UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ASSIMILATION OF ATTENUATED DATA FROM X-BAND NETWORK RADARS USING ENSEMBLE KALMAN FILTER

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By
JING CHENG
Norman, Oklahoma
2013

ASSIMILATION OF ATTENUATED DATA FROM X-BAND NETWORK RADARS USING ENSEMBLE KALMAN FILTER

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

Dr. Ming Xue, Chair

Dr. Frederick Carr

Dr. Lance Leslie

Dr. Guifu Zhang

Dr. Yan Zhang

Acknowledgements

I would like to sincerely thank my advisor, Dr. Ming Xue, for his supervision, patience, and support he provided to me over years. His profound insight and sophisticated knowledge guided me to solve many scientific problems. His untiring passion and enthusiasm drove me to pursue my academic goals. He inspired me to develop independent and critical thinking during my graduate studies. Moreover, his kindness and hospitality made my life in US joyful.

Though I benefitted immensely from the wisdom of numerous professors and mentors throughout my graduate studies, there are a few individuals who stand out among the rest to whom I wish to express my gratitude. I wish to thank Dr. Frederick Carr and Dr. Guifu Zhang. Their recommendations and suggestions have been invaluable to my work. Thank Dr. Lance Leslie to provide as much help as he can provide even when he had healthy issues. Hope everything goes well with him. Thank Dr. Brian Fiedler to stand up for me every time when I needed petitions from Graduate College.

I owe a great deal of thanks to my colleagues in CAPS. I wish to thank Dr. Youngsun Jung, who offered countless insights into the inner-working of the ARPS-EnKF system, provided many valuable suggestions to my research topic, and helped me resolve technical problems; and Dr. Alex Schenkman, who hold CASA research dinner for research discussions many times to broaden my vision, and helped to greatly improve my dissertation and journal articles, as well as led me to great food in many places; and Dr. Nathan Snook, who kindly share his python scripts for post-processing the experiment results, as well as helped to review my dissertation and lots of valuable suggestions.

I extend my sincere thanks to the staff at the School of Meteorology and CAPS – Celia Jones, Marcia Pallutto, Eileen Hasselwander, and Debra Farmer — for all of their help that resolved every puzzle I encountered during PhD studying and made my student life in OU simple. I also thank Yunheng Wang and Scott Hill for helping me with the ARPS model and computers.

My deepest gratitude goes to my parents, Cheng Wanshou and Jing jianyun, for their unconditionally love and faith in every decision I made. I also wish to thank my sister, Cheng Ying, and brother-in-law, Lv Gang, for taking care of my parents during these years and standing up for me when I needed help.

I want to thank my good friends for their support throughout my time at OU. I thank Jili Dong for his help in study and living when I first came to US. I also thank Dr. Dan Dawson and Dr. Robin Tanamachi to take me to chase tornado for many times. I would like to acknowledge other colleagues, Yujie, Mingjun, Shizhang, Wei, Ling, Qiwei, Xiaobin and many other friends for sharing their learning and encouraging each other and keeping me sane.

I greatly value my friendship with Shanshan Wu, who is the best mentor and great supporter in my life, and always bear with my bad temper. I deeply appreciate her helping me when I went through the toughest time in my life and I value every piece of time with her. I also wish to thank Yanhua Teng, the best listener so far in my life, who always was there when I wanted to talk. I also want to acknowledge my friends Ying Lu

and Xiong Ming, for feeding me many times with wonderful food while I was busy preparing graduation.

Financial support for the research contained within this dissertation was provided by the Engineering Research Center (ERC) for Collaborative Adaptive Sensing of the Atmosphere (CASA) through NSF grant EEC-0313747. Supercomputing resources were provided by the Center for Analysis and Prediction of Storms, the OU Supercomputing Center for Research and Education (OSCER), and the National Institute for Computational Sciences (NICS) on their Kraken supercomputer.

Table of Contents

Acknowledgements iv

List of Tables ix

List of Figures x

Abstract xviii

Chapter 1: Introduction 1

1.1 Background and Motivations 1

1.2 Dissertation Overview 3

Chapter 2: Historical Review and Description of Techniques 6

2.1 Overview of traditional attenuation algorithms 6

2.2 Overview of Simultaneous attenuation correction and state estimation 11

Chapter 3: OSSEs Assimilating Attenuated Data from an X-band Radar Network for a Squall Line and testing with multi-error sources observation 19

3.1 The truth simulation 19

3.2 Simulation of radar observations 22

3.3 Data assimilation experiments 24

3.4 Results of experiments 27

3.6 Summary and discussion 52

Chapter 4: Optimal Design of Radar Networks: Expansion of CASA IP1 Network and Impacts on Squall line Analysis and Forecast 57

4.1 The truth simulation 57

4.2 Simulation of radar observations 59

4.3 Data assimilation experiments 60

4.4 Assimilation Results 64

4.5 Forecast Results 74

4.6 Summary and Discussion 84

Chapter 5: Application of Attenuation-Correcting EnKF to Real CASA X-band Radar Data 88

5.1 Case background 89

5.2 Experiment design and data 92

5.3 Assimilation results 102

5.4 Forecast results 110

5.5 Summary and Discussion 134

Chapter 6: Summary and Future Work 139

6.1 General summary 139

6.2 Concluding remarks and future work 144

References 148

Appendix A: Calculation of Z_e-W and k-W relations 159

A.1 Rainwater 159

A.2 Dry snow and hail 159

A.3 Melting snow and hail 160

List of Tables

Table 3. 1 Lists of perfect model OSS experiments 26

Table 3. 2 Lists of imperfect model OSS experiments 27

Table 3. 3 List of experiments in multi-error sources experiments 50

Table 4. 1 List of radar network names and corresponding radars 63

Table 5. 1 Lists of experiments 95

Table 5. 2 List of radars used in data assimilation experiments 99

Table 5. 3 Average relative humidity observed and forecasted from 2200 UTC to 2300 UTC 130

List of Figures

FIG 3. 1 (a) Map of 288 km × 192 km computational domain. The small solid circles represent the maximum 40 km range of the CASA IP1 radars. (b) Environmental soundings used to generate the truth simulation of a squall line. 21

FIG 3. 2 Simulated (a) non-attenuated and (b) attenuated reflectivity observations at 5 ° elevation at 135 min of the squall line simulation time. 23

FIG 3. 3 The horizontal wind vectors (m/s, plotted every third grid point) and computed reflectivity (shaded at 5 dB interval, starting from 5 dBZ) for (a) Truth simulation; and experiments (b) NAC; (c) NAC10Z ; (d) VRONLY; (e) ATCZ; (f) ATC at model time 135 mins 30

FIG 3. 4 The ensemble mean analysis RMSEs averaged over points inside radar coverage for (a) u, (b) v, (c) w, (d) perturbation potential temperature θ', (e) q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h for experiments NAC (dotted), NAC10Z (dashed), VRONLY(thin solid), and ATC(thick solid) 31

FIG 3. 5 Same as FIG 3. 3 but for ensemble mean forecast (left panels) and ensemble mean analysis (right panels) from ATC at model times 130 min, 135 min and 140 min 33

FIG 3. 6 Empirical relations between observation error and radar reflectivity. Green (solid), red (dashed), blue (dotted) lines represent the observation error reflectivity relation 1, 2, 3 respectively. Black solid line indicate the observation error used by ATC, ATCZ.. 36

FIG 3. 7 The ensemble mean analysis RMSEs averaged over points inside radar coverage for (a) w, (b) q_c, (c) q_r, (d) q_i, (e0q_s, (f)q_h for experiments OBSE1 (reflectivity relation 1, dotted), OBSE(reflectivity relation 2, solid), OBSE3 (reflectivity relation 3, dashed) from 125 min to 170 min. 37

FIG 3. 8 Corresponding observation error deviation (using obserr-reflectivity relation 2) for radar (a) KCYR (b) KSAO (c) KLWE (d) KRSP at 4th tilt at model time 135 min. 38

FIG 3. 9 Same as FIG 3. 3 but for Truth (upper panel), ATC (middle panel) and OBSE (lower panel) from model time 15 min to 145 min 40

FIG 3. 10 As FIG 3. 3, but for experiments ATCZ (upper panel) and OBSEZ (lower panel) 41

FIG 3. 11 As in FIG 3. 4, but for experiments ATC (thin solid), ATCZ (dotted), OBSE(thick solid), OBSEZ (dashed) 42

FIG 3. 12 As in FIG 3. 4, but for experiments ATC (dashed), VRONLY(thin solid), OBSE (thick solid) and OBSEONLY (dotted)... 43

FIG 3. 13 As in FIG 3. 4, but for experiments HIG_ORG(dashed), HIG_OBSE(thick solid), LOW_ORG(dotted), LOW_OBSE(thin solid) 45

FIG 3. 14 Reflectivity model error as a function of reflectivity original designed (blue) and used in experiments (red).. 47

FIG 3. 15 Attenuation model error as a function of Path-Integral-Attenuation (PIA) ... 48

FIG 3. 16 Signal-to-noise ratio related error as a function of signal-to-noise ratio original designed (blue) and used in experiments (red) 48

FIG 3. 17 Simulated radar observations on elevation 1.0 ° at 120 mins model time. Character ‘R’ indicates the location of radar.. 50

FIG 3. 18 The ensemble mean analysis RMSEs averaged over points where true Z is greater than 10 dBZ for (a) u, (b) v, (c) w, (d) perturbation potential temperature θ', (e)

q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h for experiments ATTC (gray), ATTC_AEM (blue), ATTC_MSEM(red), and ATTC_LGERR(green) 52

FIG 4. 1 Map of the computational domain. The interior black rectangle denotes the domain over which quantitative verification statistics are calculated. The diamonds are locations of CASA IP1 radars and the small squares represent locations of proposed additional CASA radars. The circles represent the 40-km-range rings of CASA radars 58

FIG 4. 2 The composition of radar networks of (a) EWEX, (b) EWRI, (c) EWUN, (d) RILR, (e) RIUN and (f) LRUN. The diamonds are locations of CASA IP1 radars and the small squares represent locations of proposed additional CASA radars. The circles represent the 40-km-range rings of CASA radars. 62

FIG 4. 3 Time line of assimilation and forecast for (a) slow moving squall-line experiments and (b) fast moving squall-line experiments. (Note that the starting times of two sets of experiments are different). 64

FIG 4. 4 Radar reflectivity (shaded; dBZ) of slow moving squall-line simulation at (a) T=9900s (165 min); (b) T=10800s (180 min) and fast moving squall-line simulation at (c) T=8100s (135 min); (d) T=9000s (150 min) at 500 m AGL 66

FIG 4. 5 The RMS errors of the ensemble-mean analysis, averaged over points at which the reflectivity is greater than 10 dBZ inside verification domain for (a) u, (b) v, (c) w, and (d) θ', (e) q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h, for experiments slwEWEX (gray), slwEWRI (blue), slwEWUN (red), slwRILR (green), slwLRUN (light purple), slwRIUN (sky). Units are shown in the plots.. 67

FIG 4. 6 Difference of reflectivity between truth and (a) slwEWEX, (b) slwEWRI, (c) slwEWUN, (d) slwRILR, (e) slwRIUN, (f) slwLRUN at 165 min on level z=500m 69

FIG 4. 7 As in FIG 4. 5, but for fstEWEX (gray), fstEWRI (blue), fstEWUN (red), fstRILR (green), fstLRUN (light purple), fstRIUN (sky). 72

FIG 4. 8 Difference of Reflectivities between truth and (a) fstEWEX, (b) fstEWRI, (c) fstEWUN, (d) fstRILR, (e) fstRIUN, (f) fstLRUN at 165 min on level z=500m 73

FIG 4. 9 The RMS errors of forecasts averaged over the verification domain for (a) u, (b) v, (c) w, and (d) θ', (e) q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h. The forecasts begin from ensemble-mean analysis at t=180 min (15 cycles) of experiments slwEWEX (gray), slwEWRI (blue), slwEWUN (red), slwRILR (green), slwLRUN (light purple), slwRIUN (sky).. 75

FIG 4. 10 As in FIG 4. 9 but the forecasts begin from ensemble mean analysis at t=165 min (12 cycles) 76

FIG 4. 11 As in FIG 4. 10 but the forecasts begin from ensemble-mean analysis at t = 150 min (15 cycles) of experiments fstEWEX (gray), fstEWRI (blue), fstEWUN (red), fstRILR (green), fstLRUN (light purple), fstRIUN (sky) 78

FIG 4. 12 As in FIG 4. 11 but the forecasts begin from ensemble-mean analysis at t = 135 min (12 cycles) 79

FIG 4. 13 ETS computed over the verification domain for reflectivity threshold at (a) 15dBZ, (b) 25 dBZ, (c) 35 dBZ and (d) 45 dBZ. The forecasts begin from ensemble-mean analysis at t=165 min (12 cycles) of experiments slwEWEX (black), slwEWRI (blue), slwEWUN (red), slwRILR (green), slwLRUN (light purple), slwRIUN (sky).. 81

FIG 4. 14 As in FIG 4. 13 but the forecasts begin from ensemble-mean analysis at t=180 min (15 cycles) 82

FIG 4. 15 As in FIG 4. 13 but the forecasts begin from ensemble-mean analysis at t=135 min (12 cycles) of experiments fstEWEX (black), fstEWRI (blue), fstEWUN (red), fstRILR (green), fstLRUN (light purple), fstRIUN (sky blue) 83

FIG 4. 16 As in FIG 4. 15 but the forecasts begin from ensemble-mean analysis at t=150 min (15 cycles) 84

FIG 5. 1 Map of tornado tracks over Central Oklahoma from the National Weather Service with tornado ratings shown below each reported tornado (Tornado tracks are marked with red lines; Yellow shaded denote the metropolitan area) (http://www.srh.noaa.gov/images/oun/wxevents/20110524/maps/overview-800.jpg) .. 89

FIG 5. 2 Satellite image of the GOES-13 0.63 μm visible channel at 18:32 UT on 24 May 2011 from CIMSS 90

FIG 5. 3 Sketch of the mesoscale discussion analysis (#925) at 1718 UTC from the Storm Prediction Center showing various key environmental factors of the 24th May 2011 tornado outbreak 91

FIG 5. 4 Timeline of the analysis and forecast periods 94

FIG 5. 5 Map of the computational domain. The interior rectangle denotes the domain where the CASA radar data are assimilated. The small squares represent locations of CASA radars and WSR-88D radars. Small circles represent CASA radar 40 km range, while large circles represent WSR-88D radar 60 km range. 96

FIG 5. 6 Analytical relations between observation error deviation and radar reflectivity 97
FIG 5. 7 Observed (a) CASA pre-corrected reflectivity on 2 °; (b) CASA un-corrected reflectivity on 2 °; (c) KTLX reflectivity on 0.92 ° and KFDR reflectivity on 0.53 °. Light blue colors indicate the clear air echoes 101
FIG 5. 8 (a) The horizontal wind vectors (m/s), and computed reflectivity from outer domain analyses, and interpolated reflectivity from (b) KTLX and (c) KFDR at 2000 UTC May 24, 2012 103
FIG 5. 9 RMS ensemble spread of *u* for 40 ensemble members at 2100 UTC 24 May 2011 on outer domain 104
FIG 5. 10 RMS ensemble spread of *v* for 40 ensemble members at 2100 UTC 24 May 2011 on outer domain 104
FIG 5. 11 RMS ensembles spread of potential temperature θ of 40 ensemble members at 2100 UTC 24 May 2011 on outer domain 105
FIG 5. 12 (a) Observed reflectivity of KTLX interpolated to 2 km MSL; (b) calculated reflectivity in (b) CTRL; (c) ATTC; (d) PreAC_20DBZ; (e) PreAC_OEM on 2 km MSL at 2135 UTC 24 May 2011 108
FIG 5. 13 Calculated reflectivity in (a) CTRL; (b) ATTC; (c) PreAC_20DBZ and (d) PreAC_OEM on 500 m MSL at 2200 UTC 24 May 2011 110
FIG 5. 14 (a) Observed reflectivity of KTLX interpolated to 2 km MSL; the horizontal wind vectors (ms^{-1}), calculated reflectivity in (a) ATTC; (b) PreAC_20DBZ; (c) PreAC_OEM on 2 km MSL at 2210 UTC 24 May 2011 112
FIG 5. 15 Same as FIG 5. 14 but at 2220 UTC 24 May 2011 114

FIG 5. 16 Same as FIG 5. 14 but at 2230 UTC 24 May 2011 116

FIG 5. 17 Same as in FIG 5. 14 but at 2240 UTC 24 May 2011 117

FIG 5. 18 Same as FIG 5. 14 but at 2250 UTC 24 May 2011 118

FIG 5. 19 Same as FIG 5. 14 but at 2300 UTC 24 May 2011 120

FIG 5. 20 Deviation of 2 m temperature forecasted from surface observation (OK Mesonet) for (a) CTRL, (b) ATTC, (c) PreAC_20DBZ and (d) PreAC_OEM at 2230 UTC 122

FIG 5. 21 Same as in FIG 5. 20 but at 2300 UTC 24 May 2011 123

FIG 5. 22 Meteogram of temperature (C) observed (black) and simulated (blue for CTRL, red for PreAC_20DBZ, green for PreAC_OEM and purple for ATTC) at Minco mesonet sites from 2200 UTC -2300 UTC 24 May 2011. 124

FIG 5. 23 Same as in FIG 5. 22 but at Norman mesonet site 125

FIG 5. 24 Deviation of 2 m relative humidity forecasted from surface observation (OK Mesonet) for (a) CTRL, (b) ATTC, (c) PreAC_20DBZ and (d) PreAC_OEM at 2230 UTC 126

FIG 5. 25 Same as in FIG 5. 24 but at 2300 UTC 24 May 2011 127

FIG 5. 26 Meteogram of relative humidity (percent) observed (black) and simulated (blue for CTRL, red for PreAC_20DBZ, green for PreAC_OEM and purple for ATTC) at Minco mesonet site from 2200 UTC -2300 UTC 24 May 2011. 129

FIG 5. 27 Same as in FIG 5. 26 but at Norman Mesonet site 130

FIG 5. 28 Ensemble frequency count when vorticity > 0.012 s^{-1} in (a) CTRL, (b) ATTC, (c) PreAC_20DBZ and (d) PreAC_OEM from 2200 UTC to 2300 UTC. Red line in the north indicates the observed track of tornado C1, while the one in the south indicates the observed track of tornado D1. Circles represent the 40 km range of CASA radars..... 133

FIG 5. 29 Same as in FIG 5. 28 but when vorticity > 0.008 s^{-1}.................................... 134

Abstract

To use reflectivity data from X-band radars for quantitative precipitation estimation and storm-scale data assimilation, the effect of attenuation must be properly accounted for. Traditional approaches try to make correction to the attenuated reflectivity first before using the data. An alternative, theoretically more attractive approach builds the attenuation effect into the reflectivity observation operator of a data assimilation system, such as an ensemble Kalman filter (EnKF), allowing direct assimilation of the attenuated reflectivity and taking advantage of microphysical state estimation using EnKF methods for a potentially more accurate solution.

This study first tests the approach for the CASA (Center for Collaborative Adaptive Sensing of the Atmosphere) X-band radar network configuration through observing system simulation experiments (OSSE) for a quasi-linear convective system (QLCS) that has more significant attenuation than isolated storms. To avoid the problem of potentially giving too much weight to fully attenuated reflectivity, an analytical, echo-intensity-dependent model for the observation error (AEM) is developed and is found to improve the performance of the filter. By building the attenuation into the forward observation operator and combining it with the application of AEM, the assimilation of attenuated CASA observations is able to produce a reasonably accurate analysis of the QLCS inside CASA radar network coverage. Compared with foregoing assimilation of radar data with weak radar reflectivity or assimilating only radial velocity data, our method can suppress the growth of spurious echoes while obtaining a more accurate analysis in the terms of root-mean-square (RMS) error. Sensitivity experiments are designed to examine the effectiveness of AEM

by introducing multiple sources of observation errors into the simulated observations. The performance of such an approach in the presence of resolution-induced model error is also evaluated and good results are obtained.

The same EnKF framework with attenuation correction is used to test different possible configurations of 2 hypothetical radars added to the existing network of 4 CASA radars through OSSEs. Though plans to expand the CASA radar network did not materialize, such experiments can provide guidance in the site selection of future X-band or other short-wavelength radar networks, as well as examining the benefit of X-band radar networks that consist of a much larger number of radars. Two QLCSs with different propagation speeds are generated and serve as the truth for our OSSEs. Assimilation and forecast results are compared among the OSSEs, assimilating only X-band or short-wavelength radar data. Overall, radar networks with larger downstream spatial coverage tend to provide overall the best analyses and 1-hour forecasts. The best analyses and forecasts of convective scale structure, however, are obtained when Dual- or Multi-Doppler coverage is preferred, even at the expense of minor loss in spatial coverage.

Built-in attenuation correction is then applied, for the first time, to a real case (the 24 May 2011 tornadic storm near Chickasha, Oklahoma), using data from the X-band CASA radars. The attenuation correction procedure is found to be very effective—the analyses obtained using attenuated data are better than those obtained using pre-corrected data when all the values of reflectivity observations are assimilated. The effectiveness of the procedure is further examined by comparing the deterministic and ensemble forecasts started from the analysis of each experiment. The deterministic

forecast experiment results indicate that assimilating un-corrected observations directly actually retains some information that might be lost in the pre-corrected CASA observations by forecasting a longer-lasting trailing line, similar to that observed in WSR-88D data. In the ensemble forecasts, assimilating un-corrected observations directly, using our attenuation-correcting EnKF, results in a forecast with a more intense tornado track than the experiment that assimilates all values of pre-corrected CASA data.

This work is the first to assimilate attenuated observations from a radar network in OSSEs, as well as the first attempt to directly assimilate real, uncorrected CASA data into a numerical weather prediction (NWP) model using EnKF.

Chapter 1: Introduction

1.1 Background and Motivations

In many areas over the continental United States, the current operational Weather Surveillance Radar-1988 Doppler (WSR-88D) radar network lacks an effective means of providing dense, comprehensive, lower-troposphere observations. To address this shortcoming, the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA;McLaughlin et al. 2009) has developed a network low-cost, high-spatial-density, adaptively-scanning X-band dual-polarization Doppler radars. The Doppler and polarimetric capabilities of the CASA network, operating at high spatial and temporal resolutions, can be used to detect, track, analyze, and predict tornadoes or processes leading to tornadogenesis. CASA radars are designed and deployed to operate as a network, providing dense, overlapping coverage to overcome the horizon problem common to large, long-range radars. This network approach to scanning, known as distributed collaborative adaptive sensing (DCAS) (McLaughlin et al. 2009), optimizes the low-level volume coverage scanning and maximizes the utility of each scanning cycle for users. The data from the first CASA test-bed, known as Integrated Project One (IP1, active from 2006 to 2011), are used to drive real-time surface analyses, nowcasting, and dynamic numerical weather prediction (NWP) models (Brotzge et al. 2010).

Many studies assimilating Doppler radar observations into NWP models have shown reasonable success in analyzing and forecasting convective storms (Sun et al. 1998; Weygandt et al. 2002; Xue et al. 2003; Xiao et al. 2005; Schenkman et al. 2010; Schenkman et al. 2011; Snook et al. 2011). Compared to the 3-dimensional variational

(3DVAR) and 4-dimensional variational (4DVAR) methods, the ensemble Kalman filter (EnKF) has the notable advantage of incorporating flow-dependent error covariance information (Evensen 2003), as well as providing a set of ensemble member analyses suitable for initializing ensemble forecasts. For convection-resolving NWP, the microphysics scheme is one of the most important physical processes and has a profound impact on the forecast. Considering the complexity and highly nonlinear nature of microphysical processes, the linearization required by 4DVAR in the minimization process often encounters difficulties, particularly when ice species are involved (Xu 1996). However, the EnKF method, which uses the full nonlinear model to propagate the ensemble state, appears to be more attractive (Tong and Xue 2005).

The impact of attenuation due to precipitation poses an additional challenge for accurate quantitative and qualitative interpretation of shorter wavelength X-band radar data. To successfully use reflectivity observations from X-band radars for quantitative precipitation estimation and storm-scale data assimilation, the effect of attenuation must be properly accounted for. As a result, attenuation correction is a significant area of research in utilizing reflectivity observations from X-band and other shorter-wavelength radars. Numerous attenuation algorithms have been studied in the observation space, and results are promising (Bringi et al. 2001; Gorgucci and Chandrasekar 2005; Park et al. 2005a; Park et al. 2005b). Although attenuation can be estimated, and its effects on reflectivity and differential reflectivity fields can be compensated for, little can be done in cases of complete attenuation where the signal drops to near the noise floor (Snyder et al. 2010).

More recently, there have been efforts to include attenuation in the forward observation operator of the data assimilation system. Attenuation correction can be achieved using a variational approach by estimating the attenuation at each range gate and iterating this process until an accurate estimate of attenuation is reached (Hogan 2007); the expected attenuation can also be calculated using the estimated atmospheric state, which can be obtained through ensemble-base data assimilation (Xue et al. 2009; XTZ09 hereafter). The latter approach, as proposed by XTZ09, does not require any prior assumption about the specific hydrometeor types at particular grid points, and it is possible to include error or uncertainty from all data sources in the assimilation framework, as well as allowing for close coupling of attenuation correction with the dynamical model. Such a procedure was demonstrated to be effective in XTZ09 using a set of observing system simulation experiments (OSSEs), in which simulated radar observations of a supercell were collected using a single simulated radar. This work establishes a new direction in dealing with attenuation correction during radar reflectivity data assimilation.

1.2 Dissertation Overview

The primary objective of this work is to test the effectiveness of, and further improve, attenuation correction built into an EnKF system, in particular for more complex convective storm systems and for observations from CASA-like X-band radar networks. For a thorough evaluation, the capability of the EnKF system will be tested in OSSEs by assimilating simulated radar observations for QLCSs and in real data experiments by assimilating observations collected from the CASA IP1 network during the 24 May 2011 tornado outbreak happened in southwest Oklahoma. The results of analyses and subsequent deterministic and ensemble forecasts will be evaluated.

Chapter 2 contains a brief overview of traditional attenuation algorithms and the current attenuation correction methods used by CASA radar network. A detailed description of the attenuation-containing forward observation operator used in the research as well as an observation error model (OEM) that is designed to improve the performance of the EnKF system are included, too.

In Chapter 3 we describe the truth simulation, observation simulation, and data assimilation experiment configurations of the OSSEs. As this is the first study that directly assimilates attenuated observations from a radar network for a linear convective system, OSSEs are preferred to test the new methodology (Lord et al. 1997). The results from perfect and imperfect model experiments are discussed. In the latter, model resolution errors are introduced into the OSSEs to examine the robustness of the attenuation correction procedure in the presence of model error. The situation where multiple error sources present in observations is also examined. Another type of observation error model is proposed and examined, too.

The same EnKF framework with attenuation correction is used in OSSEs to test the possible configurations of an expanded CASA IP1 radar network in Chapter 4. CASA had previously planned to increase the size of the radar network, positioning the added radars in a way that would cover existing blind spots for the most common storm modes, especially for cases where severe attenuation happened (Brewster 2005). Recently, the CASA Dallas Fort Worth (DFW) Urban Demonstration Network also has plan to expand the existing DFW Test-bed (Philips 2012). OSSEs can be an effective way to evaluate trade-offs in the design of observing systems and provide the quantitative basis for a rational design of observing systems that will be used primarily for numerical weather prediction (Lord et al. 1997). In our OSSEs, the impact of storm motion speed is considered by designing experiments with slow-moving and fast-moving convective lines. The main focus is on evaluating the accuracy of analyses and predictions assimilating data collected by radars with different site arrangements. These OSSEs can provide guidance to future X-band radar network site selection for the purpose of data assimilation or convective storm NWP.

Tests of the attenuation correction procedure using real observations from the CASA IP1 network are discussed in Chapter 5. A brief case description of the 24 May 2011 tornado outbreak is presented. The experiment set-ups are then introduced, and preliminary results are discussed. Comparisons are made between the experiment assimilating pre-corrected observations using the original EnKF algorithm (without built-in attenuation correction) and one assimilating un-corrected observations using our attenuation-correcting EnKF system.

A summary of the dissertation and outlines of future work are presented in Chapter 6.

Chapter 2: Historical Review and Description of Techniques

2.1 Overview of traditional attenuation algorithms

2.1.1. Existing attenuation algorithms

Weather radar is one of few remote sensing platforms that can provide high spatial and temporal resolution measurements of precipitation in weather systems, including convective storms. By sending directional microwave pulses with wavelengths of 1 to 10 cm (approximately ten times the diameters of the droplets of liquid, ice crystals, snow and graupel particles of interest), Rayleigh or Mie scattering occurs and part of the energy of each pulse bounces off of these particles, back in the direction of the radar station. Theoretically, shorter wavelengths with superior angular resolution are useful for smaller particles. Shorter wavelength electromagnetic energy, however, is more strongly absorbed by water or ice droplets. For example, 3 cm wavelength X-band radars suffer echo power loss up to100 times larger than that suffered by the 10 cm S-band radars of the WSR-88D radar network (Doviak and Zrnic 1993). Significant errors can be introduced during quantitative and qualitative interpretation of X-band radar data if the effects of attenuation are not considered and corrected.

An electromagnetic wave suffers power loss both from energy absorption and scatter. Each hydrometeor absorbs an amount of power P_L from the incident power with power density S_i, that can be expressed as

$$P_L = (\sigma_a+\sigma_s)\, S_i\,, \qquad (1)$$

where σ_a is the absorption cross section, an apparent area that intercepts from the incident radiation a power equal to the power dissipated as heat in the drop, and σ_s is the total scatter cross section.

If we use the Born approximation, which neglects the scattering of the scattered field, drops within a volume element $\Delta V(r)$ do not significantly alter S_i within this volume. The power density change ΔS_i in a wave propagating a short distance Δr through the volume is

$$\Delta S_i = -\frac{\Delta r}{\Delta V}\sum_{n=1}^{N}(\sigma_{an}+\sigma_{sn})S_i \ , \qquad (2)$$

where the negative sign signifies loss, the summation extends over all N drops within ΔV, and σ_{an} and σ_{sn} are the absorption and scattering cross section, respectively, of the nth particle. In the limit $\Delta r \to 0$, the rate of change in power density is then

$$\lim_{\Delta r\to 0}(\frac{\Delta S_i}{\Delta r}) = \frac{dS_i}{dr} = -kS_i \ , \qquad (3)$$

and the power density at any range r is the integral solution of Eq.(3),

$$S_i(r_2) = S_i(r_1)\exp(-\int_{r_1}^{r_2} kdr) \ , \qquad (4a)$$

where

$$k = \lim_{\Delta r\to 0}\sum\frac{\sigma_{an}+\sigma_{sn}}{\Delta V} = \int_0^{\infty} N(D,\mathbf{r})\sigma_e(D)dD \qquad (4b)$$

is the specific attenuation, or the attenuation coefficient. $N(D, \mathbf{r})$, is the drop size distribution (DSD), which is the expected number density of hydrometeors per unit diameter. The product $N(D, \mathbf{r})d$D gives the number of hydrometeors per unit volume having diameters in the interval dD about D. The specific attenuation expressed in decibels per kilometer is

$$K \equiv \frac{d}{dr_2}[10\log\frac{S(r_1)}{S(r_2)}] = 4.34\times10^3 k \quad \text{dB km}^{-1} \qquad (5)$$

where k has units of m^{-1}.

It has been observed, for a wide range of rainfall rates and rain types, that a consistent relationship exists between specific attenuation of microwaves and the rainfall rate measured with rain gauges along the propagation path. Burrows and Attwood (1949) have used the drop size data of Laws and Parsons (1943) to compute a power law relation between specific attenuation K and rainfall rate R at various wavelengths and temperatures.

Existing attenuation correction algorithms include (a) the Hitschfeld and Bordan (H-B) solution/algorithm, and its modified versions, for correcting single polarization radar reflectivity (Hitschfeld and Bordan 1954), (b) the method based on measurement from dual-polarization radar and (c) the networked approach (Lim et al. 2010); the latter of these is the approach mainly used by the CASA radars.

The H-B method uses a reflectivity-attenuation relation to solve for true reflectivity from attenuated reflectivity. This relation, as outlined by Hitschfeld and Bordan (1954), is numerically unstable and extremely sensitive to calibration errors and partial beam blockage. The H-B solution can be made stable by using total path-integrated attenuation (PIA) as a constraint; such a method has been applied successfully to the space/air-borne radar measurement of rain in the TRMM project (Meneghini and Kozu 1990) where the PIA is determined using the surface reference method. Independent estimates of PIA, however, are not available in general. Also, in case with multiple coexisting hydrometeor species, typical attenuation correction methods, including that of H-B, usually have difficulties.

The specific attenuation-differential phase parameterization method (DP method) uses the specific propagation differential phase K_{DP} provided by dual-polarization weather radars (Bringi et al. 1990; Jameson 1992; Park et al. 2005a). This is a more stable approach, whereby specific horizontal and differential attenuation (A_H and A_{DP}) can be estimated through relations with K_{DP}. The coefficients in the relations must be supplied a priori, and they vary as a function of the drop size distribution (DSD), temperature, and drop shape relation.

Testud et al.(2000) constrains the two-way path-integrated horizontal attenuation (PIA_H)by the total change in Φ_{DP} along a radial through a rain cell, termed the ZPHI. The attenuation is then apportioned according to the distribution of reflectivity factor at horizontal polarization (Z_H) along the radial, making this technique significantly more stable than the H-B method. The correction is done either by directly adding to reflectivity and differential reflectivity using correction amounts determined from the measured differential propagation phase Φ_{DP} (Matrosov et al. 2002; Anagnostou et al. 2006), or by adjusting coefficients in the attenuation-reflectivity and attenuation-differential phase relations used in the attenuation correction procedure such that the system is self-consistent(Bringi et al. 2001; Park et al. 2005a).

The pseudo-dual-frequency (PDF) method (Tuttle et al. 1983) estimates PIA_H at the end of a ray by the dual-wavelength ratio (DWR, defined as the difference between the S- and X-band horizontal reflectivity factor at the end of the cell) and apportions attenuation similarly to the ZPHI method. The PDF method requires radar data from systems operating at two different frequencies and is highly dependent upon the assumption of Z_H being equal at two wavelengths.

2.1.2 Attenuation correction with CASA radar network

For the X-band CASA radars, attenuation is corrected in one of two ways:

(i) Single radar data are corrected using the dual-polarization variable specific differential phase (Gorgucci et al. 2005); a self-consistency check is performed between K_{dp}, Z_h, Z_{dr}, and specific attenuation or differential attenuation. The estimated K_{dp} profile is then integrated in range to build an estimated profile of differential phase (Φ_{DP}), and Φ_{DP} is iteratively solved to obtain the best match to the observations. The initial condition for the iterative solution is obtained from the estimates of the specific attenuation profile from the differential phase constraint algorithm. This attenuation correction was evaluated and showed good performance in Gorgucci et al. (2006);

(ii) Data within multi-Doppler regions can be corrected by processing data from multiple radars. The real-time network-based reflectivity retrieval system (NBRR) involves collecting data from multiple remote radars and performing digital signal processing (Lim et al. 2010). By employing the methodology for reflectivity retrieval in a networked radar environment proposed by Chandrasekar and Lim (2008), the NBRR system works robustly in real-time while retrieving attenuation-corrected reflectivity.

While the algorithm CASA used has been shown to accurately retrieve un-attenuated reflectivity values, it can only do so when the reflected power is above the noise floor of the radar receiver. When total signal extinction occurs, the resulting radar data cannot be objectively distinguished from true clear-air data. To avoid erroneous assimilation of completely attenuated reflectivity data, some studies only assimilate radial velocity data. Snook et al. (2011) assimilates CASA reflectivity and radial velocity data only in regions where attenuation-corrected reflectivity exceeds 20 dBZ.

Unfortunately, the above constraint prevents the use of CASA clear-air reflectivity data for suppressing spurious storms that can develop within the numerical model. Tong and Xue (2005) showed that the assimilation of clear-air reflectivity data is very beneficial in suppressing spurious storms.

2.2 Overview of Simultaneous attenuation correction and state estimation

As the foundation of modern data assimilation for the atmosphere (Daley 1991; Kalnay 2002), the optimal estimation theory (Leith 1974) optimally combines different sources of information together with their error or uncertainty (usually in the least square sense) to obtain the best estimate of the state and/or parameters. Variational techniques and the ensemble Kalman filter (EnKF) (Evensen 1994) are advanced data assimilation (DA) methods based on optimal estimation theory and have been effectively applied to convective-scale model initialization with radar data (Sun et al. 1994; Snyder et al. 2003; Tong et al. 2005; Hu et al. 2006; Schenkman 2008; Putnam et al. 2010).

Modern data assimilation techniques such as variational and EnKF approaches are able to assimilate observations directly using the forward observation operators that link the model state variables to the observations (Kalnay 2002). For weather radars, the forward observation operators link the atmospheric state variables, including velocity, and hydrometeor species and amount, to the observed radial velocity and reflectivity (Xue et al. 2006). Accurate observation operators should take into account radar beam propagation (Gao et al. 2006), beam pattern weighting (Xue et al. 2006; Xue et al. 2007), thermodynamic effects such as bright band effects (Jung et al. 2008b), and attenuation (XTZ09). Additional observational parameters available from polarimetric Doppler radars, including differential reflectivity and differential phase measurements, provide helpful information about the density, shape, orientation, and drop size distributions (Doviak and Zrnic 1993; Gorgucci et al. 2001) and can also be connected through forward observation operators (Jung et al. 2008a).

With proper observation operators, variational and EnKF methods seek to minimize the difference between the observed quantities, which may be attenuated, and the model presentation of those quantities, subject to their respective uncertainties. Information with a smaller uncertainty will be weighted more heavily in the minimization/estimation process, and prior estimate together with its uncertainty information can also be readily used.

2.2.1 Overview of observation operator with attenuation

Hogan (2007) estimates rain rate using dual-polarization radar data through the variational approach, where attenuation correction is built directly into the forward observation operator using explicit treatment of errors, and attenuation is included straightforwardly without the instability problem encountered by H-B method. Their scheme is tested on S-band radar data and found to be robust and stable, even in the presence of differential phase shift on backscatter. However, the retrieval in low-rain-rate regions has been found to be very sensitive to the calibration of Z_{DR}. With such standalone analysis procedures, it is difficult to directly couple rain rate estimation with precipitation microphysics employed in numerical models, assuming the model microphysics is accurate enough for the model to benefit from such coupling.

Because flow-dependent background-error covariance derived from a forecast ensemble can be used to 'retrieve' unobserved state variables, EnKF is particularly useful for radar data assimilation. XTZ09 builds the attenuation effect into the forward observation operator of the ARPS EnKF system. The attenuation is calculated based on the estimate of atmospheric state, including the hydrometeor species. Such a procedure does not require any prior assumption about the specific hydrometeor types at particular grid points. It is possible to include error or uncertainty from all sources of information in the assimilation framework and allow for a close coupling of the calculation of attenuation with the dynamic model. As the model state estimate improves through data assimilation, the attenuation estimate also improves. The effectiveness of their procedure was demonstrated using a set of observation system simulation experiments (OSSEs), in which data from a single X-band radar that covered a supercell storm was simulated.

The observation operator with attenuation proposed by XTZ09 is used in this study; it is described below. The equations for reflectivity listed in (Xue et al. 2006;XTD06 hereafter) and the attenuation calculation introduced in XTZ09 form the forward observation operator for radar reflectivity data. These equations are used in both radar data simulation and assimilation.

The measured equivalent reflectivity factor in the presence of attenuation at a given range *r* can be expressed as

$$Z_e'(r) = Z_e(r)A(r), \qquad (6)$$

where $Z_e'(r)$ is the attenuated equivalent reflectivity factor,

$A(r) = \exp(-0.46\int_0^r k(s)ds)$ is the two-way Path-Integrated Attenuation (PIA) factor for equivalent reflectivity, and *k* is the attenuation coefficient (dBkm^{-1}). The attenuated reflectivity in dBZ can be obtained by taking $10\log_{10}()$ of Equation (6) so that,

$$Z'(r) = Z(r) - 2\int_0^r k(s)ds \ , \qquad (7)$$

where *Z(r)* and *Z'(r)* are reflectivity in dBZ before and after attenuation, i.e. the intrinsic reflectivity and attenuated reflectivity, respectively. It can be seen that the total PIA in dB, i.e. $PIA = -10\log_{10}(A(r))$, is equal to twice the integral of *k* between range 0 and *r*, reflecting the effects of two-way attenuation. For the purpose of data assimilation, the effect of attenuation and its correction can be achieved by including Equation (7) in the observation operator for reflectivity. The equation for the reflectivity calculation in the observation operator therefore becomes,

$$Z'(r) = 10\log_{10}(\frac{Z_e}{1mm^6m^{-3}}) - 2\int_0^r k(s)ds \, , \qquad (8)$$

The radar reflectivity factor Z_e and the attenuation coefficient *k* are linked to hydrometeor mass content (*W*, in mass per unit volume of air) through an exponential form DSDs of rain, snow and hail/graupel, consistent with the DSD assumptions in the 5-class single-moment microphysics scheme of Lin et al.(1983) used in the ARPS prediction model used in this study:

$$N(D) = N_0 \exp(-\Lambda D) \, , \qquad (9)$$

where N_0 is the intercept parameter and Λ is the slope parameter. The intercept N_0 is a constant, and the slope parameter is then uniquely linked to *W* (=$\rho_a q$, where ρ_a is the air density and *q* is the mixing ratio), through the assumptions about the DSDs and hydrometeor density. The effect of DSD uncertainty has been tested in a set of sensitivity experiments in XTZ09; the attenuation correction procedure as part of the EnKF data assimilation system appears to be less sensitive to the DSD model or radar calibration error than conventional methods.

The hydrometeor content and radar variables are represented by weighted integrals over the DSDs as follows:

$$W = \frac{\pi}{6}\rho\int D^3 N(D)dD, \qquad (10)$$

$$Z_e = \frac{\lambda^4}{\pi^5 |K_W|^2}\int \sigma_b(D)N(D)dD, \qquad (11)$$

$$k = 4.343\int \sigma_e(D)N(D)dD, \qquad (12)$$

where ρ is the density of hydrometeors, $K_W = \frac{\varepsilon_r - 1}{\varepsilon_r + 2}$ is the dielectric factor of water, ε_r is the relative dielectric constant of water, σ_b is the backscattering radar cross-section and σ_e the extinction cross-section for hydrometeor particles. Either Mie theory or a T-matrix method are chosen to calculate the cross-sections, depending on the experiment. In this study, we mainly use Mie theory.

For computational efficiency, we perform calculations within the possible range of water content beforehand, and use curve fitting to obtain formulae that can be used efficiently during data assimilation. We derive parameterized relations of model-predicted W with Z_e and K through equations (9), (10), (11) and (12), in which the Mie theory is used to calculate the backscattering radar cross-sections and the attenuation or extinction cross-sections. Details of the coefficients for power-law relations for all hydrometeor species are described in XTZ09 and can be found in Appendix A.

2.2.2 The observation error model (OEM)

Attenuated radar data inherently leads to spatially non-uniform observation errors. Severely or completely attenuated observations often contain significant observation errors, while un-attenuated observations tend to be more accurate. Unfortunately, in practice, we have little knowledge of the distribution of observation error associated with attenuated radar observations. We can, however, try to capture the main structure empirically by noting that when attenuation is occurring along the path, the reflectivity observations become small or even zero. The observation error model (OEM) is proposed to serve this purpose.

As the first attempt, an analytical relation between observation error and reflectivity is designed based on the observed value of reflectivity. In this relation, larger observation errors will be assigned when observed reflectivity is smaller, or even zero, under the assumption that small reflectivity values indicate the possibility of attenuation. Smaller observation errors (decreasing to a small constant value) will be used when the reflectivity observation value is larger. In real CASA radar data, a flag can be included to distinguish between clear air echo and fully-attenuated echo. Thus, in our OSSEs, the simulated observations contain such information as well. When an observation is marked as clear air, a constant small value (e.g. 2 dB) will be specified as the reflectivity observation error during data assimilation. We name this relation the analytical observation error model (hereafter, AEM).

The AEM is designed as below: in regions where reflectivity is below a specified minimum threshold (e.g., 10 dBZ), the observation error deviation is assigned a constant, large value (e.g., 8 dB), while in other regions a logarithmic relation is used to model the observation errors for the observations. Equation (13) gives such a relation:

$$\begin{aligned}
&\mathit{if}\ \mathrm{R}_e < \mathrm{R}_{e\_\min} \\
&\qquad \sigma_{obs} = \mathrm{C} \\
&\mathit{else} \\
&\qquad \sigma_{obs} = \sigma_{set} \cdot \max(\log(\frac{U}{R_e}), 1.0)\,. \qquad (13)
\end{aligned}$$

Here, σ_{obs} denotes the modified observation error variance; σ_{set} denotes the original preset observation error variance. R_e is the value of observed reflectivity and R_{e_min} is the threshold used to define whether a constant value *C* should be set as observation error, or a logarithmic form should be used. *U* inside the logarithmic form can be varied in the equation. The constant value *C* and *U* are chosen based on experimentation, and can be varied from case to case to obtain the best analyses. In Chapter 3, we will further discuss the possible settings of *C* and *U* and corresponding results.

Other than the analytical forms of observation error model, another possible form of observation error model that considers multiple possible error sources and calculates corresponding error based on estimation obtained through EnKF is proposed in section 3.5. According to the process of obtaining radar observations and assimilating radar observations in EnKF, the observations can be decomposed into different parts and parameterized differently. The possible error components are: observation estimation error; reflectivity model error; attenuation mode error; signal-to-noise-ratio related bias or error; and inhomogeneity related error. During data analysis, those error components can be calculated based on the information estimated from the background, thus providing us another possible way to specify proper observation error for each observation. And this type of observations error model is named as the multi-sources error model (MSEM). Tests and result comparisons between the two types of observation error models (MSEM and AEM) will be discussed in section 3.5.

Chapter 3: OSSEs Assimilating Attenuated Data from an X-band Radar Network for a Squall Line and testing with multi-error sources observation

3.1 The truth simulation

A numerically-simulated QLCS was generated by the Advanced Regional Prediction System (ARPS;Xue 2000; 2001; 2003) to serve as the truth simulation for OSSEs. The ARPS is used in a 3D cloud model mode in which cumulus parameterization is not used and surface physics and radiation processes are ignored. The prognostic variables include three velocity components u, v, w, potential temperature θ, pressure, p, and six moisture variables, i.e., water vapor specific humidity $q_{v,}$, and mixing ratios for cloud water, q_c, cloud ice q_i, snow q_s and hail q_h. In addition, turbulence kinetic energy is also predicted and used to determine turbulent mixing coefficients based on a 1.5-order turbulence closure scheme. The microphysical processes are parameterized using the Lin et al (1983) scheme with two categories of liquid water and three categories of ice.

The physical domain of all experiments is $293 \times 197 \times 16$ km^3 and the horizontal grid spacing is 1.5 km in most experiments (FIG 3. 1a). To better resolve the lower atmosphere, a vertically stretched grid with a minimum vertical grid-spacing of 100 m near the surface is used. The initial homogeneous storm environment is defined by a two-layer wind shear profile (FIG 3. 1b), in which the westerly wind increases linearly from zero to 15 m s^{-1} at 2.5 km and remains at 15 m s^{-1} above 2.5 km. To initialize the truth simulation, a line of thermal bubbles with 4 K maximum potential temperature perturbation, along with additional random perturbations of 5%, is used to trigger the storm, located at x=50 and z=1.5 km from the northern domain boundary to southern domain boundary. The radii of these bubbles are 10 km in x-direction, 20 km in y-direction and 1.5 km in the z-direction, while the distance between each bubble is 9.6 km, and the southernmost bubble is 9.6 km from the southern boundary. A wave radiation condition is applied at the west and east boundaries while a periodic condition is used at the north and south boundaries. Free-slip conditions are applied to the top and bottom boundaries. The length of the simulation is four hours.

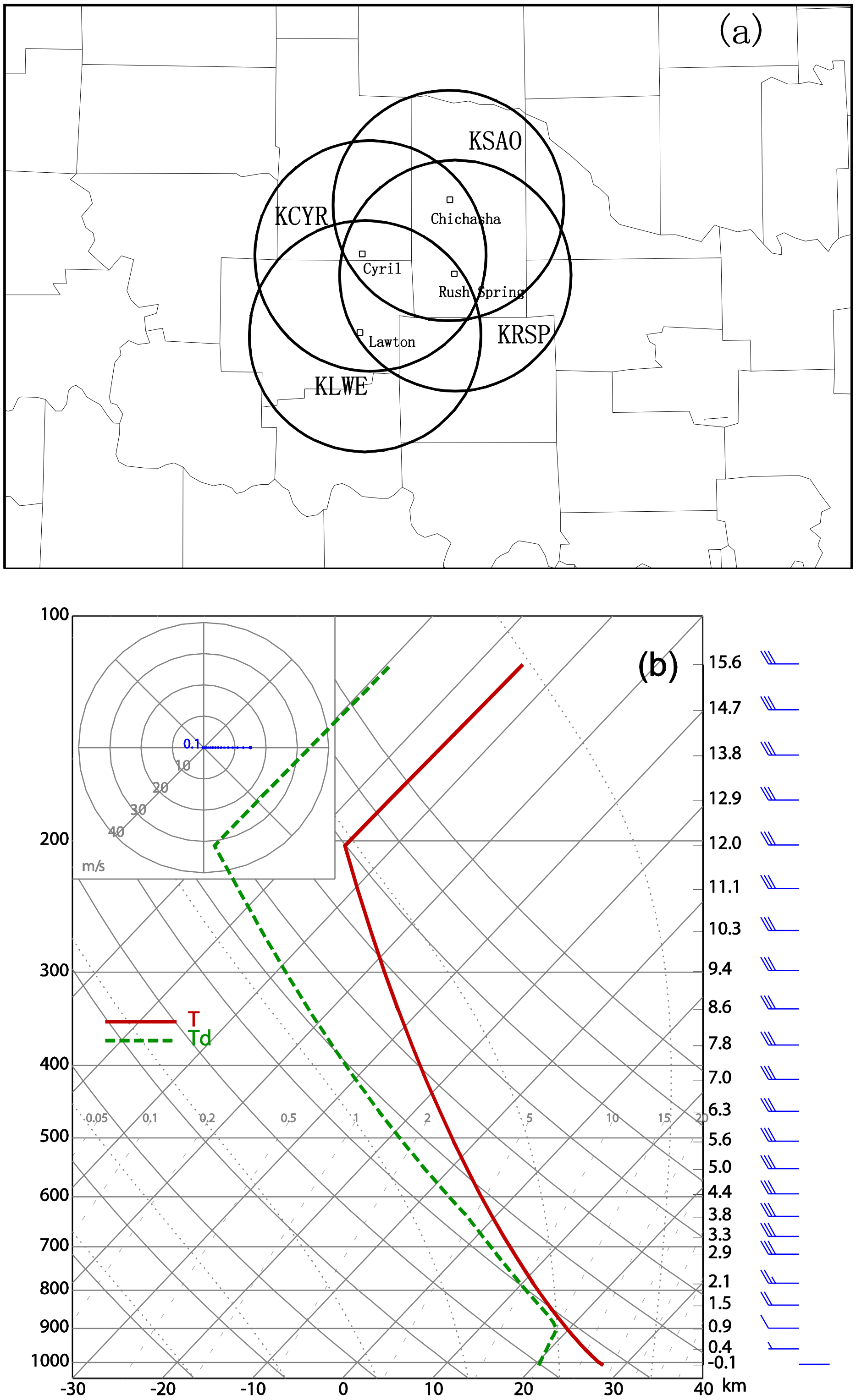


FIG 3. 1 (a) Map of 288 km × 192 km computational domain. The small solid circles represent the maximum 40 km range of the CASA IP1 radars. (b) Environmental soundings used to generate the truth simulation of a squall line.

3.2 Simulation of radar observations

The four X-band polarimetric CASA IP-1 radars used in OSSEs are located near Chickasha, Rush Springs, Lawton and Cyril (McLaughlin et al. 2009) in Oklahoma (FIG 3. 1a). Simulated radar data are generated, following XTD06, using a Gaussian power weighting function in the vertical for observations simulated on radar elevation levels (PPI or plan position indicator planes). Data are assumed to have been interpolated to the model Cartesian coordinates horizontally, but remain on the radar elevations vertically. The effect of Earth curvature and beam bending due to vertical change of refractivity are taken into account using the 4/3 effective Earth radius model discussed in Doviak and Zrnic (1993). The velocity is projected to the direction of radar beam locally to give the simulated radial velocity, and the terminal velocity effect is properly taken into account using the hydrometeor state variables. The radars are assumed to operate in one of the CASA radar storm scan modes, having 10 elevations with one volume scan every 5 minutes and a 1.8° beamwidth. The attenuated reflectivity is calculated by integrating along the path of each radar beam using Eq. (7), where the reflectivity before attenuation (in dBZ) is given by $Z = 10\log_{10}[Z_{er} + Z_{es} + Z_{ews} + Z_{eh} + Z_{ewh}]$, (where subscripts *r, s*, and *h* denote rain, snow and hail, while *ws* and *wh* denote wet snow and wet hail, respectively).

As an example, FIG 3. 2 shows the simulated radar reflectivity obtained from the four radars at an elevation of 5.0°, with and without attenuation, at 135 min of model time. Each radar only partially covers the QLCS. The high reflectivity regions, mainly associated with high mixing ratios of rainwater and hail, including melting hail, tend to be completely attenuated at large ranges from the observing radars, resulting in wedges of zero or near-zero reflectivity behind heavy precipitation cores. The maximum reflectivity in the convective cores is reduced by more than 50 dBZ. Thus, the pattern and magnitude of attenuation appear realistic. Since this radar network has many overlapping regions, some of the heavily-attenuated reflectivity regions from one radar can be detected by other radars, which in the following experiments helps model correct attenuation.

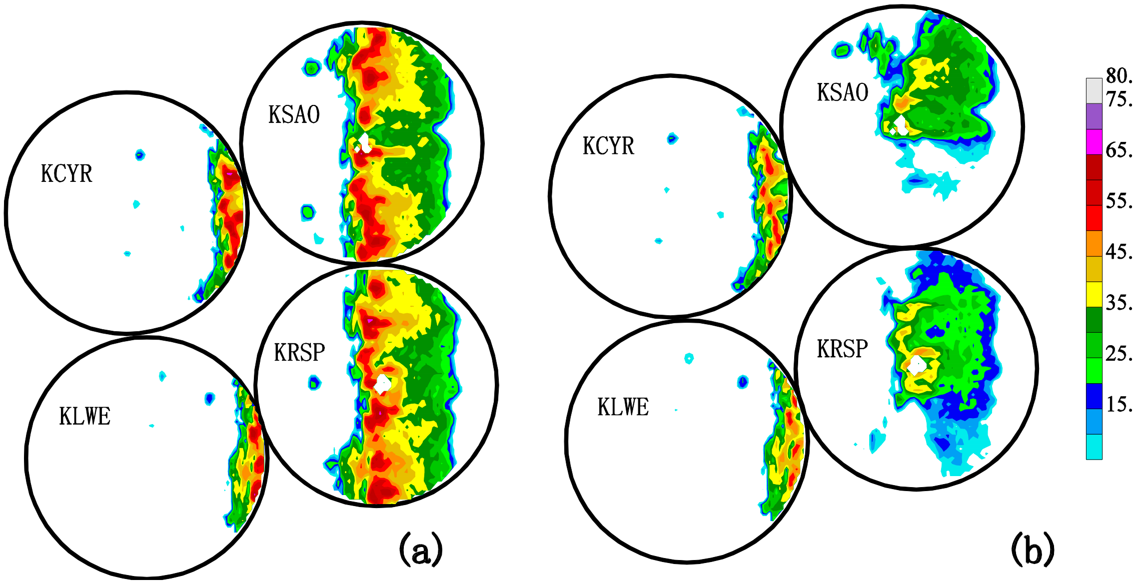


FIG 3. 2 Simulated (a) non-attenuated and (b) attenuated reflectivity observations at 5 ° elevation at 135 min of the squall line simulation time.

In all experiments, Gaussian-distributed random errors with a mean of zero and standard deviations of 1 m s^{-1} and 2 dB for radial velocity V_r, and reflectivity Z, respectively, are added to the simulated data sampled from the truth simulation. These values are also used to specify the observation error variances during data assimilation. This is not the case in the observation error model series of experiments to be discussed later.

3.3 Data assimilation experiments

The basis of our data assimilation is the EnKF radar data assimilation framework of XTD06, which was based on Tong and Xue (2005) and further enhanced in Tong and Xue (2008) to include the terminal velocity effect in the radial velocity observation operator. The observation operator including the attenuation effect, as described in section 2.2.1, is used for reflectivity.

An ensemble of 40 members is initialized at t = 105 min of model time by adding random perturbations to a horizontally homogeneous ensemble mean defined by the environmental sounding used by the truth simulation. Observations are assimilated every 5 min beginning at t=110 min. The smoothed random perturbations are sampled from Gaussian distributions with zero mean and standard deviations of 2 ms^{-1} for u, v, and w, 2 K for potential temperature θ, and 0.6 gkg^{-1} for q_v, q_c, q_r, q_i, q_s, and q_h. Pressure and microphysical variables are not perturbed. The variables u, v, θ, and q_v at the first grid level above ground are not perturbed, following XTD06, as doing so was found to introduce noise into the analyzed pressure field. The covariance localization procedure follows Houtekamer and Mitchel (2008a), applying a Schur product of the background error covariance calculated from the ensemble and a correlation function with local support. Covariance inflation is limited to the grid points where observed $Z >$ 10 dBZ and all points within 4 grid points in the horizontal (6 km in perfect model experiments) and/or vertical directions. The multiplicative inflation factor, β, is set to 1.07.

As described in Table 3. 1 a set of standard experiments is first performed to test the effectiveness of the reflectivity observation operator for the QLCS case. We compare the performance of performing attenuation correction during EnKF data assimilation (experiments name start with ‘ATC’) with the experiments that do not perform attenuation correction (experiment names start with ‘N’). Suffix ‘Z’ indicates that only reflectivity data is assimilated, otherwise both Z and V_r data are assimilated. In VRONLY, only radial velocity data are assimilated. Simulated attenuated radar data are used in all experiments except in experiment ‘NAC10Z’, in which only reflectivity data above 10 dBZ are assimilated. Experiment names starting with “OBSE” indicate the application of the AEM during data assimilation. We also tested three types of possible reflectivity relations in OBSEs series experiments. The assimilation of V_r data is conduct once read-in Z exceeding 10 dBZ in all experiments. Details of the observation error model will be discussed further in section 3.4.2.

Table 3. 1 Lists of perfect model OSS experiments

Experiment	Observation operator with attenuation effect	Observations assimilated	Observations error standard deviation
NAC	No	Z & V_r	2dB
NAC10Z	No	Z (>10 dBZ) & V_r	2dB
VRONLY	No (not necessary)	V_r	2dB
ATC	Yes	Z & V_r	2dB
ATCZ	Yes	Z	2dB
OBSE	Yes	Z & V_r	reflectivity relation 2
OBSEZ	Yes	Z	reflectivity relation 2
OBSEONLY	No	Z & V_r	reflectivity relation 2
OBSE1	Yes	Z & V_r	reflectivity relation 1
OBSE3	Yes	Z & V_r	reflectivity relation 3

Table 3. 2 lists all imperfect model experiments and contains specific descriptions of the experiment settings. In this set of experiments, the truth simulation is generated in the same way as the truth used in the perfect model experiments except that the grid spacing is 500 m. When the EnKF OSSEs are performed using a 1.5 km grid spacing, model error due to the resolution difference arises. The impact of model resolution error is examined in experiments starting with “LOWRES”. For comparison, experiments with prefix “HIRES” are run at the full 500-m grid-spacing (i.e. without model resolution error). The observation operator with attenuation effect is used in all imperfect model experiments; the suffix “OBSE” denotes whether the

Observation Error Model is applied during data assimilation. Comparison between these experiments will enable us to evaluate the impact of observation error and model error when assimilating attenuated radar observations.

Table 3. 2 Lists of imperfect model OSS experiments

Experiments	Observations assimilated	Observation error deviation	Localization radius	Model resolutions (model error contained)
HI_ORG	Z & V_r	fixed observation error (2dB)	4dx (2 km)	dx=dy=500 m, dz=500 m (no)
HI_OBSE	Z & V_r	reflectivity relation 2	4dx (2 km)	dx=dy=500 m, dz=500 m (no)
LOW_ORG	Z & V_r	fixed observation error (2dB)	4dx (6 km)	dx=dy=1500 m, dz=500 m (yes)
LOW_OBSE	Z & V_r	reflectivity relation 2	4dx (6 km)	dx=dy=1500 m, dz=500 m (yes)

3.4 Results of experiments

A key question our study seeks to answer is whether or not the EnKF system can successfully recover the attenuated signal in areas with near-complete signal extinction. We also examine the impact of the fact that the QLCS in the truth simulation is much larger than the radar network observing it. XTD06 shows that when the radar does not provide full coverage of the storm system, significant errors could develop in the analysis that cannot be effectively corrected due to lack of observational information. Including data from WSR-88D radars can help increase the radar coverage. However, because the purpose of this research is to examine the performance of a short-wavelength observation operator with attenuation effects, we will focus on observations from X-band radars only. As such, it is expected that the root-mean-square (RMS) errors of the results will be higher than those in XTZ09.

Furthermore, in this study, calculation of RMS errors is confined to grid points inside of the IP1 radar network coverage region since the model cannot accurately analyze storm features outside the radar coverage area. Calculating RMS error only inside radar network coverage allows us to focus on evaluating the capability of the DA system in recovering the model state subjecting to reflectivity attenuation.

3.4.1. Perfect model experiments—simultaneous attenuation correction in EnKF for a squall line

First we evaluate the effectiveness of the observation operator with attenuation (hereafter called the attenuation observation operator) proposed by XTZ09 in the simulated squall-line case. The analysis of NAC (FIG 3. 3b) shows a very weak storm system, which is mainly due to assimilating data, containing large error caused by attenuated observations and an improper observation operator which does not account for attenuation (hereafter called the original observation operator). Especially for the fully attenuated reflectivity from KSAO and KRSP (FIG 3. 2b), the observations provide incorrect information regarding the storm, which result in nearly echo-free regions in the south-east portion of radar network (this region is located around 150-170 km in x direction, and 64-82 km in y direction). Discarding radar reflectivity below 10 dBZ during data assimilation (NAC10Z) could improve the analysis quality by omitting bad observations; however, doing so will also remove the clear air data that could aid in suppressing spurious echoes. NAC10Z shows minor improvement on storm reflectivity structure (FIG 3. 3c). Assimilating radial velocity only can improve the wind field and produce a reflectivity pattern closer to the truth in the squall line area inside the radar network coverage region (FIG 3. 3d). There is still, however, a large amount of spurious echo surrounding the storm system due to lack of clear air data assimilation. On the other hand, the application of the attenuation observation operator allows the model to correct attenuation based on simultaneous state estimation inside the storm system. Even assimilating reflectivity data alone (as in ATCZ) can help capture several convective cells of the storm (FIG 3. 3e) compared to NAC (FIG 3. 3b) and NAC10Z (FIG 3. 3c). However, some of the attenuation correction seems erroneous, especially in the area near 70km in the x direction and 170 km in the y direction in FIG 3. 3e, possibly due to inaccurate state variable estimation affected by the spurious echoes located just outside of the southeast edge of KRSP range. The addition of radial velocity data in experiment ATC makes the storm look much like the truth storm inside the radar coverage area both in terms of storm structure and intensity at 135 min of model time (FIG 3. 3f), except for the southern part of the squall line. Here, in the overlapping region of KLWE and KRSP, the model is still able to recover some small echoes, likely because data from KLWE are less attenuated at the time (FIG 3. 2b). When only observations from KRSP are available, the filter failed to correct this area because it was assimilating fully attenuated observations. However, the model does not produce erroneous echoes as in ATCZ due to assimilation of radial velocity data. Still, more work is needed to improve our system of EnKF which includes attenuation correction.

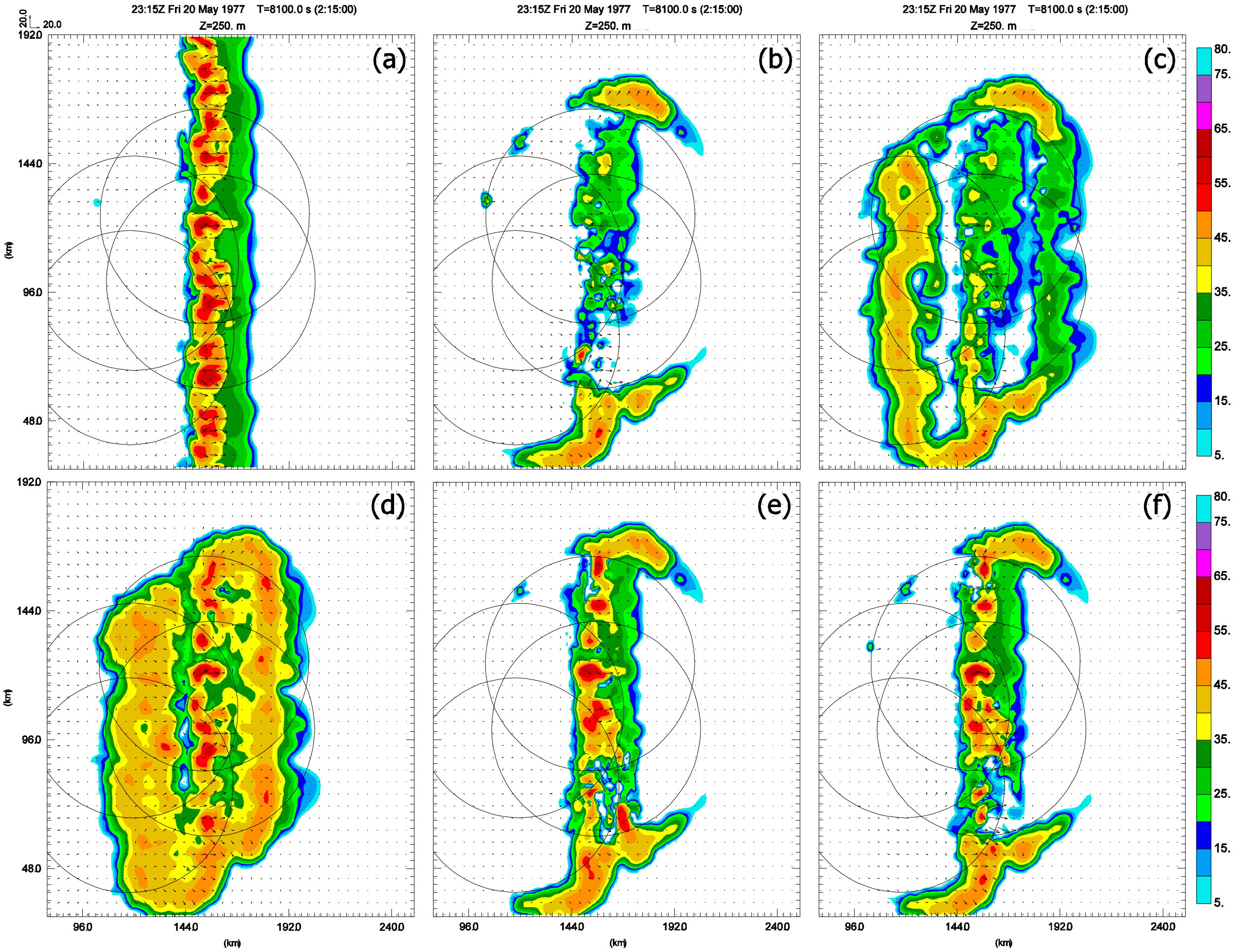


FIG 3. 3 The horizontal wind vectors (m/s, plotted every third grid point) and computed reflectivity (shaded at 5 dB interval, starting from 5 dBZ) for (a) Truth simulation; and experiments (b) NAC; (c) NAC10Z ; (d) VRONLY; (e) ATCZ; (f) ATC at model time 135 mins

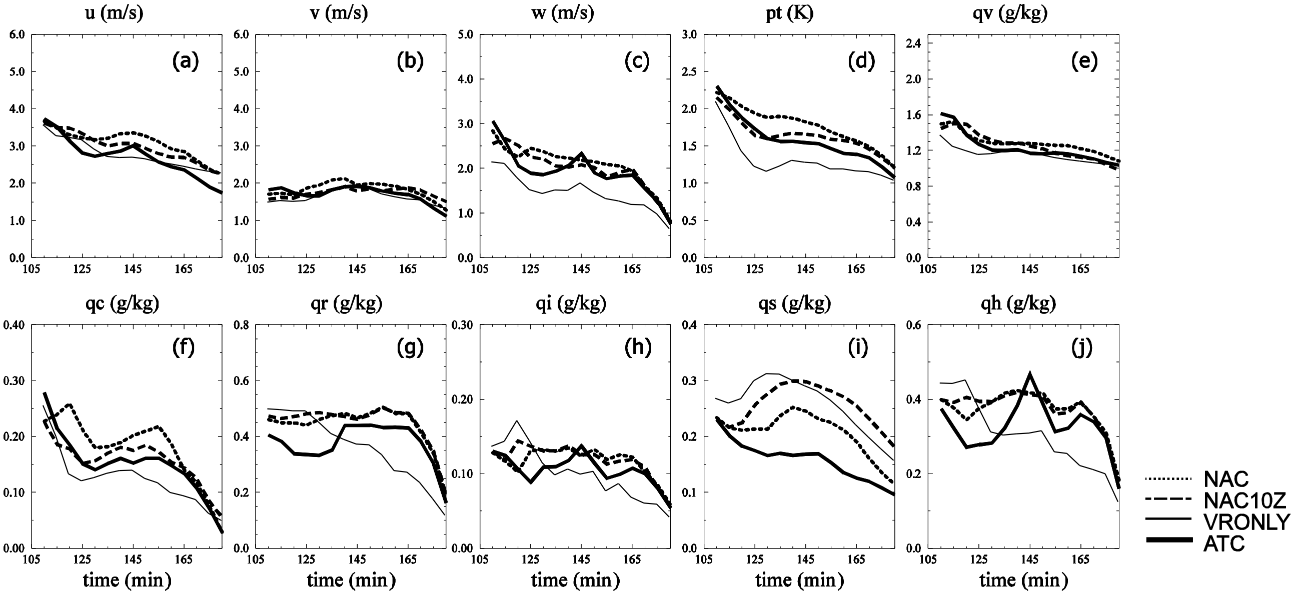


FIG 3. 4 The ensemble mean analysis RMSEs averaged over points inside radar coverage for (a) u, (b) v, (c) w, (d) perturbation potential temperature θ', (e) q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h for experiments NAC (dotted), NAC10Z (dashed), VRONLY(thin solid), and ATC(thick solid)

We examine the quality of state estimation, i.e. the analyzed individual model state variables, by looking at the RMS errors of ensemble mean analyses during the analysis cycles. As mentioned in the beginning of section 3.4, the calculation of RMS error will be only inside the radar network coverage. These RMS errors are calculated against the truth fields in regions where the truth reflectivity is greater than 0 dBZ only, i.e. the verification is not performed when truth reflectivity is equal or less than 0 dBZ. In another word, the spurious echoes that do not exist in truth are not verified. Thus although VRONLY produces quite a lot spurious echoes, the RMS error levels are lower than other experiments (FIG 3. 4). For clarity, we show the RMS errors for the analyses only, not for the background forecasts.

FIG 3. 4 compares the RMS errors from four expermients (NAC, NAC10Z, VRONLY and ATC). The error levels of ATC are lower than other experiments in hydrometer fields especially for q_r, q_i, q_s and q_h before model time 130 min. Later, however, a sudden increase in error levels occurs around 135 min. FIG 3. 5 reveals that during this period the storm system is moving away from the radar network center and entering an area where fewer Multi-Doppler observations are available. Ingesting fully-attenuated observations appears to deteriorate the analysis. To better explain this result we can examine the EnKF analysis equation, given by:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}[y_j^o - H(\overline{\mathbf{x}}^b)], \qquad (14)$$

where $\overline{\mathbf{x}}^a$ is the ensemble mean analysis vector containing all model state variables. The amount of attenuation can only be correct when the estimate states are accurate, but at this time the state estimation is poor (FIG 3. 4), thus attenuation calculations based on the estimated states are not very accurate either. Thus, when fully attenuated observations read in, y_j^o has a value of zero, but $H(\overline{\mathbf{x}}^b)$ is not (instead, it has a positive reflectivity value). The analyzed model state variables $\overline{\mathbf{x}}^a$ become smaller than the background state variable $\overline{\mathbf{x}}^b$; in other words, bad observations result in deterioration of the analysis. As seen in FIG 3. 5, the decrease of storm intensity or even complete disappearance of storm echoes occurs in the ensemble mean analysis compared to the ensemble mean forecast after each cycle of data assimilation, especially in the southern part of the QLCS inside the radar network coverage area where only observations from KRSP are available.

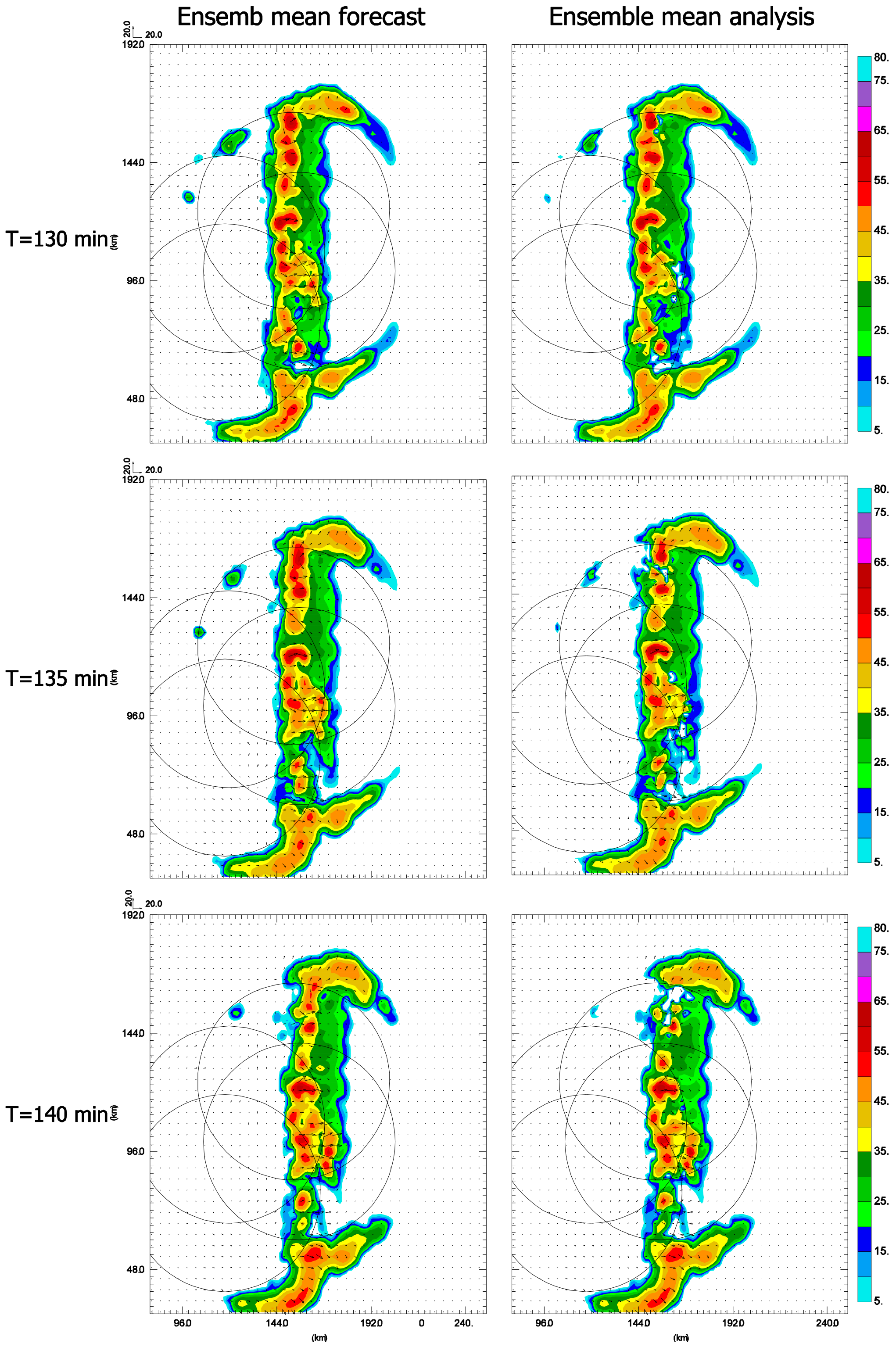


FIG 3. 5 Same as FIG 3. 3 but for ensemble mean forecast (left panels) and ensemble mean analysis (right panels) from ATC at model times 130 min, 135 min and 140 min

FIG 3. 4 also indicates that VRONLY has the smallest errors in the wind field compared to the other experiments, and that VRONLY also most accurately analyzes the potential temperature, water vapor, and cloud water distributions; however, prior to 130 minutes of model time, RMS errors are worse for q_r, q_i, q_s, q_h are worse in VRONLY than in other experiments due to the spurious storms found within the radar network coverage area that could not be effectively suppressed without the help of reflectivity data.

3.4.2 Perfect model experiments— the analytical observation error model (AEM)

As discussed in section 2.2.2, observations with severe attenuation usually contain larger errors than data with little or no attenuation, and zero reflectivity has uncertainties (because it is unclear if such data are fully attenuated or truly clear-air returns). The observation error model (AEM), based on a reflectivity relation, is designed to specify the observation error in the EnKF DA as introduced in section 2.2.2. Since this is an empirical technique, we first tested several possible observation error variance relation curves and show 3 of them in FIG 3. 6. Eq. 13 is used to generate these relation curves. Here, R_{e_min} is set to 10 dBZ and σ_{set} is set to 2.5 dB in all test experiments. The three relations are generated by changing the C and U in Eq. 13. C is set to 8.0 dB in relation 1 and 2, and 16.0 dB in relation 3. U is set to 220 dBZ in relation 2 and 3, and 160 dBZ in relation 1. FIG 3. 7 shows the RMSEs of the ensemble mean analysis of the experiments using the 3 relations. Differences among the experiments exist mainly in w and hydrometer fields, because improvement from assimilating reflectivity data is mainly evident in hydrometeor fields, and estimation of mixing ratio is critical to w retrieval (Jung et al. 2008b). Although relation 3 has the lowest error level after 165 min, during the period from 140 min to 155 min, when less Multi-radar observation coverage are available, errors grow quickly and the analysis becomes the worst (especially for rain water) compared to other two relations (FIG 3. 7c). This is likely due to the relatively large observation error assigned when reflectivity is less than 10 dBZ. Such large observation error variance (16 dB) can limit the usage of zero reflectivity data caused by full attenuation. We have a clear air flag to prevent us from erroneously treating clear air observation as fully attenuated data. When weak reflectivity (<10dBZ) that are not fully attenuated occur and the correct attenuation amount can be calculated, adding overly large observation errors (relation 3) reduces the positive impact of this reflectivity data. In comparison, experiments employing relation 2 show better analyses during the same period (relation 2 has the smallest error level from 140 min to 155 min and larger error than relation 3 but smaller error than relation 1 after 165 min). Hence, relation 2 is chosen as the AEM used in our remaining experiments.

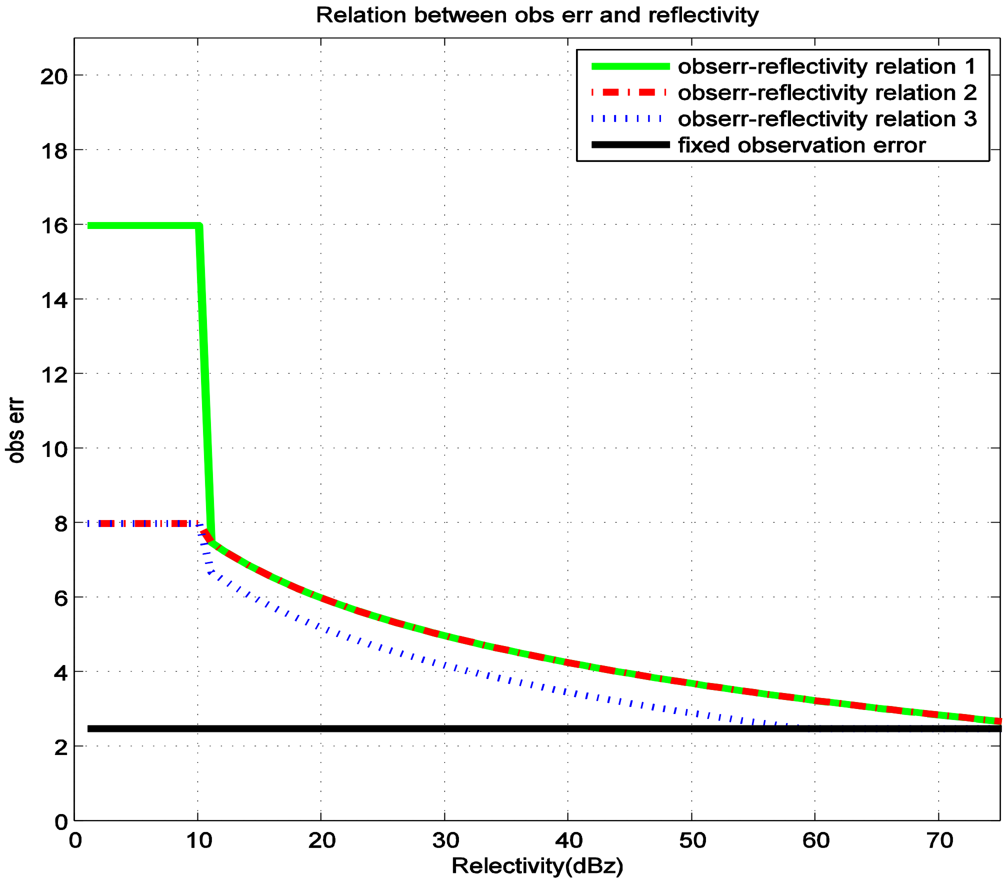


FIG 3. 6 Empirical relations between observation error and radar reflectivity. Green (solid), red (dashed), blue (dotted) lines represent the observation error reflectivity relation 1, 2, 3 respectively. Black solid line indicate the observation error used by ATC, ATCZ

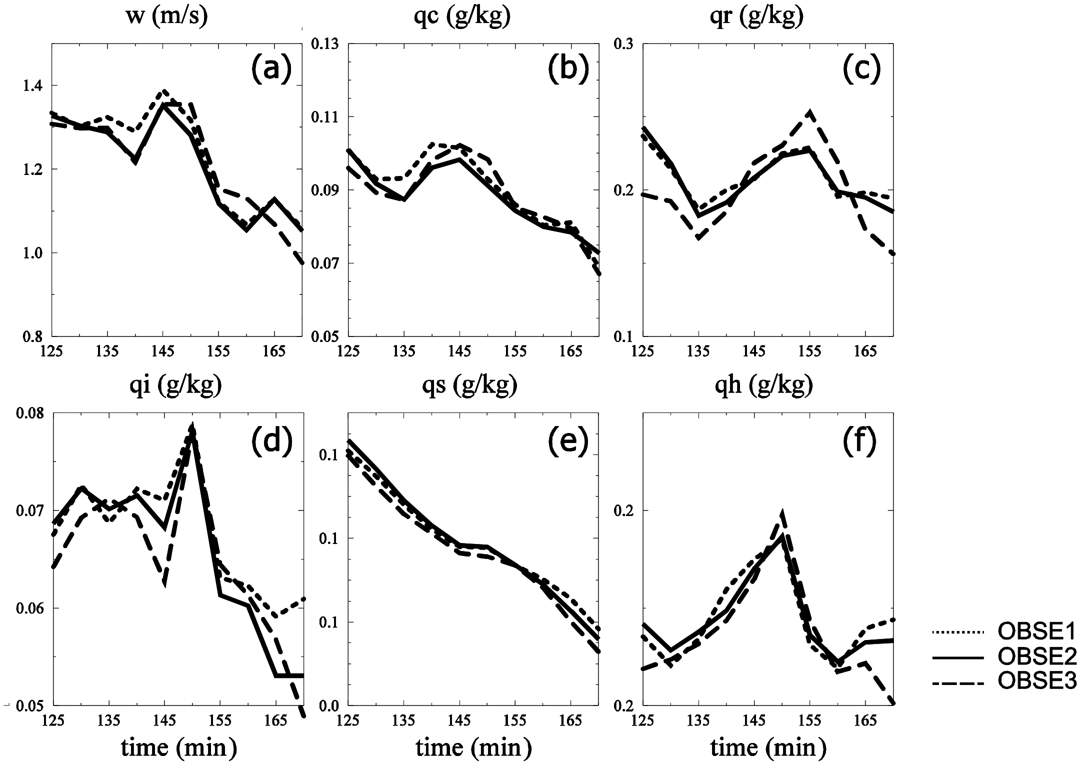


FIG 3. 7 The ensemble mean analysis RMSEs averaged over points inside radar coverage for (a) w, (b) q_c, (c) q_r, (d) q_i, (e0q_s, (f)q_h for experiments OBSE1 (reflectivity relation 1, dotted), OBSE(reflectivity relation 2, solid), OBSE3 (reflectivity relation 3, dashed) from 125 min to 170 min.

The observation error structures after applying AEM relation 2 to radar observations on 4th elevation at model time 135 min are shown in FIG 3. 8. It is easy to tell that the small constant value (2.0 dB) of calculated observation error as a denotation for clear air echo appears in center/western part of the coverage of KCYR and KLWE, since at this time, storm located at far east portion of this two radar coverage. Areas behind radar echoes are specified with large values, while the storm regions (where large reflectivities are) are filled with varied observation errors especially in north part of the storm that observed by KSAO and KRSP.

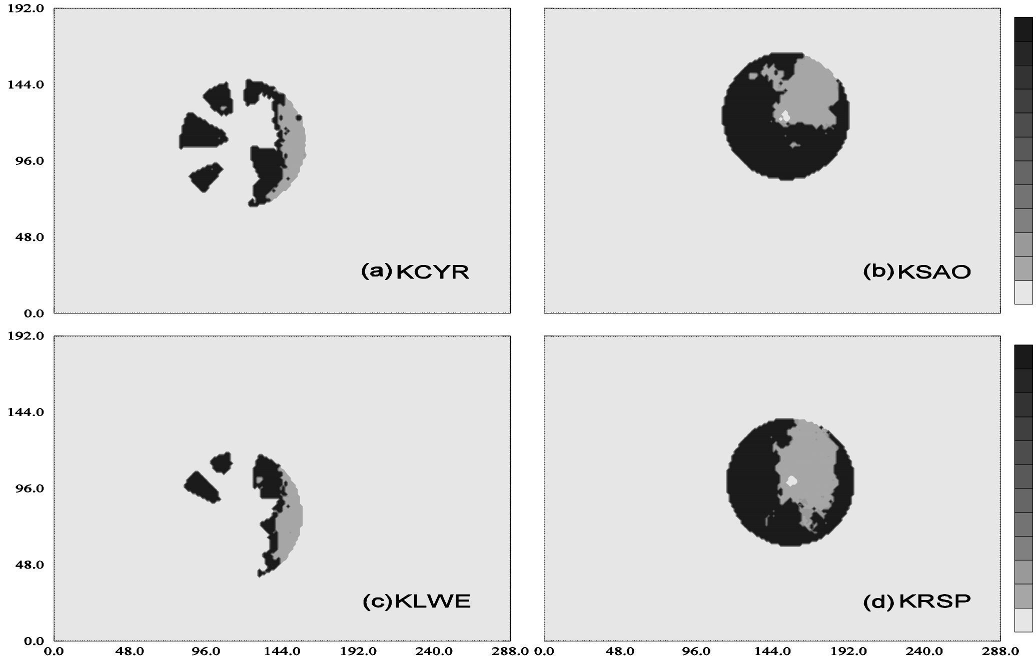


FIG 3. 8 Corresponding observation error deviation (using obserr-reflectivity relation 2) for radar (a) KCYR (b) KSAO (c) KLWE (d) KRSP at 4th tilt at model time 135 min

Applying AEM relation 2 leads to much improved analyses, especially for the fully attenuated reflectivity that is essentially treated as clear air signal in experiments such as ATCZ. For example, the analyzed storm echoes from ATCZ decay rapidly from 135 min to 145 min, and mainly exist in the center of radar network coverage at 145 min. In contrast, the result of OBSEZ shows that some strong convective cells still appear in the south part of radar network coverage area at this time (FIG 3. 10). When the result of OBSEZ (FIG 3. 10) is compared to the truth at 145 min (FIG 3. 9), the convective cells located in the center of radar network coverage and the southern part of KRSP coverage are similar to the truth in reflectivity fields. However a strong cell that does not exist in truth also appears in the southeast part of the KRSP radar coverage region at this time. Moreover, the north part of squall line of OBSEZ is still much weaker than in the truth simulation at 145 min.

Adding high-quality V_r data into our EnKF DA system shows a larger positive impact than assimilating reflectivity data alone (FIG 3. 9). Looking at V_r , the comparisons between using constant observation error and using AEM are more obvious. As the analyzed QLCS shrinks in to middle in ATC from 135 min to 145 min, the analysis from OBSE successfully recovers most of the squall line, especially at later times (145 min). Plots of RMS errors from experiments ATCZ, ATC, OBSEZ, and OBSE (FIG 3. 11) confirm the advantage of using the AEM. Overall, the analysis error levels of OBSEZ and OBSE are lower than those of ATCZ and ATC. The RMS errors of the analyzed model fields are about 1/3 less than those not using AEM, i.e., when assuming a constant error variance for reflectivity observations.

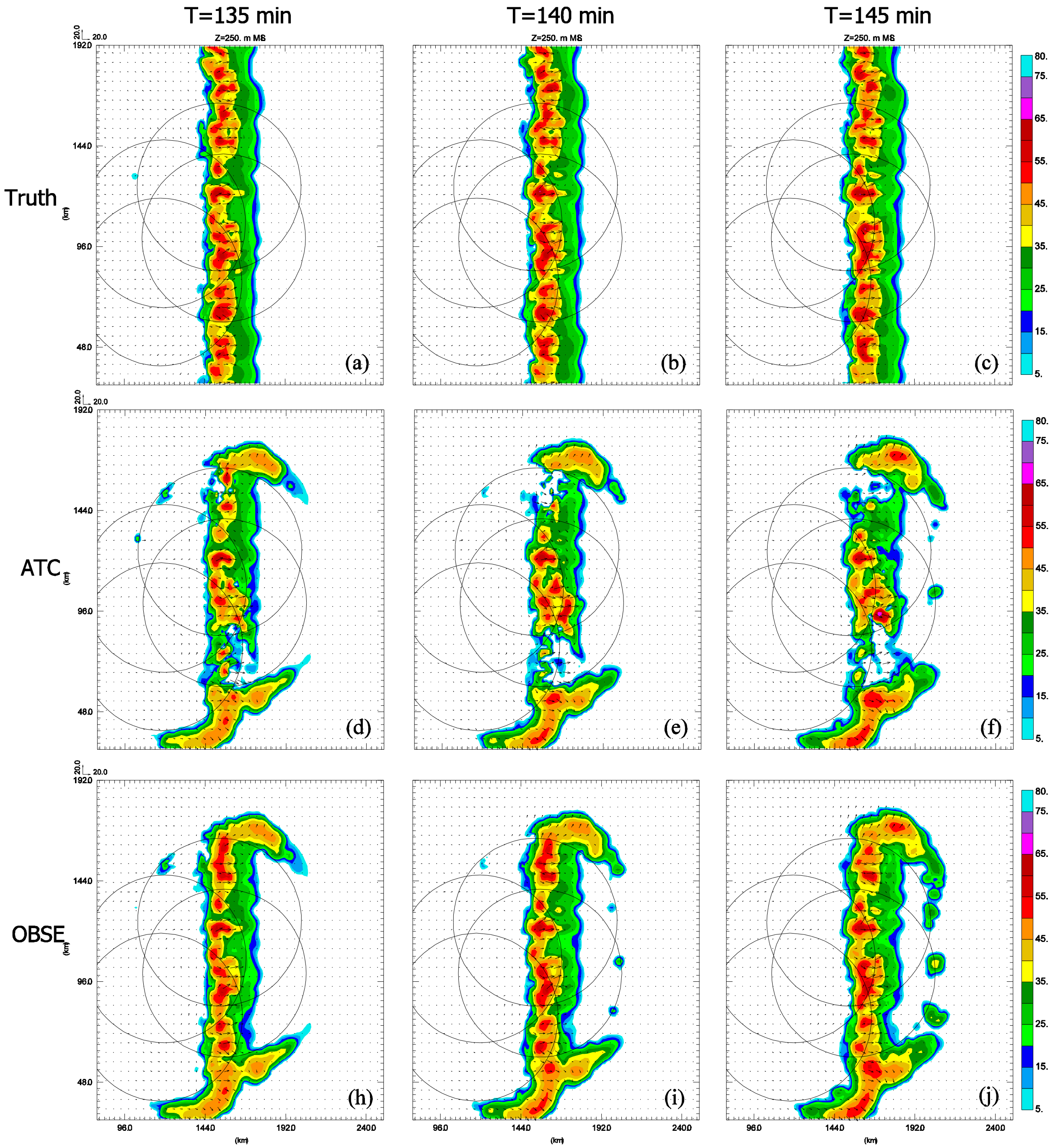


FIG 3. 9 Same as FIG 3. 3 but for Truth (upper panel), ATC (middle panel) and OBSE (lower panel) from model time 15 min to 145 min

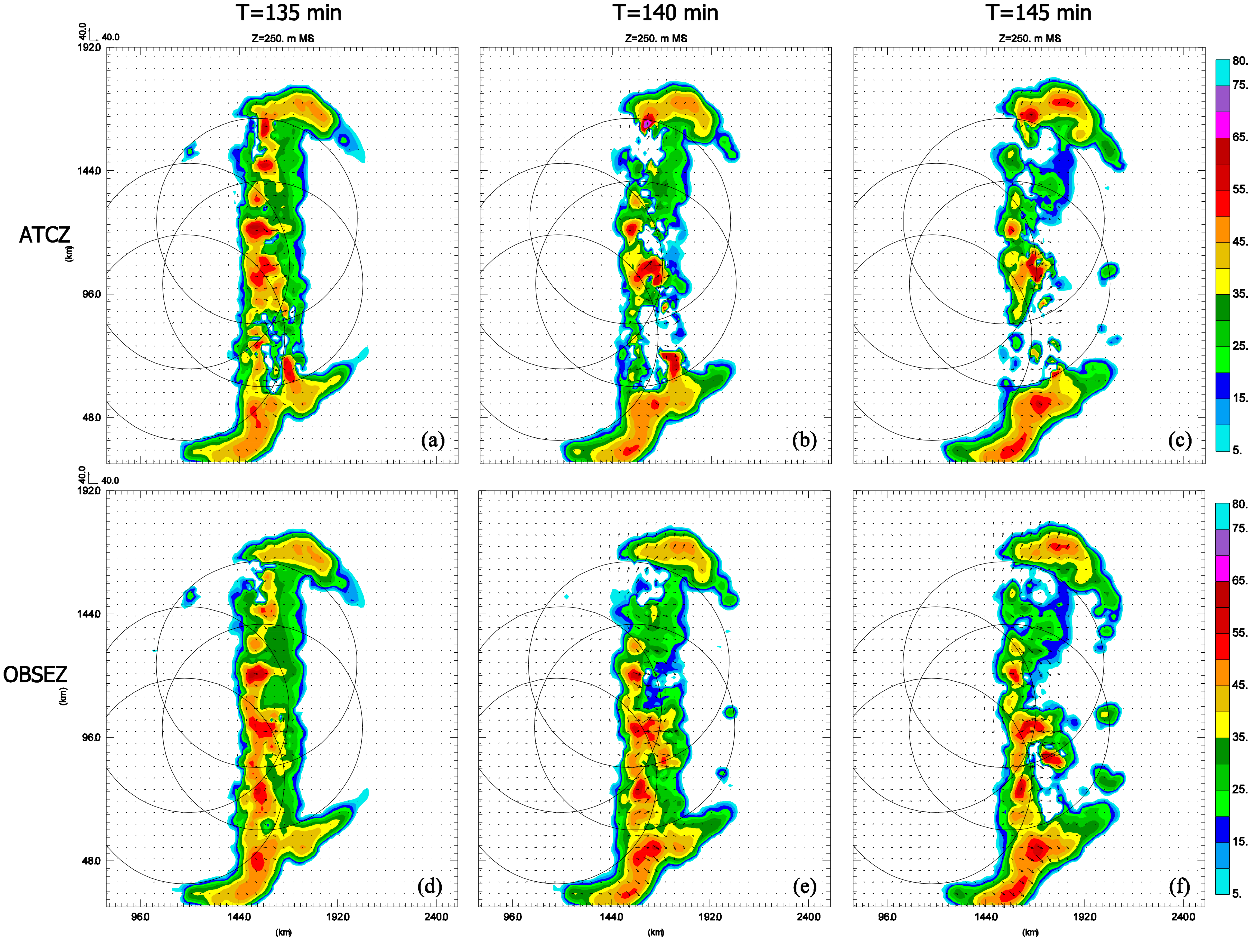


FIG 3. 10 As FIG 3. 3, but for experiments ATCZ (upper panel) and OBSEZ (lower panel)

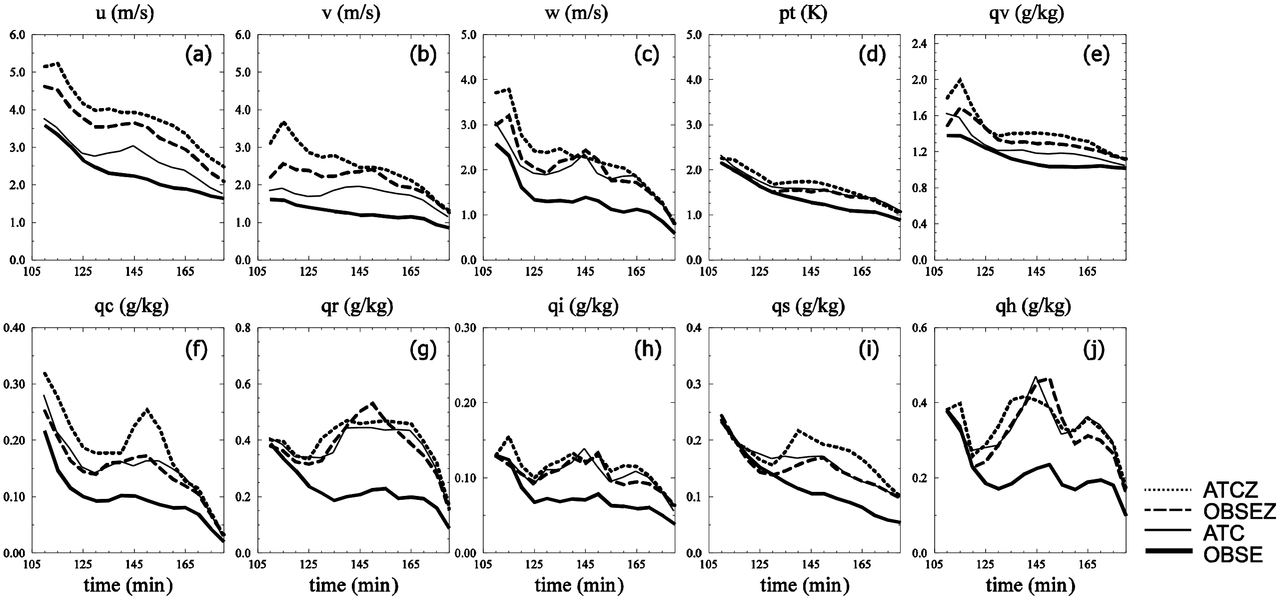


FIG 3. 11 As in FIG 3. 4, but for experiments ATC (thin solid), ATCZ (dotted), OBSE(thick solid), OBSEZ (dashed)

Through OBSEONLY, we also examine the situation when the AEM is used without correcting the attenuation through the observation operator. Can the DA system still produce a quality analysis under such a restriction? The resulting analysis RMS errors (FIG 3. 12) indicate that the AEM has a smaller positive impact on the model solution when attenuation is not included in the observation operator. This result confirms the importance of including attenuation in the observation operator. Overall, best results are obtained when an attenuation-correcting observation operator is used in combination with the AEM.

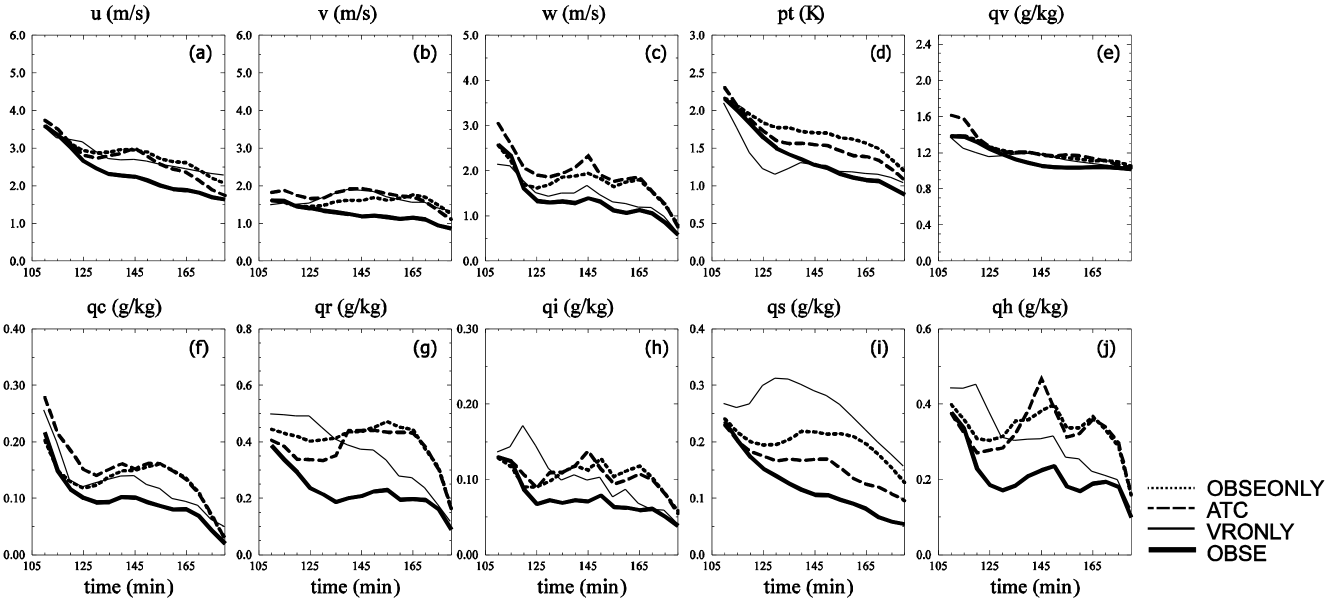


FIG 3. 12 As in FIG 3. 4, but for experiments ATC (dashed), VRONLY(thin solid), OBSE (thick solid) and OBSEONLY (dotted)

3.4.3 Experiments with imperfect assimilating model

In reality, numerical prediction models are never perfect. In fact, in many cases model error is likely one of the largest sources of error. In XTZ09, which examined a super cell storm, the robustness of the attenuation correction process was tested by introducing error into the intercept parameters of the rain, snow, and hail size distributions assumed in the ice microphysics parameterization scheme of the prediction model. These intercept parameters were also involved in the reflectivity and attenuation calculations. Neglecting the effect of non-spherical raindrop shapes in the observation operator is found to have minimal impact. Further, the procedure was not sensitive to moderate-sized systematic radar calibration errors. In this study, model resolution error is introduced into the OSSEs to examine the impact of an attenuation-correcting observation operator and the AEM under imperfect model conditions.

As in the perfect model experiments, the truth simulation used here is created by adding the same perturbations to the same environmental sounding to generate the ensemble initial conditions for the first analysis cycle, using a finer horizontal grid spacing of 500 m. The same vertically-stretched grid with minimum grid-spacing of 100 m near the surface is used. Radar observations are simulated from this high-resolution truth simulation, while the EnKF DA cycles are performed on model grids with dx = 500 m and dx = 1500 m in different experiments (see.Table 3. 2). Prefix "HI" in the experiment names indicates DA on the 500 m resolution grid, while "LOW" indicates DA on the 1500 m grid that has model resolution error (compared to the truth simulation). Suffix "ORG" indicates that a fixed value of observation error variance was used during data assimilation, while "OBSERR" indicates use of the AEM (the observation error model using relation 2).

We note that the observations assimilated on the 1500 m grid (exp. LOW series) are interpolated in the horizontal directions to the 1500 m grid. Because our simulated radar data are defined at the model columns, the number of observations assimilated in the LOW experiments is one-ninth of the number in the HI experiments. The localization radius is set to 4 grid intervals in all directions in both sets of experiments, so that each state variable is updated by a similar number of observations in the two sets of experiments.

In the 500 m truth simulation, the QLCS propagates faster than in the 1500 m truth simulation. To keep the QLCS inside the radar network during the data assimilation period, the assimilation cycles are started earlier, at 80 min of model time. Assimilation cycles are run from 80 min to 160 min at 5 min intervals, and in both sets of experiments (listed in Table 3. 2), both reflectivity and radial velocity data are assimilated.

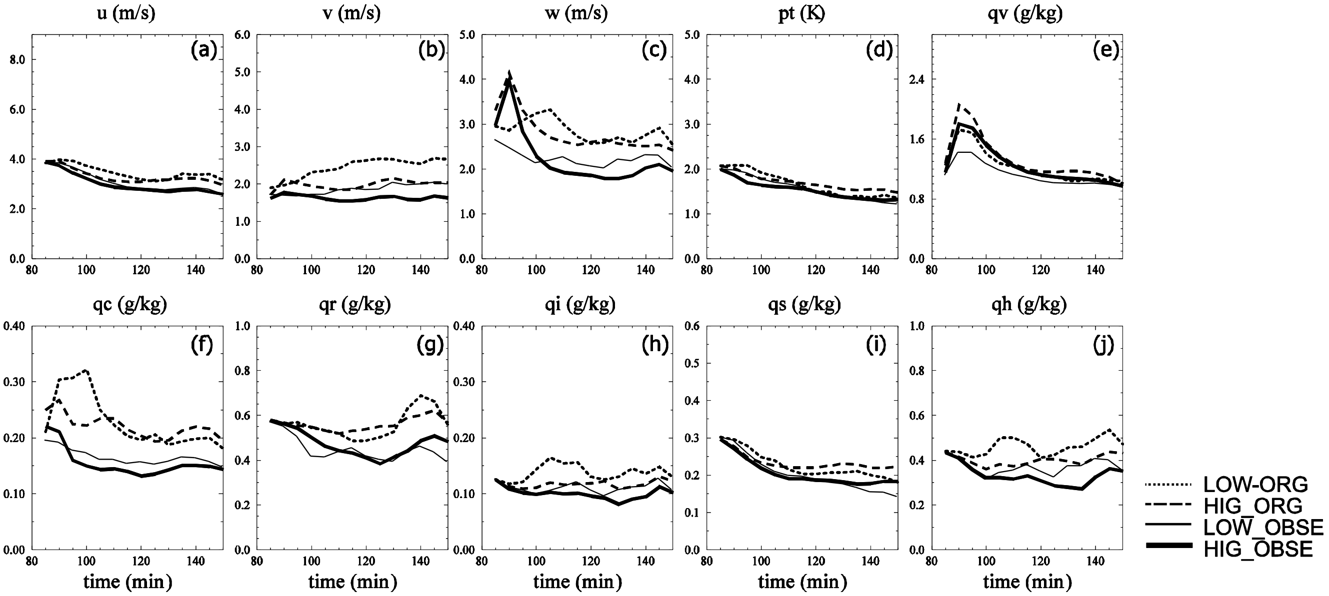


FIG 3. 13 As in FIG 3. 4, but for experiments HIG_ORG(dashed), HIG_OBSE(thick solid), LOW_ORG(dotted), LOW_OBSE(thin solid)

In order to quantitatively compare the experiments results against the same truth, a 1500 m resolution truth is generated by using every third point from the 500-m truth simulation. Both HI experiments and LOW experiments are verified against this 1500 m resolution truth. FIG 3. 13 presents the RMSEs for results of the imperfect model experiments. Overall, HIG_OBSE gives the lowest error levels while the worst analyses in terms of RMSEs are from LOW_ORG. When both model error and large observation error exists in experiment LOW_ORG, significant errors develop in the analysis that cannot be effectively corrected by data assimilation. However, the EnKF data assimilation system appears to be less sensitive to resolution-related model error than to observation error when we compare HI_ORG, HI_OBSE and LOW_OBSE. The differences in error levels between LOW_OBSE and HIG_OBSE indicates the impact of model resolution error, which are smaller than the differences between HIG_ORG and HIG_OBSE that represents observation error impact. Such differences are obvious, especially in *w* and hydrometeor fields. In other words, the AEM is still effective even when model error exists. The results of the experiments indicate the robustness of our attenuation correction procedure as part of EnKF data assimilation.

3.5 Testing OEM with simulated observations contain multiple error sources

In above experiments, Gaussian-distributed random errors with a mean of zero and a standard deviation of 2 dB are added to the simulated reflectivity (*Z*) data sampled from the truth simulation. In reality, observation errors can come from various sources.

According to the process of obtaining radar observations and assimilating radar observations in EnKF, the observations can be decomposed into different parts and parameterized differently. The possible error components are: observation estimation (measurement) error (σ_{Ze}); reflectivity model error (σ_{Zm}); attenuation model error (σ_{Za}); signal-to-noise-ratio (SNR) related bias or error (σ_{ZSNR}); and inhomogeneity related error ($\sigma_{Zin\mathrm{homo}}$).

The observation estimation (measurement) error represents the error caused by instruments. The error standard deviation σ_{Ze} is considered as a constant value of 2 dB. The reflectivity model used in experiments is imperfect, thus a concave function with minimal error for median reflectivity represents the reflectivity model error standard deviation σ_{Zm} (FIG 3. 14). Similarly, a monotonically-increasing function with path-integral-attenuation (PIA) defines the error standard deviation σ_{Za}, related to attenuation model (FIG 3. 15). SNR is another source for error and we consider the error standard deviation σ_{ZSNR} as a monotonically-decreasing function with SNR (FIG 3. 16). Finally, radar observations are not uniformly distributed. Dense observations can be obtained close to the radar location while sparse observations usually exist in far from the radar. Interpolation of radar data from observation space to model space can produce inhomogeneity related error. But in OSSEs, since radar observations are produced on the model grid horizontally, inhomogeneity-related error is not considered in experiments in this section.

Gaussian distribution is assumed for each error term, and the sum of all the errors is added to Z_e, producing the observations to be assimilated during the EnKF experiments. The average error deviation generated using this method is around 6 dB.

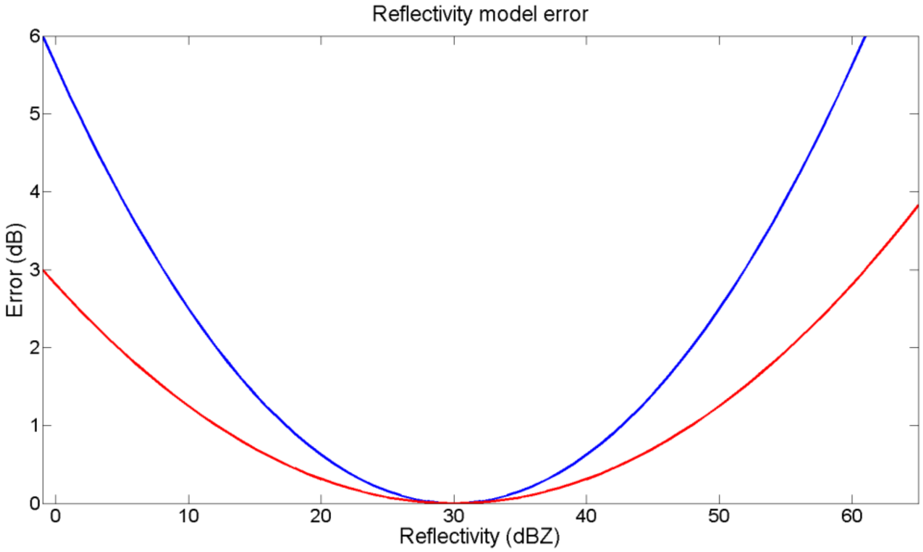


FIG 3. 14 Reflectivity model error as a function of reflectivity original designed (blue) and used in experiments (red)

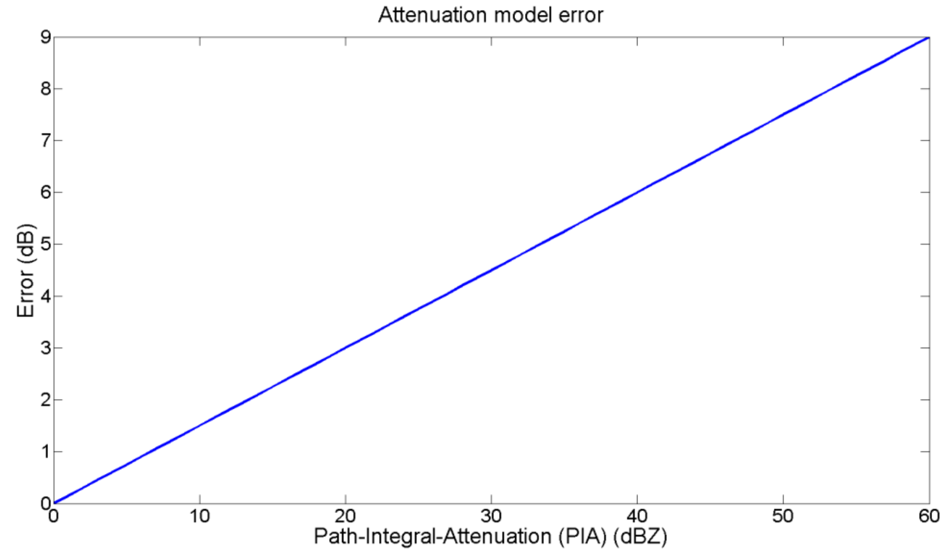


FIG 3. 15 Attenuation model error as a function of Path-Integral-Attenuation (PIA)

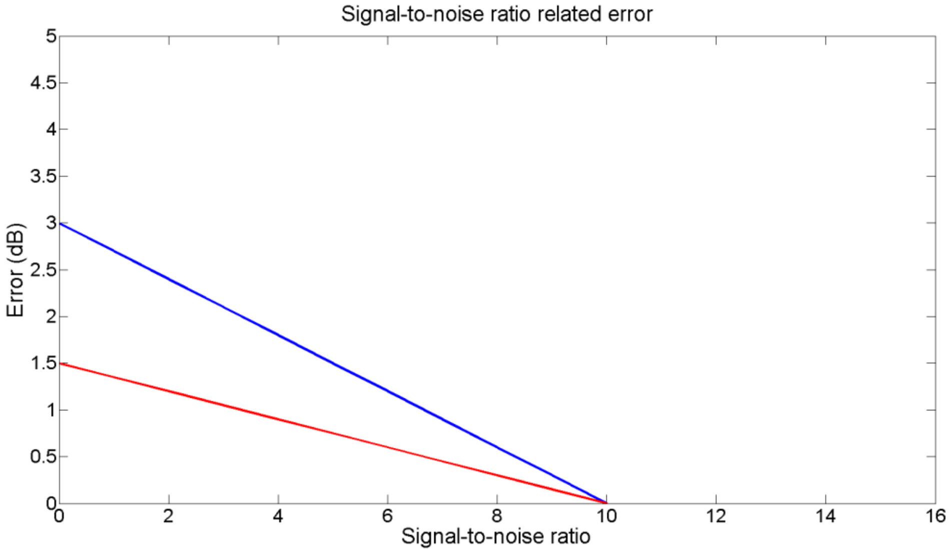


FIG 3. 16 Signal-to-noise ratio related error as a function of signal-to-noise ratio original designed (blue) and used in experiments (red)

The purpose of this set of experiments is testing the performance of our attenuation-correcting system when observation errors come from multiple sources. To keep the experiment simple, we produce a supercell storm (instead of a QLCS) using a sounding from May 20, 1977. An X-band radar with 60 km range is placed at the southwest corner of a domain with a physical size of 70 km $\times$ 70 km $\times$ 16 km. Doing so avoids error caused by lack of radar coverage. The truth simulation is generated on a grid with 500 m horizontal and 500 m vertical resolution, while EnKF analysis and cycles forecast are run on a 1.5 km horizontal model grid. The model resolution error is included in this set of experiments. Other configurations are the same as in the OSSE experiments from the previous sections. The same method used in section 3.2 is adopted to generate the simulated reflectivity data sampled from the truth simulation, but added with the sum of above observation errors. FIG 3. 17 shows an example of the simulated observation after adding errors. Compared with the simulated observation with additive perturbations using 2dB standard deviations Gaussian-distributed random errors, the severe attenuation area contains more noise due to attenuation model error. The average standard deviation of error added onto Z_e at every time step based on this method is between 5~6 dB.

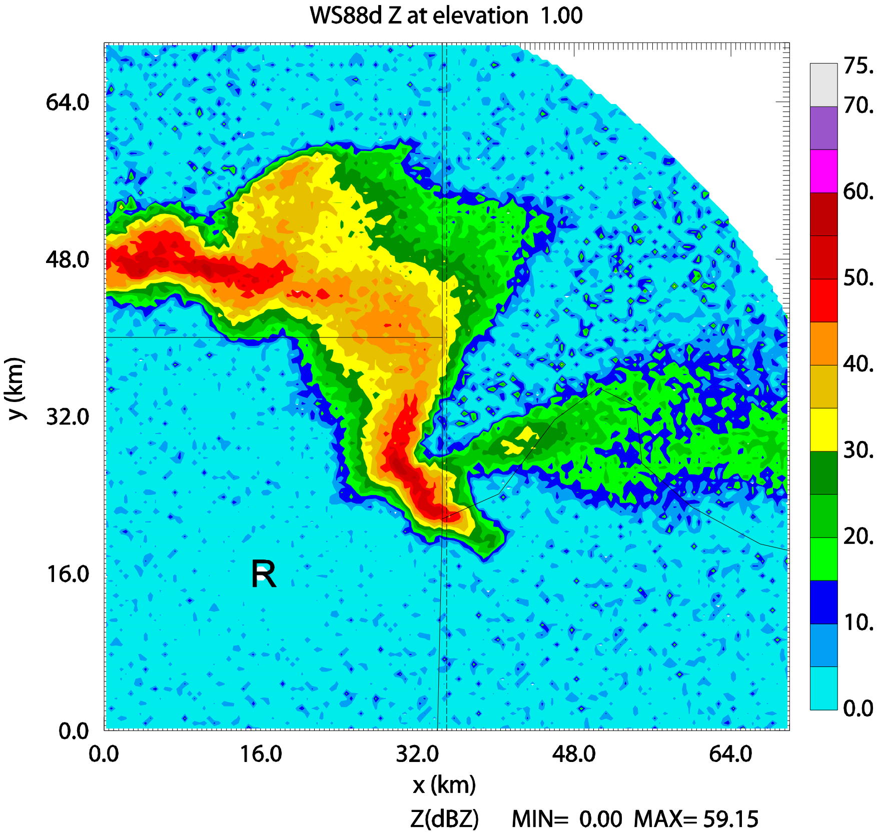


FIG 3. 17 Simulated radar observations on elevation 1.0 °at 120 mins model time. Character ‘R’ indicates the location of radar

Table 3. 3 List of experiments in multi-error sources experiments

Experiments	Observation error deviation
ATTC	2 dB
ATTC_AEM	Empirical analytical function
ATTC_MSEM	Use background information to estimate error using the same equation as in generating observation
ATTC_LGERR	8 dB

The experiments in this section all use the forward observation operator with attenuation (Table 3. 3), but differ in the way that observation errors in EnKF are specified. ATTC uses a constant (2 dB) value for observation error deviation, while ATTC_AEM applys the same AEM as other experiments in previous sections. In ATTC_MSEM, the observation error also varies, but is not specified by an empirical analytical function, which actually is another type of OEM; for this reason we name it the multi-sources error model (MSEM). A large constant (8 dB) observation error deviation is used in ATTC_LGERR. The background information is used to estimate observation error deviation; in other words, we calculate each term of the observation error deviation ($\sigma_{Zm}, \sigma_{Za}, \sigma_{ZSNR}$) through the same equations, but do so based on the estimated reflectivity, PIA, and SNR in Equation (15).

$$R = \sigma_{Ze}{}^2 + \sigma_{Zm}{}^2 + \sigma_{ZA}{}^2 + \sigma_{ZSNR}{}^2 \qquad (15)$$

RMSE level in ATTC is high since the value used for observation error deviation is too small. And when the same equations are used to calculate the error deviation that used in EnKF (ATTC_MSEM), the error levels are reduced slowly, which might be due to the calculation of errors mainly relying on the background estimation, putting too much weight on the background. The AEM, on the other hand, specifies the observation error deviation based on the observation value. The RMSEs levels in ATTC_AEM decrease faster than those in ATTC_MSEM. When both large observation error and large model error exist, using a large constant value (ATTC_LGERR) is another way to reduce the error level in the analysis. However, we also conducted a set of experiments in perfect model scenario, and simply increasing the constant value of observation error deviation in EnKF is not as effective as in imperfect model cases. However, at the end of assimilation time, the error in the three experiments (ATTC_MSEM, ATTC_AEM, ATTC_LGERR) reaches the same level, which may suggests a saturation error level in this experiment.

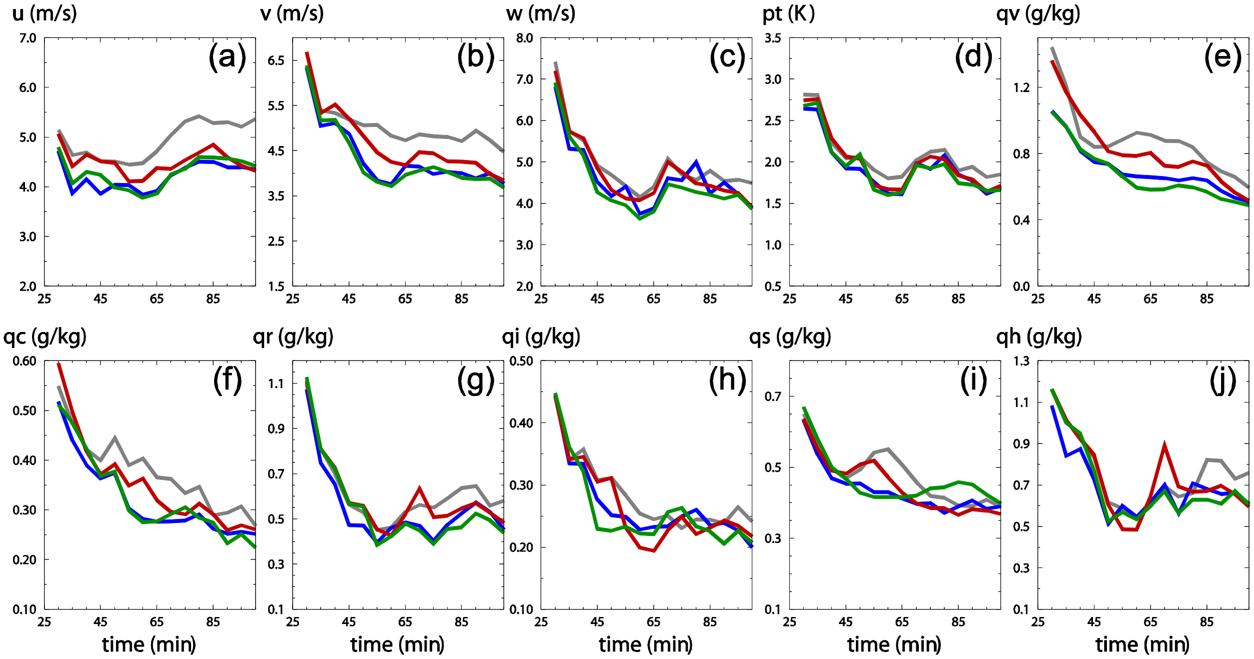


FIG 3. 18 The ensemble mean analysis RMSEs averaged over points where true Z is greater than 10 dBZ for (a) u, (b) v, (c) w, (d) perturbation potential temperature θ', (e) q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h for experiments ATTC (gray), ATTC_AEM (blue), ATTC_MSEM(red), and ATTC_LGERR(green)

3.6 Summary and discussion

An ensemble Kalman filter system that can directly assimilate attenuated radar reflectivity observations is examined for a simulated squall line in an OSSE framework. A forward observation operator with attenuation included is adopted to deal with attenuated radar observations. The calculation of attenuation is based on the current estimate of the atmospheric state, including the hydrometeor species. The estimated state is obtained through an ensemble-based data assimilation system, using attenuated data. This implies that we are performing simultaneous “attenuation correction” and state estimation. Xue et al. (2009) first demonstrated the effectiveness of such a procedure through a set of OSSEs for a relatively isolated supercell storm. In this study, we apply this procedure to a simulated QLCS that has a broader precipitation regions than an isolated supercell storm and hence more attenuation problems; a network of four X-band radars positioned according to the CASA IP1 radar network is assumed in the OSSEs. The overlapping observation coverage afforded by multiple radars is very helpful in producing more accurate reflectivity analyses, as multiple viewing angles reduce the areas of complete attenuation.

Severe or complete attenuation in the reflectivity data do lead to larger analysis errors than those in earlier studies when no attenuation exists and/or when radar provides more complete coverage on the convective storms. Despite these added difficulties, the EnKF system with the attenuation effects built into the reflectivity observation operator is shown to significantly improve the analyses, compared to the case when the attenuation effect is ignored. In the latter case, the model hydrometeor fields tend to be substantially under-estimated. When assimilating radial velocity data alone, extensive spurious echoes develop in the analyses that cannot be effectively suppressed. As a simple attempt to avoid using fully attenuated reflectivity, only reflectivity data exceeding 10dBZ are assimilated in one experiment. A problem with excessive spurious echoes also arises, because of the lack of clear-air observational information in those areas that could help suppress spurious storms. This points to the importance of effectively utilizing all available reflectivity information to the maximum extent feasible. Since, in the EnSRF, observations are assimilated sequentially, to prevent fully attenuated observations from negatively impacting previous analyses during DA, an attempt to rearrange the sequence in which observations are read in using a different radar order is also examined. However, the improvements to analyses are still minor. One possible solution is to use the reflectivity data whenever they are available, but it would be necessary to estimate their uncertainty, including possible attenuation effect, so that negative impacts from erroneous use of fully attenuated data are reduced.

Considering that observations with severe attenuation usually contain larger errors than data with little or no attenuation, and that zero reflectivity has larger uncertainty (because we cannot be sure if the zero reading is due to complete attenuation or actual clear air), an empirical analytical observation error model or AEM based on a reflectivity relation is designed and used to specify the observation error variance in the EnKF DA. In such a relation, larger observation errors are assigned to lower observed reflectivity values, while higher reflectivity can be weighted more by specifying a smaller error. The clear-air data are properly taken care of by flagging them during the process of producing simulated radar observations. Such data can be generated in a real case too, which will be discussed in Chapter 5. A small error variance is assigned to known clear air data. With the help of the AEM, the analysis results are greatly improved. The RMS errors of the analyzed model fields are about 1/3 less than those assuming a constant error variance for the reflectivity observations.

The effectiveness of the attenuation correction procedure and the utilization of a reflectivity-dependent observation error model are further tested in the presence of the resolution-related error in the assimilating model. This error is introduced by using a truth simulation that is produced at 3 times the resolution as that of the assimilating model. It was shown that while the analysis errors increase when such model error exists, the model error impact is smaller than the positive impacts of the attenuation correction procedure and the use of the AEM.

When multiple error sources are present in the observations, AEM is still effective in reducing the RMSEs levels quickly. Using MSEM can also reduce the error levels, however the reduction is slower than that obtained using AEM. Theoretically, the same equations used to produce observation errors are used in MSEM to estimate the observation error deviation in EnKF. MSEM should be able to produce similar observation error structure to the observations. However, the PIA related error we generated is based on a Gaussian distribution (which is unrealistic); this might be the key reason that MSEM did not exhibit better performance than AEM. In the future, we plan to add PIA related error with spatial correlations and test how the MSEM and AEM will perform under that circumstance.

Meanwhile, using large constant observation error deviation in EnKF is another way to obtain a good analysis when model error is considered. Since the average error deviation produced in the observations is around 6 dB, which is close to the large constant observation error deviation (8 dB) used in EnKF, this might be a coincidence. In another set of perfect-model experiments for this multiple error source case (not shown here), using a large constant observation error actually leads to worse results compared to AEM and MSEM. Thus, the inclusion of model error could be another reason why simply using large constant observation error in the EnKF works. This results also suggest that when the observation error structure is unknown, using large observation error can help to improve the analysis.

The OEM designed here aimed at specifying different observation errors used in the EnKF since the attenuated observation can contain nonuniform errors. It is mainly based on an analytical empirical reflectivity observation error relation. More realistic error models are possible. Because attenuation occurs along the radar beam propagation path, a more realistic OEM could consider the beam propagation path and radar range. In the current AEM, smaller observation errors are assigned to large reflectivity values. In reality, large observation error can exist in strong echoes because of uncertainties with the microphysical processes, including uncertainties with the observation operator. In current MSEM, the PIA related error is designed as simply Gaussian distribution, but in reality it should be spatial related. More sophisticated OEMs that address these issues should be investigated in the future.

Chapter 4: Optimal Design of Radar Networks: Expansion of CASA IP1 Network and Impacts on Squall line Analysis and Forecast

CASA radars are designed and deployed to operate as a network. It had been previously planned to expand the IP1 network with more radars to probe the benefit of X-band radar network consisting of a much larger number of radars. How to design the radar network appropriately and maximize the benefits is an important issue to consider before actually deploying the radars. In this study, different possible configurations with 4 IP1 radars with 2 hypothetical radars are tested through OSSEs using the same EnKF framework with attenuation correction capability discussed in Chapter 2 and examined in Chapter 3 to provide guidance for site selection.

4.1 The truth simulation

The truth simulations (nature runs) are initialized in almost the same way as in Chapter 3, except that the domain is larger to keep the QLCS inside the model domain during the assimilation and forecast. The model domain is therefore increase to 267×195 horizontal points with 1.5 km resolution, corresponding to 395×288 km in physical size (FIG 4. 1). The evaluation of the impact of possible radar network configuration under the effects of storm motion is desired. Thus two sets of squall-line systems with different propagation speed are created. We create a slow-moving QLCS with propagation speed of 15 ms^{-1} using an initially homogeneous storm environment defined by a two-layer shear wind profile (identical to that in Chapter 3). The fast-moving QLCS (propagation speed = 25 ms^{-1}) is generated by adding 10 m/s u wind to the slow-moving QLCS. To trigger the storms, a line of 4 K thermal bubbles sit from north to south located at x=50 and z=1.5 km, along with additional perturbations of 5%.

The radii of these bubbles are 9.6 km in x-direction, 20 km in y-direction, and 1.5 km in the z-direction, and the distance between bubbles is 9.6 km. The southernmost bubble is 9.6 km from the southern boundary. The number of vertical layers is 35. The horizontal grid spacing is 1.5 km, and a vertically stretched grid with a minimum vertical resolution of 100 m is used to better resolve the lower atmosphere. Open conditions are used at the lateral boundaries. A wave radiation condition is also applied at the west and east boundaries, while a periodic condition is used at the north and south boundaries. Free-slip conditions are applied to the top and bottom boundaries. The length of the simulation is four hours.

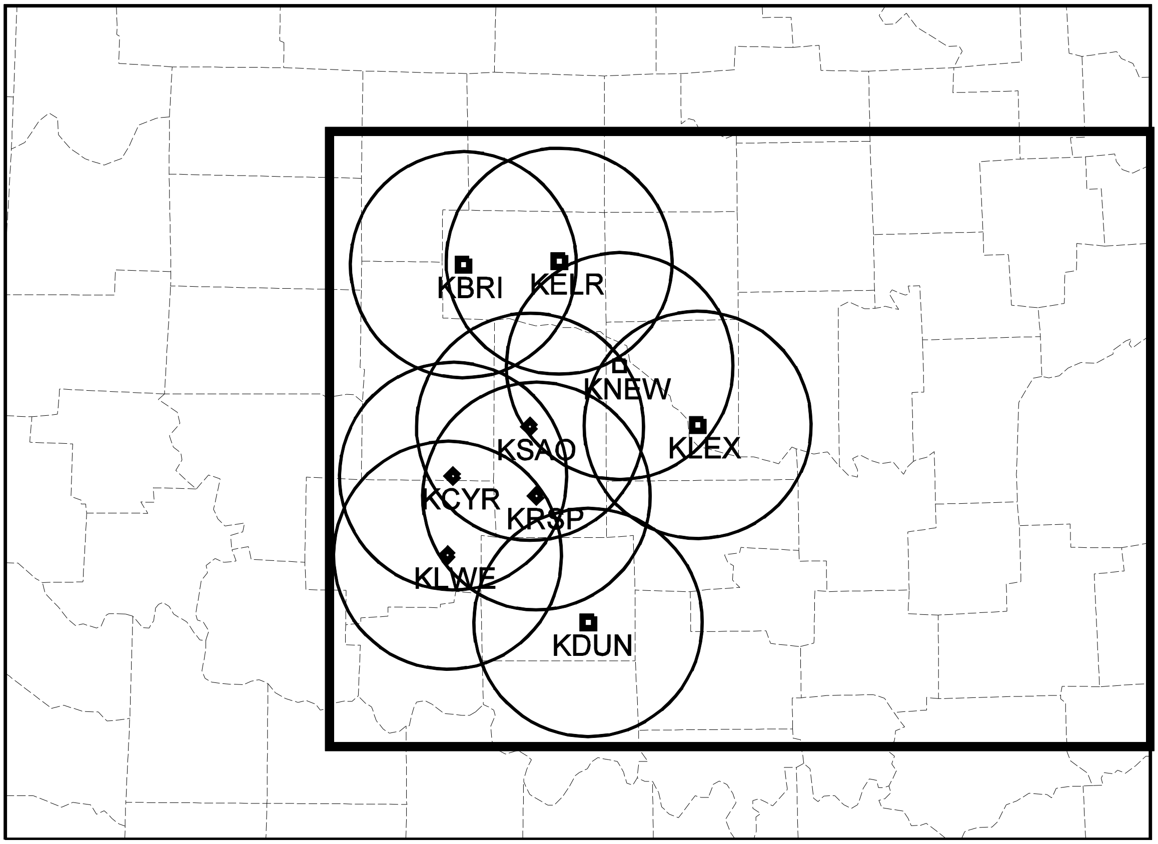


FIG 4. 1 Map of the computational domain. The interior black rectangle denotes the domain over which quantitative verification statistics are calculated. The diamonds are locations of CASA IP1 radars and the small squares represent locations of proposed additional CASA radars. The circles represent the 40-km-range rings of CASA radars

4.2 Simulation of radar observations

The four X-band polarimetric CASA IP1 radars used in OSSEs are located near Chickasha (KSAO), Rush Springs (KRSP), Lawton (KLWE) and Cyril (KCYR) in Oklahoma. The proposed new locations for additional radars are mainly north or east of the IP1 network, like Newcastle (KNEW), El Reno (KELR), Bridgeport (KBRI), Lexington (KLEX), and Duncan (KDUN), as shown in FIG 4. 1. The new radars will have a maximum range R_{max}= 40 km, which is the same as the IP1 radars. The possible configuration of expanded radar network will be 4 IP1 radars plus any 2 radars selected from the new proposed locations.

The simulation of radar data follows Chapter 3, using a Gaussian power weighting function in the vertical for observations simulated on radar elevation levels (plan position indicator planes). In the horizontal, the data are assumed to have been interpolated to the model Cartesian coordinates (the horizontal locations of model grid columns). The effects of Earth curvature and beam bending due to vertical change of refractivity are taken into account using the 4/3 effective Earth radius model discussed in Doviak and Zrnic (1993). The velocity is projected to the direction of radar beam locally to give the simulated radial velocity. Radars are assumed to operate in CASA radar storm scan mode, having 10 elevations with one volume scan every 5 minutes and a 1° beam width. The attenuation effect of X-band radar data is properly taken into account by calculating the expected attenuation within the forward observation operators using the estimated atmospheric state following Chapter 3.

Gaussian-distributed random errors with a mean of zero and standard deviations of 1 m/s and 2 dBZ are added to the simulated radial velocity, V_r, and reflectivity, Z, data sampled from the truth simulation in linear domain, and then become non-Gaussian when they are transformed to the log domain.

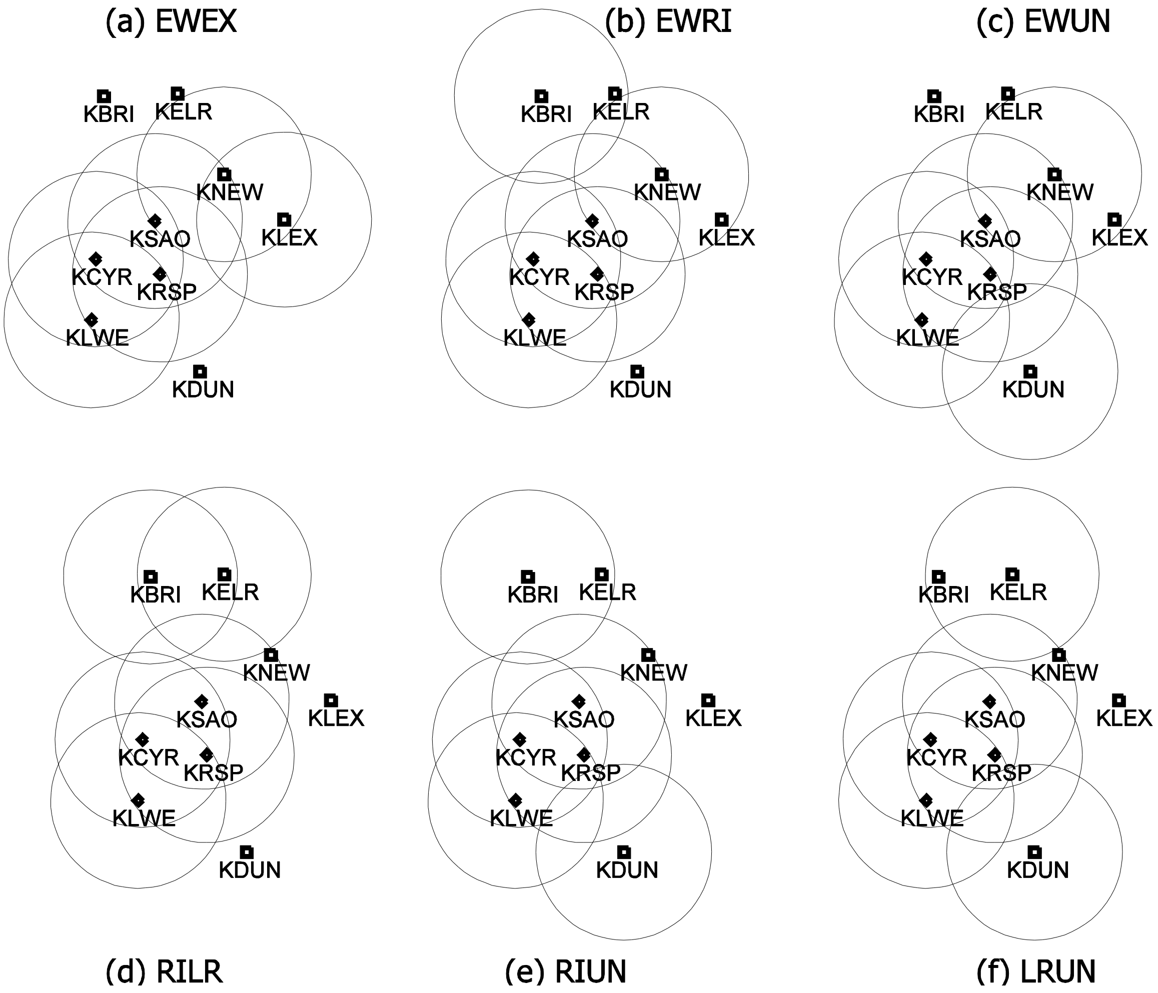
4.3 Data assimilation experiments

A 40 member ensemble is used all experiments. For slow-moving squall line experiments (FIG 4. 3a), the assimilation of reflectivity and radial velocity starts at 110 min of model time and is repeated at 5 min intervals until 165 min. While for fast-moving squall line experiments (FIG 4. 3b), the assimilation starts at 80 min and ends at 135 min, since the simulated storm moves faster than the one in slow-moving squall line experiments.

Since observations with severe attenuation usually contain larger errors than data with little or no attenuation, and zero reflectivity measurements contain large uncertainty (because we are not sure if they are due to complete attenuation), Chapter 3 has shown that using an empirical analytical observation error model or AEM to specify the observation error variance in the EnKF DA can substantially improve the analyses. Thus the AEM, which assigns large observation errors for small reflectivity values, is employed in all experiments. The covariance localization radius used in the filter is 6.0 km, which is 4 times of the horizontal grid spacing and determined based on prior experimentation. The multiplicative inflation initially proposed by Anderson (2001) and modified by Tong and Xue (2005) is used, which increases the ensemble spread by multiplying the ensemble perturbations by a factor greater than 1 in regions within and close to observed precipitation echo. The covariance inflation factor used in experiments is 7%.

The initial ensemble first guesses are generated by adding spatially smoothed stochastic perturbations with a horizontal correlation scale of 6.0 km into a horizontally homogeneous environment defined by the soundings extracted from corresponding truth simulation at 105 min model time (for slow-moving storm experiments) and 75 min model time (for fast-moving storm experiments) respectively, to make sure the lengths of forecast times are the same before the first analyses are taken. Perturbations are added to u, v, w, θ, q_v, q_c, q_r, q_i, q_s, and q_h at the grid points located within 6 km of significant observed reflectivity (i.e. where reflectivity exceeds 10 dBZ). The standard deviations of those perturbations are 2 m s^{-1}, 2 K, and 0.6 g kg^{-1}, respectively.

As mentioned earlier, the expanded radar network will be 4 IP1 radars plus any 2 radars selected from the list of new proposed locations. The experiments are named after the combination of the last two characters of the two new radars. For example, EWEX means the radar network contains new radars KNEW and KLEX, as well as the 4 existing CASA IP1 radars. FIG 4. 2 details each possible radar network tested in this research. Table 4. 1lists the radar network chosen following assimilation experiments and the corresponding names.



(a) EWEX (b) EWRI (c) EWUN

(d) RILR (e) RIUN (f) LRUN

FIG 4. 2 The composition of radar networks of (a) EWEX, (b) EWRI, (c) EWUN, (d) RILR, (e) RIUN and (f) LRUN. The diamonds are locations of CASA IP1 radars and the small squares represent locations of proposed additional CASA radars. The circles represent the 40-km-range rings of CASA radars.

Table 4. 1 List of radar network names and corresponding radars

RADAR NETWORK	RADARS
EWEX	CASA IP1 + KNEW + KLEX
EWRI	CASA IP1 + KNEW + KBRI
EWUN	CASA IP1 + KNEW + KDUN
RILR	CASA IP1 + KBRI + KELR
LRUN	CASA IP1 + KELR + KDUN
RIUN	CASA IP1 + KBRI + KDUN

Since two sets of storm at different speed are produced here, we have two sets of assimilation and forecast experiments. The prefix ‘slw’ indicates the observations are obtained from slow-moving squall line, while ‘fst’ represents observations generated by fast-moving squall line. For example, experiment ‘fstEWEX’ indicates that in this experiment, fast-moving squall line is used as the truth and attenuated observations from radar KNEW, KLEX, and the 4 CASA IP1 radars are assimilated. FIG 4. 3 shows the timeline of assimilation for each set of experiment. The impact of the length of assimilation time is examined too, by evaluating the quality of 1-hour forecasts starting from different assimilation cycles (12 cycles and 15 cycles, i.e. 60 minutes and 75 minutes). To make the comparison as fair as possible, the same number of assimilation cycles is used in both slow-moving and fast-moving squall-line experiments. All the experiments here assimilate both V_r and Z data.

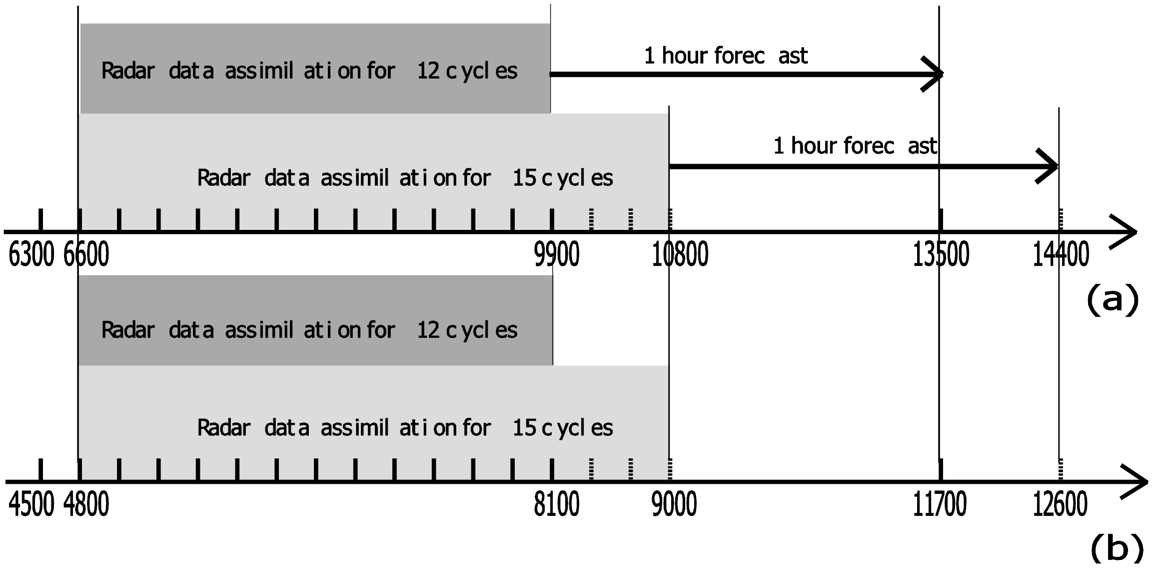


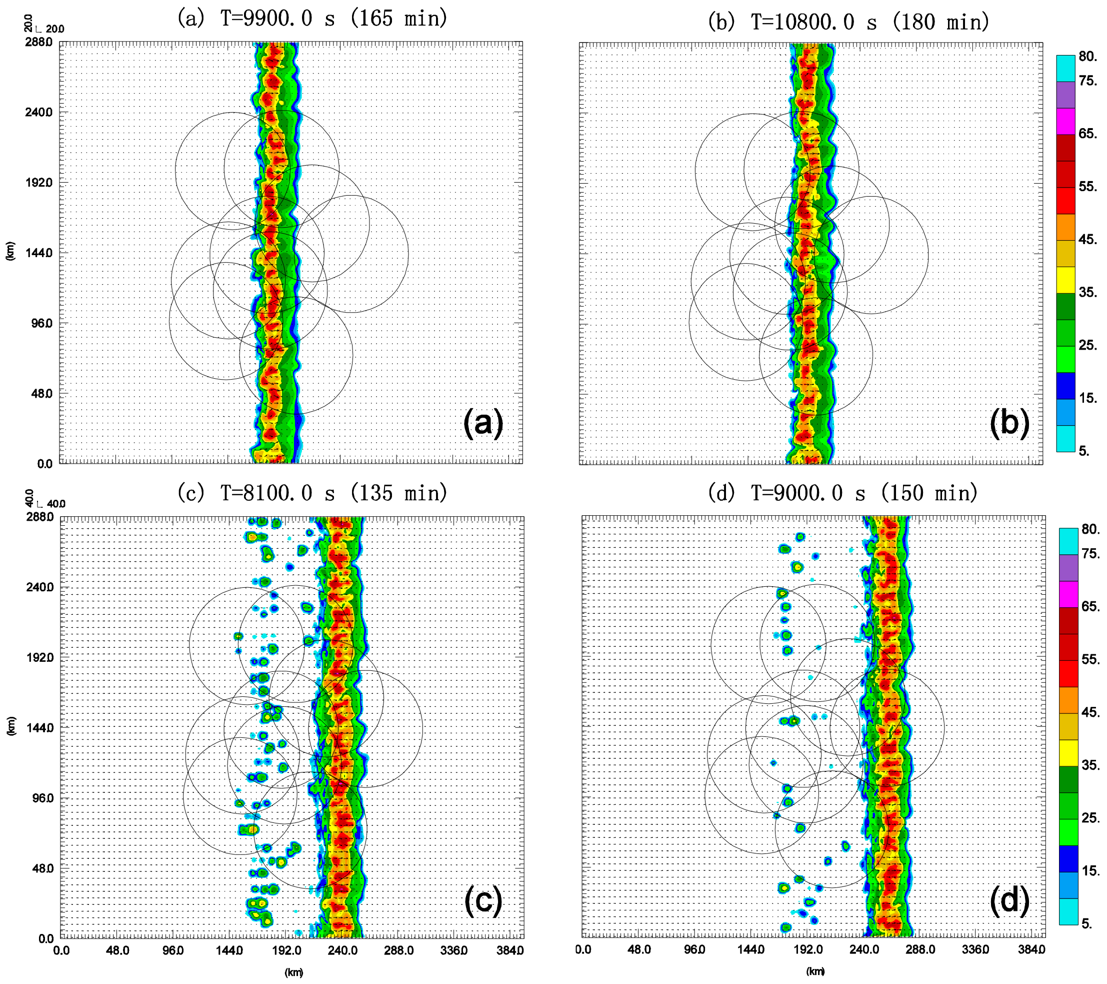
FIG 4. 3 Time line of assimilation and forecast for (a) slow moving squall-line experiments and (b) fast moving squall-line experiments. (Note that the starting times of two sets of experiments are different).

4.4 Assimilation Results

First we examine the quality of state estimation, i.e. the analyzed individual model state variables, by looking at the root-mean-square (RMS) errors of ensemble mean analyses during the analysis cycles. The verification domain is smaller than model domain, as shown in FIG 4. 1, and is a rectangular zone that covers the radar network and downstream area, designed to avoid large errors from lack of radar coverage, thus allowing us to focus on the difference between each type of radar network while still having enough room for forecast verification. The RMS errors are calculated against the truth fields in regions where the truth reflectivity is greater than 10 dBZ inside the verification domain. When describing the results, for convenience and clarity, the west side of the network coverage will be called the “upstream area”, while the east side of the network coverage will be called the “downstream area”, since the simulated storms move from west to east.

4.4.1. Slow-moving squall-line experiments

As mentioned in section 4.3, for slow-moving squall line experiments, radar observations are assimilated every 5 min and the first analysis is performed at 110 min. The length of assimilation window is 60 minutes (assimilation ends at 165 min, 12 cycles, short assimilation window hereafter) or 75 minutes (assimilation ends at 180 min, 15 cycles, long assimilation window hereafter), thus the impact of length of assimilation time can be examined as well. As shown in FIG 4. 4b, the squall-line is still located inside radar network at the end of the long assimilation window, thus the issue of the squall line moving out of radar network coverage can be neglected.


(a) T=9900.0 s (165 min)

(b) T=10800.0 s (180 min)

(c) T=8100.0 s (135 min)

(d) T=9000.0 s (150 min)

FIG 4. 4 Radar reflectivity (shaded; dBZ) of slow moving squall-line simulation at (a) T=9900s (165 min); (b) T=10800s (180 min) and fast moving squall-line simulation at (c) T=8100s (135 min); (d) T=9000s (150 min) at 500 m AGL

Overall, the error levels of model state variables in each experiment decrease (FIG 4. 5) during assimilation window. The RMSEs of slwEWRI (blue), slwRIUN (sky), slwRILR (green) start from lower values and decrease more rapidly during earlier cycles than the other three experiments, especially for u, θ', q_r, q_i, q_s, and q_h. These three types of network include radar KBRI, which increases the number of observations in the upstream area. In the early cycles, the squall line is mainly located in the coverage area of KBRI, KCYR, and KLWE. More observations are available for EWRI, RIUN and slwRILR, which helps the model to quickly reach a better estimation of the state variables. For estimation of potential temperature, as shown in FIG 4. 5d, error levels of potential temperature of slwEWRI (blue), slwRIUN (sky), slwRILR (green) stay low even at the end of assimilation window. This is mainly because the observations from upstream radars are in locations where the cold pool is still useful for potential temperature estimation even at the later assimilation cycles.

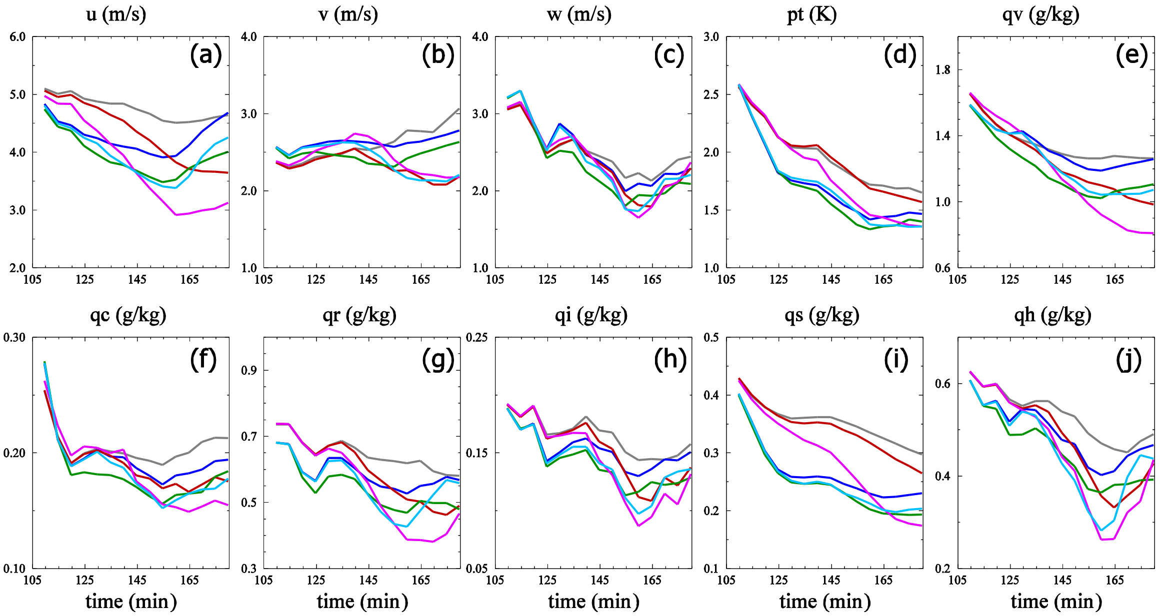


FIG 4. 5 The RMS errors of the ensemble-mean analysis, averaged over points at which the reflectivity is greater than 10 dBZ inside verification domain for (a) *u*, (b) *v*, (c) *w*, and (d) θ', (e) q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h, for experiments slwEWEX (gray), slwEWRI (blue), slwEWUN (red), slwRILR (green), slwLRUN (light purple), slwRIUN (sky). Units are shown in the plots

As storm approaches the eastern edge of the radar network coverage, acquiring more observations downstream starts to become important for better results. The sudden decrease in error levels of slwEWUN (red) and slwLRUN (light purple) is quite striking from 140 min to 160 min. The comparison between these two experiments reveals that under such circumstances, increased radar coverage (slwLRUN) is more important for improved data analysis than greater dual-Doppler coverage (slwEWUN), at least in terms of RMS errors. Downstream radar network coverage at this time can suppress spurious echoes effectively. In FIG 4. 6, the differences of reflectivity between truth and each experiment at the end of the short assimilation window (12 cycles) are presented. The calculation of differences is made in the areas where analyzed storm reflectivity exists, thus clear air regions shown in the north of EWEX coverage (FIG 4. 6a) and also in the northern portion of the EWUN coverage area (FIG 4. 6c) are not included in the calculation. Experiments with observations from radar KDUN or KELR or both show less differences than those without these radar observations. Overall, slwLRUN produces the best analyses, exhibiting the lowest error levels and smallest differences from truth at the end of the long assimilation window (15 cycles).

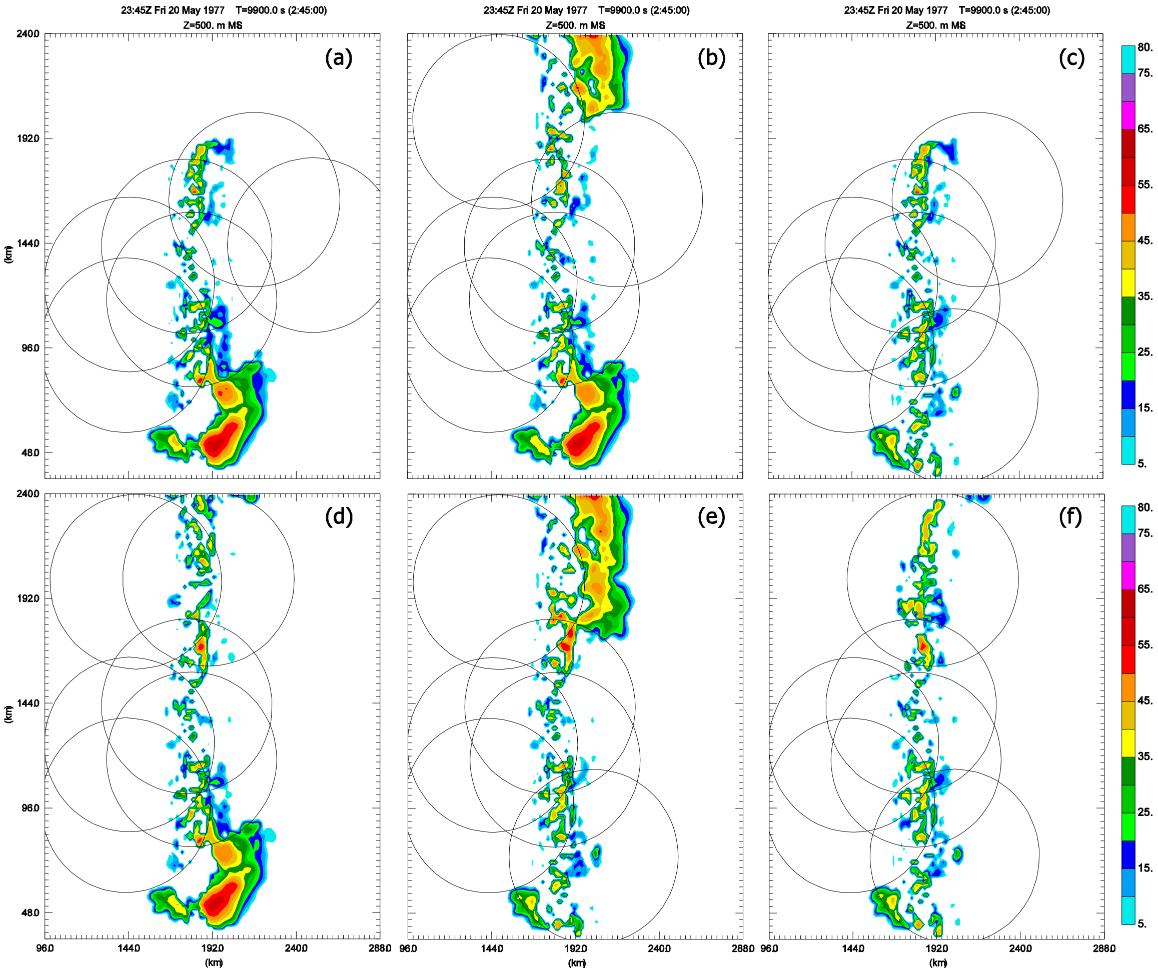


FIG 4. 6 Difference of reflectivity between truth and (a) slwEWEX, (b) slwEWRI, (c) slwEWUN, (d) slwRILR, (e) slwRIUN, (f) slwLRUN at 165 min on level z=500m

The increase in error levels during later assimilation cycles (from 165 min to 180 min) in almost all the experiments is also notable. Such increases are most obvious in *w* and hydrometeor fields, indicating the state estimation become worse during this period. The rise of RMSEs occurs primarily starting from 165 min, when the storm system begins moving out of the multi-Doppler coverage area. Since attenuation effects are considered in our experiment, as discussed in Chapter 3, error levels can increase due to lack of multi- or dual-Doppler observations available. Fully attenuated observations from one radar can be supplemented by observations from another radar with a different detection angle. For this reason, radar networks with more multi- or dual-Doppler radar coverage (EWEX and EWUN in this case) downstream have relatively smaller error increases.

The advantage of assimilating radar observations upstream is more clear when the attenuation effect is excluded (not shown). As discussed in Chapter 3, the accuracy of state variable estimation improves through data assimilation, and the estimate and correction of attenuation also improve. Once attenuation is no longer an issue for DA, storm features can be captured and recovered in early assimilation cycles by assimilating observations with no attenuation (and thus less observation error). For most of the model variables, large error reduction could be found in slwRILR, slwEWRI and slwRIUN up to 12 cycles of assimilation time. Among them, slwRIUN shows the best performance, again suggesting that the larger coverage of the radar network, and thus the increased number of observations, the better the analysis will become. As the storm moves toward the eastern edge of the radar network, error levels of slwLRUN tend to decrease rapidly; slwLRUN has the best performance at the end of assimilation time (15 cycles).

4.4.2. Fast-moving squall-line experiments

As the simulated storm system moves through the radar network coverage rapidly, placing radars upstream gives only a minor advantage in earlier data assimilation cycles, except for relatively good estimates of potential temperature and water vapor at early cycles (FIG 4. 7). For instance, fstEWRI (blue) shows the lowest error level in potential temperature during entire assimilation time (FIG 4. 7d), which indicates that observations at upstream of the storm are vital to the estimate of potential temperature.

Networks with radars at downstream, such as LRUN (light purple), EWUN (red) or EWEX (gray), show the benefits of obtaining more observations during the passage of the storm. Comparison among these three experiments can give us a good understanding of impacts of overlapping radar coverage and spatial coverage (along the storm) on analyses and subsequent forecasts in such fast-moving storm cases. Having more spatial coverage, network LRUN (light purple) tends to produce the lowest RMSEs during most of the assimilation period, up to a model time 150 min (15 cycles) for u ,q_v, q_i, q_s, and q_h (FIG 4. 7a, e, h, i, j). On the contrary assimilating observations from EWEX(gray), which has the smallest spatial coverage but most overlapping radar coverage, does not reduce error much during assimilation for almost all model variables examined here. However, fstEWEX(gray) has the smallest RMSEs at the end of the long assimilation window (15 cycles) for w, q_c, q_r and q_h (FIG 4. 7c,f,g,h). The difference of error level between fstLRUN (light purple) and fstEWEX (gray) at end of assimilation time (15 cycles) is not as big as the difference in the middle of the assimilation window. Considering that during the last several assimilation cycles the squall-line storm system is almost out of the coverage area of LRUN but still partially captured by EWUN (red) and EWEX (gray), the performance of fstLRUN (light purple) is promising.

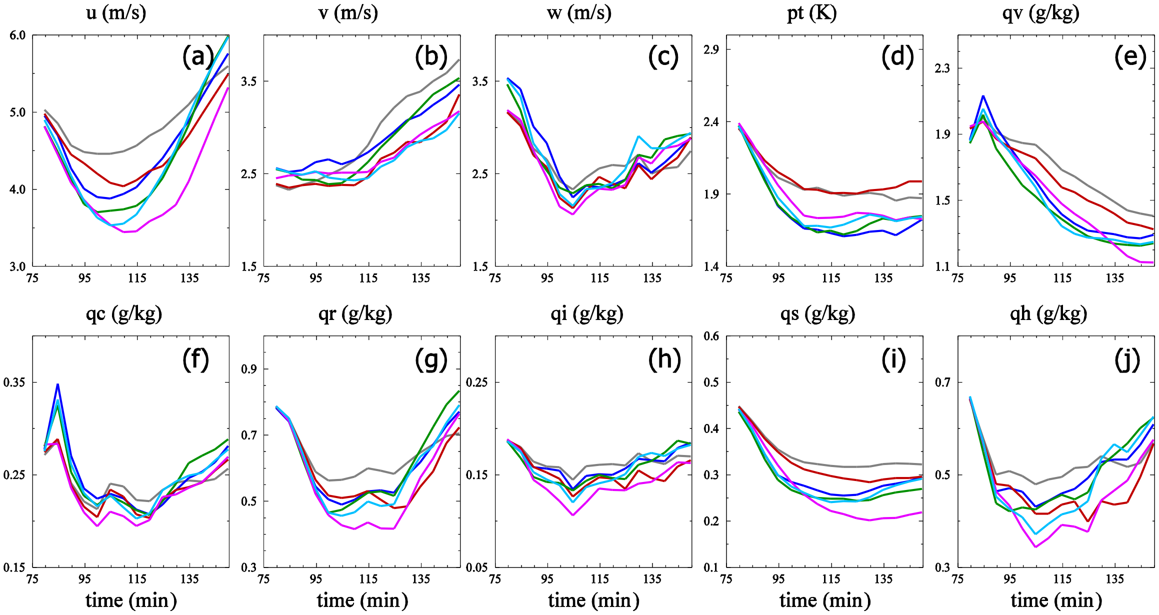


FIG 4. 7 As in FIG 4. 5, but for fstEWEX (gray), fstEWRI (blue), fstEWUN (red), fstRILR (green), fstLRUN (light purple), fstRIUN (sky).

As in the slow-moving squall-line experiments discussed in section 4.4.1, adding radars downstream can effectively suppress spurious echoes. Upstream observations still have a positive impact even at the end of the assimilation window when the storm has already moved out of the upstream radar coverage area. For example, in FIG 4. 8, we note a smaller difference inside the coverage of KNEW for fstEWRI (FIG 4. 8b) compared to fstEWEX (FIG 4. 8a) and fstEWUN (FIG 4. 8c). Although not obvious, the same situation can be found in the comparison between fstRILR (FIG 4. 8d), fstRIUN (FIG 4. 8e), fstLRUN (FIG 4. 8f) (LRUN does not have radar KBRI to provide upstream observations of the northern part of the QLCS). Since among these three types of radar network EWRI is the only one has the upstream radar KBRI in the northern part of the domain, we believe that upstream radars still can have a positive impact even when the storm quickly passes through the radar network coverage area.

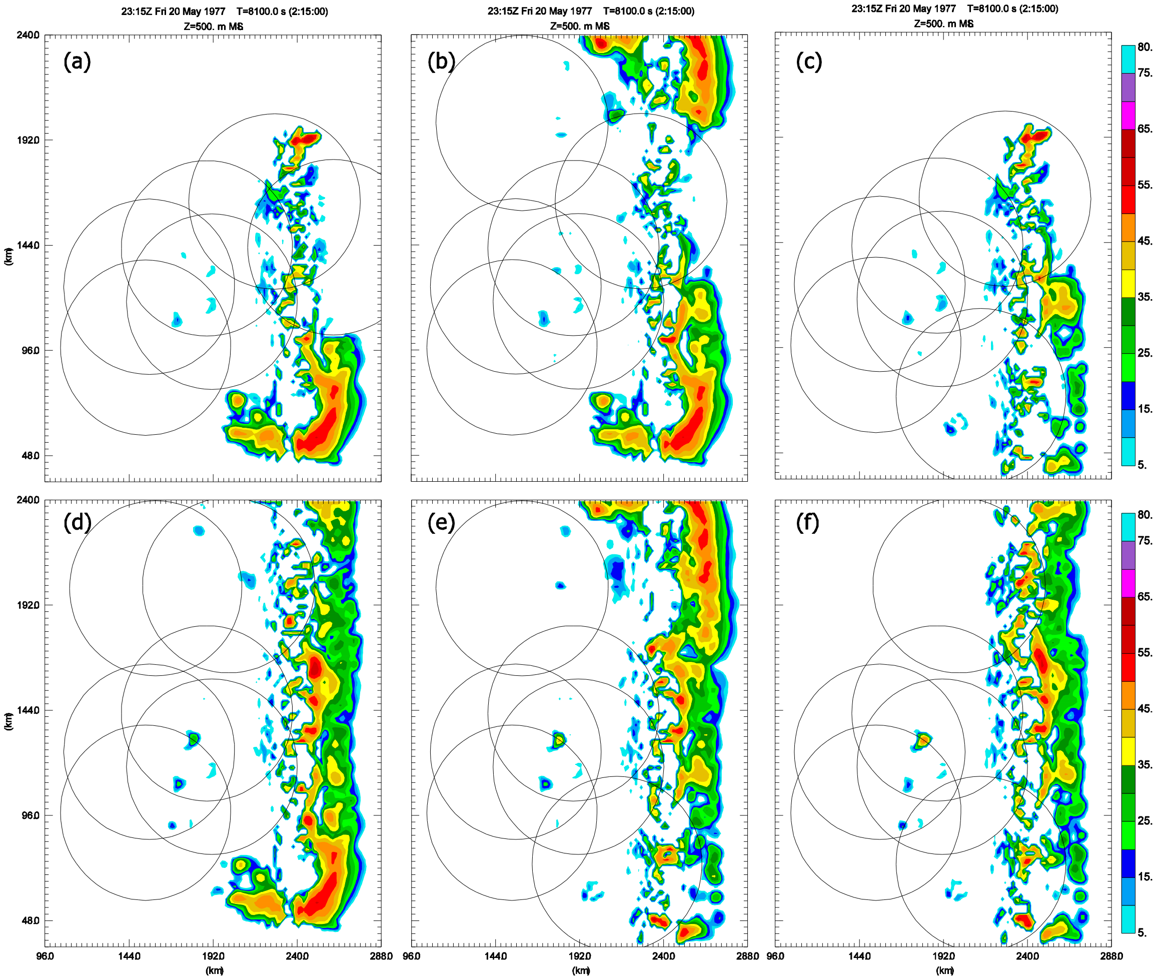


FIG 4. 8 Difference of Reflectivities between truth and (a) fstEWEX, (b) fstEWRI, (c) fstEWUN, (d) fstRILR, (e) fstRIUN, (f) fstLRUN at 165 min on level z=500m

Attenuation is a key factor limiting the performance of upstream radar observations. For example, radar networks RILR, RIUN, and EWRI can still have a positive impact on the analysis during earlier assimilation times when no attenuation exists in the observations. The difference in error level between fstLRUN and fstEWEX becomes significant at the end of assimilation. Another interesting phenomenon is that without attenuation in the observations, fstEWEX and fstEWUN show more advantage in the w and q_r fields.

4.5 Forecast Results

The quality of forecasts produced from ensemble mean analysis of the experiments described in section 4.4 is examined here.

4.5.1. Slow-moving squall-line experiments

In general, assimilating radar data 15 minutes longer does not result in a better 1-hour forecast RMSEs of all experiments grow quickly (FIG 4. 9, FIG 4. 10), and reach the same level at the end of the 1-h forecast no matter whether the forecast is started from the earlier or the later time for both slow-moving and fast-moving squall-line experiments. However, differences in the forecast error among those experiments are still worthwhile to discuss.

In section 4.4.1, we noted that slwLRUN (light purple) tends to produce the best analysis results by providing overall lowest error levels in most model state variables during later cycles of assimilation. Such situation continues in forecast experiments. Although at the end of long assimilation window (FIG 4. 5), slwLRUN (light purple) does not have the lowest error for *w*, it manages to maintain a relatively slow increase of error levels during the 1 hour forecast and has the lowest error levels for *u*, *v*, and hydrometeor fields (FIG 4. 9). Thus, the best 1hour forecast started from model time 180 min is still given by slwLRUN in terms of RMSEs.

Another experiment worth mentioning is slwEWUN (red). As seen in FIG 4. 5, slwEWUN(red) tends to have good performance during later analysis cycles, but not as good as slwLRUN (light purple). When the forecast starts, error levels of slwEWUN(red) have the slowest rate of increase, making the performance of slwEWUN(red) almost as good as that of slwLRUN(light purple), except for v and q_s (FIG 4. 9b,i). As shown in FIG 4. 9, the RMSEs lines of these two experiments remain close and often overlap later in the forecast for u, w ,q_v, q_c ,q_r, q_{i}, and q_h when the forecast is initialed at 180 min (15 cycles) of model time. For θ and q_s, although error levels of slwEWUN (red) increase slowly at the beginning of the forecast, error levels do not reach the level of slwLRUN (light purple) because initial differences in errors level are too large (FIG 4. 9d,i). For forecasts of v, slwEWUN(red) starts with a smaller error, but error increases after almost half an hour (FIG 4. 9b). A possible reason is that the lack of coverage in y-direction results in insufficient information for the model to accurately analyze and forecast the v wind field.

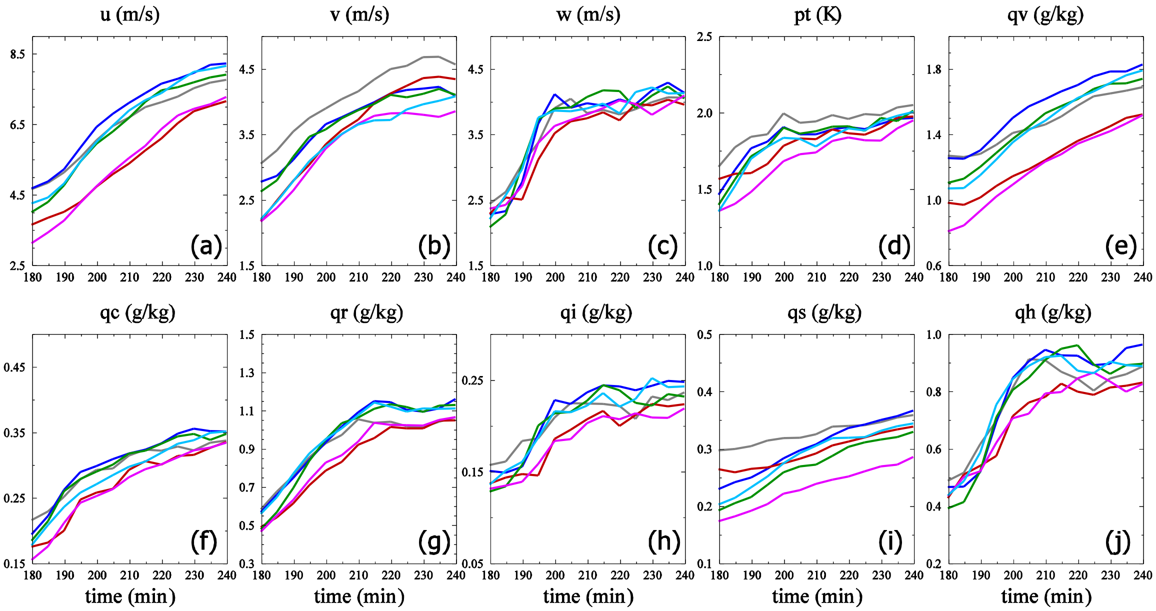


FIG 4. 9 The RMS errors of forecasts averaged over the verification domain for (a) u, (b) v, (c) w, and (d) θ', (e) q_v, (f) q_c, (g) q_r, (h) q_i, (i) q_s, (j) q_h. The forecasts begin from ensemble-mean analysis at t=180 min (15 cycles) of experiments slwEWEX (gray), slwEWRI (blue), slwEWUN (red), slwRILR (green), slwLRUN (light purple), slwRIUN (sky).

When the forecast starts from the end of short assimilation window (12 cycles, 165 min) (FIG 4. 10), the best 1-hour *v* wind forecast is given by slwRIUN. Experiment slwEWEX, the one that has the highest error levels during analysis, surprisingly, produces good *w*, q_{h}, q_i forecasts, and the best rain water forecast (FIG 4. 10c,h,j), probably benefiting from its use of multi- or dual-Doppler observations. In the set of slow-moving experiments, multi- or dual-Doppler observations have more impact on smaller, thunderstorm-scale disturbances in *w* and q_r. Experiment slwEWUN, which assimilates more observations from overlapping radar coverage than slwLRUN does, also manage to give a better forecast of *w* and q_r. However, slwLRUN more accurately forecasts the storm environment by providing smaller errors of *u*, *v*, θ and q_v.

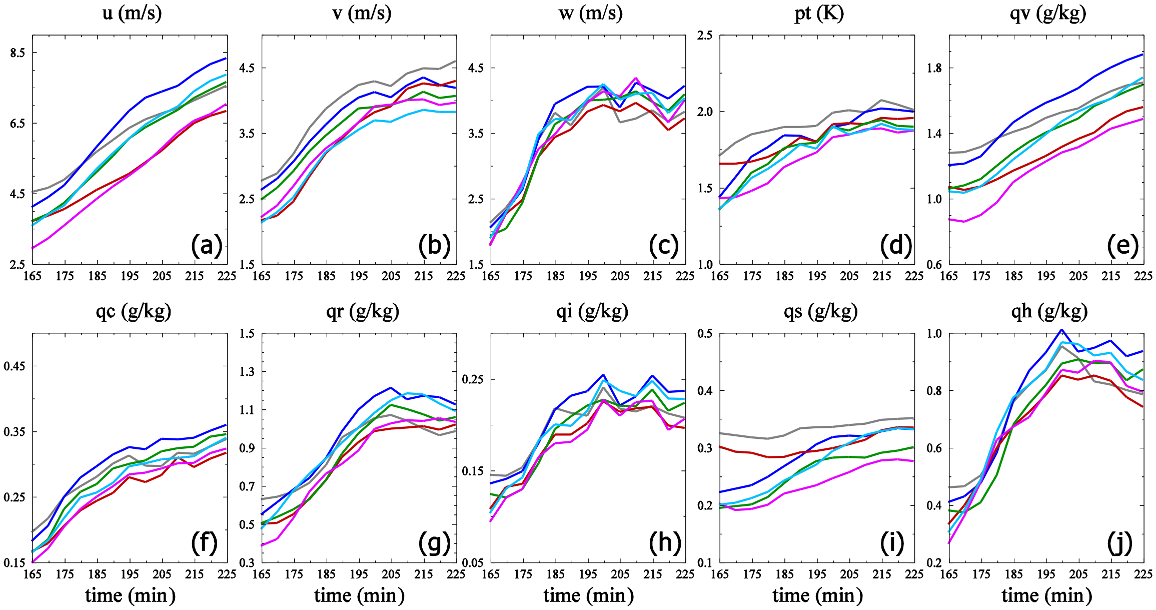


FIG 4. 10 As in FIG 4. 9 but the forecasts begin from ensemble mean analysis at t=165 min (12 cycles)

4.5.2. Fast-moving squall-line experiments

As discussed in 4.4.2, overall the best analysis is produced by fstLRUN, but lowest errors of w and q_r are seen in fstEWUN and fstEWEX which assimilate more downstream and multi-radar observations at the end of the long assimilation window (15 cycles) (FIG 4. 7).These observations not only improve the analysis results during data assimilation but also help forecast smaller thunderstorm-scale disturbances. This time, for both the short assimilation window and the long assimilation window, FIG 4. 11 and FIG 4. 12 shows that fstEWUN (red) and fstEWEX (gray) maintain relatively low increases of RMSE for w and q_r during the forecast. By assimilating more x-direction observations than other experiments, fstEWEX(gray) maintains an even lower error growth rate for forecasts of u. However, due to lack of spatial coverage along the squall line system, fstEWEX(gray) also has the highest error levels for v, θ, and q_v, which means spatial coverage is still critical for forecasting environment conditions.

When starting from different assimilation times, slwLRUN tends to produce a worse forecast of θ when starting forecast the 15th analysis cycle (FIG 4. 11d) than when starting from the 12th cycle (FIG 4. 12d). This did not happen in slwRIUN, in which observations from KBRI were assimilated. This means the impact from upstream observation not only exists in analyses but also in forecasts.

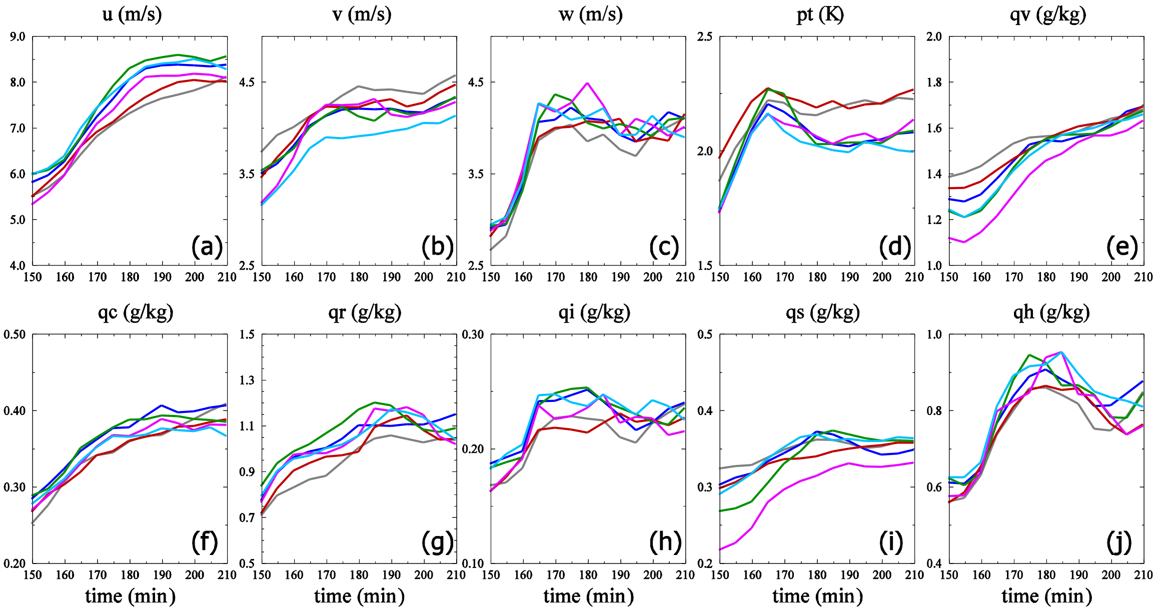


FIG 4. 11 As in FIG 4. 10 but the forecasts begin from ensemble-mean analysis at t = 150 min (15 cycles) of experiments fstEWEX (gray), fstEWRI (blue), fstEWUN (red), fstRILR (green), fstLRUN (light purple), fstRIUN (sky)

Just as in the slow-moving squall-line experiments, assimilating observations from radar network RIUN(sky) which considered both upstream and downstream coverage along y direction improves forecasts of *v*.

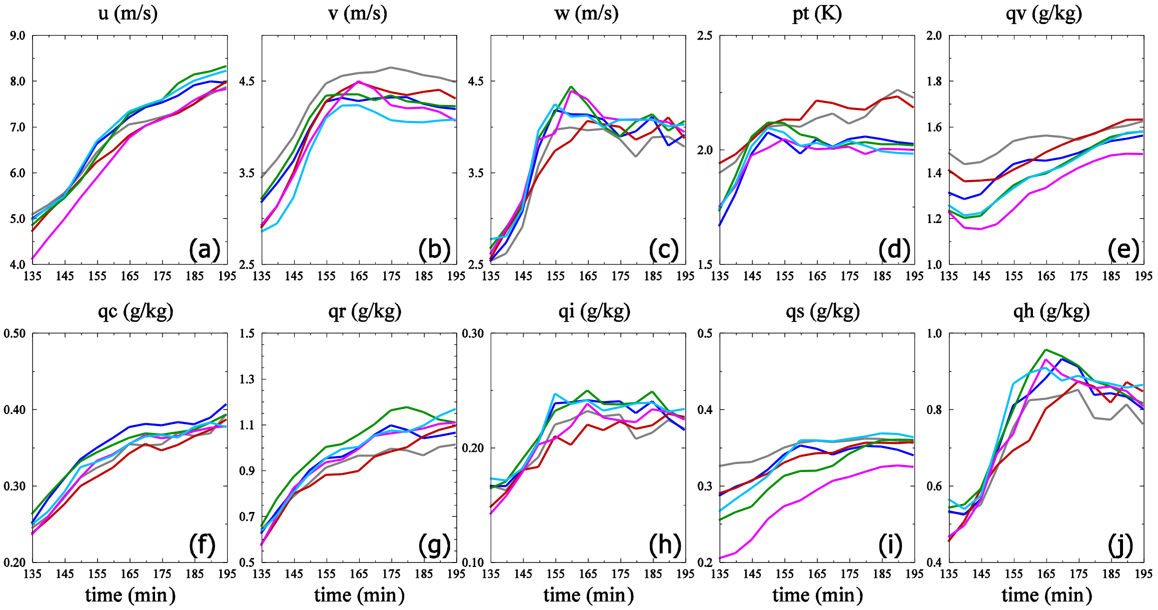


FIG 4. 12 As in FIG 4. 11 but the forecasts begin from ensemble-mean analysis at t = 135 min (12 cycles)

4.5.3. Equitable threat score (ETS)

The RMSEs shown above are used for verification of forecast model state variables. According to its formulation, RMSE is sensitive to large errors rather than being representative of the forecasts as a whole (Jolliffe and Stephenson 2003). Also, the point-by-point calculation is too strict for the storm-scale, especially when spatial displacement errors exist. Alternatively, the continuous values can be categorized, for example, by threshold, and statistics based on the resulting contingency tables can be calculated (Gilleland et al. 2010). The Gilbert skill score (Schaefer 1990), a statistic more commonly known as the equitable threat score (ETS), is a method that measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance. We will now examine the 1-h forecast from every experiment using ETS for radar reflectivity verification.

The ETS is measured at each grid point within the verification domain where reflectivity exceeds a threshold of 15, 25, 35, or 45 dBZ. We calculate ETS at 500 m MSL, which can reflect the benefits from CASA radar low levels detection.

Generally, the best 1-hour forecast is from the experiment assimilating the observations of radar network LRUN (light purple) in both slow-moving and fast-moving squall-line experiments. In detail, when the forecast starts from the end of the short assimilation window (12 cycles) (FIG 4. 13), slwRILR(green) can achieve a good ETS for the 15 dBZ threshold, but it shows no advantage for higher thresholds, indicating that it could not provide much information on storm details at these thresholds. On the contrary, radar network EWUN (red), which provides more overlapping observations than LRUN (light purple), generates a better forecast in strong convection details no matter how long the assimilation windows are. Such good performance exists in both slow-moving and fast-moving squall-line experiments. Assimilating radar observations 15 minutes longer (15 total cycles) (FIG 4. 14, FIG 4. 16) results in insignificant improvement in the forecast for all experiments.

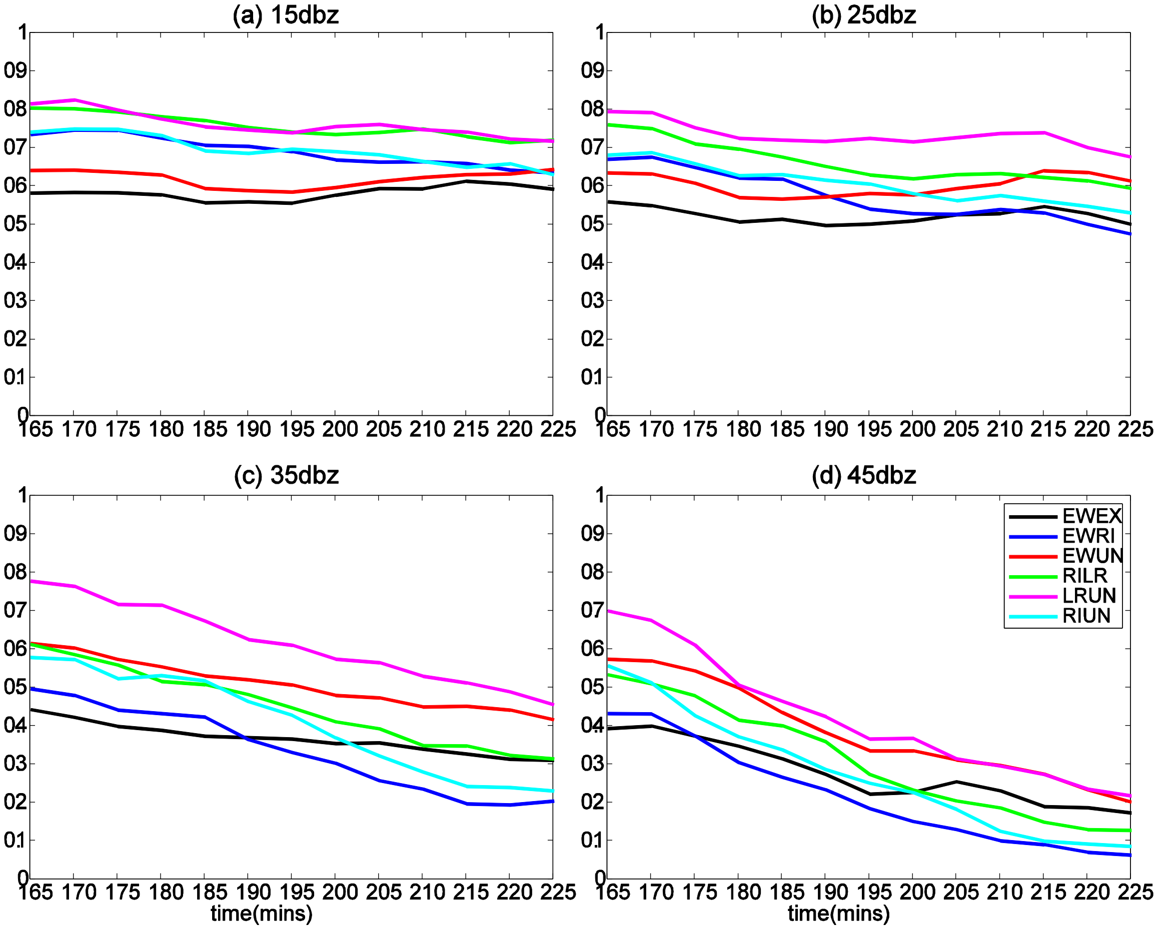


FIG 4. 13 ETS computed over the verification domain for reflectivity threshold at (a) 15dBZ, (b) 25 dBZ, (c) 35 dBZ and (d) 45 dBZ. The forecasts begin from ensemble-mean analysis at t=165 min (12 cycles) of experiments slwEWEX (black), slwEWRI (blue), slwEWUN (red), slwRILR (green), slwLRUN (light purple), slwRIUN (sky).

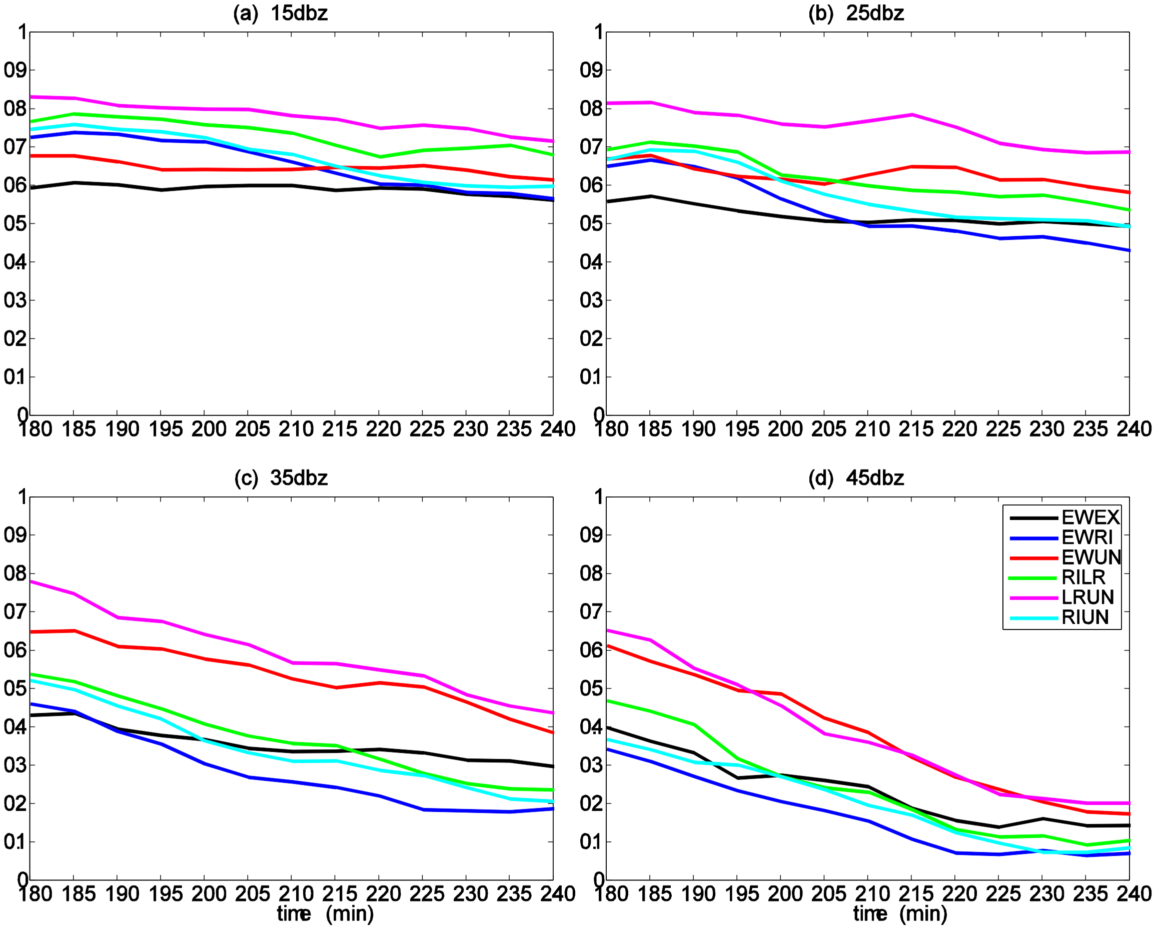


FIG 4. 14 As in FIG 4. 13 but the forecasts begin from ensemble-mean analysis at t=180 min (15 cycles)

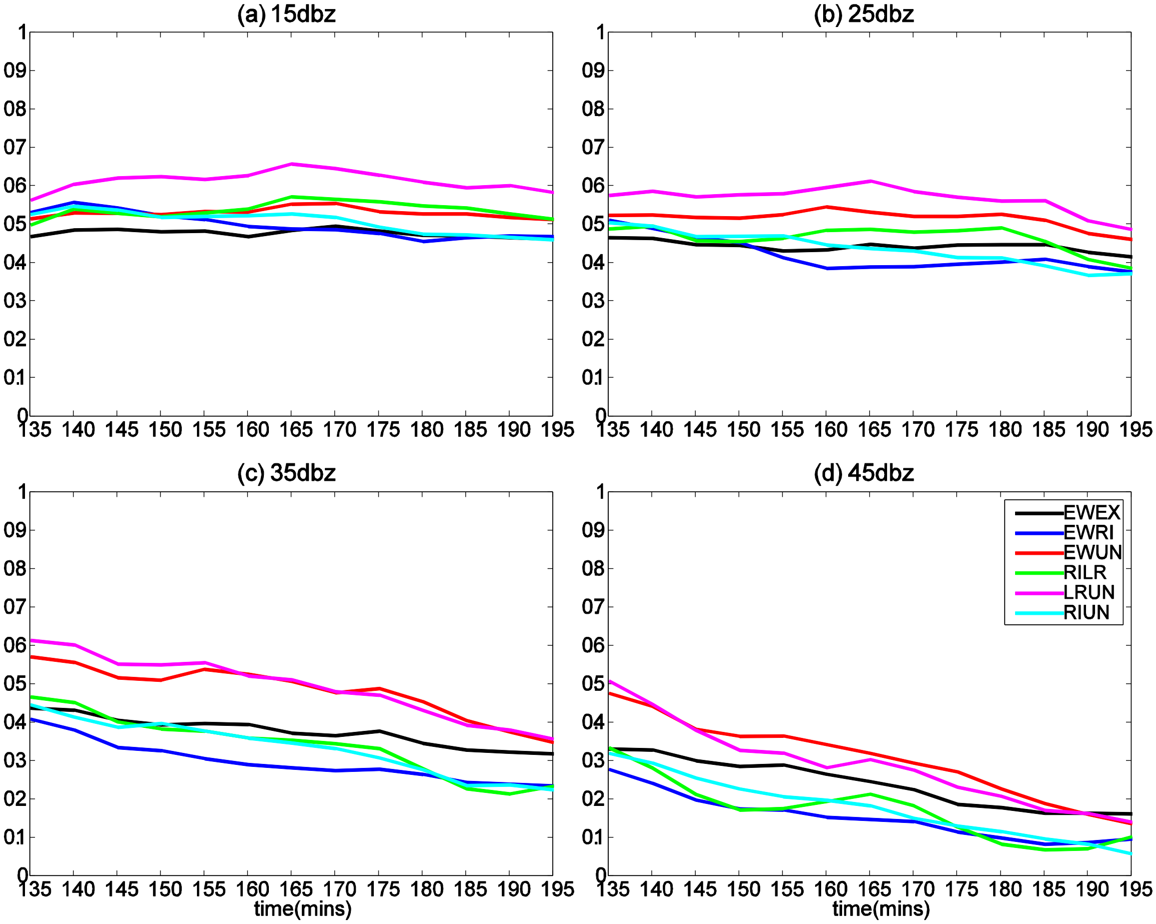


FIG 4. 15 As in FIG 4. 13 but the forecasts begin from ensemble-mean analysis at t=135 min (12 cycles) of experiments fstEWEX (black), fstEWRI (blue), fstEWUN (red), fstRILR (green), fstLRUN (light purple), fstRIUN (sky blue)

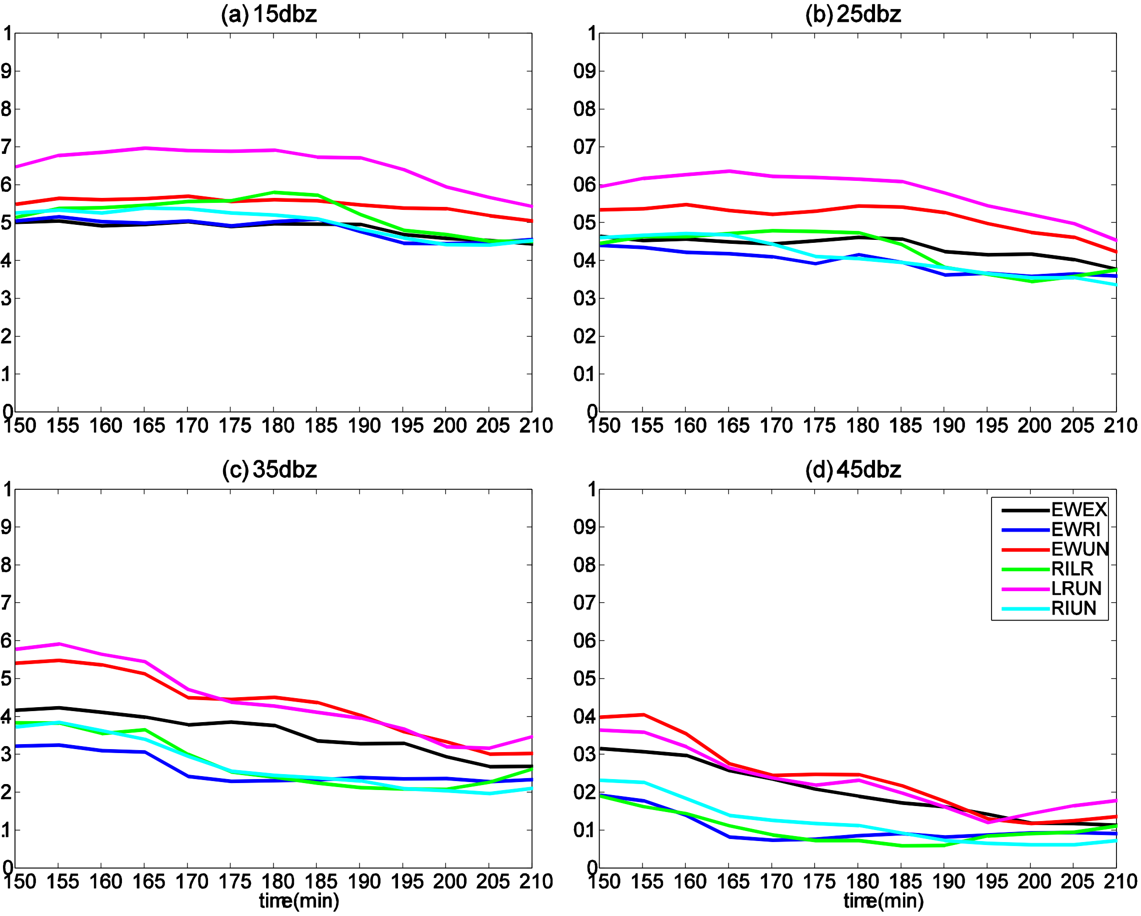


FIG 4. 16 As in FIG 4. 15 but the forecasts begin from ensemble-mean analysis at t=150 min (15 cycles)

4.6 Summary and Discussion

In this study, possible configurations of extended an CASA radar network are tested using OSSEs assimilating data from the 4 existing CASA Integrated Project 1 (IP1) radars plus 2 new, hypothetical radars, using an ensemble square-root Kalman filter. Such experiments can be used to provide guidance for site selection in future X-band radar networks. The unavoidable attenuation problem inherent to X-band radar must be properly accounted for. The method proposed by Xue et al.(2009) and further developed in Chapter 3 proves very effective in dealing with attenuation in observations, thus the same attenuation correction EnKF system is adopted in all experiments. Recent work of Zhang et al.(2011) shows that lower quality data with better coverage is beneficial to hurricane track forecasting. The trade-off between Dual-Doppler coverage and spatial coverage in a convective squall line system is also worth evaluating in this study. To make such an examination, the impacts of storm propagation speed and assimilation time window length on the quality of analysis and forecast are also examined.

For a slow moving quasi-linear convective storm system, with severe attenuation in the observations, a radar network with larger spatial coverage and observations is preferable to one with smaller spatial coverage; though the latter can provide more Dual- or Multi-Doppler observations, during both analysis and forecast. Since the slow-moving storm system is nearly stationary within the radar network, assimilating radar observations every 5 min for an hour is sufficient to obtain the main structure of the storm. More Dual- or Multi-Doppler features can provide details in smaller storm structure, but results in the loss of more information by limiting spatial coverage. Radars located upstream of a storm can aid the model in obtaining a better estimate and forecast of potential temperature. Assimilating observations 15 minutes longer does not improve the forecast significantly because the error levels increase rapidly to saturation levels no matter how low they were initially, which indicates that model error quickly dominates once the forecast starts.

In experiments where the storm system moves quickly through the radar network, radars located downstream become important, providing sufficient observations to suppress spurious echoes during later assimilation cycles. However, the best analysis and forecast of potential temperature is still given by experiments assimilating more upstream observations. Although reducing the spatial coverage, a radar network with more overlapping coverage shows distinct improvements, especially in smaller, storm-scale forecasts. When downstream coverage is present, ETS shows that experiments with more Dual- or Multi-Doppler observations generate a more accurate forecast of convective activities. Because a longer assimilation period has the problem of the storm system moving out of the network, assimilating data longer does not improve the analysis and may let spurious echoes grow and cause deterioration in the forecast. Thus, in our case of both slow-moving and fast-moving squall line systems, 1 hour DA should be sufficient for analysis and forecast.

In general, as guidance to future X-band network site selection, radar network LRUN (CASA IP1 plus KELR and KDUN), the one with larger downstream spatial coverage, tends to provide the overall best analysis and 1 hour forecast. However, in our experiment design, observations from only the X-band radar network are used, and the storm system is larger than radar network coverage. Once other long range radars (e.g. WSR88D) are taken into account to provide full coverage of storm system, spatial coverage will not be an important issue. However, dual-Doppler coverage is always essential to the convective scale structure analysis and forecast, especially when attenuation exists in X-band or other short-wavelength radar networks.

The location of additional radars were proposed by a former IP5 plan, thus the networks examined here are mainly in Oklahoma. Due to the limitation of physical domain size, the impact of different radar network on 1hour warn-on-forecast is evaluated. Also, the weather system considered here is a quasi-linear convective system.

Future studies could use more radars and more extreme situation for radar network design, e.g. more overlapping radar coverage versus larger spatial coverage, or more upstream radars versus more downstream radars. The forecast could also be increased, up to 3 hours long.

Chapter 5: Application of Attenuation-Correcting EnKF to Real CASA X-band Radar Data

Having explored and examined the vulnerabilities and effectiveness of the technique assimilating radar observations directly using EnKF within OSSEs, the technique is next tested with real observations of a CASA X-band radar network. The chosen case is that of the 24 May 2011 Oklahoma tornado outbreak, which is characterized by a synoptic and mesoscale set-up with the development of long-lived supercells with large, destructive tornadoes. A total of 12 tornadoes were reported during the severe weather event, and two EF-4 tornadoes (C1 and D1, FIG 5. 1) developed within two supercells that initiated within the CASA IP1 radar network. The settings and results of the data assimilation and forecast experiments are presented in this chapter.

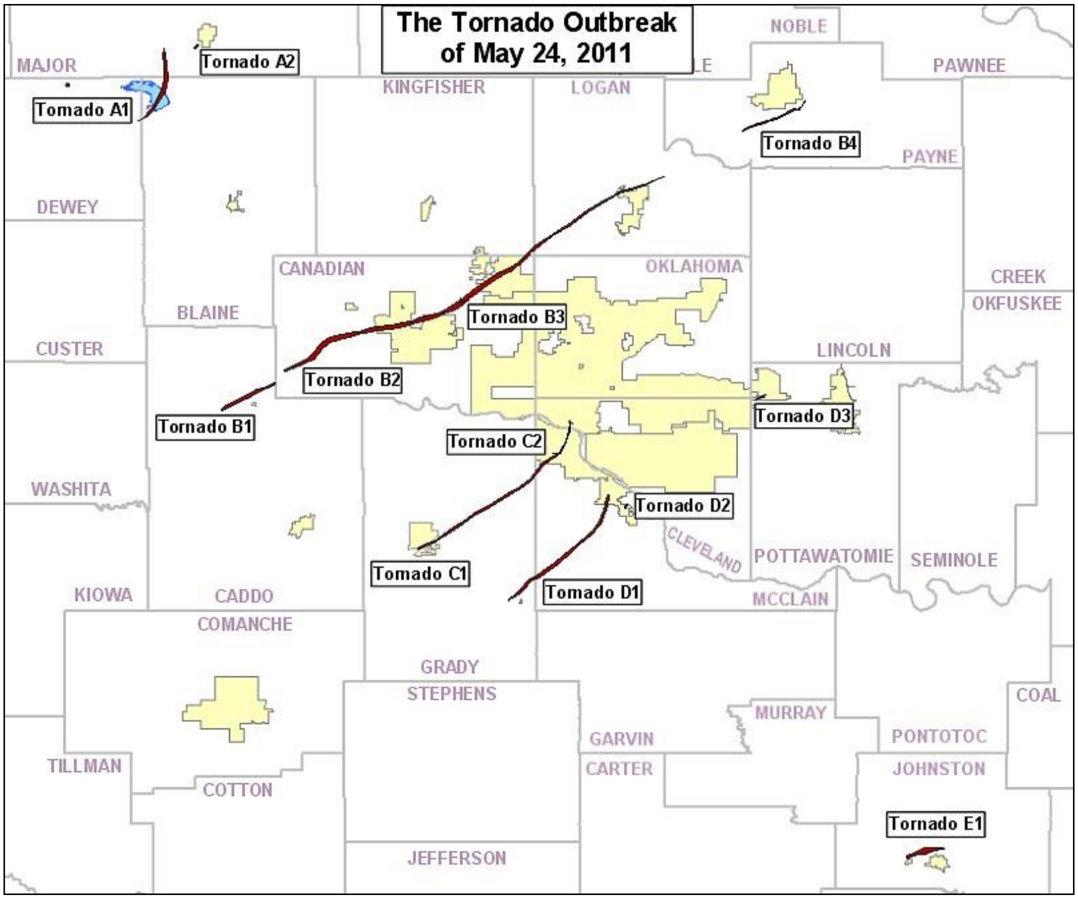


FIG 5. 1 Map of tornado tracks over Central Oklahoma from the National Weather Service with tornado ratings shown below each reported tornado (Tornado tracks are marked with red lines; Yellow shaded denote the metropolitan area) (http://www.srh.noaa.gov/images/oun/wxevents/20110524/maps/overview-800.jpg)

5.1 Case background

On 24 May 2011, a major outbreak of severe weather occurred across much of the southern Great Plains region of the US. One of the ingredients for this severe weather scenario was the approach of a strong jet stream, which was rounding the base of a broad upper level trough located over the Rocky Mountains (CIMSS Satellite Blog http://cimss.ssec.wisc.edu/goes/blog/archives/8237). A deepening low (995 hPa) associated with a negatively tilted shortwave trough emerging from the Rockies facilitated significant advection of moisture northward from the Gulf of Mexico ahead of a pronounced dryline in the Texas panhandle (Fierro et al. 2012). The dryline started to advance east-west direction with visible satellite imagery (FIG 5. 2) showing an expanding field of cumulus from I-40 in western OK south-west direction to near San Angelo in west central Texas after 1830 UTC. Dewpoints ahead of the dryline were as high as 294K (72F). There was high static instability in the low to mid troposphere with mixed layer CAPE (convective available potential energy) values around 3000 Jkg^{-1} (FIG 5. 3) over a broad area. Extreme low-level wind shear resulted in 0-3km storm relative helicity value in excess of 250 m^2s^{-2}, while the threshold for the occurrence of significant tornadoes is 100 m^2s^{-2}.(Fierro et al. 2012)

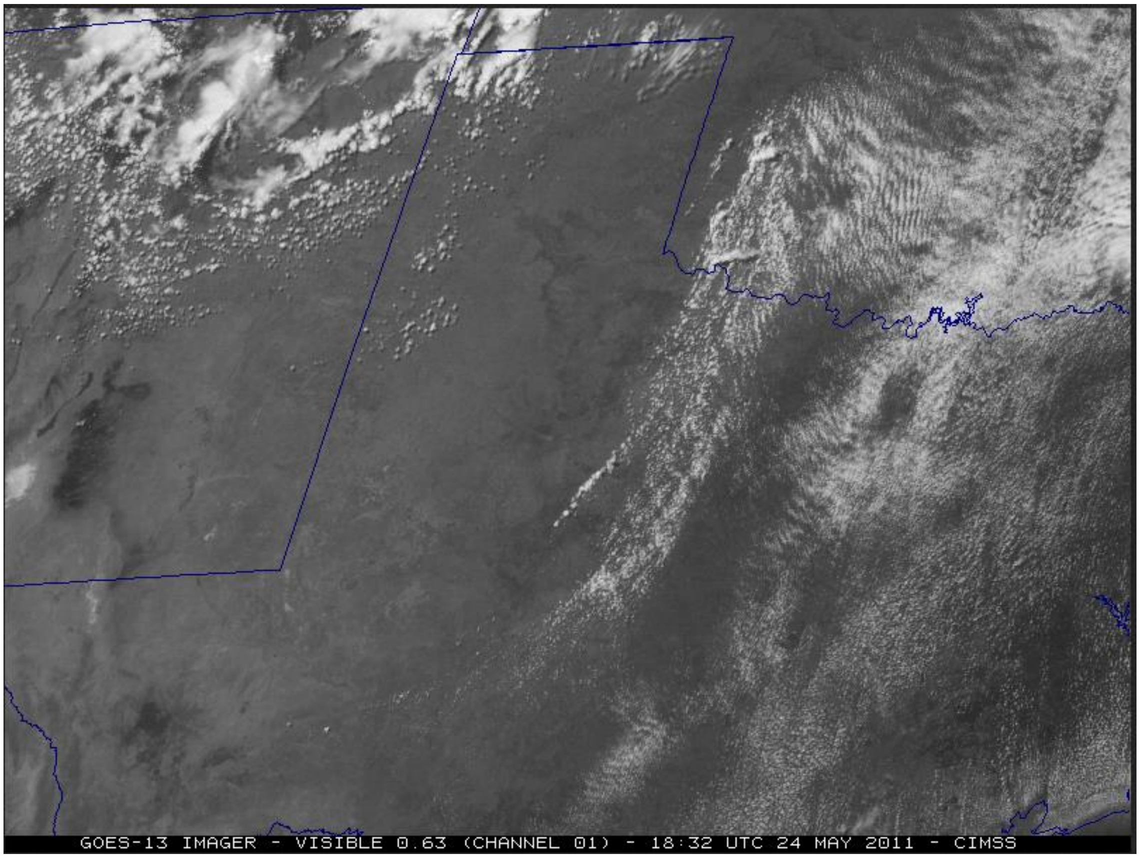


FIG 5. 2 Satellite image of the GOES-13 0.63 µm visible channel at 18:32 UT on 24 May 2011 from CIMSS

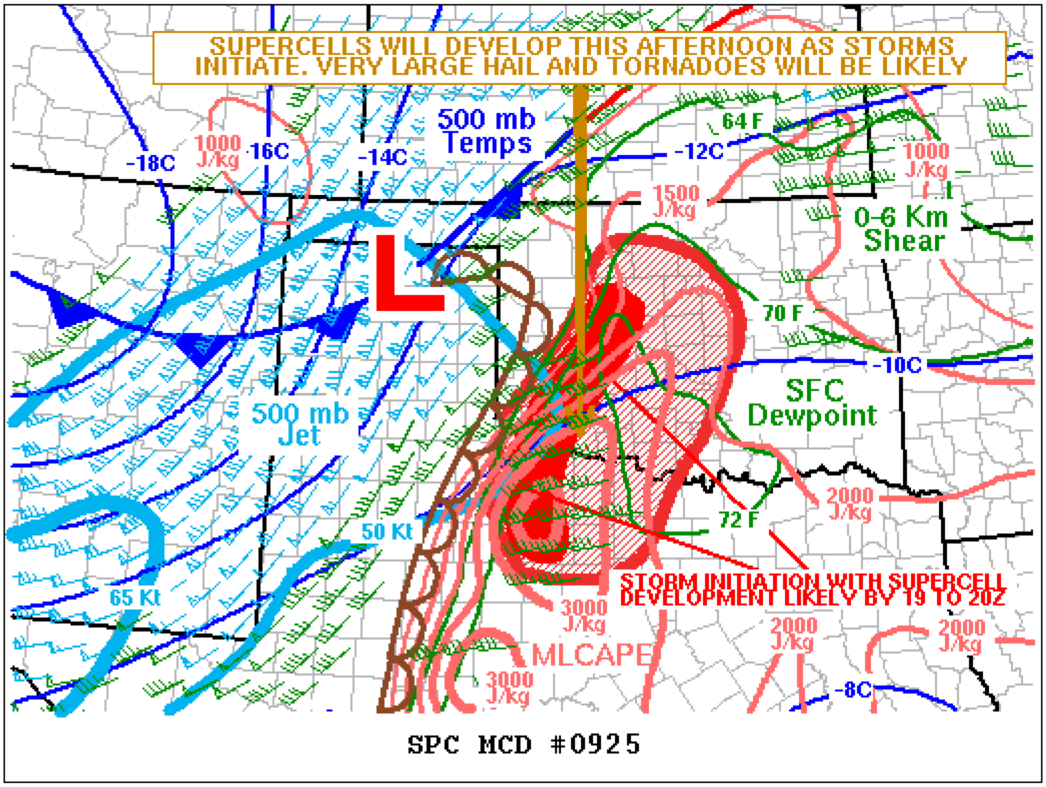


FIG 5. 3 Sketch of the mesoscale discussion analysis (#925) at 1718 UTC from the Storm Prediction Center showing various key environmental factors of the 24th May 2011 tornado outbreak

Within this exceptional thermodynamic and dynamic environment, thunderstorms began to develop along the dryline across western parts of Oklahoma and north Texas around 1900UTC. These thunderstorms intensified rapidly and became severe and tornadic quickly. Tornado C1, the so-called Chickasha-Blanchard-Newcastle tornado, formed around 3km south of Chichasha at 2206UTC and then moved northeast until reaching the SW 149th and Portland Ave in Oklahoma City (FIG 5. 1). This tornado was persistent, lasting 55 minutes, and the damage path length was over 50 km long. Tornado D1, known as Washinton-Goldsby tornado of May 24, 2011, was spawned at 2 miles west of Bradley at 2226 UTC and proceeded to 1 miles west of Goldsby, lasting 49 minutes (FIG 5. 1). Fortunately tornado D1 dissipated just prior to hitting the National Weather Center, the building that houses School of Meteorology of University of Oklahoma and Storm Prediction Center in Norman, OK on May 24, 2011. The storms that spawned these two tornadoes originated inside CASA IP1 radar network coverage, thus the experiments are designed according to these two tornadoes' locations and lifetime.

5.2 Experiment design and data

In Chapter 3, it is noted that incomplete radar coverage of a widespread storm system can lead to large analysis error in OSSEs. Things are generally even worse in real data experiments. Although S-band WSR-88D radar data may reduce the impact of attenuation within the CASA X-band radar observations, we assimilate both WSR-88D radar and CASA radar data to help suppress the growth of spurious radar echoes outside of the CASA IP1 network coverage. The 88D radar observations used are from KTLX in Oklahoma City, OK and KFDR in Frederick, OK.

The experiments are designed as follows: ARPS is initialized from NAM 12 km analyses at 1800 UTC, 24 May 2011, on a 363×363×43 with 1500 m horizontal spacing and stretched vertical grid spacing of 50 m near the ground with an average vertical grid spacing of 500 m. Meanwhile, the NAM 12 km forecasts at 1900 UTC, 2000 UTC and 2100 UTC on 24 May 2011 are used to provide external boundary conditions. To generate initial ensemble, we add 2 m s^{-1} standard deviation Gaussian random perturbations to u, and v, and 0.5 K perturbations to θ' for the first guess interpolated from NAM analyses at 1800 UTC. The extent of the domain is shown below in FIG 5. 5. The small rectangle box in FIG 5. 5 corresponds to the nested/inner domain with 500 m horizontal resolution, which mainly covers the CASA IP1 radar network and the downstream area. After a1 hour “spin up” forecast, Oklahoma Mesonet (OK Mesonet) data are assimilated at 1900 UTC and 2000 UTC in the outer domain. Then the WSR-88D radar data from KTLX and KFDR, together with OK Mesonet data, are assimilated every 15 minutes from 2000 UTC to 2100 UTC. Two hour ensemble forecasts are conducted to provide ensemble boundary conditions for experiments on inner domain. The ensemble members of the inner domain experiments are generated from interpolation from the analyses on the outer domain at 2100UTC. Observations from CASA radar are assimilated every 5 minutes from 2105UTC to 2200UTC.After that, 1 hour (from 2200 UTC to 2300 UTC) deterministic forecasts and ensemble forecasts are conducted, respectively. This period will cover the tornado C1 (formed at 2206 UTC) and tornado D1 (formed at 2206 UTC) genesis. FIG 5. 4 illustrates the timeline for the analysis and forecasting experiments.

Four inner-domain experiments are run: a control experiment (hereafter referred to as “CTRL”) using all pre-corrected reflectivity and radial velocity data provided by CASA; an experiment using radial velocity data and pre-corrected reflectivity data exceeding 20 dBZ (hereafter referred as “PreAC_20DBZ”) from CASA; an experiment using all values of pre-corrected CASA radar observations and applying the AEM (hereafter referred as “PreAC_OEM”) to reflectivity observations; and an experiment assimilating all values of uncorrected reflectivity and radial velocity provided from CASA but using forward observation operator with attenuation and applying the AEM (hereafter referred as “ATTC”). In CTRL, PreAC_20DBZ, the observation error variances are set to 5 dB for reflectivity and 3 m s^{-1} for radial velocity. The radial velocity data are assimilated when reflectivity data exceed 10 dBZ, even in PreAC_20DBZ. The analytical observation error-reflectivity relation in AEM used in PreAC_OEM and ATTC is shown in FIG 5. 6. Besides the relation, the clear air/full attenuation flag is another important part of the AEM, and will be discussed later in the data description. The differences in model setup between these four experiments are summarized in Table 5. 1

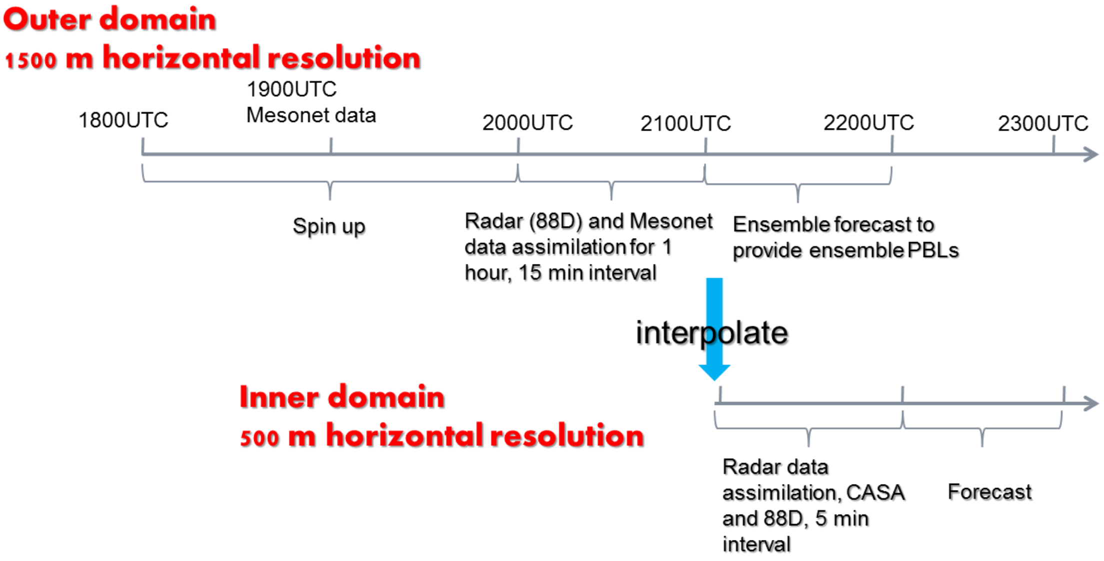


FIG 5. 4 Timeline of the analysis and forecast periods

Table 5. 1 Lists of experiments

Experiments	Observations Assimilated	Attenuation Correction in EnKF	Observation Error
CTRL	R_e and V_r from 88D Pre-corrected R_e and V_r from CASA	N	3 m/s for V_r 5 dB for R_e 88D and CASA
PreAC_20DBZ	R_e and V_r from 88D Pre-corrected R_e and V_r from CASA R_e < 20 dBZ is discard in CASA data	N	3 m/s for V_r 5 dB for R_e 88D and CASA
PreAC_OEM	R_e and V_r from 88D Pre-corrected R_e and V_r from CASA	N	3 m/s for V_r 5 dB for R_e for 88D AEM applied to CASA
ATTC	R_e and V_r from 88D Un-corrected R_e and V_r from CASA	Y	3 m/s for V_r 5 dB for R_e for 88D AEM applied to CASA

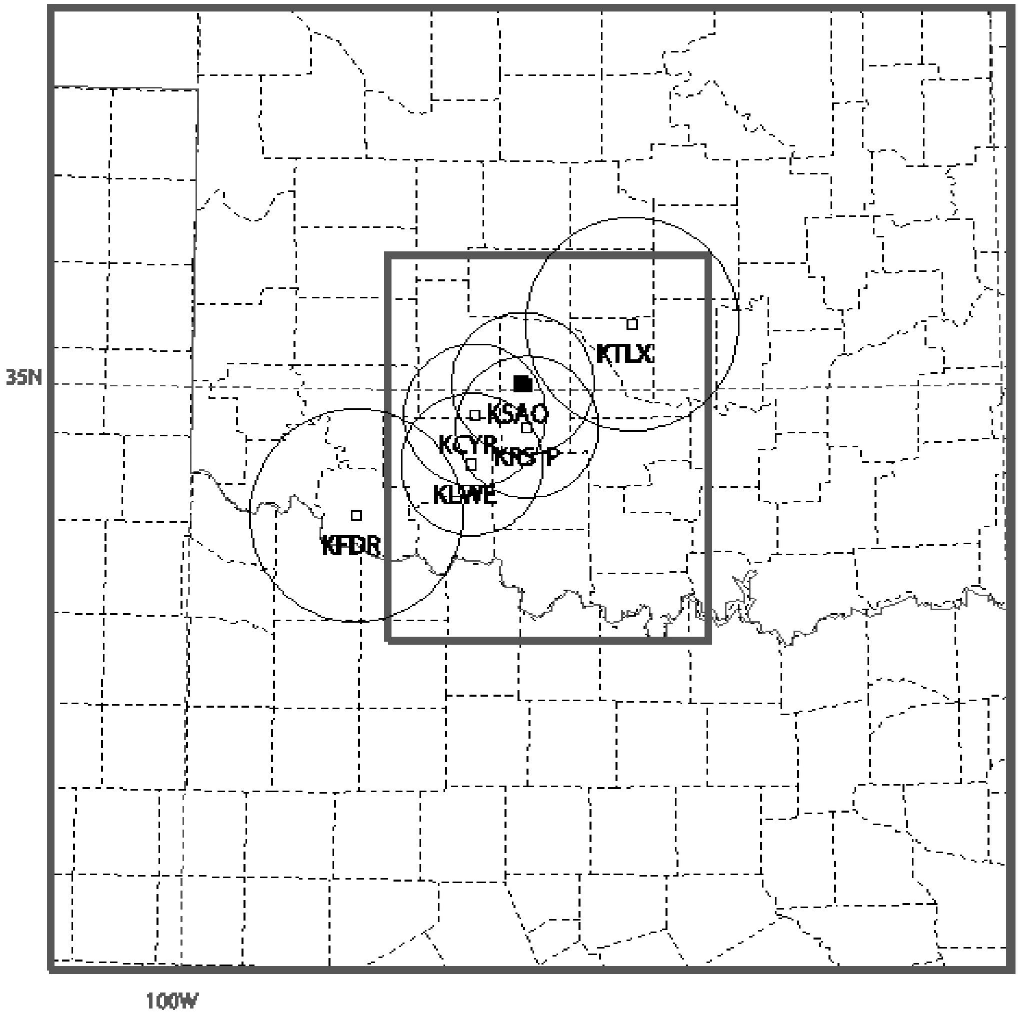


FIG 5. 5 Map of the computational domain. The interior rectangle denotes the domain where the CASA radar data are assimilated. The small squares represent locations of CASA radars and WSR-88D radars. Small circles represent CASA radar 40 km range, while large circles represent WSR-88D radar 60 km range.

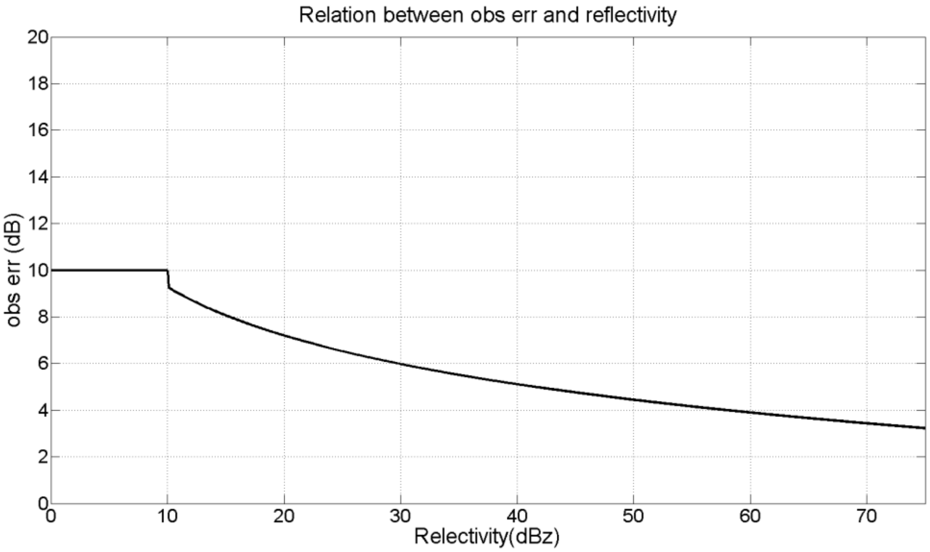


FIG 5. 6 Analytical relations between observation error deviation and radar reflectivity

To counteract the inherent tendency for the ensemble to converge on a solution that is different from the true state of atmosphere because of model error, a method for maintaining ensemble spread is needed (Anderson and Anderson 1999). A covariance inflation factor of γ=1.15 is applied to the prior deviation of each ensemble member from the ensemble mean (Tong and Xue 2005) when assimilating Oklahoma mesonet data on the outer domain. This value of γ was chosen to be large enough to broaden the ensemble distribution, but not so large as to result in an unstable or unrealistic model state. Whitaker et al. (Whitaker 2010; Zhu 2013) proposed an adaptive inflation algorithm to apply to posterior that can restored the ensemble spread lost during the assimilation period,

$$x'_a = x'_a \sqrt{\alpha \frac{\sigma_b^2 - \sigma_a^2}{\sigma_a^2} + 1} \qquad (16)$$

Such a method does not add any inflation where there are no increments, and add more inflation where there are dense/accurate observations. However potentially large spatial gradients in inflation may disrupt growing structures. Thus we use α=0.85 and apply this adaptive inflation only once on the outer domain to increase ensemble spread, at 2000UTC when the background covariance is more reliable than the initial model time. For inner domain experiments, after several tests, only γ=1.20 is used, which generates sufficient spread and does not cause model instability. The localization radius used in mesonet data assimilation is 75km, while those used for radar observations are 6 km for V_r and 4 km for R_e.

In all experiments, level-II volume scans of WSR-88D radial velocity and reflectivity from two WSR-88D radars are assimilated at 15 minute intervals on the outer domain, and at 5 minute intervals on the inner domain. The WSR-88D radars used are those located at Oklahoma City, OK (KTLX) and Frederick, OK (KFDR). These two radars can provide enough coverage for the storms that developed in Oklahoma and northern part of Texas from 1900 UTC to 2100 UTC in the outer domain, and from 2100 UTC to 2300 UTC in the inner domain. In experiments using CASA data, aggregate volumes of radial velocity and reflectivity data are assimilated, also at 5 minute intervals. They are from four CASA IP1 network radars: Cyril, OK (KCYR); Lawton, OK (KLWE); Chickasha, OK (KSAO); and Rush Springs, OK (KRSP). Aggregate CASA radar volumes are created by interpolating raw CASA scan data on observed elevations to a uniform radial grid with azimuthal spacing of 1 degree. For each radial, the nearest data before and after the radial are linearly interpolated to obtain the corresponding radial in the aggregate volume scan. If only one scan is available for a given radial, that scan is used with interpolation. If no scans are available, that radial is marked as missing. For both WSR-88D and CASA data, we pick the scan closest to the assimilation time. No time interpolation is applied.

Since observations are assimilated one by one in EnSRF, a different sequence may results in a different analysis. However, the changes of sequence of using radar data show a minor difference on analysis results after several tests. The results presented here are obtained by assimilating radar in the sequence of: KCYR, KLWE, KSAO, KRSP, KTLX and KFDR. A summary of the radars used and their locations are given in Table 5. 2.

Table 5. 2 List of radars used in data assimilation experiments

	Radar Type	Latitude	Longitude	Elevation
KCYR	CASA	34.8739 N	98.2522 W	448 m
KLWE	CASA	34.6239 N	98.2708 W	396 m
KSAO	CASA	35.0314 N	97.9562 W	356 m
KRSP	CASA	34.8128 N	97.9306 W	436 m
KTLX	WSR-88D	35.3331 N	97.2778 W	384 m
KFDR	WSR-88D	34.3521 N	98.9839 W	383 m

FIG 5. 7 shows examples of radar reflectivity from CASA and 88D that are used in four experiments at 2135 UTC, after being interpolated to the aggregate radar volume. The pre-corrected observations are obtained by applying attenuation correction using a method employing differential phase measurements from dual-polarization radar (Gorgucci et al. 2006).The attenuation correction algorithm used had been shown to be quite accurate in retrieving true reflectivity values. However, traditional attenuation correction method can only make correction to where signals are above noise floor as

we mentioned in Chapter 2 except the pseudo-dual-frequency (PDF) method, which requires radar data from systems operating at two different frequencies and is highly dependent upon the assumption of Z_H being equal at two wavelengths. In pre-corrected CASA data, fully attenuated data cannot be distinguished from clear air echo. The same area may have strong echoes when observed by other radars. For example, at 2135 UTC, the northern part of the storm is strong in KSAO, with observed reflectivity as high as 55 dBZ. However, in KCYR, the largest echo is around 45 dBZ, and the northern part of the storm is shown as clear air (FIG 5. 7a). And in KTLX and KFDR, the north part of the storm is even stronger than in KSAO (FIG 5. 7c). In the uncorrected CASA observations (FIG 5. 7b), the storm is even weaker and the inconsistency is even larger. This severe inconsistency that exists in different radar observations at the same time not only decreases the performance of EnKF but also causes imbalances during analysis, which requires many tests to tune the model to obtain a stable analysis, especially for CTRL.

CASA data are obtained using a dynamically-adaptive scanning strategy that is designed to focus on certain storm features. However, when processing data for 24 May 2013, we found that with the exception of the second full azimuth scan, the section scans sometimes were not aimed at the target storms, or only provide partial coverage of the target storms. Since CASA focuses at low level detection, a limited number of elevations (most of the time only 3) are available. The number of effective observations assimilated from a single WSR-88D could be 100 times more than the effective observations provided by a single CASA radar at one assimilation step.

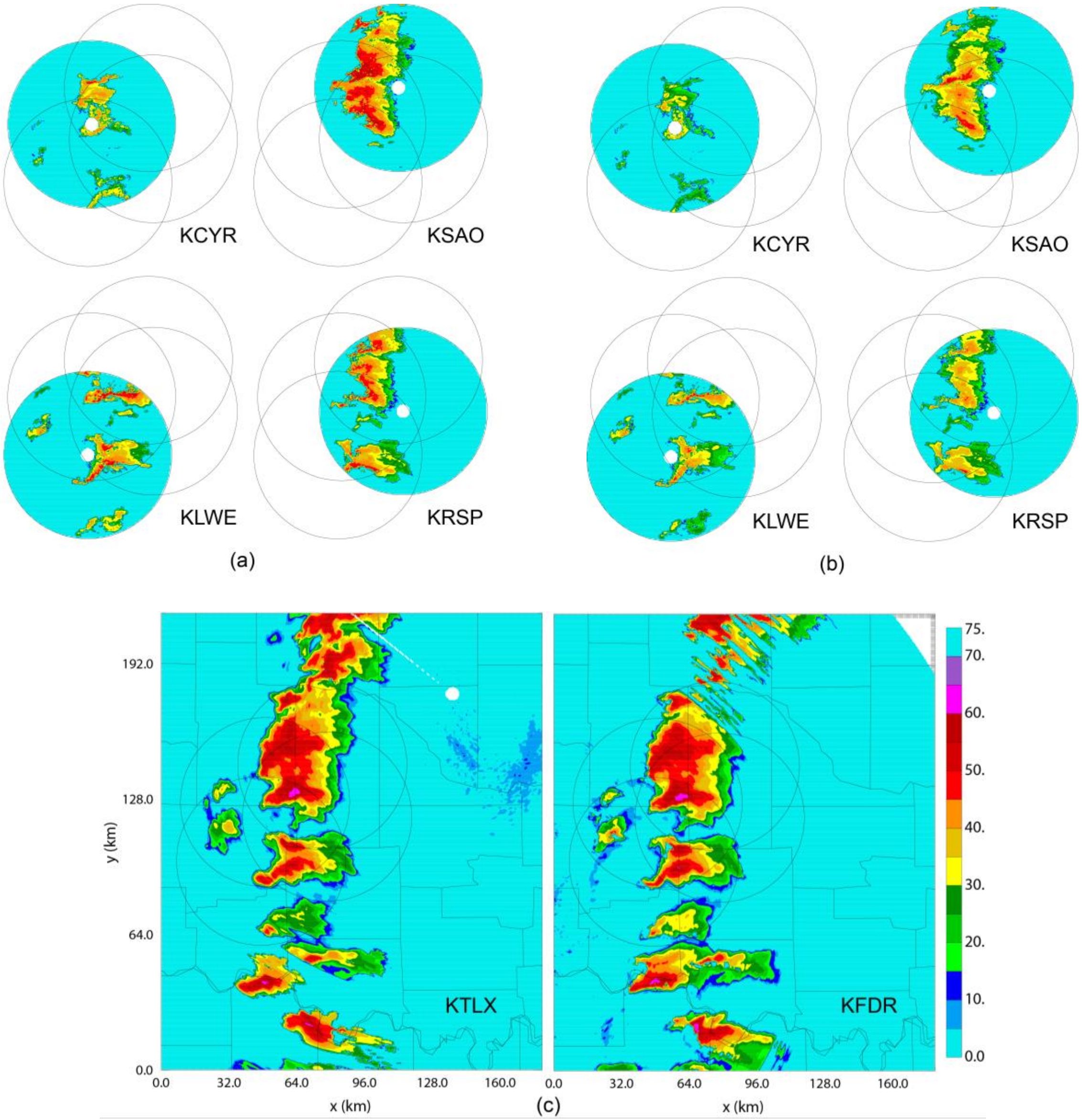


FIG 5. 7 Observed (a) CASA pre-corrected reflectivity on 2 °; (b) CASA un-corrected reflectivity on 2 °; (c) KTLX reflectivity on 0.92 ° and KFDR reflectivity on 0.53 °. Light blue colors indicate the clear air echoes

It is also challenging to make flags to distinguish between clear air echo and completely attenuated echo in real observations. After many tries, we decided to search reflectivity value gate by gate along each beam and flag the all the data once a threshold of reflectivity value is met (in this case, 35 dBZ). Since such a flag is only used in AEM to judge whether the observation is clear air or complete attenuation, reflectivity observations larger than 0 dBZ are not affected by this flag during data assimilation.

5.3 Assimilation results

Before proceeding to experiments to test attenuation-correcting EnKF on the inner domain grids, the analysis on outer domain grids must be promising enough to provide the initial state of ensemble members. FIG 5. 8 shows the reflectivity and total wind field of ensemble mean analysis at 2100 UTC on the outer domain grids (FIG 5. 8a) and the corresponding observations from KTLX and KFDR (FIG 5. 8b,c), which are interpolated to 2.0 km AGL for comparison purpose. At this time, the WSR-88D radar data and the OK Mesonet data have been used four times from 2000 UTC to 2100 UTC. Compared to the observation, a qualitatively accurate portrayal of the state of the convective system has been produced. The experiment produces storm systems with reflectivity structure very similar to that observed by KTLX and KFDR. Assimilation of radar data may result in decreasing ensemble spread. FIG 5. 9, FIG 5. 10 and FIG 5. 11 show the horizontal RMS ensemble spread for u, v, and θ of 40 ensemble members at the end of analysis time on outer domain. The RMS spread of u can reach as high as 10 ms^{-1} in the storm region, and 2.5 in the western part of the outer domain. This number is a little low in eastern Oklahoma. However, since the downstream area is not included in the inner domain, this is acceptable. Also, the ensemble spread is 2 to 3 ms^{-1} near the west and south boundaries, which is good since this is the upstream area (the background wind field is mainly southwesterly). Similar patterns can be found in FIG 5.

10 for *v* and in FIG 5. 11 for *θ*. Thus, it is acceptable to interpolate the ensemble analyses to the higher-resolution inner domain(500 m horizontal grid spacing).

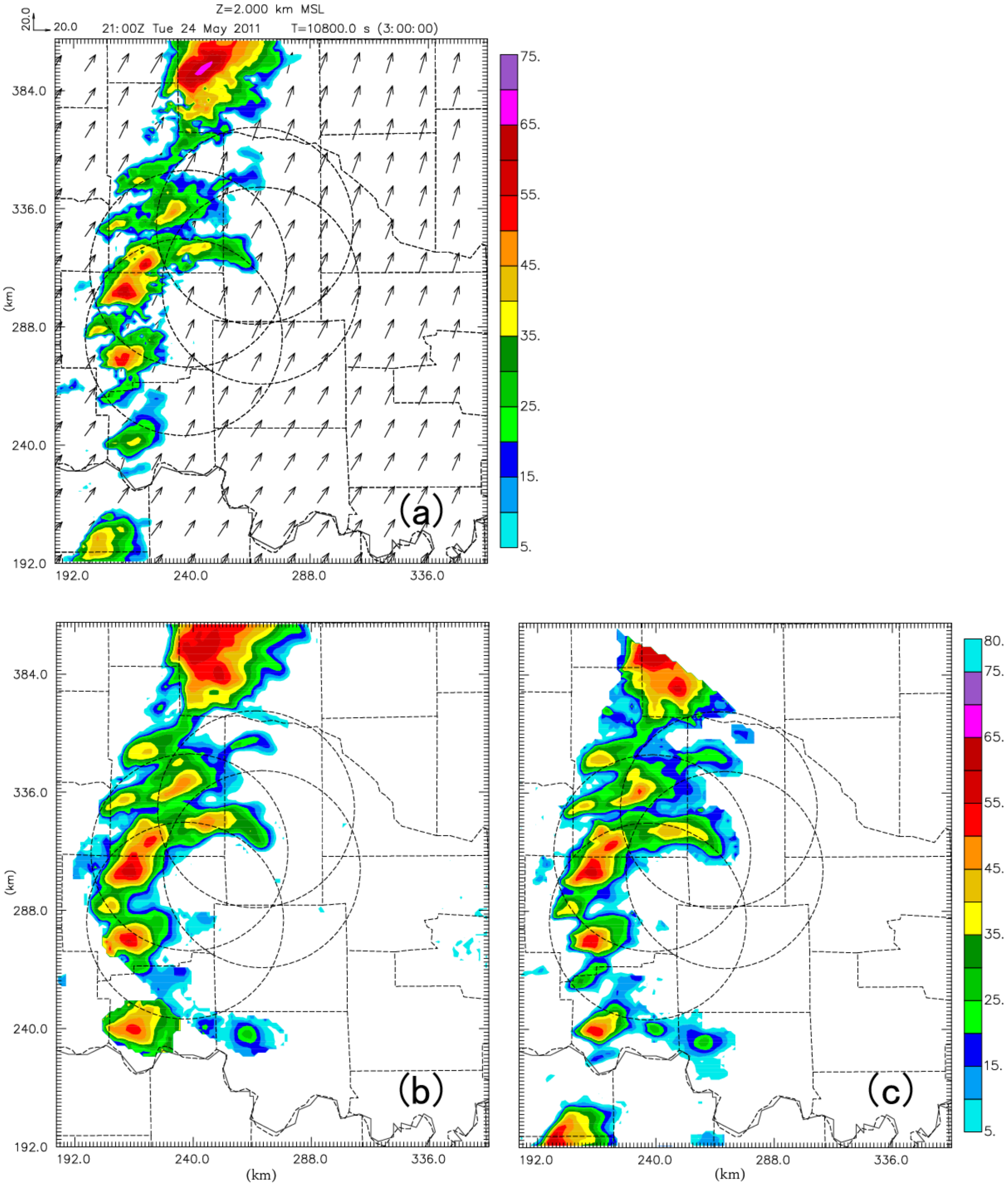


FIG 5. 8 (a) The horizontal wind vectors (m/s), and computed reflectivity from outer domain analyses, and interpolated reflectivity from (b) KTLX and (c) KFDR at 2000 UTC May 24, 2012

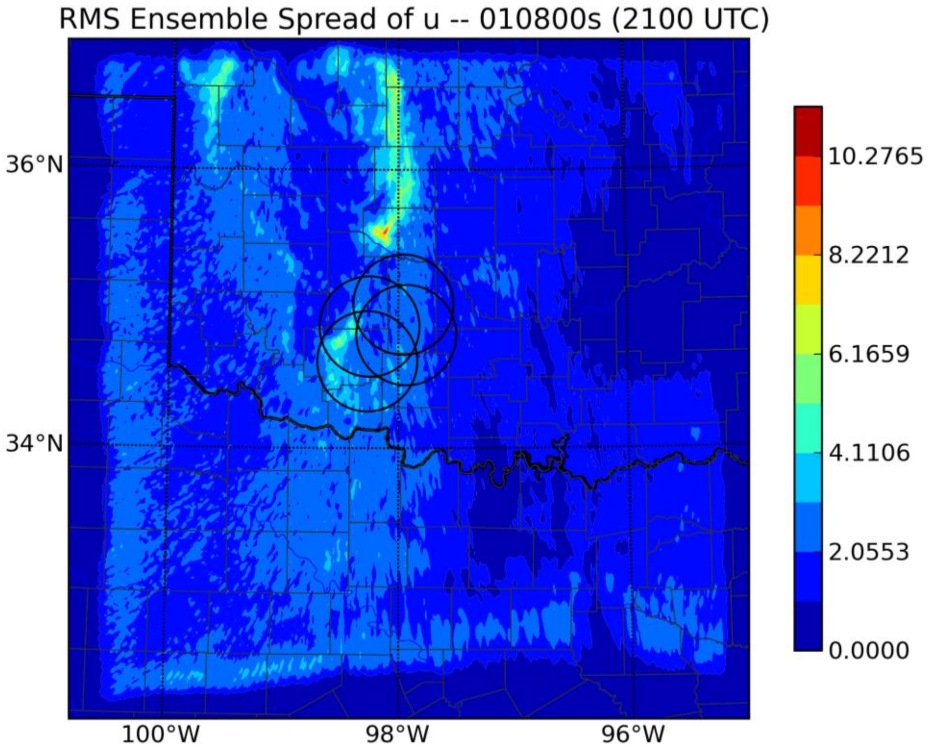


FIG 5. 9 RMS ensemble spread of *u* for 40 ensemble members at 2100 UTC 24 May 2011 on outer domain

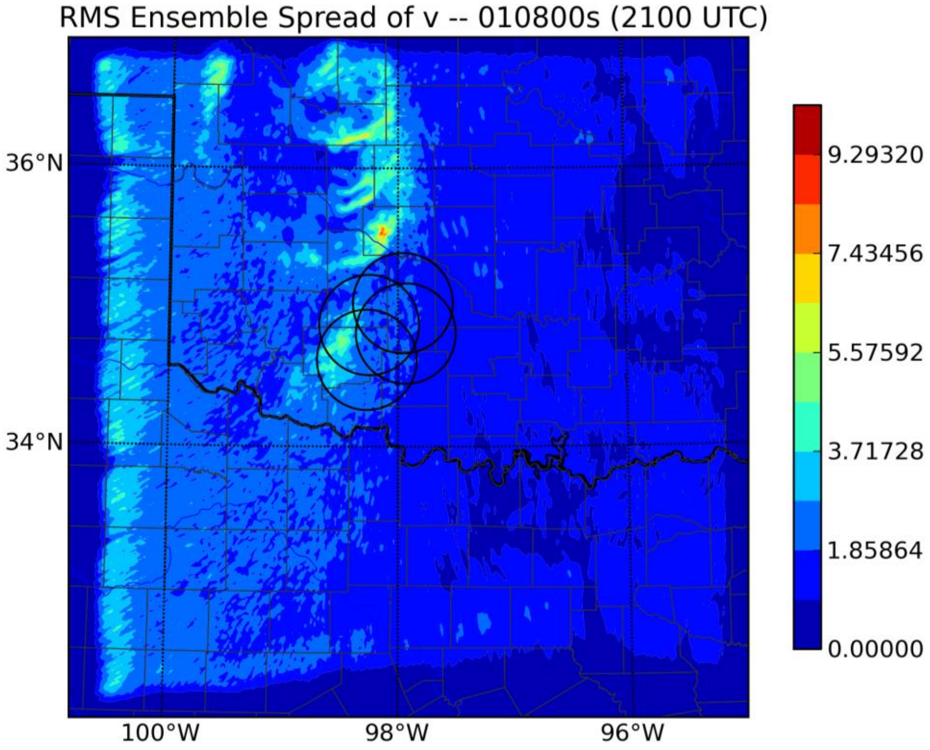


FIG 5. 10 RMS ensemble spread of *v* for 40 ensemble members at 2100 UTC 24 May 2011 on outer domain

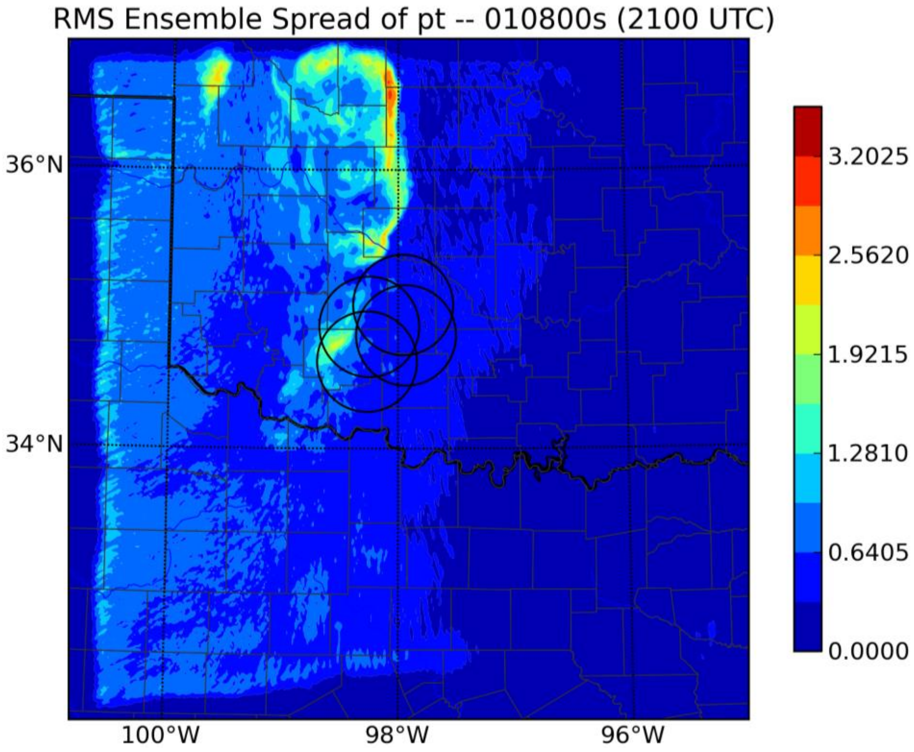


FIG 5. 11 RMS ensembles spread of potential temperature θ of 40 ensemble members at 2100 UTC 24 May 2011 on outer domain

The major difference among the design of the four experiments on the inner domain is the way the reflectivity data are assimilated. Making comparisons of the reflectivity field calculated from the results can give us a direct vision of how these methods work. We mainly look at the calculated reflectivity field of each experiment plotted at 2 km altitude, which is a level at which every radar has data available. In those plots, only the observations from KTLX are interpolated onto the same level for comparison convenience, since the coverage of KTLX is enough for this storm system.

Though CASA data can provide low-level detection and higher resolution compared to 88D data when 88D is far from the feature of interest, attenuation in X-band observations is so severe that even pre-corrected reflectivity could only partially capture the storm system. FIG 5. 12 shows the reflectivity calculated from analyses of four experiments and the observed reflectivity from KTLX interpolated onto 2 km MSL at 2135 UTC, a time when the location of the storm can be covered by most radars.

The storm system analyzed in CTRL is much weaker inside the CASA network than in the observations (FIG 5. 12b), mainly due to the fully-attenuated observations that could not be effectively corrected. Because we assimilated all values of reflectivity from CASA, the zero reflectivity areas that might be caused by fully attenuation play an important role in suppressing storm development. In fact, during the assimilation window, storms tend to grow in forecast steps, but get suppressed in assimilation steps by zero reflectivity from CASA. The storms outside of the CASA network, for example in northern part of the inner domain, are stronger than those inside the CASA coverage area, which further suggests that using all values of CASA reflectivity results in deterioration of the analysis.

It is expected that discarding observed reflectivity smaller than 20dBZ in the CASA observations may reduce the negative impact. The analyzed storm systems inside the CASA IP1 network in PreAC_20DBZ are much stronger than in CTRL (FIG 5. 12d). In Chapter 3, when we discarded reflectivity smaller than 10 dBZ, spurious echoes grew with no way to suppress them. Here, since we also use observations from WSR-88D radars which can suppress spurious echoes efficiently due to the effective absence of attenuation at S-band, the analyzed storm is much stronger than in CTRL.

Applying AEM to CASA observations during analysis is another way to reduce the impact of fully-attenuated observations. The weight of weak reflectivity is reduced by assigning small observation errors, and zero values of reflectivity are ignored where the clear air flag is negative. Doing so puts more weight on less-attenuated WSR-88D observations, background estimation, and high reflectivity CASA data, while also using information from weak reflectivity CASA data. FIG 5. 12e indicates that PreAC_OEM also obtains a relatively stronger storm compared to CTRL.

Different from the other three experiments, the un-corrected reflectivity data from CASA are used in ATTC and EnKF with built-in attenuation correction is applied. Combined with AEM, the analyzed storm structures are closer to observations from KTLX at this time, especially in the area where reflectivity is greater than 50 dBZ inside CASA the network. In PreAC_20DBZ and PreAC_OEM, a similar storm system can be obtained, too, but it is slightly weaker in the region where reflectivity exceeds 50dBZ.

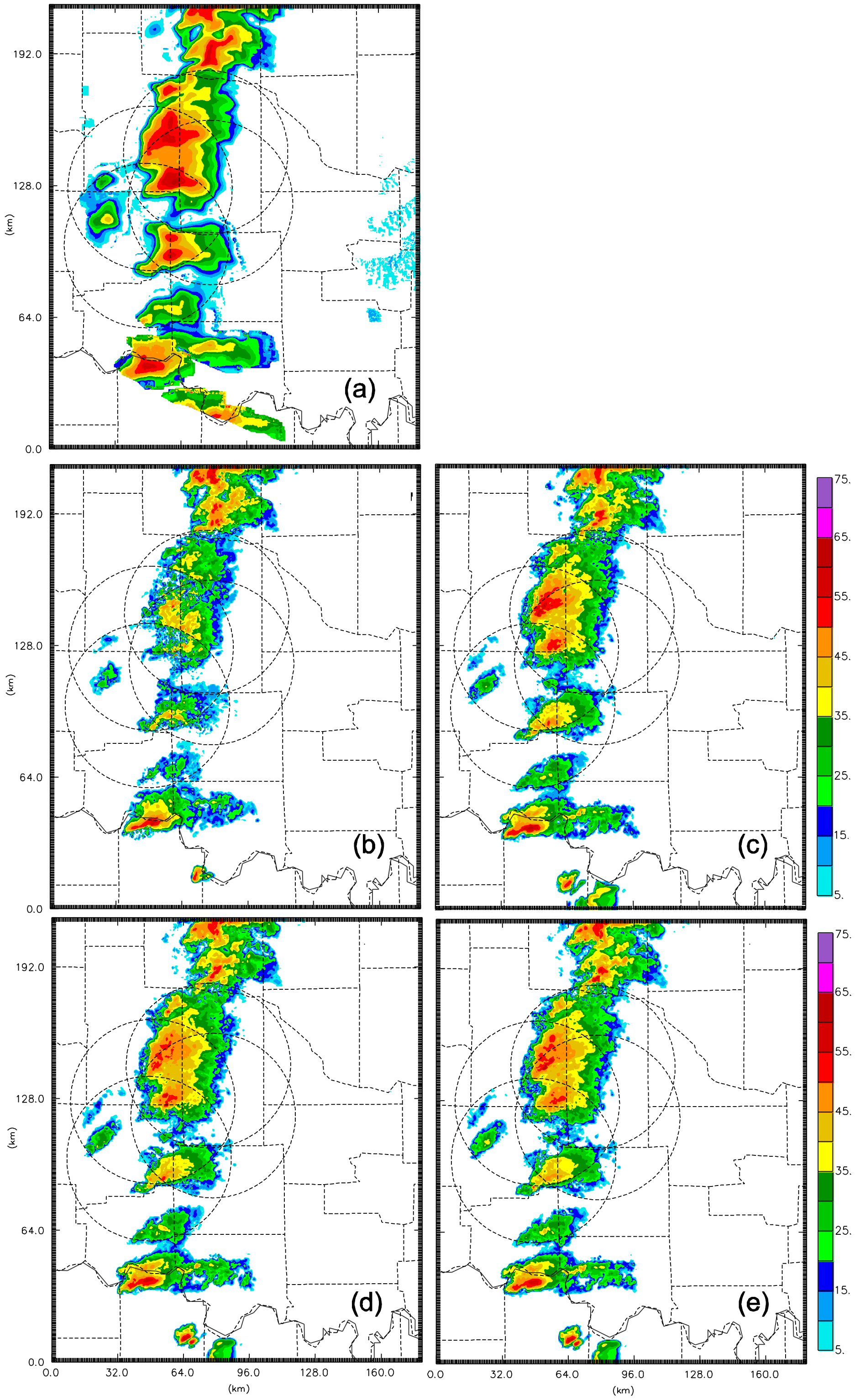


FIG 5. 12 (a) Observed reflectivity of KTLX interpolated to 2 km MSL; (b) calculated reflectivity in (b) CTRL; (c) ATTC; (d) PreAC_20DBZ; (e) PreAC_OEM on 2 km MSL at 2135 UTC 24 May 2011

Analyses in CTRL are still bad at the end of the assimilation window in the terms of calculated reflectivity. The analyses at 2200UTC show minor differences among experiments ATTC, PreAC_20DBZ, and PreAC_OEM at 2 km. We decide to plot results on 500m since CASA data feature low level detection which can benefit the analyses at low MSL. Unfortunately, WSR-88D radar observations are not available at this level, and CASA pre-corrected data are not good enough for verification. However, comparing storm A and storm B as denoted in FIG 5. 13, we see that these two storms are stronger in ATTC than in PreAC_20DBZ or PreAC_OEM since there are more areas with reflectivity exceeding 50 dBZ in ATTC than in PreAC_20DBZ or PreAC_OEM. Considering that storm A produces a tornado (C1) 6 minutes later, and that storm B spawns tornado D1 26 minutes later, the storm systems analyzed in ATTC seem more realistic.

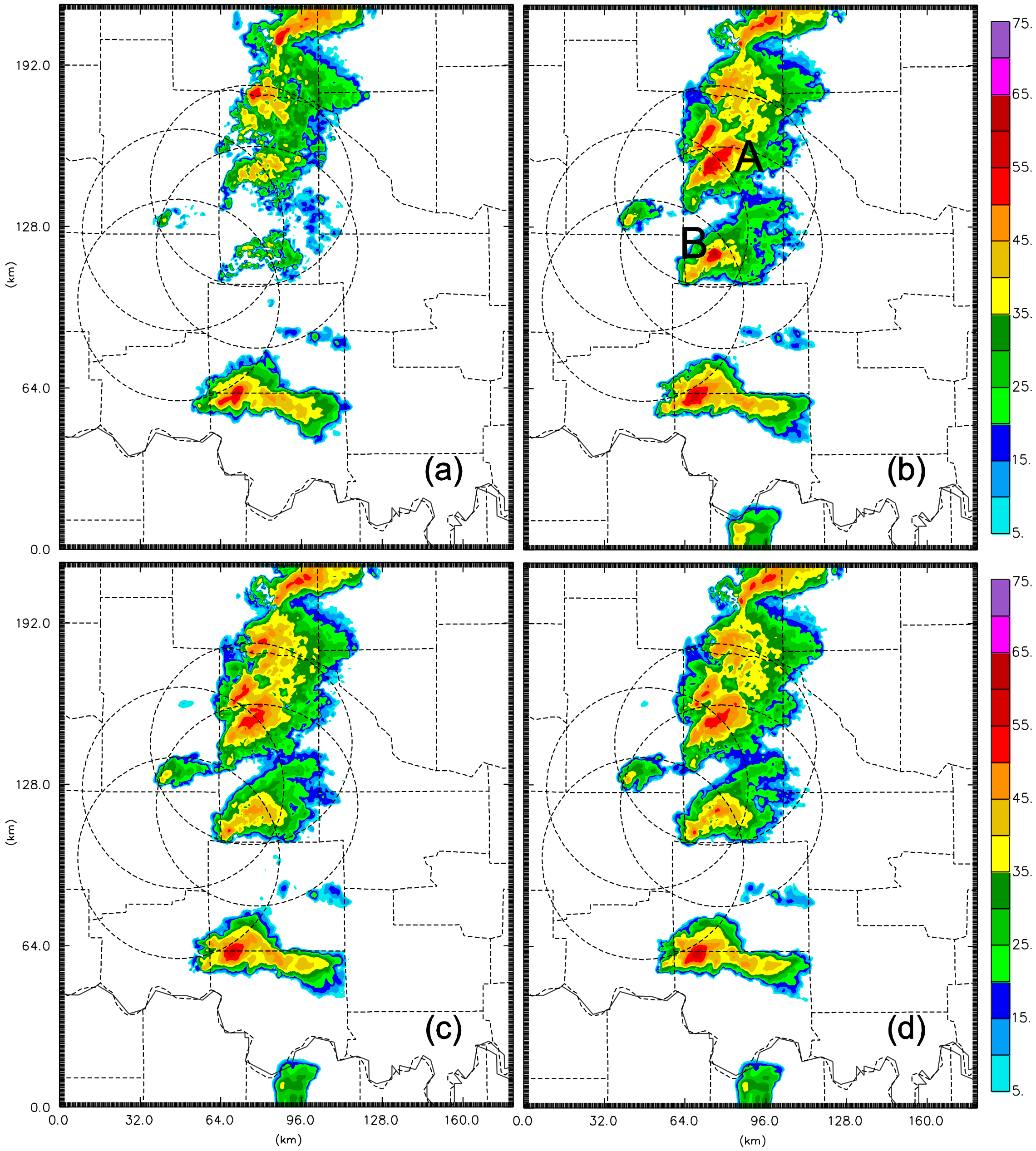


FIG 5. 13 Calculated reflectivity in (a) CTRL; (b) ATTC; (c) PreAC_20DBZ and (d) PreAC_OEM on 500 m MSL at 2200 UTC 24 May 2011

5.4 Forecast results

Examination of the forecasts is another way to verify the results of these experiments. We have performed 1 hour deterministic forecasts starting from the ensemble mean analysis at the end of the assimilation window, and 1 hour ensemble forecasts starting with the 40 member analyses at the same time. The deterministic forecasts on the outer domain 1500m grid from 2200 UTC to 2300 UTC are used as the boundary conditions for the deterministic forecasts on inner domain 500 m grid, while the ensemble forecasts on the outer domain serve as the ensemble boundary conditions for the ensemble forecasts on inner domain for the same period. Discussion of the forecasts is provided here.

5.4.1. Deterministic Forecast

Since the analyzed storm obtained in CTRL at the end of assimilation time is weak compared with the other three experiments, we do not show the forecast reflectivity field of CTRL in FIG 5. 14, FIG 5. 15, FIG 5. 16, FIG 5. 17, FIG 5. 18, and FIG 5. 19, in which the calculated reflectivity fields are plotted in 10 minute intervals from 2200 UTC to 2300 UTC, for the sake of brevity.

In addition to the development of storm A and storm B, a trailing line T starts to grow from 2200 UTC to 2300 UTC, according to KTLX and KFDR observations. FIG 5. 14 shows the forecasted reflectivity at 2 km MSL at 2210 UTC. All three experiments (ATTC, PreAC_20DBZ, and PreAC_OEM) forecast solid storm A and B structures and even a “hook echo” feature similar to the observations. The trailing line (T) can be found in the three experiments, too. In PreAC_OEM T is the weakest with maximum reflectivity below 25 dBZ. In PreAC_20DBZ, the forecasted reflectivity in T is as large as 35 dBZ, however the length of the trailing line is shorter than the one forecasted in ATTC; the reflectivity reaches above 50 dBZ in the T forecasted in ATTC.

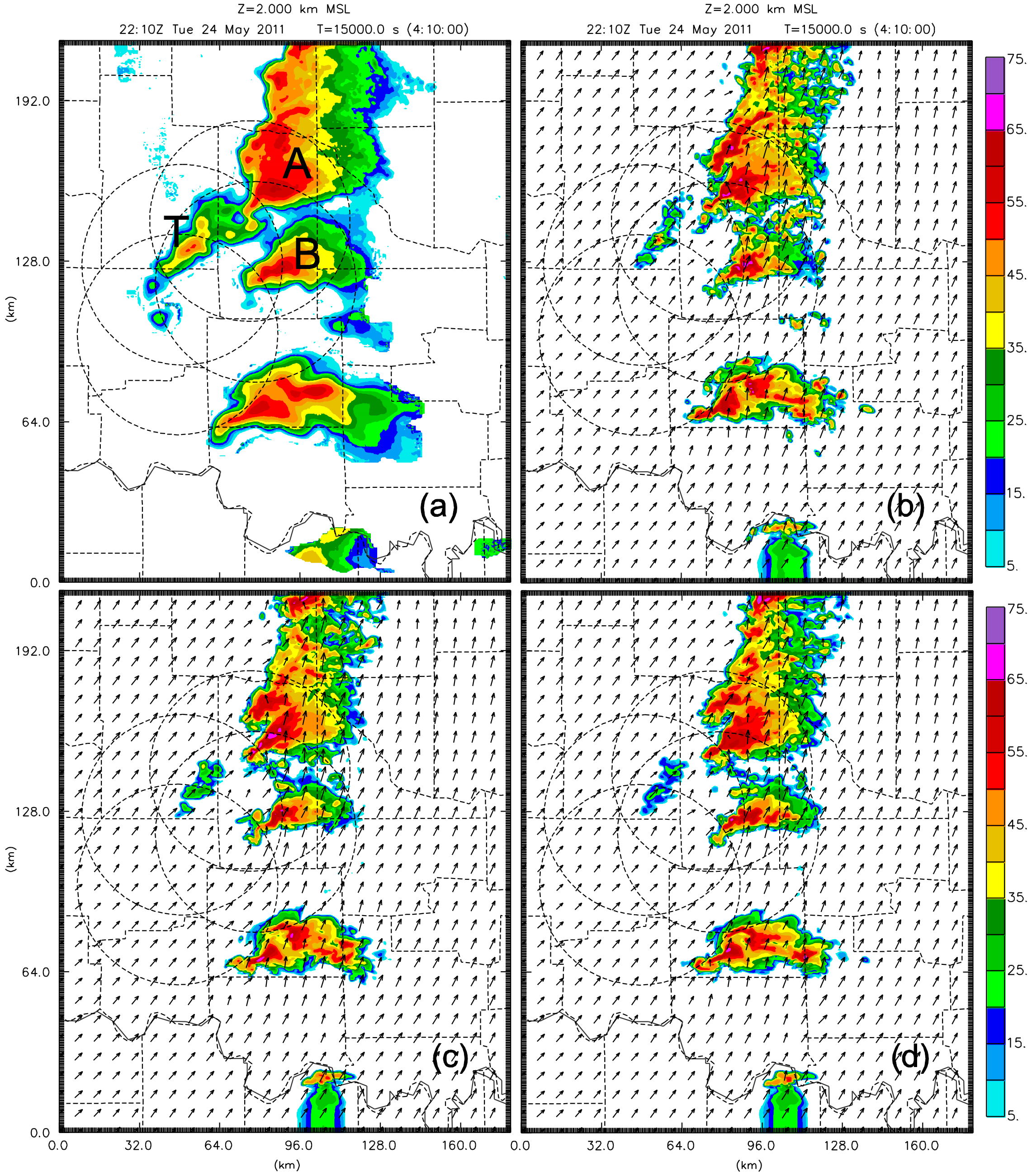


FIG 5. 14 (a) Observed reflectivity of KTLX interpolated to 2 km MSL; the horizontal wind vectors (ms^{-1}), calculated reflectivity in (a) ATTC; (b) PreAC_20DBZ; (c) PreAC_OEM on 2 km MSL at 2210 UTC 24 May 2011

At 2220UTC, 20 minutes after the final analysis, ATTC, PreAC_20DBZ, and PreAC_OEM continue to forecast the "hook echo" features of storms A and B that are observed in KTLX (FIG 5. 15). However the trailing line T shrinks in PreAC_OEM and is almost gone in PreAC_20DBZ. In ATTC, though T is not as long as in KTLX observations, the largest echoes can still reach 50 dBZ. This may benefit from the small values of reflectivity (less than 20 dBZ in CASA data) that contain some useful information, because small values are still used in ATTC and PreAC_OEM but not PreAC_20DBZ. Also, T is stronger in ATTC than in PreAC_OEM, which indicates that pre-corrected data also misses some information, motivating the use raw observation directly in data assimilation.

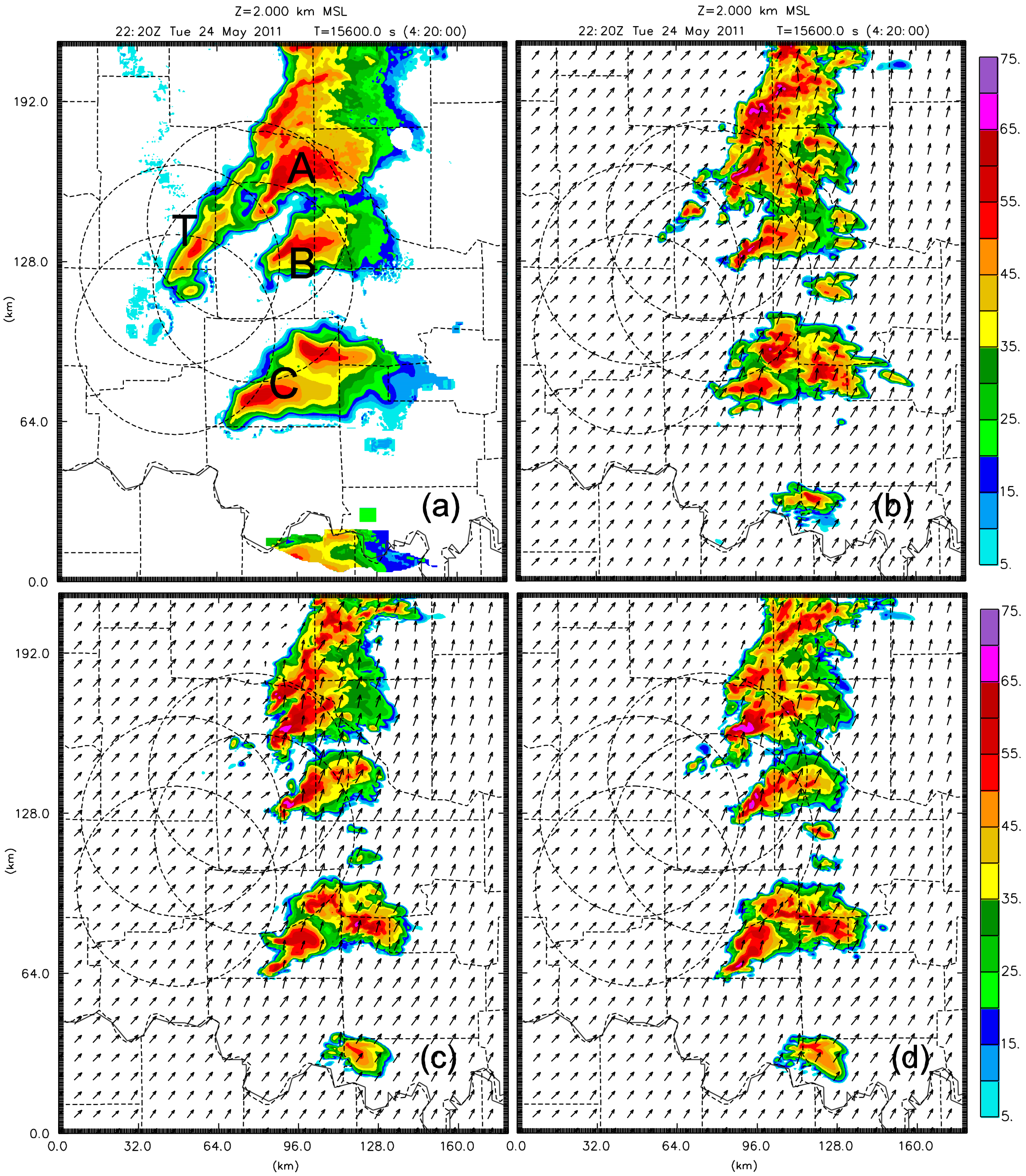


FIG 5. 15 Same as FIG 5. 14 but at 2220 UTC 24 May 2011

The trailing line T shrinks in ATTC at 2130 UTC (FIG 5. 16b), while remaining strong in KTLX observations (FIG 5. 16a). In PreAC_20DBZ and PreAC_OEM, T has completely dissipated (FIG 5. 16c, d). The model error likely dominates, and information obtain from raw observations is unable to confer benefits after 30 minutes of forecast time. The structure of storms A and B, especially the “hook echo” region, is still similar to the observations in ATTC, PreAC_20DBZ, and PreAC_OEM.

After 40 minutes of forecast (FIG 5. 17), storm A starts to dissipate in PreAC_20DBZ and PreAC_OEM. In observations, storm A is still intense at this time. The “hook echo” feature of storm A is lost in ATTC, but the storm predicted in ATTC looks more intense than in other two experiments.

Although the trailing line T is absent, in ATTC storm B still has similar structure as KTLX observations after 50 minutes of forecast (FIG 5. 18b), while in PreAC_OEM, storm B shrinks to a small size. In PreAC_20DBZ, the forecasted reflectivity in storm B is as high as 55 dBZ, however, the size of the storm is smaller than in the observations. Storm C developed outside of CASA radar coverage, thus the forecast of storm C in the three experiments is much worse than the forecast of storm A and B. In PreAC_20DBZ and PreAC_OEM storm C is stronger than in ATTC, but is shifted to the southeast of the location of the observed storm C, while the storm C location is close to observations in ATTC .

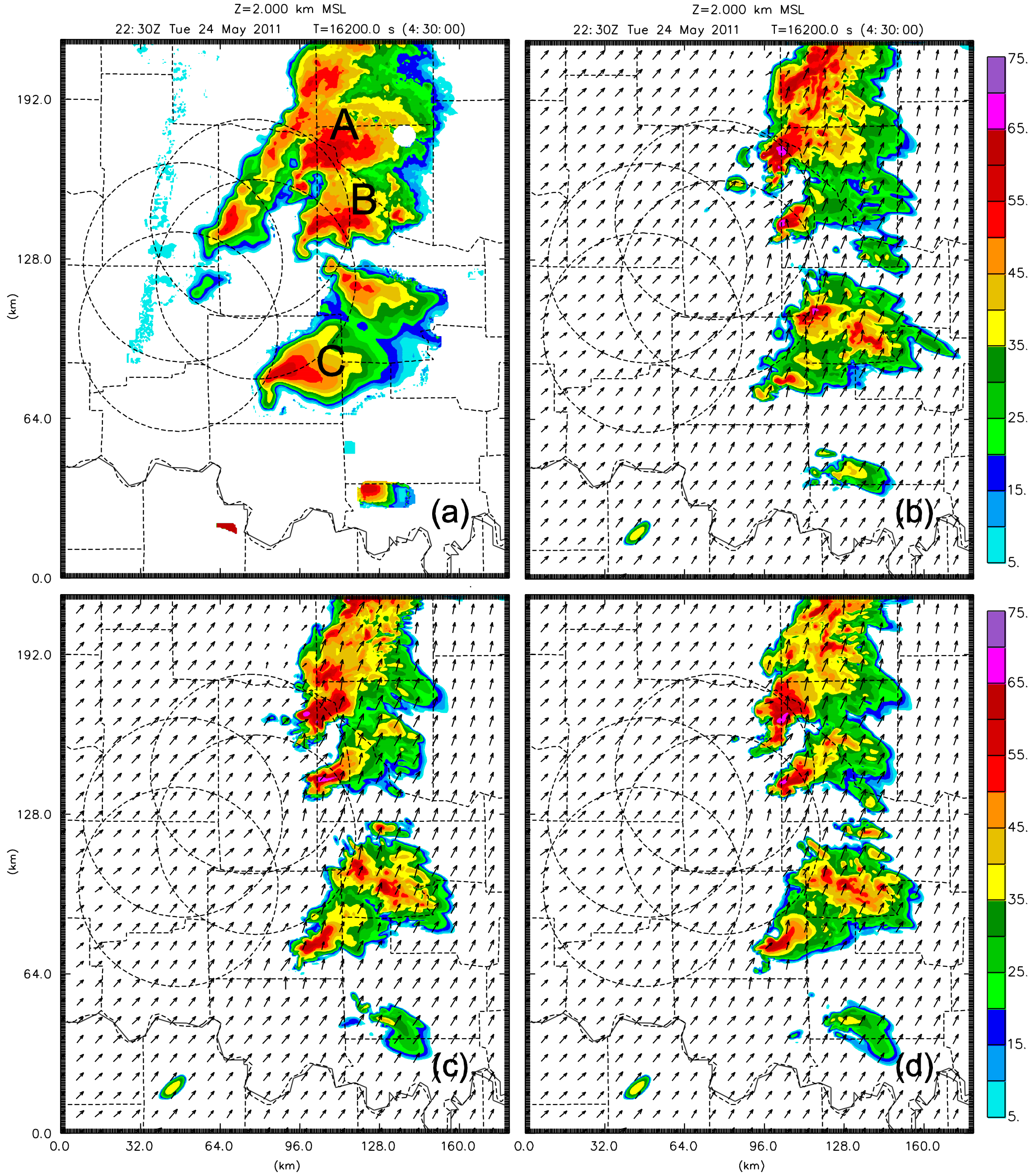


FIG 5. 16 Same as FIG 5. 14 but at 2230 UTC 24 May 2011

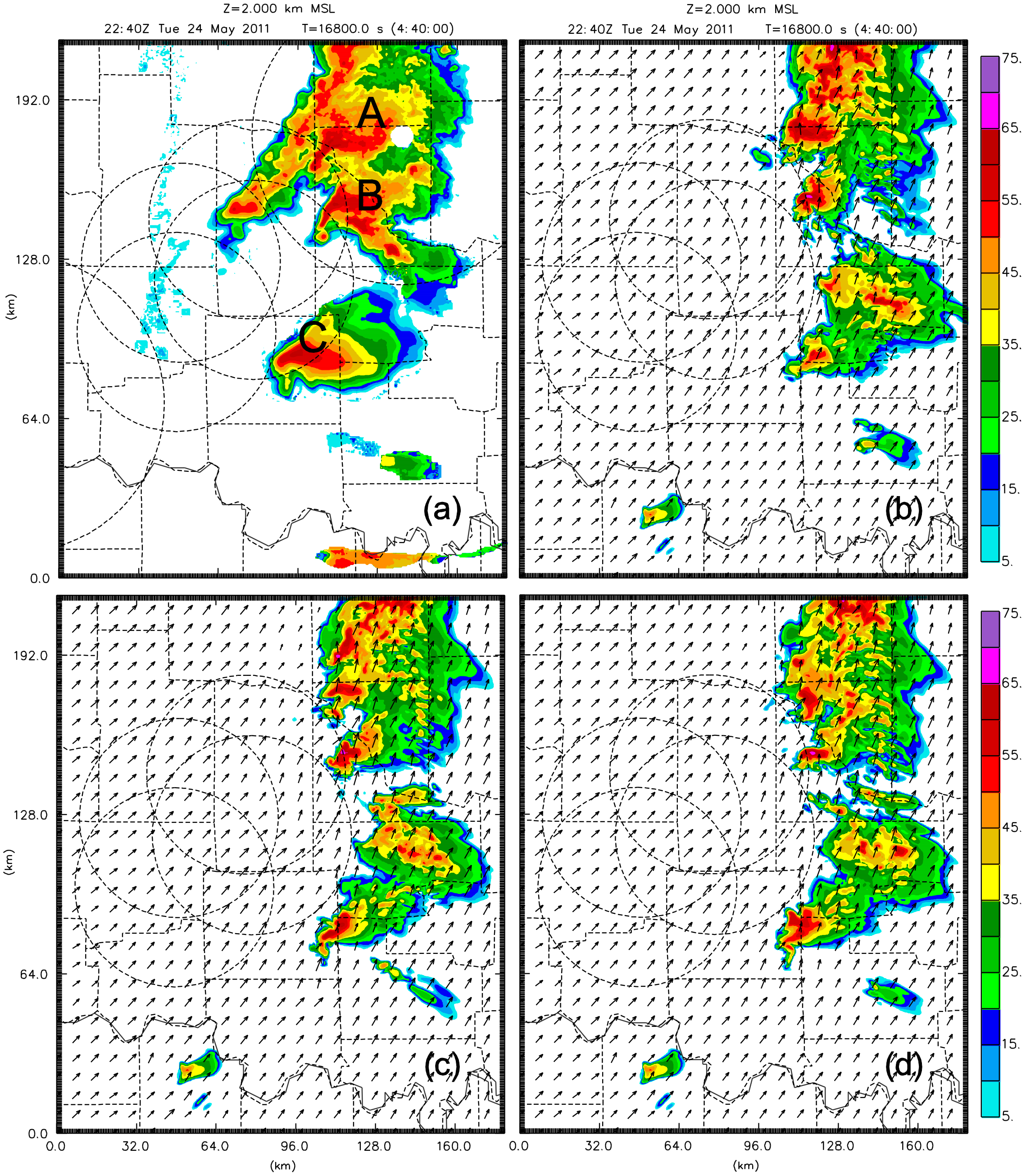


FIG 5. 17 Same as in FIG 5. 14 but at 2240 UTC 24 May 2011

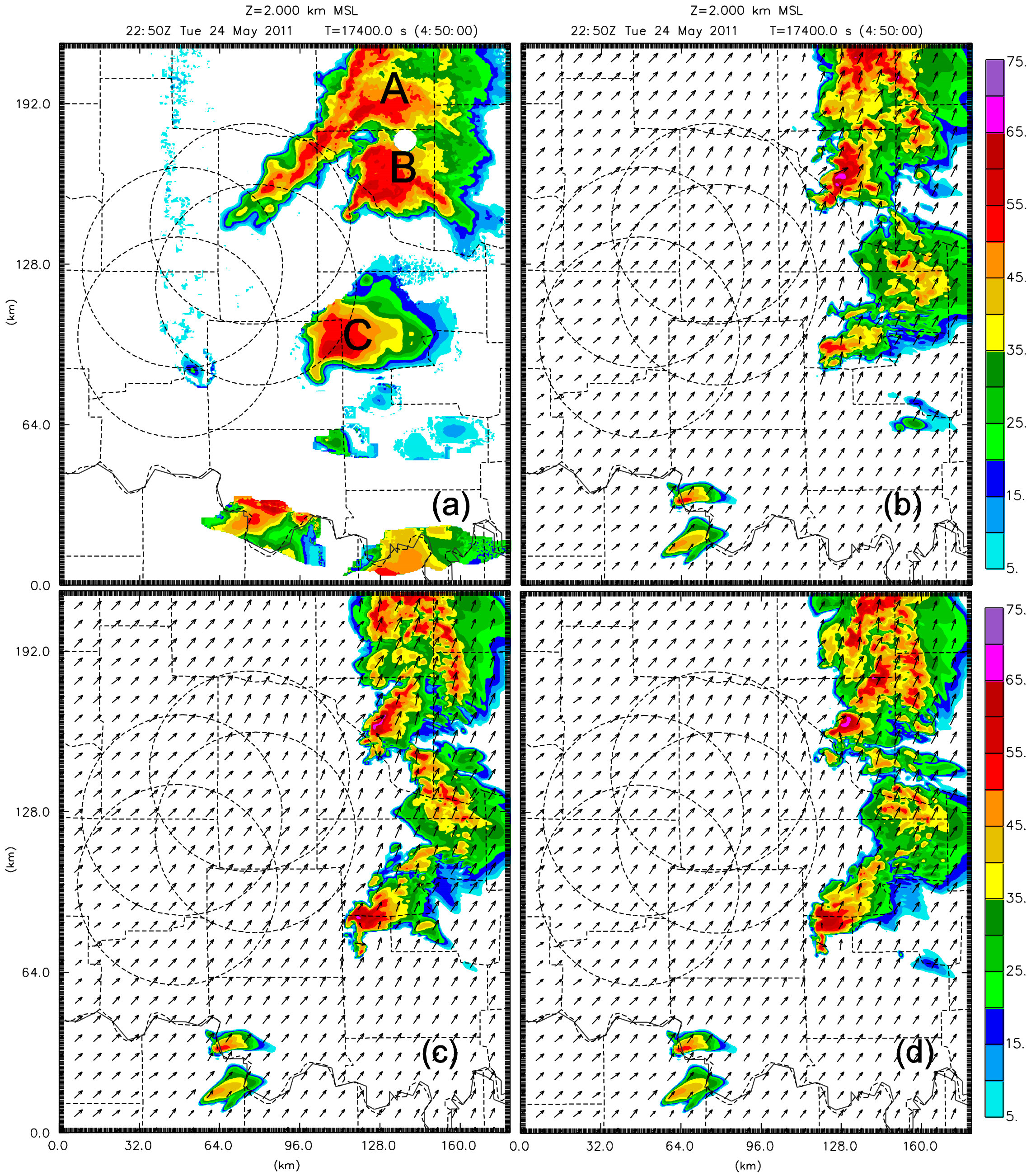


FIG 5. 18 Same as FIG 5. 14 but at 2250 UTC 24 May 2011

One issue with the forecast is that the storm systems predicted in all three experiments are phase shifted, though these shifts are not severe. At 2300 UTC, storm A is shrank to the northeast corner of the inner domain. Storm B is shifted about 15 ~ 20 km to the east in ATTC, PreAC_20DBZ, and PreAC_OEM. In ATTC, the location of storm C is about 25 km east of the same storm in the observations, while in PreAC_20DBZ and PreAC_OEM, the predicted storm C moves to the southeast of the observed one, further displaced than in ATTC. There is another storm that developed between storm B and storm C in all the three experiments, which does not exist in observations. The development of this spurious storm may be the reason that storm C becomes so weak in ATTC, PreAC_20DBZ, and PreAC_OEM. Overall, most of the precipitation areas could still be captured in all three experiments, except in the southern part of the domain where forecasting is difficult because of interaction with the southern domain boundary.

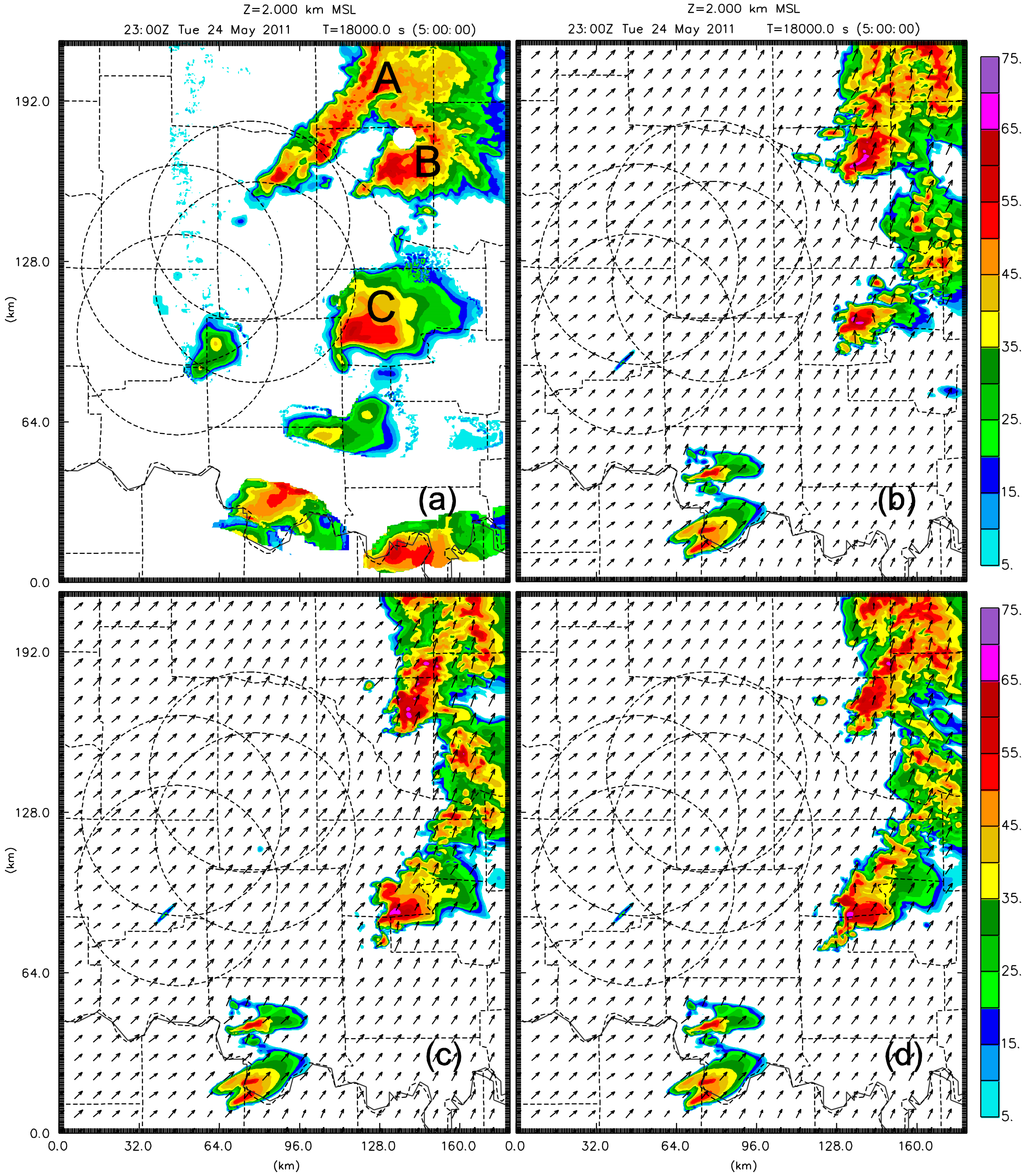


FIG 5. 19 Same as FIG 5. 14 but at 2300 UTC 24 May 2011

The forecast results of all four experiments (CTRL, PreAC_20DBZ, PreAC_OEM and ATTC) are also verified against OK Mesonet observations. In FIG 5. 20 and the following FIG 5. 21, FIG 5. 24 and FIG 5. 25, labels ‘A’, ‘B’, ‘C’ and ‘T’ are included in the figures to indicate the location of storm A, B, C and trailing line T.

Since those figures show the deviation of forecasted variables from the surface observations obtained from the OK Mesonet, the lighter the color is, the closer the forecast is to the observation.

Comparison with OK Mesonet 2-m temperature observation suggests that at 2230 UTC (FIG 5. 20), all four experiments (CTRL, PreAC_20DBZ, PreAC_OEM, and ATTC) produce relatively strong cold pools near the western boundary of the domain and warm temperatures in the east part of domain. Small differences exist between every experiment. For example, it is too warm in the trailing line area as well as in the cold pool of storm A in CTRL compared to the other three experiments. The cold pool produced by storm A and C is too strong (widespread and colder) in PreAC_20DBZ and PreAC_OEM than in ATTC. In the eastern part of storm C, the temperature is much lower in CTRL than in the other three experiments.

At 2300 UTC, the end of forecast period (FIG 5. 21), the strong regions of low temperature that exist in all four experiments move into the northwest portion of the domain, propagating with the background wind. A large amount of warm air enters the domain via the southern boundary in all four experiments. This may all come from the boundary conditions. In CTRL, the area east of storm A as well as the area near storm B and T are warmer than in the other three experiments. And in ATTC, the cold pool produced by storm C is still closer to the observation than the other three experiments, which may explain why the location of storm C forecasted in ATTC is closer to the observed position.

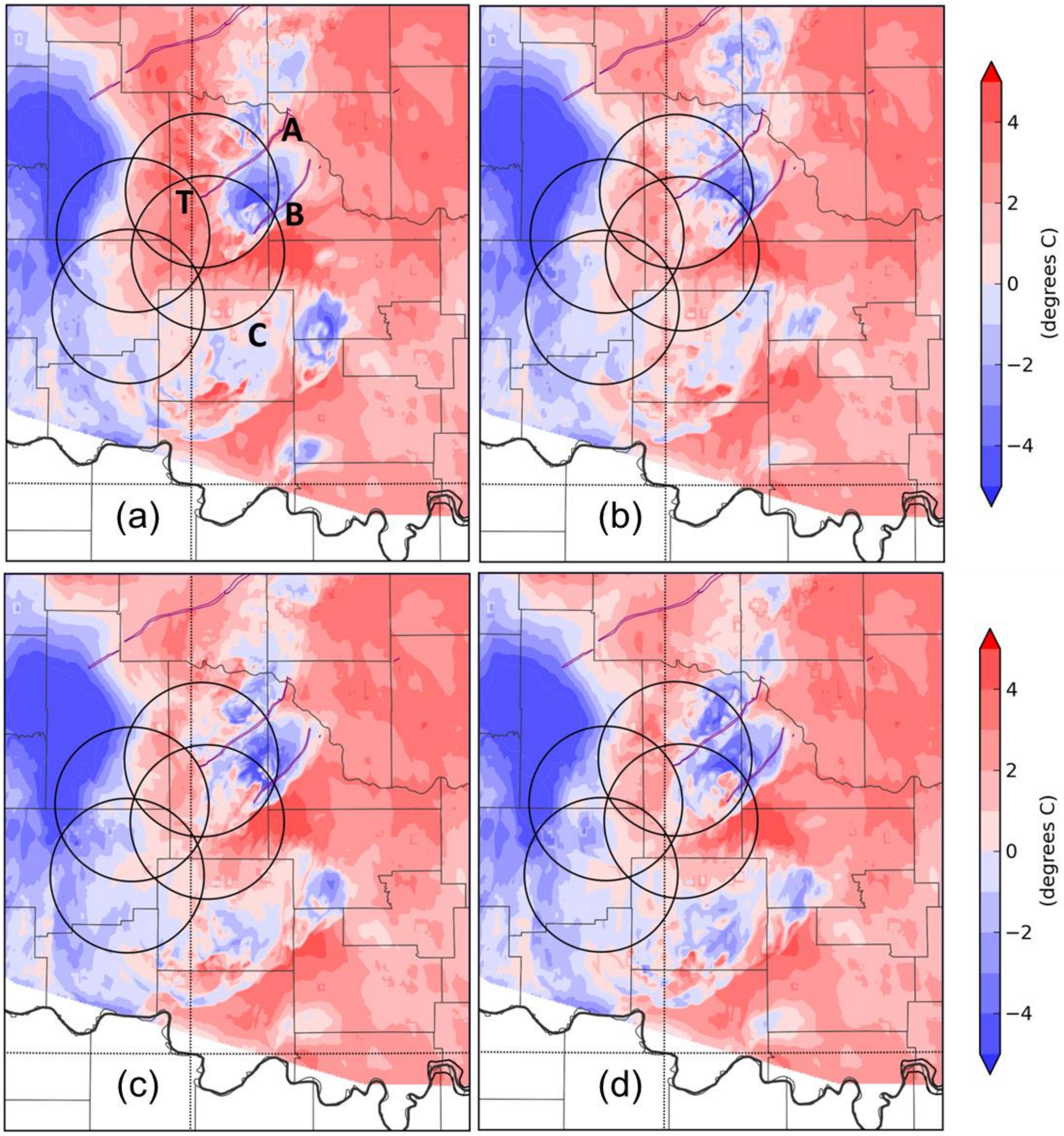


FIG 5. 20 Deviation of 2 m temperature forecasted from surface observation (OK Mesonet) for (a) CTRL, (b) ATTC, (c) PreAC_20DBZ and (d) PreAC_OEM at 2230 UTC

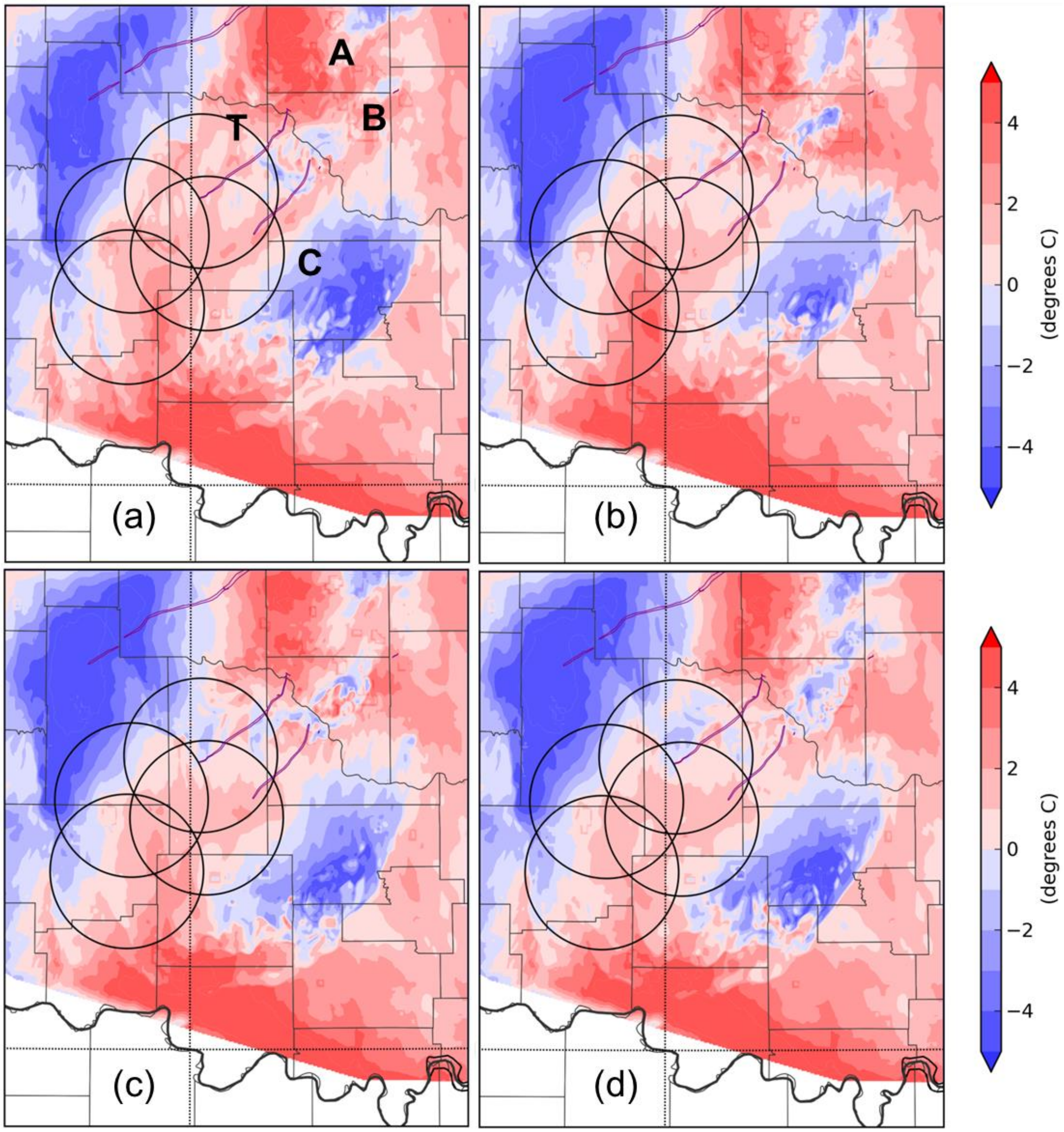


FIG 5. 21 Same as in FIG 5. 20 but at 2300 UTC 24 May 2011

Further comparison with time series plots of observed temperature T from Minco (FIG 5. 22) and Norman OK Mesonet site (FIG 5. 23) shows that the temporal evolution of temperature represents part of the development of trailing line, and storm A and B evolution respectively. The temperature evolution in ATTC shows the smallest difference and the closest match with observations at the Minco mesonet station, which reminds us that only in ATTC do the trailing line features last longer than in PreAC_20DBZ and PreAC_OEM. At the Norman station, the temperatures predicted in ATTC also have a similar pattern to that of the observations, as well as the smallest temperature departure among those experiments.

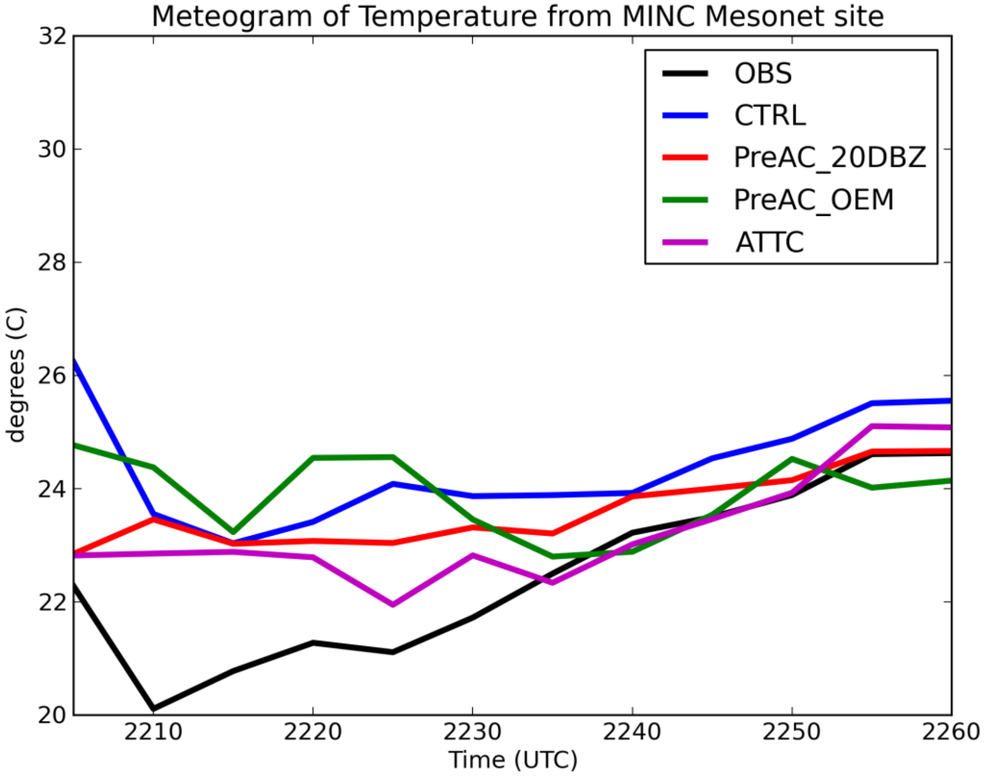


FIG 5. 22 Meteogram of temperature (C) observed (black) and simulated (blue for CTRL, red for PreAC_20DBZ, green for PreAC_OEM and purple for ATTC) at Minco mesonet sites from 2200 UTC -2300 UTC 24 May 2011.

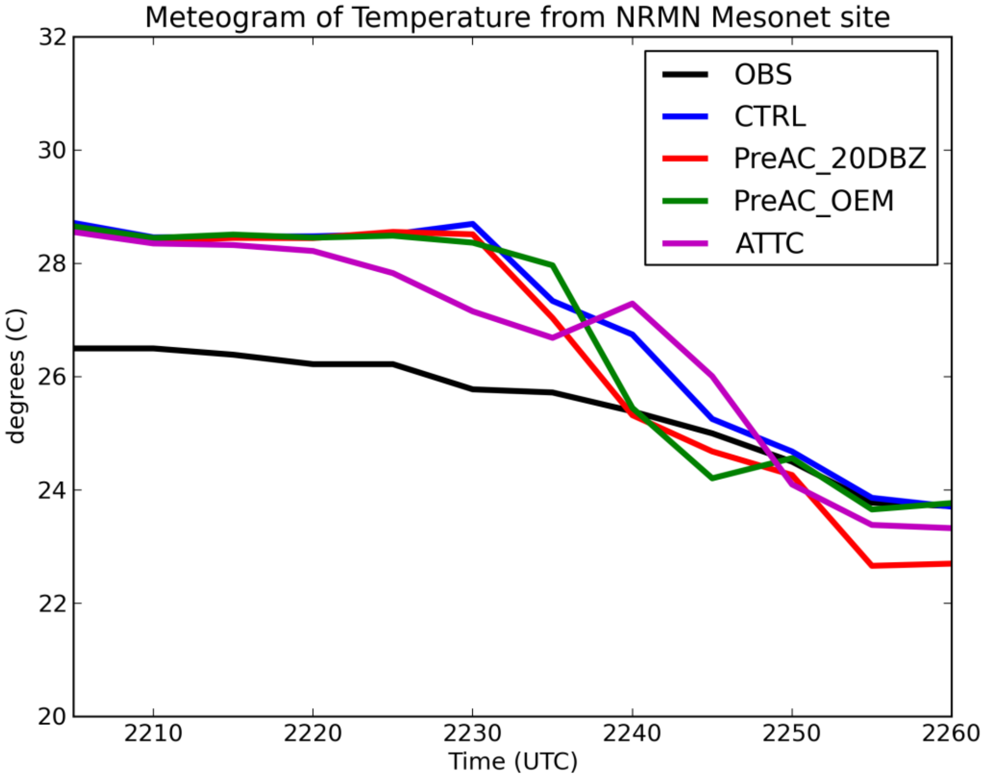


FIG 5. 23 Same as in FIG 5. 22 but at Norman mesonet site

We also calculated the deviation of relative humidity forecasts from the observations. Again, the lighter the color is, the closer the forecast is to the observations. After 30 minutes of forecast, widespread moisture advects in through the western boundary (FIG 5. 24). The most obvious difference among those four experiments is near the trailing line (T) area, where the air is much drier in CTRL, PreAC_20DBZ, and PreAC_OEM. Considering that the temperature in ATTC is also close to observation (FIG 5. 20b, FIG 5. 22) in this area, this explains why the trailing line T lasts longer in the forecast of ATTC than in other three experiments. At the end of the forecast, it is drier in the area of T and storm A and C in CTRL than in the other three experiments. The difference among PreAC_20DBZ, PreAC_OEM, and ATTC at

2300 UTC, is not significant except that the deviation in relative humidity near storm B is a little bit smaller in ATTC than in other three experiments.

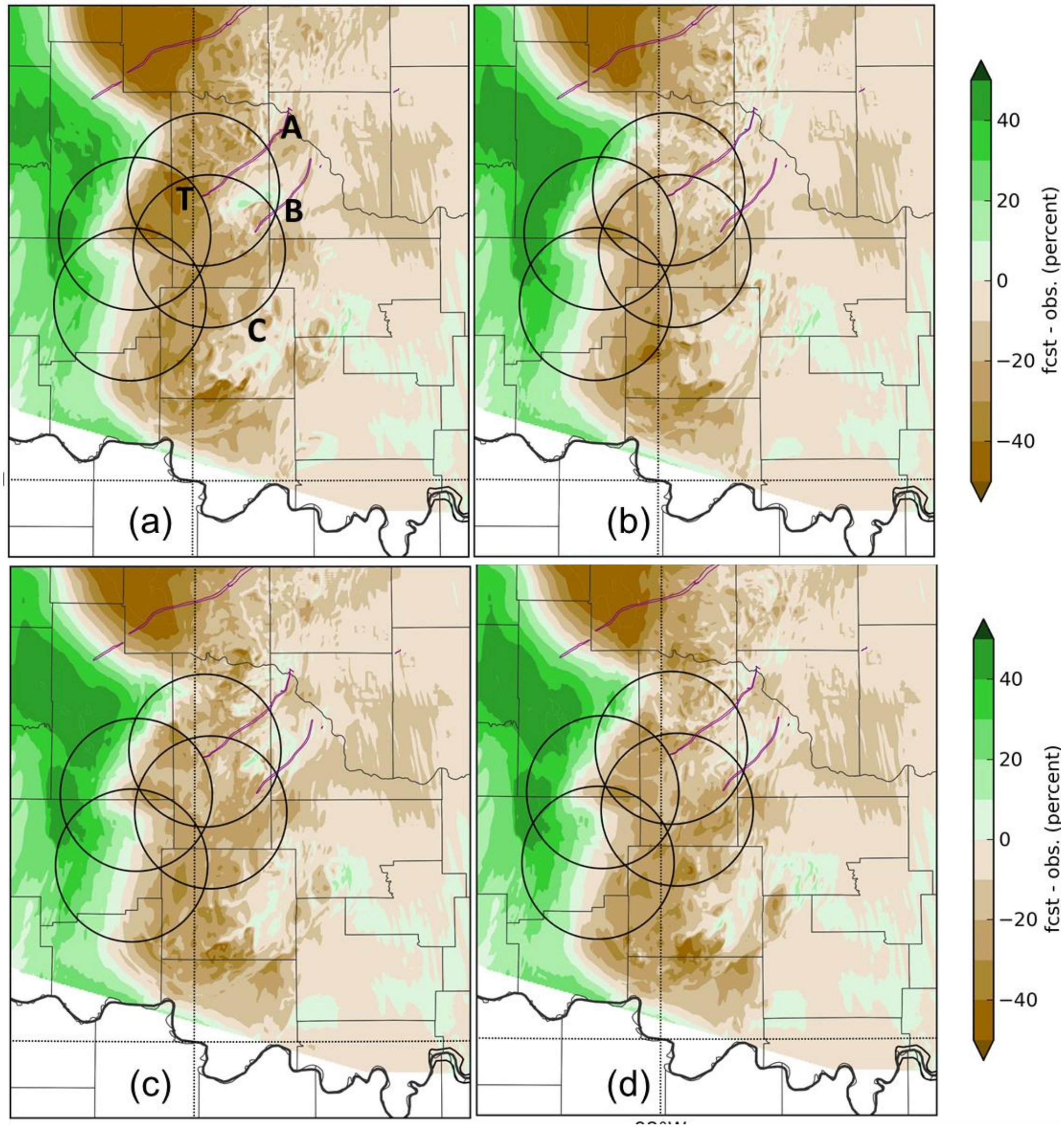


FIG 5. 24 Deviation of 2 m relative humidity forecasted from surface observation (OK Mesonet) for (a) CTRL, (b) ATTC, (c) PreAC_20DBZ and (d) PreAC_OEM at 2230 UTC

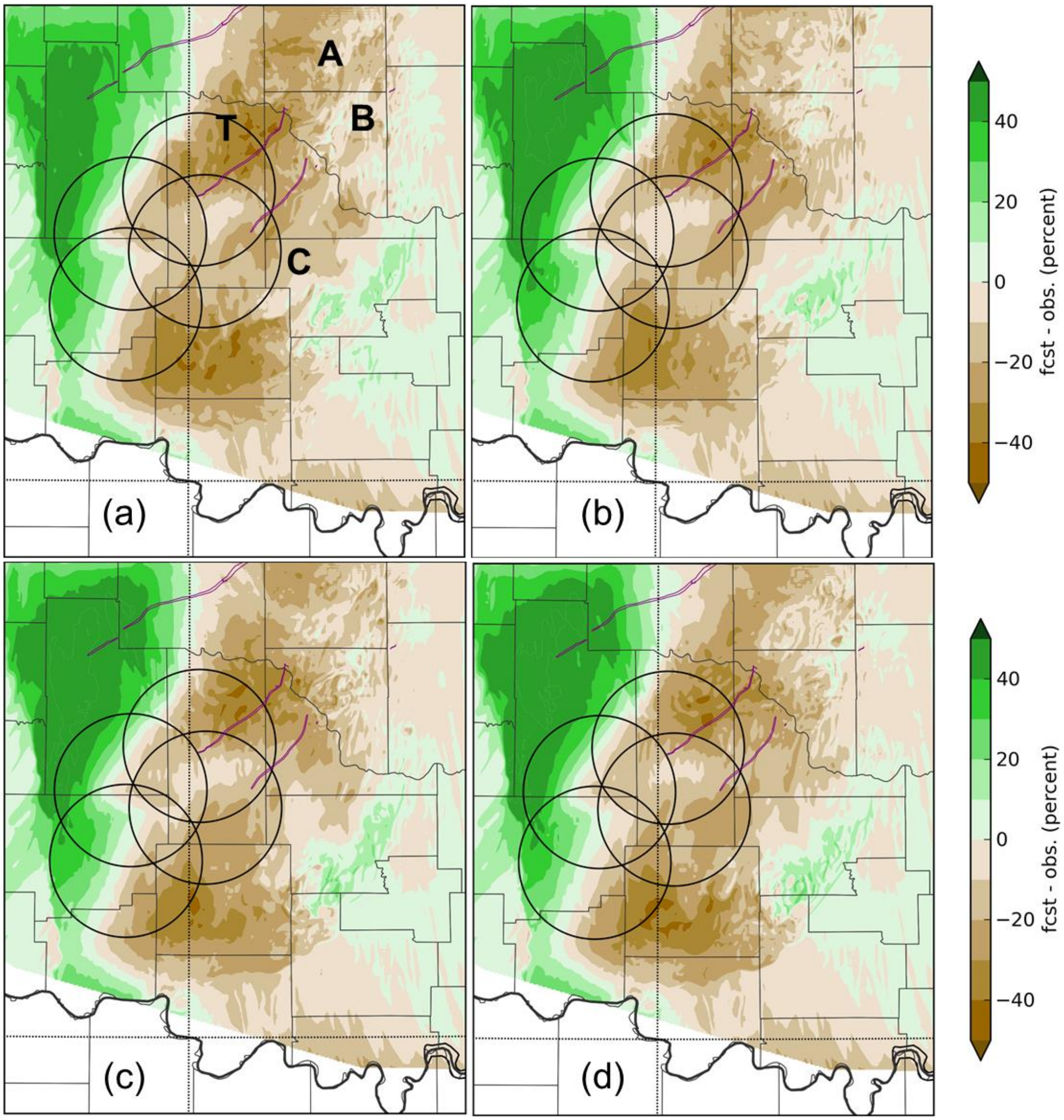


FIG 5. 25 Same as in FIG 5. 24 but at 2300 UTC 24 May 2011

The difference in the time evolution of relative humidity is small at the Minco mesonet site (FIG 5. 26). The value of calculated relative humidity of the four experiments is similar. The forecast from CTRL has the lowest value for most of the time, furthest from the observations.

At the Norman mesonet site (FIG 5. 27), for the first 20 minutes the forecasted relative humidity is almost the same in all four experiments. After this, in ATTC the forecasted value starts to rise and remains close to the observations until the end of forecast time, while in CTRL the value starts to decrease and becomes the lowest among the four experiments until 2300 UTC. We also calculate the average relative humidity every experiment from 2200 UTC to 2300 UTC and compare to the observations from Minco mesonet site and Norman mesonet site, respectively (Table 5. 3). At Minco site, the closest number of relative humidity to observation is forecasted by PreAC_20DBZ, but ATTC actually forecasted a close number that has less than 1% deviation from PreAC_20DBZ. At Norman site, the best forecast is given by ATTC, followed by PreAC_20DBZ with over 3% deviation from ATTC.

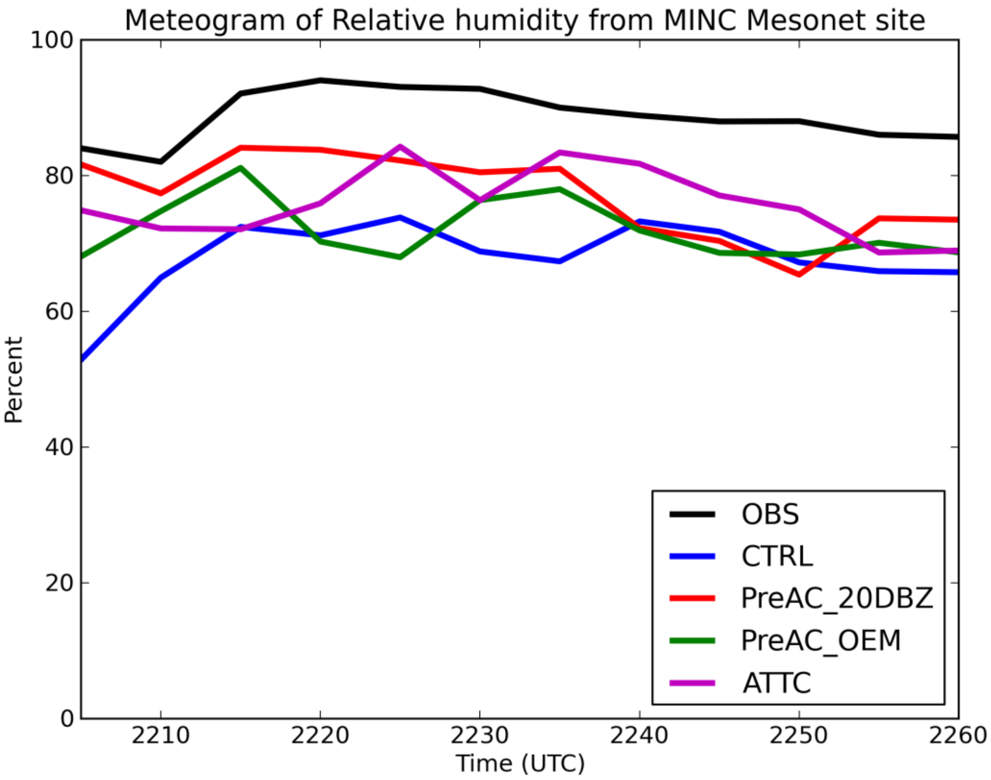


FIG 5. 26 Meteogram of relative humidity (percent) observed (black) and simulated (blue for CTRL, red for PreAC_20DBZ, green for PreAC_OEM and purple for ATTC) at Minco mesonet site from 2200 UTC -2300 UTC 24 May 2011.

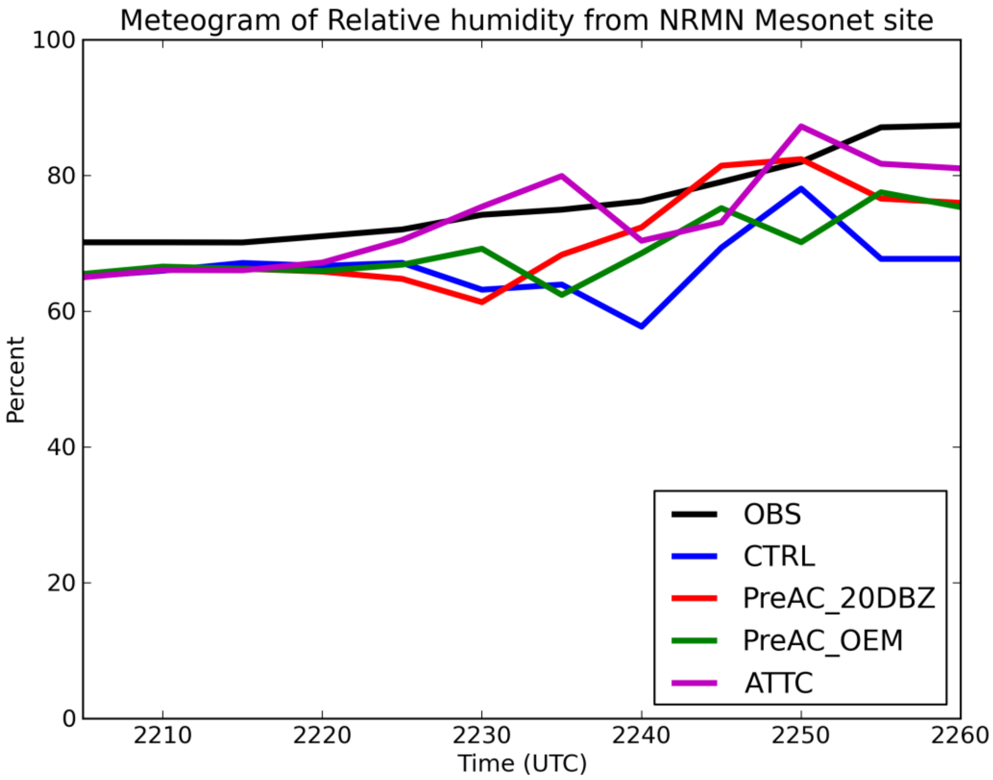


FIG 5. 27 Same as in FIG 5. 26 but at Norman Mesonet site

Table 5. 3 Average relative humidity observed and forecasted from 2200 UTC to 2300 UTC

RH(%)	Minco	Norman
OK Mesonet	88.51	76.42
CTRL	67.85	66.66
PreAC_20DBZ	77.03	70.16
PreAC_OEM	71.08	67.69
ATTC	76.08	73.20

5.4.2. Ensemble Forecast

Ensemble forecasts can provide the probability of intense low-level vertical vortices, which are usually associated with tornadoes. As in section 5.2, ensemble forecasts are started from the final analyses of the 40 ensemble members and use the ensemble boundary conditions produced by the 40 member forecasts on outer domain grid. Here we count the numbers of vortices that exceed a certain threshold at the first level above ground in all ensemble members from 2200 UTC to 2300 UTC and plot those numbers in FIG 5. 28 and FIG 5. 29 .

The threshold is set at 0.012 s^{-1} in FIG 5. 28; this is a relatively high value. All four experiments produce the frequency count lines along the track (green line in the north of domain) of tornado C1, spawned by storm A, which moved from southwest to northeast. In CTRL, the frequency count line is thin (and even broken in many places) (FIG 5. 28a) compared to other three experiments, indicating that the number of low-level vortices above 0.012 s^{-1} forecasted by ensemble members in CTRL is less than in the other three experiments, probably due to the forecast starting from the weak storm A in the CTRL analyses. The frequency count lines in ATTC (FIG 5. 28b), PreAC_20DBZ (FIG 5. 28c), and PreAC_OEM (FIG 5. 28d) along the tornado C1 track are all rather intense. The difference among the three experiments is small. Only the line in ATTC is more intense in the southwest part of the track, which indicates improved prediction of the early stage of tornado C1 compared to PreAC_20DBZ and PreAC_OEM.

Tornado D1, produced by storm B, starts 26 minutes after the final analysis time, thus it is more difficult to forecast than tornado C1. As shown in FIG 5. 28, in which the green line in the south indicates the observed track of tornado D1, not too many members forecast low-level vortices near its track at this threshold. Some ensemble members forecast relatively large low-level vortices embedded in Storm B before tornado D1 actually touches down. FIG 5. 29 shows the ensemble frequency count plot when threshold is reduced to 0.008 s^{-1}, which is a bit smaller but still strong enough to indicate tornado potential. Although the field is much noisier, there is still a line overlaid on and much longer than the track of tornado D1, especially in ATTC (FIG 5. 29b), PreAC_20DBZ (FIG 5. 29c), and PreAC_OEM (FIG 5. 29d). The difference among the three experiments is not significant. The line overlaid on the tornado D1 track in CTRL (FIG 5. 29a) is shorter than the other three.

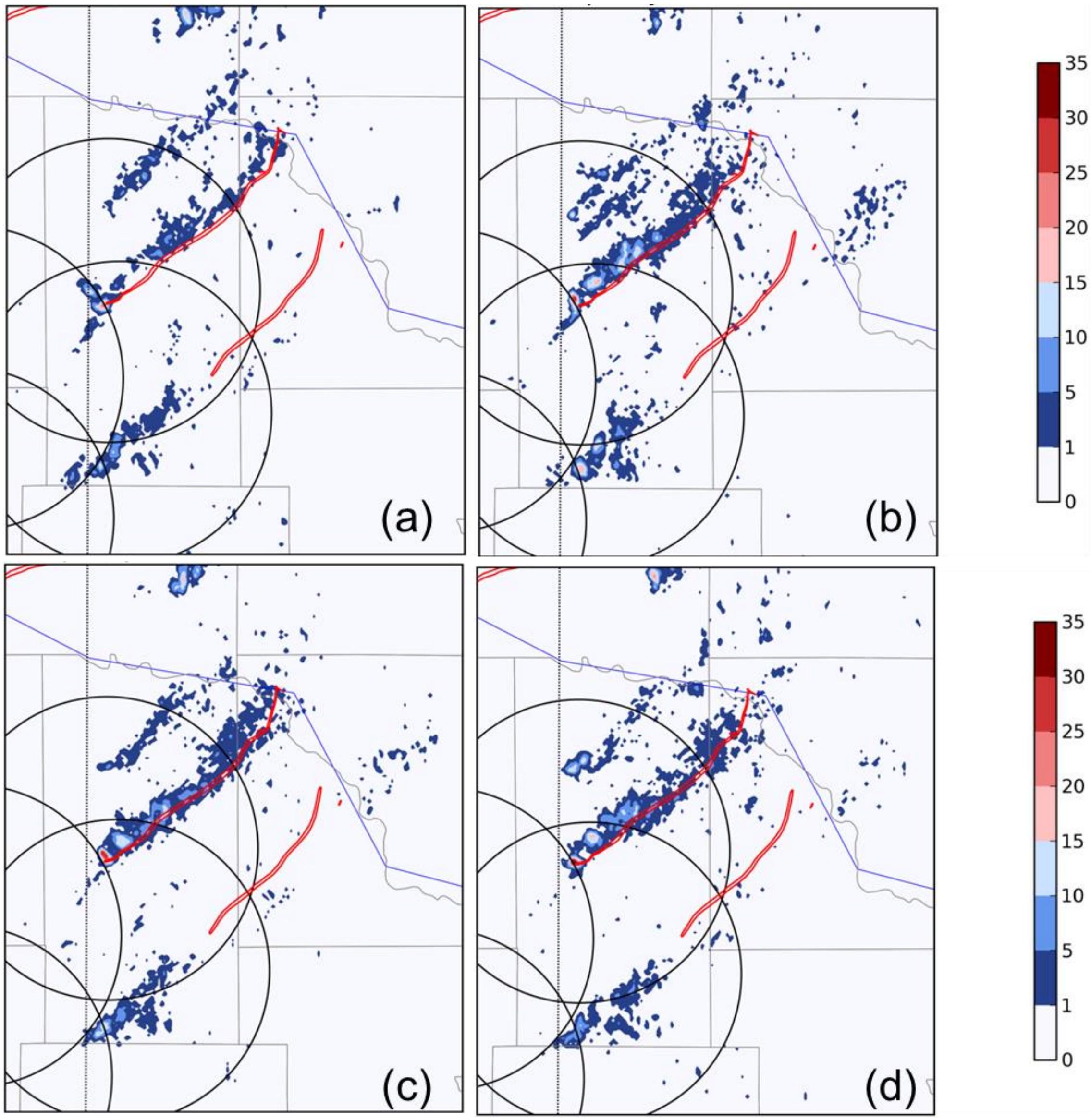


FIG 5. 28 Ensemble frequency count when vorticity > 0.012 s^{-1} in (a) CTRL, (b) ATTC, (c) PreAC_20DBZ and (d) PreAC_OEM from 2200 UTC to 2300 UTC. Red line in the north indicates the observed track of tornado C1, while the one in the south indicates the observed track of tornado D1. Circles represent the 40 km range of CASA radars.

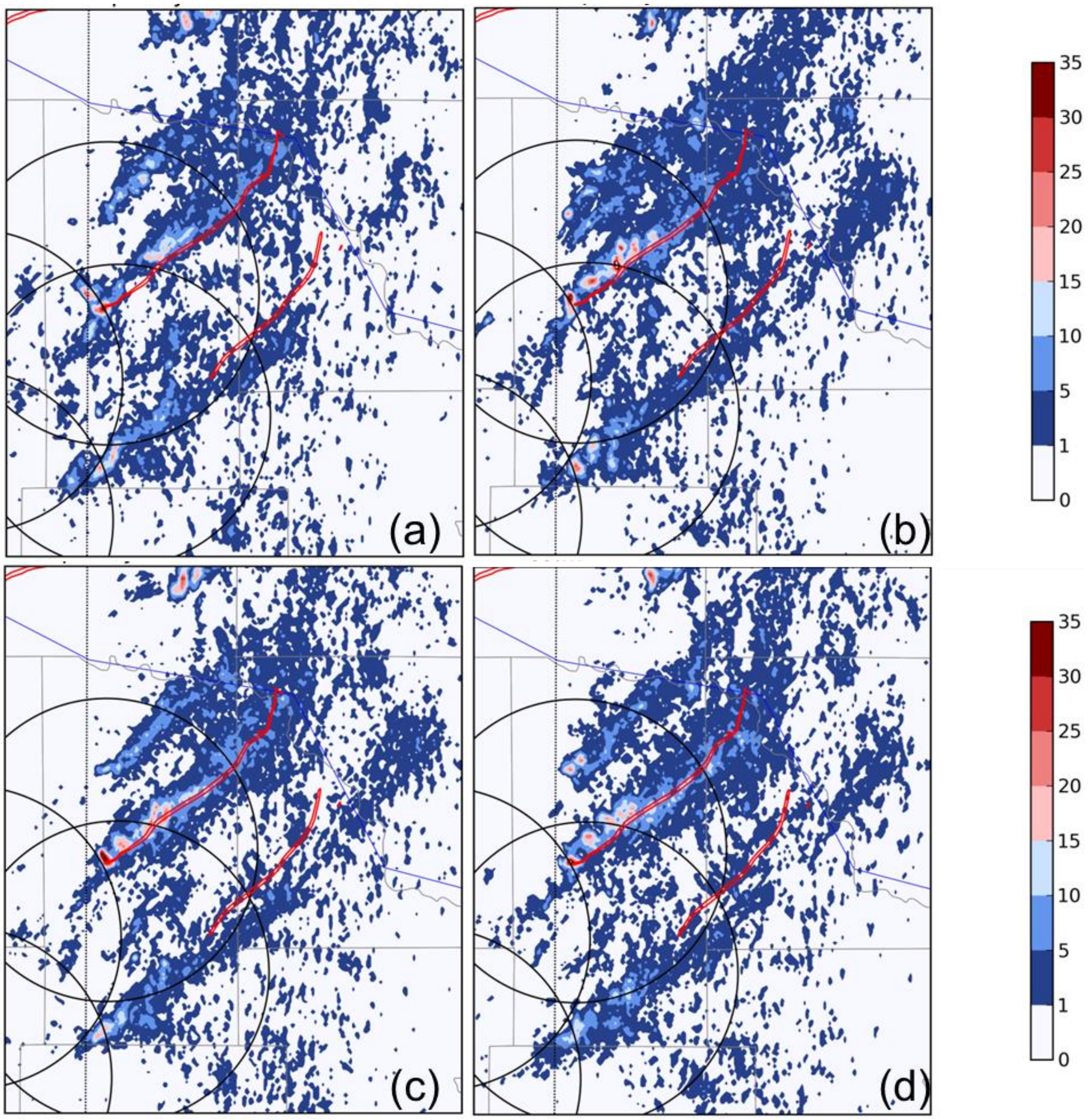


FIG 5. 29 Same as in FIG 5. 28 but when vorticity > 0.008 s^{-1}

5.5 Summary and Discussion

Testing the technique using real radar observations is more difficult than running OSSEs. The true state of atmosphere is not known; uncertain observation error amounts and sources exist in radar data processing (due to clutter removal; aliasing; bias; interpolating to model grids); and the DSD model used in microphysics scheme may not be accurate enough. The purpose of running real case experiments is to examine the effectiveness of our EnKF system with capability of assimilating attenuated radar observations directly. Thus compared to other similar EnKF works (Snook et al. 2011), we tend to assimilate reflectivity data in all thresholds, which can seriously affect the quality of analyses. Fortunately, our approach still produces similar convective storm structure in the data assimilation and forecast.

The 24 May 2011 tornado outbreak is selected as a case study since the storms spawned two EF4 tornadoes (C1 and D1) that developed inside the CASA IP1 radar network. To make the results comparable with each other, all experiments on the 500 m resolution grids use the same first guess interpolated from 1500 m resolution analysis at 2100 UTC. WSR-88D observations from KTLX and KFDR and the radial velocity observed from CASA IP1 network are assimilated in all experiments. The difference among those experiments are whether the reflectivity used in EnKF is pre-corrected or not, and the corresponding methods to deal with the reflectivity observations.

Despite the instability brought by the inconsistency existing in WSR-88D and CASA observations, assimilating reflectivity data at all thresholds do harms the quality of analysis. Although the observations provided by CASA are pre-corrected in observation space through traditional approach, the zero value reflectivity which might contain full attenuation that could not be corrected effectively tends to suppress the development of the storms and leads to a much weaker storm system. One way to avoid the impact of zero or near-zero reflectivity is omitting reflectivity below a threshold (e.g. 20 dBZ). Another way is applying the AEM that we proposed and tested in OSSEs to limit the impact brought by zero or near-zero reflectivity. In this real CASA case, the two methods, discarding weak reflectivity or applying AEM, are proven effective with pre-corrected CASA observations by presenting a more solid storm system than simply assimilating all values of pre-corrected CASA observations.

Once the attenuation correcting procedure is built into EnKF, it is possible to assimilate the un-corrected (attenuated) CASA reflectivity data directly. Combining with AEM, the analysis results are quite promising by analyzing a strong storm system with similar reflectivity structure as in the observations. The deterministic forecast experiment indicates that assimilating un-corrected observations directly actually retains some information that might be lost in the pre-corrected CASA observations. The experiments assimilating un-corrected CASA observations forecast a longer-lasting trailing line feature than experiments assimilating pre-corrected observations. The storm system forecasted in experiment using un-corrected CASA reflectivity manages to maintain strong radar echoes even after 50 minutes of forecast, while the storm system forecasted by experiments assimilating pre-corrected observations dissipates sooner. The comparison with 2 m temperature and relative humidity observed by Oklahoma Mesonet shows a small advantage when using un-corrected CASA data over the experiments using pre-corrected CASA observation. The difference is not very large, though.

The results of ensemble forecasts indicate that experiments using un-corrected observations have more members that predict low-level vortices, a signature of possible tornadoes, above a specified (large) threshold than the experiments using pre-corrected observations early in the lifetime of tornado C1. Thusly, we believe that directly assimilating raw observations using EnKF can preserve more information than using pre-processed observations. Again, the difference between the experiment using un-corrected CASA data, the experiment using pre-corrected observations above 20 dBZ, and the experiment assimilating pre-corrected observation using AEM is not obvious. But the above three experiments all do a better job than the experiment simply assimilating all values of pre-corrected CASA data.

The assimilation window we chose here is a period when many thunderstorms had begun to develop and intensify rapidly. Limited by the 40 km range, using CASA data alone could not provide enough coverage of those storms and let spurious echoes grow. For this reason, we use S-band WSR-88D observations to suppress those spurious echoes, though doing so may mask the impact of attenuated data on the EnKF analyses. One may tempted to think that WSR-88D data alone can also provide a promising analysis. However, after simple tests, we found that the cold pool analyzed in the experiment that use WSR-88D only spread out too wide and quick. Similar results are also obtained in Schenkman (Schenkman et al. 2011), in which CASA and WSR-88D observations are assimilated using a 3DVAR method.

As a part of AEM, the way to distinguish fully attenuation from zero reflectivity used in this research is simply searching each beam gate by gate. Once certain large values are found in one gate, all the bins after it will are flagged as possibly containing attenuation. Combined with the judgment inside the AEM, fully attenuation can be recognized and skipped. This method may, however, treat some real clear air echoes as fully-attenuated. The differential propagation phase shift Φ_{DP} could also assist in flagging the attenuation. When SNR is low, Φ_{DP} is difficult to use, though. It would be very helpful if CASA and other CASA-like X-band radar network could provide such information to tell attenuation has happened or not.

Chapter 6: Summary and Future Work

6.1 General summary

In this study, a simultaneous state estimation and attenuation correction method for convective storm systems within an ensemble Kalman filter (EnKF) data assimilation (DA) system has been improved and tested using Observing System Simulation Experiments (OSSEs) and a real case. Advanced data assimilation technologies, such as ensemble Kalman filter (EnKF) methods, offer flexibility to do simultaneous attenuation correction within the data assimilation procedure. By calculating the expected attenuation within the forward observation operators using the estimated atmospheric state, attenuation correction is built into the ensemble square-root Kalman filter (EnSRF) data assimilation system.

We first apply this method to a simulated convective squall line system that has a much broader precipitation region and hence substantially more attenuation than an isolated supercell storm examined in (Xue et al. 2009) . A network of four X-band radars positioned according to the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) integrated project 1 (IP1) radar network is assumed in the OSSEs. Overlapping observation coverage afforded by multiple radars is very helpful in producing more accurate reflectivity analyses as multiple viewing angles reduces the areas of complete attenuation.

Considering that observations with severe attenuation usually contain larger errors than data with little or no attenuation, and zero reflectivity has large uncertainty (maybe an area of complete attenuation or simply an area with no scatters), an empirical analytical observation error model (AEM) based on a reflectivity relation is designed and used to specify the observation error variance in the EnKF data assimilation. In such a relation, larger observation errors are assigned to lower observed reflectivity values, while higher reflectivity can be weighted more by specifying a smaller error. The clear-air data are properly taken care of by flagged out during the process of producing simulated radar observation. Such flag can be read in during EnKF and the observation marked with fully attenuation will be discarded. A small error variance is assigned to relatively small value of observed data during analysis.

The effectiveness of the attenuation correction procedure and the utilization of a reflectivity-dependent AEM are further tested in the presence of the resolution-related error in the assimilating model. This error is introduced by using a truth simulation that is produced at 3 times the resolution as that of the assimilating model. It was shown that while the analysis errors increase when such model error exists, the model error impact is smaller than the positive impacts of the attenuation correction procedure and the use of the AEM.

When multiple error sources are present in the observations (e.g. the instrumental error, the reflectivity model error, the attenuation model error, the Signal-to-noise ratio related error, and the inhomogeneity error) AEM is still effective and can reduce the RMSEs levels quickly. Meanwhile, another type of observation error model (multiple sources error model, or MSEM) that can use the same calculation equations to estimate the observation error deviation during EnKF is proposed and tested, too. The preliminary results show MSEM helps to improve the analysis, but not as much as the AEM does. In this set of experiments the attenuation model error that related to the PIA was designed as a Gaussian distribution, but in reality, this type of error should be spatially correlated. Thus, further experiments in which the attenuation model error has spatial correlation are planned. MSEM is actually an attempt to design a more sophisticated observation error model than the AEM. And we see there are more potential in this type of observation error model.

The attenuation correction EnKF system is used to explore the problem of expanded X-band radar network site selection, since CASA had previously planned to expand the IP1 radar network. Experiments are designed based on assimilating observations from possible radar network configuration, that is, 4 CASA IP1 radars with additional 2 hypothetical radars. The trade-off between radar overlapping coverage and spatial coverage has been evaluated. The impact of storm propagation speed and assimilation window length on the quality of analysis and forecast are also examined.

For a slow moving squall line storm system, with severe attenuation effects, the radar network with larger spatial coverage is preferable than the one with smaller spatial coverage and more overlapping coverage, even though the latter can provide more Dual- or Multi-Doppler observations, in both analysis and forecast. Since the slow-moving storm system is nearly stationary within radar network during DA, assimilating radar observation every 5 min for an hour is sufficient to obtain the main structure of the storm. Dual- or Multi-Doppler observations can provide details in smaller-scale storm structures but lose more information of the storm system in whole owing to more limited spatial coverage. Radars located upstream of a storm can aid model to a better analysis of potential temperature and generated a better stratiform precipitation forecast. Assimilating observations 15 minutes more can produce better estimate of potential temperature and water vapor but worse *w* field and hydrometeor fields, which are mainly due to the storm moving out of the network center, and lack of Dual-Doppler observations could not aid DA system to correct attenuation effectively. Forecasts starting from different assimilation windows do not show significant differences, because the error levels increase rapidly to saturation levels no matter how low they started. One possible explanation is that the lack of spatial coverage cannot provide enough observations to model to capture the correct dynamic information of storm system, no matter how long the assimilation windows are. Model error starts to dominate the forecast, and error increases very quickly to counteract the positive impact of data assimilation.

In experiments where the storm system moves quickly through the radar network, radars located downstream become important because they provide sufficient observations to suppress spurious echoes during later assimilation cycles. Although spatial coverage is reduced, the radar network with more Dual- or Multi-Doppler features shows more improvement in smaller-storm scale structure, especially in the forecast. When combined with sufficient downstream coverage, ETS shows that experiments with more Dual- or Multi-Doppler observations generate a more accurate forecast of convective storm evolution. Since assimilating observations 15 minutes longer faces the situation that squall line system is about to move out of some radar networks, starting the forecast from longer assimilation windows means starting from a much worse analysis. Experiments with less upstream observations tend to forecast worse potential temperature and water vapor fields when the forecast starts from the longer assimilation window than from the shorter window. The forecast of hydrometeor fields is not affected much by the length of assimilation window though, as their error reaches a saturation level quickly.

In general, on the condition of only X-band or short-wavelength radar network being available, a radar network with larger downstream spatial coverage tends to provide the best overall analysis and 1-hour forecast. However, if the emphasis is on convective scale structure analysis and forecast, more Dual- or Multi-Doppler coverage is preferred despite losing some spatial coverage.

Because of the lack of knowledge of the true atmosphere state, observational uncertainty amount, and even the sources of error in the data assimilation system (e.g. observation error, model error), applying our attenuation correction EnKF system to a real case is quite challenging. However, our procedure shows promising results. Once the attenuation correcting procedure is built into EnKF, it is possible to assimilate the un-corrected CASA reflectivity data directly. Combining with AEM, the analysis results in a much stronger storm system, close to the WSR- 88D observations, compared to the experiment that simply assimilates all value of pre-corrected observation without AEM applied. Subsequent deterministic forecast experiments indicate that assimilating un-corrected CASA observations directly has retained some information that might be lost in the pre-corrected CASA observations. The experiments assimilating un-corrected CASA observation forecast a longer lasting trailing line feature that disappears quickly in experiments assimilating pre-corrected observations. Also, 40 member ensemble forecast experiments reveal that the experiment using un-corrected observations has more members predict low-level vortices stronger than a specified threshold, a signature of possible tornadoes, than the experiment using all values of pre-corrected observations without the application of AEM. And the prediction of the track of tornado D1, although not accurate, is still more intense than the one predicted by the experiments using all values of pre-corrected observations. Thus, we believe that directly assimilating raw observations using EnKF, is a competitive way to use observations from X-band radars, and has the potential to preserve more information than using pre-processed observations.

6.2 Concluding remarks and future work

In this research we test the forward observation operator with attenuation and develop an analytical observation error model designed especially for the attenuated observation data assimilation problem. This is the first time this method tested using OSSEs with a radar network and applied to real data cases, with promising results.

Our method is mainly based on model state variables estimation in EnKF. It does not require dual-Polarization radar measurements and does not need to estimate PIA independently thus avoiding problems with numerically instability which are found in traditional attenuation correction algorithms.

The capability of the technique allows us to use attenuated radar observations directly, particularly when the observations come from CASA X-band network, which is an attractive direction for radar data assimilation. More and more short-wave radars are used for scientific research purposes (e.g. mobile radars, airborne radars). Applying our method to those radar observations can provide another way to ingest information. More case studies should be made in the future in order to get more robust conclusions.

The application of AEM is a way to limit the possible negative impact brought by near-zero reflectivity observations while retaining useful information contained in those weak reflectivity observations. Currently, the AEM we used in all the experiments is mainly depends on the value of reflectivity. The only way to flag fully attenuated data is to use information from along the beampath. We would like to further improve AEM by including more information that could be obtained from radar observations. For example, φ_{DP} could be used to recognize attenuation. Large errors are also contained in the melting layer, which means large values of reflectivity may also contain errors. The MSEM is another potential direction to improve the observation error model that is designed especially for use in attenuated radar data assimilation.

As another essential part of AEM, the procedure to identify clear air returns needs to be improved, too. We will not only flag the data where raw reflectivity and corrected reflectivity are not equal, but also search for sudden reflectivity value changes gate by gate along each beam during processing CASA radar. Doing so may help to identify areas where complete attenuation has occurred; large observation error can be used in these areas thus improving the data assimilation of X-band radar data.

Currently, the calculation of attenuation in EnKF is performed sequentially, consuming a lot of computation time (around 160 minutes in total to assimilate observation from 2 WSR-88D radars and 4 CASA radars for one analysis step, while the time used in the MPI version and without the attenuation correction procedure inside EnKF is only around 36 minutes).This attenuation correcting procedure needs to be merged into the distributed-memory parallel version (DMP) of ARPS-EnSRF (Wang 2013) to save computational cost. Also, it would be interesting to see how the attenuation correction procedure performs in ARPS LETKF system. Different from the EnSRF system, all the observations are assimilated at the same time in LETKF. Under that circumstance, will AEM be still useful? Merging the current AEM into the LETKF system will be another challenge.

The forward observation operator designed for attenuated radar observations are non-local, just like those for satellite radiance data. As suggested by the so-called hybrid algorithm, which uses a combination of the static and ensemble derived flow-dependent covariance, i.e. the hybrid covariance (Lorenc 2003; Buehner et al. 2010), the state-space-based covariance localization used in the hybrid formulation is potentially advantageous for non-local forward observation operator. Thus, testing the attenuation correcting procedure in the hybrid EnKF framework is another interesting research direction.

The parameterized relations between model-predicted total mass content W with equivalent reflectivity factor Z_e and the attenuation coefficient k are derived based on the exponential form Drop size distributions (DSDs) in the 5-class single-moment microphysics scheme of Lin et al. (1983). We can also derive the relations based on different DSDs forms, and see how this approach works in two-moment or multi-moment microphysics parameterization. Accurate representation of microphysical processes is especially important for attenuated radar data assimilation. Polarimetric radars observations could be assimilated into storm-scale numerical models to improve the estimation of the shape of DSD. The correct shape of DSD itself can affect other microphysical processes, which in turn can influence the thermodynamic and kinematics of the storm. Improving the DSD can also help the estimation of the attenuation in our EnKF. Additionally, dual-pol observations can produce the "hydrometeor classification" product. The current version of the hydrometeor classification algorithm classifies radar echoes as light/moderate ran heavy rain, "big drops", rain/hail mixture, biological scatterers, ground clutter/AP, dry snow, wet snow, ice crystals and graupel. This classification product could be used to identify potential attenuation areas versus clear air echoes, which is another possible way to improve our AEM.

Given the general challenges facing storm-scale data assimilation using radar data, and as the first step to using attenuated observation directly and try to extract as much information as possible, much research is still needed in these area.

References

Anagnostou, M. N., E. N. Anagnostou, and J. Vivekanandan, 2006: Correction for rain path specific and differential attenuation of X-band dual-polarization observations *IEEE Trans. Geosci. Remote Sensing*, **44,** 2470-2480.

Anderson, J. L., 2001: An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, **129,** 2884-2903.

Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127,** 2741-2758.

Brewster, K., L. White, B. Johnson, and J. Brotzge, 2005: Selecting the Sites for CASA NetRad, a Collaborative Radar Network. *9th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), AMS Conf.,*.

Bringi, V. N., T. D. Keenan, and V. Chandrasekar, 2001: Correcting C-band radar reflectivity and differential reflectivity data for rain attenuation: A self-consistent method with constraints, . *IEEE Trans. Geosci. Remote Sensing*, **39,** 1906-1915. .

Bringi, V. N., V. Chandrasekar, N. Balakrishnan, and D. S. Zrnic ́, 1990: An examination of propagation effects in rainfall on radar measurements at microwave frequencies. *J. Atmos. Oceanic Technol.*, **7,** 829-840.

Brotzge, J., K. Hondl, B. Philips, L. Lemon, E. J. Bass, D. Rude, and D. L. Andra, 2010: Evaluation of Distributed Collaborative Adaptive Sensing for Detection of Low-Level Circulations and Implications for Severe Weather Warning Operations. *Weather and Forecasting,* **25,** 173-189.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.***,** doi: 10.1175/2009MWR3157.1.

Burrows, D. R. a. A., S. S., 1949: Radio Wave Propagation. Academic Press.

Chandrasekar, V. and S. Lim, 2008: Retrieval of Reflectivity in a Networked Radar Environment. *Journal of Atmospheric and Oceanic Technology*, **25,** 1755-1767.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

Doviak, R. and D. Zrnic, 1993: *Doppler Radar and Weather Observations*. 2nd ed. Academic Press, 562 pp.

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.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53,** 343-367.

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. *Monthly Weather Review*, **140,** 2609-2627.

Gao, J., K. Brewster, and M. Xue, 2006: A comparison of the radar ray path equations and approximations for use in radar data assimilation. *Adv. Atmos. Sci.*, **23,** 190-198.

Gilleland, E., D. A. Ahijevych, B. G. Brown, and E. E. Ebert, 2010: Verifying Forecasts Spatially. *Bulletin of the American Meteorological Society*, **91,** 1365-1373.

Gorgucci, E. and V. Chandrasekar, 2005: Evaluation of Attenuation Correction Methodology for Dual-Polarization Radars: Application to X-Band Systems. *Journal of Atmospheric and Oceanic Technology*, **22,** 1195-1206.

Gorgucci, E., V. Chandrasekar, and L. Baldini, 2006: Correction of X-Band Radar Observation for Propagation Effects Based on the Self-Consistency Principle. *Journal of Atmospheric and Oceanic Technology,* **23,** 1668-1681.

Gorgucci, E., G. Scarchilli, V. Chandrasekar, and V. N. Bringi, 2001: Rainfall Estimation from Polarimetric Radar Measurements: Composite Algorithms Immune to Variability in Raindrop Shape-Size Relation. *J. Atmos. Oceanic Tech.*, **18,** 1773-1786.

Hitschfeld, W. and J. Bordan, 1954: Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *J. Meteor.*, **11,** 58-67.

Hogan, R. J., 2007: A variational scheme for retrieving rainfall rate and hail reflectivity fraction from polarization radar. *J. App. Meteor. Climatology*, **46,** 1544-1564.

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

Jameson, A. R., 1992: The effect of temperature on attenuation-correction schemes in rain using polarization propagation differential phase shift. *J. Appl. Meteor.,* **31,** 1106-1118.

Jolliffe, I. T. and D. B. Stephenson, 2003: *Forecast Verification*. WILEY, 240 pp.

Jung, Y., G. Zhang, and M. Xue, 2008a: Assimilation of simulated polarimetric radar data for a convective storm using ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Wea. Rev.*, **136,** 2228-2245.

Jung, Y., M. Xue, G. Zhang, and J. Straka, 2008b: Assimilation of simulated polarimetric radar data for a convective storm using ensemble Kalman filter. Part II: Impact of polarimetric data on storm analysis. *Mon. Wea. Rev.*, **136,** 2246-2260.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, 341 pp.

Laws, J. O. a. P., D. A., 1943: The relationship of raindrop size to intensity. *Trans. Am. Geophys.***,** 452-460.

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

Lim, S., V. Chandrasekar, P. Lee, and A. P. Jayasumana, 2010: Real-Time Implementation of a Network-Based Attenuation Correction in the CASA IP1 Testbed. *Journal of Atmospheric and Oceanic Technology,* **28,** 197-209.

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

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

Lorenc, A., 2003: Modelling of error covariances by 4DVAR data assimilation. *Qart. J. Roy. Meteor. Soc.*, **129,** 3167-3182.

Matrosov, S. Y., K. A. Clark, B. E. Martner, and A. Tokay, 2002: X-band polarimetric radar measurements of rainfall. *J. Appl. Meteor.*, **41,** 941-952.

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

Meneghini, R. and T. Kozu, 1990: Spaceborne Weather Radar. Ar-tech House. *199*.

Park, S. G., V. N. Bringi, V. Chandrasekar, M. Maki, and K. Iwanami, 2005a: Correction of Radar Reflectivity and Differential Reflectivity for Rain Attenuation at X Band. Part I: Theoretical and Empirical Basis. *Journal of Atmospheric and Oceanic Technology*, **22,** 1621-1632.

Park, S. G., M. Maki, K. Iwanami, V. N. Bringi, and V. Chandrasekar, 2005b: Correction of Radar Reflectivity and Differential Reflectivity for Rain Attenuation at X

Band. Part II: Evaluation and Application. *Journal of Atmospheric and Oceanic Technology*, **22,** 1633-1655.

Philips, B., 2012: CASA DFW URBAN DEMONSTRATION NETWORK. Approaches for determining economic benefits. *AMS Washington Forum*.

Putnam, B. J., M. Xue, G. Zhang, Y. Jung, N. Snook, and A. D. Schenkman, 2010: Comparison of the structural evolution and polarimetric variable fields of a forecasted MCV using a single and two-moment microphysics scheme. *25th Conf. Severe Local Storms*, Amer. Metero. Soc., Paper P7.6.

Schaefer, J. T., 1990: The Critical Success Index as an Indicator of Warning Skill. *Weather and Forecasting*, **5,** 570-575.

Schenkman, A., 2008: High-Resolution Assimilation of CASA and WSR-88D Data for the Prediction of a Tornadic Convective System, School of Meteorology, University of Oklahoma, 197.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011: Impact of CASA Radar and Oklahoma Mesonet Data Assimilation on the Analysis and Prediction of Tornadic Mesovortices in an MCS. *Monthly Weather Review*, **139,** 3422-3445.

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. *Monthly Weather Review*, **139,** 3446-3468.

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

Snyder, J. C., H. B. Bluestein, G. Zhang, and S. J. Frasier, 2010: Attenuation Correction and Hydrometeor Classification of High-Resolution, X-band, Dual-Polarized Mobile Radar Measurements in Severe Convective Storms. *Journal of Atmospheric and Oceanic Technology*, **27,** 1979-2001.

Sun, J. and N. A. Crook, 1998: Dynamical and Microphysical Retrieval from Doppler Radar Observations Using a Cloud Model and Its Adjoint. Part II: Retrieval Experiments of an Observed Florida Convective Storm. *J. Atmos. Sci.*, **55,** 835-852.

Sun, J., N. A. Crook, and J. Verlinde, 1994: Dynamical and mirophysical retrieval from Doppler radar observations using a cloud model and its adjoint. *Prec. 10th Conf. on Numer. Wea. Predition*, Portland, OR, Amer. Meteor. Soc., 466-468.

Testud, J., 2000: The Rain Profiling Algorithm Applied to Polarimetric Weather Radar. *J. of Atmos. and Oceanic Technol.*, **17,** 332-356.

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

Tong, M. and M. Xue, 2008: Simultaneous Estimation of Microphysical Parameters and Atmospheric State with Simulated Radar Data and Ensemble Square Root Kalman Filter. Part I: Sensitivity Analysis and Parameter Identifiability. *Monthly Weather Review,* **136,** 1630-1648.

Tuttle, J. D. and R. E. Rinehart, 1983: Attenuation Correction in Dual-Wavelength Analyses. *Journal of Climate and Applied Meteorology,* **22,** 1914-1921.

Wang, Y., Y. Jung, T. A. Supinie, and M. Xue, 2013: A hybrid MPI/OpenMP parallel algorithm and performance analysis for an ensemble square root filter suitable for dense observations. *J. of Atmos. and Oceanic Technol.,* **30,** 15.

Weygandt, S. S., A. Shapiro, and K. K. Droegemeier, 2002: Retrieval of Model Initial Fields from Single-Doppler Observations of a Supercell Thunderstorm. Part I: Single-Doppler Velocity Retrieval. *Mon. Wea. Rev.,* **130,** 433-453.

Whitaker, J. S. a. T. M. H., 2010: A simple state-dependent multiplicative inflation algorithm. *The 4th EnKF Workshop*.

Xiao, Q., Ying-Hwa Kuo, J. Sun, W.-C. Lee, E. Lim, Y.-R. Guo, and D. M. Barker, 2005: Assimilation of Doppler radar observations with a regional 3DVAR system: Impact of Doppler velocities on forecasts of a heavy rainfall case. *J. App. Meteor.,* **44,** 768-788.

Xu, Q., 1996: Generalized adjoint for physical processes with parameterized discontinuities - Part II: Vector formulations and matching conditions. *J. Atmos. Sci.*, **53,** 1143-1155.

Xue, M., 2000: High-Order Monotonic Numerical Diffusion and Smoothing. *Monthly Weather Review*, **128,** 2853-2864.

Xue, M. and S.-J. Lin, 2001: Numerical Equivalence of Advection in Flux and Advective Forms and Quadratically Conservative High-Order Advection Schemes. *Monthly Weather Review*, **129,** 561-565.

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

Xue, M., S. Liu, and T. Yu, 2007: Variational analysis of over-sampled dual-Doppler radial velocity data and application to the analysis of tornado circulations. *J. Atmos. Ocean Tech.*, **24,** 403-414.

Xue, M., M. Tong, and G. Zhang, 2009: Simultaneous state estimation and attenuation correction for thunderstorms with radar data using an ensemble Kalman filter: Tests with simulated data. *Q. J. Roy. Meteor. Soc.*, **135,** 1409-1423.

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

Zhang, L., Z. Pu, W.-C. Lee, and Q. Zhao, 2011: The influence of airborne Doppler radar data quality on numerical simulations of a tropical cyclone. *Weather and Forecasting*.

Zhu, K., Y. Pan, M. Xue, X. Wang, J. S. Whitaker, S. G. Benjamin, S. S. Weygandt, and M. Hu, 2013: A regional GSI-based EnKF system for the Rapid Refresh configuration: Results with a single, reduced resolution. *Monthly Weather Review*.

Appendix A: Calculation of *Z_e-W* and *k-W* relations

The parametrized relations between model-predicted W with Z_e and k are derived through the following procedure:

1) Use Equation (9) in (10) and solve for Λ as a function of *W*, yielding the slope parameter $\Lambda = (\frac{N_0 \pi \rho}{W})^{1/4}$.

2) Vary *W* in its possible range then calculate Λ and the corresponding Z_e and k through Equation (11) and (12). The backscattering radar cross-sections and the attenuation or extinction cross-sections are calculated through Mie theory method.

3) Performing least-square fitting to the data for Z_e-*W* and *k*-*W* in log domain, leading to power-law relations

$$Z_e = \alpha_Z W^{\beta_Z} \text{ and } k = \alpha_k W^{\beta_k} \qquad (17)$$

The units for *W*, Z_e and *K* are $\mathrm{gm^{-3}}$, $\mathrm{mm^6m^{-3}}$ and $\mathrm{dBkm^{-1}}$, respectively. The above procedure is applied to all hydrometeor species and the results are given below.

A.1 Rainwater

Based on Equation (17), we are assuming Mie scattering and a 10 ℃ temperature. The relative dielectric constant used is $\varepsilon_r = 55.43 - 37.85i$, a complex number. Rain intercept parameter assumes the default value, $N_{0r} = 8\times10^6$ m^{-4}, of the Marshal-Palmer DSD. The resulting parameters in (17) are $\alpha_{Zr} = 2.53\times10^4$, $\beta_{Zr} = 1.84$, $\alpha_{kr} = 0.319$ and $\beta_{kr} = 1.83$, where subscript r denotes rain.

A.2 Dry snow and hail

The calculation and fitting procedures for dry snow and dry snow are the same as those for rain. Mie scattering theory is used for hail and snow because they have little polarization signatures. For snow, $N_{0s} = 3\times10^6$ m^{-4}, $\rho_s = 0.1$ gcm^{-3} (s is for snow), and for hail $N_{0h} = 4\times10^4$ m^{-4}, $\rho_h = 0.917$ gcm^{-3}(h is for hail). The resulting parameters in (17) are $\alpha_{Zds} = 3.48\times10^3$, $\beta_{Zds} = 1.66$, $\alpha_{Zdh} = 8.18\times10^4$, $\beta_{Zdh} = 1.50$, $\alpha_{kds} = 0.00483$, $\beta_{kds} = 1.28$, $\alpha_{kdh} = 0.159$ and $\beta_{kdh} = 1.64$.

A.3 Melting snow and hail

The melting ice model of Jung *et al*. (2008) is used to derive the formulae for melting or wet snow and hail. Using the same procedure in Jung *et al*. (2008), we can obtain the coefficients for the power-law relations for the wet snow:

$$\begin{aligned}
\alpha_{Zms} &= (0.00491 + 5.75 f_{ws} - 5.58 f_{ws}^2)\times10^5, \\
\beta_{Zms} &= 1.67 - 0.202 f_{ws} + 0.398 f_{ws}^2, \\
\alpha_{kms} &= 0.0413 + 22.7 f_{ws} - 50.5 f_{ws}^2 + 28.6 f_{ws}^3, \\
\beta_{kms} &= 1.06 - 0.579 f_{ws} + 2.03 f_{ws}^2 + -1.24 f_{ws}^3,
\end{aligned} \qquad (18)$$

and for wet hail.

$$\begin{aligned}
\alpha_{Zmh} &= (0.809 + 10.13 f_{wh} - 5.98 f_{wh}^2)\times10^5, \\
\beta_{Zmh} &= 1.48 + 0.0448 f_{wh} - 0.0313 f_{wh}^2, \\
\alpha_{kmh} &= 0.256 + 6.28 f_{wh} - 11.36 f_{wh}^2 + 6.01 f_{wh}^3, \\
\beta_{kmh} &= 1.26 - 0.659 f_{wh} + 1.44 f_{wh}^2 - 0.817 f_{wh}^3.
\end{aligned}$$

f_w is the melting percentage that can be calculated. Detail information for calculating f_w can be found in Jung *et al*. (2008).