UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SENSITIVITIES OF 1-KM FORECASTS OF TORNADIC SUPERCELLS TO MICROPHYSICS PARAMETERIZATIONS, ASSIMILATED RADAR DATA, AND ASSIMILATION TECHNIQUES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

DEREK RAY STRATMAN
Norman, Oklahoma
2016

SENSITIVITIES OF 1-KM FORECASTS OF TORNADIC SUPERCELLS TO MICROPHYSICS PARAMETERIZATIONS, ASSIMILATED RADAR DATA, AND ASSIMILATION TECHNIQUES

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

_____
Dr. Ming Xue, Chair

_____
Dr. Keith A. Brewster

_____
Dr. Frederick H. Carr

_____
Dr. Michael B. Richman

_____
Dr. Christopher E. Weaver

Acknowledgements

After five years of dissertation work and eleven total years of undergraduate and graduate school, I have numerous people to thank for helping me get to this point in my career and life. First, I'd like to thank my doctoral committee members, Drs. Ming Xue, Keith Brewster, Fred Carr, Michael Richman, and Chris Weaver, for their constructive criticisms and suggestions, which led to the improvement of my dissertation. I'd especially like to thank my doctoral advisor, Dr. Keith Brewster, for his help and guidance over the past four years as I settled on and completed my dissertation research. I'd like to also thank everyone else who provided feedback on my research during seminars and scientific meetings. This work is supported by the National Science Foundation via CASA ERC grant EEC 03-13747 and related support provided by the University of Oklahoma (OU). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Most of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER). This work utilized data from the Oklahoma Mesonet provided by the Oklahoma Climatological Survey.

This work would not have been completed without the support of my family and friends. I'd like to thank both my grade school and college friends for keeping me sane through the entire graduate school process. I'd like to thank my parents, Gerald and Doris, and my sister, Jenna, for fostering my love of weather since I was quite young and providing unconditional loving support. Finally, I'd like to thank my caring, lovely wife, Stephanie, for the endless love and support, even during my stressful times.

Table of Contents

List of Tables

List of Figures

Abstract

The aim of this study is to examine the impact of using five different microphysics parameterization schemes, including single-, double-, and triple-moment microphysics, in an efficient high-resolution data assimilation system suitable for nowcasting and short-term forecasting with low latencies. In addition to testing the sensitivity to microphysics, the impact of gap-filling radars and variations in analysis cycling and incremental analysis updating (IAU) techniques are explored using a variety of verification methods.

On 24 May 2011, Oklahoma experienced an outbreak of tornadoes, including one rated EF-5 and two rated EF-4. The extensive observation network in this area, including the WSR-88D radars, Collaborative Adaptive Sensing of the Atmosphere (CASA) IP-1 X-band radars, Oklahoma Mesonet, and standard surface data, makes this an ideal case for these tests. Additionally, the real-time configuration of the 1-km ARPS, which used 3DVAR with cloud analysis via IAU, had success providing a good baseline forecast. ARPS forecasts of 0-2h are verified using point-to-point, neighborhood, and object-based verification techniques. The object-based verification technique uses updraft helicity fields to represent mesocyclone centers, which are verified against tornado locations from three supercells of interest. Varying levels of success in the forecasts are found and appear to be dependent on the complexity of storm interaction, with early forecasts of isolated storms exhibiting the most success. Verification scores indicate the multi-moment schemes tend to produce better forecasts, assimilating CASA radar data can improve forecasts for storms within the CASA radar network, and analysis cycling and modified IAU techniques generally contribute to better forecasts.

Chapter 1: General Introduction

1.1 Background and Motivation

On the afternoon of 24 May 2011, an outbreak of twelve tornadoes, including two EF-4 tornadoes and one EF-5 tornado, afflicted northern and central Oklahoma within the National Weather Service (NWS) Norman, OK Weather Forecast Office's (WFO) county warning area (CWA). Unfortunately, this outbreak caused numerous deaths and injuries along with considerable damage. An extensive observation network was in place in this area during the spring of 2011, so despite the tragic loss of life, this is an ideal case to explore aspects of the evolving decadal goal of providing real-time, thunderstorm-resolving forecasts using high-resolution (i.e., < 4-km grid spacing) numerical weather prediction (NWP).

The tight clustering of the tornadic and non-tornadic supercells and storms on this date made forecasting of supercell tracks difficult for storm-scale models, but the Center for Analysis and Prediction of Storms (CAPS) real-time forecasting system had good success at simulating these storms. An accumulation of 1–6-km updraft helicity (UH; Kain et al., 2008) is used to identify simulated mesocyclones every minute in eight consecutive real-time 130-minute forecasts, which were initiated every 20 minutes from 2050 UTC to 2250 UTC, predicted the potential severity of this event (Figure 1). However, advances in forecasting supercell evolution and motions (i.e., location and timing) could be gained with improved data assimilation techniques and better parameterizations of physical processes.

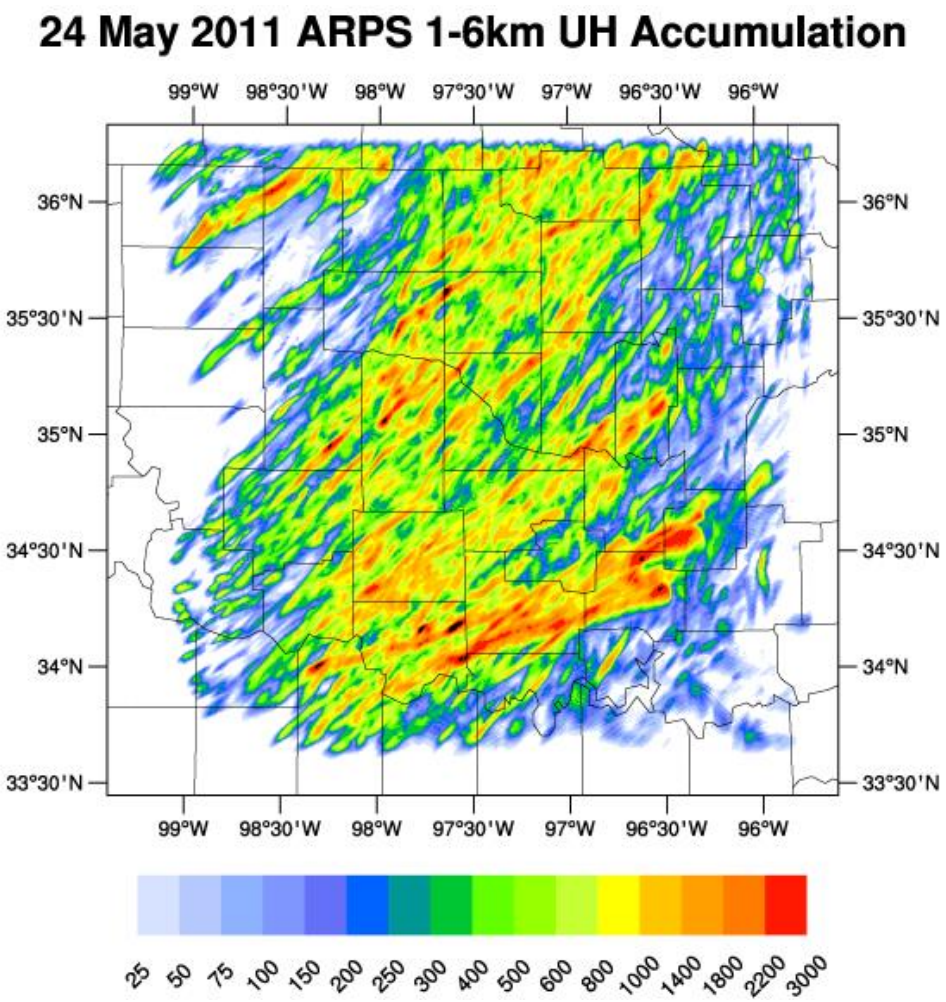

Figure 1. Real-time ARPS forecast system's Control member's 130-minute forecasts of 1–6-km UH from eight consecutive initiation times, including 2050, 2110, 2130, 2150, 2210, 2230, and 2250 UTC, on 24 May 2011.

1.2 Current State of the Science of Storm-Scale NWP

Recent storm-scale NWP studies have largely concentrated on improving forecasts of severe weather hazards (i.e., tornadoes, damaging wind, large hail, flooding, and lightning) by testing the efficacy of various numerical analysis and data assimilation techniques and parameterizations of sub-grid-scale processes and by examining new scale-appropriate verification and evaluation metrics. The motivation of many of these studies is centered on a long-term goal of real-time thunderstorm resolving NWP, as articulated in the Warn-on-Forecast (WoF) concept (Stensrud et al., 2009, 2013). As opposed to the current system of issuing warnings based on the observation or detection of severe weather, the WoF paradigm is founded on the idea that one day NWS forecasters will have the ability to issue severe thunderstorm and tornado warnings based on short-term, storm-scale forecasts with the aim of increasing warning lead times.

Almost all storm-scale NWP studies have been completed using either a three-dimensional variational data assimilation (3DVAR; Gao et al., 2004) system (e.g., Dawson et al., 2015, Gao et al., 2013; Schenkman et al., 2011a,b; Stensrud and Gao, 2010) or a variant of the ensemble Kalman filter (EnKF; Evensen, 1994) data assimilation system (e.g., Dawson et al., 2012; Dowell et al., 2004, 2011; Jones et al., 2015; Putnam et al., 2014; Snook et al., 2011, 2012; Wheatley et al., 2014, Xue et al., 2006; Yussouf et al., 2013, 2015) to produce an initial analysis. Other data assimilation techniques [e.g., four-dimensional variational data assimilation (4DVAR) and hybrid data assimilation systems] are also employed in NWP systems to produce initial analyses, but 3DVAR and EnKF (and related variants, e.g., ensemble square-root

Kalman filter, ensemble adjustment Kalman filter, and local ensemble transform Kalman filter) have recently become the most popular data assimilation techniques for the generation of storm-scale analyses, which are then used by storm-scale forecasting systems or for storm simulations.

In general, 3DVAR seeks to minimize a cost function to find the optimal analysis fields using all available observation data. More specifically, the 3DVAR process begins by calculating the cost function,

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}[\mathbf{y}_o - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}_o - H(\mathbf{x})] + J_C \qquad (1)$$

which is the distance between the analysis, $\mathbf{x}$, and background fields, $\mathbf{x}_b$, weighted by the inverse of the background error covariance, $\mathbf{B}$, plus the distance between the observations, $\mathbf{y}_o$, and the analysis fields interpolated to observation locations, $H(\mathbf{x})$, weighted by the inverse of the observational error covariance, $\mathbf{R}$, plus an optional penalty term, which can apply dynamic constraints to enforce balance flow conditions (Gao et al., 2004). The background field (i.e., usually a prior forecast) is used as the initial first guess. Next, the gradient of the cost function, $\nabla J(\mathbf{x})$, is computed with respect to the analysis variable or analysis increment. The gradient of the cost function is then passed through a minimization algorithm (e.g., conjugate gradient or quasi-Newton methods) to determine new analysis fields. The norm of the cost function gradient or the satisfaction of a sufficiently-minimized cost function determines if the process continues iterating until an optimal analysis is produced. Forecasts can start directly from the analyses produced by the 3DVAR system or gradually "spin up" by introducing analysis increments during a short time window (e.g., 5 or 10 min for storm-scale fields) using incremental analysis updating (IAU; Bloom et al., 1996).

As the name suggests, the EnKF data assimilation technique benefits from using ensembles to simultaneously perform multiple data assimilation cycles. First, an ensemble of analyses is produced by adding random perturbations to the observations. The ensemble of analyses is integrated forward in time until the next data assimilation time. New analysis fields are determined for each ensemble member using the equation,

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}_i\{\mathbf{y}_i^o - H[\mathbf{x}^f(t_i)]\}, \quad (2)$$

where the forecast fields, $\mathbf{x}^f(t_i)$, are summed with the observational innovations, $\mathbf{y}_i^o - H[\mathbf{x}^f(t_i)]$, weighted by the Kalman gain matrix, $\mathbf{K}_i$, which is derived using the updated forecast error covariance, $\mathbf{P}^f(t_i)$, and observations error covariance, $\mathbf{R}_i$, using the equation,

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^T[\mathbf{R}_i + \mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}^T]^{-1}. \quad (3)$$

Most importantly, the forecast error covariances are estimated from the ensemble of forecasts (Evensen 1994). This allows for the adjustment of unobserved variables in the analysis assuming the ensemble can properly represent the error covariances. The EnKF process repeats itself until a desired analysis is generated for use as a reliable analysis or in a forecast system.

Both data assimilation techniques have advantages and disadvantages with respect to producing analyses. For example, EnKF has the advantage of using an ensemble of forecasts to estimate flow-dependent forecast error covariances, and 3DVAR uses a time-invariant background (or forecast) error covariance matrix. 3DVAR has the advantage of being computationally more efficient [O(10-100)] than EnKF, offering significantly lower cost to implement as well as lead-time advantages due to reduced latency. However, the majority of storm-scale NWP retrospective studies

use EnKF likely due to EnKF affording itself to ensemble forecasting and its documented success over 3DVAR (e.g., Stensrud et al., 2013; Johnson et al., 2015). Even so, both data assimilation techniques can produce suitable analyses for storm-scale forecasting applications (e.g., WoF; Stensrud et al., 2013).

Numerous studies have investigated the sensitivities of cloud and precipitation microphysics parameterization schemes (e.g., Dawson et al., 2010, 2015; Putnam et al., 2014; Wainwright et al., 2014; Wheatley et al., 2014; Yussouf et al., 2013), model grid spacing (e.g., Potvin and Flora 2015; Dowell et al., 2015; Schwartz et al., 2009; Kain et al., 2008; Johnson et al., 2013), and assimilating various sources of data (e.g., Snook et al., 2011; Jones et al., 2015; Xue et al., 2006; Dawson et al., 2012; Sobash and Stensrud, 2015) on storm-scale forecasts. All microphysics-related studies comparing different microphysics schemes come to the general conclusion that multi-moment microphysics schemes produce better forecasts of idealized and real supercells (Dawson et al., 2010, 2015; Yussouf et al., 2013) and mesoscale convective systems (MCSs; Wheatley et al., 2014) than single-moment microphysics schemes.

Kain et al. (2008) and Schwartz et al. (2009) concluded that using a smaller grid spacing (e.g., 4 km vs 2 km) doesn't appreciably improve forecast value and skill at smaller scales for 18–36-h forecasts. However, Johnson et al. (2013) found that 1-km forecasts of precipitation features added some skill on the scales not resolvable by 4-km forecasts (i.e., objects > ~16 km in size) but resolvable by 1-km forecasts (i.e., objects > ~4 km in size but < ~16 km in size). Additionally, Potvin and Flora (2015) and Dowell et al. (2015) both determine using simulations of idealized and real cases, respectively, that 1-km forecasts exhibit better representations of supercell features (e.g., low-level

rotation tracks and intensity) than 2-, 3-, or 4-km forecasts, but the differences among 1-, 2-, and 3-km forecasts of supercells are smaller than between 3- and 4-km forecasts.

A plethora of studies have investigated the impact of assimilating data from in-situ (e.g., Schenkman et al., 2011; Sobash and Stensrud, 2015) and remote sensing (e.g., Jones 2015; Dawson et al., 2012; Dowell et al., 2011; Schenkman et al., 2011; Xue et al., 2006; Stratman et al., 2013; Potvin and Wicker, 2013) observations. Schenkman et al. (2011) concluded that assimilating Oklahoma Mesonet observations (Brock et al., 1995) substantially improves storm-scale analyses and forecasts when low-level Collaborative Adaptive Sensing of the Atmosphere (CASA) IP-1 X-band (McLaughlin et al., 2009) radar radial velocity data are withheld from the assimilation process, but there is still at least some benefit to assimilating Mesonet data even when CASA radar data are also assimilated. Sobash and Stensrud (2015) found that convective initiation (CI) times and locations were improved for their case study by utilizing the spatially and temporally higher-resolution Mesonet data. From Dowell et al. (2011), simulated storms develop more quickly when both Weather Surveillance Radar-1988 Doppler (WSR-88D) S-band radar reflectivity and radial velocity data are employed in the assimilation process. Stratman et al. (2013) demonstrated that assimilating radar data yields some benefit at mesoscales, but not at convective scales. The assimilation of both cloud water path retrievals from satellites and WSR-88D radar reflectivity and radial velocity data helps analyses and forecasts spin storms up faster, suppress spurious convection, and better predict CI (Jones et al., 2015). Dawson et al. (2012) showed the substantial impact of using varying WSR-88D velocity-azimuth display (VAD) wind profiles on forecasting 4 May 2007 Greensburg, Kansas tornadic mesocyclones. Also,

both the location of radars relative to supercells (e.g., Potvin and Wicker, 2013) and the number and type of radars in a network used in the assimilation process (Xue et al., 2006) can considerably alter forecasts of storms.

The accumulation of findings from these recent storm-scale NWP studies has yielded several positive steps toward realizing the substantial benefits to short-term, storm-scale forecasting, including operational implementation of real-time forecasting systems, but more research is needed to further improve current forecast systems. This study will address some questions related to using various microphysics schemes, assimilating CASA radar data, and exploring assimilation techniques with the aim of benefiting convection-allowing, storm-scale models.

1.3 Hypotheses and Contributions

This dissertation presents a series of experiments addressing specific topics related to storm-scale NWP using data collected during the 24 May 2011 tornado outbreak in Oklahoma. Although this work focuses on results from a single day, there are multiple storms during the day representing an isolated supercell and then other supercells with increasing levels of storm interaction. The design of each experiment is planned with a potential future real-time storm-resolving forecasting system in mind. The hypotheses to be tested include:

- Multi-moment microphysics schemes yield smaller forecast errors than single-moment microphysics schemes (confirming prior studies).
- Assimilating CASA radar data produces better forecasts of convection within the CASA radar network than not assimilating CASA radar data.
- When radial velocity and reflectivity data from the Twin Lakes (KTLX) WSR-88D radar are withheld from data assimilation, the CASA radar network, along with the other surrounding WSR-88D radars, are able to prevent the loss of forecast skill.
- IAU with analysis cycling, as opposed to IAU with no analysis cycling, leads to better forecasts of low-level circulations.
- Introducing temperature, wind, and water vapor field increments before cloud and precipitation microphysics field increments during IAU windows better maintains initial simulated storms than inserting increments using the same temporal distribution for all variables.

- Using a new verification technique (as described in Section 3.3) for low-level rotation centers in conjunction with other objective verification metrics can quantitatively indicate forecast skill to help elucidate sources of error or possible improvement.

By testing these hypotheses, a number of new contributions can be made to the greater understanding of storm-scale NWP. A new objective verification technique is utilized to determine location and timing errors for simulated circulation centers. All of the projects employ for the first time the recently updated Advanced Regional Prediction System's (ARPS; Xue et al., 2000, 2001, 2003) data assimilation system (ADAS) complex cloud analysis package (Hu et al., 2006a,b; Brewster and Stratman, 2015). This is the first known study to examine the impact of assimilating CASA radar data on simulated supercells using real data and full model physics, implementing IAU and analysis cycling, and introducing variable-dependent temporal distributions of increments (Brewster et al., 2015; Brewster and Stratman, 2016) while utilizing a microphysically-diverse set of simulations.

1.4 Outline of Dissertation

In Chapter 2, an overview of the 24 May 2011 Oklahoma tornado outbreak is presented with an introduction of the event, a synoptic and mesoscale overview, a tornado overview, and a section on event-related studies. Verification metrics and methods are described in Chapter 3 with an introduction to verification, a detailing of the types of verification metrics, and the verification methodology for this study. The sensitivities of various microphysics schemes in storm-scale forecasts are presented in Chapter 4. In Chapter 5, the impacts of assimilating CASA radar data on storm-scale forecasts are demonstrated. The impacts of analysis cycling, variable-dependent IAU, and the combination of both data assimilation techniques on storm-scale forecasts are presented in Chapter 6. All three results chapters include subsections on introduction and background, experiment design, and verification results for near-surface variables, simulated composite reflectivity, and low-level circulations from three supercells of interest. Finally, Chapter 7 features a general summary of the dissertation, summaries of the results, and a discussion of the findings and potential future work.

Chapter 2: Overview of 24 May 2011 Oklahoma Tornado Outbreak

2.1 Introduction of Event

On 24 May 2011, a well-forecasted severe weather outbreak tormented parts of Kansas, Missouri, Oklahoma, Arkansas, and Texas with all facets of severe weather (Figure 2). Within the NWS Norman, OK WFO's CWA, the outbreak was responsible for a dozen tornadoes comprised of one EF-5 (B2), two EF-4 (C1, D1), two EF-3 (A1, B1), two EF-2 (B4, D3), two EF-1 (D2, E1), and three EF-0 (A2, B3, C2) tornadoes (Figure 3) and, unfortunately, 11 deaths and 342 injuries [National Climatic Data Center's (NCDC) Storm Events Database (https://www.ncdc.noaa.gov/stormevents/)]. This portion of the tornado outbreak spanned from near Fairview, OK to near Tishomingo, OK (i.e., nearly 280 km) with 14 out of 56 counties in Norman, OK WFO's CWA experiencing at least one tornado and lasted about 4 hours and 25 minutes with the first tornado forming around 2020 UTC and the last tornado dissipating around 0045 UTC. Damaging wind and large hail up to 3 inches in diameter also wreaked havoc on many areas (Figure 2).

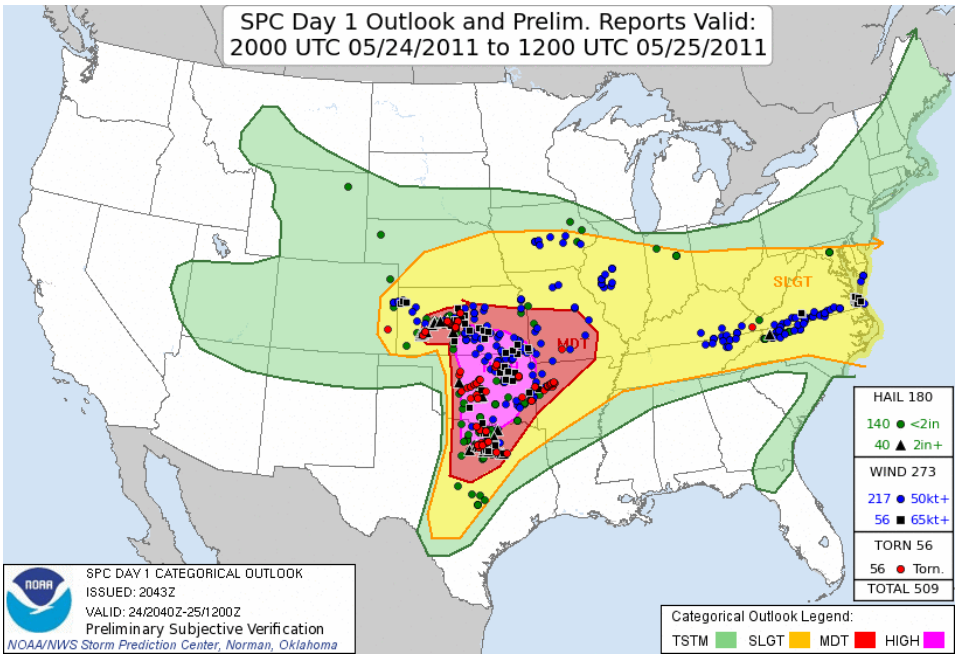

Figure 2. NWS SPC's 2000 UTC convective outlook for 24 May 2011 with preliminary storm reports from this severe weather event overlaid. Figure taken from http://www.spc.noaa.gov/exper/archive/events/.

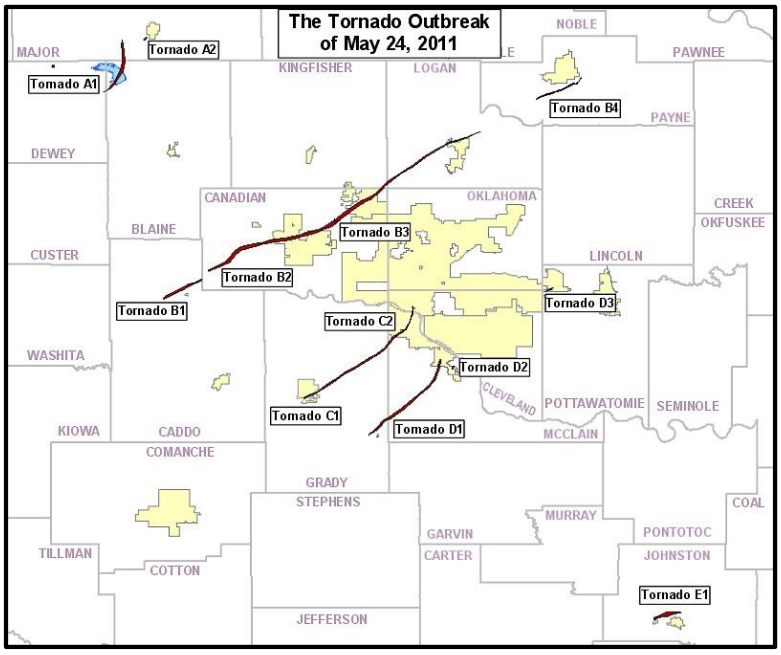

Figure 3. Map of all twelve 24 May 2011 tornado tracks for NWS Norman, OK WFO's CWA. Map taken from http://www.srh.noaa.gov/oun/?n=events-20110524.

2.2 Synoptic and Mesoscale Overview

Severe weather episodes tend to transpire when atmospheric ingredients (i.e., shear, lift, instability, and moisture) come together in space and time, and on the morning of 24 May 2011, little doubt existed among forecasters that a significant severe weather event loomed in the coming hours due to forecasts and observations depicting several key ingredients needed for the development and maintenance of severe storms. The 1200-UTC sounding out of Norman, OK (OUN) indicated nearly dry-adiabatic, mid-level lapse rates contributing to ~1800 J/kg of convective available potential energy (CAPE) with veering winds with height (Figure 4), but a capping inversion just below 800 hPa prevented the development of convection prior to the early afternoon hours. By 1800 UTC, CAPE increased to ~2900 J/kg due to warming and moistening at the low levels and slight cooling around 800 hPa (Figure 4). However, the strong capping inversion remained for most of the warm sector. Soundings with temperature and dewpoint temperature profiles like OUN's 1800-UTC sounding are commonly described as "loaded-gun" soundings because, if moist parcels are able to be lifted to the level of free convection, which exists above the capping inversion, then ample CAPE is available for intense development of convection.

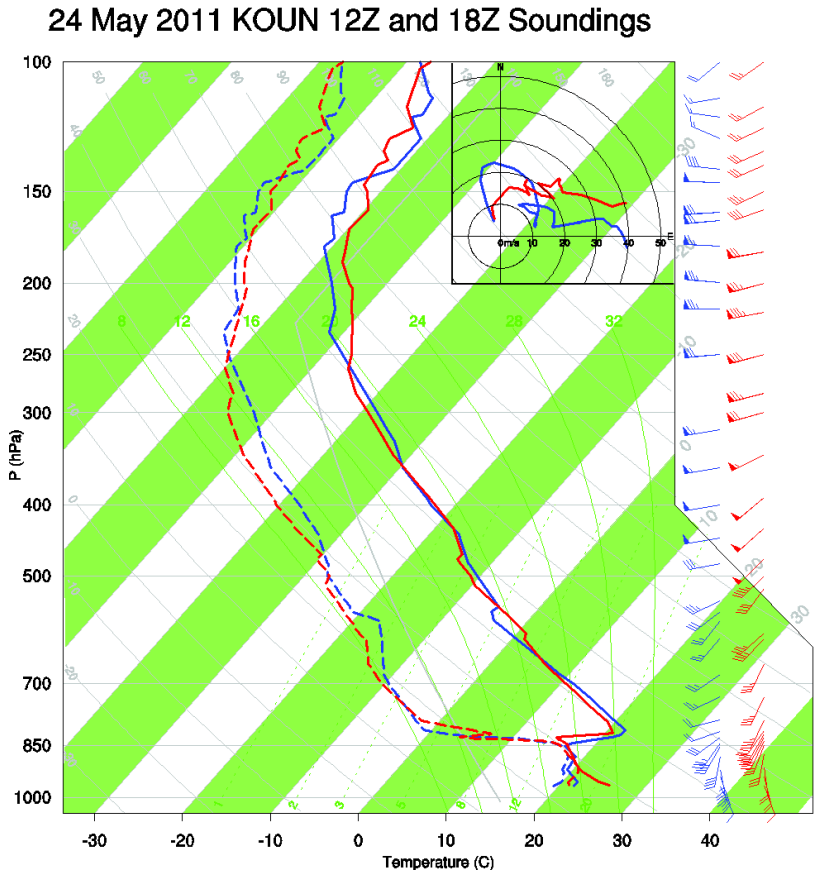

Figure 4. 1200 UTC (blue) and 1800 UTC (red) vertical sounding profiles of temperature (solid line), dewpoint temperature (dashed line), and wind vectors on a skew-t plot. Hodograph in top-right corner of skew-t showing reported wind vectors up to 12-km AGL. Sounding data provided by University of Wyoming's sounding archive webpage (http://weather.uwyo.edu/upperair/sounding.html).

Between 1800 UTC and 2100 UTC, the approach of a negatively-tilted trough increased upper-level divergence, which contributed to large-scale ascent, over Oklahoma (Figure 5a,b). At 500 hPa, wind speeds increased while temperatures slightly decreased, which increased the mid-level lapse rates and, thus, instability (Figure 5c,d). At 850 hPa, the wind directions backed from southwesterly to southerly, aiding in better directional shear for rotating storms, while moisture increased across the entire warm sector (Figure 5e,f). From 1800 UTC to 2000 UTC, the dryline sharpened, which increased low-level convergence and lift, and became more perpendicular to the deep-

layer mean wind vector, which supports isolated storm modes (Bluestein and Weisman 2000; Dial and Racy 2004; Figure 6). Also, by 1900 UTC, convection began to develop along the dryline in southwest Oklahoma, where the highest surface temperatures coexisted with ample moisture (Figure 6b).

Overall, this severe weather outbreak was a classic high-shear, high-instability event with > 25 m/s of 0–6-km bulk shear and > 3000 J/kg of CAPE, which collectively support supercells, and significant tornadoes were possible given > 200 m^2/s^2 of 0–1-km storm-relative helicity (Thompson et al., 2003; Markowski et al., 2003). Upper-level divergence and the dryline provided large-scale ascent and focused lift, respectively, while a stout ~800-hPa warm nose mostly prevented widespread convection across the warm sector, especially during the first couple of hours after CI. Ample low-level moisture aided in low lifting-condensation levels (i.e., < 1000 m), which benefit the generation of tornadoes (Rasmussen and Blanchard, 1998). With these severe weather ingredients forecasted and observed, NWS and SPC forecasters were confident that the potential for an outbreak of tornadic supercells existed across the Southern Plains region, especially Oklahoma, as reflected in the convective outlooks leading up to the event (e.g., the 0600 UTC Day 1 Convective Outlook contained a High Risk for a majority of the eastern two-thirds of Oklahoma; see Figure 2 and http://www.spc.noaa.gov/exper/archive/event.php?date=20110524). With such great potential over a wide area it is important to try to more precisely identify areas at risk of damage and/or life threatening conditions.

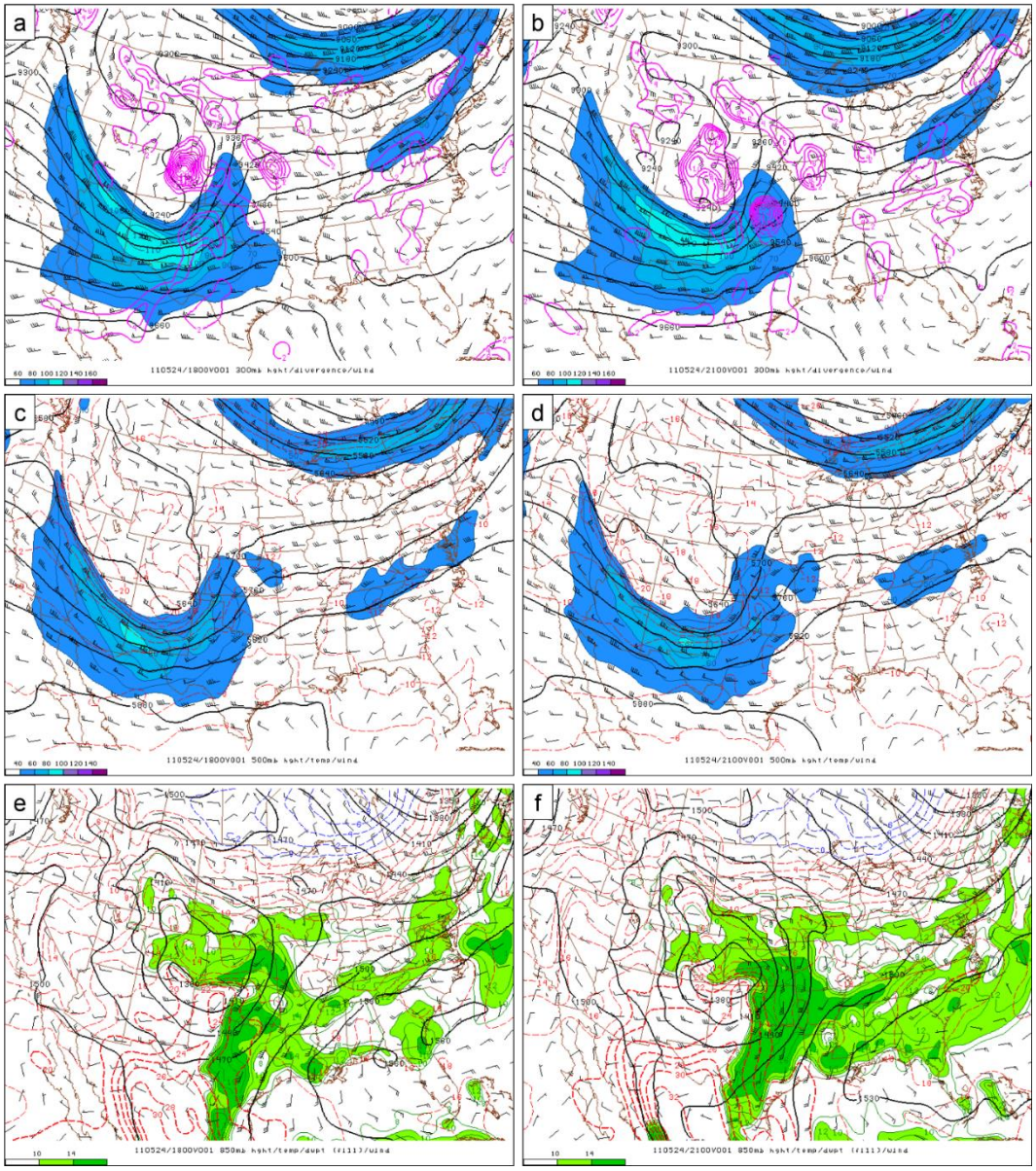

Figure 5. 1800 UTC (a,c,e) and 2100 UTC (b,d,f) analysis plots of (a,b) 300-hPa heights (m; black contours), divergence (s^{-1}; magenta contours), wind barbs (kt), and wind speeds (kt; blue fill), (c,d) 500-hPa heights (m), temperatures (°C; red dashed), wind barbs (kt), and wind speeds (kt), and (e,f) 850-hPa heights (m), temperatures (°C), wind barbs (kt), and dewpoint temperature (°C; green fill) from SPC's mesoscale analysis archive.

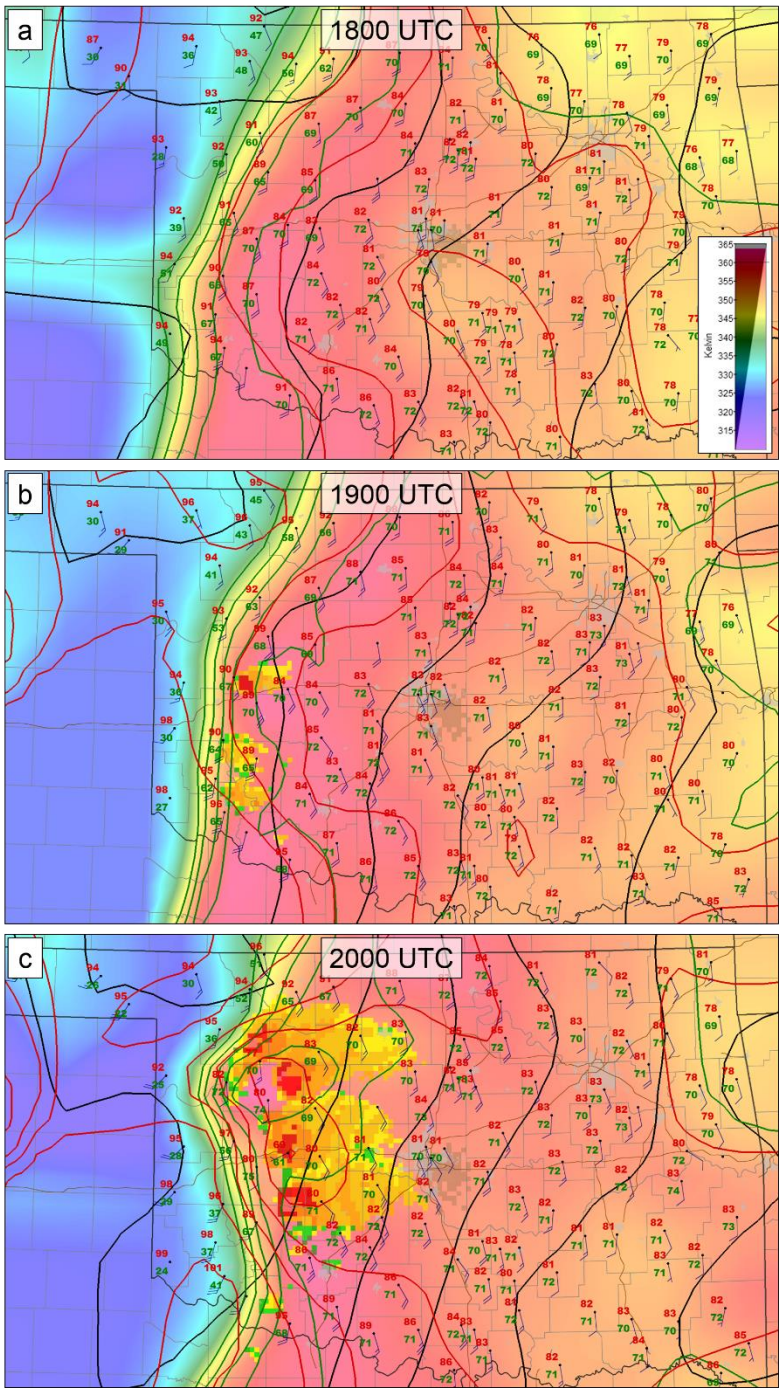

Figure 6. (a) 1800 UTC, (b) 1900 UTC, and (c) 2000 UTC Oklahoma Mesonet station plots of 1.5-m AGL temperature (°F; red text), 1.5-m AGL dewpoint temperature (°F; green text), and 10-m AGL wind barbs (kt; blue); analyses of sea-level pressure (hPa; black contours every 2 hPa), 1.5-m AGL temperature (°F; red contours every 5°F starting at 65°F), 1.5-m dewpoint temperature (°F; green contours every 5°F starting at 50°F), and equivalent-potential temperature (K; background shading); and derived echo tops (kft; colored blocks) using the KTLX radar. Plots created using WeatherScope 1.9.6 created by UCAR/Unidata and OCS.

2.3 Tornado Overview

Tornadic and non-tornadic supercells developed along the dryline and trekked northeast across Oklahoma, which lead to the issuance of 34 tornado warnings by the Norman, OK WFO (Figure 7). Only five storms were responsible for the twelve tornadoes documented within NWS Norman, OK's CWA. Tornado path lengths (max widths) ranged from ~0.8 km (~40 m) to ~101.4 km (~1610 m) with an average path length (max width) of ~21.6 km (~575 m) and the sum of all path lengths of ~259.7 km. The twelve tornadoes lasted from less than 1 minute to around 105 minutes with an average tornado life span of ~23 minutes. The tornadoes rated EF-3 or higher all had average translational velocities around 16 m/s.

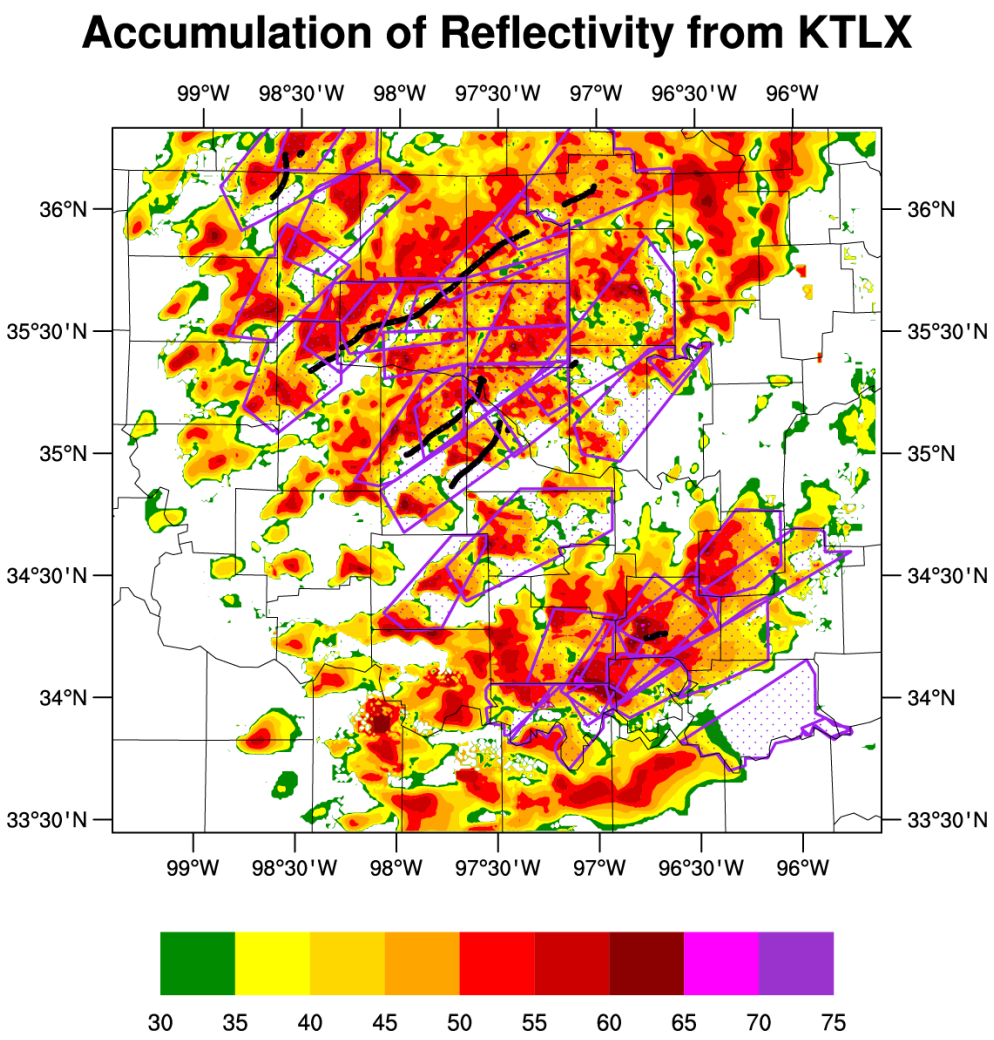

Figure 7. Max 0.5°-tilt reflectivity (dBZ) above 30 dBZ from KTLX for output times 1902, 1932, 2002, 2032, 2102, 2132, 2201, 2231, 2301, and 2331 UTC on 24 May 2011 and 0000, 0030, and 0101 UTC on 25 May 2011. Black dots represent estimated tornado locations every minute. Tornado warnings issued by the NWS Norman, OK WFO during the 24 May 2011 tornado outbreak are represented by magenta polygon outlines and stipples.

2.4 Event-Related Studies

Various sources of observational data and model output from the 24 May 2011 severe weather outbreak have been utilized in several observationally- (Brotzge and Luttrell, 2015; French et al., 2015; Heymsfield et al., 2013; Houser et al., 2015; Xu et al., 2015) and NWP-based (Clark et al., 2013; Fierro et al., 2012; Jones et al., 2015; Shapiro et al., 2015; Tanamachi et al., 2015) studies. All of the observational-based studies used different sources of radar data to analyze different facets of this outbreak. Heymsfield et al. (2013) analyzed measurements collected from flying a dual-frequency (Ku and Ka band) nadir-pointing Doppler radar through storms containing hail and graupel. Brotzge and Luttrell (2015) documented the genesis of one of the EF-4 tornadoes using Oklahoma Mesonet and CASA radar data to show the potential benefits of using spatially and temporally high-resolution observations in an operational forecasting situation. Houser et al. (2015) used data collected by a mobile, rapid-scan, X-band, polarimetric, Doppler radar (RaXPol) to understand the dynamical processes leading to the genesis of the EF-5 tornado a few minutes after the dissipation of an EF-3 tornado. French et al. (2015) utilized three stationary S-band radars [i.e., two WSR-88D radars and one multifunction phased-array radar (MPAR)] and one mobile phased-array radar to explore the merger process between the EF-5 tornado and a short-lived, satellite tornado. Xu et al. (2015) assessed the performance of a new method to analyze radar-based vortex winds using examples from the 24 May 2011 outbreak.

Fierro et al. (2012) showed the potential benefits of assimilating lightning data into initial analyses, which may lead to better convective forecasts, by using the 24 May 2011 case. Shapiro et al. (2015) simulated a supercell storm using an analysis sounding

from the 24 May 2011 environment to assess the impact of advection correction in trajectory calculations. As in Houser et al. (2015), Tanamachi et al. (2015) also investigated the Lookeba-El-Reno-Piedmont tornadic supercell's dissipation of the EF-3 tornado to the genesis of the EF-5, but MPAR data and a numerical model were used to explore the role of a storm merger on the tornado handoff. Finally, Clark et al. (2013) assessed the tornado path-length forecasts (via the UH field) from the Storm Scale Ensemble Forecast (SSEF) system for several cases in 2011, including 24 May.

Chapter 3: Verification Metrics and Methodology

3.1 Introduction

Subjective evaluation and objective verification have played an integral part in validating improvements and diagnosing potential problematic areas of forecast systems. This information can then be utilized in the operational forecasting process by forecasters to improve forecasts or in NWP studies by model researchers and developers to improve forecast systems. The annual Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE) in Norman, OK is one example where both forecasters and researchers provide subjective evaluation (e.g., comparing forecasted reflectivity with observed reflectivity in a side-by-side visual comparison) of forecasts and forecast tools (Clark et al., 2012). Objective verification (e.g., bias for forecasted rainfall) should aim to complement subjective evaluation by confirming and quantifying the researchers and forecasters' visual evaluations of forecasts and by exposing hidden problems researchers and forecasters may not have noticed through visual evaluation. Conversely, subjective evaluation can be employed to help explain and understand the results found from using objective verification. While subjective evaluation usually consists of side-by-side or overlay comparisons, a growing spectrum of objective verification techniques is required to isolate areas where forecasts are less skillful for both spatially-continuous variables (e.g., surface dewpoint temperature) and spatially-discontinuous fields (e.g. thunderstorm reflectivity), especially in the case of rare events (e.g., supercells).

3.2 Types of Verification Metrics

NWP forecasts are verified using a range of methods, and the best forecast verification method to use is dependent on what specific questions are being asked about a forecast's performance. Typically, gridded forecasts are verified with predictands (e.g., observations) using traditional and spatial techniques. Traditional verification techniques consist of point-by-point metrics, which are based on verifying gridded model output with either non-gridded points of observational data or gridded observational data. Verification statistics, such as mean absolute error (MAE), can be used to assess model forecast performance for various variables (e.g., temperature) by interpolating gridded model output to observation points (e.g., Mesonet locations). If observational data are interpolated to model grid points, contingency tables consisting of hits, misses, false alarms, and correct negatives can be used to compute additional verification statistics, such as false alarm ratio (FAR) and probability of detection (POD). While these traditional verification techniques can provide assessment of forecast skill for non-rare events, they tend to suffer in their assessment of forecast skill for rare events (Ahijevych et al., 2009). For example, if a simulated storm isn't in the exact location of an observed storm, traditional verification metrics would indicate poor forecast skill by double counting (i.e., miss and false alarm) the displacement error even though a forecaster or researcher might deem the forecast to have some value in predicting the existence and various characteristics of the storm (e.g., intensity, structure, and mode).

One way to assess the forecasts of rare events is to utilize spatial verification techniques. In general, these techniques can be classified as neighborhood, scale-

separation, object-based, or field-deformation methods (Figure 8; Gilleland et al., 2009, 2010b). The neighborhood (or fuzzy) and scale-separation (or scale-decomposition) methods are forms of spatial smoothing and bandpass filters, respectively, and the object-based (or featured-based) and field deformation methods can be categorized as spatial displacement methods (Gilleland et al., 2009, 2010b). Ebert (2008) and Ebert (2009) outline a plethora of neighborhood-based methods, which all act to reward forecasts that are "close enough" to the observations. One popular neighborhood-based method is the fractions skill score (FSS; Roberts and Lean 2008; Roberts 2008; Mittermaier and Roberts 2009), which determines a forecast to have useful skill when the number of forecasted events is similar to the number of observed events within a neighborhood window. Unlike neighborhood-based verification, scale-separation methods provide assessment of forecast performance on individual scales (Gilleland et al., 2009, 2010b). Casati et al. (2004) introduced the intensity-scale (IS) verification technique, which utilizes a 2-D discrete Haar wavelet decomposition to assess forecast skill and isolate errors at varying spatial scales and variable intensity (e.g., simulated reflectivity values), and the associated IS skill score (ISS). To assess bias at varying scales and intensities, Casati (2009) adjusted the IS method of Casati et al. (2004) by not recalibrating the forecasts to be unbiased.

While the filtering methods can provide information about scale and intensity errors, the displacement methods can directly evaluate location and structure errors (Gilleland et al., 2009). Object-based verification methods attempt to define features within a spatially-discontinuous model field (e.g., simulated reflectivity) usually using thresholds with the goal of determining attributes, such as shape, size, centroid location,

and mean intensity (Gilleland et al., 2009, 2010b). From these attributes, various errors, such as displacement, orientation, and intensity errors, can be computed to assess model performance. An example of an object-based verification method is the method for object-based diagnostic evaluation (MODE) developed by Davis et al. (2006). MODE uses convolution to smooth forecast and observation fields prior to using thresholds to define objects. Additional studies have utilized feature-based techniques to ascertain convective attributes (e.g., size, intensity, frequency, and timing) for climatologies of observed convective systems (Hitchens et al., 2012) and from model output to better assist forecasters in evaluating real-time simulated storms (Carley et al., 2011). As opposed to evaluating individual features, field-deformation verification methods essentially morph the forecast field to somewhat match the observation field and then produce a field of distortion vectors (Gilleland et al., 2009, 2010a,b).

Even though several verification techniques have been mentioned and briefly described, numerous other verification methods are continually being developed to answer specific model performance questions. While the different types of verification techniques can be used exclusively to evaluate model performance, verification techniques can be merged together to garner additional information about model performance. For example, a field-deformation method can be used for individual features defined by an object-based method (Gilleland et al., 2009). By using these objective verification metrics with subjective evaluation, valuable information is exposed and made available for forecasters and researchers to use in forecasting and model developing, respectively.

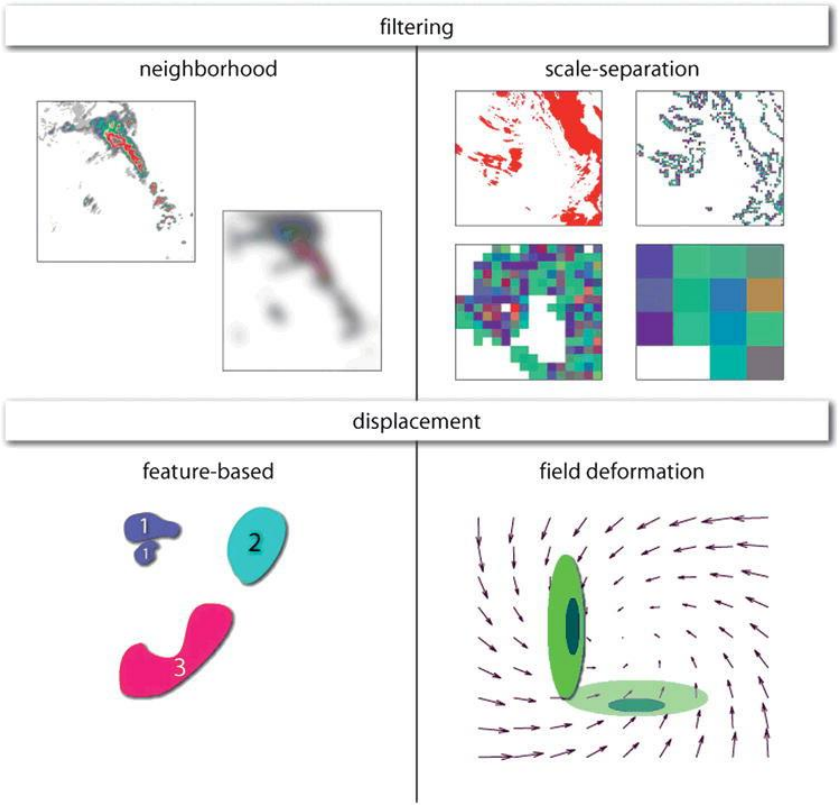

Figure 8. Schematic of spatial verification categories. (Top row) Neighborhood (or fuzzy) and scale-separation (or scale-decomposition) methods are classified as filtering verification methods, and (bottom row) features-based (or object-based) and field-deformation methods are classified as displacement verification methods. Schematic is from Gilleland et al. (2009).

3.3 Verification Methodology

In this study, a few different verification techniques are employed to properly test the hypotheses outlined in Section 1.3. First, root-mean-squared errors (RMSE) for five near-surface variables are computed using the equation,

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(f_k - o_k)^2}\,, \tag{4}$$

where n is the total number of observation locations, f_k and o_k are the forecasted and observed variables, respectively, at the k^{th} observation location. For consistency of observing instruments, the RMSEs are calculated using only Oklahoma Mesonet data, which uses 5-minute averages for 1.5-m temperature (T), 1.5-m dewpoint temperature (T_d), 10-m u- and v-component wind (u and v, respectively), and 0.75-m atmospheric pressure (p) (Brock et al. 1995). The forecasted values, f_k, for T, T_d, u, v, and p are first linearly interpolated in the vertical to 2-m, 2-m, 10-m, 10-m, and 2-m AGL, respectively, and then bi-linearly interpolated to the n mesonet locations. In addition to RMSE, linear regression slopes, coefficients of determination, and probability distribution functions (PDFs) are calculated for further comparisons. The linear regression slopes are computed using the equation,

$$m = \frac{\sum_{k=1}^{n}[(o_k-\bar{o})(f_k-\bar{f})]}{\sum_{k=1}^{n}(o_k-\bar{o})^2}\,. \tag{5}$$

The coefficients of determination, which is the square of the Pearson product-moment correlation coefficient, are computed using the equation,

$$R^2 = \frac{m^2\sum_{k=1}^{n}(o_k-\bar{o})^2}{\sum_{k=1}^{n}(f_k-\bar{f})^2}, \tag{6}$$

where the numerator is the regression sum of squares and the denominator is the total sum of squares.

To determine how much a forecast of composite reflectivity needs to be "blurred" to exhibit "useful" skill, the FSS is computed for 30-, 40-, and 50-dBZ thresholds using the equation,

$$FSS = 1 - \frac{\frac{1}{N}\sum_N (P_f - P_o)^2}{\frac{1}{N}\left(\sum_N P_f^2 + \sum_N P_o^2\right)}, \quad (7)$$

where P_f and P_o are the fractional coverages of forecasted and observed values, respectively, exceeding a threshold within a neighborhood window and N is the number of neighborhood windows used for each neighborhood size (Ebert, 2008, 2009). Since FSS ranges from 0 (no skill) to 1 (perfect skill), the equation, $FSS_{useful} = 0.5 + f_o/2$, where f_o is the base rate of the observed values, from Roberts and Lean (2008) is used in this study to determine the smallest scales with "useful" skill, which is considered to be halfway between a random forecast ($FSS_{random} = f_o$) and a perfect forecast. Square neighborhood windows are employed, and when grid points within neighborhood windows extend beyond the edge of the domain, those grid points are assigned the value of 0. The observed composite reflectivity from the National Severe Storm Laboratory's national mosaic and quantitative precipitation estimation (NMQ; Zhang et al., 2011) system is used as the observation field.

A new object-based verification technique was developed to assess model performance by verifying simulated mesocyclone centers, via the UH field, with estimated tornado points. The locations of the six tornadoes (i.e., B1, B2, B4, C1, D1, and D2 in Figure 3) associated with the three storms of interest are estimated every minute based on NWS damage surveys, radar data, and high-resolution aerial photos from Google Maps. Two adjacent layers of UH (namely, 0–1 km and 1–6 km) are used in the verification process to verify the simulations and are derived using the equation,

$$UH = \int_{z_1}^{z_2} w\zeta dz, \tag{8}$$

which is the integral through the depth, z_1 to z_2, of the product of vertical velocity, w, and the vertical component of relative vorticity, ζ. These two layers are intended to represent simulated low-level and mid-level mesocyclones, respectively. Kain et al. (2008) used UH from 2–5-km AGL to signify mid-level mesocyclones, but for this study, a deeper layer of UH is utilized to give more robust UH values by capturing more of the simulated mid-level mesocyclones, while the 0–1-km UH value is used to pinpoint rotation near the ground, expected to be more closely associated with damage reports.

Since UH is a 2-D field and not point data, a simple 2-D object-based technique is utilized to find UH-weighted centers (analogous to mass-weighted centers), which will be compared to the estimated tornado points. A search radius of 4 km (i.e., 4 grid points) is used to isolate 1–6-km (0–1-km) UH maxima that are greater than or equal to 300 m^2 s^{-2} (15 m^2 s^{-2}) and their surrounding grid point values. A max UH value is considered a UH-center candidate if 4 out of 8 (1 out of 8) of the adjacent grid point values equals or exceeds 150 m^2 s^{-2} (10 m^2 s^{-2}). Once the UH-center candidates are determined, the UH-weighted center is computed using a radius of 3 km (2 km) extending from the grid point with the max UH value. The 0–1-km UH-weighted centers are filtered by requiring a 1–6-km UH-weighted center to concurrently exist within 5 km. This is meant to identify centers of strong low-level updraft rotation that are also associated with a significant mid-level rotating updraft.

With the filtered 0–1-km UH-weighted center locations, an objective verification technique is used to quantify location and timing errors. This technique is

similar to the method used to verify simulated tropical cyclone tracks and intensities (e.g., Xue et al., 2013), but in addition to displacement errors, timing errors are estimated as well since the lead time for tornadic circulations is relatively more important than for tropical cyclones (i.e., small vs. large time scales, respectively). First, distance errors are computed between the estimated tornado point locations and the nearest 0–1-km UH (0-1UH) center locations at coincident times (referred to as "same time", or ST, for rest of paper). Second, distance and timing errors are calculated between the estimated tornado point locations and the nearest 0-1UH center locations at any time during the life of the tornadoes of interest (referred to as "any time", or AT, for rest of paper). The average max 0-1UH value for each center and the total number of 0-1UH centers are also computed for further evaluation.

Chapter 4 – Impact of Various Microphysics Schemes

4.1 Introduction and Background

Cloud and precipitation microphysics parameterizations play a large role in all scales of NWP. For example, microphysics parameterizations have an impact on latent heating and cooling associated with water phase changes (i.e., freezing, melting, evaporation, condensation, deposition, and sublimation), precipitation type and amount, and surface variables (e.g., varying types of downdrafts can alter near-surface temperatures and/or wind speeds and directions and cloud cover can indirectly alter near-surface temperatures through blocking or trapping of long-wave and short-wave radiation). From a short-term, small-scale convective forecasting perspective, these microphysics parameterization scheme impacts can affect simulated storm intensity, motion (i.e., speed and direction), and mode (e.g., supercell vs. MCS), which are all attributes of severe weather forecasting.

Microphysical parameterization schemes are generally characterized as bin and bulk schemes. Bin schemes use discrete bins to form particle size distributions for various particle sizes and types, and bulk schemes use continuous functional forms to describe particle size distributions. Bulk schemes are largely preferred for NWP and storm simulation studies over bin schemes due to bin schemes being computationally more expensive and having less success at predicting changes in ice particle concentrations (Stensrud 2007). Therefore, this study will focus on the impact of various bulk schemes on storm-scale forecasts. Specifically, research experiments are done using five different microphysics parameterization schemes: Lin 3-ice microphysics scheme (LIN3; Lin et al., 1983), Weather Research and Forecasting (WRF) single-

moment 6-class microphysics scheme (WSM6; Hong and Lim, 2006), Milbrandt and Yau single-moment bulk microphysics scheme (MYSM), Milbrandt and Yau double-moment bulk microphysics scheme (MYDM), and Milbrandt and Yau triple-moment bulk microphysics scheme (MYTM; Milbrandt and Yau 2005a,b).

All five of these schemes are based on particle size distributions most generally described by a gamma distribution:

$$N_x(D) = N_{ox} D^{\alpha_x} e^{-\lambda_x D}, \tag{9}$$

where N_x is the number of particles per unit volume (m^{-4}) for species x, D is the particle diameter (m), N_{ox} is the intercept parameter, α_x is the spectral shape parameter, and λ_x is the slope parameter. LIN3, WSM6, and MYSM only predict the mixing ratios for each species, namely water vapor (q_v), cloud water (q_c), rain (q_r), snow (q_s), ice (q_i), hail (q_h; LIN3 and MYSM only), and graupel (q_g; WSM6 and MYSM only). For these single-moment schemes, $\alpha_x = 0$, N_{ox} is constant, and λ_x varies via the mixing ratio (i.e., inverse-exponential distribution). MYDM predicts both mixing ratios (q_x) and number concentrations (N_x) for the same species as MYSM, so while $\alpha_x = 0$ for this study, both N_{ox} and λ_x can vary with q_x and N_x. Finally, MYTM predicts q_x, N_x, and reflectivities (Z_x) for the same species as MYSM. The addition of the third moment allows for α_x to vary along with N_{ox} and λ_x (Milbrandt and Yau 2005a).

4.2 Experiment Design

Since this experiment intends to explore the capabilities of the forecast system in a realistic setting, the numerical simulations use data from multiple observing platforms. Surface observations from NWS and FAA METAR and Oklahoma Mesonet stations along with radial wind and reflectivity data from the WSR-88D [Dallas/Fort Worth (KFWS), Dodge City (KDDC), Frederick (KFDR), Tulsa (KINX), Twin Lakes (KTLX), Vance (KVNX), and Wichita (KICT)] and CASA IP-1 X-band [Chickasha (KSAO), Cyril (KCYR), Lawton (KLWE), and Rush Springs (KRSP); see Figure 9] radar networks (McLaughlin et al., 2009) are ingested into the initial analyses of the numerical simulations.

Recently, the ARPS's (ARPS; Xue et al., 2000; Xue et al., 2001; Xue et al., 2003) ADAS complex cloud analysis package (Hu et al., 2006a,b) was updated for several microphysics schemes, including the five in this study (Brewster and Stratman, 2015). The goal of this update was to improve analyses of hydrometeors using scheme-specific reflectivity inversion equations. The 1800-UTC 12-km NAM (North American Mesoscale) model's 3-hour forecast is used as a background field in the 3DVAR data assimilation (Gao et al., 2004) and complex cloud analysis (Hu et al, 2006a, Brewster et al., 2015) to produce an initial analysis on a 323x353-km domain with 1-km horizontal grid spacing (Figure 9) and 53 vertically-stretched levels with a minimum vertical grid spacing of 20 m at the bottom. Three analysis passes with 20, 50, and 50 iterations for minimizations and horizontal influence radii of 45, 2, and 1 km, respectively, are used to produce the 3DVAR analysis through the minimization of the cost function. The surface in-situ data are utilized in the first and third passes, while the

radar data are applied in the second and third passes. In addition, a 3-D mass divergence constraint is utilized to couple the wind components together (Hu et al., 2006b).

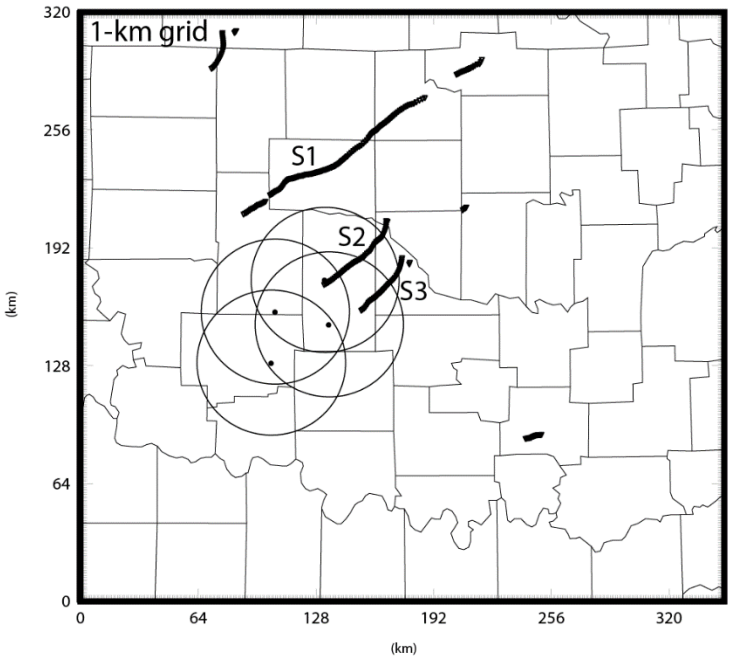

Figure 9. Domain of numerical simulations with CASA radar locations and 40-km range rings, estimated tornado points, and storm IDs.

After the 3DVAR analysis is produced, an ARPS's model simulation is integrated forward to produce forecasts out to 125 minutes. During the first 5 min, an incremental analysis update (IAU, Bloom et al., 1996) assimilation is performed by introducing fractional analysis increments every 20 s (i.e., the fractional increments are added directly to the various variable fields). The fractional increments have a triangular time weighting pattern to slowly ramp up and then down with the maximum around the mid-point of the assimilation window (Figure 10). The increments are applied to all fields except for vertical velocity and pressure since those two fields are not directly observed in 3-D and will quickly respond to the other fields to create a balanced state.

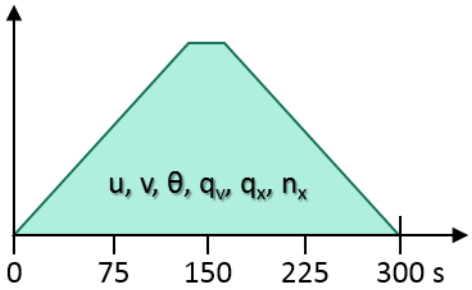

Figure 10. Schematic of the temporally-weighted distributions of increments for the Control runs.

The simulation proceeds on its own for the remaining 120 min. During this integration of ARPS, a big and small time step of 2.0 s and 0.5 s, respectively, are employed in the leapfrog time formulation. In addition, the 1800-UTC 12-km NAM forecasts are used for the lateral boundary conditions. Some other model details include: 4th-order momentum advection in both the horizontal and vertical directions, scalar advection using Zalesak's multi-dimensional version of flux-corrected transport (Zalesak, 1979), 1.5-order TKE closure based on Sun and Chang (1986), 4th-order computational mixing, Rayleigh damping beginning at 12-km AGL, National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) atmospheric parameterization of long- and short-wave radiation processes (Chou, 1990, 1992 and Chou and Suarez, 1994, respectively) , surface fluxes calculated from stability-dependent surface drag coefficients using predicted surface temperature and volumetric water content, and two-layer force-store soil model based on Noilhan and Planton (1989).

ARPS Begin – End	S1 2031 Z - 2046 Z 2050 Z - 2235 Z	S2 2206 Z - 2301 Z	S3 2226 Z - 2305 Z 2302 Z - 2303 Z
1900 Z – 2100 Z	+1:31		
1930 Z – 2130 Z	+1:01		
2000 Z – 2200 Z	+0:31		
2030 Z – 2230 Z	+0:01	+1:36	+1:56
2100 Z – 2300 Z	-0:29	+1:06	+1:26
2130 Z – 2330 Z	-0:59	+0:36	+0:56
2200 Z – 0000 Z	-1:29	+0:06	+0:26
2230 Z – 0030 Z	-1:59	-0:24	-0:04

Table 1. Storm ID names and associated tornado and ARPS-simulation forecast times. Positive (green) and negative (red) values indicate the time difference (hour:minute) between the start of the simulations and first tornadogenesis for each of the storms.

In addition to microphysics diversity, simulations are run using a potential future real-time, storm-scale forecasting framework. Eight simulations are integrated with start times every 30 minutes beginning at 1900 UTC (with a 5-minute IAU performed prior to the simulation start times) and ending at 2230 UTC (Table 1). With this framework, four simulations provide forecasts encompassing each observed storm's first tornadogenesis time (Table 1). The first storm (S1; storms depicted in Figure 9) developed and stayed outside the CASA radar network and produced two tornadoes, including the outbreak's only EF-5 tornado. The second and third storms (S2 and S3, respectively) developed in the CASA radar network and both produced EF-4 tornadoes, which dissipated before impacting the Oklahoma City metro area.

4.3 Observation Point-Based Results

Since microphysics parameterization schemes can have an impact on near-surface variables, the RMSE is computed for each forecast time for near-surface temperature, dewpoint temperature, pressure, and u- and v-wind components. MYTM has the smallest average RMSE for temperature for all simulations and forecast times with 1.63°C, but MYDM and MYSM's average RMSE for temperature are within 0.03°C (Figure 11a). LIN3 and WSM6 both have average RMSEs for temperature around 0.5°C larger than the MY schemes. The RMSEs for dewpoint temperature are approximately twice the RMSEs for temperature, but once again, the MY schemes have the smallest average RMSEs even though the differences are smaller than for temperature (Figure 11b). For pressure, the MY schemes again having the smallest average RMSEs for pressure, but the differences between the average RMSEs are less than 0.3 hPa (Figure 11c).

The RMSEs for u-wind gradually increase with later model initialization times while the RMSEs for v-wind largely remain within the 2 to 4 m/s range (Figure 11d,e). Interestingly, the average RMSEs for u-wind are about 0.5 m/s larger than the RMSEs for v-wind. This difference could partially be due to boundary layer and turbulence parameterization schemes improperly mixing westerlies aloft down to the surface in areas of downdrafts or deep, well-mixed boundary layers. As with the other variables, the MY schemes generally have the smallest RMSEs for both wind components. For all five near-surface variables, MYDM and MYTM exhibit very similar RMSEs with MYSM having only a slightly larger average RMSE. Compared to WSM6, LIN3 has

smaller average RMSEs for all variables except for temperature, which the difference is less than 0.1°C.

For further comparison, linear regression slopes, coefficients of determination, R^2, and PDFs are computed to assess biases and correlations, respectively, for the same variables, except for pressure. For temperature, MYDM and MYTM have the largest R^2 values while having the smallest slopes (Figure 12a). This is due to MYDM and MYTM having a warm bias for observed temperatures below 25°C. WSM6's slope for temperature is closest to 1 with LIN3 having a slightly smaller slope, but their R^2 values are substantially smaller than the MY schemes (Figure 12a). MYSM's slope falls in between the two pairs of slopes, but the R^2 value is closer to MYDM and MYTM. All schemes shift the distribution peak to the bin that is 1°C warmer than the bin containing the observed distribution peak. Also, LIN3 and WSM6 exhibit higher probabilities in the bins less than 23°C, indicating larger and colder cold pools. Average base rates for temperatures below 20–25°C reveal that is indeed the case for LIN3 and WSM6 (Figure 13). On average, all microphysics schemes lead to forecasted temperatures cooler than observed temperatures for temperatures greater than about 30°C. For dewpoint temperature, all microphysics schemes tend to be too moist for observed dewpoint temperatures less than 15°C (i.e., the dry side of the dryline; Figure 12b). The MY schemes have the largest R^2 values, and their slopes are slightly closer to 1 than the other microphysics schemes. This is due to their better performance near the observed distribution peak, which was well forecasted by all schemes.

The forecasted u-wind speeds tend to be too slow for observed u-wind speeds greater than about 0 m/s and are too fast for observed u-component wind speeds less

than about -5 m/s (Figure 12c). The PDFs are similar among the observations and schemes' forecasts with the majority of the u-wind values existing within the -5 to 0 m/s range, which is due to the relatively expansive warm, moist sector. The MY schemes' simulations produce the steepest slopes and largest R^2 values for u-wind, but both values remain near or below 0.5 (Figure 12c). Similar to u-wind, the forecasted v-wind speeds are generally too slow (too fast) for observed v-wind greater than (less than) about 5 m/s (Figure 12d). All schemes' simulations are substantially worse at predicting v-wind than u-wind based on shallow slopes (i.e., < 0.35), small R^2 values (< 0.20), and forecasted distributions of v-wind being too narrow.

Overall, the comparison between Mesonet observations and forecasted values using various statistics has highlighted some model successes and failures. The MY schemes, especially MYDM and MYTM, exhibit the smallest RMSEs for all five near-surface variables, but the largest RMSE differences between the MY schemes and LIN3 and WSM6 are for the near-surface temperatures likely due to LIN3 and WSM6 producing cold pools that are too cold and too large. All simulations result in cooler tropical air masses (i.e., continental and maritime), but slightly less so for the MY schemes. The large differences between the observed and forecasted wind fields (i.e., direction and magnitude) largely stems from areas of observed and forecasted convection and areas on the dry side of the dry line (Figure 14). These various air masses within the model domain likely contribute to the nonlinear nature of the biases (i.e., overforecasting and underforecasting), especially for the near-surface dewpoint temperature field. Numerous parameterizations and assumptions (e.g., microphysics schemes, turbulence schemes, PBL schemes, soil types, and roughness lengths) likely

contribute to not only these problematic near-surface wind forecasts, but also the other near-surface variables.

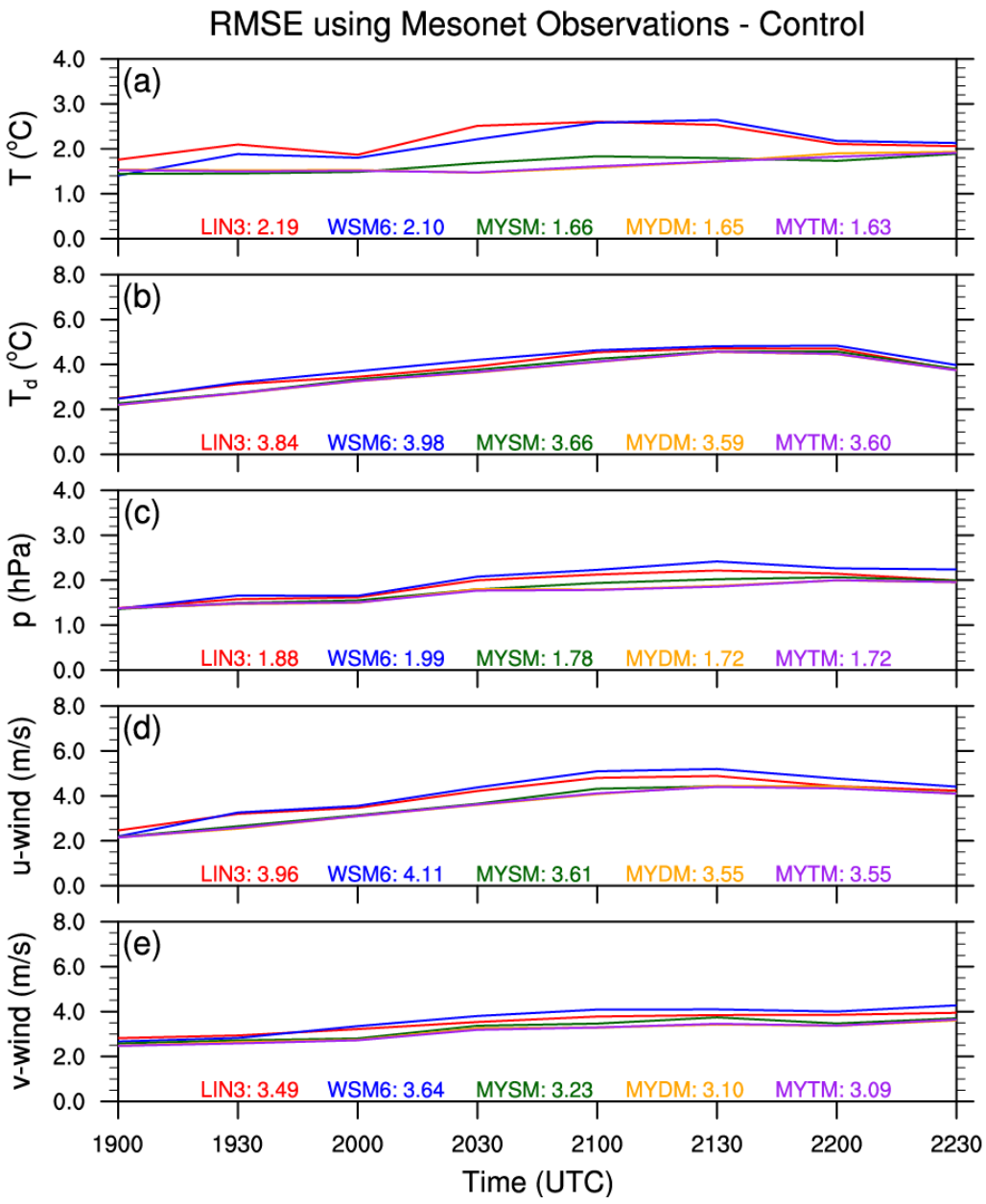

Figure 11. Average RMSEs plotted for each simulation for near-surface (a) temperature (°C), (b) dewpoint temperature (°C), (c) pressure (hPa), (d) u-component wind (m/s), and (e) v-component wind (m/s). Average RMSEs for all simulations for each microphysics scheme are annotated on the bottom of the plots.

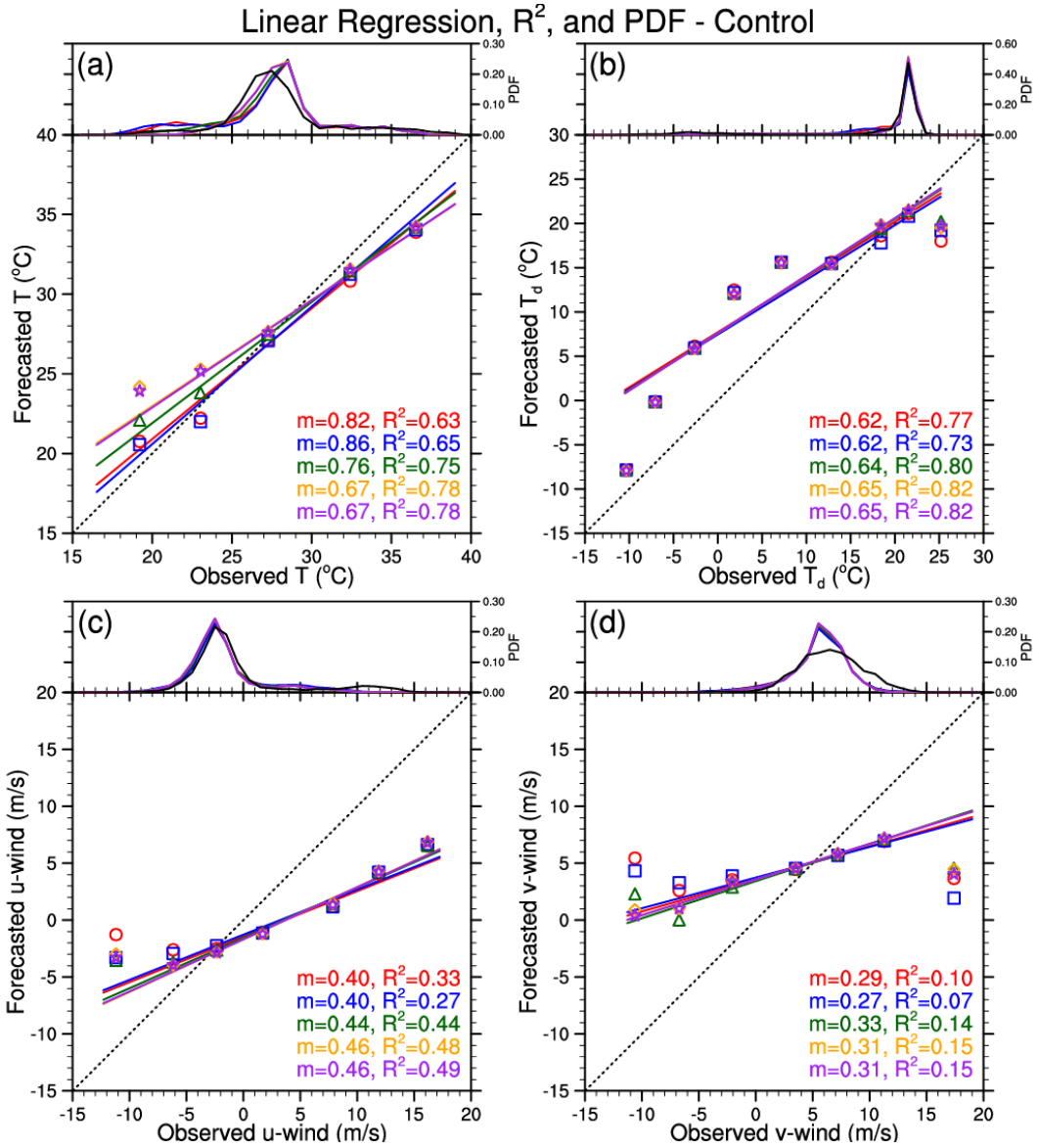

Figure 12. Linear regression slopes, *m*, and R^2 for all paired observations and forecasts for (a) temperature (°C), (b) dewpoint temperature (°C), (c) u-component wind (m/s), and (d) v-component wind (m/s) are plotted and annotated for LIN3 (red), WSM6 (blue), MYSM (green), MYDM (orange), and MYTM (purple). Colored shapes depict averages for observations in distinct 5-unit bins and their associated forecast values. Probability distribution functions for the number of forecasted and observed values within 1-unit bins are plotted in the top portion of each variable's plot window.

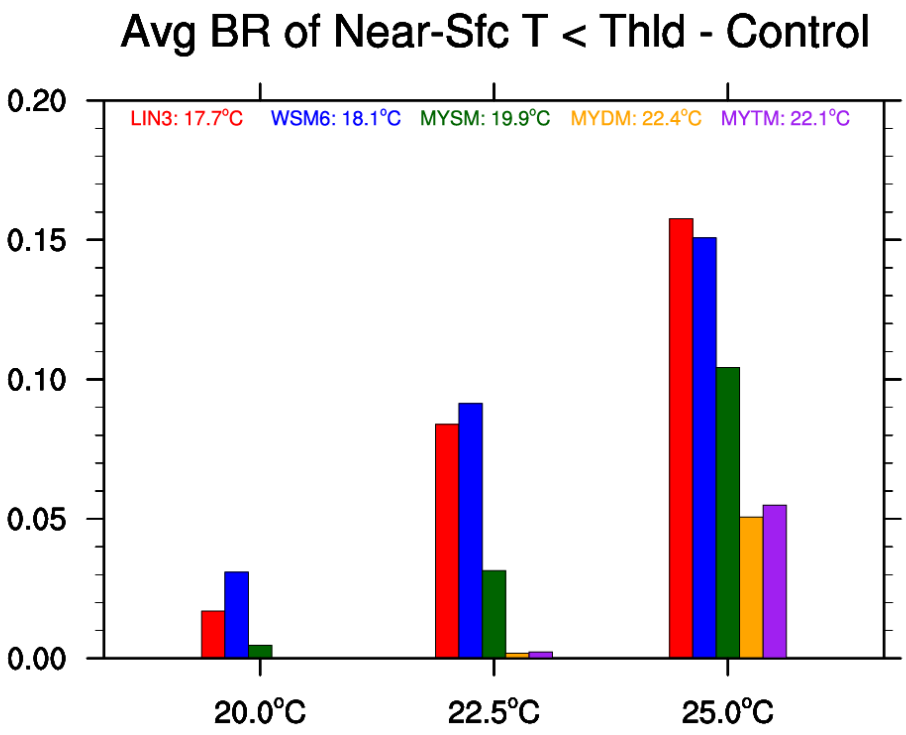

Figure 13. Bar graph of average base rates of temperatures less than 20.0°C, 22.5°C, and 25.0°C for the different microphysics schemes. Average minimum temperatures are annotated near the top of the plot window.

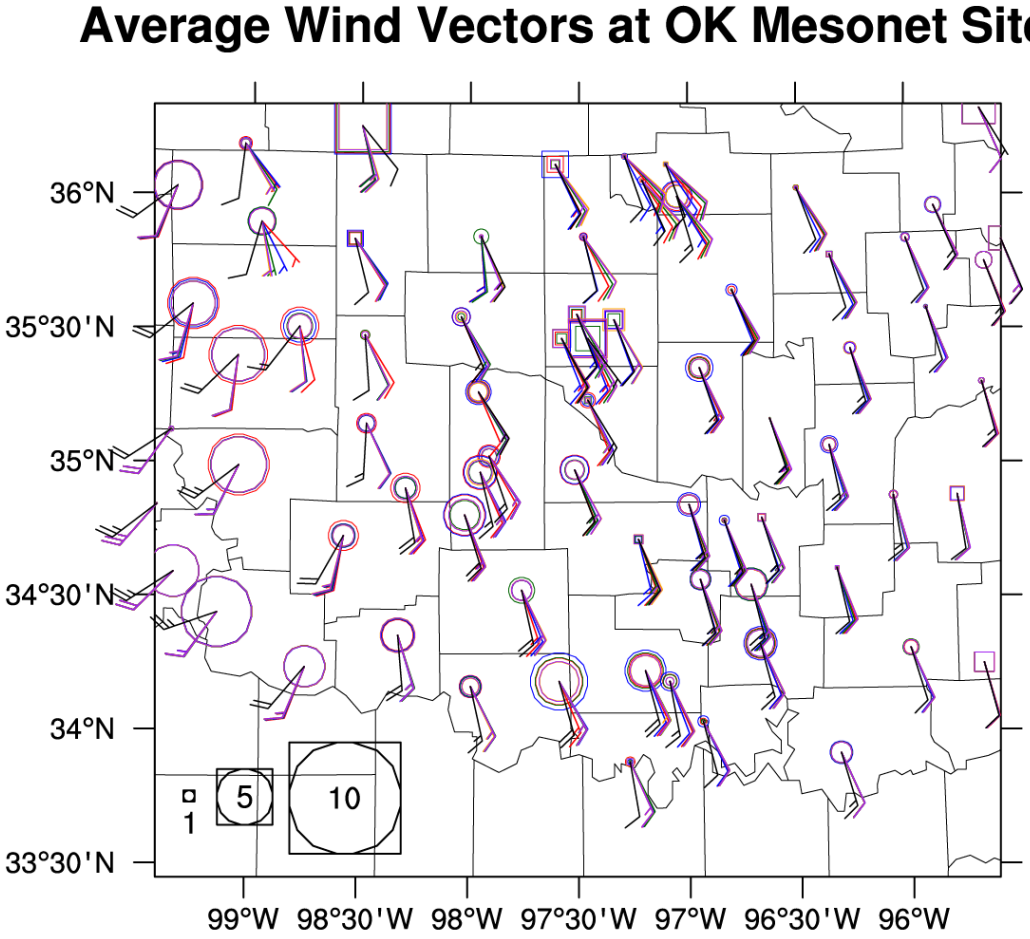

Figure 14. Keychain plot of average wind vectors at all Oklahoma Mesonet stations within the model domain. Wind barbs depict average wind speeds and directions for observed values (black) and forecast values from the Control runs using LIN3 (red), WSM6 (blue), MYSM (green), MYDM (orange), and MYTM (purple). On the wind barbs, a long (short) "feather" represents 10 m/s (5 m/s). Circles (squares) represent when average observed wind speeds (m/s) are smaller (larger) than average forecasted wind speeds (m/s).

4.4 Neighborhood-Based Results

Observed composite reflectivity is a spatially-dense, remotely-sensed field and, thus, provides a unique opportunity to spatially verify simulated convective systems on the entire grid domain. At the 30-dBZ threshold, LIN3 transitions to useful skill at the smallest scale (~16 km) with MYDM exhibiting useful skill beginning around ~19 km (Figure 15a). MYSM performs the worst with useful skill only at scales larger than ~41 km. Interestingly, MYTM is more similar to MYSM than MYDM at this threshold. LIN3 depicts the smallest scales for the transition from no useful skill to useful skill for five out of eight sets of simulations with average useful skill beginning at scales less than 9 km twice (Figure 15b). MYSM performs the worst with average useful skill never occurring at scales less than 17 km. In general, the earlier initiated simulations exhibit useful skill at smaller scales than the later initiated simulations likely due to the increasing number of storms and the complex nature of storm interactions.

At the 40-dBZ threshold, MYDM begins to show useful skill at scales around ~12 km (Figure 15c). Once again, MYSM performs the worst with useful skill on average not existing for scales below ~41 km. MYTM transitions to useful skill around ~18 km, which is now closer to MYDM than MYSM. While not considered useful skill, MYDM and MYTM both depict more skill than the other microphysics schemes at the smaller scales, but at larger scales, LIN3 and WSM6 match and/or beat MYDM and MYTM. For all initiation times, MYDM on average has the smallest scale at which useful skill begins with over half of the runs falling at or below 9 km (Figure 15d). Conversely, MYSM starts exhibiting useful skill at scales larger than the other microphysics schemes for all initiation times except for the 1900 UTC simulations.

Once again, there's a slight upward trend in scales where useful skill begins with later initiation times likely due to an increase in the areal coverage of reflectivity.

For the 50-dBZ threshold, MYDM and MYTM both depict useful skill starting around ~19 km and ~22 km, respectively, while the other three microphysical schemes lead to useful skill beginning around ~40 km (Figure 15e). For the first six model initiation times, MYDM and MYTM depict similar average scales where useful skill begins while WSM6 and MYSM usually perform the worst (Figure 15f). As an overview, all of the schemes generally become less skillful with increasing reflectivity threshold at the smaller scales, but at the larger scales, LIN3, WSM6, and MYSM remain similar with all thresholds while MYDM and MYTM increase in skillfulness with increasing thresholds (Figure 15a,c,e). Also, the differences among the microphysics schemes' simulations are greater at larger thresholds, but this is likely due to variability in the number of events exceeding the thresholds among the different schemes.

Overall, MYDM and MYTM performed the best at the 40- and 50-dBZ thresholds, and MYSM consistently performed the worst at this verification metric. As compared to the observed base rates, all of the microphysics schemes, especially the MY schemes, yield larger forecast base rates for each threshold (Figure 16). However, LIN3's base rates have the smallest biases, so this success likely results in similar FSSs at the 30- and 40-dBZ thresholds as MYDM and MYTM, respectively. Conversely, MYSM's base rates are the largest for each threshold, and this fact likely results in useful skill being first achieved at larger scales. Even with larger forecast base rates than LIN3 and WSM6, MYDM and MYTM still exhibit the most forecast skill at the smaller

scales at the 40- and 50-dBZ thresholds, which are usually associated with strong to severe weather when deep, moist convection is present.

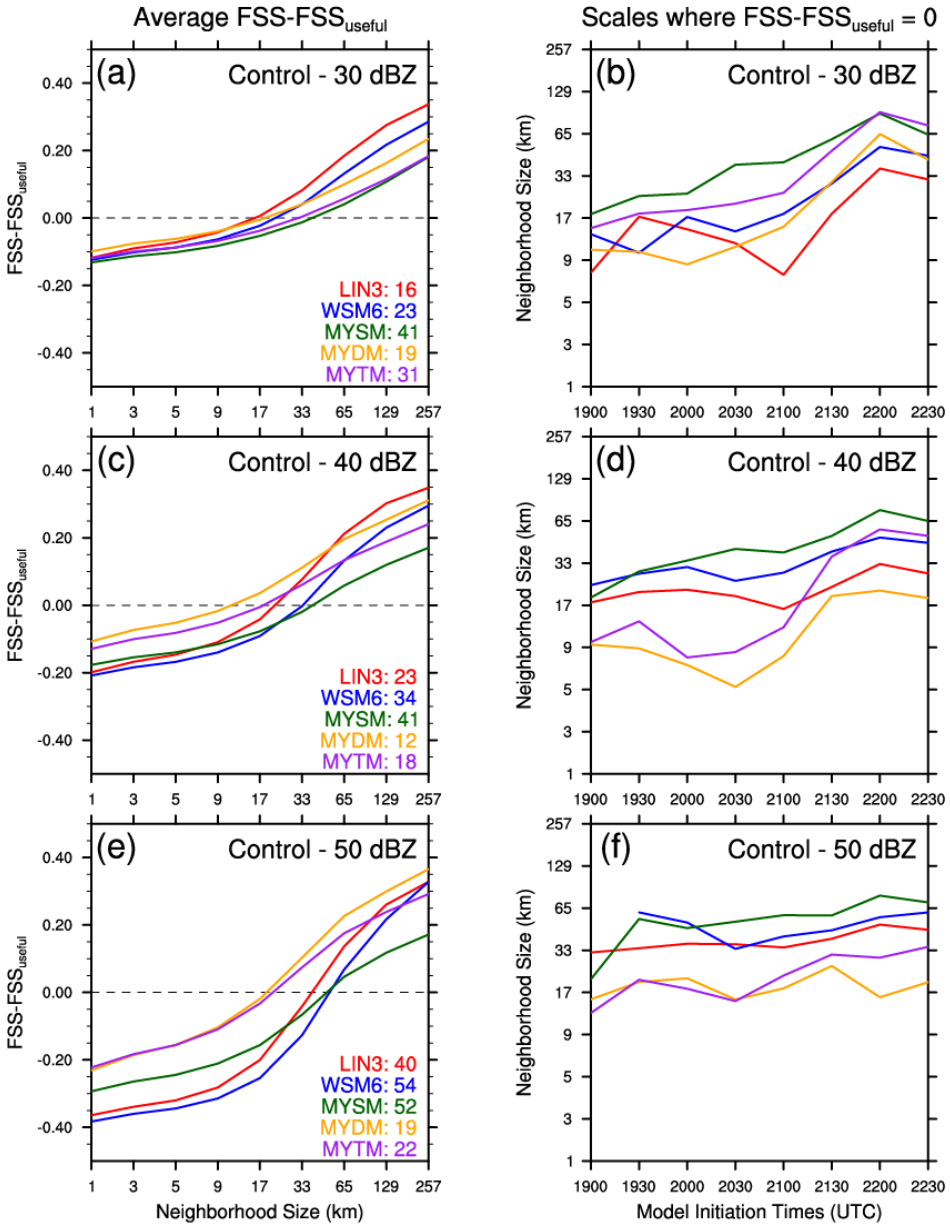

Figure 15. Average FSS-FSS$_{useful}$ across all simulations for LIN3 (red), WMS6 (blue), MYSM (green), MYDM (orange), and MYTM (purple) are plotted for thresholds of (a) 30 dBZ, (c) 40 dBZ, and (e) 50 dBZ. Average neighborhood sizes (km) at which FSS-FSS$_{useful}$ = 0 for each model initiation time (UTC) and microphysical scheme are plotted for thresholds of (b) 30 dBZ, (d) 40 dBZ, and (f) 50 dBZ. Average neighborhood size (km) at which FSS-FSS$_{useful}$ = 0 for each microphysics scheme is annotated in (a), (b), and (c).

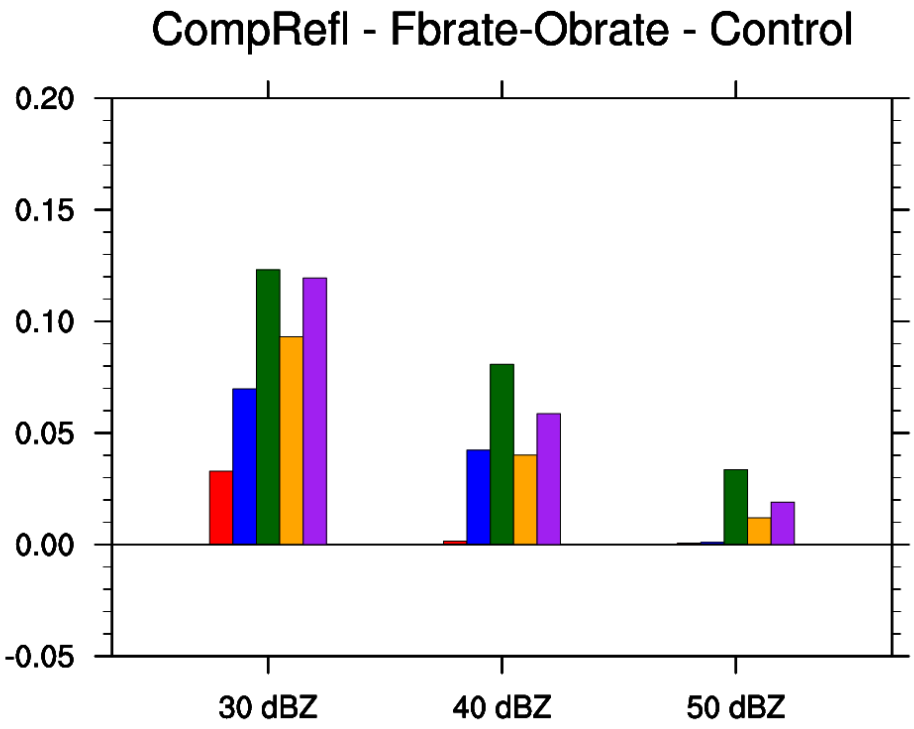

Figure 16. Bar graph of the forecasted base rate of composite reflectivity minus the observed base rate of composite reflectivity for the 30-dBZ, 40-dBZ, and 50-dBZ thresholds. Positive values indicate overforecasting, and negative values indicate underforecasting.

4.5 Object-Based Results

Supercell mesocyclones are mostly discontinuous in space, so simulated mesocyclones can be treated as objects using the UH field for verification purposes. The thresholds detailed in Section 3.3 define what constitutes a simulated mesocyclone, so the number of simulated mesocyclones can be greatly altered by adjusting the thresholds. For this study, less stringent thresholds are utilized to potentially produce more robust results and to agree with visual assessment of what constitutes a mesocyclone. However, the thresholds have minimal impact on the overall results. As stated before, 0-1UH centers are used to represent low-level mesocyclones and are filtered by the 1–6-km UH centers, which represent mid-level mesocyclones, to insure that detected features have deep support.

Using all simulations' output from every 5 min, several hundred 0-1UH centers were identified as objects for each microphysics scheme (Figure 17a). Both S1 and S2 have a plethora of 0-1UH centers near their tornado points, but S3 has substantially fewer 0-1UH centers nearby, which indicates relatively poor forecasting of S3's low-level circulation by the forecast system. 0-1UH centers generally surround S1's tornado points, but are largely clustered to the north of S2's tornado locations. All of the simulations in this set of experiments produce similar numbers of 0-1UH centers with LIN3 yielding the most and MYDM yielding the least (Figure 17b). The simulations initialized at 2130 UTC produce the largest number of 0-1UH centers with 150 to 200 centers. MYDM has the largest average max 0-1UH for each center, and WSM6 has the smallest average max 0-1UH for each center (Figure 17c). For most initialization times, the MY schemes, especially for MYDM and MYTM, have distinctly larger average max

0-1UH values for each center than LIN3 and WSM6, which are very similar in this respect. To briefly summarize, the MY schemes produce stronger and fewer 0-1UH centers than LIN3 and WSM6.

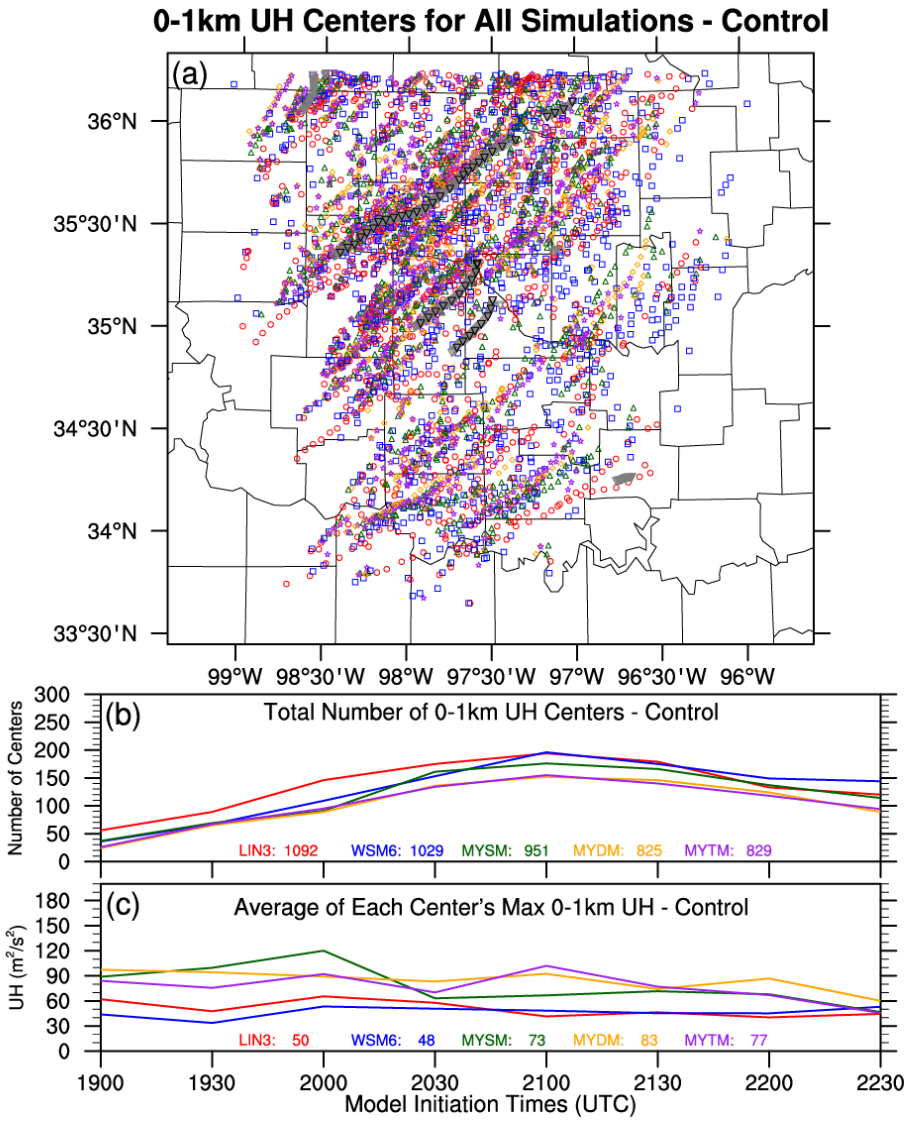

Figure 17. (a) Filtered 0-1UH centers for LIN3 (red circle), WSM6 (blue square), MYSM (green triangle), MYDM (orange diamond), and MYTM (purple star) for all simulations. Grey upside-down triangles respresent tornado locations every minute, while black upside-down triangles represent tornado locations occurring at the same times of the forecast output. (b) Line graph of the total number of 0-1UH centers for all simulations with the total number of centers across all simulations annotated at the bottom of the plot. (c) Line graph of average max 0-1UH (m^2/s^2) for each simulation's centers with the overall average max 0-1UH annotated at the bottom of the plot.

4.5.1 Storm 1

As seen previously, numerous 0-1UH centers are generally close to S1's tornado locations with many same-time centers within 10 km. Use of a 1-km model with the 3DVAR and IAU produced accurate simulations in a computationally efficient manner. Digging into the details, LIN3 and WMS6 lead to same-time 0-1UH centers that tend to be too fast and too far south of the tornado locations (Figure 18a). The MY schemes have same-time 0-1UH centers that are generally much closer to the tornado locations and each other (Figure 18a). For any-time 0-1UH centers, all schemes contribute to a similar number of centers, which are evenly spread north and south of S1's tornado locations (Figure 18b).

The MY schemes have average ST distance errors less than 11 km for 0-1UH, but LIN3 and WSM6 have average ST distance errors 5–10 km larger (Figure 18c). The ST distance errors for each microphysics scheme are fairly consistent across all simulations. It's worth noting that for the 1900-UTC simulations all of the microphysical schemes, except for LIN3, have average ST distance errors around or less than 5 km, which is considered to be very good given the nearly 90 minute lead time. As an example, the simulated reflectivity and 0-1UH fields from the MYDM run are plotted in Figure 19 and depict very successful forecasts of S1 at one and two hour lead times. The differences in the AT distance errors among the five schemes is smaller than for the ST distance errors, and the average AT distance errors are less than 9 km for all of the schemes (Figure 18d). Even though LIN3 and WSM6 exhibit AT distance errors similar to the MY schemes, those schemes' AT timing errors are on average mostly 15–20 min too fast (Figure 18e). The MY schemes have substantially smaller AT timing

errors with most 0-1UH centers occurring within ±10 min (i.e., too fast and too slow) of the estimated tornado times.

Overall for S1, all of schemes contributed to successful forecasts of low-level circulations near S1's tornado points. The MY schemes tend to produce the smallest distance and timing errors, and WSM6 generally has the largest errors with LIN3 not too different. The largest issue with LIN3 and WSM6 is their tendency to produce 0-1UH centers that are too fast. As previous studies and the near-surface variables have alluded to, this is likely due to those schemes producing cold pools that are too cold.

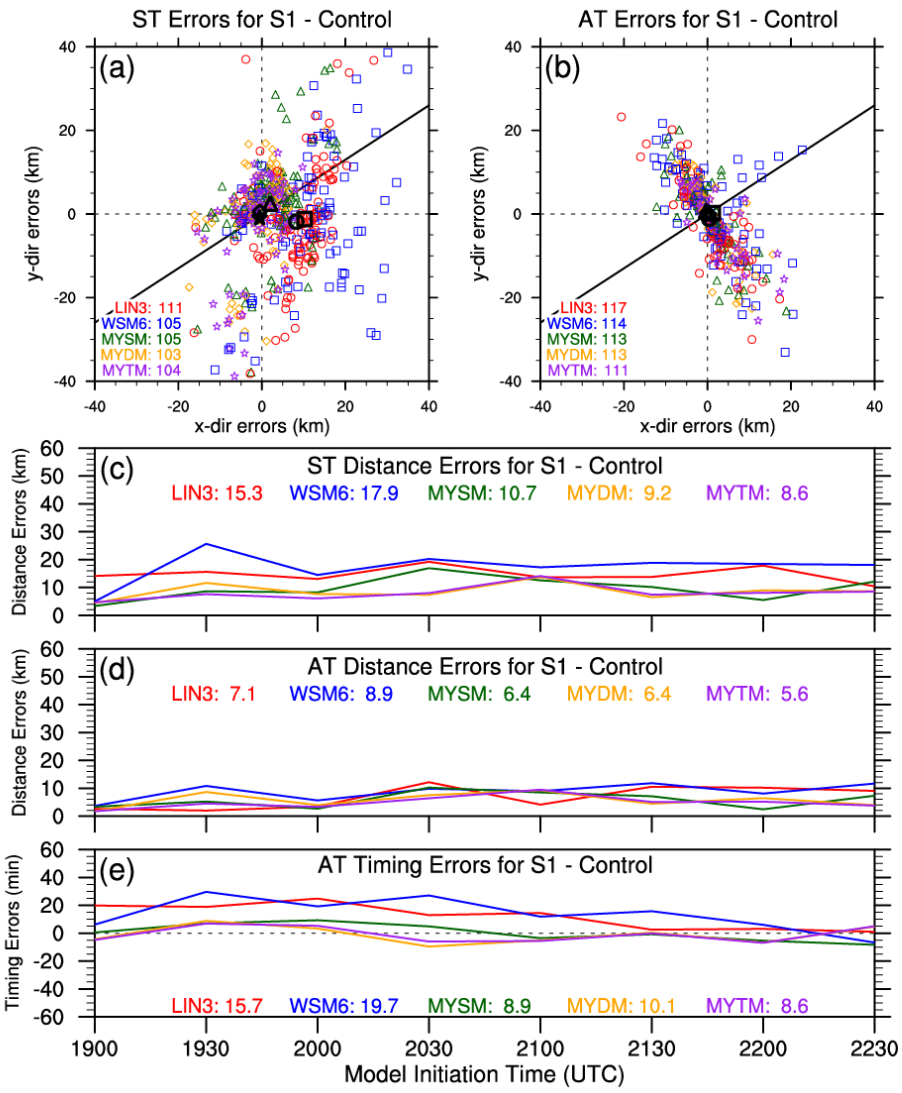

Figure 18. (a) Same-time and (b) any-time 0-1UH centers for Storm 1 from all simulations with LIN3 (red circle), WSM6 (blue square), MYSM (green triangle), MYDM (orange diamond), and MYTM (purple star). Black shapes represent the average locations of the microphysical schemes' 0-1UH centers. The number of 0-1UH centers within the plot window are annotated in (a) and (b). Line graphs of average (c) ST distance errors (km), (d) AT distance errors (km), and (e) AT timing errors (min) for all simulations' 0-1UH centers, and the respective averages across all simulations for each of the microphysical schemes is annotated within the plots.

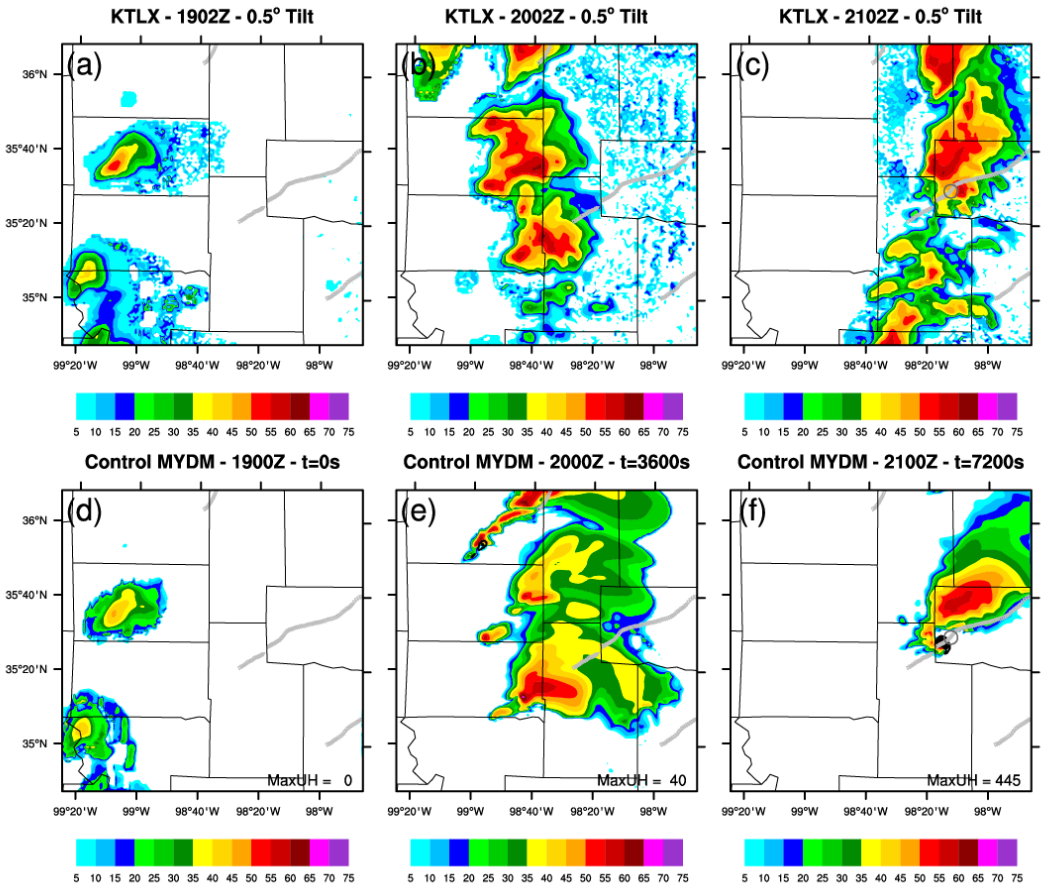

Figure 19. Observed 0.5°-tilt reflectivity (dBZ) from KTLX for (a) 1902 UTC, (b) 2002 UTC, and (c) 2102 UTC are interpolated and plotted to the 1-km model grid. Forecasts of 1-km AGL simulated reflectivity (dBZ; colored) from the 1900-UTC Control run using MYDM are plotted for (d) 1900 UTC (t = 0 s), (e) 2000 UTC (t = 3600 s), and (f) 2100 UTC (t = 7200 s). 0-1UH is contoured in black in (d), (e), and (f) from 10 m^2/s^2 to 210 m^2/s^2 with an interval of 25 m^2/s^2, and the max UH value (m^2/s^2) in each plot window is annotated near the bottom of each plot. Light gray upside-down triangles depict the estimated tornado locations every minute, and darker gray circles in (c) and (d) indicate the estimated location of the tornado occurring at 2100 UTC.

4.5.2 Storm 2

While ~70% fewer 0-1UH centers occur near S2 than S1, a plethora of 0-1UH centers still exist near S2's tornado points for all schemes. Given that a 1-km model is trying to accurately forecast the locations of tornadic circulations, these forecasts are considered successful. The MY schemes' ST 0-1UH centers are generally 5–10 km north of the tornado locations, and while LIN3's 0-1UH centers are mostly just too fast, WSM6's 0-1UH centers are too fast and too far south (Figure 20a). For any-time centers, the majority of the 0-1UH centers occur near the tornado locations with a bias to the north, except for WMS6 (Figure 20b). The earlier initiated sets of simulations exhibit more variability in ST distance errors among the different microphysics schemes than the later simulation runs, and average ST distance errors for 0-1UH centers range from about 15–20 km for the MY schemes to ~29 km for LIN3 and WSM6 (Figure 20c). The 0-1UH AT distance errors are substantially smaller than the ST distance errors with average errors less than 15 km for all schemes, and the differences among the schemes are smaller, as well (Figure 20d). However, timing errors range from ~14 min with MYDM and MYTM, which tend to be too slow, to nearly 30 min with LIN3, which is generally too fast (Figure 20e).

Overall for S2, the MY schemes mostly outperform the LIN3 and WSM6 with mostly smaller distance and timing errors. As mentioned before, there are fewer 0-1UH centers near S2 than near S1 even though S2 developed within the CASA radar network. However, S1 was, at least initially, a more isolated storm than S2, which in reality was influenced by storm interactions and mergers. This complication likely lead to larger

distance and timing errors, but the forecast system still managed to produce successful forecasts of low-level circulations.

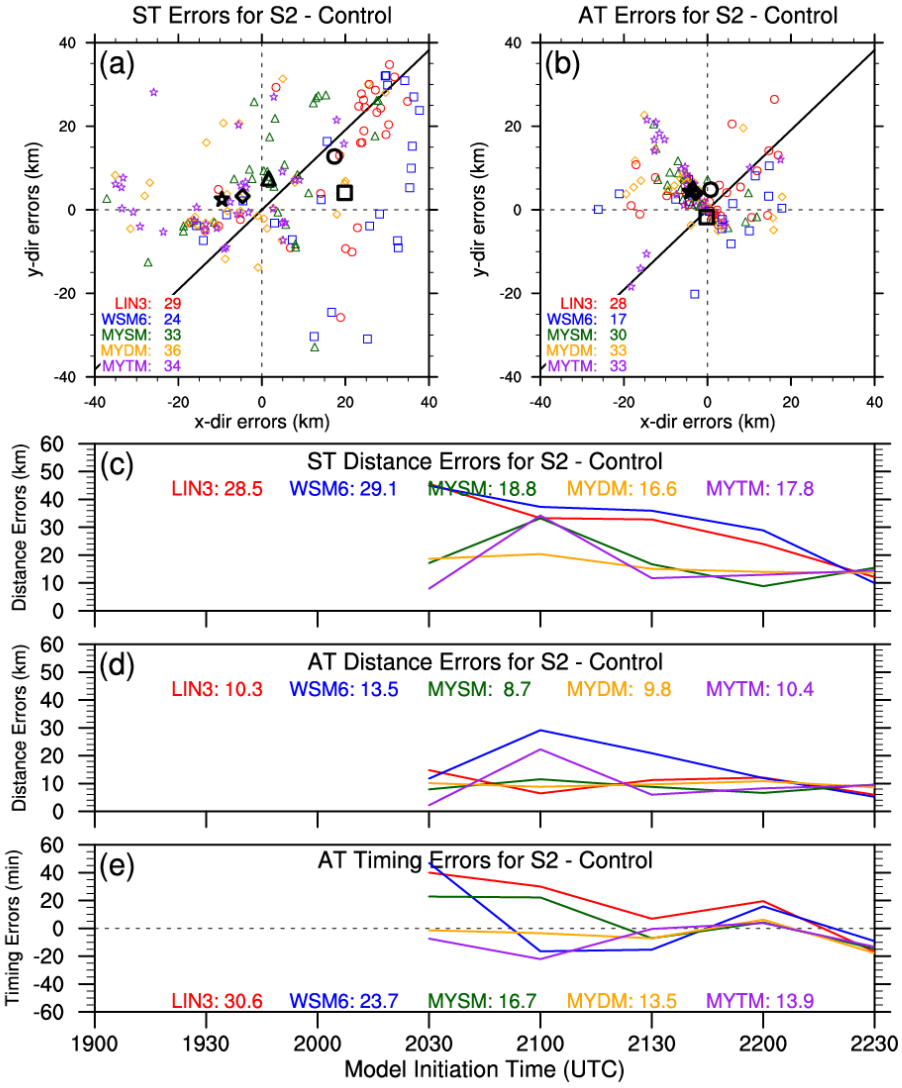

Figure 20. Same as for Figure 18, but for Storm 2.

4.5.3 Storm 3

While the microphysically-diverse set of simulations had good success with forecasting 0-1UH centers for S1 and S2, most of the simulations struggle with forecasting 0-1UH centers near S3's tornado locations, as depicted in Figure 17. Few same-time 0-1UH centers exist within 10 km of S3's tornado points, and average 0-1UH centers indicate a southeast bias due to a simulated supercell existing ~30 km to the southeast of S3 (Figure 21a). This bias is even more evident when timing differences are ignored (Figure 21b). Not surprisingly, ST and AT distance errors for 0-1UH are generally greater than 20 km for all microphysics schemes (Figure 21c,d). All simulations produce 0-1UH centers that are mostly within 20 minutes of the tornado occurrence times, and WSM6 is the only scheme that is consistently too fast (Figure 21e).

Since very few 0-1UH centers were forecasted near S3, the results above aren't too meaningful except that the forecast system as a whole struggled with S3 even though it developed and propagated through the CASA radar network. Other model configurations, besides using more advanced microphysics schemes, need to be explored to improve the forecasts of S3, which was directly affected by storm interactions and mergers.

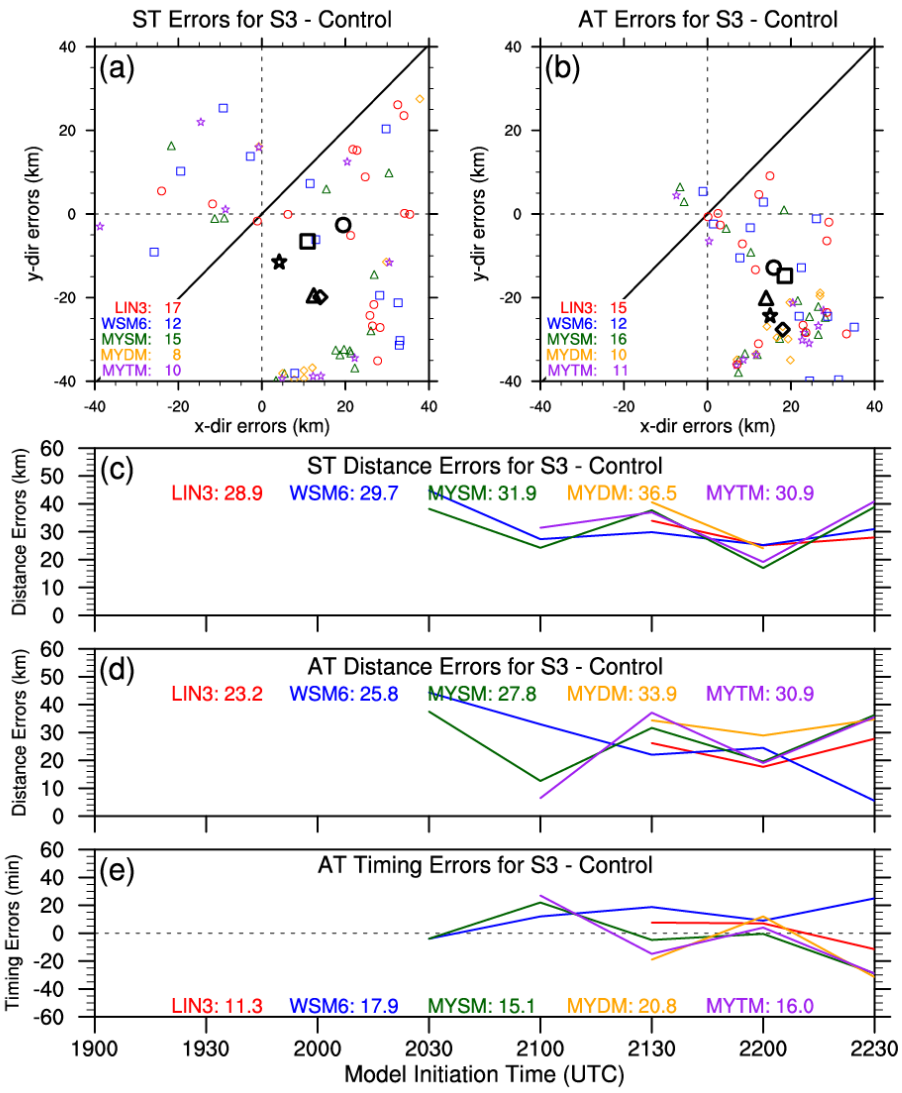

Figure 21. Same as for Figure 18, but for Storm 3.

Chapter 5 – Impact of Assimilating CASA Radar Data

5.1 Introduction and Background

The most popular way to remotely observe phenomena related to the evolution of severe storms is with radars. A national network of WSR-88D S-band radars sufficiently covers most of the heavily populated areas across the United States with a range of ~230 km. However, if operational meteorologists want to "see" what is occurring below 1-km AGL within a storm, the coverage is drastically reduced from near 100% coverage at 3 km to < 20% coverage at 500 m east of the Rocky Mountains (McLaughlin et al., 2009). X-band radars are a potential solution to filling these radar-observing gaps wherever they occur. For example, CASA radars, such as those used in CASA IP-1, have a range of ~40 km and provide higher-resolution reflectivity and radial velocity data with a range increment around 100 m (note, beam width is wider than WSR-88D radars' beam width), so that atmospheric processes leading to genesis, maintenance, and dissipation of tornadic circulations, which typically occur on scales *O*(10–100) meters and within the lowest couple of kilometers AGL, can be further observed and detected (e.g., Brotzge and Luttrell 2015).

A few studies have explored the impact of assimilating CASA radar data on the forecasts of a tornadic MCS, which occurred on 8–9 May 2007 and moved across the CASA radar network. Schenkman et al. (2011a) used 3DVAR and ADAS complex cloud analysis and found that assimilating CASA radar data led to small improvements, which yielded more accurate forecasts of the tornadic MCS. Schenkman et al. (2011b) expanded on this research, but instead of using 2-km horizontal grid spacing, they decreased the horizontal grid spacing to 400 m. This difference resulted in the CASA

radial velocity data having a larger positive impact on the forecasts of the low-level winds, gust fronts, and mesovortices in both strength and location. Snook et al. (2011) and Snook et al. (2012) used EnKF and also discovered that the forecasted low-level vorticity maximum associated with the mesovortex was stronger when the CASA radial velocity data was also assimilated due to improved wind fields in the lowest few kilometers AGL. This research will also examine the impact of assimilating CASA radar data, but this study is the first known study to explore this impact using an outbreak of tornadic supercells.

5.2 Experiment Design

This part of the study uses the same model configurations and forecast system framework as detailed in Chapter 4. To explore the impact of assimilating CASA reflectivity and radial velocity radar data on the forecasts of the 24 May 2011 tornado outbreak, three additional analyses and associated simulations are completed by withholding data from different radars during the data assimilation process. For one of the sets of simulations, all CASA radar data are withheld during the data assimilation process (NoCASA). In this case, the KTLX (Twin Lakes/Oklahoma City) WSR-88D radar provides fairly good low-level coverage in the domain of interest, so to investigate the ability of the CASA radars to act as gap-filling radars, KTLX radar data are withheld during the data assimilation process for another set of simulations, while all other WSR-88D radars are retained (NoKTLX). For completeness, additional simulations are completed by withholding both CASA and KTLX radar data during the data assimilation process (NoCASAKTLX). All simulations are completed using the five microphysics parameterization schemes described in Chapter 4.

5.3 Observation Point-Based Results

For RMSE, there are no substantial differences (i.e., a difference of 0.01 units is < 1%) between the Control and NoCASA runs (Figure 22). For NoKTLX, the average RMSEs are smaller for LIN3 and WSM6 for all five near-surface variables (Figure 23). The MY schemes show mixed changes (e.g., larger RMSEs for T and T_d and smaller RMSEs for p), but the changes are generally smaller than for LIN3 and WSM6. This could be due to less convection being assimilated by withholding KTLX radar data, so the cold pools are smaller in areal size. The changes in the RMSEs for NoCASAKTLX are similar to the changes for NoKTLX (Figure 24). This result indicates the withholding of KTLX radar data has more of an impact on RMSEs for near-surface variables than the withholding of CASA radar data during the data assimilation process.

Once again, there are no substantial differences between the NoCASA and Control runs for the linear regression slopes and R^2 values for the near-surface variables (Figure 25). The differences between the NoKTLX and Control runs for the slopes and R^2 values are slightly larger than the differences between the NoCASA and Control runs (Figure 26). Most of differences between the NoKTLX and Control runs are not substantial, but LIN3 and WSM6 mostly show improved R^2 values, which could be due to the smaller cold pools. As for the RMSEs, the NoCASAKTLX runs depict slopes and R^2 values more similar to NoKTLX than NoCASA (Figure 27).

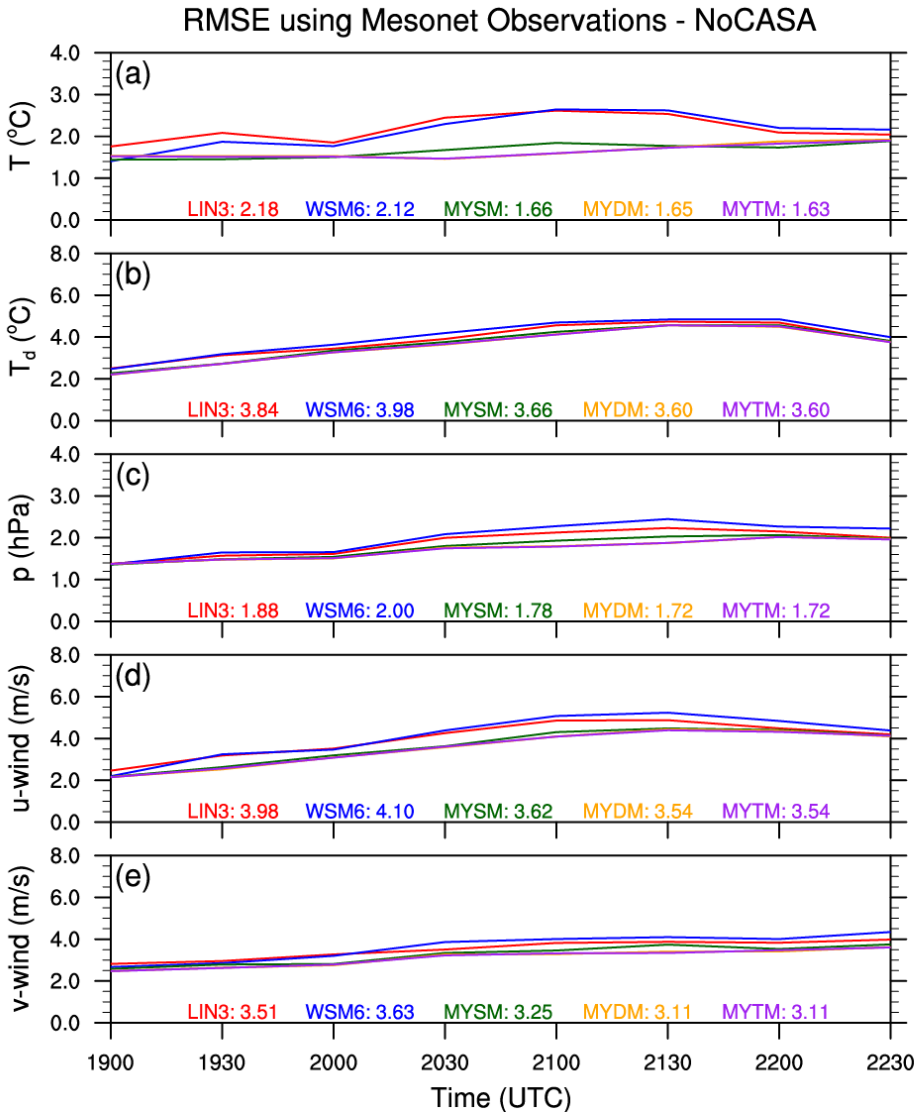

Figure 22. Same as for Figure 11 in Chapter 4, but for NoCASA.

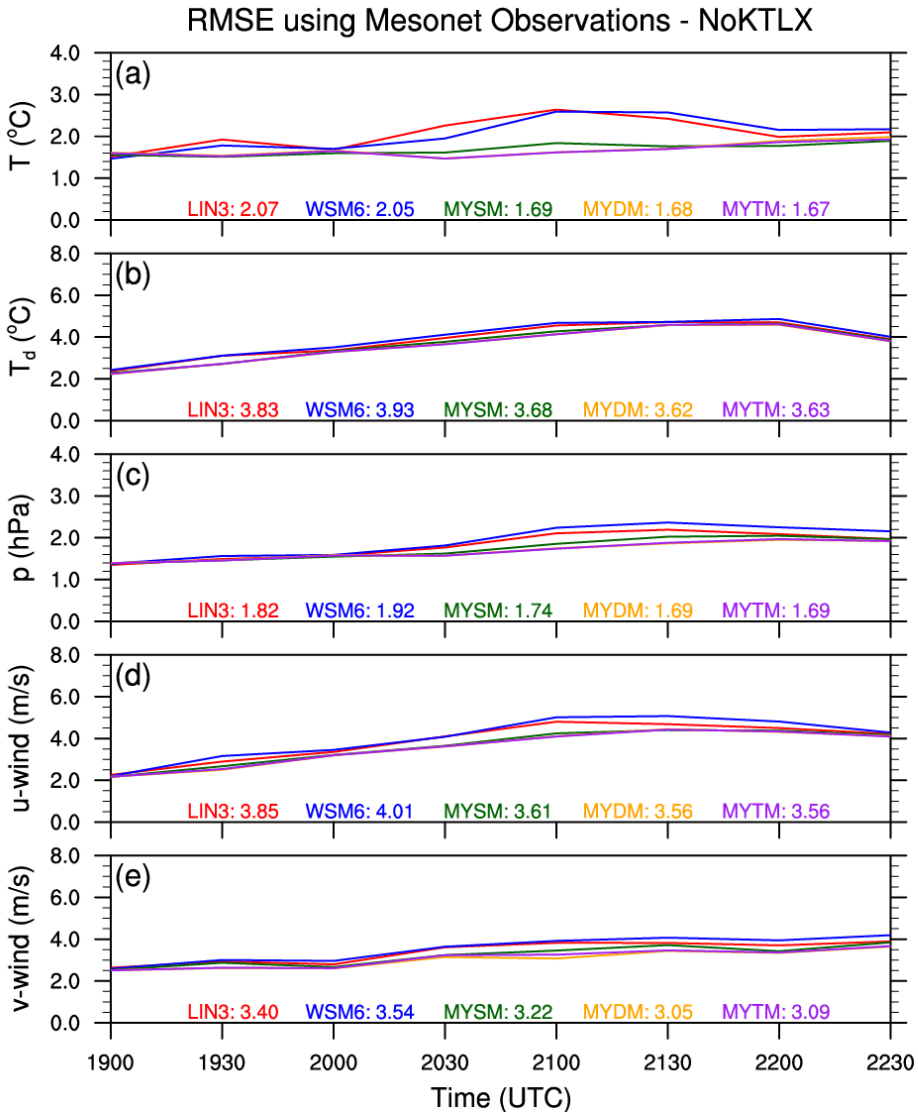

Figure 23. Same as for Figure 11 in Chapter 4, but for NoKTLX.

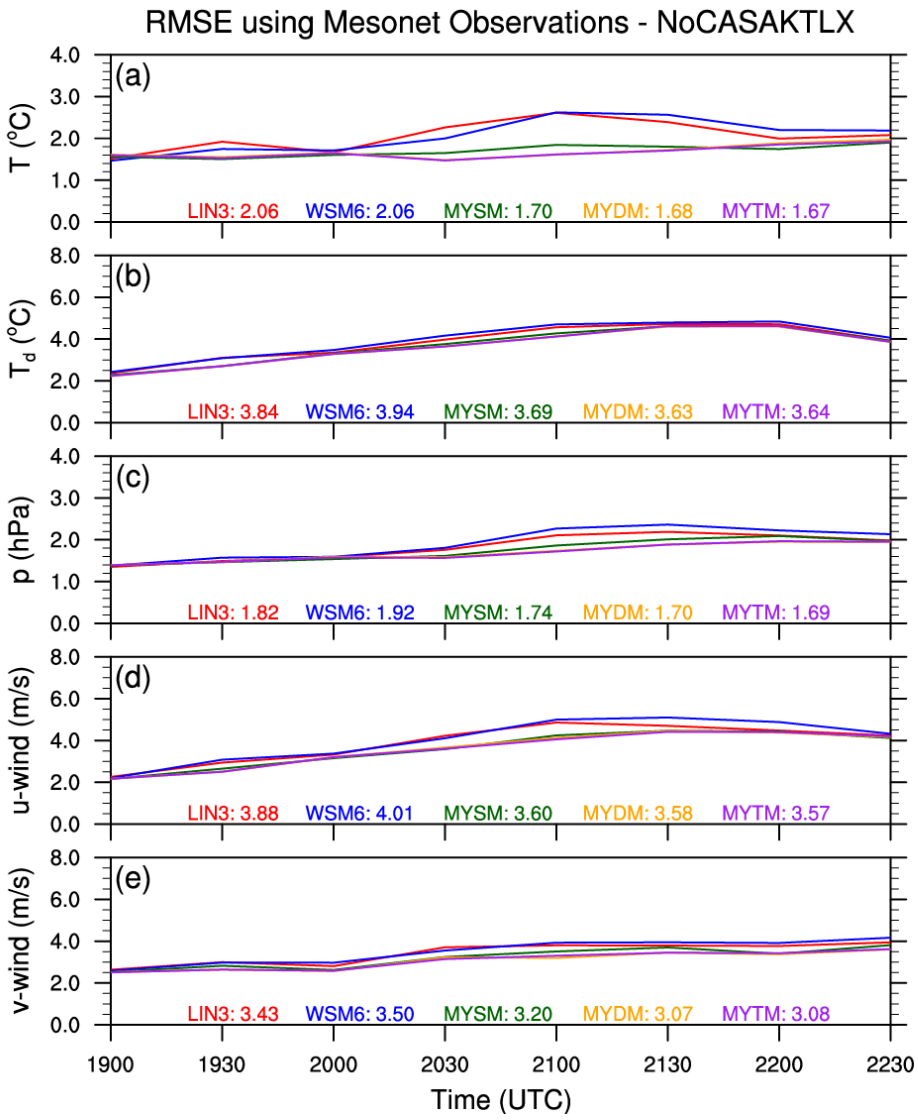

Figure 24. Same as for Figure 11 in Chapter 4, but for NoCASAKTLX.

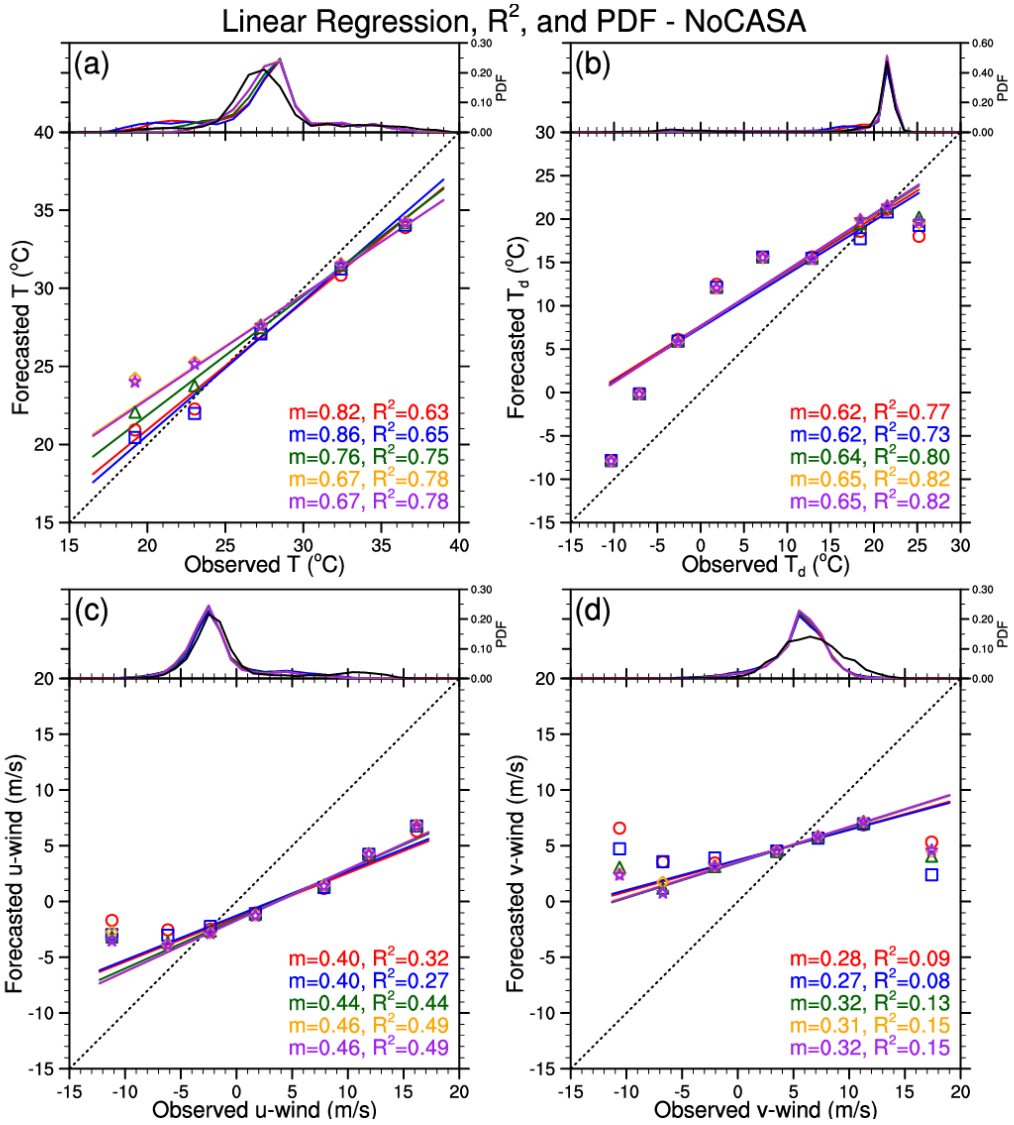

Figure 25. Same as for Figure 12 in Chapter 4, but for NoCASA.

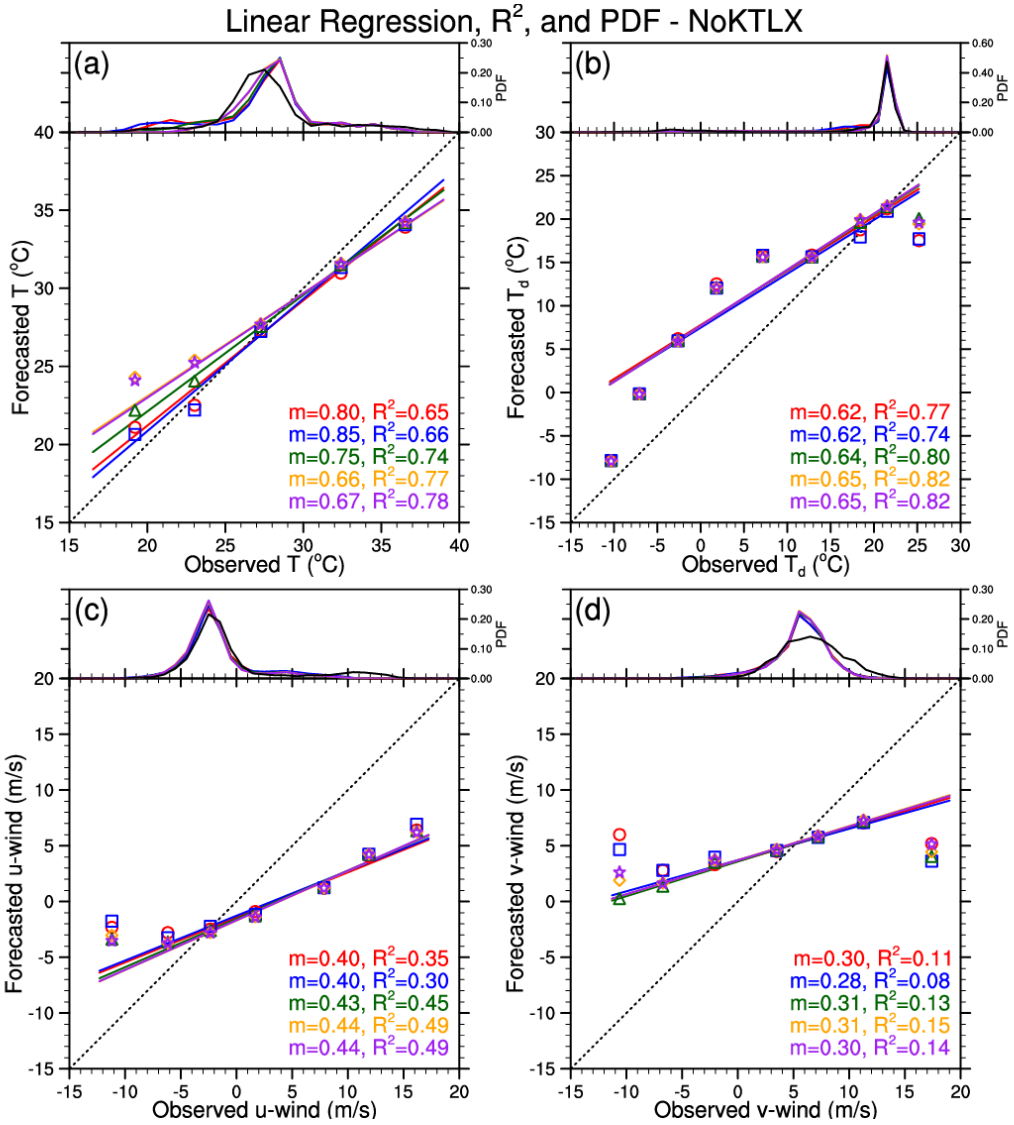

Figure 26. Same as for Figure 12 in Chapter 4, but for NoKTLX.

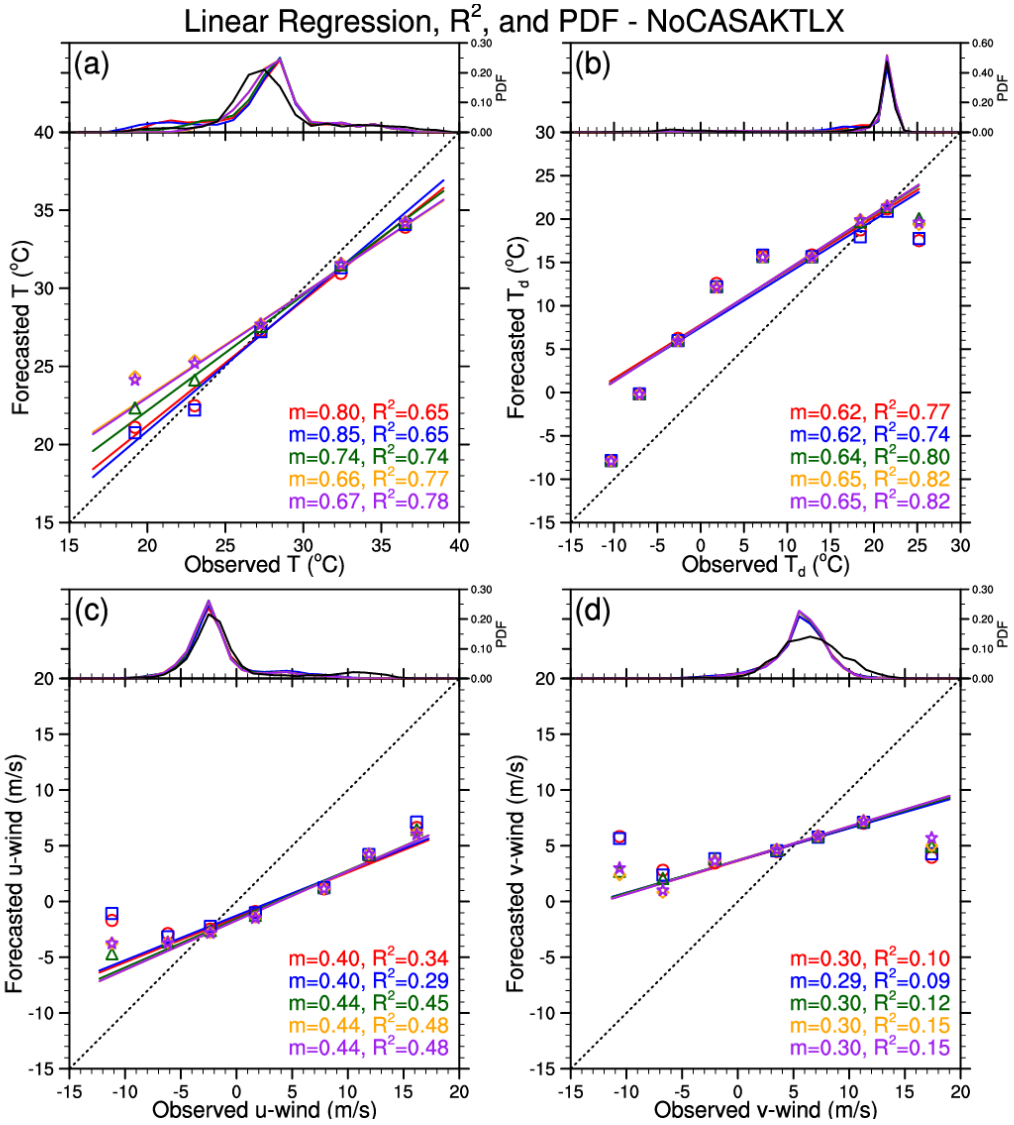

Figure 27. Same as for Figure 12 in Chapter 4, but for NoCASAKTLX.

5.4 Neighborhood-Based Results

Withholding CASA radar data from the data assimilation process has negligible impact on the skill of the simulations' composite reflectivity field (Figure 28). However for NoKTLX, all of the microphysics schemes experience at least a slight improvement in overall skill for the 30- and 40-dBZ thresholds, with MYSM exhibiting the largest improvements at all three thresholds (Figure 29). These improvements can be mostly explained by a decrease in overforecasting convection due to initially assimilating less convection owing to the withholding of KTLX radar data from the data assimilation process. Also with NoKTLX, LIN3 and WSM6 flip from slightly overforecasting to slightly underforecasting 50+ dBZ simulated reflectivity. This fact has little effect on LIN3's results at the 50-dBZ threshold, but WSM6 is noticeably worse at scales > 17 km. The results for NoCASAKTLX are very similar to the results for NoKTLX (Figure 30).

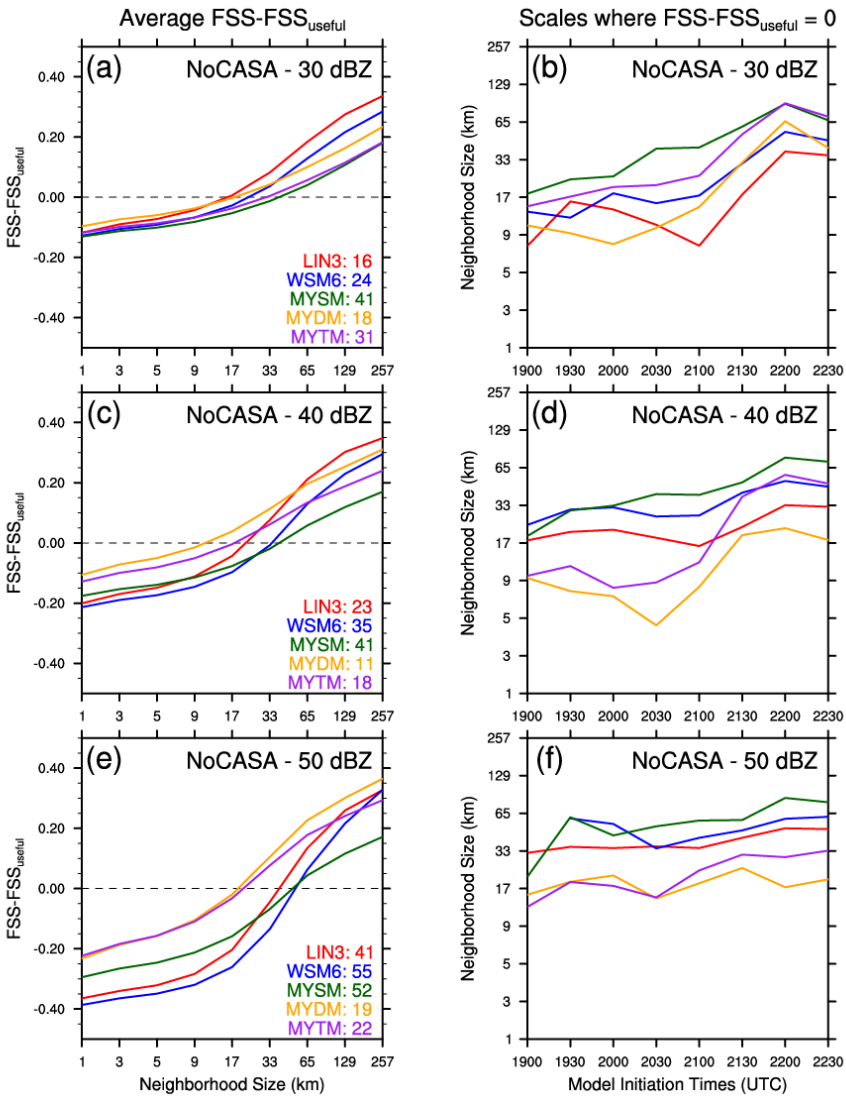

Figure 28. Same as for Figure 15 in Chapter 4, but for NoCASA.

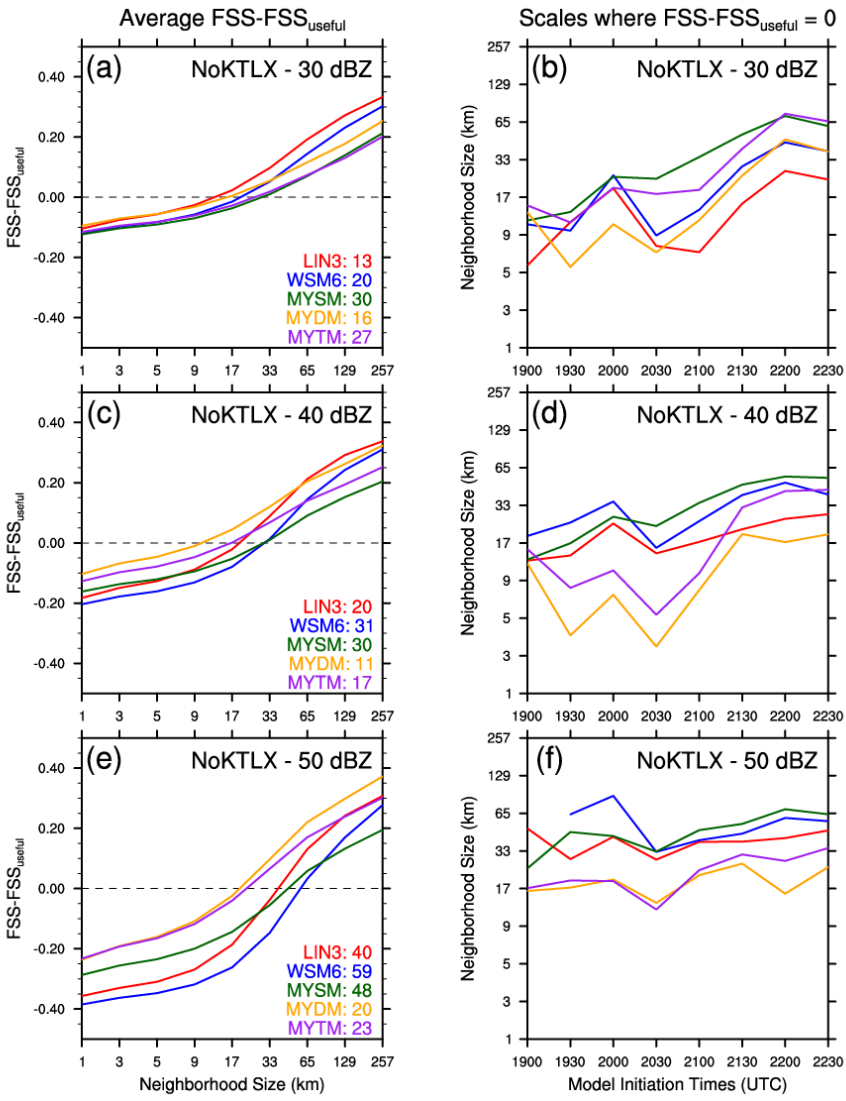

Figure 29. Same as for Figure 15 in Chapter 4, but for NoKTLX.

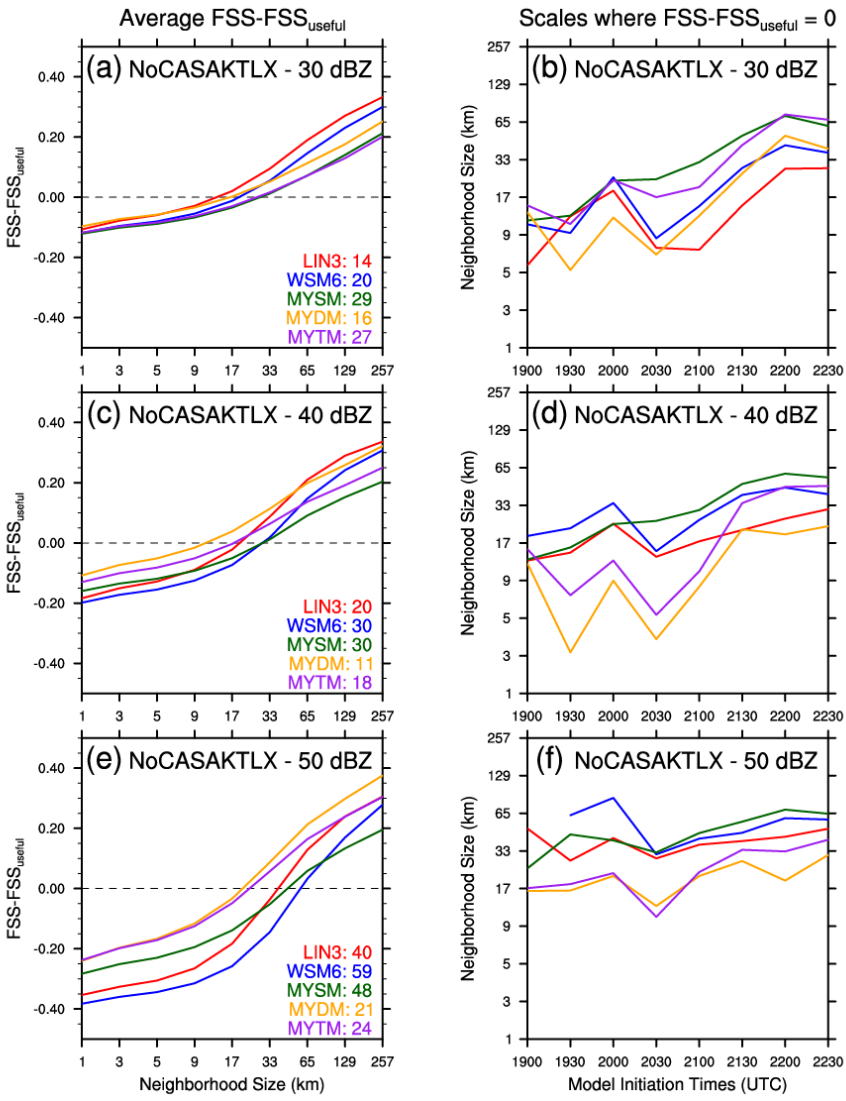

Figure 30. Same as for Figure 15 in Chapter 4, but for NoCASAKTLX.

5.5 Object-Based Results

Compared to the Control runs, the NoCASA simulations produce similar (WSM6 and MYDM) or slightly fewer (~35–55 for LIN3, MYSM, and MYTM) 0-1UH centers, and there are no substantial differences in the average max 0-1UH values (Figure 31). When KTLX radar data are withheld from the data assimilation process, all of the microphysics schemes produce ~100 fewer 0-1UH centers, and the average max 0-1UH values are slightly weaker, especially for the MY schemes, than for the Control runs (Figure 32). The simulations initialized early in the afternoon (1930 UTC – 2030 UTC) experience the largest decreases in the number of 0-1UH centers, and for the MY schemes, the 1900-UTC simulations produce weaker 0-1UH centers. Both of these findings could be due to insufficient assimilation of pre-mature storms (i.e., weakly rotating, shallower storms) from the withholding of KTLX radar data. Similar to the NoKTLX simulations, the NoCASAKTLX simulations generally result in fewer 0-1UH centers from 48 for WSM6 to 117 for MYTM (Figure 33). MYDM and MYTM experience the largest decrease in average max 0-1UH values, but the MY schemes' 0-1UH centers are still more than 20 m^2/s^2 stronger than LIN3 and WSM6's 0-1UH centers.

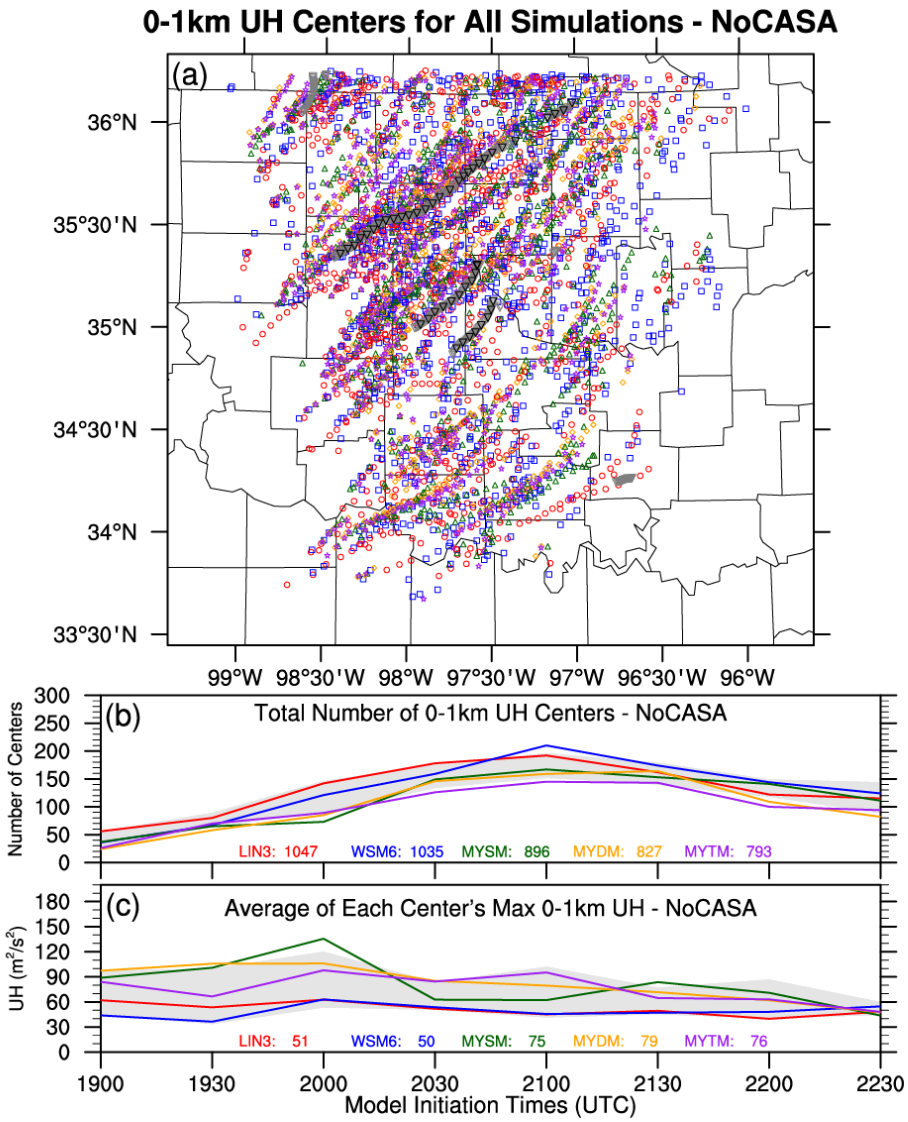

Figure 31. Same as for Figure 17 in Chapter 4, but for NoCASA. Light gray shading in (b) and (c) represents the max/min bounds of the Control runs for comparison purposes.

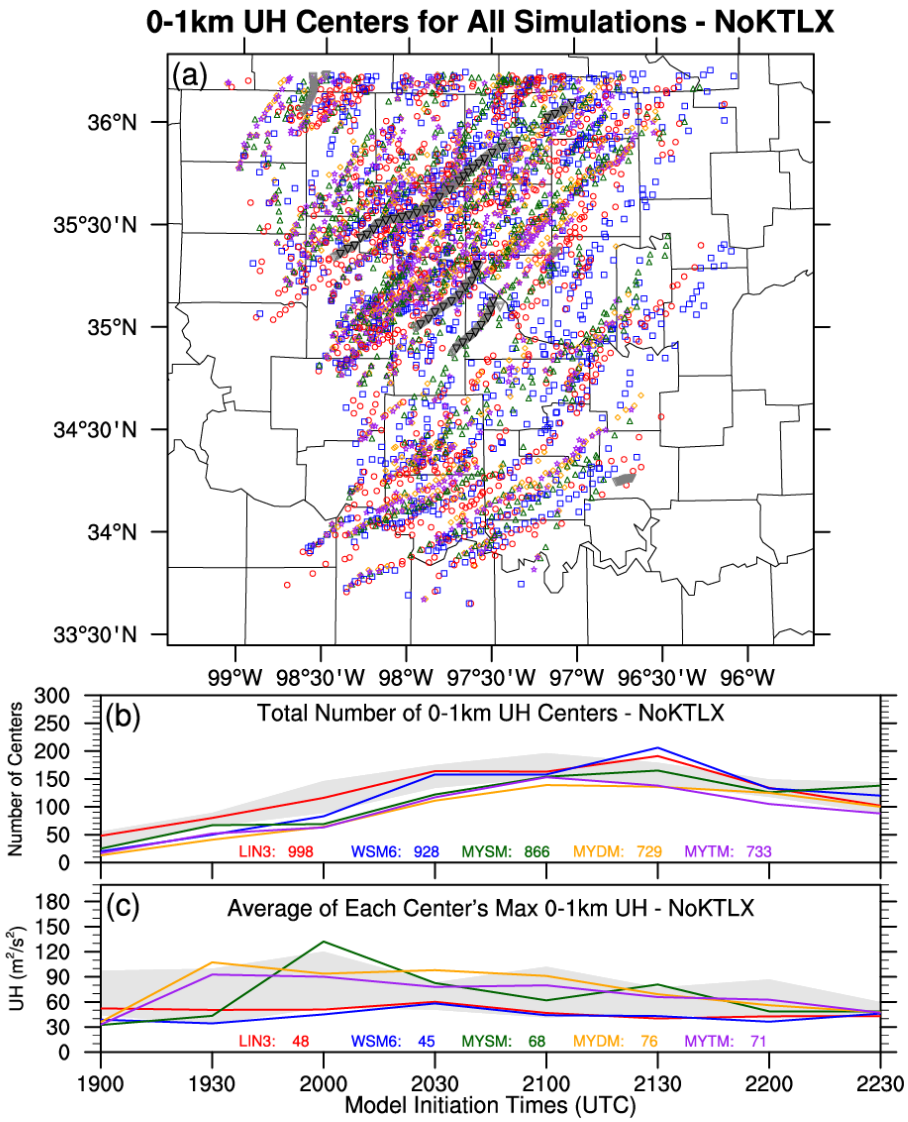

Figure 32. Same as for Figure 17 in Chapter 4, but for NoKTLX. Light gray shading in (b) and (c) represents the max/min bounds of the Control runs for comparison purposes.

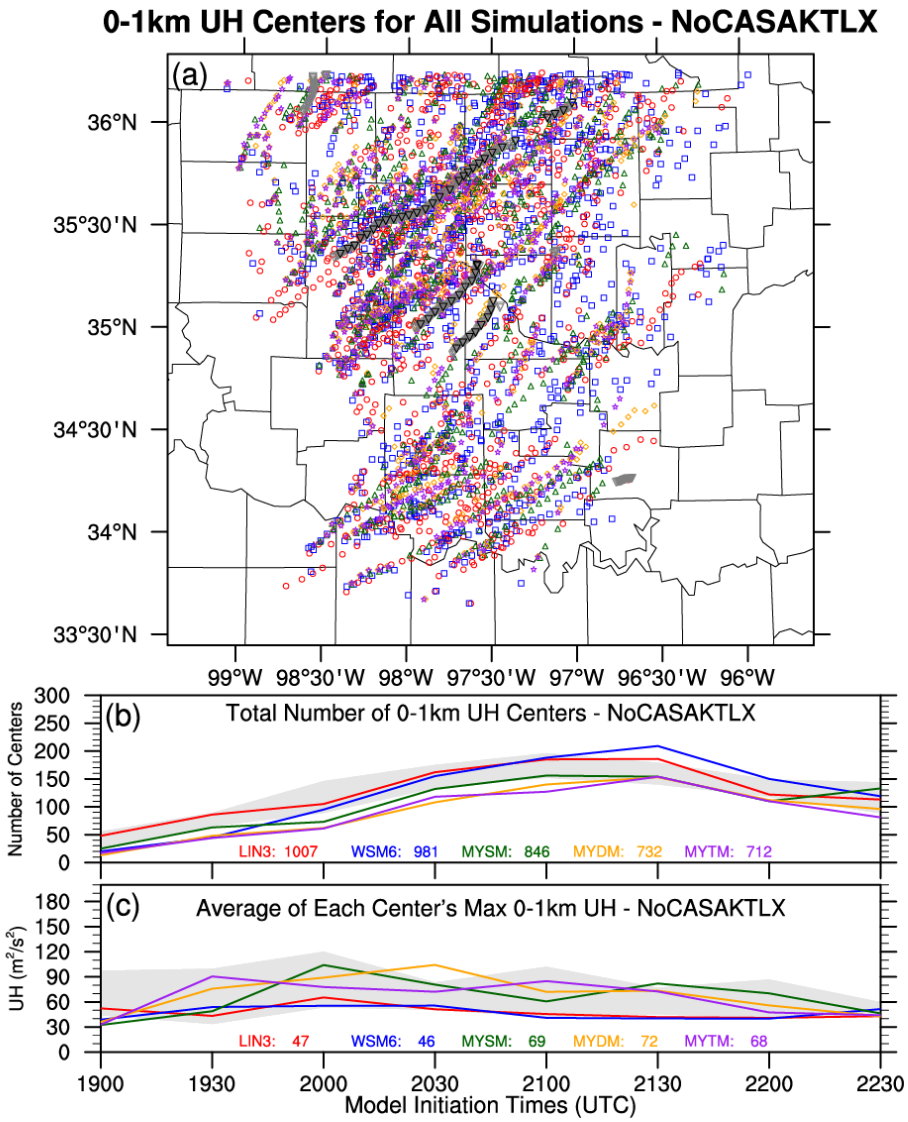

Figure 33. Same as for Figure 17 in Chapter 4, but for NoCASAKTLX. Light gray shading in (b) and (c) represents the max/min bounds of the Control runs for comparison purposes.

5.5.1 Storm 1

Compared to the Control runs, the NoCASA simulations produce similar spatial distributions of ST and AT 0-1UH centers for S1 (Figure 34a,b). Hence, the differences in ST and AT distance and timing errors between the Control and NoCASA runs are relatively minor (Figure 34c,d,e). This result is not surprising since S1 developed and stayed well outside of the CASA radar network. Conversely, the NoKTLX simulations yield fewer ST and AT 0-1UH centers near S1, and for all microphysics schemes, the average ST centers are at least slightly further away from S1's tornado points (Figure 35a,b). Generally, all of the schemes' NoKTLX simulations, especially MYDM and MYTM, result in larger ST and AT distance and timing errors, which are more apparent for the 1900-UTC and 2230-UTC runs (Figure 35c,d,e). In fact, the NoKTLX 1900-UTC simulation using MYDM produces no 0-1UH centers near S1. This result is in stark contrast to the success of the Control runs using the MY schemes in predicting S1. Similar to the NoKTLX runs, the MY schemes' NoCASAKTLX runs experience relatively large increases in ST and AT distance and timing errors compared to the Control runs (Figure 36c,d,e). Conversely, WSM6's NoCASAKTLX simulations result in slightly more 0-1UH centers near S1 and slightly smaller ST and AT distance and timing errors (Figure 36). However, even with larger errors, the MY schemes still generally performed better than LIN3 and WSM6.

Overall for S1, the NoCASA runs performed similar to the Control runs. The lack of impact of assimilating CASA radar data on storms outside the CASA radar network is apparent and not surprising. Even though the NoKTLX and NoCASAKTLX

runs generally performed worse than the Control runs, their forecasts still had good success at forecasting S1.

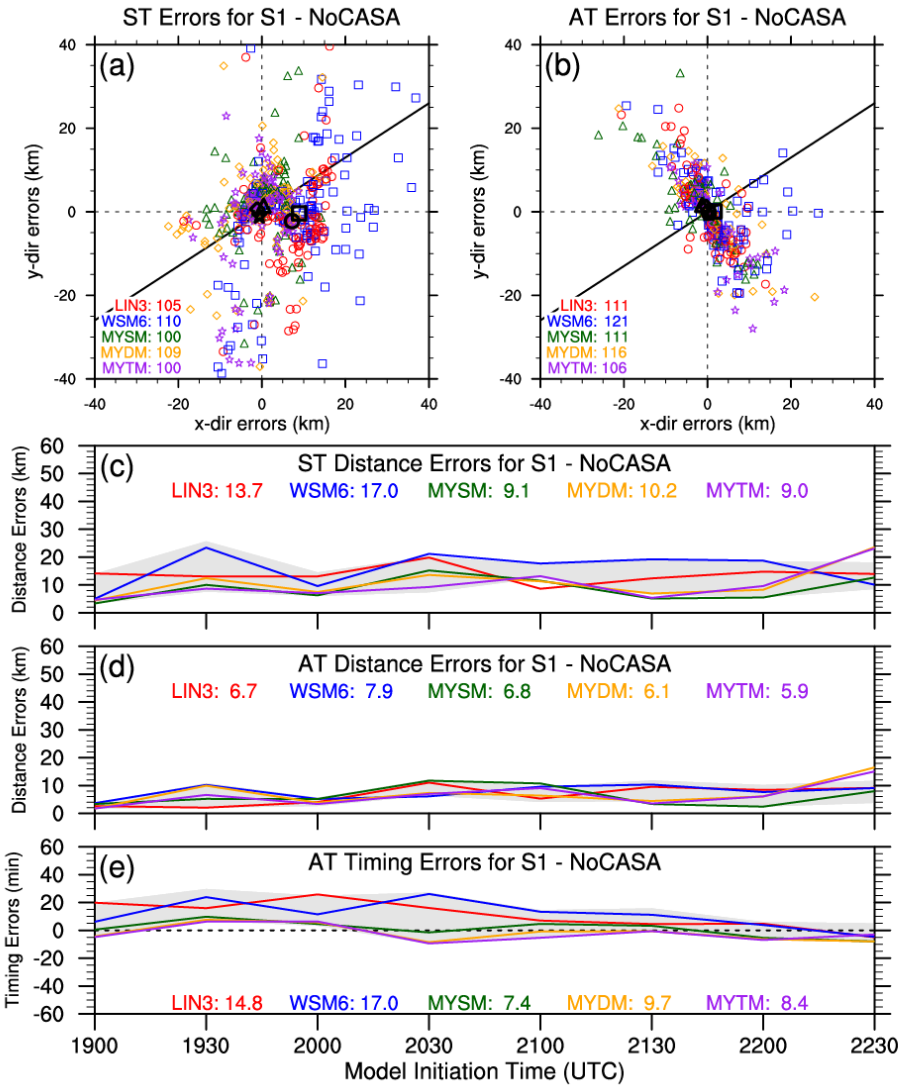

Figure 34. Same as for Figure 18 in Chapter 4, but for NoCASA. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

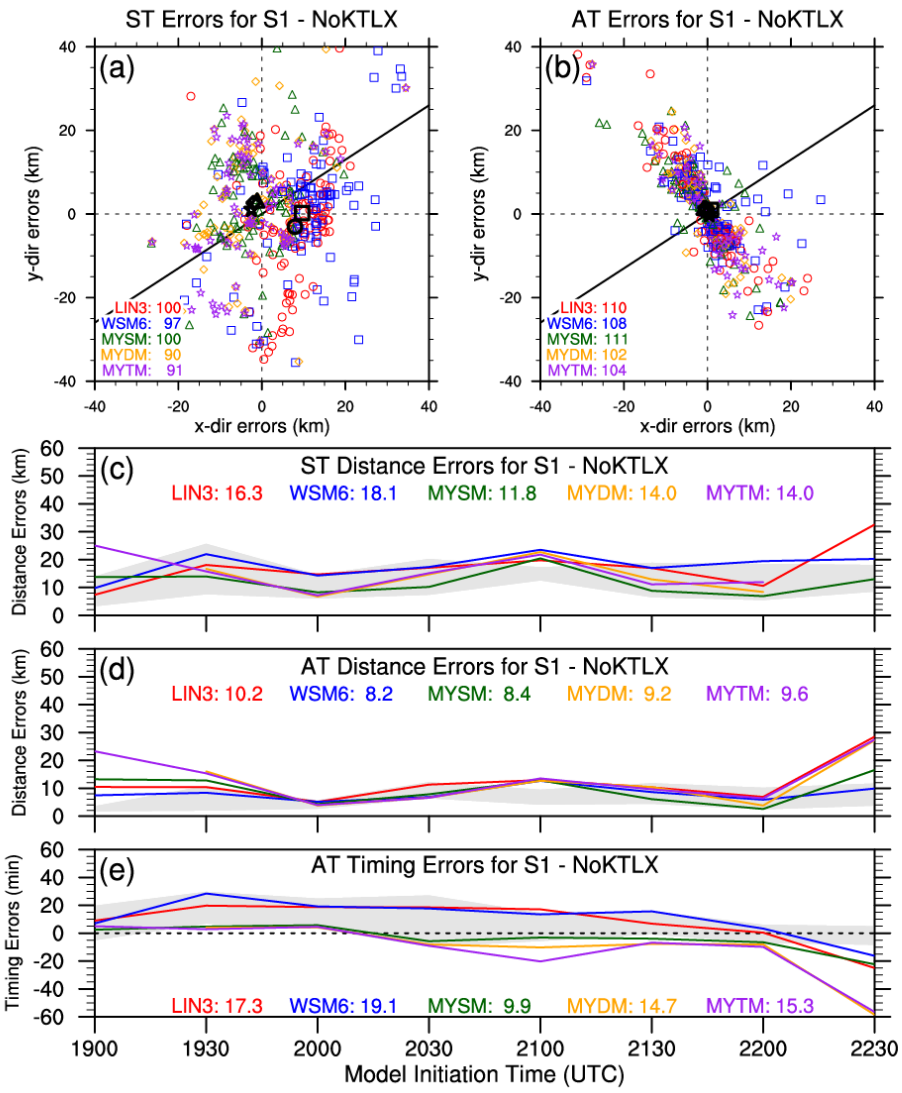

Figure 35. Same as for Figure 18 in Chapter 4, but for NoKTLX. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

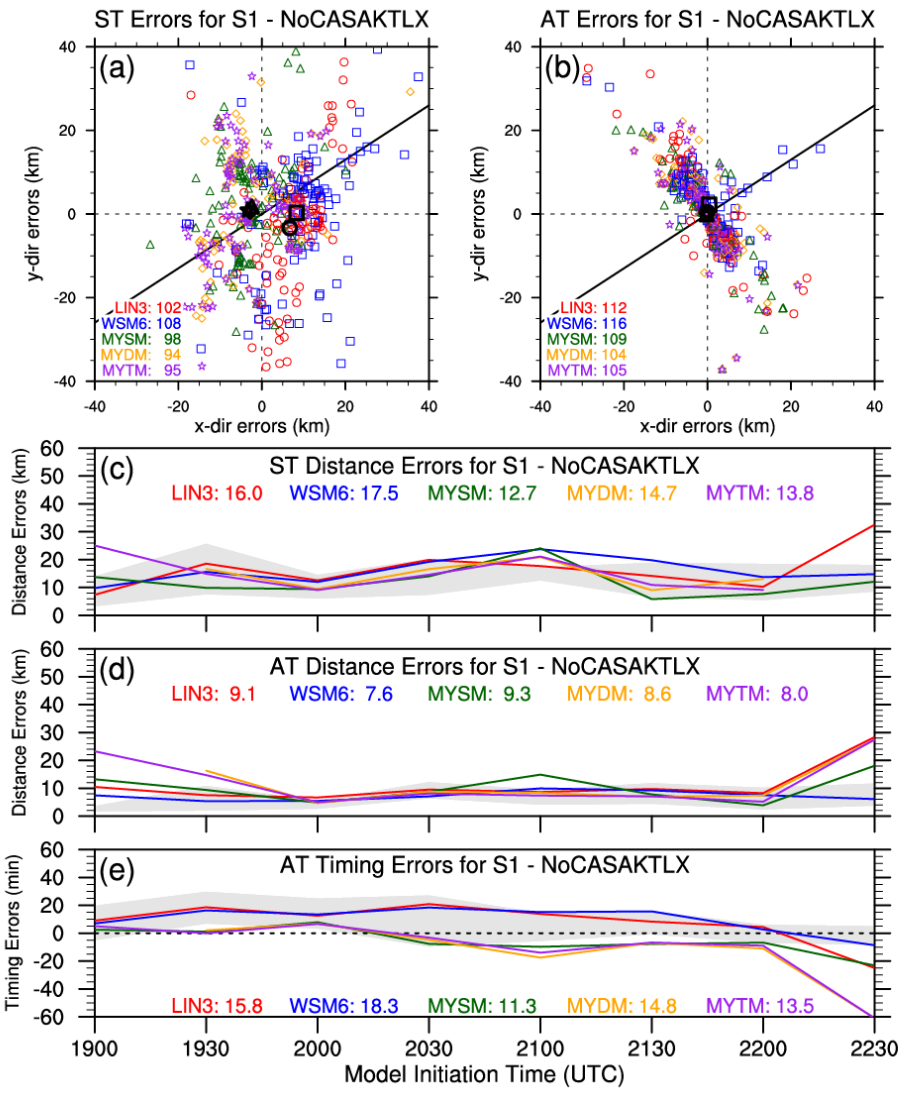

Figure 36. Same as for Figure 18 in Chapter 4, but for NoCASAKTLX. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

5.5.2 Storm 2

Unlike S1, S2 developed and propagated through the CASA radar network, so the impact of assimilating CASA radar data should be more evident. While the number of 0-1UH centers from the NoCASA simulations are similar (or a little less than) to what the Control runs produced, the cluster of 0-1UH centers north of S2's tornado points is more spaced out for both ST and AT (Figure 37a,b). This translates to larger ST and AT distance and timing errors for the MY schemes, but the error changes are mixed for LIN3 and WSM6, which also experiences a substantial increase in the timing error (Figure 37c,d,e). It should be noted that the relative impact of one 0-1UH center on the different errors is larger for S2 than for S1 due to S2 having ~70% fewer centers nearby.

Similar to the NoCASA runs, the NoKTLX runs produce similar numbers of 0-1UH centers as the Control runs (Figure 38a,b). The ST and AT distance and timing errors for NoKTLX's MYDM and MYTM are larger than they are for the NoCASA runs, but the results are mixed for the single-moment schemes (Figure 38c,d,e). For example, LIN3 has larger ST distance errors and smaller AT distance and timing errors compared to the Control runs. The clustering and number of 0-1UH centers for the NoCASAKTLX runs is somewhat similar to both the NoCASA and NoKTLX runs, but in different respects (Figure 39a,b). For example, the average location of MYTM's 0-1UH centers is closer to NoCASA than NoKTLX, but the average location of LIN3's 0-1UH centers is closer to NoKTLX than NoCASA. The ST and AT distance and timing errors for LIN3 and WSM6 are more similar to their errors with the NoKTLX runs than the NoCASA runs, but MYSM's errors are more similar to its errors with the NoCASA

runs than the NoKTLX runs (Figure 39c,d,e). The AT and ST distance and timing errors for MYDM and MYTM are similar and or greater than the corresponding errors for the NoKTLX and NoCASA runs.

Summarizing for S2, the simulations generally perform worse when CASA and/or KTLX radar data are withheld during the data assimilation process. This is especially true for the MY schemes. These results indicate that withholding CASA radar data during the data assimilation process has a similar impact to withholding KTLX radar data, so while the ability of the CASA radar data to provide gap-filling benefits to 1-km forecasts is still unclear, the results are clear that assimilating CASA radar data has a similar impact to assimilating KTLX radar data for rotating storms developing and propagating within the CASA radar network.

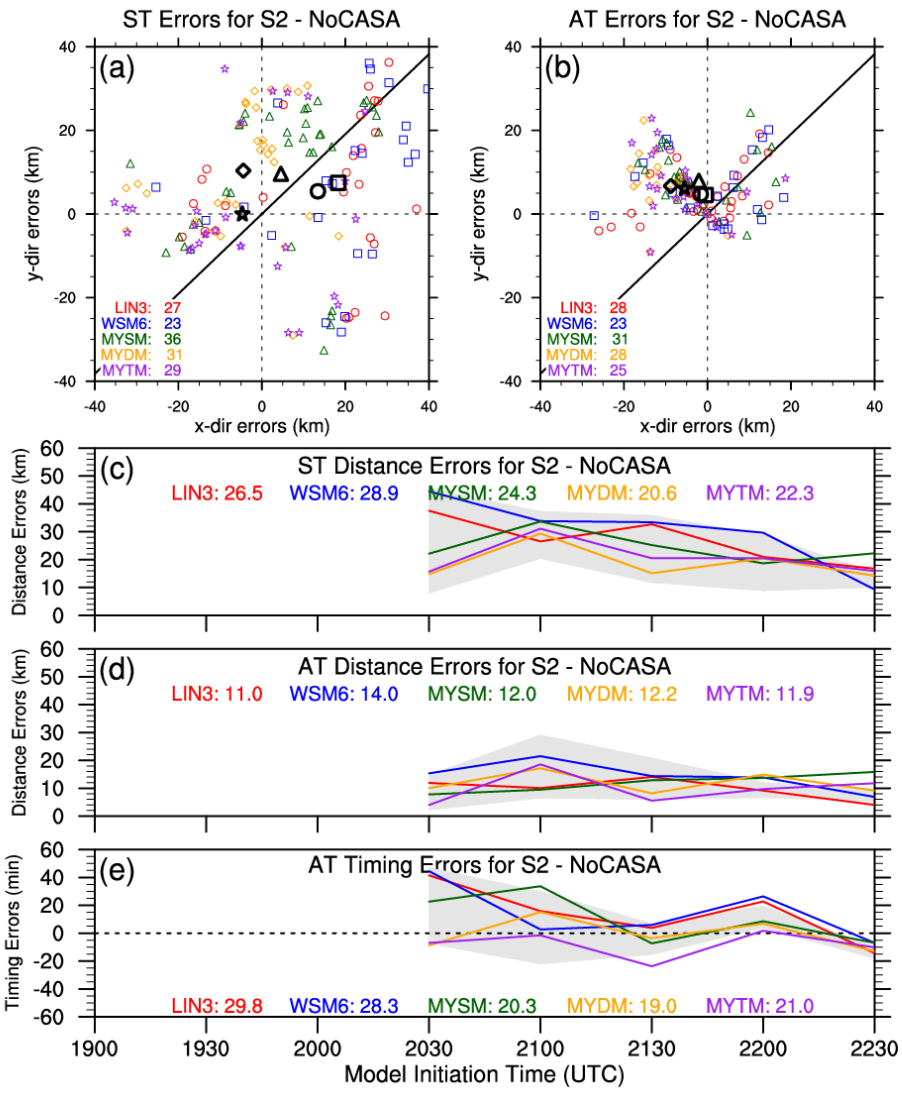

Figure 37. Same as for Figure 20 in Chapter 4, but for NoCASA. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

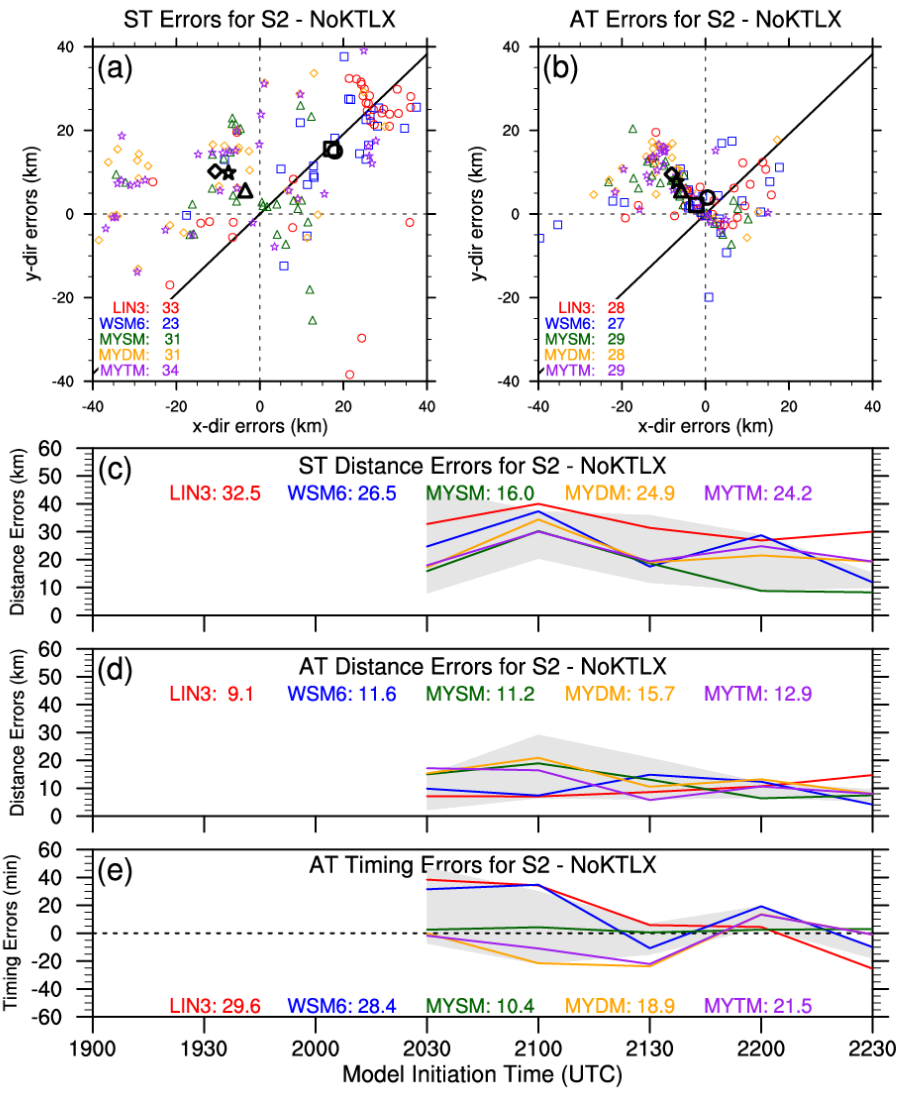

Figure 38. Same as for Figure 20 in Chapter 4, but for NoKTLX. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

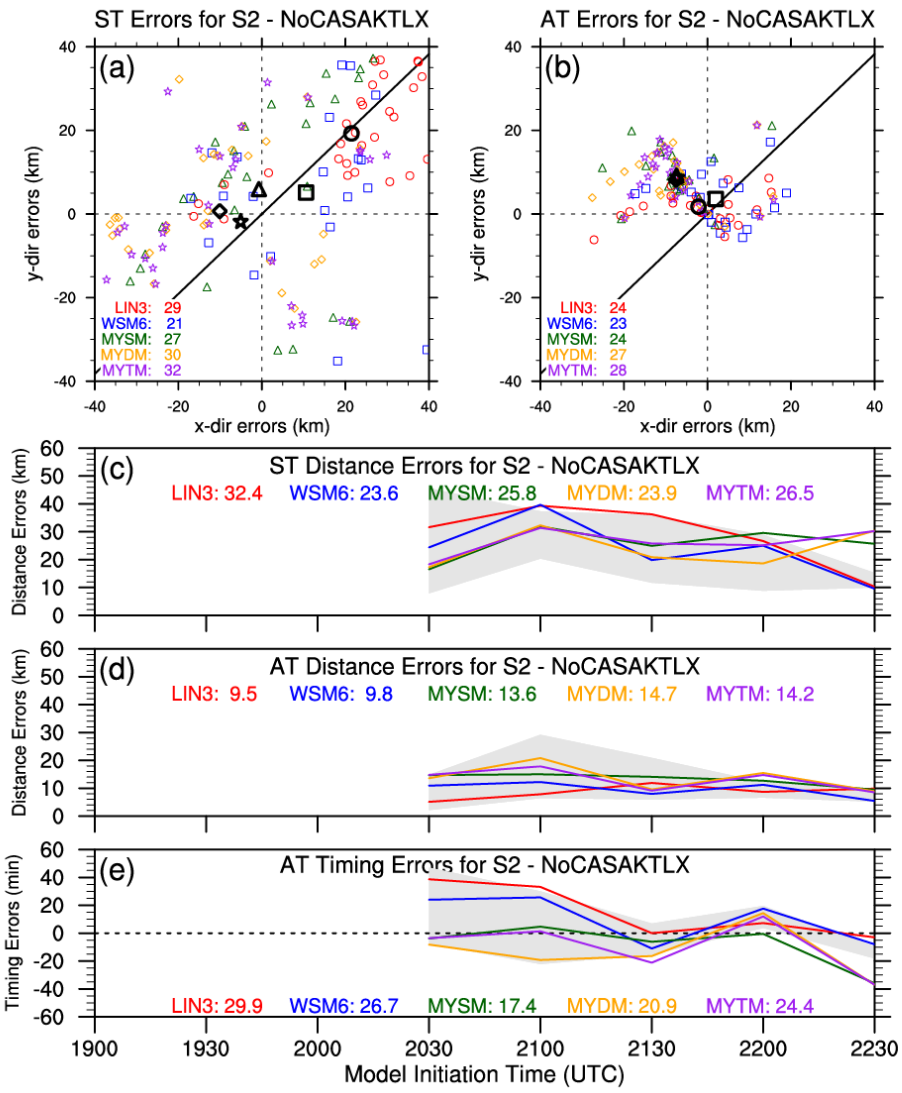

Figure 39. Same as for Figure 20 in Chapter 4, but for NoCASAKTLX. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

5.5.3 Storm 3

As depicted in Chapter 4, S3 is not well forecasted by the Control runs. When CASA radar data are withheld from the data assimilation process, similar forecasts of 0-1UH centers are produced with very few centers near S3's tornado points and a spray of centers ~30 km to the southeast (Figure 40a,b). The ST and AT distance and timing errors for NoCASA have mixed results as compared to the Control runs, but MYSM and MYDM show reduced errors for all three error types (Figure 40c,d,e). However, as mentioned in Section 5.5.2, a single 0-1UH center can have a substantial impact on the average errors due to the small number of centers (i.e., < 20). Interestingly, when KTLX radar data are withheld from the data assimilation process, a small cluster of 0-1UH centers appears near S3's tornado points (Figure 41a,b). This result contributes to mostly smaller ST and AT distance and timing errors for all of the microphysics schemes, especially the MY schemes (Figure 41c,d,e). The NoCASAKTLX simulations exhibit similar results to the NoKTLX runs (Figure 42).

Overall for S3, the NoCASA, NoKTLX, and NoCASAKTLX runs struggled to successfully forecast S3 as in the Control runs. The small cluster of 0-1UH centers near S3 in the NoKTLX runs might indicate that the KTLX radar data are hampering the CASA radar data during the data assimilation process, but this needs more investigation in order to conclusively determine the cause. What is conclusive, though, is the need to explore other data assimilation techniques to potentially improve the forecasts of S3. The next chapter will do exactly that.

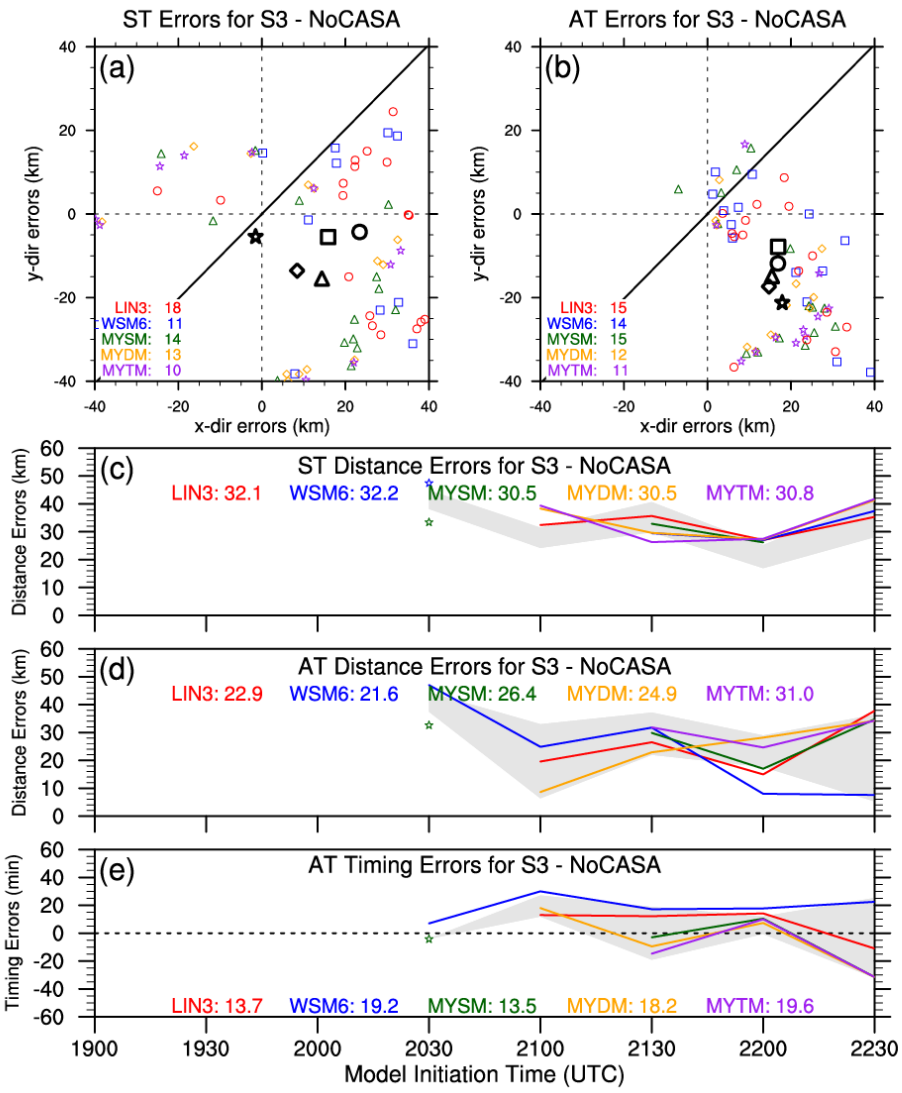

Figure 40. Same as for Figure 21 in Chapter 4, but for NoCASA. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

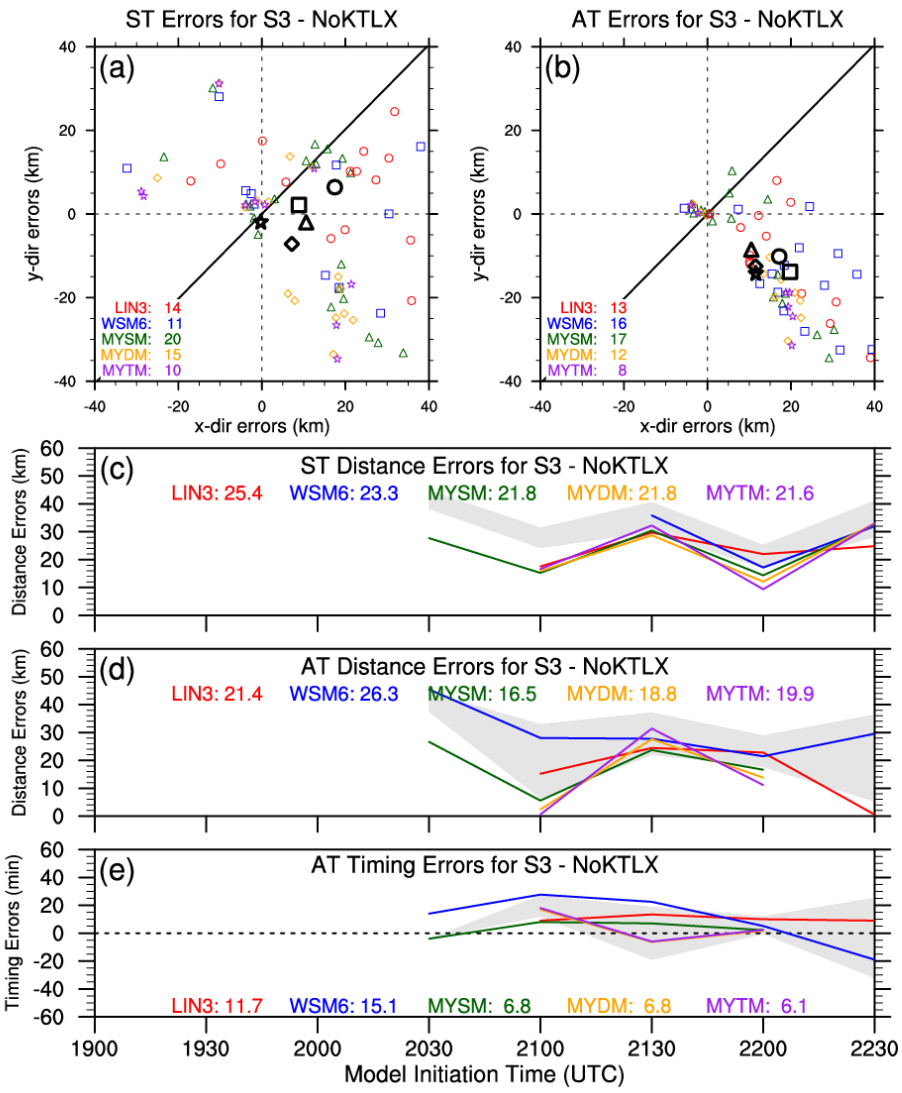

Figure 41. Same as for Figure 21 in Chapter 4, but for NoKTLX. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

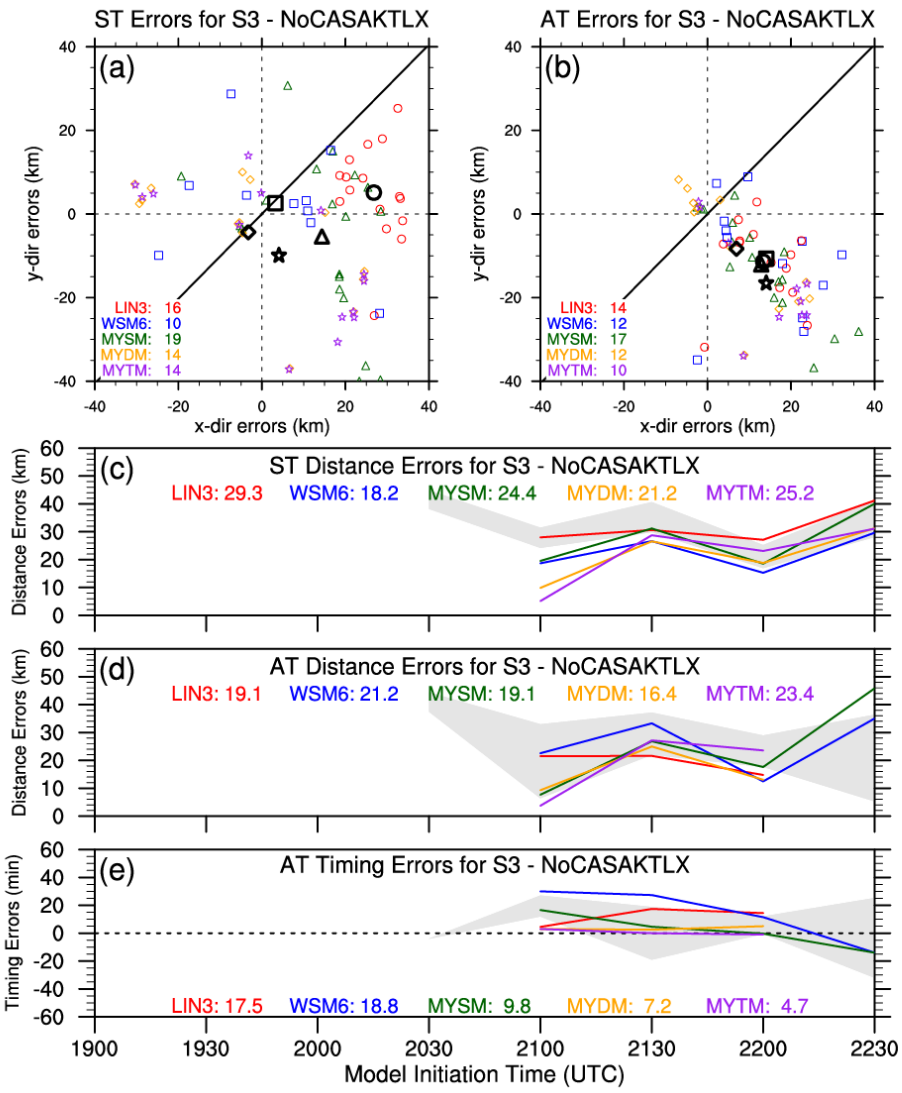

Figure 42. Same as for Figure 21 in Chapter 4, but for NoCASAKTLX. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

Chapter 6 – Impact of Other Data Assimilation Techniques

6.1 Introduction and Background

A primary goal of various data assimilation techniques is to produce accurate analyses by minimizing deviations from observations while also minimizing imbalances that can lead to, for example, spurious gravity waves, within the initial model state. In Chapter 1, a brief overview of the 3DVAR (Gao et al., 2004) and EnKF (Evensen, 1994) data assimilation systems was discussed, and a few benefits can be attributed to each system. However, additional data assimilation techniques within each data assimilation system may further increase forecast skill through improved analyses. One such technique, as discussed and used before, is IAU (Bloom et al., 1996). Another data assimilation method somewhat similar to the IAU technique is Newtonian relaxation, or nudging, which provides an additional way to foster a dynamically-balanced model state by making use of the observational innovation (Hoke and Anthes 1976). Both methods act to minimize the "shock" to the model state by gradually updating the analysis during the data assimilation process.

Brewster et al. (2015) and Brewster and Stratman (2016) expand on the original IAU method by allowing for variable-dependent, temporally-weighted distributions of the increments of the model state variables. Both studies determine that introducing the majority of the hydrometeor increments after the temperature, water vapor, and wind increments results in stronger vertical velocities and preserves larger graupel and hail mixing ratios. The CAPS' real-time 1-km forecast system for the Dallas-Fort Worth Urban Testbed utilizes the modified IAU method along with another data assimilation technique, analysis cycling. The majority of storm-scale NWP studies use analysis

cycling to bring in observations from multiple time periods during the data assimilation process. While some forecast systems use continuously cycled analysis systems (e.g., Sobash and Stensrud, 2015), the CAPS' real-time 1-km forecast system uses two assimilation cycles with 10-min windows. Also, as described in Brewster and Stratman (2016), each assimilation cycle employs a different set of increment distributions. In this chapter, the analysis cycling and variable-dependent IAU techniques will be examined using the various forecast verification techniques.

6.2 Experiment Design

For this set of numerical experiments, the same model configurations as in Chapter 4 are used with the same microphysically-diverse set of simulations. To examine the potential benefits of cycling, a set of simulations is initialized from analyses created from two 5-min analysis cycles (Cycling) using the same IAU method as the Control runs (Figure 43a). The first analysis cycle begins 10 min prior to the start of the simulations (e.g., 2050 UTC for simulations starting at 2100 UTC). To explore the potential benefits of varying the temporal distributions for each state variable's increments, a set of simulations is initialized from the modified IAU (ModIAU) using the temporally-weighted distributions in Figure 43b, where the hydrometeor fields' increments are introduced during the second half of the IAU window. Unlike the IAU process used in the experiments in Chapters 4 and 5, where increments were introduced every 20 s, the modified IAU process used in these experiments introduces increments every 26 s to better accommodate the temporally-weighted distributions. Finally, as in Brewster and Stratman (2015), a set of simulations is initialized using both the analysis cycling and modified IAU techniques (CyModIAU). The first 5-min analysis cycle uses the same temporally-weighted distributions in the ModIAU runs, and the second 5-min analysis cycle uses a slightly different set of distributions (Figure 43c). The use of the analysis cycling and IAU techniques require very little additional computational cost, so even small improvements in forecast skill will be considered worthwhile for future implementation in real-time systems.

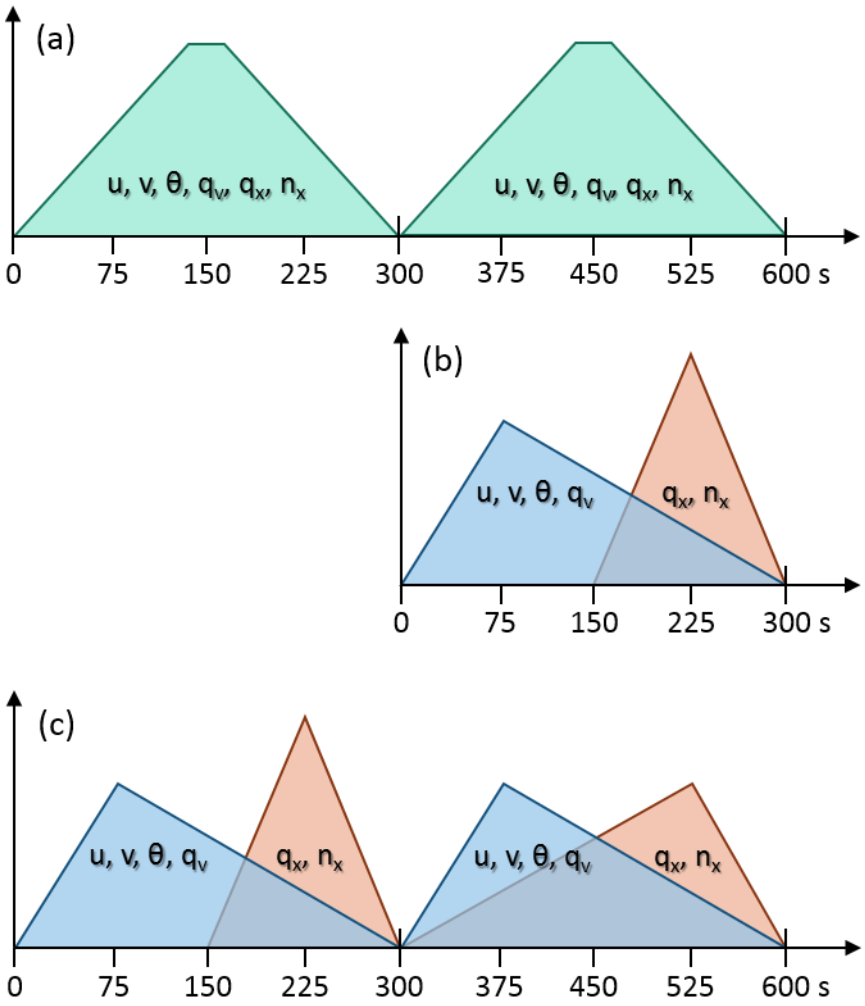

Figure 43. Schematic of the temporally-weighted distributions of increments for the (a) Cycling runs, (b) ModIAU runs, and (c) CyModIAU runs.

6.3 Observation Point-Based Results

For the Cycling runs, LIN3 and WSM6 mostly have larger RMSEs than the Control runs, especially for temperature (i.e., more than 0.25°C larger), but MYDM and MYTM have smaller RMSEs for all variables, except for pressure (Figure 44). Besides the MY schemes' RMSEs for temperature, the ModIAU runs generally experience smaller RMSEs than the Control runs (Figure 45). All of the microphysics schemes' CyModIAU runs result in larger temperature RMSEs than the Control runs, especially WSM6, but all of the schemes' CyModIAU runs result in substantially smaller RMSEs for dewpoint temperature, especially LIN3 and the MY schemes (Figure 46). Also, the MY schemes' RMSEs for u-wind and v-wind are substantially smaller.

For all microphysics schemes and near-surface variables, the Cycling runs have linear regression slopes closer to 1 than the Control runs (Figure 47). However, the R^2 values depict mixed results with smaller R^2 values for temperature and u-wind for the single-moment schemes and larger R^2 values for v-wind for the MY schemes. Also of note, all of the simulations appear to be colder in the Cycling runs than in the Control runs for temperatures less than 30°C, and the runs using LIN3 and WSM6 depict an increase in the number of forecasted near-surface temperatures between 18 and 21°C (Figure 47a). Interestingly, the PDFs for v-wind for all of the schemes is less narrow and closer to the observed v-wind PDF (Figure 47d). The ModIAU runs exhibit small differences from the Control runs for both slopes and R^2 values (Figure 48). Except for the MY schemes' slopes for temperature, the CyModIAU runs result in slopes closer to 1 for all variables, and the R^2 values are mostly larger for the MY schemes and mixed for LIN3 and WSM6 (Figure 49).

Overall, analysis cycling improves some aspects of the forecasts of the near-surface variables, and the modified IAU technique improves similar and other aspects of the same forecasts. The CyModIAU runs' forecasts of the near-surface variables resemble a combination of both the Cycling and ModIAU runs. Analysis cycling in both the Cycling and CyModIAU runs leads to colder and larger cold pools, especially for LIN3 and WSM6, likely due larger areas of convection and cloud cover, as will be shown in the Section 5.4. Additionally, both analysis cycling and the modified IAU contribute to better forecasts of the near-surface winds for the MY schemes in the CyModIAU runs.

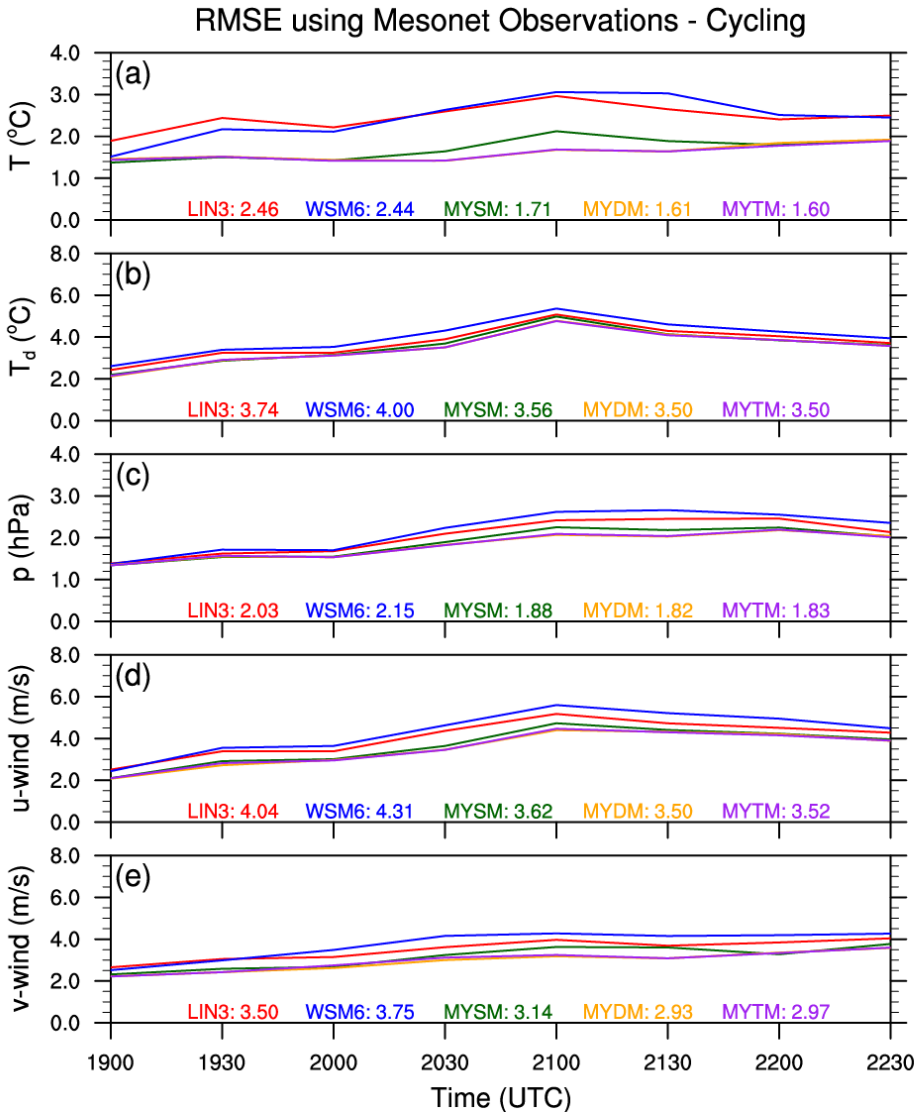

Figure 44. Same as for Figure 11 in Chapter 4, but for Cycling.

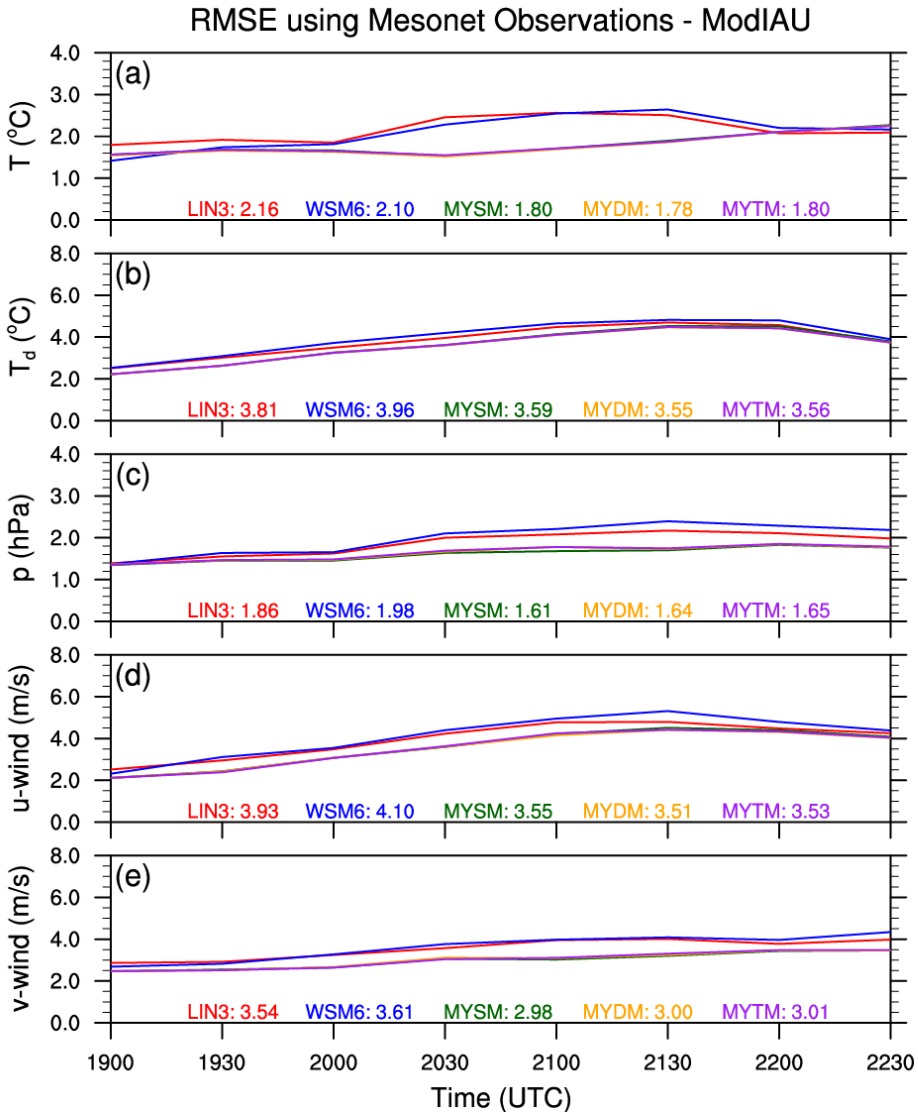

Figure 45. Same as for Figure 11 in Chapter 4, but for ModIAU.

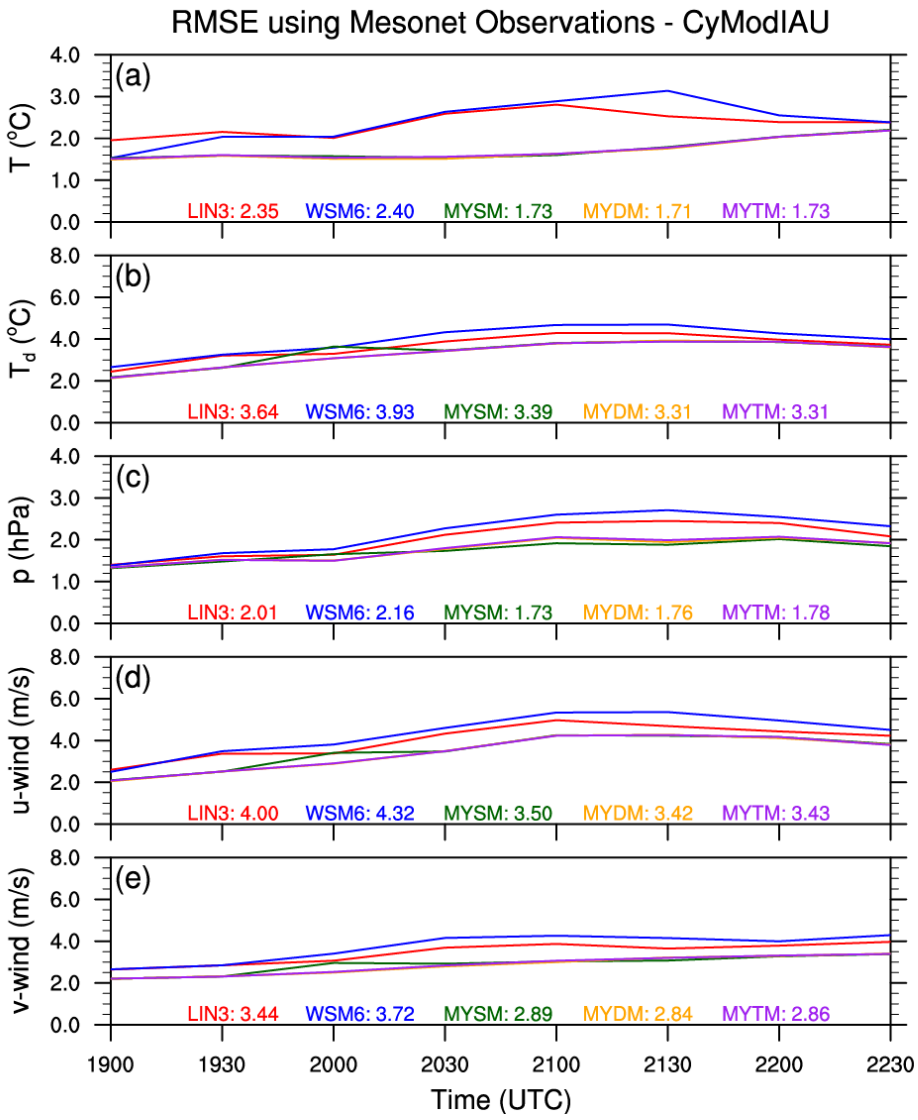

Figure 46. Same as for Figure 11 in Chapter 4, but for CyModIAU.

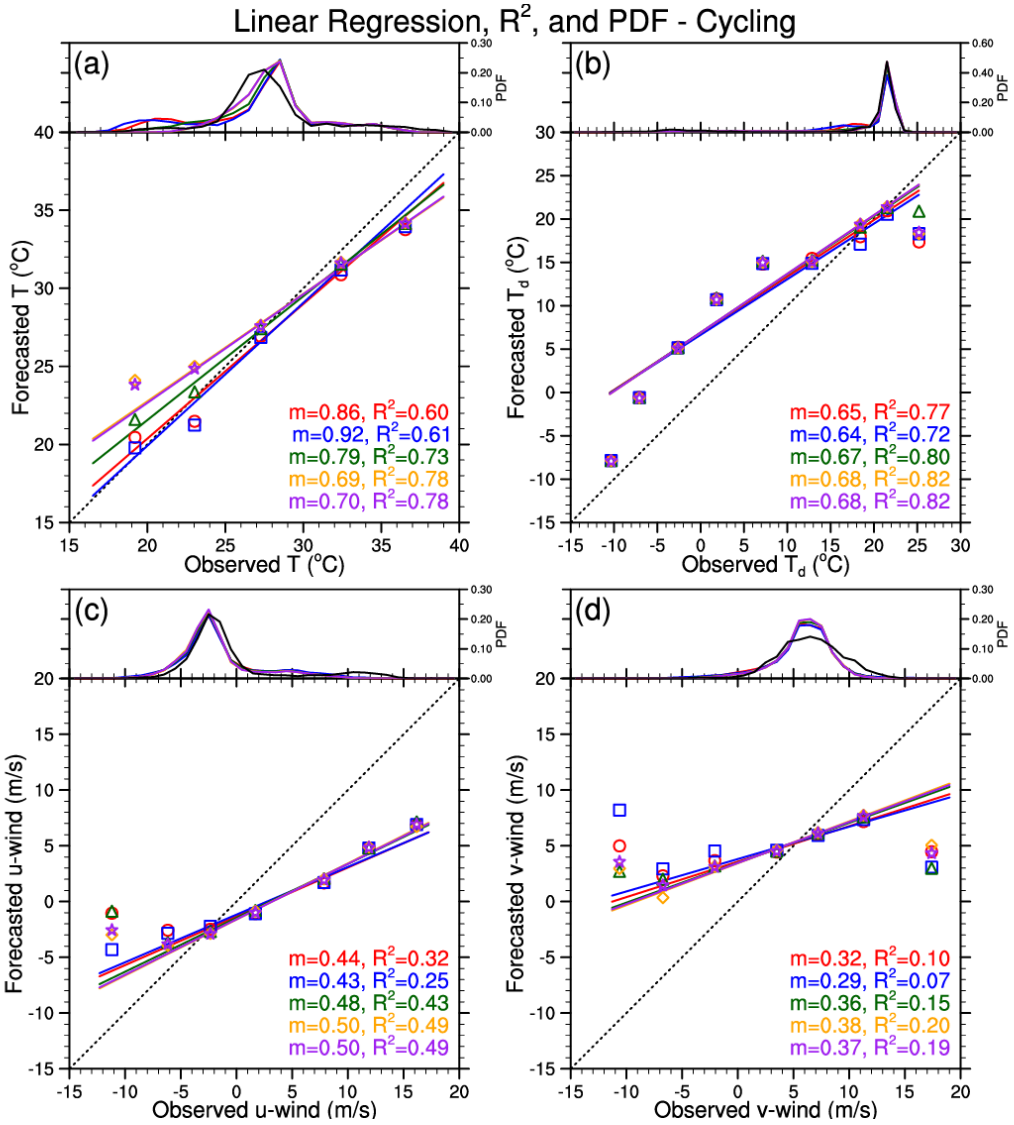

Figure 47. Same as for Figure 12 in Chapter 4, but for Cycling.

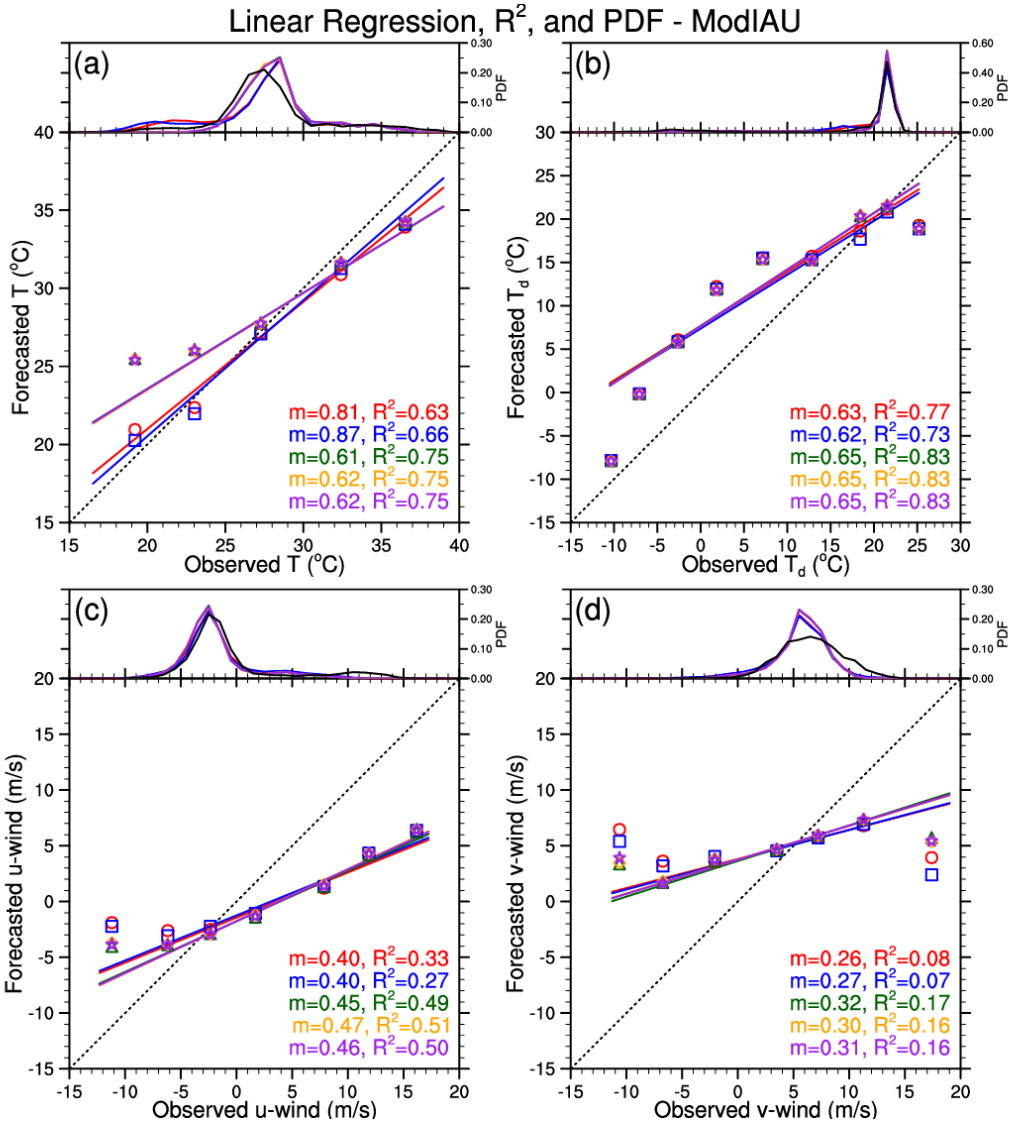

Figure 48. Same as for Figure 12 in Chapter 4, but for ModIAU.

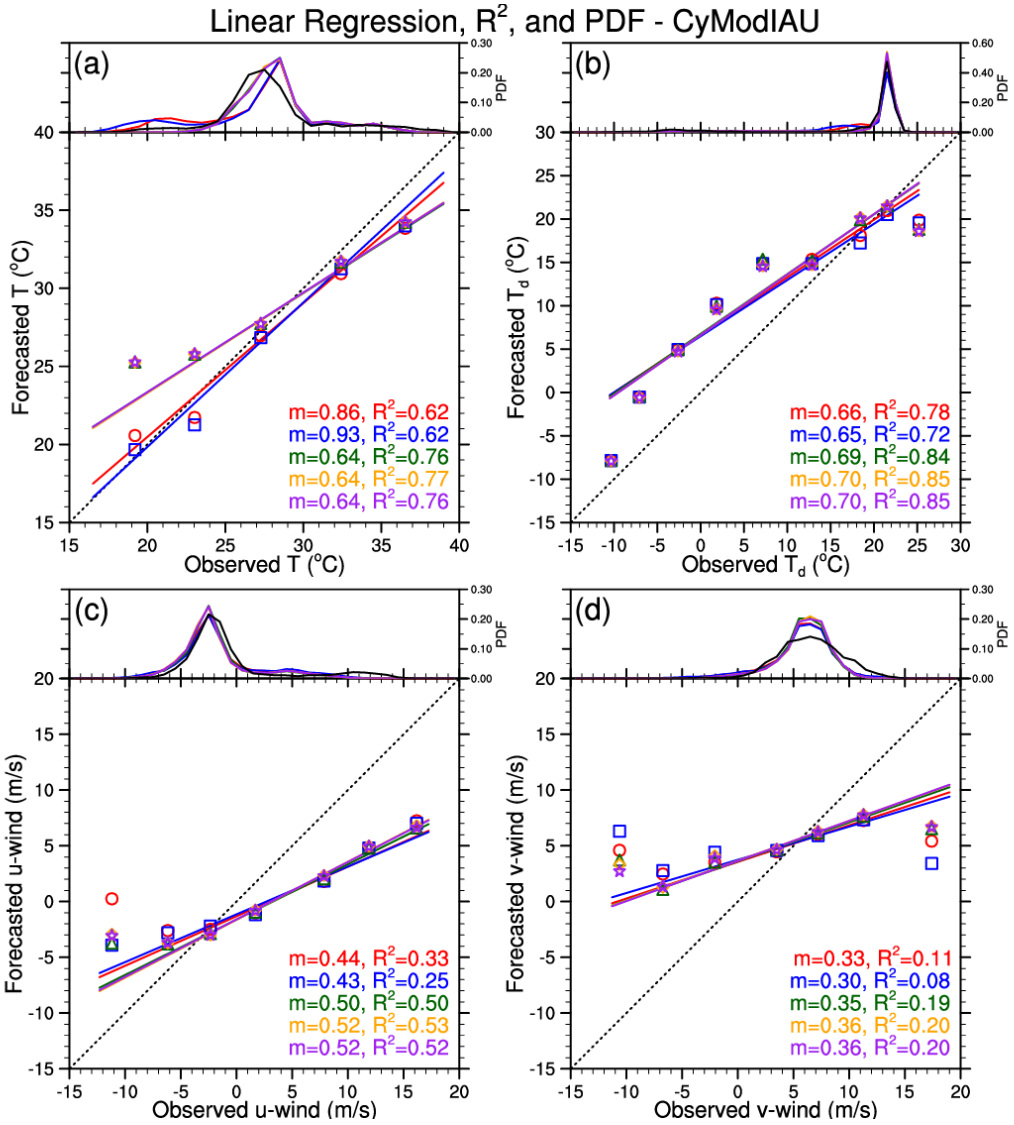

Figure 49. Same as for Figure 12 in Chapter 4, but for CyModIAU.

6.4 Neighborhood-Based Results

All of the Cycling runs experience a decrease in skill at all scales, especially those using MYSM (Figure 50). Not surprisingly, this loss in skill can likely be attributed to the increase in areal coverages of the simulated composite reflectivity. The results are mixed for the ModIAU runs (Figure 51). For example, the MYTM runs lost quite a bit of skill at the 30-dBZ threshold, but gained skill at the large scales at the 50-dBZ threshold. The spread across the spectrum of microphysics schemes is larger at the 30-dBZ threshold (Figure 51a), but the 40- and 50-dBZ threshold results are fairly similar to the Control runs' results (Figure 51b,c). Interestingly, the ModIAU runs using MYDM are the only simulations to exhibit a decrease in forecasted base rates at all three thresholds (not shown). Similar to the Cycling runs, the CyModIAU runs generally depict smaller FSSs at most scales (Figure 52). When the ModIAU runs perform better than the Control runs and the Cycling runs perform worse than the Control runs, the CyModIAU runs typically perform somewhere in between the ModIAU and Cycling runs. However, when both the ModIAU and Cycling runs perform worse than the Control runs, the CyModIAU generally perform worse than both the ModIAU and Cycling runs.

Overall, analysis cycling has the greatest negative impact on the FSS results for the Cycling and CyModIAU runs. This finding is likely tied to the increase in areal coverage of the simulated composite reflectivity, except for the CyModIAU runs using MYDM. Conversely to analysis cycling, the modified IAU technique has some beneficial impact for some of the microphysics schemes and reflectivity thresholds, but

as indicated before, these changes are largely related to the changes in the forecasted base rates.

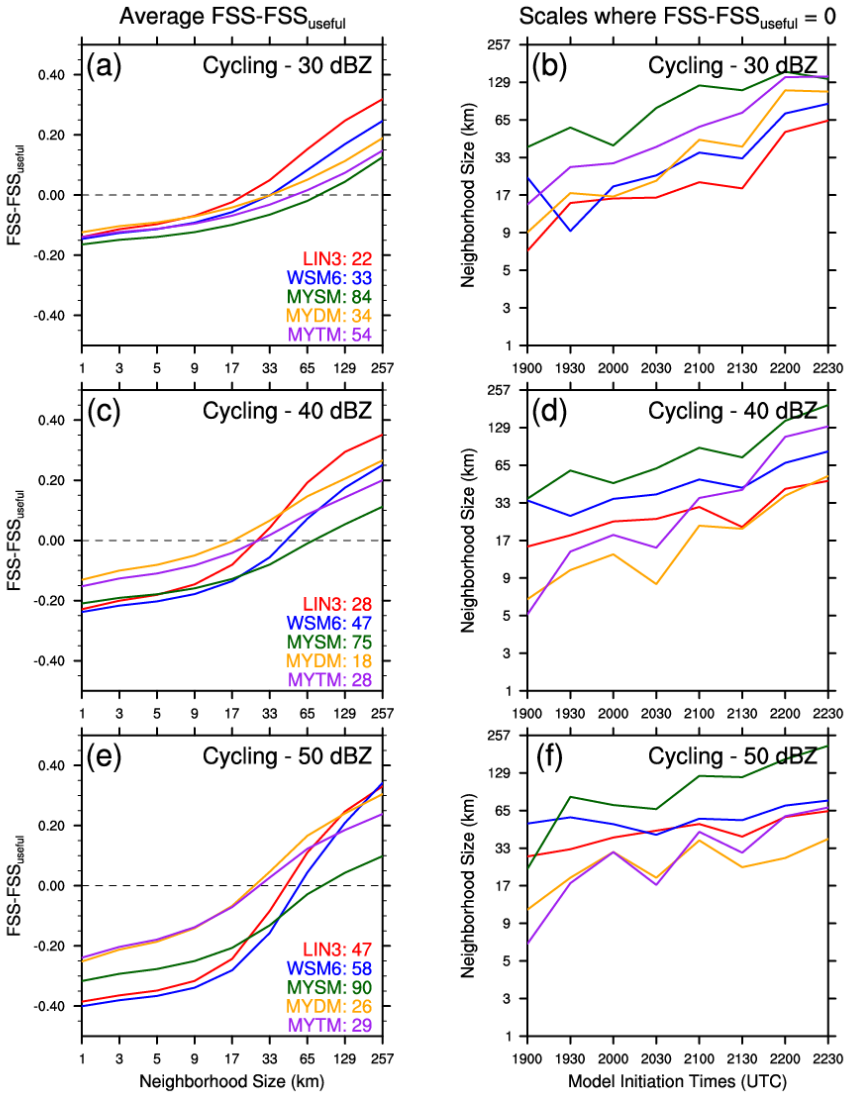

Figure 50. Same as for Figure 15 in Chapter 4, but for Cycling.

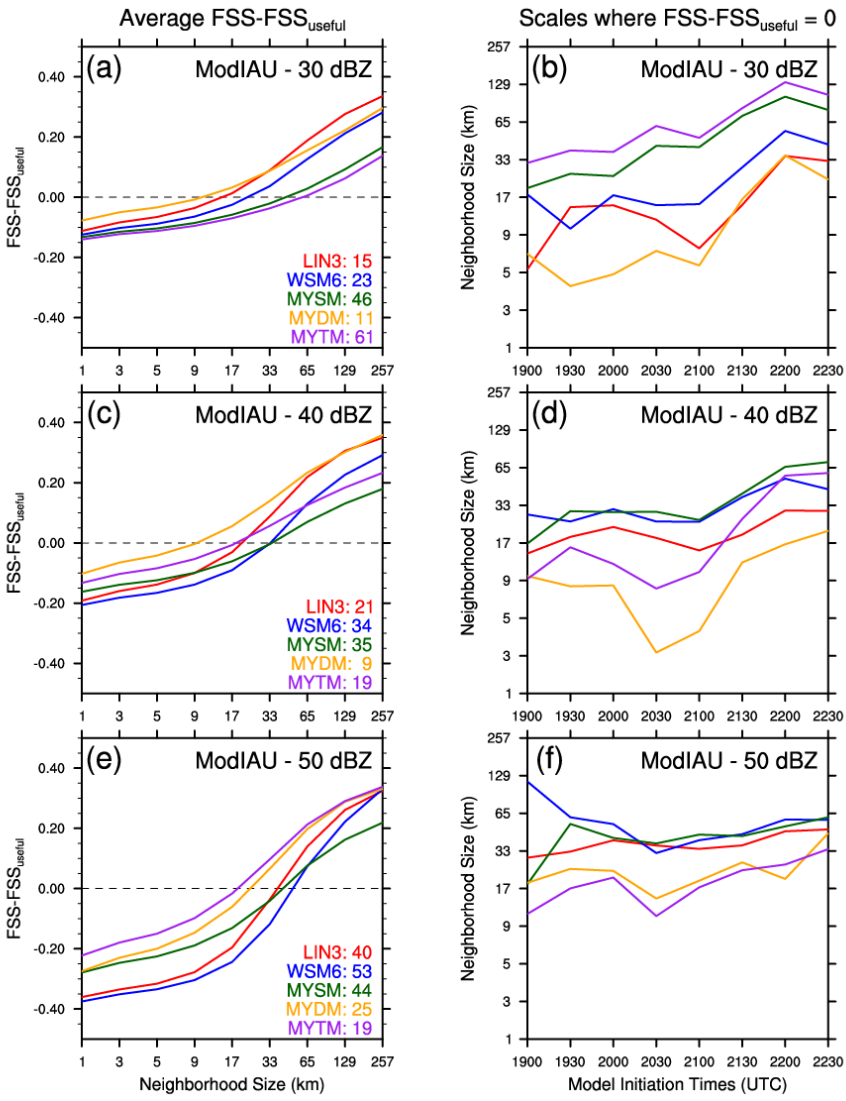

Figure 51. Same as for Figure 15 in Chapter 4, but for ModIAU.

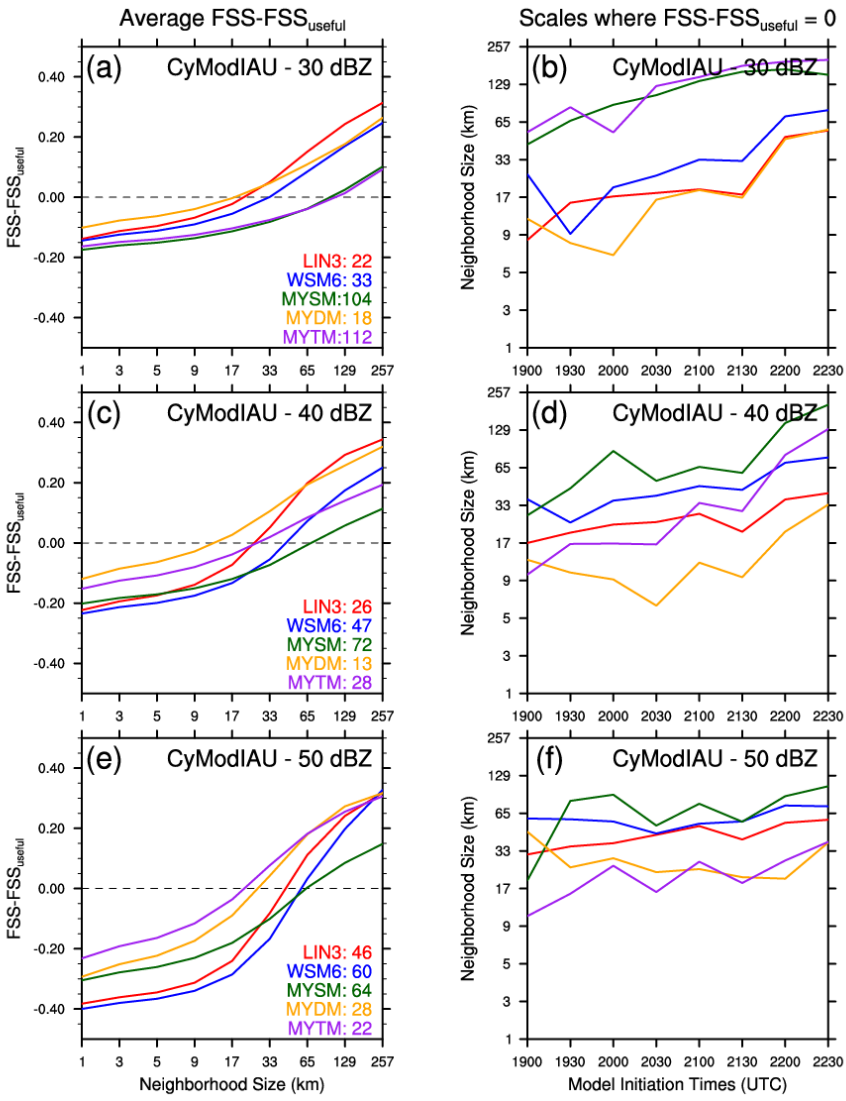

Figure 52. Same as for Figure 15 in Chapter 4, but for CyModIAU.

6.5 Object-Based Results

Subjective evaluation of the locations of all 0-1UH centers from the Cycling runs reveals similar densities of centers near S1's tornado locations as the Control runs, but now a higher density of 0-1UH centers surrounds S2 and, to lesser extent, S3's tornado points (Figure 53a). Also, the Cycling runs yield nearly 150 or more 0-1UH centers across all eight simulations for each microphysics scheme with little change to the average of each center's max 0-1UH value, except for MYDM (Figure 53b,c). For the ModIAU runs, the density of 0-1UH centers are similar to the Control runs' density of centers for all three storms of interest (Figure 54a). The ModIAU simulations using the MY schemes exhibit fewer 0-1UH centers than the Control runs, but the runs with LIN3 and WSM6 experience more 0-1UH centers (Figure 54b). Even with this contrast, all simulations, especially the MYDM runs, generally depict weaker max 0-1UH values for each center (Figure 54c).

While the CyModIAU runs exhibit similar densities of 0-1UH centers near S1 and S2's tornado locations as the Cycling runs, the density of 0-1UH centers near S3's tornado locations appears to be less than the Cycling runs, but more than the ModIAU runs (Figure 55a). The CyModIAU runs using LIN3 and WSM6 result in substantially more 0-1UH centers than the Control runs, and although the runs with MYDM and MYTM have about the same number of 0-1UH centers, the runs with MYSM have substantially fewer 0-1UH centers than the Control runs (Figure 55b). The CyModIAU simulations using the single-moment microphysics schemes exhibit little change in the average of each center's max 0-1UH value, but the runs using MYTM and, especially, MYDM mostly experience substantially weaker max 0-1UH values (Figure 55c).

Overall, analysis cycling improves the densities of 0-1UH centers near S2 and S3's tornado locations while not greatly affecting the density around S1's tornado locations. All of the Cycling, ModIAU, and CyModIAU runs with LIN3 and WSM6 result in an increase in the number of 0-1UH centers, while not substantially impacting the average max strength of each 0-1UH center. However, for the simulations using the MY schemes, analysis cycling leads to considerably more 0-1UH centers, while the modified IAU technique yields fewer 0-1UH centers than the Control runs. Combining the analysis cycling with the modified IAU results in 0-1UH center counts somewhere in between the center counts for the Cycling and ModIAU runs. Also, the runs using the MY schemes, especially MYDM, experience weaker 0-1UH centers for all three types of simulations.

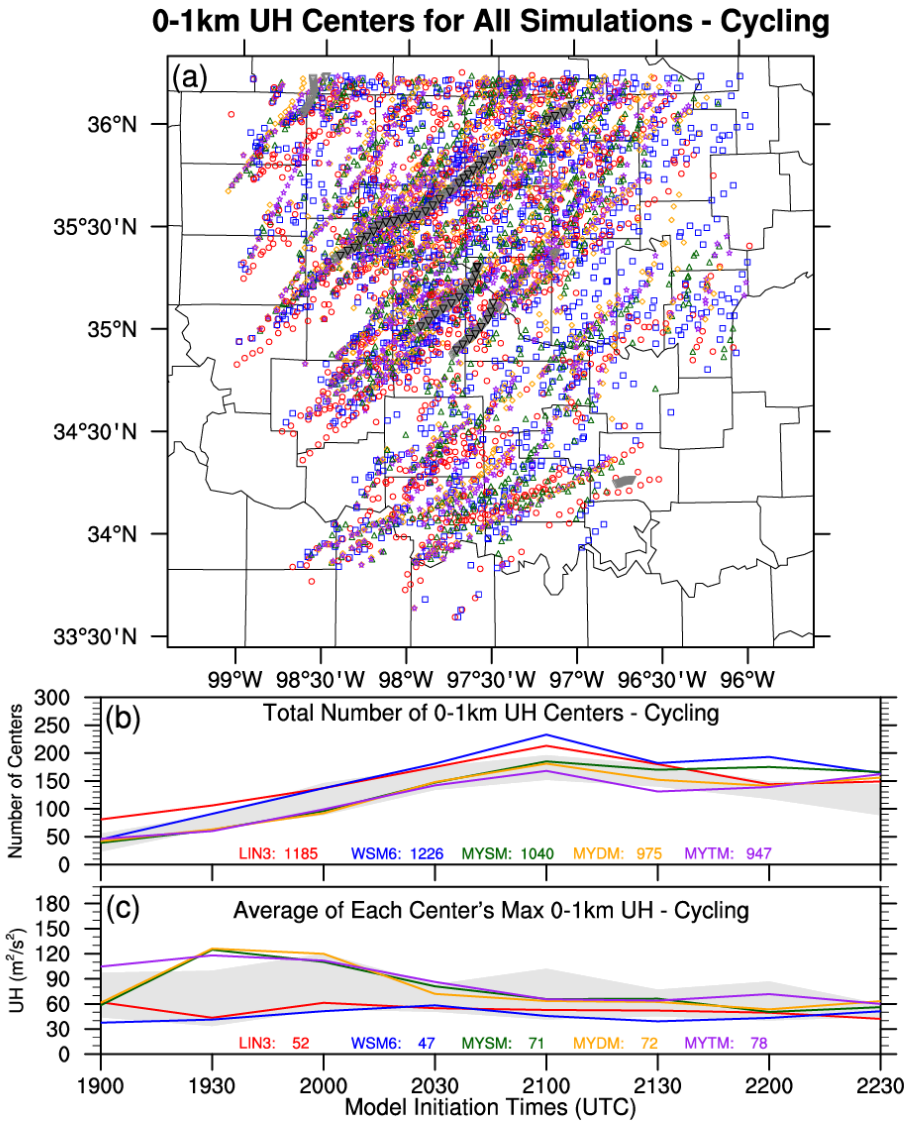

Figure 53. Same as for Figure 17 in Chapter 4, but for Cycling. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

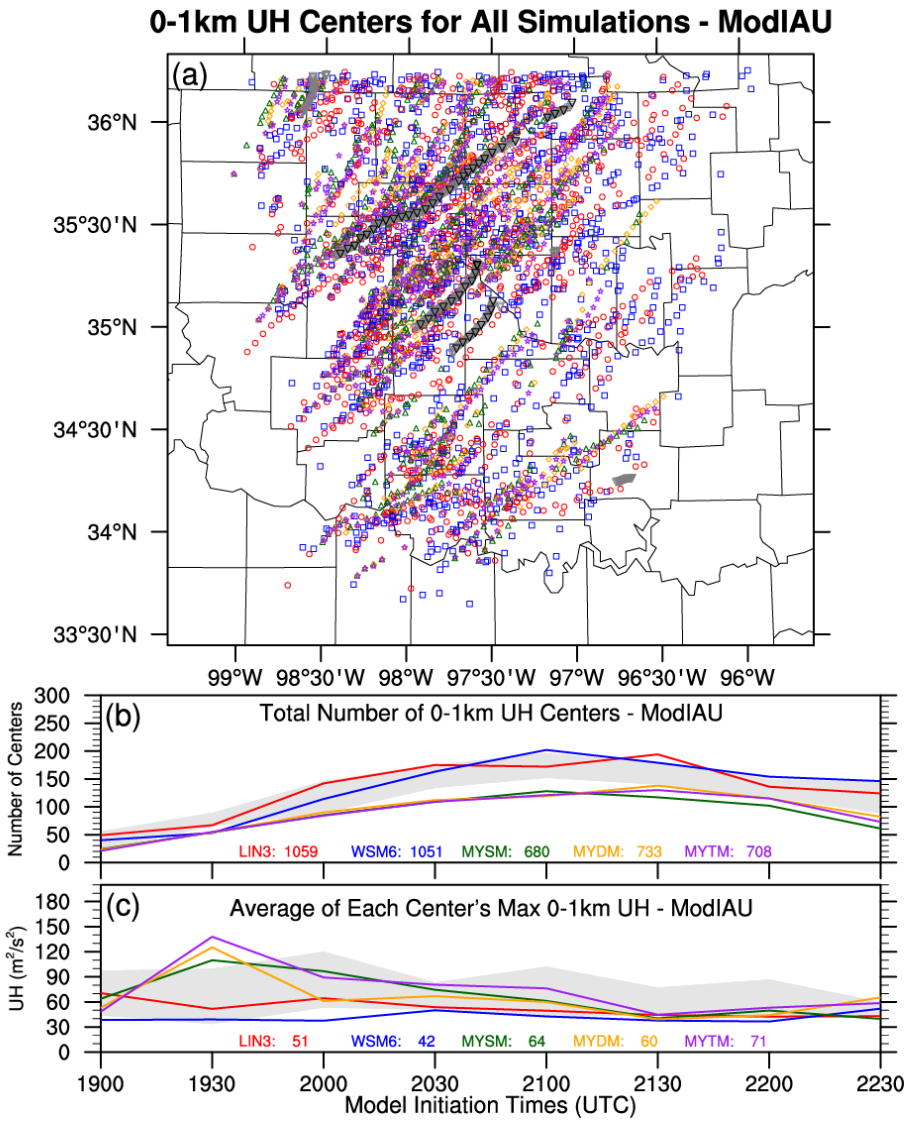

Figure 54. Same as for Figure 17 in Chapter 4, but for ModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

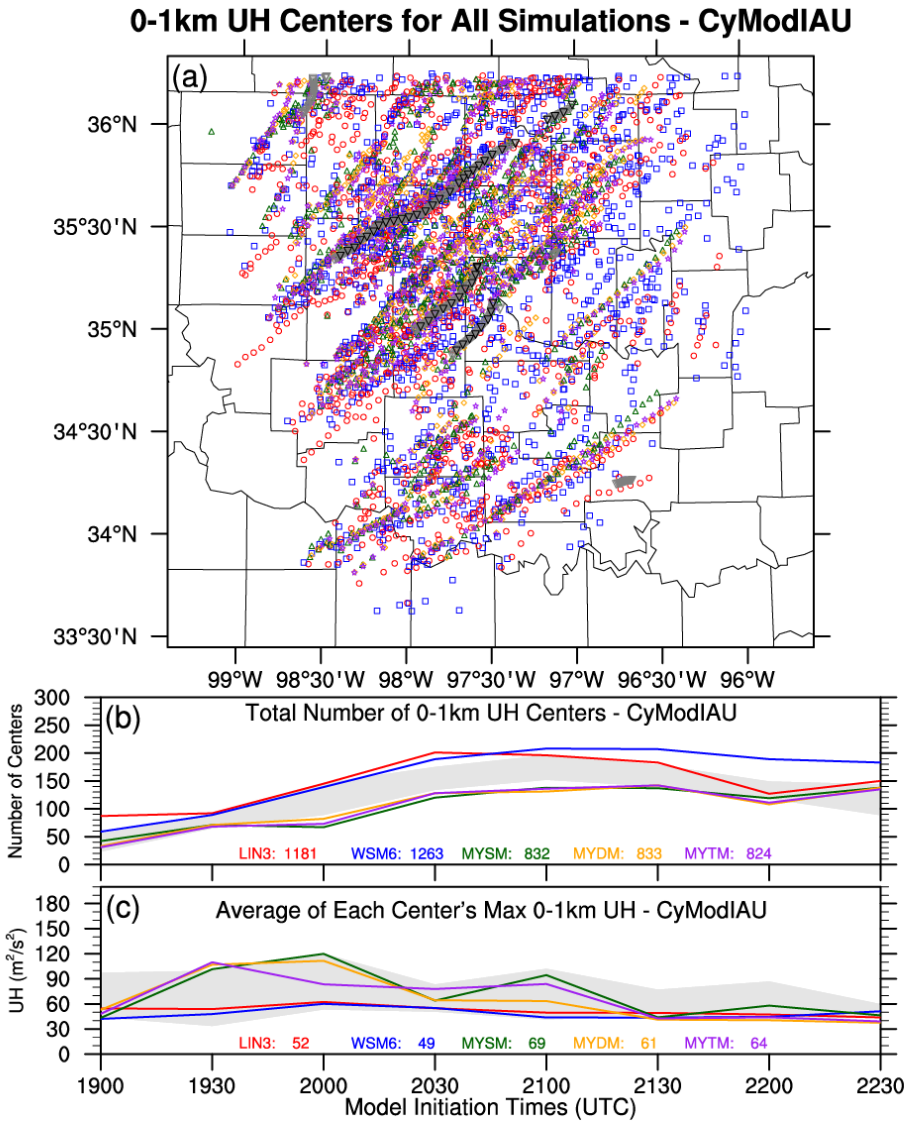

Figure 55. Same as for Figure 17 in Chapter 4, but for CyModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

6.5.1 Storm 1

As shown in Chapter 4, the Control runs had remarkably good success at forecasting S1's circulation tracks, so it will be difficult for the analysis cycling and modified IAU techniques to improve upon the Control runs' forecasts. For the Cycling runs, the number of ST 0-1UH centers is slightly larger (i.e., 2–14) for all microphysics schemes, and the average ST 0-1UH center location for each microphysics scheme is slightly further to the south, especially for the single-moment schemes (Figure 56a). There are also slightly more AT 0-1UH centers (i.e., 3–10) for all microphysics schemes, and there appears to be a little more spread in the AT 0-1UH centers than for the Control runs (Figure 56b). These small changes translate into minimal ST and AT distance and timing error changes between the Cycling and Control runs (Figure 56c,d,e).

Except for WSM6, the other microphysics schemes' ModIAU runs experience a small reduction (i.e., 2–8) in the number of ST 0-1UH centers, and there is a general shift of the collective mass of 0-1UH centers to the north for all microphysics schemes as compared to the Control runs (Figure 57a). This slight shift to the north is also evident in the AT 0-1UH centers and is more pronounced with the MY schemes' ModIAU runs, which also result in fewer AT 0-1UH centers. (Figure 57b). These slight northward shifts in the 0-1UH centers contribute to small increases in ST and AT distance and timing errors for the MY schemes and mixed results for LIN3 and WSM6, but even with the increases in errors for the MY schemes, the ModIAU runs' forecasts for S1 are considered to be just as successful as the Control runs' forecasts (Figure 57c,d,e).

The CyModIAU runs' ST 0-1UH centers are slightly further south than the Control runs' 0-1UH centers just like for the Cycling runs, but the MY schemes' ST 0-1UH centers are a little slower than the ST 0-1UH centers for both the Cycling and Control runs (Figure 58a). The slight shift to the south is also evident with the AT 0-1UH centers (Figure 58b), but this change is less noticeable than for the ST 0-1UH centers. Also, most of the various microphysics schemes' runs result in more ST and AT 0-1UH centers than the Control runs (Figure 58a,b). The CyModIAU runs using LIN3, MYDM, and MYTM experience the largest increases in ST and AT distance errors, while WSM6 has smaller distance errors (Figure 58c,d). All of the CyModIAU simulations generally involve increases in AT timing errors, but the MY schemes exhibit the largest increases in AT timing errors due to mostly being slower than the Control runs (Figure 58e).

Overall for S1, analysis cycling leads to an increase in the number of 0-1UH centers and a slight shift to the south, and the modified IAU technique contributes to a reduction in the number of 0-1UH centers and a slight shift to the north. The combination of the analysis cycling and the modified IAU technique results in mostly an increase in the number of 0-1UH centers and a slight shift south with a slow down for the MY schemes. Thus, analysis cycling has more of an impact than the modified IAU when both are used together. However, most of these changes are relatively small, and the success of the forecasts from the Cycling, ModIAU, and CyModIAU runs is comparable to the Control runs' successful forecasts of S1.

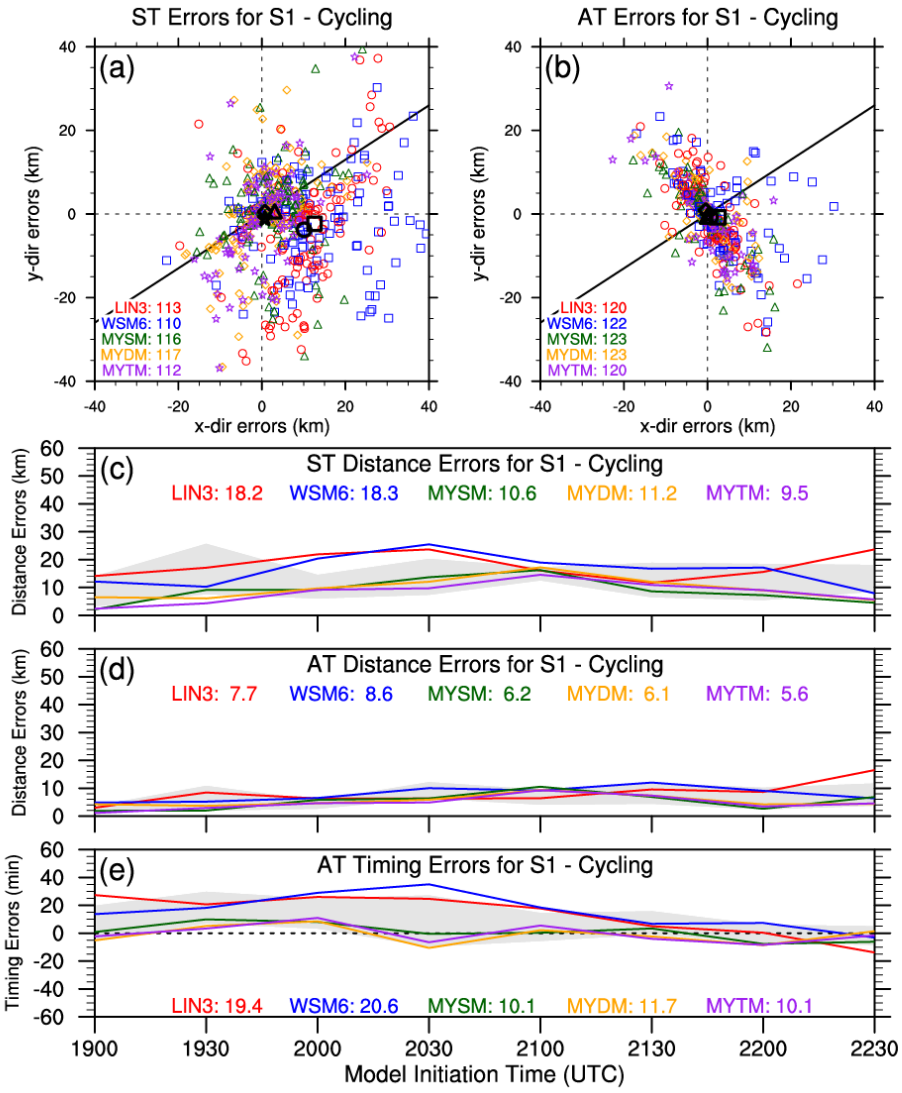

Figure 56. Same as for Figure 18 in Chapter 4, but for Cycling. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

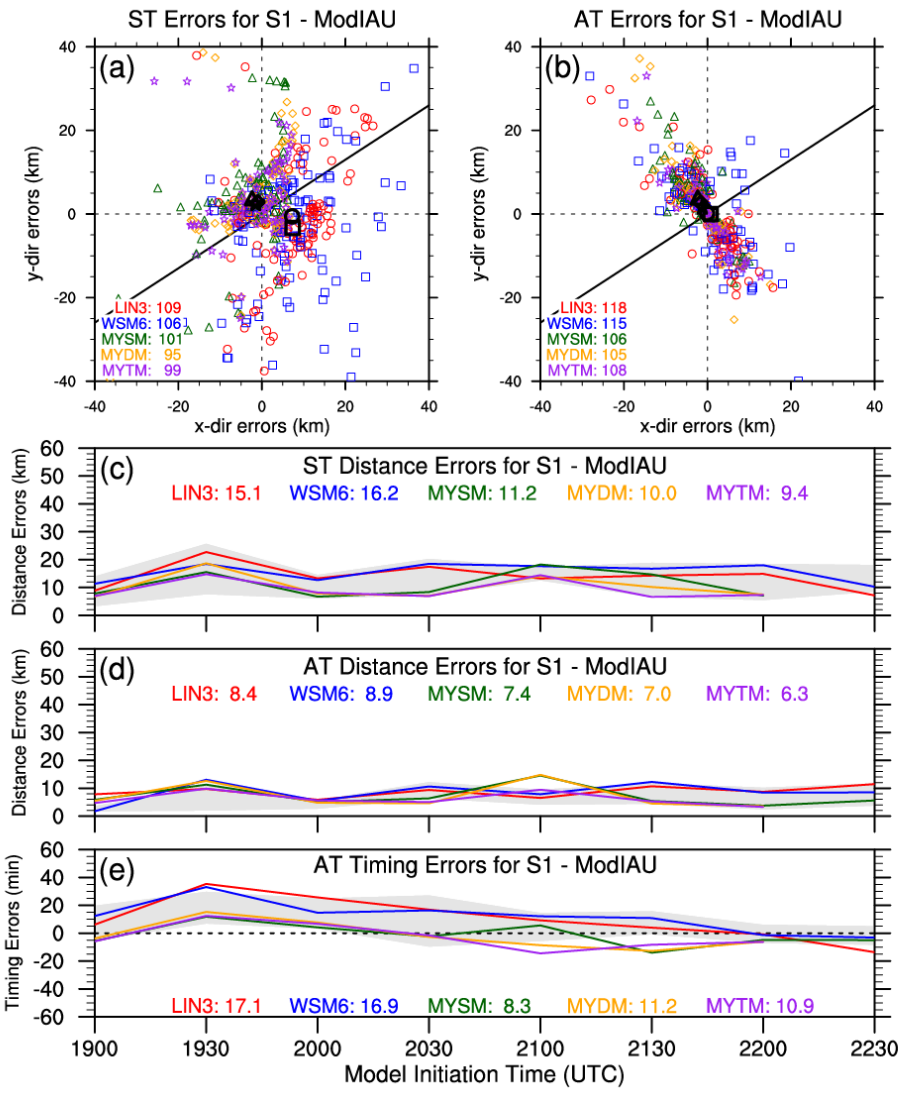

Figure 57. Same as for Figure 18 in Chapter 4, but for ModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

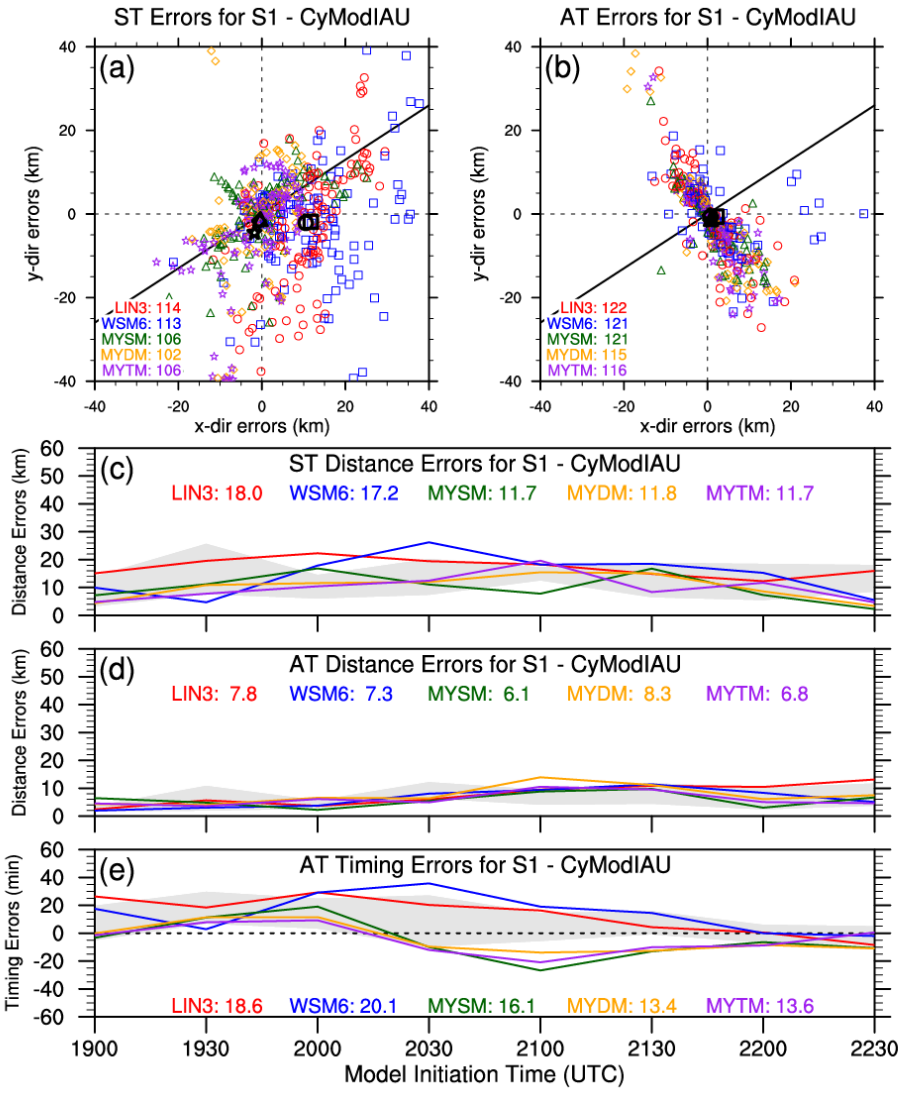

Figure 58. Same as for Figure 18 in Chapter 4, but for CyModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

6.5.2 Storm 2

The Control runs' forecasts for S2 were fairly good, but perhaps improvements can be attained through the use of different data assimilation techniques. Except for MYDM, the Cycling runs yield a small increase in the number of ST 0-1UH centers near S2 compared to the Control runs, and the average ST 0-1UH center locations for each microphysics scheme is generally closer to S2's tornado locations, except for LIN3, which is further south (Figure 59a). Besides the MYSM runs, there are fewer AT 0-1UH centers near S2 than the Control runs, and the average AT 0-1UH center locations for each microphysics scheme are slightly faster, especially the WSM6 runs, and slightly further south, especially the LIN3 runs, than the AT 0-1UH centers from the Control runs (Figure 59b). This results in the MY schemes' AT 0-1UH centers being somewhat closer to S2's tornado locations. The Cycling runs using LIN3 and WSM6 yield a reduction in the ST distance errors, but the runs using the MY schemes depict an overall increase in ST distance errors (Figure 59c). Most of the Cycling runs, especially those using WSM6, experience a reduction of AT distance errors (Figure 59d), but they generally result in an increase in AT timing errors, except for LIN3, which exhibits the largest AT timing errors (Figure 59e).

Each set of ModIAU simulations for the different microphysics schemes experiences an increase of up to 5 additional ST 0-1UH centers near S2 as compared to the Control runs, and except for LIN3, the other microphysics schemes' ModIAU runs' average ST 0-1UH center locations are slightly further north (Figure 60a). Interestingly, two distinct clusters of ST 0-1UH centers from the ModIAU runs using the MY schemes are depicted with one too slow and the other too fast and too far north (Figure 60a).

Except for LIN3, the number of AT 0-1UH centers near S2 is larger and slightly further north (Figure 60b). Besides the runs using LIN3, the other ModIAU runs generally experience a reduction in the ST distance errors (Figure 60c). The ModIAU runs using WSM6 or MYDM result in smaller AT distance errors (Figure 60d), and the runs using LIN3, MYSM, or MYDM have smaller AT timing errors (Figure 60e).

For the CyModIAU runs, the number of ST 0-1UH centers is larger than for the Control runs, except for the runs using WSM6, and while the MY schemes' CyModIAU runs' average ST 0-1UH center locations are similar or closer to the Control runs' centers, the ST 01-UH centers for LIN3 and WSM6 are generally further away (Figure 61a). The AT 0-1UH centers produced by the simulations using MYDM and MYTM are generally slightly closer to S2's tornado locations, but the CyModIAU runs using LIN3 and WSM6 are slightly faster (Figure 61b). However, the spread in AT 0-1UH centers for the runs using WSM6 is smaller. Except for the runs using LIN3, the CyModIAU runs all mostly yield smaller ST and AT distance errors (Figure 61c,d). Just like for the ModIAU runs, the CyModIAU runs using LIN3, MYSM, and MYDM all have smaller AT timing errors, while the other two have larger AT timing errors (Figure 61e).

Overall for S2, analysis cycling results in a general southward shift for a majority of the 0-1UH centers, and the modified technique appears to have consolidated a majority of the 0-1UH centers into somewhat distinct clusters. Combining the analysis cycling with the modified IAU technique yields a result somewhere in between the results of the Cycling and ModIAU runs. Therefore, the CyModIAU runs exhibit the largest overall improvement over the Control runs.

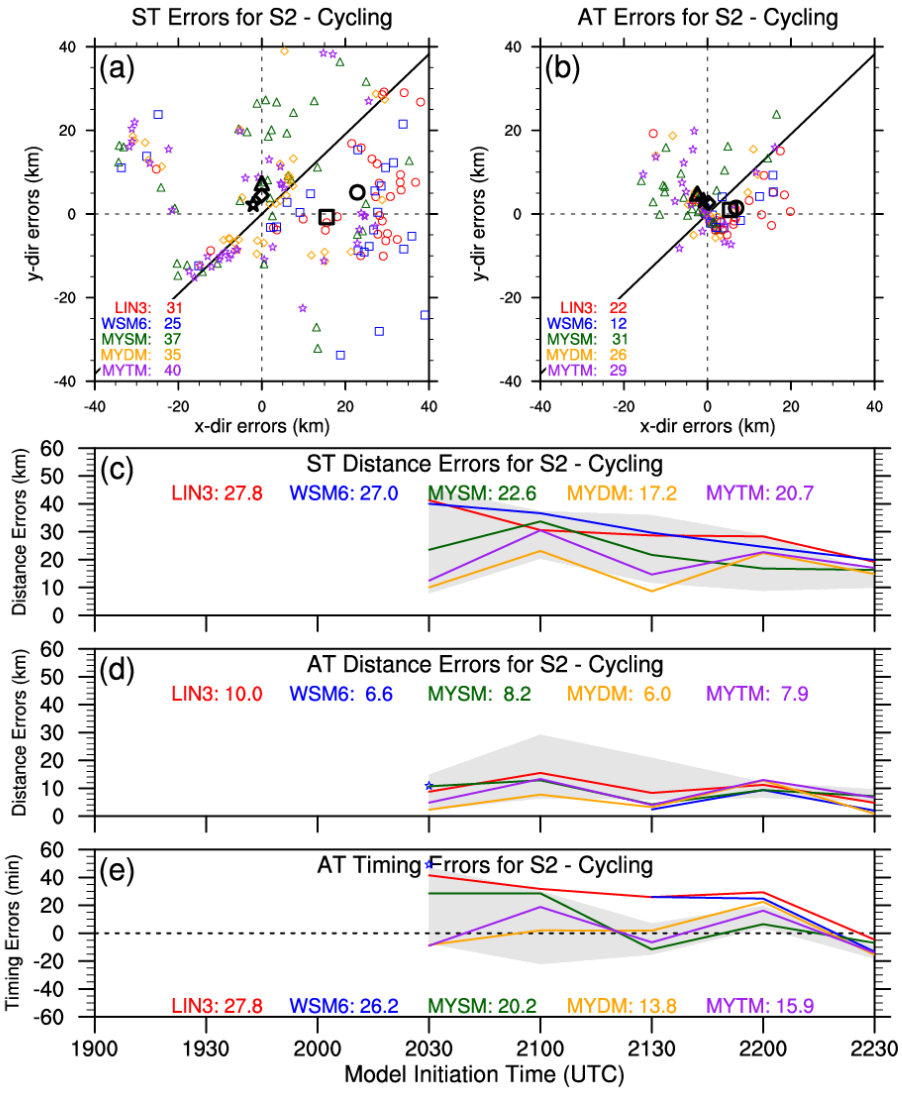

Figure 59. Same as for Figure 20 in Chapter 4, but for Cycling. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

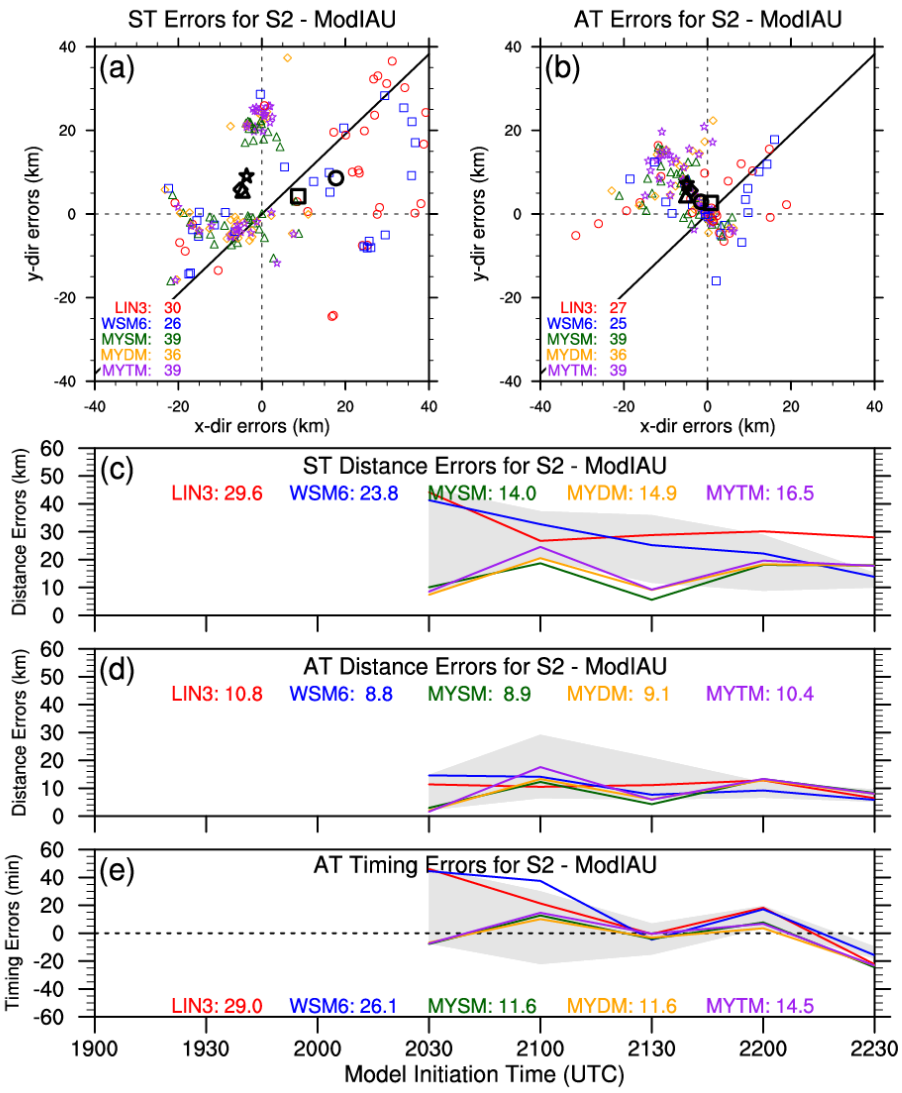

Figure 60. Same as for Figure 20 in Chapter 4, but for ModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

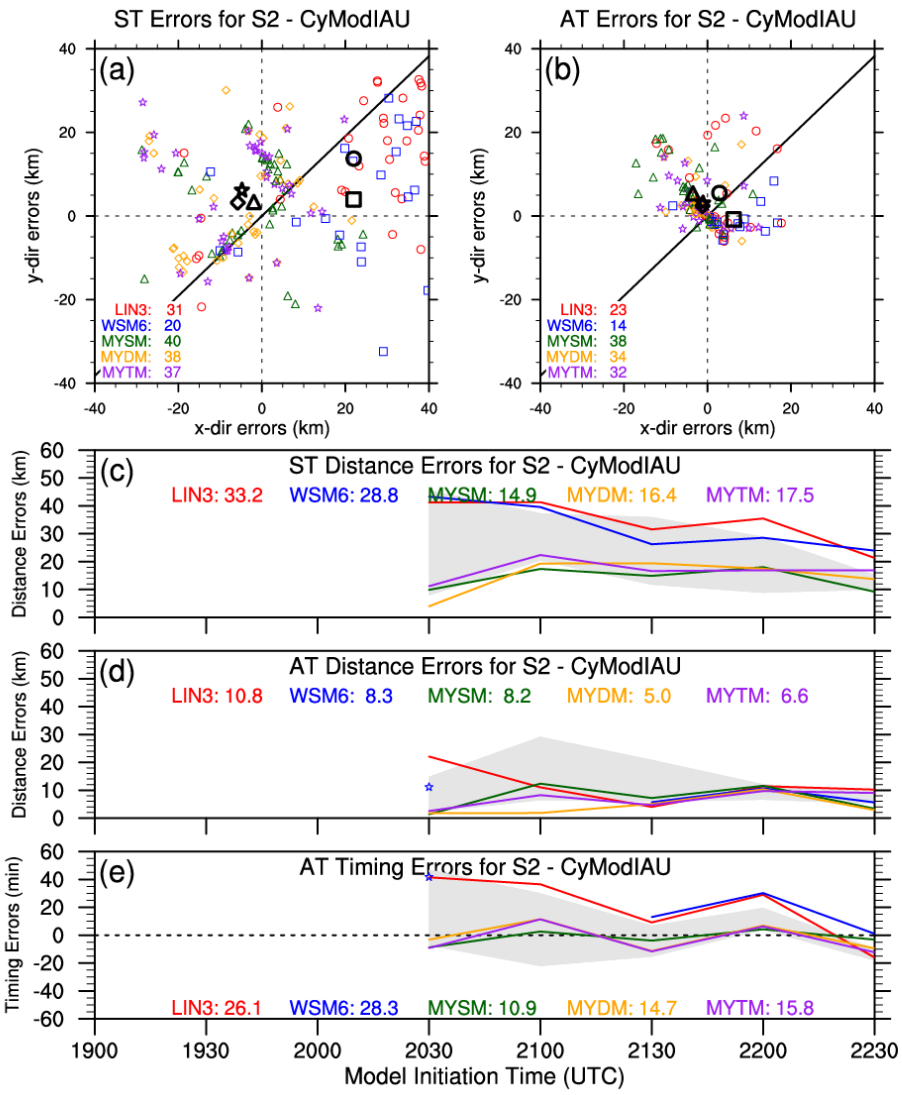

Figure 61. Same as for Figure 20 in Chapter 4, but for CyModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

6.5.3 Storm 3

Because of the numerous storm interactions and mergers ongoing at the start of S3's tornadoes, the tested forecast systems, including the Control runs, struggle to appropriately depict S3's circulation tracks, so perhaps introducing other data assimilation techniques can improve the forecasts of S3. Except for the Cycling runs using LIN3, the runs using the other microphysics schemes generally produce more ST and AT 0-1UH centers (Figure 62a,b). Even though the only Cycling runs with average ST 0-1UH center locations closer to S3's tornado locations use MYSM and MYDM, a cluster of ST 0-1UH centers exists near S3's tornado locations and includes centers from runs associated with each microphysics scheme (Figure 62a). Although a cluster of AT 0-1UH centers still exists ~30 km to the south of S3's track, a cluster of AT 0-1UH centers also exists near S3's tornado locations (Figure 62b). Because the Cycling runs have this cluster of 0-1UH centers near S3 and the Control runs do not, ST and AT distance and timing errors are substantially reduced for most of the Cycling simulations, especially the ones using MYDM or MYTM (Figure 62c,d,e).

Similar to the Control runs, the ModIAU runs tend to produce ST 0-1UH centers too far south and no distinct clusters within ~20 km of S3's tornado locations (Figure 63a). Except for ModIAU runs using WSM6, there is an overall decrease in the number of AT 0-1UH centers, but the cluster of AT 0-1UH centers ~30 km south of S3's tornado locations in the Control runs' forecasts is ~10 km closer in the ModIAU runs' forecasts (Figure 63b). Most of the ModIAU runs exhibit smaller ST and AT distance errors (Figure 63c,d). The ModIAU runs using the single-moment microphysics schemes yield larger AT timing errors, while the runs with the multi-moment microphysics schemes

experience slightly smaller AT timing errors compared to the Control runs (Figure 63e). However, it is important to note that most of these results are based on fewer than 15 0-1UH centers, so the results might not be robust.

Unlike the Control runs, the CyModIAU runs, except for the ones using LIN3, produce a cluster of ST and AT 0-1UH centers within ~10 km north of S3's tornado locations, while reducing the number of 0-1UH centers > 20 km to the south (Figure 64a,b). The result of these 0-1UH center distributions is an overall substantial reduction in ST and AT distance errors for all CyModIAU runs (Figure 64c,d). Additionally, the CyModIAU simulations using WSM6, MYDM, and MYTM generally yield smaller AT timing errors (Figure 64e).

Overall for S3, analysis cycling greatly improves the predictions of S3 for some of the forecast runs. Even though the modified IAU technique didn't have as large of a positive impact on the forecasts of S3, the combination of analysis cycling with the modified IAU technique yields slightly better forecasts of S3 than just using analysis cycling with the original IAU technique.

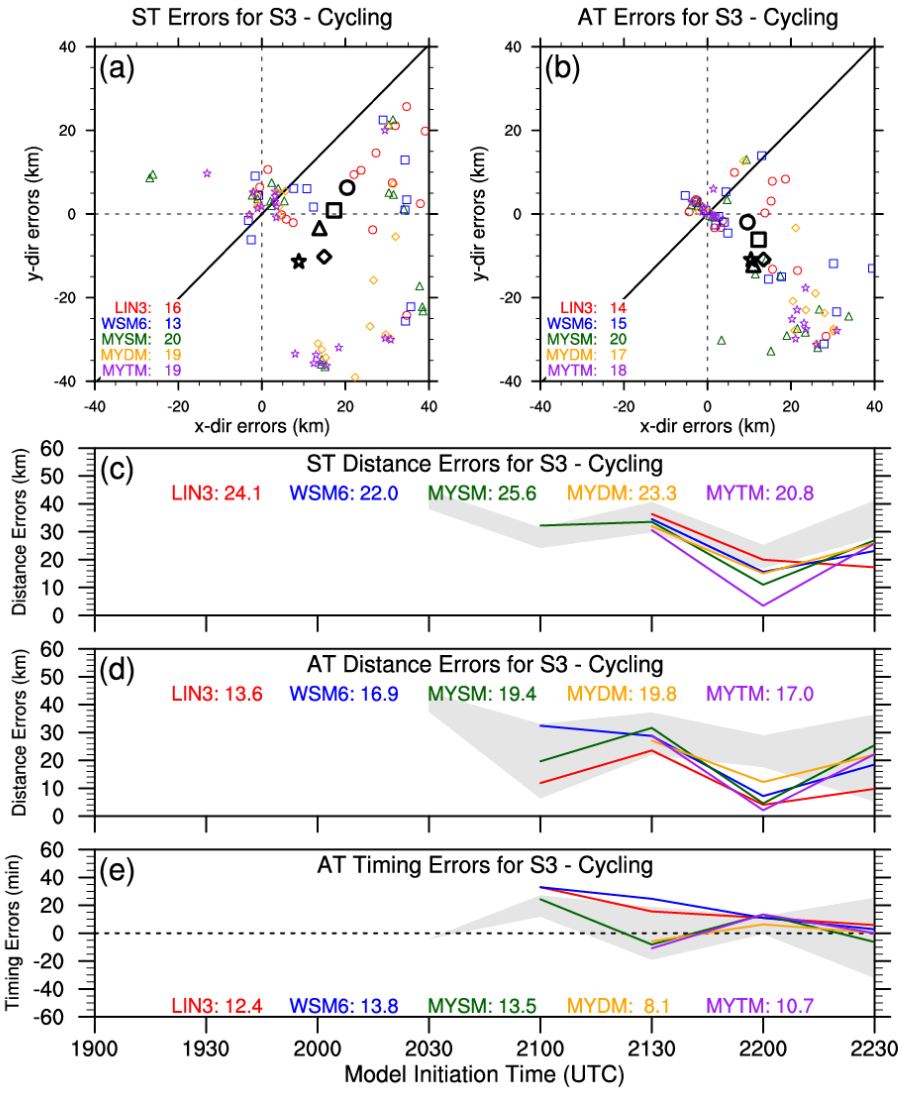

Figure 62. Same as for Figure 21 in Chapter 4, but for Cycling. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

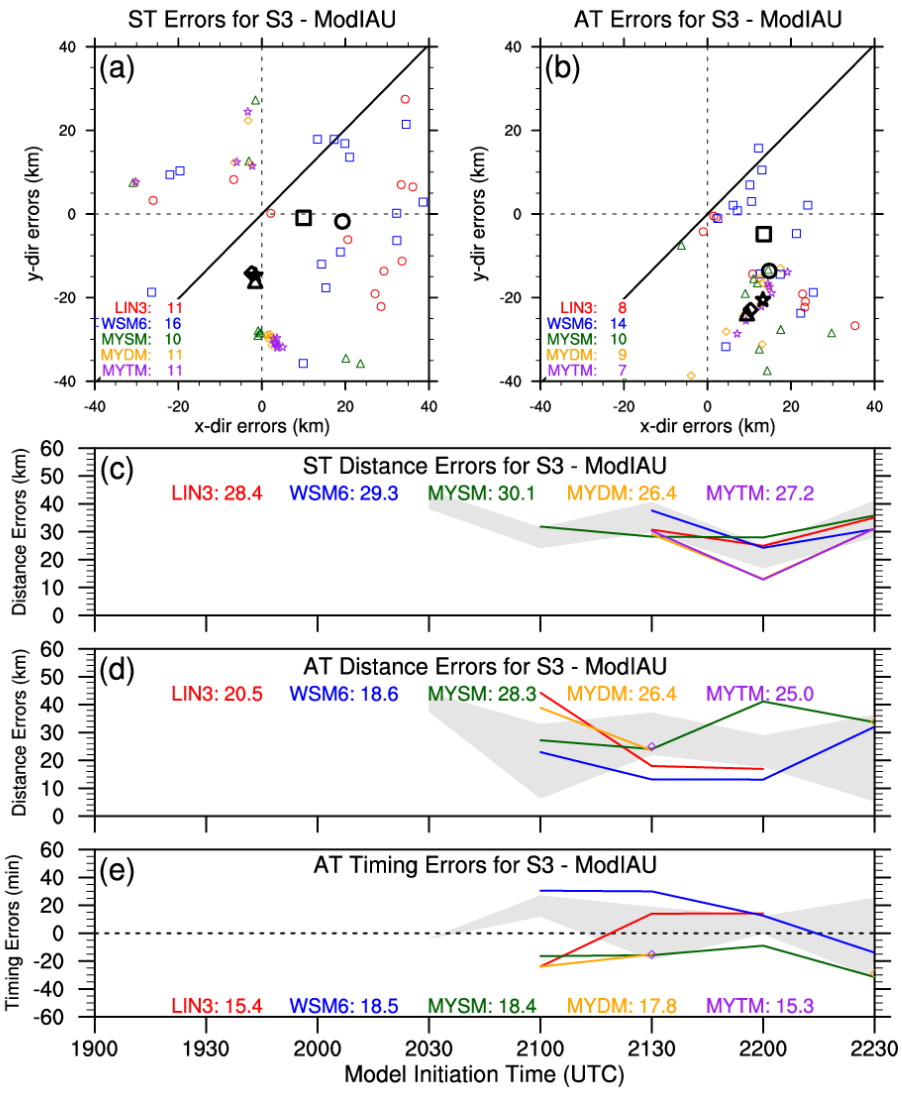

Figure 63. Same as for Figure 21 in Chapter 4, but for ModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

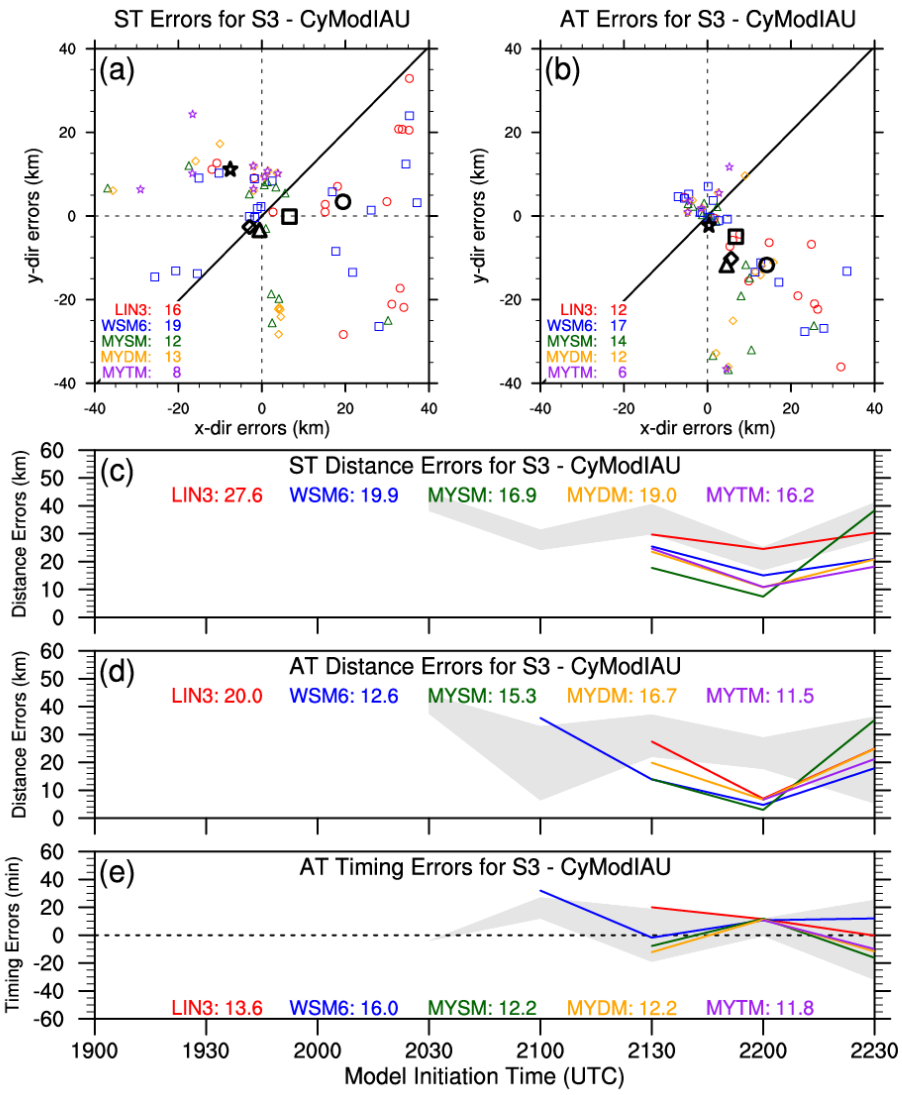

Figure 64. Same as for Figure 21 in Chapter 4, but for CyModIAU. Light gray shading in (c), (d), and (e) represents the max/min bounds of the Control runs for comparison purposes.

Chapter 7 – Summary and Discussion

7.1 General Summary of Dissertation

On 24 May 2011, an outbreak of tornadic supercells propagated across Central Oklahoma producing several strong to violent tornadoes. Unfortunately, the tornadoes and supercells caused numerous deaths and injuries along with widespread damage. However, the extensive observation network across Central Oklahoma during the spring of 2011 provides an opportunity for data assimilation and forecast system sensitivity studies, such as this one, to explore ways to potentially improve short-term, storm-scale forecasts of severe convection. The extensive observation network consisted of several NWS/FAA METAR locations, ~70 Oklahoma Mesonet stations, 4 CASA IP-1 X-band radars, and 7 WSR-88D S-band radars. The data from these observation sources were used in the data assimilation process to produce initial analyses for an ARPS 1-km model. The data assimilation process consisted of using 3DVAR and ADAS complex cloud analysis to produce model state variable increments, which were introduced during an IAU window. In conjunction with this study, the ADAS complex cloud analysis (Brewster and Stratman, 2015) and IAU (Brewster et al., 2015; Brewster and Stratman, 2016) packages within the ARPS system were both updated for use in this study.

To emulate a real-time, storm-scale forecasting system, eight microphysically-diverse sets of simulations were initialized every 30 min starting at 1900 UTC and integrated out to 2 hours. This forecast framework allowed for in-depth examinations of sensitivities related to microphysics parameterization schemes, radar data assimilation, and various data assimilation techniques. For the first experiment,

simulations with different microphysics schemes (i.e., LIN3, WSM6, MYSM, MYDM, and MYTM) were compared to each other to assess the characteristics of each scheme and to determine which scheme leads to better forecast results. To assess the ability of CASA radars being used as gap-filling radars in data assimilation, the second set of experiments examined the impact of assimilating CASA radar data relative to the impact of assimilating KTLX radar data. For the last set of experiments, analysis cycling and variable-dependent IAU techniques were employed separately and together to determine any potential benefits they may contribute to the forecasts.

To quantitatively assess these experiments, observation point-based, neighborhood-based, and object-based verifications techniques were utilized in the evaluation process. For the observation point-based verification metrics, RMSEs, linear regression slopes, and R^2 values were used to evaluate near-surface variables. For the neighborhood-based verification metric, the FSS was utilized to determine the scales at which forecasts of simulated composite reflectivity exhibited "useful" skill. For the object-based verification technique, objectively-determined 0-1UH centers were verified with estimated tornado locations for three storms of interest by calculating distance and timing errors. A summary of the verification results from these experiments are presented in the next couple of sections followed by a discussion and potential future work section.

7.2 Summary of Microphysics Schemes Sensitivities

Prior studies investigating the differences between single-moment and multi-moment microphysics schemes' impact on forecasts of supercells (Dawson et al., 2010, 2015; Yussouf et al., 2013) and MCSs (Wheatley et al., 2014) have generally found that the multi-moment microphysics schemes yield better forecasts of storm intensity, structure, and evolution. Hence, the expectation was that MYDM and MYTM would outperform LIN3, WSM6, and MYSM. For near-surface temperatures, it could be argued that MYSM outperformed MYDM and MYTM due to having slopes closer to 1 (Figure 65). However, MYDM and MYTM have smaller RMSEs and larger R^2 values, so MYSM's success over MYDM and MYTM is limited. For the other near-surface variables, MYDM and MYTM do show improvement over the single-moment schemes, but it's worth noting that MYSM's forecasts of near-surface variables end up closer to MYDM and MYTM's forecasts than LIN3 and WSM6's forecasts, which tend to have colder and larger cold pools (see Figure 13). By these metrics it appears the MY schemes, as a group, outperformed LIN3 and WSM6.

Due to having the largest forecasted base rates for composite reflectivity (see Figure 16), MYSM exhibited the least "useful" skill among the five microphysics schemes (Figure 65). LIN3 has a better overall score than MYTM for the neighborhood-based verification technique likely due to MYTM's larger forecasted base rates, but generally, MYTM depicted more skill at the 40- and 50-dBZ thresholds and at the smaller scales. Even with commonly higher forecasted base rates than LIN3 and WSM6 (but less than MYSM and MYTM), MYDM usually exhibited the most skill at all three reflectivity thresholds.

In regards to 0-1UH centers, the MY schemes resulted in substantially fewer and stronger 0-1UH centers than LIN3 and WSM6, so in this respect, no inference can be made that separates the single-moment microphysics schemes from the multi-moment microphysics schemes. For S1, the MY schemes predominately performed similar to each other and largely outperformed LIN3 and WSM6, but overall, the entire set of simulations had remarkably good success at forecasting S1's circulation locations (Figure 65). Even though the simulations had somewhat less success at forecasting S2's circulation locations, the MY schemes still substantially outperformed LIN3 and WSM6 (Figure 65). Conversely, all of the microphysics schemes using the base data assimilation scheme struggled with forecasting the more complex S3 and mostly yielded somewhat similar errors (Figure 65).

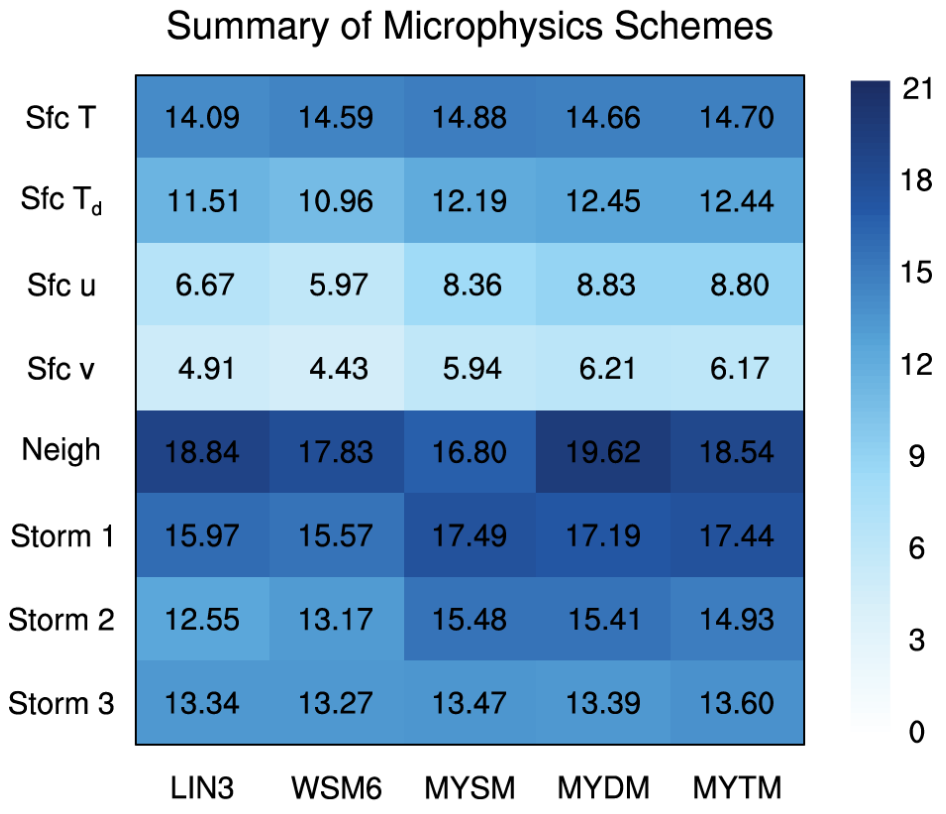

Figure 65. First, unity-based normalization is used to normalize all of the verification metrics' average values for each simulation run with a value of 1 being perfect. The

perfect skill (no skill) values used for this normalization include 5 (0) for RMSE, 1 (0) for slopes and R^2 values, 1 (257) for FSS–FSS_{useful} = 0 scales, 0 (56.6) for the distance errors, and 0 (60) for the timing errors. Second, each verification variable's three normalized values from all run types are summed together. For example, the surface temperature's normalized values for RMSE, slope, and R^2 value from all of the runs (Control, NoCASA, etc.) are summed together. This summation gives a score out of 21 with 21 indicating perfect forecast skill. Comparisons can only be made among the different microphysics schemes and variables with the same units.

7.3 Summary of Data Assimilation Sensitivities

Withholding CASA and/or KTLX radar data during the data assimilation process has little to no impact on the forecasts of near-surface variables (Figure 66). Even though the NoCASA runs exhibited similar FSS values as the Control runs, the NoKTLX and NoCASAKTLX runs yielded at least some increase in "useful" skill (Figure 66). However, this increase in skill is likely due to the decrease in forecasted base rates of composite reflectivity caused by the withholding of KTLX radar data from the data assimilation process. The total number of 0-1UH centers and their average max intensities for the NoCASA runs were similar to the Control runs, but the NoKTLX and NoCASAKTLX runs typically resulted in weaker and fewer 0-1UH centers (not shown).

The NoCASA runs generally forecasted S1's circulation locations better than the Control runs, but the differences are relatively small (Figure 66). When KTLX radar data are withheld from the data assimilation process, the forecasts errors of S1's circulation locations are mostly larger, especially for the runs using MYDM and MYTM (Figure 66). However, S1 formed and tracked well north of the CASA radar network, so the S1 results are not too surprising. Withholding CASA radar data during the data assimilation process had more of a negative impact on the forecasts of S2's circulation tracks than the withholding of KTLX radar data (Figure 66). The negative impact is even larger when both CASA and KTLX radar data were withheld from the data assimilation process. Therefore, with the absence of KTLX radar data, the assimilation of CASA radar data was able to prevent some of the loss of forecast skill, consistent with the results of Xue et al. (2006). Interestingly, the NoCASA, NoKTLX and

NoCASAKTLX runs all resulted in decreased errors for the forecasts of S3's circulation locations (Figure 66). However, small changes in the number and locations of 0-1UH centers near S3 can have a substantial impact on the forecast errors, as these results indicate.

Analysis cycling had little to no effect on the near-surface temperature forecasts, but the modified IAU technique generally improved the near-surface temperature forecasts even though the simulations using the MY schemes depicted larger RMSEs (Figure 66). Both analysis cycling and the modified IAU technique resulted in improved forecasts of the other near-surface variables, and the use of both data assimilation techniques lead to the largest improvements for those same variables (Figure 66). Analysis cycling in both the Cycling and CyModIAU runs caused a substantial increase in forecasted base rates of composite reflectivity (not shown), so those runs exhibited substantially less "useful" skill (Figure 66). The modified IAU technique in the ModIAU runs had little impact to this verification technique's result compared to the Control runs.

The Cycling runs produced substantially more 0-1UH centers than the Control runs, but the results were more mixed for the ModIAU and CyModIAU runs; the runs using the MY schemes generally had fewer 0-1UH centers, and the runs using LIN3 and WSM6 had more 0-1UH centers (not shown). Even though Brewster et al. (2015) and Brewster and Stratman (2016) demonstrated the modified IAU technique aids in stronger updrafts immediately following the assimilation window, the ModIAU and CyModIAU runs using the MY schemes mostly had weaker max 0-1UH intensities. Analysis cycling and the modified IAU technique had a negligible impact on the

forecasts of S1's circulation locations, except for the ModIAU runs using WSM6 (Figure 66). The CyModIAU generally had larger forecast errors than the Control runs for S1, except for the runs using WSM6. The Cycling, ModIAU, and CyModIAU runs all depicted substantial improvements to the forecasts of S2's circulation locations with the modified IAU technique having the largest positive impact (Figure 66). Substantial improvements were also the result for all three data assimilation techniques for S3's tornado locations (Figure 66). For example, the 2130-UTC Control and CyModIAU MYDM runs' simulated reflectivity and 0-1UH fields are depicted in Figure 67, and the CyModIAU run's forecasts show more and stronger 0-1UH centers near S2 and S3 than the Control run's forecasts. Even though the NoCASA, NoKTLX, and NoCASAKTLX runs also depicted similar decreases in forecasted errors, the error reductions in the Cycling and CyModIAU runs were legitimately due to forecasted clusters of 0-1UH centers near S3's tornado locations.

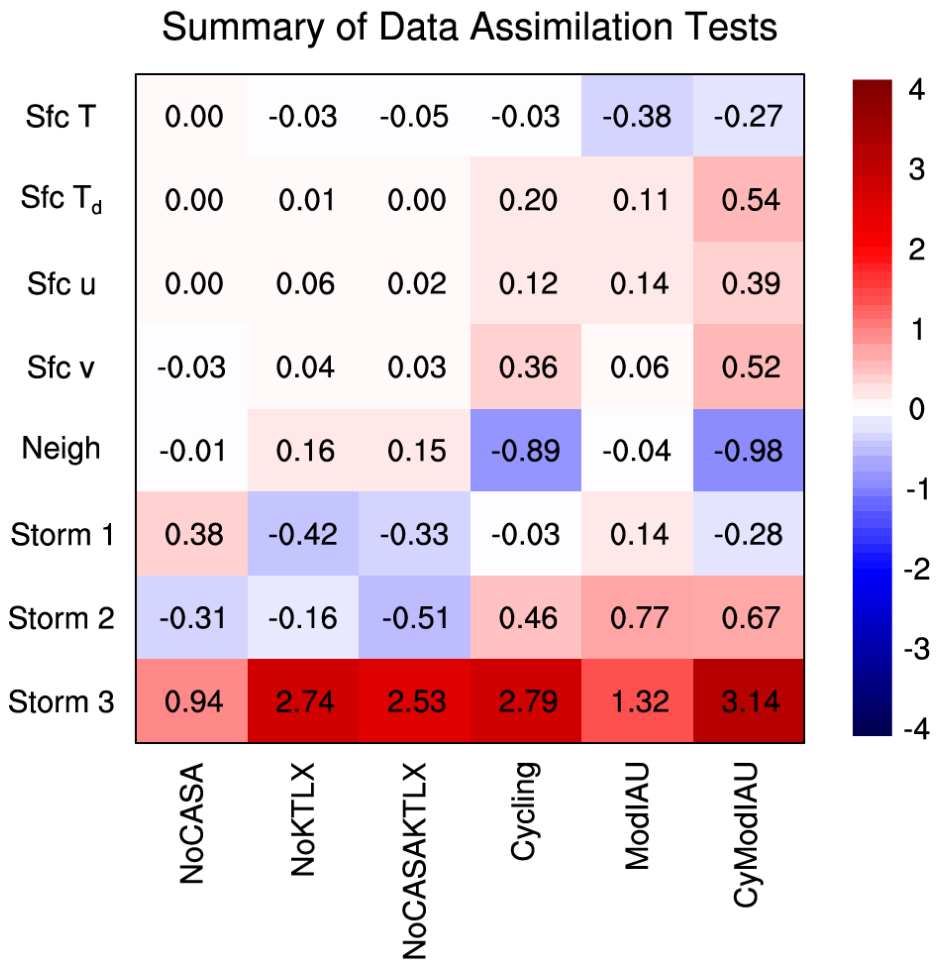

Figure 66. Similar to Figure 65, but instead of summing the normalized values across all run types for each microphysics scheme, the normalized values are summed across all microphysics schemes for each run type. To highlight the sensitivities, the Control run's normalized values are subtracted from the six non-Control runs' normalized values. Positive (negative) values indicate where the non-Control run performed better (worse) than the Control run.

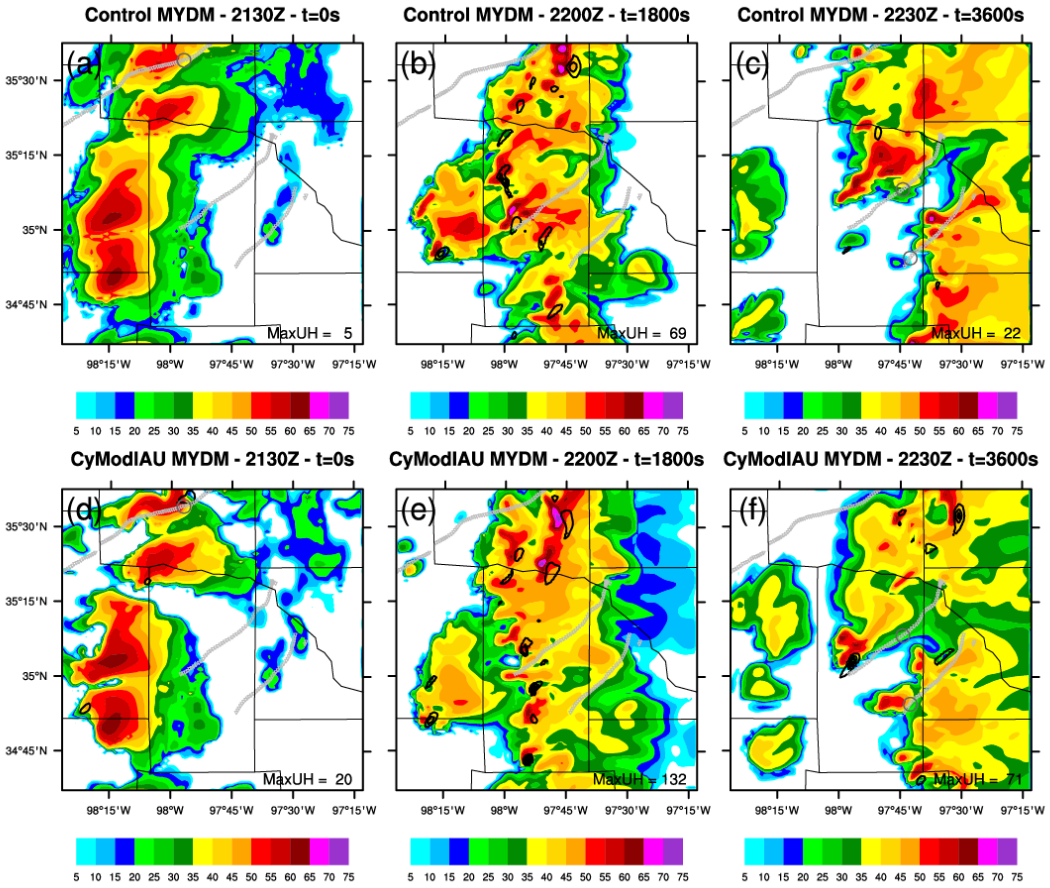

Figure 67. Forecasts of 1-km AGL simulated reflectivity (dBZ; colored) from the 2130-UTC Control (a,b,c) and CyModIAU (d,e,f) MYDM runs are plotted for (a,d) 2130 UTC (t = 0 s), (b,e) 2200 UTC (t = 1800 s), and (c,f) 2230 UTC (t = 3600 s). 0-1UH is contoured in black from 10 m^2/s^2 to 210 m^2/s^2 with an interval of 25 m^2/s^2, and the max UH value (m^2/s^2) in each plot window is annotated near the bottom of each plot. Light gray upside-down triangles depict the estimated tornado locations every minute, and darker gray circles indicate the estimated location of the tornado occurring at 2100 UTC.

7.4 Discussion and Future Work

When designing a real-time forecast system, a balance exists between computational cost and using potentially better data assimilation and modeling techniques, which impacts not only the computational resources required but the real-time latency of the forecast. For example, changing the microphysics parameterization scheme in a model (e.g., from single-moment to multi-moment) might substantially improve forecasts, but that change might create a large increase in computational cost (though not approaching the 1-2 orders of magnitude cost increase for using ensemble based methods). Conversely, changing the way variable increments are introduced during the IAU window might also contribute to substantial forecast improvements while having very little impact on the overall computational cost. Intensive studies need to be carried out to determine what modeling methods yield the biggest forecast improvements for the least amount of additional computational cost. Furthermore, improvements in forecasts can vary depending on what and how a model field is being examined, so subjective evaluations and objective verification techniques need to be tailored to the forecasts' objective (e.g., forecasting low-level circulations). Therefore, for this study, multiple objective verification techniques and metrics were utilized to determine the differences in model skill among the various data assimilation and model configurations with respect to forecasts of tornadic supercells.

Overall, the simulations using the MY schemes outperformed the simulations using LIN3 and WSM6, so based on this finding, the differences between the various types of microphysics schemes (i.e., LIN3 and WSM6 vs. the MY schemes) are larger than the differences between single-moment and multi-moment microphysics schemes,

which doesn't support the first hypothesis. Even though the MY schemes exhibited similar abilities to forecast the tornado locations from the three storms of interest, the multi-moment MY schemes do tend to better forecast the near-surface variables and simulated reflectivity coverage, which does support the first hypothesis. Additionally, the computational cost of the MY schemes is almost double that for LIN3 and WSM6, and while not as large, the differences in computational cost among the MY schemes are also noticeable (Figure 68). All of these factors must be taken into account when designing a model configuration best suited for storm-scale, short-term forecasts of severe convection within the confines of computational resources.

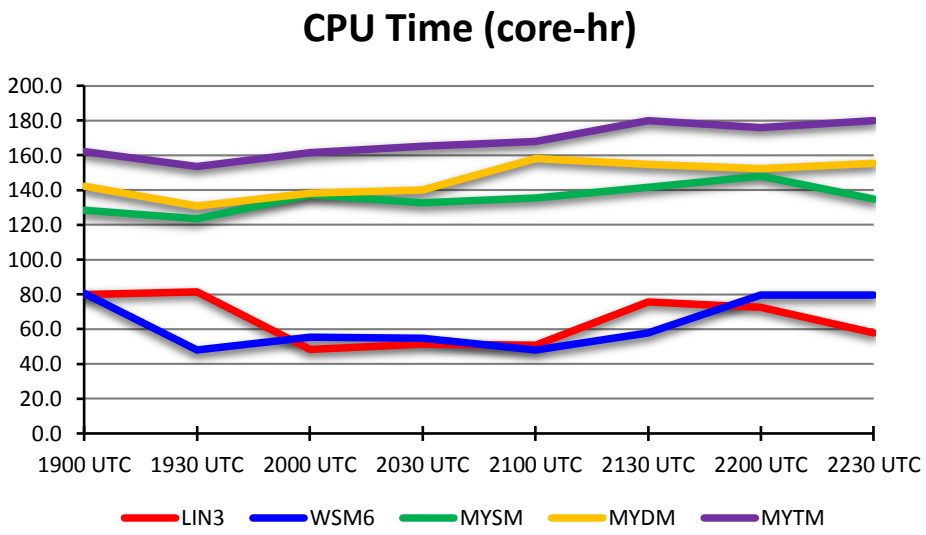

Figure 68. Line graph of the total CPU time (core-hr) used for each Control simulation. Note, for these experiments, 50 cores of Intel Xeon "Sandy Bridge" processors were used, so 100 core-hrs represents approximately 2 hours of wall-clock time. The forecasts are generally scalable by core. With current ARPS real-time system using 200 cores, LIN3 and WSM6 would have wall-clock times ~15–25 min for the 2-h Control runs, and the MY schemes would have wall-clock times ~35–55 min to complete the same 2-h forecast.

The assimilation of CASA radar data wasn't able to fully prevent the loss of forecast skill for S2 when KTLX radar data was withheld from the data assimilation

process. However, the assimilation of CASA radar data did have a larger positive impact than the assimilation of KTLX radar data for S2, so gap-filling radars, such as CASA radars, have the potential to improve forecasts of storms and low-level circulations when storms develop and/or propagate through a network of these relatively inexpensive, low-power radars. The Dallas/Ft. Worth Urban Testbed, which will eventually include eight CASA X-band radars, will continue to collect cases to further explore this potential benefit to forecasting severe convection (Brewster and Stratman, 2016).

Analysis cycling and the variable-dependent IAU technique each largely improved forecasts for both near-surface variables and S2 and S3's circulation tracks while maintaining the Control runs' success with S1, which supports two of the hypotheses. Combining the analysis cycling with the modified IAU technique resulted in similar or even larger forecast improvements, especially for near-surface moisture and wind components and S3's circulation locations. These findings support the theory that introducing temperature, wind, and water vapor increments prior to the microphysics increments allows for updrafts to strengthen before being loaded with hydrometeors, which can cause initial simulated storms to collapse.

Even though meaningful results were revealed in this study using just one case, the lack of knowledge of the statistical significance is a major caveat. To address this concern, additional ARPS 1-km simulations of tornadic supercell events, such as the 26 December 2015 tornadoes near Dallas, TX, using the same data assimilation techniques and model configurations need to be executed to fully understand the sensitivities tested in these experiments. In addition to continuing to examine the impact of assimilating CASA X-band radar data, research should also focus on exploring the sensitivities to

assimilating Terminal Doppler Weather Radar (TDWR) C-band data by comparing its impact relative to the assimilation of other radar data. Also, perhaps the assimilation of CASA radar data has a larger impact with smaller model grid spacings (e.g., 250 m), as Schenkman et al. (2011b) found, so this is something to test in future research studies, as well.

Although the five microphysics schemes used in this study provided a decent spread of solutions, other single-moment and multi-moment schemes (e.g., Thompson microphysics scheme, Thompson et al., 2008) currently not in the ARPS package should be added and tested alongside the current set of microphysics schemes. Other factors associated with analysis cycling and the IAU technique should also be explored. For example, the sensitivity to widening the IAU window from 5 min (as in this study) to 10 min (as in the real-time ARPS 1-km simulations) should be examined along with the impact of using different variable-dependent temporally-weighted distributions of increments.

The object-based verification technique utilized in this study to determine distance and timing errors proved to be highly effective in discerning the model's skill at forecasting circulation locations. However, perhaps other observed or analyzed objects, such as radar-detected rotation tracks (Lakshmanan et al., 2012), analysis locations of low-level vorticity or UH, or storm reports, can be used instead of the estimated tornado locations to streamline the verification process. Furthermore, the object-based verification technique can be employed to verify, for example, forecasted hail and graupel mixing ratios with radar-detected Z_{DR} columns (Kumjian and Ryzhkov, 2009). With respect to the neighborhood-based verification technique, most of the

simulations tended to substantially overforecast the composite reflectivity field, which is likely due to both the microphysics scheme and the relatively large 1-km grid spacing (i.e., coarser than reality), so other precipitation fields, such as precipitation rates, should possibly be used instead to determine "useful" skill by alleviating most of the overforecasting bias. However, an over-moistening in downdrafts in the cloud analysis scheme used in this work has recently been identified, and modifications to the cloud analysis scheme to address this issue have been made in the last couple weeks. These may reduce the reflectivity overforecast bias when the model is initialized with ongoing convection.

This study highlights the ability of a 1-km forecast system to successfully predict potentially tornadic supercells on short-time scales using analyses derived from 3DVAR and various data assimilation techniques, so when considering the computational cost of real-time data assimilation systems, less computationally intensive systems, such as the one used in this study, should be considered alongside more computationally expensive systems, such as EnKF. A study comparing both data assimilation systems is needed to better assess any potential benefits each data assimilation system may provide to short-term, storm-scale forecasts of severe convection.

References

Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert, 2009: Application of spatial verification methods to idealized and NWP gridded precipitation forecasts. *Wea. Forecasting*, **24**, 1485–1497.

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

Bluestein, H. B., and M. L. Weisman, 2000: The interaction of numerically simulated supercells initiated along lines. *Mon. Wea. Rev.*, **128**, 3128–3149.

Brewster, K. A. and D. R. Stratman, 2015: An updated high resolution hydrometeor analysis system using radar and other data. Preprints, *27th Conf. on WAF and 23rd Conf. on NWP*, Chicago, IL, Amer. Meteor. Soc., 2 July 2015.

Brewster, K. A., F. H. Carr, K. W. Thomas, and D. R. Stratman, 2015: Utilizing heterogeneous radar systems in a real-time high resolution analysis and short-term forecast system in the Dallas/Ft Worth Testbed. Preprints, *37th Conf. on Radar Meteor.*, Norman, OK, Amer. Meteor. Soc., 18 Sept. 2015.

Brewster, K. A. and D. R. Stratman, 2016: Tuning an analysis and incremental analysis updating assimilation for an efficient high resolution forecast system. Preprints, *20th Conf. on IAOS-AOLS*, New Orleans, LA, Amer. Meteor. Soc., 14 January 2016.

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

Brotzge, J. A., and C. M. Luttrell, 2015: Genesis of the Chickasha, Oklahoma, tornado on 24 May 2011 as observed by CASA radar and Oklahoma Mesonet. *J. Operational Meteor.*, 3 (6), 59–69.

Carley, J. R., B. R. J. Schwedler, M. E. Baldwin, R. J. Trapp, J. Kwiatkowski, J. Logsdon, and S. J. Weiss, 2011: A proposed model-based methodology for feature-specific prediction for high-impact weather. *Wea. Forecasting*, **26**, 243–248.

Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.*, **11**, 141–154.

Casati, B., 2010: New developments of the intensity-scale technique within the Spatial Verification Methods Intercomparison Project. *Wea. Forecasting*, **25**, 113–143.

Chou, M.-D., 2009: Parameterizations for the absorption of solar radiation by O2 and CO2 with application to climate studies. *J. Climate*, **3**, 209–217.

Chou, M.-D., 1992: A solar radiation model for use in climate studies. *J. Atmos. Sci.*, **49**, 762–772.

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

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

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

Davis, C., B. G. Brown, and R. Bulloks, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.

Dawson II, 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. Wea. Rev.*, **140**, 696–716.

Dawson II, D. T., M. Xue, J. A. Milbrandt, and A. Shapiro, 2015: Sensitivity of real-data simulations of the 3 May 1999 Oklahoma City tornadic supercell and associated tornadoes to multimoment microphysics. Part I: Storm- and tornado-scale numerical forecasts. *Mon. Wea. Rev.*, **143**, 2241–2265.

Dial, G. L., and J. P. Racy, 2004: Forecasting short term convective mode and evolution for severe storms initiated along synoptic boundaries. Preprints, *22nd Conf. on SLS*, Hyannis, MA, Amer. Meteor. Soc., October 2004.

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

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. Wea. Rev.*, **139**, 272–294.

Dowell, D. C., N. Yussouf, and C. K. Potvin, 2015: Sensitivity of storm-scale analyses and forecasts to radar-observation and model-grid resolution. *Presentation at 37th Conf. on Radar Meteorology*, Norman, OK, 18 September 2015.

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

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*, **24**, 1498–1510.

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

Fierro, A. O., E. R. Mansell, C. L. Ziegler, and D. R. MacGorman, 2012: Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. *Mon. Wea. Rev.*, **140**, 2609–2627.

French, M. M., P. S. Skinner, L. J. Wicker, and H. B. Bluestein, 2015: Documenting a rare tornado merger observed in the 24 May 2011 El Reno–Piedmont, Oklahoma, Supercell. *Mon. Wea. Rev.*, **143**, 3025–3043.

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 Tech.*, **21**, 457-469.

Gao, J., T. M. Smith, D. J. Stensrud, C. Fu, K. Calhoun, K. L. Manross, J. Brogden, V. Lakshmanan, Y. Wang, K. W. Thomas, K. Brewster, and M. Xue, 2013: A real-time weather-adaptive 3DVAR analysis system for severe weather detections and warnings. *Wea. Forecasting*, **28**, 727–745.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430.

Gilleland, E., J. Lindström, and F. Lindgren, 2010a: Analyzing the image warp forecast verification method on precipitation fields from the ICP. *Wea. Forecasting*, **25**, 1249–1262.

Gilleland, E., D. Ahijevych, B. G. Brown, and E. E. Ebert, 2010b: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.*, **91**, 1365–1373.

Heymsfield, G. M., L. Tian, L. Li, M. McLinden, and J. I. Cervantes, 2013: Airborne radar observations of severe hailstorms: Implications for future spaceborne radar. *J. Appl. Meteor. Climatol.*, **52**, 1851–1867.

Hitchens, N. M., M. E. Baldwin, and R. J. Trapp, 2012: An object-oriented characterization of extreme precipitation-producing convective systems in the midwestern United States. *Mon. Wea. Rev.*, **140**, 1356–1366.

Hoke, J. E., and R. A. Anthes, 1976: The initialization of numerical models by a dynamical initialization technique. *Mon. Wea. Rev.*, **104**, 1551–1556.

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

Houser, J. H., H. B. Bluestein, and J. C. Snyder, 2015: Rapid-scan, polarimetric, Doppler radar observations of tornadogenesis and tornado dissipation in a tornadic supercell: The "El Reno, Oklahoma" storm of 24 May 2011. *Mon. Wea. Rev.*, **143**, 2685–2710.

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

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

Johnson, A., X. Wang, F. Kong, and M. Xue, 2013: Object-based evaluation of the impact of horizontal grid spacing on convection-allowing forecasts. *Mon. Wea. Rev.*, **141**, 3413–3425.

Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108.

Jones, T. A., D. Stensrud, L. Wicker, P. Minnis, and R. Palikonda, 2015: Simultaneous radar and satellite data storm-scale assimilation using an ensemble Kalman filter approach for 24 May 2011. *Mon. Wea. Rev.*, **143**, 165–194.

Kain, J. S., S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. L. Weisman, K. K. Droegemeier, D. B. Weber, and K. W. Thomas, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952.

Kumjian, M. R., and A. V. Ryzhkov, 2009: Polarimetric signatures in supercell thunderstorms. *J. Appl. Meteor. Climatol.*, **47**, 1940–1961.

Lakshmanan, V., M. Miller, T. Smith, 2012: Quality control of accumulated fields by applying spatial and temporal constraints. *J. Atmos. Oceanic Technol.*, **30**, 745–758.

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

Markowski, P., C. Hannon, J. Frame, E. Lancaster, A. Pietrycha, R. Edwards, and R. L. Thompson, 2003: Characteristics of vertical wind profiles near supercells obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1262–1272.

McLaughlin, D., D. Pepyne, B. Philips, J. Kurose, M. Zink, E. Knapp, D. Westbrook, E. Lyons, A. Hopf, A. DeFonzo, R. Contreras, T. Djaferis, E. Insanic, S. Frasier, V. Chandrasekar, F. Junyent, N. Bharadwaj, Y. Liu, and Y. Wang, K. Droegemeier, M. Xue, J. Brotzge, F. Carr, K. Kloesel, K. Brewster, S. Cruz-Pol, and K. Hondl, 2009: Short-wavelength technology and the potential for distributed networks of small radar systems. *Bull. Amer. Meteor. Soc.*, **90**, 1797-1817.

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

Milbrandt, J. A. and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62**, 3065–3081.

Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354.

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

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

Potvin, C. K., and M. L. Flora, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for Warn-on-Forecast. *Mon. Wea. Rev.*, **143**, 2998–3024.

Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2014: The analysis and prediction of microphysical states and polarimetric radar variables in a mesoscale convective system using double-moment microphysics, multinetwork radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 141–162.

Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164.

Roberts, N. M., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169.

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

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

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011b: Impact of CASA radar and oklahoma mesonet data assimilation on the analysis and prediction of tornadic mesovortices in an MCS. *Mon. Wea. Rev.*, **139**, 3422–3445.

Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, and M. C. Coniglio, 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.

Shapiro, A., S. Rahimi, C. K. Potvin, and L. Orf, 2015: On the use of advection correction in trajectory calculations. *J. Atmos. Sci.*, **72**, 4261–4280.

Sobash, R. A., and D. J. Stensrud, 2015: Assimilating surface mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 May 2012. *Mon. Wea. Rev.*, **143**, 3700–3725.

Snook, N., M. Xue, and Y. Jung, 2011: Analysis of a tornadic mesoscale convective vortex based on ensemble Kalman filter assimilation of CASA X-band and WSR-88D radar data. *Mon. Wea. Rev.*, **139**, 3446–3468.

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

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

Stensrud, D. J., L. J. Wicker, K. E. Kelleher, M. Xue, M. P. Foster, J. T. Schaefer, R. S. Schneider, S. G. Benjamin, S. S. Weygandt, J. T. Ferree, and J. P. Tuell, 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, D. J., L. J. Wicker, M. Xue, D. T. Dawson, N. Yussouf, D. M. Wheatley, T. E. Thompson, N. A. Snook, T. M. Smith, A. D. Schenkman, C. K. Potvin, E. R. Mansell, T. Lei, K. M. Kuhlman, Y. Jung, T. A. Jones, J. Gao, M. C. Coniglio, H. E. Brooks, K. A. Brewster, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2-16.

Stratman, D. R., M. C. Coniglio, S. E. Koch, and M. Xue, 2013: Use of multiple verification methods to evaluate forecasts of convection from hot- and cold-start convection-allowing models. *Wea. Forecasting*, **28**, 119–138.

Sun, W.-Y., and C.-Z. Chang, 1986: Diffusion model for a convective layer. Part I: Numerical simulation of convective boundary layer. *J. Climate App. Meteorol.*, **25**, 1445–1453.

Tanamachi, R. L., P. L. Heinselman, and L. J. Wicker, 2015: Impacts of a storm merger on the 24 May 2011 El Reno, Oklahoma, tornadic supercell. *Wea. Forecasting*, **30**, 501–524.

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

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

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

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman filter analyses and forecasts of a severe mesoscale convective system using different choices of microphysics schemes. *Mon. Wea. Rev.*, **142**, 3243–3263.

Xu, Q., L. Wei, and K. Nai, 2015: Analyzing vortex winds in radar-observed tornadic mesocyclones for nowcast applications. *Wea. Forecasting*, **30**, 1140–1157.

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic 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 nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Physics*, **76**, 134-165.

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

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

Xue, M., J. Schleif, F. Kong, K. W. Thomas, Y. Wang, and K. Zhu, 2013: Track and intensity forecasting of hurricanes: Impact of convection-permitting resolution and global ensemble Kalman filter analysis on 2010 Atlantic season forecasts. *Wea. Forecasting*, **28**, 1366–1384.

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.

Yussouf, N., D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in alabama. *Mon. Wea. Rev.*, **143**, 3044–3066.

Zalesak, S. T., 1979: Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.*, **31**, 335–362.

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, S. Van Cooten, K. Kelleher, D. Kitzmiller, F. Ding, D. J. Seo, E. Wells, and C. Dempsey, 2011: National Mosaic and Multi-sensor QPE (NMQ) system:–Description, results and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338.