UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ENSEMBLE KALMAN FILTER METHODS FOR CONVECTIVE-SCALE RADAR DATA ASSIMILATION AND MULTI-SCALE DATA ASSIMILATION OF THE 13 JUNE 2010 TORNADIC SUPERCELL ENVIRONMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSPHY

By

THERESE THOMPSON
Norman, Oklahoma
2014

ENSEMBLE KALMAN FILTER METHODS FOR CONVECTIVE-SCALE RADAR DATA ASSIMILATION AND MULTI-SCALE DATA ASSIMILATION OF THE 13 JUNE 2010 TORNADIC SUPERCELL ENVIRONMENT

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

Dr. Xuguang Wang, Chair

Dr. Louis J. Wicker, Co-Chair

Dr. Micheal I. Biggerstaff

Dr. Alan Shapiro

Dr. David J. Stensrud

Dr. S. Lakshmivarahan

Acknowledgements

While the research presented in this dissertation is the product of many years' effort on my part, it would not have been possible without the generous assistance of a large number of individuals and organizations. Contributions ranged from the tangible (e.g. financial support, computing facilities, code, office resources) to the intangible (e.g. proofreading, administration, mentorship, moral support). Herein, I attempt to recognize those who are most deserving of my gratitude, but the order in which they appear does not necessarily relate to the significance of their contributions.

I would like to express my sincere gratitude to my advisor Dr. Louis Wicker of the National Severe Storms Laboratory (NSSL). He has been my supervisor and chief research mentor since I became a University of Oklahoma (OU) graduate student in the summer of 2007. His grants and research funding have supported me during my seven years as a masters and doctoral student. His insight and guidance have significantly improved my work, and his limitless enthusiasm is motivating. Under his mentorship, my skills as a scientist have matured and flourished. I am thankful he gave me the opportunity to participate in field projects, especially VORTEX2. I am deeply indebted to him for supporting me through numerous challenges, and for allowing me to complete my dissertation as a visitor at the National Center for Atmospheric Research (NCAR). I know he has truly wanted me to succeed, and has done everything he can to help me along the way.

Dr. Xuguang Wang has also served as my adviser since the summer of 2010. I am thankful for her guidance and encouragement throughout my dissertation work. Her reliable input on my research plans and results ensured my progress towards degree completion. I aspire to achieve her level of expertise in the field of data assimilation.

I am also grateful for Dr. Michael Biggerstaff, who has served on my Ph.D. committee. His passion for discovery is inspiring. He taught me how to operate the Shared Mobile Atmospheric Research and Teaching Radars (SMART-Rs) and passed on his passion and extensive knowledge of fieldwork. Many of my fondest memories as an OU graduate student are from adventures inside SMART-R1. I truly appreciate his faith in me as a radar operator and instrument scientist during VORTEX2 and several other field projects.

I also must thank the three additional members of my Ph.D. committee, who have provided me with many important suggestions and encouragement. Dr. Alan Shaprio's unwavering support for students drove me to pursue my academic goals. I enjoyed his exceptional teaching in advanced atmospheric dynamics 1 and 2. I am lucky to have Dr. David Stensrud as a member of my Ph.D. committee. He exemplifies outstanding leadership and has always provided a valuable perspective on the larger picture of my work. Dr. S. Lakshmivarahan has also been helpful.

One additional individual, Dr. Glen Romine, might as well have served on my Ph.D. committee for all his efforts helping me with this research. He served as a mentor during my first visit to NCAR as an Advanced Studies Program (ASP) graduate student visitor (3-months), secured funding from the NCAR Institute for Mathematics Applied to Geosciences (IMAGe) for my second visit (4-months), and welcomed me into NCAR's Mesoscale & Microscale Meteorology (MMM) division as a long-term visitor for the last year of my Ph.D. studies. He provided an enormous amount of voluntary assistance, computing resources and support. He also provided one-on-one instruction on the use of the Data Assimilation Research Testbed (DART) toolkit and the Weather Research and Forecasting (WRF) model. I could always count on him to provide honest feedback, and his recommendations have been vital to the success of my multi-scale data assimilation work. I also enjoyed our hiking trips.

Discussions with many scientists at NSSL and OU, including Dr. Corey Potvin, Dr. Dusty Wheatley, Dr. Ted Mansell, Dr. Mike Coniglio, Dr. Ryan Sobash, Dr. Nusrat Yussouf, Kent Knopfmeier, Dr. Robin Tanamachi, Dr. Dan Dawson, Don Burgess, Dr. Adam Clark, Dr. Thomas Jones, Dr. Kim Elmore, and Dr. Sean Crowell helped refine ideas and overcome challenges of this dissertation research. Many of them also provided comedic relief during lunch and hallway chats.

I owe a great deal of thanks to my fellow OU graduate students, including Amanda Kis, Charlotte Wainwright, David Bodine, Dan Bettan, and numerous others. In particular, Amanda Kis has listened to my troubles, celebrated my achievements, and helped me survive the perils of graduate student research.

I am happy to have had the opportunity to spend a year and half as an NCAR visitor. I am thankful that NCAR's MMM division has welcomed me as a long-term visitor. Stimulating conversations with Dr. Chris Snyder, Dr. Jeff Anderson, and Nancy Collins furthered the robustness of my dissertation research. I also owe a special thank you to Dr. David Dowell of the Earth Systems Research Laboratory (ESRL) for providing helpful feedback during the early stages of this research and for manually edited NOXP radar images for 13 June 2010. I am thankful for my new friendships with other young scientists at NCAR including Kate Fossell, Dr. Dan Stern, Mike Kavulich, Kelly Keene, and several other visitors and post-docs. Visiting NCAR has also allowed me to discover my love for the great outdoors. I am thankful for the relaxation and clarity that trips to the mountains provide and friends like Danny Cheresnick, Aaron Bansemer, and George Bryan to go exploring with me.

I extend my sincere thanks to staff at the OU School of Meteorology, the Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), NSSL and NCAR. Celia Jones, Marcia Pallutto, Tracy Reinke, and Linda McGuckin have patiently and happily helped me conquer many administrative tasks. My School of Meteorology professors and directors, Dr. Fred Carr and Dr. David Parsons, have taught me so much, and make me proud to be a Sooner Meteorologist. It may seem odd, but I would also like to thank the University of Oklahoma. I have enjoyed the campus life, football games, and National Weather Center so much that I decided to remain a student for 11 years.

I would like to acknowledge the data and funding providers used in this study. The WSR-88D data were provided by the National Climatic Data Center (NCDC). The conventional observations were provided by the National Oceanic and Atmospheric Administration (NOAA) through the Meteorological Assimilation Data Ingest System (MADIS). Oklahoma Mesonet data are provided courtesy of the Oklahoma Mesonet, a cooperative venture between Oklahoma State University and The University of Oklahoma and supported by the taxpayers of Oklahoma. The VORTEX2 data were provided by NCAR's Earth Observing Laboratory (EOL) and the hundreds of scientist and students who participated in the VORTEX2 data collection. I appreciate Takemasa Miyoshi for making his LETKF core code publicly available. Warn-on-Forecast funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce.

Many colleagues, classmates, friends, and family have helped me achieve my dream of becoming a meteorologist. In addition to those mentioned above, I would like to thank close friends, Erin Watson, Scott and Michelle Ganson, Tonya and Kenneth McCallister, and Renee Curry who have shared laughter and many good times with me.

I am fortunate to be the product of a very supportive family that has stood by me throughout my life. My parents have encouraged me every step of the way. They continue to ask about my work, offer ideas, and simply listen to me talk. My Mom added my first publication to the library in the law office where she works and my Dad even attended a dynamics class. Together they have shown me so much love and dedication. I am also grateful for my big brother RC and sister Megan. They may tease me about being a weather nerd, but they are truly supportive of my passion and challenge me to explain and predict the weather around them.

Last, and most importantly, I would not be a productive scientist without a loving and supportive atmosphere at home. During my tenure as a doctoral student, Bill Ladwig took the bold step of asking me to marry him. In fact, we will do just that, exactly one week after my dissertation defense. He is the love of my life and he has made this dissertation possible. He has helped me manage my stress in many ways, including exercising with me on a regular basis. He has provided me with a sounding board for ideas and has spent numerous hours proofreading this text. I know he believes in me and he reminds me to believe in myself. Bill, I would not have been able to finish this without your encouragement, patience, and love. Thank you.

Table of Contents

List of Tables

List of Figures

Abstract

Ensemble data assimilation at convective-scales will need to solve a number of scientific and technical issues prior to being usable for operational numerical weather prediction. This research contributes to this goal by first comparing the Local Ensemble Transform Kalman Filter (LETKF) to the Ensemble Square Root Filter (EnSRF) to examine whether either method consistently produces more accurate analyses and forecasts. Second, multi-scale data assimilation strategies are explored to improve the analysis of complex environmental conditions and subsequent convective forecasts.

While theoretically the LETKF and EnSRF filters should behave the same for ideal systems, a comparison between the serial and simultaneous filters has not previously been explored at the convective-scale where significant non-linear effects are present. Observing System Simulation Experiments (OSSEs) are first used to compare the assimilation systems for the analysis and forecast of a supercell thunderstorm. Both the EnSRF and LETKF produce reasonable analyses from the Doppler velocity and reflectivity observations of the true supercell. Small improvements in analysis errors and system noise from the LETKF simultaneous update do not significantly impact the subsequent forecasts. This result is consistent across a range of localization length scales and is independent of the manner in which localization is applied. Tests comparing the EnSRF and LETKF for a real-data case also have small differences. The magnitudes of these differences are similar to those that arise from the sampling variability associated with a finite ensemble. Overall, the results suggest the EnSRF and LETKF approaches are equally capable methods for radar data assimilation at convective-scales.

A multi-scale data assimilation framework is developed for an ensemble assimilation and prediction system using the Weather Research and Forecasting (WRF) model and the Data Assimilation Research Testbed (DART). A CONUS mesoscale model domain, with 15 km horizontal grid spacing, provides the boundary conditions for a nested convective-allowing model grid (3 km horizontal grid spacing). The WRF grids and the DART data assimilation toolkit are used to create analyses and forecasts of the convective environment and a tornadic storm on 13 June 2010. This case was chosen because it represents a complex mesoscale environment and storm evolution that was not captured well with conventional observations or WSR-88D radars. Thus, this case presents a challenging event to analyze and predict, and demonstrates the benefits of multi-scale data assimilation for generating initial conditions for convective-scale ensemble forecasts. Several aspects of multi-scale DA cycling are investigated through comparisons of ensemble forecast performance relative to a control 6-hourly cycled analysis system. Results indicate that increased cycling frequency improves forecasts of the mesoscale storm environment and convection. The addition of radar observations in hourly DA cycling leads to further improvement in forecast skill, which is tied to better forecasts of the outflow boundary from overnight convection and subsequent convective evolution. Lastly, the initial conditions for the multi-scale data assimilation cycling system are found to have an impact on the characteristics of the near-storm environment.

Chapter 1: Introduction

1.1 Motivation

Accurate and reliable forecasts of convection remain an elusive challenge. There is an obvious societal need for convective forecasts, particularly for severe convection that has the potential to cause damage. The National Oceanic and Atmospheric Administration (NOAA) is currently researching the operational potential of convective-scale numerical weather prediction via the assimilation of Doppler radar and other fine-scale observations into high-resolution models. This program, called Warn-on-Forecast, focuses on explicit prediction of convective storms to extend severe weather warning lead times using analyses and forecasts from convection-allowing ensemble modeling systems (Stensrud et al. 2009, 2013).

In the past decade, advances in data assimilation for convection have substantially improved the initial conditions and subsequent forecasts (e.g., Snyder and Zhang 2003; Aksoy et al. 2009). However, a number of challenges limit the success of data assimilation on convective-scales and progress toward overcoming them is required for the Warn-on-Forecast mission to be successful (as discussed in Stensrud et al. 2013). The focus of this dissertation is on two data assimilation challenges: determining the data assimilation method that produces the best analyses; and improving the analysis of the environmental conditions to include the accurate representation of mesoscale features.

Since Doppler radar data assimilation is a relatively new research area (Snyder and Zhang 2003), the methods used to assimilate observations are still being developed, and the efficiency and accuracy of different data assimilation techniques on the convective-scale need to be investigated. The Ensemble Kalman filter (EnKF; Evensen 1994) methods are obvious candidates because of their demonstrated ability to assimilate radar and other observations of convection. For example, EnKF assimilation of synthetic Doppler radar observations has been successful in retrieving unobserved variables (e.g., Snyder and Zhang 2003; Zhang et al. 2004; Tong and Xue 2005; Potvin and Wicker 2013). The EnKF has also been successfully used to assimilate real Doppler velocity and radar reflectivity factor (referred to hereafter as reflectivity) observations (e.g., Dowell et al. 2004; Aksoy et al. 2009; Dawson et al. 2012; Potvin et al. 2013). Major advantages of EnKF methods include the production of initial conditions for ensemble forecasts, and ease of implementation relative to four-dimensional variational techniques (Caya et al. 2005; Wang et al. 2008a; Wang 2011). Both data assimilation studies in this dissertation employ EnKF methods.

The essence of the EnKF data assimilation method is the use of an ensemble to provide flow-dependent estimates of the background error covariances. There are several different approaches to implementing an EnKF algorithm. EnKF approaches (that do not perturb observations) are equivalent when the ensemble priors and observations have Gaussian errors and no localization is applied. In real applications, however, these conditions are not met (due, e.g., to model error and non-linear observation operators) and the specific implementation of the EnKF can impact the analysis. The choice of "assimilation pattern", meaning the choice of whether to assimilate all observations simultaneously (Evensen 1994) or sequentially (Houtekamer and Mitchell 1998) is an important part of the EnKF algorithm. The sequential and simultaneous assimilation patterns are not equivalent when covariance localization is applied (Ehrendorfer 2007) and/or when the observation operators are nonlinear. The method used to apply covariance localization is also an important consideration. The two most common approaches are "B-localization" or "covariance localization", applying a localization function to the background error (Houtekamer and Mitchell 2001), and "R-localization" or "observation localization", applying an inverse localization function to the observation error covariance (Hunt et al. 2007). Several previous studies have compared the effects of the assimilation pattern and/or the method of localization (see Section 3.1). Previous studies indicate that complications in real-world applications such as observation type and number, characteristic ratio between background and observation errors, ensemble size, and application of a digital filter might lead to different EnKF performance for specific applications (Miyoshi and Yamane 2007; Nerger et al. 2012; Holland and Wang 2013, hereafter HW2013). This consideration motivates the continued exploration of the relative strengths and weaknesses of different assimilation patterns and localization methods in the present study, including for the real-world application of EnKF radar data assimilation at convective-scales.

The first focus of this dissertation is on the comparison of two EnKF variants: the Ensemble Square Root Filter (EnSRF; Whitaker and Hamill 2002; Section 2.2.4) and the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al. 2007; Section 2.2.6). The EnSRF is the variation that has been most often used in Doppler radar assimilation studies (e.g., Zhang et al. 2004; Dowell et al. 2004; Tong and Xue 2005; Caya et al. 2005; Aksoy et al. 2009; Dawson et al. 2012; Potvin and Wicker 2013). The EnSRF assimilates observations sequentially and typically uses B-localization. The LETKF, on the other hand, assimilates observations simultaneously and uses R-localization. Unlike the EnSRF, the LETKF has not been thoroughly tested using Doppler radar data assimilation at convective-scales. This study applies the LETKF to convective-scale radar data assimilation to examine if (1) it performs with similar accuracy as the EnSRF, and (2) to identify any impacts of the simultaneous update relative to the sequential update used in the EnSRF. If the LETKF compares favorably to the EnSRF when applied to convective scales, then other considerations such as parallelization, algorithmic flexibility, etc., can be considered by the user when choosing a particular method.

To gain understanding of these issues, the EnSRF and LETKF are applied to both an observing system simulation experiment (OSSE) and a real-data experiment. The experiments assimilate Doppler radar data into a three-dimensional cloud model at convective allowing resolution. The OSSE allows the LETKF and EnSRF analyses/forecasts of kinematic and thermodynamic storm structure to be compared quantitatively through the difference total energy and hydrometer difference total energy and qualitatively through two-dimensional reflectivity and vertical motion plots. The analyses/forecasts of storm structure using the LETKF and EnSRF are also compared qualitatively for the real-data assimilation case (8 May 2003 Moore Oklahoma supercell; Burgess 2004; Hu and Xue 2007; Romine et al. 2008; Dowell and Wicker 2009; Dowell et al. 2011; Yussouf et al. 2013). In addition, low-level vorticity forecasts are used as proxy for tornadic circulation and compared to the observed tornado track. These results will be presented in Chapter 3.

The second focus of this dissertation is on the challenge of including environmental conditions in convective forecasts. Since the mid 20th century, severe weather forecasters have linked environmental conditions to the risk of storms developing and the nature of the storms that develop (see Section 2.1.1). Many modeling and severe storm environment studies have shown that the development and evolution of severe convection is sensitive to the environment (e.g., Weisman and Klemp 1982; Thompson et al. 2003; Ziegler et al. 2010). Thus, environmental conditions, its variability and its uncertainty need to be incorporated into convective forecasts. However, previous ensemble radar data assimilation studies have primarily focused on isolated convective events within small domains (~100 squared km) by initializing the data assimilation system with horizontally homogeneous environments (e.g. Dowell and Wicker 2009; Dawson et al. 2012). The success of a convective forecast depends on the identifiable and more predictable environmental forcing features (Weisman et al. 2008) and the use of an inhomogeneous environment can lead to improved ensemble forecasts (Stensrud and Gao 2010). Thus, future ensemble forecasts for convection (including those being developed for the Warn-on-Forecast mission) must include cross-scale interactions that simulate the environmental conditions.

A framework for multi-scale data assimilation that includes analyses and forecasts of the mesoscale environment and the convection is developed. The term multi-scale is used to indicate that multiple spatial and temporal scales are included in the forecasts, analyses, and observations. For example, both the synoptic-scale flow pattern and the convective-scale flow pattern are of interest. Thus, multiple model grids and observations of both radiosondes and radars are included. Herein, multi-scale data assimilation is defined as the combination of radar data assimilation with simultaneous assimilation of conventional observations to analyze/forecast the primary synoptic and mesoscale convective forcing features and the convection itself. There are many open questions regarding multi-scale data assimilation and the following overarching questions will guide the experiments in this study: (1) Can multi-scale data assimilation improve convective forecasts compared to mesoscale or storm-scale data assimilation alone? (2) What type of observations and data assimilation cycling is needed to analyze and forecast the storm environment? (3) How important is an accurate storm environment for obtaining accurate convective forecasts?

This study is among the first attempts to analyze and predict the mesoscale environment and the convection within the same analysis and forecast system. The multi-scale data assimilation techniques will be applied to the 13 June 2010 tornadic supercell environment. On 13 June 2010 a cold pool from overnight convection in southern Kansas created an east-west oriented outflow boundary that was located near the Oklahoma-Texas border in the afternoon. New convection developed in the afternoon along a southwest to northeast oriented cold front located in the Oklahoma and Texas Panhandles. The sub-severe convection slowly moved to the northeast and a cell moved over the intersection of the two boundaries around 2000 UTC. This storm intensified, gained supercell characteristics, and became tornadic. This case represents a complex mesoscale environment and storm evolution that was not captured well with conventional observations or operational radars. Thus, it presents a challenging event to analyze and predict and will demonstrate the benefit of multi-scale data assimilation in generating initial conditions for ensemble forecasts. The impact of radar data assimilation, cycling frequency, and background ensemble initialization on the mesoscale environment and convective forecasts on 13 June will be investigated. The surface temperature, dew point, and winds are used to determine if the surface boundaries are analyzed/forecasted. The reflectivity and updraft helicity analyses/forecasts are used to evaluate the storms location, intensity, and evolution. These results will be discussed in Chapter 4.

1.2 Statement of hypotheses

This dissertation has two hypotheses. First, I hypothesize that the simultaneous update in the LETKF could slightly improve the dynamic balance in analyses of deep convection (as measured by the surface pressure tendency) compared to a sequential filter such as the EnSRF, because the state is updated simultaneously in the LETKF. Although the expectation is that the differences between the LETKF and the EnSRF are small, any differences that do exist could potentially lead to superior forecast performance. The goal of this research is to demonstrate that the LETKF can be applied to radar data assimilation on the convective-scale and to document any differences in filter performance compared to EnSRF.

Second, I hypothesize that for a complex storm environment with surface boundaries playing a significant role, multi-scale data assimilation is required to analyze the boundaries. Further, without an accurate storm environment that contains the boundaries, the convection and the tornadic potential of the convection in this environment cannot be accurately forecasted. The goal of this research is to apply a multi-scale data assimilation approach to 13 June 2010 and investigate the impact of this approach on the storm environment and convective forecasts. In particular, the focus will be on how the design choices of a multi-scale data assimilation system change the storm environment.

1.3 Overview of dissertation

This dissertation is organized as follows. In Chapter 2, introductory material and a review of the literature regarding numerical weather prediction and ensemble data assimilation is presented. Chapter 3 discusses the implementation of LETKF for storm-scale radar data assimilation. The OSSE and real-data comparisons of LETKF and EnSRF are also presented in Chapter 3. In Chapter 4, the investigation of multi-scale data assimilation is discussed. A meteorological and VORTEX2 operations overview on 13 June 2010 are also included in Chapter 4. Further, Chapter 4 contains the examination of the impact of radar data assimilation, cycling frequency and background ensemble initialization on the mesoscale environment and convective forecasts. Principal conclusions are reiterated and future work is outlined in Chapter 5.

Chapter 2: Background

The material in this chapter reviews the definitions, explanations, and previous work related to the experiments performed in the following chapters. A brief review of the history of Numerical Weather Prediction (NWP) for convection, predictability, and ensemble forecasting is presented in Section 2.1. Data assimilation techniques including the algorithms used in this study and Doppler radar data assimilation for convection are discussed in detail in Section 2.2. Section 2.3 briefly summarizes the chapter.

2.1 Numerical Weather Prediction (NWP) for convection

2.1.1 Historical perspective

The concept of numerical weather prediction was shaped by a handful of dedicated meteorologists in the early 20th century. In 1901, Cleveland Abbe called for meteorologists to move beyond empirical rules and use mathematical, graphical and numerical structure to follow the general, and possibly detailed phenomena of the atmosphere (Abbe 1901). Abbe was the first head of the U.S. Weather Bureau, which became the modern day National Weather Service. Vilhelm Bjerknes shared Abbe's desire to make meteorology an exact science. He defined a two-step plan for weather forecasting. First, the "diagnostic" step is the process of determining a sufficiently accurate state of the atmosphere at the initial time. Second, the "prognostic" step uses sufficiently accurate laws to describe how one state of the atmosphere develops from the other (Bjerknes 1904, translated by Mintz 1954). Bjerknes developed a qualitative, graphical method to solve for seven basic variables: pressure, temperature, density, humidity, and three components of velocity. The method used seven independent equations: the three hydrodynamic equations of motion, the continuity equation, the equation of state and the first and second laws of thermodynamics (in fact, he should have specified a continuity equation for water rather than the second law of thermodynamics; Lynch 2008). Although Bjerknes approach was the first attempt to mathematically predict weather patterns, he was not able to put his ideas to practical use. Influenced by Bjerknes ideas, Lewis Fry Richardson attempted the direct solution of the equations of motion using a finite differencing method. His first forecast resulted in a totally unrealistic surface pressure change of 145 hPa in 6 hours (Richardson 1922; Lynch 2008). Richardson speculated that the glaring error resulted from an unnatural initial distribution, which was later shown to be true (Lynch 2006). Implementing Richardson's method during the 1910's was impractical due to the lack of observations, lack of physical understanding of the dynamics, and it preceding the age of digital computers. However, his finite difference method became the foundation for modern numerical forecasting.

Building upon the ideas of Abbe, Bjerknes, and Richardson, Numerical weather prediction (NWP) on digital computers began in 1950 (i.e. "ENIAC Computations of 1950 - Gateway to Numerical Weather Prediction", Platzman 1979). Early NWP focused on forecasting the large-scale weather pattern. At the time NWP was beginning, the conceptual building blocks of a unified theory of large-scale motion were already in place (Brooks et al. 1992). Thus, the advances in the large-scale NWP capabilities paralleled the rapid advancement in computing capabilities. However, convective forecasts have not progressed as quickly because a comprehensive theory of motion on the mesoscale or cloud-scale does not exist (Brooks et al. 1992).

Mesoscale NWP began in the 1970's (see Dudhia 2014 for a thorough review). Mesoscale models have grid sizes in the tens of kilometers, and include several physical parameterizations for cloud/microphysics, convective/cumulus physics, surface physics, mixing/planetary boundary layer physics, and radiative physics. The physical parameterizations are the largest source of uncertainty in mesoscale models (Dudhia 2014) and limit their ability to forecast convection. To study convection, idealized cloud models were developed in the 1970's (e.g. Miller and Pearce 1974; Cotton and Tripoli 1978; Klemp and Wilhelmson 1978). Cloud models have limited physical processes and are run for short time scales over small areas (covering a single convective system). Thus, cloud models are useful research tools but cannot provide operational forecast guidance. Without numerical guidance on the convective-scales, forecasters relied on empirical rules. For example, on March 25, 1948, Air Force Captain Robert C. Miller and Major Ernest J. Fawbush correctly identified that the environment was similar to previous environments in which tornadoes occurred. This lead to the first ever tornado warning being issued for Tinker Air Force Base, Oklahoma (Maddox and Crisp 1999).

The concept of numerically predicting thunderstorms was proposed by Lilly (1990) 25 years ago. The prospect of developing a numerical prediction system at convective-scale resolution was made possible by the development of the national network of Weather Surveillance Radars, 1988-Doppler (WSR-88D; Crum and Albert

1993) in the U.S., advances in the techniques for retrieving unobserved quantities from single-Doppler data, and the advent of increasingly powerful parallel-processing supercomputers (Xue et al. 2000). In the 1990's, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma was established for the mission of demonstrating "the practicability of storm-scale numerical weather prediction and to develop, test, and validate a regional forecast system appropriate for operational, commercial, and research applications" (e.g. Droegemeier et al. 1996a,b; Xue et al. 1996). CAPS performed the first real-time explicit convective-allowing forecasts with radar data assimilation in the Spring of 1996 (Droegemeier et al. 1996b; Xue et al. 1996). The development of numerical forecasts of convective storms embedded within the evolving larger-scale environment has been ongoing since this time.

2.1.2 Predictability

Predictability is a roadblock to accurate NWP. Two types of predictability exist. *Intrinsic* predictability is defined as "the extent to which prediction is possible if an optimum procedure is used in the presence of infinitesimal initial errors" (Lorenz 1969; Zhang et al. 2006). *Practical* predictability is defined as "the ability to predict based on the procedures currently available" (Zhang et al. 2006; Melhauser and Zhang 2012). An understanding of predictability limits is essential to manage expectations of the potential gains that can be achieved through improvements in data assimilation and forecast systems.

Intrinsic predictability limits are unavoidable. Using a low-order model of convection, Lorenz (1963) found that slightly differing initial states could lead to considerably different future states. From these results, Lorenz inferred that "prediction of the sufficiently distant future is impossible by any method, unless the present conditions are known exactly." (Lorenz 1963). Eliminating errors in the initial conditions is impossible, even with a fine observing network. Thus, errors due to intrinsic predictability will always be present in numerical forecasts.

In addition to the intrinsic predictability limit, practical predictability limits the skill of NWP forecasts. Errors in the NWP model, observations, and methods used to create initial conditions, contribute to the practical predictability limit of NWP forecasts. Improving the estimate of the initial atmospheric state can extend the lead-time of a skillful forecast by mitigating practical predictability limits, which is the focus of this dissertation. Obtaining an accurate three-dimensional estimate of the initial atmospheric state is the principal goal of data assimilation. Advances in data assimilation methods can help extend the practical predictability limits.

The forecast sensitivity to initial condition errors can be flow dependent. Forecast errors will grow more rapidly as successively smaller scales are resolved (Lorenz 1969). Previous studies have confirmed that the rapid upscale cascade of small-scale initial error imposes limits on predictability (Lorenz 1969; Leith 1971; Leith and Kraichnan 1972). Lorenz (1969) compared an experiment with initial errors at small wavelengths to an experiment with initial errors at the longest retained wavelength, and found that the predictability was lost rapidly in both experiments. Thus, there is no guarantee that the large-scales can be specified with sufficiently small errors to ensure the correct mesoscale response (Durran and Gingrich 2014). Further, Durran and Gingrich (2014) showed that initial small-scale errors do not matter when relatively minor errors are present in the largest scales. They explain that downscale error propagation is much more rapid than upscale propagation (for turbulent flows whose kinetic energy spectrum is proportional to the -5/3 power of the horizontal wavenumber). Thus, convective-scale NWP forecasts are sensitive to errors in the initial conditions of the mesoscale storm environment.

Zhang et al. (2002; 2003) investigated the predictability limits for a real-data case, the "surprise" snowstorm in 2000. The snowstorm on 24-25 January 2000, along the east coast of the U.S. (between North Carolina and Washington, D.C.) is referred to as a surprise because it was not predicted by the operational numerical models. The operational Eta Model, which had 32 km horizontal grid spacing, missed most of the observed precipitation. However, the research, nonhydrostatic Mesoscale Model (MM5; Dudhia 1993), with nested 10 and 3.3 km horizontal grid resolution produced a precipitation forecast that closely fit the observations. Studying the MM5 model, Zhang et al. (2002) found that the forecast error growth was much smaller when the effects of latent heat release was turned off, implicating that convective processes contribute to rapid error growth in forecasts. In addition, Zhang et al. (2003) found that error growth at scales of 100-200 km over the first 6-hour forecast spreads to larger scales in the subsequent 12-hour forecasts. This places severe constraints on the accuracy of mesoscale forecasts, particularly when convection is ongoing. Further, Zhang et al (2006) investigated an extreme precipitation event and found that realistic initial condition uncertainty can result in large forecast errors for a heavy rain event. When initial condition and model errors are large, forecast errors arising from small-scale, small-amplitude random noise are of secondary importance (though not negligible) (Zhang et al. 2006). Thus, higher forecast accuracy can potentially be obtained through improving the initial analysis with better data assimilation methods or enhanced observations.

The current practical predictability limits can be extended via improvements to the initial analysis from better data assimilation techniques. Practical atmospheric predictability can also be extended through the use of ensemble forecasts, which will be discussed in the following section.

2.1.2 Ensemble forecasting

An ensemble of concurrently valid forecasts can be created from different initial conditions, different models, through the use of different physical parameterizations, and/or the same model initiated at different times. Ensemble forecasts created from slightly perturbed initial conditions that represent the uncertainty in the analysis, can be averaged to filter out some of the unpredictable components of the forecast (Leith 1974). An ensemble mean can be used to represent the best estimate of the true state in a least square error sense (Toth and Kalnay 1993; Stensrud et al 2000; Wang and Bishop 2003). The ensemble spread (standard deviation about the mean) can provide information on the uncertainty of the forecasts (Leith 1974; Toth and Kalnay 1993; Stensrud et al 2000; Wang and Bishop 2003). Ensemble forecasts can also be used to generate probabilistic forecasts (i.e. ensemble frequencies) of future weather events.

The additional products that ensemble forecasts provide (mean, spread, probabilities) extend the practical predictability of a skillful numerical forecast compared to a single deterministic forecast.

An ensemble of forecasts is particularly important for convective forecasts due to the difficulties in modeling convective initiation, precipitation structure, and convective flows, which lead to rapid model error growth (e.g., Brooks et al. 1992; Stensrud et al. 2000; Dawson and Xue 2006). For example, Brooks and Doswell (1993) pointed out that numerical models typically perform the best when large-scale, quasi-geostrophic forcing dominates, which is also when the forecasting situation is the easiest. Brooks and Doswell (1993) also suggested that numerical guidance is most needed when mesoscale and storm-scale details dominate the forecast situation. However, the numerical guidance is more likely to be wrong when mesoscale and storm-scale details dominate. Thus, an ensemble of forecasts that provide guidance on the uncertainties is desired.

Short-range ensemble forecasting (SREF) refers to applying ensemble NWP techniques to regional scale (1-3 days). Advancements in computational capabilities have facilitated NWP progress of SREF systems (e.g., Du et al. 2006; Clark et al. 2012a). Further, advances in the understanding of mesoscale and storm-scale phenomena, NWP models, initial conditions, and observations have contributed to the ongoing development of SREF systems.

The ensemble forecasts created for the Spring Experiments held in the NOAA Hazardous Weather Testbed (preceded by the Science Support Area) have been the cornerstone of SREF development (Kain et al. 2003; Kain et al. 2005; Kain et al.

2008a,b; Xue et al. 2008; Coniglio et al. 2010; Clark et al. 2011; Johnson and Wang 2012; etc.). The real-data, real-time Spring Experiment forecasts have begun to address the challenge of providing useful information on storm attributes for severe weather forecasting. The early Spring Experiment output with parameterized-convection resolution (grid spacing coarser than 4 km; Kain et al. 2003) laid the groundwork for the more recent convective-allowing resolution models (Clark et al. 2011).

Convective-allowing models (CAMs) have sufficient resolution to develop convective circulations explicitly, precluding the need for convective parameterization. The horizontal grid spacing of CAMs is between 1 and 4 km. Kong et al. (2006) advanced SREF via a case study with multiple nested grids including a CAM (spacings of 24, 6, and 3 km), terrain, horizontally varying initial conditions, and the assimilation of real Doppler radar observations. Kong et al. (2006,2007) showed that for a 3-km horizontal grid spacing forecast, an ensemble has greater value than a single deterministic forecast, and the 3-km ensemble has greater value than an ensemble or deterministic forecast at coarser spacing. Clark et al. (2009) also found that the use of a CAM with limited size ensemble (5 member) outperforms a larger ensemble (15 member) of parameterized-convection resolution, indicating the importance of resolving finer scales. Further, CAMs improve the simulation of convective diurnal characteristics (Done et al. 2004; Lui and Moncrieff 2007) and convective mode and structure (Clark et al. 2007; Kain et al. 2008a,b; Schwartz et al. 2009; Sobash et al. 2011). Thus, the use of CAM for severe weather forecasts is superior to mesoscale models.

Despite the progress in SREF, significant scientific and technical challenges regarding the details of data assimilation methodology, quality of observations, model parameterizations, storm environment representation, and physical understanding need to be addressed. The importance of these issues will continue as NWP moves from CAMs to convective-resolving scales (horizontal grid spacing on the order of 100 m). For CAMs, it may not be possible to accurately predict details of the convection, such as precipitation distribution and amount and propagation speed (Bryan et al. 2003). Further, the current formulation of convective-resolving models with traditional subgrid large-eddy simulation are not suitable for grid spacing of the order 1 km, rather grid spacing on the order of 100 m is appropriate (Bryan et al. 2003). Thus, it is likely that the movement towards progressively small grid spacings will continue.

CAMs are difficult to initialize due to the coarse resolution of traditional observational data sources. If coarse-resolution initial and boundary conditions are used, there is a “spinup” period of approximately 3-hours for smaller-scale circulations to develop (Skamarock 2004). Initialization that includes observations on scales that commensurate with model resolution could significantly improve the CAM forecasts. Thus, improvements to data assimilation including the assimilation of Doppler radar data and other fine-scale observations are needed. “Data assimilation, in the overall process of forecasting convective precipitation, may be the most critical path through which the pace of forecast advances will be modulated.” (Fritsch and Carbone 2004).

2.2 Data assimilation

As convective forecasts have progressed, so have the techniques used to initialize forecasts. The diagnostic step, introduced by Bjerknes (1904, translated by Mintz 1954) is the process of creating initial conditions for subsequent forecasts. It is vital for producing accurate weather forecasts of the future state. Data assimilation is the essential tool for this step. Data assimilation is the process through which all of the available information (observations and the physical laws that govern the evolution of the flow) is used to determine as accurately as possible the state of the atmosphere (Talagrand 1997). For example, the best estimate of the true temperature at a given point uses information from both observations and their errors with numerical forecasts and their errors.

Doppler radars represent the only regularly available observations with the spatial and temporal resolution to sample convective structures. Thus, throughout the following sections there is an emphasis on radar data assimilation. However, the techniques discussed are also used for assimilation of all observation types.

The following discussion does not include a description of all analysis systems. Methods such as the successive corrections method (SCM) or observation nudging, optimal interpolation (OI; Gandin 1963), state estimation through retrieval (Gal-Chen 1978; Sun et al. 1991; Shapiro et al. 1995; Shapiro et al. 2003), adjoint (Qiu and Xu 1992; Xu et al. 1994; Gao et al. 2001), and dual-Doppler (Doviak et al. 1976; Gal-Chen and Kropfli 1984; Ray et al. 1975; Ray et al. 1980; Dowell and Bluestein 1997) techniques are not discussed because the focus of this work is on the forefront of advanced ensemble data assimilation techniques.

2.2.1 Variational approaches

Variational data assimilation methods are actively used for convective-scale data assimilation (e.g. Gao et al. 2004). They use a cost function composed of terms that represent the departure between an analysis and various pieces of information (i.e. observations and physical laws) (Sasaki 1970). Variational methods are different from other single Doppler retrieval techniques because the model variables are determined simultaneously in a dynamically consistent way (Sun and Crook 1998). Variational methods seek to obtain a solution to the cost function through an iterative approach. In three-dimensional variational assimilation (3DVar) the cost function includes a static estimate of background errors. However, the true background error covariance structure is flow-dependent (especially for meso- and convective-scales).

Four-dimensional variational assimilation (4DVar) includes observations within an assimilation window, rather than observations at a single time (Talagrand and Courtier 1987; Sun and Crook 1998, Caya et al. 2005). 4DVar minimizes the difference between the observations and the model predictions as in 3DVar, but 4DVar evolves the background error covariance matrix and compares the observations to the analysis state valid at the same time (Lorenc and Rawlins 2005). Although 4DVar is a valuable technique, it is limited due to the requirement of developing a tangent-linear and adjoint model. Thus, physics parameterization schemes cannot easily be changed or replaced. 4DVar also does not provide an ensemble of initial conditions for ensemble forecasts. Further, Caya et al. (2005) compared 4DVar and Ensemble Kalman filtering (EnKF;

discussed in the following sections) at the convective scale and found that EnKF produced better analyses after at least 10 minutes of data assimilation. Major advantages of EnKF methods include the production of initial conditions for ensemble forecasts, and ease of implementation relative to four-dimensional variational techniques (Caya et al. 2005; Wang et al. 2008a; Wang 2011).

2.2.2 Kalman filter

The Kalman Filter (KF; Kalman 1960) is a data assimilation technique that estimates the state and the uncertainty. Consider a discrete representation of the atmosphere on a three-dimensional numerical model grid with a large number of observations. The state of the system, **x**, consists of every variable (θ, u, v, w, π, q_c, etc.), at every gridpoint, concatenated into a single vector of length N_x. The true state of the atmosphere, $\mathbf{x}^t$, is considered a random variable because it cannot be exactly determined. Thus, the KF will use observations and governing equations to estimate and forecast the probability distribution function (hereafter PDF), denoted p( ), of $\mathbf{x}^t$.

Begin with a background forecast, which is the first guess at the true state and a set of N_y observations, $\mathbf{y}^o$, which try to measure the true state, $\mathbf{x}^t$. The KF assumes the observations are unbiased and linearly related to $\mathbf{x}^t$:

$$\mathbf{y} = \mathbf{H}\mathbf{x}^t + \varepsilon \, , \tag{1.}$$

where **H** is a $N_y \times N_x$ matrix , the observation operator, mapping the state variables onto the observations. **H** often includes the interpolation to the observation location as well as the conversions to the observed quantity. For example, the observation operator will convert model winds to Doppler velocities. The observational error, $\boldsymbol{\varepsilon}$, is a random error vector of dimension N_y that is independent of $\mathbf{x}^t$ and includes instrument errors and representativeness errors. The KF assumes that the PDF of $\boldsymbol{\varepsilon}$ is Gaussian, with zero mean and known covariance $\mathbf{R}$. The PDF of the true state, $\mathbf{x}^t$, given the observations, $\mathbf{y}^o$, denoted $p(\mathbf{x}^t \mid \mathbf{y}^o)$, is also assumed Gaussian. The KF provides formulas for calculating $p(\mathbf{x}^t \mid \mathbf{y}^o)$, which have mean $\overline{\mathbf{x}^a}$ and covariance $\mathbf{P}^a$,

$$\overline{x^a} = \overline{x^f} + \mathbf{K}(y^o - \mathbf{H}\overline{\mathbf{x}^f}), \qquad (2.)$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f, \qquad (3.)$$

where the weight is called the Kalman gain, $\mathbf{K}$,

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1}. \qquad (4.)$$

The optimal gain matrix is found using error covariances. In general, an error covariance matrix, $\mathbf{P}$, is obtained by multiplying a vector error by its transpose and averaging over many cases to obtain an expected value,

$$\mathbf{P} = \overline{\varepsilon\varepsilon^T}, \qquad (5.)$$

where $\boldsymbol{\varepsilon}$ is a vector error and the overbar represents expected value (Kalnay 2006).

$\mathbf{P}^f\mathbf{H}^T$ in Equation (4.) is the forecasted (or background or prior) covariance of the state and observed variables, and Houtekamer and Mitchell (1998) define $\mathbf{P}^f_{xy} = \mathbf{P}^f\mathbf{H}^T$. The forecasted covariance is, for example, the covariance between the forecasted potential temperature (state variable) and the forecasted Doppler velocity (observed variable) and is expressed as,

$$\mathbf{P}^f_{xy} = \mathbf{P}^f\mathbf{H}^T = \mathrm{cov}(x^t, \mathbf{H}(x^t)). \qquad (6.)$$

In the KF the forecast error covariance, $\mathbf{P}^f$, is advanced for each analysis cycle using the model itself. Thus, it is assumed that the forecast model error covariances are known. The forecast model is also assumed to be linear and unbiased. Lastly, the KF assumes the forecast errors and the observation errors are uncorrelated so their covariance is zero (Kalnay 2006). For Doppler radar data assimilation with an advanced atmospheric model the KF assumptions are not valid making the analyses suboptimal. Also, it is difficult to implement the error covariance calculation because of the computational cost, nonlinear dynamics, and poorly characterized error source (Tippett et al. 2003).

2.2.3 Ensemble Kalman Filter

Using an ensemble representation of the forecast and analysis error covariances in the KF decreases the computation cost and mitigates problems with the nonlinear dynamics (Evensen 1994). Evensen (1994) introduced the Ensemble Kalman filter (EnKF), which is a Monte Carlo approximation to the KF. In other words, instead of estimating and forecasting $p(\mathbf{x}^t \mid \mathbf{y}^o)$, an ensemble is used to represent a sample of $p(\mathbf{x}^t \mid \mathbf{y}^o)$ (Snyder and Zhang 2003). In EnKF, the model error covariances are an approximation that typically underestimates the true covariances and assimilating observations reduces the ensemble spread. This underestimation leads to a systematic underweighting of the observations in the Kalman gain, which then leads to filter divergence. Therefore, artificial means of maintaining ensemble spread are required.

Houtekamer and Mitchell (1998) applied EnKF to atmospheric data assimilation and took a stochastic approach to maintain ensemble spread. They used perturbed sets of observations to update each member. Perturbed observations are the actual observations plus random noise. However, the use of perturbed observations reduces the accuracy of analysis error covariances and increases the probability of the ensemble underestimating the analysis error covariance (Whitaker and Hamill 2002). Alternatively, deterministic EnKF algorithms first update the ensemble mean, then the ensemble members are updated in a manner that maintains the analysis error covariance. The deterministic algorithms used in this study are discussed in the following sections.

2.2.4 Ensemble Square Root Filter

Whitaker and Hamill (2002) developed a deterministic EnKF algorithm that avoids the systematic underestimation of the analysis covariance, by using a “reduced” Kalman gain to update the deviations from the ensemble mean. This method involves the square root of observation error covariance, making it a Monte Carlo implementation of a square root filter. Thus, Whitaker and Hamill (2002) called this method the Ensemble Square Root Filter (EnSRF). Their EnSRF method requires the observations to be processed one at a time to avoid the computation of matrix square roots. This is called sequential assimilation.

The EnSRF was the first EnKF variation applied to the assimilation of Doppler radar observations (Snyder and Zhang 2003). Algorithmically, EnSRF loops over each observation valid at the assimilation time, applying the observation operator to the prior ensemble state and then updating the analysis using the Kalman gain and innovation.

Using the notation from Holland and Wang (2013; hereafter HW03), the following equation is used to update the ensemble mean state vector, $\overline{\mathbf{x}}^{\mathrm{a}}$,

$$\overline{\mathbf{x}}^{\mathrm{a}}=\overline{\mathbf{x}}^{\mathrm{b}}+\mathbf{K}\left(\mathbf{y}-\overline{H(\mathbf{x}^{\mathrm{b}})}\right), \tag{7.}$$

where $\overline{\mathbf{x}}^{\mathrm{b}}$ is the background or forecasted ensemble mean state vector. Both $\overline{\mathbf{x}}^{\mathrm{a}}$ and $\overline{\mathbf{x}}^{\mathrm{b}}$ have dimensions of nx1, where n is the number of state variables. $\mathbf{y}$ contains the observations. In the serial EnSRF, $\mathbf{y}$ is a single, scalar observation. H is the nonlinear observation operator. The Kalman gain $\mathbf{K}$, is given by:

$$\mathbf{K}=\boldsymbol{\rho}_{\mathbf{B}}\circ\left(\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\right)\left(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}+\mathbf{R}\right)^{-1}, \tag{8.}$$

where $\mathbf{P}^{\mathrm{b}}$ is the background error covariance matrix estimated by the ensemble sample covariance. $\mathbf{R}$ is the observations' error covariance. In the serial EnSRF, $\mathbf{R}$ is a scalar representing observation error variance. $\circ$ represents a Schur product (element-by-element multiplication), and $\boldsymbol{\rho}_{\mathbf{B}}$ is the localization matrix. Since $\boldsymbol{\rho}_{\mathbf{B}}$ is applied on the background error covariances, it is denoted as B-localization. In the sequential EnSRF, $\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ is a scalar, and $\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ and $\boldsymbol{\rho}_{\mathbf{B}}$ are nx1 matrices. As in Houtekamer and Mitchell (1998), the full $\mathbf{P}^{\mathrm{b}}$ matrix is not computed. Instead, the nonlinear observation operator is applied on the state vectors before calculating the ensemble covariances.

The ensemble perturbations are updated according to the following equation:

$$\mathbf{X}^{a}=\mathbf{X}^{\mathrm{b}}-\widetilde{\mathbf{K}}\left(\mathbf{H}\mathbf{X}^{\mathrm{b}}\right), \tag{9.}$$

where $\mathbf{X}^{\mathrm{a}}$ is the nxk analysis ensemble perturbation matrix (ensemble members minus ensemble mean), and $\mathbf{X}^{\mathrm{b}}$ is the nxk background ensemble perturbation matrix. $\mathbf{H}\mathbf{X}^{\mathrm{b}}$ is calculated by first applying the nonlinear observation operator on the state vectors, and then taking the ensemble mean out. The “reduced” Kalman gain matrix, $\widetilde{\mathbf{K}}$, has dimensions of nx1, and is computed as:

$$\widetilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{HP}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}+\mathbf{R}}}\right)^{-1} \mathbf{K}. \qquad (10.)$$

The reduction to the Kalman gain arises because the scheme does not use perturbed observations (Whitaker and Hamill 2002). In the sequential EnSRF, $\mathbf{HP}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ and $\mathbf{R}$ reduce to scalars.

After the first observation is assimilated, the resulting analysis state is used as the background for the second observation, and so on. In the EnSRF implementation in this study, reflectivity and terminal fall velocity are treated as state variables in the filter (but not in the forecast model). This means that they are updated by each observation and then the updated values are used for subsequent observation assimilation (observation operator is interpolation only) (Anderson and Collins 2007). The alternative is to apply the full observation operator to re-compute the reflectivity and terminal fall velocity from the updated prognostic state variables after each observation is assimilated. Both strategies were compared and no significant differences were produced between the two methods (not shown). EnSRF will be used in Chapter 3.

2.2.5 Parallelized Ensemble Adjustment Kalman Filter

Anderson (2001) developed a deterministic EnKF algorithm that is mathematically equivalent to the EnSRF (Tippett et al. 2003), called the Ensemble Adjustment Kalman Filter (EAKF). Anderson and Collins (2007) realizes parallelization for the sequential EAKF algorithm by calculating the prior ensemble estimates for each observation initially, then the observation increment updates both the state vector and the subsequent prior ensemble estimates[1].

The parallel algorithm presented in Anderson and Collins (2007) proceeds as follows. First, the observation operator is applied to the background ensemble for all available observations. Next, the increment is calculated from the first observation and it's observation error variance. The increments are applied to update the state vector and the prior ensemble estimates for all subsequent observations. The parallel algorithm allows the computations to be partitioned onto any number of processors and produce identical answers to a single-processor implementation (for the low-latency implementation). When the observation operator is a complicated, high cost calculation, time can be saved by updating the prior ensemble estimates directly. Further, assimilation of derived variables becomes much easier with the parallel method. For example, rainfall is derived from a physical parameterization package over sequence of model time steps. Thus, in order to assimilate accumulated rainfall observations you must make a model forecast, calculate observation operator, calculate increment, update model state vector, and re-run model forecast. This process must be repeated for each observation. Thus, it is impractical in large models. With the parallel algorithm re-computing the forecast is unnecessary because the prior ensemble is updated directly. One caveat is that the parallel algorithm is only identical to the sequential algorithm for linear observation operators. But, derived variables like

---

[1] The Ensemble Square Root Filter can also be implemented as a parallel algorithm. However, the 'EnSRF' refers to the sequential algorithm in this text.

rainfall often have highly nonlinear observation operators. Thus, the results will not be the same as the sequential algorithm.

The Anderson and Collins (2007) parallel algorithm can be implemented for several flavors of EnKF to optimize performance for a variety of parallel computing environments. Depending on the computing system the low-latency or high-latency implementation can be used to speed up the assimilation and decrease computational cost. The parallel algorithm does require additional calculations but in some cases it can substantially decrease assimilation time. In general, the Anderson and Collins (2007) algorithm parallelizes as much as possible the serial EnSRF method. The parallel version of EAKF will be used in Chapter 4.

2.2.6 Local Ensemble Transform Kalman Filter

The Ensemble Transform Kalman Filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003) is another deterministic square root EnKF method (Tippett et al. 2003). The ETKF ensemble is used to estimate the forecast error covariance for predicting the analysis error covariance but it is not used for updating the mean (Wang and Bishop 2003). Therefore, the control analysis may not be as accurate as the control analysis in the EnSRF but the computational expense of the ensemble generation in ETKF is considerably less. The Local Ensemble Transform Kalman Filter (LETKF) is a local implementation of the ETKF (Hunt et al. 2007).

Hunt et al. (2007) developed the LETKF with the primary goals of ease of use and improved computational speed compared to previous EnKF variations. The latter goal is achieved by generating local analyses for each grid point independently through simultaneously assimilating local observations. The LETKF has been used as a data assimilation scheme for both global and regional atmospheric models (e.g., Ott et al. 2004; Szunyogh et al. 2005; Miyoshi and Yamane 2007; Miyoshi 2010; Miyoshi and Kunii 2012). The LETKF is beginning to be applied to assimilating Doppler radar data into convection-permitting models to improve rainfall forecasts (e.g., Tsai et al. 2012). Chapter 3 focuses on using the LETKF for Doppler radar data assimilation on scales where storm structures are important.

Algorithmically, the LETKF begins by applying the observation operator to the prior ensemble state for all of the observations valid at the time of the assimilation. For each grid point, all of the observations within the localization cutoff length are identified and used to simultaneously update the model state at that point. Using the notation from HW2013 and following Hunt et al. (2007), the mean update at a grid point is given by

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{b}} + \mathbf{X}^{\mathrm{b}}\overline{\mathbf{w}}^{\mathrm{a}}, \tag{11.}$$

where the dimensions of $\overline{\mathbf{x}}^{\mathrm{a}}$ and $\overline{\mathbf{x}}^{\mathrm{b}}$ are 1x1. The ensemble perturbation matrix, $\mathbf{X}^{\mathrm{b}}$**,** has dimensions of 1x*k*, where *k* is the number of ensemble members. The *k*x1 "weight" vector for the observations within the local localization radius is given by

$$\overline{\mathbf{w}}^{\mathrm{a}} = \left\{(k-1)\mathbf{I} + \left(H\mathbf{X}^{\mathrm{b}}\right)^{\mathrm{T}}(\boldsymbol{\rho}_{\mathbf{R}} \circ \mathbf{R})^{-1}\left(H\mathbf{X}^{\mathrm{b}}\right)\right\}^{-1}\left(H\mathbf{X}^{\mathrm{b}}\right)^{\mathrm{T}}(\boldsymbol{\rho}_{\mathbf{R}} \circ \mathbf{R})^{-1}\left(\mathbf{y} - \overline{H(\mathbf{x}^{\mathrm{b}})}\right), \tag{12.}$$

where **I** is a *k*x*k* identity matrix. $\mathbf{HX}^{\mathrm{b}}$ has dimensions of *p*x*k*, where *p* is the number of local observations. **R** has dimensions of *p*x*p*. $\boldsymbol{\rho}_{\mathbf{R}}$ is a *p*x*p* diagonal localization matrix with non-zero elements equal to the inverse of the corresponding elements of $\boldsymbol{\rho}_{\mathbf{B}}$ in

Equation (8.). Since the localization is applied through modifying the observation error covariance, this method of localization is called R-localization. The perturbation update is given by

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{b}}\mathbf{W}^{\mathrm{a}}, \tag{13.}$$

where the dimensions of $\mathbf{X}^{\mathrm{a}}$ and $\mathbf{X}^{\mathrm{b}}$ are 1x*k*. The *k*x*k* weight matrix, $\mathbf{W}^{\mathrm{a}}$, is given by

$$\mathbf{W}^{\mathrm{a}} = \left[(k-1)\mathbf{I}\left\{(k-1)\mathbf{I} + \left(\mathbf{HX}^{\mathrm{b}}\right)^{\mathrm{T}}(\boldsymbol{\rho}_{\mathbf{R}} \circ \mathbf{R})^{-1}\left(\mathbf{HX}^{\mathrm{b}}\right)\right\}^{-1}\right]^{1/2}. \tag{14.}$$

The core LETKF algorithm used in this study, based on Miyoshi (2010), was obtained from http://code.google.com/p/miyoshi/.

2.2.7 Hybrid data assimilation methods

Data assimilation systems that merge ensemble based methods and variational methods are referred to as hybrid methods. Since hybrid techniques have gained increasing interest in the research and operational NWP communities, they are briefly introduced here for completeness. Instead of using static covariance in a variational system, hybrid techniques employ a variational system and estimate the background error covariance flow-dependently from an ensemble of background states. The ensemble background states are typically produced by an EnKF variant. The potential advantage of the coupled ensemble-variational system compared to a stand alone variational system is the flow-dependent covariances (Wang et al. 2008a,b). Compared to a stand alone EnKF system, the hybrid method is more robust for small ensemble sizes or large model errors, and benefits from dynamic constraints during the variational minimization (Wang 2010; Wang and Lei 2014). Future studies should consider the use of hybrid techniques.

2.3 Summary

The development of advanced data assimilation techniques for the initialization of convective-scale ensemble forecasts are possible today due to the extensive previous research in the areas of numerical weather prediction, predictability, storm dynamics, and Doppler radar. The origins of numerical weather prediction were connected to the onset of digital computers. However, the prospect of numerically predicting thunderstorms did not exist until the 1990s and is related to the availability of Doppler radar observations. Predictability limits can restrict the length of a skillful numerical forecast. However, improving data assimilation techniques used to initialize forecasts can limit the practical predictability constraints. Ensemble forecasts can provide an estimate of uncertainty and extend the practical predictability of a skillful numerical forecast compared to a single deterministic forecast. This study builds on the previous attempts to numerically predict thunderstorms and works to extend the predictability of a convective event using advanced data assimilation for initial conditions of ensemble forecasts.

The ensemble Kalman Filter techniques developed and implemented in many previous studies are used in this study. The Ensemble Square Root Filter and the Local Ensemble Transform Kalman Filter are compared for the convective-scale assimilation of Doppler radar observations in Chapter 3. Chapter 4 uses the Ensemble Adjustment Filter in the development of a multi-scale data assimilation system.

Chapter 3: A comparison between the Local Ensemble Transform Kalman Filter and the Ensemble Square Root Filter for the assimilation of radar data in convective-scale models

3.1 Introduction

Exploring the efficiency and accuracy of different EnKF data assimilation techniques on the convective-scale is part of the Warn-on-Forecast project (as discussed in Chapter 1). The essence of the EnKF data assimilation method is the use of an ensemble to provide flow-dependent estimates of the background error covariances (as discussed in Chapter 2). Deterministic EnKF approaches (observations are not perturbed) are equivalent when the ensemble priors and observations have Gaussian errors and no localization is applied. In real applications, however, these conditions are not met (due, e.g., to model error and non-linear observation operators) and the specific implementation of the EnKF can impact the analysis. The choice of "assimilation pattern", meaning the choice of whether to assimilate all observations simultaneously (Evensen 1994) or sequentially (Houtekamer and Mitchell 1998) is an important part of the EnKF algorithm. The sequential and simultaneous assimilation patterns are not equivalent when covariance localization is applied (Ehrendorfer 2007) and/or when the observation operators are nonlinear. The method used to apply covariance localization is also an important consideration. The two most common approaches are "B-localization" or "covariance localization", applying a localization function to the background error (Houtekamer and Mitchell 2001), and "R-localization" or "observation localization", applying an inverse localization function to the observation error covariance (Hunt et al. 2007). Several previous studies have compared the effects of the assimilation pattern and/or the method of localization. For example, Whitaker et al. (2008) found a simultaneous scheme with R-localization and a sequential scheme with B-localization had similar performance. Greybush et al. (2011) obtained similar analysis error and balance from a sequential scheme with B-localization, a simultaneous scheme with B-localization, and a simultaneous scheme with R-localization. HW2013 compared the choice of assimilation pattern and the choice of B- or R-localization in a two-layer primitive equation model. They found that the combination of the simultaneous assimilation with the R-localization method produced the smallest analysis errors. Their diagnostic experiments show that such differences were associated with the different amounts of dynamical imbalance in the analysis as a result of systematic differences between the schemes in mass and wind increments. Previous studies indicate that complications in real-world applications such as observation type and number, characteristic ratio between background and observation errors, ensemble size, and application of a digital filter might lead to different EnKF performance for specific applications (Miyoshi and Yamane 2007; Nerger et al. 2012; HW2013). This consideration motivates the continued exploration of the relative strengths and weaknesses of different assimilation patterns and localization methods in the present study, including for the real-world application of EnKF radar data assimilation at convective scales.

This chapter compares two EnKF variants: the Ensemble Square Root Filter (EnSRF; Whitaker and Hamill 2002; see Section 2.2.4) and the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al. 2007; see Section 2.2.6). The EnSRF is the variation that has been most often used in Doppler radar assimilation studies (e.g., Zhang et al. 2004; Dowell et al. 2004; Tong and Xue 2005; Caya et al. 2005; Aksoy et al. 2009; Dawson et al. 2012; Potvin and Wicker 2013). The EnSRF assimilates observations sequentially and typically uses B-localization. The LETKF, on the other hand, assimilates observations simultaneously and uses R-localization. Unlike the EnSRF, the LETKF has not been thoroughly tested using Doppler radar data assimilation at convective-scales. This chapter applies the LETKF to convective-scale radar data assimilation to examine if it performs with similar accuracy as the EnSRF, and to identify any impacts of the simultaneous update relative to the sequential update used in the EnSRF. If the LETKF compares favorably to the EnSRF when applied to convective scales, then other considerations such as parallelization, algorithmic flexibility, etc., can be considered by the user when choosing a particular method.

Given the chaotic and unbalanced nature of convection, and the complexities of Doppler radar observations, the Observing System Simulation Experiments (OSSE) framework is used to make initial comparisons of the filters at convective scales. Simulated Doppler velocity and reflectivity data are assimilated using both algorithms in a set of perfect model OSSEs. Filter analysis and forecast performance are compared (Section 3.5), as well as the sensitivity to localization length (Section 3.4) and implementation (Section 3.6). Experiments with only Doppler velocities assimilated are also examined to compare the performance of the two filters given a quasi-linear observation operator (Section 3.7). Lastly, the filter performance is compared using a real-data assimilation case (8 May 2003 Moore Oklahoma supercell; Section 3.8) to extend the OSSE results.

3.2 EnKF algorithms

To facilitate comparisons between the EnSRF (Section 2.2.4) and LETKF (Section 2.2.6), both algorithms were written in a simple framework using a hybrid code comprised of Python and Fortran. Most aspects of the data assimilation systems are the same for both algorithms. For example, both filters create analysis increments on the unstaggered (A-), which means all variables in the state vector are at the same grid point. The use of the unstaggered grid simplifies the code and improve its performance. This necessitates that the model velocity field be destaggered (from the C-grid) and (after being updated) restaggered using the 4th-order interpolation formulas from Sanderson and Brassington [1998; see their equations (2) & (3)]. The interpolation formulas are not reversible, and therefore introduce small errors during the A-C grid conversions. However, analyses do not qualitatively change when data assimilation updates are done on the C-grid instead of the A-grid (not shown). The observation operator used to convert from model space to observation space is also the same in both filters. For Doppler velocity, a point operator is used. (The three components of the model wind and the terminal fall velocity from the microphysical parameterization are trilinearly interpolated to the observation location.) Thompson et al. (2012) showed that use of a simple point operator did not impact the data assimilation results compared to using a more realistic volumetric radar-sampling operator. The observation operator for reflectivity, simply consists of the reflectivity values computed by the microphysics scheme being trilinearly interpolated to the observation locations. Within the Ziegler

Variable Density (ZVD) microphysics scheme (Mansell et al. 2010, Ziegler 1985) used in this study, the total equivalent reflectivity is calculated from the sum of the reflectivities for all hydrometeor species (raindrops, dry ice, wet ice, snow, graupel, hail) using the equations presented by Ferrier (1994; see their Appendix C).

Both filters apply a form of localization to restrict the update of state variables to a region within a certain radius of the observation location and proportions the influence of the observation based on distance to each grid point. These restrictions account for the fact that correlations between an observation and distant grid points become small relative to the sampling errors (due to the finite ensemble size) in the background error covariance estimates (Houtekamer and Mitchell 1998; Snyder and Zhang 2003). The localization isotropic weighting function is the fifth-order correlation function (Gaspari and Cohn 1999). Due to the different localization implementations in each algorithm, the shape of the localization function is slightly different for each filter (Nerger et al. 2012; HW2013). The impacts of localization cutoff length and implementation are explored in Section 3.4 and Section 3.6, respectively.

3.3 Observing System Simulation Experiment design

3.3.1 Model and data assimilation system

The similarities and differences between the two data assimilation methods are first compared using OSSEs with the perfect model assumption. The OSSE framework is chosen to allow comparisons in a setting where the truth is known, and is implemented as in Thompson et al. (2012). The National Severe Storms Laboratory

Collaborative Model for Multiscale Atmospheric Simulation (NCOMMAS; Wicker and Skamarock 2002; Coniglio et al. 2006) with Ziegler Variable Density (ZVD) microphysics scheme (Mansell et al. 2010, Ziegler 1985) is used for the supercell ("truth") simulation. A flat earth approximation is used and no land surface or radiation effects are included. The model domain size is 100 km in the horizontal and 20 km in the vertical. The model grid moves to match the storm motion, which is 8 m s^{-1} toward the east and 7 m s^{-1} toward the north. The domain is designed to encompass the storms and their outflows throughout the simulation period. The horizontal grid spacing is 1 km, and the vertical grid spacing increases from 200 m near the surface to a maximum 600 m near the model domain top. The ensemble members in the assimilation experiments use the same grid setup except the horizontal domain is 120 km.

The initial environment is derived from a sounding that is favorable for severe convection, taken at Weatherford, Oklahoma on 29 May 2004 (Thunderstorm Electrification and Lightning Experiment; MacGorman et al. 2008). A warm bubble is used to initiate convection, and a storm develops after ~25 minutes in the truth simulation. The simulated storm quickly becomes strong with reflectivity exceeding 65 dBZ by 44 minutes (Figure 3.1b). By 56 minutes, the storm begins to exhibit supercellular structure, including a divided mesocyclone containing both updrafts and downdrafts (Lemon and Doswell 1979) and hook echo radar signature (Fujita 1958) (Figure 3.1c). At 68 minutes, additional convection develops on the left flank of the supercell (Figure 3.1d). At 80 and 92 minutes, the mature supercell coexists with a less organized storm that moves off to the north (Figure 3.1e,f).

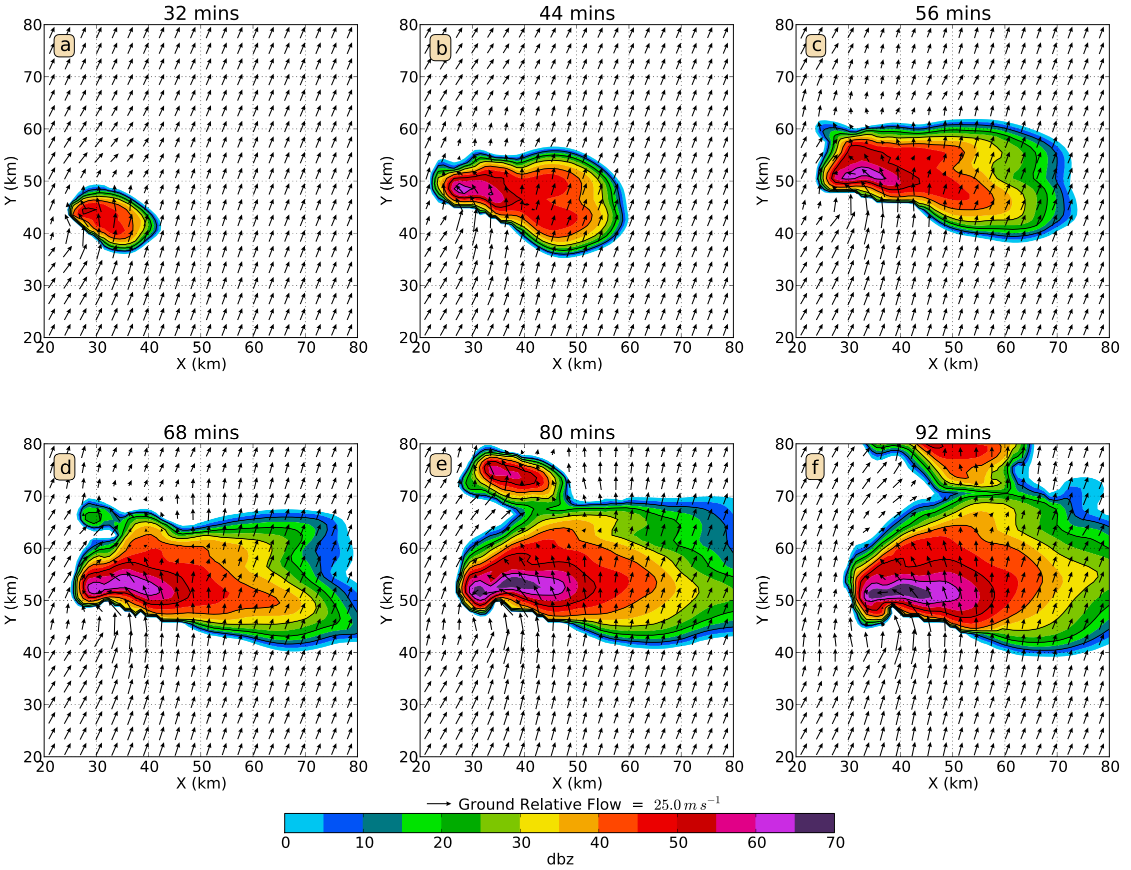


Figure 1. Reflectivity and ground relative wind vectors at 2.125 km AGL for the truth simulation every 12 minutes. The central (60 km)^2 of the domain is shown.

Synthetic Doppler velocity observations are generated from the truth simulation via a radar sampling observation operator from Wood et al. (2009) and Thompson et al. (2012). The synthetic observational resolution mimics a typical WSR-88D storm mode scan, with fourteen elevation angles ranging from 0.5° to 19.5°. One radar volume scan is generated every 4-minutes. In order to simulate observation non-simultaneity, model fields at progressively later times are used to generate observations at progressively higher elevation angles (Yussouf and Stensrud 2010). To simulate measurement errors, the observations are perturbed with Gaussian random errors having zero mean and 2 m s^{-1} standard deviation for Doppler velocity and 5 dBZ standard deviation for reflectivity.

The filter-assumed observational error variances are set to 4 m^2 s^{-2} for Doppler velocity and 25 dBZ^2 for reflectivity to match the random errors added to the synthetic observations. The simulated observation errors have no correlation. The observations are objectively analyzed to a 2 km quasi-horizontal grid on each radar sweep surface (Dowell et al. 2004) in order to thin the data to twice the model grid spacing. The synthetic Doppler velocity, positive reflectivity, and clear air reflectivity (zero reflectivity) observations are assimilated every 2-minutes. Therefore, two analysis cycles are needed to assimilate one complete radar volume and at worst, any individual observation is temporally displaced from the analysis time by one minute.

Forty-five ensemble members are used in each experiment. Ensemble members are initialized from the sounding used in the truth simulation but with random perturbations added to the horizontal wind components, temperature, and dew point temperature fields of each member. A 16-minute forecast is made prior to the first data assimilation to allow variation to develop in the ensemble background. The data assimilation begins at 32 minutes, when there are at least 500 radial velocity observations. To help maintain sufficient ensemble spread, the additive noise method of Dowell and Wicker (2009) is used. This technique adds Gaussian perturbations with standard deviations of 1.0 m s^{-1} or 1.5 K to the horizontal wind components, temperature and dew point temperature fields in regions where observed reflectivity > 20 dBZ. The perturbations are subsequently spatially smoothed. The model three-dimensional winds, potential temperature, and microphysical variables (water vapor, cloud liquid water, cloud ice, rain, snow, graupel, and hail) are updated during the assimilation.

3.3.2 OSSE Performance Evaluation

Two summary statistics are computed to compare the performance of the two filters throughout the data assimilation and forecast periods. The difference total energy (DTE) is evaluated because it includes the three-dimensional winds and potential temperature:

$$DTE = \frac{1}{2}\left\{(\delta u)^2 + (\delta v)^2 + (\delta w)^2 + \frac{C_\mathrm{p}}{T_\mathrm{r}}(\delta\theta)^2\right\}, \qquad (15.)$$

where δ denotes the difference between the ensemble mean and true values, $C_p = 1004.7$ J kg^{-1} K^{-1} is the specific heat of dry air at constant pressure, and $T_r = 270$ K is the reference temperature (Zhang et al. 2007; Wang et al. 2013). Wang et al. (2013) calculates DTE only at grid points where the true reflectivity > 10 dBZ. However, this excludes most of the main updraft in the supercell simulated in the present study. Thus, the mean DTE is computed by averaging grid points inside a verification domain extending from 20 to 80 km in the horizontal (shown in Figure 1) and from 0.125 (lowest model layer) to 9.82 km AGL in the vertical. The square root of the mean DTE (hereafter, "RM_DTE") is then computed.

The hydrometer difference total energy (HydroDTE)[2] is used to evaluate errors in the hydrometeor fields:

$$HydroDTE = \frac{1}{2}\left\{(\delta q_v)^2 + (\delta q_r)^2 + (\delta q_s)^2 + (\delta q_g)^2 + (\delta q_h)^2\right\}, \qquad (16.)$$

---

[2] The author is aware this quantity does not have units of energy. We choose to use the label "HydroDTE" to be consistent with previously published literature.

where q_v, q_r, q_s, q_g, and q_h are the mixing ratios of water vapor, rainwater, snow, graupel, and hail, respectively (Wang et al. 2013). The root-mean HydroDTE (RM_HydroDTE) is computed over the subset of the RM_DTE verification domain where true reflectivity > 10 dBZ. The reflectivity threshold accounts for the fact that the mixing ratios (with the exception of water vapor) are near zero outside the storm. RM_DTE and RM_HydroDTE statistics provide a more complete view of experimental analysis errors than comparing individual state variable (such as vertical velocity or rainwater mixing ratio) errors.

3.4 Localization sensitivity tests

3.4.1 Localization length comparisons

When comparing the EnSRF and LETKF convective-scale analyses, one issue to consider is the difference in the application of localization. As discussed previously, the EnSRF typically uses B-localization (applies localization to the background error covariance matrix; e.g., Snyder and Zhang 2003; Dowell and Wicker 2009) while the LETKF uses R-localization (applies the inverse of the localization function to the observation error variance; Hunt et al. 2007). For a given cutoff length (the distance at which the correlation function becomes zero), B-localization yields less observation influence at a given distance than does R-localization (Nerger et al. 2012; Miyoshi and Yamane 2007; HW2013). Increasing the B-localization cutoff length to be 25% larger than the R-localization cutoff length gives an observation less influence on nearer grid points but more influence on more distant grid points (see Fig. 1 in Holland 2011).

Since B- and R-localization are not equivalent, several values for the horizontal and vertical cutoff lengths are tested to optimize each assimilation methodology. Previous convective-scale radar data assimilation studies have used cutoff lengths around 6 km in both the horizontal and vertical directions (e.g. Dowell et al. 2004, see Table 1 in Sobash and Stensrud 2013). Sobash and Stensrud (2013) suggest that the vertical cutoff length should be shorter than the horizontal cutoff length in radar data assimilation due to the smaller vertical length scales of convective structures. The cutoff lengths in the sensitivity experiments that follow were therefore initially set to 6 km (3 km) in the horizontal (vertical) directions, then increased to 9 km (4.5 km) and 12 km (6 km) in the horizontal (vertical) directions. Analysis errors from each localization length test are examined to determine the sensitivity to localization length scales for the LETKF and EnSRF (Figure 2 and 3).

For both filters, the RM_DTE and RM_HydroDTE are only weakly sensitive to the localization cutoff lengths (Figure 2a and 3a). The EnSRF errors vary more than the LETKF errors, indicating the EnSRF is more sensitive to the chosen localization lengths. For both filters, increasing the localization cutoff length increases the size of the analyzed updraft core (region where $w > 10$ m s^{-1}; Figure 2). The reflectivity core (where reflectivity > 55 dBZ) also increases with localization cutoff length (Figure 3).

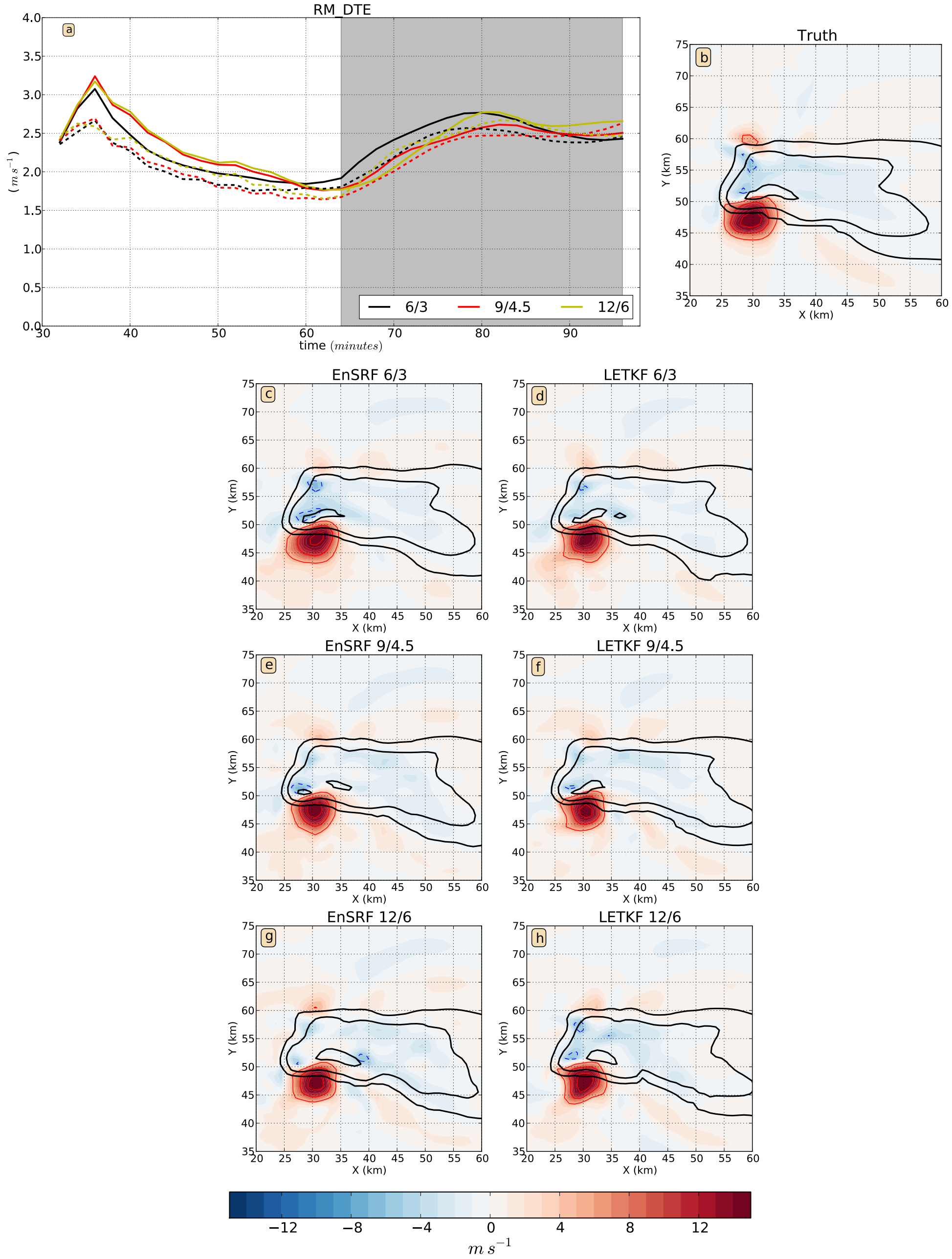


Figure 2. Progressively larger localization cutoff lengths in the horizontal/vertical are compared via RM_DTE for the mean EnSRF (solid) and LETKF (dashed) analyses and forecasts (shaded region) (a), and via cross-sections at 2.125 km AGL of vertical velocity (color filled contours every 1 m s^{-1}; red = positive, blue = negative) and reflectivity (black contours every 20 dBZ) for the mean EnSRF (c,e,g) and LETKF (d,f,h) analyses at 56 minutes (6.5 radar volumes have been assimilated). The horizontal/vertical cutoff in km, beyond which the covariance are zero, is listed in the title of each subplot. The truth cross-section is shown for reference (b).

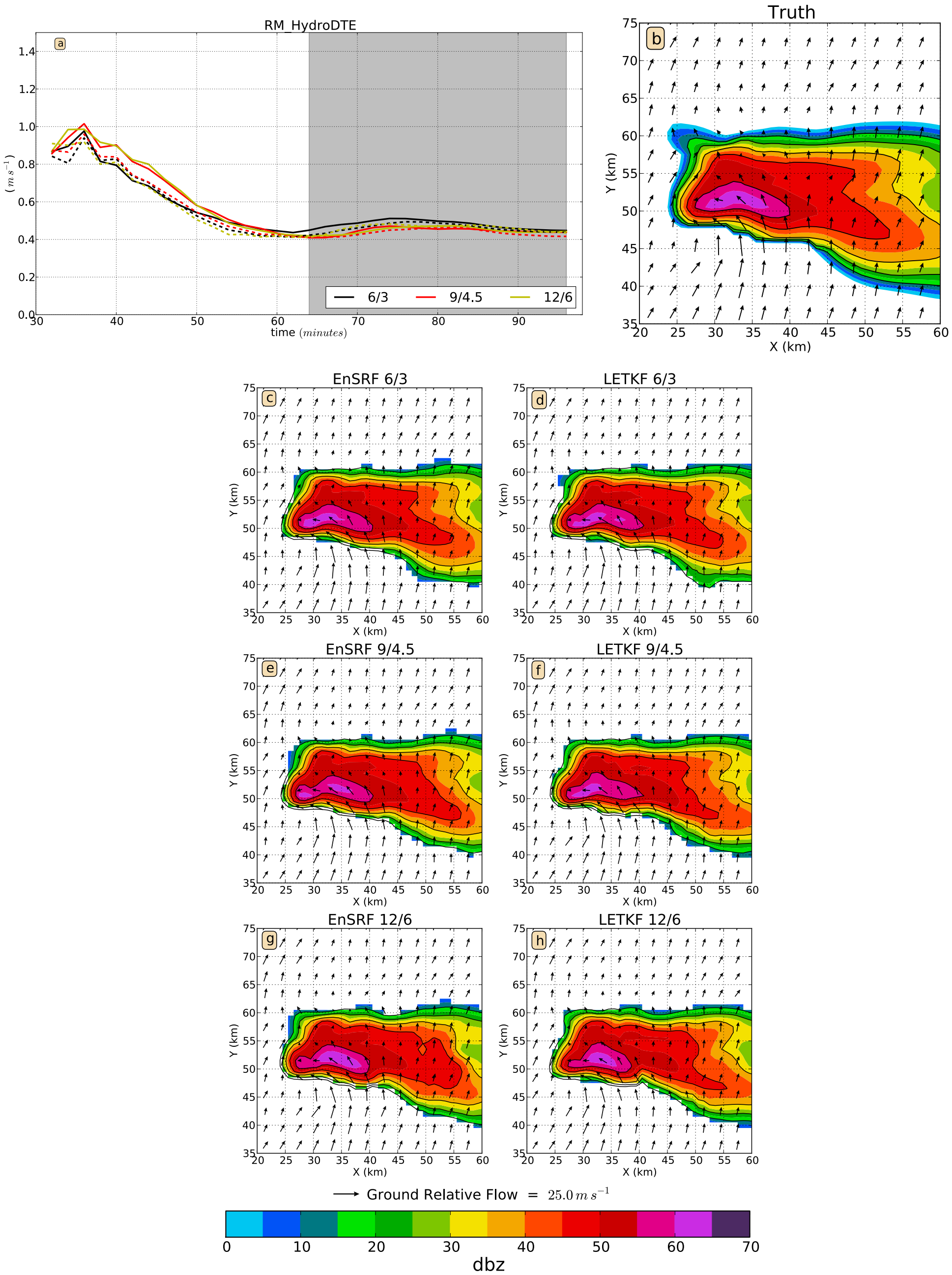


Figure 3. Progressively larger localization cutoff lengths in the horizontal/vertical are compared via RM_HydroDTE for the mean EnSRF (solid) and LETKF (dashed) analyses and forecasts (shaded region) (a), and via cross-sections at 2.125 km AGL of reflectivity (color filled contours every 5dBZ) and horizontal wind vectors for the mean EnSRF (c,e,g) and LETKF (d,f,h) analyses at 56 minutes (6.5 radar volumes have been assimilated). The horizontal/vertical cutoff in km, beyond which the covariance are zero, are listed in the title of each subplot. The truth cross-section is shown for reference (b).

The results in Figure 2 and 3 show that selecting a “best” localization cutoff length to use for filter comparisons is challenging. We chose the 12/6 km horizontal/vertical cutoff lengths for the EnSRF (i.e., B-localization) and the smaller, 9/4.5 km, lengths for the LETKF (i.e., R-localization) for the filter comparisons that will be discussed. The B-localization cutoff length is 25% longer since it yields less observation influence at a given distance than does R-localization (Nerger et al. 2012; Miyoshi and Yamane 2007; HW2013). A comparison using a 10.5/5.25 km horizontal/vertical cutoff for EnSRF (B-localization) and a smaller cutoff, 7.5/3.75 km, for LETKF (R-localization) yield similar relative filter performance. The ratio of the horizontal to vertical cutoff lengths was varied in additional experiments. When this ratio became too large (e.g., a 12 km horizontal cutoff with a 4.5 km vertical cutoff), the development of the secondary storm on the left flank of the main supercell (Figure 1e) was inhibited (not shown). The tuning experiments required to obtain an appropriate cutoff motivates the use of adaptive localization techniques as suggested by Sobash and Stensrud (2013).

3.4.2 Consistency ratio comparisons

Localization affects ensemble spread, which affects analysis and forecast metrics. Doppler velocity innovation statistics are computed to compare the ensemble spread in the experiments. To the extent that the Kalman filter assumptions are satisfied (e.g., Gaussian error distributions), the consistency ratio (CR) approximates the ratio of the actual to optimal ensemble forecast spread (variances), therefore approaching unity if the ensemble represents the forecast error statistics accurately and the observation error variances are correctly specified in the filter (Aksoy et al. 2009). CR values < 1 may indicate insufficient ensemble spread in the analysis. The CR is computed for the priors over the domain where reflectivity > 15 dBZ in order to isolate the measurement of performance to the main convective regions (Snyder and Zhang 2003, Tong and Xue 2005, Aksoy et al. 2009). Although the radar data are assimilated every 2 minutes, the statistics are aggregated over 4 minutes so that each time bin contains one complete radar data volume, similar to Aksoy et al. (2009).

Producing ensemble spread that results in comparable consistency ratios among different experiments is a challenge. Both EnSRF and LETKF appear underdispersive near the start of the analysis period (Figure 4). Suboptimal spread is a common problem in convective scale radar data assimilation (Aksoy et al. 2009). The spread increases through most of the analysis period. The reason for this trend is unclear, but may be related to the increasing areal expansion of reflectivity as the supercell matures, which results in additive noise being applied over a larger area. Less localization (longer cutoff length) leads to smaller CR values indicating less ensemble spread. This result is expected because less localization allows observations to influence more grid points (decreases spread) and the additive noise used to maintain spread was not varied. The ensemble spread during the beginning of the analysis period is most deficient for the longer cutoff lengths. This may contribute to the slightly increased analysis errors for the longer cutoff lengths shown in Figures 2a and 3a. The large CR values for the 6/3 km horizontal/vertical cutoff length experiments during the later analysis cycles may contribute to the increase in errors during the forecast (Figure 3.2a and 3.3a). The

CRs for the chosen EnSRF experiment (yellow solid line in Figure 4) and the chosen LETKF experiment (red dashed line in Figure 4) follow the same trend and are reasonably close in magnitude. Since the differences between the EnSRF and LETKF CRs are small and the values of CR are similar to those of previous work (Aksoy et al. 2009, Dowell et al. 2011), we conclude that the experiments are sufficiently tuned for the present comparisons.

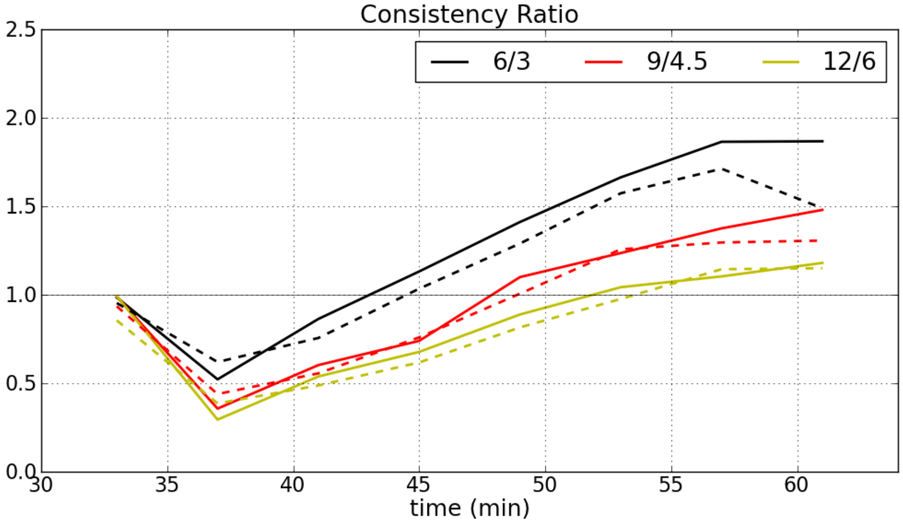


Figure 4. Progressively larger localization cutoff lengths in the horizontal/vertical (6/3: black, 9/4.5: red, 12/6: yellow) are compared via consistency ratio for radial velocity during the assimilation period for the EnSRF (solid) and LETKF (dashed) experiments.

3.5 EnSRF and LETKF OSSE comparisons

The EnSRF and LETKF are compared for the convective-scale assimilation of synthetic radar observations. The data assimilation begins at 32 minutes, while the storm is developing (Figure 1a). Eight volumes of radar data are assimilated and the final analysis time is at 62 minutes. The final analyses initialize ensemble forecasts that are run out until 96 minutes.

3.5.1 Storm analysis comparisons

The mean analyses for both the EnSRF and LETKF experiments contain the same storm features as the truth simulation. After three and a half radar volumes are assimilated, at 44 minutes, the vertical velocity patterns in the EnSRF and LETKF analyses are similar to each other (Figure 5b,c). The EnSRF analysis has a stronger updraft that is closer in magnitude to the truth than does the LETKF analysis. However, the EnSRF analysis also contains stronger downdrafts, than are present in the truth simulation. Both the EnSRF and LETKF analyses have overly strong upward motion on the left (northern) flank of the storm compared to the truth simulation. The EnSRF updraft is also slightly larger in areal extent than the LETKF updraft; both are slightly larger than the truth's updraft. The larger updraft areas are in part due to the averaging of ensemble member position errors in the ensemble means. At 68 minutes, eight volumes have been assimilated followed by a 6-minute forecast. The mean EnSRF and LETKF analyses are still very similar to each other and they capture the vertical velocity pattern in the truth (Figure 5d,e,f). The LETKF forecast's main updraft is closer to the truth, while the EnSRF captures better the small left-split cell (near x = 30 km, y = 68 km). Overall, both algorithms have the main features of the true supercell. This suggests that the LETKF is roughly as effective as the EnSRF for assimilating radar data at convective-scales.

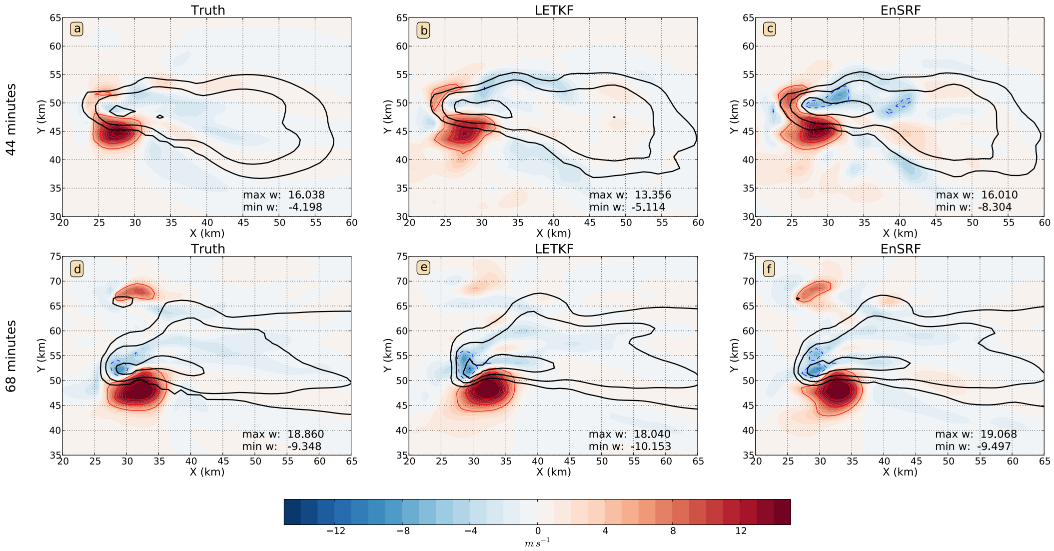


Figure 5. . Cross-sections at 2.125 km AGL of vertical velocity (color filled contours every 1 m s^{-1}; red = positive, blue = negative) and reflectivity (black contours every 20 dBZ) for the truth simulation (a), mean LETKF analysis (b) and mean EnSRF analysis (c) at 44 minutes (3.5 radar volumes have been assimilated). Cross-sections are also shown for the truth simulation (d), mean of LETKF ensemble forecasts (e) and mean of EnSRF ensemble forecasts (f) at 68 minutes (8 radar volumes have been assimilated followed by a 6 minute forecast). The maximum and minimum vertical velocity values (m s^{-1}) are printed in the lower right corner of each panel.

3.5.2 Surface pressure tendency comparisons

The surface pressure tendency is often used to measure the noise or dynamical imbalance in the model state. For example, Huang et al. (1994) used the mean absolute tendency of surface pressure as a global measure of high-frequency noise. Similarly, the surface pressure tendency in a non-hydrostatic model can be used as a measure of analysis balance (Reich et al. 2011). The magnitude of the surface Exner function time tendency (hereafter, DPDT) can also be used as a measure of noise (e.g., Wang et al. 2009). In the truth simulation, the maximum DPDT associated with the supercell storm remains below 10^{-7} sec^{-1} (not shown). The ensemble mean DPDT rapidly increases to 25-35 x 10^{-6} sec^{-1} after data assimilation begins in the EnSRF and LETKF experiments (Figure 6). The DPDT from the LETKF experiment is 10-20% smaller than the EnSRF DPDT. This suggests that the column divergence in the wind field analyses is more in balance with the forecast pressure gradients (the pressure field is not updated in our experiments) in the LETKF experiment, inducing smaller responses in the surface pressure field when the ensemble is integrated forward. That is, the LETKF increments appear to better preserve the mass balance in the forecast states than do the EnSRF increments. These results agree with Bowler et al. (2012), who found that for a simple model and a nonlinear observation operator, a simultaneous update performs better than a sequential update. The results are also consistent with those obtained in HW2013 using a large-scale primitive equation two-layer model. Interestingly, the DPDT in the LETKF forecasts remains slightly smaller than in the EnSRF forecasts through the end of the forecast period.

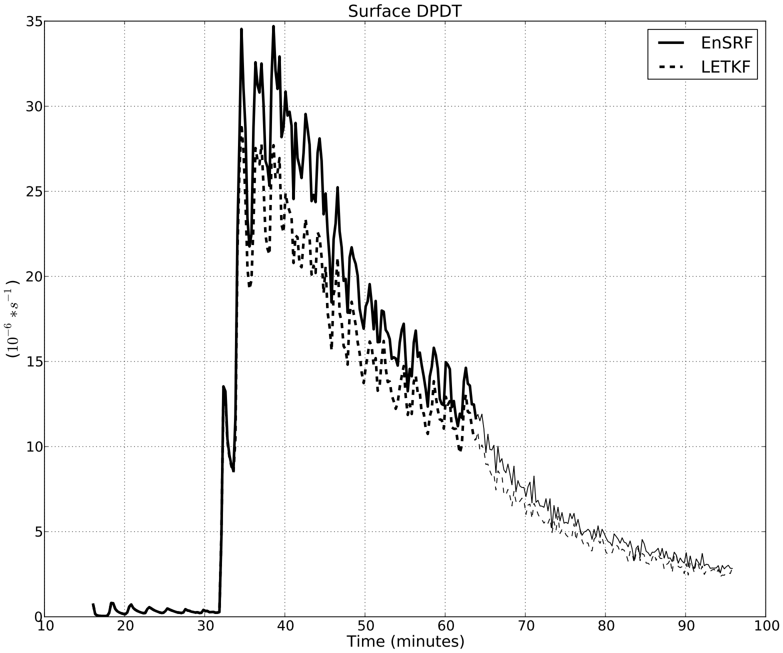


Figure 6. Change in surface perturbation Exner function every third model time step (15 seconds) throughout the EnSRF (solid) and LETKF (dashed) experiments. Thick (thin) lines indicate the statistics are valid for a mean ensemble analysis (mean ensemble forecast).

3.5.3 Analysis and forecast errors comparisons

To quantitatively compare the filters, the RM_DTE and RM_hydroDTE are computed for the analyses and forecasts. An example horizontal cross-section of DTE for each experiment at 56 minutes is shown in Figure 7b,c. The LETKF analysis has the largest maximum DTE, 46.9 $m^2 s^{-2}$. The EnSRF analysis has a larger number of grid points with DTE values greater than a given threshold. For example, EnSRF has 529 grid points with DTE greater than 4 $m^2 s^{-2}$, compared to 458 grid points for LETKF. There are spatial differences in the errors between the two filters (Figure 7c). However, in general the errors are of similar magnitude and occur over approximately the same amount of grid points.

The LETKF analysis RM_DTE are consistently less than in the EnSRF analyses (Figure 7a). The RM_DTE differences between the filters are larger early in the assimilation periods. This implies that the LETKF analysis increments initially improve the ensemble's three-dimensional winds and potential temperature faster than the EnSRF increments do. Thereafter, the LETKF errors decrease at roughly the same rate as the EnSRF errors, and therefore remain substantially lower through most of the assimilation period, indicating the LETKF ensemble mean is closer to the truth. Smaller LETKF errors are also observed in root-mean-square errors for individual state variables (three-dimensional winds and temperature; not shown). Both filters have similar RM_DTE values during the forecast period (63-96 minutes). The LETKF and EnSRF have similar RM_HydroDTE values during the analysis and forecast, indicating that both filters appear to generate hydrometeor fields of similar accuracy (Figure 7a).

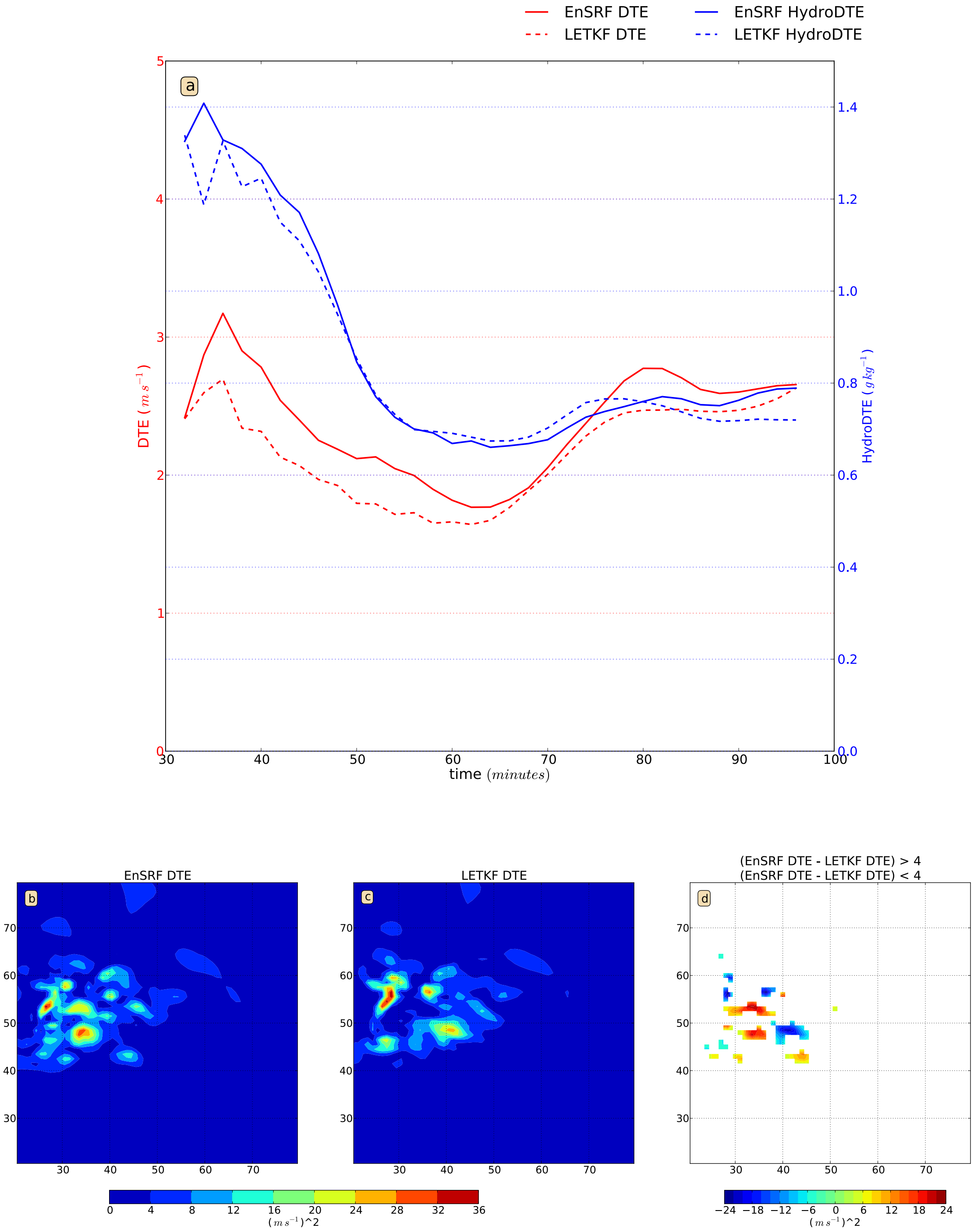


Figure 7. (a) RM_DTE (reds) and and RM_HydroDTE (blues) for the EnSRF (solid) and LETKF (dashed) experiments. Example cross-section of DTE at 2.125 km AGL at 56 minutes for (b) the EnSRF analysis, (c) the LETKF analysis, and (d) differences between the two analyses > 4 or < -4 $m^2\ s^{-2}$.

3.6 Localization method comparisons

We largely attributed the differences between the EnSRF and LETKF performance in the previous section (3.5) to the difference in how observations are processed (sequentially versus simultaneously). To test this hypothesis, we now eliminate the difference in localization formulations between the EnSRF (B-localization) and LETKF (R-localization) by implementing the R-localization in the EnSRF code. Note that the B-localization cannot be simply implemented in the LETKF because the background error covariance is not explicitly computed (Nerger et al. 2012). R-localization is implemented in the EnSRF by removing the correlation function, $\boldsymbol{\rho}_{\mathbf{B}}$ from Equation (8.) and elementwise-multiplying the observation error variance, **R**, by the correlation function $\boldsymbol{\rho}_{\mathbf{R}}$ in Equation (8.) and Equation (10.), as is done in the LETKF in Equation (12.) and Equation (14.).

Three assimilation experiments are used to test the hypothesis that the localization method is not the primary reason for the differences in filter performance. The results from the EnSRF with B-localization (12/6 km horizontal/vertical cutoff length, hereafter referred to as “EnSRF-B”) and the EnSRF with R-localization (9/4.5 km horizontal/vertical cutoff length, hereafter referred to as “EnSRF-R”) show almost no difference in analysis error (Figure 8). The control LETKF (9/4.5 km horizontal/vertical cutoff length) is shown for comparison. Both EnSRF experiments have larger errors than the LETKF experiment. Experiments using other localization cutoff lengths produce qualitatively similar results (not shown). Thus, the differences in error between EnSRF and LETKF are not primarily caused by the difference in localization. This conclusion combined with the previous results indicates that the simultaneous update in LETKF has a small positive impact on the convective-scale analysis errors.

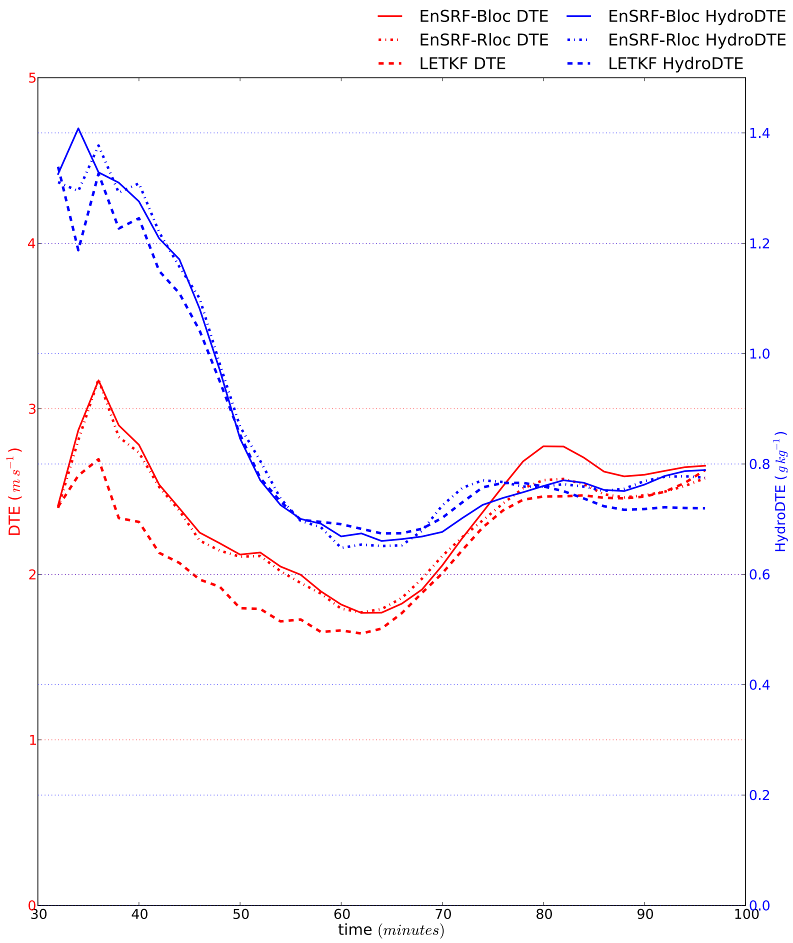


Figure 8. RM_DTE (reds, left axis) and RM_HydroDTE (blues, right axis) for the EnSRF with B localization (solid), EnSRF with R localization (dash-dot) and LETKF (dashed) experiments.

3.7 Observation type

In the results discussed thus far, the simultaneous update in LETKF produces slightly improved error performance relative to the EnSRF. To further examine the filter performance differences, the EnSRF and LETKF experiments are performed with only Doppler velocity observations assimilated. These experiments are motivated by the fact that EnKF methods assume observations are linearly related to the model state, a condition that is approximately satisfied in the case of Doppler velocity but not for reflectivity. The nonlinearity of the reflectivity observation operator makes it challenging to optimize reflectivity assimilation (Dowell et al. 2011). Due to the complexities involved with reflectivity assimilation, several EnKF studies have assimilated only Doppler velocities (e.g., Snyder and Zhang 2003, Zhang et al. 2004). The nonlinearity in the reflectivity observation operator could enhance the analysis differences that arise due to the different assimilation patterns of the two filters (recall that the LETKF uses the reflectivity priors only from the forecast state, while the EnSRF re-computes an updated reflectivity prior directly after each observation update).

Verification statistics for the velocity-only assimilation experiments are shown in Figure 9. Unlike in the previous experiments (Figure 8), the LETKF does not outperform the EnSRF when only Doppler velocity is assimilated. Rather, both assimilation methods produce analyses of similar accuracy. We hypothesize that LETKF's simultaneous update mitigates some of the approximation errors associated with the nonlinear observation operator for radar reflectivity. This would explain the lower errors of the LETKF relative to the EnSRF in the experiments that assimilate reflectivity and the similar performance of the LETKF and EnSRF in the experiments that only assimilate Doppler velocity.

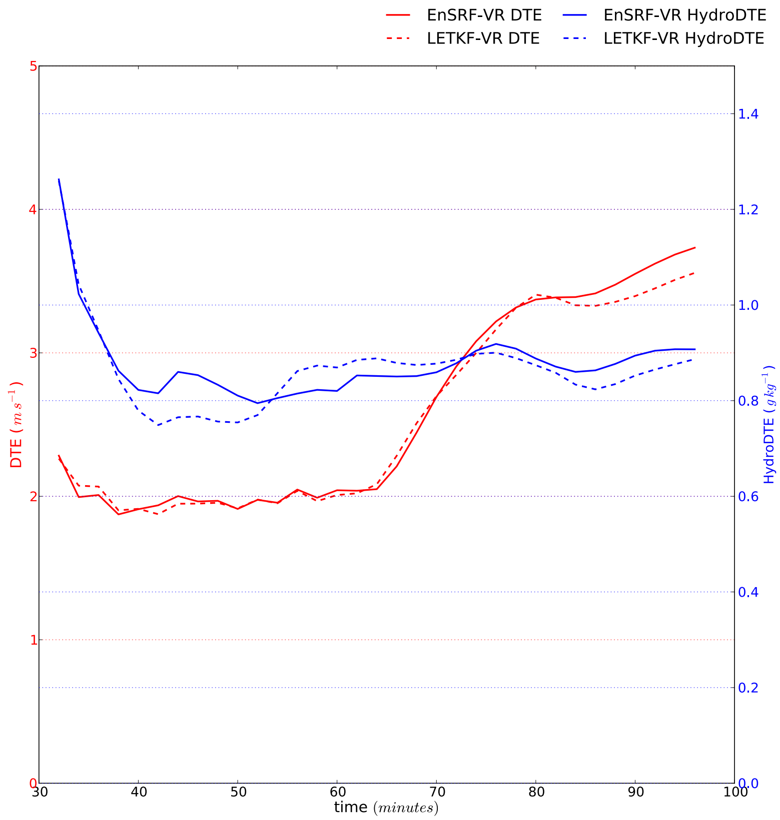


Figure 9. RM_DTE (reds, left axis) and RM_HydroDTE (blues, right axis) for the EnSRF (solid) and LETKF (dashed) experiments that assimilate only radial velocity observations.

During the first four radar volume assimilations (32 to 50 minutes), the velocity-only experiments have smaller errors than the experiments that assimilate all data (Figure 8, 9). However, examination of horizontal cross-sections of model winds and hydrometeor variables during this time show that the default EnSRF experiments develop convection more quickly than, and qualitatively improve upon, the velocity-only experiments. The velocity-only experiments analyses contain smaller mixing ratios for the hydrometeor variables indicating slower storm development. During the rest of the assimilation period (52 to 90 minutes), the RM_DTE and RM_HydroDTE indicate the reflectivity assimilation improves the analyses, particularly of the hydrometeor state variables.

3.8 Real data experiment

3.8.1 design

On 8 May 2003 a destructive tornadic storm passed over central Oklahoma, producing damage rated at category 4 on the Fujita scale (F4) in the town of Moore. This storm has been studied extensively (Burgess 2004; Hu and Xue 2007; Romine et al. 2008; Dowell and Wicker 2009; Dowell et al. 2011; Yussouf et al. 2013). Storms initiated along the dryline around 2050 UTC, and by 2130 UTC a dominant cell matured into a supercell. The supercell intensified as it moved to the northeast. A violent tornado developed around 2210 UTC and traveled approximately 30 km before dissipating at 2238 UTC. The experimental polarimetric WSR-88D KOUN radar documented the life cycle of the supercell. The data were edited manually to remove ground clutter, range folding, and spurious data and to unfold aliased Doppler velocities (Dowell et al. 2011). The data were then objectively analyzed to a 2 km quasi-horizontal grid and assimilated every 2 minutes from 2100 UTC to 2200 UTC. This helps to decrease observational error correlations caused by the relatively coarse radar effective beamwidth. Any observation error correlations that exist in the assimilated observations are ignored, which allows the **R** matrix to remain diagonal. The EnSRF and LETKF filters, assumed observation variance, and additive noise technique are the same as in the OSSEs. The NCOMMAS model is used with 1 km horizontal grid spacing over a 100 km horizontal domain (as in the OSSEs). The vertical grid has 400 m spacing near the surface and stretches to 700 m near the domain top, 20 km AGL. The domain moves to match the storm motion, which is 14 m s^{-1} towards the east and 8 m s^{-1} towards the north. The initial environment is derived from the Norman, Oklahoma sounding on 0000 UTC 9 May 2003. The double moment ZVD microphysics scheme and forty-five ensemble members are used, as in the OSSEs.

3.8.2 Storm analysis comparisons

The EnSRF and LETKF filters are compared for the convective-scale assimilation of KOUN radar observations over 1 hour and for a subsequent 15-minute forecast. The mean analyses for both the EnSRF and LETKF experiments contain the supercell storm features expected within the May 8th storm. The vertical velocity analyses have a strong, comma-shaped updraft at the end of the analysis period (Figure 10a,b). The differences between the analyses are limited to a small area (Figure 10c). There are some differences within the main updraft region. For example, the updraft in the LETKF analysis extends slightly farther to the northeast along the forward flank gust front. Also, the EnSRF updraft has higher values in the comma head, despite having a lower maximum value.

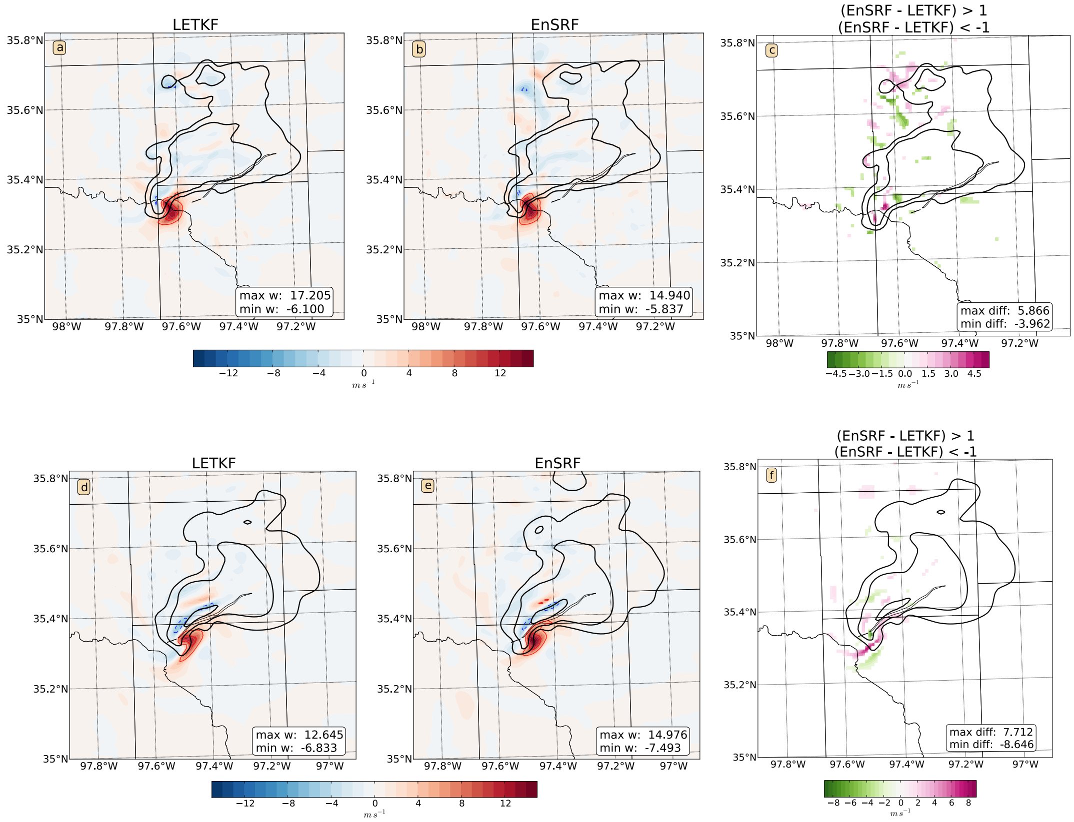


Figure 10. Cross-sections for the 8 May 2003 real-data case at 1.0 km AGL of vertical velocity (color filled contours every 1 m s^{-1}; red = positive, blue = negative) and reflectivity (black contours every 20 dBZ) for the mean LETKF analysis (a) and mean EnSRF analysis (b) at 2200 UTC (after 1 hour of data assimilation). Difference between (a) and (b) that are greater than 1.0 m s^{-1} and less than -1.0 m s^{-1} are shown in (c) along with the reflectivity from the mean LETKF analysis at 2200 UTC. The bottom row of panels (d-f) is as in the top row except for 15-minute forecasts valid at 2215 UTC. The maximum and minimum vertical velocity values or difference values (m s^{-1}) are printed in the lower right corner of each panel. Overlain in each panel is the NWS-observed tornado damage track that starts at 2210 UTC and ends at 2238 UTC.

The means of the ensemble forecasts for both filters exhibit a common problem with forecasts of the May 8th supercell; the storm decays too rapidly and is shifted to the southeast compared to the observations (Figure 10d,e). The portion of the updraft that extends along the rear flank gust front is farther south in the LETKF forecast than in the EnSRF forecast. The vertical velocity maximum is larger in the EnSRF forecast indicating that the LETKF forecast has decayed faster (since it’s maximum was larger in the last analysis). Overall, however, the 15-minute forecasts initialized from the two filters are quite similar. In longer forecasts (up to an hour), initialized from the EnSRF and the LETKF, the storm decays at a similar rate.

The OSSE surface pressure tendency comparisons indicated that the LETKF analyses contain less noise than the EnSRF analyses (Figure 6). The 8 May 2003 EnSRF and LETKF analyses, however, contain roughly the same amount of noise (Figure 11), indicating that the LETKF simultaneous update does not substantially decrease the dynamical balance disturbance in this real-data case. It is possible that any small improvement from the simultaneous update is undetectable given the relatively large values of DPDT (double the OSSE values), which are likely caused by the larger and/or more correlated errors in the real observations being assimilated.

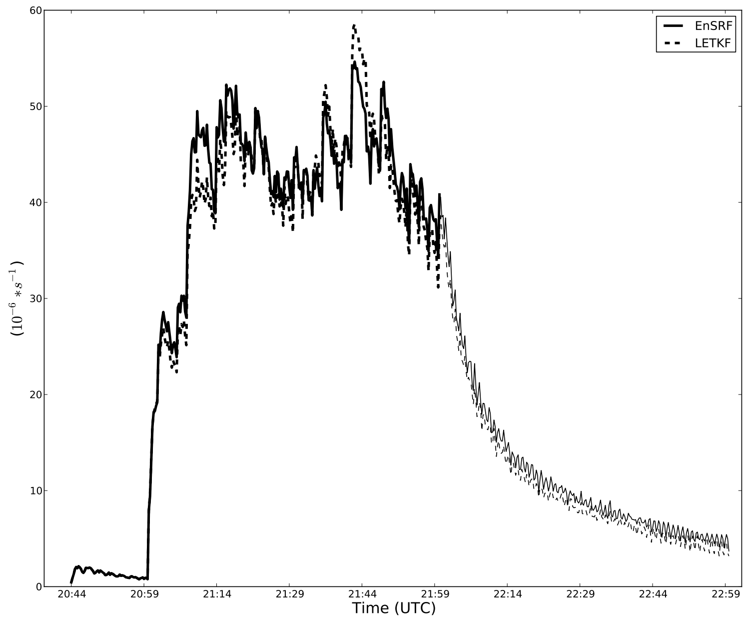


Figure 11. Change in surface perturbation Exner function every third model time step (15 seconds) throughout the EnSRF (solid) and LETKF (dashed) real-data 8 May 2003 experiments. Thick (thin) lines indicate the statistics are valid for a mean ensemble analysis (mean ensemble forecast).

3.8.3 Ensemble probabilistic forecasts of low-level vorticity

Forecasts for the real data experiments can be evaluated by examining the predicted tornadic potential compared to the observed tornado track. Since the 1 km horizontal grid spacing used in the May 8th experiments is too coarse to explicitly resolve a tornadic circulation a proxy must be used. The presence of significant low-level rotation (vorticity) can be used as a proxy for tornadic potential because it is indicative of a significant mesocyclone (Stensrud and Gao 2010; Dawson et al. 2012; Stensrud et al. 2013; Yussouf et al. 2013). The vorticity can be used as metric for the location of rotation, however it may not always correlate well to tornado strength or indicate whether a tornado exists (Markowski et al. 2011, Marquis et al. 2012). Forecast probabilities (i.e. ensemble frequencies) of vorticity exceeding a threshold value of 0.01 s^{-1} are calculated at each grid point during the period 2200-2230 UTC (Figure 12a-f). Recall there are 45 ensemble members in each experiment. There are small differences between the EnSRF and LETKF high probability vorticity swath locations. To explore the importance of these differences, the EnSRF and LETKF experiments are repeated twice using different random number seeds in the ensemble member initialization. The differences between the EnSRF and LETKF vorticity swaths in the original experiments (Figure 12a,b) are no larger than the differences among the EnSRF or LETKF vorticity swaths obtained by varying the random number seeds (Figure 12a,c). When the ensemble members from the three seed experiments are combined, in each case (EnSRF and LETKF), the ensemble probabilities can be calculated based on 135 members. The EnSRF and LETKF combined-ensemble vorticity swaths are very similar for low to moderate probabilities, but differences increase for higher probabilities. For example, for probabilities > 90%, the LETKF combined-ensemble vorticity swath covers a smaller area and is less continuous than the EnSRF combined-ensemble swath. Compared to the observed tornado track, all of the forecast vorticity swaths are shifted to the southeast and are not long enough. Thus, the forecasts initialized from the EnSRF and the LETKF suffer from the same model bias. Overall, both sets of forecasts are similar.

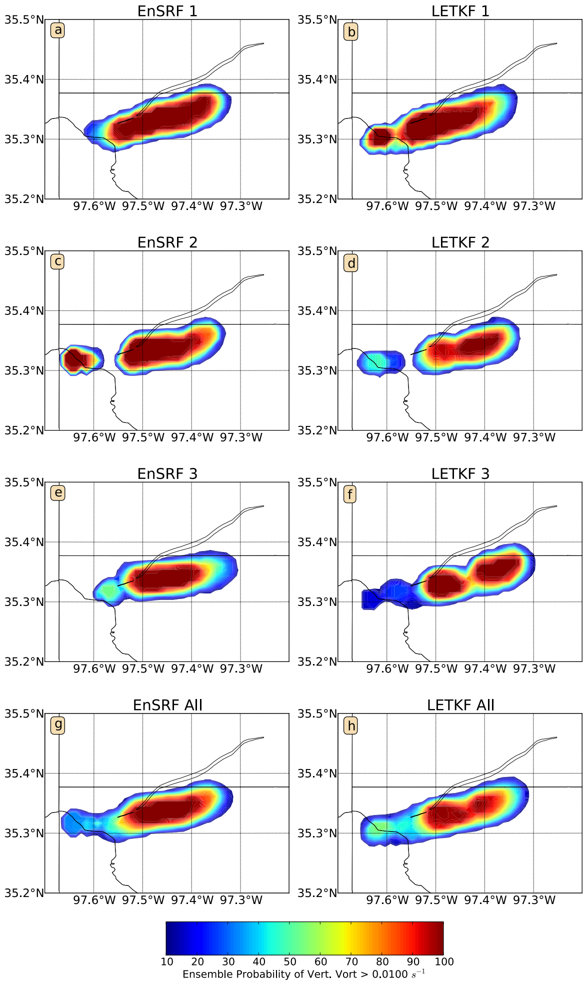


Figure 12. 8 May 2003 Ensemble probability of vorticity exceeding 0.01 s^{-1} at 1 km AGL during a 30 minute forecast period starting from the analysis at 2200 UTC and ending at 2230 UTC for the EnSRF (a,c,e) and LETKF (b,d,f) with different random number seeds. The combined-ensemble probabilities for the EnSRF (g) and LETKF (h). Overlain in each panel is the NWS-observed tornado damage track (black outline) that is from 2210 to 2238 UTC.

3.9 Conclusions and discussion

This chapter applies the LETKF to convective-scale radar data assimilation and compares it to the EnSRF, which has been the primary algorithm for many previous convective-scale radar data assimilation studies. The LETKF assimilates observations simultaneously while EnSRF uses a sequential assimilation pattern. Comparisons are first performed in an OSSE framework where synthetic WSR-88D radar observations are generated from a supercell thunderstorm ("truth") simulation and assimilated every 2-minutes using forty-five-member ensembles.

To facilitate comparisons, the filters are tuned to produce similar observation space statistics. In both filters, the choice of localization cutoff length impacts the wind and temperature analyses, and has slightly smaller impact on hydrometeor state variables. The EnSRF is slightly more sensitive than the LETKF to the localization cutoff length. However the localization method and choice of cutoff length do not prevent meaningful comparisons between the filters and do not dominate the differences between the filters. For both filters, the localization length sensitivity tests indicate a preference for longer cutoff lengths than the values commonly used in the literature for convective-scale radar data assimilation. This result agrees with Sobash and Stensrud (2013). Once the differences in the effective localization length of the B- and R-localization functions are accounted for (by using a smaller cutoff length in the latter), the performance of the EnSRF with localization applied to the background error covariance matrix (B-localization) is nearly the same as that with localization applied to the observation error matrix (R-localization). Both EnSRF methods produce larger analysis errors than the LETKF.

The LETKF produces less noise and somewhat smaller analysis errors than the EnSRF in the OSSE. The slightly improved performance is likely due to the simultaneous observation assimilation used in the LETKF as opposed to the sequential assimilation in the EnSRF. When only Doppler velocity observations are assimilated, however, the filters produce very similar errors. This indicates that analysis differences between the assimilation methodology occurs when the observation operator is nonlinear, as with radar reflectivity. This result agrees with conclusions from a simple model study (Bowler et al. 2012) and merits further investigation.

Experiments with real radar observations from 8 May 2003 help to further elucidate the similarities and differences between the filters. The EnSRF and LETKF analyses and forecasts have some small differences, however, both are able to capture the supercell. Changing the random number seed value used to initialize the perturbations that are added to each ensemble member reveals that the analysis and forecast differences between the EnSRF and the LETKF are similar in magnitude to the differences that arise from the sampling variability associated with a finite ensemble. Therefore, the difference in accuracy between the filters does not appear to be of great practical importance. The overall results strongly suggest the LETKF is an acceptable alternative to the EnSRF for convective-scale radar data assimilation.

Convective-scale numerical weather prediction may require very high-resolution grids (e.g., Dx ~ 0.2 - 1 km; Bryan et al. 2003) and therefore, very large computational resources (e.g., 10^5-10^6 cpu cores). Since the EnSRF and the LETKF show similar assimilation performance, future work can investigate the differences in scalability between the filters. The ability to interpolate the analysis perturbation weights from the LETKF analysis (Yang et al. 2009) may increase the efficiency for high-resolution grids. The application of weight interpolation to convective-scale data assimilation will be investigated in the future. Additionally, more work is needed in the future to determine if weight interpolation is also possible for the EnSRF. Further investigation is also needed to explore how the results of this study generalize to other convective modes.

Chapter 4: Multi-Scale Data Assimilation of the 13 June 2010 Tornadic Supercell Storm Environment during VORTEX2

4.1 Introduction

4.1.1 Overview

A framework for multi-scale EnKF data assimilation is developed to enable accurate analysis of both storms and their parent environment. Herein, multi-scale data assimilation is defined as a method that combines radar data assimilation with simultaneous assimilation of conventional observations, analyzing the primary synoptic and mesoscale convective forcing features and the convection itself. The development of multi-scale data assimilation techniques, with the goal of creating an accurate storm environment is discussed in this chapter. Further, the potential of this approach to improve ensemble convective forecasts is assessed, and the hypothesis that starting from an accurate storm environment will improve forecasts of severe convective storms is also tested.

Creating analyses and forecasts of a complex convective environment, with boundary interactions playing a significant role, represents a forecast challenge particularly well suited to the multi-scale approach and the Warn-on-Forecast mission (discussed in Chapter 1). The severe convective storms event on 13 June 2010, included the interaction of a storm with an outflow boundary and stationary front. Thus, 13 June is used as an example case study for multi-scale data assimilation development. Section 4.2 reviews the meteorological events that occurred on 13 June 2010.

Using EnKF to assimilate multi-scale observations, with the goal of improving both the storm environment and storm forecasts, has not been well documented in the literature (see Section 4.1.2). The experiments in this chapter are designed to address some of the outstanding research questions regarding the generation of an accurate storm environment. These questions are provided below and the motivation for these foci is discussed in the following section (4.1.2).

- Is data assimilation at convective-allowing model (CAM) resolution required to obtain an accurate storm environment?
- How does cycling frequency impact the near-storm environment and convective forecast?
- Can infrequent (hourly) assimilation of radar observations improve the near-storm environment and convective forecast?
- Is multi-scale data assimilation cycling sensitive to the background ensemble used for initialization?
- What are the challenges and limitations for producing analyses/forecasts of a complex convective event?
- Does improving the storm environment via multi-scale data assimilation result in improved convective forecasts? How much improvement and for how long are these improvements realized?

The EnKF data assimilation software, forecast model, and observation processing used in the data assimilation and forecast system are described in Section 4.3. All of the experiments are overviewed in Section 4.4. Mesoscale data assimilation and forecast experiments are discussed in Section 4.5. The results of the multi-scale data assimilation and forecast experiments are presented in Sections 4.6-4.8. Section 4.6 discusses the frequency of data assimilation cycles. Section 4.7 focuses on the impact of radar observations. Section 4.8 examines the development of and subsequent forecast sensitivity to the initial background ensemble. Lastly, a summary and discussion are presented in Section 4.9.

4.1.2 Previous work and motivation

This study will leverage previous findings on mesoscale (e.g. Fujita et al. 2007; Torn and Hakim 2008; Romine et al. 2013) and storm-scale (e.g. Aksoy et al. 2009, 2010; Dowell and Wicker 2009) ensemble-based data assimilation in a combined approach to achieve a multi-scale system. This study will also build on previous efforts to produce convective storm forecasts with fully complex heterogeneous environments (e.g., Lei et al. 2009; Stensrud and Gao 2010; Dowell et al. 2010; Yussouf et al. 2013). The advantages and limitations of the existing mesoscale and storm-scale assimilation techniques are discussed below to motivate the need for a multi-scale analysis system and the associated the sensitivity experiments used to develop it.

Current operational analysis systems use mesoscale grid spacing (O(15 km) horizontal grid spacing), which is often downscaled to convective-allowing model resolution (CAM; O(3 km) horizontal grid spacing) for real-time convective forecast applications (e.g. WRF-NSSL, Kain et al. 2010; AFWA ensemble, Hacker et al. 2011;

HRRR, Alexander et al. 2010). Several previous case studies have shown that CAMs improve forecasts compared to the mesoscale, cumulus-parameterizing resolution (e.g. Clark et al. 2009; see Section 2.1.3). Experimental forecast systems developed at the National Center for Atmospheric Research (NCAR) (e.g., Weisman et al. 2008) indicated sensitivity of convective forecasts to the initial state drawn from external analysis systems, motivating further investigation of data assimilation systems used to provide initial conditions. Romine et al. (2013) implemented a real-time continuously cycled assimilation system that generated mesoscale analyses to initialize deterministic CAM forecasts. They found that skill in the forecasts is limited by systematic bias in the initial conditions. However, Romine et al. (2013) also found that short-term forecasts of convection benefitted from a data assimilation system that used the same model system as the forecast model. Further, results from Schwartz et al. (2014) indicate ensemble forecasts initialized from a mesoscale EnKF system are skillful, and have a minimal ‘spinup’ relative to forecasts initialized from external analysis systems. These results motivate building a data assimilation system and forecast system that share the same modeling framework. This study will expand on these findings by using a shared assimilation and modeling framework. The first experiment in this study mimics the system design of Romine et al. (2013) and Schwartz et al. (2014) to investigate the skill of a mesoscale analysis, and downscaled CAM ensemble forecasts (Section 4.5).

The alternative approach to initializing convective-scale forecasts, uses convective-scale EnKF data assimilation that focuses on assimilating Doppler radar observations using horizontally homogeneous and temporally constant environmental conditions (e.g. Dowell et al. 2004a; Tong and Xue 2005; Aksoy et al. 2009, 2010; Yussouf and Stensrud 2010; Dowell et al. 2011; Dawson et al. 2012). Many of theses previous studies have shown that the initial conditions and subsequent forecast of convection can be substantially improved with ensemble data assimilation methods using Doppler radar observations. Radar data assimilation enhances the initial state by adding or removing convection to better match observations and also by correcting errors in the representation of convection in the analysis. Radar observations can provide information about the parent convective environment, though techniques to extract this have yet to be fully developed. This work will expand on these findings by merging the assimilation of radar observations in a CAM with the assimilation of conventional observations.

Representing mesoscale environmental uncertainty is important for maintaining ensemble spread and improving fit to the observations in an analysis system (Fujita et al. 2007; Aksoy et al. 2009). Yet, only a few radar data assimilation studies have attempted to incorporate more realistic representation of mesoscale uncertainties, through the use of a heterogeneous environment. For example, Stensrud and Gao (2010) found that a more realistic inhomogeneous mesoscale environment led to substantial improvement in forecast accuracy compared to using a homogeneous, single-sounding environment in their 3DVAR, Advanced Regional Prediction System (ARPS; Xue et al. 2000, 2001, 2003) forecast system. Further, Lei et al. (2009) were able to improve a deterministic forecast of the 8 May 2003 supercell through a heterogeneous environment nested-grid strategy with EnKF data assimilation of both radar data and surface mesonet data. These results illustrate the potential value of a multi-scale approach. Yussouf et al. (2013) also employed a nested grid system (outer domain with horizontal grid spacing of 18 km and inner domain with horizontal grid spacing of 2 km) to investigate the impact of single- and double-moment microphysics schemes. Since the larger domain provided initial conditions for the inner domain, the radar data assimilation and ensemble forecasts received the benefits of a heterogeneous environment. However, the radar data assimilation was only done in the afternoon of the event of interest on the inner domain and no conventional observations were assimilated on the inner domain. Thus, the analysis cycles were not multi-scale as defined in this study. Using a similar nested grid system for another case study, Sobash (2013) found that high-frequency assimilation of surface observations led to improvements in forecasts of convective initiation. Sobash (2013) also found that analyses where both radar and surface observations are assimilated produced the best forecasts. Each of the studies suggest that a multi-scale data assimilation and forecasting system could provide improved skill relative to a system in which only radar data were assimilation.

This study builds on the previous data assimilation studies with heterogeneous environments by systematically exploring potential design configurations for a multi-scale data assimilation and forecasting system that merges mesoscale and storm-scale techniques. The data assimilation system design begins with strategies similar to real-time mesoscale analysis systems (Romine et al. 2013; Schwartz et al. 2014); then concepts from storm-scale radar data assimilation systems are incorporated. One of the main differences between the mesoscale and storm-scale data assimilation approaches is the cycling frequency; mesoscale systems typically use 6-hour cycles, and storm-scale systems typically use 5-minute cycles. Thus, to bridge this gap, this study investigates the potential value of hourly cycles. Next, since mesoscale systems typically assimilate conventional observations and storm-scale systems typically assimilate radar observations, this study examines the impact of assimilating radar observations with conventional observations. Lastly, since the focus of this work is on the storm environment, the sensitivity to ensemble initialization is explored by comparing a simple approach, using an available large-scale global analysis for initialization, and a more sophisticated continuously cycled analysis for initialization. The data assimilation techniques will be applied to the 13 June convective events.

4.2 Case overview: 13 June 2010

A complex and volatile convective environment transpired on the afternoon of 13 June 2010 in the Oklahoma-Texas panhandle, leading to several tornadoes. This event was well sampled by the Second Verification of the Origins of Rotation in Tornadoes Experiment (VORTEX2; Wurman et al. 2012) instrumentation. Therefore, 13 June is an ideal case to explore the capabilities of a multi-scale ensemble data assimilation system. The convective environment and storm evolution are discussed in the following section. The VORTEX2 field operations are reviewed in Section 4.2.2 and real-time forecasts are reviewed in Section 4.2.3.

4.2.1 Storm environment and evolution

The large-scale pattern provided weak accent supportive of convection, as a mid-level closed low over Utah evolved into an open wave (Figure 13). The approaching upper low provided 40 knots of southwesterly flow over the Texas panhandle region, contributing to more than sufficient deep-layer shear for supercells. Although the mid-level temperatures were relatively warm for a severe convective event (e.g. -6 °C was observed by the 1200 UTC Amarillo, Texas radiosonde), the lapse rates were fairly steep (e.g. approximately 7°C/km at Amarillo; not shown) and the boundary layer was extremely moist (dew points above 70 °F in the Texas panhandle; Figure 15) that contributed to between 2000-3000 J kg^{-1} of Convective Available Potential Energy (CAPE) by early afternoon. While the synoptic scale pattern was supportive of organized thunderstorms, forcing along surface boundaries would primarily drive storm development and subsequent evolution.

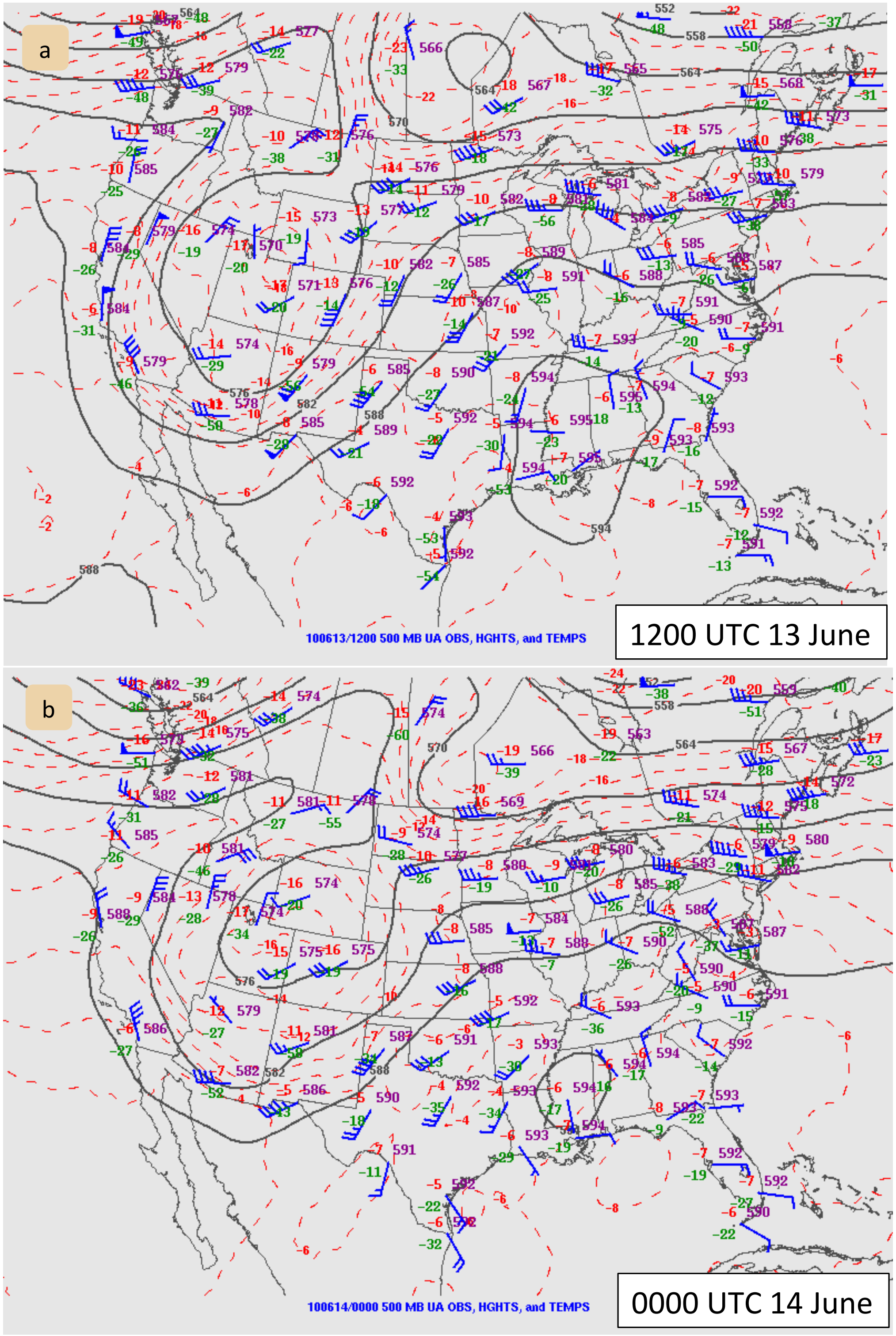


Figure 13. The 500 mb analysis chart at 1200 UTC on 13 June 2010 (a) and 0000 UTC on 14 June 2010 from the SPC archive.

Strong convection developed during the evening of 12 June 2010 and continued throughout the overnight hours into the early morning of 13 June. This is illustrated in Figure 14, which shows snapshots of the southern plains composite WSR-88D radar reflectivity throughout the 13th. The mesoscale convective systems (MCSs) that moved into Kansas overnight generated a cold pool that persisted through the afternoon. The cooler outflow air was separated from the warn sector by an east-west oriented outflow boundary that settled just south of the Oklahoma-Texas Panhandle border in the afternoon. The surface winds on the north side of this boundary were easterly, while winds south of this boundary were southerly. A southwest to northeast oriented, stalled cold front (stationary front) was also located in the Oklahoma and Texas Panhandles. The front was the primary source of lifting that lead to storm initiation. Gaps in the surface observing network in this region limited identifying the exact position of the intersection of the front and surface outflow boundary (Figure 15). Modest convection initiated just behind the surface cold front between 1700 and 1800 UTC and slowly moved to the northeast (Figure 14). The early storms approaching the Oklahoma Panhandle struggled to sustain surface-based updrafts as they crossed to the cool side of the outflow boundary, decaying as they moved towards/into southern Kansas. Further south, more intense surface-based storms developed closer to the intersection of the front and the dryline, the most northerly of which moved over the intersection of the cold front and outflow boundary around 2000 UTC. This storm rapidly intensified, gained supercell characteristics, and become tornadic at approximately 2052 UTC. Thus, the outflow boundary intersection with the front played an important role in the generation of low-level rotation. The tornado formed approximately 5 miles north of

Booker, Texas (tornado was in Oklahoma), moved to the northeast, and lasted approximately 8 minutes (hereafter referred to as the “Booker tornado” and the “Booker storm”; Figure 16). The storm produced a second rain wrapped multiple vortex tornado from approximately 2117 to 2122 UTC between Elmwood and Slapout, Oklahoma, which also moved to the northeast. Both of these tornadoes traversed over open land and no damage was reported (rated EF0). Around 2200 UTC there was no longer rotation present at low-levels in the Booker storm. Southwest of the tornadic storm, other storms had developed near the triple point and also moved to the northeast. One other storm briefly became tornadic at approximately 2226 UTC in Lipscomb County, Texas. By 0000 UTC on 14 June, the storms had evolved into several convective line segments that moved eastward into Oklahoma and south-central Kansas. The preliminary storms reports for 13 June from NOAA’s Storm Prediction Center (SPC) are shown in Figure 17, and serve as a summary of the severe events on the 13th.

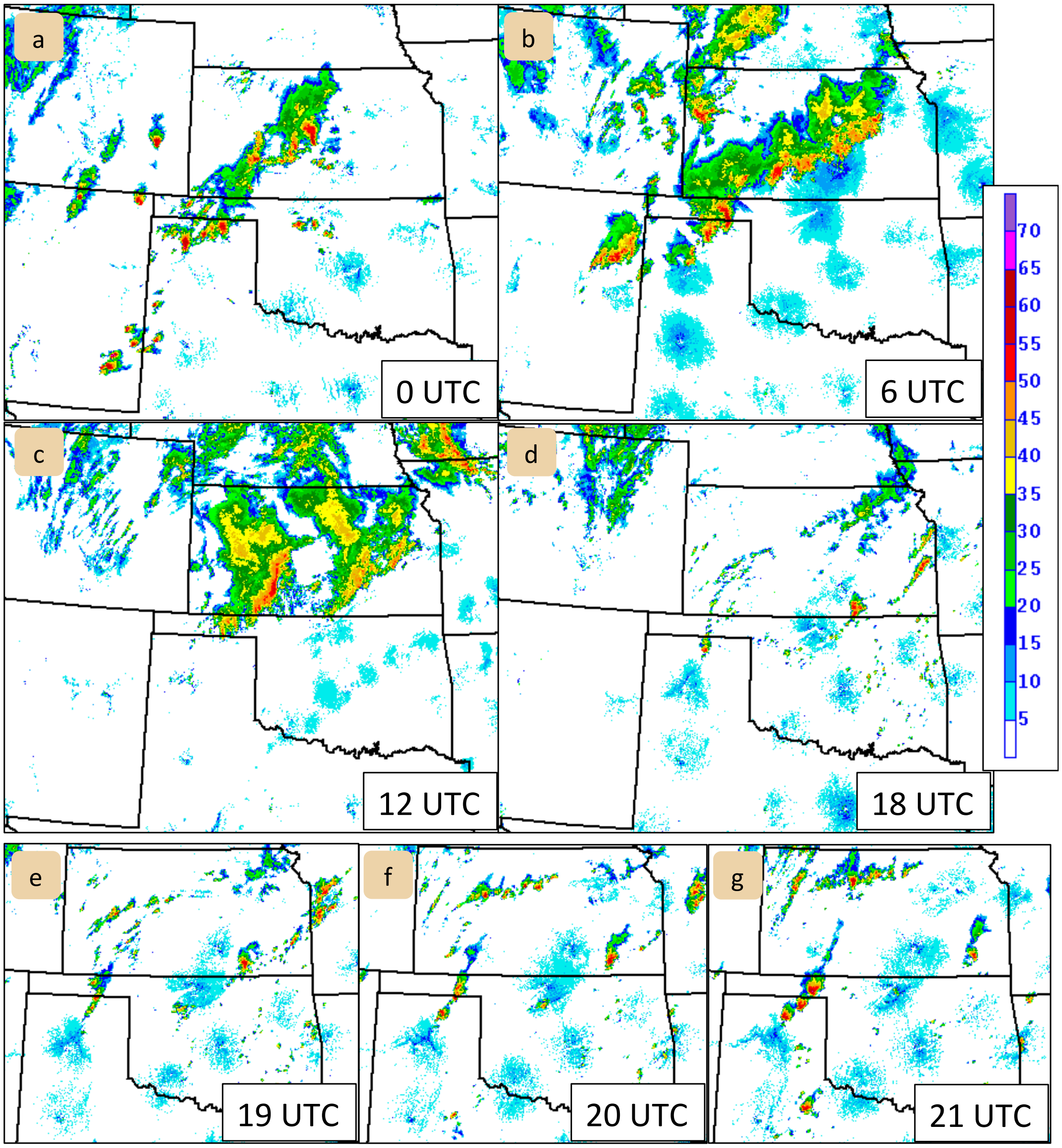


Figure 14. Composite WSR-88D reflectivity on 13 June 2010 at 0000 UTC (a), 0600 UTC (b), 1200 UTC (c), 1800 UTC (d), 1900 UTC (e), 2000 UTC (f), and 2100 UTC (g) from the VORTEX2 field catalog archive.

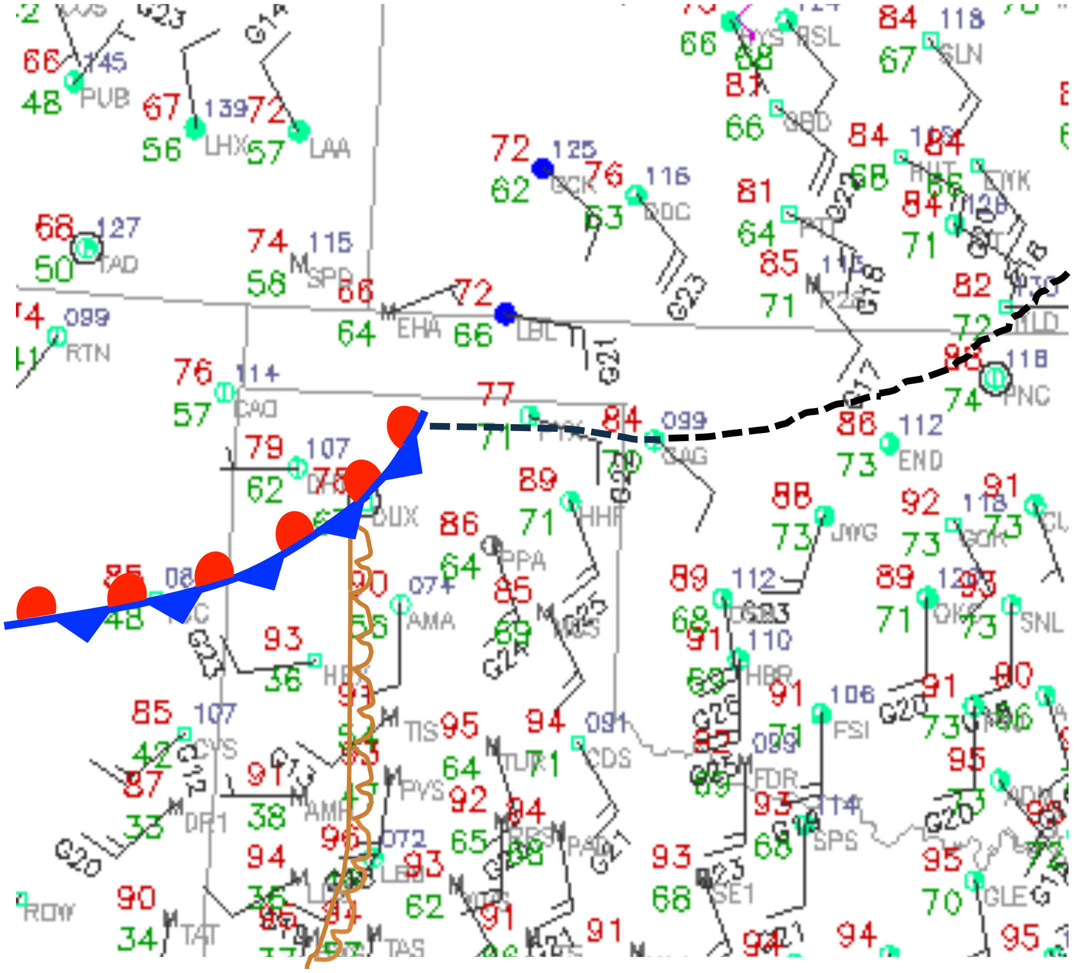


Figure 15. Surface station plot at 2000 UTC on 13 June 2010 from the VORTEX2 field catalog archive. The stationary front (red half circles and blue triangles), dryline (tan line and half circles), and outflow boundary (black dashed line) are hand analyzed in the Texas Panhandle.

Figure 16. The Booker, Texas tornado on 13 June 2010. Picture provided by Glen Romine.

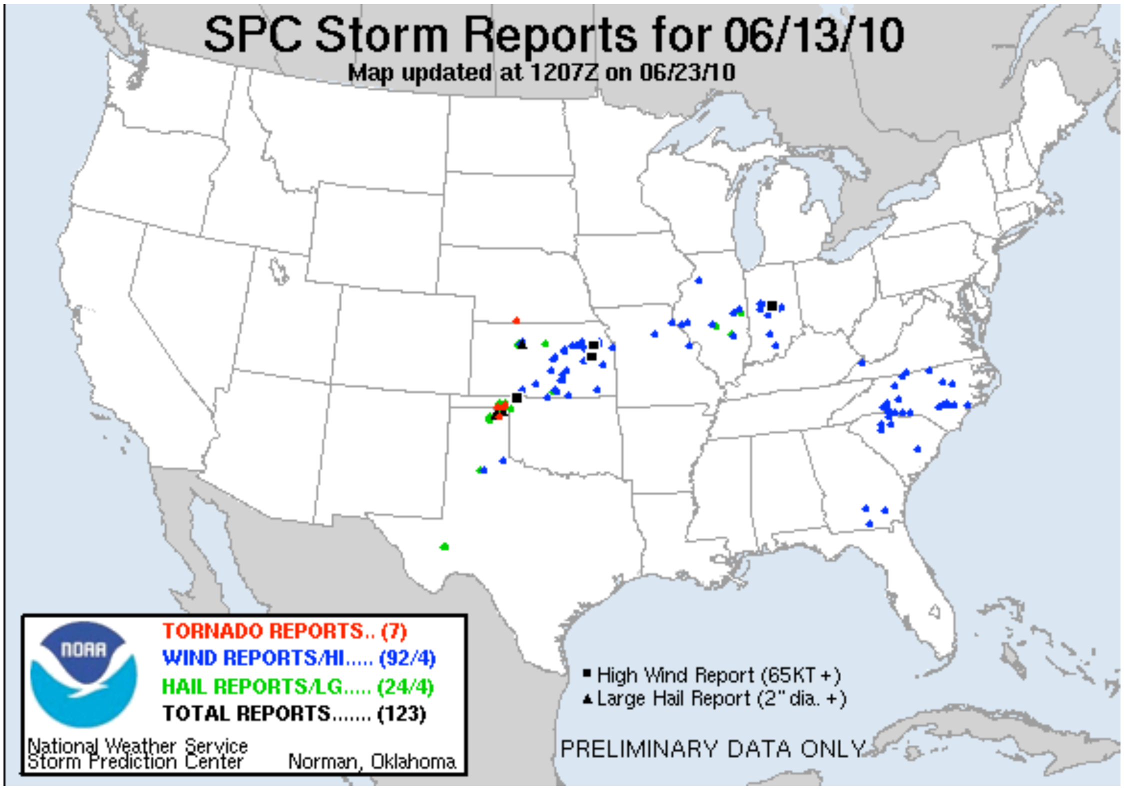


Figure 17. The preliminary storm reports on 13 June 2010 from the SPC archive.

The stationary cold front was the primary lifting mechanism for thunderstorm initiation and the presence of the outflow boundary was likely important to the development of the tornadoes. Numerous previous studies have presented evidence that tornadoes can be associated with storms that interact with low-level boundaries (e.g. Maddox et al. 1980; Markowski et al. 1998). Preexisting boundaries, such as those generated by outflow, can provide a rich source low-level horizontal vorticity, which can become vertical vorticity through tilting, and concentrated though stretching by a thunderstorm's updraft (e.g. Markowski et al. 1998; Atkins et al. 1999). The backed winds associated with the boundary also led to a more elongated hodograph enhancing storm organization. The Booker tornado likely resulted from the enhanced vorticity generated by the mesoscale outflow boundary interacting with the storm's updraft. Thus, to properly forecast the tornadic potential on 13 June, representing the outflow boundary is an essential component of the storm environment.

The 13 June event occurred in a complex mesoscale environment, in a region with sparse conventional observations, and the storms were distant from adjacent WSR-88D radars. Thus, it is hypothesized that the use of a multi-scale data assimilation approach can leverage available observations to generate an analysis with the essential components of the Booker storm environment. A successful retrieval of the storm environment for this event may hold promise that a similar approach could be applied to additional cases.

4.2.2 VORTEX2 operations

The field phase of VORTEX2 was ongoing during 13 June 2010. Teams collected data on three storms on the 13th, including the tornadic storm that produced the Booker tornado (Figure 18; VORTEX2 data archive site, http://data.eol.ucar.edu/master_list/?project=VORTEX2). The first storm VORTEX2 targeted for data collection was one of the storms that developed on the cool side of the front (Figure 18a). VORTEX2 operations abandoned that storm in favor of a storm developing to the southwest because of its more favorable position along the front, referred to as target storm #2 (Figure 18b), which produced the Booker tornado (Figure 18c). VORTEX2 teams also briefly targeted a third storm, but that storm became cutoff from warm inflow air by outflow from the previous storms (Figure 18d).

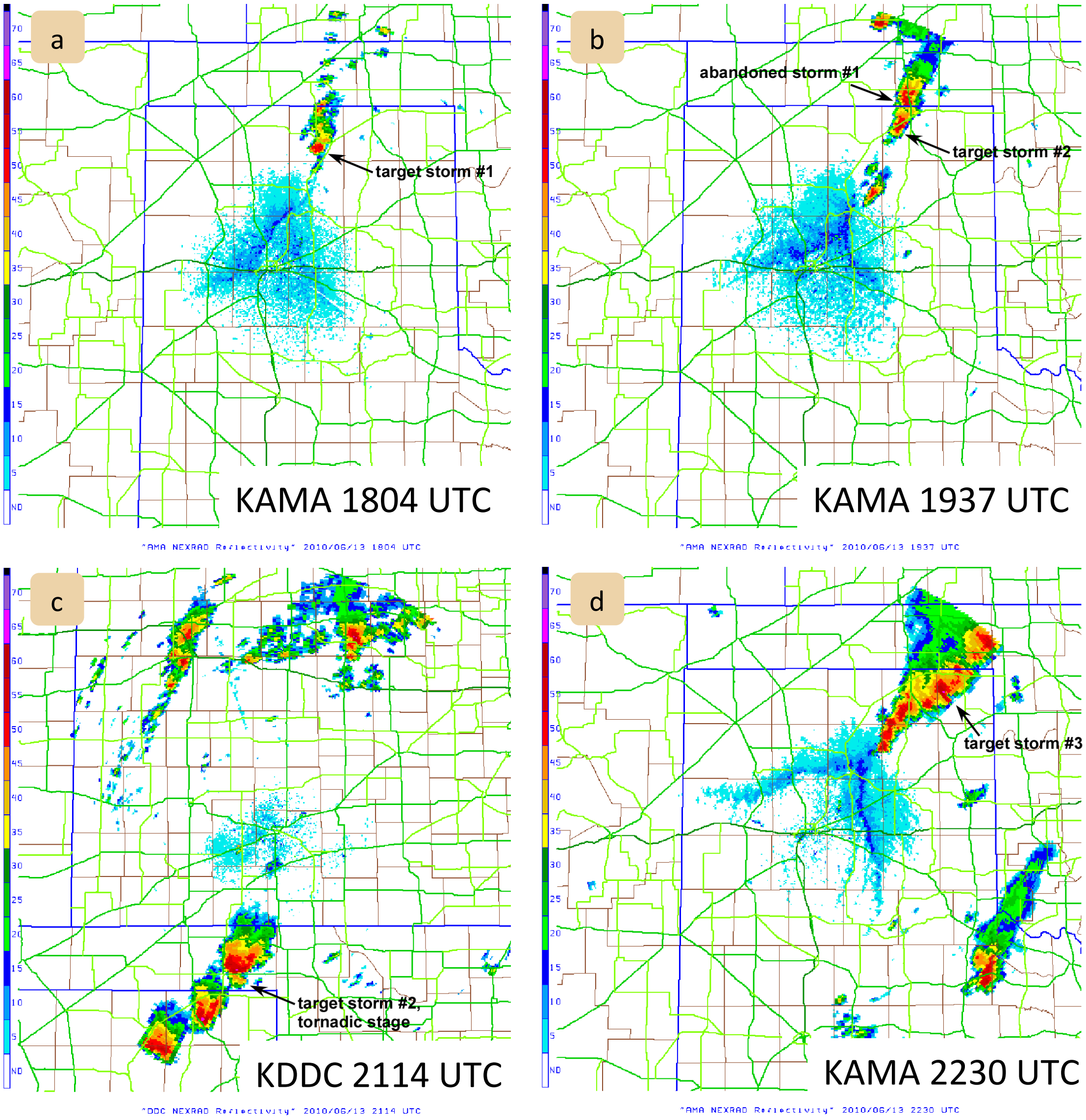


Figure 18. Field Coordinator (FC) overview the VORTEX2 field operations on 13 June 2010 via screen captures of the Situation Awareness for Severe Storms Intercept (SASSI) from the VORTEX2 field catalog archive.

The NOAA-NSSL dual-polarized X-Band mobile radar (NOXP) collected detailed observations of the Booker storm (target storm #2) interacting with the boundaries and becoming tornadic (Figure 19). At the beginning of the NOXP deployment, the Booker storm had moved over the boundary intersection and the low-level rotation began to increase (Figure 19a) leading to tornagenesis and continuing through the mature tornadic phase (Figure 19b). After approximately 2200 UTC, the Booker storm had a more linear appearance at low-levels on radar and was abandoned by VORTEX2 due to its limited tornado potential.

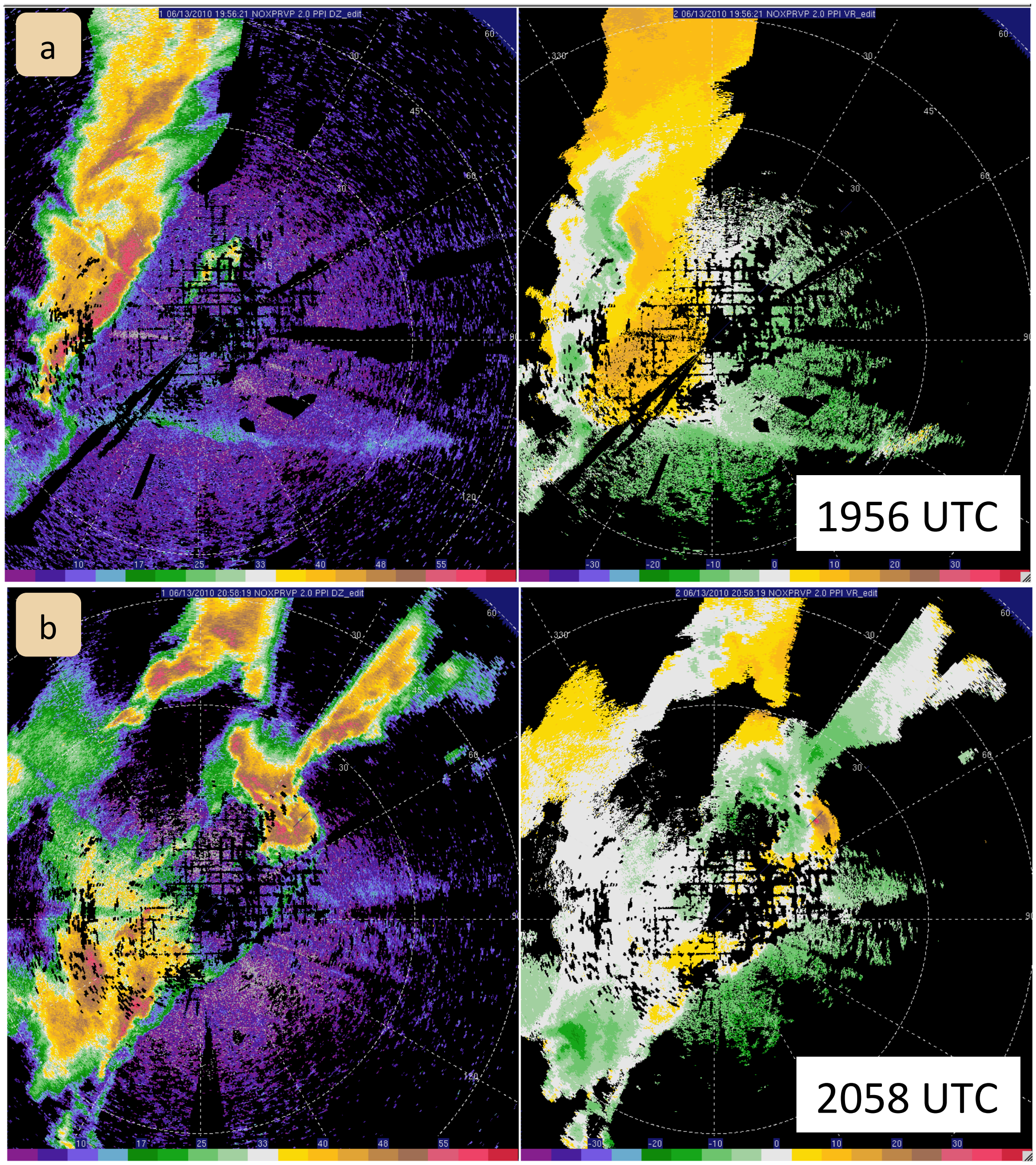


Figure 19. NOXP reflectivity and Doppler velocity data at 1958 UTC (a) and 2058 UTC (b). These data were manually quality controlled by David Dowell.

The wealth of observations collected by VORTEX2 can provide ground truth data for assimilation and verification of convective-resolving models. However, these data cannot be fully realized without including the analysis and forecast of the storm environment with the surface boundaries (the focus of this study). VORTEX2 field teams collected numerous sounding profiles of the environment on 13 June, which will be discussed in Section 4.3.5.

4.2.3 Real-time forecasts

Operational forecast products from the Storm Prediction Center (SPC) recognized the risk for severe thunderstorms to develop on 13 June in the Oklahoma and Texas Panhandles. The SPC convective outlook included discussion of the potential for enhanced storm-relative inflow and storm-relative helicity for any storms that moved over the outflow boundary in the Oklahoma and Texas panhandles. The SPC also issued a mesoscale discussion to highlight this potential tornado threat (Figure 20) and a subsequent tornado watch.

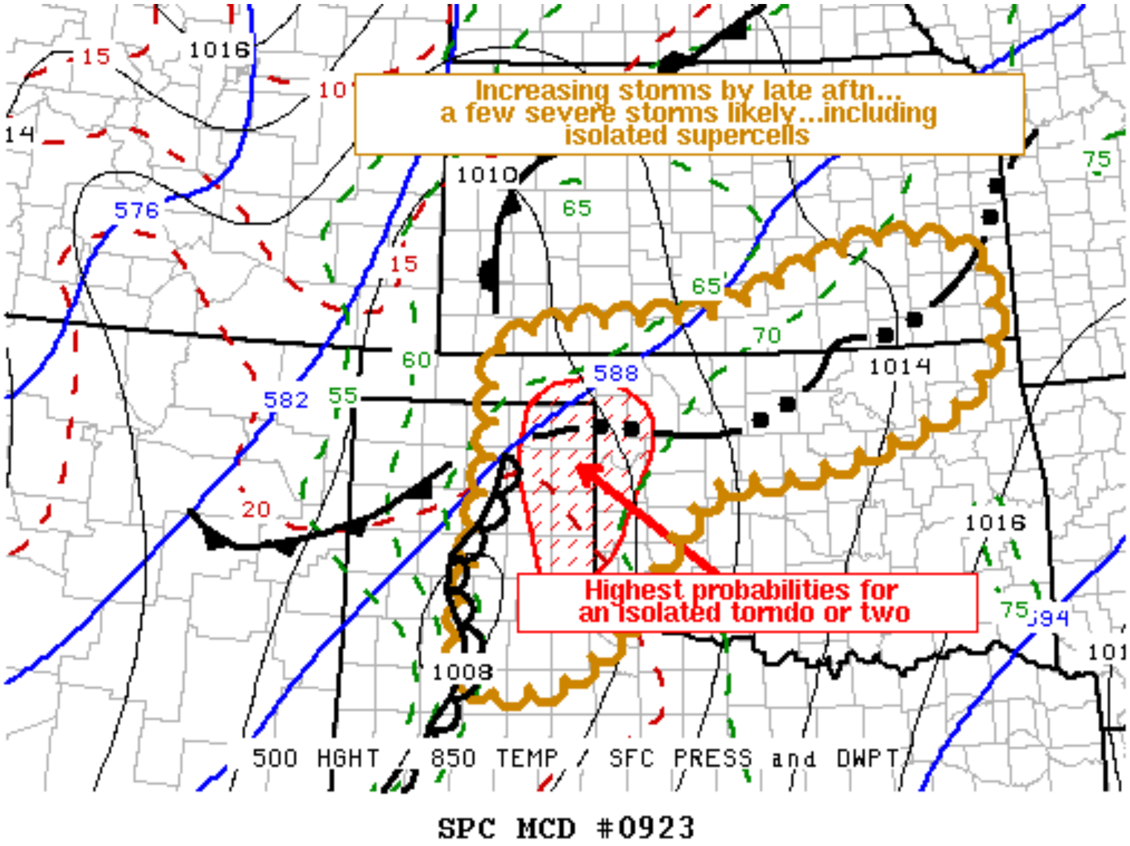


Figure 20. The mesoscale discussion #0923 graphical forecast generated by the Storm Prediction Center.

The real-time experimental convection-allowing model forecasts (CAMs) captured the presence of the stationary front and initiated convection in the Oklahoma-Texas Panhandle region. The CAMs were more accurate in terms of precipitation location than operational mesoscale models such as the North American Mesoscale Forecast System (NAM). However, neither the operational nor experimental models included an accurate representation of the observed outflow boundary. For example, the NOAA Earth System Research Laboratory's High Resolution Rapid Refresh (HRRR; Alexander et al. 2010) model forecasts contained convection in the Panhandle and Western Kansas (Figure 4.21b). However, the forecast at the surface did not contain the residual cold pool air or the easterly 10 m winds associated with the outflow boundary (Figure 21a). Nonetheless, the forecasted updraft helicity, which is a measure of rotating storms in the model (see Section 4.3.7), did indicate that some potential for rotating storms existed near Booker at 2100 UTC (Figure 21c). The updraft helicity also indicated potential for rotating storms in Southwest Kansas and North-Central Oklahoma that did not occur. The generally positive performance of the real-time HRRR indicates that this event is not beyond the predictability limit for CAMs.

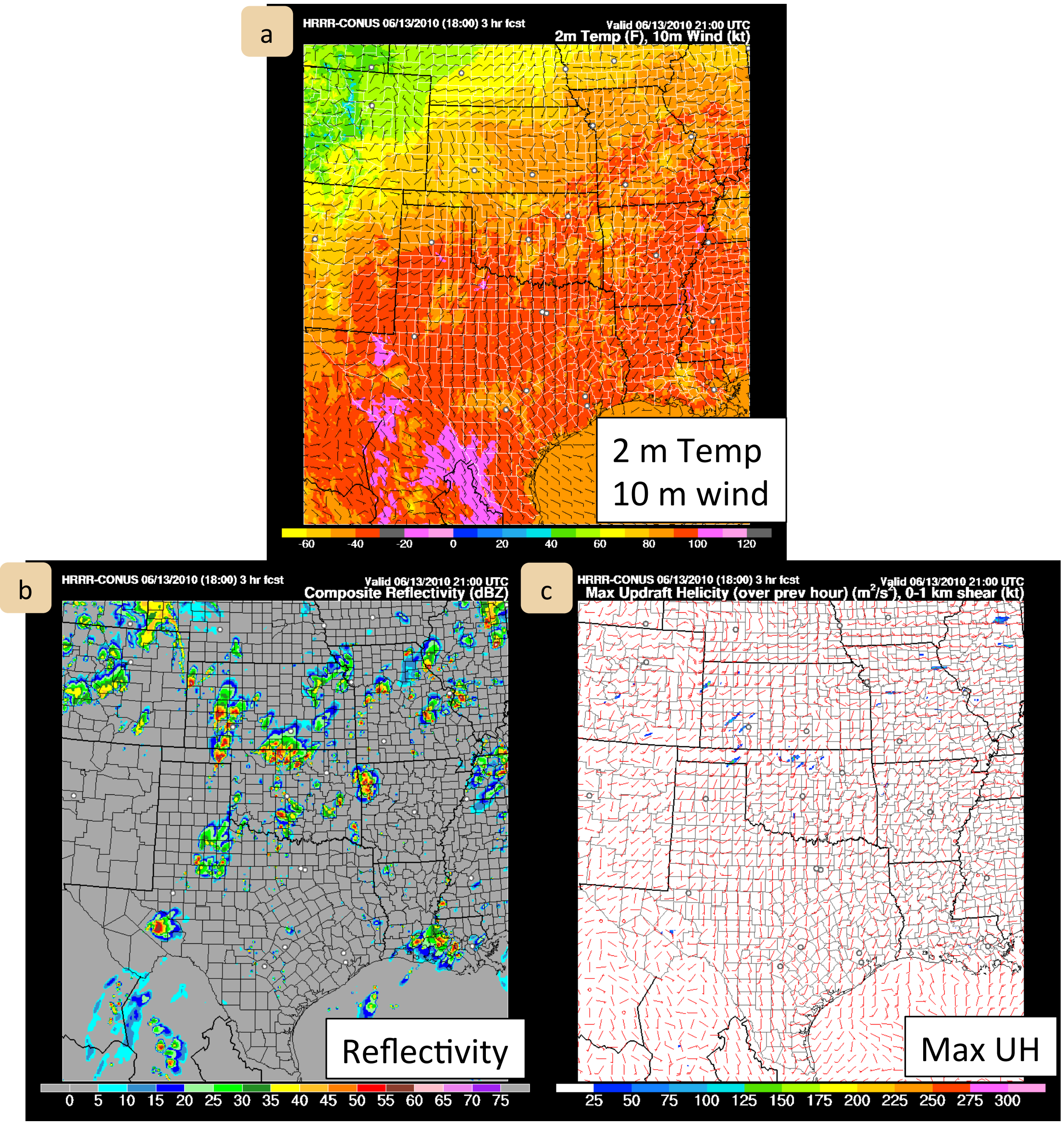


Figure 21. The High Resolution Rapid Refresh (HRRR) real-time forecast of 2 m temperature (a), composite reflectivity (b), and maximum updraft helicity and 0-1 km shear (c) valid at 2100 UTC from the VORTEX2 field catalog archive.

4.3 EnKF data assimilation and forecast system design

The EnKF data assimilation and forecast system design is based on the experience of several experimental analysis and forecast systems developed at NCAR for real-time forecasts (as discussed in Section 4.1; Romine et al. 2013, Schwartz et al. 2014). The forecast model is described in the following Section, 4.3.1. The EnKF system is reviewed in Section 4.3.2. Information about the conventional (Section 4.3.3) and radar (Section 4.3.4) observations is also presented. Section 4.3.5 explains the steps of a data assimilation and forecast cycle. The VORTEX2 sounding observation processing for verification is discussed in Section 4.3.6. Finally, Section 4.3.7 introduces the verification methods used throughout the rest of the Chapter.

4.3.1 WRF model and physical parameterizations

Since the focus of this study is on the development of a multi-scale EnKF data assimilation framework for the Warn-on-Forecast initiative (Stensrud et al. 2009, 2013), the use of an advanced, state-of-the-art atmospheric model is an essential component to generate probabilistic forecasts of weather hazards. This study uses the Advanced Research core of the Weather Research and Forecasting Model (WRF-ARW; hereafter WRF; Skamarock et al. 2008) version 3.3.1 with nested, limited-area domains for both ensemble cycled analysis and forecasts. The outer domain covers an extended area beyond the continental U.S. (CONUS) with a horizontal grid spacing of 15 km, while the one-way nested interior domain with 3 km grid spacing has an area of 810 square km centered over the convective region of interest (Figure 4.22a). 56 vertical Eta levels stretch from the surface to 50 hPa aloft (Figure 4.22b). Care was taken in selecting the vertical Eta levels to provide enhanced resolution near the surface and smooth changes in vertical spacing between levels. A time step of 30 (7.5) seconds was used on the outer (inner) domain. Positive definite moisture advection (Skamarock and Weisman 2009) was used on both domains. The ARW core uses a 3rd-order Runge-Kutta time integration scheme coupled with a split-explicit 2nd-order time integration scheme for the acoustic and gravity wave modes, and 5th-order upwind-biased advection operators are used in the fully conservative flux divergence integration (Skamarock et al. 2008).

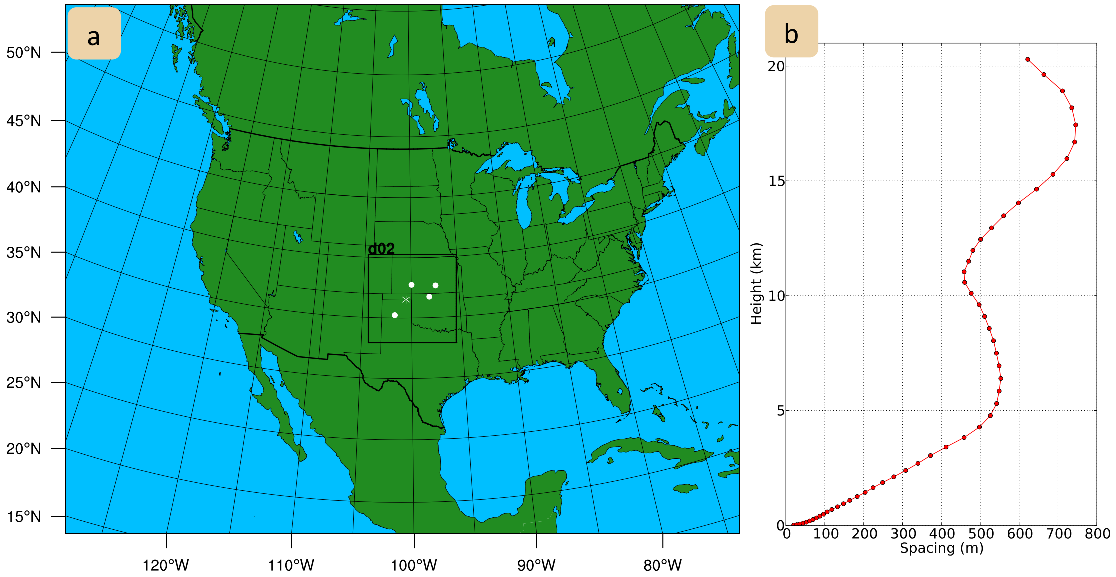


Figure 22. The map view of the horizontal coverage for the outer and inner WRF domains (a). The white dots indicate the WSR-88D locations that are utilized and the white star is the location of the Booker tornado. The vertical Eta levels approximate height and spacing between levels used for both domains (b).

The WRF model provides a wide range of physics parameterizations to represent phenomenon and processes that are not fully resolved in the model. The WRF physics settings and domain information is summarized in Table 1.

Table 1. WRF Settings Summary.

WRF setting	Domain 1 value	Domain 2 value
Horizonatl grid	415 x 325, Δx = 15 km	271 x271, Δx = 3 km
Vertical grid	50 levels, p_{top} = 50 hPa	
PBL scheme	MYJ	
Microphysics scheme	Morrison 2-moment	
Radiation (LW) scheme	RRTMG	
Radiation (SW) scheme	RRTMG	
Land-surface scheme	NOAH	
Cumulus scheme	Tiedtke	none

In this study, the Mellor-Yamada-Janjić scheme (MYJ) is used for the planetary boundary layer (PBL) parameterization (Janjić 1994). The MYJ PBL scheme uses the 1.5-order turbulence closure model of Mellor and Yamada (1982) and determines eddy diffusion coefficients from prognostically calculated turbulent kinetic energy (TKE). The MYJ PBL scheme is a local closure scheme, meaning the turbulent fluxes at each model grid point are estimated from the mean atmospheric variables and/or their gradients at that point. Thus, it assumes that the fluxes depend solely on local values and gradients of basic model variables. This assumption is least valid under convective conditions when turbulent fluxes are dominated by large eddies that transport fluid longer distances (Troen and Mahrt 1986; Stull 1984; Hu et al. 2010). However, the MYJ PBL scheme has commonly been used in convection focused forecast systems (e.g. Yussouf et al. 2013; Johnson et al. 2011). Further, Romine et al. (2013) found that when the MYJ scheme was used in a continuously cycled data assimilation system, subsequent forecasts had less bias than forecasts from an analysis system leveraging the Yonsei University PBL scheme (YSU; Hong et al. 2006). Previous studies have identified the MYJ PBL scheme to have a cool and moist bias due to too little mixing (Kain et al. 2005; Weisman et al. 2008). While the MYJ PBL scheme may be one of the better-performing schemes in WRF, the use of any PBL scheme at CAM resolution (~3 km horizontal grid spacing) could be impinging upon a "gray area" in grid spacing, in which resolved large eddies blend with the parameterized mixing from the PBL scheme (Stensrud 2007). The PBL scheme is likely one of the largest contributors to model error (Coniglio et al. 2013) and the forecast sensitivity to the PBL choice warrants future investigation.

Both domains also use the Morrison double-moment bulk microphysics scheme (Morrison et al. 2009). Double-moment bulk schemes predict both the mixing ratios and number concentrations of hydrometeor size distributions. The Morrison scheme uses five hydrometeor species: cloud droplets, cloud ice, snow, rain, and graupel. Several studies have shown that the use of double-moment microphysics provides a better representation of modeled storms (e.g. Morrison et al. 2009; Dawson et al. 2010; Yussouf et al. 2013). The Rapid Radiative Transfer Model for General circulation models (RRTMG) is used for the longwave and shortwave radiation schemes (Iacono et al. 2008). The Noah land surface model is used for the land surface parameterization (Chen and Dudhia 2001; Ek et al. 2003). The radiation and land surface scheme choices are expected to have less impact on short range (0-3 hour) forecasts compared to the PBL and microphysics schemes, but could play an increasingly important role as the cycling period extends for longer windows (Johnson et al. 2011). The outer WRF domain (15 km horizontal grid spacing) additionally requires the use of a cumulus parameterization to account for subgrid-scale vertical redistribution of heat and moisture resulting from moist convection. This study uses the Tiedtke cumulus parameterization scheme (Tiedtke 1989; Zhang et al. 2011) following the results of Torn and Davis (2012) and Romine et al. (2013), as the latter study found smaller analysis errors using the Tiedtke scheme compared to when the Kain-Fritsch scheme (Kain 2004) was used in the cycled data assimilation system.

The WRF grids are initialized starting from a downscaled 6-hourly National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS) analysis. The 50-member ensemble initial conditions are generated by adding random draws with zero mean and covariances from the global background error covariances using the WRF data assimilation software (WRFDA; Barker et al. 2012) random 'CV3' option, as in Torn and Hakim (2008). These samples are added to the horizontal components of wind, water vapor mixing ratio, and temperature of the analysis state. The lateral boundary condition for the analysis and target states (perturbed boundary conditions) are updated using the fixed covariance technique (Torn et al. 2006). The perturbations are an attempt to account for uncertainties in both the initial and boundary conditions. Further discussion of the ensemble initialization and its impact can be found in Section 4.7. Since the edges of the interior domain are spatially distant from the outer domain boundary edges, the imposed state on the outer domain lateral boundaries does not have a significant influence on the interior domain in short duration forecasts (e.g., Romine et al. 2013). The soil state is initialized identically for all ensemble members from the GFS analysis, but is allowed to evolve freely in each ensemble member thereafter. Tests in which the soil state is re-initialized from the GFS at every analysis time led to slightly larger errors in the analysis fit to surface observations (not shown).

4.3.2 DART EAKF data assimilation system

This study uses the ensemble adjustment Kalman filter (EAKF; Anderson 2001; see Section 2.2.5) option within the Data Assimilation Research Testbed toolkit (DART; Anderson and Collins 2007; Anderson et al. 2009). In order to maintain ensemble spread, adaptive spatially and temporally varying inflation (Anderson 2009) is applied to the prior (background first guess) state. The initial inflation has a mean of 1.0 and standard deviation of 0.8. Sampling error correction (Anderson 2012) is also applied to help reduce influence from spurious correlations due to a limited ensemble size. Horizontal and vertical localization is used to reduce impact of sampling errors using the isotropic weighting function from Gaspari and Cohn (1999). The cutoff length (weight becomes zero) in the horizontal (vertical) for conventional observations is 1020 (13) km away from the observation location and 24 (12) km for radar observations. Observations are rejected when the squared difference between the observation and the prior ensemble mean exceeded 3 times the sum of the prior ensemble variance and observation error variance. The analysis is updated using DART from a 50-member ensemble of WRF forecasts. The parameters used here are based on Romine et al (2013) and additional real-time cycling experiments using DART performed at NCAR (e.g. Schwartz et al. 2014). Table 2 summarizes the DART settings.

Table 2. DART settings summary.

DART setting	Conventional Obs. Value	Radar Obs. Value
Filter type	EAKF	
Ensemble members	50	
Outlier threshold	3	
Sampling error correction	TRUE	
Adaptive prior inflation	Initial = 1.0, SD = 0.6	
Localization type	Gaspari and Cohn (1999)	
Horizontal localization cuttoff	1020 km	24 km
Vertical localization cuttoff	13 km	12 km
Additive noise	none	1 m s^{-1}, 1.0 K

When high-density observations are assimilated, there is a tendency for the ensemble to become under dispersive, which can lead to filter divergence. In particular, for the assimilation of radar observations, additional spread is often provided by using an additive noise technique (Dowell and Wicker 2009). As described in Section 3.2.1, the technique employed in this study adds Gaussian perturbations with standard deviations of 1.0 m s^{-1} or 1.0 K respectively for the horizontal wind components, temperature and dewpoint temperature fields. This random noise is applied each analysis cycle, in regions where observed reflectivity > 40 dBZ, to the model state variables immediately before the updated ensemble is integrated forward in time.

4.3.3 Conventional observation sources and processing

Routinely available observations are obtained from the NOAA's Global Systems Division (GSD) Meteorological Assimilation Data Ingest System (MADIS), which includes automated quality control (described at http://madis.noaa.gov/madis_qc.html). The observations used in this study include mandatory and significant level rawinsondes [u,v,T,Td, altimeter (Alt)], standard aviation routine weather report (METAR) and maritime reports [u,v,T,Td,Alt], and Aircraft Meteorological Data Relay (AMDAR) reports [u,v,T]. AMDAR report density is reduced by averaging observations over boxes of dimension 30 km in the horizontal and 25 hPa in the vertical, following Torn (2010). Atmospheric motion vectors (AMVs; Velden et al. 2005) [u,v] are obtained from the Cooperative Institute for Satellite Studies Space Science and Engineering Center. AMVs are also averaged spatially but over 60 km in the horizontal and are excluded over land. Global Positioning System (GPS) radio occultation observations (Kursinski et al. 1997) are obtained from the Constellation Observing System for Meteorology Ionosphere and Climate (COSMIC) with profiles thinned to 15 levels in the vertical.

MADIS also provides surface observations [u,v,T,Td,Alt] from various local mesoscale networks, collectively referred to as mesonets. These data can be valuable due to the increased observation coverage that they provide at the surface. However, the instruments and instrument sighting are not standardized, with limited quality control by MADIS, and therefore these observations are less reliable for use in the assimilation system. Separately, Oklahoma Mesonet data [u,v,T,Td,Alt] are obtained from the Oklahoma Climatological Survey. Given the questionable reliability, mesonet data are only used for the data assimilation and forecast verification on June 13th.

A complete list of the observation types and assumed observation errors appear in Table 3. The moisture observation type from all platforms is dew point temperature. Dew point observation errors are assigned following Lin and Hubbard (2004), with the errors increasing for decreasing relative humidity. Surface observations are excluded when the model terrain and station height differ by more than 300 m, in order to reduce potential observation quality errors near steep terrain. To enhance system stability, analysis increments adjacent to the grid lateral boundary edges are minimized; all observations within three grid lengths of the lateral boundaries are excluded, and observation errors for observations within five grid lengths of lateral boundaries are inflated.

Table 3. Observation types and errors.

Platform	Variable	Observation error
Radiosonde	Temperature E-W, N-S winds Dew Point	NCEP statistics NCEP statistics Lin & Hubbard (2004)
AMDAR (30 km, 25 hPa)*	Surface altimeter Temperature E-W, N-S winds	2 hPa NCEP statistics NCEP statistics
METAR	Altimeter Temperature E-W, N-S winds Dew Point	0.75 hPa 2 K 1.75 m s^{-1} Lin & Hubbard (2004)
Maritime reports	Altimeter Temperature E-W, N-S winds Dew Point	1 hPa 2 K 1.75 m s-1 Lin & Hubbard (2004)
Mesonet	Altimeter Temperature E-W, N-S winds Dew Point	1 hPa 2 K 1.75 m s-1 Lin & Hubbard (2004)
AMV (60 km, 25 hPa)*	E-W, N-S winds	50 % NCEP statistics
GPS (thinned to 15 levels)	RO refractivity	Kuo et al. (2004)

* Superobs (horizontal, vertical)

All the available observations at 2000 UTC on 13 June are plotted in Figure 23, as an example of the observational coverage for the interior WRF domain. The majority of the available observations are located at the surface (METAR and mesonets).

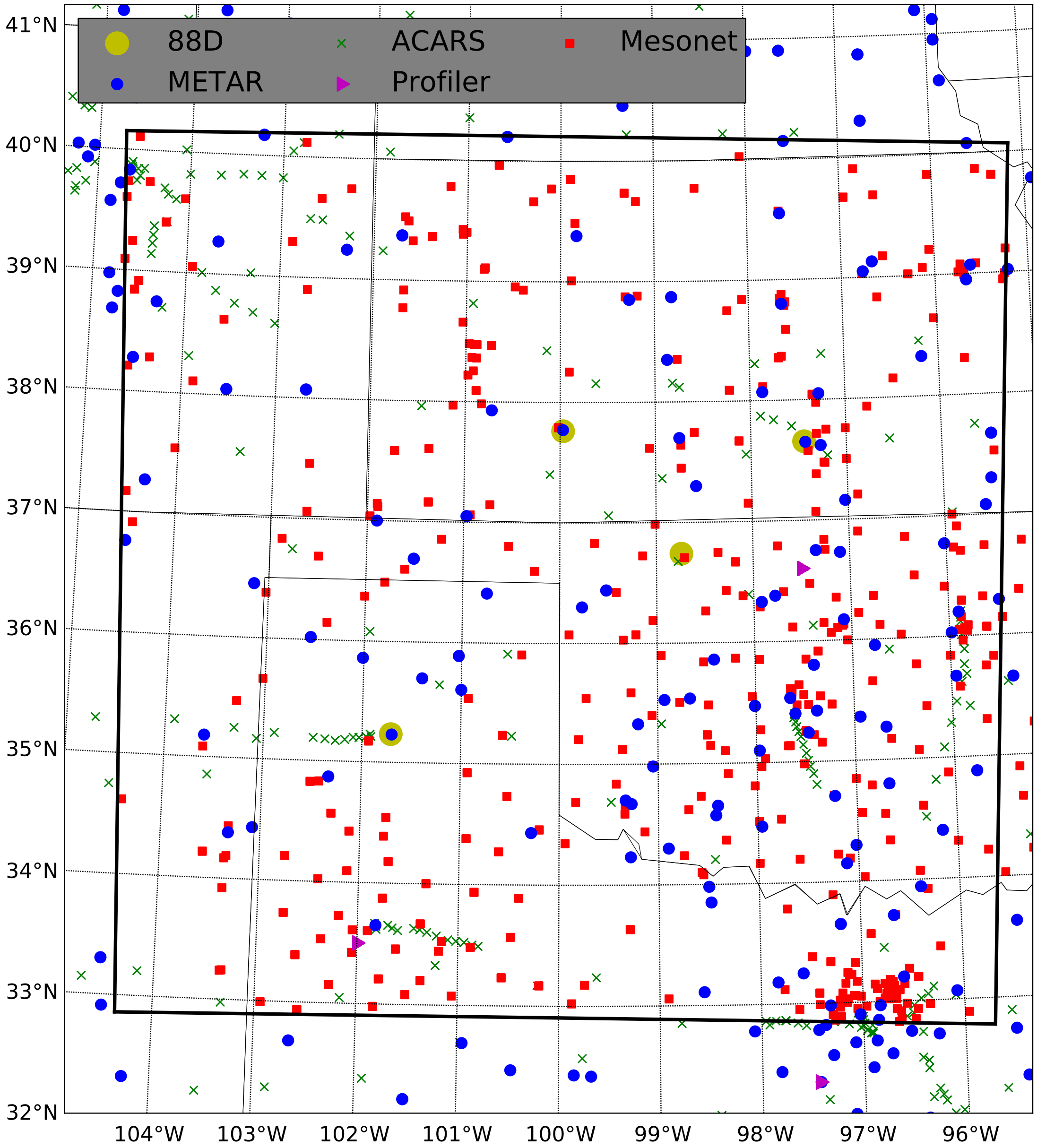


Figure 23. The available conventional observations at 2000 UTC. The inner WRF domain is the black box and the observation locations are marked with symbols for each observation platform. The locations of the four WSR-88D radars are shown for reference.

4.3.4 WSR-88D radar observations and processing

Doppler velocity, radar reflectivity factor above 10 dBZ (hereafter referred to as reflectivity), and clear air reflectivity (less than or equal to zero reflectivity) data from four operational WSR-88D sites are simultaneously assimilated. The WSR-88D sites are the Amarillo, Texas (KAMA), Dodge City, Kansas (KDDC), Wichita, Kansas (KITC), and Vance, Oklahoma (KVNX) (white circles in Figure 22; yellow circles in Figure 23, 24). The WSR-88D data is obtained from the National Climatic Data Center (NCDC). A single volume of data that is closest to the assimilation time is selected from each radar site. The reflectivity data are automatically edited using the Quality Control Neural Network (QCNN; Lakshmanan et al. 2007) method to remove non-meteorological echoes, anomalous propagation, and ground clutter. The Doppler velocity is dealiased using the method from Eilts and Smith (1990) and the built-in DART quality control check that uses ensemble estimates of the radial observation at several Nyquist velocity offsets to determine the appropriate unfolding, if needed. The edited reflectivity and velocity observations are objectively analyzed using the Observation Processing and Wind Synthesis (OPAWS; http://code.google.com/p/opaws/; Majcen et al. 2008) software. To reduce spatial error correlation and improve computational efficiency, radar observations are analyzed onto a regularly spaced 6 km grid in the horizontal, but on the original conical scan surfaces (Sun and Crook 2001; Dowell et al. 2004; Dowell and Wicker 2009) using a two-pass Barnes (1964) scheme. Clear air reflectivity observations are also assimilated to help suppress spurious convection that may develop in the model (Tong and Xue 2005; Aksoy et al. 2009; Dowell et al. 2011), but are analyzed on a coarser 12 km grid (limited testing has found reduced observation density of clear-air reflectivity can be used to suppress spurious convection while reducing the computational cost).

Radar observation error standard deviations are uniformly assumed to be 5 dBZ and 2 m s^{-1} for reflectivity and Doppler velocity, respectively. Reducing observation errors would increase confidence (weighting) of observations on the analysis, but has the potential to enhance bias. An example of the WSR-88D radar reflectivity observation coverage is shown in Figure 24 for 2000 UTC on 13 June, 2010. While the observations cover much of the horizontal interior domain, vertical coverage varies by distance from radars.

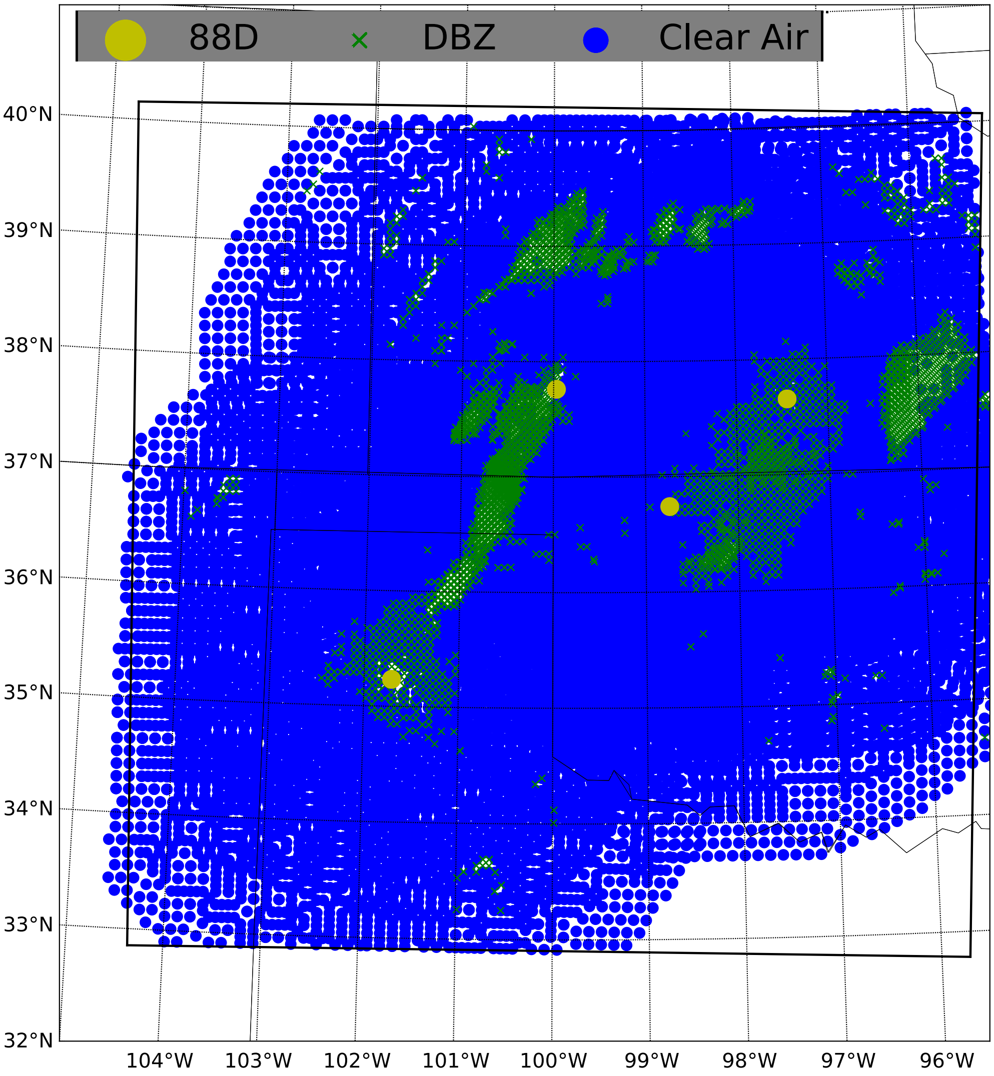


Figure 24. The available reflectivity observations at 2000 UTC from the KAMA, KDDC, KITC, and KVNX WSR-88D radars.

4.3.5 Data assimilation and forecast cycles

A *cycle* refers to the two-step process of creating an analysis and a forecast, which was originally introduced by Bjerknes (Bjerknes 1904, translated by Mintz 1954; see Section 2.1.1). Here, the diagnostic step is provided by the DART data assimilation system and the prognostic step is the WRF model forecast. This process is further broken down to explain the flow of a cycle in more detail.

For the assimilation of conventional observations, a cycle proceeds as follows:

(1) Run a set of WRF forecasts to create the prior ensemble.

(2) Apply adaptive inflation to the state, based on conventional observations.

(3) Assimilate the conventional observations with DART to generate a new analysis state, providing ensemble initial conditions

These three steps are repeated for each time interval (cycle; Figure 25a).

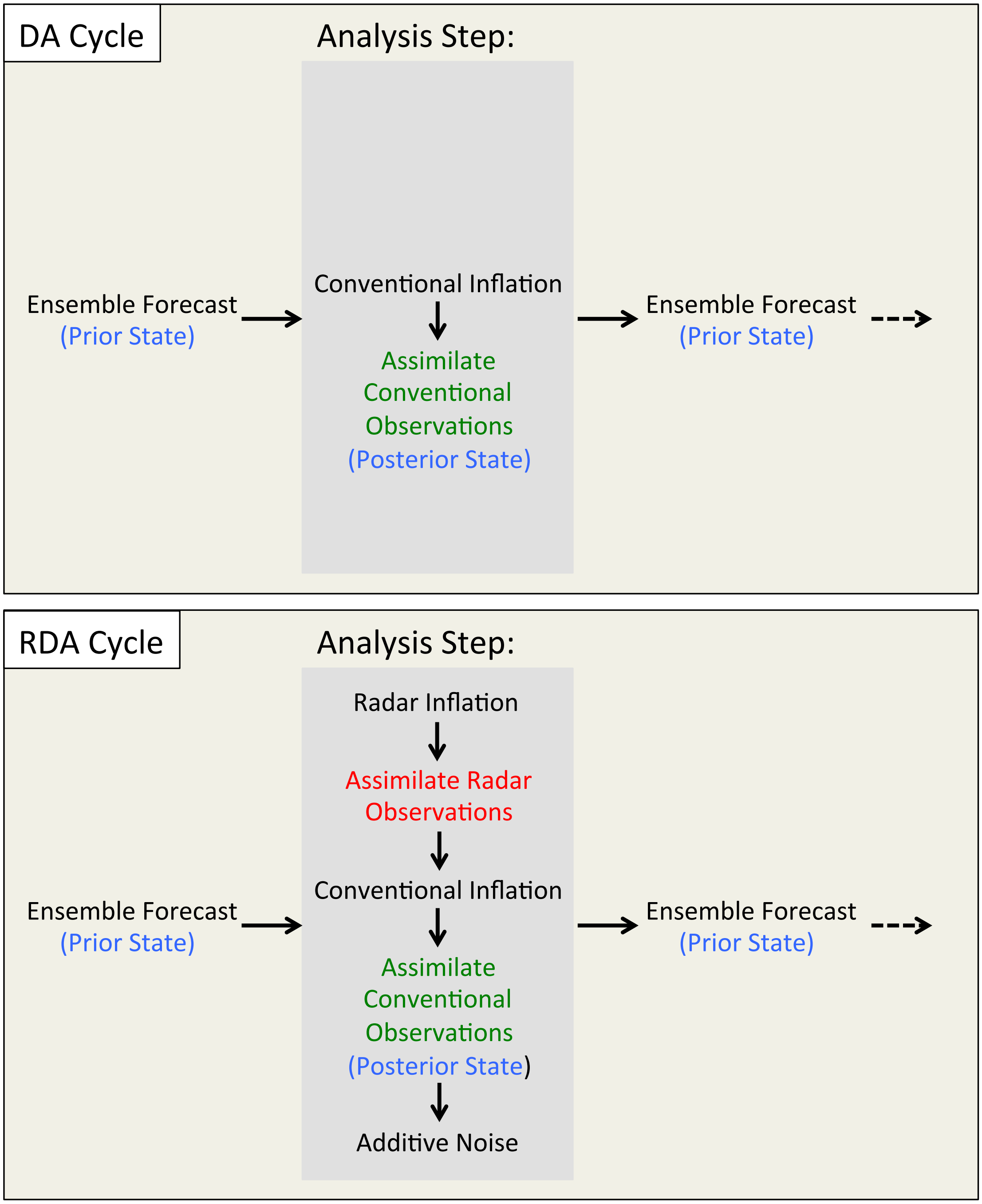


Figure 25. The flow chart of a data assimilation – forecast cycle for the assimilation of conventional observations (a) and for the assimilation of both radar and conventional observations (b).

To allow for the assimilation of both conventional and radar observations, the above analysis step is modified into two sub-steps based on observation platform type for practical reasons owing to runtime performance. The DART EAKF algorithm searches for nearby grid points an observation should update using the largest specified localization cutoff length. Recall, the horizontal cutoff length for conventional (radar) observations is 1020 (24) km away from the observation location, with radar observations typically available every 6 km in the horizontal. The unnecessarily large search radius for radar observations, combined with the large number of available observations, required excessive computational resources at the time of this study[3]. In fact, the wall clock time required is so excessive that it is impossible to run on some computing systems. The difficulties were avoided by doing the data assimilation step in two sub-steps, first the radar assimilation followed by assimilation of conventional observations. Thus, a single cycle for radar and conventional observation assimilation proceeds as follows:

1. Run a set of WRF forecasts to create the prior ensemble.
2. Apply adaptive inflation to the state, based on the radar observations.
3. Assimilate the radar observations.
4. Apply adaptive inflation to the state, based on conventional observations.
5. Assimilate the conventional observations.
6. Apply additive noise to the state.

These steps are repeated for each time interval (Figure 25b).

If the observations are uncorrelated, the order observations are assimilated does not matter. However, since real observations are correlated, the order they are

---

[3] Subsequent to this study, an alternate approach was developed with the DART toolkit enabling significantly improved computational performance for assimilation of observation sets with vastly different horizontal cutoff lengths.

assimilated can have an impact. Further, since adaptive inflation is separated into two parts, along with the use of an outlier threshold, the order observations are assimilated is even more likely to have an impact. The radar observations are assimilated first to allow them to take advantage of the largest ensemble spread that exists in the background. When the ensemble spread is larger, the observations are less likely to be rejected by the analysis system. The opposite assimilation order (conventional observations first) was also tested, revealing that the increments from the radar observations are slightly smaller when they are assimilated after the conventional observations. Ideally, the radar and conventional observations would be assimilated during the same analysis step. Errors introduced from using a two-step process are not well known at this time, but are not expected to significantly impact results. This topic will be investigated further in the future.

4.3.6 VORTEX2 sounding observations and processing for verification

The VORTEX2 mobile sounding teams launched 17 balloons on 13 June 2010 with Vaisala RS92 radiosondes. The Earth Observing Laboratory (EOL) at NCAR provided a quality-controlled version of the data (additional information is available in a "readme" document on the EOL VORTEX2 data archive site; http://data.eol.ucar.edu/master_list/?project=VORTEX2). The quality controlled sounding data are usually available every second while the balloon is ascending. Thus, there are more than 2500 individual observations within a single sounding. This far exceeds the vertical resolution of the model estimates of the atmospheric state, so prior to using the sounding data for verification, the data are thinned to contain approximately 10% of the observations below 700 hPa, 5% of the observations between 700 hPa and 500 hPa, and 2% of the observations above 500 hPa. Any observations above 200 hPa are removed. The thinned data sets contain approximately 110 observation locations for each sounding profile, which is still about twice as dense as the vertical levels in the model (the model extends to 50 hPa, not 200 hPa).

Radiosondes can travel up to 200 km horizontally during the hour-long ascent. As such, observation locations are updated based on the location of the ascending balloon. Further, observations are split into 15-minute time windows centered on the forecast output times to better match the time of the observations with the forecast model state. To evaluate the skill of ensemble forecasts, DART is used in evaluation mode to interpolate the forecast model state to the observation type and location. This allows quantitative measures such as the root mean squared error to be computed. The 10 sounding launch sites used for forecast verification from 1800 to 2100 UTC across the region of interest are shown in Figure 26.

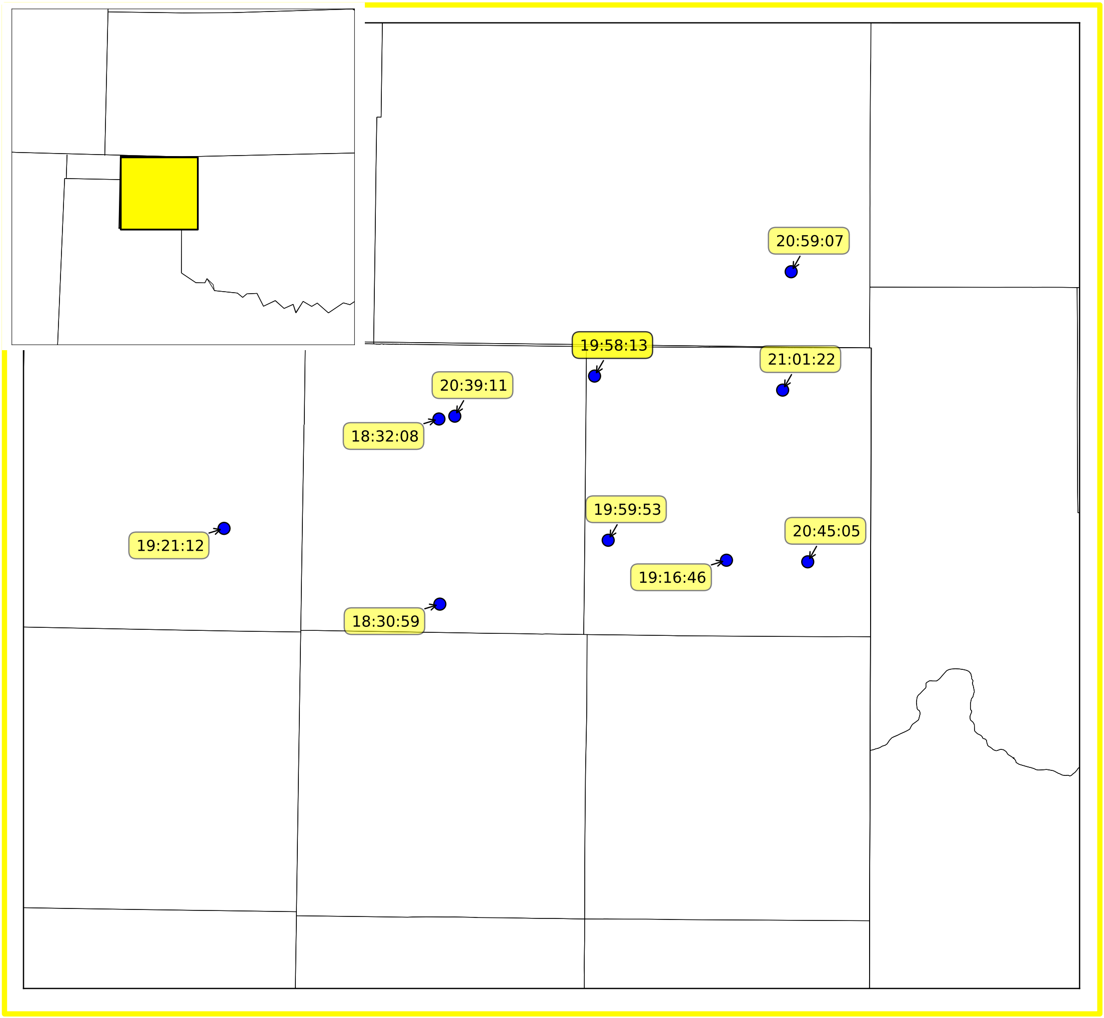


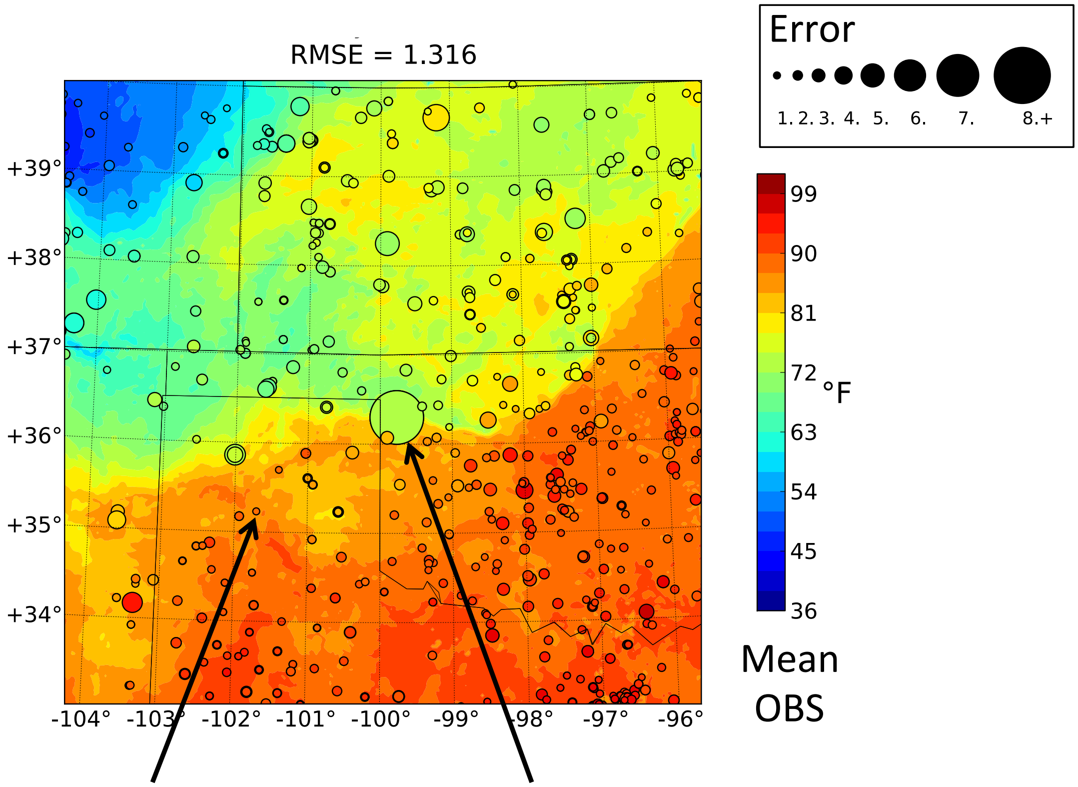
Figure 26. The VORTEX2 sounding launch sites on 13 June 2010.

The VORTEX2 sounding launched at 1958 UTC, was released just north of the outflow boundary location near Booker, Texas. Thus, it collected observations of the environment in the inflow region of the Booker storm. The ability to reproduce the inflow environment in the model forecasts is believed to be important for the reproduction of the Booker storm in the model forecast. Comparisons between the observed profile and the model forecast profiles are examined to investigate the differences in the forecast experiments.

4.3.7 Verification metrics

Analyses and forecasts are compared against available observations to assess their relative skill. In this study, the term '*analysis fit'* refers to the quantitative difference between an observed quantity (**O**) and the model analysis (**A**) value of the same quantity at the observation location, (**O** - **A**). Also, model '*forecast errors' refers to* subtracting the forecasted value of an observed quantity (**F**) at the observation location from the observation itself (**O** - **F**). Another quantitative evaluation metric is the Root Mean Squared Error (RMSE). RMSE is computed by squaring the analysis fits (or forecast errors), computing the mean value for all observations, and then computing the square root. The ensemble mean analysis (forecast) can be used to represent the best estimate of the true state in a least squares sense (Toth and Kalnay 1993; Stensrud et al 2000).

Model, Error, and Observation Weather (MEOW) plots will be used to display the results of the data assimilation and forecasts. An example MEOW plot for temperature is shown in Figure 27. A MEOW plot has two components; a contour plot, and an overlain scatter plot with markers of varying sizes. First, the ensemble mean analysis or forecast state is contoured (the model). Then, circular markers are overlain at the locations of available observations, which are also referred to as *dots*. The color of each dot represents the value of the observation using the same color scale as the model field (the observation). For example, the bright green color indicates temperatures between 69 °F and 72 °F (Figure 27). The dot sizes vary based on the magnitude of the absolute value of the analysis fit or forecast error (the error). For example, if the analysis fit is close the dot is small, while if the analysis fit is poor the dot is large. Thus, differences in the dot color relative to the background color indicates where the analysis (forecast) is warmer or cooler (drier or more moist) than observations, with the locations having greater disagreement highlighted by large markers. The RMSE is shown in the title of the MEOW plots. MEOW plots for temperature and dew point temperature are shown throughout the result sections of this chapter.



Analysis fit = +1°F
T_{ob}= 87-90 °F

Analysis fit = -7°F
T_{ob} = 69-72 °F

Figure 27. Model, Errors, and Observations Weather plot example. The contours are the mean model temperature. The dot sizes are the model error. The dot colors are the temperature observation.

The forecast experiments can also be evaluated by comparing model proxies for storm tracks and severe weather events to the observed storm locations and tornado tracks. For example, updraft helicity is a diagnostic quantity indicating rotating updrafts in simulated storms. Updraft helicity is computed by taking the integral of the vertical vorticity times the updraft velocity between 2 and 5 km AGL (Kain et al. 2008a; Clark et al. 2013). The 50-member ensemble forecasts are used to generate grid point probabilistic forecasts (i.e. ensemble frequencies) of strong rotation, as in Clark et al. (2012b; 2013) where updraft helicity exceeds the 100 m^2 s^{-2} threshold. To account for point displacement errors and ensemble variability, a "neighborhood approach" (e.g. Ebert 2008; Romine et al. 2013) is applied to the probabilities; a three by three grid point stencil is specified around each grid box and a hit at any grid point in the stencil is counted as a hit for that grid box.

In addition to updraft helicity, the presence of significant low-level rotation (probability of near surface vorticity tracks) can be used as a proxy for tornadic potential because it is indicative of a strong low-level mesocyclones in model forecasts (see Section 3.8.3; Stensrud and Gao 2010; Dawson et al. 2012; Stensrud et al. 2013; Yussouf et al. 2013). The surface vorticity tracks serve only to indicate where strong low-level mesoscyclones are forecast. The magnitude of the surface vorticity in model forecasts at 3 km grid spacing is not expected to be correlated with tornado strength or even indicate whether a tornado exists (Markowski et al. 2011, Marquis et al. 2012), but may indicate a model forecast storm with enhanced risk to produce a tornado event. Herein, the maximum vorticity below 1 km AGL is found at every grid point, for each ensemble member forecast. Then the neighborhood probability for maximum 0-1 km vorticity exceeding 0.0025 s^{-1} is computed in the same manner as the updraft helicity probabilities. The surface vorticity probabilities are used as an additional forecast metric to differentiate between forecast experiments that contain similar updraft helicity probabilities (Section 4.8).

4.4 Experiment overview

This section overviews all of the data assimilation experiments to provide a holistic view of the research. The experiments start with a mesoscale assimilation approach, and proceed towards storm-scale assimilation techniques.

The first experiment is simple in that it performs mesoscale data assimilation (hereafter referred to as "MesoOnly"; Table 4, row 1). Herein, mesoscale data assimilation is defined as assimilating conventional observations, every 6-hours, on a domain with 15 km horizontal grid spacing (Domain 1, Figure 22). A CAM (3 km horizontal grid spacing; Domain 2, Figure 22) is not included in the data assimilation system. The complex near-storm environment preceding the Booker storm is not well represented in the ensemble analysis using this mesoscale configuration (Section 4.5). When the MesosOnly analysis is downscaled to CAM, subsequent forecasts exhibit notable shortcomings in storm development and intensity relative to the observed thunderstorm event (Section 4.5). The use of more sophisticated data assimilation strategies is expected to yield more accurate analyses and improved forecast skill. Thus, the following experiments explore how to improve the analysis and forecasts of the 13 June event with multi-scale data assimilation.

The remaining experiments are separated into three sets of multi-scale data assimilation experiments, and each set investigates a research question regarding the development of a multi-scale data assimilation and forecast system. The first experiment set investigates the potential value of more frequent data assimilation cycling (Section 4.6); hourly cycling is compared to 6-hour analysis cycles (6Hourly), which are typically used in mesoscale analysis systems (Table 4, red and blue rows). In addition, the lead-time for hourly cycles is evaluated (HourlyFrom12utc, HourlyFrom0utc) to examine the differences in the overnight convection preceding the Booker storm, which played a vital role in setting up the convective environment. The second experiment set investigates the use of Doppler radar observations with conventional observations (Table 4, red and purple rows). Previous studies have typically focused on either the assimilation of conventional observations or the assimilation of radar observations (Section 4.1.2). This study determines the impact of radar observations on the mesoscale environment and the convective forecast by comparing an experiment in which both radar and conventional observations are assimilated (RDA) to an experiment in which only conventional observations are assimilated (DA). Further, an experiment where radar data assimilation is limited to observations of the storms of interest (AddRDA) provides additional insight on the impact of the radar observations on the environment. The radar experiment set is discussed in Section 4.7. The final experiment set investigates the sensitivity to the background ensemble initialization (Table 4, red and green rows). A simple approach to ensemble initialization, initializing the ensemble from a global analysis on the day of interest (DayOfStart) is compared to the use of a continuously cycled system for ensemble initialization, motivated by the success of real-time continuously cycled analysis systems (Romine et al. 2013; Schwartz et al. 2014). In addition, the impact of model resolution on the continuous cycling is investigated in Section 4.8 (Cycled15Add3, Cycled).

Table 4. Data assimilation experiments.

Experiment name	Starts From	Cycling Frequency	Assimilated Observations	DA on Domain
MesoOnly	8 June - 1800 UTC GFS analysis	6 hours	Conventional	1
6Hourly	8 June - 1800 UTC GFS analysis	6 hours	Conventional, + Radar on 13 June	1 & 2
HourlyFrom12utc	13 June - 1200 UTC 6Hourly forecast	1 hour	Conventional + Radar	1 & 2
HourlyFrom0utc/ RDA/Cycled	13 June - 0000 UTC 6Hourly forecast	1 hour	Conventional + Radar	1 & 2
DA	13 June - 0000 UTC 6Hourly forecast	1 hour	Conventional	1 & 2
AddRDA	13 June - 1800 UTC DA forecast	1 hour	Conventional + Radar	1 & 2
Cycled15Add3	D1: 8 June - 1800 UTC GFS analysis D2: downscaled from D1 at 12 June - 1800 UTC	6 hours 1 hour on 13 June	Conventional, + Radar on 13 June	1, 1 & 2
DayOfStart	12 June - 1800 UTC GFS analysis	1 hour	Conventional + Radar	1 & 2

The three experiment sets are conducted in parallel, because in order to carry out an experiment set that tests one data assimilation parameter, all other parameters are held constant, including the parameters that are tested in the other experiment sets. For example, cycling frequency cannot be tested without a choosing which observations to assimilate. The default data assimilation parameter choices for initialization, cycling frequency, and observation type are based on previous studies (Section 4.1.2) and intuition about which setting might give the best convective forecast performance. The red highlighted row in Table 4 show the default choices. A continuously cycled analysis is used as the default initialization. Hourly assimilation cycles is the default cycling frequency, and both radar and conventional observations are the default observations.

4.5 Mesoscale data assimilation

The skill of a mesoscale data assimilation system is investigated, to further verify the need for a multi-scale data assimilation system being developed in this chapter. As discussed in Section 4.1.2 and 4.4, current operational data assimilation systems use mesoscale grid spacing (~15-20 km horizontal grid spacing), and the analyses are often downscaled to CAMs for real-time convective forecast applications. In this section, results from data assimilation on a WRF domain defined at mesoscale (cumulus-parameterizing resolution) are discussed, and downscaled CAM ensemble forecasts are examined.

The MesoOnly analysis is on a 15 km horizontally spaced grid, and is updated every 6-hours with conventional observations using continuous EAKF cycling from 1800 UTC on 8 June through 1800 UTC on 13 June. The 1200 UTC mean analysis captures the observed synoptic pattern (Figure 13), such as the 500 mb trough position and orientation across the western U.S. (Figure 28). The eastward progression of this trough from 1200 UTC on 13 June to 0000 UTC on 14 June provides weak ascent over the Oklahoma-Texas Panhandle region and contributes to favorable deep layer shear for organized thunderstorms in that region. The 1200 UTC mid-level wind analysis above the Oklahoma-Texas Panhandle indicates a uniform 40 knots of southwesterly flow similar to that observed by the 1200 UTC sounding released from Amarillo, Texas.

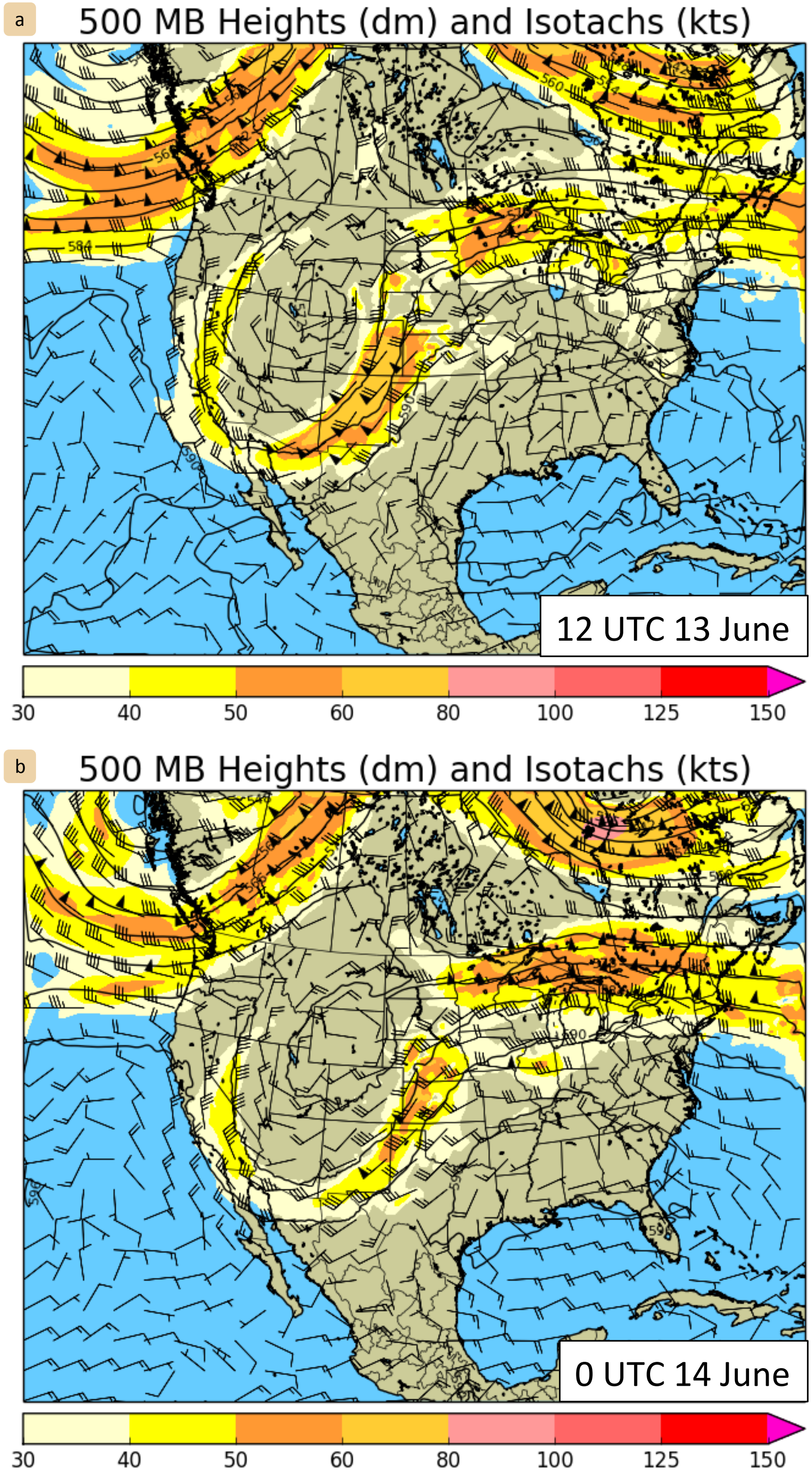


Figure 28. The 500 mb heights and isotachs at 1200 UTC on 13 June 2010 (a) and 0000 UTC on 14 June 2010 (b).

By 1800 UTC, the mean MesoOnly surface analysis includes the stationary front across the target region (Figure 29a), but it is further northwest relative to the observed frontal location (Figure 29b). The observed outflow boundary from overnight convection is not well resolved in the 1800 UTC analysis, although the temperature analysis is cooler (~70 °F) in a small region just north of the Oklahoma-Kansas border (approximately located at latitude 37.5° and longitude -99.5°) and in Central Kansas (approximately located at latitude 38.5° and longitude -98.5°). Still, the cold pool in the analysis was considerably smaller in spatial extent and weaker in intensity relative to the observed cold pool and outflow boundary. Summarizing, the 6-hourly cycled MesoOnly analysis contains some representation of the core elements believed to be important for this event, but these key features are spatially displaced and contained weaker gradients relative to the observations. Thus, the near-storm environment for the Booker storm is not well represented.

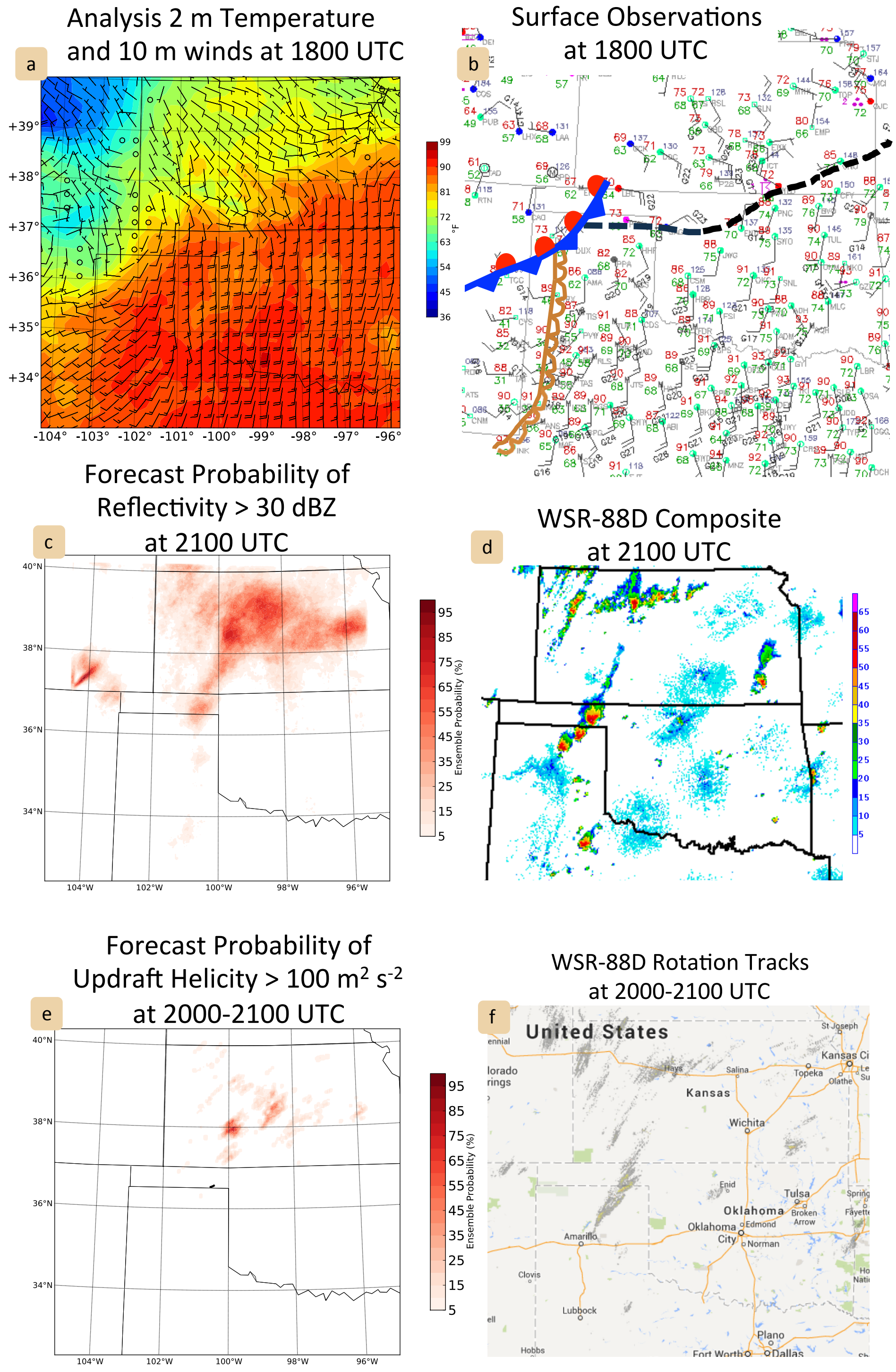


Figure 29. The 2 m temperature MesoOnly analysis (a) and observations (b) at 1800 UTC. The MesoOnly 2100 UTC forecasted ensemble probability of reflectivity exceeding 30 dBZ (c), and the observed WSR-88D reflectivity (d). The MesoOnly 2000-2100 UTC forecasted ensemble probability of updraft helicity exceeding 100 $m^2 s^{-2}$ (e), and the WSR-88D derived rotation tracks from NSSL (f).

To further evaluate the storm environment provided by the MesoOnly analysis, the 13 June 1800 UTC ensemble analysis is downscaled to an inner WRF domain (3 km horizontal grid spacing; See Figure 22) to initialize CAM forecasts. The 50-member ensemble forecasts with the inner domain are integrated for 3-hours (2100 UTC) and compared to the observed radar reflectivity valid at the forecast time. The ensemble forecasts indicate convection developing along the stationary front from Kansas into the Oklahoma Panhandle (Figure 29c). Ensemble probabilities of reflectivity greater than 30 dBZ indicate that the ensemble forecast generated from downscaling the MesoOnly analysis has a high probability of convection in central Kansas, where little was observed, with only low probabilities of storms in the Texas Panhandle region, where the most intense storms were observed at this time (Figure 29d). Ensemble probabilities for updraft helicity exceeding 100 m^2 s^{-2} indicate rotating storms in the forecast (Figure 29e), but displaced well northeast of the observed Booker storm rotation (Figure 29f). These rotating storms are associated with the cold front and the outflow boundary in Central Kansas.

The CAM forecasts initialized from the mesoscale domain provide general guidance for the potential for organized severe storms in the region, but are not sufficient for capturing the observed evolution of storms in the target region. These results motivate the need for more accurate ensemble initial conditions, which is the goal of the multi-scale assimilation approaches that follow. The MesoOnly downscaled forecasts discussed here will serve as a control and provide a baseline for the assimilation experiments presented in the following sections.

4.6 Data assimilation cycling frequency

To improve the deficiencies of the convective forecasts initialized with the MesoOnly analysis in the previous section, a CAM grid is included in the data assimilation cycling and more frequent observations are provided to the analysis system. The goal of this section is to investigate the impact of the cycling frequency on the storm environment and convective forecasts with a combined mesoscale and CAM analysis system.

To investigate the impacts of cycling frequency, three experiments are compared. The first experiment mimics the cycling frequency of mesoscale analysis systems; conventional, mesonet, and radar observations are assimilated every 6-hours on 13 June, using a mesoscale WRF domain (15 km horizontal grid spacing) and a nested WRF domain (3 km horizontal grid spacing; Figure 22). It is referred to as the "6Hourly" experiment (Figure 30). The analysis and forecast results discussed below indicate deficiencies with the 6Hourly multi-scale cycling. Thus, the potential value of hourly analysis cycles is examined. Hourly cycles from 1200 to 1800 UTC are added to the 6Hourly experiment, which is referred to as "HourlyFrom12utc" (Figure 30). HourlyFrom12utc is designed to investigate the advantages or disadvantages of starting hourly assimilation cycling in the morning of the day of interest. Since the overnight convection preceding the Booker event played a vital role in setting up the convective environment, the HourlyFrom12utc forecasts also have limited skill. Thus, the third cycling frequency experiment begins hourly cycling at 0000 UTC on 13 June, "HourlyFrom0utc" (Figure 30).

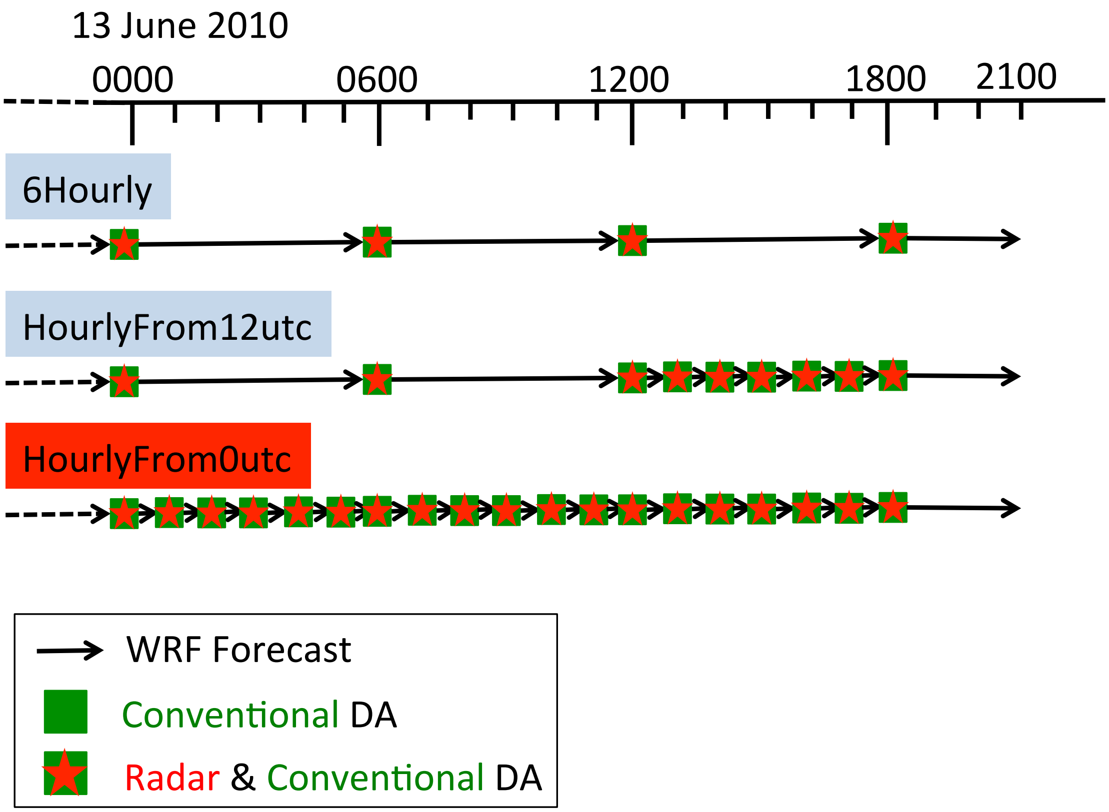


Figure 30. Data assimilation cycle timeline on 13 June. Times are in UTC. The experiment name textboxes are colored to match the Table 4 (experiment overview) row colors.

The system design and data assimilation cycling methodology will be identical in each experiment (see Section 4.3.5; Figure 25), with the exception of the number of observations assimilated. Both experiments are initiated from the same background ensemble forecast (the initial ensemble will be discussed in Section 4.8), use the same boundary conditions, and employ the same model physics. Data assimilation is performed on both the outer WRF domain (15 km horizontal grid spacing) and the nested WRF domain (3 km horizontal grid spacing). The conventional observations for the 6Hourly experiment are selected using a 90-minute window centered at the assimilation time, which allows more observations to be included for one cycle compared to the 30-minute window on either side of the assimilation time for hourly assimilation. However, the mesonet observations for both 6-hour and hourly are the same; a 15-minute window on either side of the assimilation time is used for the mesonet data. One volume of radar observations from the four WSR-88Ds is assimilated during each assimilation cycle of the experiments.

4.6.1 Analysis results

First, the analysis of the overnight convection on 13 June is examined. The prior mean reflectivity at 0600 and 1200 UTC are plotted in Figure 31. During the overnight hours the 6Hourly reflectivity priors appear smoother than the HourlyFrom0utc priors. At 0600 UTC, the high reflectivities (>45 dBZ) in the mean 6Hourly prior do not match the location of the high reflectivity cores in the observed storms. The 6Hourly prior also does not capture the broken line convective mode. The area that is covered by low reflectivities (< 20 dBZ) in the 6Hourly prior does generally match the observed storms, with the exception of the western extent of the reflectivities in the Texas Panhandle and into New Mexico. At 1200 UTC, the 6Hourly forecasted reflectivity is much less intense as the observed MCSs and does not represent the convective mode. Further, the 6Hourly mean prior reflectivities do not extend into the Oklahoma Panhandle. The HourlyFrom0utc mean reflectivity prior qualitatively matches the observations better than the 6Hourly prior.

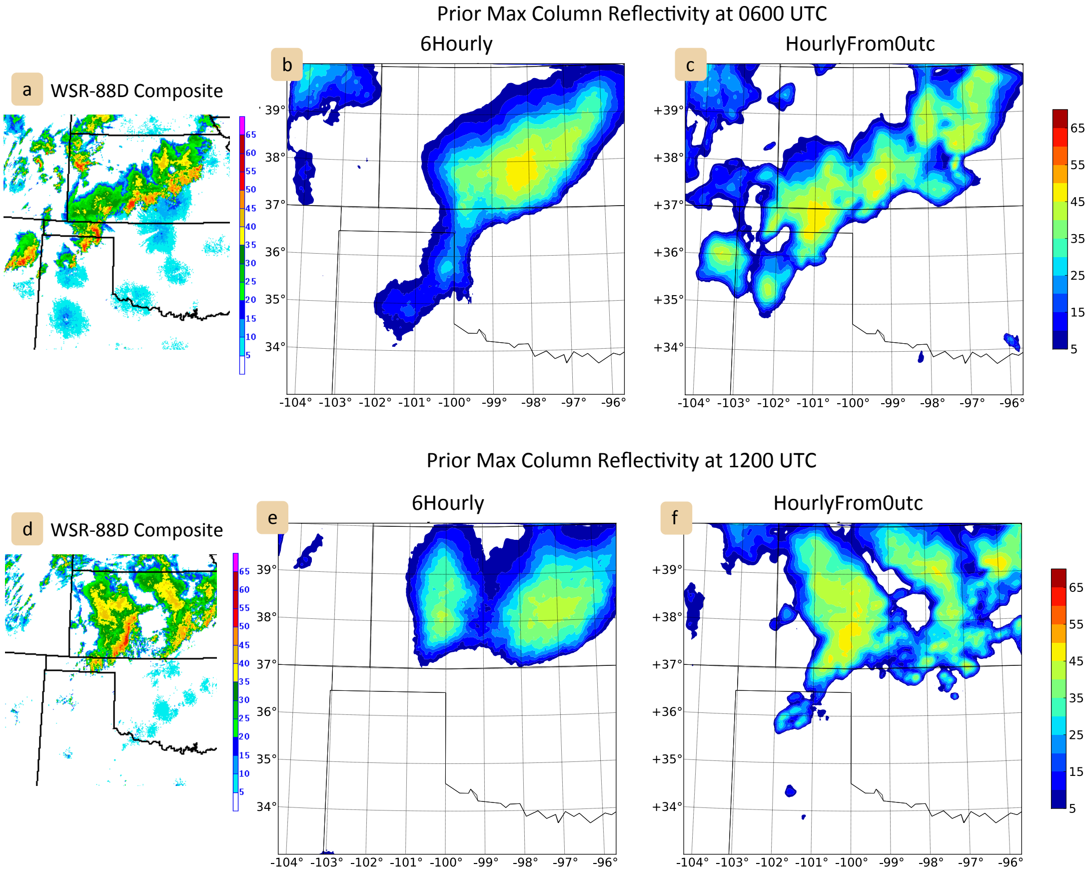


Figure 31. WSR-88D composite reflectivity at 0600 UTC (a) and 1200 UTC (b). Mean prior maximum column reflectivity at 0600 UTC and 1200 UTC for 6Hourly (b,e) and HourlyFrom0utc (c,f) experiments.

Despite the reflectivity errors, the 6Hourly analysis fit to the surface temperature observations is only slightly worse than the hourly data assimilation (Figure 32a,b). The 6Hourly surface temperature analysis at 1200 UTC does not contain the cold pool in Southwest Kansas that is present in the HourlyFrom0utc analysis, but the depiction of the cold front is similar in both analyses. Both experiments have a cool bias across the body of Oklahoma and Texas (dots are warmer colors than the contours). This could be caused by the MYJ BPL scheme, which has been found to have a cool bias in previous studies (Kain et al. 2005; Weisman et al. 2008). It could also be caused by a under estimation of the morning heating from solar radiation. The 6Hourly surface dew point analysis errors are slightly smaller than the HourlyFrom0utc experiment (Figure 32c,d). Although, higher dew points (e.g. 70 °F isodrosotherm) do not extend as far to the west in Western Oklahoma and the Texas Panhandle in the 6Hourly analysis as in the HourlyFrom0utc analysis, which is important for the development of afternoon convection in the Panhandle.

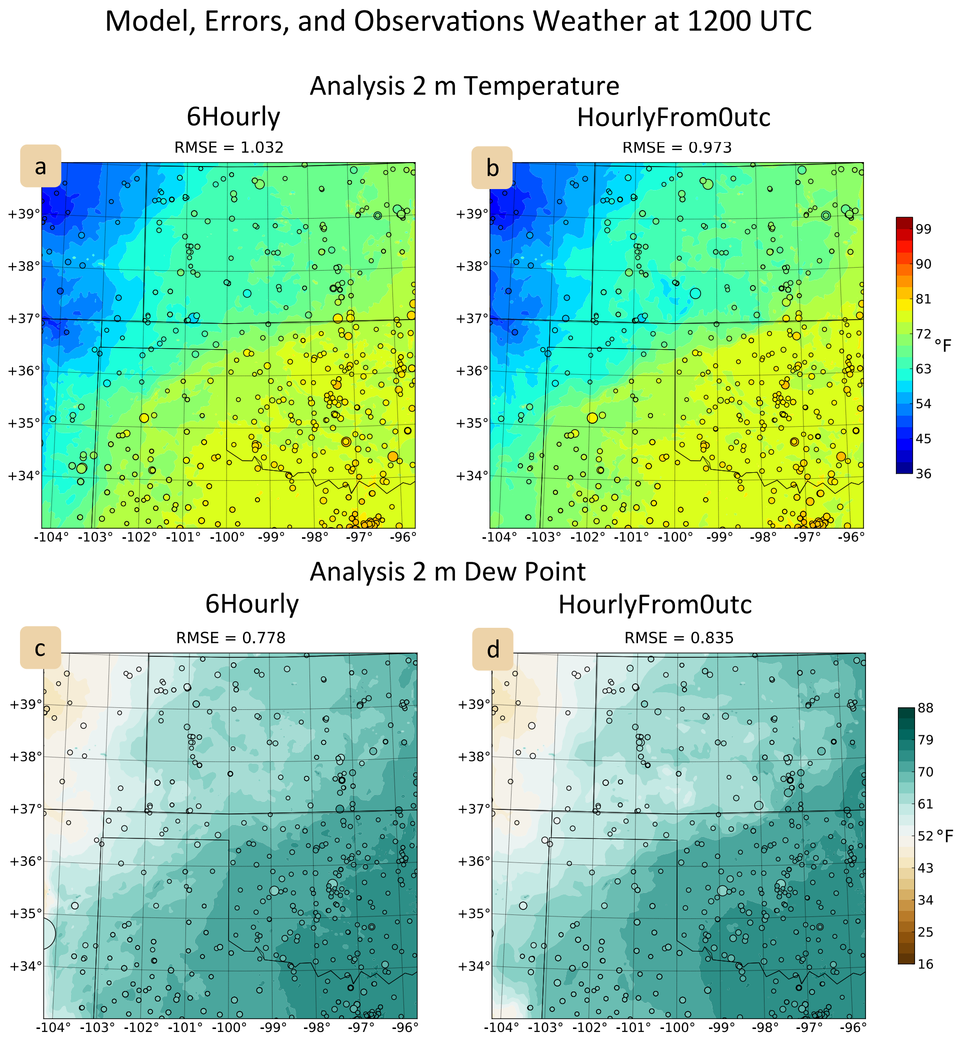


Figure 32. Model, Errors, and Observations Weather at 1200 UTC. The contours are the analysis mean surface temperature for 6Hourly (a) and HourlyFrom0utc (b), and the analysis mean surface dew point for 6Hourly (c) and HourlyFrom0utc (d). The dot

sizes are the model errors. The dot colors are the temperature or dew point observations.

The surface analysis differences that exist at 1200 UTC are smaller than expected, given that only three assimilation cycles have occurred in the 6Hourly experiment (0000,0600,1200 UTC) compared to 13 assimilation cycles (0000-1200 UTC) in the HourlyFrom0utc experiment. This result can be further examined by comparing the data assimilation increments and ensemble spread. An increment is the prior subtracted from the posterior. The surface temperature increments in the 6Hourly experiment are larger than the HourlyFrom0utc experiment (Figure 33). Specifically, at 1200 UTC the total temperature increments in 6Hourly have greater magnitude, cover larger spatial area, and are less noisy (Figure 33d,f). The larger increments are moving the state closer to the observations, which improves the analysis surface fit in Figure 32. The larger increments are partially explained by the larger surface spread (Figure 34a,b). Since there is little surface temperature spread in the HourlyFrom0utc experiment, the increments are limited compared to the 6Hourly increments. While the ensemble surface temperatures have less spread in the HourlyFrom0utc analyses, above the surface the spread is much larger. The column mean and column maximum perturbation potential temperature ensemble spread show that the HourlyFrom0utc experiment contains more spread throughout the domain than in the 6Hourly experiment (Figure 34c-f). This indicates that the hourly data assimilation cycles do not collapse the ensemble spread through the depth of the domain. But the hourly data assimilation does limit the surface spread. This is due to the relatively large number of surface observations compared to the number of observations above the surface. The adaptive inflation used to maintain spread contains much larger values at the surface, and throughout the domain, in the 6Hourly experiment than in the HourlyFrom0utc experiment (Figure 35). Thus, the selected inflation settings seem to be more appropriate for 6-hourly data assimilation cycles. The adaptive inflation employed here was developed and tested with 6-hourly cycles on global grids and the lack of ensemble spread is a known problem (Anderson 2009; Romine et al. 2013; Schwartz et al. 2014). Thus, the inflation technique needs to be more carefully tuned for hourly assimilation or new techniques need to be employed for ensemble spread maintenance. This is outside the scope of the present work.

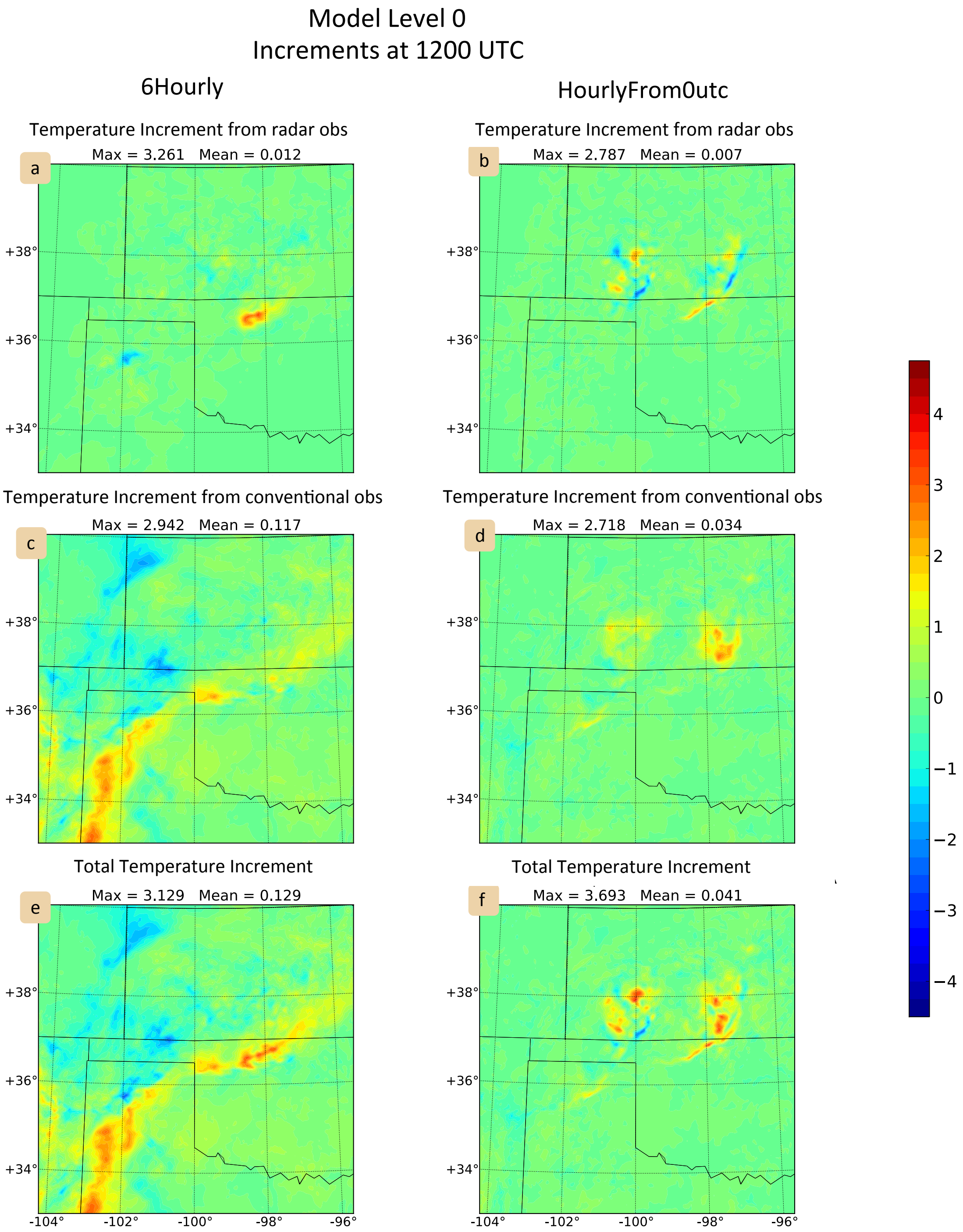


Figure 33. Temperature increments for the lowest model layer at 1200 UTC, after the radar observations are assimilated for 6Hourly (a) and HourlyFrom0utc (b), after the conventional observations are assimilated for 6Hourly (c) and HourlyFrom0utc (d), and the total increment after all observations are assimilated for 6Hourly (e) and HourlyFrom0utc (f).

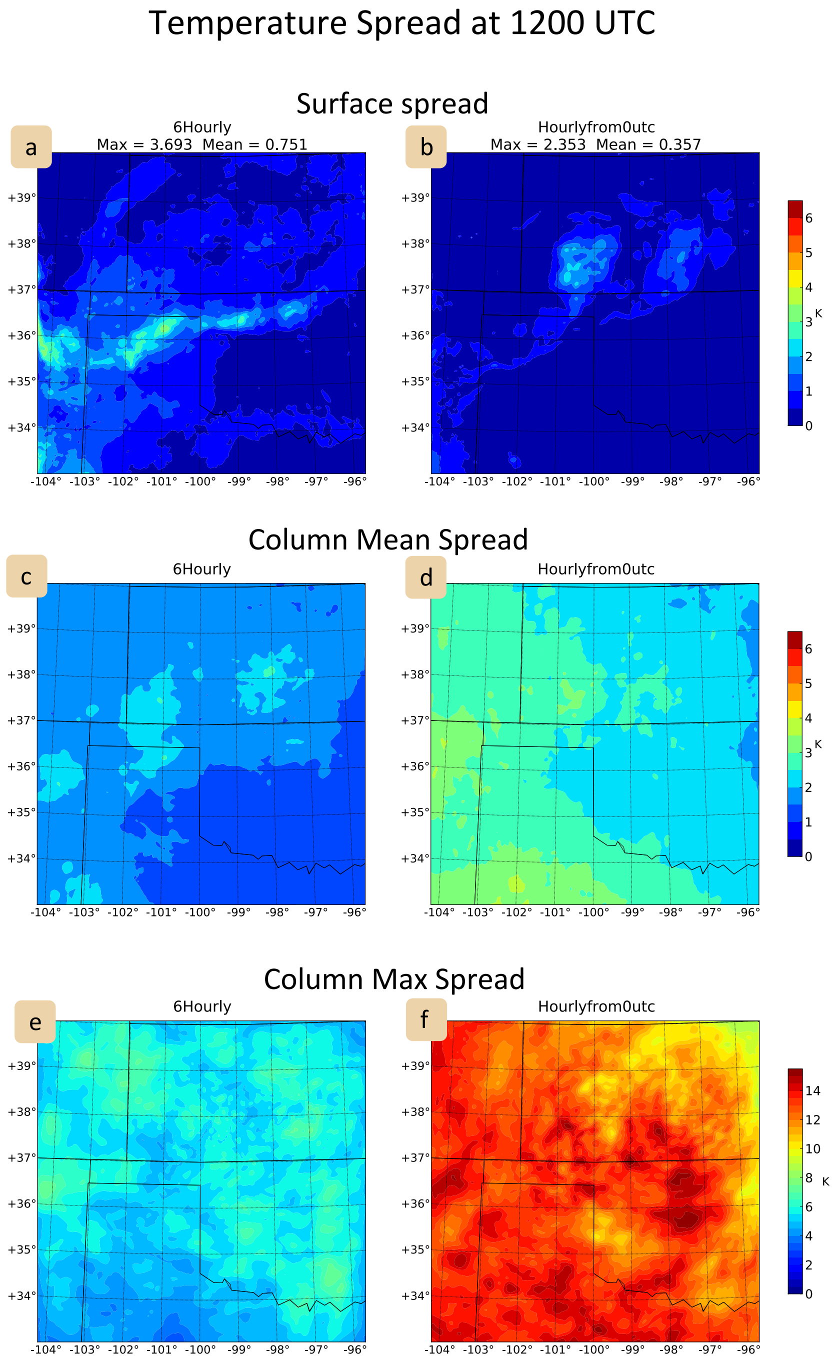


Figure 34. Temperature spread at 1200 UTC at the surface for 6Hourly (a) and HourlyFrom0utc (b). Vertical column mean temperature spread for 6Hourly (c) and HourlyFrom0utc (d). Vertical column maximum temperature spread for 6Hourly (e) and HourlyFrom0utc (f).

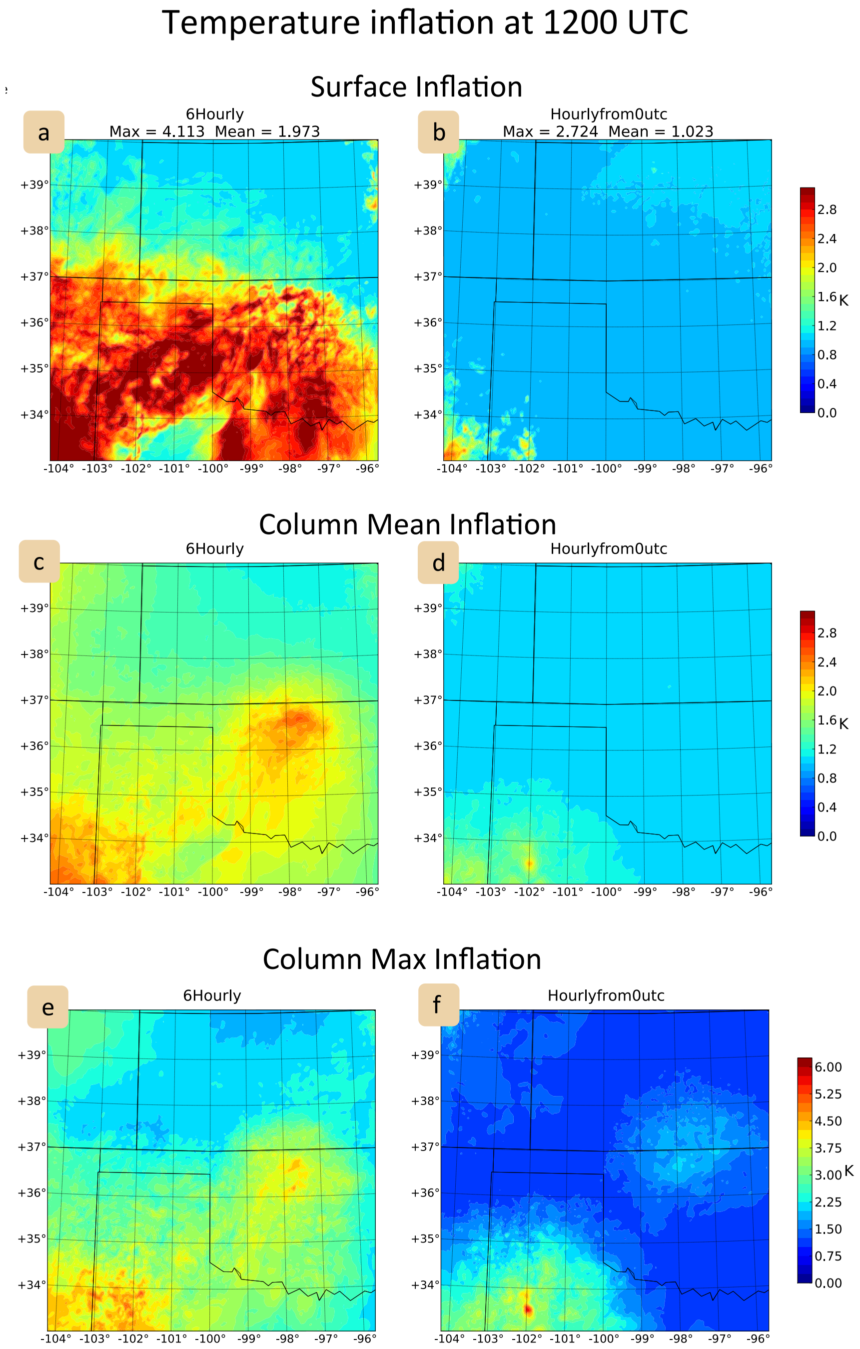


Figure 35. Temperature inflation at 1200 UTC at the surface for 6Hourly (a) and HourlyFrom0utc (b). Vertical column mean temperature inflation for 6Hourly (c) and HourlyFrom0utc (d). Vertical column maximum temperature inflation for 6Hourly (e) and HourlyFrom0utc (f).

The analysis fit to the radiosondes observations is another measure of the data assimilation performance. The analysis fit to radiosonde temperature across all of the available 1200 UTC sounding temperature observations is shown in Figure 36a,b. The 6Hourly and HourlyFrom0utc have similar RMSE values and appear to fit the observations with similar accuracy, which indicates good agreement with the large-scale temperature pattern. The analysis fit to radiosonde temperature observations within the interior domain is shown in Figure 36c,d. The HourlyFrom0utc temperature analysis has a lower RMSE than the 6Hourly analysis, which indicates the hourly cycling improves the temperature profiles in the region of interest. HourlyFrom0utc also has lower RMSE for the interior domain analysis fit to radiosonde dew point, u wind component, and v wind component (Table 5). Thus, within the convective region of interest the HourlyFrom0utc analysis has an improved mesoscale environment.

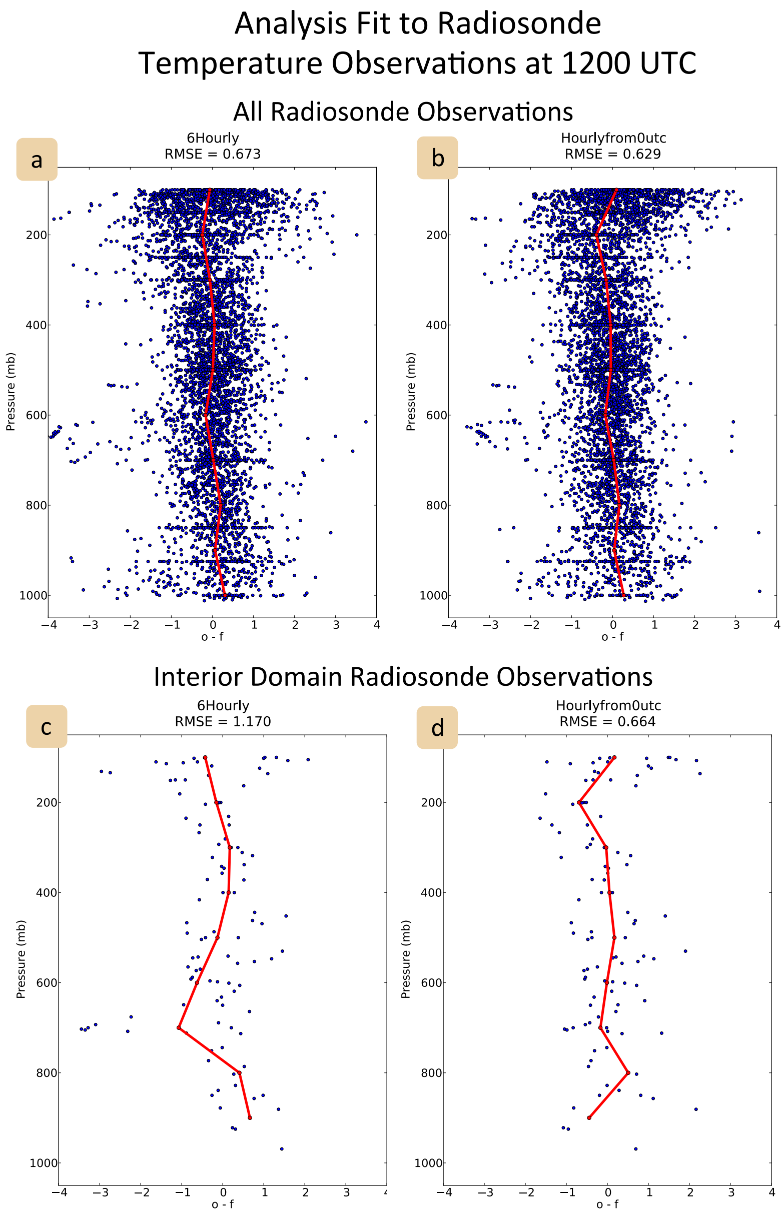


Figure 36. Analysis fit to Radiosonde temperature observations. Each blue marker is an observation minus analysis value and the red line is the mean. The RMSE is shown above the each panel. All observatiosn are plotted for 6Hourly (a) and HourlyFrom0utc (b). Only the observation within the interior domain for 6Hourly (c) and HourlyFrom0utc (d).

Table 5. RMSE values for analysis fit to Radiosonde observations in the interior domain at 1200 UTC.

Variable	6Hourly	HourlyFrom0utc
Temperature	1.170	0.664
Dew point	6.883	5.308
u wind component	1.369	0.995
v wind component	11.924	6.511

The final surface temperature and dew point analyses at 1800 UTC are shown in Figure 37. The 6Hourly temperature and dew point analysis RMSE is the lowest. Similar to the comparisons at 1200 UTC, the 6Hourly surface spread is larger (not shown), which leads to larger surface increments (not shown) and the closer analysis fit to the surface observations. The HourlyFrom12utc surface temperature analysis has less RMSE than the HourlyFrom0utc analysis. However, the HourlyFrom12utc has slightly higher RMSE than the 6Hourly analysis. Consistent with the results shown above, these comparisons indicate that the hourly assimilation is decreasing the spread too much at the surface, which leads to an under dispersive ensemble and worse fit. This result is counter-intuitive, because assimilating additional surface observations should improve the analysis.

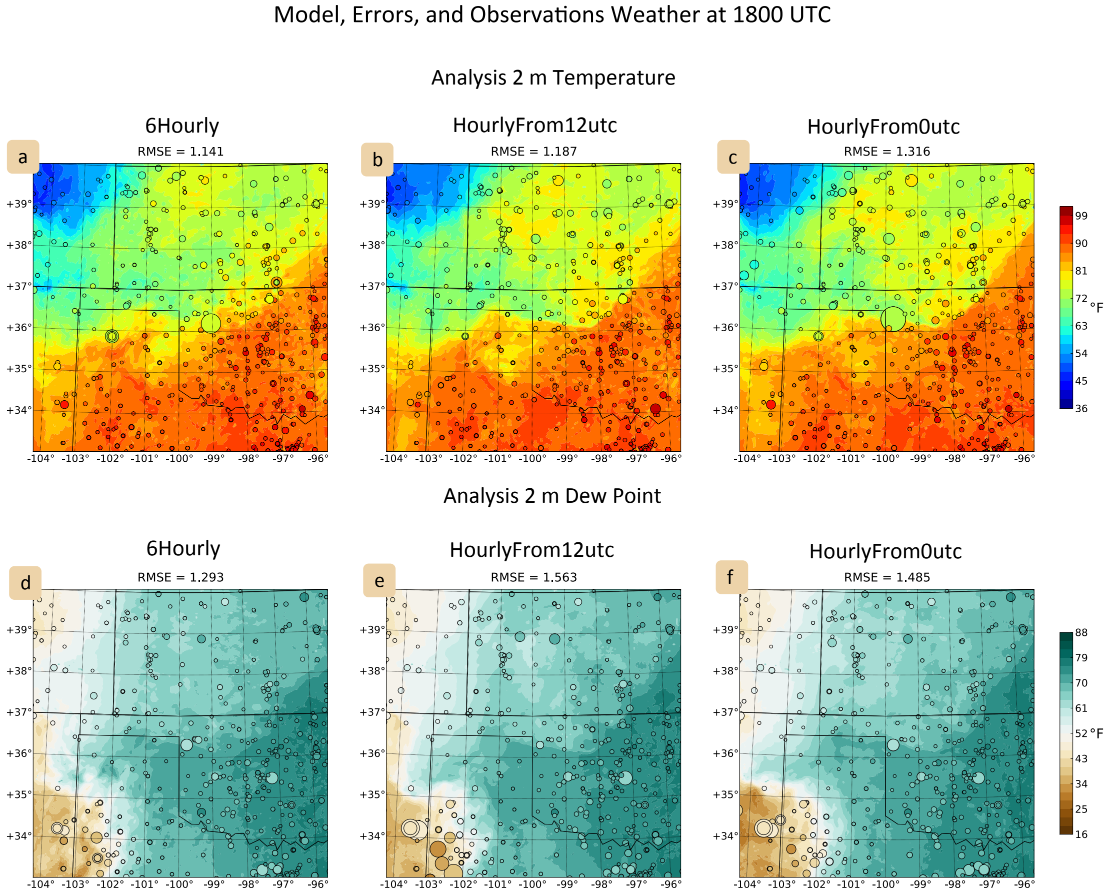


Figure 37. Model, Errors, and Observations Weather at 1800 UTC. The contours are the analysis mean surface temperature for 6Hourly (a), HourlyFrom12utc (b), and HourlyFrom0utc (c), and the analysis mean surface dew point for 6Hourly (d), HourlyFrom12utc (e), and HourlyFrom0utc (f). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

The near surface wind analyses at 1800 UTC contain a wind shift associated with the observed outflow boundary in each experiment (Figure 38). However, in the HourlyFrom0utc analysis the easterly winds in the Texas Panhandle (just south of the Oklahoma border) extend slightly farther west, to the point where they intersect the northerly cold front winds. The horizontal wind RMSE is almost the same for all three experiments.

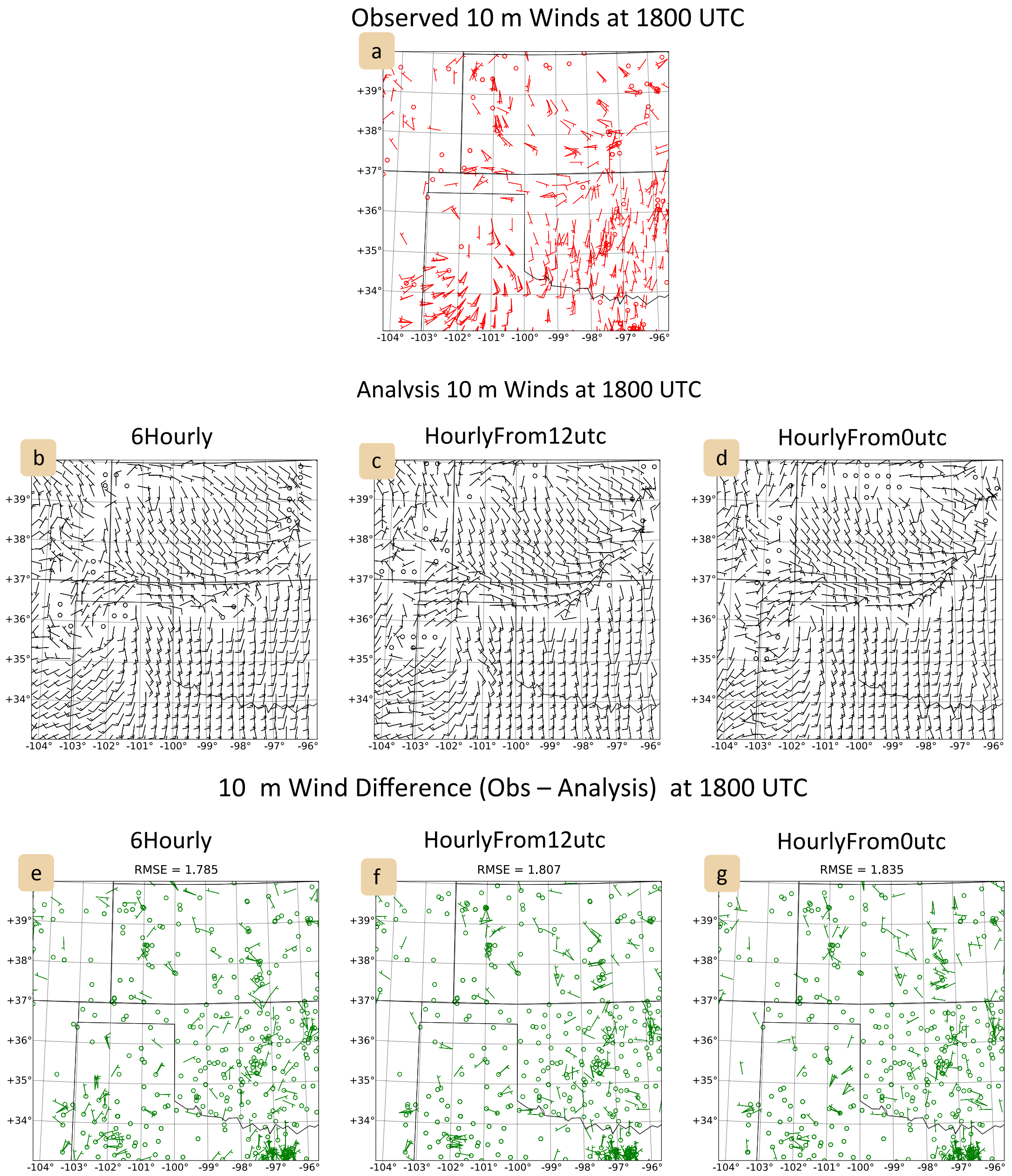


Figure 38. At 1800 UTC, the 10 m wind observations (a), and mean analysis for 6Hourly (b), HourlyFrom12utc (c), and HourlyFrom0utc (d). Half barb = 5 m s^{-1}, full barb = 10 m s^{-1}. The observation minus analysis wind difference for 6Hourly (e), HourlyFrom12utc (f), and Hourly From0utc (g). Half barb = 2.5 m s^{-1}, full barb = 5 m s^{-1}.

The HourlyFrom0utc ensemble mean reflectivity analysis at 1800 UTC matches the observed storm in the Texas Panhandle (Figure 39c,g). However, the 6Hourly and

Hourlyfrom12utc ensemble mean analyses do not contain the developing storm at the correct location (Figure 39b,e&c,f). Thus, despite the worst surface temperature and dew point fit, and similar near surface wind fit, the HourlyFrom0utc analysis has the best reflectivity fit.

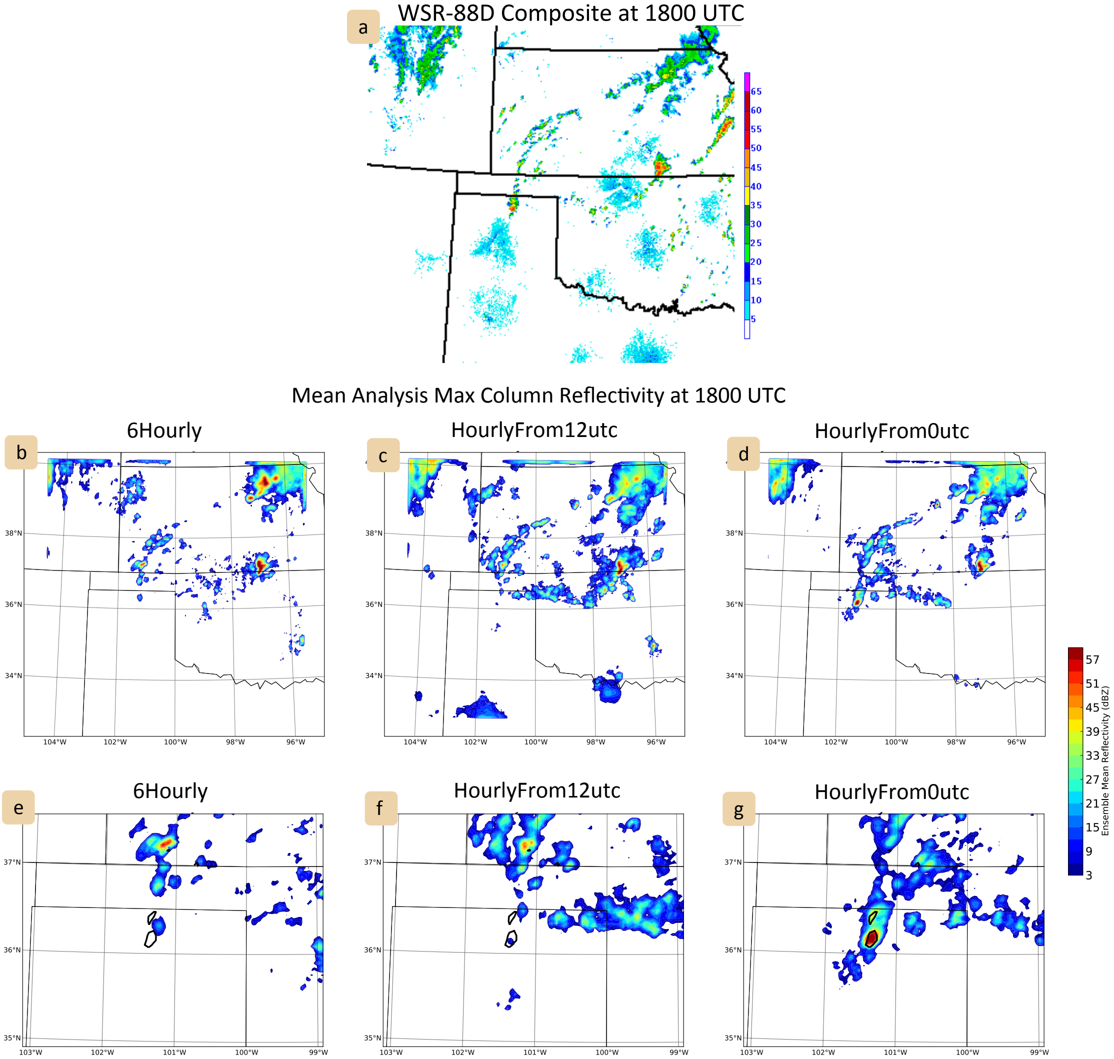


Figure 39. At 1800 UTC, the WSR-88D composite reflectivity (a), and the mean analysis maximum column reflectivity for 6Hourly (b,e), HourlyFrom12utc (c,f), and HourlyFrom0utc (d,g). The KAMA WSR-88D reflectivity above 45 dBZ is contoured in black in the lower panels (e,f,g).

In summary, the hourly assimilation cycles provide more details on the storm structure of the overnight MCSs compared to the 6-hourly cycles. The temperature and dew point analyses of the mesoscale forcing features at the surface are similar for the 6-hourly and hourly cycles, due to differences in surface ensemble spread. However, the experiments with hourly analyses cycles have lower RMSE values when evaluated against the VOTREX2 radiosonde observations. When the hourly cycling begins at 1200 UTC, the storm environment RMSE lies between the 6-hourly and hourly cycles analyses. Finally, the hourly cycling beginning at 0000 UTC is the only experiment containing convection that approximately matches the observed storms at 1800 UTC.

4.6.2 Forecast results

Ensemble forecasts (50 members) are initiated from the analyses at 1800 UTC. First, the forecasted storm environment in the three experiments is evaluated. The MEOW plots (see Section 4.3.7) for 2-hour forecasts indicate that each experiment has similar fit to the surface observations (Figure 40). The 6Hourly forecast has the lowest RMSE for temperature and the HourlyFrom0utc forecast has the lowest RMSE for dew point. The HourlyFrom0utc dew point forecast also qualitatively appears to be the closest match to the observed location of the dry line. The MEOW temperature plot is also used to qualitatively assess the location of the cold air associated with the outflow boundary, which is an important feature of the near-storm environment. The cold air associated with the outflow is farther north in the 6Hourly temperature forecast compared to the HourlyFrom0utc temperature forecast. For example, in the

HourlyFrom0utc surface temperature forecast, the 81 °F isotherm does not extend north of the Oklahoma Panhandle, and the fit to the observations is slightly better than the 6Hourly or HourlyFrom12utc forecast in the Panhandle region (smaller dots). The cold air associated with the outflow is farther south in the HourlyFrom12utc forecast compared to the 6Hourly forecast despite being shifted to the north compared to the observations. Thus, the both experiments with hourly analysis cycles improve the forecasts of the outflow air.

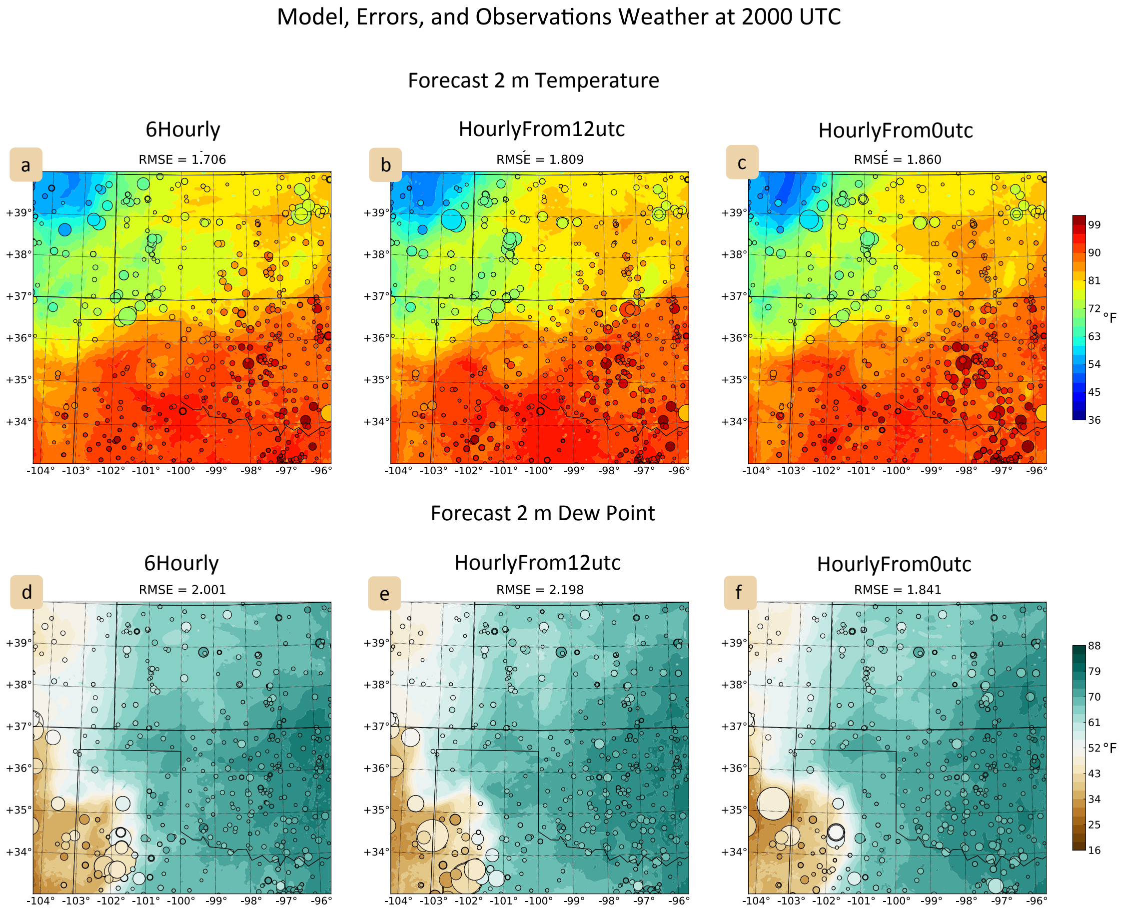


Figure 40. Model, Errors, and Observations Weather at 2000 UTC. The contours are the mean forecast surface temperature for 6Hourly (a), HourlyFrom12utc (b), and HourlyFrom0utc (c), and the mean forecast surface dew point for 6Hourly (d),

HourlyFrom12utc (e), and HourlyFrom0utc (f). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

The forecasted near-surface winds at 2000 UTC (2-hour forecast) are similar in the three experiments (Figure 41). The HourlyFrom12utc forecast has the lowest RMSE and appears to qualitatively match the observed winds the closest. Each of the experiments have a wind shift associated with the outflow boundary. However, the forecasted wind shift is too far to the north compared to the observed wind shift. The 6Hourly wind forecast does not represent the calm winds along stationary front in the Panhandle. Thus, the 6Hourly forecast does not contain the observed boundary intersection. The boundary intersection is present in the HourlyFrom0utc and HourlyFrom12utc wind forecasts. However, the forecasted intersection in both of the hourly experiments is shifted to the northwest of the observed intersection.

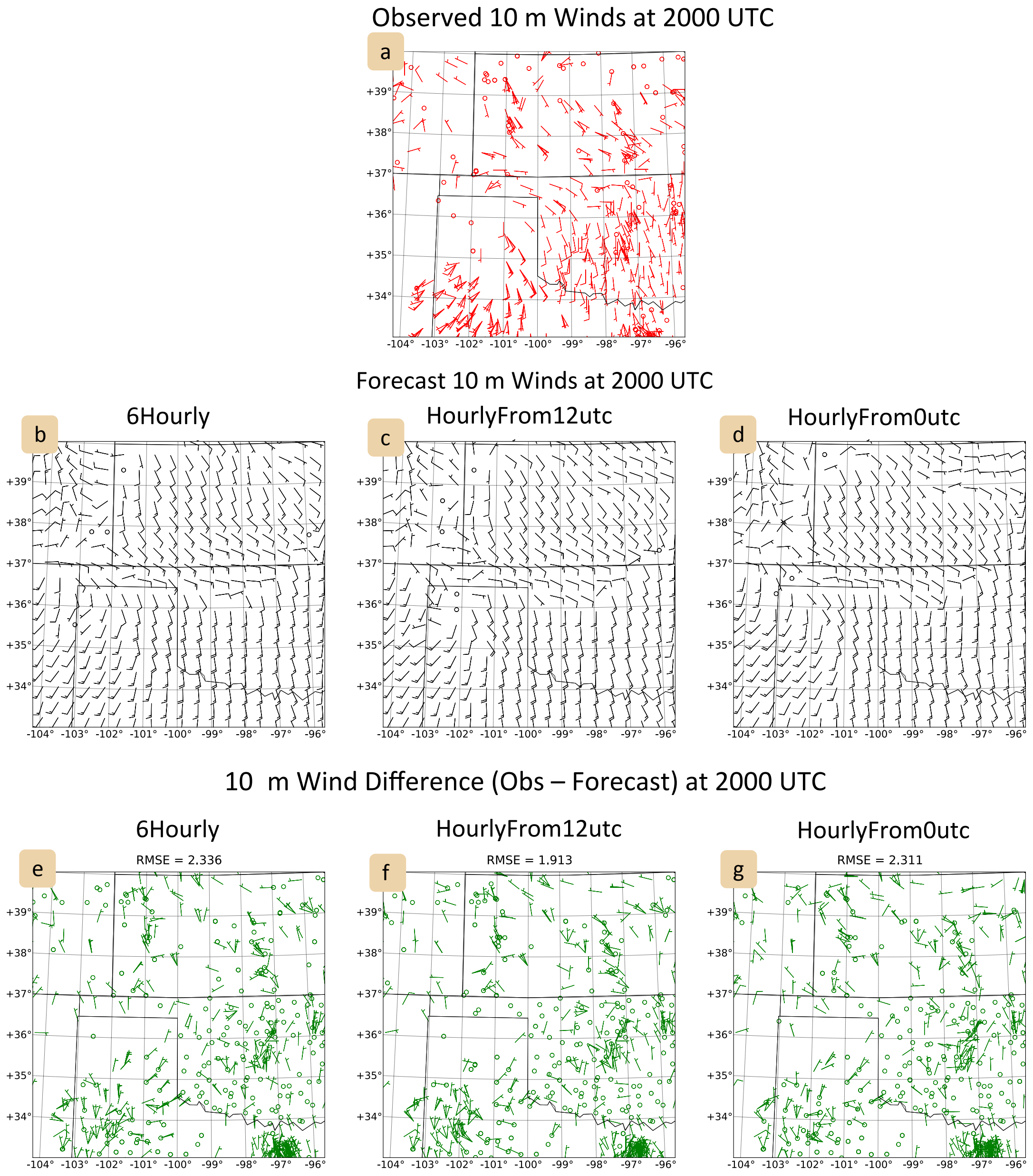


Figure 41. At 2000 UTC, the 10 m wind observations (a), and mean forecast for 6Hourly (b), HourlyFrom12utc (b), and Hourly From0utc (c). Half barb = 5 m s^{-1}, full barb = 10 m s^{-1}. The observation minus forecast wind difference for 6Hourly (e), HourlyFrom12utc (f), and Hourly From0utc (g). Half barb = 2.5 m s^{-1}, full barb = 5 m s^{-1}.

The forecast fit to the VORTEX2 sounding observations (see Section 4.3.6) is used as a measures how accurately the vertical profiles of wind, temperature, and humidity are being forecasted. The HourlyFrom0utc forecast has the lowest RMSE values below 800 mb for temperature and below 500 mb for the v component of the winds (Figure 42a,d). The sounding RMSE values in the mid- and upper-levels is similar among the forecast experiments (Figure 42). The HourlyFrom0utc storm environment forecast is quantitatively more accurate than the HourlyFrom12utc and 6Hourly forecasts.

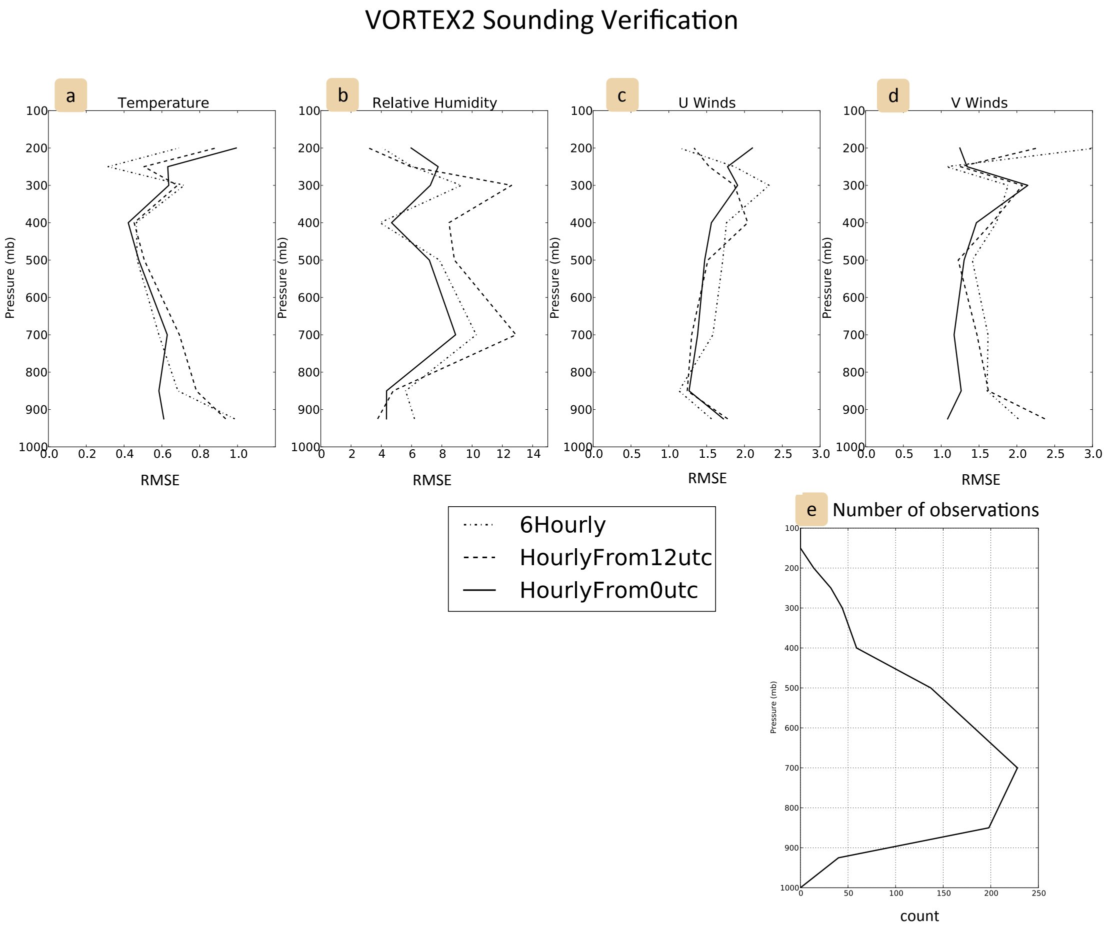


Figure 42. The VORTEX2 sounding temperature RMSE (a), relative humidity RMSE (b), u-wind component RMSE (c), v-wind component RMSE (d), and number of observations (e) for the 6Hourly (dashed-dot), HourlyFrom12utc (dashed), and HourlyFrom0utc (solid) forecasts.

The forecasts at 2100 UTC (3-hour forecast) are compared to investigate how the differences in the near-storm environment forecast affect the location and character of the thunderstorms. The probabilistic reflectivity forecasts at 2100 UTC show that although each of the three experiments' ensemble forecasts have convection close to the observed storms, none of the experiments forecast the location of the observed storms (Figure 43). The HourlyFrom0utc forecasts have ensemble reflectivity probabilities of approximately 25% co-located with the three observed storms and 25% probabilities that extend to the northeast of the observed storms. These forecasted storms evolve from the analyzed storms that were added to the state during the last assimilation cycle. The HourlyFrom0utc forecast also includes higher probabilities, ~60%, to the west of the observed storms, which initiate along the cold front during the forecast. Thus, the HoulryFrom0utc ensemble forecast is over-predicting the spatial extent of the convection. The ensemble reflectivity probabilities in the 6Hourly forecast indicate relatively high confidence of convection, >60% probability for reflectivity above 30 dBZ. However, the forecasted storm probabilities are concentrated to a small area west of the observed Booker storm location. Thus, the 6Hourly forecast does not contain reflectivity probabilities associated with the all of the observed storms; it is under-predicting the spatial extent of the observed convection. The HourlyFrom12utc ensemble reflectivity probabilities cover less spatial area than the HourlyFrom0utc forecast but more area than the 6Houlry forecast. The highest probabilities (near 60%) in the HourlyFrom12utc forecast are also west of the observed storms and there are zero probabilities co-located with the observed Booker storm.

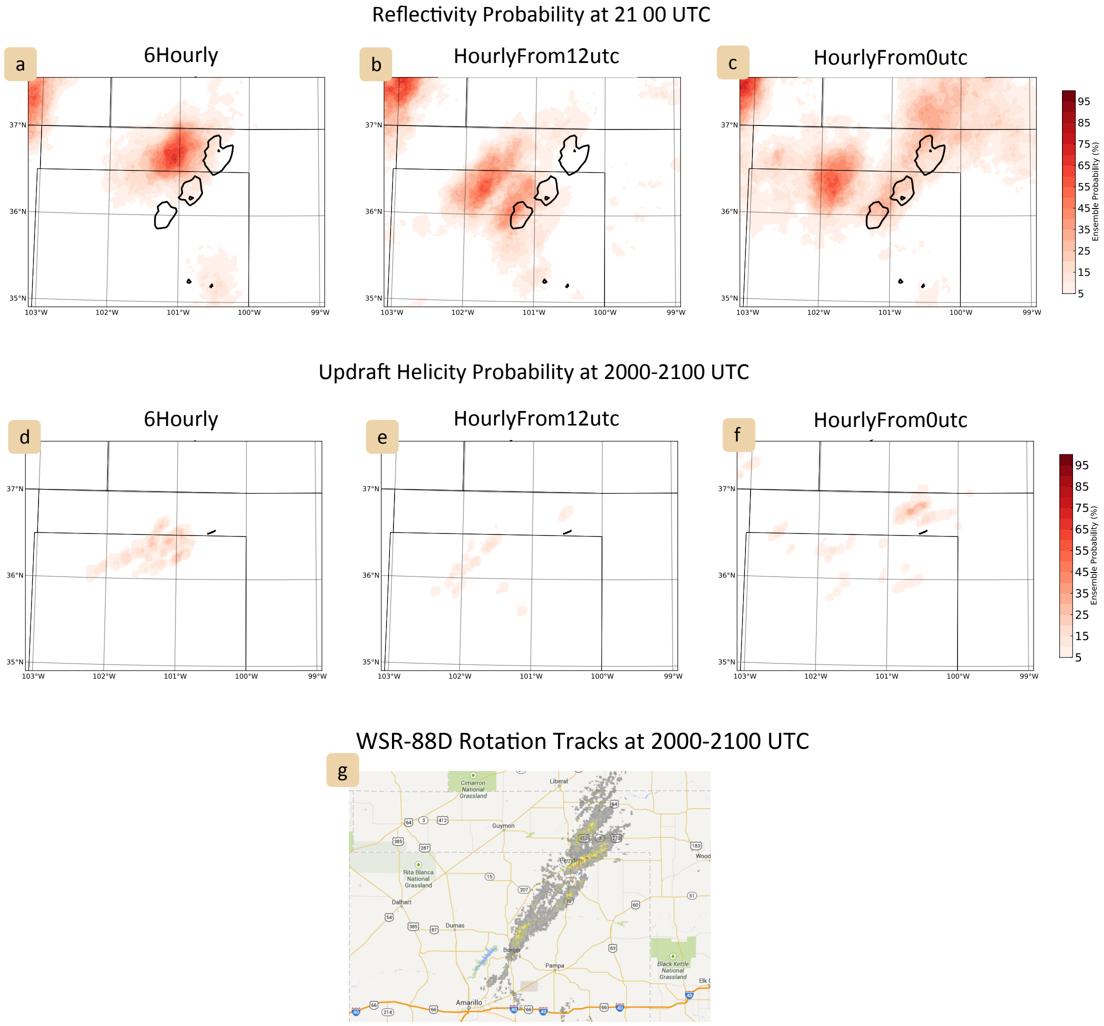


Figure 43. The 2100 UTC forecast ensemble probability of reflectivity greater than 30 dBZ for the 6Hourly (a), HourlyFrom12utc (b), and HourlyFrom0utc (c) forecasts. The KAMA 30, 60 dBZ reflectivity observations are contoured in black. The 2000-2100 UTC forecast ensemble probability of updraft helicity greater than 100 m^2 s^{-2} for the 6Hourly (d), HourlyFrom12utc (e), and HourlyFrom0utc (f) forecasts. The observed tornado track is marked in black. The 2000-2100 UTC WSR-88D derived rotation tracks (g) from the NSSL archive.

In order to determine if the forecasted storms are rotating, the updraft helicity is compared. The forecasted ensemble probability of updraft helicity above 100 m^2 s^{-2} from 2000 UTC to 2100 UTC is plotted in Figure 43d-f. The WSR-88D derived rotation tracks (Smith and Elmore 2004) are shown in Figure 43g for reference. All three experiments contain rotation tracks indicating rotating storms. The HourlyFrom0utc forecast contains the highest probabilities of the three experiments, 30%, within a short swath in the Oklahoma Panhandle. The location of the swath is shifted to the north of the observed Booker tornado track. The scattered nature of the additional low probabilities across the Panhandle region indicate ensemble variability. The 6Hourly forecast has a longer swath of updraft helicity, however it appears to be caused by the front instead of by the boundary intersection. Recall, that the wind forecast in the 6Hourly experiment does not have a boundary intersection. The HourlyFrom12utc updraft helicity probabilities indicate low probability for storm rotation.

In summary, forecasts initialized from HoulryFrom0utc qualitatively matched the observed mesoscale environment better than the HourlyFrom12utc and the 6Hourly forecasts. Further, the HourlyFrom0utc forecasts quantitatively fit the VORTEX2 sounding observations more closely than the HourlyFrom12utc and the 6Hourly forecasts. Despite errors in each of the experiments' forecasts, the HourlyFrom0utc forecast is the closest to forecasting the Booker storm in the correct location and with rotation. The HourlyFrom0utc forecast has more skill, indicating the value of the hourly analysis cycles and the need for hourly cycling during the overnight hours in this case.

4.6.3 Conclusions

The cycling frequency does impact the quality of the analysis for the storm environment. Hourly assimilation cycles improve the fit to radiosonde observations.

The analyses of the mesoscale forcing features at the surface are similar for the hourly and 6-hourly cycles, due to differences in surface ensemble spread. The ensemble spread is larger for the 6-hourly cycles, which allows the increments to be larger and the surface fit to be close to the hourly analyses. Future work should investigate techniques to improve the maintenance of ensemble spread for hourly data assimilation cycles, which will likely lead to better analysis and forecast performance for hourly data assimilation.

When hourly cycling begins at 1200 UTC, the storm environment is similar to the 6-hourly analyses. However, beginning the hourly assimilation cycles from 0000 UTC results in a more accurate representation of the reflectivity structure in the overnight MCSs and improves the location of the outflow boundary. Further, the hourly cycling beginning at 0000 UTC is the only experiment with analyzed storms that match the observed storms at 1800 UTC.

The differences in the storm environment analyses lead to differences in the character of the forecasted storms. Hourly cycling from 0000 UTC produces an ensemble forecast that is closest to predicting the Booker storm location and rotation. Thus, hourly assimilation cycles are used to investigate the impact of radar observations in the following section.

4.7 Hourly radar data assimilation

Due to the success of previous Doppler radar data assimilation studies (see Section 4.1.2), the impact of assimilating radar observations in a multi-scale data assimilation framework is examined in this section. The goal of this section is to determine if radar data assimilation can contribute to creating an accurate storm environment and improve convective forecasts.

To investigate the impact of radar data assimilation on the storm environment and storm forecasts, two data assimilation experiments are compared. The HourlyFrom0utc experiment discussed Section 4.6, which assimilates radar and conventional observations every hour, is hereafter referred to as the "RDA" experiment in this section. The RDA experiment uses the radar and conventional observation cycle described in Section 4.3.5, which includes both the additive noise technique and the adaptive inflation technique to maintain ensemble spread (Figure 25b). One volume of radar observations from the four WSR-88D radars is assimilated each hour in the RDA experiment. The RDA experiment is compared to an experiment that assimilates only conventional observations, hereafter referred to as "DA". The DA experiment uses the conventional observation assimilation cycle described in Section 4.3.5 (Figure 25a).

The difference between the RDA and DA experiments is the inclusion of the radar data assimilation in RDA; otherwise the experiments are identical. Both experiments are initiated from the same background ensemble forecast (the initial ensemble will be discussed in Section 4.8), use the same boundary conditions, and employ the same model physics. Hourly data assimilation is performed on both the outer WRF domain (15 km horizontal grid spacing) and the nested WRF domain (3 km horizontal grid spacing) from 0000 UTC through 1800 UTC on 13 June (Figure 44). The conventional observations used in both experiments are the same, and include mesonet observations.

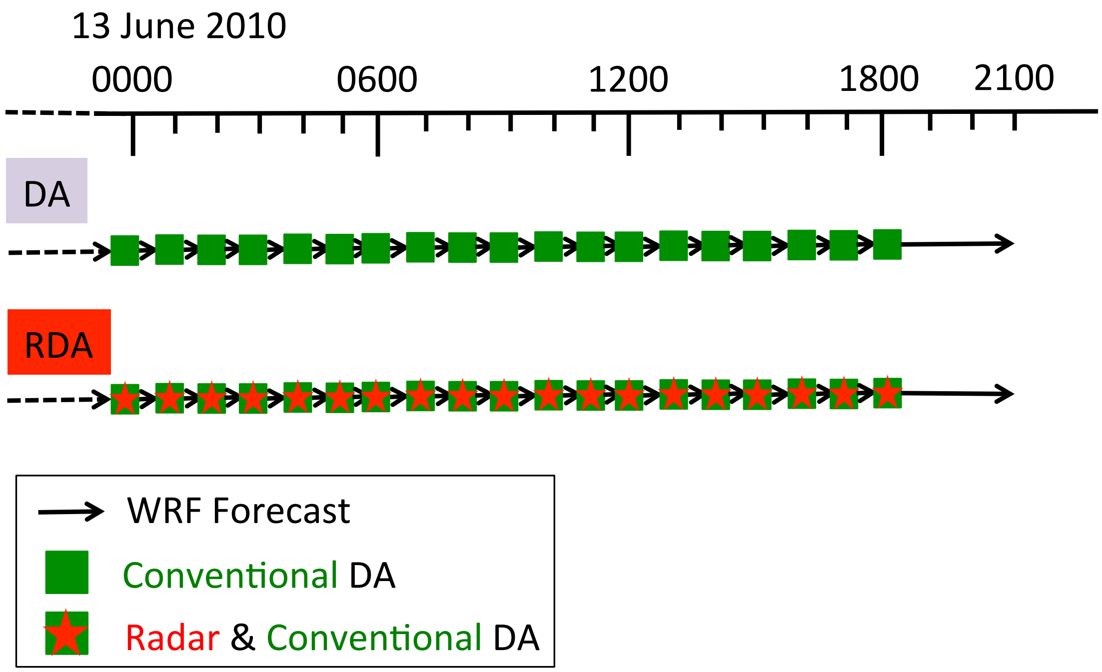


Figure 44. Data assimilation cycle timeline on 13 June. Times are in UTC. The experiment name textboxes are colored to match the Table 4 (experiment overview) row colors. RDA is in red to indicate it is the same experiment discussed in the previous section (HourlyFrom0utc).

A third experiment, in which only WSR-88D observations are assimilated (no conventional observations are assimilated) every hour, was briefly investigated. The mesoscale environment is very poor in terms of the fit to surface and sounding observations and the mesoscale forcing features (front, dry line, outflow boundary) are not well represented. The subsequent forecasts did produce convection however, the location and evolution of the storms does not resemble the observed storms. The results from this experiment indicate that the assimilation of conventional observations is essential and they will not be shown.

4.7.1 Analysis results

Early in the analysis period differences can be seen between the DA and RDA experiments. For example, after 6-hours of data assimilation the RDA reflectivity prior more accurately represents the observed convection compared to the DA prior (Figure 45a-c). Specifically, the RDA experiment is able to forecast the convection close to the observed location, including capturing the stronger intensity convection measured by reflectivity > 40 dBZ, from Northeast New Mexico into the Texas Panhandle and across Kansas. In addition, the RDA reflectivity prior captures the observed convective mode, which is a broken line of heavy convective cells. However, the DA reflectivity prior consists of a slightly bowed line with trailing precipitation. The DA > 40 dBZ reflectivity is also shifted to the southeast of the observed storms. Further differences between the reflectivity priors are evident at 1200 UTC (Figure 45d-f). The RDA mean prior is able to represent the strong MCS in the western half of Kansas with reflectivities > 45 dBZ. However, the DA mean prior has weak storms in Western Kansas with reflectivities of approximately 35 dBZ, and the spatial area of the convection is larger than observed, extending westward into Colorado. Further, the reflectivity associated with the western MCS does not extend into the Oklahoma Panhandle in the DA prior. The RDA mean prior is also able to represent the MCS in Central Kansas with convective cells of 40 dBZ. The DA mean prior of the Central Kansas MCS is too broad in the east-west direction and is not decaying as quickly as observed. Overall, the radar assimilation is improving the model’s representation of the overnight convection.

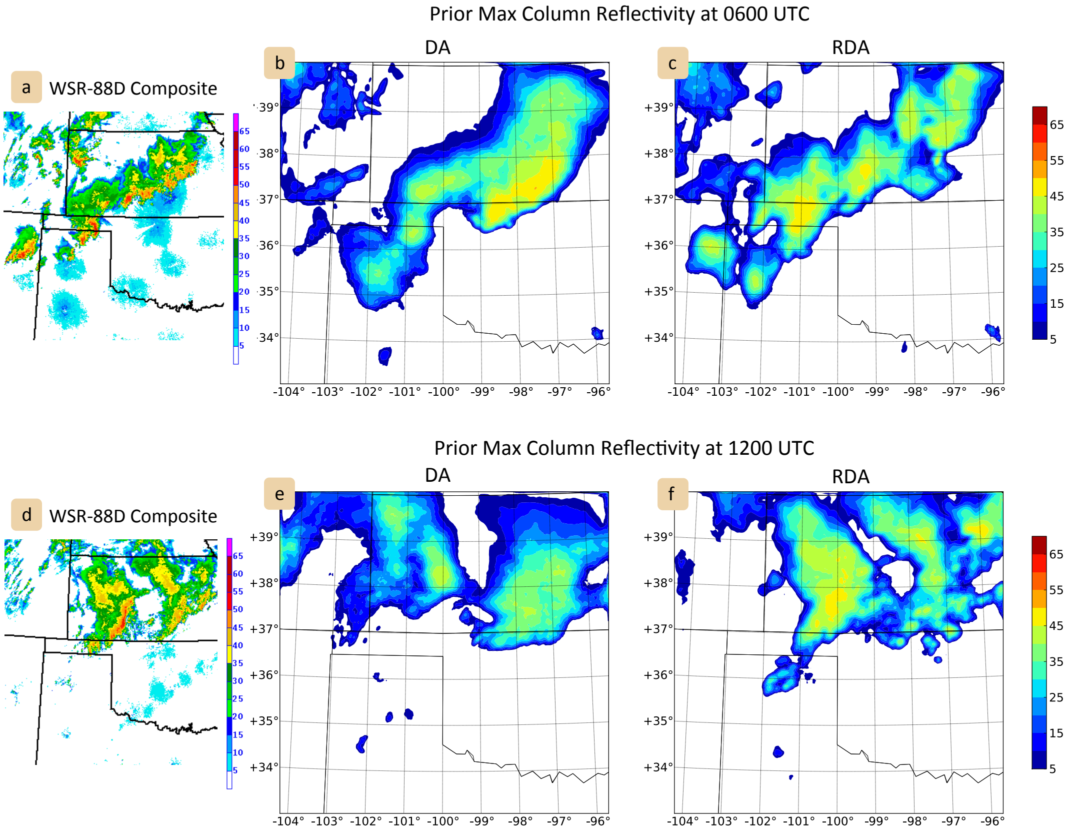


Figure 45. WSR-88D composite reflectivity at 0600 UTC (a) and 1200 UTC (b). Mean analysis maximum column reflectivity at 0600 UTC and 1200 UTC for DA (b,e) and RDA (c,f) experiments.

The radar assimilation also has an impact on unobserved variables, such as temperature. Improvements to the representation of convection in the model lead to improvements for the rest of the state variables via the dynamic and thermodynamic relationships in the model. For example, increased convection represented by higher values of hydrometeor mixing ratios can lead to a larger sized and/or colder cold pool at the surface. The radar observations also have a direct impact on unobserved variables through the EAKF increments, which are investigated by examining the increments to the state. For example, Figure 46 shows the temperature increments at model level 5 (approximately 135 m AGL) at 0600 UTC. The increments from the radar observation assimilation, conventional observation assimilation, and the total increments (sum of increments from radar and conventional observations) in the RDA experiment are shown in Figure 46b,c,d. The total temperature increment (from conventional observations) in the DA experiment is shown in Figure 46a. Even though the majority of the radar observations are located farther above the ground than the 5th model level, the radar assimilation increments show that the radar observations are changing the unobserved temperature state. Positive temperature increments (yellows and reds) occur near the edges of the MCS in the Panhandle and in Kansas and negative temperature increments (blues) occur under the convective cores of the MCS. The total increments in the DA experiment are on average smaller than the total increments in the RDA experiment.

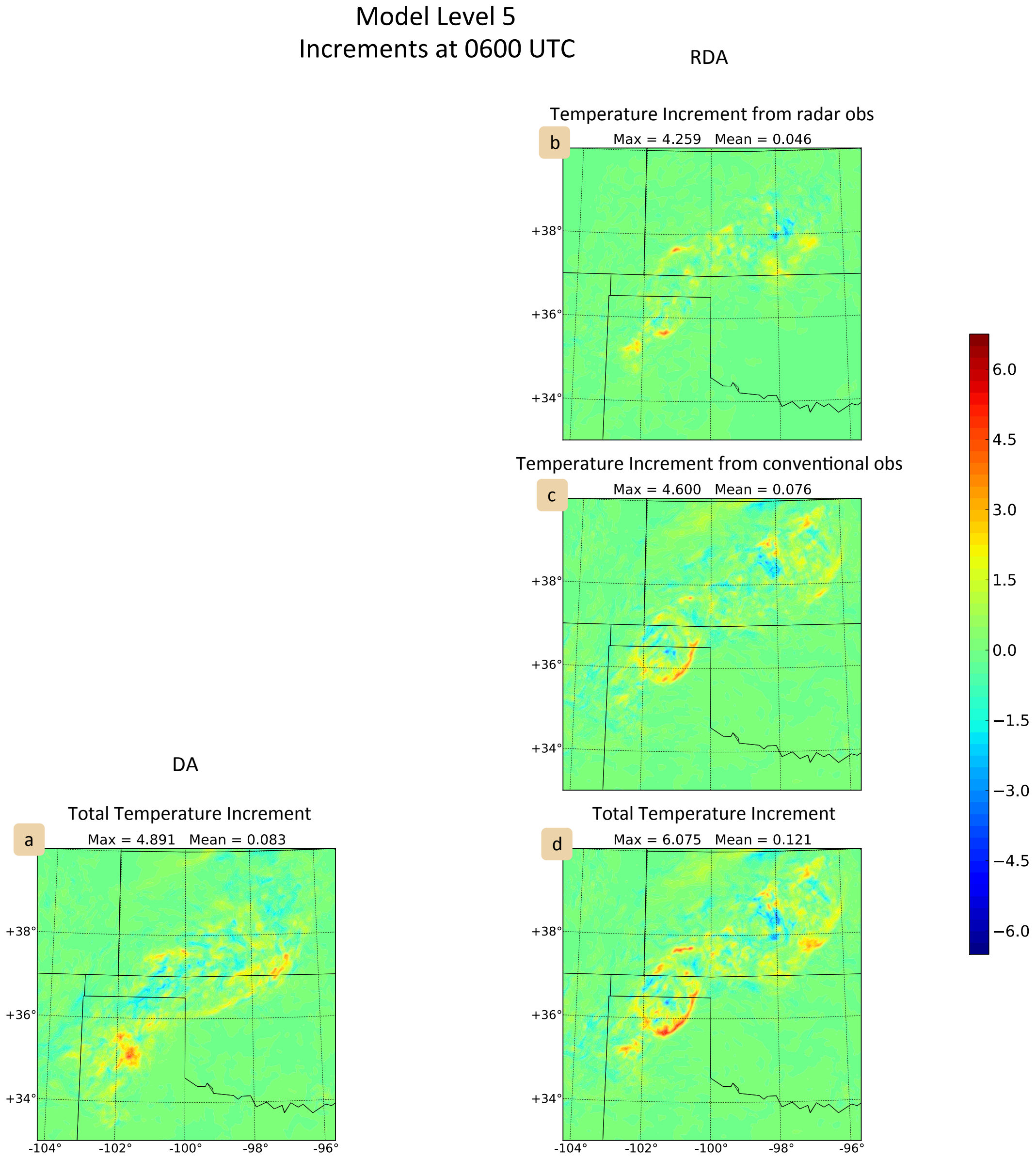


Figure 46. Temperature increments for model layer 5 (approximately 135 m AGL) at 1200 UTC, after the radar observations are assimilated for RDA (b), after the conventional observations are assimilated for RDA (c), and the total increment after all observations are assimilated for DA (a) and RDA (d).

The direct and indirect impacts of radar observations in the RDA analyses of the overnight convection lead to superior fit to surface observations compared to the DA analyses at 1200 UTC (Figure 47). For example, the RDA surface temperature analysis in the morning contains a cold pool in Southwest Kansas, and the cool air associated with the front extends farther to the southeast in the Northeast Texas Panhandle. Further, the RMSE for the RDA temperature and dew point analysis is lower than the DA analysis RMSE. The surface analysis differences persist into the afternoon on 13 June (Figure 48). The RDA surface analysis has lower RMSE and a better fit to the temperature and dew point observations (smaller dots) than the DA analysis. The RDA surface temperature analysis at 1800 UTC contains cooler air (~72 °F) in the northeast corner of the Texas Panhandle (bright green contours). This cooler air indicates the north side of the outflow boundary. However, the DA surface temperature analysis fit is the worst along the outflow boundary, which is indicated by the large dots along approximately 36° - 37° latitude in Figure 4.36a. The outflow boundary can also be seen in the RDA 10 m horizontal wind analysis at 1800 UTC (Figure 49c,e). The RDA wind analysis shows an abrupt shift from southerly to easterly winds associated with the outflow boundary in the Northeast Texas Panhandle. The DA analyzed winds in the Northeast Texas Panhandle have a small easterly component, but the abrupt wind shift is not present (Figure 49b,d). Despite limited wind observations (Figure 49a), a shift from southerly to easterly winds associated with the outflow boundary was observed, and the RDA analysis has lower RMSE and a better fit to the observations than the DA analysis. Therefore, the inclusion of radar observations has generated a more accurate storm environment that better represents the surface outflow boundary critical to the afternoon's convective evolution.

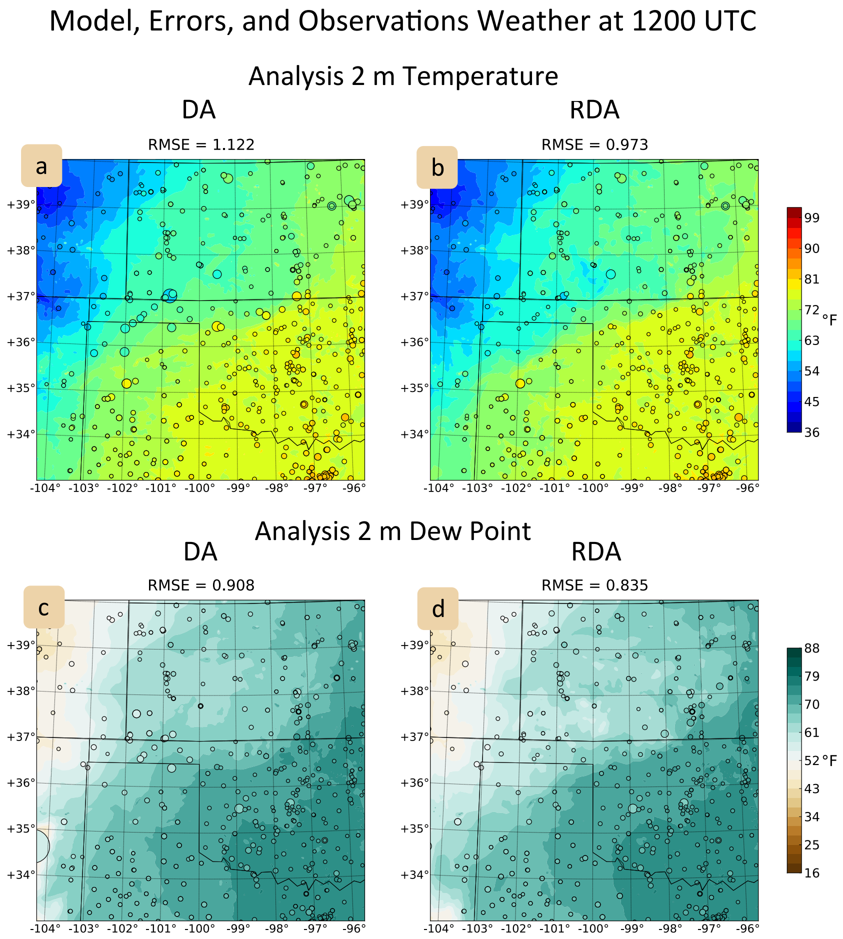


Figure 47. Model, Errors, and Observations Weather at 1200 UTC. The contours are the analysis mean surface temperature for DA (a) and RDA (b), and the analysis mean surface dew point for DA (c) and RDA (d). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

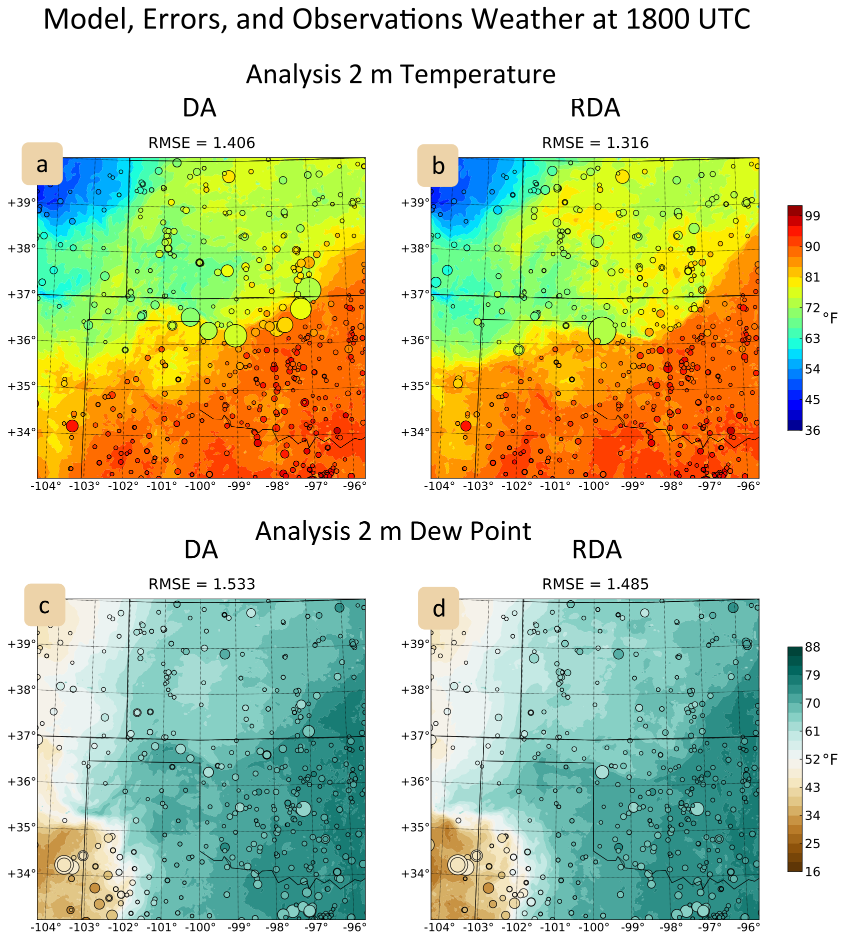


Figure 48. Model, Errors, and Observations Weather at 1800 UTC. The contours are the analysis mean surface temperature for DA (a) and RDA (b), and the analysis mean surface dew point for DA (c) and RDA (d). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

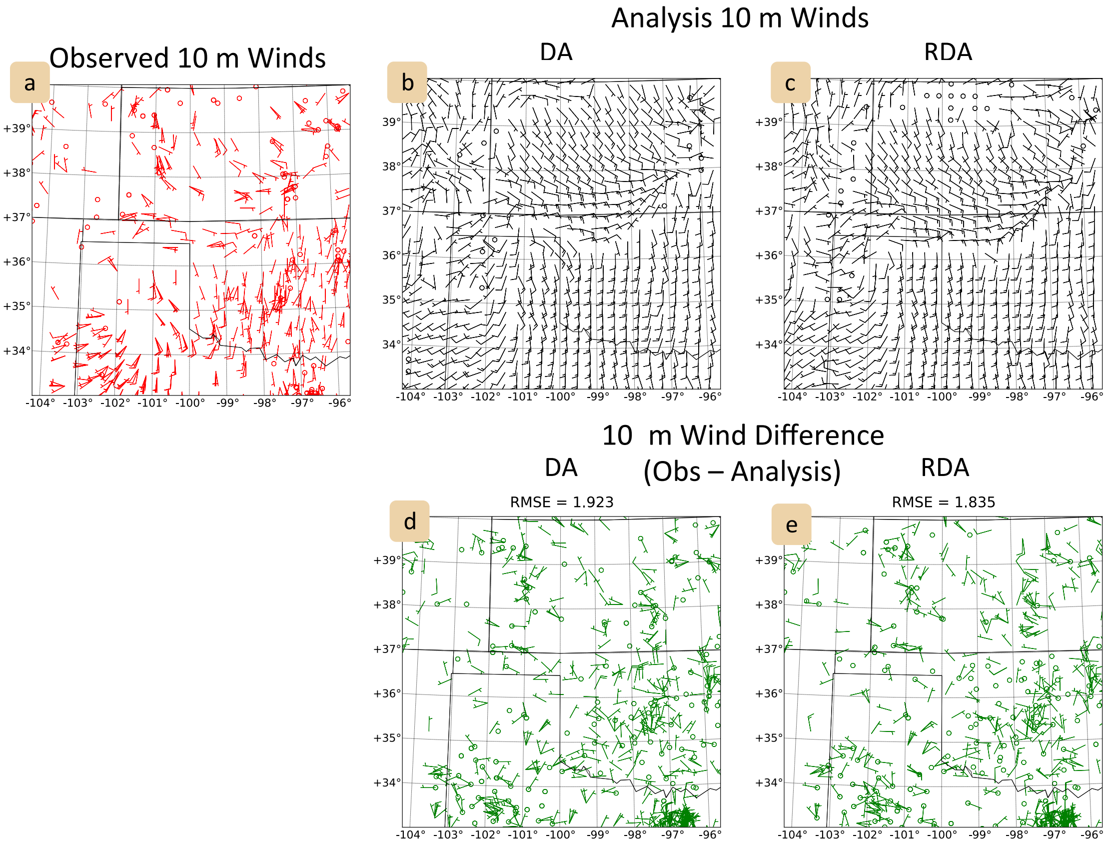


Figure 49. At 1800 UTC, the 10 m wind observations (a), and mean analysis for DA (b) and RDA (c). Half barb = 5 m s^{-1}, full barb = 10 m s^{-1}. The observation minus analysis wind difference for DA (d) and RDA (e). Half barb = 2.5 m s^{-1}, full barb = 5 m s^{-1}.

The dry line and the stationary front are also better analyzed in the RDA experiment compared to the DA experiment at 1800 UTC (Figure 48), which is important for convective initiation because storms initiate near the triple point. The initiation of storms in the Texas Panhandle northeast of the triple point can be seen in the RDA reflectivity analyses (Figure 50c,e). Specifically, near 36° latitude in the Texas Panhandle, the RDA mean analysis has maximum column reflectivity values > 50 dBZ within the black contour that indicates where the KAMA observed reflectivity is ≥ 45 dBZ. However, the DA reflectivity analysis does not match the observed development (Figure 50b,d). Instead, the DA reflectivity analysis has scattered storms in the Oklahoma Panhandle and Southwest Kansas. These spurious storms are likely caused by the northward shift in the surface frontal boundary. In general, the RDA experiment has a more accurate analysis of the developing convection than the DA experiment.

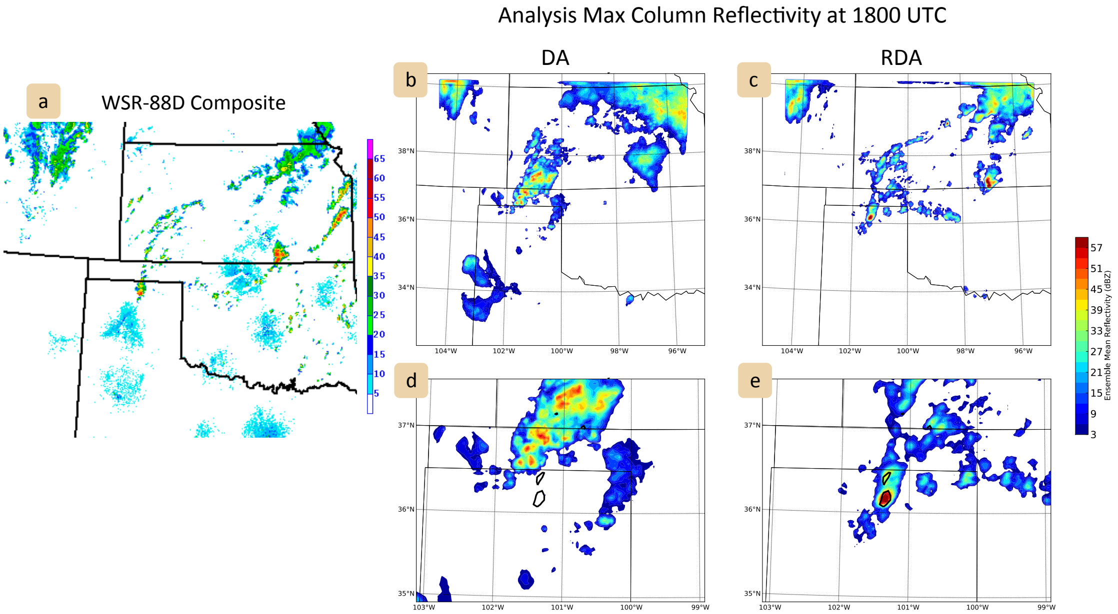


Figure 50. At 1800 UTC, the WSR-88D composite reflectivity (a), and the mean analysis maximum column reflectivity for DA (b,d) and RDA (c,e). The KAMA WSR-88D reflectivity above 45 dBZ is contoured in black in the lower panels (d,e).

In summary, the RDA analyses contain overnight convection that is similar to the observed MCSs, whereas the DA analyses struggle with the location and intensity of the convection. Further, the RDA representation of the storm environment at the surface includes the observed mesoscale forcing features (front, dry line, and outflow boundary) with more accuracy than the DA experiment. The RDA reflectivity analysis in the afternoon also resembles the observed storms, unlike the DA analysis. Hourly assimilation of radar data improves the analyses of the storm environment and the thunderstorms themselves.

4.7.2 Forecast results

The more accurate analysis of the storm environment in the RDA experiment leads to improved forecast performance when compared to the DA experiment. The RDA surface temperature and dew point forecasts are a better fit to the observations (Figure 51). At 2000 UTC, the cool air associated with the both the front and the convective outflow have been mixed out in the Texas Panhandle. However, the RDA forecasts contain sharper temperature gradients associated with the boundaries compared to the DA forecasts. For example, the dots in Figure 51a are larger then those in Figure 51b near 37° latitude, -101.5° longitude, which indicates larger 2 m temperature forecast errors in the DA forecast associated with the cold front and residual cold pool. In addition to the thermodynamic environment differences, the forecasted winds differ in the RDA and DA experiments. The 10 m horizontal winds in the RDA forecasts have the southerly-to-easterly wind shift associated with the outflow boundary, although it is slightly farther to the north, into the Oklahoma Panhandle than the observed boundary (Figure 52a,c,e). In the DA forecast, the easterly component of the 10 m winds increases from south to north in the Texas Panhandle to Kansas region (Figure 53b,d). However, the wind forecast RMSE and fit to the observations is worse than the RDA forecast.

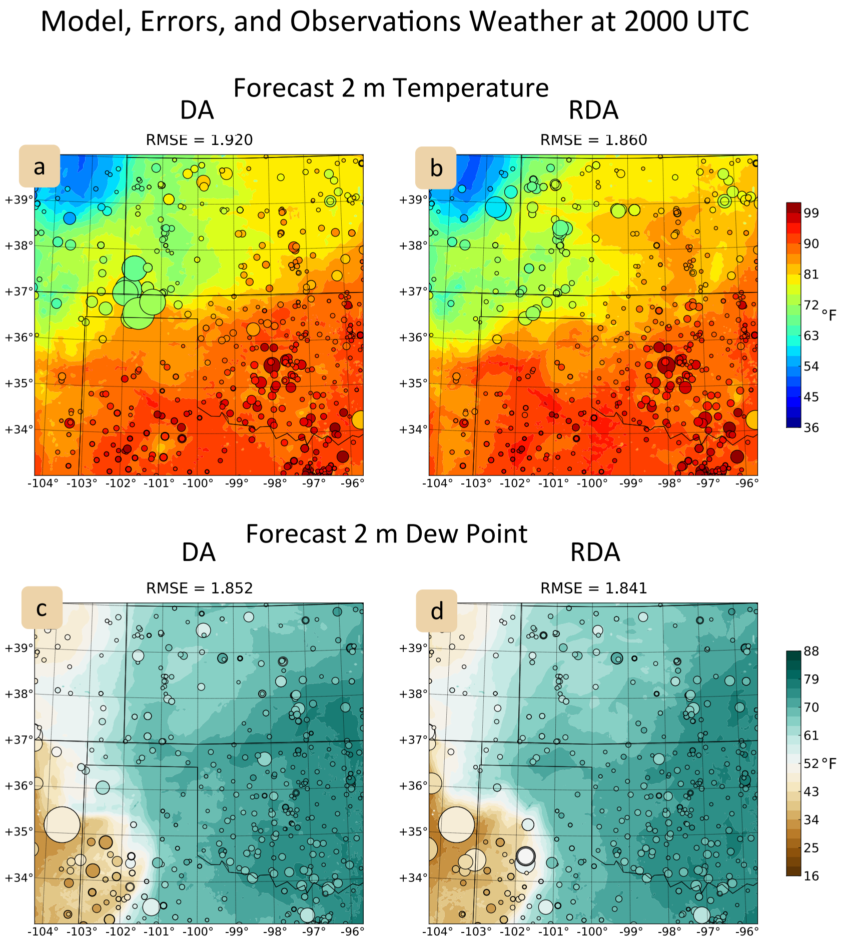


Figure 51. Model, Errors, and Observations Weather at 2000 UTC. The contours are the mean forecast surface temperature for DA (a) and RDA (b), and the mean forecast surface dew point for DA (c) and RDA (d). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

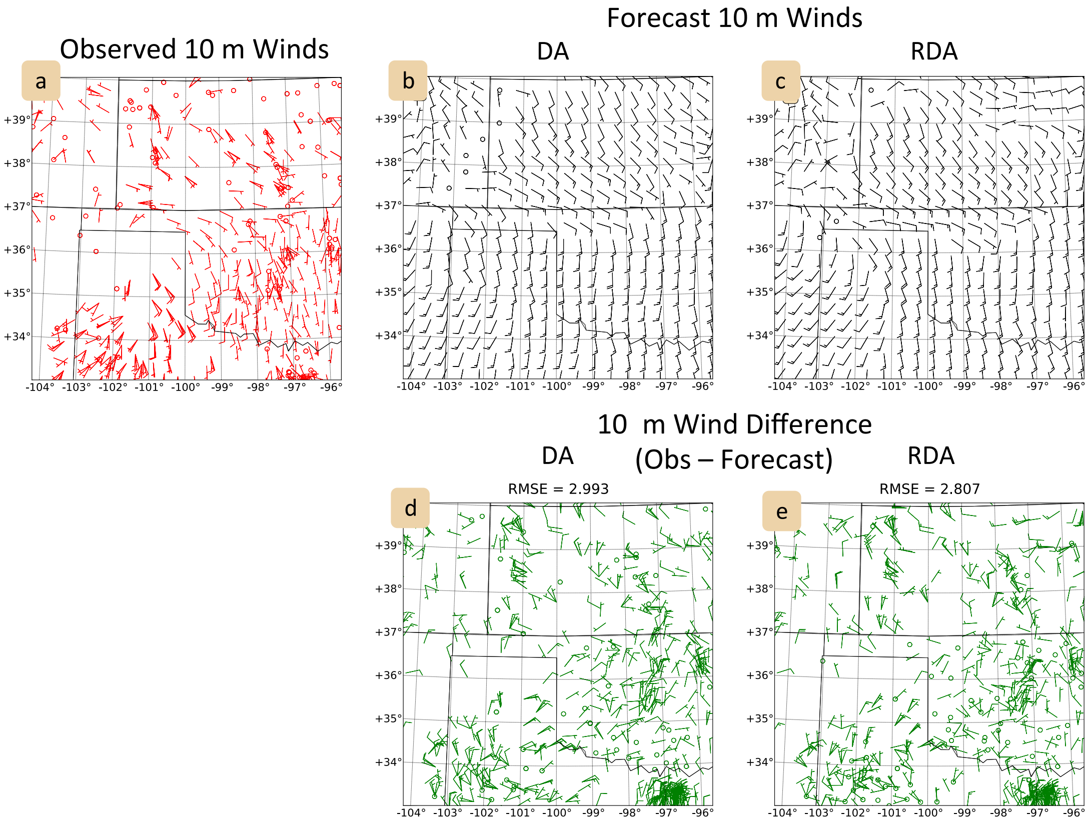


Figure 52. At 2000 UTC, the 10 m wind observations (a), and mean forecast for DA (b) and RDA (c). Half barb = 5 m s^{-1}, full barb = 10 m s^{-1}. The observation minus forecast wind difference for DA (d) and RDA (e). Half barb = 2.5 m s^{-1}, full barb = 5 m s^{-1}.

The differences in the forecasted storm environment, lead to differences in the forecasts of the storms. At 2100 UTC, the probabilistic reflectivity forecasts for the DA experiment have only a 15% probability of reflectivity > 30 dBZ co-located with the observed storms in the Panhandle (Figure 53a). The DA forecast has 40% reflectivity probabilities to the northeast of the Booker storm, near 37° latitude and 99.7° longitude. The RDA forecast has 30% reflectivity probabilities co-located with the observed storms in the Panhandle (Figure 53b). The RDA forecast also has high probabilities (>60%) to the west of the observed storms, which is caused by the shifted location of the forecasted cold front compared to the observed cold front. Despite the location offset, the RDA 3-hour forecast has an enhanced signal for storms near the observed storms relative to the DA forecast. The forecasts are also compared for the presence of rotating storms, measured by the ensemble probability for updraft helicity exceeding 100 m^2 s^{-2} from 2000 to 2100 UTC (Figure 53c,d). Although, neither experiment has a probability swath for updraft helicity co-located with the Booker tornado (36.5°, -100.5°), the RDA forecast has an enhanced signal for the risk of supercells compared to the DA forecast. The WSR-88D derived rotation tracks (Smith and Elmore 2004) are shown in Figure 53e for reference. In summary, with 18 hours of hourly radar data assimilation, the ensemble forecasts produce storms that resemble the observed severe convection.

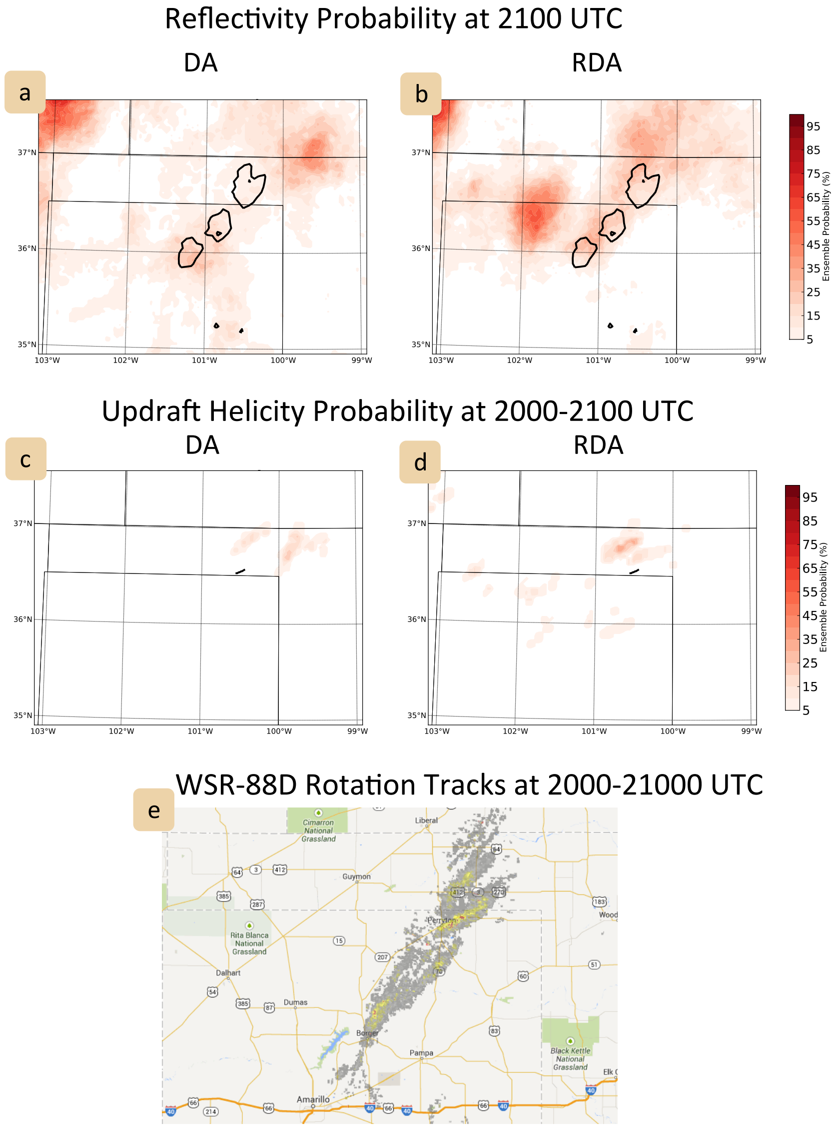


Figure 53. 2100 UTC forecast ensemble probability of reflectivity greater than 30 dBZ for the DA (a) and RDA (b) forecasts. The KAMA 30, 60 dBZ reflectivity observations are contoured in black. The 2000-2100 UTC forecast ensemble probability of updraft helicity greater than 100 m^2 s^{-2} for the DA (c) and RDA (d) forecasts. The observed tornado track is marked in black. The 2000-2100 UTC WSR-88D derived rotation tracks (e) from the NSSL archive.

4.7.3 Radar assimilation in the afternoon

The previous section indicated that the assimilation of radar and conventional observations throughout the day provide superior forecasts compared to only conventional observation assimilation. In this section, the impact of assimilating only the afternoon radar observations is investigated. Thus, the radar observations provide information about the storms of interest on the 13th, but do not include the overnight MCSs. This investigation will determine if the radar observations in the RDA analyses/forecasts had an impact on the storm environment, or if the radar observation simply forced the convection to look like the observations.

Three experiments will be examined (Figure 54). The DA and RDA experiments previously discussed are cycled until 2000 UTC. The third experiment begins from the DA analysis at 1700 UTC and assimilates both radar and conventional observations at 1800, 1900, and 2000 UTC (as in RDA). The third experiment is referred to as the “AddRDA” experiment.

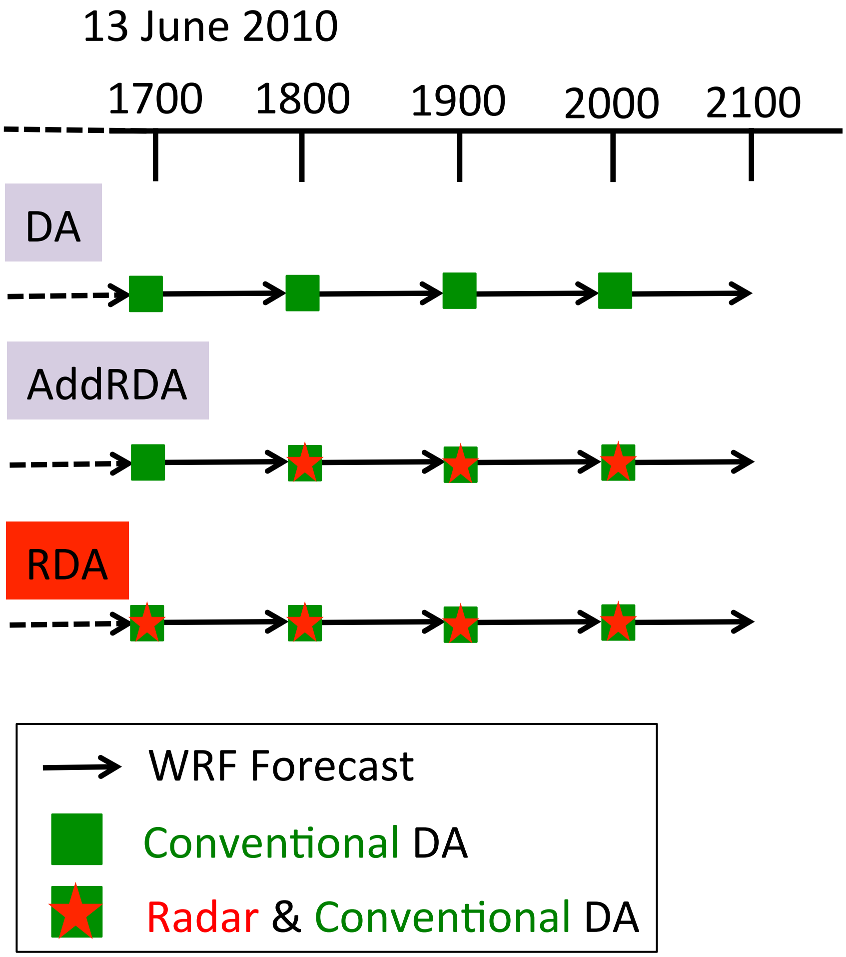


Figure 54. Data assimilation cycle timeline on the afternoon on 13 June. Times are in UTC. The experiment name textboxes are colored to match the Table 4 (experiment overview) row colors. RDA is in red to indicate it is the same experiment discussed in the previous section (HourlyFrom0utc).

The surface analyses at 2000 UTC show that the RDA experiment has a better representation of the outflow boundary than the AddRDA or DA experiments (Figure 55, 56). The RDA surface temperature analysis maintains the cooler air, < 80 °F, along the Eastern Oklahoma-Texas Panhandle border. Further, the RDA 10 m wind analysis contains the abrupt shift from southerly-to-easterly winds near the Eastern Oklahoma-Texas Panhandle border, which indicates the outflow boundary. The AddRDA experiment surface analysis is very similar to the DA surface analysis, which is expected because the experiments only differ by three cycles of radar data assimilation. Thus, the afternoon radar assimilation appears to have minimal impact on the storm environment.

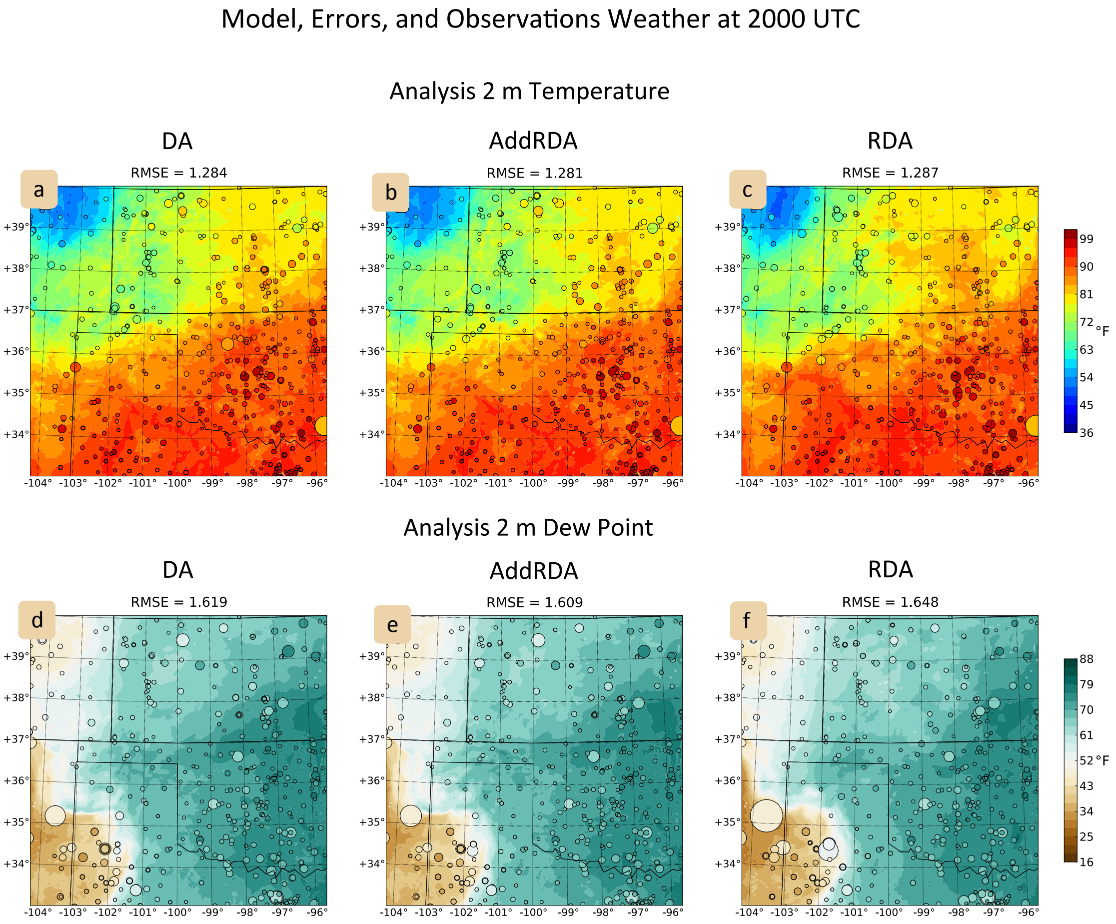


Figure 55. Model, Errors, and Observations Weather at 2000 UTC. The contours are the mean forecast surface temperature for DA (a), AddRDA (b), and RDA (c), and the mean forecast surface dew point for DA (d), AddRDA (e), and RDA (f). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

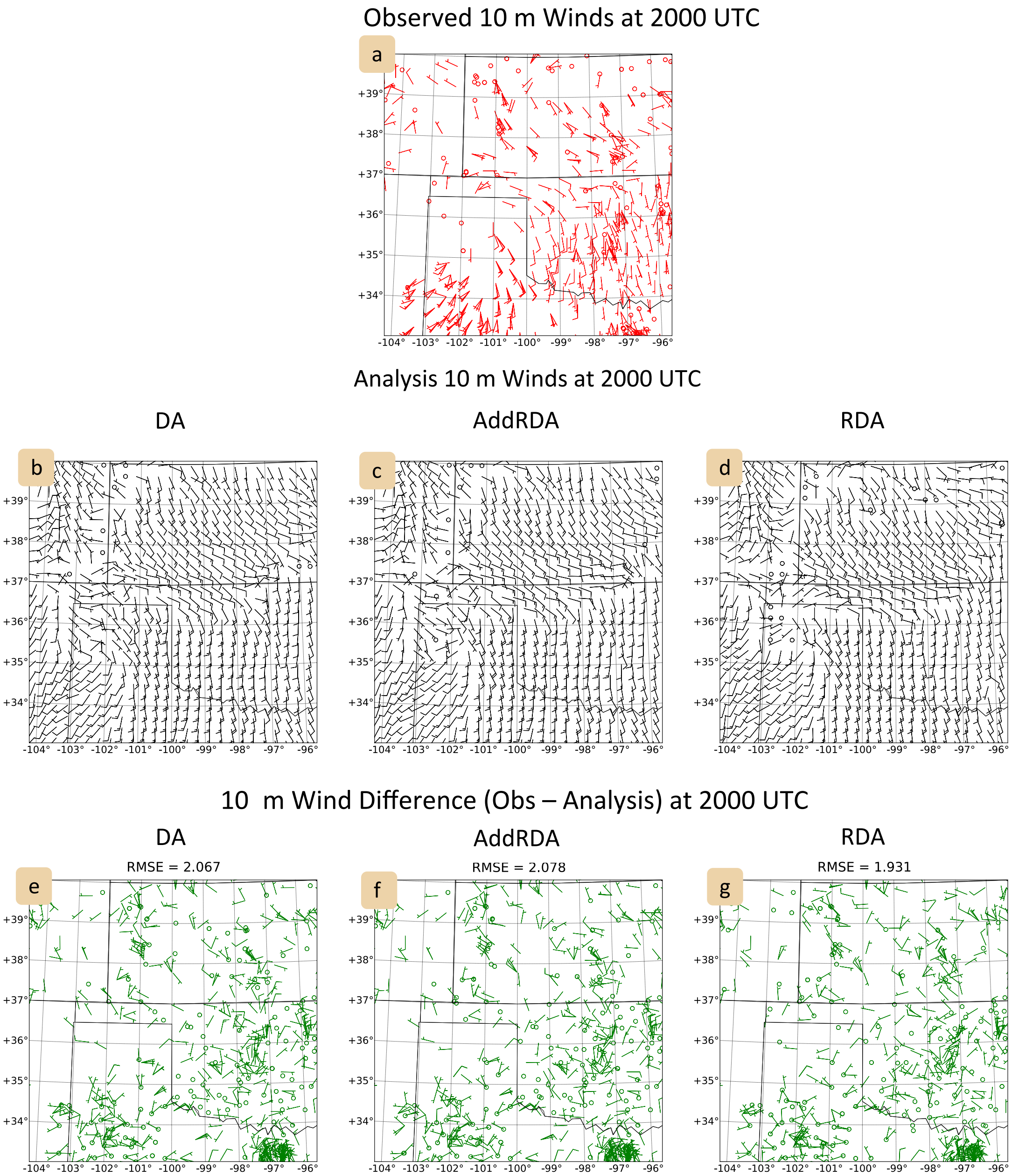


Figure 56. At 2000 UTC, the 10 m wind observations (a), and mean forecast for DA (b), AddRDA (b), and RDA (c). Half barb = 5 m s^{-1}, full barb = 10 m s^{-1}. The observation minus forecast wind difference for DA (e), AddRDA (f), and RDA (g). Half barb = 2.5 m s^{-1}, full barb = 5 m s^{-1}.

At 2000 UTC, the RDA and AddRDA ensemble mean reflectivity analyses are similar (Figure 57). They both have intense thunderstorms in the Texas Panhandle that fit the KAMA WSR-88D observed storms (Figure 57f,g). The RDA and AddRDA analyses also capture the observed storms in Kansas (Figure 57a,c,d). However, the DA analysis does not fit the observations well (Figure 57a,b,e). It is missing the intensity of the observed thunderstorms, and is not as accurate in terms of the location of the thunderstorms across the region. Thus, the addition of the afternoon radar observations makes a positive impact on the analyzed storms.

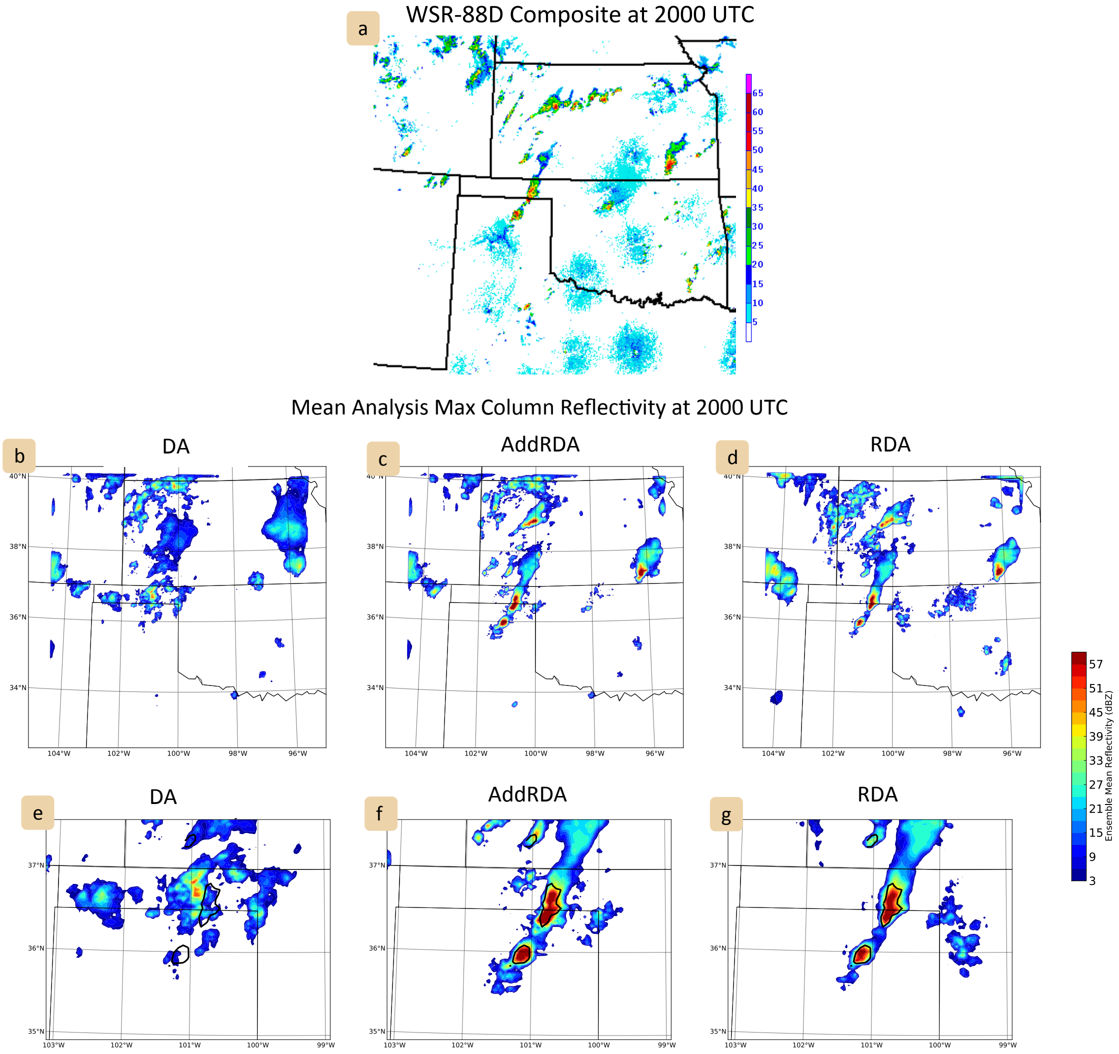


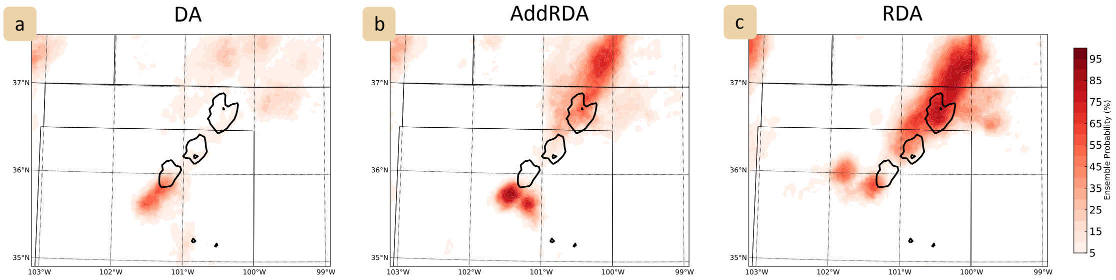
Figure 57. At 2000 UTC, the WSR-88D composite reflectivity (a), and the mean analysis maximum column reflectivity for DA (b,e), AddRDA (c,f), and RDA (d,g).

The KAMA WSR-88D reflectivity above 45 dBZ is contoured in black in the lower panels (e,f,g).

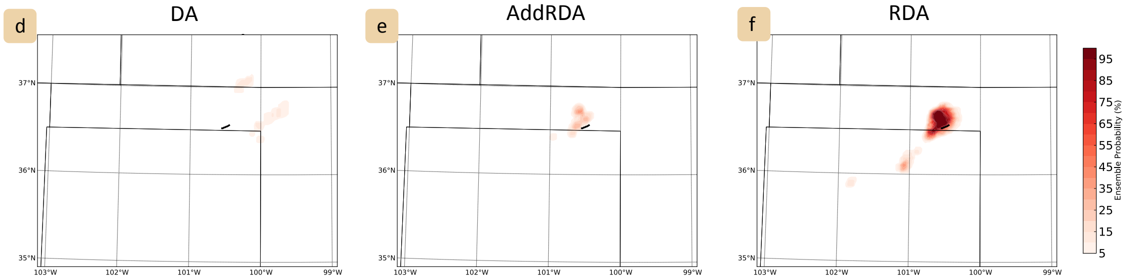
4.7.4 1-Hour forecast results

Ensemble forecasts (50 members) are initiated from the analyses at 2000 UTC. The 1-hour RDA forecasts have the highest reflectivity ensemble probabilities, approximately 85% (Figure 58a-c). The RDA probabilities do not exactly match the KAMA observed storm locations. But the RDA forecast probabilities are a better fit than the other experiments. The AddRDA forecasts contain two regions of enhanced ensemble reflectivity probabilities, > 50%. However, the probabilities are lower than and the location errors are worse than the RDA forecasts. The DA forecasts have one area of low probabilities of approximately 40% in the Panhandle region and are the worst fit to the observations.

Reflectivity Probability at 2100 UTC



Updraft Helicity Probability at 2000-2100 UTC



WSR-88D Rotation Tracks at 2000-2100 UTC

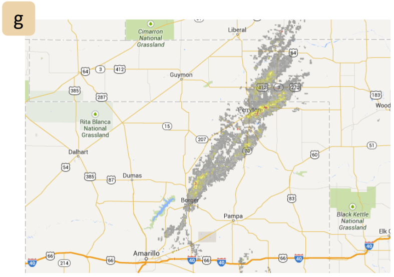


Figure 58. The 2100 UTC forecast ensemble probability of reflectivity greater than 30 dBZ for the DA (a), AddRDA (b), and RDA (c) forecasts. The KAMA 30, 60 dBZ reflectivity observations are contoured in black. The 2000-2100 UTC forecast ensemble probability of updraft helicity greater than 100 m^2 s^{-2} for the DA (d), AddRDA (e), and RDA (f) forecasts. The observed tornado track is marked in black. The 2000-2100 UTC WSR-88D derived rotation tracks (g) from the NSSL archive.

The RDA forecasted ensemble probability of updraft helicity exceeding 100 m^2 s^{-2} indicates strong rotation associated with the Booker storm (Figure 58f). The AddRDA forecasted probabilities shows a weak signal of rotation in the Oklahoma Panhandle (Figure 58e). The DA probabilities are lower than the AddRDA forecast, and misplaced to the northeast of the observed rotation. This result indicates that the superior storm analysis in the RDA experiment, which includes the outflow boundary, leads to improved forecasts of the potential for updraft rotation co-located with the observed Booker storm. Thus, the assimilation of radar observations throughout the day on 13 June is beneficial to forecast skill.

4.7.5 Conclusions

The analyses and forecasts using conventional observation data assimilation is compared to the analyses and forecasts using Doppler radar and conventional observation data assimilation for hourly cycles on 13 June. Even though only one volume of WSR-88D observations (from 4 radars) is assimilated every hour, the data still has an impact on the analyses and subsequent forecasts. The RDA priors match the observed overnight convection, whereas the DA priors struggle to capture the observed location, intensity, and mode of the convection. The RDA analyses contain the mesoscale surface boundaries; stationary front, dry line, and outflow boundary. The representation of the boundaries leads to improved convective forecasts in terms of storm location, intensity, and updraft rotation. Without radar data assimilation, the ensemble forecasts do not resemble the observed storm environment and severe convection in the Panhandle region. Thus, the simultaneous assimilation of radar and conventional observations does contribute to creating a more accurate storm environment and improved convective forecasts.

The addition of radar observations in the afternoon hours improves the analyses and forecasts of the storms. However, it makes little impact on the storm environment. The RDA experiment provides a better storm environment than the DA or AddRDA experiment, which leads to improved forecasts of storm rotation. Thus, the addition of radar observations in the afternoon can force the model to have convection resembling the observations. However, the forecasted storm evolution is dependent on the mesoscale environment. For this case, the overnight assimilation of radar observations is important to the accurate analysis of the mesoscale environment. The comparisons presented here suggest the need for radar observations to be included in multi-scale data assimilation systems.

4.8 Initial conditions for multi-scale data assimilation

In the experiments previously discussed, the focus was on the data assimilation on 13 June, the day of the event of interest. However, another open question is, how should a multi-scale data assimilation system be initialized? The use of data assimilation cycling for the initialization of ensemble forecasts was discussed in Section 4.1.2. However, it is unclear how far in advance of the event of interest data assimilation cycling is needed, which is the focus of this section. Previous studies have ranged from including forecasts from 10 days prior to the event (Jung et al. 2012), to starting from a sounding or analysis within a few hours of the event (e.g. Dawson and Xue 2006; Aksoy et al. 2009). Further, the forecast sensitivity to the background environment has been discussed in the literature (e.g. Dawson and Xue 2006).

However, most previous studies have focused on the background for convection, rather than the background for the storm environment. In this section, three choices for background ensemble initialization are investigated. The experiments described below do not represent all of the possible options for ensemble initialization, nor will they be able to solve all of the problems discussed here. However, the goal is to compare three reasonable choices for ensemble initialization, and to investigate the sensitivity of the forecasts to the initialization choice.

4.8.1 Experiment design

In the first experiment, the background ensemble is initialized from the GFS analysis at 1800 UTC on 12 June, and is referred to as the "DayOfStart" experiment because it starts data assimilation on the day of interest. The GFS analysis contains information from observations at 1800 UTC and from previous GFS forecasts. Thus, it should be a reasonably accurate representation of the large-scale weather pattern. The GFS analysis is downscaled to both the outer WRF domain (15 km horizontal grid spacing) and the nested WRF domain (3 km horizontal grid spacing). A 6-hour forecast is made from 1800 UTC on the 12th to 0000 UTC on the 13th. Then, the hourly cycling with both radar and conventional observations is executed on the 13th (as discussed previously). A schematic of the experiment start time and data assimilation cycling is shown in Figure 59.

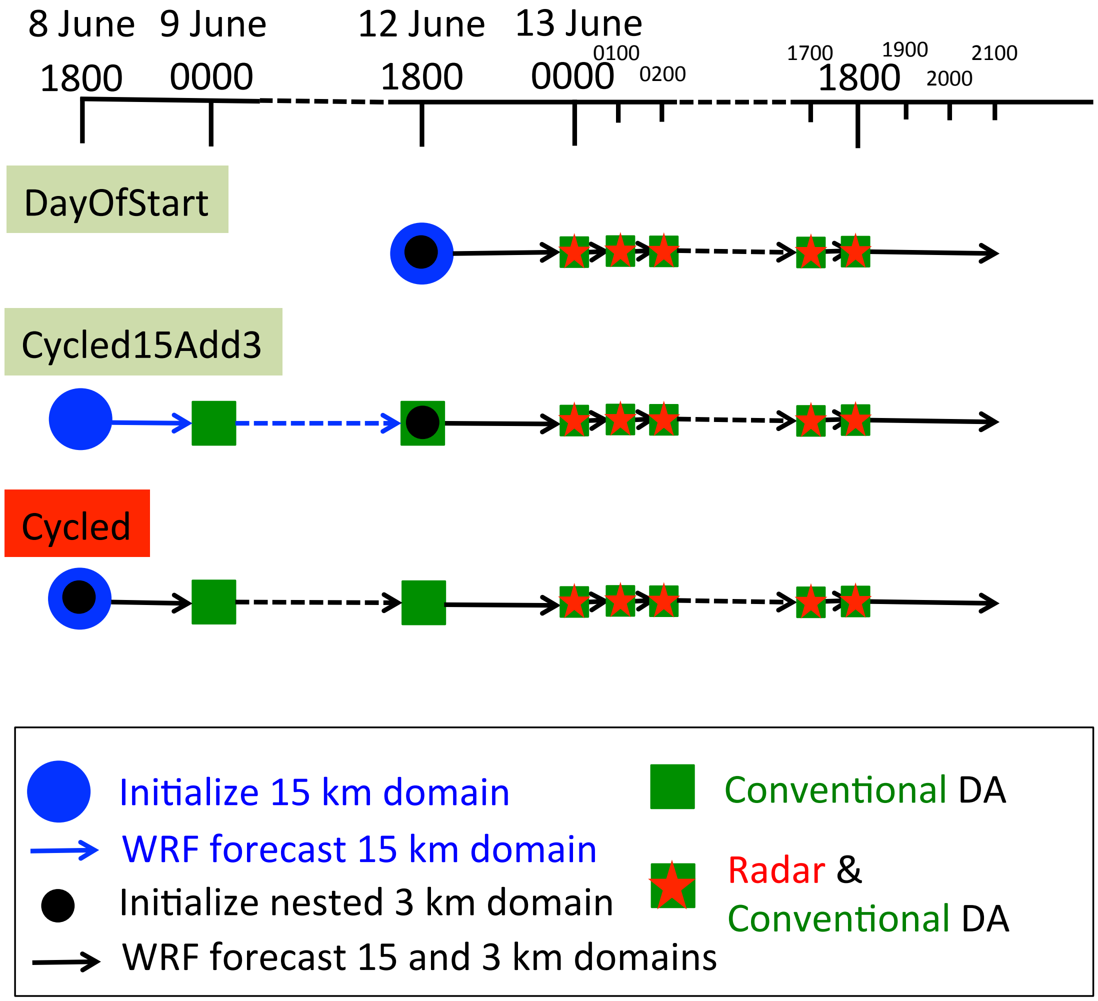


Figure 59. Data assimilation cycle timeline for the full assimilation period. Times are in UTC. The experiment name textboxes are colored to match the Table 4 (experiment overview) row colors. Cycled is in red to indicate it is the same experiment discussed in the previous sections (RDA/HourlyFrom0utc).

Another approach to initialize the background ensemble is to start from an analysis created with data assimilation cycling, which is referred to as the "Cycled" experiment. The Cycled experiment is also initialized from a GFS analysis but data assimilation cycles begin 4-days prior to the day of interest. On 8 June at 1800 UTC, both the outer WRF domain (15 km horizontal grid spacing) and the nested WRF domain (3 km horizontal grid spacing) are initialized from the GFS analysis. Next, a 6-

hour WRF forecast is run. Then, conventional observations are assimilated every 6-hours leading up to 0000 UTC on the 13th. Finally, the hourly data assimilation with radar and conventional observations is executed on the 13th (Figure 59). The Cycled experiment is identical to the "HourlyFrom0utc" experiment in Section 4.6 and the "RDA" experiment in Section 4.7. Each of the experiments in Sections 4.6 and 4.7 is initialized from the cycled background ensemble at 0000 UTC on 13 June.

The third background ensemble experiment is a combination of the aforementioned approaches. The outer WRF domain (15 km horizontal grid spacing) is initialized 4-days prior to the event and has 6-hourly conventional assimilation, as in the Cycled experiment. The nested WRF domain (3 km horizontal grid spacing) is initialized through downscaling of the outer WRF domain at 1800 UTC on 12 June. Finally, both grids are used for the hourly data assimilation with radar and conventional observations on the 13th. This experiment is referred to as the "Cycled15Add3" experiment (Figure 59). The observations assimilated in the Cycled15Add3 experiment are identical to those assimilated in the Cycled experiment. The nested grid takes advantage of the ensemble variability and finer resolved scales from the outer domain compared to being initialized from the GFS as in the DayOfStart experiment. However, the adaptive inflation for the data assimilation on the 13th is not spun up like it is in the Cycled experiment. Also, due to technical complications, the soil state for the nested grid is initialized from the GFS analysis at 1800 UTC on the 12th. (The WRF interpolation routine used to downscale to a nested grid is only implemented for the atmospheric state variables.) The difference between using the downscaled cycled soil state and the GFS analysis soil state appear to be small (Romine personal communication). The initial differences in the three experiments are discussed in the following section.

4.8.2 Initial analysis

As mentioned in the previous section, the Cycled experiment has an atmospheric state, soil state, and adaptive inflation field that have evolved through the previous 4-days of cycling. Comparisons at 0000 UTC on 13 June between the Cycled and the DayOfStart experiment (and Cycled15Add3 experiment for the atmospheric state) are investigated to examine the similarities and differences in the ensemble after the nested CAM WRF domain is initialized.

At 0000 UTC on 13 June the Cycled experiment has an adaptive inflation field that has evolved over 4-days of cycling. For example, the perurbation temperature inflation varies across the domain (Figure 60c). The inflation field already contains knowledge of the available observations. However, the inflation for the DayOfStart (and Cycled15Add3) experiment was initialized at 0000 UTC to a constant value of 1.0 with a standard dievation of 0.8 (Figure 60a,b). A spun-up inflation field might be an advantage because the inflation acts to increase ensemble spread when observations are assimilated. Thus, for the following assimilation cycles the ensemble spread should be maintained more effectively for the Cycled experiment.

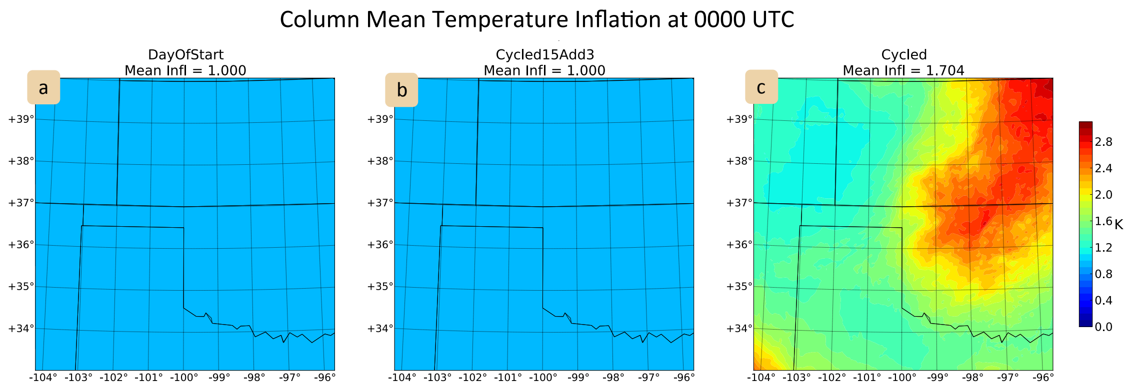


Figure 60. Column mean temperature inflation at 0000 UTC for DayOfStart (a), Cycled15Add3 (b) and Cycled (c).

At 0000 UTC on 13 June, the soil state for the Cycled experiment has been freely evolving for 4-days. However, the soil state for the DayOfStart (and Cycled15Add3) experiment was initialized only 6-hours prior to 0000 UTC. Despite this difference, the Cycled and DayOfStart top layer soil moisture is very similar (Figure 61). The DayOfStart ensemble members forecasts have a smoother soil moisture pattern across the domain, but they still contain the increased moisture in the regions where precipiation occurred during the forecast. The top layer soil moisture is likely the part of the soil state that impacts the atmosphere the most through surface exchanges (Romine personal communication). The general similarity between these soil states indcates that the 4-days of cycling does not have a significant impact on the soil state.

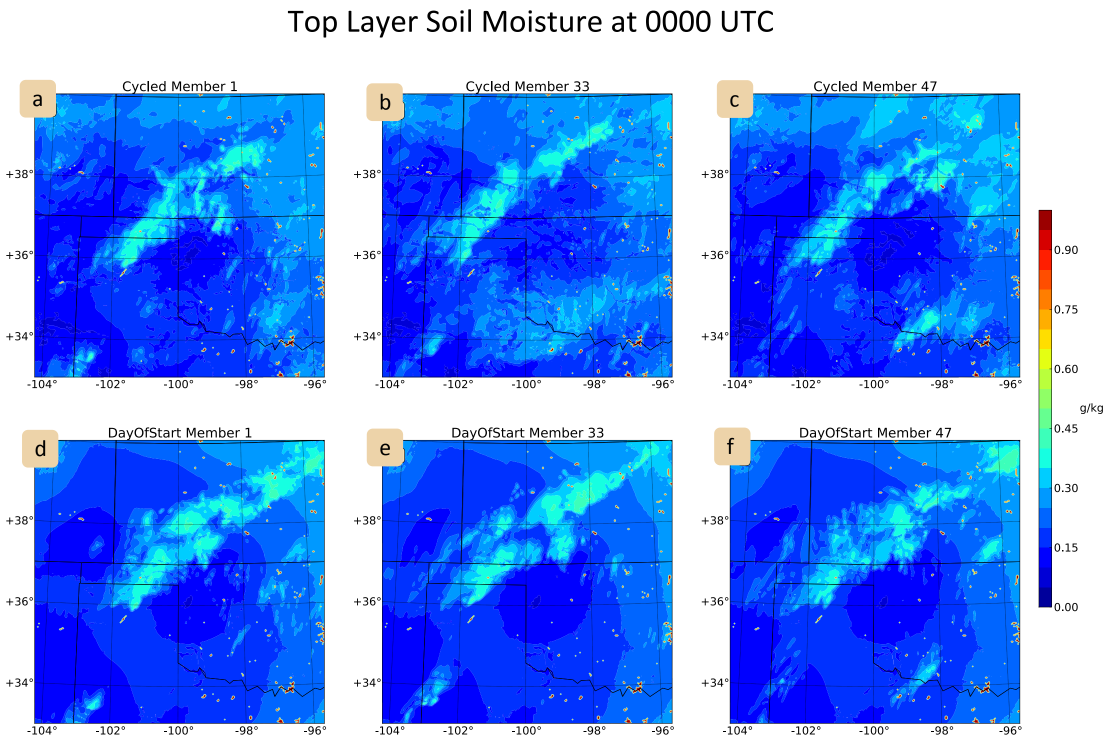


Figure 61. Top layer (closest to the atmosphere) soil moisture at 0000 UTC for member 3 (a), 33 (b), 47 (c) in the Cycled experiment and for member 3 (d), 33 (e), 47 (f) in the DayOfStart experiment. The members are randomly chosen.

The 6-hour forecasts valid at 0000 UTC on 13 June are used to indicate differences in the ensemble initialization. The three experiments have similar surface temperature forecasts (Figure 62a-c). The cold front is shifted to the northwest in the DayOfStart experiment compared to the Cycled and Cycled15Add3 forecasts. Also, the DayOfStart surface temperature is not as cool in the northwest corner of the domain and warmer along the southern third of the domain compared to the Cycled and Cycled15Add3 forecasts. Thus, it appears that the DayOfStart has a warm bias at the surface and/or the Cycled/Cycled15Add3 have a cool bias. A cool bias could be caused by the MYJ PBL scheme (see Section 4.3.1).

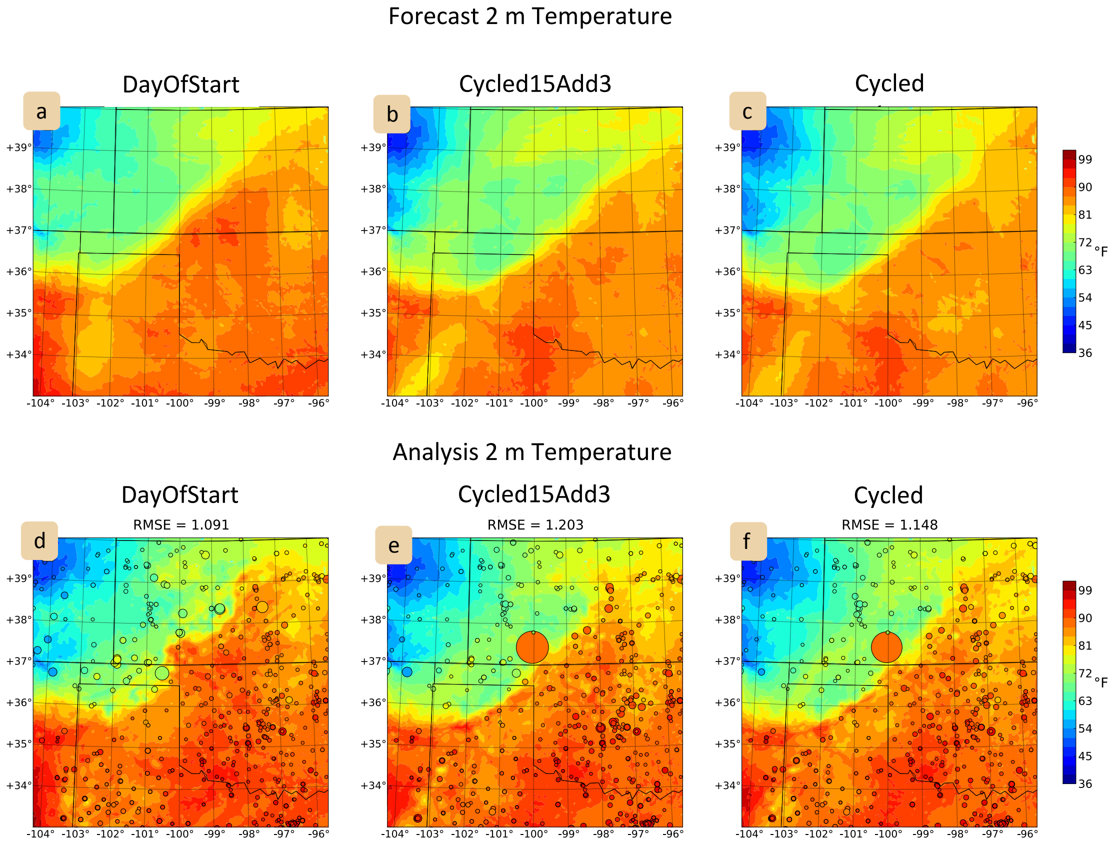


Figure 62. Ensemble mean surface temperature forecast valid at 0000 UTC on 13 June for the DayOfStart (a), Cycled15Add3 (b), and Cycled (c) experiments. Model, Errors, and Observations Weather at 0000 UTC. The contours are the analysis mean surface temperature for DayOfStart (d), Cycled15Add3 (e), and Cycled (f). The dot sizes are the model errors. The dot colors are the temperature observations.

The 0000 UTC analysis indicates that the DayOfStart forecast was not as cold as the observations in the northwest corner of the doamin. Further, the Cycled/Cycled15Add3 forecasts were too cool in most of Oklahoma and North Texas compared to the observations (Figure 62d-f). Thus, there is evidence of bias in each of the experiments. The DayOfStart analysis has the lowest RMSE of the three analyses and might indicate that the analysis is overfitting the surface observations. Although, the fit to the observations appears to be better (smaller circles) in the Oklahoma and Texas Panhandle for the Cycled analysis.

The 6-hour surface dew point forecast (valid at 0000 UTC on 13 June) are similar in each experiment (Figure 63). The biggest differences are for the shape and location of the dryline near the wester edge of the domains. Unfortunately, comparisons of the modeled drylines to observations are difficult due to a limilted number of observations in this region. The Cycled15Add3 dew point analysis at 0000 UTC has the lowest RMSE.

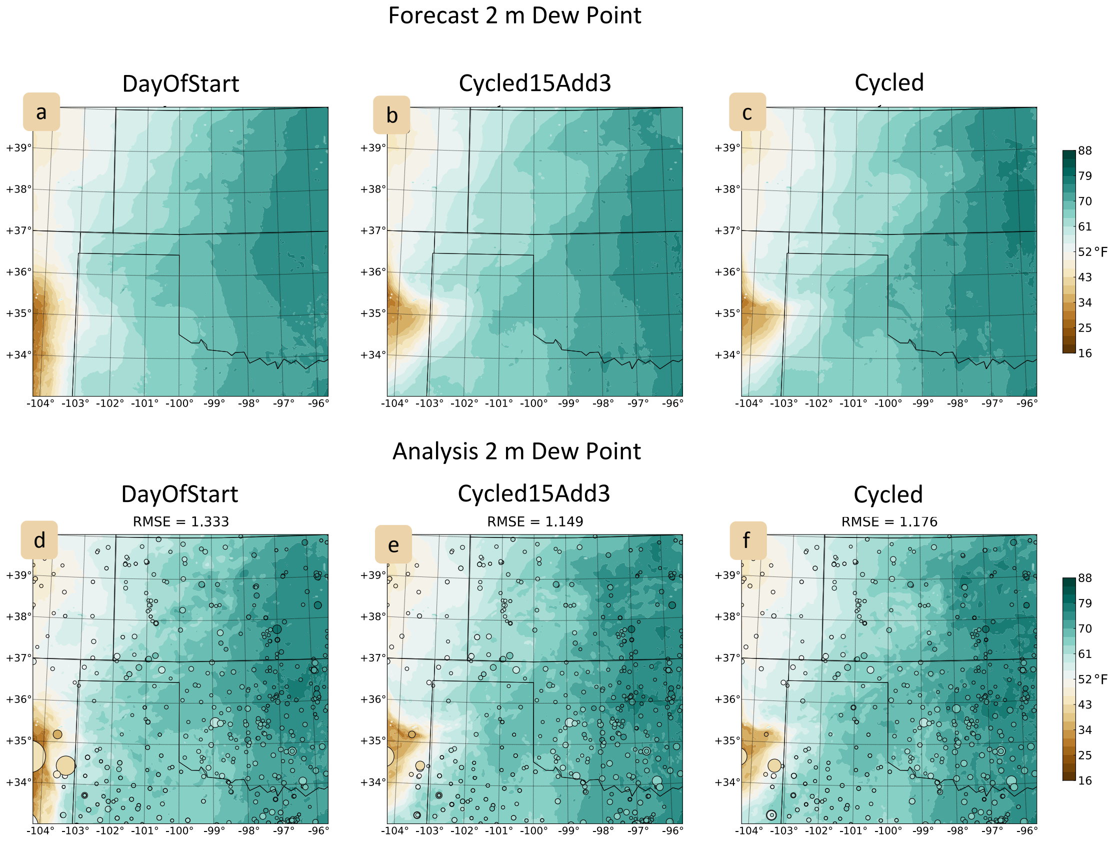


Figure 63. Ensemble mean surface dew point forecast valid at 0000 UTC on 13 June for the DayOfStart (a), Cycled15Add3 (b), and Cycled (c) experiments. Model, Errors, and Observations Weather at 0000 UTC. The contours are the analysis mean surface dew

point for DayOfStart (d), Cycled15Add3 (e), and Cycled (f). The dot sizes are the model errors. The dot colors are the dew point observations.

The differences between the general weather pattern for each background initialization is small at 0000 UTC on the 13th, which indicates that each of the ensemble initialization methodologies are reasonable choices. However, the details of the mesoscale pattern, including the location of the cold front and dryline, will be shown to have an impact on later data assimilation cycles and forecasts in the following section.

4.8.3 Analysis results

This section evaluates if the initial differences between the DayOfStart, Cycled15Add3 and Cycled analyses lead to differences in the hourly analysis cycles on 13 June.

The reflectivity priors in all three experiments are able to capture the convective mode and approximate location/orientation of the MCSs during the overnight hours (Figure 64). The details in the forecast at 0600 and 1200 UTC differ slightly, but fit the observations reasonably well. Specifically, the low reflectivities (< 25 dBZ) are over the same area in each prior. However, the higher reflectivies (> 40 dBZ) vary between the mean priors. Since these are mean forecasts, the reflectivity values are below the observed peak intensity. At 0600 UTC, the DayOfStart reflectivity prior does not contain the observed storms in Northeast New Mexico, but it does contain the observed storms near Goodland, Kansas.

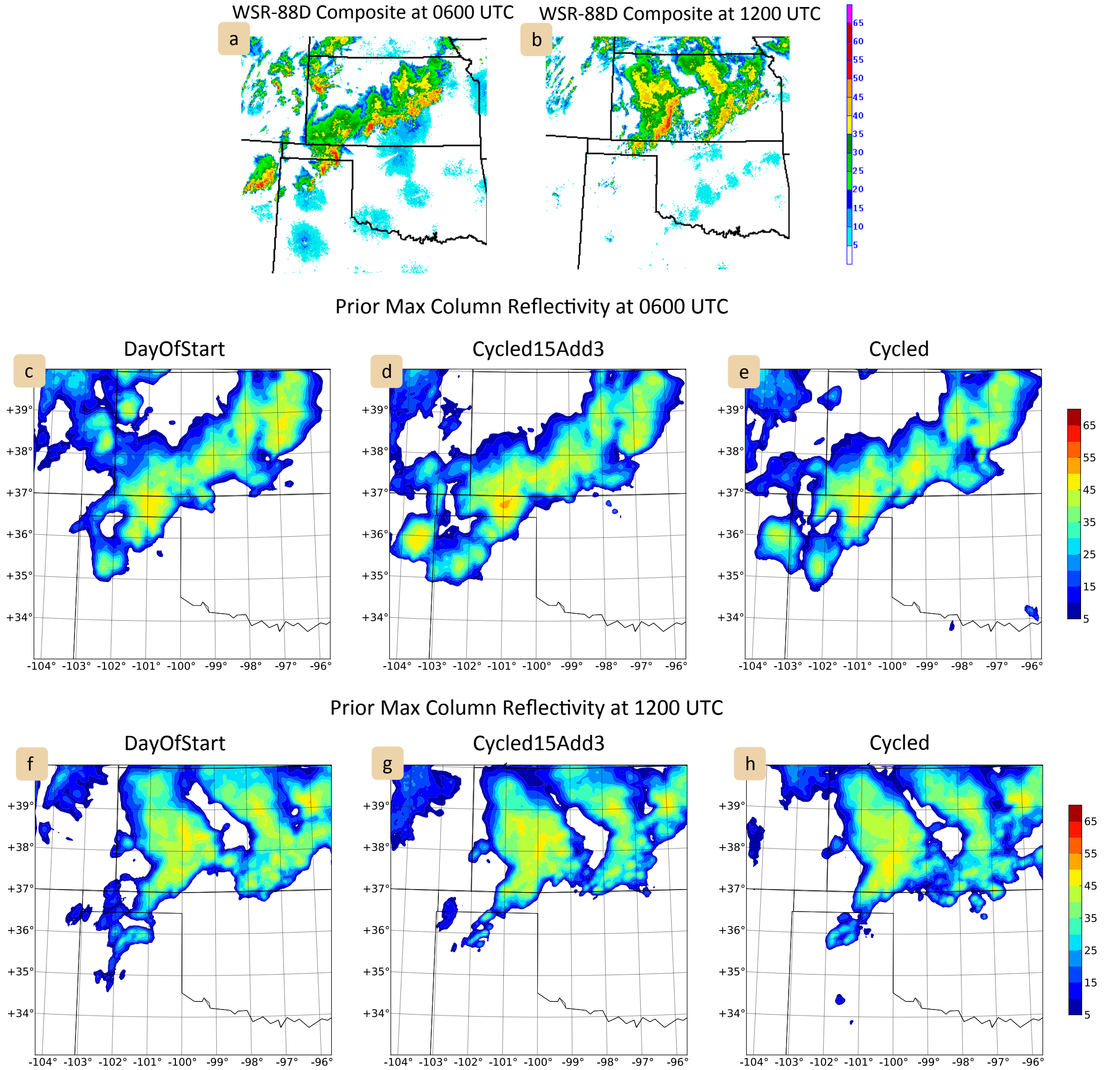


Figure 64. WSR-88D composite reflectivity at 0600 UTC (a) and 1200 UTC (b). Mean prior maximum column reflectivity at 0600 UTC and 1200 UTC for DayOfStart (c,f), Cycled15Add3 (d,g), and Cycled (e,h) experiments.

At the last analysis time (1800 UTC), the surface analyses of all three experiments have a close fit to the temperature and dew point observations (Figure 65). The errors in the three experiments are of similar magnitude, but the DayOfStart has the lowest RMSE value for temperature. The Cycled15Add3 surface temperature analysis appears to have the closest fit to the observations near the front and outflow boundary, as shown by the smaller sized dots in Figure 65b. The DayOfStart and Cycled surface analyses are more similar to each other at 1800 UTC (Figure 65a,c) than they were at 0000 UTC (Figure 32d,f and Figure 63d,f). Thus, hourly data assimilation cycles with radar and conventional observations converge the analyses. The Cycled15Add3 wind analysis has the lowest RMSE value, and appears to have the closest fit to the observations as shown by the calm and low magnitude winds in the wind difference plot (Figure 64). Future work can focus on seeing the errors get low (Lil' Jon 2002).

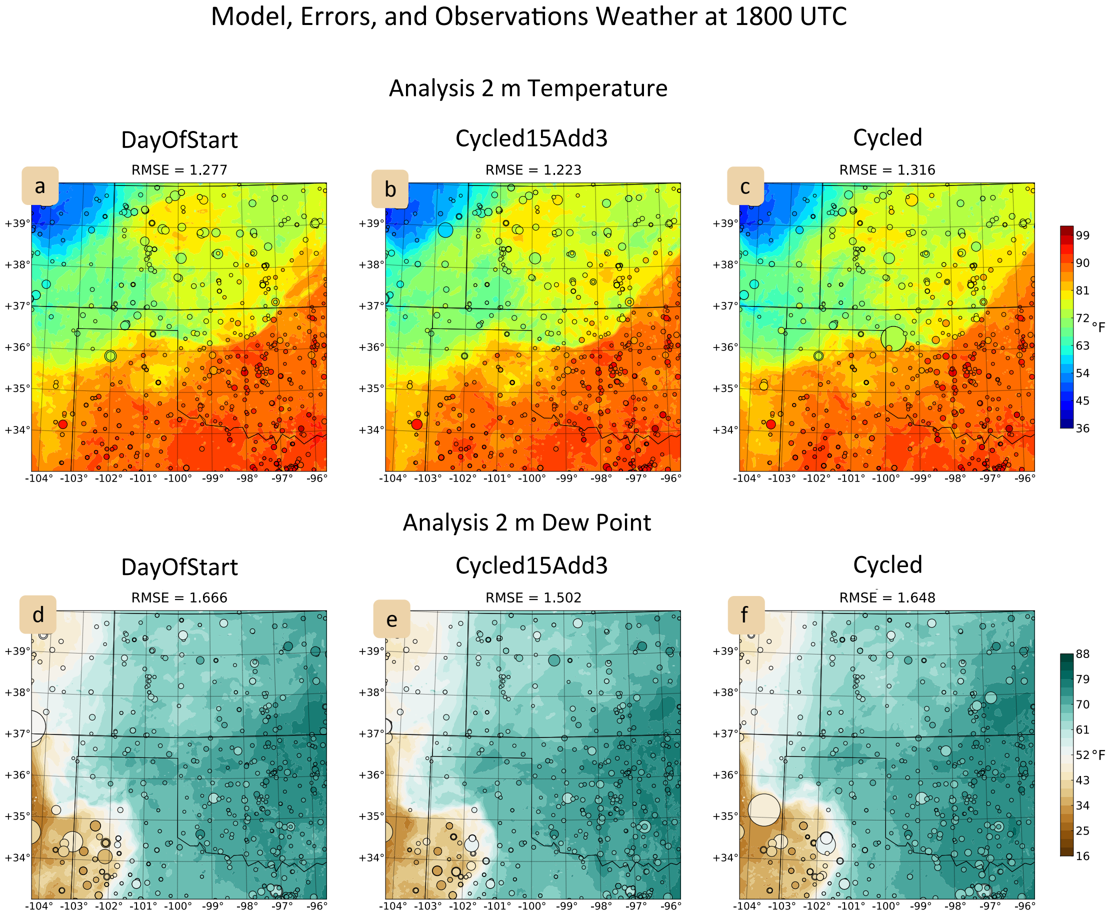


Figure 65. Model, Errors, and Observations Weather at 1800 UTC. The contours are the analysis mean surface temperature for DayOfStart (a), Cycled15Add3 (b), and Cycled (c), and the analysis mean surface dew point for DayOfStart (d), Cycled15Add3 (e), and Cycled (f). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

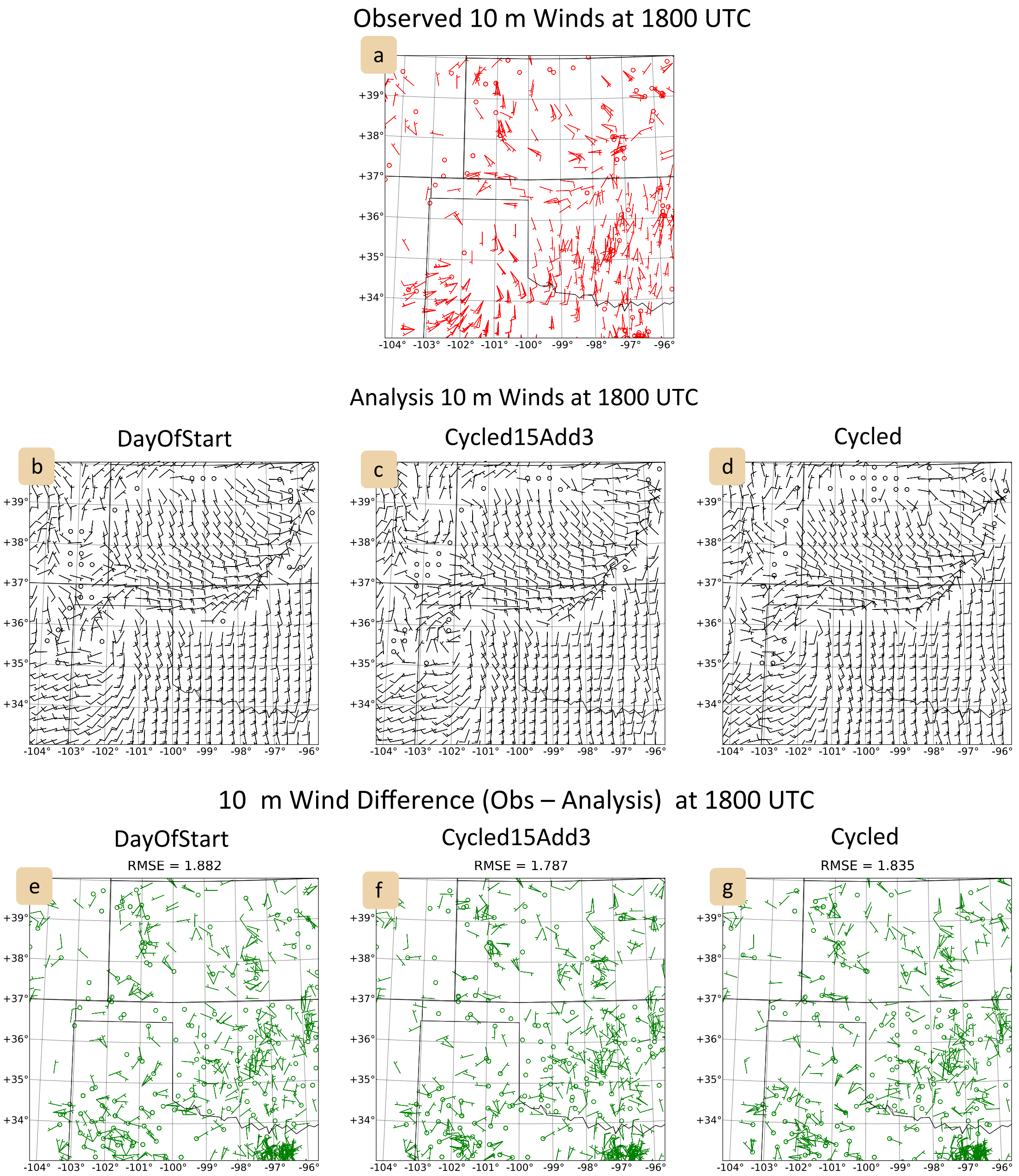
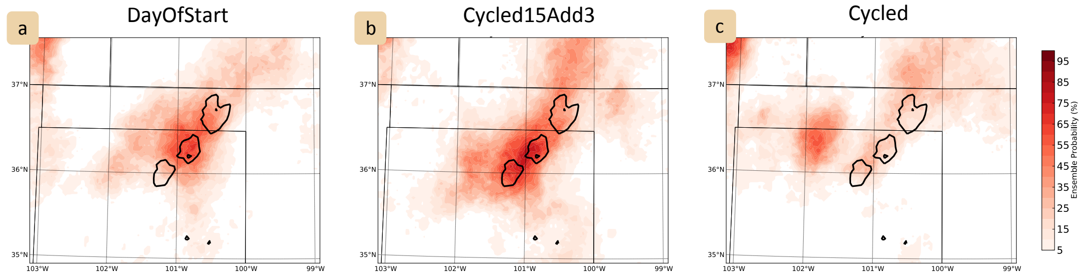


Figure 66. At 1800 UTC, the 10 m wind observations (a), and mean analysis for DayOfStart (b), Cycled15Add3 (c), and Cycled (d). Half barb = 5 m s^{-1}, full barb = 10 m s^{-1}. The observation minus analysis wind difference for DayOfStart (e), Cycled15Add3 (f), and Cycled (g). Half barb = 2.5 m s^{-1}, full barb = 5 m s^{-1}.

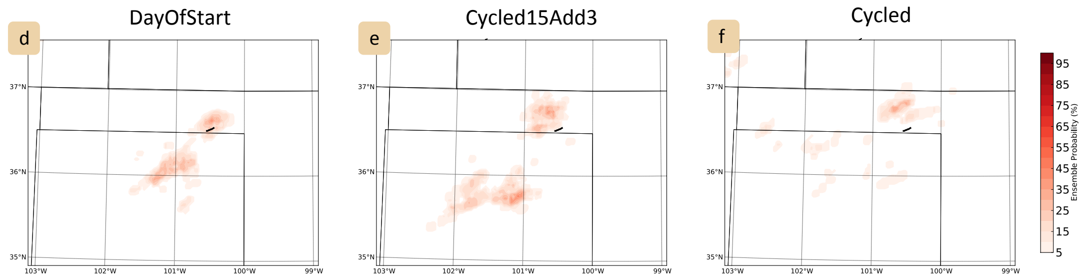
4.8.4 Forecast results

Ensemble forecasts are initiated from the analyses at 1800 UTC. Each experiment forecasts storms in the Oklahoma and Texas Panhandle at 2100 UTC (3-hour forecast), which indicates that the environments are supportive of convection (Figure 67). The highest ensemble reflectivity probabilities in the Cycled experiment are approximately 60%, and are shifted to the west of the three observed storms (Figure 67c). The Cycled15Add3 forecasted reflectivity probabilities are in a swath co-located with the observed storm locations. The highest reflectivity probabilities in the Cycled15Add3 forecast are greater than 70 %, which indicates that the Cycled15Add3 forecast has greater ensemble agreement than the Cycled forecast. The highest DayOfStart forecasted ensemble reflectivity probabilities are ~65%, and are co-located with the middle storm of the three observed storms. Thus, the DayOfStart forecast also has less ensemble agreement, compared to the Cycled15Add3 forecast.

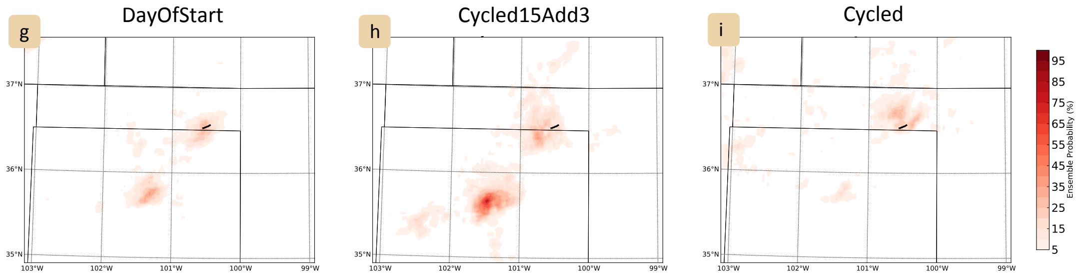
Reflectivity Probability at 2100 UTC



Updraft Helicity Probability at 2000-2100 UTC



0-1 km Vorticity Probability at 2000-2100 UTC



WSR-88D Rotation Tracks at 2000-2100 UTC

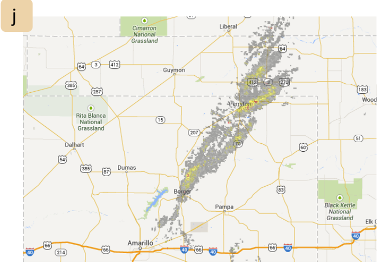


Figure 67. The 2100 UTC forecast ensemble probability of reflectivity greater than 30 dBZ for the DayOfStart (a), Cycled15Add3 (b), and Cycled (c) forecasts. The KAMA 30, 60 dBZ reflectivity observations are contoured in black. The 2000-2100 UTC forecast ensemble probability of updraft helicity greater than 100 m^2 s^{-2} for the DayOfStart (d), Cycled15Add3 (e), and Cycled (f) forecasts. The 2000-2100 UTC forecast ensemble probability of 0-1 km maximum vorticity greater than 0.0025 s^{-1} for the DayOfStart (g), Cycled15Add3 (h), and Cycled (i) forecasts. The observed tornado track is marked in black (d-i). The 2000-2100 UTC WSR-88D derived rotation tracks (j) from the NSSL archive.

The updraft helicity forecasts are also different for each experiment (Figure 67d-f). The Cycled15Add3 probability of updraft helicity has three clusters of rotating updrafts, and probabilities of 35% exist in same location as the observed Booker storm. The DayOfStart updraft helicity is orientated along the cold front, and probabilities of 30 % exist near the Booker storm. The Cycled forecast has some small areas of updraft helicity probability > 20% scattered across the region, and the highest probabilities in the forecast, 30%, are near the observed Booker storm.

The ensemble probabilistic forecasts of maximum vorticity exceeding 0.0025 s^{-1} in the 0-1 km layer are compared to examine the tornadic potential. The Cycled15Add3 forecast has probabilities of 35% near the observed tornado track, and has the highest probabilities, 55%, near the triple point. The Cycled forecast has scattered low probabilities (~20%), including an area of low probabilities just north of the outflow boundary and co-located with the location of the tornado. The DayOfStart forecast also has low probabilities (~20%) near the triple point and the observed Booker tornado. The 3-hour probabilistic vorticity forecast results show that the storm environments are supportive of storms with low-level rotation. However, the ensemble agreement for rotation near the observed tornado is fairly low and additional regions (triple point) are forecasted to have low-level vorticity values above the 0.0025 s^{-1} threshold as well.

4.8.5 Forecast sounding verification

The VORTEX2 soundings (see Section 4.3.6) can be used for verification of the near-storm environment. The temperature and relative humidity profiles in the Cycled15Add3 forecast have the lowest RMSE values and the Cycled forecast has the highest temperature and relative humidity RMSE values (Figure 68). Further, the Cycled15Add3 forecast has the smallest RMSE values (by a small margin) for u and v winds, except for the low-level v winds, which has a higher RMSE value than the other two experiments. The lower RMSE values indicate that the Cycled15Add3 forecast has a more accurate storm environment. This result provides evidence for why the probabilistic forecasts in the Cycled15Add3 are more accurate in terms of storm location and storm rotation (as discussed in the previous section).

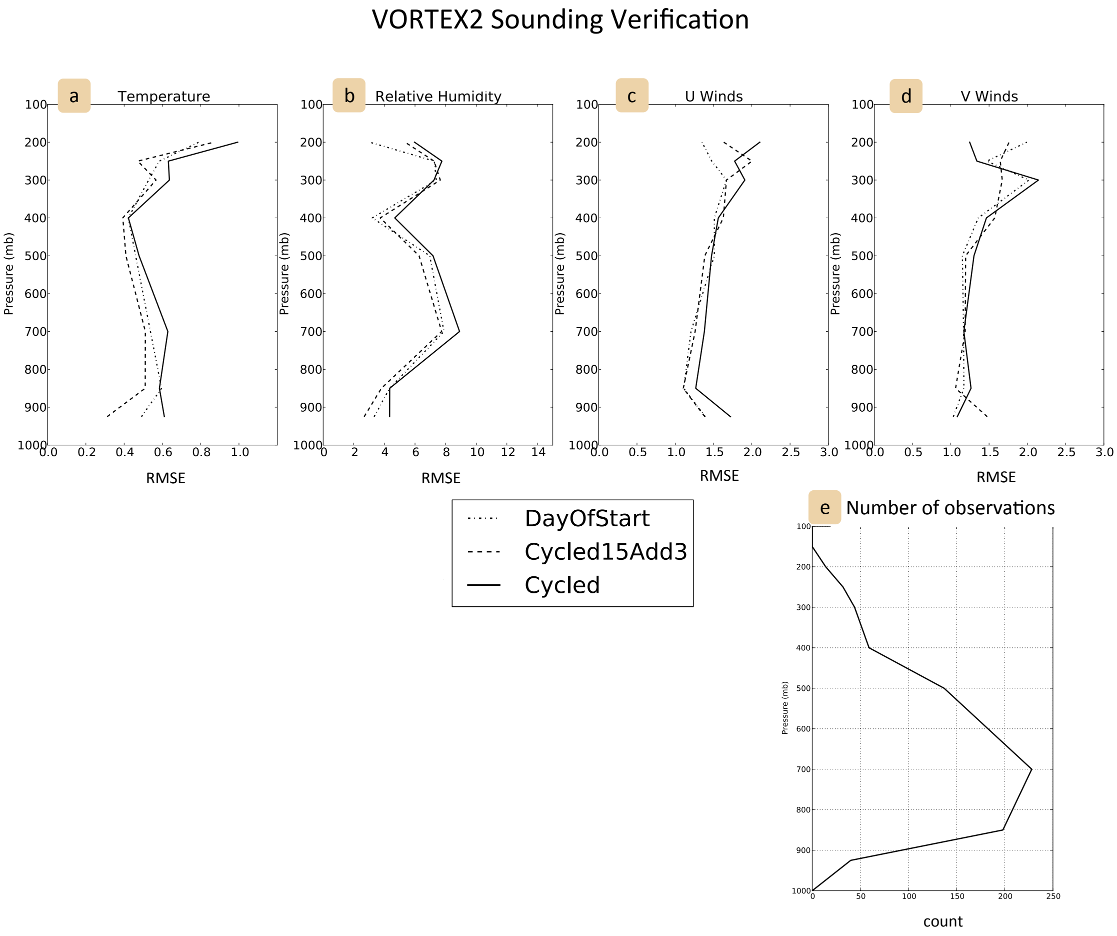


Figure 68. The VORTEX2 sounding temperature RMSE (a), relative humidity RMSE (b), u-wind component RMSE (c), v-wind component RMSE (d), and number of observations (e) for the DayOfStart (dashed-dot), CycledAdd3 (dashed), and Cycled (solid) forecasts.

The forecast errors compared to the VORTEX2 sounding launched at 1958 UTC is examined to investigate environmental differences in the inflow region of the Booker storm (see Section 4.3.6). The Cycled forecast, for most of the 50 members, does not contain the observed cool temperature profile near the surface (Figure 69e). The temperature error (observation minus forecast) profile clearly shows the warm bias in the cycled forecasts (Figure 69f). However, the DayOfStart and Cycled15Add3 ensemble forecasts do have ensemble members with cool low-level temperature profiles and a mean temperature profile that is close to the observed profile (Figure 69a,c). The Cycled15Add3 temperature forecast has less ensemble spread near the surface than the DayOfStart temperature forecast (Figure 69b,d).

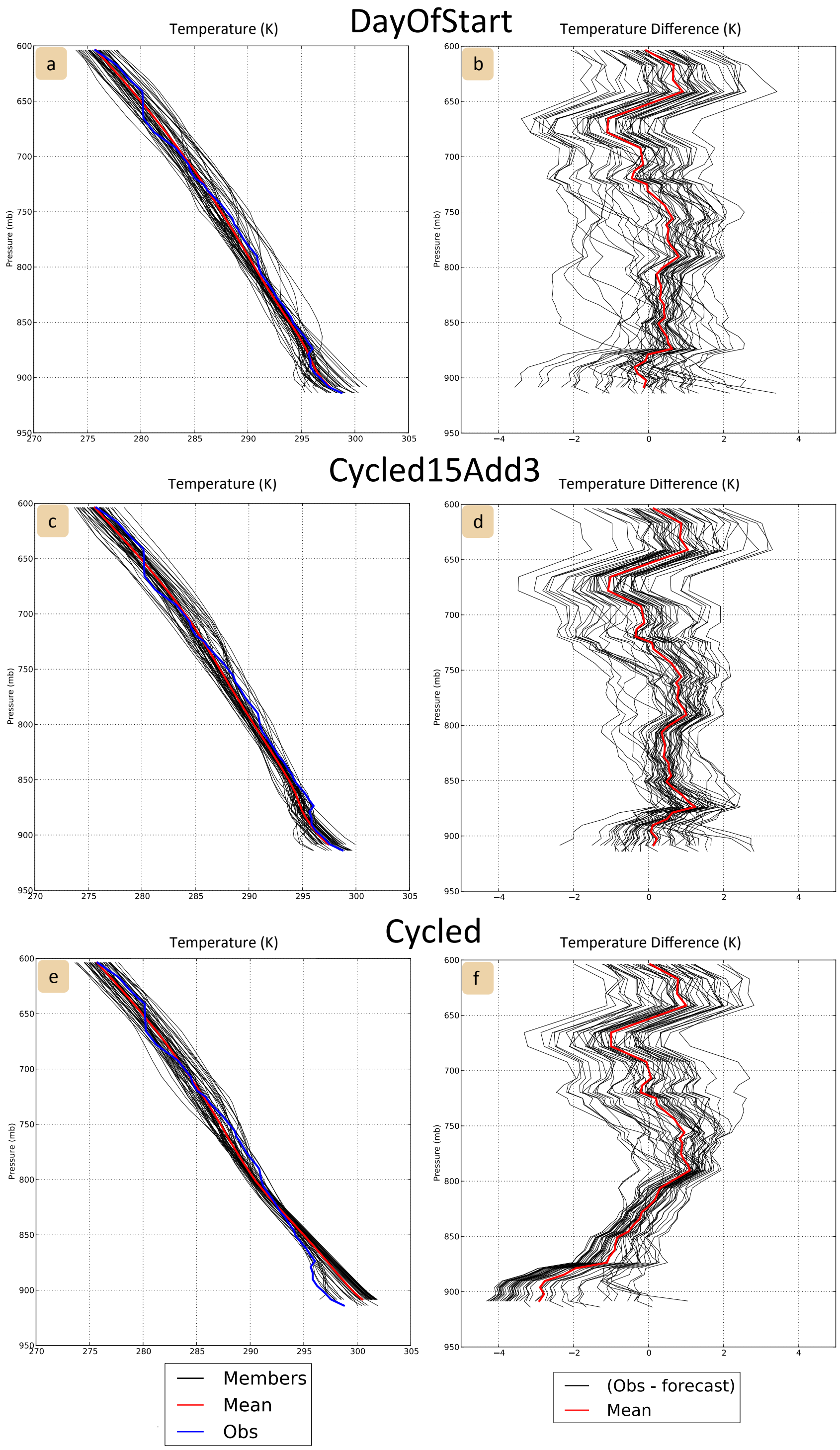


Figure 69. The low-level temperature observations from the 1958 UTC VORTEX2 sounding (blue), and the ensemble members (black) and ensemble mean (red) estimate of the low-level temperature at 2000 UTC for the DayOfStart (a), Cycled15Add3 (c) and Cycled (e) forecast. The low-level temperature difference (observation minus forecast) at 2000 UTC for the DayOfStart (b), Cycled15Add3 (d) and Cycled (f) forecast.

The observed wind profile for the 1958 UTC VORTEX2 inflow sounding has easterly surface winds, and backing winds with height (red lines in Figure 70). The Cycled ensemble forecast does not capture the magnitude of the easterly winds near the surface in the u wind profile (Figure 70e). Further, approximately half of the ensemble members in the Cycled forecast have v wind component profiles that are close to uniform magnitudes with height, 5 m s^{-1} from the surface to 800 mb (Figure 70f). Thus, the Cycled mean v wind profile does not follow the observed trend of magnitude increasing with height; v observations go from 0 to 13 m s^{-1} from the surface through 850 mb. The trend with height in the Cycled u and v profile does not clearly indicate backing winds. The Cycled15Add3 and DayOfStart ensemble forecast wind profiles are a better fit to the wind observations (Figure 70a-d). The Cycled15Add3 forecast contains the easterly surface winds in all of the ensemble member u profiles (Figure 70c). The DayOfStart forecasted profiles also have easterly low-level winds (Figure 70a). However, the magnitudes of some DayOfStart members are weak (< 5 m/s), which causes the mean u profile to have weaker easterly winds when compared to the observed and the Cycled15Add3 profile. The v winds for approximately half of the ensemble members in the DayOfStart forecast are southerly at the surface, and increase in magnitude slowly with height to 5 m s^{-1} at approximately 850 mb. The Cycled15Add3 v wind values increase in magnitude with height closer to the ground. Most Cycled15Add3 ensemble members v wind values is > 5 m s^{-1} at 875 mb. Thus, the Cycled15Add3 ensemble mean v profile is closer in magnitude to the observed profile than the DayOfStart mean v profile. The mean Cycled15Add3 wind profile is the closest to the observed low-level winds and the backing winds with height.

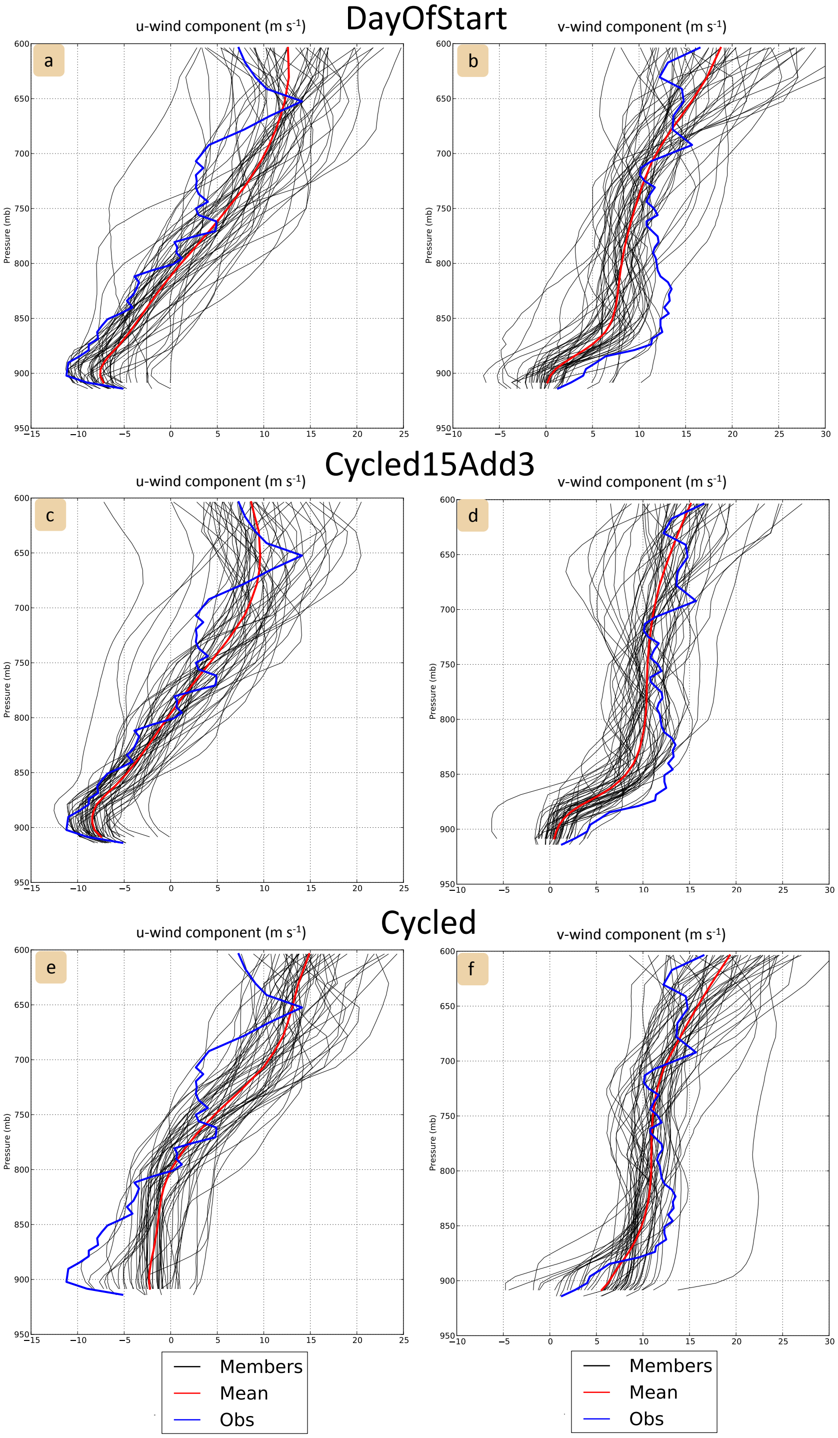


Figure 70. The low-level wind component observations from the 1958 UTC VORTEX2 sounding (blue), and the ensemble members (black) and ensemble mean (red) estimate of the low-level wind components at 2000 UTC for the DayOfStart (a,b), Cycled15Add3 (c,d) and Cycled (e,f) forecast.

In summary, the Cycled15Add3 forecast has the most accurate near-storm environment and lowest errors for the low-level profile of the Booker storm inflow air. The errors in the Cycled experiment forecast are the largest near the surface. This result suggests accumulated bias near the surface, which may be caused by too little ensemble spread.

4.8.6 1-Hour forecast results

The Warn-on-Forecast project is focused on the 0- to 1-hour forecast timeframe. This section compares the background ensemble initialization experiments for 1-hour forecasts to investigate the differences for short lead times. Ensemble forecasts are initiated from the analyses at 2000 UTC.

Similar to the analyses at 1800 UTC (previously discussed), the 2000 UTC analyses of the three experiments have small differences in the location and character of the mesoscale forcing features (front, outflow boundary, and dryline). The Cycled15Add3 surface temperature and dew point analysis has the lowest RMSE values (Figure 71). Near the forcing features, the Cycled15Add3 fit to the observations is better than the fit in the Cycled and DayOfStart analyses, which is represented by the smaller dots in Figure 71. All three analyses have easterly winds associated with the outflow boundary in the northeast corner of the Texas Panhandle at 2000 UTC (Figure 72). The Cycled wind analysis has the lowest RMSE. However, the front – outflow boundary intersection depicted by the shift from easterly winds to calm winds is well represented in the Cycled15Add3 wind analysis near 36.5° latitude and -101.5° longitude.

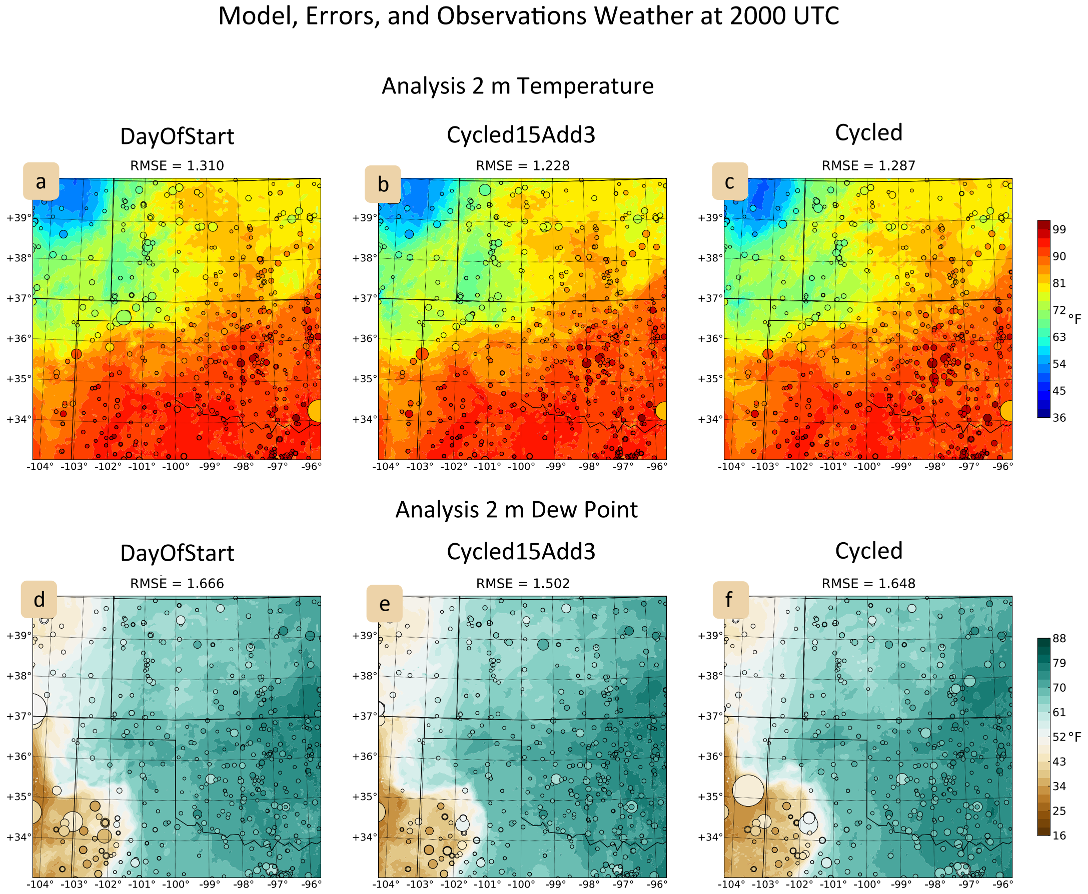


Figure 71. Model, Errors, and Observations Weather at 2000 UTC. The contours are the mean forecast surface temperature for DayOfStart (a), Cycled15Add3 (b), and Cycled (c), and the mean forecast surface dew point for DayOfStart (d), Cycled15Add3 (e), and Cycled (f). The dot sizes are the model errors. The dot colors are the temperature or dew point observations.

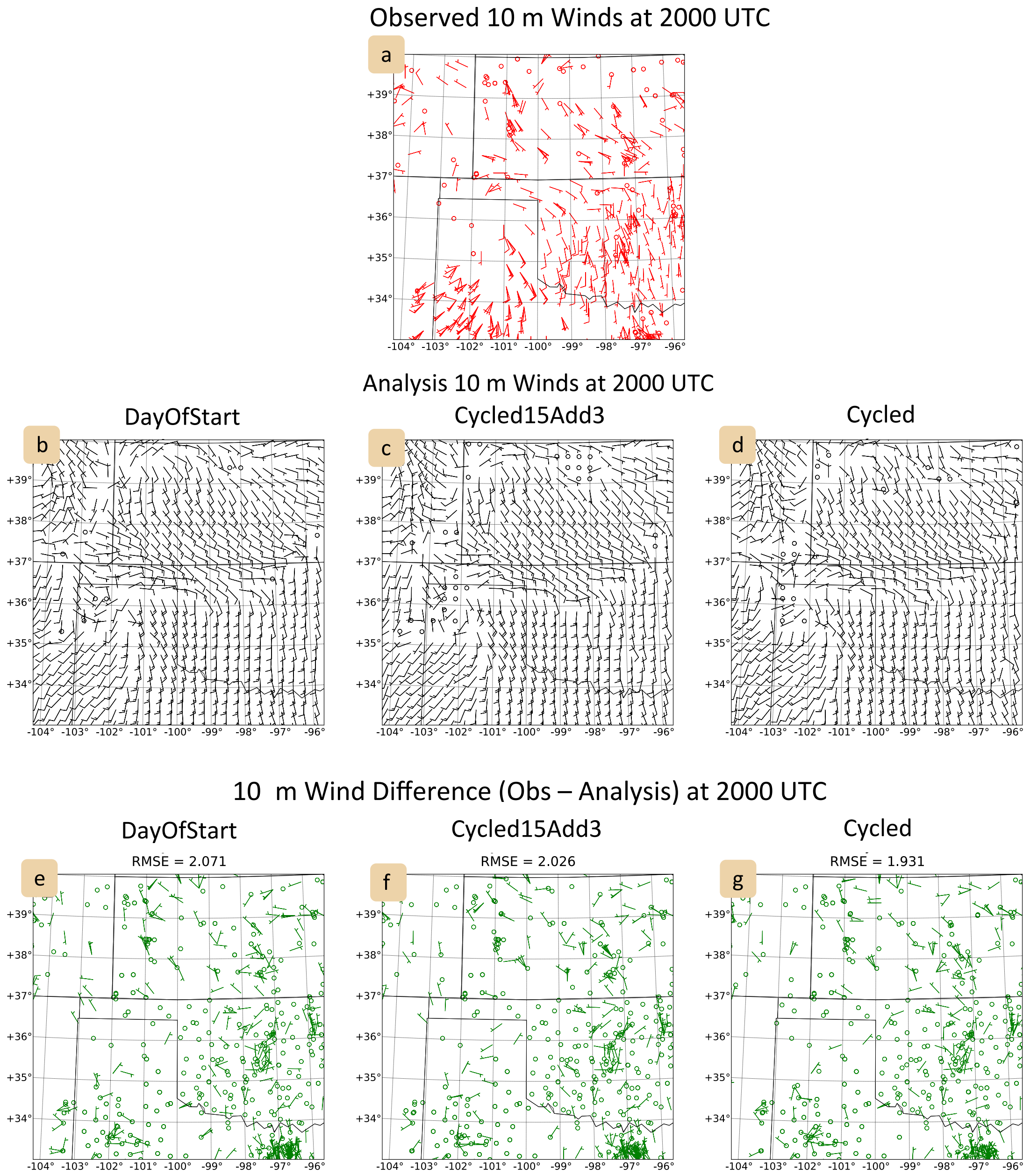


Figure 72. At 2000 UTC, the 10 m wind observations (a), and mean forecast for DayOfStart (b), Cycled15Add3 (c), and Cycled (d). Half barb = 5 m s^{-1}, full barb = 10 m s^{-1}. The observation minus forecast wind difference for DayOfStart (e), Cycled15Add3 (f), and Cycled (g). Half barb = 2.5 m s^{-1}, full barb = 5 m s^{-1}

The 1-hour ensemble probabilistic reflectivity forecasts have better ensemble agreement and more accurate storm locations than the 3-hour forecast discussed in the previous section (Section 4.7.4). All three experiments have high ensemble probabilities (>75 %) for storms near the observed storm locations (Figure 73). Thus, each forecast indicates the potential for storms near the Booker storm. However, the forecasted high ensemble reflectivity probabilities also extend to the northeast, beyond the location of the observed storms. The displacement of the high probabilities to the northeast of the observed storms is most pronounced in the Cycled forecast. The Cycled forecast also has probabilities for storms in Central Oklahoma, which did not occur. The ensemble reflectivity probabilities in the Cycled15Add3 forecast reach 100% co-located with the southern most of the observed storms. The DayOfStart forecast also has 100% probabilities for reflectivity > 30 dBZ, slightly offset to the southwest of the southern most storm. Each of the experiments has reflectitivy probabilities associated with the observed storms in Northwest Kansas, although they are missing the orientation of the storms.

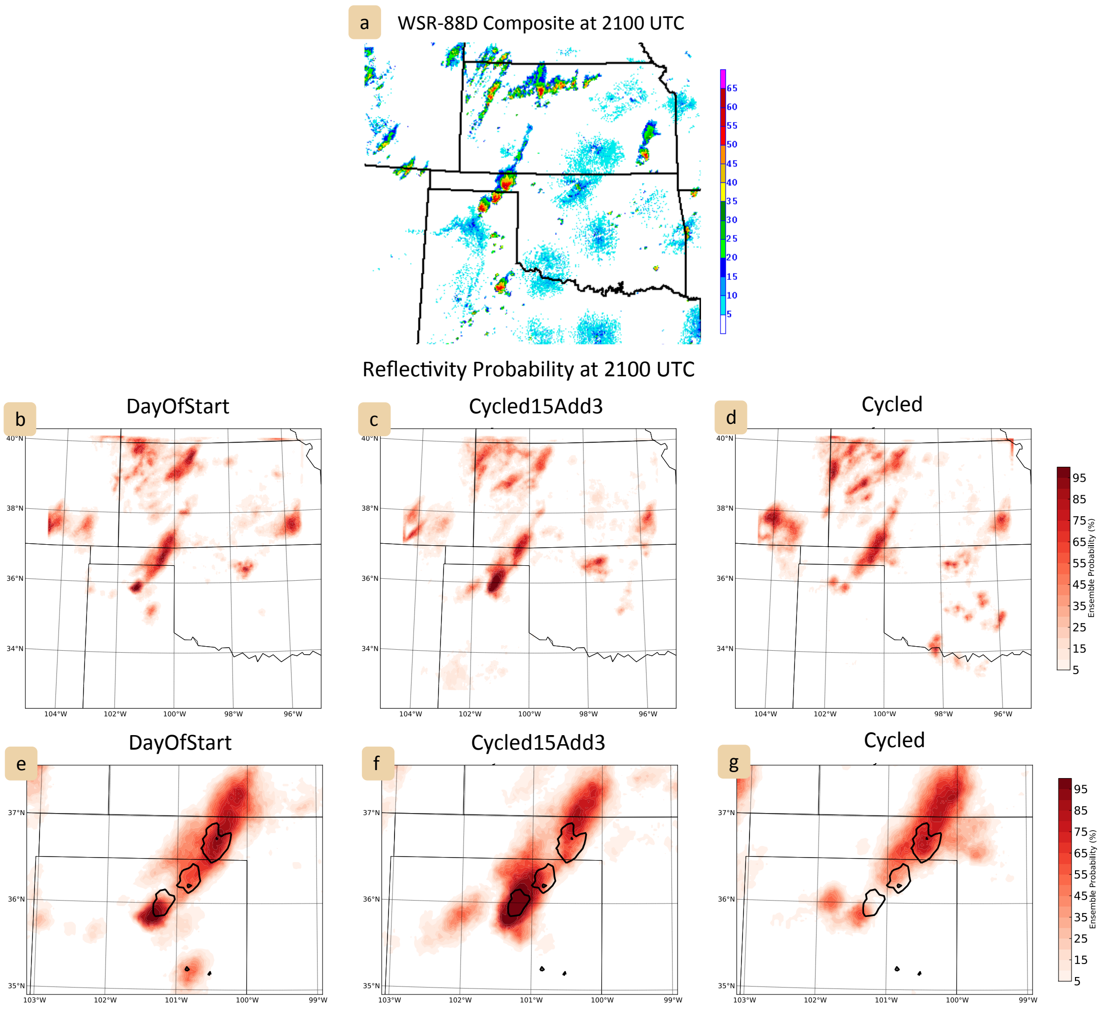


Figure 73. WSR-88D composite reflectivity at 2100 UTC (a). The 2100 UTC forecast ensemble probability of reflectivity greater than 30 dBZ for the DayOfStart (b,e), Cycled15Add3 (c,f), and Cycled (d,g) forecasts. The KAMA 30, 60 dBZ reflectivity observations are contoured in black (e-g).

The ensemble probability for updraft helicity indicate the presence of rotating storms in each model forecast over the location of the Booker storm (Figure 74). Further, the probabilistic forecasts for low-level vorticity, indicate forecasted tornadic potential co-located with the Booker tornado track. The Cycled forecast has the highest probability (~85%) of 0-1 km vorticity exceeding 0.0025 s^{-1} near the observed Booker tornado track compared to the DayOfStart (~80%) and Cycled15Add3 (~75%) (Figure

74f). Thus, the slight differences in the mesoscale environment have less impact in the 1-hour forecasts compared to the 3-hour forecasts. The other area highlighted by the vorticity probabilities in all three experiments is associated with the triple point. In particular, the Cycled15Add3 forecast low-level vorticity probabilities are 70 % near the triple point.

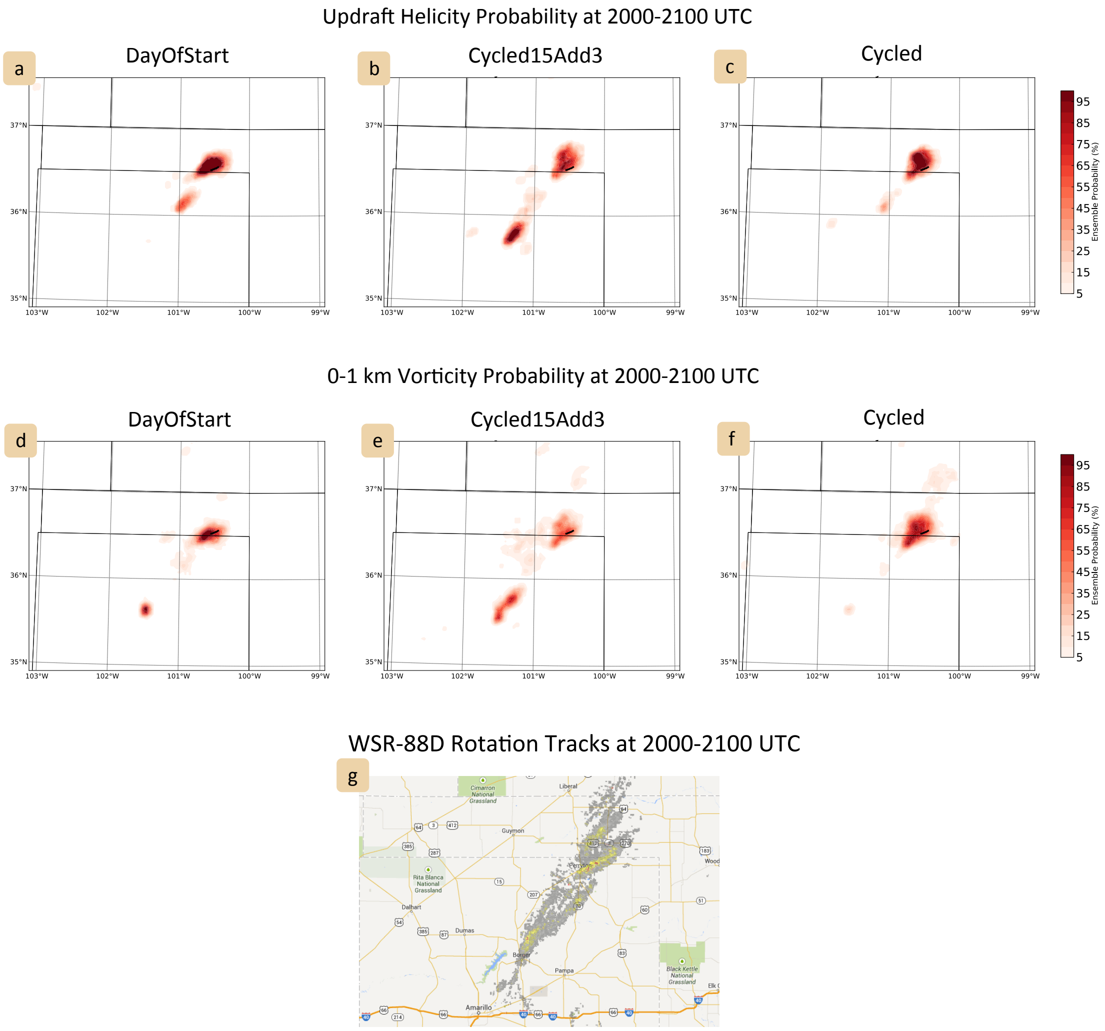


Figure 74. The 2000-2100 UTC forecast ensemble probability of updraft helicity greater than 100 m^2 s^{-2} for the DayOfStart (a), Cycled15Add3 (b), and Cycled (c) forecasts. The 2000-2100 UTC forecast ensemble probability of 0-1 km maximum vorticity greater than 0.0025 s^{-1} for the DayOfStart (d), Cycled15Add3 (e), and Cycled

(f) forecasts. The observed tornado track is marked in black (a-f). The 2000-2100 UTC WSR-88D derived rotation tracks (g) from the NSSL archive.

4.8.7 Conclusions

Three ensemble initialization choices are compared in this section. Despite the ensemble initializations taking place prior to 18 (or 20) hours of data assimilation, the differences in the ensemble background impact the storm environment and thunderstorm forecasts. Thus, the data assimilation analyses and forecasts are sensitive to the ensemble initialization.

The analyses of the storm environment at 1800 UTC show that the Cycled15Add3 fits the observation the best. The 3-hour forecast initiated at 1800 UTC indicate the potential for storms in the Panhandle region in all three experiments. The Cycled15Add3 ensemble forecasts of reflectivity, updraft helicity and vorticity match the observed Booker storm slightly better than the other experiments. The more accurate depiction of the convection can be attributed to a more accurate depiction of the storm's mesoscale environment. The VORTEX2 soundings verifications indicated that the Cycled15Add3 forecast has the lowest RMSE. Further, the 1958 UTC inflow sounding comparison shows the Cycled15Add3 forecast has the observed cool air near the surface and easterly low level winds that back with height. The DayOfStart forecast contains the same features but has slightly larger profile errors. The Cycled forecast is missing the key features in the observed inflow sounding profile. Thus, the Cycled15Add3 3-hour forecasts produce the most accurate representation of the 13 June storm environment. Differences between the initialization experiments' 3-hour ensemble forecasts are larger than the differences in the 1-hour ensemble forecasts. However, the 1-hour ensemble forecasts initialized with the Cycled15Add3 analysis is also a closer match to the observed Booker storm location, intensity, and rotation, relative to the DayOfStart and Cycled 1-hour ensemble forecasts.

4.9 Summary and Discussion

A multi-scale EnKF data assimilation and forecast system for the analysis and prediction of the environmental conditions and the severe convective storms is developed in this chapter. The multi-scale system is applied to 13 June 2010. The severe thunderstorms on 13 June occurred within a complex mesoscale environment, which developed subsequent to overnight mesoscale convective systems. Two boundary intersections, between a stationary front and a dryline, and the stationary front and an outflow boundary, played a significant role in the initiation and evolution of the severe thunderstorms during the afternoon of 13 June. One thunderstorm moved over the front – outflow boundary intersection and became tornadic near Booker, Texas.

The WRF model and DART data assimilation toolkit are used for the analysis and forecast of the storm environment and convection on 13 June. WRF forecasts are made with a mesoscale domain (15 km horizontal grid resolution), covering an area beyond the CONUS, and a convective allowing model (CAM) interior nest (3 km horizontal grid spacing), covering the Oklahoma-Texas Panhandle region. Conventional observations (including mesonets) collected from the MADIS database (and Oklahoma Climatological Survey), and quality-controlled observations from four WSR-88D radars are assimilated.

The questions that motivated this research can now be addressed:

- **Is data assimilation at convective-allowing model (CAM) resolution required to obtain an accurate storm environment?**

To investigate this question, a continuously cycled mesoscale data assimilation system is examined. The mesoscale assimilation uses only the 15 km horizontally spaced domain and executes 4-days of 6-hourly data assimilation cycles to assimilate conventional observations. The mesoscale analysis captures the observed synoptic pattern, and contains some representation of the core mesoscale elements: the stationary cold front, dryline, and outflow boundary. However, these key boundaries lacked the sharpness and correct placement to match the observed near-storm environment for the Booker storm. When the mesoscale analysis is used to initialize convective-allowing (3 km horizontal grid spacing) ensemble forecasts, the forecasts have rotating thunderstorms developing along the stationary front in Central Kansas, where few storms were observed, and does not forecast storms in the Texas Panhandle, where the most intense storms were observed. Thus, the CAM forecasts initialized by downscaling the mesoscale analyses provide general guidance on the potential for organized severe storms in the region, but are not sufficient for capturing the observed location and evolution of storms in the target region. These results indicate the need for multi-scale data assimilation within the CAM.

- **How does cycling frequency impact the near-storm environment and convective forecast?**

A multi-scale data assimilation system, which uses both the 15 and 3 km horizontally spaced domains and assimilates conventional and radar observations, is examined at two cycling frequencies: 6-hourly and hourly. Hourly cycling constrains the storm environment more than 6-hourly cycling. However, a large number of available conventional observations are located at the surface, which leads to little ensemble spread at the surface, and the surface analysis fit to the observations is similar for hourly and 6-hourly assimilation cycles. Using hourly cycling during the overnight hours on the 13th, qualitatively and quantitatively improves the forecast of the mesoscale environment, compared to starting the hourly cycling in the morning on the 13th. In addition, continuous hourly cycling on the 13th leads to the best forecast skill in terms of predicting the Booker storm location and rotation. Thus, hourly assimilation cycles are used to investigate the impact of radar observations and ensemble initializations.

- **Can infrequent (hourly) assimilation of radar observations improve the mesoscale environment and convective forecast?**

Yes, despite only assimilating one volume of WSR-88D observations (from 4 radars) every hour, the simultaneous assimilation of radar and conventional observations improves the accuracy of analyses and subsequent forecasts, compared to the assimilation of conventional observations alone. When radar data is assimilated with conventional observations, the analyses fit the observed overnight convection and the representation of the mesoscale surface boundaries (stationary cold front, dry line, and outflow boundary) are more accurate. Subsequent forecasts have more skill compared to forecasts initialized with the conventional observation assimilation analysis.

The addition of radar observations in the afternoon hours improves the analyses and forecasts of the storms, compared to the assimilation of conventional observations alone. However, it makes little impact on the storm environment. For this particular case, the assimilation of radar observations during the overnight hours is needed to generate an accurate storm environment.

- **Is multi-scale data assimilation cycling sensitive to the background ensemble used for initialization?**

Yes, differences in the background ensemble initialization impact the storm environment and thunderstorm forecasts. For this case, the most accurate forecast of the storm environment and Booker storm are obtained with the CAM that is initialized 24-hours prior to the event from a mesoscale domain that has been cycled for 4-days (Cycled15Add3). When the CAM and the mesoscale domain are initialized from a global analysis on the day of interest, the storm environment is slightly worse (DayOfStart), and when the CAM is cycled for 4-days with the mesoscale domain the forecasts are worse (Cycled).

- **What are the challenges and limitations in producing analyses/forecasts of a complex convective event?**

Maintaining appropriate values of ensemble spread is the biggest limitation encountered in this study. Deficiencies in ensemble spread at the surface limit the benefit from of hourly assimilation of dense observations. The development of techniques to improve ensemble spread, particularly in regions with dense observations is an important research area that needs to be addressed in future work.

The density of available observations is highly variable, which in addition to impacting the ensemble spread, leads to several data assimilation challenges associated with localization, observation errors, and observation operators. In particular, the high-resolution Doppler radar observations are smoothed via objective analysis for their use in the multi-scale data assimilation. However, much work is needed to improve radar observation processing and how the assimilation techniques use the radar observations.

The quality control of observations is also a challenge for data assimilation systems. In this study, analyses indicated the presence of a few bad conventional observations that passed the quality control. The ability to find and remove bias and errors in the observations before they are assimilated will improve the analyses.

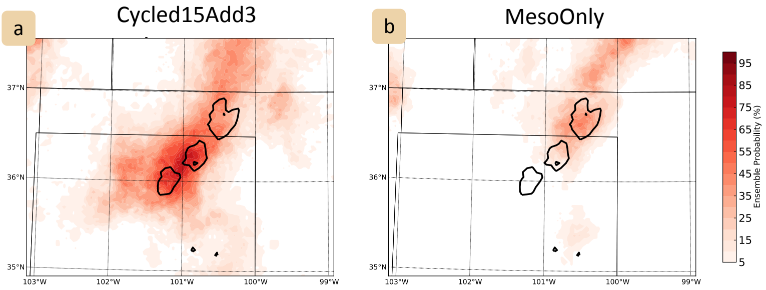
Verification of ensemble analyses and forecasts is another challenge. In this study, surface observations and sounding observations are used for verification of the storm environment and radar observations are used for the verification of the modeled storms. However, it is difficult to determine which metrics provide the most insight. Additional techniques to take advantage of the ensemble statistics could allow more detailed comparisons or expedite the verification process.

- **Does improving the storm environment via multi-scale data assimilation result in improved convective forecasts? How much improvement and for how long are these improvements realized?**

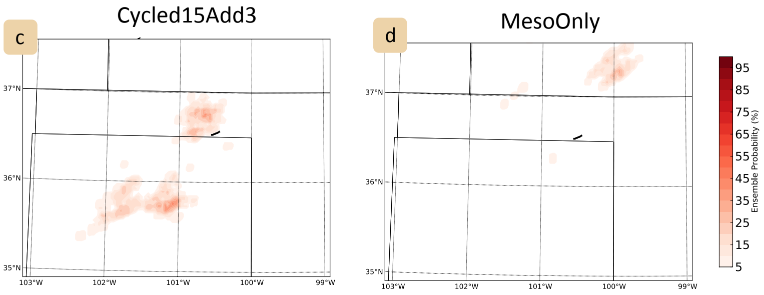
To address this question, the 13 June forecasts with the most skill and the worst skill are reviewed. The MesoOnly experiment (Section 4.5), which does not use multi-scale data assimilation (the CAM is initialized through downscaling), has the least skillful convective forecast. The Cycled15Add3 experiment (Section 4.8), which initialized the CAM from the mesoscale cycling analysis 24-hours prior to the Booker storm and executed multi-scale cycling on the 13th (hourly cycling with radar and conventional observations) has the most skillful convective forecast.

The 3-hour forecast valid at 2100 UTC for reflectivity, updraft helicity, and low-level vorticity probabilities from the MesoOnly and Cycled15Add3 experiments are shown in Figure 75. The Cycled15Add3 forecast probability of reflectivity exceeding 30 dBZ clearly indicates the potential for convection where the severe storms were observed. However, the MesoOnly forecast has low reflectivity probabilities and they are shifted to the northeast of the observed storms. The Cycled15Add3 updraft helicity and low-level vorticity probabilities indicate the potential for rotation associated with the modeled storms. However, the signal for rotation is weak in the MesoOnly forecast and shifted to the northeast of the Booker storm.

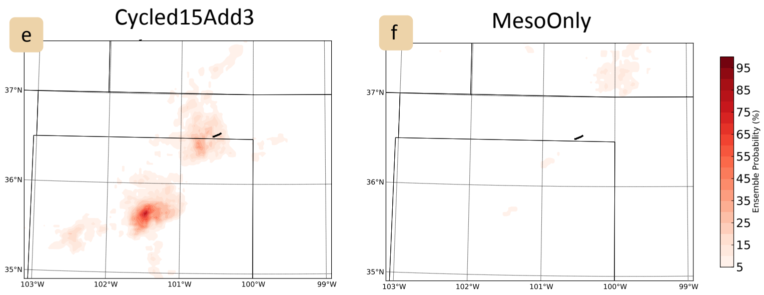
Reflectivity Probability at 2100 UTC



Updraft Helicity Probability at 2000-2100 UTC



0-1 km Vorticity Probability at 2000-2100 UTC



WSR-88D Rotation Tracks at 2000-2100 UTC

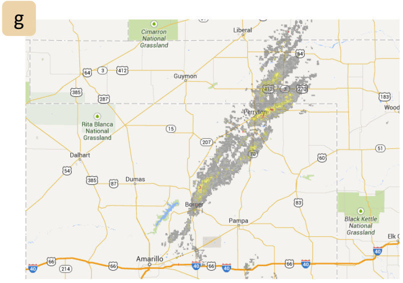


Figure 75. The 2100 UTC forecast ensemble probability of reflectivity greater than 30 dBZ for the Cycled15Add3 (a) and MesoOnly (b) forecasts. The KAMA 30, 60 dBZ reflectivity observations are contoured in black. The 2000-2100 UTC forecast ensemble probability of updraft helicity greater than 100 m^2 s^{-2} for the Cycled15Add3 (c) and MesoOnly (d) forecasts. The 2000-2100 UTC forecast ensemble probability of 0-1 km maximum vorticity greater than 0.0025 s^{-1} for the Cycled15Add3 (e) and MesoOnly (i) forecasts. The observed tornado track is marked in black (d-i). The 2000-2100 UTC WSR-88D derived rotation tracks (j) from the NSSL archive.

These comparisons highlight the advantages of multi-scale data assimilation compared to mesoscale data assimilation for predicting severe storms. All of the multi-scale analysis experiments investigated in this chapter demonstrate superior forecast skill than the mesoscale analysis system. Further, the multi-scale results provide evidence that improved analysis of the storm environment leads to improved convective forecast skill. This result emphasizes the importance of the data assimilation system used to initialize forecasts.

The experiments compared in this chapter show that multi-scale data assimilation improves convective forecasts on 13 June, which indicates the potential of a mulit-scale approach. Building on the results presented, more work is needed to implement a multi-scale analysis and forecast system for the Warn-on-Forecast mission. In particular, future studies should investigate more cases that have complex environments and continue to improve the multi-scale techniques.

Chapter 5: Summary

Two ensemble data assimilation challenges are investigated in this dissertation to improve the initial conditions for convective-allowing model forecasts. Chapter 3 compares the Local Ensemble Transform Kalman Filter (LETKF) and the Ensemble Square Root Filter (EnSRF) for convective-scale Doppler radar data assimilation, to determine if one of the methods consistently produces more accurate analyses and forecasts. Second, data assimilation strategies are investigated to improve the analysis of a complex mesoscale near-storm environment. Several configurations of a multi-scale data assimilation and forecast system are implemented for 13 June 2010. Special observations from the VORTEX2 field program are used to help quantify the accuracy of the forecasts. A brief summary of the primary findings of this dissertation is provided in this chapter.

5.1 LETKF versus EnSRF

Chapter 3 describes a quantitative and qualitative comparison between the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al. 2007) and the Ensemble Square Root Filter (EnSRF; Whitaker and Hamill 2002). The major difference between these two approaches is the assimilation pattern and the application of observations localization. The LETKF assimilates all observations that impact the state vector at a given location simultaneously, and applies localization via inflating the observation error covariance (i.e., "R-localization"). The EnSRF assimilates observations sequentially, to generate analysis increments for multiple state vector locations from an individual observation, and applies a localization function to the background error covariance (i.e. “B-localization”). The analyses and forecasts from each method are compared using a convective-scale (1 km horizontal grid spacing) cloud model with Doppler radar data assimilation for an Observing System Simulation Experiment (OSSE) and a real-data case, the 8 May 2003 Oklahoma City, Oklahoma tornadic supercell (Burgess 2004; Hu and Xue 2007; Romine et al. 2008; Dowell and Wicker 2009; Dowell et al. 2011; Yussouf et al. 2013).

The major findings for this study are:

- The OSSE is used to perform localization length sensitivity tests, and when comparing the LETKF and the EnSRF a smaller localization cutoff length is used for the R-localization to account for the difference in the effective localization length between R- and B-localization (HW2013).
- The OSSE results suggest longer localization cutoff lengths should be used compared to the values commonly used in the literature for convective-scale radar data assimilation. For B-localization (R-localization), a horizontal cutoff length of 12 km (9 km) and a vertical cutoff length of 6 km (4.5 km) are used.
- The choice of localization cutoff length impacts the wind and temperature analyses, and has slightly smaller impact on hydrometeor state variables. The EnSRF is slightly more sensitive than the LETKF to the localization cutoff length.
- The OSSE performance of the EnSRF with B-localization is nearly the same as that with R-localization.

- The EnSRF method produces larger analysis errors than the LETKF, as measured by the difference total energy and hydrometer difference total energy in the OSSE.
- The LETKF appears to provide slightly better dynamic balance, as measured by the noise seen in surface pressure tendencies in the OSSE.
- When only Doppler velocity observations are assimilated in the OSSE, the filters produce very similar analyses. This indicates that analysis differences arising from the different assimilation algorithms is larger when the observation operator is nonlinear, as with radar reflectivity.
- Assimilation of real radar observations from the 8 May 2003 Oklahoma City supercell indicates the LETKF and the EnSRF analyses and forecasts are qualitatively similar.
- Changing the random number seed value used to initialize the numerical perturbations added to each ensemble member reveals that the analysis and forecast differences between the EnSRF and the LETKF in the real data case are similar in magnitude to the differences that arise from the sampling variability associated with a finite ensemble. Therefore, the difference in accuracy between the filters does not appear to be of great practical importance.
- **The results strongly support the use of either the LETKF or the EnSRF as viable convective-scale radar data assimilation algorithm.**

5.2 Multi-scale data assimilation

Chapter 4 describes the development of a multi-scale data assimilation and forecast system, which is applied to the complex convective environment from 13 June 2010. The multi-scale system simultaneously assimilates Doppler radar and conventional observations, to analyze the primary synoptic and mesoscale convective forcing features and the convection itself. Convective forecasts initialized with a mesoscale analysis system are compared to the convective forecast initialized from various configurations of a multi-scale analysis system. Cycling frequency, observation type, and initialization of the background ensemble are varied to determine optimal settings for a multi-scale analysis system. A mesoscale WRF model domain (15 km horizontal grid spacing), and a nested convective-allowing WRF model domain (3 km horizontal grid spacing), are used along with the DART data assimilation toolkit for the analysis and forecast of the storm environment and convection on 13 June 2010.

The major findings for this study are:

- A mesoscale-only data assimilation system captures the observed synoptic pattern. However, the location and amplitude of mesoscale surface boundaries in the pre-storm environment have significant errors. Thus, downscaling the mesoscale analysis grid (15 km) to a CAM grid (3 km) leads to inaccurate CAM forecasts.
- In a multi-scale data assimilation system, hourly cycling improves the convective environment as measured by the fit to VORTEX2 soundings. Further, the hourly reflectivity analysis and subsequent forecasts qualitatively match the WSR-88D observations better compared to 6-hourly analysis cycles.

- Hourly multi-scale cycling that begins at 0000 UTC on the 13th outperforms hourly multi-scale cycling that begins at 1200 UTC on the 13th, which emphasizes the importance of accurately representing the overnight convection for the 13 June case.
- The hourly simultaneous assimilation of radar and conventional observations on 13 June improves the mesoscale environment when compared to the assimilation of conventional observations alone. The outflow boundary is poorly represented in the conventional observation assimilation experiment. Further, the forecasts initialized with the hourly simultaneous assimilation of radar and conventional observations have a much stronger signal for rotating storms near the observed tornadic Booker storm.
- When radar observations of the storms of interest are added to the conventional observation assimilation only during the afternoon of 13 June (after 1800 UTC), the storm analyses are somewhat improved. However, the radar data have little impact on the convective environment, and the forecast skill is worse than the experiment that assimilated radar and conventional observations hourly beginning at 0000 UTC on 13 June.
- The multi-scale analysis and forecast system is sensitive to the background ensemble initialization. For this study, the most accurate forecast of the storm environment and Booker storm are obtained with the CAM that is initialized 24-hours prior to the event via downscaling of a mesoscale domain that has been cycled for several days, and uses multi-scale hourly cycling with radar and conventional observations leading up to the forecasts.

- **A multi-scale data assimilation system produces a more accurate analysis and superior forecasts compared to a mesoscale-only data assimilation system.**
- **Convective forecasts are sensitive to the storm environment, and forecast skill is improved when a more accurate storm environment is used for initialization.**

5.3 Future Work: Implications for a Warn-on-Forecast system

The results discussed above have implications for the proposed data assimilation analysis and short-term (approximately 1-hour) forecast Warn-on-Forecast system (Stensrud et al. 2009, 2013).

Since the EnSRF and the LETKF show similar assimilation performance, either filter can be used for convective-scale data assimilation and future work can investigate the differences in scalability between the filters. The ability to interpolate the analysis perturbation weights from the LETKF analysis (Yang et al. 2009) may increase the efficiency for high-resolution grids and should be investigated for radar data assimilation in the future.

Since the 13 June convective forecasts are sensitive to the storm environment, future work should focus on using fully heterogeneous environments in a Warn-on-Forecast system. Specifically, the use of a multi-scale data assimilation approach will improve convective forecast skill. Implementing hourly analysis cycles, which assimilate both conventional and Doppler radar observations should provide an accurate mesoscale environment and lead to improved convective forecasts. The difficulty in maintaining ensemble spread in regions with dense observations needs to be addressed in future studies. Given the large number of configuration parameters in a multi-scale data assimilation prediction system, as well as the wide variety of pre-storm environments that generate severe weather in the U.S. each year, the testing and development of this and more sophisticated multi-scale data assimilation systems will remain an active area of research for scientists during the coming decade.

References

Abbe, C., 1901: The physical basis of long-range weather forecasts, *Mon. Weather Rev.*, **29,** 551–561.

Aksoy, A., D. C. Dowell, and C. Synder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Weather Rev.*, **137,** 1805-1824.

Aksoy, A., D. C. Dowell, and C. Snyder, 2010: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts. *Mon. Wea. Rev.*, **138**, 1273-1292.

Alexander, C. R., S. S. Weygandt, T. G. Smirnova, S. Benjamin, P. Hofmann, E. P. James, and D. A. Koch, 2010: High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2010 convective season. Preprints, *25th Conf. on Severe Local Storms, Denver, CO,* 9.2.

Anderson, Jeffrey L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Anderson, J. L., and N. Collins, 2007. Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61**, 72–83.

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, and R. Torn, 2009: The data assimilation research testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296.

Anderson, J. L., 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359– 2371.

Atkins, N. T., M. L. Weisman, and L. J. Wicker, 1999: The influence of preexisting boundaries on supercell evolution. *Mon. Wea. Rev.*, **127**, 2910–2927.

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

Barker, D. M., and Coauthors, 2012: The Weather Research and Forecasting model's community variational/ensemble Data Assimilation system: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Bowler, N. E., J. Flowerdew, and S. R. Pring, 2012: Tests of different flavours of EnKF on a simple model. *Q. J. Meteor. Soc.*, **139**, 1505-1519.

Bjerknes, V., 1904: Das problem der wettervorhersage, betrachtet vom standpunkte der mechanik und der physik. *Meteor. Zeit.*, **21**, 1–7, Translation by Y. Mintz, 1954: The problem of weather forecasting as a problem in mechanics and physics. Los Angeles. Reprinted in Shapiro and Grøna˚s (1999) 1–4.

Brooks, H. E., C. A. Doswell, and R. A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. Forecasting*, **7**, 120–132.

Brooks, H. E., and C. A. Doswell. 1993: New technology and numerical weather prediction – a wasted opportunity? *Weather*, **48**, 173-177.

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

Burgess, D. W., 2004. High resolution analyses of the 8 May 2003 Oklahoma City storm. Part I: Storm structure and evolution from radar data. Preprints, 22nd Conf. on Severe Local Storms, Hyannis, MA, Amer. Meteor. Soc., 12.4.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Weather Rev.*, **133**, 3081-3094.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface– hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585.

Clark, A. J., W. A. Gallus, and T-C Chen, 2007: Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models. *Mon. Wea. Rev.*, **135**, 3456–3473.

Clark, A. J., W. A. Gallus, M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140.

Clark, A. J., and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.*, **139**, 1410–1418.

Clark, A. J., and Coauthors, 2012a: An overview of the 2010 hazardous weather testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

Clark, A. J., J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012b: Forecasting tornado pathlengths using a three-dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113.

Clark, A. J., J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia, M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387–407.

Coniglio, M. C., D. J. Stensrud, and L. J. Wicker, 2006: Effects of upper-level shear on the structure and maintenance of strong quasi-linear mesoscale convective systems. *J. Atmos. Sci.* **63**, 1231–1252.

Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF model output for severe weather forecasting from the 2008 NOAA hazardous weather testbed spring experiment. *Wea. Forecasting*, **25**, 408–427.

Coniglio, M. C., J. Correia, P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862.

Cotton, W. R., and G. J. Tripoli, 1978: Cumulus convection in shear flow- three-dimensional numerical experiments. *J. Atmos. Sci.*, **35**, 1503-1521.

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

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**, 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**, 1152–1171.

Dawson II, D. T., L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2012: Impact of the environmental low-level wind profile on ensemble forecasts of the 4 May 2007 Greensburg, Kansas, tornadic storm and associated mesocyclones. *Mon. Weather Rev.*, **140**, 696–716.

Done, J., C. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett*., **5** (6), 110–117.

Doviak, R. J., P.S. Ray, R. G. Strauch, and L. J. Miller, 1976: Error estimation in wind fields derived from dual-Doppler radar measurement. *J. of Appl. Meteor.*, **15**, 868-878.

Dowell, D. C., and H. B. Bluestein, 1997: The Arcadia, Oklahoma, storm of 17 May 1981: Analysis of a supercell during tornadogenesis. *Mon. Wea. Rev*., **125**, 2562-2582.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and thermodynamic retrievals in the 17 May 1981 Arcadia, Oklahoma supercell: Ensemble Kalman filter experiments. *Mon. Weather Rev*. **132**, 1982-2005.

Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble forecasting and data assimilation. *J. Atmos. Ocea. Tech*., **26**, 911–927.

Dowell, D., G. S. Romine, and C. Snyder, 2010: Ensemble storm-scale data assimilation and prediction for severe convective storms. Preprints, *25th Conf. on Severe Local Storms, Denver, CO,* Amer. Meteor. Soc., 9.5.

Dowell, D. C., L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Weather Rev.*, **139**, 272–294.

Droegemeier, K. K., M. Xue, A. Sathye, K. Brewster, G. Bassett, J. Zhang, Y. Liu, M. Zou, A. Crook, V. Wong, R. Carpenter, and C. Mattocks, 1996a: Realtime numerical prediction of storm-scale weather during VORTEX '95: goals and methodology. Preprints, *18th Conf. on Severe Local Storms*, 19-23 Feb., Amer. Meteor. Soc., San Francisco, CA, 6-10.

Droegemeier, K. K., and Coauthors, 1996b: The 1996 CAPS spring operational forecasting period: Realtime storm-scale NWP, Part I: Goals and methodology. *Preprint, 11th Conf Num. Wea. Pred.,* Norfolk, VA., 194-296.

Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New dimension of NCEP Short-Range Ensemble Forecasting (SREF) system: Inclusion of WRF members. *Preprint*, WMO Expert Team Meeting on Ensemble Prediction System, Exeter, UK, 2 pp.

Dudhia, Jimy, 1993: A nonhydrostatic version of the Penn State–NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493–1513.

Dudhia, Jimy. 2014: A history of mesoscale model development. *Asia-Pacific J. of Atmos. Sci.,* **50** 121-131.

Durran, D. R., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not of practical importance. *J. Atmos. Sci.* In press.

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

Ehrendorfer, M., 2007. A review of issues in ensemble-based Kalman filtering. *Meteorol Z.*, **16**, 795-818.

Ek, M., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851.

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

Evensen, G. 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys Res-Oceans*, **99**, 10143-10162.

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

Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85**, 955–965.

Fujita, T. 1958: Mesoanalysis of the Illinois tornadoes of 9 April 1953. *J. Meteorology*, **15**, 288-296.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135**, 1846–1868.

Gal-Chen, T., 1978: A method for the initialization of the anelastic equations: Implications for matching models with observations. *Mon. Wea. Rev.*, **106**, 587-606.

Gal-Chen, T., and R. A. Kropfli, 1984: Buoyancy and pressure perturbations derived from dual-Doppler radar observations of the planetary boundary layer: Applications for matching models with observations. *J. Atmos. Sci.,* **41**, 3007-3020.

Gandin, L. S., 1963: Objective analysis of meteorological fields. *Gidrometeor. Isdaty., Leningrad.* [Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp.]

Gao, J., M. Xue, A. Shapiro, Q. Xu, and K. K. Droegemeier, 2001: Three-dimensional simple adjoint velocity retrievals from single-Doppler radar. *J. Atmos. Ocea. Tech.*, **18**, 26-38.

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

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Meteor. Soc.,*, **125**, 723-757.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522.

Hacker, J., S. Ha, C. Snyder, J. Berner, F. Eckel, E. Kuchera, M. Pocernich, S. Rugg, J. Schramm, and X. Wang, 2011: The U.S. Air Force weather agency's mesoscale ensemble: Scientific description and performance results. *Tellus A*, North America, **63**.

Holland, B., 2011: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter. MS Thesis. School of Meteorology, University of Oklahoma.

Holland, B. and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter. *Quart. J. Meteor. Soc.*, **139**, 758-770.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hu, M., and M. Xue, 2007: Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. *Mon. Wea. Rev.*, **135**, 507–525.

Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844.

Huang, X., A. Cederskov, E. Källén, 1994: A comparison between digital filtering initialization and nonlinear normal-mode initialization in a data assimilation system. *Mon. Wea. Rev.*, **122**, 1001–1015.

Hunt, B. R., E. J. Kostelich and I. Szunyogh, 2007: Efficient data assimilation of spatiotemporal chaos: A local ensemble transform Kalman filter. *Physics D*, **230**, 112-126.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res*., **113**, D13103.

Janjić, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Johnson, A., X. Wang, M. Xue, and F. Kong, 2011: Hierarchical cluster analysis of a convection-allowing ensemble during the hazardous weather testbed 2009 spring experiment. Part II: Ensemble clustering over the whole experiment period. *Mon. Wea. Rev.*, **139**, 3694–3710.

Johnson, A., and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Mon. Wea. Rev.*, **140**, 3054–3077.

Jon, Lil', and E. S. Boyz, 2002: Get low. *Kings of Crunk.* **1**, 19.

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

Kain, J. S., P. R. Janish, S. J. Weiss, R. S. Schneider, M. E. Baldwin, and H. E. Brooks, 2003: Collaboration between forecasters and research scientists at the NSSL and SPC: The spring program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806.

Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181.

Kain, J. S., S. J. Weiss, M. E. Baldwin, G. W. Carbin, D. A. Bright, J. J. Levit, and J. A. Hart, 2005: Evaluating high-resolution configurations of the WRF model that are used to forecast severe convective weather: The 2005 SPC/NSSL spring program. Preprints, *21st Conf. on Weather Analysis and Forecasting and 17th Conf. on Numerical Weather Prediction, Washington, DC*, Amer. Meteor. Soc., 2A.5.

Kain, J. S., S. J. Weiss, M. E. Baldwin, G. W. Carbin, J. J. Levit, D. R. Bright, and J. A. Hart, 2005: Evaluating high-resolution configurations of the WRF model that are used to forecast severe convective weather: The 2005 SPC/NSSL Spring Program. Preprints, *21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction, Washington, DC,* Amer. Meteor. Soc., 2A.5.

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

Kain, J. S., and Coauthors, 2008b: Severe-weather forecast guidance from the first generation of large domain convection-allowing models: Challenges and opportunities. Preprints, *24th Conf. on Severe Local Storms, Savannah, GA,* Amer. Meteor. Soc., 12.1.

Kain, J. S., and Coauthors, 2010: Assessing advances in the assimilation of radar data and other mesoscale observations within a collaborative forecasting–research environment. *Wea. Forecasting*, **25**, 1510–1521.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. of Basic Enginerring*, **82**, 35-45.

Kalnay, E., 2006: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 341 pp.

Klemp, J. B., and R. B. Wilhelmson, 1978: The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci*., **35**, 1070-1096.

Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2006: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparsion of coarse- and fine-grid experiments. *Mon. Wea. Rev.*, **134**, 807–833.

Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2007: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part II: Storm-scale experiments. *Mon. Wea. Rev.*, **135**, 759–782.

Kursinski, E. R., G. A. Hajj, J. T. Shofield, R. P. Linfied, and K. R. Hardy, 1997: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. *J. Geophys. Res.,* **102** (D19), 23 429–23 465.

Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. Stumpf, 2007: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor. Climatol.*, **46**, 288–305.

Lei, T., M. Xue, and T. Yu, 2009: Multi-scale analysis and prediction of the 8 May 2003 Oklahoma City tornadic supercell storm assimilating radar and surface network data using EnKF. Preprints, *13th Conf. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, Phoenix, AZ,* Amer. Meteor. Soc., 6.4.

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.*, **28**, 145–161.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Leith, C. E., R. H. Kraichnan, 1972: Predictability of turbulent flows. *J. Atmos. Sci.*, **29**, 1041–1058.

Lilly, D. K. 1990: Numerical prediction of thunderstorms – has its time come? *Quart. J. Meteor. Soc.,* **139.** 779-798.

Lin, X., and K. G. Hubbard, 2004: Sensor and electronic biases/ errors in air temperature measurements in common weather station networks. *J. Atmos. Oceanic Technol*., **21**, 1025–1032.

Liu, C., and M. W. Moncrieff, 2007: Sensitivity of cloud-resolving simulations of warm-season convection to cloud microphysics parameterizations. *Mon. Wea. Rev.*, **135**, 2854–2868.

Lemon, L. R., and C. A. Doswell, 1979: Severe thunderstorm and mesocyclone structure as related to tornadogensis. *Mon. Wea. Rev.,* **107**, 1184-1197.

Lorenc, A. C., and F. Rawlins, 2005: Why does 4D-Var beat 3D-Var? *Quart. J. Meteor. Soc.,* **613.** 3247-3257.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

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

Lynch, P., 2006: *The Emergence of Numerical Weather Prediction: Richardson's Dream*, Cambridge University Press, Cambridge, 279pp.

Lynch, P., 2008. The origins of computer weather prediction and climate modeling. *J of Computational Physics*, **227**, 3431-3444.

MacGorman, D., R., and Coauthors, 2008: TELEX The Thunderstorm Electrification and Lightning Experiment. *Bull. Amer. Meteor. Soc.*, **89**, 997–1013.

Maddox, R. A., L. R. Hoxit, and C. F. Chappell, 1980: A study of tornadic thunderstorm interactions with thermal boundaries. *Mon. Wea. Rev.,* **108** , 322–336.

Maddox, R. A., and C. A. Crisp, 1999: The Tinker AFB tornadoes of March 1948. *Wea. Forecasting*, **14**, 492–499.

Majcen, M., P. Markowski, Y. Richardson, D. Dowell, and J. Wurman, 2008: Multipass objective analyses of Doppler radar data. *J. Atmos. Oceanic Technol*., **25**, 1845–1858.

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194.

Markowski, P. M., E. N. Rasmussen, and J. M. Straka, 1998: The occurrence of tornadoes in supercells interacting with boundaries during VORTEX-95. *Wea. Forecasting*, **13**, 852–859.

Markowski, P. M., M. Majcen, Y. Richardson, J. Marquis, and J. Wurman, 2011: Characteristics of the wind field in a trio of nontornadic low-level mesocyclones observed by the Doppler on Wheels radars. *Electron. J. Severe Storms Meteor*., 6 (3).

Marquis, J., Y. Richardson, P. Markowski, D. Dowell, and J. Wurman, 2012: Tornado maintenance investigated with high-resolution dual-Doppler and EnKF analysis. *Mon. Wea. Rev*., **140**, 3–27.

Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys*., **20**, 851–875.

Miller, M. J., and R. P. Pearce, 1974: A three-dimensional primitive equation model of cumulonimbus convection. *Quart. J. Roy. Meteor. Soc*., **100**, 133-154.

Miyoshi, T., and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135**, 3841–3861.

Miyoshi, T., 2010: The gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535.

Miyoshi, T., and M. Kunii, 2012: The local ensemble transform Kalman filter with the weather research and forecasting model: Experiments with real observations. *Pure and Appl. Geophys.*, **169**, 321-333.

Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007.

Nerger, L., T. Janjic, J. Schroter, and W. Hiller, 2012: A regulated localization scheme for ensemble-based Kalman filters. *Quart. J. Meteor. Soc.*, **138**, 802-812.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56a**, 415–428.

Platzman, G. W., 1979: The ENIAC Computations of 1950—Gateway to Numerical Weather Prediction. *Bull. Amer. Meteor. Soc.*, **60**, 302–312.

Potvin, C. K., L. J. Wicker, D. Betten, M. I. Biggerstaff, and A. Shapiro, 2013: Comparison between dual-Doppler and EnKF storm-scale wind analyses: The 29-30 May 2004 Geary, Oklahoma, supercell thunderstorm. *Mon. Weather Rev.* **141**, 1612-1628.

Potvin, C. K., and L. J. Wicker, 2013: Assessing ensemble forecasts of low-level supercell rotation within an OSSE framework. *Wea. and Forecasting*, **28**, 940-960.

Qui, C.-J., and Q. Xu, 1992: A simple Adjoint method of wind analysis for single-Doppler data. *J. Atmos. Oceanic Technol.,* **9**, 588-598.

Ray, P. S., R. J. Doviak, G. B. Walker, D. Sirmans, J. Carter and B. Bumgarner, 1975: Dual-Doppler observations of a tornadic storm. *J. Appl. Meteor.,* **14**, 1521-1530.

Ray, P. S., C. L. Ziegler, W. Bumgarner, and R. L. Serafin, 1980: Single- and multiple-Doppler radr observations of tornadic storms. *Mon. Wea. Rev.*, **108**, 1607-1625.

Reich, H. A. Rhodin, and C. Schraff, 2011: LETKF for the nonhydrostatic regional model COSMO-DE. *Consortium for Small-Scale Modelling Newsletter,* **11**, 27-31.

Richardson, L. F., 1922: *Weather prediction by numerical process*, Cambridge University Press, Cambridge, xii+236pp. Reprinted by Dover Publications, New York, 1965, with a new Introduction by Sydney Chapman, xvi+236pp.

Romine, G. S., D. W. Burgess, and R. B. Wilhelmson, 2008: A dual- polarization-radar-based assessment of the 8 May 2003 Oklahoma City area tornadic supercell. Mon. Wea. Rev., **136**, 2849–2870.

Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284.

Sanderson, B. G., G. Brassington, 1998: Accuracy in the context of a control-volume model. *Atmos. Ocean*, **36**, 355-284.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev*., **98**, 875–883.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372.

Schwartz, C. S., G. Romine, K. R. Smith, and M. L. Weisman. 2014: Characterizing and optimizing precipitation forecasts from a convective-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting,* In press.

Shapiro, A., S. Ellis, and J. Shaw, 1995: Single-Doppler velocity retrievals with Phoenix II data: Clear air and microburst wind retrievals in the planetary boundary layer. *J. Atmos. Sci.,* **52**, 1265-1287.

Shapiro, A., P. Robinson, J. Wurman, J. Gao, 2003: Single-Doppler velocity retrieval with rapid-scan radar data. *J. Atmos. Ocea. Tech*. **20**, 1758-1775.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev*., **132**, 3019–3032.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the advanced research WRF version 3. NCAR Technical Note TN- 475+STR, 113 pp.

Skamarock, W. C., and M. L. Weisman, 2009: The impact of positive-definite moisture transport on NWP precipitation forecasts. *Mon. Wea. Rev*., **137**, 488–494.

Smith, T. M., K. L. Elmore, Shannon, and A. Dulin, 2004: A damaging downburst prediction and detection algorithm for the WSR-88D. *Wea. Forecasting*, **19**, 240–250.

Snyder, C., and F. Zhang, 2003. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Weather Rev.* **131**, 1663-1677.

Sobash, R., A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728.

Sobash, R. A., and D. Stensrud, 2013: The impact of covariance localization for radar data on EnKF analyses of a developing MCS: Observing system simulation experiments. *Mon. Wea. Rev.,* **141**, 3691-3709.

Sobash, R., 2013: Assimilation of radar and surface observations of a developing convective system: Observing system simulation and real-data experiments. PhD Thesis. School of Meteorology, University of Oklahoma.

Stensrud, D. J., J-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107.

Stensrud, D. J., 2007: *Parameterization schemes: Keys to understanding numerical weather prediction models.* Cambridge University Press, 459 pp.

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

Stensrud, D. J., and J. Gao, 2010: Importance of horizontally inhomogeneous environmental initial conditions to ensemble storm-scale radar data assimilation and very short-range forecasts. *Mon. Wea. Rev.*, **138**, 1250–1272.

Stensrud, David J., and Coauthors, 2013. Progress and challenges with Warn-on-Forecast. *Atmos. Research*, **12** , 2-16.

Stull, R. B., 1984: Transilient turbulence theory. Part I: The concept of eddy-mixing across finite distances. *J. Atmos. Sci.*, **41**, 3351–3367.

Sun, J., D. W. Flicker, and D. K. Lilly, 1991: Recovery of three-dimensional wind and temperature fields from simulated single-Doppler radar data. *J. Atmos. Sci.,* **48**, 876-890.

Sun, J., and N. A. Crook, 1998: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part II: Retrieval experiments of an observed Florida convective storm. *J. Atmos. Sci.*, **55**, 835–852.

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

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assesing a local ensemble Kalman filter: Perfect model experiments with the national centers for environmental prediction global model. *Tellus*, **57a**, 528–545.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Meteor. Soc.,* **113.** 1311-1328.

Talagrand, O., 1997: Assimilation of observations, an introduction. *J. of the Meteor. Soc. Of Japan,* **75**, 191-209.

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

Thompson, T. E., L. J. Wicker, and X. Wang, 2012: Impact from a volumetric radar-sampling operator for radial velocity observations within EnKF supercell assimilation. *J. Atmos. Oceanic Technol.*, **29**, 1417–1427.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square Root Filters*. *Mon. Wea. Rev.*, **131**, 1485–1490.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Weather Rev.* **133**, 1789-1807.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502.

Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963.

Torn, R. D., 2010: Performance of a mesoscale ensemble Kalman filter (EnKF) during the NOAA high-resolution hurricane test. *Mon. Wea. Rev.*, **138**, 4375–4392.

Torn, R. D., and C. A. Davis, 2012: The influence of shallow convection on tropical cyclone track forecasts. *Mon. Wea. Rev.*, **140**, 2188– 2197.

Toth, Z., E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Troen, I., and L. Mahrt, 1986: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. *Bound.- Layer Meteor.*, **37**, 129–148.

Tsai, C-C, S-C. Yang, and Y-C. Liou, 2012: Improving short-term QPFs over complex terrain with a WRF-LETKF radar data assimilation system: OSSEs on typhoon Morakot (2009). *5th EnKF Workshop, Albany, NY.*

Velden, C., and Coauthors, 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.*, **86**, 205–223.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.

Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131.

Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008b: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147.

Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2009: A comparison of the hybrid and EnSRF analysis schemes in the presence of model error due to unresolved scales. *Mon. Wea. Rev.*, **137**, 3219-3232.

Wang, X., 2011: Application of the WRF hybrid ETKF–3DVAR data assimilation system for hurricane track forecasts. *Wea. Forecasting*, **26**, 868–884.

Wang, S., M. Xue, and J. Min, 2013: A Four-Dimensional Asynchronous Ensemble Square-Root Filter (4DEnSRF) algorithm and tests with simulated radar data. *Q. J. R. Meteo. Soc.*, **139**, 805-819.

Wang, X., 2010: Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117.

Wang, X. and T. Lei, 2014: GSI-based Four-Dimensional Ensemble-Variational (4DEnsVar) data assimilation: Formulation and single resolution experiments with real data for NCEP global forecast system. *Mon. Wea. Rev.*, In press.

Weisman, M. L., and J. B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504–520.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.*, **130**, 1913-1924.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463–482.

Wicker, L. J., and W. C. Skamarock, 2002: Time-splitting methods for elastic models using forward time schemes. *Mon. Wea. Rev.*, **130**, 2088–2097.

Wood, V. T., R. A. Brown, and D. C. Dowell, 2009: Simulated WSR-88D velocity and reflectivity signatures of numerically modeled tornadoes. *J. Atmos. Oceanic Technol.*, **26**, 876–893.

Wurman, J., D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The Second Verification of the Origins of Rotation in Tornadoes Experiment: VORTEX2. *Bull. Amer. Meteor. Soc.*, **93**, 1147–1170.

Xu, Q., C.-J. Qui, and J.-X. Yu, 1994: Adjoint-method retrievals of low-altitude wind fileds from single-Doppler wind data. *J. Atmos. Oceanic Technol.,* **11**, 579-585.

Xue, M., K. Brewster, K. K. Droegemeier, F. Carr, V. Wong, Y. Liu, A. Sathye, G. Bassett, P. Janish, J. Levit, and P. Bothwell, 1996: Real time prediction of storm-scale weather during VORTEX- 95, Part II: Operation summary and example cases. Preprints, *18th Conf. on Severe Local Storms, San Francisco, CA*, Amer. Meteor. Soc., 178–182.

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

Xue M, K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D.-H. Wang, 2001: The Advanced Regional Prediction System (ARPS) – A multiscale non-hydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor Atmos Physics*, **76**, 143–165.

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

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, J. S. Kain, S. J. Weiss, D. R. Bright, M. C. Coniglio, and J. Du, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *Preprints, 24th Conference on Severe Local Storms, Savannah, GA.,* Amer. Meteor. Soc., 12.2

Yang, S-C., E. Kalnay, B. Hunt, and N. E. Bowler, 2009: Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Q. J. R. Meteo. Soc.*, **135**, 251-262.

Yussouf, N., and D. J. Stensrud, 2010: Impact of phased array radar observations over a short assimilation period: Observing system simulation experiments using an ensemble Kalman filter. *Mon. Wea. Rev.*, **138**, 517–538.

Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412.

Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Mon. Wea. Rev.*, **139**, 3489–3513.

Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the "surprise" snowstorm of 24–25 January 2000. *Mon. Wea. Rev.*, **130**, 1617–1632.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253.

Zhang, F., A. M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149–166.

Zhang, F., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594.

Ziegler, C. L., 1985. Retrieval of thermal and microphysical variables in observed convective storms. Part 1: Model development and preliminary testing. *J. Atmos. Sci.*, **42**, 1487–1509.

Ziegler, C. L., E. R. Mansell, J. M. Straka, D. R. MacGorman, and D. W. Burgess, 2010: The impact of spatial variations of low-level stability on the life cycle of a simulated supercell storm. *Mon. Wea. Rev.*, **138**, 1738–1766.