data should result in an improved simulation of a supercell, let alone a tornado-like vortex (TLV). Substantially more information about the kinematic fields within the supercell is available from the assimilation of Doppler radar velocities, especially given the proximity to the KFWS site. In addition, as outlined in the introduction, Schenkman et al. (2011) find that the assimilation of radial velocities is of greater importance than Oklahoma Mesonet surface data in the simulation of a mesoscale convective vortex. Finally, this is—to the best of our knowledge—the first time an attempt has been performed to both quantify and qualitatively assess the impacts of non-conventional surface observations on high-resolution simulations. We believe this is a unique contribution to the broader scientific literature, especially given the unexpected profound effects these data have on our simulations. In an attempt to improve this manuscript, we have laid out more clearly the novelty of this work at the end of the introduction.

I find the paper lacks focus [and] tries to cover too much ground. Is it really necessary to go into such detailed in the difference between the simulations? The paper seems to veer too much into descriptions of supercell morphology. Are these details necessary to make your point that the inclusion of these unconventional sources is important? Many of the simulated storms appear to not show any rotation aloft, suggesting that they are not supercellular.

We believe this level of discussion is warranted. One of the main points of this paper is to show clearly the distinct differences in the simulated evolutions of the various supercells, a task that requires a thorough discussion of the morphologies of each simulated supercell. Even with the in-depth analysis of the distribution of theta-e, visualization of vertical vorticity, and plotting of trajectories, we've only begun to scratch the surface of an entirely more complex parameter space that goes into producing a tornado.

Finally, we do see deep and intense rotation within our simulated supercells as evidenced by plots of vertical vorticity (see point to follow below which provides more detail). The time-height plot shown in Fig. 16 was taken in a straight column, and therefore begins to exit the area of maximum vorticity aloft as the vortex tube is tilted from the vertical amidst strong deep-layer shear, which may give the appearance of a vortex tube of limited vertical extent.

The paper is highly descriptive perhaps at the expense of a more valuable analysis.

This is a fair point and a previous reviewer voiced some concerns about the level of detail that went into the ARPS3DVAR section. We point out, however, that several important changes are made to the analysis code for this work, which necessitates a more thorough description of the analysis system. In addition, those readers unfamiliar with data assimilation may find some of the information useful in facilitating a better understanding of this work's methodology.

None of the simulations seemed to get the tornado track right at all, as compared to observations (shown in Fig. 11). The different simulations are compared to each other, but not to the observed storm (not sure

whether this is due to lack of observational data?)

Getting the tornado track exactly right was not a primary goal of this research, nor is it one that should be expected at such a large grid length relative to the dimensions of a tornado. Instead, this work is an attempt to share with the broader scientific community an important result that may improve tornado simulations in the future, points that we have laid out clearly in the introduction and conclusion from a bigger-picture, mesoscale environmental perspective.

There are too many figures, and many figures lack adequate annotations (more detail below).

We believe the figures are all useful in facilitating the reader's understanding of this work, particularly those who may not be familiar with high-resolution modeling, data assimilation, or tornadogenesis. However, we have taken many of your suggestions into account and used them to improve the quality of several figures. We have additionally removed what was previously Fig. 8, as another reviewer found it added little substance to the manuscript.

It is clear that the authors got a positive result by assimilating surface data from unconventional sources. It would have been more interesting if they hadn't gotten that result! In some regions, surface θ_e is off by 20K! So is it any surprise that a supercell ingesting this kind of near-surface-θ_e air will be more vigorous than one that ingests dry/cool air? That is the main problem I have with the paper, that as far as pushing the field forward, I think this paper reads too much like a paper on exploring the effects of supercell morphology to changes in environmental conditions, and many papers have already been written on this.

We address these concerns in point 2 above and in the first response in the "Paper Conclusions" section below, but provide some additional discussion here. The purpose of this work is not to show that increasing theta-e within a supercell's inflow results in a stronger storm, but that more properly characterizing the near-storm environment is accomplished only with the inclusion of non-conventional observations, which then results in improved simulations of the supercell.

I understand the practical nature of their approach, categorizing based upon network/data provider, not location but Fig. 4 indicates the vast majority of the nonconventional surface obs are clustered in what would be the (far?) forward-flank region of the supercell. One wonders the relative importance of the nonconventional obs as a function of location, not as a function of network. I think a more valuable approach would have been to selectively remove [and] add assimilated surface data in regions as a function of location relative to the mesocyclone and the air that feeds it. It may turn out that only a tiny percentage of the assimilated data matters at all (i.e., 1 or 2 points determine whether you get a vigorous circulation or not), and the rest is of secondary importance.

While it's true that a majority of the non-conventional observations shown in Fig. 4 fall in the forward-flank region of the supercell at the final analysis time of 0145 UTC, this is at the end of a 45-minute assimilation period. During the beginning of the 5-min intermittent assimilation cycle, the supercell is located in northern Hood County (see Fig. 4a for reference) where the cluster of non-conventional sites in Hood and Johnson Counties would be in or near the storm inflow.

To address your second point about selectively adding back observations in particular regions, we have done this with the WXBUGADD, WXBUGTHERMO, and WXBUGWIND experiments. These experiments add back observed variables to the analysis in/near the storm inflow and we find that they do in fact play a fundamental role in developing a TLV in the correct location.

It seems that in many of the simulations, we just don't even get a supercell. When I look at the 10-m vorticity and the surface wind speed, it seems we have marginal supercells at best. Is it necessary to analyze these storms in such detail? [Then] Figure 16 is pretty clearly indicating storms with little to no rotation.

We appreciate this observation; however, we offer an argument indicating that we do simulate rapidly rotating storms. Weisman and Trapp (2003) consider vertical vorticity greater than 0.025 s^{-1} to be the

level defining a mesovortex in their 1-km grid spacing simulations. Scaling this to 400 m roughly equates to 0.01 s^{-1}, an order of magnitude less than we see in our results. In addition, Xue et al. (2014) show simulations of a tornadic supercell at 1-km and 50-m grid spacing and found maximum vertical vorticity values on the order of 0.015 s^{-1} and 1–1.5 s^{-1} respectively (see Figs. 6 and 13 in their paper). These values scale comparatively well to our values (~0.1–0.2 s^{-1}) at 400-m grid spacing. To address your last point, the time-height plots of vertical vorticity were taken in a straight column through the maximum surface-level vertical vorticity (ζ). Thus the column quickly begin to exit the primary vertical vortex tube, which has been tilted, and therefore do not capture the full scale of vertical vorticity through the entire depth of the storm. However, our main goal with Fig. 16 was to show the magnitude of ζ associated with the simulated storm's low-level updraft in Fig. 15, not the full breadth of the vortex tube.

References:

Weisman, M. L., and R. J. Trapp, 2003: Low-level mesovortices within squall lines and bow echoes. Part I: Overview and dependence on environmental shear. Mon. Wea. Rev., ***131***, *2779–2803.*

Xue, M., 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.*

Do the authors have any explanation for the rapid spinup of the control simulation, and lack thereof for all other simulations?

This could be the result of a number of phenomena: initial organization of low-level vertical vorticity, the actual theta-e values in both the inflow and RFD, differences in the final analysis developed during the 45-min assimilation-forecast window, etc. We can only speculate as to the exact reasoning here, and further understanding would require a much more detailed investigation outside the scope of this work. CTL, WXBUGTHERMO, and WXBUGADD all spin up within 200–400 s of one another, while WXBUGTHERMO and WXBUGADD simply maintain a higher level of ζ for an extended period of time compared to CTL.

Getting thetae right at the surface in regions important to supercell morphology leads to dramatic differences in numerical simulation morphology, a not surprising result, and appears to be the main conclusion of the paper.

We likely could have artificially inflated the theta-e in and around the supercell's inflow and produced similar results, a finding that would not be surprising. The fact that we produce these results via the inclusion of previously untested, potentially less accurate and representative sensors is what is fundamentally new to the field of high-resolution data assimilation and modeling. In other words, the "end" is not the most surprising part, the "means to that end" are.

At the end of the second to last paragraph you state that "...lending support to the idea that the low-level updraft plays a key role in the stretching of vertical vorticity leading to the development of the Cleburne mesovortex". While this may indeed be true, it is a conclusion that comes about without any previous discussion from the paper text, which is ostensibly not about supercell dynamics. I'm not sure it really belongs here, as it appears to be little more than speculation, not the result of any analysis.

We completely agree with your concerns here and have altered the last two sentences of that paragraph to make the speculative nature of that argument more clear.

I think a more valuable contribution would look at the relative importance of assimilating surface data in a more generic sense, grouping not by network but by location. Then you could go a step beyond saying that having more assimilated surface data improves the simulation, and could explore just how much you need to make an impact and why. Grouping by network/provider company is not useful in a case-by-case basis; with the next storm, who knows how it would have panned out? It would have been more valuable, I think, to deny data in a more selective way to see the relative value of assimilated surface data with respect to

relative storm location.

To some extent, we have addressed this issue already in the present work. In typical Observing System Experiment (OSE)-fashion, we denied observation platforms by type, but the WXBUGADD, WXBUGTHERMO, and WXBUGWIND experiments added stations (albeit from a single data source) near the storm inflow back into the analysis for just this purpose. However, a more thorough location-specific data denial would require substantial additional investigation, which is largely tangential to the main scope of this work. We do agree that a location-specific OSE could be an interesting topic for future work and have added a reference to this in the conclusion.

[Minor comments omitted...]

Second review:

Recommendation: Accept.

General comment: I am satisfied with the paper as revised, and offer no further suggestions for improvement. They have done a fine job addressing my concerns and I recommend it for publication in EJSSM.

REVIEWER C (Brice E. Coffer):

Initial Review:

Reviewer recommendation: Accept with major revisions.

General comments: This is a nice paper that demonstrates the positive impacts of assimilating non-conventional surface observations into high-resolution convection forecasts, especially in regions where the conventional data sources are sparse. The research is clearly of interest to the readers of the Electronic Journal of Severe Storms Meteorology (EJSSM), especially as the Warn-on-Forecast initiative continues to improve in the coming years. Overall, my comments are mostly minor-to- moderate, but I had a few issues which were slightly more major. Thus, I would like to see the paper again before full acceptance. If there are any questions on my review, feel free to email me at becoffer@ncsu.edu.

Thank you for providing such clear and constructive comments. These reviews have added tremendously to this paper.

Major comments:

The same supercell that produced the Cleburne EF3 tornado discussed herein also produced two tornadoes (EF4 & EF1) in the preceding hour and within 30 km. Can you please comment on the model forecast during this time period? Was the convective mode and timing similar to what was observed? Were any 400-m forecasts attempted for these tornadoes? Readers may be interested in this.

Several 400-m forecasts were attempted roughly during this timeframe to simulate the earlier evolution of the tornadic supercell. However, these simulations were much less successful at producing a rotating thunderstorm in the first place. In order to capture the initial thunderstorm organization and development, several experiments were run with an assimilation window running from 0000–0045 UTC using the same methodology as described in this manuscript. A possible explanation for the poor simulation skill may be related to the dearth of surface observations in the areas of interest during this timeframe across eastern portions of Erath and western Hood counties, where the thunderstorm was initially maturing.

We agree that this was an oversight to not mention the earlier tornadoes in more detail and as such, have added discussion about the 0000–0045 UTC runs and their inability to produce comparatively accurate forecasts (see section 4c, 400-m forecasts).

The discussion in section 5a solely focuses on the low-level thermodynamic fields (represented by θ_e). However, on at least two occasions, the text states that the kinematic fields also are compared. Please either include similar analyses to Fig. 7 and Table 4 regarding the low- level wind field or remove the implication that differences in the kinematic fields "lead to differences in the evolution of the Cleburne supercell".

The references to the analysis of the kinematic fields have been removed in those two instances since discussion of changes to the low-level wind fields is outside the scope of this paper. We have, however, left previous references to the kinematic fields in the discussion about the WXBUGWIND experiment.

The description and presentation of the trajectory analysis in Section 5b (2) needs more work:

a. Interpolation: How were the parcels integrated backward? What was the interpolation order? How much time elapsed between output used for interpolation? Dahl et al. 2012 showed that backward trajectories were prone to large interpolation errors in supercells, especially in confluent flow regimes and as the output interval increased above 30 s. How many parcels were tracked? What was the grid spacing between the parcels?

Citation:

Dahl J. M. L., M. D. Parker, and L. J. Wicker, 2012: Uncertainties in trajectory calculations within near-surface mesocyclones of simulated supercells. *Mon. Wea. Rev.*, **140**, 2959–2966.

Thank you for pointing out this oversight. We have added additional details to the trajectory analysis subsection detailing the use of (1) a 4th order Runge-Kutta scheme using (2) all available data, which was output at 30-s intervals with (3) 72 parcels in a circular ring of 300-m radius (every 5°).

b. [Then] Figures 14, 15: It took me a considerable amount of time to orient myself with these figures. Firstly, the figures should show an outline of the storm (e.g. the 20-dB*Z* reflectivity contour). Secondly, it makes more sense to me to exclusively focus on the trajectories originating in the low-level updraft, since the buoyancy of the updraft is the main point of this analysis. Removing the forward flank trajectories would also help simplify the figure.

These two figures have been improved by adding in a 25-dBZ reflectivity contour, which, in this case, yields a cleaner figure than the lower 20-dBZ threshold that provides too much noise unrelated to the storm. Additionally, the trajectories from the forward-flank gust front have been removed.

[Minor comments omitted.]

Second review:

Recommendation: Accept with minor revisions.

General comment: The authors have sufficiently addressed my major concerns. The goal of the paper is to provide a first attempt at assimilating non-conventional surface observations in high-resolution convection forecast. In doing so, this work reiterates the need to have improved lower tropospheric observations if the Warn-on-Forecast initiative is to be feasible. Upon rereading the manuscript, I have recommended a few more minor tweaks, but I believe the manuscript is in publishable form, pending these relatively minor revisions.

[Minor comments omitted.]