UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ASSESSING THE IMPACT OF NON-CONVENTIONAL OBSERVATIONS ON HIGH-RESOLUTION ANALYSES AND FORECASTS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN METEOROLOGY

By

MATTHEW THOMAS MORRIS
Norman, Oklahoma
2017

ASSESSING THE IMPACT OF NON-CONVENTIONAL OBSERVATIONS ON HIGH-RESOLUTION ANALYSES AND FORECASTS

A THESIS APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

_____
Dr. Frederick Carr, Chair

_____
Dr. Keith Brewster

_____
Dr. Xuguang Wang

Acknowledgements

I would like to thank my advisors, Dr. Frederick Carr and Dr. Keith Brewster, for their guidance with this project; their knowledge and expertise were a valuable resource. Dr. Xuguang Wang provided useful guidance as a member of my thesis committee. I also would like to thank Jonathan Labriola, a Ph.D. student in the School of Meteorology, for his assistance with the hail verification portion of this research. I would also like to thank Nicholas Gasperoni, Andrew Moore, and Andrew Osborne for providing feedback and guidance throughout the research process. I would also like to thank Eric Hewitt and Nicole Homeier of Understory Weather for their assistance using the data from their network in the DFW Testbed. Experiments were performed using supercomputing resources from the OU Supercomputing Center for Education and Research (OSCER). I would also like to thank the OSCER staff members for providing assistance when I faced difficulties using the supercomputing resources.

Table of Contents

List of Tables

List of Figures

Abstract

A key recommendation of a 2009 report by the National Research Council (NRC) was for new mesoscale networks to be integrated with existing ones to form a nationwide "network of networks". This recommendation originated in response to noted deficiencies in the U.S. mesoscale observing network. The report also recommended that research testbeds be established, such as the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) DFW Urban Demonstration Network, to ascertain the potential benefit of proposed observing systems.

In this work, non-conventional surface observations from Global Science & Technology (GST) Mobile Platform Environmental Data (MoPED), WeatherBug, Citizen Weather Observer Program (CWOP), and Understory Weather in the DFW Testbed are considered. Radar data include Terminal Doppler Weather Radars (TDWRs) and CASA X-band radars. The Advanced Regional Prediction System (ARPS) model is used to perform observing system experiments (OSEs) that are designed to assess the impact of the aforementioned networks. The three-dimensional variational (3DVAR) analysis system is used, along with the complex cloud analysis, to produce analysis increments every 10 minutes, which are then applied to the model forecast using incremental analysis updating (IAU). Experiments are performed on a supercell thunderstorm that impacted the DFW metroplex on 11 April 2016 with large, damaging hail. The analysis includes qualitative and quantitative comparisons of the forecast reflectivity fields, quantitative comparisons of model-derived hail with radar-observed hail, and surface-level verification of the temperature and dew point fields. The CASA radial velocity data offer positive benefit to the forecasted storm structure as noted in the

simulated reflectivity, along with model-derived hail. However, the data appear to be detrimental when considering quantitative comparisons of the simulated reflectivity with observations. The inclusion of dew point temperature measurements from the non-conventional CWOP and WeatherBug networks resulted in a degradation in the forecasted dew point field. The analysis concludes with a brief comparison of the results for single-moment versus double-moment microphysics scheme sensitivity. Future work should assess the impacts of the non-conventional observations on a wider array of cases.

Chapter 1

1.1 A Brief History of Numerical Weather Prediction

In 1904, Norwegian meteorologist Vilhelm Bjerknes described the problem of numerical weather prediction (NWP; Bjerknes 1904). With an accurate initial depiction of the atmosphere (i.e., initial conditions), along with the corresponding boundary conditions, one should, in theory, be able to predict the future state of the atmosphere by integrating the equations of motion forward in time.

Almost two decades later in 1922, Lewis Fry Richardson proposed numerical integration as a means of forecasting the future state of the atmosphere (Richardson 1922). Integrating the primitive equations of motion by hand, Richardson predicted an inordinately large 6-hour pressure tendency of 146 hPa, a value that is unobservable in the real atmosphere. Despite the apparent failure, Richardson's work provided the first evidence of the importance of accurately sampling the initial state of the atmosphere. The wind and pressure were out of balance owing to a scarcity of upper-air observations at the time; as such, the meteorological signal was largely masked by gravity waves attempting to restore geostrophic balance (Lynch 2008).

The combination of high-performance computing capabilities and increased surface and upper-air observations revived interest in NWP during the late 1940s (Kalnay 2003). Beginning in the 1950s, operational model forecasts have been produced by the National Center for Environmental Prediction (NCEP; formerly the National Meteorological Center, or NMC), with these forecasts becoming global in 1973. As model resolution and computing capabilities have continued to improve, the simulation of mesoscale features such as thunderstorms has become an area of research focus (Lilly

1990). However, it is widely recognized that, in order for these advances in numerical models and computing systems to be fully realized, there must be corresponding improvements in observations.

1.2 Research Motivation

In 2003, the United States Weather Research Program (USWRP) organized a workshop to discuss ways of alleviating deficiencies in the current observational network (Dabberdt et al. 2005). Although forecast skill has improved over time with improved model resolution, the full potential of advances in numerical modeling has not been realized. High spatiotemporal resolution mesoscale observations, in concert with improved data assimilation techniques and parameterization schemes, have the ability to improve forecasts of wind and precipitation. Mesoscale phenomena, such as frontal boundaries and mountain flows, along with planetary boundary layer (PBL) structures, are particularly difficult to analyze and predict with the current observational network.

The primary recommendation of the workshop was to establish a nationwide network of mesoscale surface stations that collect observations at a higher spatiotemporal resolution. These stations would complement the existing observational network by providing additional data in the lowest levels of the atmosphere where the greatest observational need exists. The committee recommended that these mesoscale surface observations be collected at least every 5 minutes and have an average station separation distance of 25 km in flat terrain. The average station separation distance should be reduced to roughly 10 km in areas of greater observational need, such as in coastal, mountainous, or urban areas.

The current Weather Surveillance Radar-1988 Doppler (WSR-88D; Crum and Alberty 1993) network of S-band (10-cm wavelength) radars is unable to observe roughly 70% of the PBL, missing important low-level features such as convective outflows and mesoscale cyclones and anticyclones. This deficiency could be remedied by integrating additional radars into the WSR-88D network, such as Terminal Doppler Weather Radars (TDWRs) and privately-owned radars operated by television stations. Furthermore, low-power, short-range radars could be strategically placed to fill in the gaps of the WSR-88D network and improve observational coverage (Dabberdt et al. 2005). This concept has been demonstrated by the Collaborative Adaptive Sensing of the Atmosphere (CASA) consortium, which installed a testbed of four X-band (3-cm wavelength) radars in southwest Oklahoma in 2006 (McLaughlin et al. 2009).

A 2009 report by the National Research Council entitled *Observing Weather and Climate from the Ground up: A Nationwide Network of Networks* expanded upon the findings of the 2003 workshop (National Research Council 2009). The report noted that while the United States has a respectable synoptic-scale observing network, the quantity, quality, and accessibility of mesoscale observations varies considerably, with a rather poor network of three-dimensional observations. The report proposed that existing and new mesoscale networks be integrated to form a nationwide "network of networks" in order to maximize the observational benefit of the disparate networks. These networks should include comprehensive metadata in order to maximize the value of the observations; in fact, it is recommended that complete metadata be a requirement for membership in the network of networks. The integration process should include collaboration from academic, public, and private partners, with the federal government

acting as the central authority and overseeing the resulting network. Some examples of high-priority observations include tropospheric profiles of temperature and moisture and PBL structure. An additional recommendation of the report is that the United States Department of Transportation should oversee the future deployment of high-density mobile observations, such as temperature and rain rate (from wiper speed) collected by fleets of commercial vehicles.

Testbeds have been recommended as a means of collaboration between federal, private, and academic partners (Dabberdt et al. 2005 and National Research Council 2009). These testbeds should be established in regions that present operational challenges, such as urban areas and mountainous regions, with the goal being to objectively assess the future benefit of proposed observing systems (National Research Council 2009). As detailed in Section 1.4, observing system experiments (OSEs) and observing system simulation experiments (OSSEs) can be used to objectively demonstrate whether a specific set of observations within the testbed helps improve forecast skill.

1.3 DFW Urban Demonstration Network

One such testbed has been established in the Dallas-Fort Worth (DFW) metroplex, known as the DFW Urban Demonstration Network (National Research Council 2012). The testbed is being managed by the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) and represents a joint endeavor among academic institutions, private companies, local governments, and the National Weather Service (NWS) forecast office in Fort Worth. The Dallas-Fort Worth metroplex was chosen as a suitable location for the demonstration testbed as it is a large urban area with a population in excess of 6

million people that has two major airports and several highly-traveled interstate highways. Most importantly, the region experiences a wide array of significant weather impacts, such as tornadoes, severe wind, and localized flash flooding.

A network of eight closely-spaced X-band radars is proposed to supplement the existing WSR-88D radar (KFWS) in Fort Worth by providing increased low-level coverage in regions that are poorly observed by the existing KFWS radar (McLaughlin et al. 2009). Currently, seven out of the eight planned radars have been deployed. The radar coverage in the testbed will be further supplemented by the inclusion of the TDWRs at the aforementioned passenger airports. Other data sources include satellites, radiosondes, aircraft data (e.g., take-off and landing soundings), SODARs, and various conventional and non-conventional surface observation networks. More details on the individual data sources can be found in the following chapter.

1.3.1 CASA X-band Radars

CASA, an NSF Engineering Research Center (ERC), developed and deployed a testbed of four densely-spaced X-band radars in southwest Oklahoma in 2006, known as Integrated Project One (IP1; McLaughlin et al. 2009). This region was chosen as it experiences numerous severe thunderstorms and tornadoes annually and is located roughly halfway between the existing Oklahoma City (KTLX) and Frederick (KFDR) S-band radars, resulting in poor coverage in the lowest levels of the atmosphere (Brewster et al. 2005b). Research studies that have incorporated the CASA radar data from the IP1 deployment into storm-resolving numerical models have shown positive forecast impact (Brewster et al. 2007; Schenkman et al. 2011a, b; and Snook et al. 2012). Given these promising results, a network of eight similar radars is being deployed within the DFW

Urban Demonstration Network. With continued promising results, further expansions of the X-band radar technology are certainly within the realm of possibility. These radars could either be deployed nationwide or in specified regions where there are increased observational needs, such as mountainous regions or urban areas. Roughly 10,000 radars would be required to maintain the current 30-km average separation distance in a nationwide deployment. More details on the CASA X-band radars can be found in section 2.3.

1.4 Observing System Experiments

According to Dabberdt et al. (2005), the decision-making process related to the nationwide "network of networks" should include atmospheric models, as models have the ability to quantify the greatest observational needs for analysis and prediction applications. Moreover, models are able to determine the minimum spacing and resolution requirements for NWP applications, which is important in maintaining economic viability of future nationwide observing systems. Historically, these goals have been accomplished using either observing system experiments (OSEs) or observing system simulation experiments (OSSEs). In an OSSE, the impact of proposed observing systems can be ascertained by using simulated observations. First, a high-resolution NWP model is used to generate a "nature run," which acts as the assumed atmospheric "truth" (Atlas 1997). Simulated observations are then created from this "nature run" by, for instance, interpolating the nature run values to the observation locations. Numerical simulations using these synthetic observations are then compared to the nature run to determine the impact the proposed observations would have on numerical simulations if the system were implemented (e.g., Arnold and Dey 1986; Lord et al. 1997; Atlas 1997).

On the other hand, in an OSE, the impact of currently-deployed observing systems can be determined. Traditionally, in an OSE, an analysis and resulting forecast are computed for a control experiment in which all available real-data are assimilated. The control forecast can then be compared to data denial experiments, in which observations from a particular class of observations are denied (e.g., all aircraft observations or observations from a particular sensor type), to determine the magnitude of resulting improvements (or degradations) in the analyses and forecasts attributable to the denied dataset. It is important to note that OSEs may reveal negligible or negative value of particular observational datasets. For example, McNally et al. (2014) evaluated the impact of geostationary and polar-orbiting satellite data in the forecasted track of Hurricane Sandy, which made a sharp left turn into the coast of New Jersey in October 2012. This left hook was correctly predicted by the European Centre for Medium-Range Weather Forecasts (ECMWF) well before it was forecasted by other operational centers. The authors found that the denial of geostationary satellite data did not significantly degrade the forecasted turn; however, polar-orbiting satellite data was shown to have a more significant role in capturing the left turn in this event.

Coincident with improvements in numerical models, computational power, and data assimilation systems, increased research has been focused on determining observation impact using OSEs. These studies have considered the impact of sounding and profiler data (e.g., Graham et al. 2000; Benjamin et al. 2010; Agustí-Panareda et al. 2010), GPS-derived precipitable water (e.g., Smith et al. 2007; Benjamin et al. 2010), aircraft data (e.g., Benjamin et al. 2010), satellite radiances and satellite derived winds

(Bouttier and Kelly 2001; Zapotocny et al. 2002, 2005, 2007; Kazumori et al. 2008; Bi et al. 2011), and radar radial winds and reflectivity (Schenkman et al. 2011a,b).

Despite this upturn in OSE-related research, determining observation impact using OSEs can prove to be both time-consuming and computationally expensive, due to a large number of experiments that must be run in order to test the denial of numerous combinations of observations. In recent years, a new diagnostic tool has emerged to overcome these issues, known as Forecast Sensitivity to Observation (FSO; Cardinali 2009). In this adjoint approach, the observation impact is determined using a single experiment in which all observational data are assimilated using a four-dimensional variational (4DVAR) analysis system.

1.4.1 OSEs using Radar Data

Early efforts to determine the impact of radar data on high-resolution analyses and forecasts of convection began at the University of Oklahoma during the late 1980s and early 1990s with the inception of the Center for Analysis and Prediction of Storms (CAPS; Lilly 1990). These efforts coincided in large part with the nationwide deployment of the WSR-88D network (Crum and Alberty 1993). Assimilation of Doppler radar data is crucial for modeling ongoing thunderstorms, as Doppler radar is the only system capable of observing convective storms with the requisite spatiotemporal resolution.

It is theorized that the dense network of X-band CASA radars in the DFW Urban Demonstration Network will better observe the lowest levels of the atmosphere, filling in observation gaps in the widely-spaced WSR-88D radar network. Several studies have looked at the impact of CASA radar data from the IP1 deployment in southwest

Oklahoma during the spring of 2007. In Schenkman et al. (2011a), a tornadic mesoscale convective system (MCS) and associated line-end vortex (LEV) are simulated using the Advanced Regional Prediction System (ARPS) model. Reflectivity and radial velocity data from the WSR-88D and CASA IP1 networks are assimilated, with the ARPS 3DVAR and complex cloud analysis (Brewster et al. 2005c; Hu et al. 2006a) using these data to adjust the cloud and hydrometeor fields, along with in-cloud temperature to account for latent heating. When CASA radar data were assimilated alongside WSR-88D data, the squall line structure is improved at the end of the data assimilation window, resulting in an improved simulation. The radial velocity data from CASA were particularly important in accurately analyzing the gust front. In a closely-related study, Schenkman et al. (2011b) examined the influence of CASA radial velocity data on the prediction of tornadic mesovortices. Experiments in which low-level radial velocity data were assimilated yielded the most accurate forecast evolution, owing to improved depictions of the low-level shear profile and cold pool development. Snook et al. (2012) found that assimilating CASA and WSR-88D radar data into a forecast ensemble resulted in improved probabilistic forecasts of mesovortices, which serve as a proxy of tornado potential. Stratman and Brewster (2015) examined the influence of assimilating CASA radar data for a cluster of supercell thunderstorms on 24 May 2011 using diverse microphysics parameterization schemes. It was found that the CASA data afforded little, if any, value for a storm located outside the radar coverage area. The value added for storms inside the radar coverage area was less clear, perhaps due to the complex interactions with neighboring supercells.

Dawson and Xue (2006) demonstrated that forecasts of a strong, bow-shaped MCS most closely matched the observed system when the complex cloud analysis package was used, thus resulting in the elimination of the 2 to 3 hour model "spin-up" time. The authors also found that the use of intermittent assimilation cycles was beneficial. Hu et al. (2006a) were able to successfully reproduce a tornadic thunderstorm in the Fort Worth area, including reductions in timing and location errors, when the complex cloud analysis procedure was used in conjunction with radar reflectivity data. In addition, the model "spin-up" time was reduced with the usage of the cloud analysis package. Hu et al. (2006b) found additional forecast improvements with the assimilation of radial velocity data using the ARPS three-dimensional variational (3DVAR) data assimilation system, although a larger improvement was found with the addition of clouds and latent heat. Zhao and Xue (2009) also used the ARPS cloud analysis package to examine the impacts of reflectivity and radial velocity data from coastal WSR-88D radars on the forecasted track and intensity of landfalling Hurricane Ike in 2008. The assimilation of radial velocity data was found to be most impactful for improving the track forecast, while reflectivity data was most useful for improving the intensity forecast. Xiao and Sun (2007) found that assimilation of multiple-Doppler data resulted in improved simulations of a squall line, owing to a better initial depiction of a cold pool. Moreover, radial velocity data afforded the most benefit to wind and vertical velocity analyses, whereas radar reflectivity was most beneficial in improving hydrometeor analyses. The authors also note that cycling of the Doppler radar data results in a better analysis than when radar data is assimilated just once. Xue et al. (2013) found that assimilating radial velocity and reflectivity data from the WSR-88D network yielded a

positive impact on forecasts of convection throughout a domain covering a majority of the continental United States for a period of at least 24 hours.

It is also worth noting that improved numerical simulations of convective storms are a cornerstone of the proposed "warn-on-forecast" paradigm (Stensrud et al. 2009; Stensrud et al. 2013). In the "warn-on-forecast" paradigm, observations of convective storms and their ambient environment are assimilated into an ensemble of convection-allowing models, providing NWS forecasters and end-users with probabilistic forecast information concerning storm evolution. This information could then result in increased lead times for severe thunderstorm, tornado, and flash flood warnings, furthering the NWS mission of protecting life and property. In order for "warn-on-forecasting" to become a reality, the forecast model must accurately depict and support ongoing precipitation in the short-term forecasts. As such, continued improvements in the assimilation of radar data and other mesoscale data are necessary. To that end, this research will examine the impacts of auxiliary CASA and TDWR radar data for a case study in the DFW Urban Demonstration Network to determine if the additional data affords improved analyses and forecasts.

1.4.2 OSEs using Surface Data

As described in the previous section, numerous studies have demonstrated the value of radar data in generating useful forecasts of ongoing convection. Despite this, the use of radar data in forecasting convective initiation is fundamentally limited in that radar data only provides precipitation information and radial velocity data (i.e., no thermodynamic information is directly provided by the radar). Moreover, operationally available radar data are unable to fully observe the lowest levels of the atmosphere due

to their spacing and the Earth's curvature. In response to this limitation, increased research focus has been placed on understanding the benefit of surface observations in relation to NWP analyses and forecasts.

Mukhopadhyay et al. (2005) utilized the Regional Atmospheric Modeling System (RAMS) to study the effect of surface observations on model forecasts of three monsoon low-pressure systems in the vicinity of India. Surface data inclusion resulted in an improved forecast of heavy rainfall throughout the region, when equitable threat score (ETS) and bias are used as forecast verification metrics. Additionally, the authors noted that the surface data appear to be particularly beneficial due to the highly-varied terrain in the region, with surface observations identifying differential heating effects. More specifically, regions of enhanced surface heating experienced a corresponding mass response of convergence and upward motion, which resulted in improved precipitation forecasts. This study underscores the need for increased density of observations in regions of complex terrain.

Alapaty et al. (2001) developed a continuous surface data assimilation technique and found that this new technique consistently improved boundary layer structure. Interestingly, the authors noted that surface data yielded the most benefit when assimilated alongside upper-level radiosonde data. In essence, the full potential of the surface data would not be realized without the additional upper-air data. One goal of this research is to identify similar relationships among observational datasets in the DFW Urban Demonstration Network in order to maximize observation benefit. In a similar study, Ha and Snyder (2014) assimilated surface observations using the Ensemble Kalman Filter (EnKF) and found improvements in subsequent Weather Research and

Forecasting (WRF) model forecasts of a squall line. Not only did the surface observations result in a better representation of the boundary layer structure, they improved horizontal gradients of both temperature and moisture that would prove crucial in properly forecasting thunderstorm development.

Knopfmeier and Stensrud (2013) compared surface analyses generated using EnKF to those produced by the NCEP's Real-Time Mesoscale Analysis (RTMA). Surface mesonet data were assimilated in the EnKF analyses, which overall were fairly similar to the RTMA analyses, albeit with a somewhat smoother appearance. Most notably, denying up to 75% of the mesonet data resulted in only minor differences in the analyses. The authors speculate that this result is attributable to background error covariance scales that are significantly larger than the average station separation distance, thus allowing for enhanced observational increment spreading throughout the domain. This research demonstrates the value in determining optimal observational density, as suggested by Dabberdt et al. (2005).

Until recently, most studies focused on the impact of observational data have been centered on conventional observations, with only limited research focused on non-conventional datasets. Tyndall and Horel (2013) considered the impact of nearly 20,000 surface observations and found that observation impact was largely dependent on the observation location. For instance, observations located in metropolitan areas with widespread observations tended to have lower observational impact than observations in more remote locations with fewer overall observations. In addition, high-impact observations tended to be found in regions with more local variability, such as coastal regions. It is important to note that this study did not consider the impacts of observing

systems individually. Perhaps the first study to do so was Hilliker et al. (2010), which considered the impact of surface observations from Automated Weather Services, Inc. (AWS). These observations, more commonly known as "WeatherBug," are commonly taken from the tops of buildings such as schools. The authors found improvements to National Digital Forecast Database (NDFD) forecasts of temperature and dew point. However, the observations offered limited improvements for wind speed forecasts, perhaps as a result of biases in wind speed measurements due to siting concerns (sheltering by nearby buildings and trees). NDFD forecasts use numerical model output as the starting point, but are modified by forecasters at the NWS Weather Forecast Offices (WFOs) to generate the final forecast.

More recently, Carlaw et al. (2015) examined the impact of several non-conventional data sources on ARPS forecasts of a tornadic supercell that impacted Cleburne, TX on 15 May 2013. The non-conventional surface data sources used included AWS WeatherBug, Citizen Weather Observer Program (CWOP), and Global Science and Technology (GST) Mobile Platform Environmental Data (MoPED). Given Cleburne's location in the southwestern fringe of the DFW metroplex, it is poorly observed by conventional surface observing systems. Thermodynamic measurements from the WeatherBug stations were able to capture increased levels of moisture in the lowest levels, as compared to the model background field. Enhanced instability due to the increased humidity resulted in increases in both updraft velocity and vertical vorticity in the resultant storm, and thus produced simulated storms more closely matching observations. Carlaw (2014) examined the impacts of the aforementioned non-conventional observations on hourly analyses for a month-long period in March 2014.

Dew point errors were reduced owing to the non-conventional observations, when averaged over the entire month. Additionally, separate cases of a dryline and cold front were examined, with improvements to the analyzed boundaries in both cases. The non-conventional observations degraded the wind speed analysis, likely due to the siting issues noted in Hilliker et al. (2010).

In the framework of the nationwide "network of networks," it is important to understand the advantages and disadvantages of various data sources in a research testbed prior to a potential nationwide deployment. The purpose of this research is to perform OSEs to determine the value of several new observing systems within the DFW Urban Demonstration Network, including non-conventional surface data and CASA radar data. Chapter 2 will describe the observational datasets used in this study, including both conventional and non-conventional observations, along with pre-processing and quality control procedures applied to these datasets. Chapter 3 will detail the ARPS model and its associated 3DVAR analysis system that are used for simulations presented in this work. Chapter 4 will present the results of OSEs performed for a high-impact hail case in the Dallas-Fort Worth metroplex on 11 April 2016. Finally, concluding remarks will be presented in Chapter 5, including implications of this work to the "network of networks" vision and suggestions for future research avenues.

Chapter 2

2.1 Conventional Observations

In this work, conventional surface data sources refer to those that are available in the federal observing network and assimilated into operational forecast models. The Automated Surface Observing System (ASOS) is one of these conventional networks, and serves as the main surface observing network in the United States (NWS 1999). The ASOS network represents a joint venture between the Federal Aviation Administration (FAA), National Weather Service (NWS), and Department of Defense (DoD), and provides valuable information for forecasting and aviation applications. The ASOS network was developed in the 1980s and deployed in the 1990s. The Automated Weather Observing System (AWOS) is a closely-related automated network that is operated by the FAA and reports data every 20 minutes at secondary airports. Sensors for both automated networks are positioned in a region that provides a representative observation for the entire airport complex (i.e., within 2 to 3 miles of the sensor location); for most airports, this location is near the touchdown zone of the main runway (NWS 1999). ASOS stations are monitored by the ASOS Operations and Monitoring Center (AOMC), with site maintenance performed by NWS technicians, as needed. Mesoscale networks (or "mesonets") provide additional surface observations throughout the domain, and include the Oklahoma Mesonet (Brock et al. 1995; McPherson et al. 2007) and the West Texas Mesonet (Schroeder et al. 2005). In the case of the Oklahoma Mesonet, stations are sited so as ensure that the physical characteristics of a site are representative of the surrounding area (e.g., minimal terrain slope, minimal obstructions that preclude proper ventilation, and minimal influences from urban areas, forests, and bodies of water).

Sensors are calibrated prior to deployment and replaced at regular intervals, while observations are subject to automated and manual quality assurance techniques. Additional information on the Oklahoma Mesonet can be found in McPherson et al. (2007). Upper-air observations are obtained from multiple sources, including the Meteorological Data Collection and Reporting System (MDCRS), which provides observations of flight-level temperature, dew point, and wind for assimilation in forecast models. NWS radiosonde data are not available for the time period considered for this study, generally being available at 0000 and 1200 UTC, only. Data from eight radars in the WSR-88D network fall within the domain used for the 11 April 2016 case study, with the most notable being the KFWS radar in Fort Worth, Texas. Additional WSR-88D radars assimilated include Dyess Air Force Base, TX (KDYX), Frederick, OK (KFDR), Ft. Hood, TX (KGRK), Ft. Polk, LA (KPOE), Shreveport, LA (KSHV), Fort Smith, AR (KSRX), and Oklahoma City, OK (KTLX). Lastly, visible and infrared data from the Geostationary Operational Environmental Satellite (GOES) are incorporated in the complex cloud analysis, which is described in detail in Section 3.2.2.

2.2 Non-Conventional Observations

With the inception of the National Mesonet Pilot Program, Global Science & Technology (GST) was selected to develop a new system known as Mobile Platform Environmental Data (MoPED), collecting observations from sensors developed by Weather Telematics and mounted on mobile fleets of trucks and other transportation vehicles (Dahlia 2013). Since its beginning, MoPED has rapidly grown to provide more than two million observations daily, with the majority of observations originating from a fleet of over 1500 Con-way freight trucks. These trucks have been fitted with sensors

that measure such variables as temperature, humidity, pressure, and precipitation. One limitation of the MoPED system is that wind cannot be measured due to contamination from the motion of the vehicles. Pressure measurements are corrected to account for these effects. Since active vehicles collect data roughly every ten seconds, a data thinning algorithm has been developed to reduce the spatiotemporal resolution of these data. Observations are grouped together based upon truck identifier and thinned once either a five-minute time threshold or a one kilometer distance threshold are met. Appendix A presents the results of statistical testing for various averaging schemes that only considered the effects of time, with these results helping motivate the final thinning methodology.

The Automatic Position Reporting System as a Weather Network (APRSWXNET) represents an additional non-conventional data source in the DFW Urban Demonstration Network (CWOP 2014). Electronic weather stations owned by ham radio operators and private citizens collect weather observations, which are ingested into the Meteorological Assimilation Data Ingest System (MADIS). These observations are subjected to the MADIS Quality Control and Monitoring System (QCMS), which performs a variety of quality control checks. This data source is more commonly known as the Citizen Weather Observer Program (CWOP).

A third non-conventional data source is the Automated Weather Services (AWS) Convergence Technologies, Inc. WeatherBug network, which is operated by Earth Networks. There are roughly 8,000 WeatherBug weather stations throughout the United States, the majority of which are located atop schools and public buildings. These data are used daily in local weather broadcasts owing to partnerships with over 100 television

stations. Additionally, educational tools utilizing these data are made available to K-12 school children. The expansive network of WeatherBug observations offers the potential for improvements to NWP forecasts, which has been demonstrated in Hilliker et al. (2010) and Carlaw et al. (2015).

A recently deployed network of solar-powered weather stations from Understory Weather represents the fourth non-conventional data source. This dense network was originally focused on the immediate Dallas area, with ten stations deployed at the time of the 11 April 2016 case study. Since this time, the number of deployed stations has increased considerably, including increases in spatial coverage, with the full network of about 120 stations largely in place by April 2017. Temperature, pressure and humidity variables are measured using standard sensors, while wind, rain, and hail impacts are calculated based upon the forces acting on a metallic ball, or sonde. While these data are mainly intended for insurance companies responding to weather-related insurance claims, they could offer improvements to NWP forecast models.

It is important to note that these non-conventional surface data sources are not subjected to the same siting standards as the ASOS and AWOS networks. Thus, these data may be subject to bias and representativeness errors, such as the low wind speed bias noted previously by Hilliker et al. (2010) and Carlaw et al. (2015).

Since the NWS radiosonde network typically only samples the vertical profiles of wind, temperature, and dew point at 12 hour intervals, additional instruments can be used to provide more continuous coverage of vertical profiles. For instance, two SecondWind (now part of Vaisala) SODARs (SOnic Detection And Ranging) have been installed by WeatherFlow in the DFW Urban Demonstration Network to fill in these temporal gaps,

one at the Fort Worth NWS forecast office and the other in Midlothian. Wind speed and direction in the PBL (up to about 2 km) are derived from these ground-based remote sensing instruments by measuring the Doppler shift of acoustic sinusoidal pulses that are scattered back by turbulence resulting from the atmosphere's thermodynamic structure (Lang and McKeogh 2011). Figure 2.1 displays a typical distribution of both conventional and non-conventional surface data sources, along with the locations of two SODARS, in the DFW Urban Demonstration Network. The impact of the non-conventional surface observations on analyses and forecasts will be examined in this study.

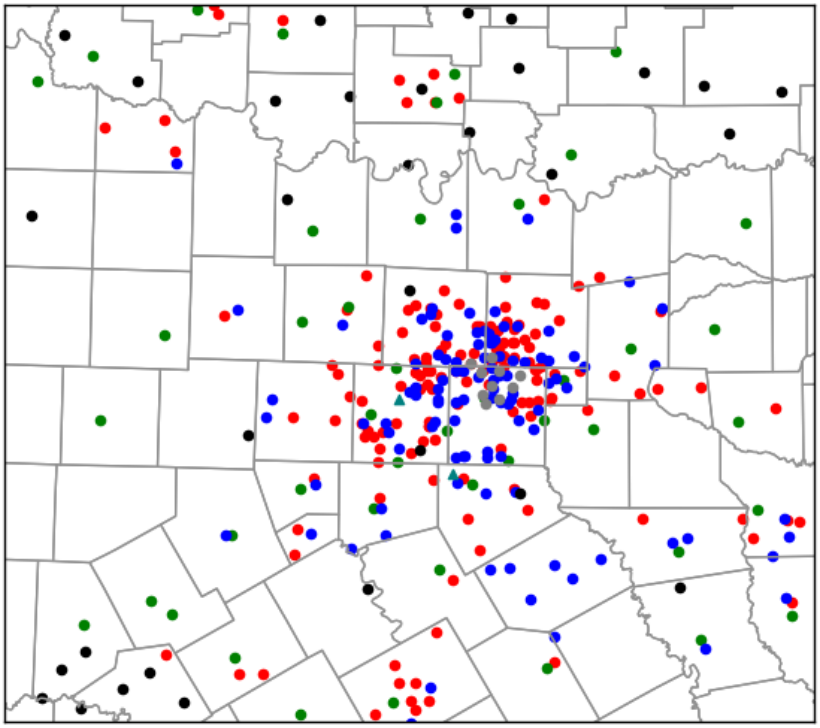


Figure 2.1: Spatial distribution of the conventional and non-conventional surface data assimilated at the first analysis time (2150 UTC). Observations shown include CWOP (red – 148), METAR (green – 44), WeatherBug (blue – 105), Understory (gray – 10), mesonet (black – 32), and SODAR (teal triangles – 2).

2.3 Non-Conventional Radar Data

The CASA Integrated Project One (IP1) testbed from the spring of 2007 consisted of four dual-polarization X-band Doppler radars located in southwest Oklahoma, a region susceptible to severe thunderstorms (McLaughlin et al. 2009). The radars were spaced 30 km apart, on average, and have a range of 40 km. These radars have a wavelength of 3.2 cm, requiring an antenna size of roughly 1 m, significantly smaller than the 8.5 m antennas required for the 10-cm WSR-88D radars. This allows the radar antennas to be placed on existing infrastructure, such as cell towers and buildings. The short wavelength, however, makes these X-band radars susceptible to attenuation in regions where the radar reflectivity factor exceeds 40 dBZ (Brewster et al. 2005a). Thus, these radar networks are designed to provide overlapping radar coverage, whenever possible (Brewster et al. 2005b). In recent years, the CASA testbed has been relocated to the Dallas-Fort Worth metroplex, with radars located in Addison, Arlington, Denton, Midlothian, Fort Worth, and Johnson County at the time of the 11 April 2016 case study. Since this time, an additional radar has been deployed in Mesquite, with a further radar planned for McKinney (Brewster et al. 2017). This network is comprised of the four radars from the original IP1 network, along with additional radars from EWR Weather Radar, Ridgeline Instruments, Furuno, and Enterprise Electronics Corporation. This network is the result of a multisector partnership between CASA and the North Central Texas Council of Governments (NCTCOG; Bajaj and Philips 2012). The beam width for each of the CASA X-band radars used in this work is shown in Table 2.1.

Table 2.1: Beam Width for CASA X-band Radars

Radar	**Beam Width**
Addison (XADD)	2.3 degrees
Arlington (XUTA)	1.8 degrees
Denton (XUNT)	2.7 degrees
Fort Worth (XFTW)	1.8 degrees
Johnson County (XJCO)	1.8 degrees
Midlothian (XMDL)	1.4 degrees

One of the more notable features of the CASA IP1 radar network is the ability to scan the atmosphere both collaboratively and adaptively. Collaborative sensing occurs when the radar control architecture from multiple radars coordinate with one another to observe the same volume simultaneously, which allows for radar-based detection algorithms such as multiple-Doppler wind retrievals. Scanning strategies of the radars can also be modified by the radar control architecture based on the current highest priority observational needs, referred to as adaptive sensing. Together, these features allow the radars to provide improved horizontal resolution and faster update times. One example of a meteorological phenomenon in which these adaptive scanning strategies would prove useful is a supercell thunderstorm with rapidly evolving low-level rotation. Rapidly forming and dissipating tornadic signatures could be observed by the X-band radars, but be missed if they occurred between scans of the WSR-88D or below the lowest elevation scan in the low-level data coverage gap. To date, these collaborative adaptive scanning strategies have not been implemented in the Dallas-Fort Worth testbed. Rather, the radars follow a traditional "sit-and-spin" scanning strategy, with pre-determined scanning

angles (see Figure 2.2) concentrating on low-level scans. Despite this, these radars afford improved spatial coverage and faster update times than the existing WSR-88D network.

In addition, two Terminal Doppler Weather Radars (TDWRs) are available from the two major passenger airports in the DFW metroplex (Istok et al. 2008). These C-band (5-cm wavelength) radars are operated by the FAA. There are 45 TDWRs operational at selected airports, with these radars mainly designed for the detection of precipitation and hazardous wind shear near airports. Figure 2.3 displays the spatial distribution of the WSR-88D radars used in this study, while Figure 2.4 displays the spatial distribution of the two TDWR and six CASA radars deployed as of 11 April 2016, with the locations of the as yet deployed McKinney and Mesquite radars shown, as well.

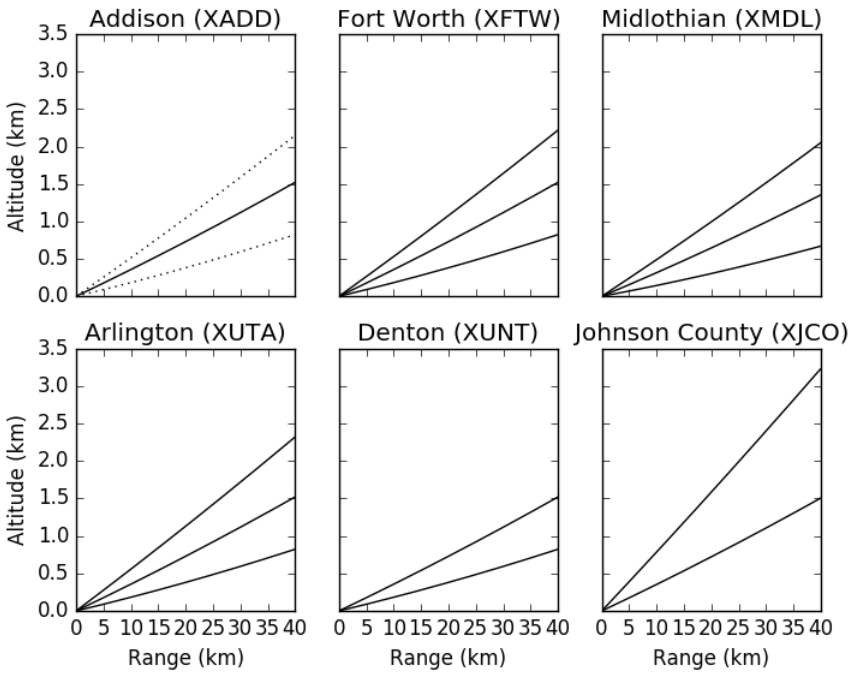


Figure 2.2: Radar beam heights vs. range for the six CASA X-band radars used in this study. Beam spreading is illustrated in the upper left panel for the Addison radar, using a representative beam width of 1.8 degrees.

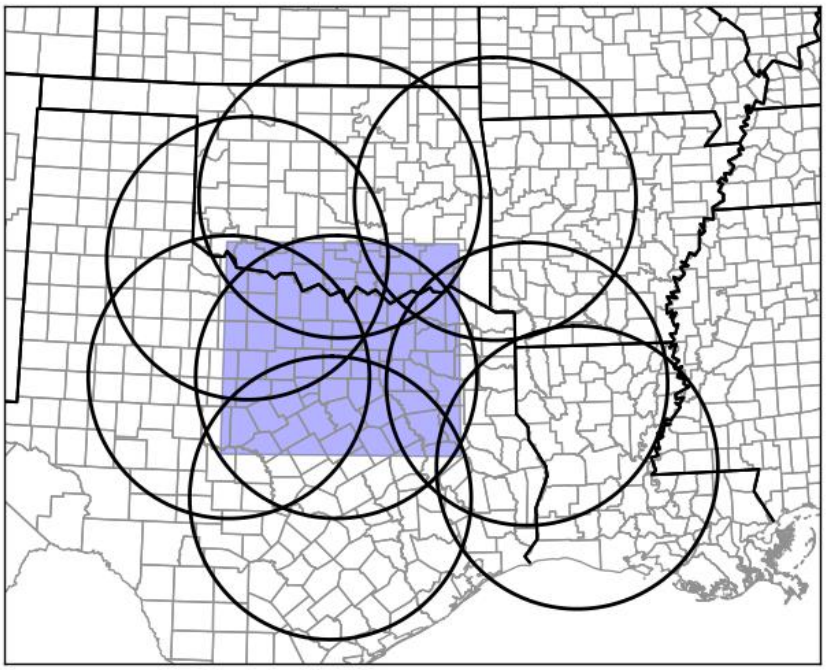


Figure 2.3: Locations of the 8 WSR-88D radars whose data are used in this work. The blue shaded region represents the model domain used.

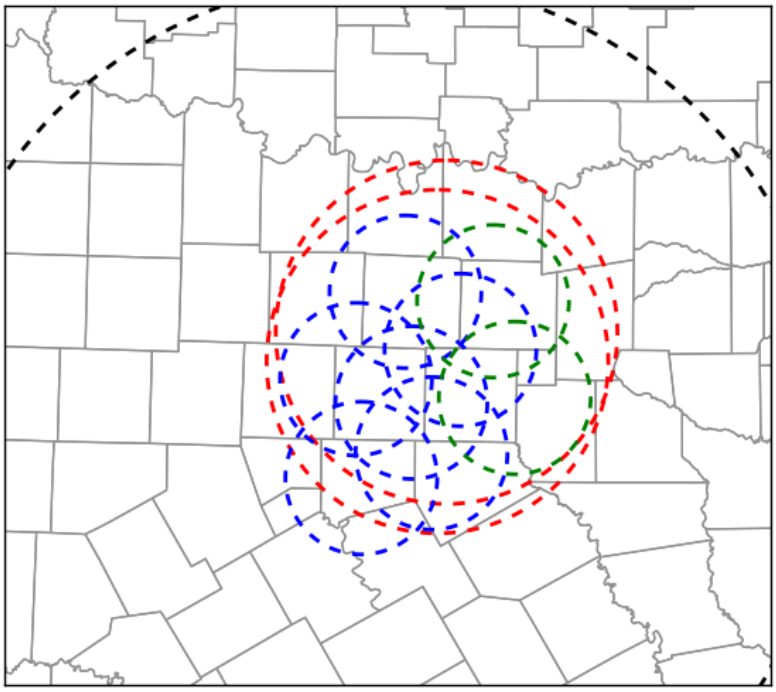


Figure 2.4: Locations of the radars used in this study. CASA X-band range rings are shown in blue (active for this case study) and green (proposed), TDWR range rings are shown in red, and the WSR-88D KFWS range ring is in black. The range rings for seven additional WSR-88D radars are not shown.

2.4 Quality Control Procedures

Observations assimilated in this work are subject to several quality control procedures. Observations acquired from MADIS undergo internal quality control checks, the details of which are outlined in the NWS Techniques Specification Package (NWS 1994). Radar data are also subject to quality control procedures in the ARPS radar remapping program, which is described in the following chapter.

Furthermore, the ARPS 3DVAR analysis program also employs several quality control checks to remove inaccurate observations. Observations undergo a temporal consistency check, which compares each observation to a preceding observation at the same location, typically one hour earlier. When the difference in these observations exceeds a user-defined difference threshold, the observation is deemed to be unreliable and is not assimilated. Similarly, observations are discarded when the difference between the observed value and the background value interpolated to the observation location via the forward operator exceeds a user-defined threshold. Finally, a local Barnes (Barnes 1964) analysis to each observation site is used to check for spatial consistency among nearby observations.

Chapter 3

3.1 Advanced Regional Prediction System (ARPS)

The Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma developed the first version of the Advanced Regional Prediction System (ARPS) model during the early 1990s (Xue et al. 1995, 2000, 2001). ARPS is a compressible, non-hydrostatic model with a terrain-following vertical coordinate on an Arakawa C-grid. The vertical coordinate is stretched using a hyperbolic tangent function. Simulations of tropical cyclones (Zhao and Xue 2009), MCSs (Dawson and Xue 2006) and tornadoes (Xue et al. 2014) have been performed using ARPS. The ARPS model is used to perform the OSEs presented in this research. Details on the parameterization schemes and model configurations used in these experiments can be found in Table 3.1.

Table 3.1: Model parameterizations and configurations

Microphysics	Single-Moment (Milbrandt and Yau 2005)
Radiation	NASA atmospheric radiation transfer
Planetary Boundary Layer (PBL)	1.5 order TKE (Deardorff 1980)
Advection	Fourth-order in the vertical and horizontal
Convection	Explicitly resolved
Soil Model	Two-layer diffusive soil model (Noilhan and Planton 1989)

3.2 ARPS Three-Dimensional Variational (3DVAR) Analysis System

The ARPS three-dimensional variational (3DVAR; Gao et al. 2004) analysis system produces an analysis by combining information from the background field and

observations. The analysis is found by minimizing a scalar cost function, which is given by:

$$J(x) = \frac{1}{2}(x - x_b)^{\mathrm{T}}\mathbf{B}^{-1}(x - x_b) + \frac{1}{2}(H(x) - y_o)^{\mathrm{T}}\mathbf{R}^{-1}(H(x) - y_o) + J_c \quad (3.1)$$

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

The final term in equation (3.1) is a penalty term, and represents a weak anelastic mass continuity constraint:

$$J_c = \frac{1}{2}\lambda_c D^2 \quad (3.2)$$

where D is given by:

$$D = \alpha\left(\frac{\partial\bar{\rho}u}{\partial x} + \frac{\partial\bar{\rho}v}{\partial y}\right) + \beta\left(\frac{\partial\bar{\rho}w}{\partial z}\right) \quad (3.3)$$

Here, λ_c represents a weighting coefficient for the mass continuity constraint, α and β correspond to weighting terms for the horizontal and vertical terms, respectively, and $\bar{\rho}$ is the mean air density at a given height. The anelastic mass continuity constraint acts to derive non-radial wind information from the observed radial velocities (Gao et al. 2004; Hu et al. 2006b). It is a weak constraint, meaning that the mass divergence does not have to strictly equal zero. When the grid aspect ratio is near unity (i.e., the horizontal and

vertical grid spacing are nearly the same), the anelastic mass divergence constraint is found to result in accurate analyses of vertical and horizontal velocity (Hu et al. 2006b). However, when the horizontal grid spacing is much larger than the vertical grid spacing (i.e., the aspect ratio is over 100), which is often true in the lowest levels of the model, adjustments to the vertical velocity dominate adjustments to the horizontal component of the wind. This work follows that of Carlaw et al. (2015), which uses a horizontal weighting coefficient (α) that is an order of magnitude larger than the vertical weighting coefficient (β).

The ARPS 3DVAR system numerically minimizes an incremental form of the 3DVAR cost function using a conjugate-gradient minimization algorithm. Furthermore, preconditioning is used to reduce the computational cost by reducing the number of iterations necessary for the minimization algorithm to converge to the final analysis. More information on the ARPS 3DVAR analysis system can be found in Gao et al. (2004).

3.2.1 Incremental Analysis Updating

When numerical models are forced to adjust to large volumes of information, all applied at the initial time, nonphysical adjustment processes such as gravity waves (or noise) often occur (e.g., Bloom et al. 1996; Brewster 2003). To combat this issue, this research utilizes Incremental Analysis Updating (IAU), which is a method that applies analysis increments computed at the initial time gradually as a constant forcing for the model throughout an integration period (Bloom et al. 1996). The general procedure of IAU is to apply the analysis increments during the model's large time-step after all of the other forcing terms have been applied. The analysis increments are generally applied

using a triangular distribution in time, thus applying the largest portion of the observation increment during the middle of the time window. Increments are generally not applied to the pressure and vertical velocity fields during IAU at storm-scales, as these fields are not well observed and rapidly respond to changes in other model fields. Recently, the ARPS IAU code has been updated to allow users to specify more than one shape for IAU to make the distribution in time different for each variable (Brewster et al. 2015). More specifically, one can apply a larger portion of the wind and latent heat increments at the start of the assimilation window, while applying a more significant portion of the hydrometeor increments at the end of the window. This has been shown to mitigate difficulties maintaining an updraft in a convective system by allowing some time for the model to establish wind and mass fields that are capable of supporting the weight of precipitation species before introducing additional precipitation.

Additional information on the ARPS IAU with variable dependent timing (IAU-VDT) can be found in Brewster et al. (2015), while the theoretical basis can be found in Bloom et al. (1996).

3.2.2 Complex Cloud Analysis

The variational assimilation of radar reflectivity data is rather challenging owing to nonlinearities in the microphysical models and complex cross-correlations among variables. To account for these issues, the ARPS complex cloud analysis package is used in lieu of variational assimilation to account for radar reflectivity data (Brewster et al. 2005c; Hu et al. 2006a). The cloud analysis procedure uses satellite, radar, and surface observations of cloud layers to modify hydrometeor fields by using equations that relate hydrometeor mixing ratio values and observed radar reflectivity (e.g., Ferrier 1994;

Rogers and Yau 1989), recently updated to allow inversion of hydrometeor-to-reflectivity equations for all the microphysics schemes used in ARPS and WRF (Brewster and Stratman 2015). The complex cloud analysis is performed after the 3DVAR minimization is completed.

The background hydrometeor mixing ratio values are replaced by reflectivity-derived values in regions where radar reflectivity is above a user-defined threshold (typically 10 to 20 dBZ). This is based upon the belief that at this scale the radar observations, after quality control to remove non-precipitation echoes, are superior to the model background field. On the other hand, precipitation in the model background field is removed in regions where there is radar coverage and radar reflectivity is below the prescribed threshold, thus removing spurious convection from the model field. Finally, the cloud analysis procedure adjusts the temperature profile in regions where clouds and updrafts are present to account for the latent heat released during condensation processes. This has been shown to be important in maintaining updrafts in non-hydrostatic models, such as ARPS. To calculate the temperature adjustment due to latent heating, a moist adiabatic ascent is calculated from the cloud-base, with entrainment in areas of analyzed ascent, and the resulting temperature values replacing the 3DVAR analysis value in regions where the analyzed temperature is colder. More details on the complex cloud analysis procedure can be found in Brewster et al. (2005c) and Hu et al. (2006a).

3.2.3 Radar Remapping

Prior to being utilized in 3DVAR or the complex cloud analysis package, radar data must first be quality-controlled to account for radar artifacts. First, the raw radar data are checked for beam blockage effects (e.g., from tall buildings and trees) and sun

strobes during sunrise and sunset. Then, the raw radar data are checked for anomalous propagation effects, in which the radar beam is refracted towards the earth's surface, by identifying regions of large vertical reflectivity gradients, reflectivity texture, and low radial velocities. Isolated non-meteorological echoes are removed using a "despleckling" algorithm. Finally, the raw radar data are checked for velocity aliasing. This is performed by first converting the radial velocity data into increments from the mean wind, where the mean wind field represents an average of nearby data points in the background wind field. This mitigates effects of the vertical shear of the mean wind and helps pinpoint isolated regions of aliased velocities. Horizontal consistency checks are then performed across neighboring radials by calculating gate-to-gate shear; this is applied to the perturbation radial velocities following the method described in Eilts and Smith (1990). Once all of the quality checks are performed, the radar data are remapped from the polar coordinate system to the Cartesian grid used by ARPS via a least squares fit to a quadratic function in the horizontal and to a linear function in the vertical. In addition to reflectivity and radial velocity data, the remapping program has the capability of producing velocity azimuth display (VAD) wind profiles. Additional details concerning the ARPS radar remapping algorithm are found in Brewster et al. (2005c).

Chapter 4

4.1 Case Study

During the afternoon and early evening hours of 11 April 2016, a prolific hail-producing supercell thunderstorm affected north-central Texas, including the northern portion of the Dallas-Fort Worth metropolitan area. The supercell thunderstorm formed around 1900 UTC (2:00 PM CDT) just southwest of Wichita Falls and quickly became severe as it tracked to the east-southeast.

Severe storm reports from the Storm Prediction Center (SPC) are shown in Figure 4.1a, with numerous significant severe hail reports (diameter in excess of 2 inches, 5 cm) occurring along the track of this storm. Figure 4.1b zooms in on the severe storm reports occurring in the northern portion of the Fort Worth NWS forecast office's area of responsibility. The first significant severe hail report occurred around 2000 UTC in Archer County, just south of Wichita Falls. Significant severe hail was reported in Wise County, Texas beginning around 2130 UTC. The storm continued into Denton County around 2210 UTC, with grapefruit sized hail (4.00 inch diameter, 10 cm) reported around 2220 UTC. Significant severe hail also occurred in Plano, Allen, and Wylie in Collin County, with an additional report of grapefruit sized hail occurring in Rockwall County around 2310 UTC. The largest hail associated with the storm was reported in Wylie, with 5.25 inch (13.3 cm) diameter hail reported. The storm then gradually weakened as it moved out of the metropolitan area. This hail storm, in conjunction with a separate storm in San Antonio, Texas the following day resulted in an estimated total of $3.5 billion in damage (NOAA 2017).

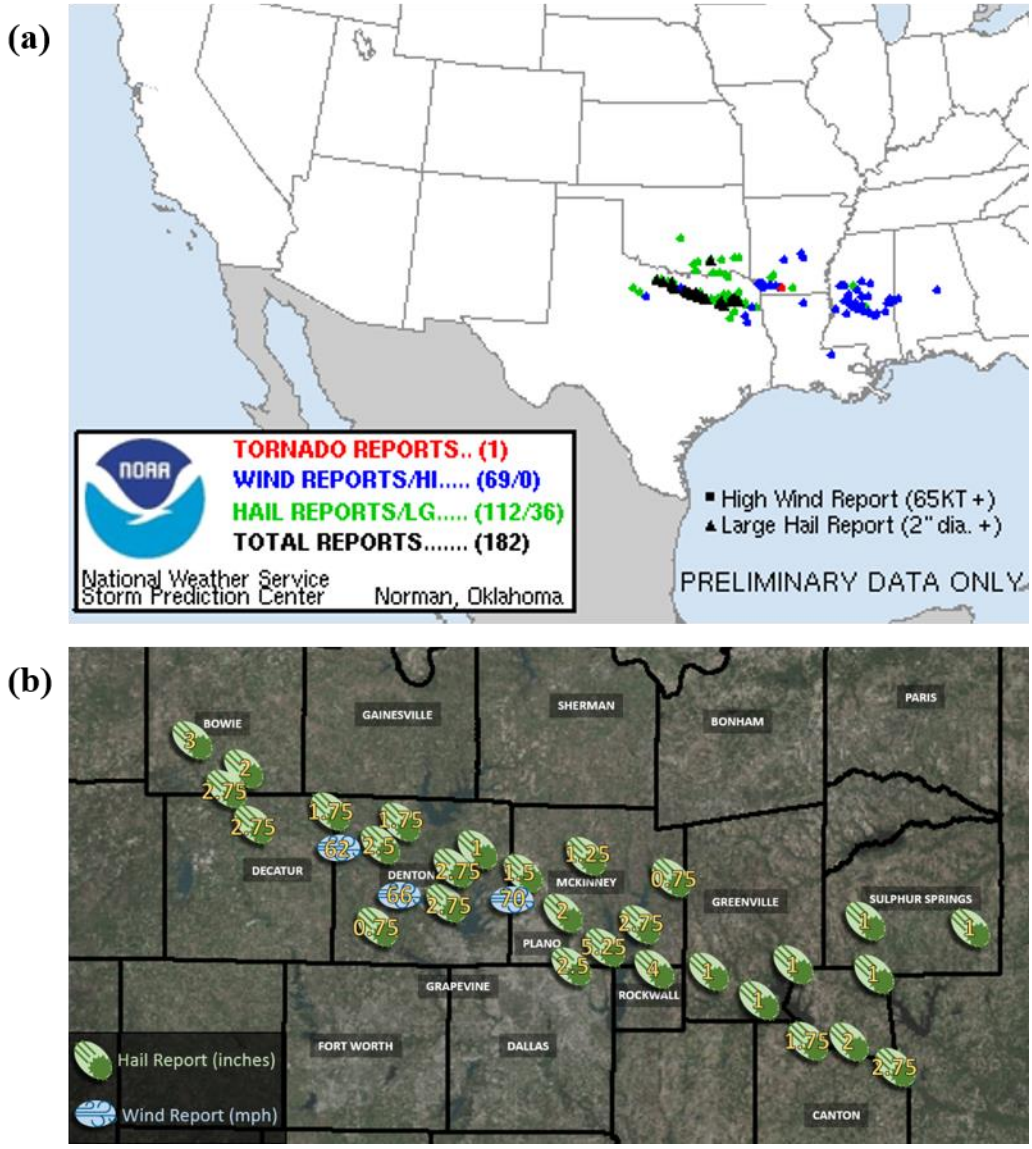


Figure 4.1: (a) Storm Prediction Center (SPC) severe storm reports for 11 April 2016 and (b) zoomed in severe storm reports for the storm of interest. Image credit: NWS Fort Worth, Texas (obtained online at http://www.weather.gov/fwd/20160411).

4.1.1 Synoptic Setup

A shortwave trough was present over the southern Plains, extending from the Texas Panhandle through New Mexico at 1200 UTC on 11 April 2016, as shown in the 500-mb upper-air chart (Figure 4.2). This trough deepened and moved to the east during the day, with a region of differential cyclonic vorticity advection (DCVA) present

downstream of the trough axis. In addition, a region of upper-level divergence is evident in the 300-mb chart, in response to the incoming subtropical jet stream maximum (Figure 4.3). Together, these features resulted in the development of a surface low pressure system, centered over northwest Texas. Southerly winds in the low-levels, as seen in the 925-mb analysis (Figure 4.4), afforded a rich moisture return, which, along with mid-level westerly winds, allowed for the formation of a dryline (e.g., Schaefer 1974; McCarthy and Koch 1982).

By 1800 UTC, the surface low pressure system was centered over northwestern Texas, just south of Wichita Falls (Figure 4.5). The aforementioned dryline extended south from this low pressure system through the Hill Country and Big Bend Regions of Texas. A cold front was present just south of Wichita Falls, Texas, with a cold front extending west from the low pressure system into New Mexico. A surface low pressure system centered north of the Great Lakes was associated with an additional cold front, which extended southwest to the Red River.

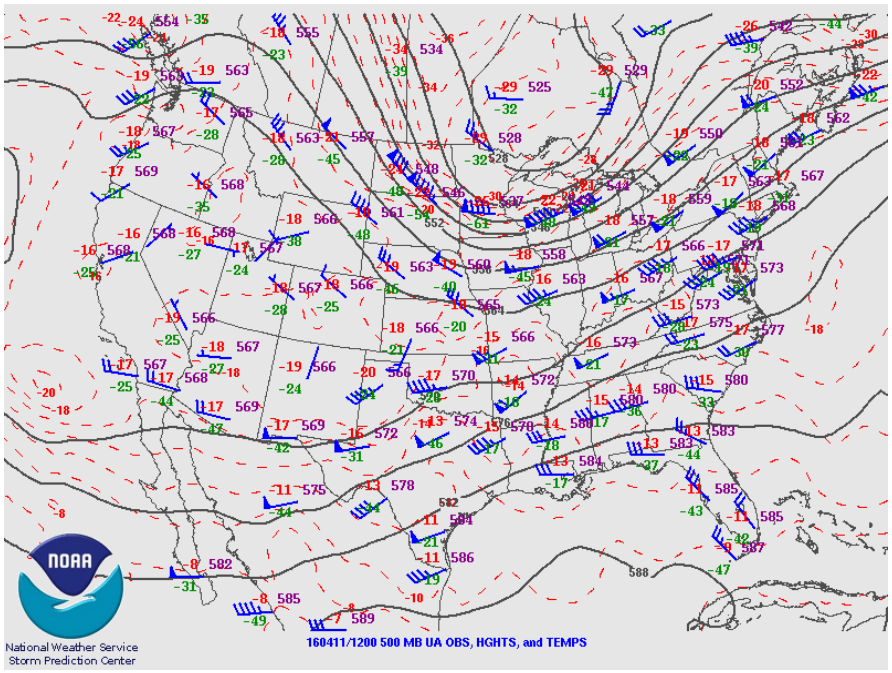


Figure 4.2: 500-mb upper-air analysis valid 1200 UTC on 11 April 2016. Solid black lines represent geopotential height contours (isohypses), while dashed red lines are isotherms.

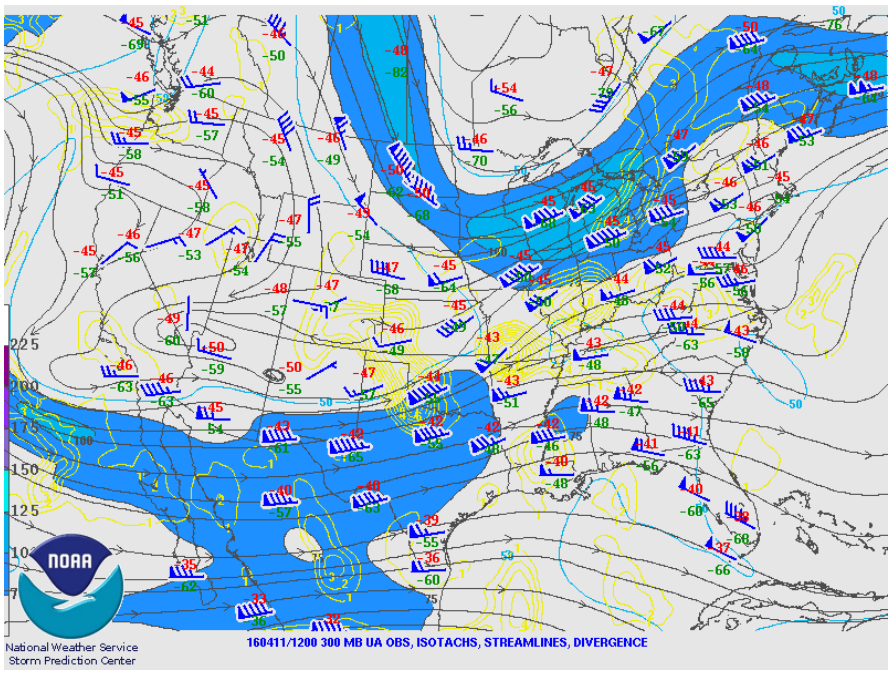


Figure 4.3: 300-mb upper-air analysis valid 1200 UTC on 11 April 2016. Isotachs are shaded, while streamlines are represented by solid arrows, and divergence is shown by solid yellow lines.

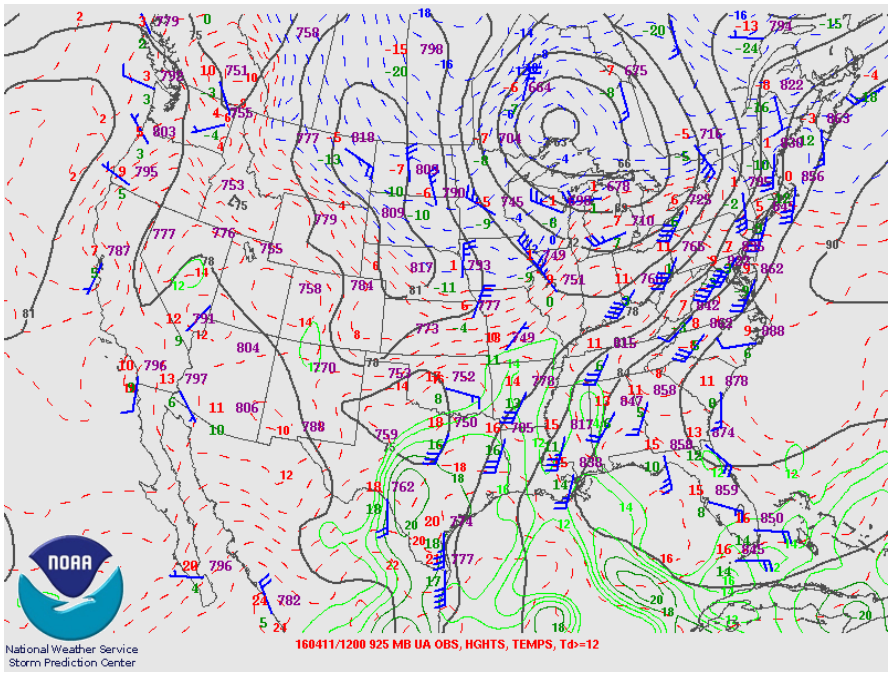


Figure 4.4: 925-mb upper-air analysis valid 1200 UTC on 11 April 2016. Isohypses are shown by solid black lines, isotherms by dashed red lines, and isodrosotherms by solid green lines.

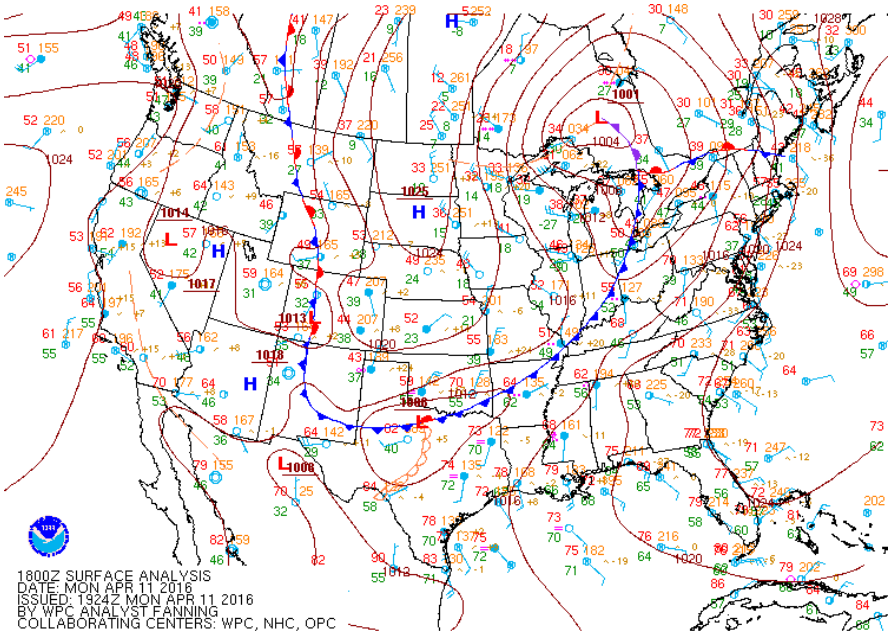


Figure 4.5: Surface analysis from the Weather Prediction Center (WPC) valid 1800 UTC on 11 April 2016.

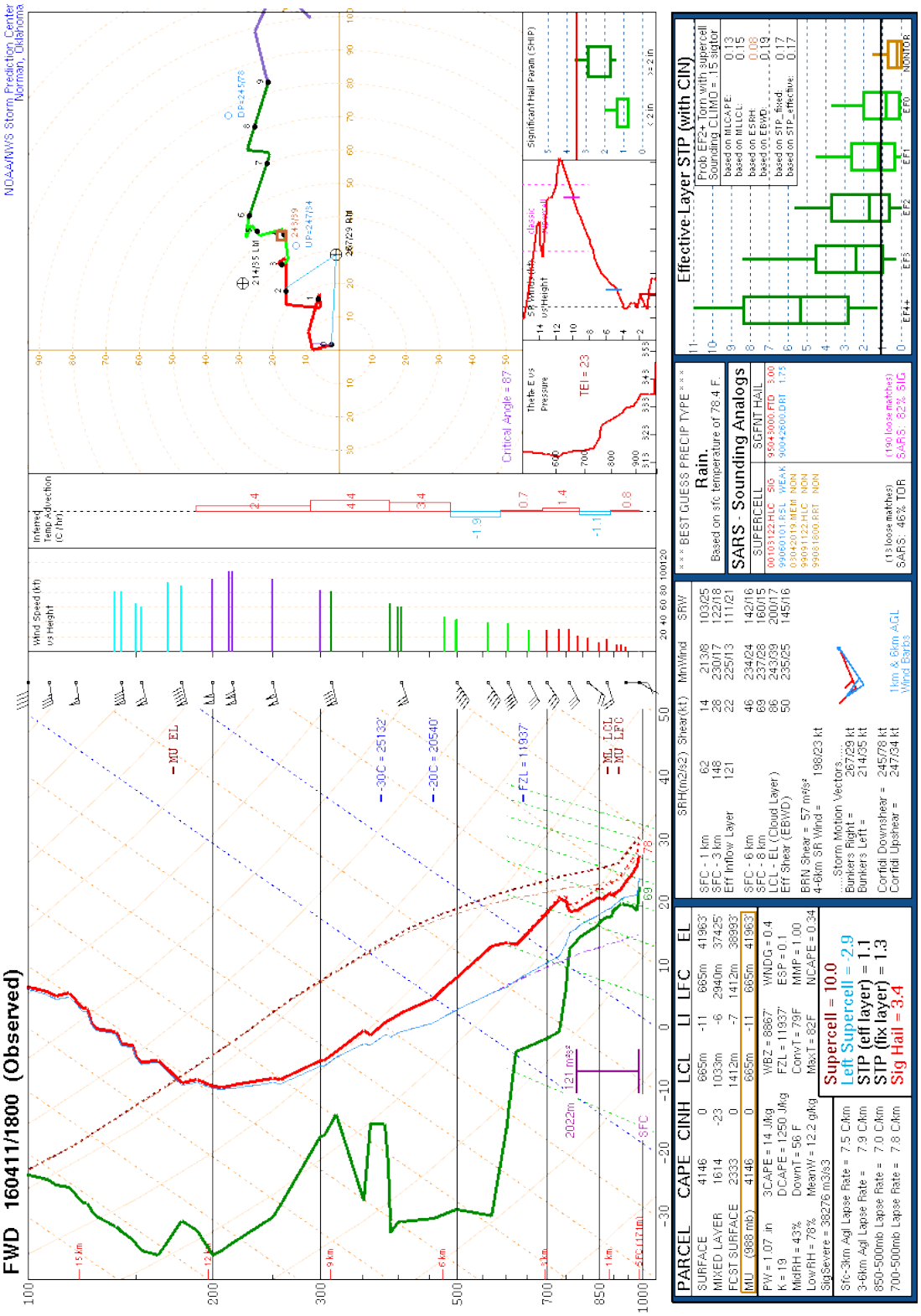


Figure 4.6: Observed sounding from Fort Worth (FWD) at 1800 UTC on 11 April 2016. The temperature and dew point profiles are shown in red and green, respectively. Analyzed variables are derived using the NSHARP program.

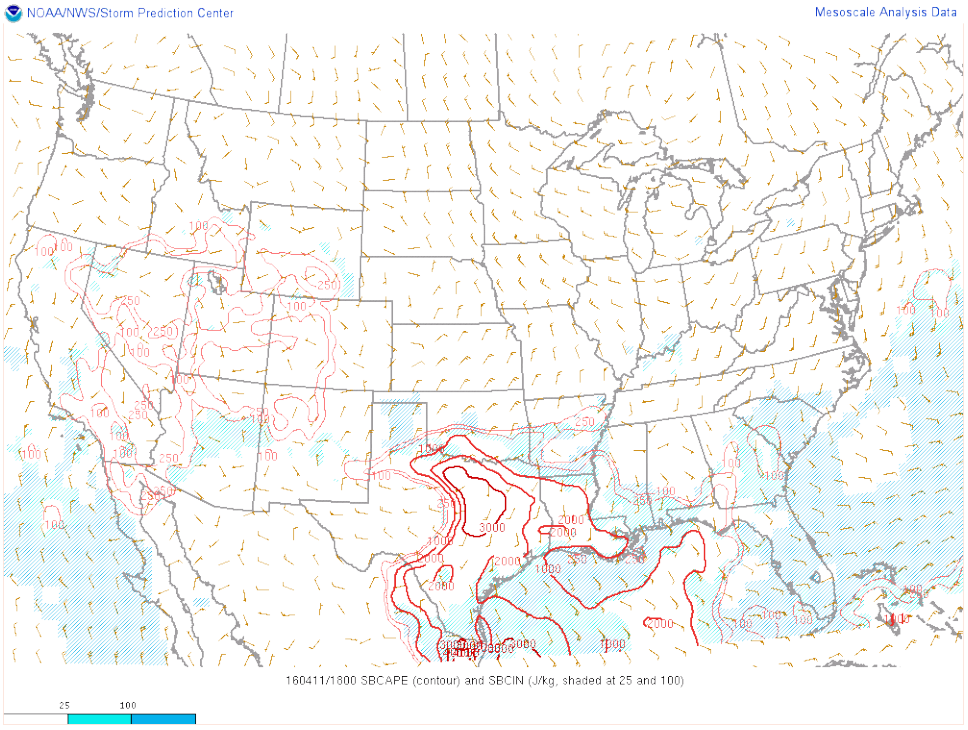


Figure 4.7: Surface-based CAPE (SBCAPE; contoured) and surface-based convective inhibition (SBCIN; shaded) at 1800 UTC on 11 April 2016.

The special 1800 UTC sounding from Fort Worth is shown in Figure 4.6. Steep mid-level lapse rates in excess of 7.5 °C per kilometer, in association with intense surface heating, resulted in surface-based convective available potential energy (SBCAPE) in excess of 4000 J/kg and surface-based convective inhibition (SBCIN) of 0 J/kg. These sounding values were obtained from the National Center Sounding and Hodograph Analysis and Research Program (NSHARP; Hart and Korotky 1991), which is used by the SPC. SBCAPE values in excess of 3000 J/kg were present throughout the entire DFW metropolitan area (Figure 4.7). Sufficient shear in the 0-6 km layer (i.e., in excess of 40 knots) was present for the development of supercells, but with the wind profile largely characterized by a straight hodograph, any thunderstorms that developed would likely need to move off of the hodograph in order to ingest streamwise vorticity and develop a

mesocyclone. Hail represented the dominant severe hazard on this date, owing to the ample CAPE values, deep-layer shear, and steep 700-to-500-mb lapse rates in excess of 7.5 °C per kilometer. Surface-to-1-km shear and storm-relative helicity (SRH) were not overly favorable for the development of tornadoes in the DFW area. Sufficient forcing for ascent owing to the dryline, frontal boundaries, and upper-level divergence would aid in convective initiation.

The supercell of interest failed to produce a tornado in the Dallas-Fort Worth metropolitan area. The storm was quickly undercut by the cold front depicted in Figure 4.5, along with an outflow boundary generated by the storm itself. Once this undercutting occurred, the supercell's supply of warm, moist surface air was restricted, and the storm became elevated in nature. As a result, the potential for tornadogenesis was significantly reduced.

4.2 ARPS Model Grid Setup and Specifications

The experiments presented here use the ARPS model on a single grid with a horizontal grid spacing of 1 km. The domain is centered on 33.0 °N, 97.25 °W, and is 400 km x 360 km in size (see Figure 2.3 and Figure 4.9 for a map of the domain). There are 50 vertical levels, with vertical grid spacing averaging 400 m increasing with height from a minimum of 20 m at the lowest model level following a hyperbolic tangent function (Xue et al. 1995). Land surface features are specified using the ARPS surface data files, while terrain elevation information is derived from the 30-second U.S. Geological Survey (USGS) terrain dataset. Additional model specifications and parameterization schemes are outlined in Table 3.1 in the preceding chapter.

This work follows that of Snook and Xue (2008) and Carlaw et al. (2015) and uses a reduced value of the rain intercept parameter in the raindrop size distribution in the Milbrandt and Yau single-moment microphysics scheme. The single-moment microphysics scheme used in this research predicts the mixing ratio for each hydrometeor species, while setting the intercept parameter to a constant value. Deep, moist convection, such as in the storm considered in this work, tends to have a drop size distribution (DSD) that favors larger raindrops. It has been shown that reducing the rain intercept parameter results in the generation of more large raindrops, while reducing the number of smaller drops (Snook and Xue 2008). The production of larger hydrometeors results in the reduction of total hydrometeor surface area, which, in combination with the faster terminal velocities associated with larger particles, results in less evaporational cooling. As a result, weaker cold pools develop, yielding stronger, more-sustained updrafts. The rain intercept parameter is reduced one order of magnitude to $8.0 \times 10^5 m^{-4}$ (Carlaw et al. 2015).

4.3 Experimental Design

For the experiments presented here, four intermittent data assimilation cycles are used to incorporate observational data in the model forecast, with the first cycle beginning at 2150 UTC. Analysis increments are determined using the 3DVAR analysis system and applied gradually during the subsequent 10 minutes using incremental analysis updating with variable-dependent timing (IAU-VDT). A larger fraction of the wind and latent heat increments are applied early in the IAU window, while the hydrometeor increments are largely applied near the end of the IAU window. Three 10-minute forecasts are performed with the forecast becoming the background field for the subsequent cycle. The

fourth cycle is used to launch a one and a half hour free forecast, beginning at 2220 UTC and ending at 2350 UTC, during which no data is assimilated beyond what is used during the first 10 minutes. Initial conditions are obtained for the model grid using the 2100 UTC and 2200 UTC Rapid Refresh (RAP) analyses, which use a 13 km grid. These model analyses are interpolated both in space and in time to produce the initial conditions valid at 2150 UTC for the model grid used. Lateral boundary conditions (LBCs) are also derived from the RAP analyses valid at 2100, 2200, 2300, and 0000 UTC, with lateral boundary forcing occurring every 5 minutes. A schematic of the assimilation and forecast procedure used is shown in Figure 4.8. Quantitative verification metrics are performed using a smaller verification grid, which is outlined in red in Figure 4.9.

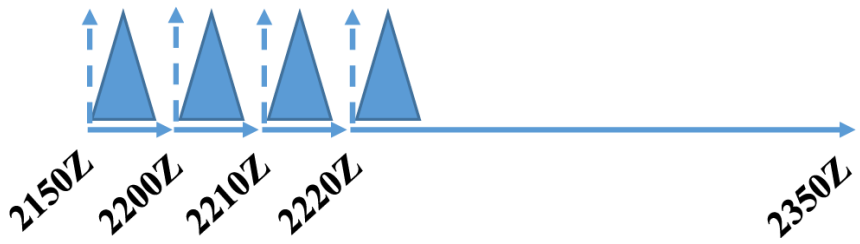


Figure 4.8: Assimilation procedure for the experiments presented. Data assimilation cycles begin at 2150Z, with a 1.5 hour free forecast beginning at 2220Z. Triangles represent the weighting of fractions of the computed analysis increment introduced during each assimilation window.

The ARPS 3DVAR analysis system used here incorporates four analysis passes, which accounts for the diverse spacing of observation types assimilated. The horizontal correlation scale distance for the first, second, third, and fourth pass is 100 km, 50 km, 10 km, and 0.8 km, respectively. Profiler data, namely from the two SODARs, are incorporated on the first and second passes, which allows this information to be spread across the model domain as the observations are considered to be representative of a larger area. Conventional surface observations (ASOS/AWOS) and MDCRS flight data

are incorporated in the second and third passes. Non-conventional surface data, along with the mesonet observations, are assimilated in the third analysis pass. Radar reflectivity data are incorporated only during the fourth and final pass using the smallest correlation scale distance. Smaller-scale details are filled into the analysis as the correlation scale is decreased on subsequent passes. The vertical correlation scale is defined to be four grid points for all four assimilation passes.

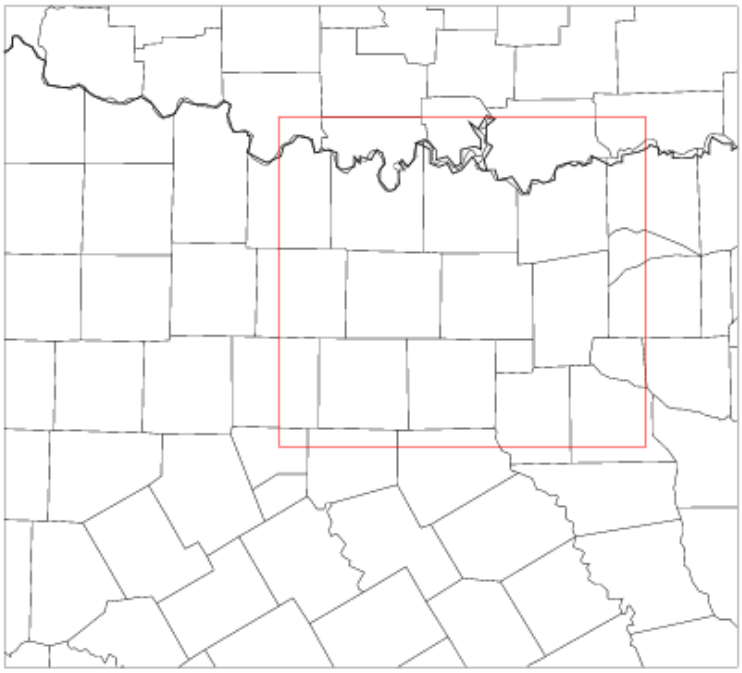


Figure 4.9: Model domain with the subdomain used for quantitative verification metrics outlined in red.

The data types assimilated for each experiment are listed in Table 4.1. In the CONTROL experiment, all available data are used, including reflectivity and radial velocity data from the WSR-88D, CASA, and TDWR radars with coverage in the domain. Conventional data sources include surface observations from ASOS/AWOS and from the Oklahoma and West Texas Mesonets, along with MDCRS aircraft data. Non-conventional surface data sources include GST MoPED, WeatherBug, CWOP, and

Understory. Finally, data from two SODARS located in the testbed are assimilated. NOTESTBED simulates what would occur if the DFW Urban Demonstration Network was not in place at the time of the case study. More specifically, the only data assimilated for this case are the data available in the federal observing network (ASOS/AWOS, Oklahoma and West Texas Mesonet, MDCRS, and WSR-88D radar data). Non-conventional surface data are denied, along with all CASA and TDWR data. NONEWSFC denies the non-conventional surface data, while retaining the CASA and TDWR radar data. NOGST, NOWXBUG, NOCWOP, and NOUNDERSTORY individually deny the GST MoPED, WeatherBug, CWOP, and Understory data, respectively. NOCWOPWXBUG denies both the CWOP and WeatherBug data. NOCASA denies reflectivity and radial velocity data from the six CASA radars active at the time of the storm of interest, whereas NOCASAVR denies only the CASA radial velocity data, but still employs the reflectivity data in the complex cloud analysis. NOTDWR denies the TDWR radar data, while 88DONLY additionally denies the CASA radar data. CASAONLY demonstrates what would occur if there were no WSR-88D radars active at the time of the storm of interest. NORADAR presents the model forecast using only surface, aircraft, and SODAR data.

Table 4.1: Observing System Experiments Performed

Experiment	Conventional surface data	Non-conventional surface data	88D data	CASA data	TDWR data	Upper-air profiles
CONTROL	All	All	All	All	All	All
NOTESTBED	All	None	All	None	None	Deny SODARs
NONEWSFC	All	None	All	All	All	All
NOGST	All	Deny MoPED	All	All	All	All
NOWXBUG	All	Deny WxBug	All	All	All	All
NOCWOP	All	Deny CWOP	All	All	All	All
NOUNDERSTORY	All	Deny Understory	All	All	All	All
NOCWOPWXBUG	All	Deny CWOP and WXBUG	All	All	All	All
NOCASA	All	All	All	None	All	All
NOCASAVR	All	All	All	Reflectivity only	All	All
NOTDWR	All	All	All	All	None	All
88DONLY	All	All	All	None	None	All
CASAONLY	All	All	None	All	None	All
NORADAR	All	All	None	None	None	All

4.4 Results

The experiment results presented here largely fall into two categories. The first category considers the impact of radial velocity and reflectivity data from the WSR-88D, TDWR, and CASA radar networks, while the second category considers the impact of non-conventional surface data. One experiment (NOTESTBED) considers the combined impacts of the CASA and TDWR data and non-conventional surface networks. The first section offers a qualitative comparison of the simulated reflectivity field for several experiments in the radar category. The second section presents a quantitative comparison of the forecasted reflectivity field. The third section details the quantitative hail verification. The fourth section performs a surface-level forecast verification. The chapter concludes with a brief foray into the impacts of the choice of microphysics parameterization scheme.

4.4.1 Qualitative Reflectivity Comparison

Figure 4.10 shows how the simulated reflectivity at model level 21 at approximately 2 km above ground level (AGL) for the CONTROL experiment compares with low-level reflectivity observations, namely the 0.5 degree tilt from the KFWS WSR-88D radar in Fort Worth. The last data are assimilated at 2220 UTC and it can be seen that the reflectivity pattern in the CONTROL experiment generally matches the observations at this time, as it captures the most intense precipitation occurring in Denton County. Furthermore, the maximum reflectivity value in both cases is about 65 dBZ. The region of precipitation that extended into southern Oklahoma was also captured in the CONTROL experiment, although the reflectivity values are somewhat larger than what was observed. The areal coverage of reflectivity in the CONTROL experiment is

more expansive than in reality, likely due to the inherent differences in the figures (2 km AGL vs 0.5 degree tilt). By 2250 UTC, 30 minutes into the free forecast period, the simulated supercell was centered in southern Collin County, close to where the storm appeared in observations, although the hook echo is difficult to discern in the 2 km AGL image. Twenty minutes later, at 2310 UTC, the observed storm was beginning to enter Rockwall County, while the simulated storm was centered in extreme northeastern Dallas County, indicating that there are slight propagation speed errors in the CONTROL experiment. This general pattern continues at 2330 UTC, as well, as the simulated storm is positioned slightly to the southwest of the observed position. The CONTROL experiment exhibits a general wet bias, with reflectivity predicted over a larger area than what was observed. Additionally, in the observations, a fine line is evident to the south of the supercell, which is indicative of the cold front and outflow boundary that undercut the storm early in its lifespan. This feature is suggested in the wind fields at this level and is more apparent at low-levels (discussed later).

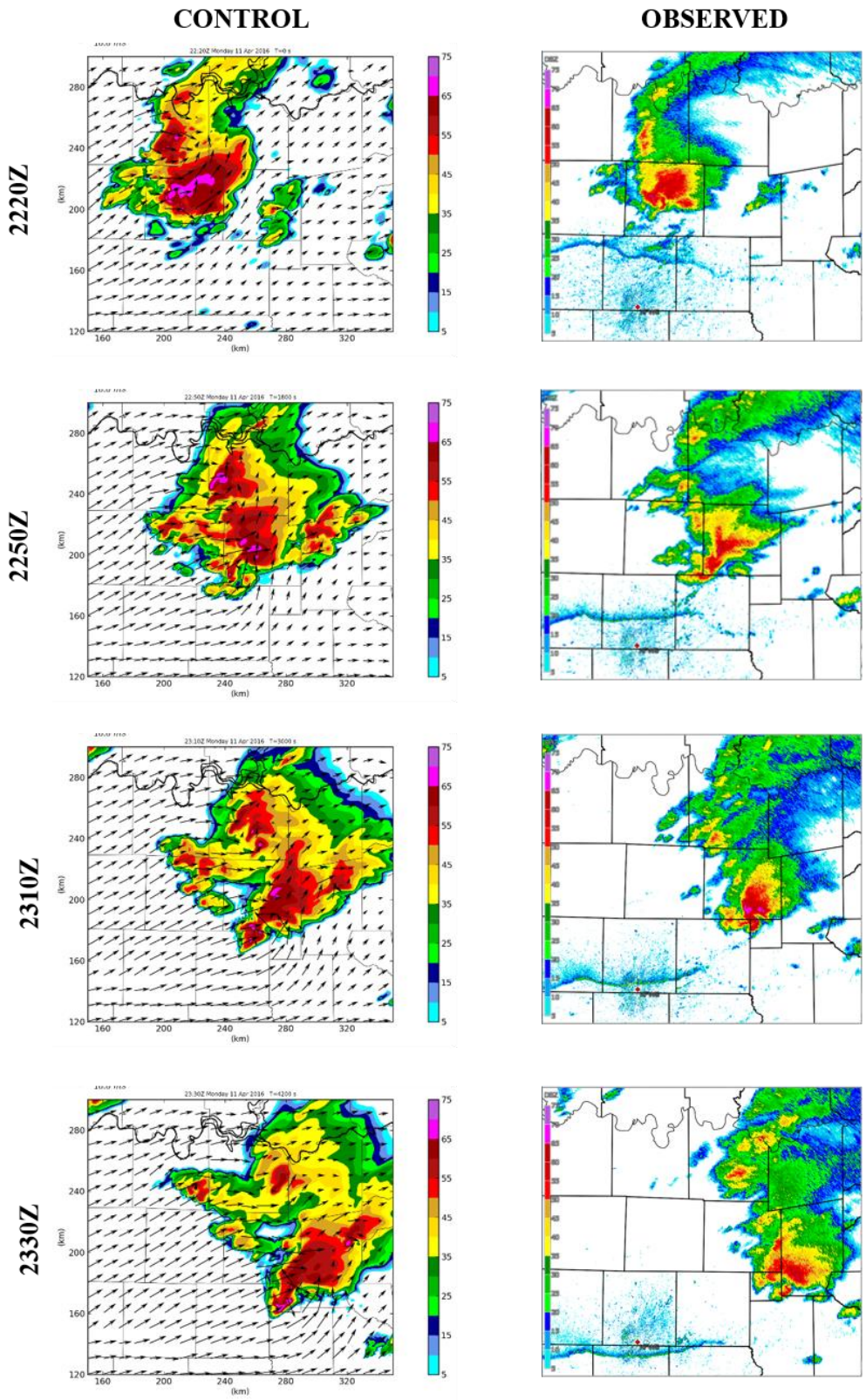


Figure 4.10: Simulated reflectivity at 2 km AGL for the CONTROL experiment (left) and reflectivity from the KFWS 0.5 degree scan (right).

Figure 4.11 shows the simulated wind vectors and 2 km AGL reflectivity fields for several of the experiments with more notable differences from CONTROL. The CONTROL experiment has an ill-defined hook echo by 2300 UTC, with a well-defined hook echo evident by 2330 UTC in eastern Rockwall County. The 88DONLY experiment does not exhibit a well-defined hook echo at 2300 UTC, which is also the case at 2330 UTC, as there appear to be two distinct hook echoes at that time. The NOCASA supercell appears less-defined than for the CONTROL experiment, with a secondary hook feature evident at 2300 UTC. By 2330 UTC, the secondary feature has weakened, while the main supercell has a well-defined hook echo and moves out of Rockwall County. Therefore, the storm is displaced to the east in the NOCASA experiment relative to the actual storm and the CONTROL run. Overall, it appears that the inclusion of the low-level radial velocity data from the CASA X-band radars afforded a better initial analysis of the wind field, which resulted in a superior simulation in the CONTROL experiment. Thus, the CASA data appear to provide some positive value for this case study. However, as can be seen by CASAONLY, the CASA radar network is insufficient as a standalone tool. Given that the CASA radars tend to only have two or three elevation scans, with radar beams approaching 2 km AGL at a range of 40 km (see Figure 2.2), the radars were incapable of observing the full depth of the ongoing storm. This can be seen in Figure 4.12, which shows cross-sections through the CONTROL and CASAONLY experiments. The depth of the storm is significantly shallower in CASAONLY. As a result, the final analysis at 2220 UTC did not include a dynamically-sound supercell, with the storm beginning to intensify by 2230 UTC, 10 minutes into the free forecast. The storm in CASAONLY was positioned west of what was observed by

2330 UTC, likely due to the model's poor initial handling of the storm. The final experiment presented here, NORADAR, does not have an analyzed storm at the beginning of the free forecast, owing to the denial of all radar data. Therefore, the storm present by 2300 UTC must be "spun up" by the model during the forecast period. Despite this, the simulation produces a supercell positioned in Rockwall County by 2330 UTC, not far from where the observed storm was located. However, the simulation also produces a band of heavy precipitation extending northwest to the Red River, which is not as prevalent in the other experiments nor the observations.

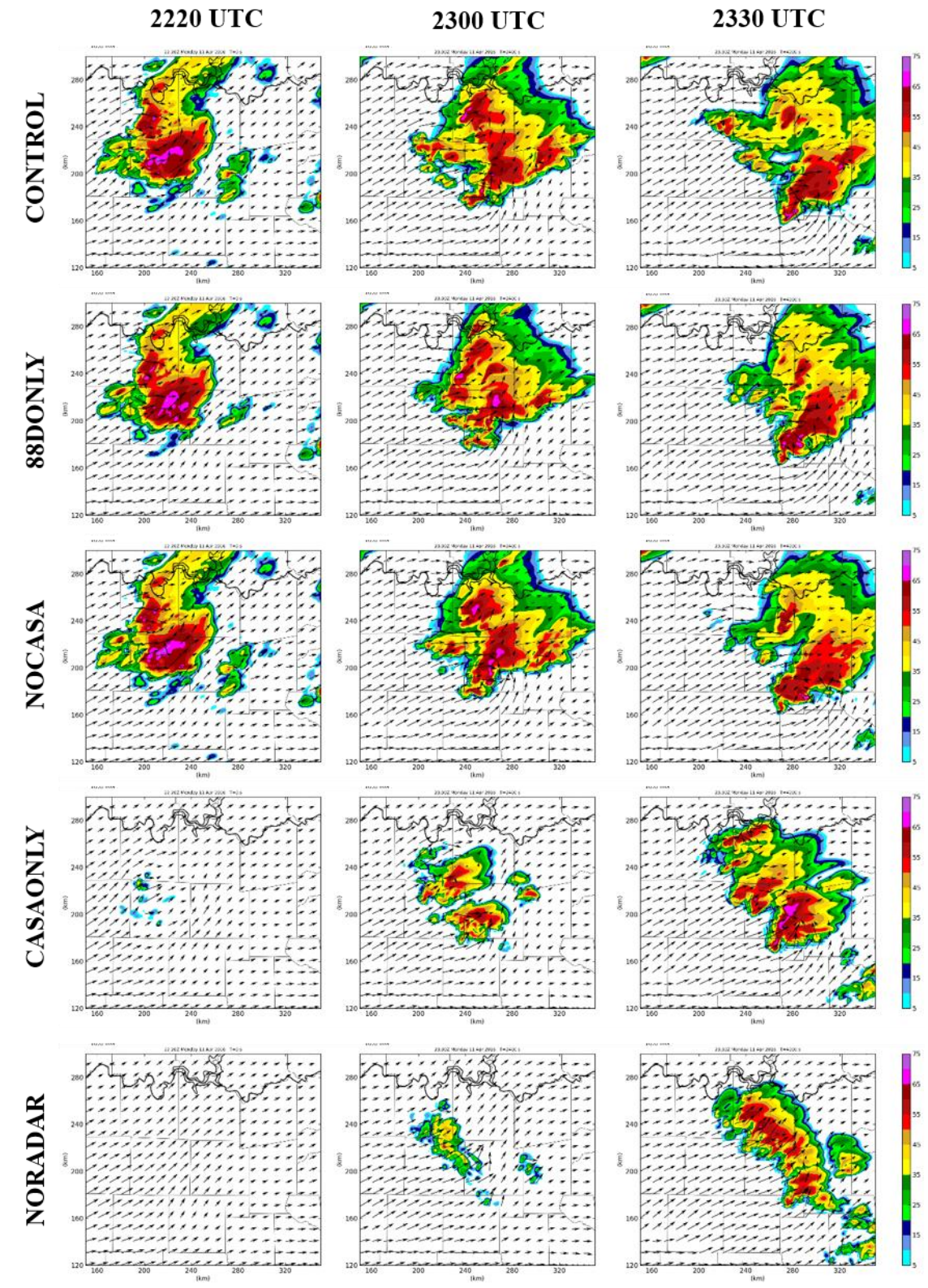


Figure 4.11: Simulated reflectivity and wind vectors at 2 km AGL for CONTROL, 88DONLY, NOCASA, CASAONLY, and NORADAR experiments.

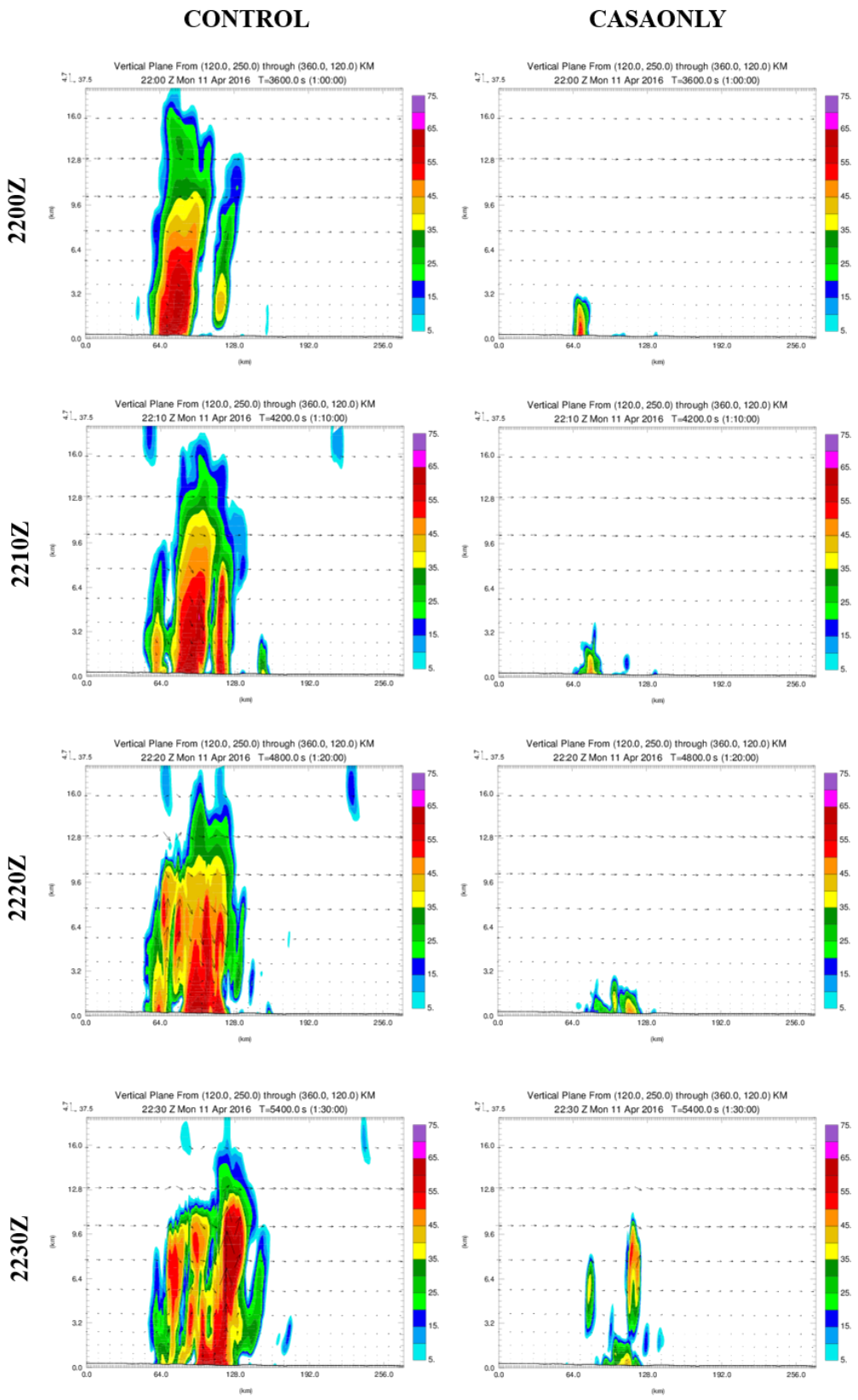


Figure 4.12: Vertical cross-sections for the CONTROL and CASAONLY experiments.

Figure 4.13 shows the simulated surface winds and 1 to 5 km updraft helicity (UH; Kain et al. 2008). Updraft helicity allows for the detection of rotating thunderstorms in numerical models and is defined by

$$UH = \int_{z_0}^{z_1} w\zeta \, dz \tag{4.1}$$

where w is the vertical component of the wind and ζ is the vertical component of vorticity. The CONTROL experiment has a clear UH center present at 2300 UTC, which strengthens further by 2330 UTC. In the 88DONLY experiment, the UH center is not as well-defined at 2300 UTC. The result of this is two distinct UH centers at 2330 UTC, with the stronger center located to the northeast of the corresponding feature in the CONTROL experiment. The weaker feature was located in western Rockwall County. This is consistent with the reflectivity field, as there appears to be two distinct hook echoes at 2330 UTC. A similar pattern is observed for the NOCASA experiment, although the stronger feature tracked further into Hunt County than in 88DONLY. The CONTROL, 88DONLY, and NOCASA experiments all do not have a clear UH center present at 2220 UTC, likely owing to the ARPS model requiring some time to build a dynamically-consistent supercell. The CASAONLY experiment fails to build a strong UH center by 2330 UTC, consistent with the less-organized supercell evident in the reflectivity field at the time. Surprisingly, the NORADAR experiment, which assimilates only surface, SODAR, and aircraft data, develops a UH center that is located in Rockwall County by 2330 UTC, not too far from where the observed storm was positioned. However, there are also several lobes of UH that are not as prevalent in the other experiments.

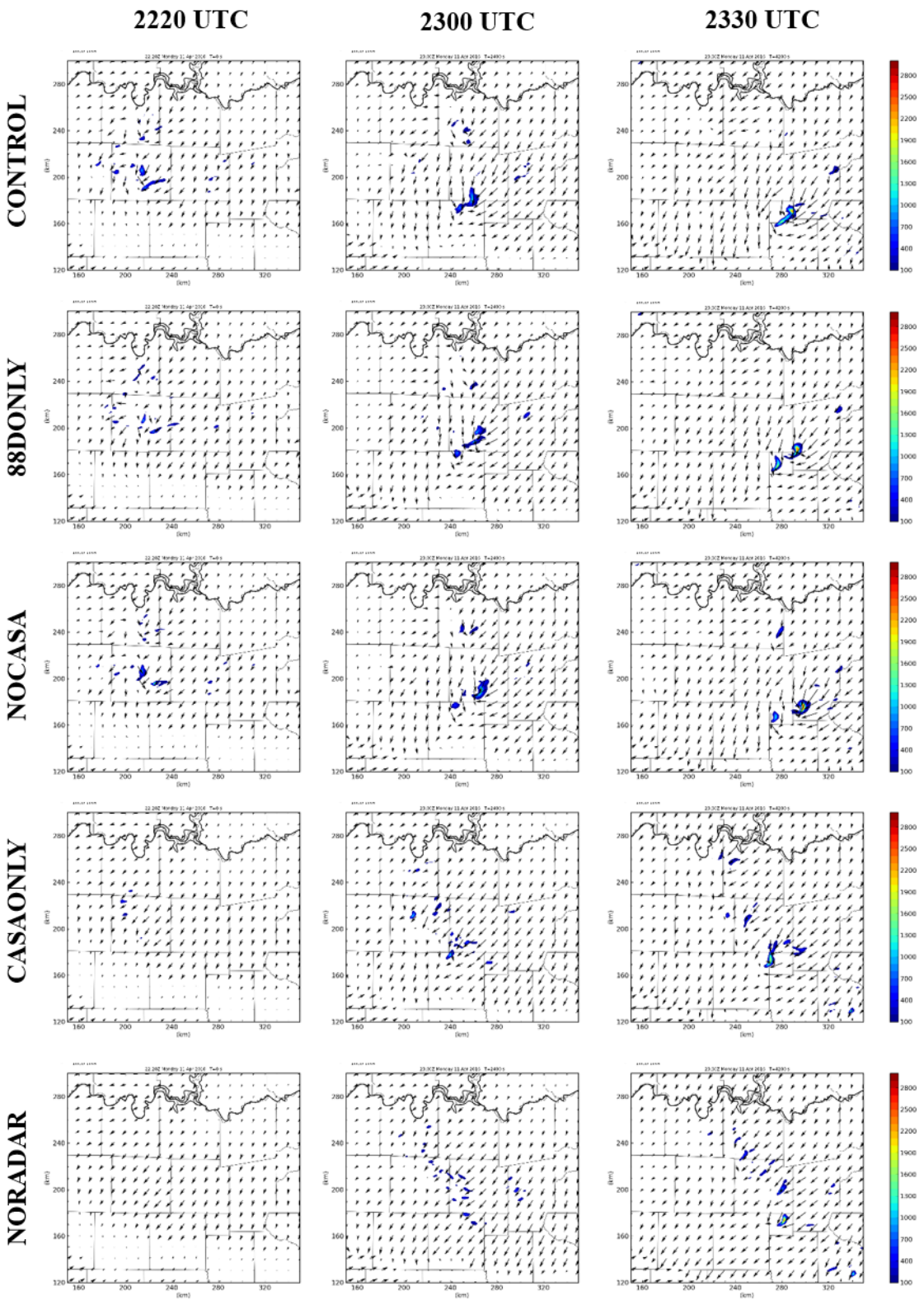


Figure 4.13: Surface winds and 1 to 5 km updraft helicity (UH) for the CONTROL, 88DONLY, NOCASA, CASAONLY, and NORADAR experiments.

4.4.2 Quantitative Reflectivity Verification

Quantitative comparisons of the forecasted reflectivity field among experiments are performed using the fractions skill score (FSS; Ebert 2008). The FSS compares the fractional coverage of forecast events to the observed fraction of the same event. The FSS is therefore a probabilistic verification technique and is defined by:

$$FSS = 1 - \frac{\frac{1}{N}\sum_N (P_f - P_o)^2}{\frac{1}{N}[\sum_N P_f^2 + \sum_N P_o^2]} \tag{4.2}$$

where P_f is the fraction of grid points within a neighborhood that have forecasted values exceeding the specified threshold, P_o is the fraction of grid points within the same neighborhood that have observed values above the same threshold, and N is the number of neighborhoods in the domain. FSS values can range from 0 to 1, with a value of 1 corresponding to a perfect forecast. Values of 0 correspond to instances when there are no events forecasted within a neighborhood and some are observed, and vice versa.

The FSS is considered more robust than traditional grid-based metrics because, unlike traditional grid-based verification metrics, the FSS does not double-penalize the forecast when there are slight spatial errors in the forecasted field.

Here, the "event" corresponds to composite reflectivity, which is determined for the WSR-88D radar in Fort Worth (KFWS) and compared to composite reflectivity for each of the experiments performed. Composite reflectivity was chosen rather than reflectivity at a single height (e.g., 1 km AGL) as the KFWS radar is unable to observe the storm completely at a single level during the entire free forecast period.

Since the FSS seeks to determine at what spatial scale the forecast and observations agree with one another, the neighborhood scale for which the forecast exhibits useful skill is first determined. A useful forecast (i.e., one in which the skill is

halfway between random and perfect) is defined as one in which $FSS \geq FSS_{uniform}$, where

$$FSS_{uniform} = 0.5 + \frac{f_o}{2} \tag{4.3}$$

and f_o is the observed fraction average over the domain (Roberts and Lean 2008). To determine this neighborhood scale, FSS is computed using a range of neighborhood scales and reflectivity thresholds and averaged over the free forecast period. Figure 4.14 shows the average FSS value using reflectivity thresholds of 20, 25, and 30 dBZ using varying neighborhood sizes. Figure 4.14 shows the results for the CONTROL, 88DONLY, and CASAONLY experiments, with the dashed line corresponding to the $FSS_{uniform}$ value. Figure 4.14 shows a general increase in the FSS as the neighborhood size is increased, although there is a slight decrease at neighborhood scales above 129 grid points (129 km) for the 25 and 30 dBZ thresholds. Additionally, it can be seen that the average FSS value decreases as the reflectivity threshold is increased from 20 dBZ (Figure 4.14a) to 25 dBZ (Figure 4.14b) to 30 dBZ (Figure 4.14c), which indicates that the simulations are better able to predict lower reflectivity thresholds. In other words, small regions of larger reflectivity values are harder to predict. Similar patterns can be seen in Figure 4.15, which shows the results for the CONTROL, NOTESTBED, and NONEWSFC experiments. Since large neighborhood sizes reduce the value of the 1 km grid spacing used in these experiments, it is desired to choose the scale that exhibits useful skill, without overly smoothing the model forecast. All experiments presented in Figures 4.14 and 4.15 exhibit useful skill for neighborhood scales at 1 km when the 20 dBZ threshold is used, with the exception of the CASAONLY experiment, which requires a 15 km neighborhood.

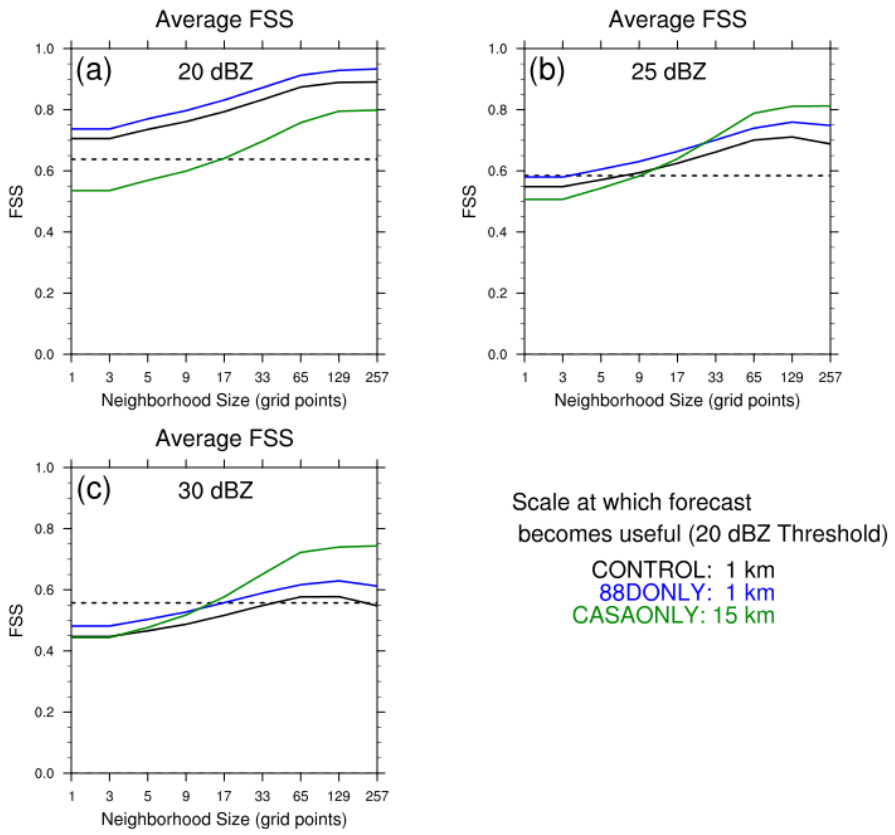


Figure 4.14: Average FSS values for composite reflectivity as a function of neighborhood size for a) 20 dBZ threshold, b) 25 dBZ threshold, and c) 30 dBZ threshold. Experiments shown include CONTROL, 88DONLY, and CASAONLY. The dashed line corresponds to the $\boldsymbol{FSS_{uniform}}$ value.

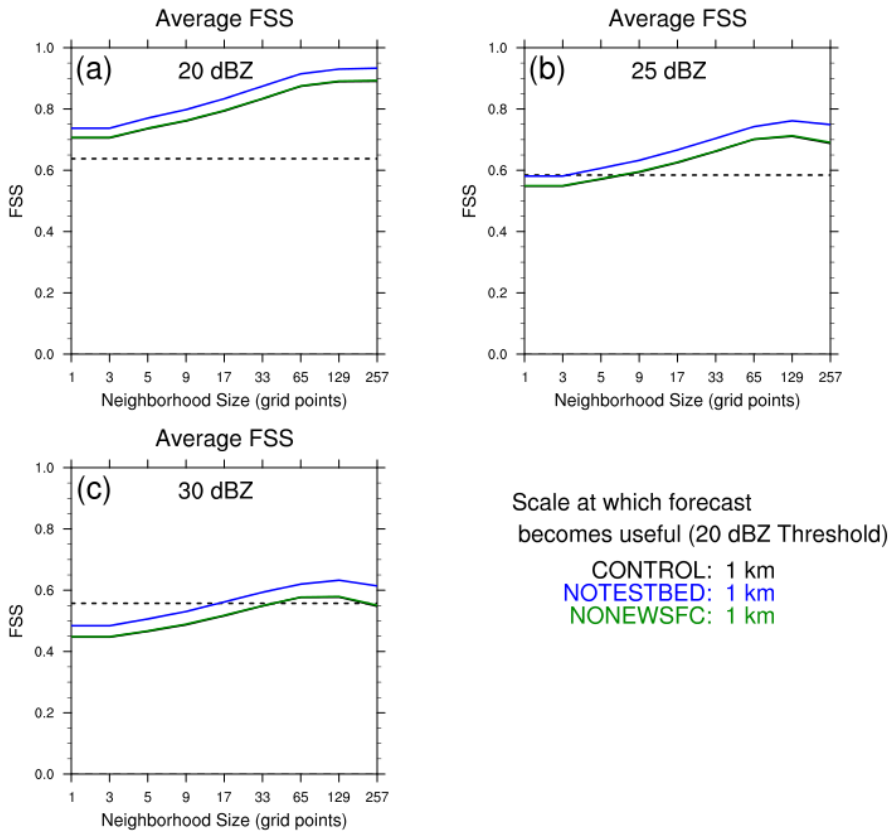


Figure 4.15: As in Figure 4.14, but for the CONTROL, NOTESTBED, and NONEWSFC experiments.

A neighborhood size of 16 km was chosen to account for increased uncertainty in timing and location of the forecasted features as the free forecast progresses forward in time. Figure 4.16 shows the time series for the CONTROL, 88DONLY, CASAONLY, and NORADAR experiments. Figure 4.16a shows the results for a 20 dBZ reflectivity threshold. The NORADAR experiment exhibits an FSS value of 0 at the beginning of the free forecast period, as it has yet to spin-up precipitation during this time, with values beginning to increase roughly 15 minutes into the free forecast as precipitation begins to develop within the model. The FSS value for NORADAR converges to the values seen for CONTROL and 88DONLY roughly one hour into the free forecast period. CASAONLY performs better than the NORADAR experiment initially, as it able to

capture some of the precipitation ongoing at the onset of the free forecast. Despite this, the NORADAR experiment has a higher FSS value than the CASAONLY experiment at the end of the free forecast period. The 88DONLY experiment has the highest FSS value for the entirety of the free forecast period, exceeding that of the CONTROL experiment. This indicates that the inclusion of the CASA and TDWR radar data reduces the FSS value when the 20 dBZ threshold is used. Some of this may be attributed to the KFWS data being used for verification rather than a mosaic of all radars as was used in initializing CONTROL. Figure 4.16b shows the results for a 25 dBZ reflectivity threshold. The NORADAR experiment exhibits a similar pattern of no skill at the beginning of the free forecast period, but this experiment again performs better than CASAONLY at the end of the free forecast period. CASAONLY has slightly better skill at the beginning of the free forecast period for this threshold. The forecast dips below the useful threshold between 2255 and 2315 UTC for the CONTROL experiment, whereas the 88DONLY experiment exhibits useful forecast skill through the entirety of the free forecast period. Figure 4.16c shows results for the 30 dBZ threshold. Besides NORADAR, the CONTROL experiment exhibits the lowest FSS value through the first half of the free forecast period. CASAONLY and 88DONLY have the highest skill through the majority of the free forecast period. The CONTROL experiment performs slightly better than CASAONLY at the end of the free forecast period.

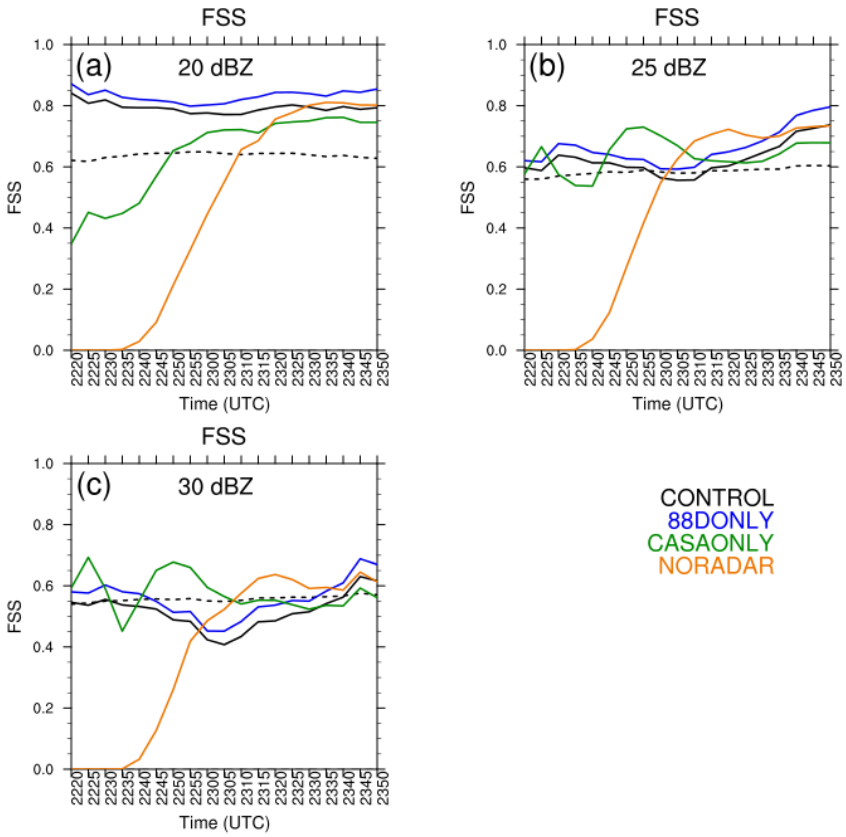


Figure 4.16: Time series of FSS values during the free forecast period using a 17 grid point (16 km) neighborhood size for a) 20 dBZ, b) 25 dBZ, and c) 30 dBZ reflectivity thresholds. Experiments shown include CONTROL, 88DONLY, CASAONLY, and NORADAR.

Figure 4.17 shows the time series for the CONTROL, NOTESTBED, and NONEWSFC experiments. The NOTESTBED experiment consistently scores better by this metric than the other experiments, including CONTROL, for the 20 dBZ threshold (Figure 4.17a). The CONTROL and NONEWSFC experiments perform similarly for this threshold, indicating that the majority of the differences seen in the NOTESTBED and CONTROL experiments are attributable to the inclusion of data from the CASA and TDWR radars. NOTESTBED consistently performs better for the 25 dBZ (Figure 4.17b) and 30 dBZ threshold (Figure 4.17c), as well.

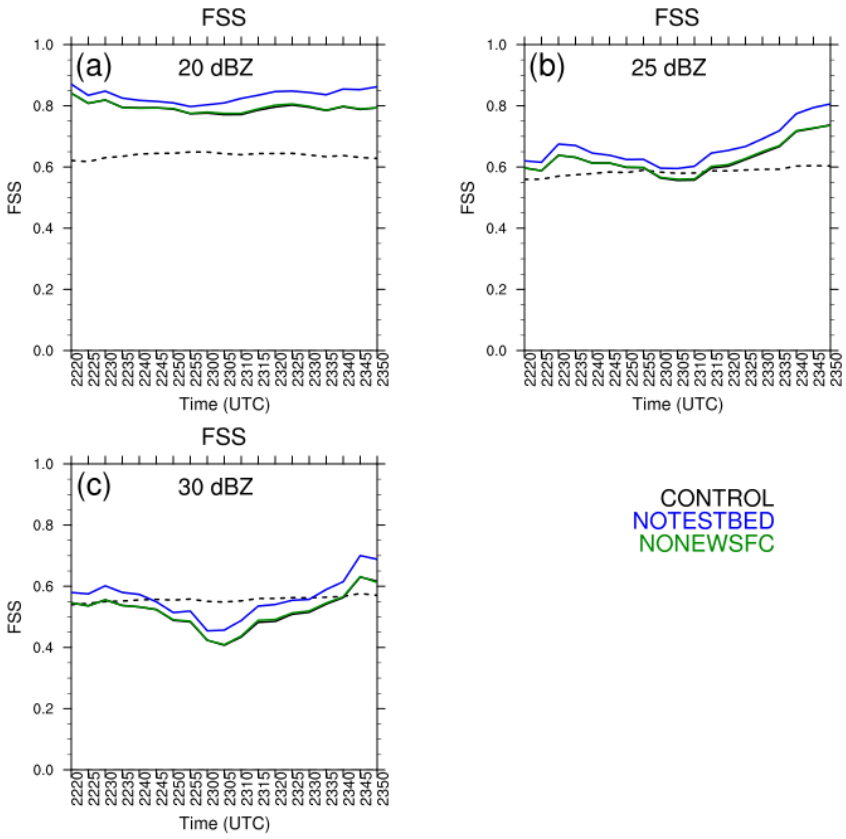


Figure 4.17: As in Figure 4.16, but for the CONTROL, NOTESTBED, and NONEWSFC experiments.

4.4.3 Hail Verification

The maximum estimated size of hail (MESH) can be determined by the WSR-88D radar network using a hail detection algorithm (HDA; Witt et al. 1998). MESH is calculated from a weighted vertical integration of the horizontal reflectivity factor (Z) exceeding 40 dBZ above the melting level. Reflectivity data are first converted into flux values of hail kinetic energy ($\dot{E}$):

$$\dot{E} = 5 \times 10^{-6} \times 10^{0.084Z} W(Z) \tag{4.4}$$

where Z is the horizontal reflectivity factor in dBZ and $\dot{E}$ is in $J\,m^{-2}s^{-1}$ (Waldvogel et al. 1978). The following reflectivity weighting function acts to filter out reflectivity values that often tend to be associated with liquid water:

$$W(Z) = \begin{cases} 0 \quad for\ Z \leq Z_L \\ \frac{Z - Z_L}{Z_U - Z_L}\ for\ Z_L < Z < Z_U \\ 1 \quad for\ Z \geq Z_U \end{cases} \tag{4.5}$$

where Z_L and Z_U are 40 dBZ and 50 dBZ, respectively. Thus, reflectivity values below 40 dBZ are assigned a weight of 0, while reflectivity values exceeding 50 dBZ are assigned a weight of 1.

The significant hail index (SHI) is defined by:

$$SHI = 0.1 \int_{H_0}^{H_T} W_T(H)\dot{E}\ dH \tag{4.6}$$

where H_T is the height of the storm cell and H_0 is the height of the environmental melting level above radar level (ARL). The assumptions are that hail growth only occurs at subzero temperatures and is maximized when the temperature is at or below -20°C, which can be represented by the following temperature-based weighting function:

$$W_T(H) = \begin{cases} 0 \quad for\ H \leq H_0 \\ \frac{H - H_0}{H_{m20} - H_0}\ for\ H_0 < H < H_{m20} \\ 1 \quad for\ H \geq H_{m20} \end{cases} \tag{4.7}$$

where H_{m20} is the height of the -20°C environmental temperature. Both H_0 and H_{m20} are determined using numerical model output for each experiment. Finally, MESH is given by:

$$MESH = 2.54(SHI)^{0.5} \tag{4.8}$$

with units of millimeters. For additional information on the development of the MESH algorithm, see Witt et al. 1998.

Observed MESH swaths are derived using remapped WSR-88D radar data and compared to forecast MESH swaths for each experiment. MESH swaths are computed for the free forecast portion of each experiment, namely from 2220 to 2350 UTC. Figure 4.18 shows the observed MESH swath (left) and forecast MESH swath for the CONTROL experiment (right) with a black contour indicating where MESH values in excess of 25 mm were observed. It can be seen that the observed MESH swath has a smaller areal extent than that of the CONTROL experiment, indicating that the microphysics scheme produced hail over a larger area than what was observed. In addition, the coverage of severe hail (diameter 25 mm or greater) is significantly larger than what was observed, with extraneous swaths of severe hail occurring both to the northwest and northeast of the main hail zone. Lastly, there is an under-prediction of the maximum hail size evident in the CONTROL experiment, which is also evident in the other experiments (Figure 4.19), which is likely a result of the microphysics scheme used. The effects of the choice of microphysics parameterization scheme will be examined in Section 4.4.5.

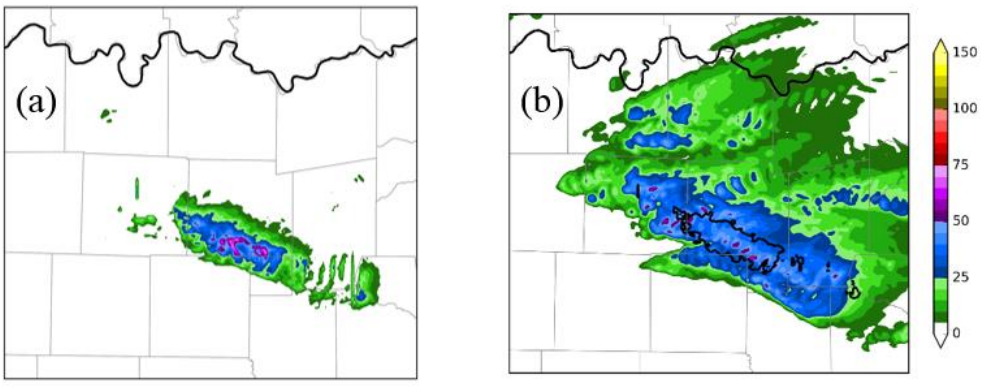


Figure 4.18: a) Observed MESH swath (mm) derived using WSR-88D radar data and b) forecast MESH swath (mm) for the CONTROL experiment. MESH swaths are shown for the free forecast period, namely 2220 to 2350 UTC.

Figure 4.19 shows the forecast MESH swaths for several experiments, with the black contour representing the region where MESH values in excess of 25 mm were observed, corresponding to severe hail. THE CONTROL experiment (Figure 4.19a) forecasts severe hail in regions where severe hail was observed, but, again, there is an over-prediction bias, with more severe hail forecasted by the model than what was actually observed. Significant severe hail (diameter in excess of 50 mm) was also forecasted in the CONTROL experiment; however, some of this hail was forecasted outside the region of observed severe hail. This general pattern holds true for the remaining experiments, with the exception of NORADAR (Figure 4.19c-f). The NORADAR experiment (Figure 4.19b) does not forecast any hail for roughly half of the region where severe hail was observed. Given that no radar reflectivity data were assimilated in this experiment, the model must "spin-up" the storm, with the storm taking roughly 15 to 20 minutes to form (not shown). Given the poor initial representation of the main storm of interest, the storm takes longer to mature, producing the majority of its severe hail later in the forecast period than for the CONTROL experiment. Moreover, little, if any, significant severe hail was forecasted for this experiment. The 88DONLY experiment (Figure 4.19c) has a somewhat smaller swath of forecasted severe hail than the CONTROL experiment, with a larger portion of the swath falling outside the region of observed severe hail. Maximum hail size is under-predicted in the western-most portion of the observed severe hail swath. Additionally, the largest hail predicted falls within the significant severe category, but this hail occurs entirely outside of the region of observed severe hail. The NOTDWR experiment (Figure 4.19d) similarly under-predicts the maximum hail size in the western-most portion of the observed severe hail

swath, but does capture some significant severe hail within the observed severe hail swath. The NOCASA experiment (Figure 4.19e) exhibits significant differences from the CONTROL experiment. Most notable is the area of significant severe hail (indicated by magenta colors) that is forecasted to occur outside the area where severe hail was observed. This pattern was also observed for the 88DONLY experiment, but to a lesser extent. In addition, a swath of significant severe hail was forecasted just north of the observed MESH contour, with this swath extending further east than that of the observations. The northward shift of the NOCASA hail is a result of the forecasted storm track being slightly north of the track in the CONTROL experiment. The eastward extension is likely a result of the supercell moving faster to the east in the NOCASA experiment than in the CONTROL experiment, which was noted previously in the reflectivity fields (Figure 4.11). The NOCASAVR experiment (Figure 4.19f) shows a very similar pattern to the NOCASA experiment, which indicates that the differences between the CONTROL and NOCASA experiment are largely attributed to the denial of the CASA radial velocity data.

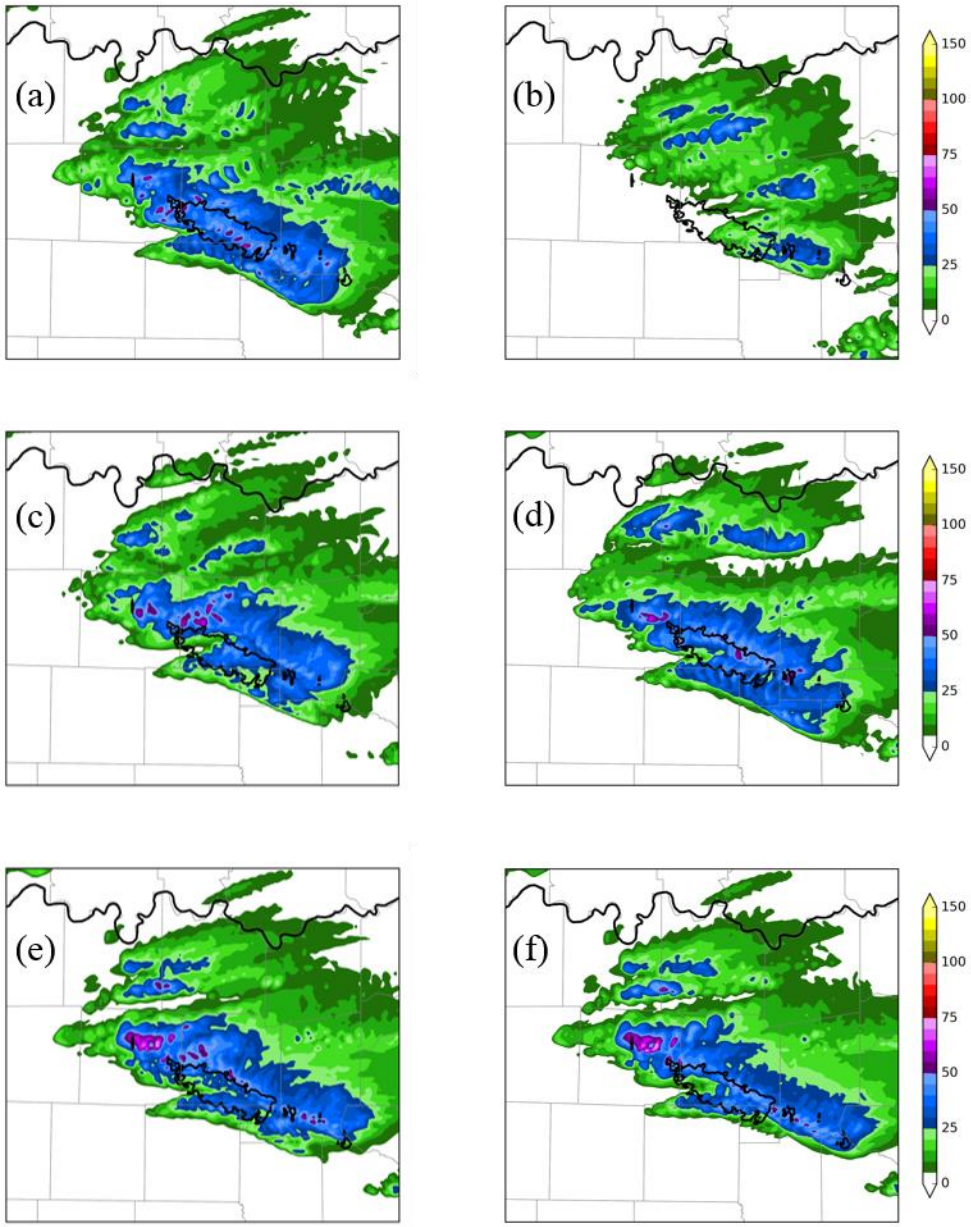


Figure 4.19: Forecast MESH swaths (mm) for the a) CONTROL, b) NORADAR, c) 88DONLY, d) NOTDWR, e) NOCASA, and f) NOCASAVR experiments.

Figures 4.20a and 4.20b show the forecasted MESH swath for the CONTROL and NONEWSFC experiments, respectively. The experiments exhibit a fairly similar spatial coverage of hail, but there is a notable region of significant severe hail that occurs to the southeast of the main observed swath in the NONEWSFC experiment.

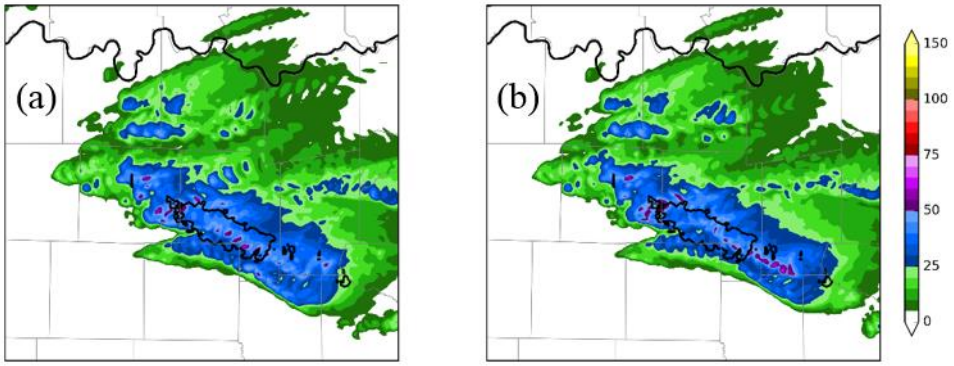


Figure 4.20: As in Figure 4.19, but for the a) CONTROL and b) NONEWSFC experiments.

Forecasts of MESH are verified quantitatively using performance diagrams (Roebber 2009) that are based on the forecast contingency table presented in Table 4.2.

Table 4.2: Contingency Table for Forecast vs. Observations

		Observed	
		Yes	No
Forecast	Yes	Hit	False Alarm
	No	Miss	Correct Null

In performance diagrams, the y-axis represents the probability of detection (POD), while the x-axis corresponds to the frequency of hits (FOH) or success ratio (SR), which is defined as $1 - FAR$, where FAR is the false alarm rate.

$$POD = \frac{Hits}{Hits+Misses} \tag{4.9}$$

$$FAR = \frac{False\ Alarms}{Hits+False\ Alarms} \tag{4.10}$$

$$FOH = SR = 1 - FAR = \frac{Hits}{Hits+False\ Alarms} \tag{4.11}$$

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

$$CSI = \frac{Hits}{Hits+Misses+False\ Alarms} \tag{4.12}$$

Since a perfect forecast has no misses or false alarms, the $POD = 1, SR = 1, and\ CSI = 1$. Therefore, a perfect forecast falls in the upper-right corner of the diagram.

Because severe hail is such a localized phenomenon, observed MESH swaths are modified using a neighborhood threshold, so the forecast is not penalized when hail is forecast within a specified radius of the observed hail. For each grid point, if hail is observed within the specified radius, observations of hail are expanded to include that grid point.

Figure 4.21 presents the performance diagram for 5 mm hail using a neighborhood threshold of 15 km for the radar denial experiments. It can be seen that the CONTROL experiment has the highest POD, indicating that the forecasted region of hail most closely encompasses the region of observed hail. It also has the highest success ratio (SR) of all experiments. This indicates that the inclusion of all available data in the CONTROL experiment results in the best forecast of hail. The NORADAR experiment has the lowest POD and SR out of all the experiments presented, which is due to the aforementioned model "spin-up" time. It can also be seen that 88DONLY has a lower POD and SR than the CONTROL experiment, indicating that the denial of CASA and TDWR radar data degraded the forecast. The NOTDWR experiment has a similar POD and SR as the CONTROL run, whereas NOCASA performs similarly to 88DONLY. Thus, the majority of the differences noted in the 88DONLY experiment are attributed to the denial of CASA

radar data. More specifically, since the NOCASAVR experiment has a similar result as the NOCASA experiment, it can be deduced that the degraded forecast is most attributable to the denial of CASA radial velocity data, which offers high-resolution sampling of the lowest-levels of the storm, below that of the WSR-88D network. The CASAONLY experiment, which indicates the value of the CASA radars in the event of all WSR-88D radars that observe the storm failing during a severe weather event, has a lower POD and SR than the CONTROL experiment.

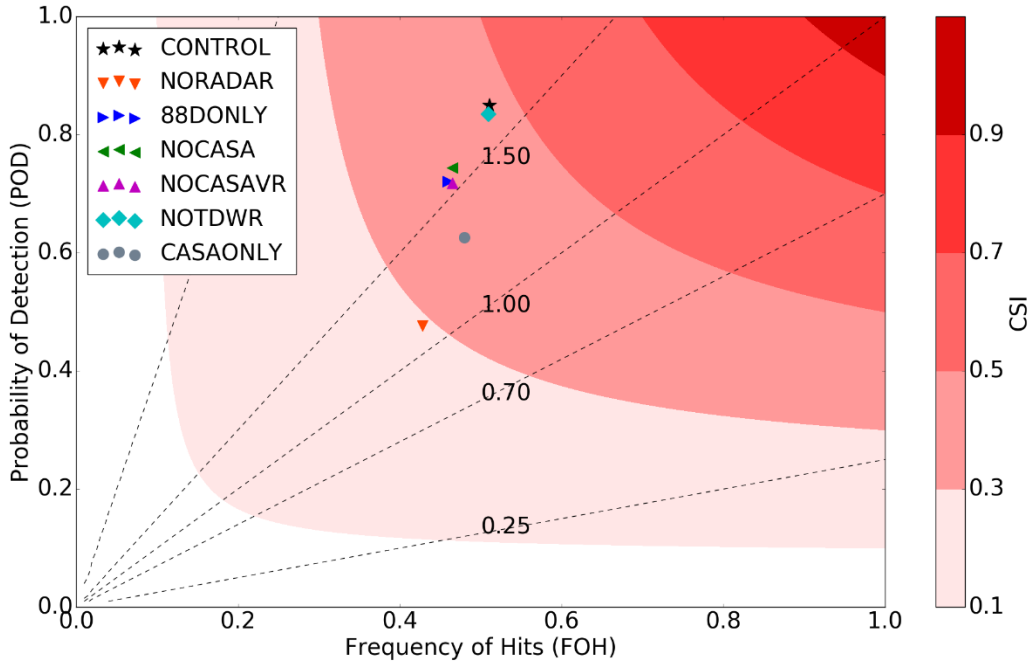


Figure 4.21: Performance diagram for radar data denial experiments, using a hail size of 5 mm and neighborhood threshold of 15 km.

Likewise, Figure 4.22 shows the performance diagram for 25 mm (severe) hail using the same 15 km threshold. As before, the CONTROL experiment has the highest POD, indicating that the inclusion of all available data results in the forecast that best encompasses the region of severe hail. NORADAR has the lowest POD, consistent with

the forecasted severe hail only marginally falling within the observed region of severe hail. NORADAR also has the smallest SR (i.e., highest FAR), which is in response to the reduced number of hits and added sensitivity to false alarms. The 88DONLY experiment has a lower POD and SR than the CONTROL experiment, indicating that the addition of CASA and TDWR data improves the prediction of this hail event. Unlike for the smaller hail size, there is a more notable decrease in POD for the NOTDWR experiment; this is counter-balanced, however, by an increase in SR. However, the resultant CSI for the NOTDWR experiment is slightly higher than for the CONTROL experiment, so overall the NOTDWR experiment performs slightly better than the CONTROL. The NOCASA experiment shows a reduced POD and reduced SR, when compared to the CONTROL experiment. However, unlike for the smaller hail size, there is a lower POD and slightly lower SR for the NOCASAVR experiment when compared to the NOCASA experiment, which indicates that the inclusion of CASA reflectivity data, without radial velocity data, results in a slight degradation of the forecast. As before, the CASAONLY experiment has a lower POD than the CONTROL experiment, suggesting that the CASA data alone is not enough to produce a simulation that captures the majority of the severe hail that occurred. It is also worth noting that the majority of experiments, with the exception of the CASAONLY and NORADAR experiments, suffer from an over-prediction bias (values above 1.0). This is consistent with the MESH swaths presented earlier, as the areal coverage of severe hail in these experiments exceeded that of the observations.

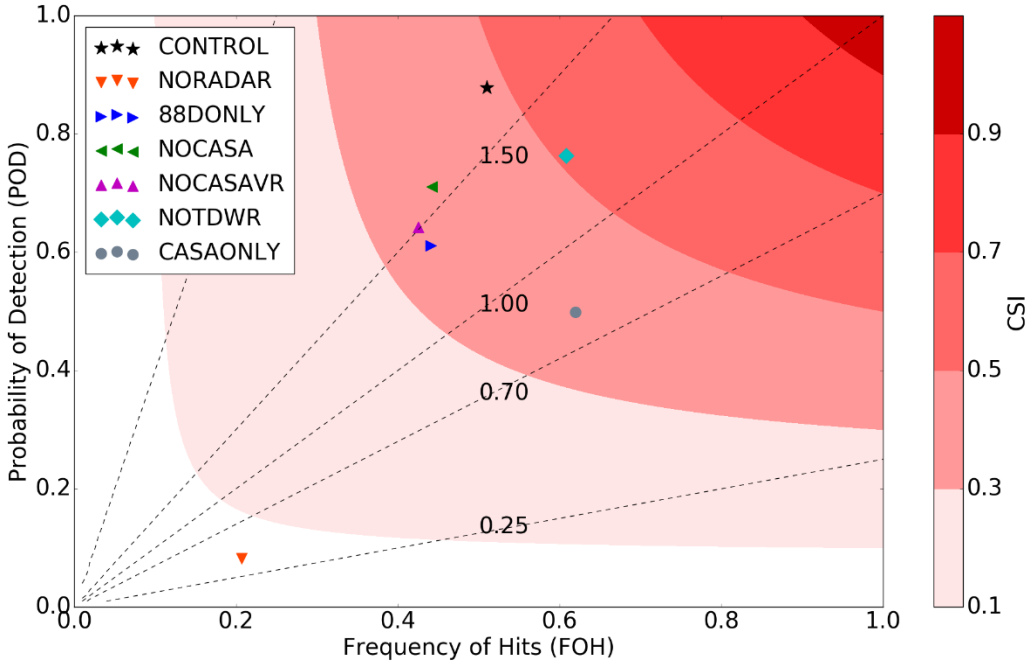


Figure 4.22: As in Figure 4.21, but for a hail size of 25 mm.

Figure 4.23 shows the performance diagram for the surface data denial experiments, using a hail size of 25 mm and neighborhood threshold of 15 km. The NOTESTBED experiment is an extension of the 88DONLY experiment in that it denies non-conventional surface data in addition to the non-conventional radar data (CASA and TDWR). Thus, this experiment has a similar performance to the 88DONLY experiment. The NONEWSFC experiment has a slightly lower POD than the CONTROL experiment, along with a slightly higher SR. Therefore, the CSI for this experiment is not significantly different from the CONTROL experiment. The individual data denial experiments (NOGST, NOUNDERSTORY, NOCWOP, NOWXBUG, and NOCWOPWXBUG) all experience fairly similar PODs and SRs, and it is difficult to definitively say which experiments exhibit the best performance due to the close clustering of points. A similar

pattern is seen in the 5 mm hail performance diagram (not shown), although the range of values seen is less disperse for that case.

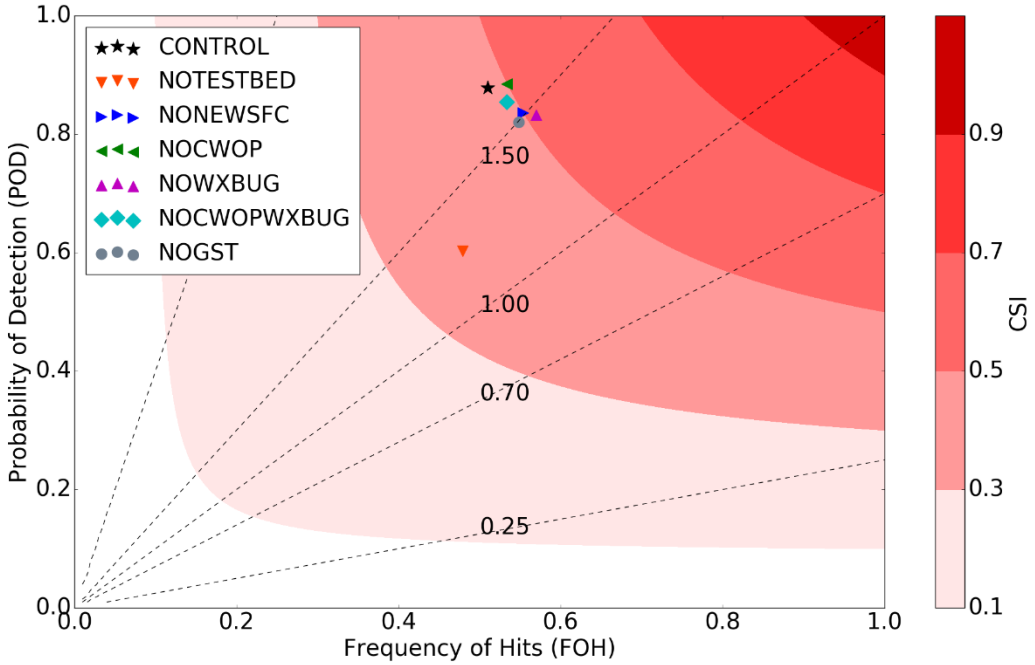


Figure 4.23: As in Figure 4.22 but for surface data denial experiments.

4.4.4 Surface-Level Forecast Verification

Forecast performance is also evaluated by considering the root mean square difference (RMSD) of surface fields such as temperature and dew point temperature. Twelve independent stations (i.e., not assimilated in the 3DVAR analysis system) are used in the RMSD calculations, including 10 ASOS stations and 2 Oklahoma Mesonet stations, which are shown in Figure 4.24. These stations were chosen because the data from these networks are considered more reliable than other surface networks. The analysis or forecast is linearly interpolated to the observation location in order to compute the RMSD value.

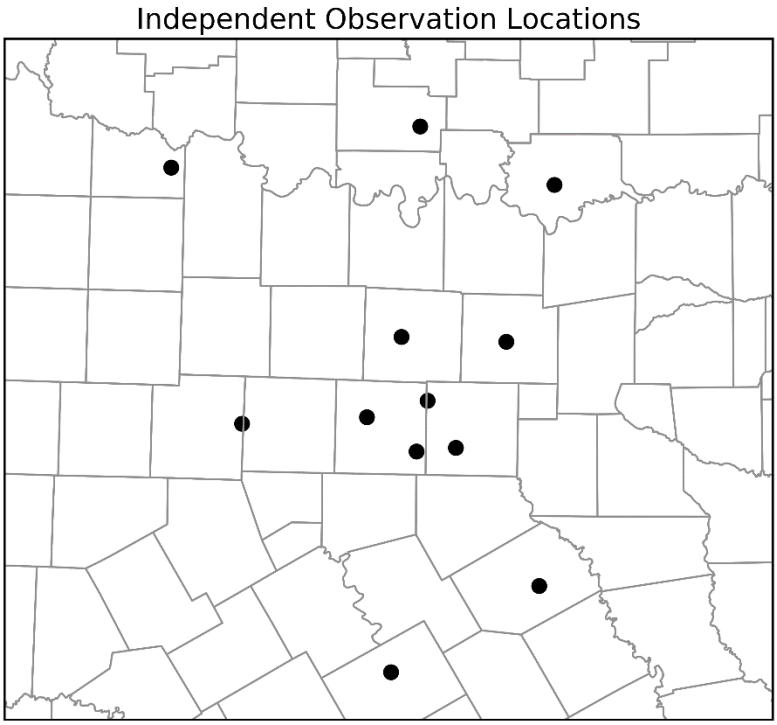


Figure 4.24: Locations of the 10 ASOS and 2 Oklahoma Mesonet stations that are denied for verification purposes.

Figure 4.25a shows the 2 meter temperature RMSD for the CONTROL, NORADAR, CASAONLY, NOCASA, and 88DONLY experiments. The RMSD at the beginning (0 min) represents the difference between the observations and the model background. The vertical line at 30 min corresponds to the beginning of the last data assimilation cycle and the free forecast period. All five experiments start out with a very similar RMSD as each experiment assimilates all available surface data. However, as the forecast progresses, differences among the experiments become more apparent. The CONTROL experiment has the lowest RMSD value for the middle of the free forecast period, roughly from 2300 to 2320 UTC (70 to 90 min). The 88DONLY experiment has the lowest RMSD at the end of the free forecast period. The RMSD is slightly higher for the NOCASA experiment, which is followed by CONTROL. Thus, the inclusion of the

CASA and TDWR radar data results in a slight increase in the 2 m temperature RMSD at the end of the forecast period. The RMSD for NORADAR increases considerably around 75 min into the forecast period (2305 UTC). CASAONLY and NORADAR have the highest RMSD at the end of the free forecast, most likely due to the poor initial handling of the storm of interest in these simulations. It is likely that these simulations do not adequately capture the strength of the cold pool and associated gust front. Figure 4.25b shows the 2 m dew point temperature RMSD. As for temperature, all experiments start out with a similar RMSD owing to the inclusion of all available surface data in this set of experiments. The differences in dew point RMSD are not as clear as for temperature at the end of the free forecast period, with the CONTROL and NOCASA experiments having the lowest RMSD values. Therefore, it appears that the radar data have a stronger influence on the resulting surface temperature field than the surface dew point field, despite the insertion of hydrometeors and humidity aloft in the cloud analysis.

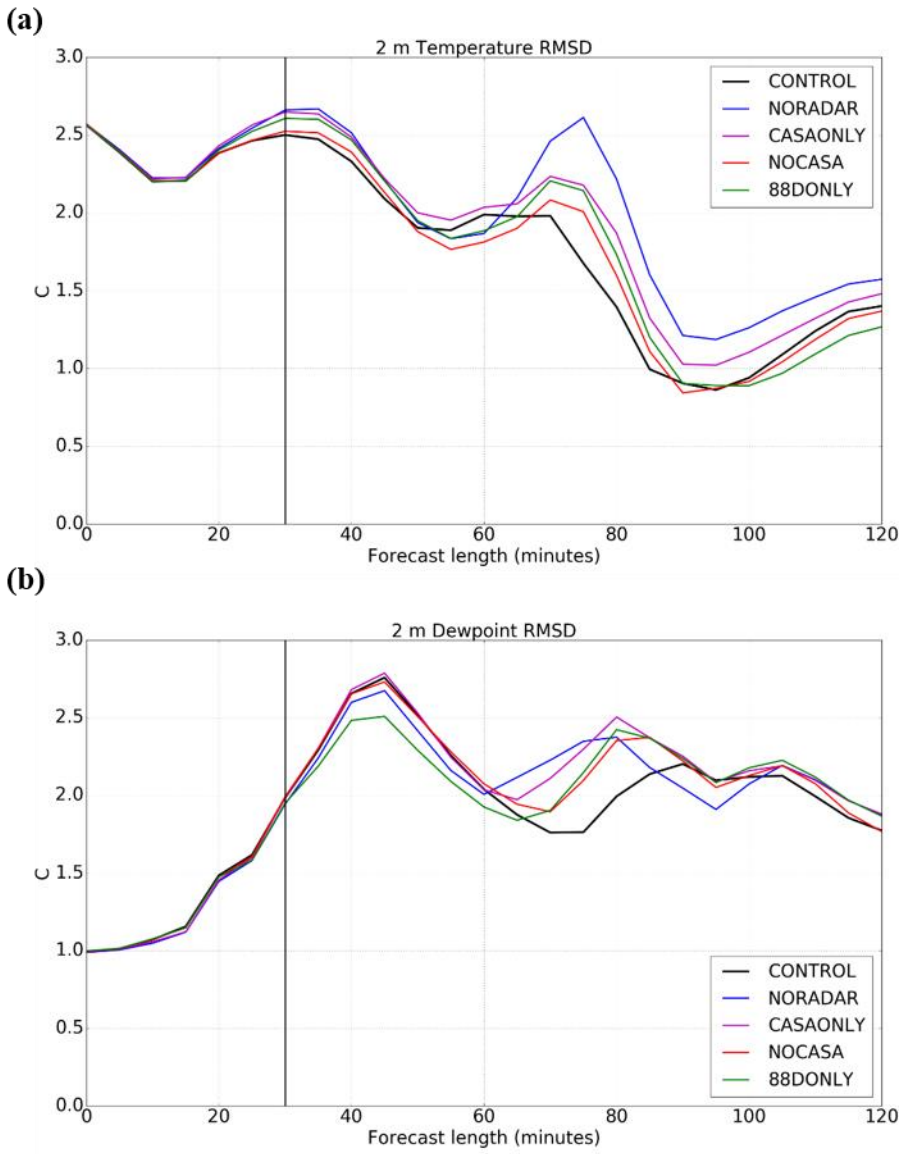


Figure 4.25: Root mean square difference (RMSD) for a) 2 m temperature and b) 2 m dew point temperature. The vertical line at 30 minutes represents the start of the free forecast.

The 2 m temperature RMSD for the CONTROL, NONEWSFC, NOCWOPWXBUG, NOCWOP, and NOWXBUG experiments are shown in Figure 4.26a. The NOCWOP experiment has the lowest RMSD at the end of the free forecast period, followed by CONTROL and NOWXBUG. NONEWSFC and NOCWOPWXBUG exhibit an increase in RMSD around 50 min into the forecast period

(2240 UTC) that is not as prevalent in CONTROL, NOCWOP, and NOWXBUG. Thus, the inclusion of the non-conventional surface observations results in a reduced RMSD at the end of the free forecast period. The RMSD begins to diverge faster for this group of experiments than for the radar experiments owing to the varied group of surface data assimilated.

There is a more distinct difference in the RMSD pattern for dew point temperature (Figure 4.26b). The dew point RMSD exhibits greater spread during the assimilation period than does temperature. The CONTROL experiment has the highest RMSD error at the beginning and end of the free forecast period, although NOCWOP is higher than CONTROL for a portion of the free forecast period. NONEWSFC has the lowest RMSD throughout the entirety of the free forecast period. The NOCWOPWXBUG experiment exhibits a similar pattern to the NONEWSFC experiment, indicating that the majority of differences seen in the NONEWSFC experiment are attributed to the CWOP and WXBUG observations. Since both the NOCWOP and NOWXBUG experiments exhibit an RMSD at the end of the free forecast period that is slightly lower than for the CONTROL experiment, the combination of these observation types is likely resulting in the differences noted in NONEWSFC. Although the NOUNDERSTORY and NOGST experiments are not shown here, these experiments have an RMSD value similar to that of the CONTROL experiment throughout the forecast period, indicating that the GST and Understory data do not result in significant differences for the dew point temperature field. Thus, the inclusion of non-conventional surface data influences the forecast dew point evolution more so than temperature.

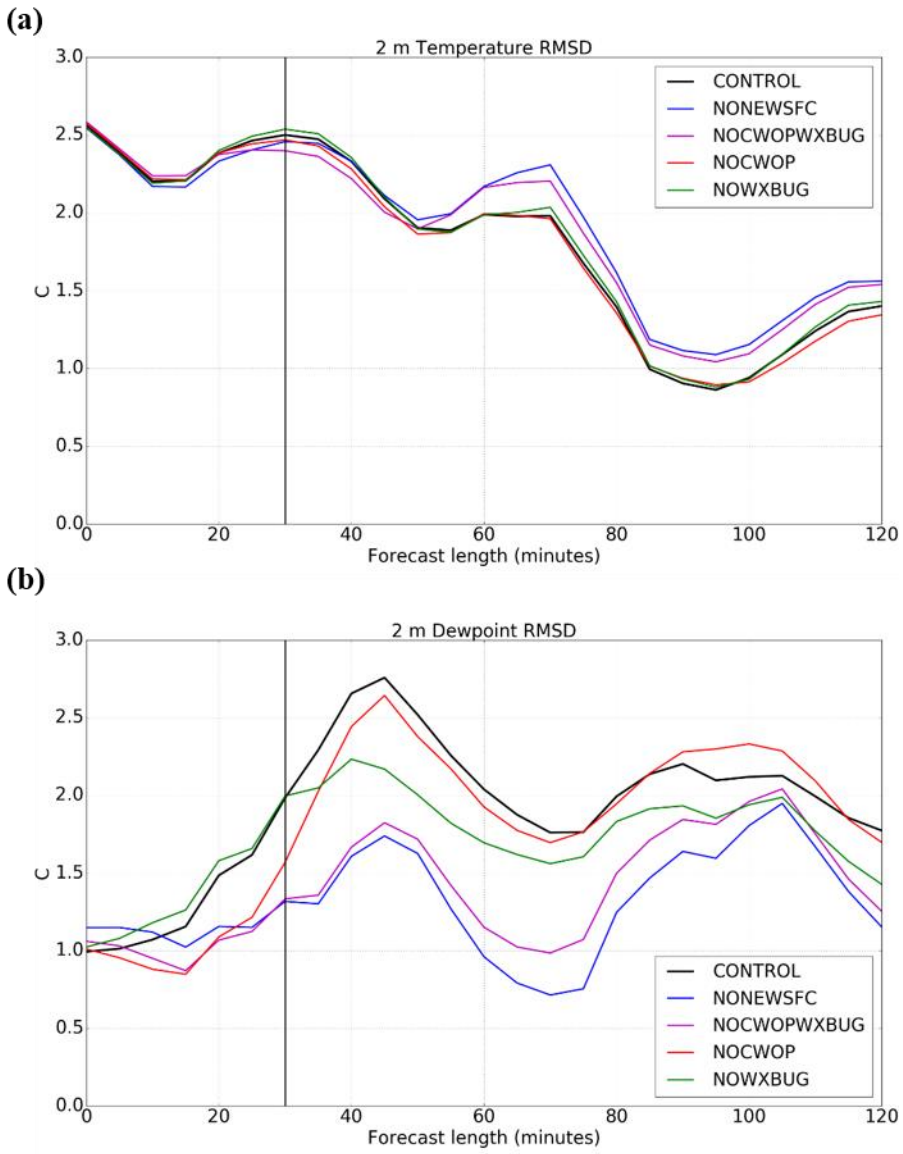


Figure 4.26: As in Figure 4.25, but for the surface data denial experiments.

Figure 4.27 shows bias for the a) 2 m temperature and b) 2 m dew point temperature fields, where bias is defined as model minus observations, for the CONTROL, NONEWSFC, NOCWOWXBUG, NOCWOP, and NOWXBUG experiments. The 2 m temperature bias is below -2.0 °C at the beginning of the forecast period. All experiments exhibit a fairly similar pattern of bias throughout the forecast period, with bias increasing towards zero throughout the forecast period. Bias for the 2

m dew point temperature field is lower for NONEWSFC throughout the forecast period, with CONTROL exhibiting a positive dew point bias through the majority of the forecast period. The bias exhibits a similar pattern for the NONEWSFC and NOCWOPWXBUG experiments, indicating that the bias is mostly in response to the WeatherBug and CWOP observations. Since NOWXBUG and CONTROL are similar throughout the forecast period, the majority of the bias in the dew point field is introduced by the CWOP observations.

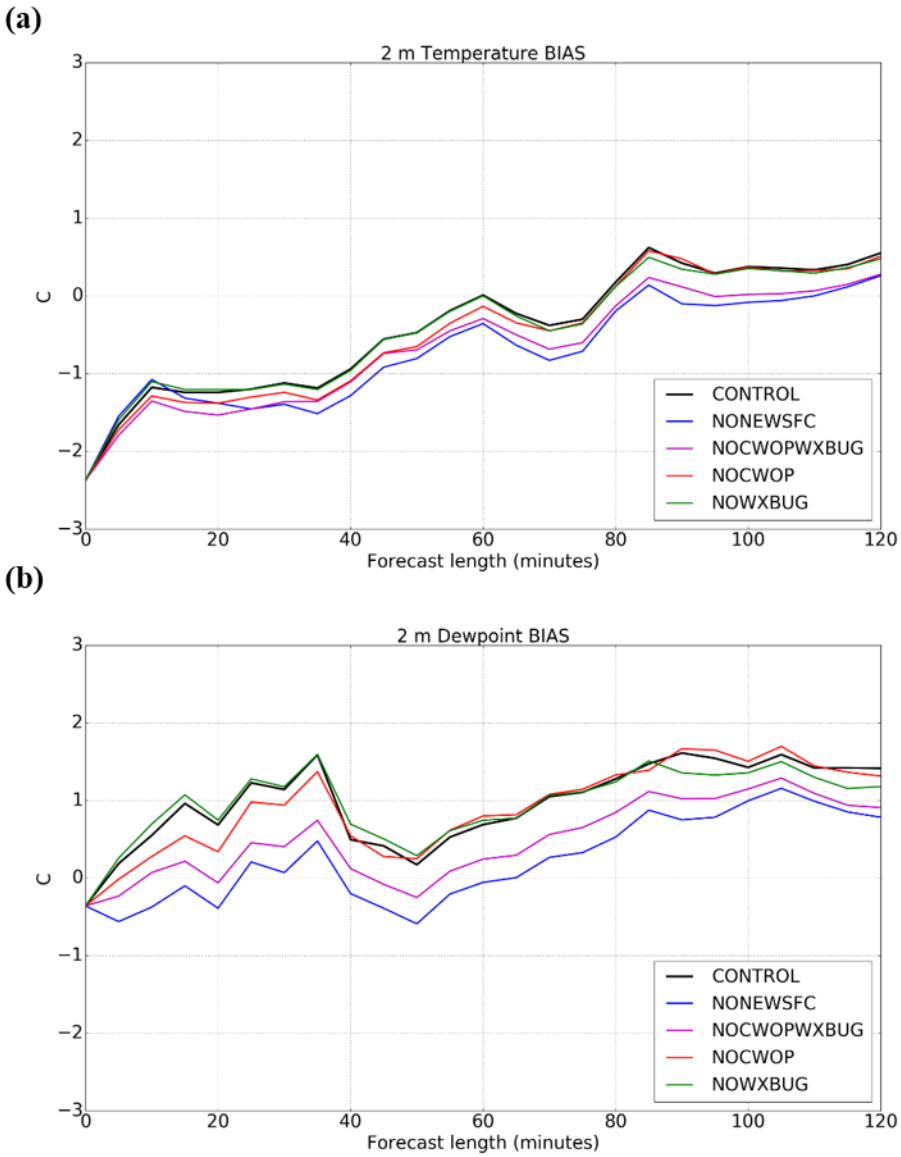


Figure 4.27: Bias for a) 2 m temperature and b) 2 m dew point temperature for the surface data denial experiments.

Figure 4.28 shows the background temperature field and wind vectors at 2150 UTC. Independent observations from the 12 verification stations are overlaid. A 31.1°C observation from KFTW (Fort Worth) coincides with a background value of roughly 25.2°C. Moreover, the observed wind at this station is southerly, while the wind in the background field surrounding the observation is northerly. This indicates that there is a phase shift in the model background field, with the front placed further south than in

reality. This phase shift is also largely responsible for the large initial RMSD value for the 2 m temperature field, as can be seen in Table 4.3. In addition, the model minus observed value is negative for all 12 verification stations, which indicates that the model background is colder than observations prior to any data assimilation. This is consistent with the cold bias seen in Figure 4.27a.

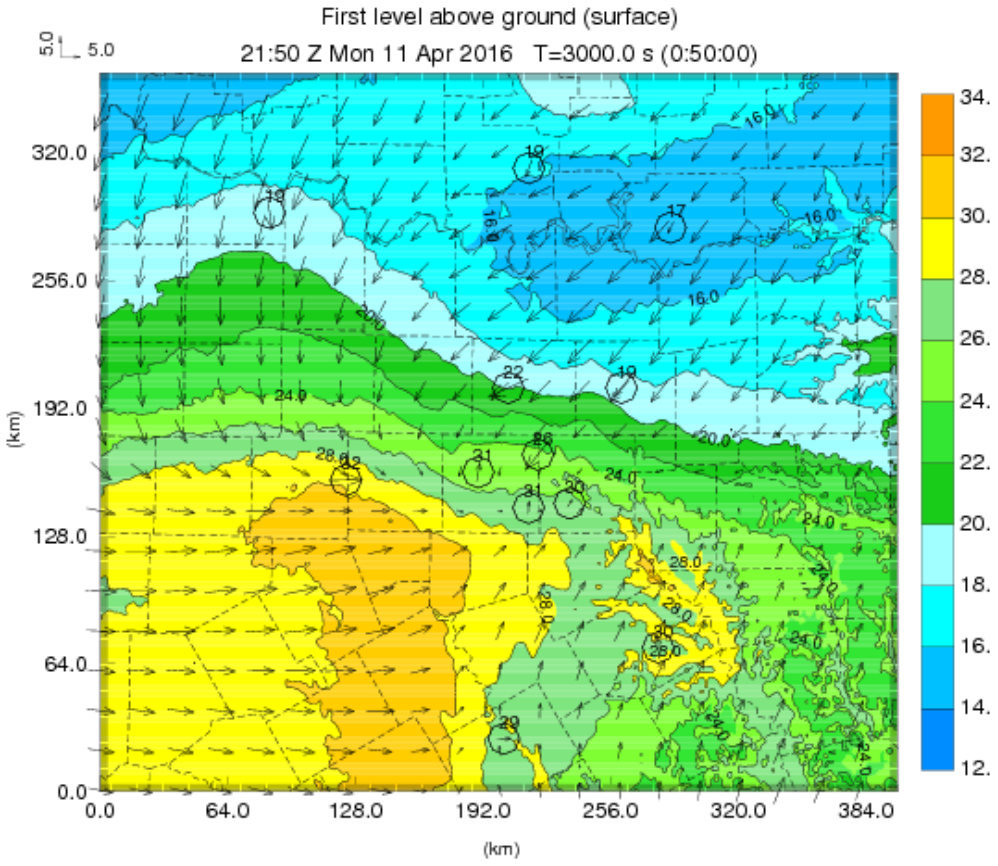


Figure 4.28: Background temperature field (°C) and wind vectors (m/s) at 2150 UTC. Independent (i.e., not assimilated) temperature and wind observations are overlaid.

Table 4.3: Model background (RAP) vs. observations at 2150 UTC.

Station	Model (°C)	Observed (°C)	Model-Observed (°C)
ARD2	16.2	19.1	-2.9
DURA	15.0	16.8	-1.8
KDFW	24.3	25.6	-1.3
KDTO	20.1	21.7	-1.6
KFTW	25.2	31.1	-5.9
KRBD	26.5	30.0	-3.5
KTKI	18.8	18.9	-0.1
KCRS	26.9	30.0	-3.1
KMWL	29.4	31.7	-2.2
KSPS	18.8	19.4	-0.7
KGKY	27.0	30.6	-3.6
KACT	27.5	29.4	-1.9

Figure 4.29 shows specific humidity of vapor (q_v) perturbations at the surface. The perturbations are determined by subtracting the q_v value from the CONTROL experiment from that of the NONEWSFC, NOCWOPWXBUG, NOCWOP, and NOWXBUG experiments, respectively. The NONEWSFC, NOCWOPWXBUG, and NOWXBUG experiments all exhibit a region of positive q_v perturbations west of the DFW metro at 2220 UTC, indicating that the CONTROL experiment is drier in this region. Since all three experiments deny the assimilation of WeatherBug observations, while CONTROL and NOCWOP do not, it can be seen that the WeatherBug observations

are introducing a dry region to the model background field in this region. Conversely, the NONEWSFC, NOCWOPWXBUG, and NOCWOP experiments exhibit a negative q_v perturbation just to the east of the aforementioned region, with this pattern not as prevalent in NOWXBUG. This indicates that a moist region is introduced to the model background field by the CWOP observations. Since more verification stations are collocated with the moist area caused by CWOP observations than the dry region caused by WeatherBug observations, the bias is largely driven by the CWOP observations, which is consistent with the moist bias seen in Figure 4.27b.

Figures 4.30 and 4.31 show the forecasted surface temperature and wind fields, along with the observed temperature and wind at the 12 verification station for 2220 UTC, 2240 UTC, 2300 UTC, and 2320 UTC. The placement of the combined cold front and gust front is indicated by the temperature gradient and wind shift. As time progresses through the forecast period, more of the verification stations are located behind the cold front, outside of the region of greatest temperature gradient. The frontal boundary passage is largely responsible for the notable decreases in RMSD evident in Figures 4.25 and 4.26.

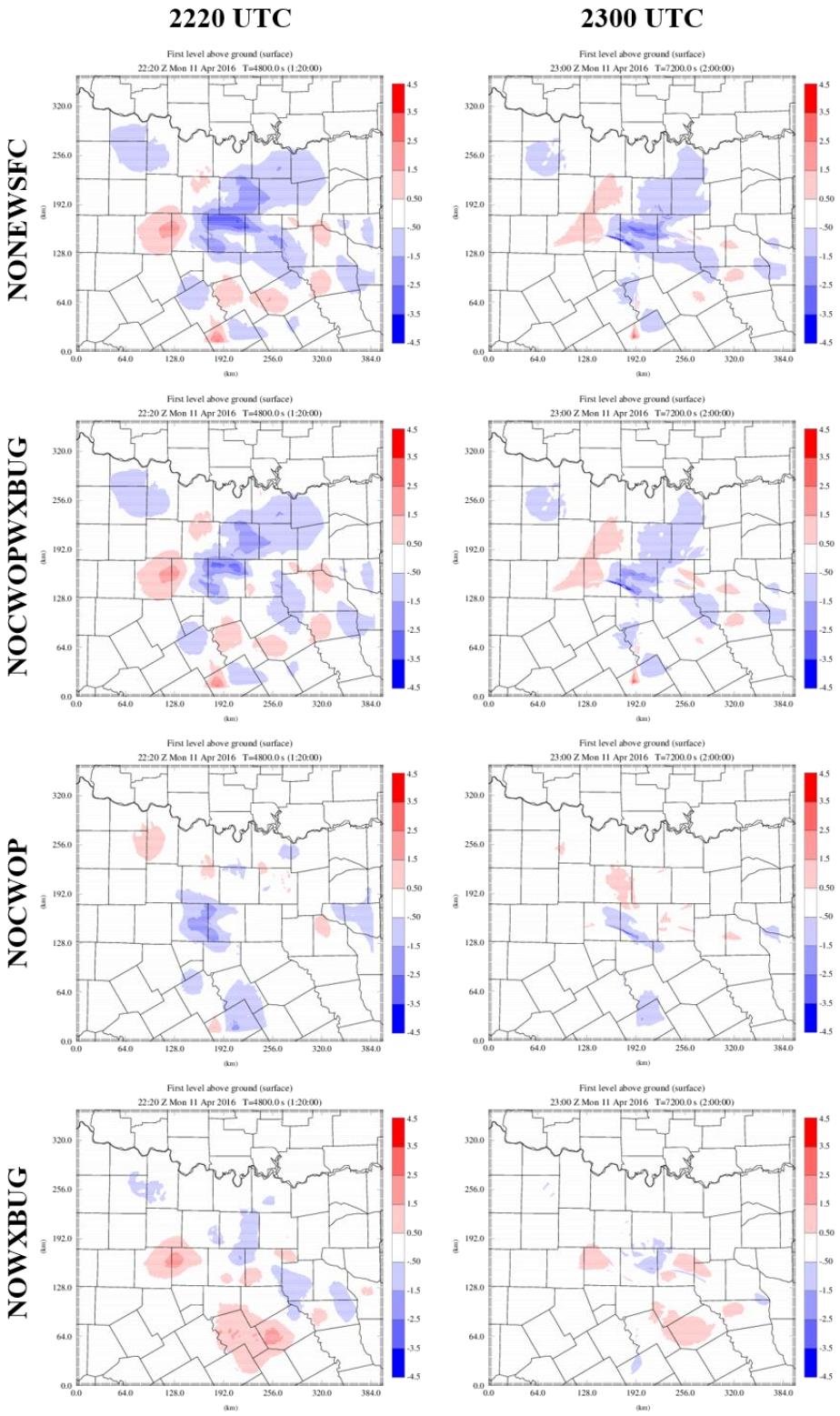
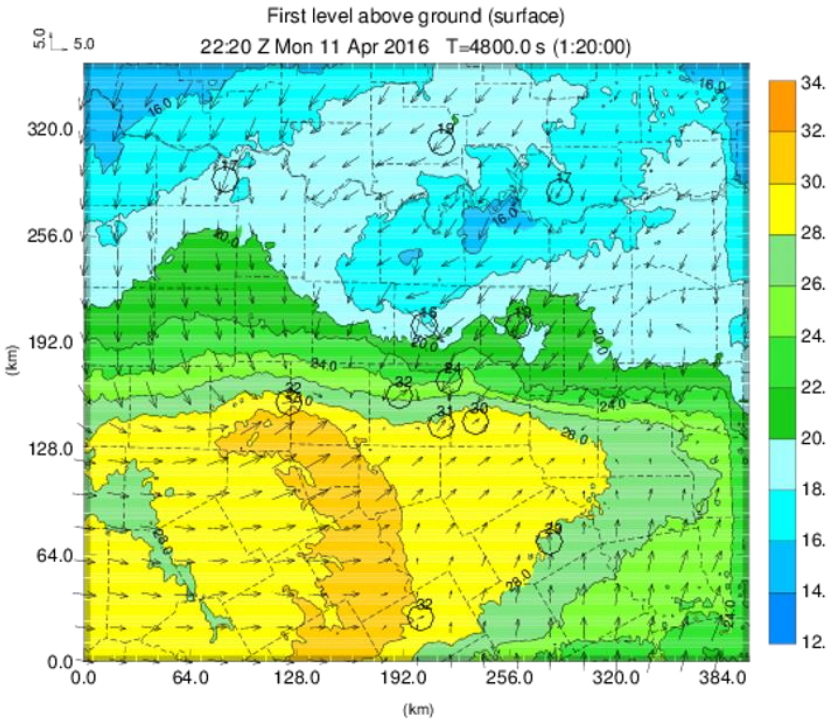


Figure 4.29: Specific humidity of vapor (q_v) differences at the surface, which are determined by subtracting the q_v value for the CONTROL experiment from the value for each experiment.

2220 UTC



2240 UTC

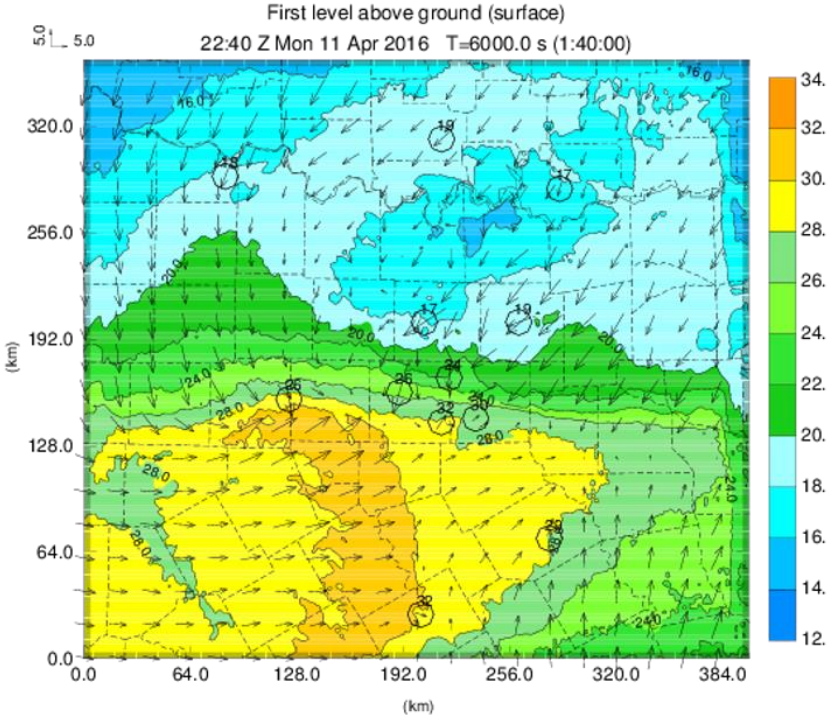
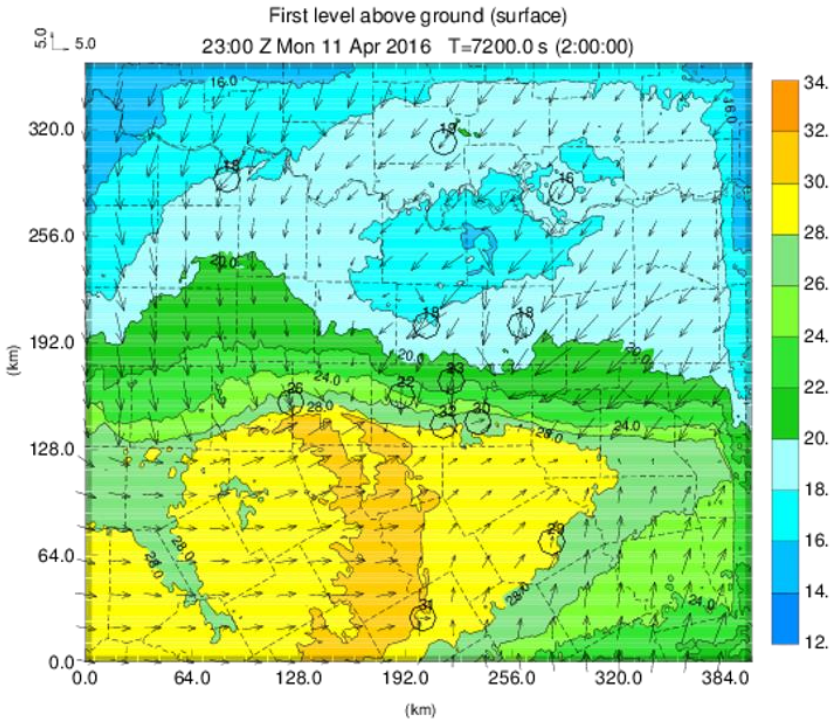


Figure 4.30: Forecasted temperature (°C) and wind (m/s) fields at the surface for 2220 UTC and 2240 UTC, with observed temperature and wind fields overlaid for the 12 verification stations.

2300 UTC



2320 UTC

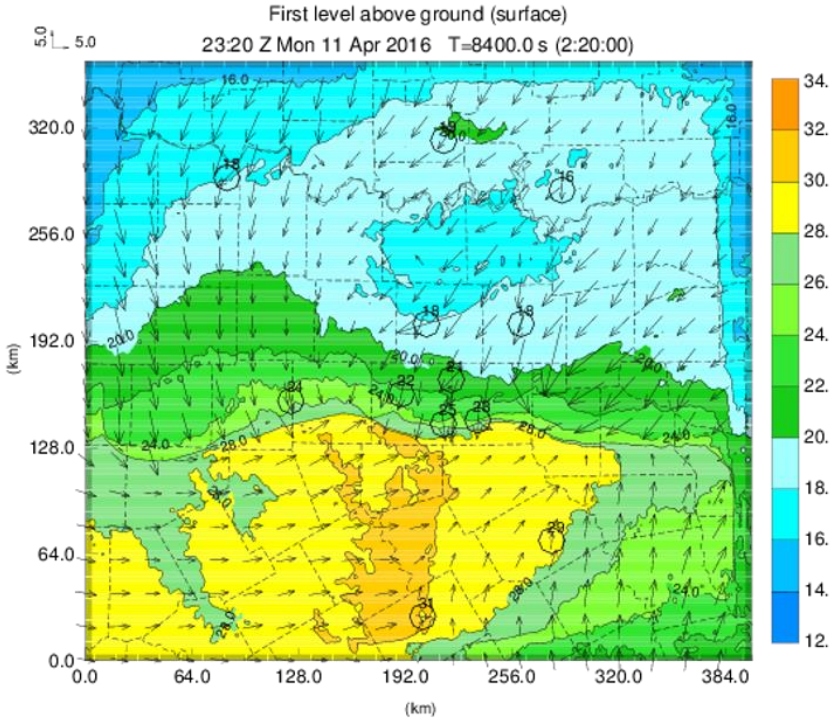


Figure 4.31: As in Figure 4.30, but for 2300 UTC and 2320 UTC.

4.4.5 Single vs. Double Moment Microphysics

The microphysics scheme used in all of the experiments presented in previous sections is the single-moment microphysics scheme of Milbrandt and Yau (2005). While the CONTROL experiment has a MESH POD above 85% for the 15 km neighborhood threshold (Figures 4.21 – 4.23), the SR is around 50% owing to the general over-prediction bias and corresponding large number of false alarms. Although evaluating the results for varying microphysics schemes was not the main focus of this work, the aforementioned over-prediction bias motivated the comparison with the double-moment Milbrandt and Yau microphysics scheme. The single-moment microphysics scheme predicts the mixing ratio (i.e., mass) for each of the hydrometeor species included, while keeping the intercept parameter to a constant value. On the other hand, the double-moment scheme predicts the number concentration, along with mixing ratio, for each hydrometeor type included. As a consequence of this increased complexity (and computational cost), it is generally expected that double-moment schemes offer superior performance to single-moment schemes. For instance, Dawson et al. (2010) considered the 3 May 1999 tornadic supercell that impacted Oklahoma and found that the use of multi-moment microphysics schemes resulted in a weaker cold pool than when the single-moment scheme was used, owing to improved handling of evaporation and size sorting. Moreover, Stratman and Brewster (2017) found that forecasts of tornadic supercells were often improved when using multi-moment microphysics schemes, although in several metrics single-moment schemes provided similar, if not superior, performance to the multi-moment schemes. Further improvements might be expected for a three-moment microphysics scheme, which predicts the shape parameter in addition to the mass field

and number concentration. However, the three-moment scheme is not considered here. The microphysics sensitivity experiments performed are described in Table 4.4. The CONTROL experiment uses the Milbrandt and Yau single-moment scheme, while CTLDOUBLE uses the double-moment scheme. Both experiments assimilate all available surface, radar, and upper-air data.

Table 4.4: Microphysics Sensitivity Experiments Performed

Experiment	Conventional surface data	Non-conventional surface data	88D, CASA, and TDWR data	Upper-air profiles	Microphysics Scheme
CONTROL	All	All	All	All	Milbrandt and Yau single-moment
CTLDOUBLE	All	All	All	All	Milbrandt and Yau double-moment

Figure 4.32 shows the wind and reflectivity field roughly 2 km AGL for the single-moment scheme (CONTROL) and double-moment scheme (CTLDOUBLE). For the final analysis time (2220 UTC), the CTLDOUBLE scheme has the same general reflectivity pattern as the CONTROL experiment, although the reflectivity values are generally lower for CTLDOUBLE. This is particularly notable in Denton County, where reflectivity values above 65 dBZ are present in CONTROL, while CTLDOUBLE has reduced reflectivity with a small area peaking at around 55 dBZ. By 2300 UTC, a hook echo is present in Collin County for both experiments, although there are still reduced reflectivity values in the CTLDOUBLE experiment. By 2330 UTC, the CTLDOUBLE experiment has a hook echo that is not as well-defined as for the CONTROL experiment.

Additionally, the reflectivity values in the hook region are again reduced from that of the CONTROL experiment.

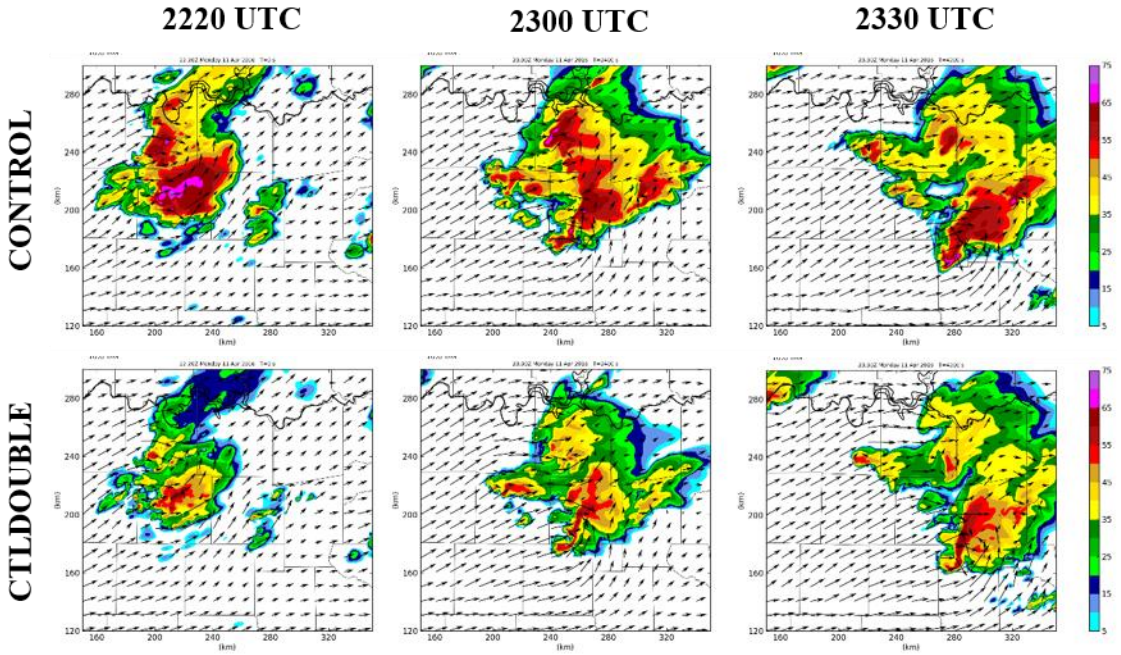


Figure 4.32: Comparison of the reflectivity field roughly 2 km AGL for the single-moment microphysics scheme (CONTROL) and double-moment microphysics scheme (CTLDOUBLE). Both experiments assimilate all available data.

Figure 4.33 shows the forecast MESH swath for the a) CONTROL and b) CTLDOUBLE experiments. It is clear that the areal coverage of the MESH swath is reduced for the CTLDOUBLE experiment, both for all hail (5 mm) and severe hail (25 mm). However, the severe hail predicted in CTLDOUBLE occurs predominantly outside of the region of observed severe hail. The CTLDOUBLE experiment has a more compact hail core, although this core is positioned about 10 to 15 km to the north of the core found in the CONTROL experiment and observations. Thus, the areal extent of the severe hail predicted in CTLDOUBLE is improved over CONTROL, although there is a displacement error present in CTLDOUBLE.

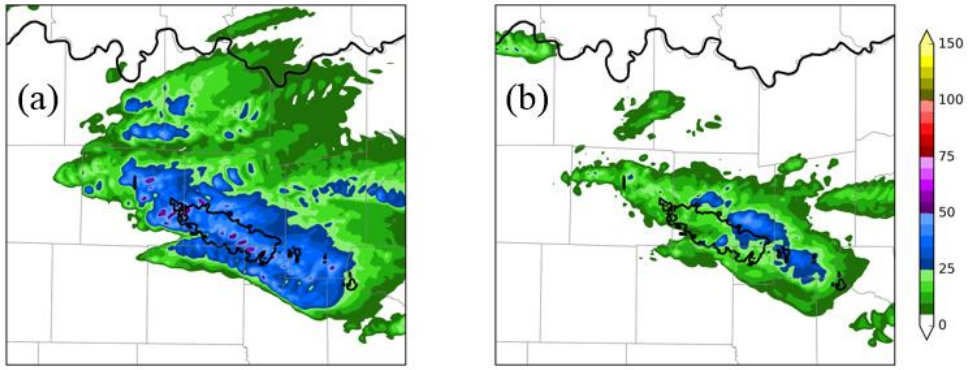


Figure 4.33: Forecast MESH swath (mm) for the a) CONTROL and b) CTLDOUBLE experiments. Observed MESH above 25 mm is contoured in black.

Figure 4.34 shows the performance diagram using a neighborhood threshold of 15 km for the CONTROL and CTLDOUBLE experiments. When considering 5 mm hail, the CONTROL experiment has a POD of roughly 0.85 with a SR of around 0.5. Conversely, the CTLDOUBLE experiment has a lower POD of around 0.55 and a higher SR of approximately 0.75 for a hail size of 5 mm. This is consistent with what is shown in the MESH swaths (Figure 4.33), as the reduced number of false alarms would increase the SR. When using a hail size threshold of 25 mm, the POD for CTLDOUBLE is significantly lower than it is for CONTROL (approximately 0.20), as the majority of the severe hail observed falls outside of the observed region with the 15 km neighborhood allowance. However, the SR is again higher for CTLDOUBLE owing to the reduced spatial coverage of severe hail in this experiment.

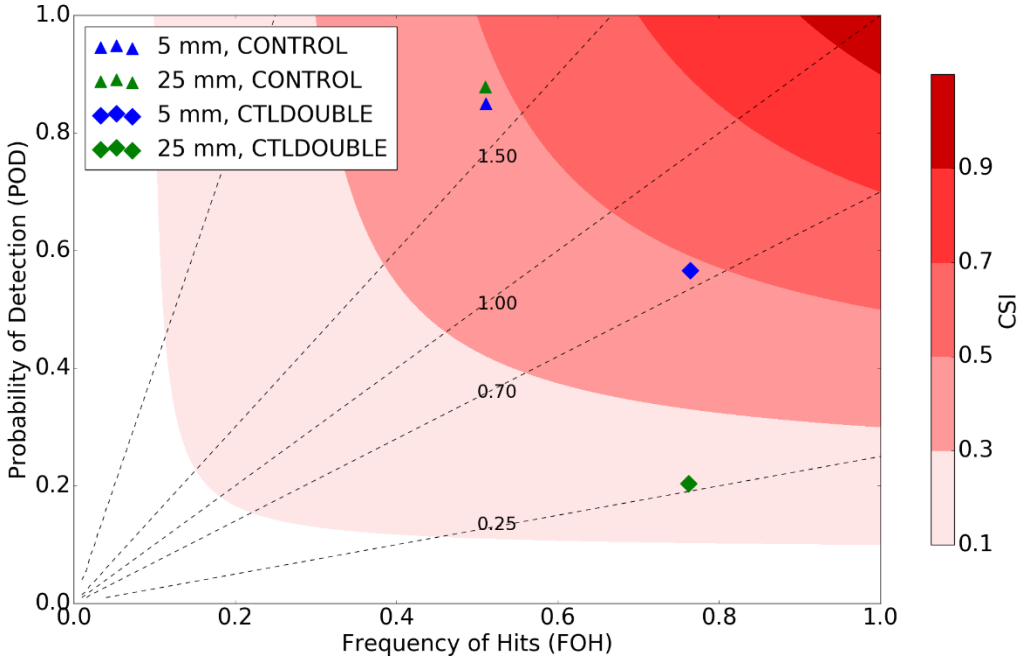


Figure 4.34: Performance diagram comparing MESH forecasts from the CONTROL and CTLDOUBLE experiments, using a neighborhood threshold of 15 km.

The RMSD metric for the CONTROL experiment was lower than the CTLDOUBLE experiment for 2 m temperature (Figure 4.35a) for the entirety of the free forecast period. In addition, the RMSD for 2 m dew point temperature (Figure 4.35b) is lower in CONTROL than CTLDOUBLE from approximately 10 min to 1 hour into the free forecast. Figure 4.36 shows potential temperature perturbations at the surface for a) 2220 UTC, b) 2300 UTC, and c) 2340 UTC. These perturbations are defined as $\theta_{double} - \theta_{single}$. It is evident that the cold pool generated by the double-moment scheme is warmer than the one produced when using the single-moment scheme. This finding is consistent with Dawson et al. (2010), which found warmer and smaller cold pools when using multi-moment microphysics schemes owing to improved handling of evaporation and drop size sorting. At 2340 UTC, the cold pool is located further south

in the CONTROL experiment than in the CTLDOUBLE experiment, which is shown by a band of positive potential temperature perturbations in c). The observed location of the gust front can be deduced from Figure 4.36d, which shows the 0.5 degree tilt from the KFWS WSR-88D radar at 2340 UTC. The gust front placement is more accurate in the CONTROL experiment, which used the single-moment scheme, although both experiments offer a good prediction of the cold pool placement.

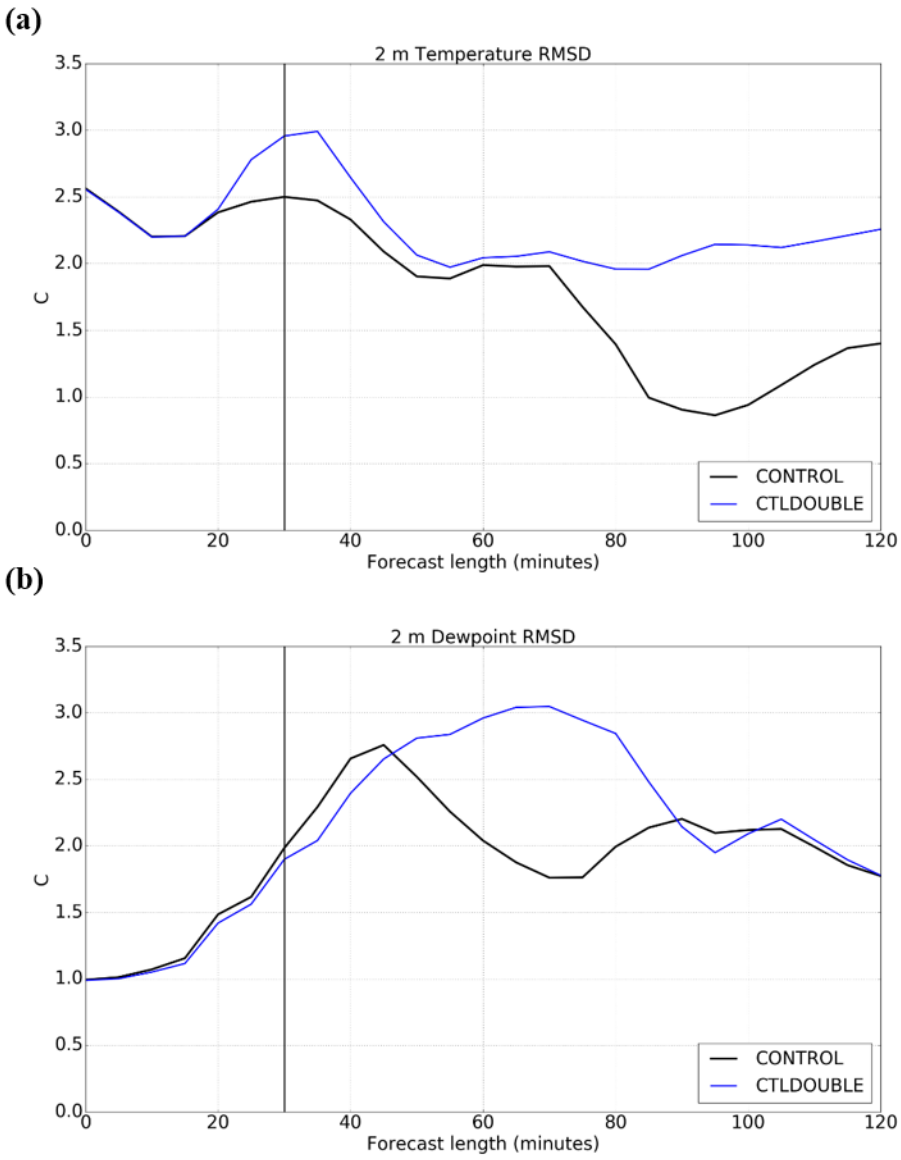


Figure 4.35: As in Figure 4.25, but for the CONTROL and CTLDOUBLE experiments.

Together, this indicates that the single-moment microphysics scheme used performed better than the double-moment version. While it is generally expected that a double-moment scheme would provide superior results to a single-moment scheme, the results were mixed for this case. While the over prediction of reflectivity outside the primary supercell and over prediction of areal extent of large hail within that cell was significantly reduced when using the double-moment scheme (MY2), the MY2 had reduced POD for the MESH field and slightly increased RMSD errors.

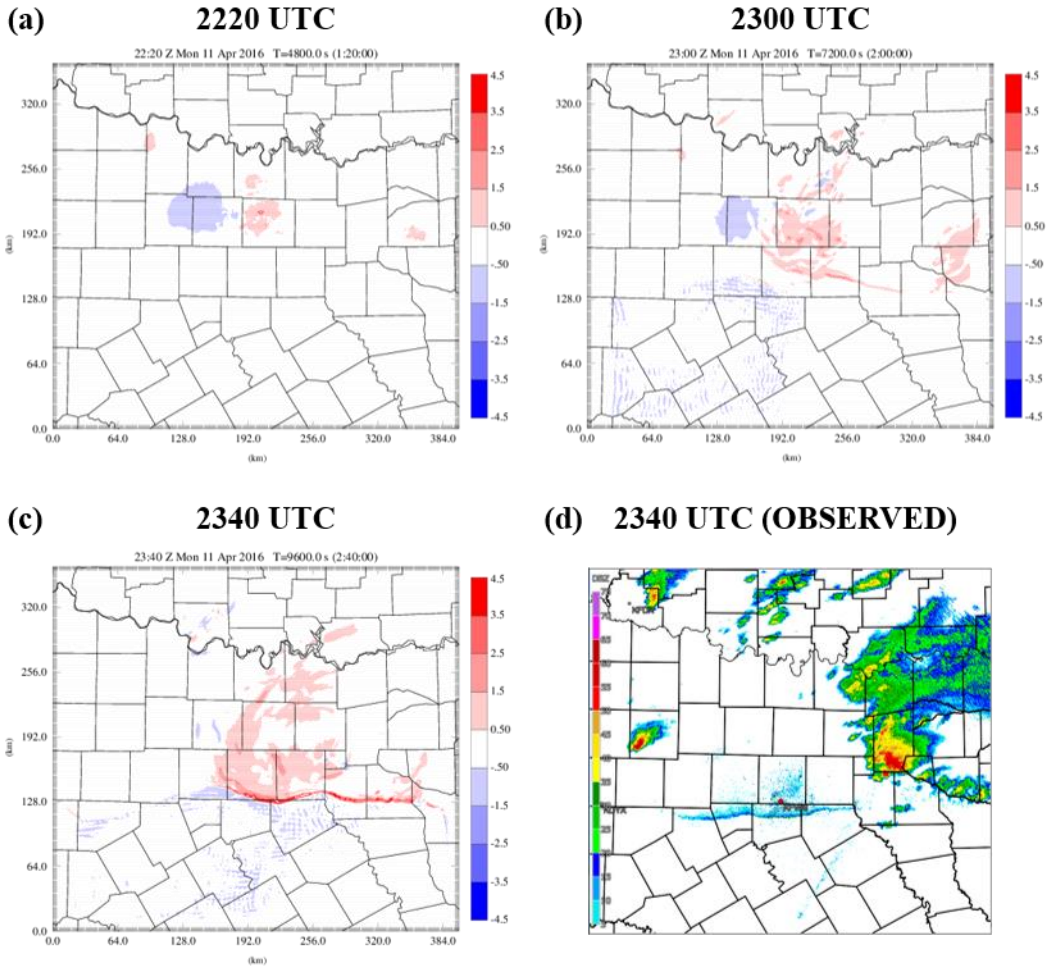


Figure 4.36: Potential temperature perturbations at the surface, which are defined as $\boldsymbol{\theta_{double} - \theta_{single}}$ for a) 2220 UTC, b) 2300 UTC, and c) 2340 UTC. The reflectivity observed by the KFWS WSR-88D radar (0.5 degree tilt) at 2340 UTC is shown in d), with a fine line indicative of the placement of the cold front.

Chapter 5

5.1 Summary and Conclusions

The United States Weather Research Program (USWRP) convened a workshop in 2003 to discuss means of improving the current observational network (Dabberdt et al. 2005). The committee recommended that a nationwide network of mesoscale surface stations be established to supplement the current observational network by providing additional observations in the lowest levels of the atmosphere. A 2009 report by the National Research Council (NRC) entitled *Observing Weather and Climate from the Ground up: A Nationwide Network of Networks* took this recommendation one step further and proposed that new mesoscale networks be integrated with existing ones to create a nationwide "network of networks" thereby maximizing the benefit of these distinct networks. An additional recommendation of this report was for the establishment of research testbeds to objectively assess the future impact of proposed observing systems.

The DFW Urban Demonstration Network was recently established in the Dallas-Fort Worth metroplex. Non-conventional surface data sources that are not available in the federal observing network include Global Science & Technology (GST) Mobile Platform Environmental Data (MoPED), WeatherBug, Citizen Weather Observer Program (CWOP), and Understory Weather. Non-conventional radar data include two Terminal Doppler Weather Radars (TDWRs) at the major passenger airports in the DFW metroplex and six Collaborative Adaptive Sensing of the Atmosphere (CASA) X-band radars. Two Sonic Detection and Ranging profilers (SODARs) have been installed in the

network, providing more frequent vertical profiles of wind than what are provided by the radiosonde network.

The purpose of this work was to evaluate the performance of the aforementioned new observing systems in the DFW testbed using observing system experiments (OSEs). In an OSE, an analysis and forecast are performed for a control experiment, in which all available data are assimilated. The control experiment can then be compared to the results of data denial experiments, in which a particular class of observations is denied, thus revealing the value of those observations.

In this work, OSEs are performed on a prolific hail-producing supercell thunderstorm that impacted the northern portion of the DFW metroplex on 11 April 2016. The Advanced Regional Prediction System (ARPS) model is used, along with the ARPS three-dimensional variational (3DVAR) analysis and complex cloud analysis. Incremental analysis updating with variable-dependent timing (IAU-VDT) is used to apply the analysis increments gradually during the assimilation window. Four 10-minute intermittent data assimilation cycles begin at 2150 UTC, with a one-and-a-half hour free forecast beginning at 2220 UTC.

The CONTROL experiment, which assimilated all available surface, radar, and aircraft data, generally captures the behavior of the observed supercell during the free forecast period, although this experiment exhibits a general wet bias in the reflectivity field as the free forecast progresses. Slight propagation speed errors are also evident as the simulated storm was positioned slightly southwest of the observed storm at 2330 UTC. The 88DONLY simulation does not produce a well-defined hook echo at the end of the forecast period, as evident in the reflectivity and updraft helicity fields, indicating

that the inclusion of the CASA and TDWR data in the CONTROL experiment aided in the production of a realistic supercell simulation. This general pattern is also observed for NOCASA, along with an eastward displacement of the storm by 2330 UTC when compared to both the CONTROL experiment and observations, indicating that the CASA data contributed to accurately capturing the storm evolution. More specifically, the NOCASAVR experiment, which still included CASA reflectivity data in the complex cloud analysis, performs similarly to NOCASA indicating that the low-level radial velocity data were most crucial for accurately analyzing and predicting this supercell. However, based upon the CASAONLY experiment, it is clear that the CASA radar data are insufficient as a standalone tool. Unlike for the 2007 Integrated Project One (IP1) testbed in southwestern Oklahoma, the CASA radars in the DFW Urban Demonstration Network do not take advantage of the collaborative adaptive scanning strategies, instead using more traditional "sit-and-spin" strategies that remain at elevation angles of 4.5 degrees or lower. As a consequence, these radars are incapable of observing the entire depth of the storm, as shown in vertical cross-sections (Figure 4.12).

The fractions skill score (FSS) is used to perform quantitative comparisons of the forecasted composite reflectivity field to the composite reflectivity field observed by the KFWS WSR-88D radar. When comparing the CONTROL, 88DONLY, CASAONLY, and NORADAR experiments using a 20 dBZ reflectivity threshold, the 88DONLY experiment has a consistently higher FSS throughout the free-forecast period than the remaining experiments. When using a 25 dBZ threshold, the 88DONLY experiment was the only forecast with skill greater than the "useful" mark by this metric during the entirety of the free forecast period. This indicates that the inclusion of the CASA radar

data degrades the forecasted reflectivity field by this metric, which is somewhat contradictory to what was noted subjectively for the reflectivity field structure, as the CONTROL experiment exhibited the best structure at the end of the free forecast period. The CONTROL and NONEWSFC experiments perform similarly at all reflectivity thresholds, indicating that the new surface networks resulted in little improvements to the forecasted reflectivity field.

Hail was an important element in this case, so forecast performance is also evaluated by comparing model-derived hail with radar-observed hail using the maximum estimated size of hail (MESH) algorithm, which incorporates a weighted vertical integration of the horizontal reflectivity factor exceeding 40 dBZ above the melting level. Deficiencies in the Milbrandt and Yau single-moment microphysics scheme (MY1) used are evident, as the areal extent of the hail in the CONTROL experiment's predicted hail swath is much larger than the observed MESH swath. Moreover, there is an under-prediction of the maximum hail size in the CONTROL experiment. The NOCASA and NOCASAVR experiments predicted a region of significant severe hail outside of the region where severe hail was observed, indicating that the inclusion of CASA radial velocity data in the lowest-levels of the atmosphere aided in the superior CONTROL experiment. This pattern of extraneous significant severe hail was also noted, although to a lesser extent, in a number of the other radar experiments. MESH forecasts were evaluated quantitatively using performance diagrams. The CONTROL experiment had the highest probability of detection (POD) and success ratio (SR) of all experiments using a hail size of 5 mm and neighborhood threshold of 15 km. The 88DONLY simulation has a lower POD and SR, which can largely be attributed to the denial of CASA radial

velocity data. The surface data denial experiments do not exhibit clear differences on the performance diagram, indicating that the radar data are most critical for accurately predicting hail. The CONTROL experiment also has the highest POD when considering severe hail (25 mm).

The root mean square difference (RMSD) of the surface temperature and dew point temperature fields are also used to assess forecast performance. Twelve stations are denied from the assimilation process for an independent comparison. When comparing the CONTROL, NORADAR, CASAONLY, NOCASA, and 88DONLY experiments, the 88DONLY experiment has the lowest 2 m temperature RMSD at the end of the free forecast period, indicating that the inclusion of the CASA and TDWR radar data degraded the resultant 2 m temperature forecast. The CASAONLY and NORADAR experiments have the highest RMSD at the end of the forecast period, likely in response to inadequate depictions of the cold pool and gust front intensity. The influence of radar data on the surface dew point temperature field was not as clear. When comparing the CONTROL, NONEWSFC, NOCWOPWXBUG, NOCWOP, and NOWXBUG experiments, the pattern is reversed, with the more pronounced differences occurring for the surface dew point temperature field. The NONEWSFC experiment has the lowest RMSD error at the end of the free forecast period, while the CONTROL experiment actually has the highest error. Based upon comparisons with NOCWOPWXBUG, NOCWOP, and NOWXBUG, it appears that the CONTROL experiment has the highest RMSD error due to the combined effects of the WeatherBug and CWOP data. The 2 m temperature RMSD value at the beginning of all experiments is in excess of 2.5°C, owing to a slight phase error in the frontal boundary placement in

the model background field. The RMSD decreases with time during the forecast period, as the frontal boundary passes the verification stations. The model background field has a cold bias initially, with the background field colder than the observed temperature for all 12 verification stations; however, the bias increases towards 0 as the forecast progresses.

Finally, the sensitivity to model microphysics parameterization schemes is explored by comparing the results using the MY1 microphysics scheme to the results when using the more sophisticated double-moment version (MY2). The reflectivity values when using the double-moment scheme (CTLDOUBLE) are generally less than when using the single-moment version (CONTROL) throughout the entirety of the free forecast. Moreover, the hook echo at 2330 UTC is not as well-defined in the CTLDOUBLE experiment. The forecasted MESH swath for the CTLDOUBLE experiment has a smaller areal extent than in CONTROL, although a large portion of the forecasted severe hail falls outside of the region of observed severe hail. The smaller areal extent (i.e., reduced number of false alarms) results in an improved success rate (SR) for the CTLDOUBLE experiment, although the probability of detection (POD) is reduced. The CONTROL experiment has a lower RMSD for 2 m temperature than the CTLDOUBLE experiment for the entirety of the free forecast period. Together, these factors indicate that the double-moment version does not provide superior results for the storm considered in this study.

5.2 Future Work

One major limitation of this work is that it only considers one convective mode (i.e., supercell) for one case. Furthermore, the storm considered here was elevated in the

cold sector, which could have potentially limited the forecast benefit of the non-conventional surface data and low-level radial velocity data from the CASA X-band radars. Similar comparisons should be performed for a more diverse array of weather events in order to obtain more comprehensive results. The CASA radar data may prove useful for simulating a quasi-linear convective system (QLCS; e.g., Schenkman et al. 2011a) and tornadic supercell (e.g., Schenkman et al. 2011b, Stratman and Brewster 2015). Moreover, the additional low-level dual-polarization radar data could provide valuable information to forecasters in a winter forecasting event. More complete results could also be obtained by considering a quasi-real-time month-long system, in which an analysis and forecast using all available data are compared to experiments using only a subset of the available data. The aggregated results over the month-long period would serve as more substantial evidence of the potential value of these observing systems.

Additionally, the seventh CASA X-band radar has recently been deployed in Mesquite, with the final radar planned for McKinney. Once the entire radar network is in place, a case should be identified to ascertain the value of the completed network. The Understory Weather observations used in this study represent a small subset of the observations available beginning in the spring of 2017. The number of stations collecting data has risen from only 10 for the case considered here to around 120 by April 2017. A case should be identified to gauge the potential forecast improvements of this recently-completed network.

Lastly, additional data denial experiments should be performed to assess the utility of the CASA X-band radars in the event that a WSR-88D radar experiences technical difficulties during a severe weather event. For the case presented here, the

KFWS radar was the most important radar for observing the storm; thus, these data denial experiments should focus on the denial of KFWS data. More specifically, two experiments should be performed, one with and one without the CASA data, to determine the ability of CASA radars to supplement upper-level radar data from neighboring WSR-88D radars. These experiments should also deny the inclusion of TDWR radar data, as these radars are not available nationwide.

References

Agustí-Panareda, A., A. Beljaars, C. Cardinali, I. Genkova, and C. Thorncroft, 2010: Impacts of assimilating AMMA soundings on ECMWF analyses and forecasts. *Wea. Forecasting*, **25 (4)**, 1142-1160.

Alapaty, K., N. L. Seaman, D. S. Niyogi, and A. F. Hanna, 2001: Assimilating surface data to improve the accuracy of atmospheric boundary layer simulations. *J. Appl. Meteor.*, **40 (11)**, 2068-2082.

Arnold, C. P., and C. H. Dey, 1986: Observing-systems simulation experiments: past, present, and future. *Bull. Amer. Meteor. Soc.*, **67 (6)**, 687-695.

Atlas, R., 1997: Atmospheric observations and experiments to assess their usefulness in data assimilation. *J. Meteor. Soc. Japan*, **75 (1 B)**, 111-130.

Bajaj, A. N. and B. J. Philips, 2012: Casting the Net – A revolutionary business model for deploying weather radar networks. *Meteorological Technology International*, 106-108.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, **3 (4)**, 396–409.

Benjamin, S. G., B. D. Jamison, W. R. Moninger, S. R. Sahm, B. E. Schwartz, and T. W. Schlatter, 2010: Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and Mesonet observations via the RUC hourly assimilation cycle. *Mon. Wea. Rev.*, **138 (4)**, 1319-1343.

Bi, L., J. A. Jung, M. C. Morgan, and J. F. Le Marshall, 2011: Assessment of assimilating ASCAT surface wind retrievals in the NCEP global data assimilation system. *Mon. Wea. Rev.*, **139 (11)**, 3405-3421.

Bjerknes, V., 1904: Das Problem der Wettervorhersage: betrachtet vom Standpunkte der Mechanik und der Physik, *Meteor. Zeits*, **21**, 1-7.

Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124 (6)**, 1256–1271.

Bouttier, F., and G. Kelly, 2001: Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **127 (574)**, 1469-1488.

Brewster, K., 2003: ADAS-ARPS data assimilation using incremental analysis updating. Accessed 10 March 2017. [Available online at http://www.caps.ou.edu/ARPS/ARPS5DOC/ADASNudging.pdf]

Brewster, K., E. Fay and F. Junyent, 2005a: How will X-band attenuation affect tornado detection in the CASA IP1 radar network? *32nd Conference on Radar Meteorology*, Albuquerque, NM, AMS, Boston. Conference CD, Paper 14R.4.

Brewster, K., L. White, B. Johnson, and J. Brotzge, 2005b: Selecting the sites for CASA NetRad, a collaborative radar network. *Ninth Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans and Land Surface (IOAS-AOLS)*, 85th Amer. Meteor. Soc. Annual Meeting CD, Paper P3.4.

Brewster, K., M. Hu, M. Xue, and J. Gao, 2005c: Efficient assimilation of radar data at high resolution for short-range numerical weather prediction. *World Weather Research Program Symposium on Nowcasting and Very Short-Range Forecasting*, WSN05, Toulouse, France, WMO, Symposium CD, Paper, Vol. 3.

Brewster, K., K. Thomas, J. Brotzge, Y. Wang, D. Weber, and M. Xue, 2007: High resolution data assimilation of CASA X-band radar data for thunderstorm forecasting. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred*.

Brewster, K.A. and D.R. Stratman, 2015: An updated high-resolution hydrometeor analysis system using radar and other data. *Preprints, 27th Conference on Wea. Analysis and Forecasting and 23rd Conf. on Numerical. Wea. Pred.*, Amer. Meteor. Soc., Paper 31.

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

Brewster, K. A., A. Bajaj, B. L. Philips, D. L. Pepyne, E. Lyons, and F. H. Carr, 2017: CASA Dallas-Fort Worth Urban Testbed observations: Network of Networks at work. *Special Symposium on Meteorological Observations and Instrumentation*, 97th AMS Annual Meeting, Seattle, WA, January 22-26, 2017, Amer. Meteor. Soc., Paper 1027.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, M. E. Eilts, 1995: The Oklahoma Mesonet: a technical overview. *J. Atmos. Oceanic Technol.*, **12 (1)**, 5-19.

Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135 (638)**, 239-250.

Carlaw, L., 2014: Evaluating the impact of assimilating surface observations on high-resolution analyses and forecasts. M.S. Thesis, School of Meteorology, University of Oklahoma, 98 pp.

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.

Crum, T. D., and R. L. Alberty, 1993: The WSR-88D and the WSR-88D operational support facility. *Bull. Amer. Meteor. Soc.*, **74 (9)**, 1669-1687.

CWOP, 2014: Hourly number of weather stations on APRS-IS. URL http://www.wxqa.com/checkservers2.html.

Dabberdt, W. F., and Coauthors, 2005: Multifunctional mesoscale observing networks. *Bull. Amer. Meteor. Soc.*, **86 (7)**, 961-982.

Dahlia, J., 2013: The national mesonet program: filling in the gaps. *Weatherwise*, **66 (4)**, 26-33.

Dawson, D. T., and M. Xue, 2006: Numerical forecasts of the 15-16 June 2002 southern plains mesoscale convective system: Impact of mesoscale data and cloud analysis. *Mon. Wea. Rev.*, **134 (6)**, 1607-1629.

Dawson, D. T., M. Xue, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms. *Mon. Wea. Rev.*, **138 (4)**, 1152-1171.

Deardorff, J. W., 1980: Stratocumulus-capped mixed layers derived from a three-dimensional model. *Boundary-Layer Meteorology*, **18 (4)**, 495–527.

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15 (1)**, 51–64.

Eilts, M. D., and S. D. Smith, 1990: Efficient dealiasing of Doppler velocities using local environment constraints. *J. Atmos. Oceanic Technol.*, **7 (1)**, 118–128.

Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51 (2)**, 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 (3)**, 457–469.

Graham, R. J., S. R. Anderson, and M. J. Bader, 2000: The relative utility of current observation systems to global-scale NWP forecasts. *Quart. J. Roy. Meteor. Soc*., **126 (568)**, 2435-2460.

Ha, S.-Y., and C. Snyder, 2014: Influence of surface observations in mesoscale data assimilation using an ensemble Kalman filter. *Mon. Wea. Rev.*, **142 (4)**, 1489-1508.

Hart, J.A., and W. Korotky, 1991: The SHARP workstation v1.50 users guide. National Weather Service, NOAA, US. Dept. of Commerce, 30 pp. [Available from NWS Eastern Region Headquarters, 630 Johnson Ave., Bohemia, NY 11716.]

Hayden, C. M., and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: applications to NESDIS operational processing. *J. Appl. Meteor.*, **34 (1)**, 3–15.

Hilliker, J. L., G. Akasapu, and G. S. Young, 2010: Assessing the short-term forecast capability of nonstandardized surface observations using the national digital forecast database (NDFD). *J. Appl. Meteor. Climatol.*, **49 (7)**, 1397-1411.

Hu, M., M. Xue, and K. Brewster, 2006a: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134 (2)**, 675-698.

Hu, M., M. Xue, J. Gao, and K. Brewster, 2006b: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3dvar. *Mon. Wea. Rev.*, **134 (2)**, 699-721.

Istok, M. J., A. D. Stern, R. E. Saffle, B. Bumgarner, B. R. Klein, N. Shen, Y. Song, Z. Wang, and W. M. Blanchard, 2008: Terminal Doppler Weather Radar for NWS Operations: Phase 3 Update. *24th Conference on Interactive Information and Processing Systems (IIPS)*.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23 (5)**, 931-952.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press, 341 pp.

Kazumori, M., Q. Liu, R. Treadon, and J. C. Derber, 2008: Impact study of AMSR-E radiances in the NCEP global data assimilation system. *Mon. Wea. Rev.*, **136 (2)**, 541-559.

Knopfmeier, K. H. and D. J. Stensrud, 2013: Influence of mesonet observations on the accuracy of surface analyses generated by an ensemble Kalman Filter. *Wea. Forecasting*, **28 (3)**, 815-841.

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

Lilly, D. K., 1990: Numerical prediction of thunderstorms—has its time come? *Quart. J. Roy. Meteor. Soc.*, **116 (494)**, 779-798.

Lord, S. J., E. Kalnay, R. Daley, G. Emmitt, and R. Atlas, 1997: Using OSSEs in the design of the future generation of integrated observing systems. Preprint volume, 1st Symposium on Integrated Observation Systems, 45-47.

Lynch, P., 2008: The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, **227 (7)**, 3431-3444.

McCarthy, J., and S. E. Koch, 1982: The evolution of an Oklahoma dryline. Part I: a meso-and subsynoptic-scale analysis. *J. Atmos. Sci.*, **39 (2)**, 225-236.

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.

McNally, T., M. Bonavita, and J.-N. Thépaut, 2014: The role of satellite data in the forecasting of Hurricane Sandy. *Mon. Wea. Rev.,* **142 (2)**, 634-646.

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

Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62 (9)**, 3051-3064.

Mukhopadhyay, P., J. Sanjay, W. R. Cotton, and S. S. Singh, 2005: Impact of surface meteorological observations on RAMS forecast of monsoon weather systems over the Indian region. *Meteor. Atmos. Phys.*, **90 (1-2)**, 77-108.

National Research Council, 2009: *Observing weather and climate from the ground up: A nationwide network of networks*. The National Academies Press, 250 pp.

National Research Council, 2012: *Urban meteorology: forecasting, monitoring, and meeting users' needs*. The National Academies Press, 176 pp.

Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev*., **117 (3)**, 536–549.

NOAA National Centers for Environmental Information (NCEI), 2017: U.S. Billion-Dollar Weather and Climate Disasters. https://www.ncdc.noaa.gov/billions/

NWS, 1994: Technique specification package 88-21-r2 for awips-90 rfp Appendix G requirements numbers: Quality control incoming data. *AWIPS Document Number TSP-032-1992R2, NOAA, National Weather Service, Office of Systems Development.*

NWS, 1999: Automated Surface Observing System (ASOS). Accessed 15 March 2017. [Available online at http://www.nws.noaa.gov/ost/asostech.html.]

Richardson, L. F., 1922: *Weather Prediction by Numerical Process.* Cambridge university press.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136 (1)**, 78–97.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24 (2)**, 601-608.

Rogers, R., and M. Yau, 1989: A short course in cloud physics, international series in natural philosophy. Butterworth Heinemann, Burlington, MA, 290 pp.

Schaefer, J. T., 1974: The life cycle of the dryline. *J. Appl. Meteor.*, **13 (4)**, 444-449.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011a: The analysis and prediction of the 8-9 May 2007 Oklahoma tornadic mesoscale convective system by assimilating WSR-88D and CASA radar data using 3DVAR. *Mon. Wea. Rev.*, **139 (1)**, 224-246.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 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 (11)**, 3422-3445.

Schroeder, J. L., W. S. Burgett, K. B. Haynie, I. Sonmez, G. D. Skwira, A. L. Doggett, J. W. Lipe, 2005: The West Texas Mesonet: a technical overview. *J. Atmos. Oceanic Technol.*, **22 (2)**, 211-222.

Smith, T. L., S. G. Benjamin, S. I. Gutman, and S. Sahm, 2007: Short-range forecast impact from assimilation of GPS-IPW observations into the Rapid Update Cycle. *Mon. Wea. Rev.*, **135 (8)**, 2914-2930.

Snook, N., and M. Xue, 2008: Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophys. Res. Lett*., 35 (24).

Snook, N., M. Xue, and Y. Jung, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140 (7)**, 2126-2146.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: a vision for 2020. *Bull. Amer. Meteor. Soc.*, **90 (10)**, 1487-1499.

Stensrud, D. J., and Coauthors, 2013: Progress and challenges with warn-on-forecast. *Atmos. Res.*, **123**, 2-16.

Stratman, D. and K. A. Brewster, 2015: Impact of assimilating CASA X-Band radar data for the 24 May 2011 tornadic storms using various microphysics schemes at 1-km resolution. *Preprints, 37th Conference on Radar Meteor.*, Amer. Meteor. Soc., Paper P.221.

Stratman, D., and K. Brewster, 2017: Sensitivities of 1-km forecasts of 24 May 2011 tornadic supercells to mirophysics parameterizations. *Mon. Wea. Rev.*, doi:10.1175/MWR-D-16-0282.1, in press.

Tyndall, D. P., and J. D. Horel, 2013: Impacts of mesonet observations on meteorological surface analyses. *Wea. Forecasting*, **28 (1)**, 254-269.

Understory Weather, 2015: Understory Sensor Data: Weather monitoring networks accurately measuring hail and wind. [Available online at http://try.understoryweather.com/white-paper-accuracy-register/]

Waldvogel, A., W. Schmid, and B. Federer, 1978: The kinetic energy of hailfalls. Part I: hailstone spectra. *J. Appl. Meteor.*, **17 (4)**, 515-520.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Elsevier, 676 pp.

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13 (2)**, 286-303.

Xiao, Q., and J. Sun, 2007: Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP 2002. *Mon. Wea. Rev.*, **135 (10)**, 3381-3404.

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

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS)-A multi-scale non-hydrostatic atmospheric simulation and prediction tool. Part I: model dynamics and verification. *Meteor. Atmos. Phys.*, **75 (3)**, 161-193.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D. Wang, 2001: The Advanced Regional Prediction System (ARPS)-A

multi-scale non-hydrostatic atmospheric simulation and prediction tool. Part II: model physics and applications. *Meteor. Atmos. Phys.,* **76 (3)**, 143-165.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, and K. K. Droegemeier, 2013: Prediction of convective storms at convection-resolving 1 km resolution over continental United States with radar data assimilation: An example case of 26 May 2008 and precipitation forecasts from spring 2009. *Adv. Meteor.*, 2013.

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 (1)**, 39-62.

Zapotocny, T. H., 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 (4)**, 887-909.

Zapotocny, T. H., W. P. Menzel, J. P. Nelson III, and J. A. Jung, 2002: An impact study of five remotely sensed and five in situ data types in the eta data assimilation system. *Wea. Forecasting*, **17 (2)**, 263-285.

Zapotocny, T. H., W. P. Menzel, 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 (2)**, 178-198.

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 (12)**.

Appendix A: Comparing Data Averaging Techniques using Permutation Testing

Since Global Science & Technology (GST) Mobile Platform Environmental Data (MoPED) observations are collected every 10 seconds, data thinning is necessary to reduce the spatiotemporal resolution of the observations. The purpose of this portion of our research is to compare various averaging techniques for thinning this data. Permutation testing is used to perform a hypothesis test that is designed to assess whether 2-min, 3-min, 4-min, and 5-min averages are statistically different from averages using a shorter 1-min window.

A severe weather event on 5 November 2015 presents a test case to examine the forecast impact of these mobile observations. Figure A.1 shows a surface analysis valid at 2100 UTC on 5 November 2015. A 992 hPa surface low pressure system was positioned over Canada, just northwest of Lake Superior. An associated cold front extended south through Kansas and into the Oklahoma panhandle, with a dryline out ahead of this cold front, extending from eastern Kansas through central Oklahoma and into central Texas. Convection formed along this surface boundary, some of which would become severe, as evidenced by severe hail and wind reports from the Storm Prediction Center (SPC; Figure A.2).

A.1 Data and Methodology

GST MoPED data were obtained for the period from 1700 to 2200 UTC for 5 November 2015. The domain considered is shown in Figure A.3, with latitude spanning 31.5 to 33.9 degrees N and longitude spanning 96.0 to 98.4 degrees W. Twelve trucks passed through the domain during the specified time period, although not all are analyzed here.

The data were then thinned for five different lengths of time. The data were grouped based on truck identifier and averages computed using data from individual trucks. One-minute averages were generated using data for a single minute, with the averaged observation assigned to the midpoint of the interval (e.g., data from 17:00:00 to 17:00:59 were averaged into a single observation for 17:00:30). A second methodology for computing 1-min averages is to consider data centered about the beginning of a minute (e.g., data from 17:00:30 to 17:01:30 were averaged and assigned to 17:01:00). Two-minute averages were computed by dividing the time period into 2-min periods (e.g., observations from 17:00:00 to 17:01:59 were assigned to 17:01:00). Four-min averages were computed in a similar fashion as the 2-min averages. Three-min averages were computed by considering 3-min windows (e.g., observations from 17:00:00 to 17:02:59 were averaged and assigned to 17:01:30). Five-min averages were calculated similarly to 3-min averages. Three-min and 5-min averages are compared to 1-min averages from the first approach, whereas 2-min and 4-min averages are compared to 1-min averages using the second methodology. Trucks with only intermittent observations were discarded from consideration here. Three distinct time periods emerged from this analysis. The first period is for truck CW0WG and spans from 1700 to 1810 UTC. The second period is for the one hour period beginning at 1700 UTC (truck CW14L). The third and final period is for truck CW14L, with a time span from approximately 1950 to 2150 UTC.

A hypothesis test is utilized to determine if the various averaging schemes produce statistically different results (Wilks 2011). More specifically, 2-min, 3-min, 4-min, and 5-min averages of temperature are compared with the corresponding 1-min

averages. The null hypothesis (H_o) is that the difference between the mean temperature (°C) when averaging over 2-min windows and 1-min windows is equal to zero. The alternative hypothesis (H_A) is that the difference between the two means is non-zero. A two-sided hypothesis test is performed, as the longer averaging periods could yield over- or under-estimates of the 1-min averages. The level of significance is set to 90%. This procedure is then repeated to compare 3-min, 4-min, and 5-min averages to the 1-min averages, as well as assess differences in averaging schemes for dew point temperature (°C) and surface pressure (hPa).

The hypothesis tests described above are performed using permutation testing (Wilks 2011). A key principle of permutation testing that holds under the null hypothesis is exchangeability, or the idea that the data from both samples originate from the same distribution. If exchangeability holds true, the labels attributing values to a given data set are arbitrary. To outline the process of permutation testing, consider the case comparing 2-min averages of temperature to the respective 1-min averages for the period from 1700 to 1810 UTC. Each data set contains 35 observations (i.e., $n_1 = n_2 = 35$), as 1-min averages without a matched observation were not included. Once all observations are pooled, the resulting sample size is $n = n_1 + n_2 = 35 + 35 = 70$. The samples are then selected without replacement from the resulting pool and placed into one of two artificial subsets, each containing $n = 35$ observations. The sample mean is recorded for each subset and the difference is computed. The permutation process is repeated 999 times for each hypothesis test that is performed.

A.2 Results

The results in Table A.1 correspond to data for the period from 1700 to 1810 UTC for truck CW0WG. The standard deviation of the differences in means of the permutated subsets is reported. A p-value is reported, which indicates how likely it is to observe the mean difference calculated from the permutation differences assuming that the samples are drawn from the same population. This p-value is then used to determine whether or not to reject the null hypothesis.

Figure A.4 shows time-series plots of temperature, dew point temperature, and surface pressure for the differing averaging schemes, with gray dots representing 1-min averages, while Figure A.5 shows histograms of the differences in sample mean among the 999 permutations. The columns correspond to temperature, dew point temperature, and surface pressure, respectively, while the rows represent the various averaging regimes, with the first row representing the 2-min averages. For this case, the null hypothesis is not rejected for each of the temperature and surface pressure averaging schemes, and no statistical difference exists between the 1-min averaging scheme and the schemes using a longer averaging time interval. The p-value for the four-minute averages of dew point temperature is 0.096, which is below the level of significance of 0.10; thus, 4-min averages of dew point temperature are statistically different from the 1-min averages and, according to this theory, the null hypothesis is rejected. The null hypothesis is not rejected for the remaining averaging time lengths. Figure A.5 shows more variability in the differences in mean for the permuted samples as the length of the averaging window increases.

Two additional time periods were considered (not shown). The second data set uses observations from 1700 to 1800 UTC for truck CW14L. In this case, 4-min averages of surface pressure are shown to be statistically different from 1-min averages. The third data set uses observations from roughly 1950 to 2150 UTC for the same truck. Three-min averages of dew point temperature are shown to be statistically different from the 1-min averages. In both cases, the other averaging schemes did not produce statistically different results.

A.3 Conclusions

For the three time periods considered in this study, no statistical difference was found between the 2-min and 1-min averaging regimes. Statistically different regimes were found for several of the 3- and 4-min averages. For this case study alone, the results indicate that statistically different averages are found once the averaging time length is increased to three minutes and above. However, it is important to note that this experiment only considers data over a limited domain for a specific date. The methodology used in this study should be extended to an increased number of trucks over a wider domain for varied dates to determine if additional patterns emerge. Also, it should be noted that the difference in means for temperature, dew point temperature, and pressure are all smaller than expected instrument error, which indicates that, while the results may yield statistically significant results, these differences would likely be unachievable using current sensor technology. Moreover, by comparing 2-, 3-, 4-, and 5-minute averages to 1-minute averages rather than the raw data, the measurement errors are largely averaged out, mitigating the effects of these errors.

While there are no major deviations in temperature and dew point temperature throughout the time periods considered, the surface analysis in Figure A.1 shows a dryline approaching the domain used in this study (Figure A.3). If a vehicle were to intersect a dryline, the dew point temperature would be subject to rapid fluctuations that may impact the applicability of longer averaging time scales, which is especially true in cases where trucks are traveling at highway speeds (up to 120 kilometers per hour).

Additionally, for the averaged results in which the null hypothesis is rejected, results averaged using a longer length of time were not rejected. For instance, in Table A.1, it can be seen that 4-minute averages of dew point temperature are deemed statistically different from the 1-minute averages to which they are compared and the null hypothesis is rejected. However, 5-minute averages are not found to be statistically different. This study utilized 999 permutations for each experiment. To test whether these findings are consistent with the data sets tested, the number of permutations considered should be increased. For example, increasing the number of permutations to 9,999 or 99,999 may result in different conclusions.

A major limiting factor of the averaging schemes used in this research is that they only rely on elapsed time and do not take the vehicle's speed into account when computing averages. Since truck speed can vary from stationary to full highway speed (upwards of 120 km/h), the distance a truck travels in the averaging window varies considerably from sample to sample. For this reason, a spatial averaging regime has been devised that takes into account both time and the accumulated distance between subsequent observations.

Table A.1: Results for truck CW0WG from 1700 to 1810 UTC

Comparison of 1 minute and 2 minute averaging schemes (n = 35):

	Temp.	**Dew point temp.**	**Sfc. pressure**
Mean of 1 minute averages	25.520	22.645	993.180
Mean of 2 minute averages	25.518	22.646	993.177
Difference in means	-0.002	0.001	-0.003
Standard deviation	0.0068	0.0046	0.0404
Two-sided p-value	0.780	0.692	1.000
Decision	Do not reject	Do not reject	Do not reject

Comparison of 1 minute and 3 minute averaging schemes (n = 23):

	Temp.	**Dew point temp.**	**Sfc. pressure**
Mean of 1 minute averages	25.533	22.641	993.165
Mean of 3 minute averages	25.541	22.655	993.230
Difference in means	0.008	0.014	0.065
Standard deviation	0.0084	0.0122	0.0706
Two-sided p-value	0.410	0.240	0.414
Decision	Do not reject	Do not reject	Do not reject

Comparison of 1 minute and 4 minute averaging schemes (n = 17):

	Temp.	**Dew point temp.**	**Sfc. pressure**
Mean of 1 minute averages	25.561	22.639	993.182
Mean of 4 minute averages	25.565	22.665	993.282
Difference in means	0.004	0.026	0.100
Standard deviation	0.0189	0.0159	0.1038
Two-sided p-value	0.776	0.096	0.376
Decision	Do not reject	Reject	Do not reject

Comparison of 1 minute and 5 minute averaging schemes (n = 14):

	Temp.	**Dew point temp.**	**Sfc. pressure**
Mean of 1 minute averages	25.498	22.639	993.236
Mean of 5 minute averages	25.517	22.645	993.179
Difference in means	0.019	0.006	-0.057
Standard deviation	0.0312	0.0202	0.1417
Two-sided p-value	0.726	0.842	0.662
Decision	Do not reject	Do not reject	Do not reject

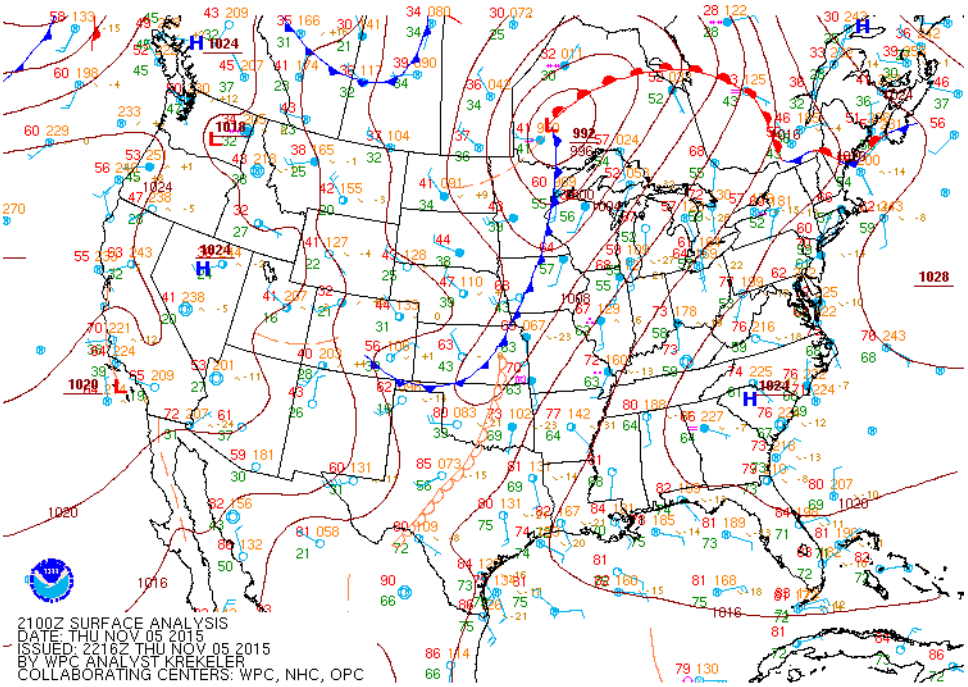


Figure A.1: Surface analysis from the Weather Prediction Center (WPC), valid at 2100 UTC on 5 November 2015.

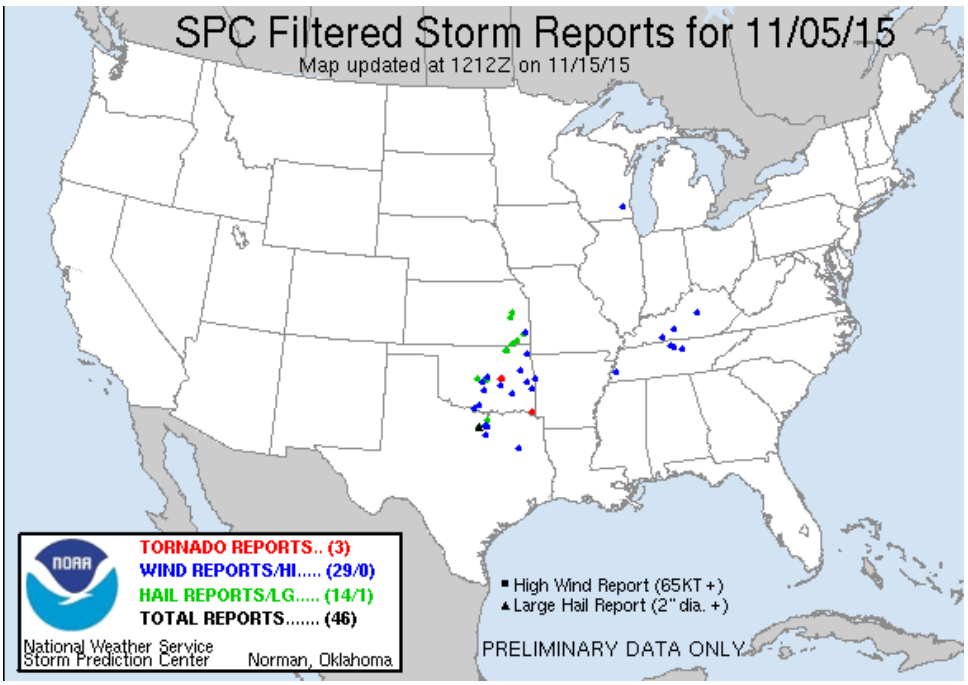


Figure A.2: Storm Prediction Center (SPC) storm reports from 5 November 2015. Severe hail and wind were both reported in the Dallas-Fort Worth metroplex.

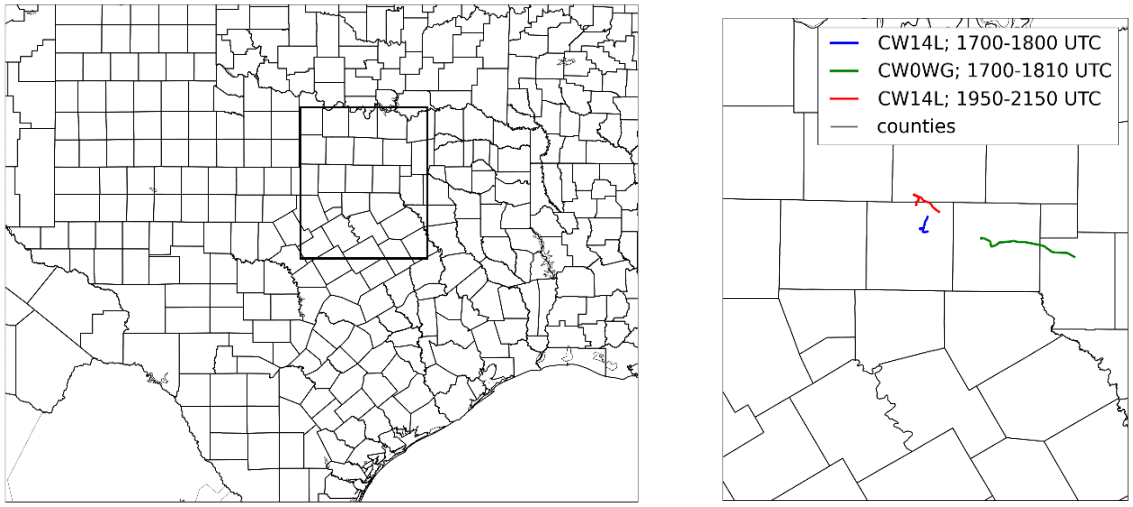


Figure A.3: (Left) Outline of the geographic area considered in this study (outlined in black), which includes Dallas-Fort Worth. (Right) The geographic location of trucks considered in this study.

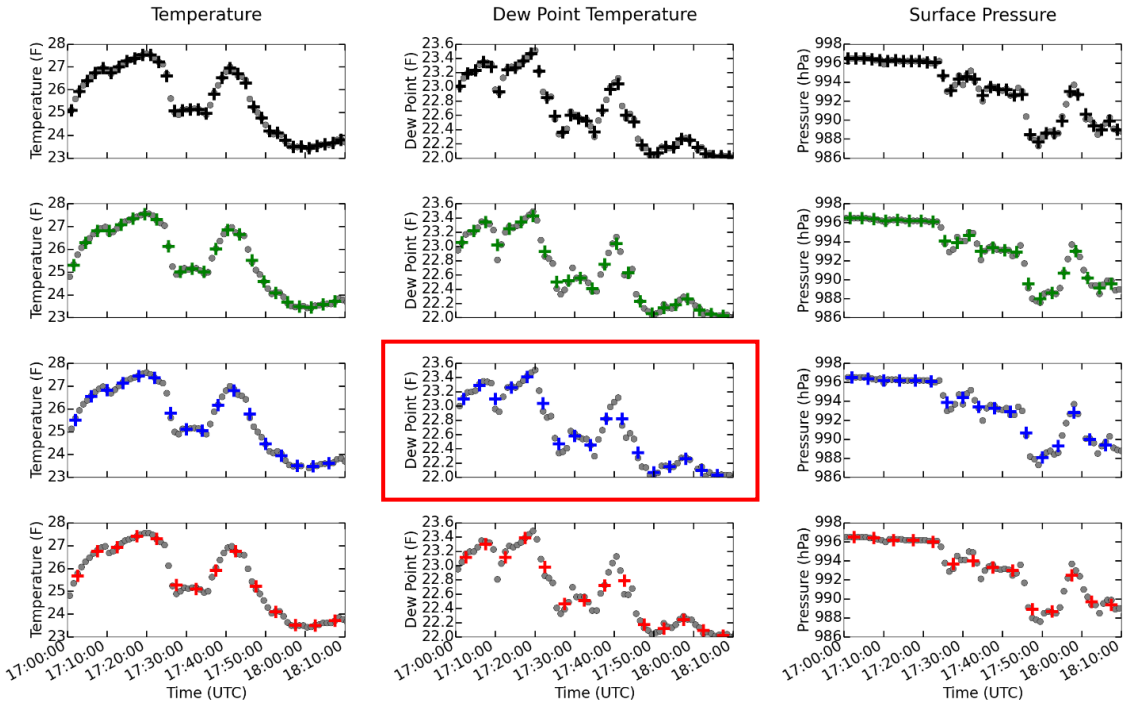


Figure A.4: Results of the thinning algorithm for truck CW0WG for the time period from 1700 to 1810 UTC. The gray dots correspond to one-minute averages. A red box indicates that an experiment resulted in statistically different results.

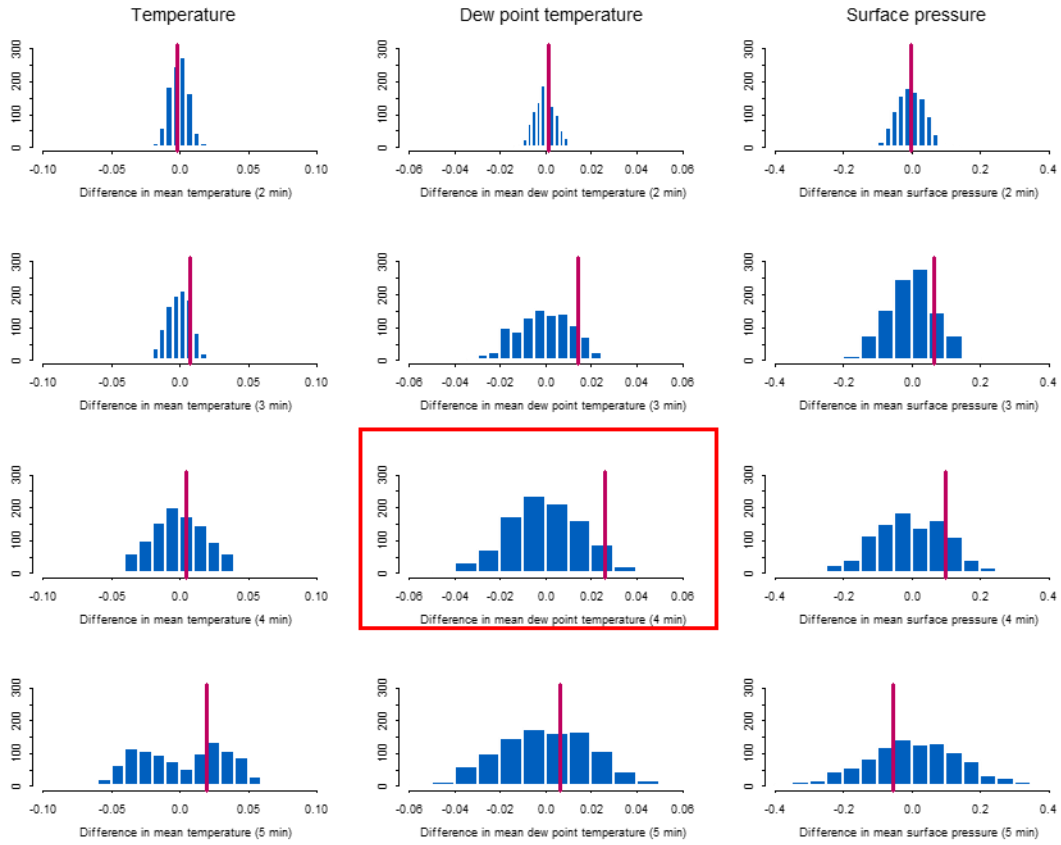


Figure A.5: Results of the permutation test for truck CW0WG for the time period from 1700 to 1810 UTC. The rows correspond to two-minute, three-minute, four-minute, and five minute averaging windows, respectively.