UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IMPROVING SHORT-TERM FORECAST OF SEVERE AND HIGH-IMPACT WEATHER EVENTS USING A WEATHER-DEPENDENT HYBRID ENSEMBLE-VARIATIONAL DATA ASSIMILATION SYSTEM WITH RADAR AND SATELLITE DERIVED OBSERVATIONS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

SIJIE PAN

Norman, Oklahoma

2023

IMPROVING SHORT-TERM FORECAST OF SEVERE AND HIGH-IMPACT WEATHER EVENTS USING A WEATHER-DEPENDENT HYBRID ENSEMBLE-VARIATIONAL DATA ASSIMILATION SYSTEM WITH RADAR AND SATELLITE DERIVED OBSERVATIONS

A DISSERTATION APPROVED FOR THE

SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Xuguang Wang, Chair

Dr. Jidong Gao, Co-Chair

Dr. Nathan Goodman

Dr. Cameron Homeyer

Dr. Thomas Jones

Dr. Greg Mcfarquhar

Acknowledgements

I would like to express my sincere gratitude and appreciation to the individuals who have been instrumental in the successful completion of my dissertation. First and foremost, I would like to extend my heartfelt thanks to my esteemed advisors, Drs. Jidong Gao and Xuguang Wang, for their invaluable guidance and unwavering support. Their patience, expertise, and encouragement have been integral in helping me achieve my research goals. I am also grateful to my committee members - Drs. Thomas Jones, Greg Mcfarquhar, Cameron Homeyer, and Nathan Goodman - for their valuable feedback and support throughout my academic journey.

In addition, I am also indebted to my friends and colleagues, including Alex Fierro, Junjun Hu, Anwei Lai, Patrick Skinner, Qingyu Wang, Yunheng Wang, and Juan Zhao, for their assistance and encouragement during the course of my studies.

My thank also go to the National Severe Storms Laboratory (NSSL) and the Cooperative Institute for Severe and High-Impact Weather Research and Operations (CIWRO) for their unwavering support and guidance, especially Drs. Pamela Heinselman, Lou Wicker, and Nusrat Yussouf, whose expertise and professionalism have been invaluable to my research. Furthermore, I am grateful for the time and effort invested by the dedicated staff members at CIWRO and the School of Meteorology for their invaluable contributions to my dissertation. Their unwavering support, encouragement, and guidance have been integral in shaping my research and academic journey.

I would like to acknowledge the generous support of the National Oceanic and Atmospheric Administration (NOAA) Warn-on-Forecast project #NA16OAR4320115 and National Science Foundation grant AGS-1341878, as well as the NOAA/Office of Oceanic and Atmospheric Research under the NOAA-University of Oklahoma Cooperative Agreement #NA08OAR4320904.

The experiments were performed on the supercomputers Odin from the NOAA NSSL and Schooner from the University of Oklahoma.

Lastly, I would like to express my heartfelt appreciation to my parents, Songhai Pan and Hongmei Cui, and my aunt, Qiao Cui, for their unending love and support throughout my academic and personal journey. Without their constant encouragement and guidance, my achievements would not have been possible.

Table of Contents

List of Tables

List of Figures

Abstract

A hybrid ensemble-variational (EnVAR) data assimilation and forecast system, namely WoF-Hybrid, was initially developed at the National Severe Storms Laboratory (NSSL) supported by the NOAA Warn-on-Forecast project (WoF). It is a real-time, deterministic, weather adaptive, dual-resolution analysis and forecasting system in parallel with the WoF ensemble data assimilation and forecast system (WoFS), which provides an on-demand, short-term, rapidly updating, convection-allowing ensemble forecasts to the end users. Some new components for the WoF-Hybrid system are developed and tested in this dissertation to address a few scientific and technical challenges in assimilating the conventional, radar and satellite derived observations to improve the short-term numerical weather prediction (NWP) of convective-scale severe and high-impact weather events.

First of all, a novel component is introduced into the WoF-Hybrid system, which extends the observation forward operators to include cloud water path (CWP) and layered precipitable water (LPW). The impact of assimilating observations from several platforms on the estimation of initial states and the short-term forecasts using the WoF-Hybrid system is firstly investigated. The findings of this research reveal that (1) assimilating CWP improves the analysis of cloud properties, downward shortwave radiation over cloudy areas, surface temperature, and atmospheric instability, (2) assimilating the LPW product intensifies the moisture gradient at the surface and saturates the moisture fields surrounding ongoing convection, thereby enhancing the convection initiation (CI), and (3) low-level vertical wind shear is also improved through flow-dependent error covariances derived from the WoFS ensemble. Furthermore, a diagnostic analysis is conducted to explain why additional satellite-derived observations can improve the positioning and intensity forecast of storm cells, illustrated with three different hazardous severe weather events occurring in 2017.

In the next part, the dissertation proceeds to further develop and evaluate an updated version of the variational component of the WoF-Hybrid system (named as WoF-Var hereafter) by incorporating weather-dependent static background error covariances derived from the WoFS ensemble forecast products. The impact of weather-dependent background error structure on WoF-Var analyses is evaluated using a single observation experiment, which reveals the heterogeneous nature of the analysis increment for the meridional wind field under different weather situations. Two severe weather events that occurred on 5 May 2020 and 10 Dec 2021 are used to evaluate the impact of the weather-dependent error covariances on the WoF-var analysis and forecast system in detail. Diagnostic analysis of both events indicates that the utilization of weather-dependent errors in the WoF-Var leads to improved analyses of in-storm structure, particularly for the three-dimensional wind, temperature, and humidity fields, resulting in significantly better subsequent 0- to 6-hour forecasts.

The last part of the dissertation investigates the impact of weather-dependent error covariances on convective-scale NWP in the framework of WoF-Hybrid. In this, the flow-dependent ensemble error covariances are incorporated by using WoFS ensemble forecasts. Quantitative evaluations, aggregated over five selected severe and high-impact weather events, demonstrate that the experiments using weather-dependent covariances, instead of homogeneous covariances, in the hybrid system performs better in 0- to 6-hour reflectivity and rotational strength forecasts, particularly for strong convection with either higher reflectivity or higher updraft helicity. A more in-depth analysis of each event indicates that the adoption of weather-dependent covariances typically results in more skillful severe weather predictions regarding the areal coverage of reflectivity, the initiation and movement of storms, as well as the position and intensity of rotation tracks, even for the event that performed the worst out of the five selected events. The findings suggest that the weather-dependent error covariances within a hybrid system have added values to short-term convective-scale NWP.

Chapter 1 Introduction and overview

Some portions of this dissertation are direct excerpts of Pan et al. (2021; © American Meteorological Society). This includes parts of Chapters 1, 2 and 3.

1.1 Background and motivation

Per the statistics released by National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NOAA National Centers for Environmental Information, 2021), severe storms in the United States have caused an average of 84 fatalities and economic losses of $16.8 billion annually over the past decade. Accurate and timely forecasting of severe and high-impact weather events is crucial in minimizing the loss of life and property. However, the accurate forecast of hazardous severe weather is often compromised due to the complex non-linear interactions between processes of different scales and the inaccurate initial conditions for convective-scale numerical weather prediction (NWP) models. The meteorological community recognizes the significance of accurate initial conditions NWP, as it directly impacts the quality of forecasts. Obtaining precise initial conditions for hazardous weather events like tornadoes, hail, flash flooding, or damaging winds is vital in determining their timing, intensity, and location, which are crucial for the government and the public to prepare for and respond to such events, reducing the risk of fatalities, injuries, and economic losses. To achieve this, optimal combination of high-spatiotemporal resolution observations, such as satellite and radar observations, and previous NWP forecasts is necessary and can be achieved through the application of data assimilation (DA) schemes.

Significant progress has been made during the past two and a half decades in convective-scale radar DA using variational methods, including both three- and four-dimensional variational methods (3DVAR and 4DVAR respectively; e.g., Sun and Crook 1997, 1998, 2001; Crook and Sun 2002; Gao et al. 2004, 2013; Hu et al. 2006; Michel et al. 2011; Fierro et al. 2016; Wang et al. 2013; Wattrelot et al. 2014; Sun et al. 2016, 2020), Ensemble Kalman filter (EnKF) methods (e.g., Snyder and Zhang 2003; Tong and Xue 2005; Dowell et al. 2004, 2011; Aksoy et al. 2009; Jung et al. 2012; Yussouf and Stensrud 2010; Yussouf and Knopfmeier 2019; Skinner et al. 2018; Zhang et al. 2018, 2019; Jones et al. 2020; Pan and Gao 2022), as well as hybrid variational and EnKF methods (e.g., Gao and Stensrud 2014; Hamrud et al. 2015; Wang and Wang 2017; Kong et al. 2020; Pan et al. 2021; Li et al. 2022).

Among the above studies, an early version of a 3DVAR was developed for radar DA for convective-scale NWP (Gao et al. 2004; Hu et al. 2006; Ge 2011; Ge et al. 2012; Stensrud and Gao 2010) mainly supported by Center for Analysis and Prediction of Storms (CAPS) and National Science Foundation. The system can assimilate Doppler radar and other conventional and remote sensing data into nonhydrostatic NWP models, including the Weather Research and Forecasting Model (WRF, Skamarock et al. 2008) and Advanced Regional Prediction System (Xue et al. 2000, 2001). Multiple analysis loops are used having multiple spatial influence scales in order to accurately represent intermittent convective storm features. Early versions of this 3DVAR system, have for several years been used to initialize high resolution (down to 1 km grid spacing) ensemble WRF model forecasts using WSR-88D radar data as part of the spring NOAA Hazardous Weather Testbed (HWT)/Experimental Forecast Program (Kain et al. 2010; Clark et al. 2011; Gao et al. 2013; Smith et al. 2014; Calhoun et al. 2014; Hu et al. 2021). In the last 12 years, this 3DVAR system was upgraded by incorporating background error covariances estimated from ensemble forecasts supported by NOAA/National Severe Storms Laboratory (NSSL) Warn-on-Forecast (WoF) project (Gao and Stensrud 2014; Gao et al. 2016), so it was renamed as WoF-Hybrid. The early version of the WoF-Hybrid was mainly developed to assimilate radar data including both radial velocity and reflectivity. Detailed documentation of the DA system and its corresponding methodologies will be provided in Chapter 2.

Despite the benefits of assimilating radar observations, challenges have also been identified. Radar observations can only detect information about wind and some hydrometeor variables within storms, while variables outside storms and other model variables, such as temperature and specific humidity, are generally not well detected (Sun 2005). Consequently, significant errors and biases in the initial conditions, particularly for moisture and related thermodynamic variables, continue to pose major challenges in accurately predicting the location, timing, and intensity of hazardous severe weather events. Additionally, the assimilation of observations at different spatiotemporal scales into the NWP model is involved when observations other than high-density radar observations are incorporated. Therefore, in this dissertation, several new components for the WoF-Hybrid system are developed to investigate a better way to handle these problems. The new upgraded WoF-Hybrid system will have multiple capabilities to perform real-time, weather-adaptive, dual-resolution, continuous DA and forecast cycles for the active period of severe weather events.

Compared to our early version of the DA system, the newly developed DA system firstly aims to address the shortcomings, specifically with regards to assimilating only radar observations to improve prediction of hazardous severe weather events. With the launch of the new generation of Geostationary Operational Environmental Satellite (GOES-R series) and the availability of satellite observations and derived products, it has become more attractive to assimilate them into convective-scale NWP models to compensate for the insufficiency of radar DA. For example, Jones et al. (2013) assimilated satellite-derived cloud water path (CWP) in addition to radar observations within an EnKF framework and showed that the assimilation of CWP resulted in improvements in cloud properties and more accurate shortwave radiation flux at the surface by comparing to the experiments without CWP assimilation. Our previous study (Pan et al. 2018) explored the benefit of assimilating another satellite product total precipitable water, and showed that the potential of assimilating total precipitable water in improving the analysis of temperature and humidity through flow-dependent error covariances from an ensemble. However, these studies have not comprehensively examined the benefits of assimilating GOES-R derived satellite observations with more real data events. Therefore, the optimal DA strategy and corresponding configurations for assimilating satellite derived observations together with radar data to better predict real severe weather events need to be further investigated, which will be discussed in Chapter 3.

The handling of observations with a wide range of spatial and temporal scales has been a well-known difficulty for the development of multi-scale DA methods and covariance localization techniques (Gustafsson et al. 2018). In convective-scale DA, it is required to obtain accurate information for model variables on synoptic-, meso- and convective-scales simultaneously. Beside this, moist physical processes and unbalanced analysis of atmospheric states also need to be considered. Static background error covariances, which were firstly developed from and used in synoptic-scale DA and forecasts, are no longer suitable for convective-scale DA, as it can lead to forecasts without appropriate convection initiation and motion. In the EnKF schemes, flow-dependent background error covariances can be estimated from an ensemble, which has potential to be properly used in convective-scale NWP for severe weather forecasts. However, one of the major sources of errors with ensemble-based DA is rank deficiency or sampling error as a result of relatively small ensemble sizes (Houtekamer and Mitchell 2001; Wang et al. 2008a,b). Although this problem can be significantly reduced with inflation and localization, it could still happen sometimes, and could be even more severe with convective-scale DA because the total degrees of freedom of the convective weather systems may be significantly larger than the practical ensemble size. Therefore, both variational DA and ensemble DA schemes have their shortcomings respectively for convective-scale NWP.

The hybrid ensemble-variational technique proposed by Hamill and Snyder (2000), Buehner (2005), and Wang et al. (2007, 2008, and 2009) combines static and flow-dependent error covariances into a single framework and has potential to improve analyses compared to either EnKF or the pure variational based methods. As stated above, based on these studies, Gao and Stensrud (2014) developed the early version of WoF-Hybrid for convective-scale radar DA. However, the static background error covariances used were very simple and are not suitable for convective-scale real-time applications. Therefore, it is worthy further developing static background error covariances for convective-scale DA. In Chapter 4, weather-dependent static error covariances are developed and implemented in WoF-Hybrid and tested in a pure variational framework. Another task of Chapter 4 is to investigate if this new scheme is effective and efficient for improving analysis and prediction of severe and high-impact weather events.

The WoF-Hybrid was formally tested in the NOAA HWT Spring Forecast experiment (SFE) from 2019 to 2021 (Gao et al. 2021, 2022; Carpenter 2022). The SFE/HWT platform provides a unique opportunity to evaluate the use of any DA and forecast system as guidance for the prediction of severe and high-impact convective weather. So, it is very important to also test our newly developed WoF-Hybrid for some severe and high-impact weather events from 2020 and 2021 during the SFE/HWT test period. Chapter 5 of this dissertation will test newly updated WoF-Hybrid with several poorly performed cases, and especially focus on evaluating the performance of using weather-dependent error covariances with these events in both qualitative and quantitative manners.

1.2 Overview of the dissertation

This dissertation first describes the early version of the WoF-Hybrid and WoFS in Chapter 2. The rest of this dissertation describes the development of several pieces of new elements for WoF-Hybrid and employs this system to assimilate satellite derived observations and to evaluate the impact of weather-dependent error covariances for a variety of severe and high-impact weather events. More specifically, the following scientific questions are mainly addressed: (i) Would assimilating satellite-derived observations in addition to radar observations help improve the analysis and forecast for severe weather events? And (ii) are the weather-dependent error covariances able to address the problem associated with simultaneous assimilation of observations with different scales?

Chapter 3 first explores whether assimilating satellite-derived observations can improve forecasts of severe weather events and their relevant convection initiation mechanisms. The three severe and high-impact weather events of interest occurred in 2017 as a result of different synoptic situations. As part of the investigation, detailed diagnostics of the analyses and forecasts are performed in this chapter to explore the relative impact of different types of observations, especially satellite-derived CWP and LPW on preparing initial conditions for the convective-scale NWP model – WRF to improve the prediction of severe and high-impact weather. This part of the dissertation was already published in AMS journal Monthly Weather Review (Pan et al. 2021).

Next, Chapter 4 explores if prediction of severe and high-impact weather events can be improved by replacing homogenous static error covariances with the new weather-dependent error covariances. The method for computing the error covariances is detailed in this chapter. To evaluate the weather-dependent error covariances, a single observation experiment and two real data experiments which were not performed well in the quasi- real-time operational environment, are conducted.

Finally, Chapter 5 of this dissertation addresses the systematic performance of WoF-Hybrid with the new weather-dependent error covariances across five poorly performed events in 2020 and 2021 real-time application, especially for the rare high-impact and severe weather event – 10 December 2021 tornado outbreak which was named as the world’s No. 3 highest catastrophic weather and climate event by Cable News Network. These events also represent the wide spectrum of mechanisms for the initiation and development of strong convective storms. The evaluation is conducted in both qualitative and quantitative manner for composite reflectivity and midlevel updraft helicity forecasts. Finally, a general summary and future plans are given in Chapter 6.

Chapter 2 The Methods and Systems Description

2.1 Overview of the Early Version of the pure 3DVAR System

The WoF-Hybrid system is a development mainly based on the ARPS 3DVAR DA system developed at CAPS and NOAA/NSSL (Gao et al. 1999, 2004; Hu et al. 2006; Stensrud and Gao 2010; Gao and Stensrud 2012), which was used to specifically assimilate radar reflectivity and radial velocity and used for high-resolution (1 km) and high frequency (every 5 minutes) real-time analysis of atmospheric states (Gao et al. 2013). The cost function J is the sum of the background and observational terms plus a penalty or equation constraint term (J_c), and can be written as:

$$J(\mathrm{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[H(\mathbf{x}) - \mathrm{y}^o]^{\mathrm{T}}\mathbf{R}^{-1}[H(\mathbf{x}) - \mathrm{y}^o] + J_c, \tag{2.1}$$

where $\mathbf{x}$ and $\mathbf{x}^b$ are the analysis and background state vectors, respectively, and $\mathbf{y^o}$ is the observation vector. **B** and **R** are the background and observation error covariance matrices (see discussion below), respectively, and $H(\mathbf{x})$ represents the forward observation operators. Any penalty or dynamic equation constraints may be included in the quantity J_c in Eq. 2.1 to serve the important role of correlating the relevant analysis variables (Gao et al. 1999, 2004; Ge et al. 2012). In our current DA system, the background field ($\mathbf{x}^b$ in Eq. 2.1) is provided by a previous WRF forecast. Observations ($\mathbf{y^o}$ in Eq. 2.1) include Doppler radar data, single-level data (e.g., from surface Mesonet observations), multiple-level observations (such as from rawinsondes, wind profilers, and aircraft), and GOES-derived datasets, such as lightning, precipitable water and atmospheric motion vectors (Fierro et al. 2016, 2019; Zhao et al. 2021a,b, 2022).

To improve conditioning of the cost function minimization and avoid the need for computing the inverse of **B**, a new control variable, **v**, is introduced and is related to the analysis increment according to,

$$\mathbf{x} - \mathbf{x}^b = \mathbf{B}^{\frac{1}{2}}\mathbf{v}. \tag{2.2}$$

In practice, **B** is decomposed into several components to avoid difficulties for explicit calculation of large dimensional problems. As noted in Gao et al. (2004), the background error covariance matrix **B** can be treated as $\mathbf{C}^{\mathrm{T}}\mathbf{C}$, where **C** has the identical eigenvectors as **B** and eigenvalues as $\mathbf{B}^{1/2}$. By letting $\mathbf{C} = \mathbf{B}^{1/2}$, the cost function becomes:

$$J_{inc}(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}[\mathbf{HCv} - \mathrm{d}]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{HCv} - \mathrm{d}] + J_c(\mathbf{v}), \tag{2.3}$$

where **H** is the linearized version of H and $\mathrm{d} \equiv \mathrm{y}^o - H(\mathbf{x}^b)$. Consequently, the minimization is performed in the space of **v**. The matrix **C** is further defined as:

$$\mathbf{C} = \mathbf{DF}. \tag{2.4}$$

Here **D**, representing the background error standard deviations, is a diagonal matrix. The standard deviations for the model variables are derived from the statistics of the Rapid Update Cycle (Benjamin et al. 2004) model 3-hour forecasts over 4 years from 2001 to 2004. A newly developed heterogeneous error profile based on reflectivity partitioning, which is used for replacing this homogeneous error profile, will be detailed described in Chapter 4. The square root of a matrix with diagonal elements equal to one and off-diagonal elements representing the background error correlation coefficients is characterized by **F**. **F** is modeled by the recursive filter (Purser et al. 2003a, b). Multiple passes (or outer loops) can be applied to the recursive filter for assimilating observations with different spatial scales. For example, surface observations are assimilated in the first loop with wider radius, while radar observations are assimilated in the second loop with narrow radius.

2.2 Overview of the Warn-on-Forecast Hybrid DA and Forecast System

The system described in Section 2.1 was extended to be implemented in a hybrid ensemble and variational DA scheme (Gao and Stensrud 2014; Gao et al. 2016) by combining the static error covariances and flow-dependent error covariances from an ensemble. To achieve this goal, a so-called alpha control method is implemented by augmentation of the state vector, from **v** to (**v**, **w**) (Hamill and Snyder 2000; Lorenc 2003; Buehner 2005; Buehner et al. 2010a,b, 2015;Wang et al. 2007, 2008 and 2009). The cost function then can be written as:

$$J = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}[H(\bar{\mathbf{x}}^b + \Delta\mathbf{x}) - \mathrm{y}^o]^{\mathrm{T}}\mathbf{R}^{-1}[H(\bar{\mathbf{x}}^b + \Delta\mathbf{x}) - \mathrm{y}^o] + J_c(\mathbf{v}), \quad (2.5)$$

$$\Delta\mathbf{x} = \beta_1\mathbf{B}^{\frac{1}{2}}\mathbf{v} + \beta_2\mathbf{P}^{\frac{1}{2}}\mathbf{w}, \quad (2.6)$$

where $\Delta\mathbf{x}$ is the analysis increment, **P** is the ensemble-based flow-dependent covariance matrix, and **w** is the augmented control vector associated with **P**. If no covariance localization is applied, **P** is simply the rectangular matrix whose columns are ensemble perturbations divided by square root of *N*-1, where *N* is the ensemble size. The localization of the ensemble covariance in a variational system with preconditioning was discussed in Lorenc (2003) and Wang et al. (2008). The parameters β_1 and β_2 control the relative weight given to each covariance matrix. The following rule should be applied:

$$\beta_1 + \beta_2 = 1. \quad (2.7)$$

By minimizing the cost function (Eq. 2.5), optimal analyses of control vectors **v** and **w** can be obtained, and then the optimal analysis increment $\Delta\mathbf{x}$ can be derived from Eq. 2.6 on the fine-scale convection-resolving grid. As discussed previously, **P** is composed of ensemble perturbations with covariance localization. In this study, the ensemble background error covariance matrix **P** is calculated based on the ensemble forecasts from the WoF ensemble data assimilation and forecast system (WoFS) that uses a customized Advanced Research WRF Model (WRF-ARW) based on version 3.8.1 (Skamarock et al. 2008), coupled with the GSI-based EnKF system (Jones et al. 2016, 2020; Yussouf and Knopfmeier 2019; Pan and Gao 2022). The WoF-Hybrid system with real data cases was evaluated by our several recent studies (Wang et al. 2019; Pan et al. 2021). The system was also formally tested in SFE/HWT in 2019, 2020 and 2021, and some updates introduced in this dissertation were implemented into the system and further tested during HWT SFE 2020 and 2021. The performance and comparisons with the WoF ensemble-based system were reasonable and promising (Gao et al. 2021, 2022; Carpenter 2022).

2.3 Overview of the Warn-on-Forecast ensemble DA and Forecast System (WoFS)

In this dissertation, the ensemble background error covariance matrix **P** employed in the WoF-Hybrid are calculated from the WoFS ensemble forecasts. The WoFS is a high-resolution, on-demand ensemble DA and forecast system. It consists of a 36-member ensemble with diverse physical parameterizations (Wheatley et al. 2015) and is usually run over a 300×300 km^2 domain with a 3-km horizontal grid spacing. The Analyses are initialized at 1700 UTC daily before 2019 and 1500 UTC after 2019 with initial and boundary conditions provided by the High-Resolution Rapid Refresh Ensemble (HRRRE; Kalina et al. 2021). The location of the domain is typically determined by the Storm Prediction Center (SPC) Day 1 outlook (Kain et al. 2003; Gallo et al. 2017). Once initial conditions are prepared, analyses are produced at a 15-minute interval with the assimilation of conventional observations, WSR-88D radar reflectivity and radial velocity data (Wheatley et al. 2015), satellite-derived cloud water path (Minnis et al. 2011; Jones et al. 2016), and clear-sky radiance (Jones et al. 2020) using the GSI-EnKF, one of the EnKF systems which is widely used for convective-scale DA in both the operational and research communities (Johnson et al. 2015; Wang and Wang 2017; Jones et al. 2018; Yussouf and Knopfmeier 2019). The conventional observations have a varied localization length from 60 km for the Mesonet to 460 km for the sparser resolution instruments (e.g., the Automated Surface Observing System and the Aircraft Communications, Addressing, and Reporting System). The radar observations and satellite observations use 18 km and 36 km horizontal radii, respectively. Localization values applied here were derived from those used by Jones et al. (2020). Some very encouraging results have been obtained using the WoFS for convective-scale radar and satellite DA in short-term forecasts in real-time operations(Guerra et al. 2022).

2.4 Overview of Radar Observation Forward Operators and GLM derived Water Vapor Used in the Early Version of WoF-Hybrid

2.4.1 Radar and conventional observation

Radial velocity and reflectivity observations are synthesized by using level II data from the NEXRAD WSR-88D radars within and surrounding the selected domain for each event, and then are interpolated onto the model grids. Radar quality control procedures (such as buddy checking, dealiasing, and anomalous returns removal) are applied before they are assimilated (Gao et al. 2013). The forward operator for radial velocity is same as that used in Gao and Stensrud (2012):

$$V_r = u \sin\phi \cos\mu + v \cos\phi \cos\mu + w \sin\mu, \tag{2.8}$$

where (u, v, w) are the wind components from the model forecast on staggered grids, μ is the elevation angle, and ϕ is the azimuthal angle of radar beam. Because we omit precipitation terminal velocity from the observed radial velocity prior to the analysis, only the preconditioned control variables (u, v, w) for wind are required to calculate model-derived radial velocity. For radar reflectivity, the forward operator is also same as that used in Gao and Stensrud (2012):

$$Z_e = f(x) = \begin{cases} Z(q_r),\ T_b > 5℃ \\ Z(q_s) + Z(q_g) + Z(q_h),\ T_b < -5℃ \\ \alpha Z(q_r) + (1-\alpha)[Z(q_s) + Z(q_g) + Z(q_h)],\ -5℃ < T_b < 5℃ \end{cases}, \quad (2.9)$$

$$Z_{dB} = 10\log_{10} Z_e, \quad (2.10)$$

where α varies linearly between 0 when temperature in background $T_b = -5℃$ and 1 when $T_b = 5℃$. By applying Eq. 2.9, it allows the system to split the hydrometeor variables and correct the weight of rain water content relative to the snow, graupel and hail content contributed to the simulated reflectivity. Conventional observations, such as wind, pressure, temperature and specific humidity may be collected from NCEP Meteorological Assimilation Data Ingest System (MADIS), PREPBUFR, or Oklahoma Mesonet (Brock et al. 1995; McPherson et al. 2007).

2.4.2 Geostationary Lightning Mapper (GLM) derived water vapor

The method of utilizing GLM derived water vapor in the WoF-Var system has been described in Fierro et al. (2019). The raw 20-s GLM data is first accumulated into a 1-hour interval and converted into gridded flash density rates. After that, the flash density observations are projected onto the NWP model grid. Pseudo observations are created for water vapor content at each grid point where the gridded flash density rate exceeds zero by adjusting the background water vapor content within a fixed depth of 3 km above the lifted condensation level. The pseudo water vapor observation is adjusted to a near-saturation value, i.e., 95% of saturated water vapor content for a given pressure and temperature at the specific grid point, if the background water vapor content does not already exceed that value. Once created, these pseudo-observations are assimilated through the WoF-Var system, which uses observation and background errors to adjust the background fields. The observation error for water vapor content is set to 3×10^{-3} kg kg^{-1} and the background error for water vapor content is set to 10×10^{-3} kg kg^{-1}. The assimilation of the pseudo-observations uses the same decorrelation length scales as the assimilation of the radar reflectivity.

Chapter 3 The Impact of Assimilating Satellite-derived Layered Precipitable Water, Cloud Water Path and Radar Data on Short-Range Thunderstorm Forecasts

3.1 Introduction

With the launch of GOES-R series, assimilating satellite observations, or derived products to improve convective-scale NWP has become more attractive. Some studies suggested that additional satellite observations can partially compensate for the limitation of assimilating only radar observations (Jones et al. 2013, 2016, 2018, 2020; Fierro et al. 2016, 2019; Zhang et al. 2019b; Hu et al. 2020). One of the applications of assimilating satellite-derived products in convective-scale NWP has been studied by Jones et al. (2013). In their study, both radar observations and satellite product CWP were assimilated into the three-dimensional compressible nonhydrostatic WRF-ARW for predicting a severe weather event that occurred on 10 May 2010. Their experiments showed that the assimilation of CWP resulted in improvements in cloud properties and more accurate shortwave radiation flux at the surface by comparing to the experiments without CWP assimilation. Although this positive impact decreases with increasing lead time, the benefits were still maintained 90 min into the model simulation.

Pan et al. (2018) explored the benefit of assimilating another satellite product, total precipitable water, in addition to CWP and radar data, in an idealized case study using a three-dimensional ensemble-variational system (3DEnVAR; Gao et al. 2016). In this idealized case study, the forecasts that assimilated total precipitable water in addition to radar data were more accurate than those that assimilated only radar observations. Wang et al. (2018) also found that the assimilation of the three-layer precipitable water product derived from the Advanced Himawari Imager in regional NWP improved heavy precipitation forecasts, especially when the assimilation was combined with appropriate cumulus and microphysical schemes (Lu et al. 2019). Based on these findings, we believe that the assimilation of satellite derived water related products has the potential to improve convective-scale NWP by providing more accurate model initial conditions especially for moisture and thermodynamic variables.

The LPW is one of the products of the newly launched GOES-16/17 satellites, and its usefulness to convective-scale NWP has not been extensively studied to our knowledge. So, in this chapter, we evaluate the added value of assimilating LPW and CWP products on short-term convective-scale NWP using the WoF-Hybrid system. Three high-impact weather events occurring in May 2017 are assessed (Table 3.1) by comparing analyses and short-term forecasts from three sensitivity experiments. The first experiment does not assimilate any observations. The second experiment uses only radar observations from WSR-88D network and surface observations. The last group of experiments use satellite derived CWP and LPW products together with radar and surface observations. The added value for the assimilation of additional satellite observations can be easily assessed by comparing the above experiments. These three events represent different weather regimes, allowing to test the applicability and effectiveness of the satellite-derived products in different weather scenarios.

Table 3.1 List of three high-impact weather events for assessing the impact of assimilating GOES-13/16 retrieval observations in convective-scale NWP. 3- to 6-hour forecasts are launched at 1-hour intervals from 2000 to 2300 UTC.

Event	Region	Period
10 May 2017	Northern TX/Southern OK	2000 – 0200 UTC
16 May 2017	TX Panhandle/Southwestern OK	2000 – 0100 UTC
26 May 2017	Eastern CO	2000 – 0100 UTC

The remainder of this chapter is organized as follows. Section 3.2 describes newly implemented CWP and LPW forward operators in the WoF-Hybrid system. The model configuration and experimental design are briefly introduced in section 3.3. Section 3.4 presents a synthesis of the current work followed by concluding remarks in the final section.

3.2 CWP and LPW Forward Operators

3.2.1 CWP from GOES-13 Imager measurements

The GOES-M series imager allows sampling of visible and infrared radiation over the continental United States (CONUS) every 5 – 15 minutes on-demand basis. The CWP is retrieved from GOES-M imager radiances with a 4-km resolution for cloudy pixels using the Visible Infrared Solar-Infrared Split-Window Technique (Minnis et al. 2011). The impact of assimilating these products in an ensemble, convection-allowing NWP system using the ensemble square root filter (EnSRF, Whitaker and Hamill 2002) assimilation method are described in detail by Jones et al. (2013, 2015, 2016). Parallax correction is performed before the assimilation process so that the locations of clouds are adjusted to their ground-relative coordinates (Jones et al. 2018).

The forward operator for satellite CWP retrievals is defined by Jones et al. (2015).The model equivalent CWP is calculated by integrating total cloud water mixing ratio over the atmospheric column from observed cloud base to cloud top:

$$CWP = -\frac{1}{g}\int_{CBP}^{CTP} \sum_{x=1}^{n} q_x(p)\, dp. \tag{2.11}$$

As shown in Eq. 2.11, CBP represents the pressure of the cloud base which is determined by the optical depth and cloud top properties, CTP represents the pressure of the cloud top, p is the atmospheric pressure, q_x is hydrometeor mixing ratio for a specific species (x), and n indicates the number of hydrometeor species based on the microphysics scheme and observed cloud phase. For example, to create a total cloud water mixing ratio $\sum_{x=1}^{n} q_x(p)$ at a specific model grid point, the mass mixing ratios of the five hydrometeor variables $(q_c, q_r, q_i, q_s, q_g)$ are used when the Thompson 1.5-moment microphysics scheme (Thompson et al. 2008) is adopted, whereas the mass mixing ratios of the six hydrometeor variables $(q_c, q_r, q_i, q_s, q_g, q_h)$ are used when the NSSL 2-

moment with CCN prediction microphysics scheme (Mansell et al. 2010) is adopted. As described in Jones et al. (2015), cloud phase at each pixel can be classified as either "liquid" or "ice" based on the cloud temperature and cloud effective particle size information. In the forward operator, CWP represents the column integration of liquid cloud and rain hydrometeors for observed liquid cloud pixels. An "ice" pixel represents the column integration of both liquid and frozen hydrometeors (such as snow, ice, graupel and hail) because the current algorithm for retrieving CWP is unable to separately identify mixed-phase clouds such as deep convective clouds that may contain both liquid and ice phase hydrometeors. The CWP product is generally used to discern cloudy and clear-sky areas in the domain. To partially address the production of spurious clouds and convection in the model domain, our DA procedure also considers the assimilation of zero CWP observations.

3.2.2 LPW derived from ABI

The advanced Baseline Imager (ABI) is designed for the GOES-R series that has been operational since late 2016. Differing from the GOES Imager carried on the GOES-M series, ABI implements more spectral bands as well as higher temporal and spatial sampling capabilities (Schmit et al. 2005, 2017). A new retrieval product, LPW, is experimentally available by taking advantage of the extra spectral bands. The retrieval algorithm was developed for deriving legacy atmospheric temperature and moisture profiles from radiance measurements from ABI. It uses the general least squares regression as a first guess followed by a one-dimensional variational approach and retrieves atmospheric temperature and moisture profiles and the derived products, including TPW and LPW from the clear sky infrared radiances (Schmit et al. 2019). The LPW retrieval algorithm mostly depends on the 6.2 μm, 6.9 μm and 7.3 μm infrared bands, which represent the water vapor distribution in upper-, mid- and lower mid-troposphere, respectively. It also requires other infrared longwave bands such as 10.3 μm, 11.2 μm, 12.3 μm, and 13.3 μm bands for cloud clearing purposes. The LPW product is currently only available over clear-sky areas.

The LPW (in cm) accumulates water vapor contained in a vertical column over a unit cross-section area between two specific levels. The forward operator used in our hybrid DA system can be written as:

$$LPW = -\frac{100}{\rho_w g}\int_{low-level}^{high-level} q_v(p)\, dp, \tag{2.12}$$

where ρ_w is the water density, and q_v is the water vapor mixing ratio. The troposphere is classified into three different PW layers: Boundary Layer (BL, Sigma Level: 1.0 – 0.9), Middle Layer (ML, Sigma Level: 0.9 – 0.7), and High Layer (HL, Sigma Level: 0.7 – 0.3). Therefore, the atmospheric pressure at a particular sigma level needs to be determined for the LPW calculation based on the following equation:

$$P_{sig} = P_{top} + sig\_index \times \left(P_{sfc} - P_{top}\right), \tag{2.13}$$

where P_{top} represents the model top pressure, P_{sfc} represents the surface pressure and sig_index represents the sigma level in the model. Once the low-level and high-level pressures for a particular layer have been determined by sigma level, then Eq. 2.12 is applied to calculate LPW within this layer. Currently, only the BL and ML of the LPW product are utilized since some sensitivity experiments indicate that large forecast biases exist within several 2017 cases when HL precipitable water is involved.

3.3 Model Configuration and Experimental Design

3.3.1 Model and DA configuration

Due to the resolution used in this study not being consistent with either the real-time WoFS resolution or an integer multiple thereof, a completely new ensemble of forecasts was generated in accordance with the real-time WoFS process but with slight modification in order to provide flow-dependent error covariances for the WoF-Hybrid. The storm events are modeled using the WRF-ARW, and then a 36-member ensemble of storm-scale forecasts is created. The initial and boundary conditions are initialized at 1200 UTC using the 1 degree 6-hour global ensemble forecast system (GEFS, Zhou et al. 2017) forecasts, and then 6-hour forecasts are launched to properly spin-up fine-scale structures in the DA system. The simulations in our DA experiments are carried out on a 2-km spaced grid of 480 × 360 horizontal points resulting in a 960 km × 720 km^2 domain, and 51 stretched vertical levels reaching up to 50-hPa with a 56m – 460m thickness between individual levels. The center of the domain for the relevant event is selected based on the storm reports provided by the National Weather Service (NWS) / SPC to ensure that the principal high-impact events, including tornadoes, large hail and high winds are within the computational domain. The initial and boundary conditions are provided by the first 18 members of the 21-member GEFS. To avoid under-dispersion between members, variations in boundary layer physics and radiation schemes are applied to create a set of ensemble member with unique combination of microphysics scheme and initial / boundary condition. This approach is consistent with previous studies Stensrud et al. (2000), Fujita et al. (2007), and Wheatley et al. (2014).

The combinations of Planetary Boundary Layer (PBL) and radiation schemes for the ensemble members are shown in Table 3.2, similar to those described in Wheatley et al. (2015) and Skinner et al. (2018). To generate the remaining 18 members, the parameterization scheme combinations are applied to the reverse order of downscaled GEFS members. For instance, the parameterization combination for member 1 is used to downscale GEFS member 18 to generate member 19. All members of the ensemble adopt the Thompson cloud microphysics scheme (Thompson et al. 2008) in this study. Because the simulations are performed on a 2-km convection-allowing grid, no cumulus parameterization scheme is used. While the multiplicative covariance inflation method is not applied to the system (Jones et al. 2018), the additive noise technique described in (Dowell and Wicker 2009) is applied to maintain the spread between ensemble members. For grid points where greater than 25 dBZ reflectivity is observed and the absolute innovation of the posterior ensemble mean reflectivity is less than 10 dBZ, local perturbations are added to temperature with standard deviation of 0.5 K, dewpoint temperature with standard deviation of 0.5 K, and wind with standard deviation of 0.5 m s^{-1}. Zero reflectivities are assimilated to suppress spurious precipitation.

In this preliminary investigation, the deterministic and ensemble components are one-way coupled, as illustrated in Figure 3.1, with a consistent spatial resolution. The aim of this setup is to evaluate the skill of the analyses and forecasts of the deterministic component. Initial and boundary conditions for the deterministic component are obtained by launching a 6-hour WRF forecast from a downscaled Global Forecast System (GFS) analysis valid at 1200 UTC. The domain used in the deterministic component exactly matches that used in the ensemble component. To maintain consistency with the High-Resolution Rapid Refresh configuration (HRRR; Dowell et al. 2022), the deterministic component uses the RRTMG scheme for both shortwave and longwave radiation and the MYJ scheme for the PBL scheme during the assimilation cycling as well as the free forecast period.

Table 3.2 PBL and radiation scheme combinations for 36 ensemble members of each 2017

event. Deterministic member indicates the extra member used for deterministic analysis and forecast. The PBL schemes include MYJ (Janjić 1994), YSU (Hong et al. 2006), and MYNN2 (Nakanishi and Niino 2004). The radiation schemes include Dudhia (Dudhia 1989), RRTM (Mlawer et al. 1997), RRTMG (Iacono et al. 2008), and Goddard (Chou et al. 1999).

Member	GEFS Member	PBL	Shortwave Radiation	Longwave Radiation
Deterministic	GFS	MYJ	RRTMG	RRTMG
1	1	YSU	Dudhia	RRTM
2	2	YSU	Goddard	Goddard
3	3	MYJ	RRTMG	RRTMG
4	4	MYJ	Dudhia	RRTM
5	5	MYNN2	Goddard	Goddard
6	6	MYNN2	RRTMG	RRTMG
7	7	YSU	Dudhia	RRTM
8	8	YSU	Goddard	Goddard
9	9	MYJ	RRTMG	RRTMG
10	10	MYJ	Dudhia	RRTM
11	11	MYNN2	Goddard	Goddard
12	12	MYNN2	RRTMG	RRTMG
13	13	YSU	Dudhia	RRTM
14	14	YSU	Goddard	Goddard
15	15	MYJ	RRTMG	RRTMG
16	16	MYJ	Dudhia	RRTM
17	17	MYNN2	Goddard	Goddard
18	18	MYNN2	RRTMG	RRTMG
19	1	MYNN2	RRTMG	RRTMG
20	2	MYNN2	Goddard	Goddard
21	3	MYJ	Dudhia	RRTM
22	4	MYJ	RRTMG	RRTMG
23	5	YSU	Goddard	Goddard
24	6	YSU	Dudhia	RRTM
25	7	MYNN2	RRTMG	RRTMG
26	8	MYNN2	Goddard	Goddard
27	9	MYJ	Dudhia	RRTM
28	10	MYJ	RRTMG	RRTMG
29	11	YSU	Goddard	Goddard
30	12	YSU	Dudhia	RRTM
31	13	MYNN2	RRTMG	RRTMG
32	14	MYNN2	Goddard	Goddard
33	15	MYJ	Dudhia	RRTM
34	16	MYJ	RRTMG	RRTMG
35	17	YSU	Goddard	Goddard
36	18	YSU	Dudhia	RRTM

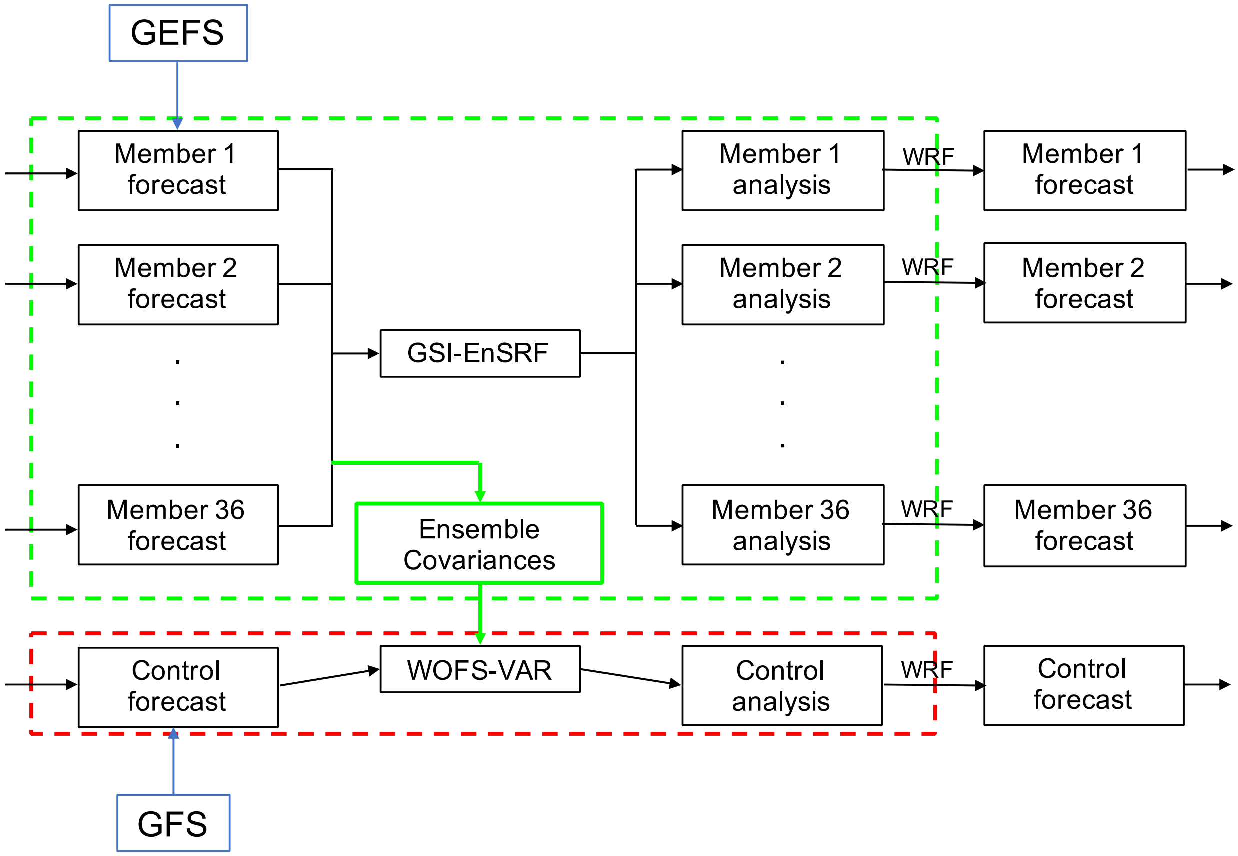

Figure 3.1 Schematic diagram of a single hybrid 3DEnVAR data assimilation cycle (Modified based on Fig. 1(a) of Wang et al. 2013).

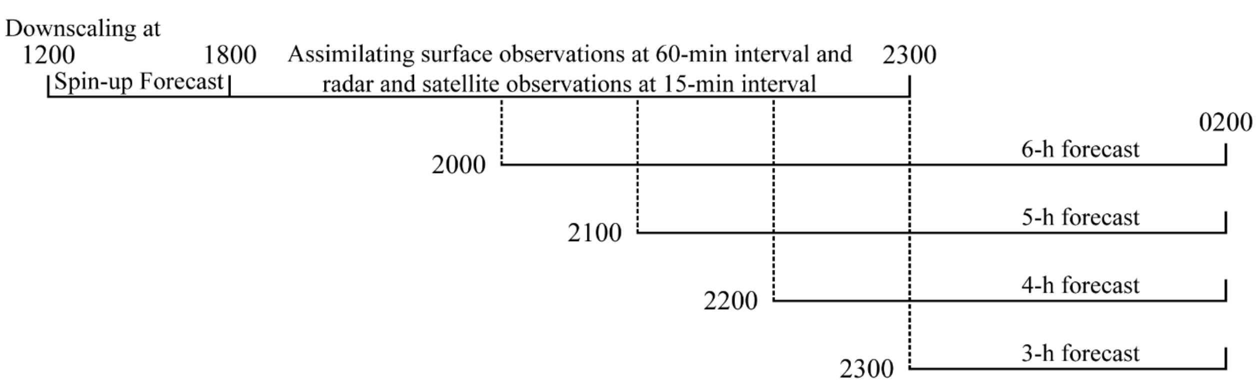

Figure 3.2 The flowchart representing assimilation cycles for experiments used by all three high-impact weather events listed in Table 3.1. Times used in this figure are in UTC.

Figure 3.2 illustrates the analysis and forecast cycling set up used in this study. Observations are continuously assimilated from 1800 UTC to 2300 UTC at 15-min intervals for all three events. The analysis variables for driving the WRF model include wind fields (U, V, W), temperature (T), water vapor mixing ratio (QVAPOR), geopotential height (PH), dry air mass (MU), hydrometeor mass (cloud water, rain, ice, snow, graupel) which are updated at each cycle in both the EnKF ensemble and hybrid 3DEnVAR deterministic systems. 3- to 6-hour forecasts are launched every hour beginning at 2000 UTC to provide performance of each experiment during the period of relevant severe weather events. All free forecasts are ended at 0200 UTC on the next day.

3.3.2 Experiment Design

Three types of experiments are performed for each event. The control experiment (CTRL) runs without assimilation of any observations so that it can provide a baseline for comparison between experiments. The second experiment, labeled radar experiment (RAD), assimilates radial velocity, reflectivity, and surface observations. By comparing the above two experiments, the impact of radar and surface observations in convection-allowing NWP can be highlighted. Considering that CWP and LPW observations generally cover different areas (i.e., cloudy area and clear-sky area) and can compensate for one another's shortage in coverage (as shown in Fig. 3.3), the third experiment assimilates both LPW from GOES-16 and CWP from GOES-13 in addition to surface and radar observations instead of separately combining CWP or LPW with the radar data, labeled as the radar + satellite experiment (RADSAT). Two additional experiments, one similar to the third experiment but without assimilating LPW (named as RADCWP), and another without assimilating CWP (named as RADLPW) will also be discussed occasionally for the comparison purpose if necessary. Observation errors and recursive filter radii for each type of observation used in the variational system are listed in Table 3.3. Analyses and forecasts are compared between three types of experiments for each event to investigate whether or not assimilating CWP and/or LPW products has positive impact on analyses and subsequent forecasts.

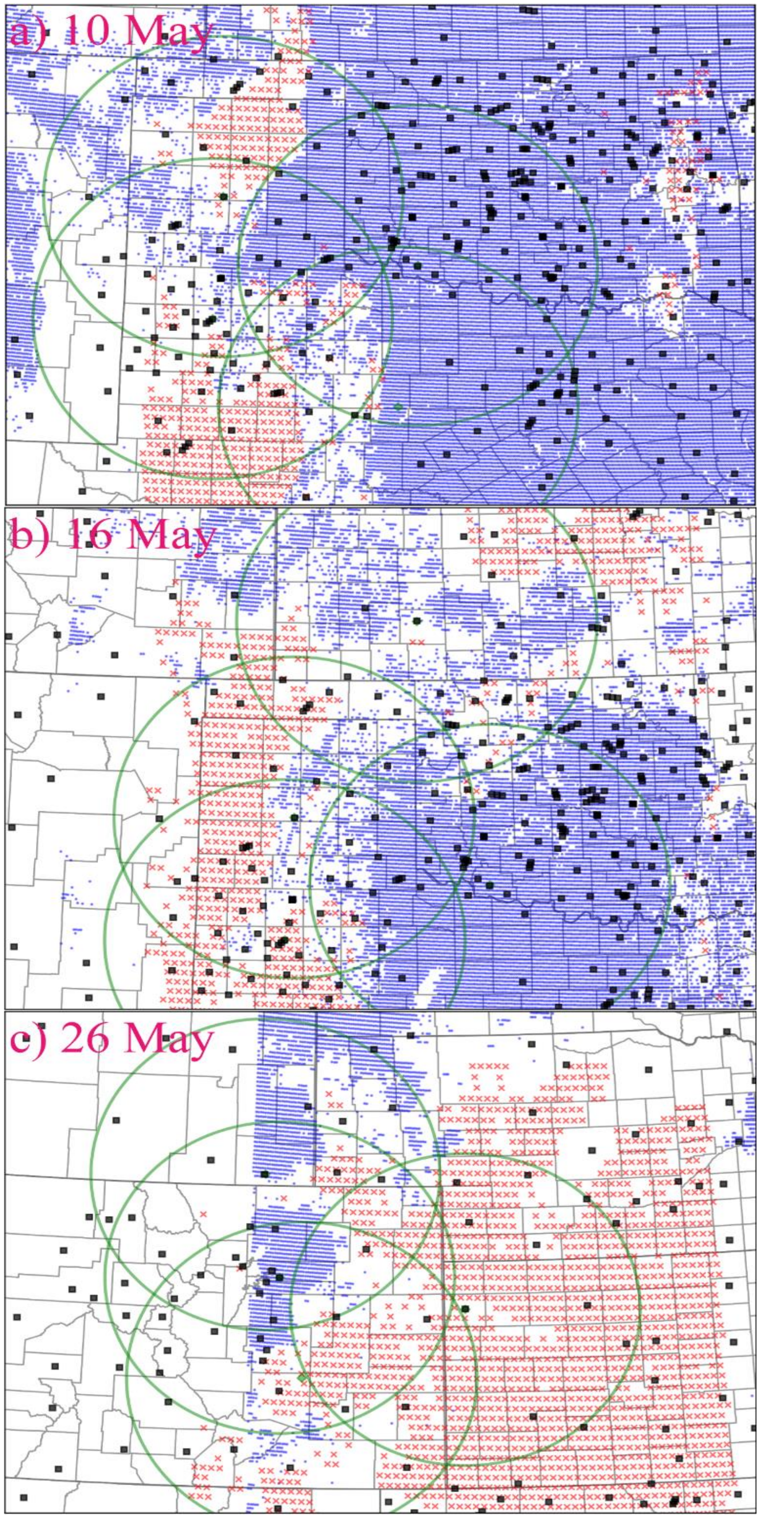

Figure 3.3 Locations of the observations assimilated into the hybrid 3DEnVAR system at 1800 UTC for (a) 10 May, (b) 16 May and (c) 26 May 2017 events. Black squares represent surface observations. Green circles represent the coverage of radars. Blue dots and red crosses are the CWP observations in cloudy area and LPW observations, respectively.

Table 3.3 Observation errors and recursive filter radius used in the variational component of the hybrid 3DEnVar system. The vertical radius of the recursive filter is given in grid points.

Observation type	Error	Horizontal radius (km)	Vertical radius (grid points)
Surface U-wind	1.5 ms^{-1}	60	4
Surface V-wind	1.5 ms^{-1}		
Surface pressure	2.0 hPa		
Surface temperature	1.5 K		
Surface specific humidity	0.5 g kg^{-1}		
Low-level LPW	0.15 cm		
Mid-level LPW	0.15 cm		
Radial Velocity	2.0 ms^{-1}	18	
Reflectivity	10.0 dBZ		
CWP	0.4 kg m^{-2}		

3.4 Analysis and Forecast Results

Since 10 May, 16 May, and 26 May 2017 each represent three different synoptic system types, analyses at some particular time when severe weather events were active on these days and their affiliated 3- to 6-hour forecasts are selected for the study. Only the results of the deterministic component of the three experiments are discussed. The subjective diagnostic analyses are provided for each case followed by the qualitative forecast evaluation including comparing the updraft helicity (UH) tracks against SPC storm reports over free forecast periods. To provide a quantitative measure of forecast skill, the frequency bias (BIAS) and fractions skill score (FSS; Roberts and Lean 2008; Roberts 2008) for composite reflectivity are computed relative to observational data from the Multi-Radar/Multi-Sensor system (MRMS; Smith et al. 2016). Additionally, neighborhood equitable threat scores (ETS; Clark et al. 2010) aggregated over all cases are also calculated relative to NCEP's Stage IV hourly precipitation product (Du 2011). Finally, information about the statistical metrics and moments of the hybrid 3DEnVAR are given.

3.4.1 Analysis diagnostics

3.4.1.1 10 May 2017 event

On 10 May 2017, the pre-storm environment over Oklahoma and Texas was becoming unstable starting from about 1700 UTC. Near the Oklahoma/Texas border, multiple severe-warned storms, including supercells, developed and affected the area throughout the day (See markers in Fig. 3.13). Figure 3.4 shows the surface temperature at 2000 UTC 10 May 2017, after two hours of DA cycling for each experiment. The temperature gradient along the Oklahoma/Texas border is not sharp or well-defined for all experiments (Figs. 3.4 a-e). A slight decrease in temperature (~ 5 F) is found near the border of Oklahoma and Kansas in RAD which is introduced due to the development of storm cells in the model (black box, Fig. 3.4 b). In contrast, with the assimilation of satellite-derived observations (Fig. 3.4 c vs Figs. 3.4 a, b), RADSAT generates a large area of cooled surface temperature (~ 65 F) extending from the eastern part of the Texas panhandle through northwestern Oklahoma to southern Kansas (ellipse area in Fig. 3.4 c). This pattern agrees with the Mesonet observations (ellipse area ~ 65 F in Fig. 3.5 a), which provides a source of independent verification for the analysis. The differences are the consequence of assimilating CWP and LPW observations which modified the cloud properties in either direct or indirect way. This interpretation is also supported by the standalone RADCWP and RADLPW experiments (Figs. 3.4 d, e).

The downward shortwave flux (SWDOWN in WRF, Figs. 3.4 f - j) represents the amount of downward shortwave radiation reaching the earth’s surface. Higher values of shortwave flux are associated with clear-sky regions and high surface temperatures, while lower values of shortwave flux are associated with the existence of clouds and cooler temperatures. In RAD, the presence of isolated and thick clouds associated with the southern Kansas MCSs is evidenced by low downward shortwave flux values at the same location where the surface temperature is reduced (the green shade in the black box of Fig. 3.4 b, and the blue shade in the black box of Fig. 3.4 g). In the RADSAT experiment, most of Oklahoma and northeastern Texas is associated with low surface temperature and SWDOWN (Figs. 3.4 c, h). This implies extensive cloud cover being present which is evident on the standalone RADCWP/RADLPW experiments (Figs. 3.4 i, j), and the corresponding GOES-16 visible image (Fig. 3.5 b). The satellite observations and shortwave flux calculated from the model variables corroborate each other and clearly depict the relationship between the cloud characteristics, downward shortwave radiation and surface temperature. Clouds prevent incoming shortwave solar energy from passing through the atmosphere by reflecting and absorbing shortwave radiation during the daytime, and hence resist surface heating and lead to a cooler surface. These changes in temperature may further stabilize the lower troposphere and restrain the development of spurious convection in later forecast times.

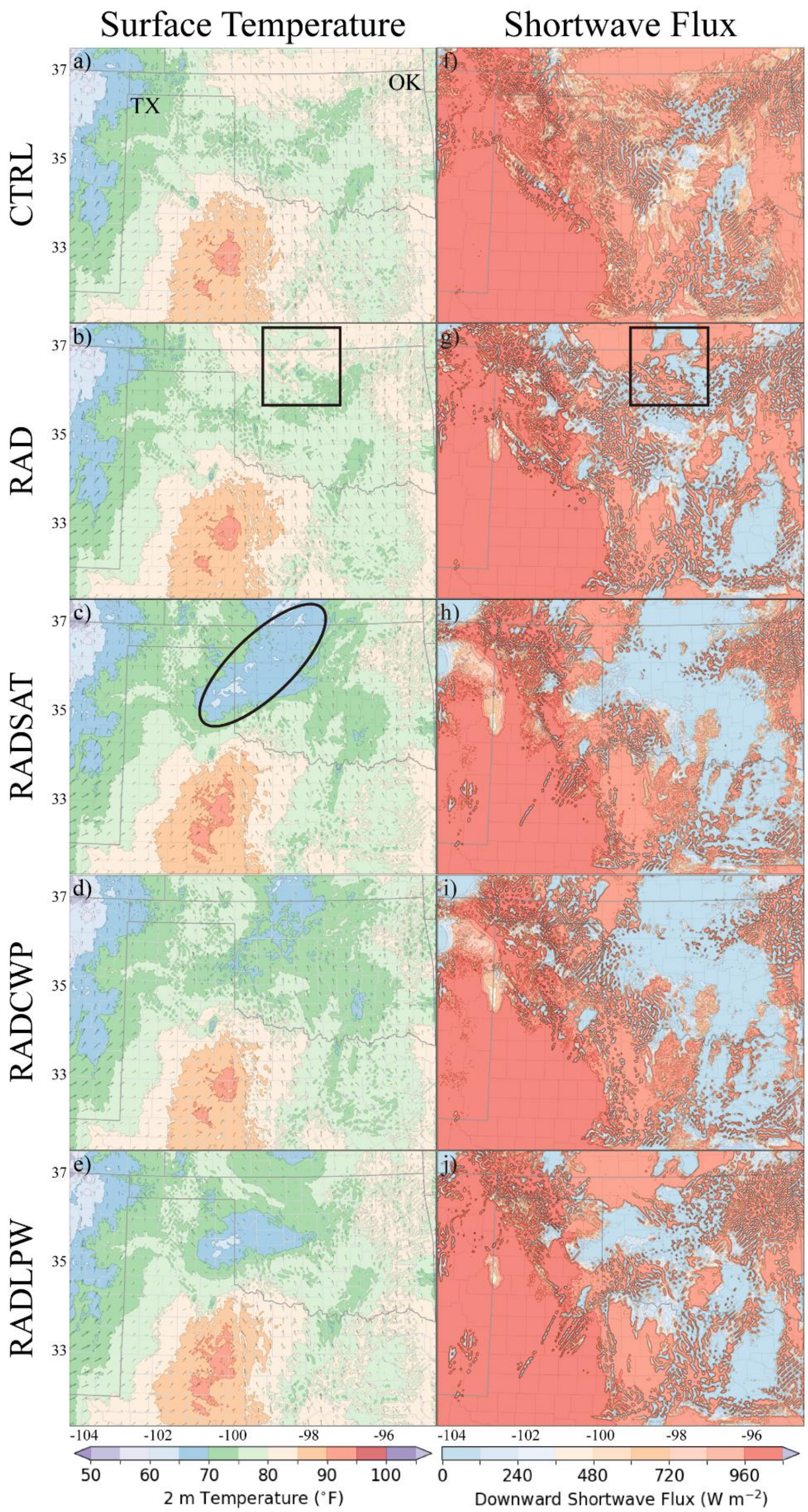

Figure 3.4 Analyses of (a) – (e) surface temperature at 2 meters above the ground, and (f) – (j) downward shortwave flux at surface for CTRL, RAD, RADSAT, RADCWP and RADLPW experiment at 2000 UTC 10 May 2017.

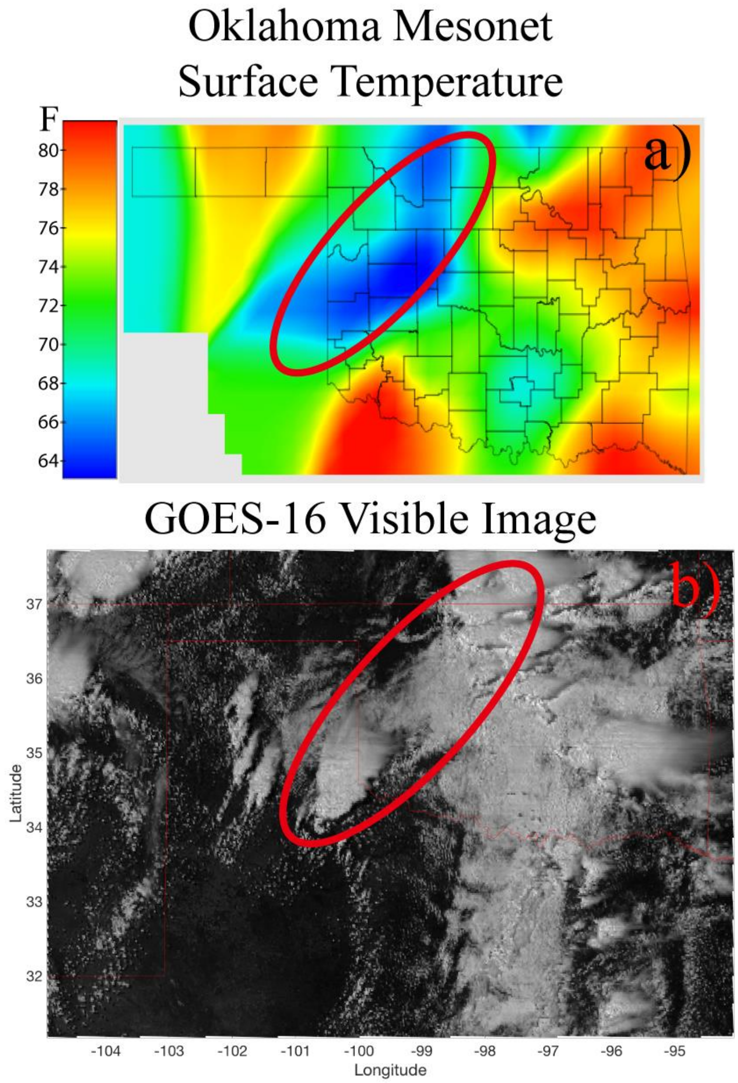

Figure 3.5 (a) Mesonet 1.5m surface temperature and (b) visible image from GOES-16 at 2000 UTC 10 May 2017.

Water vapor distribution is another critical factor that may affect convection initiation and development. Since the Eq. 2.12 indicates a linear correlation between the water vapor mixing ratio and simulated model equivalent LPW, the impact of assimilating LPW on water vapor content can be straightforwardly assessed. Differences in water vapor are not obvious between RAD and CTRL at 2000 UTC in most areas (Fig. 3.6 a). The largest difference > 2 g kg^{-1} appears near Motley, Texas. The difference is likely due to the ongoing convection. For RADSAT, the difference of water vapor mixing ratio against CTRL reaches as high as 3.0 g kg^{-1} south to the border between Oklahoma and the Texas panhandle (black box, Fig. 3.6 b). A large area with > 1.5 g kg^{-1} difference is observed over western Texas (southwest of the model domain) in RADSAT as well. This difference is obviously introduced by assimilating LPW product when we compare RAD against the standalone RADCWP/RADLPW experiments (Fig. 3.6 a vs Figs. 3.6 d, e). The water vapor content in RADSAT increases more than 30% compared to that of RAD (Fig. 3.6 c, RAD has 3 – 10 g kg^{-1} water vapor content over southern Texas panhandle [not shown]), which may lead to a more favorable moisture content for initiating deep moist convection when accompanied with a strong lifting process. In addition, an area with positive differences in water vapor is observed over southwestern Oklahoma when comparing RADSAT and CTRL/RAD (blue boxes in Figs. 3.6 b, c). In summary, a more reasonable water vapor mixing ratio field is able to either benefit the suppression of spurious convection or to promote new cell development for this case.

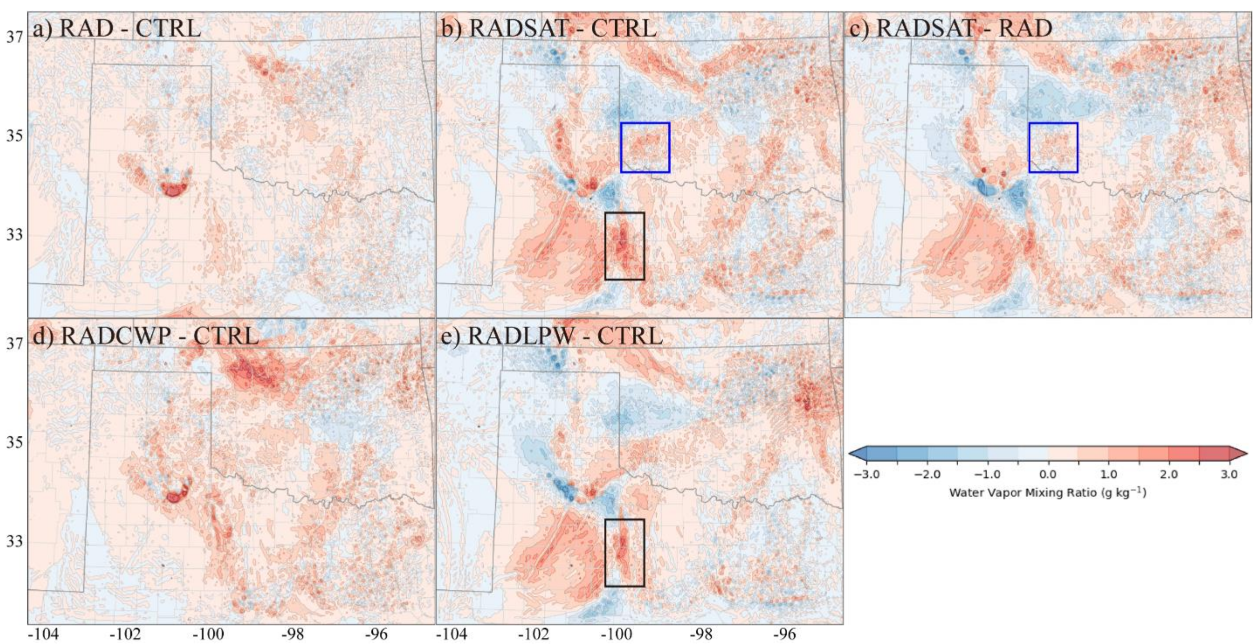

Figure 3.6 The differences of analyzed surface water vapor mixing ratio between (a) RAD and CTRL, between (b) RADSAT and CTRL, between (c) RADSAT and RAD, between (d) RADCWP and CTRL, and between (e) RADLPW and CTRL at 2000 UTC 10 May 2017. The black box highlights one of the differences of water vapor mixing ratio between

RADSAT/RADLPW and CTRL. The blue box highlights one of differences of water vapor mixing ratio between RADSAT and CTRL/RAD.

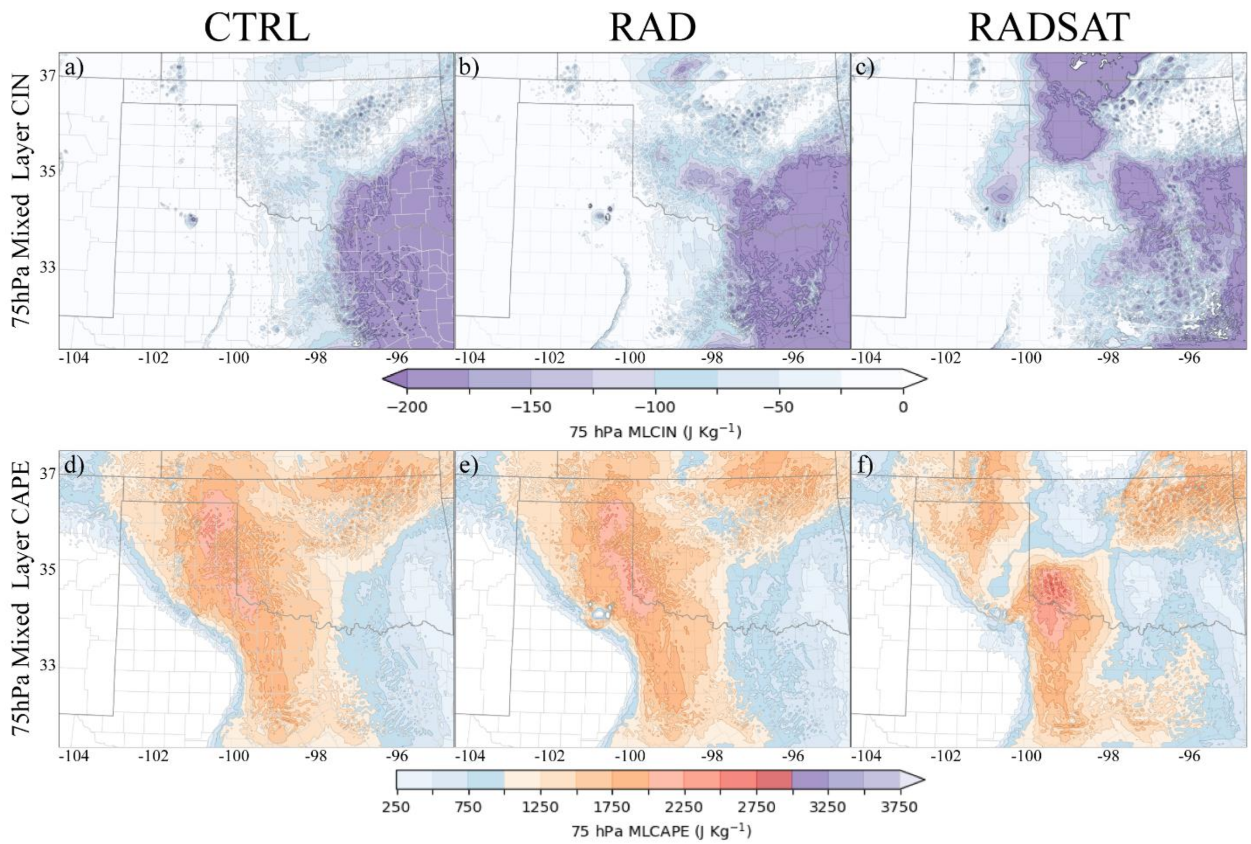

Figure 3.7 (a) – (c) The lowest 75 hPa mixed layer convective inhibition (MLCIN) and (d) – (f) the lowest 75 hPa mixed layer convective available potential energy (MLCAPE) for CTRL, RAD and RADSAT experiments at 2000 UTC 10 May 2017.

Comprehensive changes of the model's thermodynamic conditions can be evaluated by the mixed layer convective inhibition (MLCIN) and mixed layer convective available potential energy (MLCAPE) (Fig. 3.7). Essentially, high MLCIN occurs in areas of cool surface temperatures caused by clouds and precipitation. Both CTRL and RAD produce a high MLCIN over southeastern Oklahoma and eastern Texas. Differences in MLCIN between CTRL and RAD (Figs. 3.7 a, b) are evident over southern Kansas. RADSAT also generates high MLCIN over southeastern Oklahoma and eastern Texas. But apparently, it also generates high MLCIN in southern Kansas, northwestern Oklahoma and eastern Texas panhandle (Fig. 3.7 c), and low

MLCAPE near the same areas (Fig. 3.7 f). The cloud cover with CWP > 1.5 kg m^{-2} (ellipse area in Fig. 3.5 b and corresponding area in Fig. 3.4 h) allows stabilization of the low-level atmosphere and MLCIN < -200 J kg^{-1} for the same reason. Compared to the CTRL experiment, RAD does not significantly change the MLCAPE at 2000 UTC (Fig. 3.7 d, e). In contrast, the MLCAPE over the clear-sky area in southwestern Oklahoma increases to around 2750 J kg^{-1} in RADSAT. This increase corresponds well with the difference of water vapor mixing ratio at surface (Figs. 3.6 b, c). Consequently, a supercell which produced several tornadoes developed at a later time.

3.4.1.2 16 May 2017 event

The 16 May 2017 event was characterized by the presence of a dryline that extended southward from western Kansas down to the US-Mexico border at 1800 UTC. Storms formed at the intersection of the dryline and weak frontal boundary zonally crossing central Kansas (https://www.wpc.ncep.noaa.gov/archives/web_pages/sfc/sfc_archive_maps.php?arcdate=05/16/2017&selmap=2017051621&maptype=namussfc). From 1800 to 2100 UTC, the dryline remained stationary and convection initiated at several locations ahead of the dryline. Similar to the 10 May 2017 event, the differences between three experiments for surface water vapor mixing ratio are analyzed (2100 UTC). A narrow band of negative differences exceeding 3 g kg^{-1} is observed in the middle of the Texas Panhandle in the moisture difference field between RAD and CTRL (Fig. 3.8 a). The difference between RADSAT and CTRL is quite similar to that between RAD and CTRL, but positive differences > 3 g kg^{-1} are also present over the southern Texas panhandle (black box, Fig. 3.8 b). Further south, the difference becomes negative between RADSAT and CTRL, and between RADSAT and RAD (Figs. 3.8 b, c). Moreover, RADSAT generates a positive difference in surface moisture ahead of the dry line along the border of the eastern Texas panhandle and western Oklahoma. These changes promote a bulge of dry air towards the moist air to the east, intensifying the moisture gradient over this area in RADSAT.

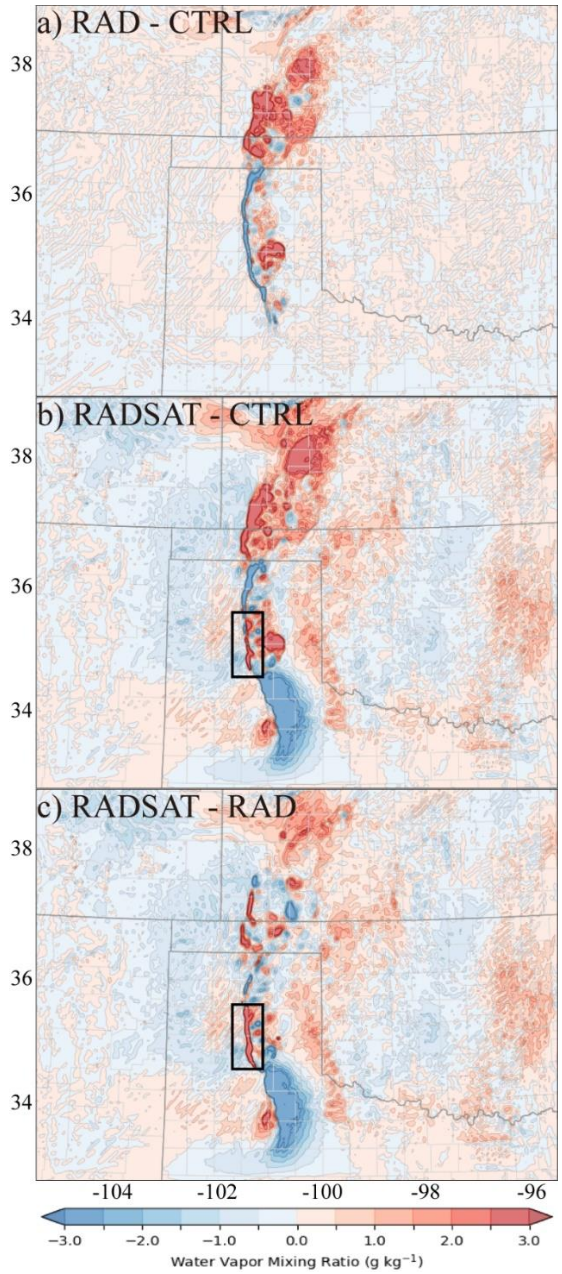

Figure 3.8 As in Fig. 3.6, but for 2100 UTC 16 May 2017. The black box indicates the positive difference between RADSAT and CTRL/RAD.

3.4.1.3 26 May 2017 event

The third analyzed event occurred on 26 May 2017 in Colorado. Based on the sounding (KDNR, not shown) at 0000 UTC 27 May 2017, this event is characterized by low an inertial instability and high vertical wind shear environment. Figure 3.9 exhibits differences of surface water vapor mixing ratio in the analyses between experiments. RAD produces more water vapor (> 0 g kg^{-1}) over southeast Wyoming and in northeastern Colorado, which is near the tornadic supercell (black box, Fig. 3.9 a). RADSAT appears to reduce water vapor (< -1 g kg^{-1}) over the area southeast of the tornadic supercell, but slightly increases water vapor (> 0.5 g kg^{-1}) in northeastern Colorado and northwestern Kansas (Figs. 3.9 b, c). More apparent differences are observed in the intensity of vertical wind shear, which is one of the most important factors that may affect the accuracy of predicting convection (Weisman and Klemp 1982). All experiments generate greater than 20 m s^{-1} 0 – 1 km vertical wind shear associated with the soon-to-be-tornadic convection in northeastern Colorado (Figs. 3.9 d - f). Both RAD and RADSAT intensify the vertical wind shear in northeastern Colorado (red boxes) while RADSAT also changes the pattern of vertical wind shear in southeastern Wyoming (Blue boxes) compared to CTRL and RAD experiments.

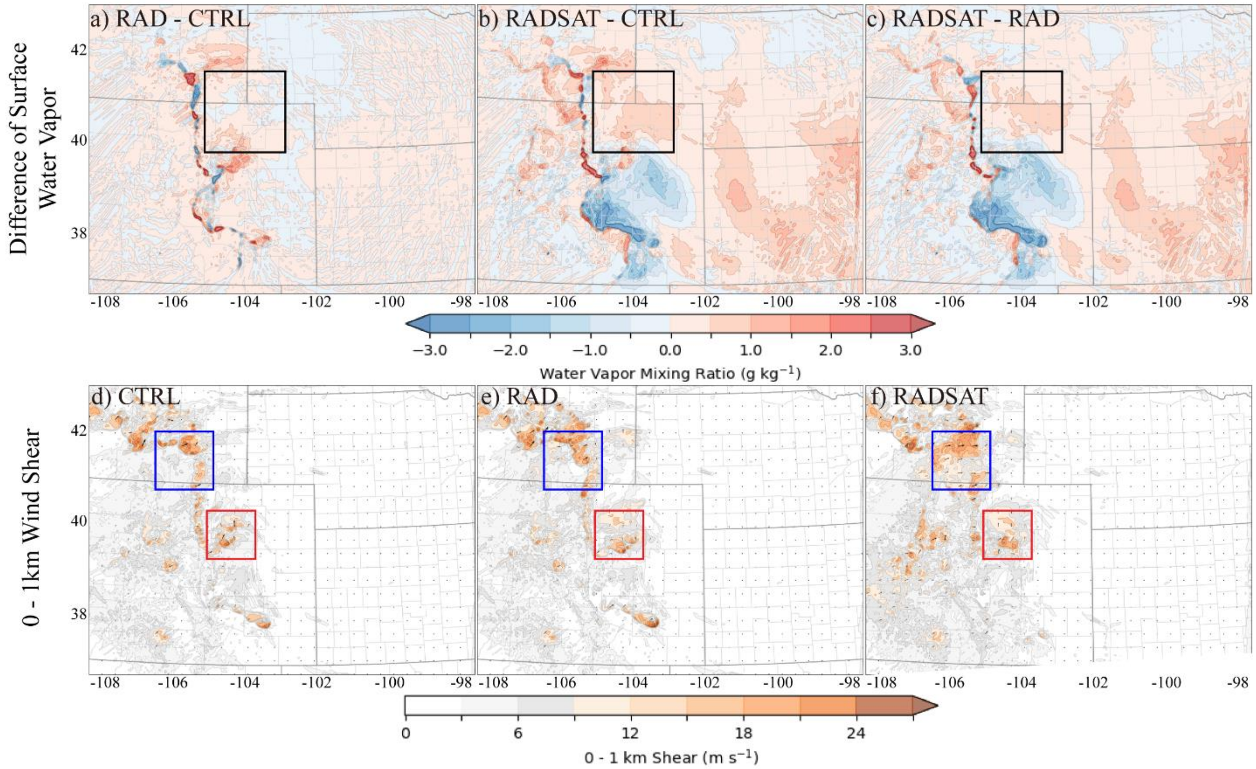

Figure 3.9 The differences of analyzed water vapor mixing at surface between experiments (a) – (c) and 0 - 1km wind shear for (d) CTRL, (e) RAD and (f) RADSAT at 2100 UTC 26 May 2017.

3.4.1.4 General discussions of reflectivity for all three cases

Some differences in the analyzed composite reflectivity are shown among the experiments depending on what kind of observations are assimilated. For the 10 May 2017 event (Fig. 3.10), CTRL misses the supercell located in the southern Texas Panhandle at 2000 UTC but produces spurious cells further south when compared with MRMS data (Fig. 3.10 a vs b). RAD analyzes the major supercell (Storm #1) but misses the storm cells in southeastern Oklahoma (Fig. 3.10 a vs c). Only RADSAT correctly places all storms near or at their observed locations, especially for the major supercell over the Texas Panhandle which produces several tornadoes at later times in the forecast near the Red River in addition to producing more storm cells near the border of Oklahoma and Kansas (Fig. 3.10 a vs d) though small spurious cells exist.

For the 16 May 2017 event (Fig. 3.11), CTRL produces both Storm #1 and Storm #2 over the Texas Panhandle at 2100 UTC which match the MRMS data (Fig. 3.11 a vs b), but the storm cluster over southwestern Kansas has significant phase errors. The pattern of the cluster prefers a southwest-northeast orientation as shown in MRMS data rather than a south-north orientation in CTRL. The pattern is somewhat corrected in RAD, but more areas over southwestern Kansas are covered by simulated composite reflectivity > 40 dBZ compared to the MRMS data (Fig. 3.11 a vs c). In RADSAT, the intensity (in dBZ) of the cluster seems match the observations well, but also produces spurious cells around this area (Fig. 3.11 a vs d).

For the 26 May 2017 event (Fig. 3.12), both CTRL and RAD produce a strong supercell at southeastern Colorado at 2100 UTC, which is not seen in the MRMS data (Fig. 3.12 a vs b, c), while RADSAT only generates a much weaker small-scale spurious cell at the same location (Fig. 3.12 d). Moreover, less spurious cells with > 40 dBZ composite reflectivity are observed over southeastern Wyoming and southeastern Colorado in the RADSAT experiment.

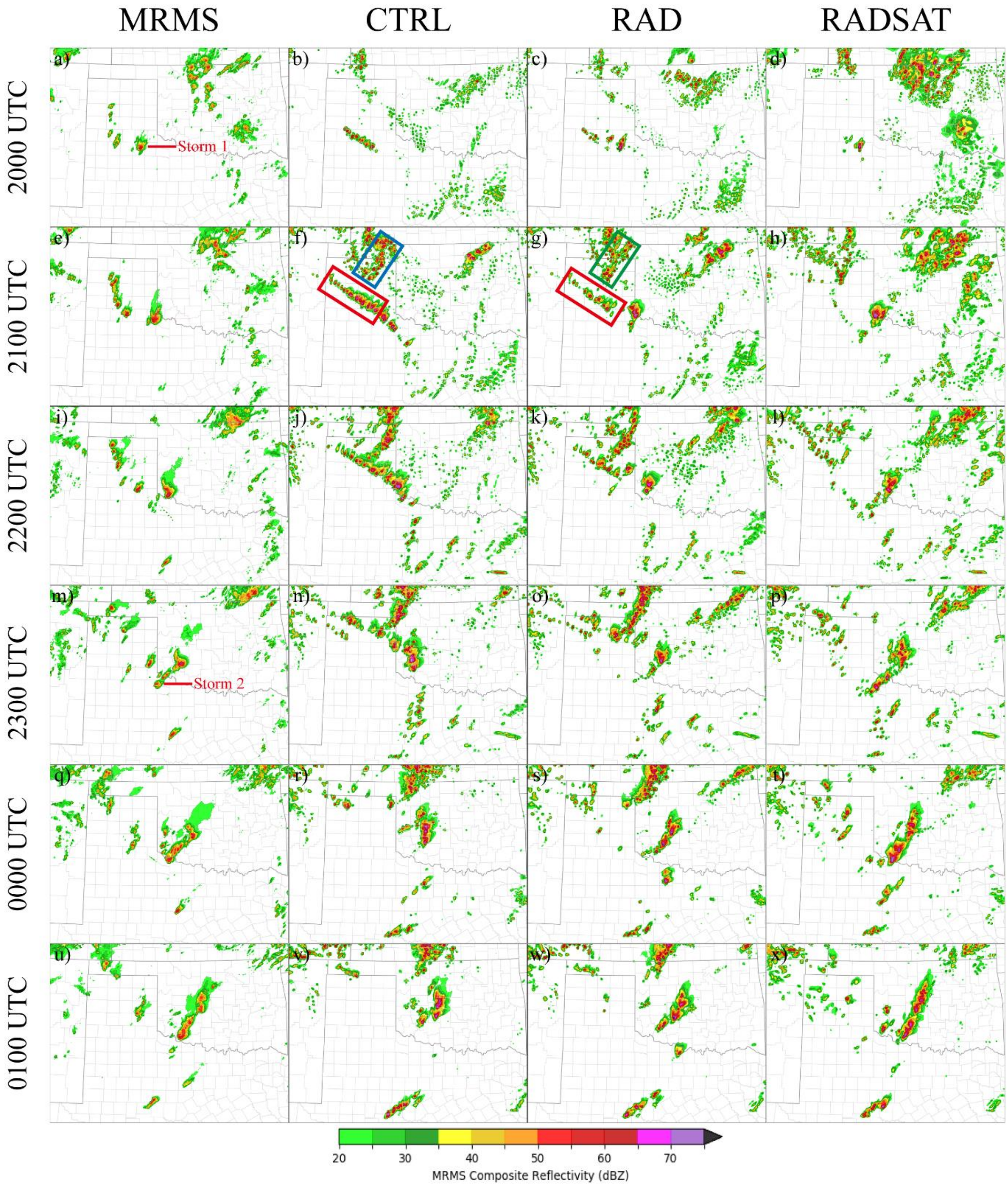

Figure 3.10 The composite reflectivity from MRMS (the first column), forecasted composite reflectivity for CTRL (the second column), RAD (the third column) and RADSAT (the fourth column). The first row (a-d) is from the analysis at 2000 UTC, 10 May 2017; the second (e-h), third (i-l), fourth (m-p), fifth (q-t), sixth (u-x) rows are from 1h, 2h, 3h, 4h, 5h forecasts initiated at 2000 UTC respectively. Boxes in (f) and (g) indicate the locations of overprediction without using satellite observations during the assimilation cycles.

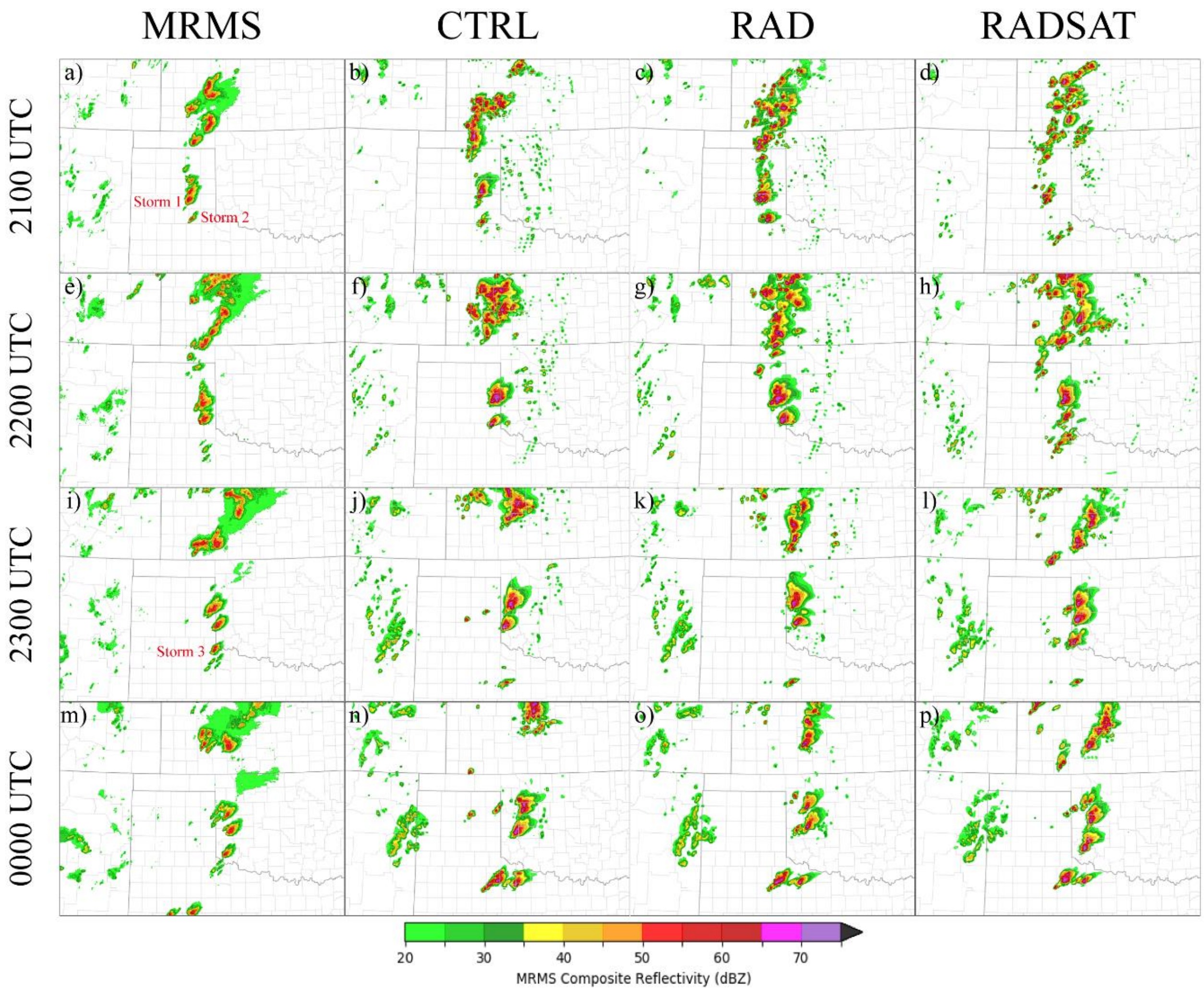

Figure 3.11 As in Fig. 3.10, but for 3-hour forecasts initiated at 2100 UTC, 16 May 2017.

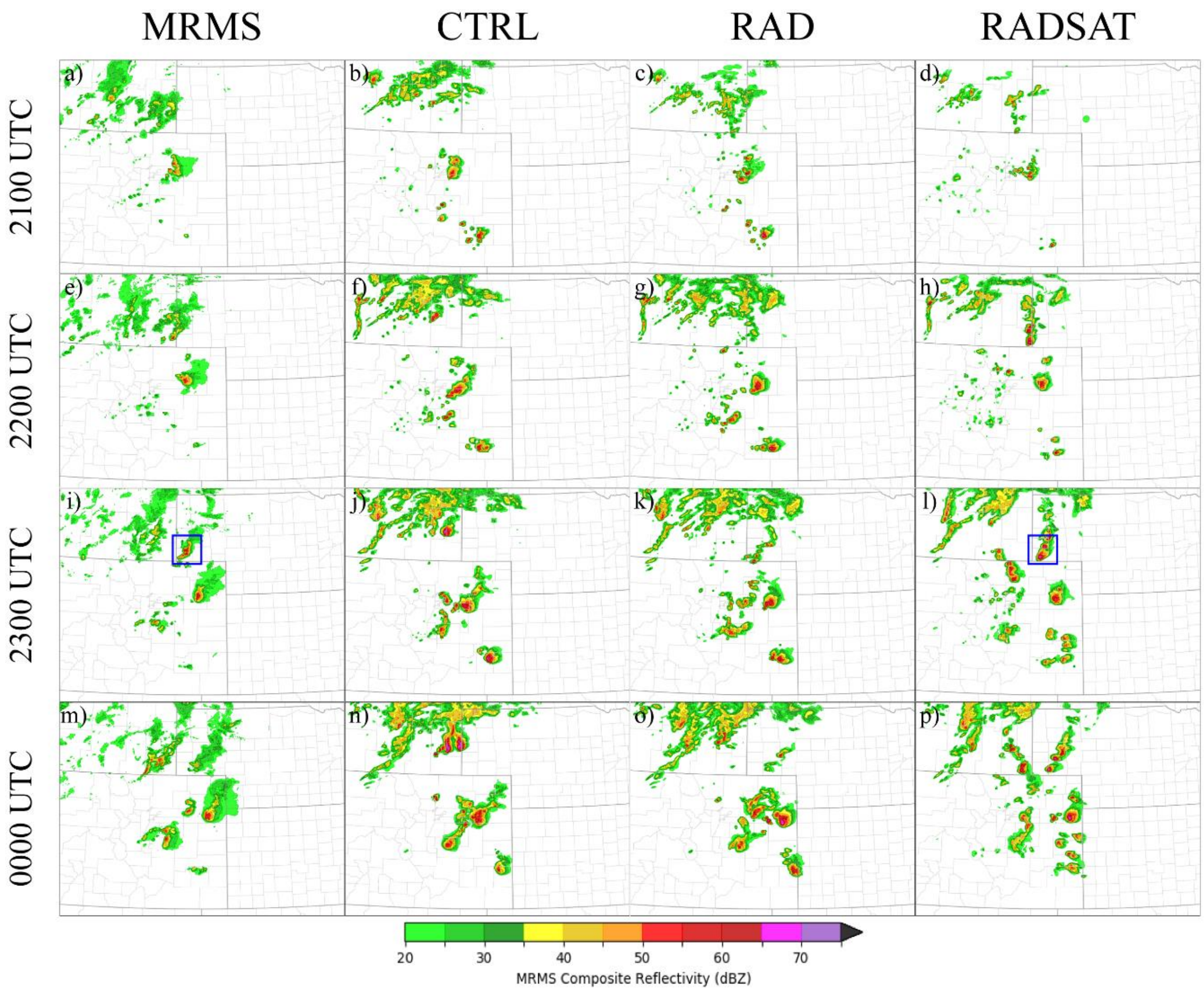

Figure 3.12 As in Fig. 3.10, but for 3-hour forecasts initiated at 2100 UTC, 26 May 2017.

3.4.2 Qualitative forecast evaluation

For the first 2-hour forecasts of the 10 May event, all experiments generate the supercell in the Texas Panhandle near the southwest corner of Oklahoma, but many spurious cells exist in both CTRL and RAD in the north Texas panhandle and Oklahoma Panhandle indicating large overprediction of convection (Figs. 3.10 f, g, j, k). RADSAT has the best prediction of the storms in general (Figs. 3.10 h, l) likely because the assimilation of satellite-derived data weakens the spurious convection in these areas. In the next 3-hour forecasts, this issue has been slightly alleviated but still exists in both CTRL and RAD. Moreover, large phase errors for the storm cluster around the southwest corner of Oklahoma are observed in both CTRL and RAD (Figs. 3.10 n, o, r, s, v, w). The storms in CTRL and RAD exhibit north biases compared to the MRMS data. The relative locations of the storms in CTRL prefer a south-north orientation instead of a southwest-northeast orientation as shown in the MRMS data. RADSAT performs better, especially for the orientation and location of the storms that produced tornadoes along the Red River (Figs. 3.10 p, t, x). The improvements observed in RADSAT are probably the consequence of combining changes in water vapor distribution over southwestern Oklahoma and its associated impact on MLCAPE and MLCIN, which are primarily due to the assimilation of LPW observations.

For the 16 May event, the result of the 3-hour forecast launched from 2100 UTC corresponds well to the analyses previously discussed. Two separate tornadic supercells are predicted by all three experiments between 2100 UTC and 0000 UTC (Figs. 3.11 a - p). However, two important improvements are present in RADSAT. The 1- and 2-hour forecasts for composite reflectivity show that all experiments maintain an MCS in western Kansas, while RADSAT generates an MCS pattern (Fig. 3.11 h) slightly closer to the observations at 2200 UTC (Fig. 3.11 e) though many small spurious cells exist. The MCS appears as a southwest-northeast linear shape at 2300 UTC in RADSAT (Fig. 3.11 *l*) which is similar to observed composite reflectivity (Fig. 3.11 i) compared to CTRL and RAD. Another improvement is related to the predictability of Storm #3 (Fig. 3.11 i). Only RADSAT produces a better forecast by generating this non-tornadic supercell associated with large hail damage while both CTRL and RAD fail to generate an isolated supercell in this region.

An improved forecast of severe convection on 26 May is also observed by comparing the simulated composite reflectivity within 3-hour forecasts launched from 2100 UTC (Figs. 3.12 a - p). All three experiments appropriately predict the tornadic supercell in northeastern Colorado during the 3-hour forecast period. CTRL differs from RAD and RADSAT in that it produces the tornadic supercell falling behind the location observed in the MRMS data. For example, RAD and RADSAT (Figs. 3.12 k - *l*) produce the tornadic supercell in Washington county, CO as in the MRMS observation (Fig. 3.12 i), while CTRL produces the tornadic supercell in Arapahoe county, CO. At 2200 UTC, 1-hour into the forecast, RADSAT overpredicts the storm initiated in the southeast corner of Wyoming (Fig. 3.12 e vs h), and the other two experiments totally miss this storm (Figs. 3.12 f, g). At 2300 and 0000 UTC, RADSAT maintains the storm similar to that in the MRMS observations. However, it produces more spurious convection near the border of Colorado and Kansas (Figs. 3.12 *l*, p) compared to the CTRL (Figs. 3.12 j, n) and RAD (Figs. 3.12 k, o).

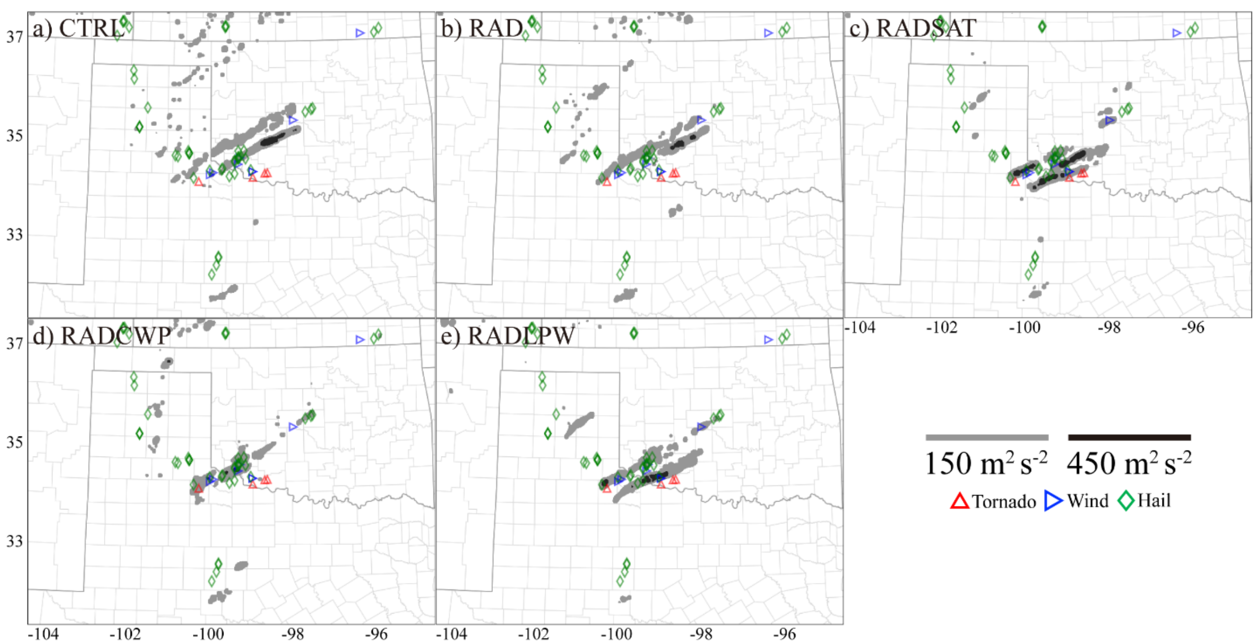

Figure 3.13 The tracks of UH between 2-5 km above ground (shaded) drawn from 6-hour forecasts initiated at 2000 UTC 10 May 2017 for the (a) CTRL, (b) RAD, (c) RADSAT, (d) RADCWP, and (e) RADLPW experiments. The red, blue and green markers indicate reports of tornadoes, damaging wind and large hail, respectively.

To illustrate the motion of convection associated with severe weather events, the 2 – 5 km UH tracks for 6-hour forecasts launched from 2000 UTC 10 May (Fig. 3.13), 2100 UTC 16 May (Fig. 3.14), and 2100 UTC 26 May (Fig. 3.15) are overlaid with the NWS/SPC storm reports during these time periods. The larger updraft helicity represents the stronger vertical motion and midlevel rotation between 2 to 5km at a given horizontal grid point.

On 10 May 2017, each experiment generates an UH track along or north of the reported severe storm events near the border of southwestern Oklahoma and North Texas. A hysteresis of strong UH associated with Storm #1 (indicated in Fig. 3.10 e) is predicted although it is imbedded in spurious convection in CTRL (Fig. 3.13 a). RADSAT best matches the track and intensity of UH to the NWS reports not only for the first storm, but also for the second storm observed around 2300 UTC in Foard County, Texas (Fig. 3.10 m). By comparing RADSAT and the standalone RADCWP/RADLPW experiments (Fig. 3.13 c vs Figs. 3.13 d, e), it is apparent that the improvement in initial conditions, and hence convection initiation and the movement of convection is due to the assimilation of LPW product.

For the 16 May event, all experiments produce similar 2 – 5 km UH tracks and intensities ($> 450\ m^2\ s^{-2}$) for the tornadic supercells. RADSAT generates a UH track almost exactly matching the NWS reports for both tornadic supercells and non-tornadic supercell (Fig. 3.14 c, Fig. 3.11 i), whereas CTRL and RAD totally miss the third storm (Figs. 3.14 a, b). RADSAT also provides a somewhat better UH track forecast for the hazardous weather events including tornadoes in southwestern Kansas between 2100 UTC and 0000 UTC.

As expected, all experiments for 26 May generate a UH track with value greater than 450 m^2s^{-2} for the tornadic supercell. However, CTRL has a south bias compared to the SPC reports (Fig. 3.15 a). Both RAD and RADSAT generate tracks better overlapping the SPC reports.

Although RADSAT has lower UH values along the track, it highlights the southeastern turn that occurred with the storm just after the first tornado formed (Fig. 3.15 c). Apparently, the storm in RADSAT is moving along the path where larger moisture gradient is observed (Figs. 3.9 b, c). Therefore, the differences of forecasts between the RADSAT and other experiments are very likely related to the changes in water vapor distribution related to the assimilation of LPW.

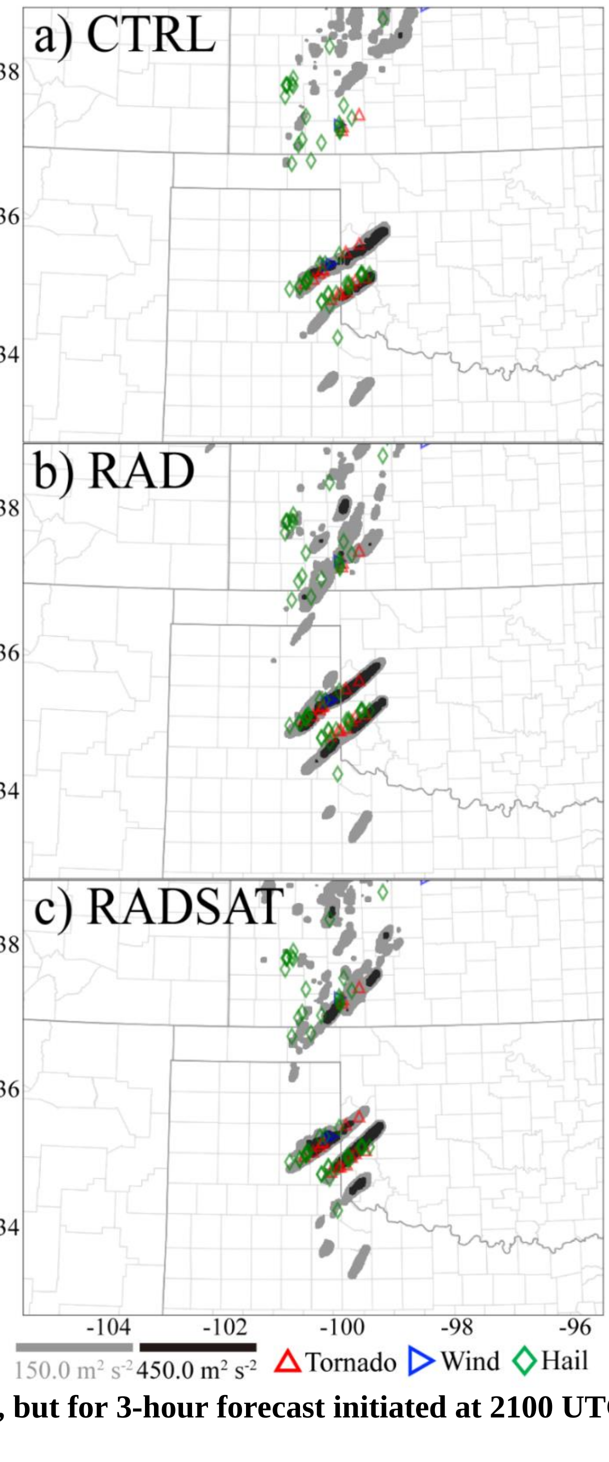

Figure 3.14 As in Fig. 3.13, but for 3-hour forecast initiated at 2100 UTC, 16 May 2017.

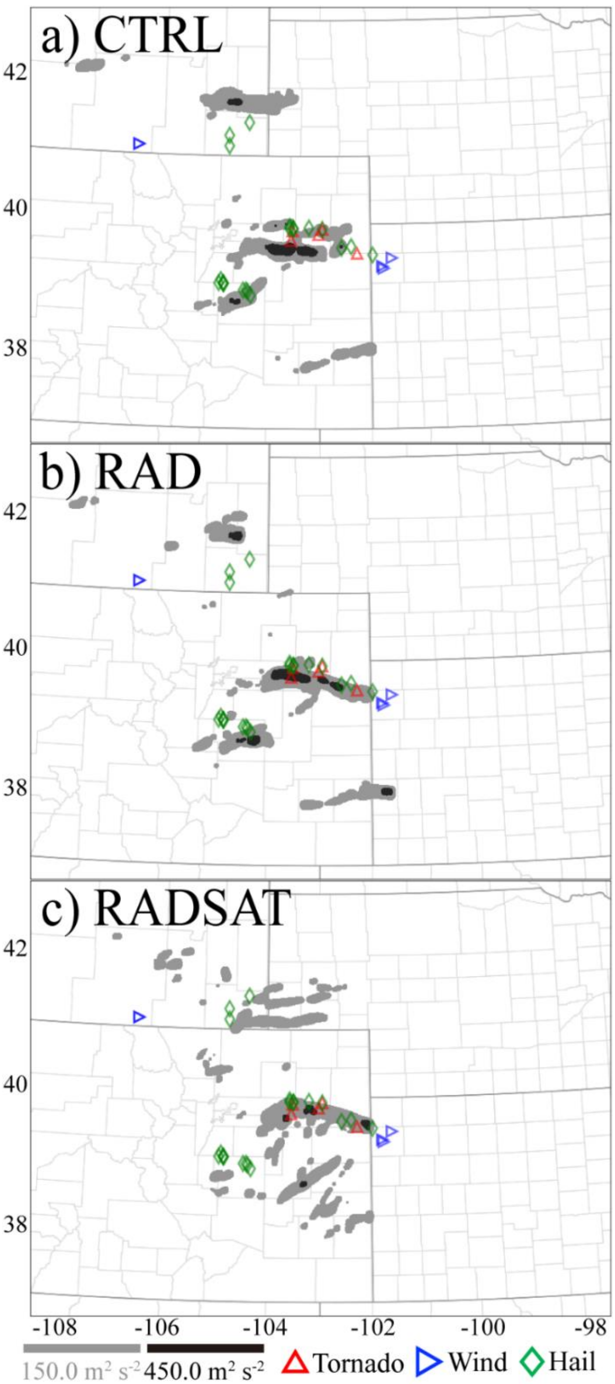

Figure 3.15 As in Fig. 3.13, but for 4-hour forecast initiated at 2100 UTC 26 May 2017.

3.4.3 Quantitative forecast evaluation

Although the subjective comparisons of composite reflectivity and UH track between the experiments suggest that assimilating additional satellite observations improves the skill of convection forecasts, the quantitative verification is essential to objectively assess the impact of assimilating different observations on the short-term forecasts. First, the simulated composite reflectivity and hourly accumulated precipitation are interpolated from model space to observation space by applying bilinear interpolation in the horizontal and linear interpolation in the vertical. Then, the BIAS and FSS of composite reflectivity and ETS of hourly accumulated precipitation are calculated relative to the MRMS composite reflectivity and the NCEP STAGE IV product, respectively. Composite reflectivity uses 20 dBZ and 40 dBZ as thresholds to assess the performance of each experiment, while hourly accumulated precipitation uses 2 mm and 10 mm thresholds. The neighborhood radius of 12 km is chosen for all these score calculations. The results are aggregated over 3-hour forecasts initiated every 1 hour between 2000 and 2300 UTC for all three cases (see Fig. 3.2). For the FSS and ETS, a value of 1.0 indicates a perfect forecast and the value of zero indicates no forecast skill. A bias of 1.0 indicates no bias, whereas values over 1.0 indicate overprediction and values below 1.0 indicate underprediction.

Throughout most of the 3-hour forecast period in all three experiments, the biases tend to be small (close to 1.0), when a reflectivity threshold of 20 dBZ is utilized (Fig. 3.16 a). However, positive biases exist during the early stage of the forecasts for both RAD and RADSAT experiments. This is due to the initiation of spurious cells at early forecast time, and they disappear along with the forecast time moving forward. For the 40 dBZ threshold, the results are different (Fig. 3.16 b). All experiments have bias scores greater than 1.0, indicating general overpredictions throughout the entire forecast period. The RADSAT maintains about 0.2 lower BIAS scores compared to the RAD and CTRL.

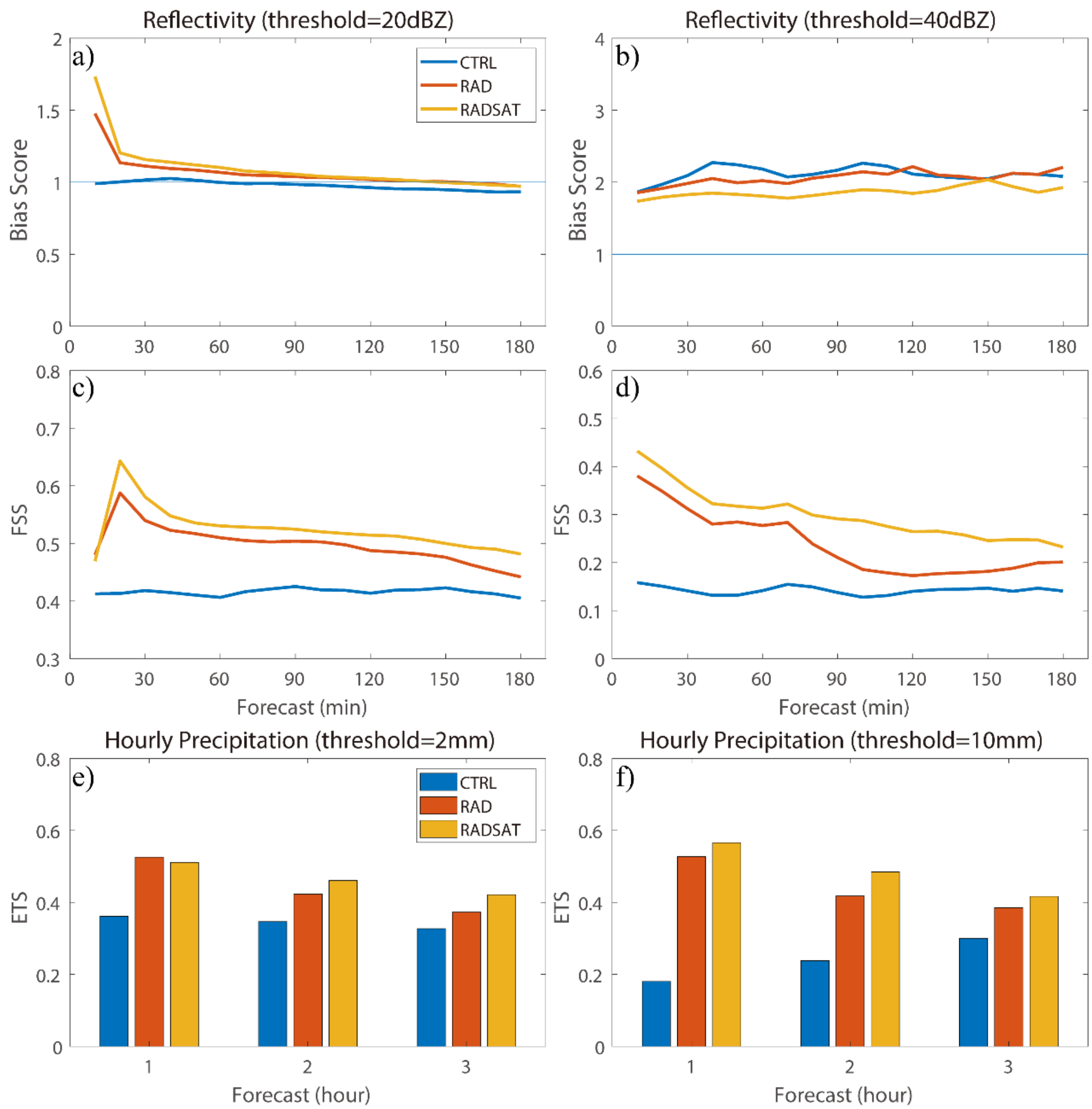

Figure 3.16 Frequency bias (a-b) and fractions skill score (FSS) (c-d) of composite reflectivity forecasts against the MRMS observations over the simulation domain. The thresholds used for verification are 20 dBZ and 40 dBZ. The results aggregate 3h forecasts starting from 2000, 2100, 2200, and 2300 UTC of all three events. Similarly, bars in (e) – (f) indicate ETS of hourly accumulated precipitation against the NCEP STAGE IV product aggregated over the same time period but using 2mm and 10mm as a threshold.

The FSS of RADSAT is always better than the other two experiments for both thresholds (Figs. 3.16 c, d) during the forecast period. Unlike the CTRL experiment which produces almost constant lowest performance in term of the FSS, RAD and RADSAT initially have higher scores (between ~ 0.4 and ~ 0.7) and decrease gradually as a function of forecast lead time. At the end of the 3-hour forecast period, the FSS of RADSAT is approximately 0.1 higher than that of CTRL and 0.05 higher than that of RAD for both thresholds. Overall, RADSAT outperforms other two experiments, and RAD outperforms CTRL against MRMS composite reflectivity.

The 1-hour accumulated precipitation performance relative to the Stage IV hourly precipitation product tells a similar story. For the 2 mm threshold, RADSAT outperforms 1-hour accumulated precipitation by 0.05 and 0.1 compared to RAD and CTRL, respectively (Fig. 3.16 e). It is worth noting that the performance of the first hour precipitation forecast in RADSAT is slightly degraded and about 0.14 lower than the RAD experiment. This overall improvement is further supported with the verification using a 10 mm threshold (Fig. 3.16 f). In general, the skill of 1-hour accumulated precipitation forecasts for all experiments decreases with time except CTRL experiment with 10 mm threshold.

3.4.4 Hybrid 3DEnVAR statistical metrics analysis

To assess the quality of the hybrid 3DEnVAR analysis cycling, some features of statistic metrics including the behavior of the normalized total functions, mean (bias) and root-mean-square (rms) for innovation and analysis residual during the cycling period are discussed. These metrics were defined in detail by Fierro et al. (2019). During the minimization process of the hybrid 3DEnVAR, the values of the cost function for each observation type generally decrease as a function of iteration number. It is also convenient to measure the distance between the observations and the background fields or the analysis by analyzing the rms innovation and the rms residual by applying the following equations:

$$RMS\_Innov = \sqrt[2]{\frac{1}{N}\sum_{i=1}^{N}\left[y_i - H\left(x_i^b\right)\right]^2}, \tag{3.1}$$

$$RMS\_Res = \sqrt[2]{\frac{1}{N}\sum_{i=1}^{N}\left[y_i - H\left(x_i^a\right)\right]^2}, \tag{3.2}$$

where N is the number of observations, y is the observation, and $H(x)$ is forward operators transferring the background fields (x^b) or the analysis field (x^a) from model space to observation space. Moreover, the mean innovation and mean analysis residual are analyzed to determine if any biases are generated for control variables by the DA process:

$$Mean\_Innov = \frac{1}{N}\sum_{i=1}^{N}[y_i - H(x_i^b)], \quad (3.3)$$

$$Mean\_Res = \frac{1}{N}\sum_{i=1}^{N}[y_i - H(x_i^a)]. \quad (3.4)$$

Because the results of all three cases are qualitatively similar, only the first event (10 May 2017) is discussed here as an example.

In general, both RAD and RADSAT experiments show a decrease in the total cost function with the RAD experiment at various analysis times producing a reduction ranging between 50% and 80% at the end of the minimization process. The largest decrease (over 75%) is observed at initial analysis time 1800 UTC (Fig. 3.17 a). Given more observations being assimilated, RADSAT generally has 10% less reduction compared to RAD at the same analysis time (Fig. 3.17 b). RAD produces greater reduction (around 40%) of the total cost function at the first 3 iterations at 1800 and 1900 UTC, while the RAD experiment for the remaining analysis times and the RADSAT experiment for all analysis times have stable reduction rates between iterations.

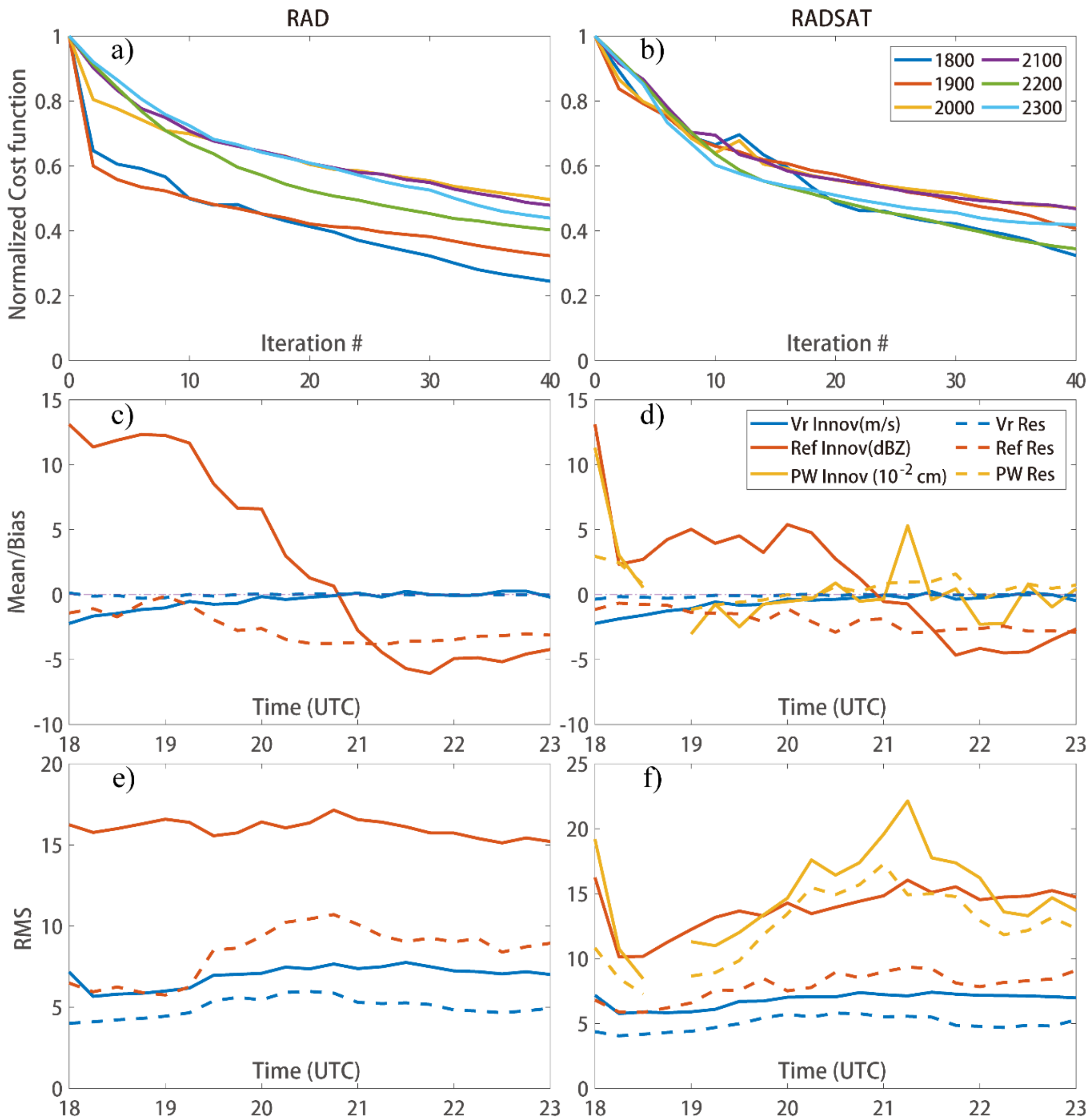

Figure 3.17 Normalized cost function (a-b) as a function of iteration number for the RAD and RADSAT experiments. The cost function curves are shown in colors for different analysis time following the legends in (b). Mean (c-d) for the innovation (solid lines) and analysis residual (dashed lines) for radial velocity (blue, in m s^{-1}), reflectivity (red, in dBZ) and precipitable water (yellow, in 10^{-2} cm) as a function of analysis time for the RAD and RADSAT experiments. Dash-dotted lines in (c) and (d) are the reference lines for measuring biases. As in (c-d), RMS for the innovation and analysis residual are shown in (e) for RAD and (f) for RADSAT.

The innovation biases for LPW are positive and largely depart from the reference line (exceeding 1 mm) for the first 30 minutes of the DA cycling (Fig. 3.17 d), indicating that the background fields initially have dry biases compared to the observations over most part of the domain; especially over western Texas and southwestern Kansas (Fig. 3.6). Since the dry biases are ameliorated within 3 analysis timestamps by assimilating LPW observations, the innovation and analysis residual biases for LPW fluctuate around the reference line starting from 1830 UTC. The innovations for radial velocity have negative biases between 1800 and 2000 UTC, but both innovation and residual for radial velocity turn to be identical and close to zero since 2000 UTC for both RAD and RADSAT experiment (Figs. 3.17 c, d). This indicates that the analyzed radial velocity fields fit the observations very well. For reflectivity, the innovations are positive with largest bias ~ 13 dBZ at 1800 UTC and decreasing at later DA cycles. They finally alter to be negative since 2100 UTC and reach lowest value (~ -5 to -6 dBZ) near 2145 UTC. In contrast, the residuals for reflectivity are close to the reference line only between 1800 and 1900 UTC for RAD and between 1800 and 2000 UTC for RADSAT. Over the whole DA cycling period, they are consistently negatively biased. This statistical feature suggests that the nonlinearity of reflectivity forward operators may prevent the DA system from readily absorbing all the information from reflectivity observations. The rms statistics (Figs. 3.17 e, f) show that the values for analysis residual are typically smaller than the innovation values. The analysis is able to reasonably fit all kinds of observations during the whole cycling period. Moreover, the residual values for all types of observations are close to their corresponding observation errors respectively, indicating these observations are properly assimilated.

3.5 Summary and Conclusion

In this chapter, three experiments are conducted to investigate the impact of assimilating radar, surface, and satellite retrieved CWP and LPW observations on short-term convective-scale

NWP with three high-impact weather events that occurred in May 2017. The CTRL experiment does not assimilate any observations. The RAD experiment assimilates radar and surface observations, and the RADSAT experiment further assimilates satellite derived CWP and LPW products in addition to the radar and surface data. Generally, the comparison of results from three types of experiments indicates the improvement on model initial conditions and short-range forecasts for three severe weather events when CWP and/or LPW are also assimilated in addition to radar and surface data.

For the 10 May 2017 event, assimilating CWP improves the cloud property analysis, the downward shortwave radiation over the cloudy areas, and therefore surface temperature and atmospheric instability. Spurious cells are mostly removed in later forecasts due to this improvement. Assimilating the LPW product also yields positive results for CI by saturating the moisture fields surrounding the ongoing convection. The improved analysis typically generates high UH swaths that are more appropriately placed when compared against the SPC damage reports.

Similar impacts of assimilating satellite products are observed for the other two cases. For the 16 May 2017, RADSAT leads to intensified moisture gradient at the surface compared to CTRL and RAD. As a result, the storms of interest are better analyzed and predicted when verified against the observations. For the 26 May 2017, RADSAT improves water vapor distribution at the surface and low-level vertical wind shear, leading to better placed storms in the model at times and locations consistent with the observations. Again, the reflectivity and UH track forecasts from RADSAT are better than those from the CTRL and RAD.

Chapter 4 The Development of Weather-dependent Background Error structures within a Convective Scale Radar Data Assimilation Scheme for Improving Short-Range Severe Weather Forecasts

4.1 Introduction

To properly spread information from observation locations to nearby model grid points and to obtain a balanced analysis among all model control variables, all DA methods require either explicit or implicit estimation of background error covariances, i.e., **B** matrix. A fundamental issue in estimating a **B** matrix, as described by Bannister (2008), is that the real state of the atmosphere is unknown. How to accurately and timely estimate background error statistics is always a major research topic in DA, particularly in variational DA. In the case of ensemble DA (e.g., EnKF), the flow-dependent **B** matrix can be calculated by assuming that the ensemble mean represents the best estimate of the atmosphere and that the spreading of the ensemble members around represents model errors (Evensen 2003). This approach allows flow-dependent background error (denoted as "**P**" in EnKF) to be calculated and used for assimilating observations without imposing homogeneous and isotropic assumptions, albeit it requires localization using the Schur product to reduce sampling noise related to spurious large-scale correlation (Houtekamer and Mitchell 1998, 2001). On the other hand, in variational DA, the static **B** matrix is commonly estimated using two different methods. The first is known as the "NMC" method (Parrish and Derber 1992). This method calculates the differences between forecasts of various lengths valid at the same time, such as a 24-hour and a 48-hour forecast. The estimated background error covariances using such forecast lengths often work for synoptic-scale DA and NWP, but not for convective-scale DA and NWP. An alternative method, known as the ensemble method or Monte Carlo method, produces background error statistics by utilizing an existing ensemble of forecasts, as presented by Buehner

(2005). The static **B** matrix established by using ensemble forecasts avoids the shortcoming associated with the NMC method (Brousseau et al. 2011), and is more suitable for short length- and time-scale, or convective-scale DA.

Despite the advantages of the ensemble technique for constructing static **B** matrix and the necessity of representing background error statistics with precision and low computing cost, limited studies have been made, particularly for convective-scale DA. Monteiro and Berre (2010) investigated the error variance of the model variables over seasons, days, and hours. Their research revealed that background error structure differs (i) between seasons, i.e., summer versus winter; (ii) under various meteorological situations, i.e., day-to-day synoptic situation changes and hour-to-hour meso- or convective-scale situation changes. As pointed out by their research, error standard deviations are much larger and horizontal correlation length scales are much smaller with a winter unstable low-pressure system, and the correlation length scales are smallest in the horizontal direction and largest in the vertical direction during convective activity in the summer afternoon. Montmerle and Berre (2010) conducted a similar study. They used a convective-scale NWP to identify differences in background error structure between precipitating and nonprecipitating areas. Larger standard deviations for variables except water vapor, smaller horizontal correlation length scale, and greater vertical correlation in the mid-troposphere were found in precipitating area compared to nonprecipitating area. Consequently, a methodology for creating heterogeneous **B** matrix using the ensemble method was suggested so that the impact of observations would be optimized for precipitating and nonprecipitating areas in the DA process. Although flow-dependent covariances can be naturally represented by the hybrid method in a variational framework, the heterogeneous static **B** matrix may also be beneficial in situations when flow-dependent error covariances are insufficiently sampled (Wang et al. 2007, 2009; Buehner et al. 2010a,b, 2015), for instance, the use of limited ensemble members in a hybrid system.

Following these efforts, Michel et al. (2011) proposed a synthetic description of calculating heterogeneous **B** matrix from short-term ensemble forecasts at convection-allowing scale, and implemented the method in a 3DVAR system for both traditional and hydrometeor variables. The method classified precipitating and nonprecipitating areas using geographic masks based on rain content, and then applied these masks to the devised **B** for precipitating and nonprecipitating areas. Wang and Wang (2021) extended Michel's work to develop a heterogeneous **B** matrix for clear-air, weak storm, and strong storm conditions using similar technique based on reflectivity thresholds in conjunction with a hybrid EnVAR DA system. The results demonstrated that the heterogeneous **B** matrix improved the accuracy of both analyses and short-term severe weather forecasts.

On top of those earlier studies, this study further explores the effectiveness of heterogeneous **B** matrix tested in the WoF-Var – variational component of WoF-Hybrid (Gao et al. 2013; Hu et al. 2021) at convective-scale by incorporating a more detailed classification of geographic masks based on composite reflectivity intensities. The improved **B** matrix, named weather-dependent background error covariance matrix, are constructed within a variational framework without hybridization with an ensemble system. This new development on analysis and short-term forecast is evaluated through a single observation experiment and two severe weather events: the 5 May 2020 linear complex and the 10 Dec 2021 long-lived supercell, both of which produced significant property damage and fatalities. Thus, improving the accuracy and efficiency of predicting these high-impact severe weather events could have valuable societal impacts.

This chapter is further organized as follows. Section 4.2 describes the method for estimating the error standard deviations, while section 4.3 outlines the experimental design and configurations. The impact of the weather-dependent **B** matrix on a single observation experiment is analyzed in section 4.4. Section 4.5 discusses the results of the analyses and forecasts of two high-impact severe weather events, both with and without the use of the weather-dependent **B** matrix, using qualitative and quantitative verification methods. The conclusion and discussions are given in the final section.

4.2 Method for Error Standard Deviation Estimation with a Reflectivity-based Bin

Some previous studies utilized ensemble spread as a proxy for the background error standard deviations, **D**, mainly for synoptic or mesoscale applications (e.g., Isaksen *et al.* 2007, Bonavita *et al.* 2012). Montmerle and Berre (2010) employed geographic masks based on vertically averaged rain content to diagnose background error statistics using a convective-scale nonhydrostatic model and found that such statistics, including error standard deviations and correlation length scales, vary considerably in precipitating and nonprecipitating areas. Other related studies at the convective scale include those by Caron and Fillion (2010), and Michel et al. (2011).These studies inspired the development of a more flexible method based upon short-term ensemble forecasts.

Specifically, the proposed approach involves the calculation of **D** by deriving vertical background error profiles using the ensemble forecasts from the experimental NSSL WoFS. This system has been in operation during the HWT SFE for several years and consists of a 36-member ensemble created using the GSI-EnSRF and the WRF-ARW model. The study uses all ensemble forecasts from 25 events of the NOAA HWT SFE 2019 to compute background error statistics. The vertical error structure for different weather situations is estimated using ensemble perturbations from 3-hour forecasts at each grid point, classified based on the strength of simulated composite reflectivity. To facilitate the use of these error standard deviation profiles for general purpose application in convective-scale DA, a new scheme is developed in this study. This scheme includes the following three steps: (1) classifying the ensemble perturbations based on reflectivity to preliminarily construct vertical error profiles, (2) filtering outliers for each reflectivity category, and (3) applying Gaussian smoothing to the vertical error profiles (Fig. 4.1).

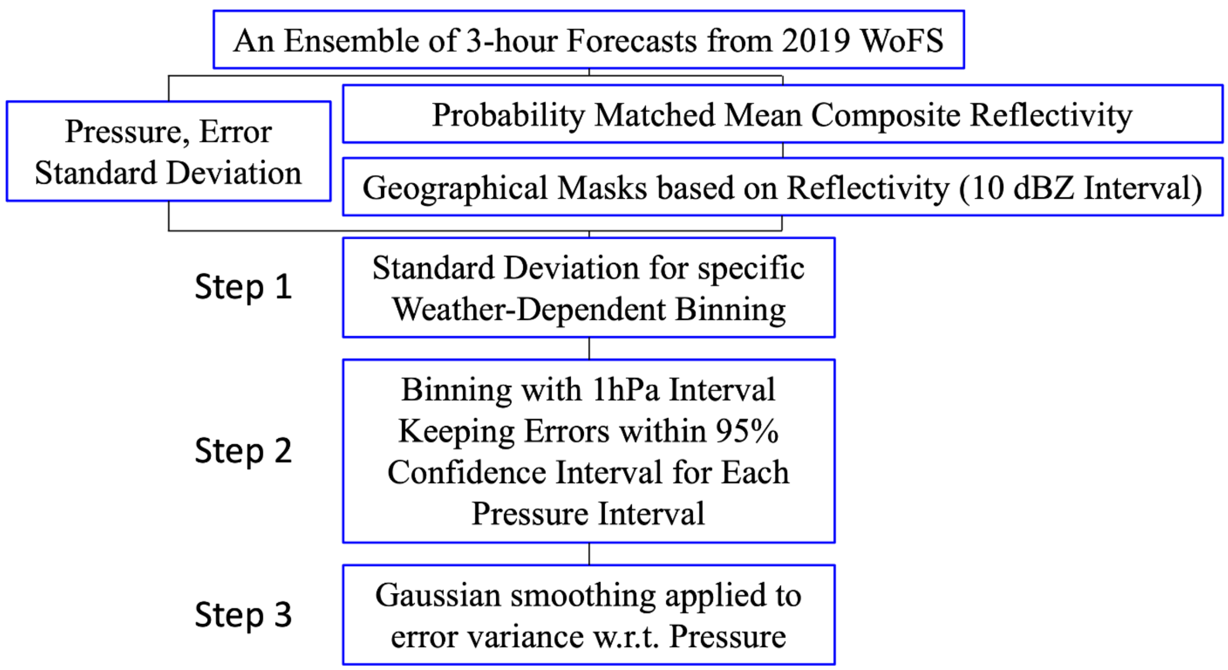

Figure 4.1 The workflow for estimating background error profiles from WoFS ensemble perturbations.

Similar to Michel et al. (2011), in the first step, the computational domain is separated based on geographic masks and then the profiles of static error statistics are retrieved for different regions. The geographic masks are defined as locations where the value of probability matched mean reflectivity (pmm_ref) from the WoFS ensemble is in between the lower and upper thresholds. More specifically, we set the thresholds at every 10 dBZ between 0 dBZ and 50 dBZ, separating the error standard deviations into 7 classes: pmm_ref = 0 dBZ (clear air); 0 < pmm_ref <=10 dBZ;

10 < pmm_ref <= 20 dBZ; 20 < pmm_ref <= 30 dBZ; 30 < pmm_ref <= 40 dBZ; 40 < pmm_ref <= 50 dBZ; and pmm_ref > 50 dBZ. Therefore, 7 error profiles will derive from those WoFS ensemble outputs.

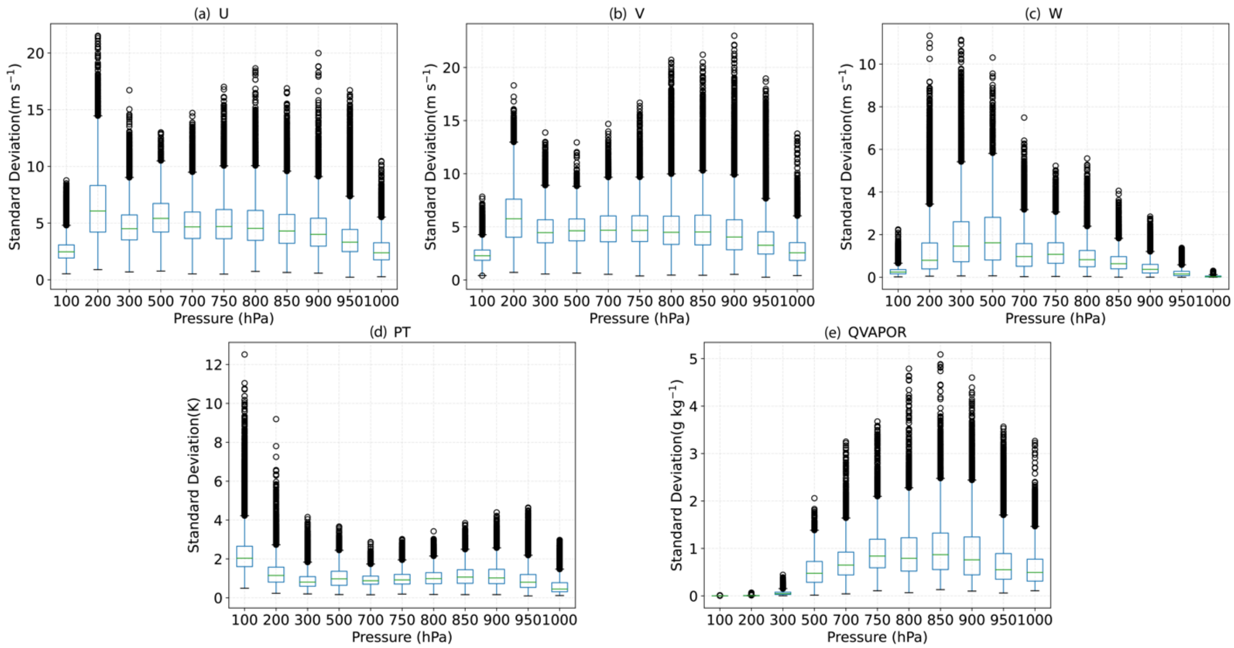

Figure 4.2 The distribution of the error variance for (a) zonal wind, (b) meridional wind, (c) vertical velocity, (d) potential temperature and (e) water vapor content within specific pressure bins for the area with 40 < Z <= 50 dBZ before outlier filtering.

The second step is to remove the outliers. To estimate the static background error profiles that typically work for all likely convective-scale circumstances, error standard deviations in each class are subsequently binned into pressure levels with 1 hPa intervals vertically. Figure 3.2 shows the distribution of the error standard deviations within specific pressure bins for the area with 40 < pmm_ref <= 50 dBZ. It is obvious that some data points could be outliers since their values may surpass the upper (lower) bound defined in the third (first) quartile plus (minus) 1.5 times interquartile range. If a dataset is normally distributed, an outlier is also considered as a data point that lies beyond ~2.7 times the standard deviation from the mean and significantly departs from other data points. As a result, these outliers are eliminated before the next step.

Finally, a Gaussian kernel is utilized to smooth the retained data points for each reflectivity class (Davies 2005). The pressure-error pairs are sorted based on their pressure values and then convolved with a one-dimensional finite impulse response kernel that uses Gaussian values. Because the size of the dataset differs remarkably between classes, the smoothing employs a Gaussian kernel with variable window size and standard deviation. For the class of decreasing reflectivity, the window size and standard deviation of the Gaussian kernel increase monotonically because of the increased amount of data points. Data points within strong convection zones (> 40 dBZ) constitute approximately 16% of the non-zero reflectivity data, while grid points in weak convection zones account for the remaining 84%. The clear-air region has roughly 4.5 times the number of grid points compared to the non-zero reflectivity region. Even in places with severe convection, over a million data points are available for providing a robust estimate of error standard deviations. As a result, the kernel for the "> 50 dBZ" reflectivity class has the smallest window size and standard deviation, whereas the kernel for "clear-air" has the biggest window size and standard deviation. The smoothed vertical profile is then linearly interpolated in the WoF-Var system to obtain an error at a given pressure level for a specific variable.

Figure 4.3 shows vertical profiles of the standard deviation of background error for traditional variables (zonal wind, meridional wind, vertical velocity, temperature and water vapor content). For comparison, standard deviations inferred from the WoFS ensemble of 2019 HWT events are displayed. For the subdomain with reflectivity greater than 0, errors in the wind fields are substantially bigger at all levels. The rise in standard deviation for wind fields may imply small-scale dynamical circulation within convection of various intensities. The profiles of horizontal wind fields (Figs. 4.3 a, b) show two extremes at 800 hPa and 200 hPa that could be related to low level convergence at cloud base and upper-level jet stream at cloud top. The greater value of vertical velocity error (Fig. 4.3 c) between 800 hPa and 200 hPa may be explained by the rotating updraft within convective clouds. For potential temperatures (Fig. 4.3 d) below 200 hPa and water vapor mass (Fig. 4.3 e) at all levels, error standard deviation derived from the area covered by convection is close to that obtained from the clear sky area. The extreme in the potential temperature profile at 150 hPa may indicate the uncertainty at the level of the tropopause due to the sudden cessation of the lapse rate, whereas water vapor error profile reveals a maximum at 800 hPa that could be connected with water vapor entrainment at cloud base.

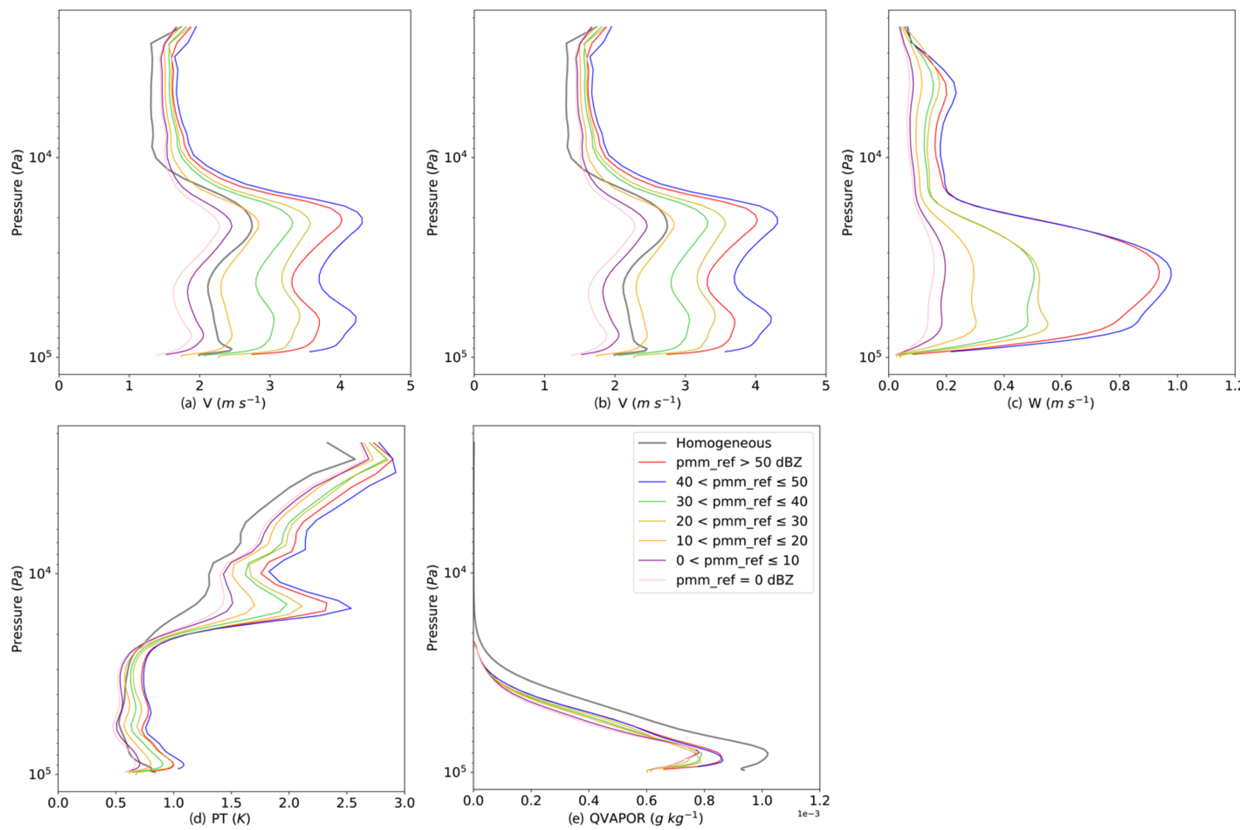

Figure 4.3 Vertical profiles of background error standard deviations for several model variables (a) zonal wind, (b) meridional wind, (c) vertical velocity, (d) potential temperature and (e) water vapor content.

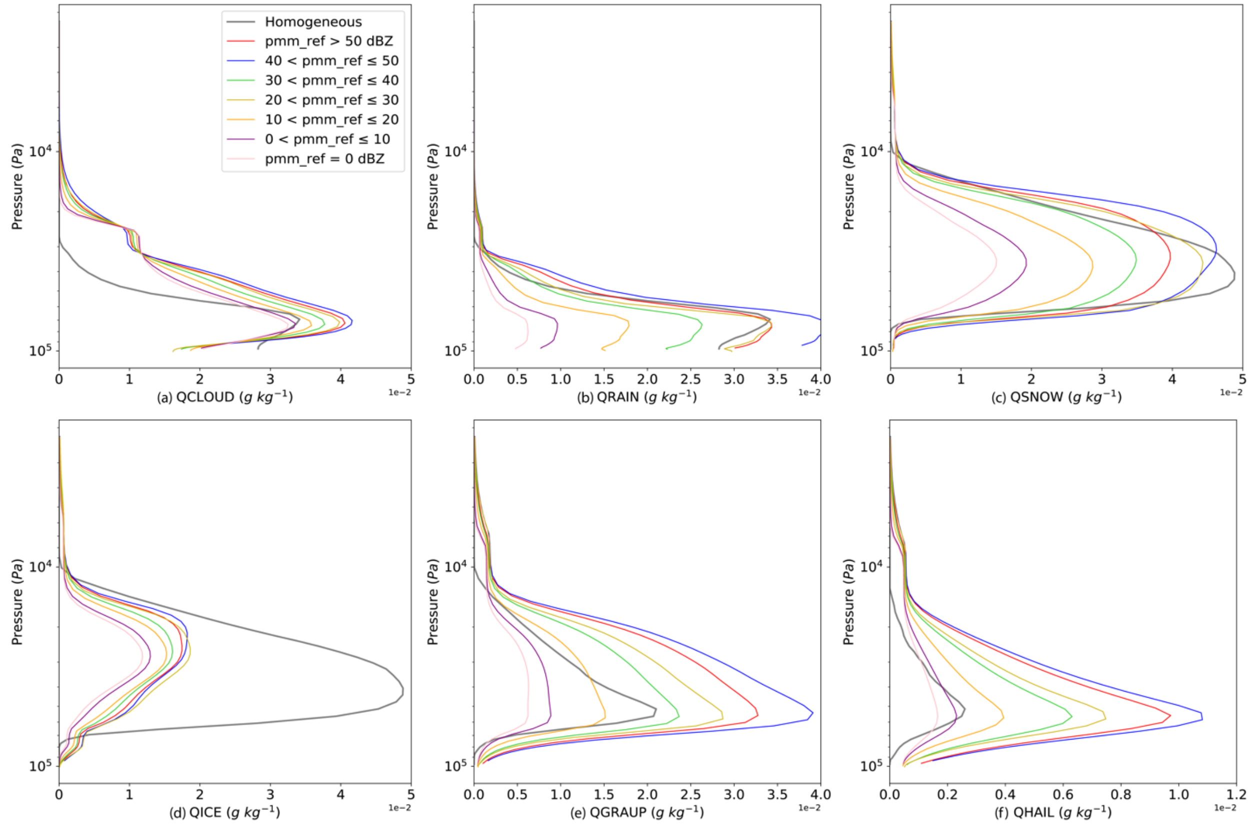

Figure 4.4 Same as Fig. 4.3, but for hydrometeor related variables (a) cloud droplet, (b) rain, (c) snow, (d) ice, (e) graupel and (f) hail.

As for the shape of the distribution of hydrometeor themselves, all standard deviations of hydrometeor errors (Fig. 4.4) have single poles at different altitudes. For example, both cloud droplets (Fig. 4.4 a) and rain (Fig. 4.4 b) errors have the maximums around 800 hPa. These maximums are typically near the level of cloud base and are below the freezing level. Snow and ice errors have the maximums above 400 hPa with a broader vertical extension and are located above the freezing level. Similar to three-dimensional wind fields, the background error standard deviations for hydrometeors are increasing with composite reflectivity. Compared to the errors in low reflectivity zones, errors in high reflectivity zones are much larger at all altitudes for rain, snow, graupel and hail (Figs. 4.4 b, c, e, f), and to a lesser extent for cloud and ice (Figs. 4.4 a, d). Nonzero error standard deviations for precipitating particles (rain, snow, graupel and hail) are also observed in the area with pmm_ref = 0, showing that phase errors of convective clouds exist in the WoFS ensemble. After these error standard deviations are derived, they can be easily incorporated into the WoF-Var system described in the last section.

4.3 Experiment Design and Configurations

The impact of the newly developed weather-dependent error standard deviations in the 3DVAR method on short-term convective NWP is evaluated through the analyses and forecasts of two severe weather events: the 5 May 2020 event over the boundary of North Carolina and South Carolina (Event #1, Figs. 4.7, 4.8) and the 10 December 2021 rare winter tornado outbreak (Event #2, Figs. 4.14, 4.15). For the first event, two EF-2 tornadoes touched down in Chester and Lancaster Counties, northern South Carolina, producing some property damage. The second event was a devastating outbreak, which included more than 30 tornado reports in six states stretching across the Mississippi Valley, Southeast and Midwest. One tornado, which is famously called the "Quad-State Tornado", ripped across four states (Arkansas, Missouri, Tennessee and Kentucky) within four hours, and produced huge damage in Mayfield, Kentucky, which was the hardest hit town in this event.

In this study, the selected forecast model is the WRF-ARW version 3.9.1. The model physics configuration includes the NSSL two-moment microphysics scheme (Mansell et al. 2010), the Dudhia shortwave radiation scheme (Dudhia 1989), the Yonsei University planetary boundary layer scheme (Hong et al. 2006), and the rapid radiative transfer model longwave radiation scheme (Mlawer et al. 1997). The model domain size for both cases is 900×900 km^2, with a horizontal grid spacing of 1.5 km. The stretched vertical grid consists of 51 levels with a top set at 20 hPa. The time step for the integration of the prognostic equations is set to 6 seconds.

The configurations used here are identical to those used in the hybrid ensemble Kalman filter and variational data assimilation schemes implemented in the HWT SFE 2021 real-time operational environment (Gao et al. 2022; Carpenter 2022), except that the assimilation of radar and satellite observations starts from 1800 UTC and ends at 0300 UTC, the next day for both cases. The 3-km HRRR forecast products are used for providing the initial background for the DA cycles and the lateral boundary conditions for forecasts. For each case, the radar observations are assimilated every 15 minutes, and the conventional observations (e.g., soundings, surface stations) are assimilated hourly. Six-hour forecasts are launched hourly until 0300 UTC. The MRMS composite radar reflectivity dataset, and NWS/SPC storm reports (https://www.spc.noaa.gov/climo/online/) are used to verify the forecast performance.

Prior to conducting real data case studies, the impact of a single observation on nearby grid points is evaluated through the analysis increments. Two single observation experiments are performed to compare the behavior of the analysis with the use of the newly developed weather-dependent background error covariances to the original scheme without using the weather-dependent error covariances. Following the single observation experiment, two types of experiments are conducted for each real-data case, one with the original homogeneous background error labeled as “Retro”, and another with the weather-dependent error structure labeled as “WD”, are performed for each case. The results of these two experiments are compared to investigate the impact of the newly constructed weather-dependent error standard deviations on short-term convective-scale severe weather forecasts within the pure variational DA framework.

4.4 Results for Single Observation Experiments

A single-observation experiment is first carried out to investigate the behavior of the storm-scale analysis using weather-dependent error profiles. The experiment is based on the event that occurred at 2330 UTC on 15 May 2020 because of the complicated weather situation provided by this particular case. In this experiment, several meridional wind observations within the analysis domain, whose locations are grid points (271, 389), (403, 369), (351, 250), (533, 324), (343, 403), (132, 159), (577, 455), (200, 371), (125, 60) at 800 hPa and 500 hPa are assimilated. These locations depict various weather conditions in the 15 May 2020 event. The pseudo-observations are generated as a result of 3 m s^{-1} meridional wind innovation at these locations.

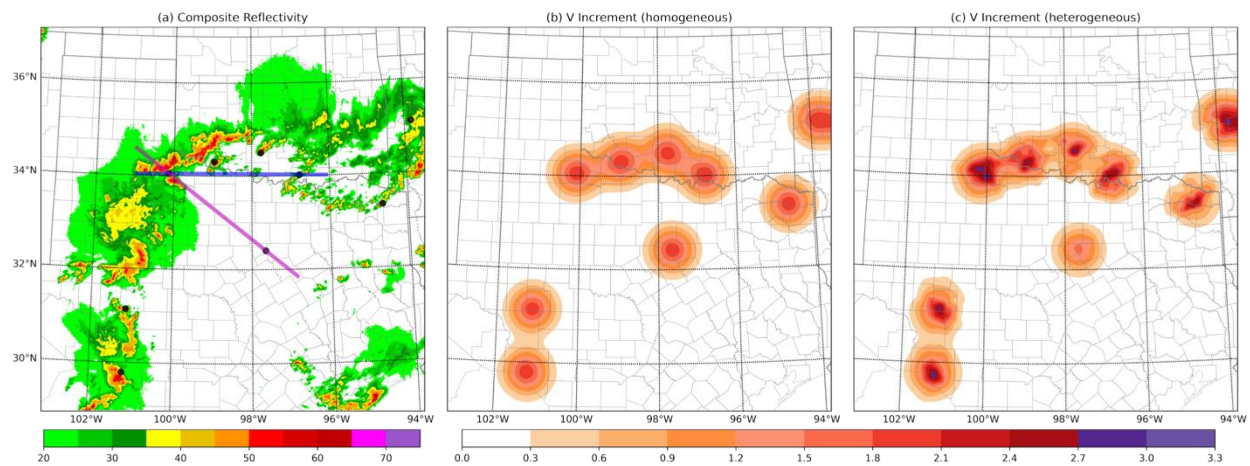

Figure 4.5 The (a) MRMS composite reflectivity and the 500 hPa analysis increments of meridional wind field for (b) the single observation experiment with homogeneous background errors and (c) the single observation experiment with weather-dependent background errors at 2330 UTC on May 15, 2020. The blue and magenta lines indicate where the vertical cross sections are plotted in Fig. 4.6.

Figure 4.5 illustrates the impact of weather-dependent background errors on the analysis increments of meridional wind field at 500 hPa when 9 pseudo meridional wind observations are assimilated using the 3DVAR approach. It is obvious that the structure of meridional wind increments all exhibit isotropic Gaussian forms (Fig. 4.5 a) when using homogeneous background error profiles. As expected, the areas of the analysis increments influenced by each observation are comparable, with a similar maximum increment in the range of 1.8 to 2.1 m s^{-1} at all pseudo-observation locations. When applying the weather-dependent background error profiles (Fig. 4.5 b), the analysis increment structures are remarkably changed resulting from the spread of the information from these pseudo-observations. It clearly shows that the analysis increment near the center of the analysis domain, which is associated with the pseudo-observation at grid point (351, 250), maintains a semi-isotropic Gaussian shape with a smaller peak value in the range of 1.5 to 1.8 m s^{-1} when compared to the single observation experiment using homogeneous background error profiles. However, with the exception of the increment at the domain's center, all of the analysis increments have weather-dependent characteristics, as indicated by the fact that higher analysis increments are extended along the higher composite reflectivity. The maximum increments at these places are generally increased to a value greater than 2.4 m s^{-1} and are typically correlated to the convective core.

Figure 4.6 displays vertical cross sections of meridional wind increment along the axis defined by the blue and magenta lines in Fig. 4.5. The blue line in Fig. 4.5 passes across two concurrent strong convective storms as well as the anvil area of another convection in the middle of the two and the magenta line passes across two adjacent intense convective cells and the clear sky area where a pseudo-observation is assimilated. In the comparison of two single observation experiments (Fig. 4.6 a vs b), it is observed that the use of weather-dependent background errors results in significantly higher increments. Additionally, the increment pattern in areas with composite reflectivity greater than 0 is found to be tighter horizontally and stretched vertically in the experiment with weather-dependent background errors. The vertical cross sections along the magenta line (Fig. 4.6 c vs d) shows similar increment variations for ongoing convection zones in both single observation experiments. Notably, the increments in the clear sky area (around 32.47 N, 97.75W) are seen to shrink in both horizontal and vertical directions, with peak values also decreasing when using weather-dependent background errors. In summary, the single observation experiments demonstrate what was expected: increments with obviously different characteristics in terms of weather intensities and patterns are shown by comparing experiments with or without using the weather-dependent background errors.

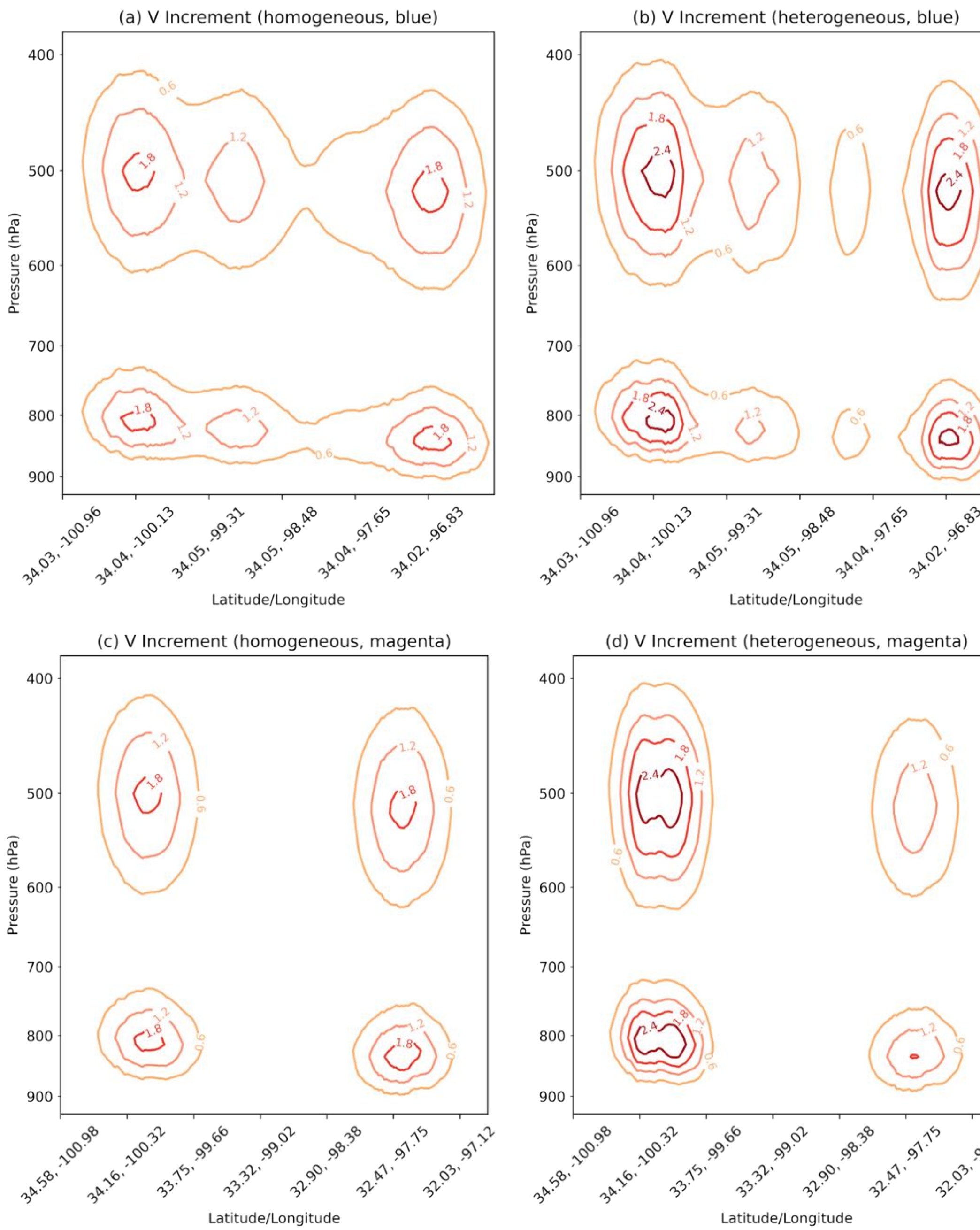

Figure 4.6 The vertical cross sections of meridional wind increment along the axis defined by the blue and magenta lines in Fig. 4.5 for (a, c) the single observation experiment with homogeneous background errors and (b, d) the single observation experiment with weather-dependent background errors.

4.5 Results

Experiments involving two events occurred on 5 May 2020, and 10 December 2021 are performed to further examine the impact of weather-dependent background errors on storm-scale DA and NWP. The results of analyses and forecasts for Retro and WD are discussed during the severe weather active periods. For each event, qualitative forecast evaluations in terms of composite reflectivity and updraft helicity (UH) tracks are conducted, followed by subjective diagnostic analyses at some initial times. Simulated composite reflectivity will be compared to observed composite reflectivity product from the NSSL MRMS and simulated UH tracks will be compared to NWS/SPC storm reports across the forecast period. Finally, quantitative verifications of forecast skill, known as neighborhood ETS and performance diagram (Roebber 2009), will be undertaken using alternative composite reflectivity thresholds. The statistical significance of the differences in skill scores is determined through a permutation test, as proposed by (Hamill 1999).

4.5.1 5 May 2020 event

On 5 May 2020, thunderstorms were developing along a weak west-to-east warm front that moved across an area ranging from northern Georgia to northern South Carolina. A buoyant airmass with temperatures in the 70s F and dewpoints in the mid 60s F heated up along the front, resulting in a 1000 to 2000 J kg^{-1} mixed-layer CAPE. Basically, the environment is considered to be favorable to initiation and persistence of convection. As a result, a mesoscale convective complex started to initiate around 1900 UTC, drifted eastward and arrived at the eastern North Carolina coast at 0300 UTC.

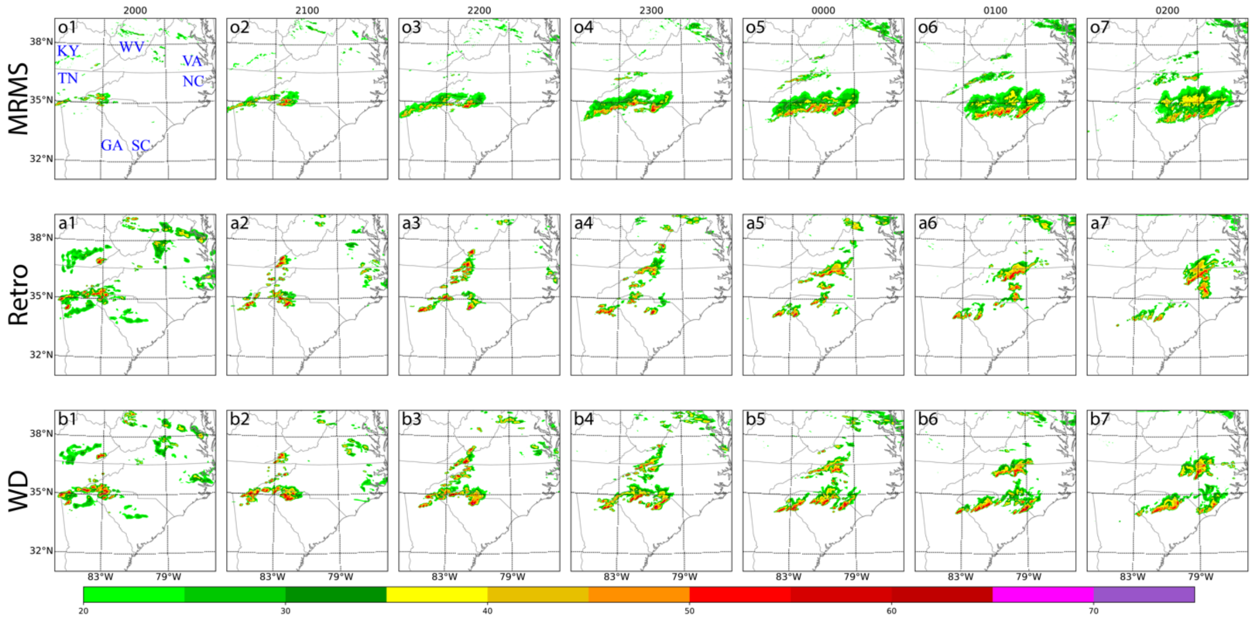

Figure 4.7 (from top to bottom) The (o) composite reflectivity from MRMS and the forecasted composite reflectivity for (a) Retro, and (b) WD. (1) The analysis at 2000 UTC 5 May 2020. The (2) 1-, (3) 2-, (4) 3-, (5) 4-, (6) 5-, and (7) 6-hour forecasts initiated at 2000 UTC.

Figure 4.7 explores the impact of using homogeneous and weather-dependent background errors in the WoF-Var system on the locations and patterns of the 5 May 2020 convective complex, which produced two EF-2 tornadoes along its way. The analyses successfully capture the mode and orientation of the storm in both experiments, as demonstrated by the analyzed composite reflectivities at 2000 UTC (Figs. 4.7 a1 and b1) compared to the observations (Fig. 4.7 o1). For the subsequent forecasts, the primary linear shape convective complex maintains and propagates pretty well in the forecast from the WD experiment (Figs. 4.7 b2 - b7 vs o2 - o7). In the Retro run, some convective storms maintain and propagate to the east direction (Fig. 4.7 a3 vs o3) that correspond well to the observations, however, they soon disorganize and then vanish after 3 hours into the forecast (Figs. 4.7 a4 - a7 vs o4 - o7). The simulated composite reflectivity values within the deeper convective storms along the front in the WD experiment are greater than the observed values in the first two hours (Figs. 4.7 b2 - b3) and become comparable to the observed values at the later forecast hours (Figs. 4.7 b4 - b7). It is worth noting that both Retro and WD produce spurious convective cells near the border of Kentucky and Virginia at 2000 UTC. Between 2000 and 0200 UTC, these convective cells are developing and propagating near the boundary of Virginia and Tennessee, then the boundary of Virginia and North Carolina. When compared to observations, these convective cells initiate earlier, propagate faster on average and exhibit a slightly northeastward displacement bias and are much stronger than observed cells after 2300 UTC. In general, the composite reflectivity values are over-forecasted. This implies that the WoF-Var analysis and forecast system is less capable of predicting weaker convection than stronger convection. This is possibly related to high biases that are easy to present in the first half hour forecast in a 1.5 km deterministic forecast using a WoFS member configuration, compared to the 3 km WoFS forecast (Miller et al. 2022).

Severe weather events, including those that can result in hail and tornadoes, are often associated with strong rotation and updrafts, which can be indicated by the UH tracks. Figure 4.8 shows the 2 – 5 km UH tracks for 6-hour forecasts released from 1900, 2000, 2100 UTC on 5 May 2020 overlaid with the NWS severe weather reports. For a given horizontal grid point, the stronger the UH values, the greater the midlevel vertical motion and midlevel rotation. In comparison to the NWS storm reports, both experiments exhibit fairly good 6-hour forecasts initialized from three different times at 1900 (Figs. 4.8 a, d), 2000 (Figs. 4.8 b, e), and 2100 (Figs. 4.8c, f) UTC respectively. Each experiment generates a UH track along or south of the reported tornadoes (purple triangles) and hail events (cyan diamonds) near the North Carolina/South Carolina border. However, several significant improvements provided in WD should be recognized below.

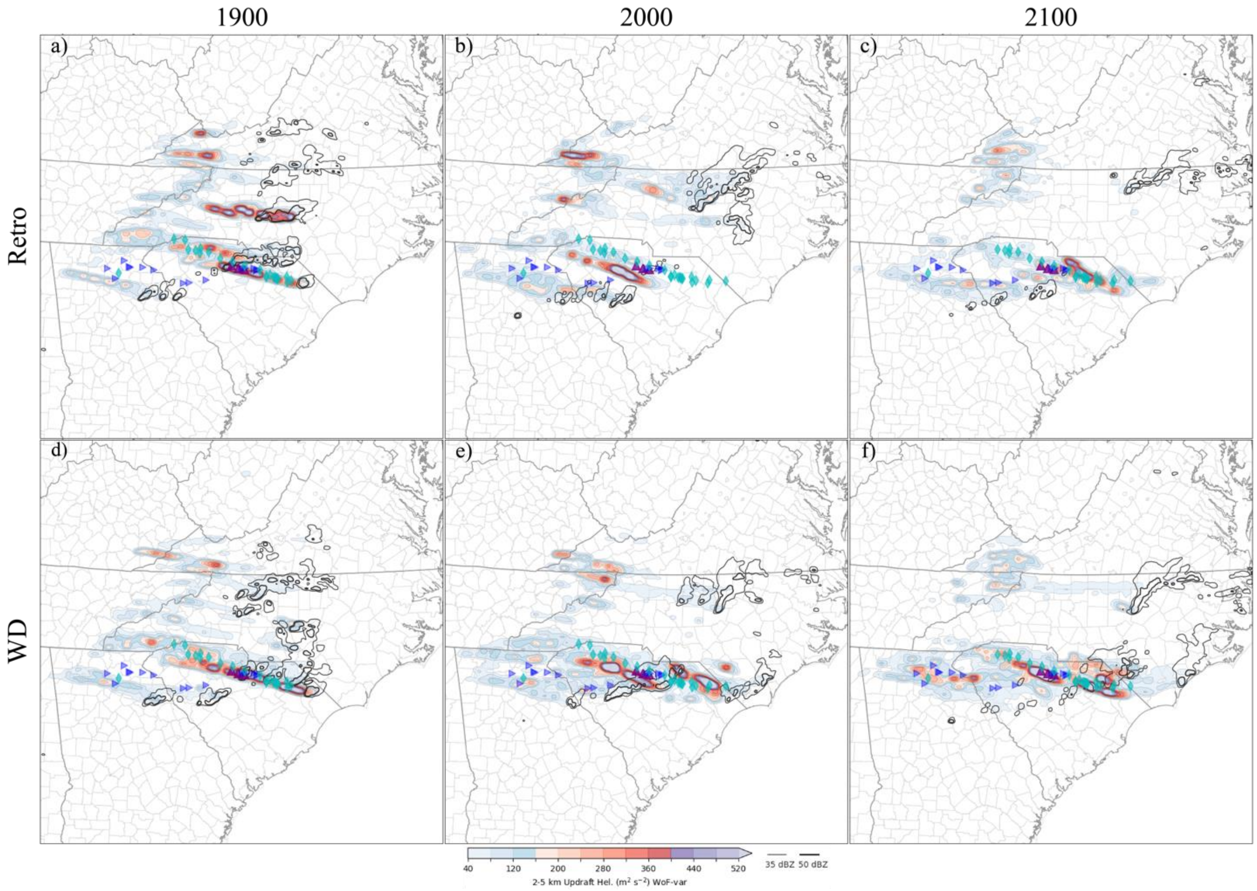

Figure 4.8 The UH tracks between 2 and 5 km above ground (filled contours) drawn from 6-hour forecasts initiated at (a, d) 1900, (b, e) 2000 and (c, f) 2100 UTC 5 May 2020 for the (top) Retro and (bottom) WD experiments. The composite reflectivities valid at the end of each forecast period are depicted by the black contour lines. The purple, cyan, and blue markers indicate reports of tornadoes, hail and damaging wind, respectively.

First, the most significant improvement is associated with the primary tornadic supercell which produced the strong UH track directly overlaid by most hail and tornado reports. WD produces a closer 2 – 5 km UH track with stronger intensity for the tornadic supercell compared to Retro, particularly for the forecasts starting from 2000 and 2100 UTC. Retro clearly misses some hail events that occur at later times in the forecast from 2000 UTC (Fig. 4.8 b vs e) and UH values are weak at earlier times in the forecast from 2100 UTC in comparison with WD (Fig. 4.8 c vs f). Furthermore, the forecast initialized from 2100 UTC (Fig. 4.8 c) and the later initiation times (not shown) in Retro yield lower UHs near tornado touchdown locations. In addition to improvements associated with the intensity of storm rotations (UH tracks), the mitigation of spurious convection is another benefit of employing weather-dependent background errors with the WD experiment. WD clearly weakens the UH swaths of the spurious convection in North Carolina in the forecasts starting from 1900 and 2000 UTC, despite producing false simulated composite reflectivity (i.e., in Fig. 4.7). Last but not least, for the forecast beginning at 2100 UTC, WD enhanced the simulated UH over northern Georgia where wind damages are reported.

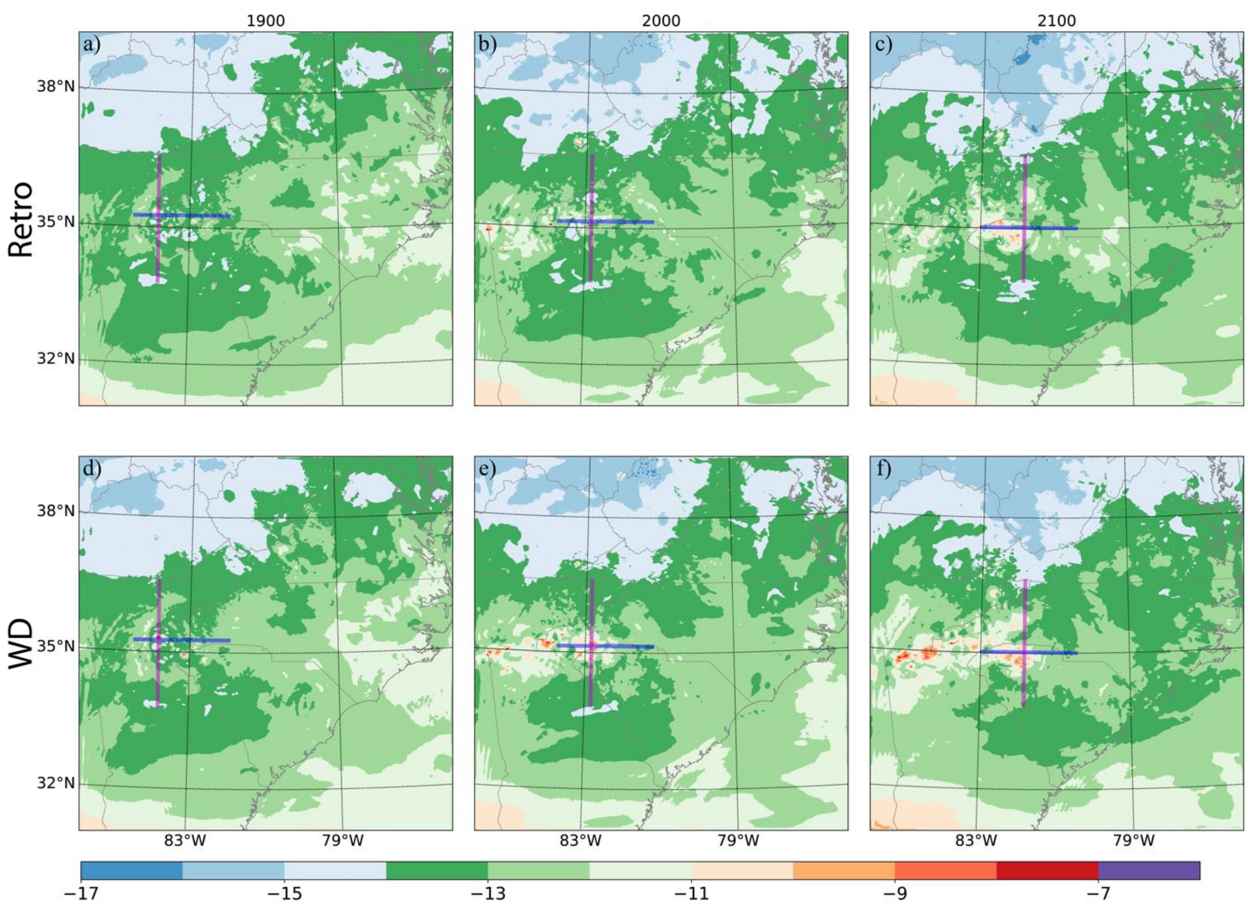

Figure 4.9 The analyzed temperature at 500 hPa valid at the analysis time (a, d) 1900, (b, e) 2000 and (c, f) 2100 UTC 5 May 2020 for the (top) Retro and (bottom) WD experiments. The magenta and blue lines are placed in zonal and meridional direction across where the major convective complex locates. The vertical cross sections are plotted in Fig. 4.10 and Fig. 4.11 along the lines.

To further understand the impact of the WoF-Var system with the weather-dependent background errors, the detailed analyses of temperature along various horizontal and vertical cross sections are given in this paragraph. Because the 20 May 2020 severe weather event is characterized by the presence of a west-to-east frontal system, temperature distribution is an important component that may affect the initiation and development of the mesoscale convective complex. Figure 4.9 illustrates the analyzed temperature at 500 hPa valid at 1900, 2000 and 2100 UTC respectively. The magenta and blue lines are placed close to where the primary convective complex initiates. The cross-section plots of moisture, temperature and wind field along the magenta and blue lines are displayed in Figure 4.10. The temperature distribution away from the intersection of two lines is comparable across two experiments at 1900 UTC, according to the 500 hPa plot (Fig. 4.9 a vs d). However, the temperature close to the intersection and east of the intersection along the front at 500 hPa, which indicate a substantial effect on buoyancy above the level of free convection, are greatly increased (~ 2 K) in WD at analysis times 2000 (Fig. 4.9 b vs e) and 2100 UTC (Fig. 4.9 c vs f).

The zonal (Fig. 4.10) and meridional (Fig. 4.11) vertical cross sections, which complement the horizontal cross section, emphasize the finding revealed by the horizontal cross section. The differences in both cross sections are highlighted by the green boxes. At 1900 UTC, the zonal cross sections of Retro and WD are similar (Fig. 4.10 a vs d), whereas the meridional cross section of WD exhibits a single amplified temperature bulge rather than two separate bulges from Retro (Fig. 4.11 a vs d). In the analyses valid at 2000 and 2100 UTC, WD exhibits an increase in temperature at levels above 700 hPa within the designated area indicated by the green boxes in Figs. 4.10 and 4.11 (b, c vs e, f). This increase is consistent with the changes observed in the 500 hPa horizontal cross section. These changes may lead to the forecast improvement with composite reflectivity and UH in the primary convective complex.

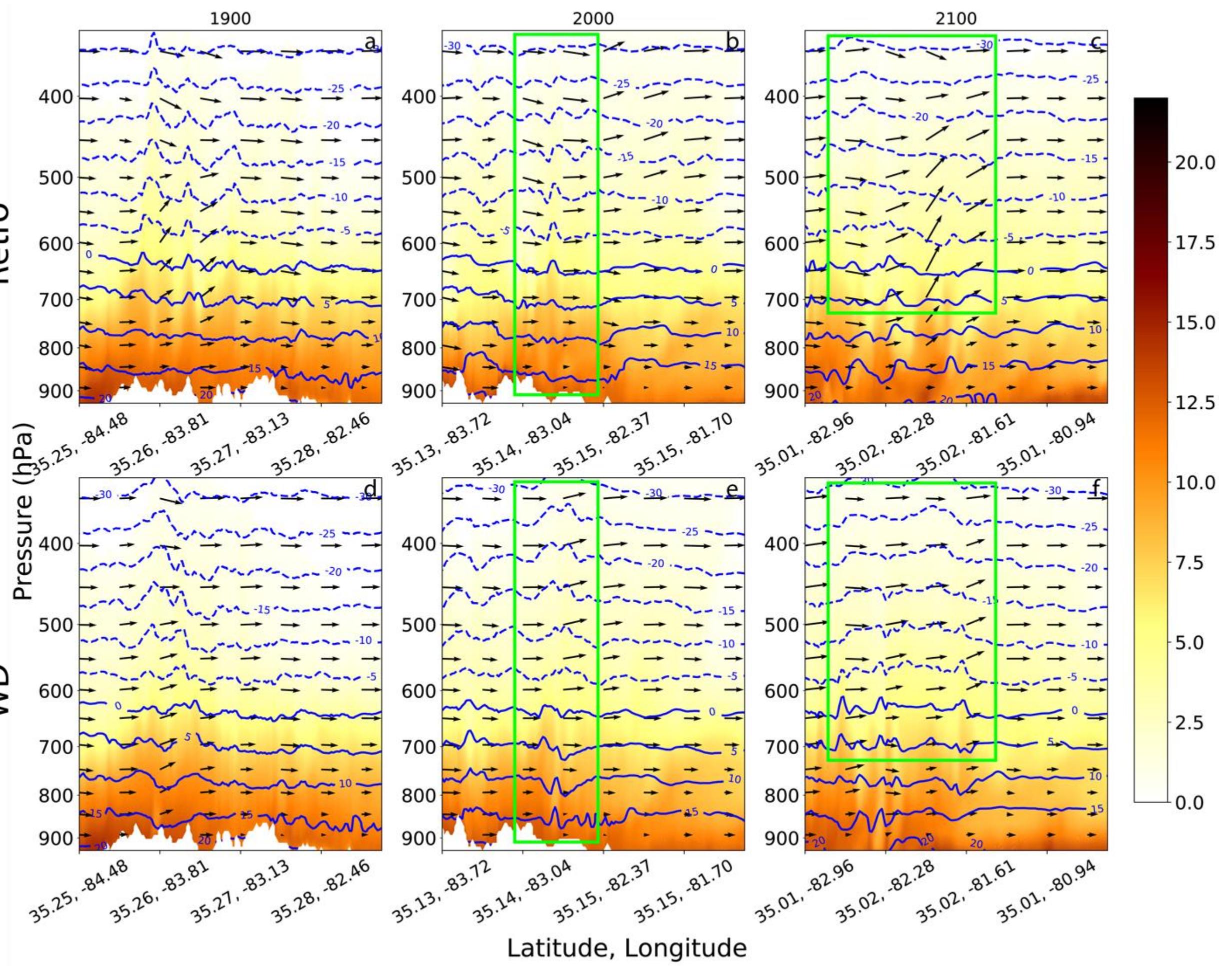

Figure 4.10 The vertical cross sections from the analysis of wind (vectors), specific humidity (filled contours, in g kg^{-1}), and temperature (blue contours) along the blue line in Fig. 4.9. The figure displays the vertical sections valid at (a, d) 1900, (b, e) 2000, and (c, f) 2100 UTC for the (top) Retro and (bottom) WD experiments.

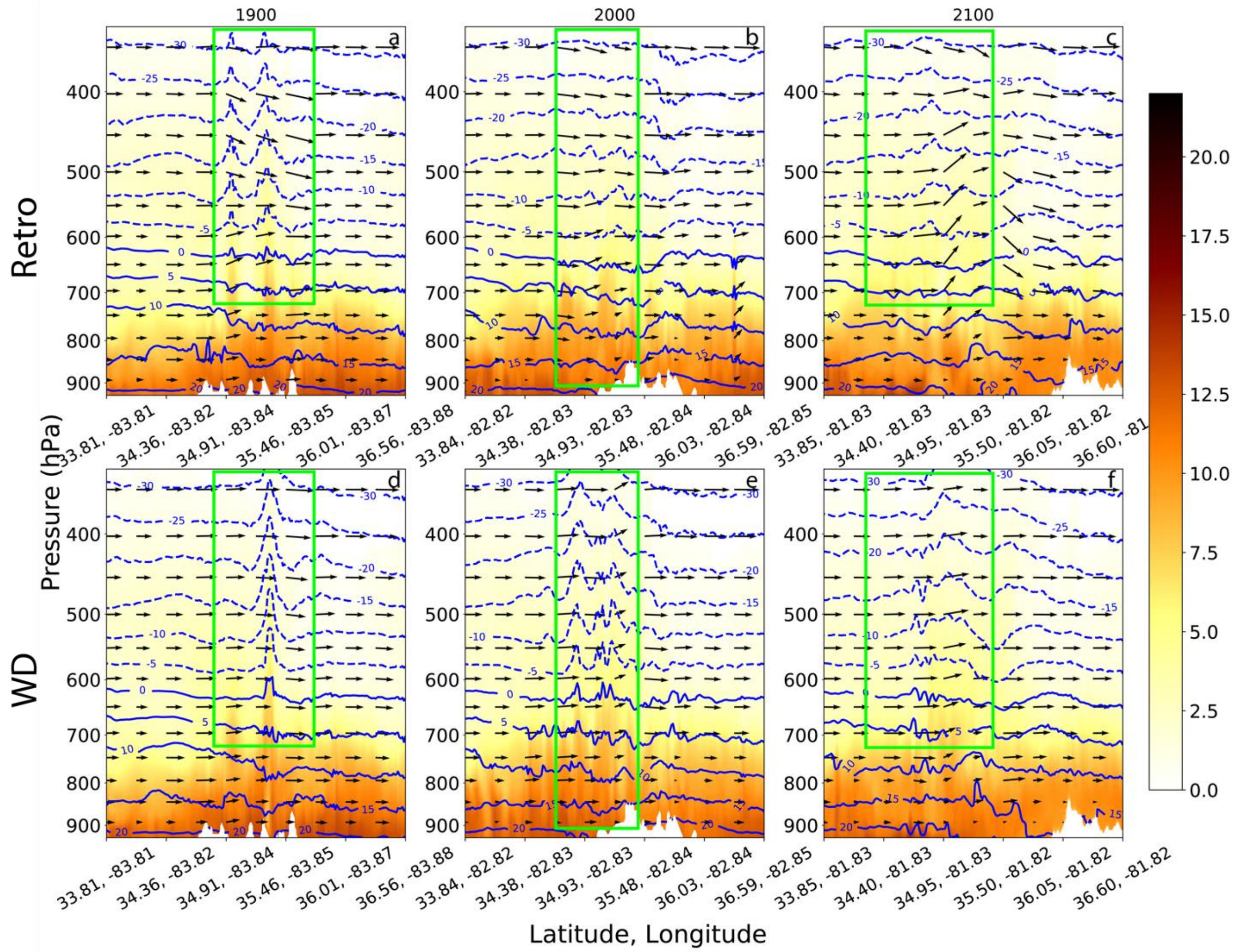

Figure 4.11 As in Fig. 4.10, but for the vertical cross section along the magenta line in Fig. 4.9.

Figure 4.12 shows the difference in the mixed-layer convective available potential energy (MLCAPE) between the Retro and WD experiments. MLCAPE is related to the environmental temperature profile and the path of a rising air parcel. Retro and WD both produce a high MLCAPE over northern South Carolina, which is the reason for the prediction of the primary storms in that area. But significant differences in MLCAPE between Retro and WD are evident in North Carolina in the 1900 UTC analysis (black boxes in Figs. 4.12 a and d) and in Tennessee in the 2000 UTC analysis (black boxes in Figs. 4.12 b and e). High MLCAPE (> 1000 J kg^{-1}) usually allows for low-level atmospheric instability, which may consequently lead to the forecast of spurious convection in both experiments. But MLCAPE in these areas are relatively low for the WD experiment, which causes the alleviation of the spurious convective cells near eastern Tennessee (Fig. 4.12 a vs d, and Fig. 4.12 b vs e) and the boundary between Tennessee North Carolina (Fig. 4.12 a vs f) in WD.

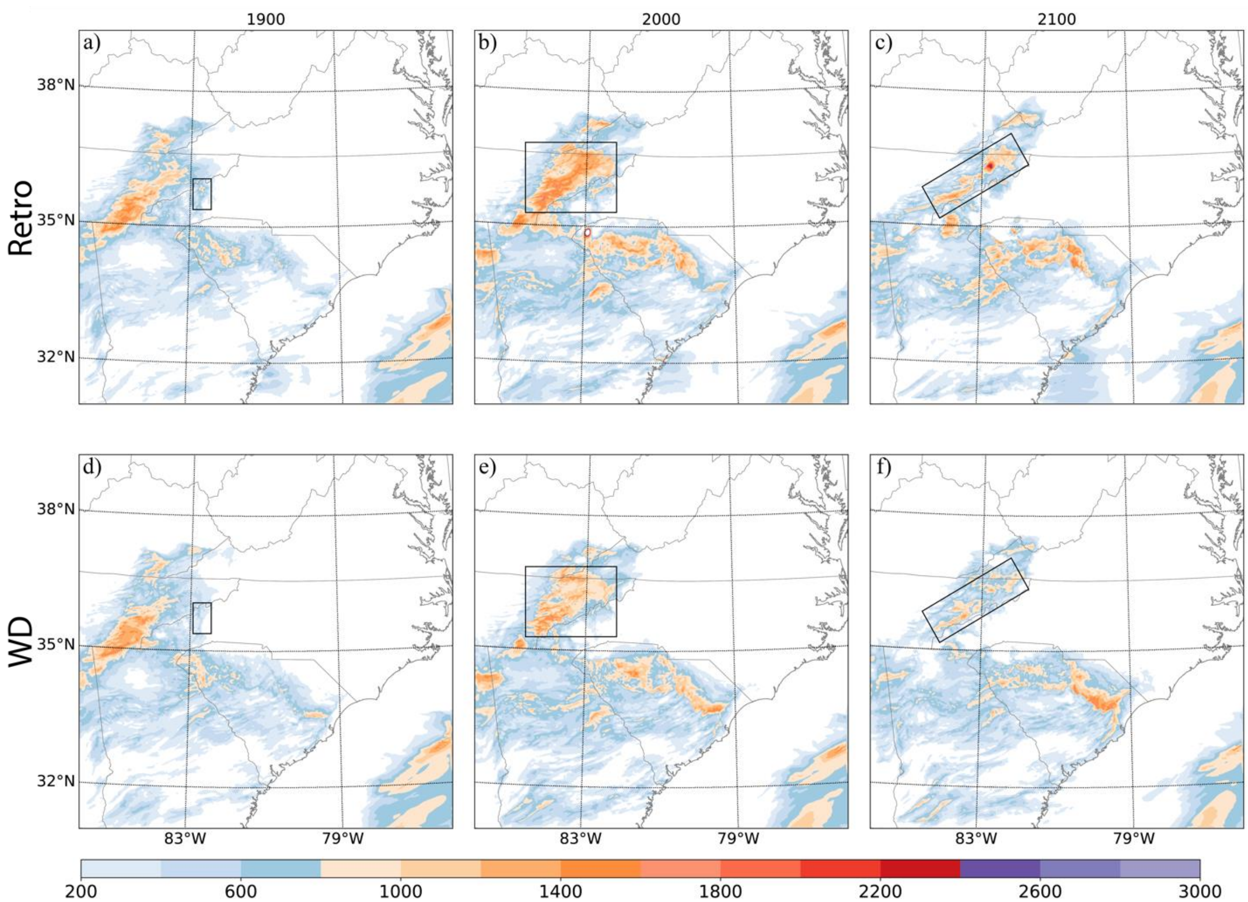

Figure 4.12 The lowest 100-hPa MLCAPE drawn from (a, d) 1900, (b, e) 2000, (c, f) 2100 analyses for the (top) Retro and (bottom) WD experiments.

Figure 4.13 shows the categorical performance diagrams and neighborhood ETS graphs of composite reflectivity forecasts for two thresholds, 20 dBZ and 40 dBZ. By aggregating the score metrics from all available initialization times, the performance of the two experiments, WD and Retro, can be compared in a comprehensive manner. As expected, the values of probability of detection (POD), success rate (SR), critical success index (CSI) and ETS decrease with respect to both the forecast lead time and reflectivity threshold values, whereas the overprediction bias increases monotonically as the threshold increases from 20 to 40 dBZ (Figs. 4.13 a, b). The reason for overprediction in these forecasts is the development of spurious convection in southeastern Kentucky, and their subsequent movement along the boundary between Virginia and North Carolina when they move into western Virginia. In general, the results show that WD, which utilizes weather-dependent background errors, outperforms Retro with larger values of POD, SR, and CSI at all forecast lead times. For the 20 dBZ threshold, WD has at least 0.1 higher CSI values compared to Retro at 1-, 3- and 6-hour forecast lead time. The underprediction bias is observed in WD at 6-hour and in Retro at 3-hour and 6-hour forecasts. For the 40 dBZ threshold, WD also produces higher POD, SR and CSI than Retro, but with slightly higher bias. The aggregated ETS values of the reflectivity forecasts indicate that the WD experiment generally has higher ETS values compared to the Retro experiment for both thresholds throughout the 0-6 hour forecast lead times (Figs. 4.13 c, d). It is important to note that as the lead times increase, the differences in ETS between the two experiments become more obvious. Based on the greater POD, SR, CSI, and ETS, it can be concluded that the 6-hour composite reflectivity forecast from WD has the best overall forecast skill.

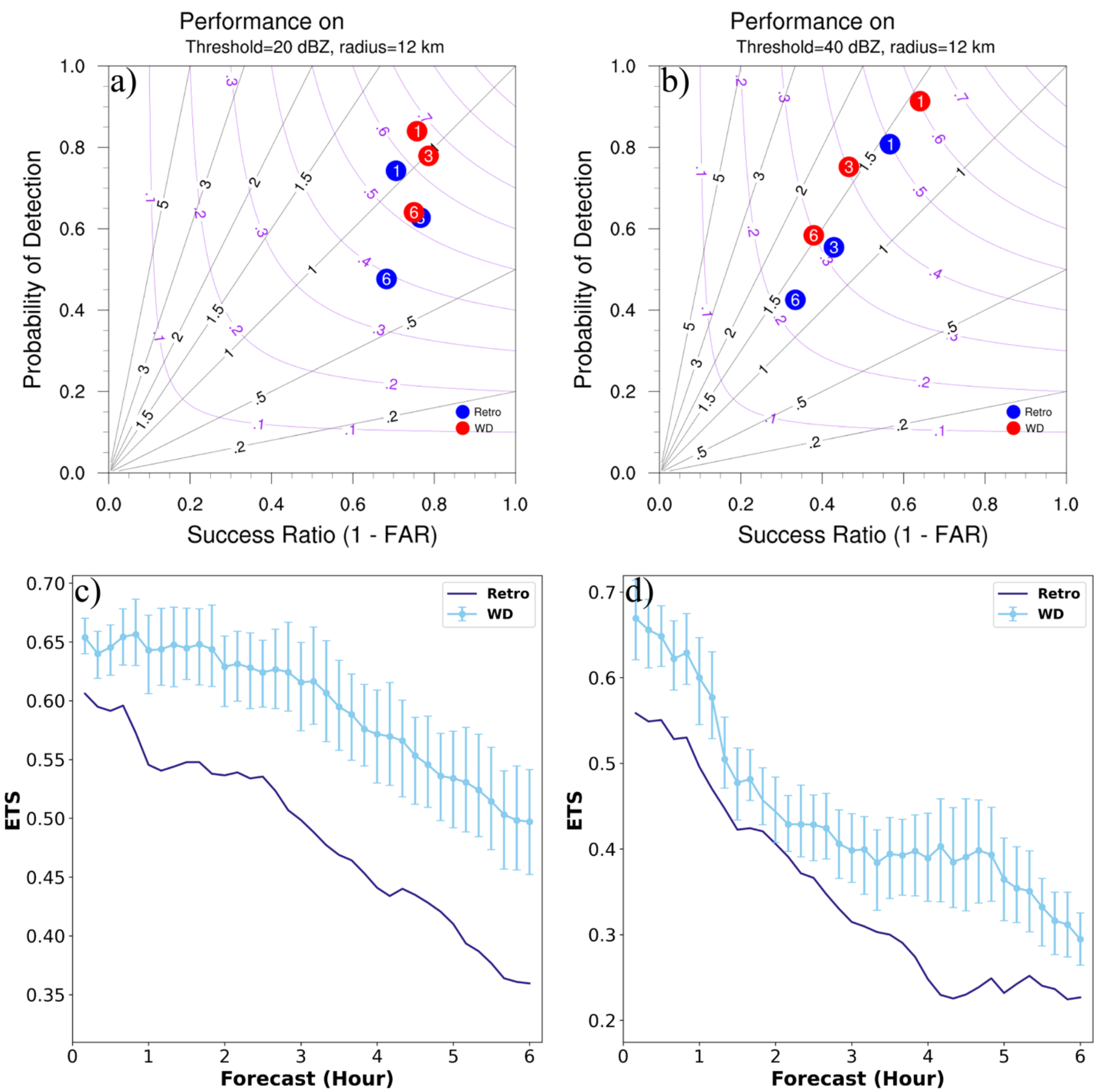

Figure 4.13 Performance diagram and equitable threat scores (ETS) for the 1- to 6-hour Composite reflectivity forecasts aggregated from each initialization hour from 1800 to 0000 UTC the next day for the 5 May 2020 event. Overplotted error bars used to indicate 2.5th and 97.5th percentiles of resampled ETSs of WD in the permutation test, referenced to the forecast from Retro. A dot on the blue lines indicates a statistically significant difference in the ETS between WD and Retro for a specific forecast time.

4.5.2 10 Dec 2021 event

On the evening of 10 Dec 2021, a powerful storm system traveling across the Central United States resulted in several tornadoes, including two EF-4 tornadoes spanning 81.2 miles and 165.7

miles long respectively, and an EF-3 tornado spanning 122.7 miles long south to the EF-4 tornadoes. The EF-4 tornadoes produced strong property loss including over 4000 residences and commercial buildings were destroyed, or damaged and 65 fatalities were confirmed and over 530 people injured in the tornado path. By bringing in cold air from the northern Midwest and mixing it with the moist, nearly record-breaking warm air from the Gulf of Mexico, the atmosphere was extremely unstable. In addition to the atmospheric instability, significant wind shear was also produced by a greater than 25.7 m s^{-1} low level jet that surged northward out of the western Gulf of Mexico and an unusually strong jet stream at 300 hPa with maximum wind speed greater than 72 m s^{-1}. Strong rotating updrafts are more likely to occur when the instability condition and significant wind shear are simultaneously present. As a result, the tornadic supercell that produced the EF-4 tornadoes started to initiate around 2100 UTC and another tornadic supercell that produced EF-3 tornado initiated around 0330 UTC. Our analysis and forecast will mainly focus on the long-lived supercell that produced the EF-4 tornadoes.

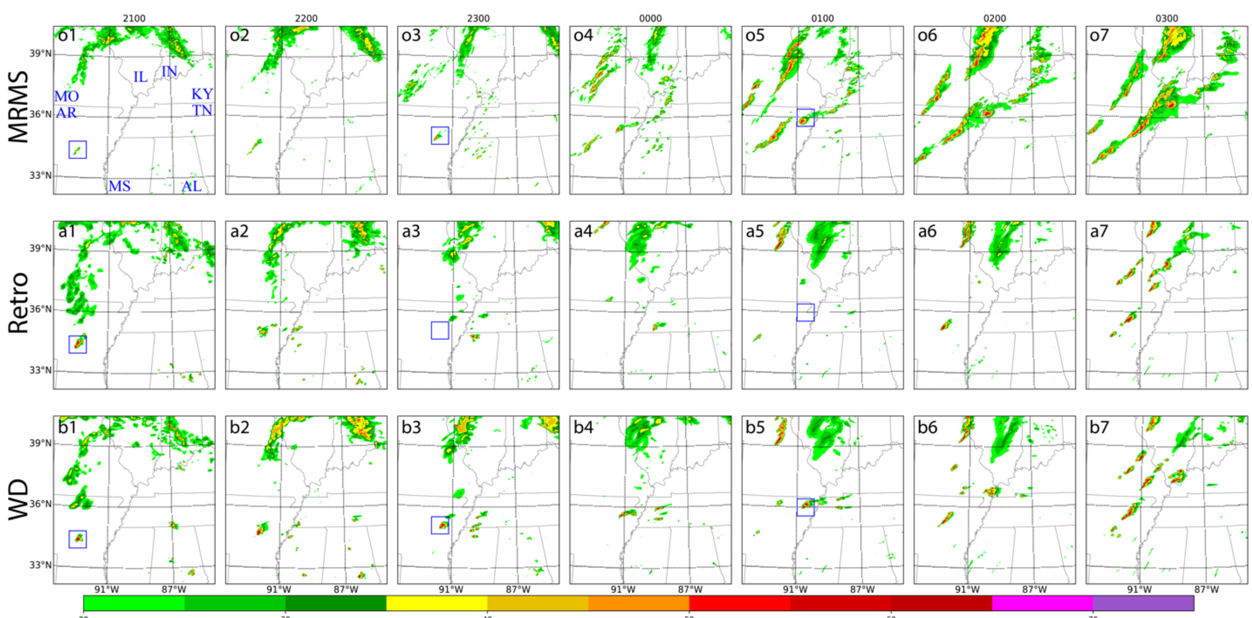

Figure 4.14 As in Fig. 4.7, but for the analysis and subsequent 6-hour forecasts initiated from 2100 UTC 10 Dec 2021.

Similar to the previous event, the benefit of employing weather-dependent background errors in the WoF-Var system is first investigated through analysis and forecast of composite reflectivities (Fig. 4.14). Both experiments show stronger convective activity in comparison to the observations in terms of composite reflectivities. The analyzed tornadic supercell in central Arkansas in Retro and WD have similar shape and position (blue box in Figs. 4.14 a1 and b1) and display a more prominent supercell signature compared to the MRMS composite reflectivities (blue box in Fig. 4.14 o1) at 2100 UTC. This discrepancy between the simulated and observed supercell can be attributed to incomplete radar coverage for this particular supercell. The forecast of the tornadic supercell in Retro completely vanishes since 2300 UTC (Figs. 4.14 a2 - a4 vs o2 - o4). In contrast, the forecast from WD (Figs. b2 - b4) shows the tornadic supercell maintains and propagates northeastward. Although the simulated tornadic supercell in WD is initially stronger than the observed one, it becomes similar to the observation 3 hours after the initialization time (Figs. 4.14 b3 - b7). The predicted supercell in WD moves slightly faster than the observed supercell, which is similar to previous research on the short-term storm-scale NWP (Stratman and Brewster 2015). The simulated result also reveals the inadequacy of both Retro and WD in predicting the development of two quasi-linear convective systems (QLCSs) that started to initiate at approximately 2300 – 0000 UTC (Figs. 4.14 o3 - o4). One spans Missouri and central Illinois and produces several weak tornadoes along with wind damages, while another spans Arkansas, southeastern Missouri, and southern Illinois and finally produces strong EF-3 tornadoes crossing northwestern Tennessee and western Kentucky. It is clear that neither Retro nor WD can predict these two QLCSs properly. Both Retro and WD can only produce isolated convection or much smaller QLCSs near the observed QLCSs at a forecast time later than 0100 UTC (Figs. 4.14 a5 - a7 and b5 – b7). In general, WD outperforms Retro in predicting the development and movement of the tornadic supercell that produced a long-lived tornado several hours in advance, while both experiments display limitations in predicting the initiation and propagation of two QLCSs.

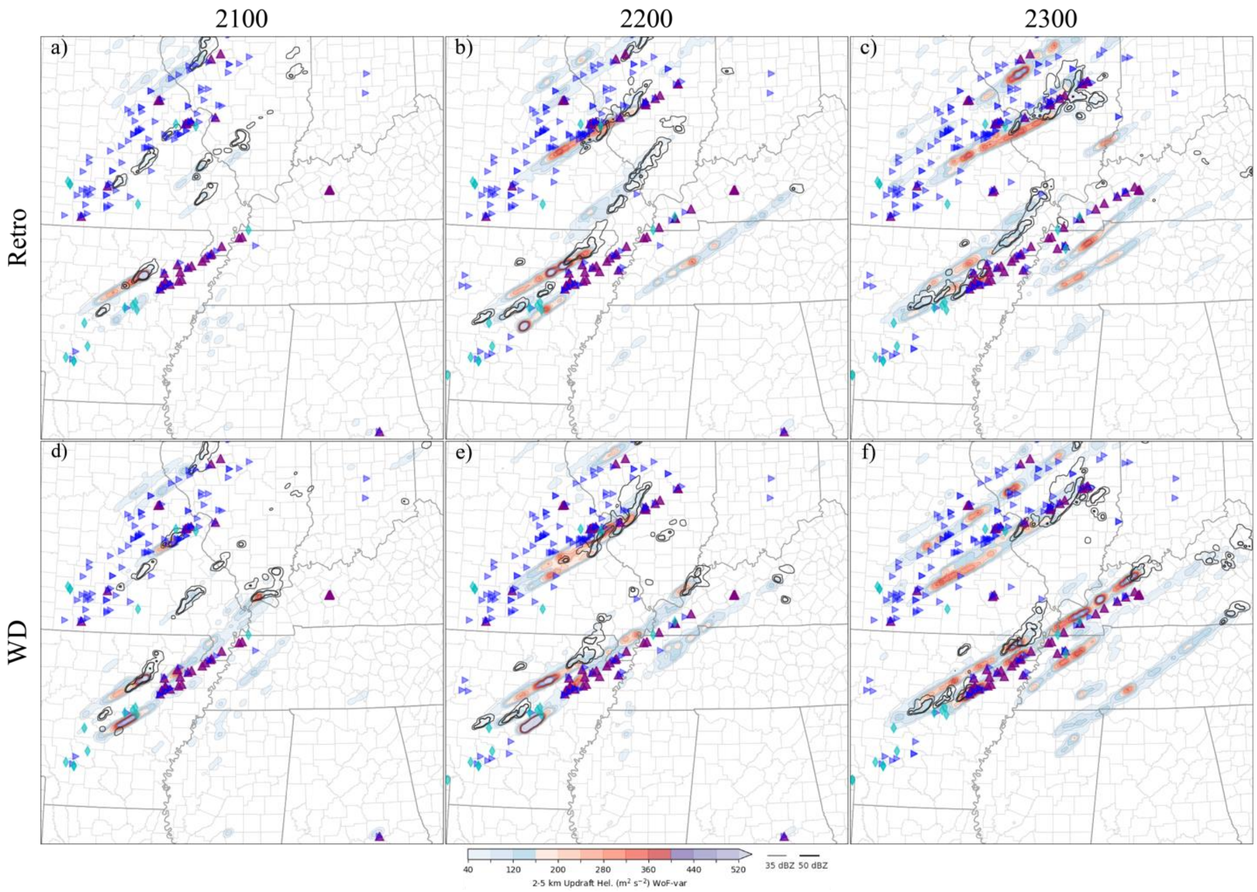

Figure 4.15 As in Fig. 4.8, but for the 6-hour forecasts initiated at (a, d) 2100, (b, e) 2200 and (c, f) 2300 UTC on 10 Dec 2021.

The 2 – 5 km UH tracks for 6-hour forecasts launched from 2100, 2200, and 2300 UTC on 10 Dec 2021, along with NWS/SPC storm reports are shown in Fig. 4.15. When compared to the NWS storm reports, both Retro and WD launched at 2200 UTC (Figs. 4.15 b, e) and 2300 UTC (Figs. 4.15 c, f) are reasonably accurate in predicting the UH tracks along the tornado and wind damage reports generated by the QLCS traversing Missouri and central Illinois. The forecast initialized at 2100 UTC, however, indicate that both experiments generally have a limited capability to generate UH swaths with values exceeding 100 m^2 s^{-2} for this QLCS during the majority of the forecast period, despite that WD offers a short UH track of roughly 300 m^2 s^{-2} in the eastern Missouri near the state.

Leaving aside the UH forecasts for the northern QLCS, the use of weather-dependent background errors in the WoF-Var system shows considerable improvements in the forecast of the tornadic supercell that initialized from Central Arkansas. In the forecast initialized at 2100 UTC (Fig. 4.15 a), Retro only produces UH values of around 100 m^2 s^{-2} at the beginning of the convection active period (roughly the first 1.5 hours into the forecast) but fails to predict the tornadic supercell that produced significant damage and fatalities at later forecast times, as indicated by the NWS storm reports. In contrast, WD predicts a UH track with intensities greater than 100 m^2 s^{-2} for the tornadic supercell for almost the entire forecast period (Fig. 4.15 d). The differences between Retro and WD become more pronounced in forecasts initialized at later times. For instance, in the Retro experiment, the UH track forms initially but quickly dissipates after 1.5 hours into the forecast (Figs. 4.15 b, c), while WD produces relatively good forecasts with even greater UH values reaching 200 m^2 s^{-2} (Figs. 4.15e and f) for the forecasts initialized at 2200 and 2300 UTC. Though WD exhibits northward biases for forecasts launched at both 2200 and 2300 UTC, which is also similar to Stratman and Brewster (2015), these forecasts are accurate than those generated by Retro, closely resembling the UH track crossing Mayfield, Kentucky which was hit hard by an EF-4 tornado. In conclusion, WD with the use of weather-dependent background errors in WoF-Var generates much better forecasts, especially in depicting the long-lived tornadic supercell that caused significant property damage and fatalities.

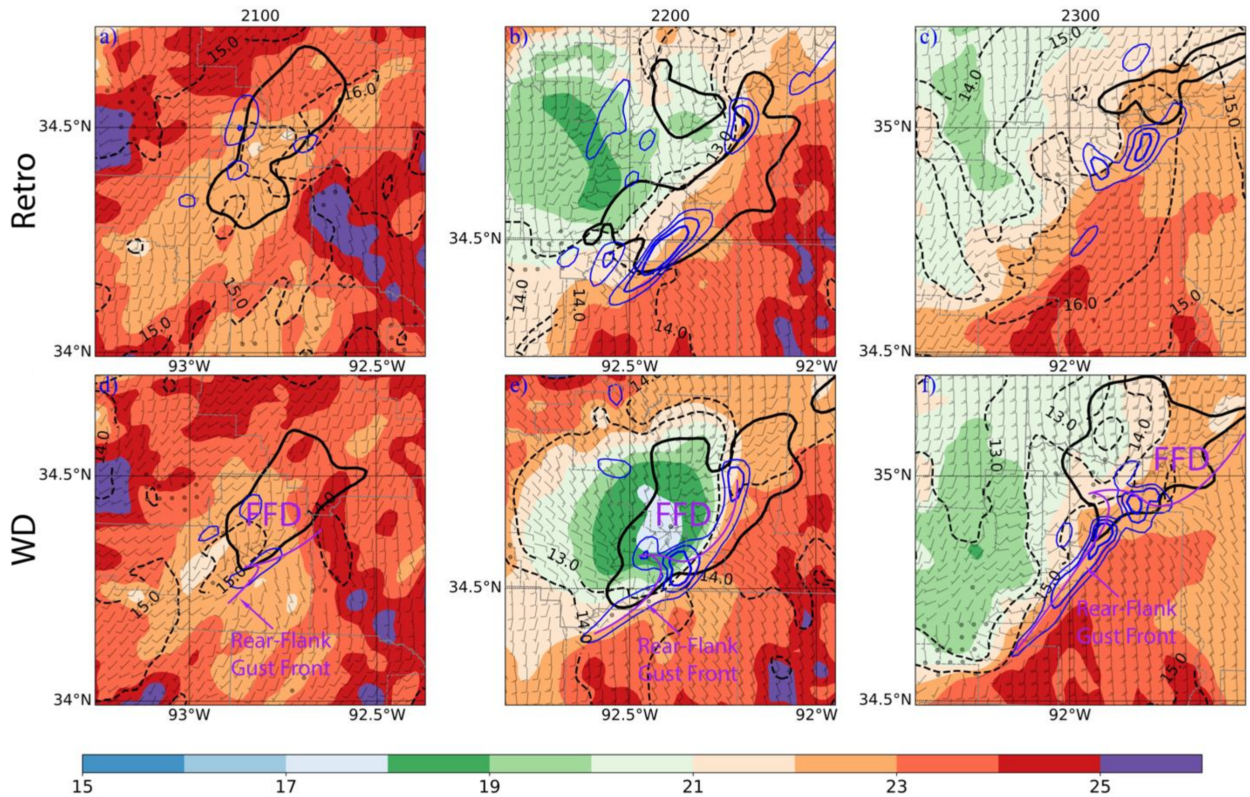

Figure 4.16 The analyzed surface temperature (shade) valid at the analysis time (a, d) 2100, (b, e) 2200 and (c, f) 2300 UTC 10 Dec 2021 for the (top) Retro and (bottom) WD experiments. The 45 dBZ composite reflectivity, surface water vapor mixing ratio and mixed-layer convergence are represented by the black solid lines, black dashed lines and blue solid lines, respectively. The interval of the water vapor mixing ratio and the convergence contours is 1 g kg^{-1} and 200 × 10^{-5} s^{-1}, respectively. The Rear-Flank Gust Front and Forward-Flank Downdraft areas are also displayed in the WD experiments.

In order to understand why the forecasts generated by Retro are worse than those generated by WD, Figure 4.16 consists of wind barbs, surface temperature (shaded), lowest 75-hPa mixed layer convergence contours (blue solid lines), surface water vapor mixing ratio contours (black dashed lines) and the contour for composite reflectivity greater than 45 dBZ (black solid lines) at the analysis times 2100, 2200, and 2300 UTC. The area shown has been zoomed in to an 80 × 80 km^2 domain centered on the site of the simulated tornadic supercell, as approximately indicated by the blue boxes in Figs. 4.14 o1 and o3, but to a smaller extent. At 2100 UTC, the analyzed temperature and water vapor mixing ratio surrounding the convergence area that supports the development of the supercell, indicated by the blue contour centered at the domain (Fig. 4.16 d), are 1 K and 1 g kg^{-1} lower than those in Retro, respectively (Fig. 4.16 a vs d). These modifications may not have a substantial impact on the supercell forecast because there is only a minimal difference in temperature gradient across the rear-flank gust front, which is about 5 K ahead and behind the front in both experiments. The most possible reason for the difference in the forecast between the experiments is due to the presence of weak convergence greater than 200×10^{-5} s^{-1} ahead of the rear-flank gust front (Fig. 4.16 d), as also suggested by the wind barbs.

For the analysis at 2200 UTC, both experiments (Figs. 4.16 b, e) show a strong convergence greater than 600 s^{-1} ahead of the rear-flank gust front. However, WD clearly produces a moderate cold pool in the forward-flank downdraft area, which is believed to contribute to the formation of a well-structured supercell. As suggested in (Khairoutdinov and Randall 2006), the dynamic effect of lifting air in the boundary layer along the outflow above a CIN layer could activate thermodynamic processes and generate free convection at the cold pool boundary regions where gust front occurs. This improvement is not as noticeable in the 2300 UTC analysis, as compared to the 2200 UTC analysis (Fig. 416 f vs e). In comparison to the Retro, WD has a well-organized reflectivity pattern that, together with a weak cold pool just west to the forward-flank downdraft area and a strong convergence ahead of the rear-flank gust front, contributes to the formation of a full, decent supercell structure. Despite the fact that the improvement of the analysis varies over time, WD consistently produces more reasonable analyses, and eventually better 6-hour forecasts compared to Retro at all three times.

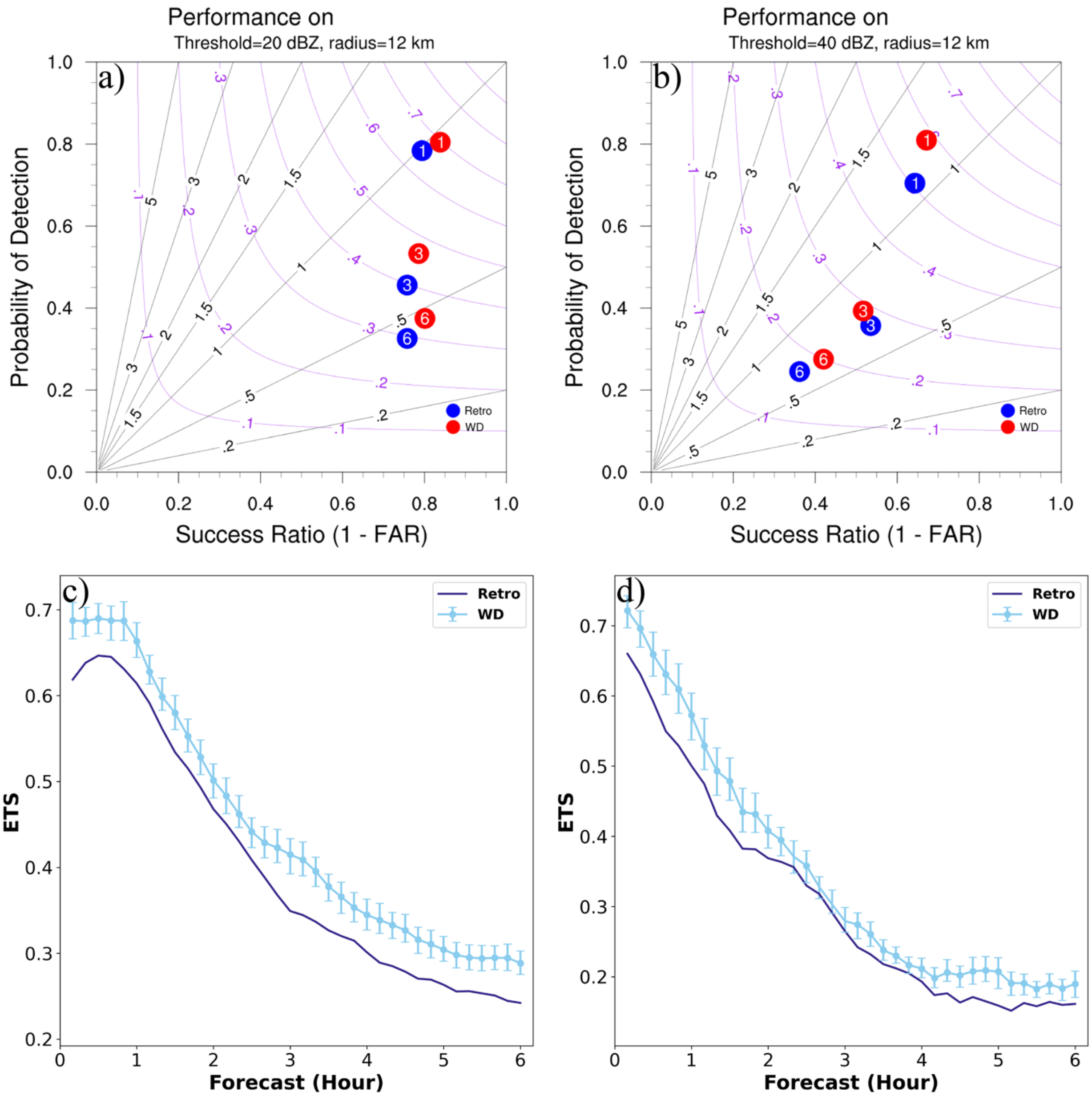

Figure 4.17 As in Fig. 4.13, but for the 1- to 6-hour forecasts initiated from 2100 to 0300 UTC the next day for the 10 Dec 2021 event.

The performance diagram and aggregated ETSs for the 10 Dec 2021 event are shown in Figure 4.17. Different from the 5 May 2020 event, both Retro and WD exhibit noticeable under-forecast performance for forecasts exceeding 3 hours, with both 20 dBZ and 40 dBZ thresholds. For the 20 dBZ threshold, POD and CSI drop as lead time grows, while SR/FAR does not show a significant change (Fig. 4.17 a). On the other hand, for the 40 dBZ threshold (Fig. 4.17 b), all the indices including POD, SR, CSI decrease over time, and both experiments have only slight under-forecast biases for 3- to 6-hour forecasts. The failure of Retro to forecast the main tornadic supercell that produced EF-4 tornadoes demonstrates its underperformance compared to WD at both thresholds and all three forecast lead times (Figs. 4.17 a, b). The aggregated ETS plots also support the conclusion that WD generally outperforms Retro (Figs. 4.17 c, d). Despite the fact that the differences in ETS between two experiments are minor, they are statistically significant as indicated by the blue dots on the lines representing WD. Considering that Retro's forecast completely misses the main tornadic supercell responsible for EF-4 tornadoes, the improvements in WD are most likely due to the increase of "hit". Therefore, we can conclude that by incorporating weather-dependent background errors, the WoF-Var system improves predictability of the hazardous event on 10 Dec 2021.

4.6 Summary and Discussion

In this chapter, weather-dependent background error structures are constructed for clear sky and six other weather classes based on the intensity of radar composite reflectivity derived from the WoF ensemble product. The procedure includes (1) classifying ensemble perturbations according to the strength of convection, (2) filtering outliers, and (3) applying Gaussian smoothing in the vertical direction. The background error statistics are generated using an ensemble of 3-hour forecasts at 3-km horizontal resolution from 25 severe weather events during the HWT SFE 2019. The sampled weather-dependent background error profiles are then integrated into the NSSL WoF-Var system.

The impact of the weather-dependent background error structure on the WoF-Var analyses is first assessed with a single observation experiment. This experiment highlights the heterogeneous nature of the analysis increment for the meridional wind field corresponding to different weather situations. This new analysis and forecast system is then evaluated using two severe weather events: the 5 May 2020 linear complex, and the 10 Dec 2021 long-lived supercell storm, the latter of which produced significant property damage and fatalities. Detailed comparisons of experiments with and without the weather-dependent background errors for the 5 May 2020 linear complex reveal that the new system using weather-dependent errors is able to better analyze the in-storm structure, especially the three-dimensional wind, temperature and humidity. Consequently, the subsequent 6-hour forecasts are significantly improved quantitatively and qualitatively. Similarly, the evaluation of the 10 Dec 2021 supercell suggests improved storm structure analysis and the ability to forecast the long-lived supercell storm several hours in advance using the new system with weather-dependent errors.

The weather-dependent background error covariance matrix proposed in this study is attractive due to its intuitive nature and simplicity. Additionally, it offers an advantage over the traditional 3DVAR method with the background error covariances are dynamic and no longer stationary and isotropic. Another most popular approach to improve static background error covariances is to incorporate ensemble covariances into the variational method. This method is also very attractive because of its flow-dependent nature. In the next chapter, the hybrid strategy (i.e., WoF-Hybrid) will be tested by coupling the above WoF-Var system with the incorporation of ensemble covariances derived from the WoFS ensemble forecasts. In so doing, the analyses and forecasts of high-impact severe weather events are expected to be further improved.

Chapter 5 Implementation of Weather-dependent Background Error within a Hybrid EnVAR System for Improving Short-term Severe Weather Forecast

5.1 Introduction

During the past two and a half decades, significant progress has been made in convective-scale radar DA using variational methods, EnKF methods, and hybrid Variational and EnKF methods. Convective-scale applications of 4DVAR have shown promising results, but their use is typically limited to simple microphysics parameterizations due to the strong nonlinearity and high computational expense involved. In contrast, EnKF, which was first proposed by Evensen (1994), has gained popularity as it shares many advantages of 4DVAR with less sensitivity to nonlinearity. EnKF has shown potential in a variety of applications including radar and satellite DA and high-resolution model applications. For instance, studies by Wilson et al. (2019), Yussouf and Knopfmeier (2019), Zhang et al. (2018, 2019a,b), Jones et al. (2020) have demonstrated its capability. However, for certain high-impact extreme weather events, EnKF suffers deficiencies because of slow model spin-up, as reported by Caya et al. (2005), Kalnay and Yang (2010), and Gao et al. (2022). Additionally, EnKF is computationally expensive when applied to high-resolution applications.

Owing to its efficiency, 3DVAR has been also used during the past two decades to assimilate conventional observations, Doppler radar and other remote sensing observations into high-resolution nonhydrostatic NWP models (Gao et al. 2004; Hu et al. 2006; Xiao et al. 2005; Xiao and Sun 2007; Wattrelot et al. 2014). Among these studies, a 3DVAR system initially developed at the CAPS (Gao et al. 1999, 2004) and then at NSSL (Gao et al. 2013), has been explored in a wide variety of ways and partially introduced in the previous chapters, including 1) assimilating satellite-derived CWP and LPW (Pan et al. 2018, 2021); 2) assimilating pseudo-observations of water vapor mixing ratio derived from vertically integrated liquid water (Lai et al. 2019), and total lightning data provided by the GLM aboard the GOES-R Series (operationally known as GOES-16 and 17; Fierro et al. 2016, 2019; Hu et al. 2020, 2021); 3) incorporating flow-dependent background error covariances estimated from ensemble forecasts (Gao and Stensrud 2014; Gao et al. 2016; Wang et al. 2019). However, the static background covariances used in the system are still similar to those used in the synoptic scale applications and do not consider the intermittent nature of convective flow.

Inspired by the studies like Michel et al. (2011) and Wang and Wang (2021), Chapter 4 introduces a more flexible method to construct static background error covariances for convective-scale DA. In it, the weather-dependent binning approach is adopted to separately construct background error structures for the clear sky and the 6 other weather classes according to the strength of probability matched mean composite reflectivity derived from an ensemble forecast product (i.e., a total of 7 classes). The sampled weather-dependent background error profiles are integrated into the NSSL WoF-Var and are evaluated using two high-impact severe weather events. The results show that the new system improves the analysis of in-storm structure, including dynamic and thermodynamic structure, as well as subsequent 6-hour forecasts.

Hamill and Snyder (2000), Buehner (2005), and Wang et al. (2007, 2008, 2009) showed that the hybrid ensemble-variational method had potential to improve analyses better than either the ensemble-based method or the pure variational method. So a hybrid ensemble and variational DA system was also developed in the NSSL supported by the NOAA Warn-on-Forecast project (WoF; Stensrud et al. 2009, 2013) and it is named as WoF-Hybrid. It is a real-time, deterministic, weather adaptive, dual-resolution analysis and forecasting system. WoF-Hybrid utilizes the flow-dependent covariances from ensemble forecasts provided by the WoFS (Wheatley et al. 2015;

Jones et al. 2016; Skinner et al. 2018; Yussouf et al. 2020a,b), which provides an on-demand, rapidly updating, convection-allowing ensemble forecasts to the end users. In WoF-Hybrid, the total static background covariances are prescribed a weight of 50% of the ensemble covariances and the static error covariances the remaining 50% (Wang et al. 2019). The weather-dependent background error covariances tested in Chapter 4 are attractive due to their intuitive nature and simplicity. The question is how this new type of weather-dependent background error covariances could be combined and used in WoF-Hybrid? Under the hybrid ensemble variational framework, can utilizing the weather-dependent static background error covariances also help improve the short-term severe weather forecasts? In this chapter, we try to answer the above questions by incorporating weather-dependent background error covariances into the WoF-Hybrid system and evaluating their impact on convective-scale analyses and short-term 0- to 6-hour severe weather forecasts with five high-impact severe weather events.

This chapter is organized as follows. Descriptions of the experimental design and the verification methods are detailed in Section 5.2. The results of the evaluation are presented in Section 5.3, which begins with a discussion of the aggregated results over all five selected severe weather events, followed by detailed comparisons and discussions of the results for three selected severe weather events. The conclusions are finally provided in Section 5.4.

5.2 Description of Experiments, Assimilated Observations and Verification Method

5.2.1 Experimental design and assimilated observations

WoF-Hybrid was operated in real-time for roughly 60 days, excluding weekends, during two separate 5-week HWT SFEs occurring between April 27th and May 29th, 2020 and between May 3rd to June 4th, 2021. Five high-impact severe weather events, each of which resulted in fatalities, injuries, or significant economic losses, are selected from the above days for this study. The events occurred on 5 May 2020, 15 May 2020, 22 May 2020, 26 May 2021, and 10 Dec 2021 respectively. The date, the highest risk assigned by NOAA SPC, the number of tornado and hail reports, the storm mode and the verification period for these events are presented in Table 5.1. For each event, the domain covers an area of 601 × 601 grid points with 1.5-km horizontal grid spacing, resulting in a 900 × 900 km^2 area, with its center determined by the Day 1 Convective Outlook from SPC. The domain has 51 vertical levels. The model physics configuration includes the NSSL 2-moment with CCN prediction microphysics scheme (Mansell et al. 2010), the Mellor-Yamada-Janjic planetary boundary layer scheme (Janjić 1994), the Rapid Radiative Transfer Model for general circulation models (RRTMG) longwave and shortwave radiation scheme (Iacono et al. 2008). No cumulus parameterization scheme is employed. The initial background field is generated from 1-h forecasts initialized from 1400 UTC HRRR analyses, and the boundary conditions are generated by using HRRR forecasts launched at 1200 UTC. Each day, analyses are produced at 15-min intervals between 1500 and 0300 UTC the next day. A 6-h forecast was launched at the top of each hour from 1800 to 0300 UTC the next day (Fig. 5.1).

Table 5.1 Events used for testing WoF-Hybrid with their associated highest risk, number of tornado and hail reports, affected states, storm mode and verification period.

Event	SPC Outlook	Tornado Reports	Hail Reports	Affected States	Storm Mode	Verification Period
5 May 2020	Slight	3	25	SC, NC	Mixed	1800 – 0000 UTC
15 May 2020	Enhanced	3	50	OK, TX, AR	Linear	1800 – 0000 UTC
22 May 2020	Enhanced	8	60	OK, TX, AR	MCS	1800 – 0000 UTC
26 May 2021	Moderate	29	122	KS, NE	Supercell	1800 – 0000 UTC
10 Dec 2021	Moderate	97	17	AR, MO, TN, KY, IL	Mixed	2100 – 0300 UTC

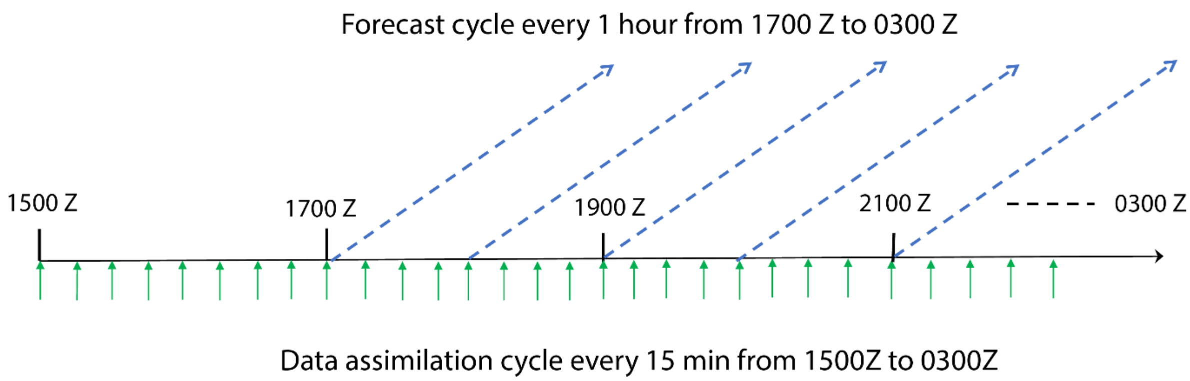

Figure 5.1 The workflow for estimating background error profiles from WoFS ensemble perturbations.

To investigate the impact of incorporating weather-dependent static **B** within WoF-Hybrid on convective-scale analyses and forecasts, two experiments are performed, namely Hyb_retro and Hyb_WD. In the Hyb_retro experiment, the original homogeneous and isotropic static covariances **B** and ensemble covariances **P** are equally blended, as was done in the real-time testing in 2019 and 2020. The Hyb_WD experiment utilizes weather-dependent static covariances **B** and ensemble covariances **P** with equal weighting, similar to Hyb_retro. Table 5.2 provides detailed descriptions of these experiments and their accompanying configurations.

Table 5.2 Experiments used for examining the weather-dependent static B in WoF-Hybrid.

Experiment	Configuration
Hyb_retro	Background error includes 50% of homogeneous static **B** and 50% of ensemble flow-dependent **P**
Hyb_WD	Background error includes 50% of weather-dependent static **B** and 50% of ensemble flow-dependent **P**

Several types of observations were assimilated into the experiments, including data from the U.S. National Weather Service's operational Weather Surveillance Radar – 1988 Doppler (WSR-88D), surface observations and pseudo-water vapor observations. The WSR-88D radar data includes radial velocity, reflectivity and Velocity Azimuth Display (VAD) winds derived from each WSR-88D radar within the analysis and forecast domains. Surface observations were obtained from PREPBUFR and the Oklahoma Mesonet. Pseudo-water vapor observations were derived from reflectivity measurements (Lai et al. 2019) and GOES-16 GLM total lightning observations (Fierro et al. 2019). All these observations were assimilated every 15 minutes.

5.2.2 Verification method

The performance of WoF-Hybrid is evaluated through a combination of qualitative and quantitative methods, utilizing several observation datasets. To assess the accuracy of the 0 – 6-h composite reflectivity forecasts, NSSL's MRMS composite reflectivity (CREF) product is used. Meanwhile, the skills for the 2 – 5 km updraft helicity (UH, Kain et al. 2008) forecasts are evaluated using NSSL's MRMS midlevel azimuthal shear (AzShear) product and NWS/SPC severe weather reports. Both MRMS CREF and AzShear products are interpolated onto the 1.5-km model grid before the verification process.

In order to quantitatively evaluate the accuracy of short-term severe weather forecasts, a series of metrics are used. These metrics are calculated for the CREF and 30-min interval 2 – 5 km UH forecasts, against MRMS CREF and midlevel AzShear products. The scores are calculated using neighborhood radii of 9, 12, and 15 km, respectively. The thresholds for CREF are set at 20 and 40 dBZ. Since UH and AzShear are two distinct measurements, the thresholds for UH and midlevel AzShear are separately determined by considering the 95th and 99.95th percentile values in the real-time WoFS ensemble forecasts and the MRMS AzShear dataset (Skinner et al. 2018) of 19 severe weather events in 2020 and 2021. Therefore, the thresholds for UH are set at 65.79 and 80.0 $m^2\ s^{-2}$; and the thresholds for AzShear are set at 0.00206 and 0.00215 s^{-1} for the 95th and 99.95th percentile respectively. Using these neighborhood radii and thresholds, contingency-table based metrics (Roebber 2009) such as the probability of detection (POD), false alarm ratio (FAR), success ratio (SR = 1 – FAR), frequency bias (BIAS), and critical success index (CSI) are calculated. The closer the values of POD, SR, BIAS, and CSI approach unity, the better forecast is achieved. Therefore, these metrics can be put into one performance diagram plot, with the perfect forecast located at the upper-right corner of the diagram. Finally, a neighborhood ETS is calculated based on the contingency table. A permutation test (Hamill 1999) is conducted as well to further examine if the variations in scores of the experiments with the use of weather-dependent static **B** in WoF-Hybrid are statistically significant.

In addition, UH swaths aggregated over a 0- to 6-hour period with a 10-minute output interval are subjectively compared against the NWS/SPC severe weather reports, which document the occurrence of hazardous weather events across the United States, including tornadoes, wind gusts of 26 m s^{-1} or greater, and hail with a diameter of 25 mm or larger, to verify the forecast qualitatively. Despite some biases in these reports (Trapp et al. 2006; Potvin et al. 2019), NWS/SPC severe weather reports are still a commonly used resource for qualitative forecast verification.

5.3 Results

5.3.1 Aggregate forecast performance

To gain an in-depth understanding of the forecasting capabilities of reflectivity and rotation track in WoF-Hybrid, which incorporates weather-dependent static **B**, various evaluation metrics such as POD, FAR/SR, CSI, BIAS, and ETS are calculated for both CREF and UH. These metrics are aggregated from forecasts initialized at hourly intervals between 2100 UTC and 0300 UTC for the event on 10 Dec 2021, as well as between 1800 UTC and 0000 UTC for the remaining four events. The reason why the aggregate period for the event on 10 Dec 2021 is different from that for other events is because severe weather was active during a different time period for this event.

Performance diagrams with varying thresholds and neighborhood radii are used to provide a clear and concise representation of forecast ability for CREF and 30-min midlevel rotation track. These diagrams, shown in Figure 5.2, are created by combining POD, FAR/SR, CSI, and BIAS into one graph.

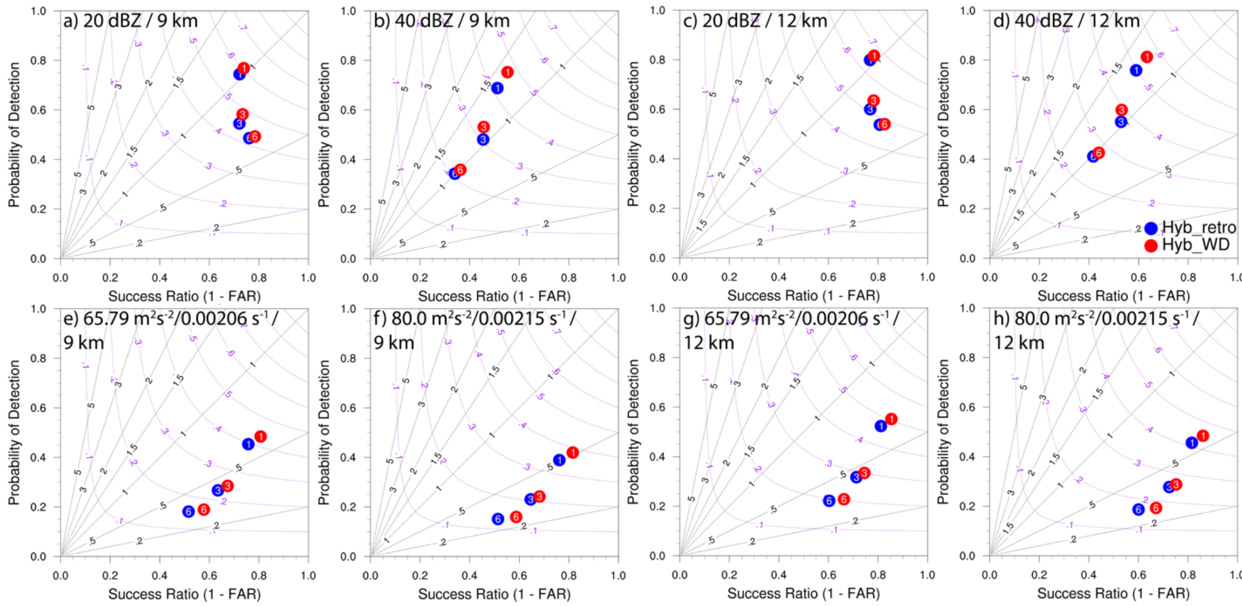

Figure 5.2 Performance diagrams for 1- to 6-hour composite reflectivity (CREF) and midlevel rotation track forecasts, aggregated from each initialization hour between 2100 and 0300 UTC the next day for 10 Dec 2021, and between 1800 and 0000 UTC the next day across the other 4 events. The numbers within the colored circles indicate the forecast hour, i.e., 1-, 3- and 6-hour forecasts. Top row shows the performance diagrams for CREF forecasts relative to MRMS CREF observations using thresholds of (a, c) 20 dBZ and (b, d) 40 dBZ. Bottom row shows the performance diagrams for middle-level UH forecasts relative to MRMS Azimuthal Shear using thresholds of (e, g) 65.79 m^2 s^{-2} / 0.00206 s^{-1} and (f, h) 80.0 m^2 s^{-2} / 0.00215 s^{-1}. The experiments Hyb_retro and Hyb_WD are represented by blue and red circles, respectively. The figure only displays the results for neighborhood radii of 9 km and 12 km.

As is frequently seen in earlier studies that evaluated short-term forecasts, the forecasting skill is observed to degrade as lead time increases (Fig. 5.2). This degradation is manifest in a general trend of decreasing POD and CSI values, alongside an escalation in FAR (or a decline in SR), and is more pronounced for higher CREF thresholds. However, it is less so for midlevel rotating updrafts, or UH. An increase in threshold values (for CREF from 20 to 40 dBZ; for UH from 65.79 to 80 m^2 s^{-2}) or a decrease in neighborhood radius (from 12 km to 9 km), results in decreased values for POD, SR and CSI. This trend occurs because, as threshold values increase or less grid points are considered within a certain neighborhood, the criteria for correct forecasts become more stringent, resulting in a greater penalty for underestimating strong convection or displacement of ongoing convections. Underprediction, represented by a bias value of less than 1.0, is consistently observed in 30-min interval midlevel rotation track forecast at all threshold values (Figs. 5.2 e - h), and reflectivity forecasts at 20 dBZ threshold (Figs. 5. 2 a, c) with the exception of the 1-hour CREF forecasts. In contrast, overestimation, represented by a bias value greater than 1.0, is only observed in 1-hour CREF forecasts with the 40 dBZ threshold (Fig. 5.2 b, d). Furthermore, the characteristic of overestimation or underestimation for a specific forecast does not significantly change when the neighborhood radius is varied from 9 km to 12 km. This suggests that the forecast model may have difficulties in accurately simulating the processes associated with rotating updrafts and stratiform precipitation.

The experiment Hyb_WD consistently demonstrates superior performance in comparison with the experiment Hyb_retro, as evidenced by metrics POD, FAR/SR, and CSI. While the superiority of Hyb_WD over Hyb_retro diminishes with increasing lead time for CREF forecasts (e.g., Figs. 5.2 b, d), the margin of superiority for rotation track forecasts remains evident for longer lead times (Figs. 5.2 e - h). It is worth noting that the improvement in WoF-Hybrid from incorporating weather-dependent static **B** does not affect the level of overestimation or underestimation for either CREF or rotation track forecasts. In general, the experiment Hyb_WD exhibits better forecasting skills for both CREF, and rotation track forecasts compared to Hyb_retro, with a similar bias value. The improvement is slightly more significant in early CREF and later rotation track forecasts.

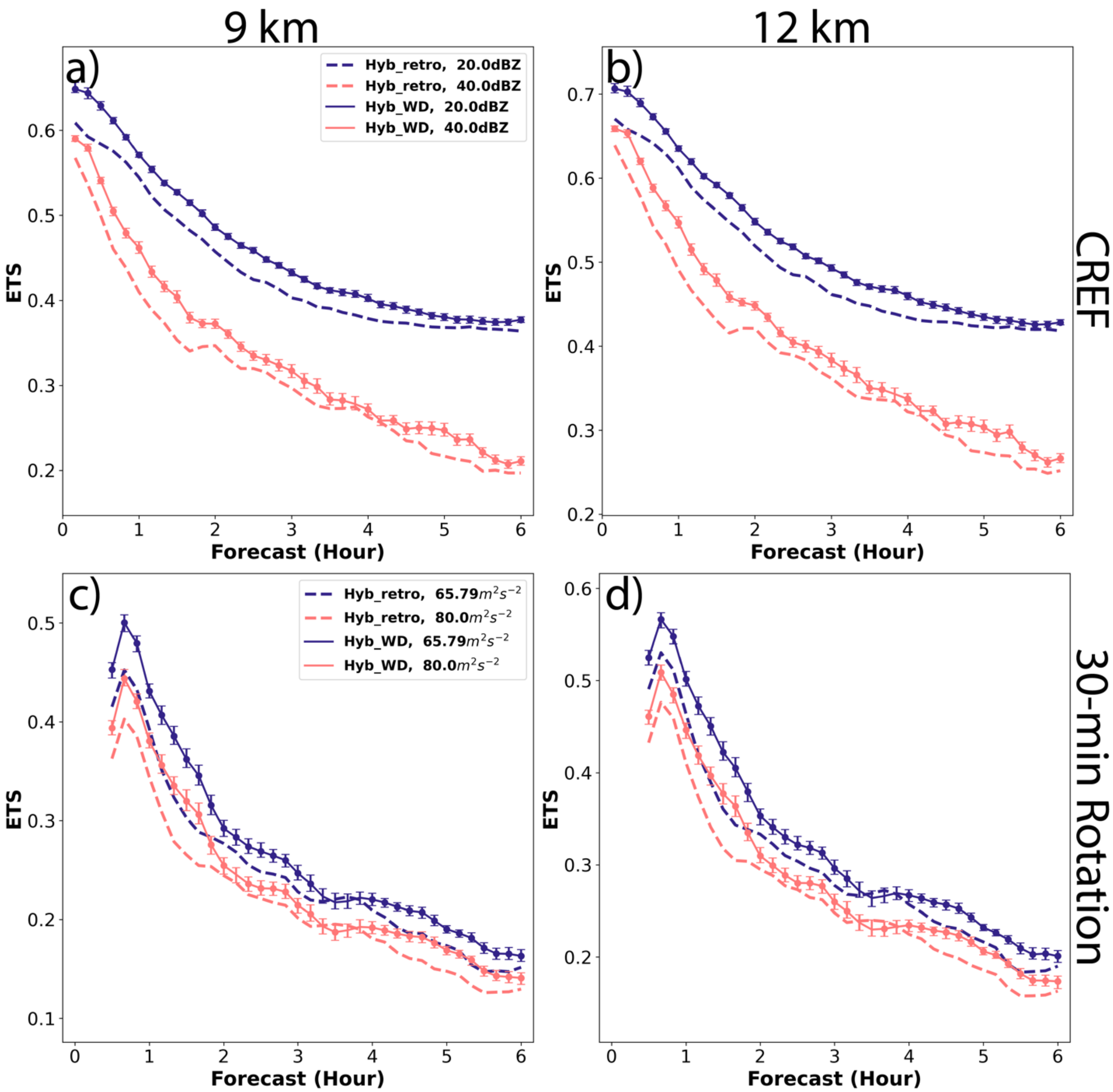

Figure 5.3 ETS for 0- to 6-hour CREF forecasts (top row), and 30-min midlevel rotation track forecasts (bottom row) aggregated from each initialization hour between 2100 and 0300 UTC the next day for 10 Dec 2021, and between 1800 and 0000 UTC the next day across the other 4 events. Top row displays the ETS for CREF forecasts with a neighborhood radius of (a) 9 km and (b) 12 km, and at thresholds of 20 dBZ (blue), (red) and 40 dBZ (red). Bottom row displays the ETS for 30-min midlevel rotation track forecasts with neighborhood radius of (c) 9 km and (d) 12 km at thresholds of (blue) 65.79 $m^2 s^{-2}$ / 0.00206 s^{-1} and (red) 80.0 $m^2 s^{-2}$ / 0.00215 s^{-1}. The experiments Hyb_retro and Hyb_WD are represented by dashed and solid lines, respectively, with dots on each solid line indicating that the changes of Hyb_WD are statistically significant in comparison to Hyb_retro.

The aggregated ETSs clearly exhibit a uniform decline trend as lead time increases (Fig. 5.3), similar to the performance diagrams. The decrease rate of rotation track forecast skill (Figs. 5.3 c - d) is comparable to that of CREF forecast skill using a 40 dBZ threshold (Fig. 5.3 b) and apparently faster than that of CREF forecast skill using a 20 dBZ threshold (Figs. 5.3 a). This agrees with the commonly observed phenomenon that high reflectivity is often indicative of a strong rotating updraft. Similar to the performance diagram, higher ETSs are noticed for both reflectivity and rotation track forecasts when using a larger neighborhood radius or lower verification thresholds, which is the result of an increase in POD coupled with a decrease in FAR due to stricter criteria for correct forecasts. Hyb_WD consistently outperforms Hyb_retro for both CREF and rotation track forecasts, regardless of the neighborhood radii or thresholds used. For CREF using lower thresholds, the improvements of Hyb_WD are more obvious within the first 4 hours (Figs. 5.3 a - b). The improvements in Hyb_WD are more pronounced within the first and last 2 hours for rotation track forecast (Figs. 5.3 c - d) and the first 2 hours for CREF forecast using a 40 dBZ threshold (Figs. 5.3 a - c). These results demonstrate that the use of weather-dependent static **B** instead of homogeneous static **B** in WoF-Hybrid has the potential to improve convective-scale forecasts.

In order to analyze forecast skills with greater specificity, contingency tables are aggregated and calculated over seven 6-hour forecasts for each individual event. As previously mentioned, forecasts initialized from 2100 to 0300 UTC the next day are aggregated for the event on 10 Dec 2021, while those initialized from 1800 to 0000 UTC the next day are aggregated for the other four events. Here, only results obtained using a neighborhood radius of 12 km are presented as an example, and similar results are found with other neighborhood radii. Despite variations in forecast skills among these different events for both CREF (Fig. 5.4) and rotation track (Fig. 5.5) forecasts, the scores for Hyb_WD are generally higher than Hyb_retro among all events. Furthermore, the forecast skills are found to exhibit more variability at longer forecast times (3-hour and 6-hour) and less variability at 1-hour forecasts. For CREF, an underpredicted bias exists for the 20 dBZ threshold for most events at 3-hour and 6-hour forecasts, while an overpredicted bias exists for the 40 dBZ threshold for all forecasts (Fig. 5.4). For rotation forecasts, an underpredicted bias exists for all events (Fig. 5.5).

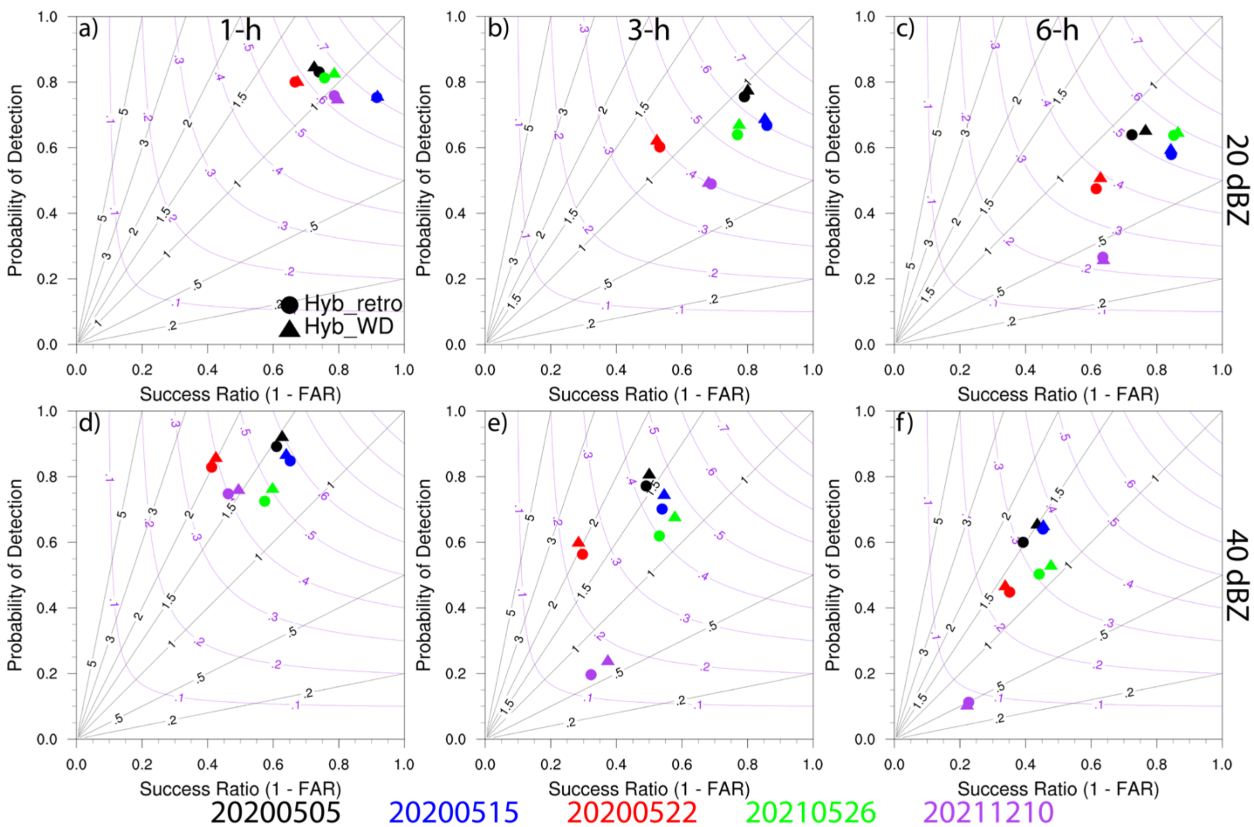

Figure 5.4 Performance diagrams for (first column) 1-, (second column) 3-, (third column) 6-hour CREF forecasts from each case relative to MRMS CREF observations at thresholds of (a-c) 20 dBZ, and (d-f) 40 dBZ. The 5 May 2020, 15 May 2020, 22 May 2020, 26 May 2021, and 10 Dec 2021 events are represented by black, blue, red, green and purple symbols, respectively. The circles and triangles indicate Hyb_retro and Hyb_WD, respectively. Scores are provided for a neighborhood radius of 12 km.

The CREF and rotation track forecasts for the event on 10 Dec 2021 (purple symbols in Figs. 5.4 and 5.5) exhibit the worst performance across all thresholds, as a result of a very complicated outbreak of tornadoes, hail and damage winds in several states during that day, which was very difficult to be handled by the WRF model. The 3-hour forecasts of CREF (Fig. 5.4 e) with a 40 dBZ threshold and rotation track with both thresholds (Fig. 5.5 b, e) show the only improvement with the use of weather-dependent static **B**, which can be attributed to an improved forecast of the supercell that produced EF-4 tornadoes in Hyb_WD.

Relatively low skill scores and high bias are observed for the CREF forecasts longer than 3 hours for 22 May 2020 for both reflectivity thresholds with both Hyb_retro and Hyb_WD experiments (Fig. 5.4). This is primarily due to the fact that the small mesoscale convective system (MCS) in southeastern Oklahoma failed to develop upscale into a large zonal MCS, as observed (details will be discussed in the next section). On the other hand, both experiments fairly portray the small MCS crossing the Oklahoma – Texas border owing to a relatively better and unbiased performance of rotation track forecasts in the first 3-hour forecast, leading to higher POD scores for this particular event over all other events (red symbols in Fig. 5.5).

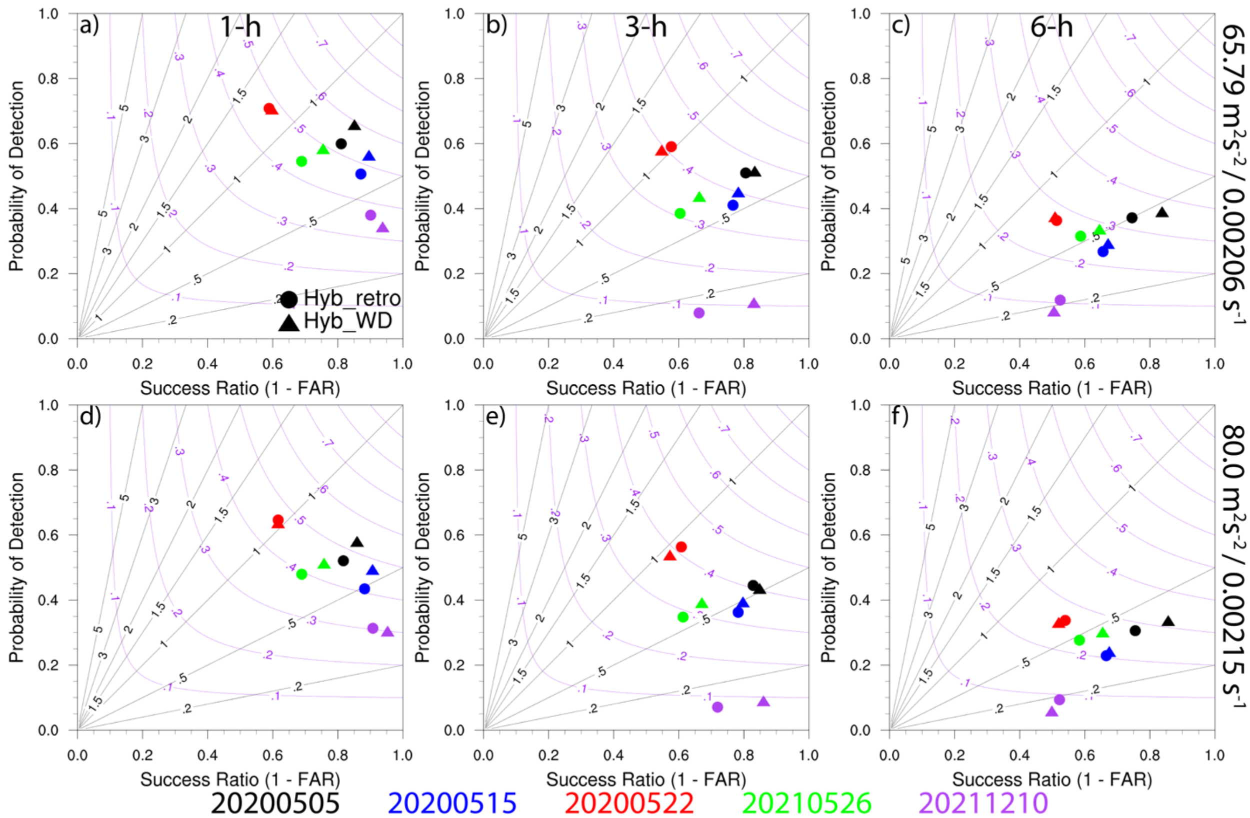

Figure 5.5 As in Fig. 5.4, but for midlevel updraft helicity forecasts relative to MRMS midlevel Azimuthal Shear observations at thresholds of (a-c) 65.79 m^2 s^{-2} / 0.00206 s^{-1} and (d-f) 80.0 m^2 s^{-2} / 0.00215 s^{-1}.

Forecasts for 5 May 2020, 15 May 2020, and 26 May 2021, generally show the highest POD and lowest FAR for CREF forecasts from the beginning of the forecast (Fig. 5.4). However, excessive misplacement of rotations associated with strong convection leads to relatively low POD and high FAR for rotation track forecasts throughout the forecast period (Fig. 5.5). Figure 5.5 also illustrates the neutral to positive impact of the implementation of weather-dependent static **B** in WoF-Hybrid on rotation track forecasts. Drawing from the aforementioned discussion, the insights will be further discussed regarding the following three events: the 5th of May, 2020 event is used as a representative of the highest-performing event, the 22nd of May, 2020 is included as an example of an event with average performance, and the 10th of December, 2021 is considered the event with the lowest performance.

5.3.2 The 5 May 2020 severe weather event in South Carolina

On 5 May 2020, a weak front was moving from the west to the east across the region spanning from northern Georgia to northern South Carolina, resulting in the development of a mesoscale convective complex. The airmass along the front originated from the Gulf of Mexico and was characterized by temperatures in the 70s F and dewpoint temperatures in the mid 60s F near the surface, creating a highly unstable atmosphere with a mixed-layer Convective Available Potential Energy (CAPE) value ranging from 1000 to 2000 J kg^{-1}. This environmental setup, with the high CAPE value, was conducive to the formation and sustenance of convective systems. As a result, the formation of the mesoscale convective complex initiated around 1900 UTC and steadily drifted eastward, eventually reaching the eastern coast of North Carolina at 0300 UTC.

To provide a general picture of the performance of two experiments, a qualitative analysis of the 0- to 6-hour forecast is conducted by comparing the 2 – 5 km UH track forecasts from both experiments Hyb_retro and Hyb_WD, initialized at 1900, 2100, and 2300 UTC with the overlaid NWS/SPC severe weather reports. The results from the forecast initialized at 1900 UTC suggest that both experiments are able to present the west-east oriented convective system and are in reasonable agreement with the NWS/SPC storm reports, more specifically for the tornado reports near the borders of South Carolina and North Carolina as well as for the subsequent hail reports (Figs. 5.6 a, b). Although the UH track forecasts for both experiments are similar at 1900 UTC, improvements can be observed in Hyb_WD for the forecast initialized at 2100 UTC due to the better estimation of initial conditions through additional assimilation cycles (Fig. 5.6 e vs b).

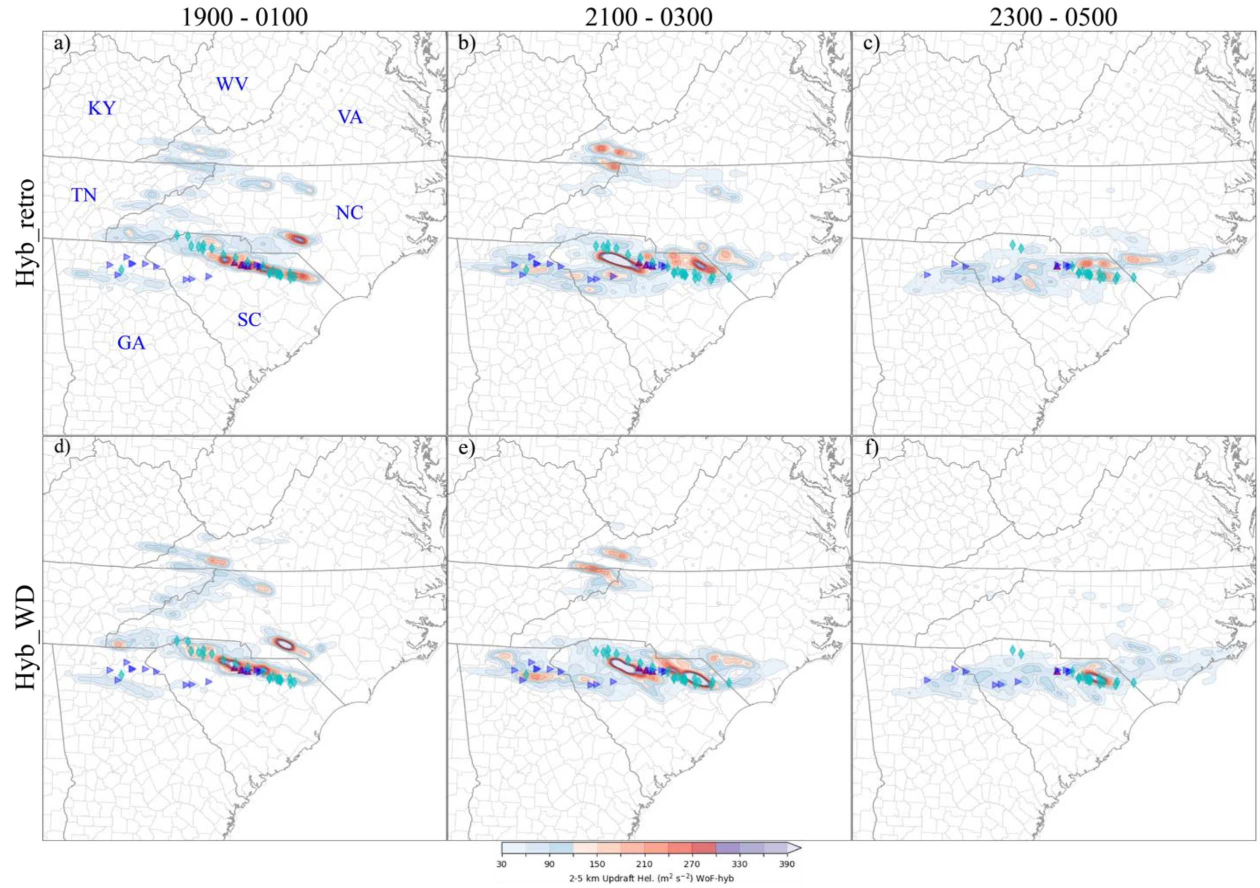

Figure 5.6 The 6-hour 2 – 5 km UH track forecasts initialized at (a, d) 1900, (b, e) 2100, and (c, f) 2300 UTC on 5 May 2020. The top row presents the results of the experiment Hyb_retro, while the bottom row presents the results of the experiment Hyb_WD. During each forecast period for each experiment, the forecast results are superimposed with the SPC storm reports, indicated by purple upward-pointing triangles for tornado events, cyan diamonds for hail events, and blue right-pointing triangles for damaging wind events.

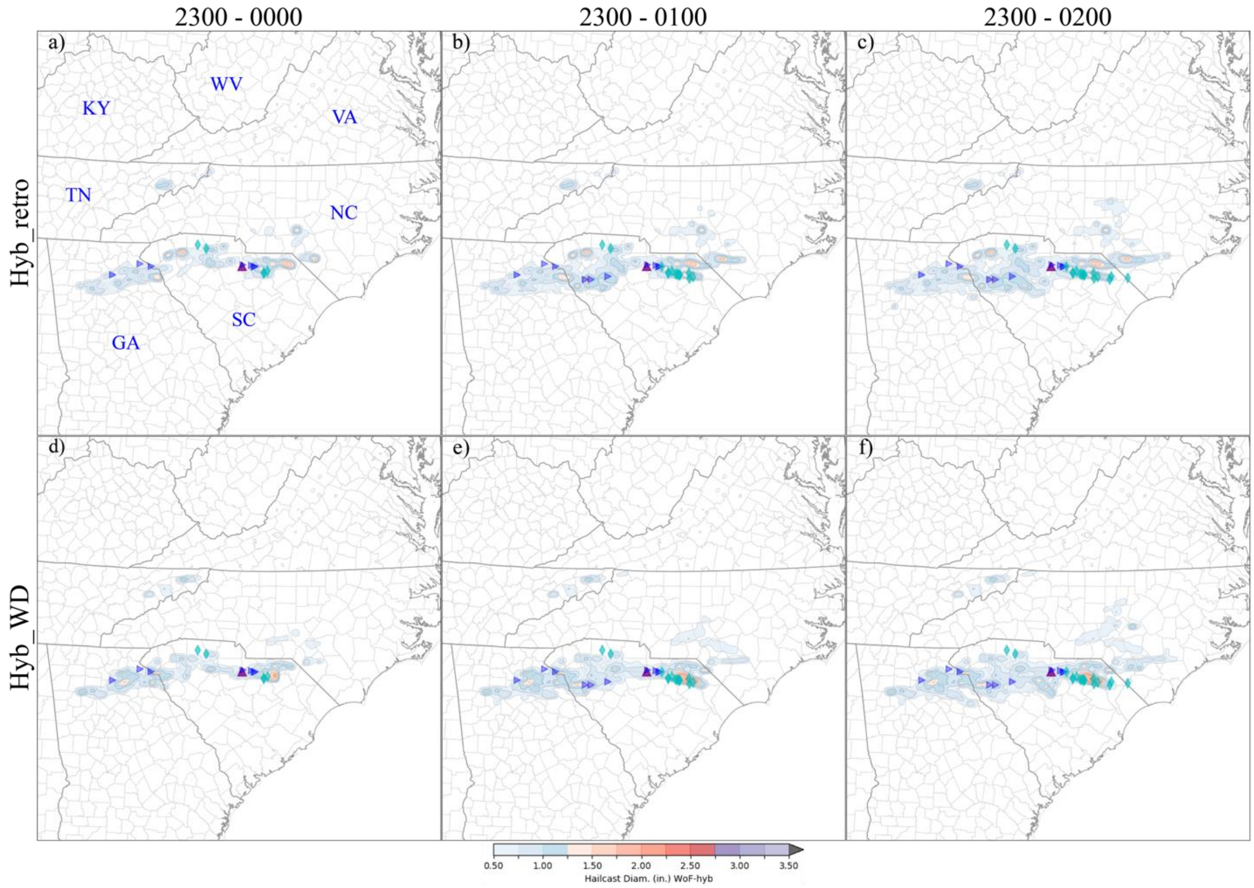

Figure 5.7 Hailcast superimposed with the SPC storm reports for (a, d) 1-, (b, e) 2-, and (c, f) 3-h forecasts initialized at 2300 UTC on 5 May 2020. The top row presents the results of the experiment Hyb_retro, while the bottom row presents the results of the experiment Hyb_WD.

Compared to the forecasts initialized at 2100 UTC, both experiments demonstrate an enhancement of the UH track intensity for hail events that occurred prior to and later than the time when the tornadoes touchdown, with Hyb_WD producing a higher UH value in northern South Carolina. However, a slight north bias in the UH track is observed in Hyb_retro compared to the NWS/SPC hail reports, and a better UH track is predicted by Hyb_WD. Additionally, Hyb_WD produces more reasonable UH values in northeastern Georgia state. In the forecast initialized from 2300 UTC, both experiments fail to predict the tornado events. Nevertheless, Hyb_WD shows a much stronger UH track, which corresponds well with the overlaid hail reports at the last moment of the active period of the convective system. Specially, the superiority of Hyb_WD in predicting hail growth can be more detailed in Figure 5.7. It is apparent that the hailcast swaths generated by both experiments align with the hail occurrence to some extent. According to NWS/SPC storm reports during this day, hail events after the tornado touchdown (2350 - 0030 UTC) generated hail with diameters greater than 2 to 3 inches. For the hail growth forecasts, it is evident that Hyb_WD more accurately predicts the hail formation with such diameters at the exact locations (Figs. 5.7 d - f), whereas Hyb_retro predicts hail with a diameter not exceeding 1.5 inches and has location errors.

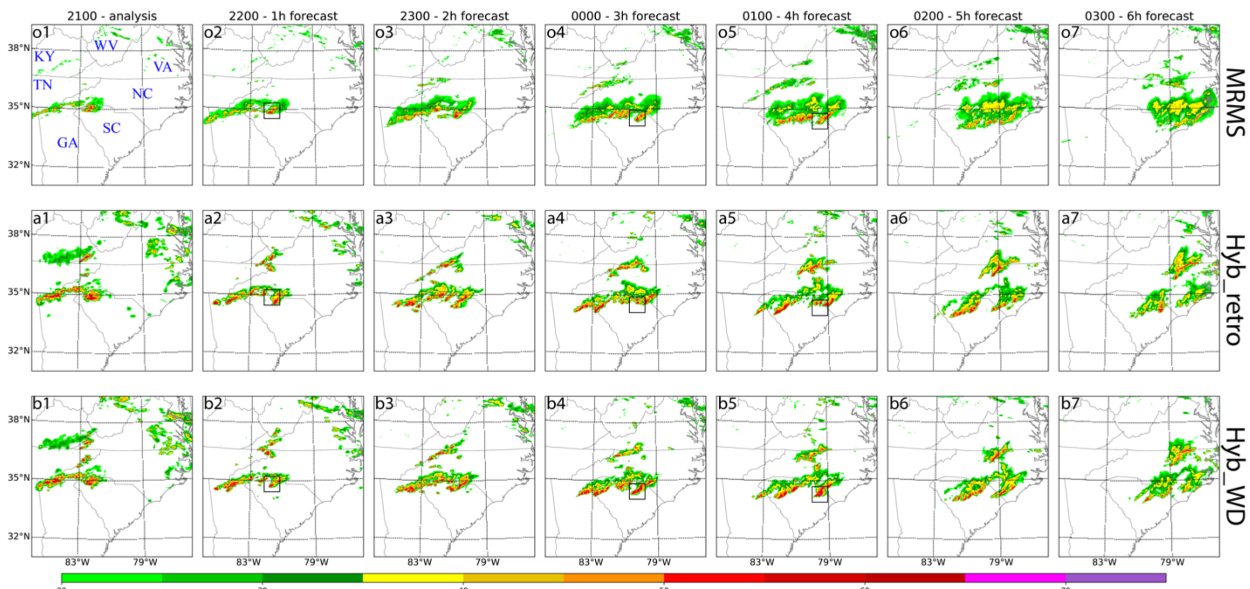

Figure 5.8 (from top to bottom) The (o) observed composite reflectivity from MRMS and the predicted composite reflectivity for (a) Retro, and (b) WD. The (1) analysis at 2100 UTC 5 May 2020. The (2) 1-, (3) 2-, (4) 3-, (5) 4-, (6) 5-, and (7) 6-hour forecasts initialized at 2100 UTC.

To further dig into the performance of both Hyb_retro and Hyb_WD experiments, the predicted and observed MRMS composite reflectivity over a 6-hour period beginning at 2100 UTC are compared (Fig. 5.8). The analyses from both experiments are found to match the observations geographically, but predicted reflectivity values are generally stronger than observed (Figs. 5.8 a, b vs o). While the forecasted CREF values for the deep convection at the leading edge of the system, as well as the developing convection within the trailing areas of the quasi-linear convective complex over Georgia and North Carolina, are generally overestimated, both experiments show reasonable CREF forecasts in comparison with the MRMS observations. In the 1-hour forecast valid at 2200 UTC, the deep convection at the east leading edge of the system displays a little southeastward displacement bias and a faster propagation in both experiments compared to the MRMS observations (Figs. 5.8 a2, b2 vs o2). However, 3- to 4-hour into the forecast (valid at 0000 UTC and 0100 UTC), displacement bias reduces in Hyb_WD and is better in line with the MRMS observations (Figs. 5.8 b4 - b5 vs o4 - o5), while the displacement bias becomes more apparent in Hyb_retro (Figs. 5.8 a4 - a5 vs o4 - o5). The final 2-hour CREF forecasts are similar for both experiments as the linear convective complex splits into two, and neither is consistent with the observations (Figs. 5.8 a6 - a7 and b6 - b7 vs o6 - o7). It should be pointed out that both experiments also produce some spurious convection starting from east Kentucky and propagated to east along the border of Virginia and North Carolina. In reality, a weak cell developed at later hours in MRMS observations (Fig. 5.8 o4 - o6), indicating that those spurious cells could be considered as the earlier stage of the convection initiation, but the predicted cells in both experiments are too early and too strong in both experiments.

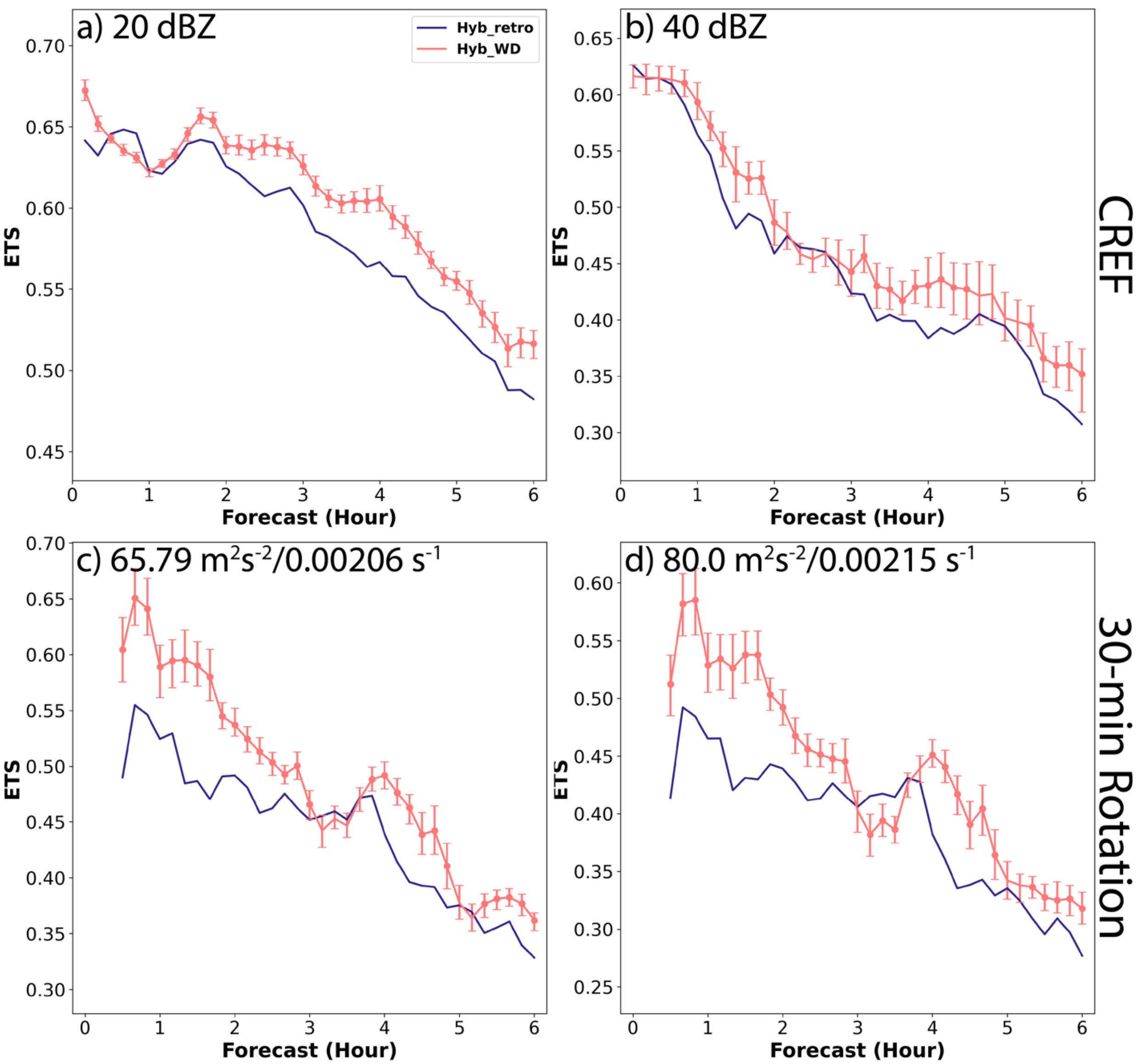

Figure 5.9 ETSs for (top row) CREF and (bottom row) 2 – 5 km UH track forecasts for 5 May 2020. The experiments Hyb_retro, Hyb_WD are indicated by blue, red lines, respectively. Dots indicate the statistical significance of changes in Hyb_WD as compared to Hyb_retro. Scores are provided for a neighborhood radius of 12 km.

In Figure 5.9, the performance of Hyb_retro and Hyb_WD for the 5 May 2020 event is evaluated using ETS for CREF and rotation track forecasts from 1800 to 0000 UTC. Same as the general evaluation, the decrease in ETS with increasing forecast lead time is observed in both CREF and rotation track forecasts for all experiments. The sharpest decline in ETS occurs during the first 3-hour forecast for the rotation track in both experiments (Figs. 5.9 c, d). The comparison of forecasts produced by Hyb_WD and Hyb_retro shows that Hyb_WD has a greater initial advantage but experiences a more rapid decline. After three hours into the forecast, Hyb_WD maintains a slight but consistent advantage over Hyb_retro. The forecast performance of 40 dBZ CREF (Fig. 5.9 b), which indicates heavy precipitation inside a strong convection, rapidly degrades during the first 3-hour forecast. In contrast, in the case of 20 dBZ CREF threshold (Fig. 5.9 a) which pertains to anvil areas away from the convection cell, the steepest decline occurs during the 3- to 6-hour forecast. During the first hour of the forecast, both Hyb_WD and Hyb_retro have equivalent performance for the 20 and 40 dBZ CREF. However, Hyb_WD exhibits a clear superiority over Hyb_retro after 1-hour into the forecast. These results are consistent with the overall performance discussed in the previous section, where Hyb_WD outperforms Hyb_retro in terms of both CREF and rotation track.

5.3.3 The 22 May 2020 severe weather event in Southern Great Plains

On May 22, 2020, a severe weather event took place in the central United States, which was predicted several days in advance by the SPC to have a moderate risk of tornadoes in southern Oklahoma and northern Texas. This event occurred as a result of a powerful trough moving from the west coast towards the central United States. The formation of a strong surface low over northwestern Texas and the presence of a warm front across southern Oklahoma facilitated the development of the event. The low-level shear was very strong with 0-1 km storm relative helicity (SRH) values reaching up to 300 $m^2 s^{-2}$. Additionally, the temperature and dew point conditions were also favorable for severe weather development, with temperatures reaching the mid 80s and dew points in the low 70s. The temperature inversion eroded in the afternoon, which led to storm development around 2200 UTC and hence the formation of a MCS at the intersection of the dryline crossing mid Texas and the warm front in southern Oklahoma (denoted by blue squares in Fig.

5.11). The severe weather event produced numerous EF-0 to EF-1 tornadoes accompanied by significant hail events, leading to damages estimated at almost $7 million. The tornado touchdown, several instances of hail with diameter greater than 2 inches and flash flooding that impacted northern Texas primarily occurred between 2200 and 0300 UTC the next day.

The 0- to 6-hour forecasts of 2 – 5 km UH track initialized at three different times, 2000, 2200, and 0000 UTC are qualitatively evaluated in comparison with the overlaid NWS/SPC severe weather reports. The main focus of this event is on tornado and hail events occurring near the state border of Oklahoma - Texas and Oklahoma - Arkansas between 2200 and 0300 UTC. The results from the forecast initialized at 2000 UTC suggest that both Hyb_retro and Hyb_WD indicate the presence of the MCS, but are not in good agreement with the NWS/SPC severe weather reports. More specifically, the forecasts from both experiments show that the large UH values are located to the south of the area where tornadoes and hail were reported in northern Texas, and to the north of the reports in eastern Oklahoma (Figs. 5.10 a, d). Additional assimilation cycles improve the analysis valid at 2200 UTC and the following 6-hour forecast (Figs. 5.10 b, e). Both experiments correct the relative northwest MCS in northern Texas to a location well-agreed with the storm reports, but Hyb_retro fails to forecast big rotation in an area where tornadoes and hail were reported (red box in Fig. 5.10 b). The experiment Hyb_WD produces a good coverage for potions of north Texas where most tornadoes and hail were reported, including the area in the red box of Fig. 5.10 e. Neither experiment improves the forecast for the MCS crossing the border of Oklahoma-Arkansas due to the north biased UH track, and the forecast for two tornadoes reported around 0300 UTC (black boxes in Figs. 5.10 b, e). Both experiments perform similarly again for the events in Texas, Oklahoma, and Arkansas in the forecast initialized at 0000 UTC. Overall, both experiments perform well in agreement with the NWS/SPC severe weather reports in terms of the rotation forecast. There is no significant improvement seen in this event by using weather-dependent error covariances, except for the MCS in northern Texas in the forecasts initialized from 2200 UTC (Fig. 5.10 e) as well as the forecast initialized from 2100 and 2300 UTC which are not shown in Figure 5.10.

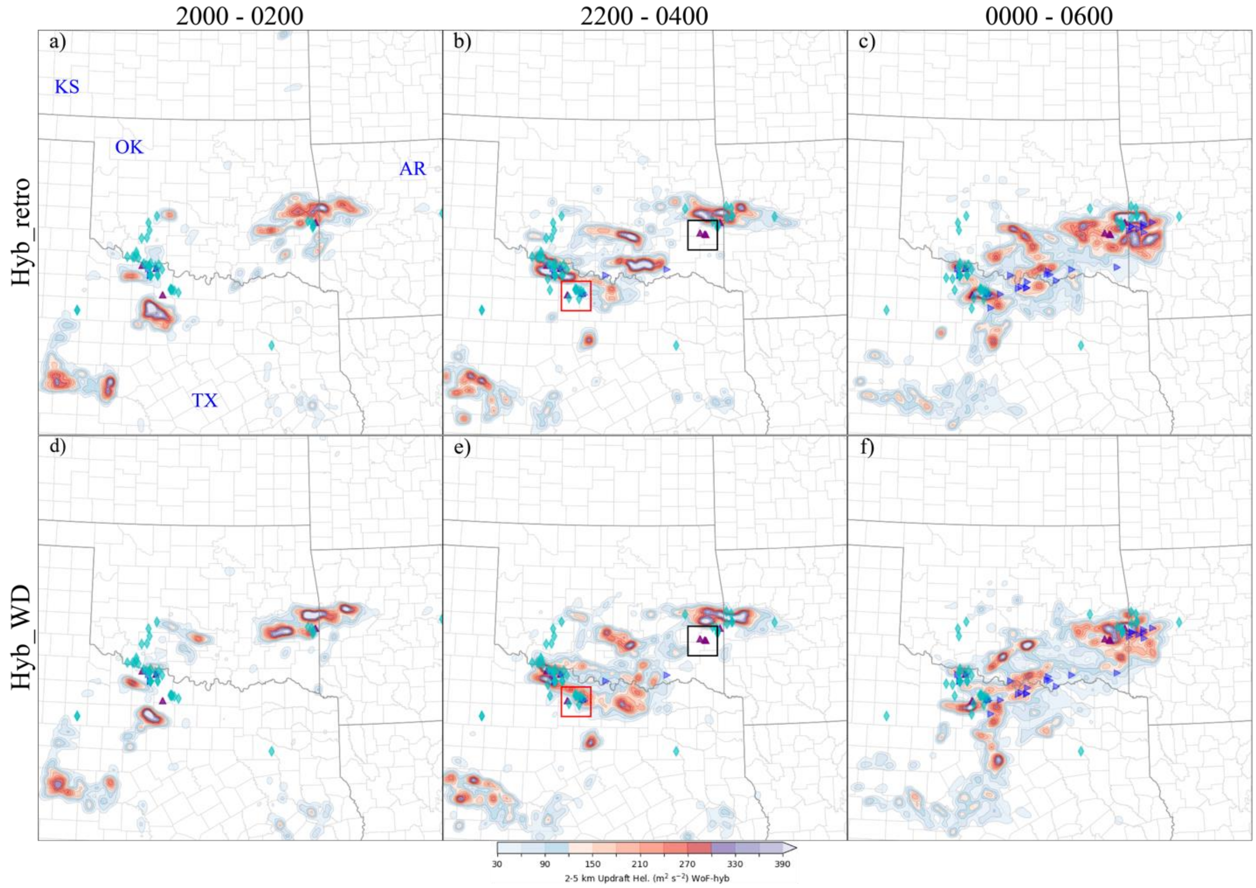

Figure 5.10 As in Fig. 5.7, but for the 6-hour 2 – 5 km UH track forecasts initialized at (a, d) 2000, (b, e) 2200, and (c, f) 0000 UTC on 22 May 2020.

In Figure 5.11, the 6-hour CREF forecasts initialized from 2200 UTC are compared with the MRMS CREF observations. During the first two hours, both the Hyb_retro and Hyb_WD experiments accurately predict the areal coverage of CREF exceeding 35 dBZ (Figs. 5.11 a1, a2, and b1, b2 vs o1, o2), although some spurious cells appear in eastern Texas. However, the prediction worsens thereafter. At 0100 UTC (the third hour forecast), Hyb_retro shows a decrease in the coverage of CREF and a more eastward drift compared to the observations. On the other hand, Hyb_WD produces a slightly south biased CREF but the areal coverage is comparable to the observations (blue boxes in Figs. 5.11 a4, b4 vs o4). The convective system in Hyb_WD continues to develop and propagate eastward until 0300 UTC (blue boxes in Figs. 5.11 b5, b6), which is in agreement with the observations (blue boxes in Figs. 5.11 o5, o6), while Hyb_retro fails to predict the corresponding cell (blue boxes in Figs. 5.11 a5, a6). This improvement is also displayed in the CREF forecasts initialized at 2100 and 2300 UTC (not shown). It is important to note that the spurious cells initiated from 2300 UTC, which develops and propagates during the following four hours and finally upscales to a quasi-linear convective system at 0300 UTC (black boxes in Figs. 5.11 a7, b7), significantly differ from the upscale process in the observations (black boxes in Fig. 5.11 o7), leading to a significant drop in the CREF ETSs during the first two hours and a consistent low value during the following period. The primary focus of our future studies is to restrain the development of spurious convection.

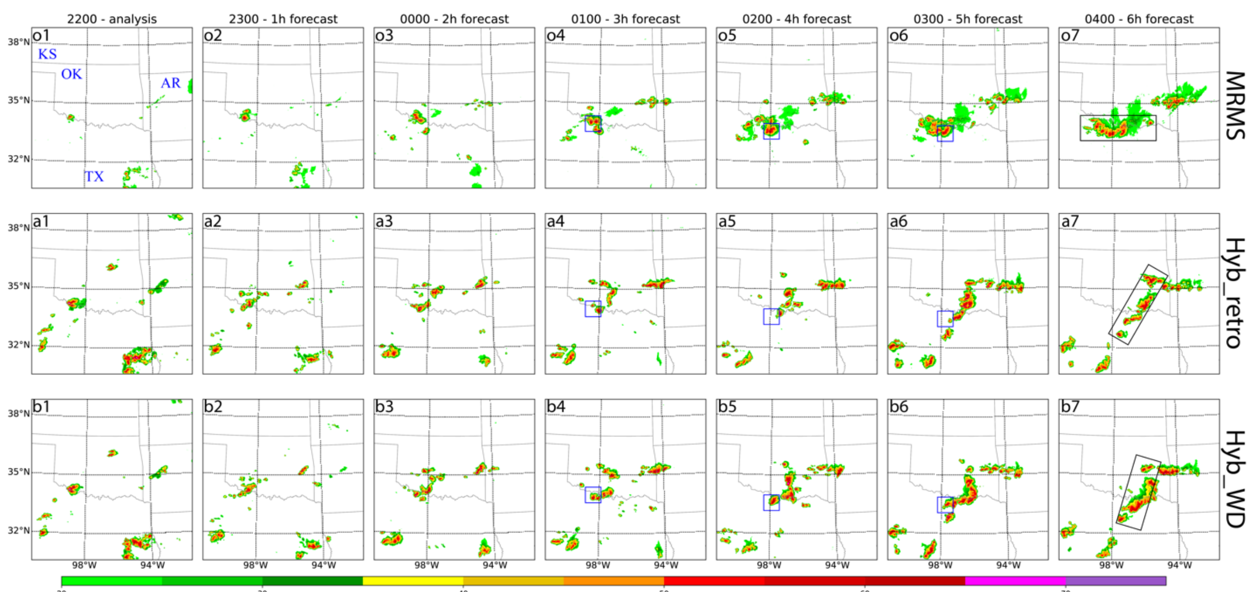

Figure 5.11 As in Fig. 5.8, but for the (1) analysis and (2 - 7) forecasts initialized at 2200 UTC on 22 May 2020.

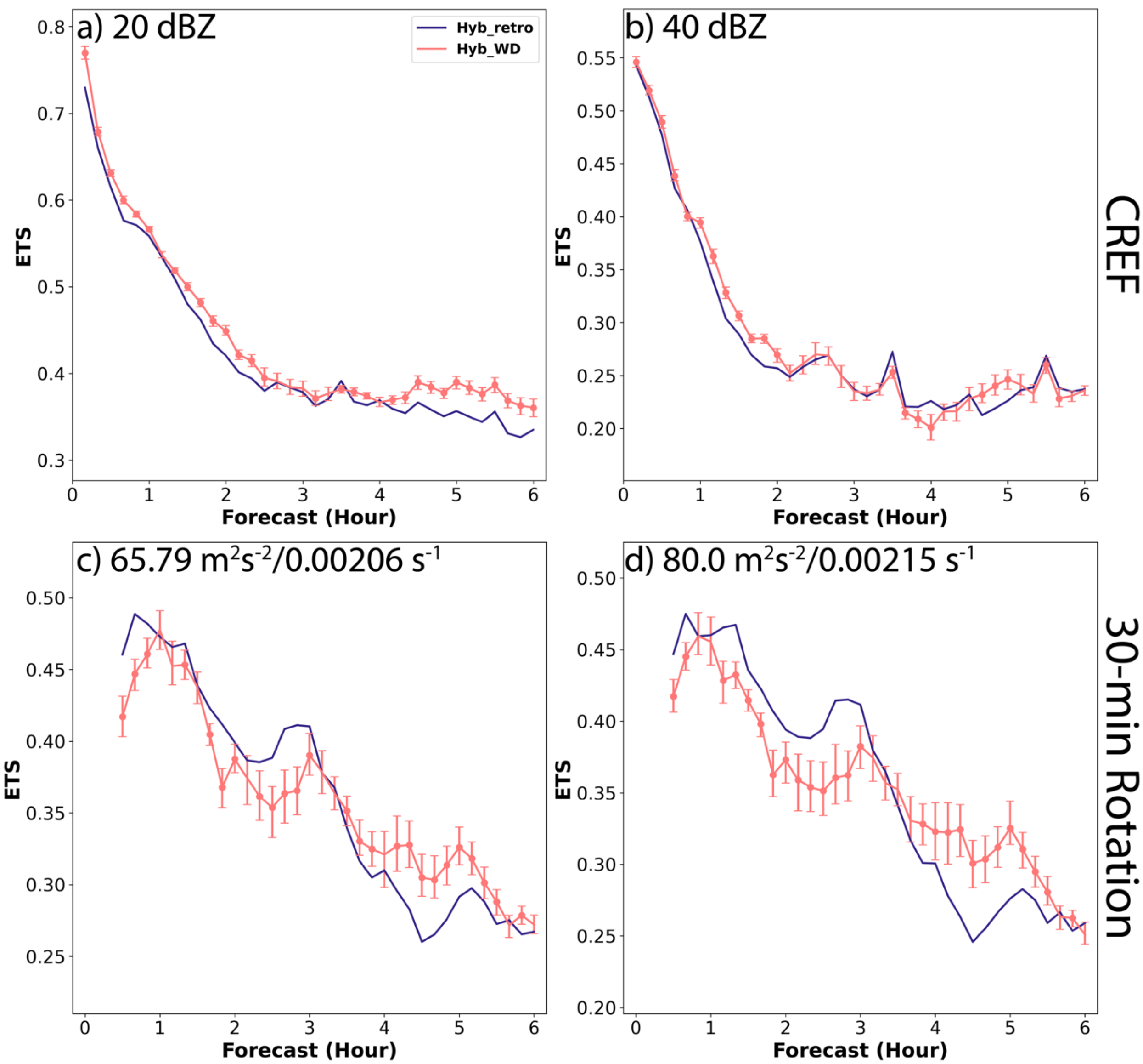

Figure 5.12 As in Fig. 5.6, but for 22 May 2020.

The variations of the ETSs for both the CREF and midlevel rotation forecasts are displayed in Figure 5.12. These results are aggregated from seven different forecast periods initialized from 1800 to 0000 UTC next day. The ETSs for CREF forecasts at 20 and 40 dBZ thresholds show a rapid decline in the first two hours, followed by a leveling off over the remaining hours (Figs. 5.12 a, b). Since ETS is a measurement highly sensitive to hits, the quick decline of ETS during the first two hours is attributed to the rapid development of spurious cells, as shown in Fig. 5.11.

The results from two experiments suggest that there are no significant differences in the performance of CREF forecasts. Hyb_WD has a constant but minor improvement over Hyb_retro for low reflectivity, which is also observed in the evaluation using the 40 dBZ threshold, but the improvement lasts for only two hours (Figs. 5.12 a, b). In contrast, the differences in ETSs of midlevel rotation forecasts between different experiments are more noteworthy than those of CREF forecasts. During the first three hours of the forecast period, Hyb_retro performs better than Hyb_WD. However, after this period, Hyb_WD outperforms Hyb_retro (Figs. 5.12 c, d).

5.3.4 The 10 Dec 2021 severe weather event in Midwest

The last case under investigation is a deadly late-season severe weather event that occurred from the evening of December 10th to the early morning of December 11st, 2021. The event caused catastrophic damage and numerous fatalities across portions of the Southern United States and Ohio Valley. The SPC had initially issued a slight risk two days prior to the event. Due to the presence of an intense upper-level trough, abundant moisture, and robust instability, however, SPC upgraded the risk to an enhanced risk on the Day 1 and eventually to a moderate risk on the morning of December 10th. Given the impressive dynamics of the event, upward trends in the 0 – 1 km and 0 – 3 km SRH values were observed, reaching a maximum of 600 m^2 s^{-2} and 1000 m^2 s^{-2}, respectively. This volatile environment resulted in formation of rare widespread winter tornadoes that caused an estimated total economic loss of $18 billion. One of the supercell storms that formed from initial mid-afternoon activity near Arkadelphia, Arkansas, was particularly devastating and produced a family of strong tornadoes. This supercell generated an EF-4 tornado that travelled 81.2 miles across Arkansas, Missouri, and Tennessee, and another EF-4 tornado that travelled 165.7 miles across Tennessee and southern Kentucky.

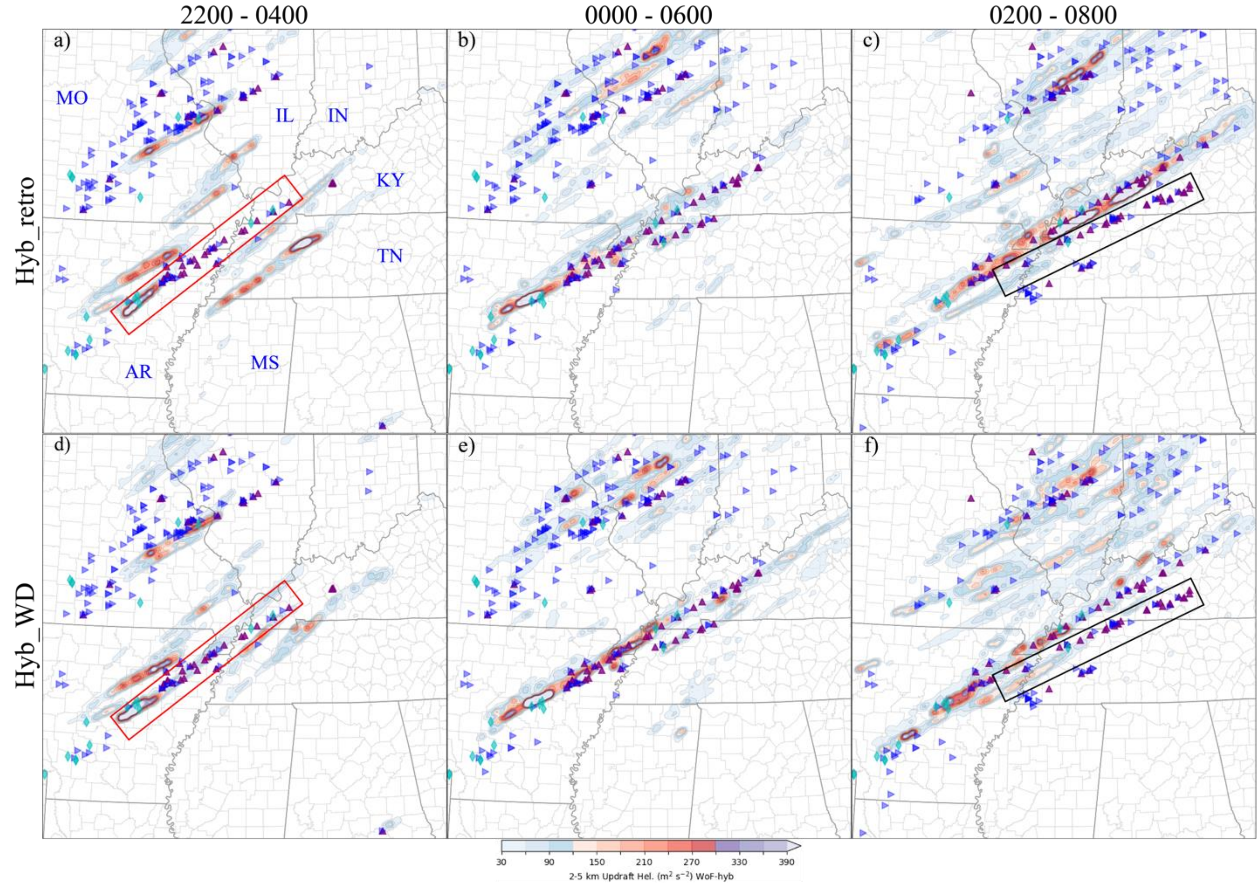

Figure 5.13 As in Fig. 5.7, but for the 6-hour 2 – 5 km UH track forecasts initialized at (a, d) 2200, (b, e) 0000, and (c, f) 0200 UTC on 10 Dec 2021.

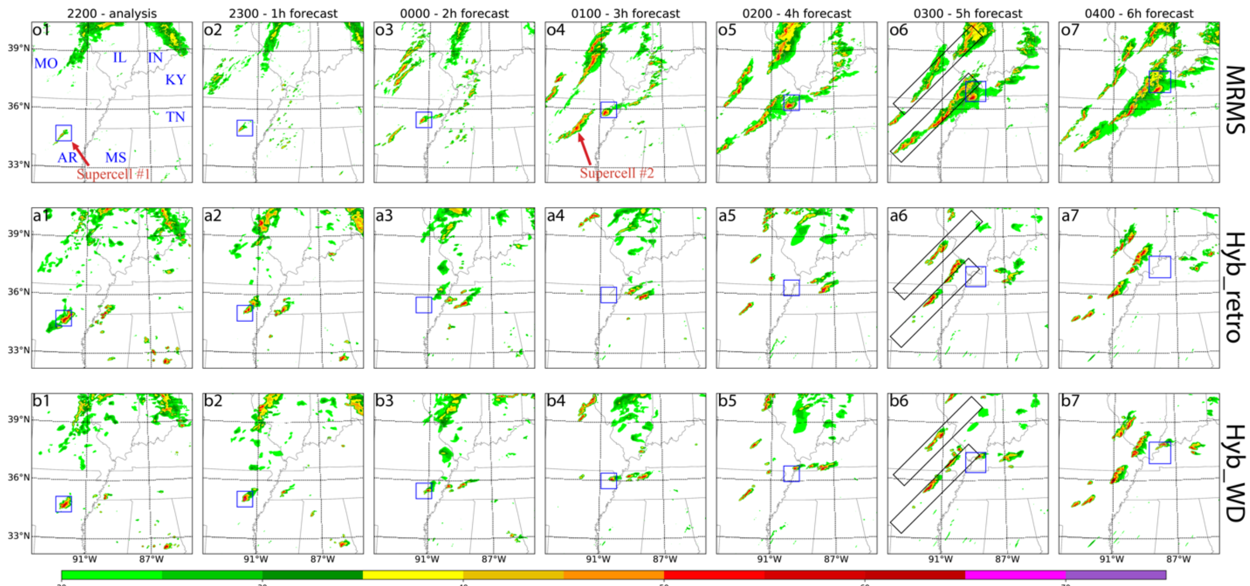

Figure 5.14 As in Fig. 5.8, but for the (1) analysis and (2 - 7) forecasts initialized at 2200 UTC on 10 Dec 2021.

During the time period from 2200 UTC on 10 Dec 2021 to 0900 UTC on 11 Dec 2021, most severe weather events including tornadoes and damaging winds were reported. The severe weather events were associated with two long-lived supercells. The first supercell was initiated in central Arkansas at 2200 UTC (Fig. 5.14 o1) and the second supercell formed near the same location at 0100 UTC the next day (Fig. 5.14 o4). The 2 – 5 km UH tracks for 6-hour forecasts launched at 2200, 0000, and 0200 UTC are compared to the NWS/SPC storm reports to evaluate the accuracy of the forecasts (Fig. 5.13). The 2200 UTC forecasts show a swath of 2 – 5 km UH greater than 200 m^2 s^{-2} in northeastern Arkansas in both Hyb_retro and Hyb_WD experiments, which rapidly downgrade in intensity to greater than 40 m^2 s^{-2} in one hour (Fig. 5.13 a, d). The swath is corresponding to the supercell that produced EF-4 tornadoes at a later time. The UH swath in Hyb_retro disappears after 2 hours into the forecast, whereas the UH swath in Hyb_WD persists until the end of the forecast, albeit with intensity only greater than 40 m^2s^{-2} (red boxes in Figs. 5.13 a vs d). This persisted UH swath corresponds well with the long-track tornado reported by SPC during the same time period. The Hyb_retro experiment also has a separate weak UH swath intersected with the track of the storm reports, but it represents a newly formed convection that is unrelated to the first long-lived tornadic supercell. For forecast launched from 0000 UTC, the UH swath in both experiments becomes progressively strong and more aligned with the tornado reports, with Hyb_retro reaching a level of 40 m^2s^{-2} and Hyb_WD reaching a level of 160 m^2s^{-2} along the tornado track indicated by the storm reports (Figs. 5.13 b, e). However, the situation in the 0200 UTC forecast is somewhat reversed. Hyb_retro provides a UH swath greater than 200 m^2 s^{-2}, while Hyb_WD provides a UH swath only greater than 80 m^2 s^{-2} (Figs. 5.13 c, f). Without considering the intensity of the swaths, both experiments provide a UH swath with little phase errors for the first long-lived supercell that produced EF-4 tornadoes. The overlap of the UH swath for the second long-lived supercell embedded in a quasi-linear convective system highlights the limitation of WoF-Hybrid in accurately simulating the motion of a complicated weather system. The UH forecast from 0200 UTC eventually established a relatively better track for the second supercell by enforcing itself to absorb velocity information from radar observation (Figs. 5.13 c, f). However, the rotation with greater than 80 m^2 s^{-2} only persisted for about 3 hours, despite its associated greater than 40 dBZ reflectivity persisting to the end of the forecast (not shown). This result indicates another shortcoming of WoF-Hybrid, which is the inability to capture the dynamic processes of finer scale convection with its current resolution. More research with higher resolution for this particular case will be pursued in the future.

The results of the 6-hour CREF forecasts initialized at 2200 UTC show that Hyb_WD is the experiment more accurate in predicting the first tornadic supercell that occurred across Arkansas, Missouri, Tennessee, and Illinois. The storm motion in this prediction is slightly faster than the observed storm, as shown by the blue boxes in Fig. 5.14 b compared to 5.14 o. On the other hand, the forecast for the same supercell in Hyb_retro only lasts one hour until 2300 UTC, and the supercell disappears after 0000 UTC, as shown by the blue boxes in Fig. 5.14 compared to 5.14 o. The second supercell embedded in the quasi-linear convective system (Fig. 5.14 o4) is initially missed by both experiments at 0100 UTC (3 hours into the 2200 UTC forecast, Figs. 5.14 a4, b4). however, it initiates after 4 hours into the forecast (Figs. 5.14 a5, b5). Nevertheless, both experiments display the supercell along with the quasi-linear convective system located northwest of their actual locations (bottom black boxes in Figs. 5.14 a6, b6 vs o6). This placement is consistent with the situation observed in the UH track forecasts (Fig. 5.13a, d). Additionally, the areal coverage of the quasi-linear convective systems producing tornadoes and another one located to the north is found to be far less than the observed coverage (Figs. 5.14 a6 - a7, b6 - b7 vs o6 -

o7). The underestimation of the quasi-linear convective systems is the main reason behind the low POD shown in Figure 5.4.

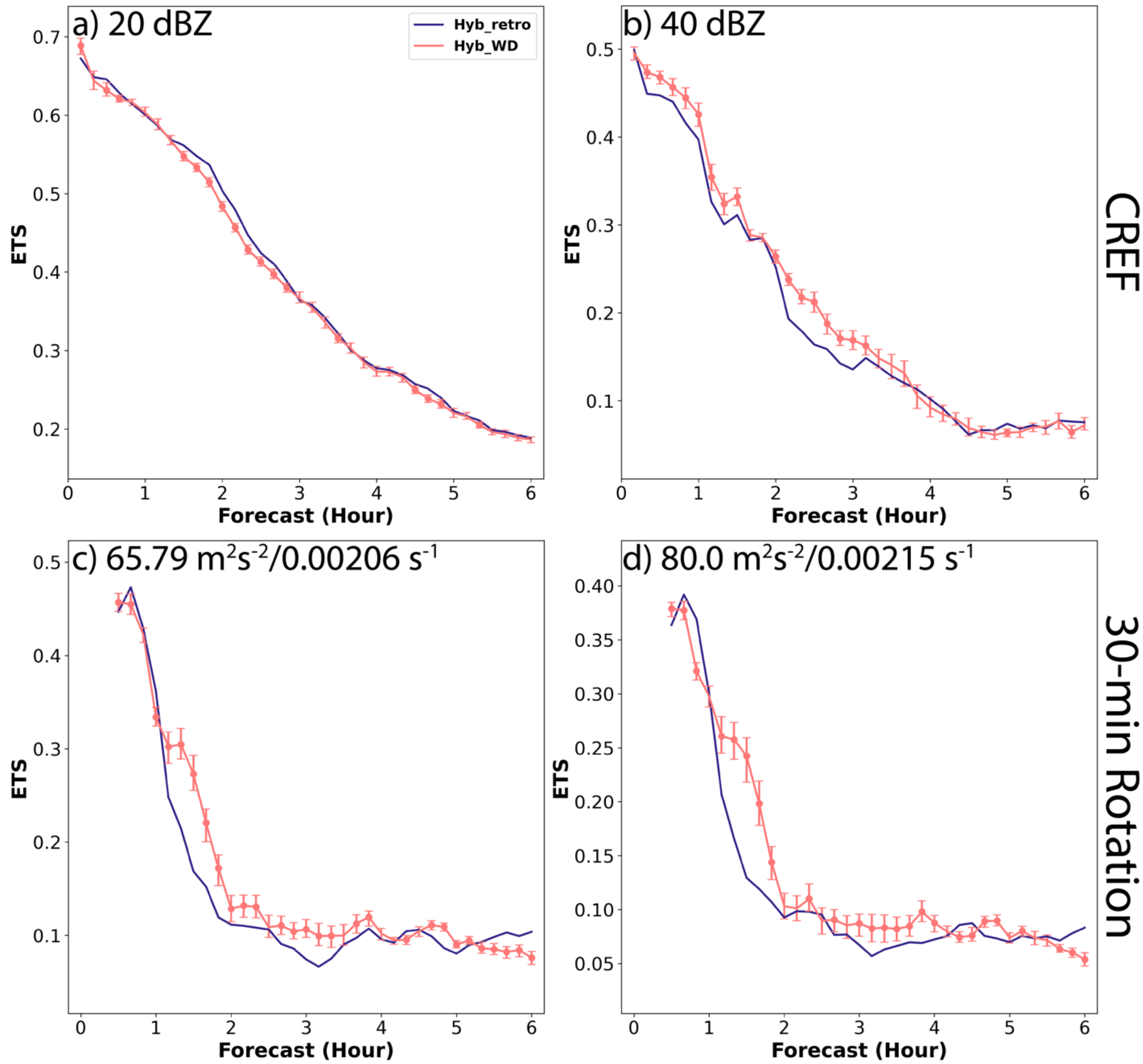

Figure 5.15 ETSs for (top row) CREF and (bottom row) 2 – 5 km UH track forecasts for 10 Dec 2021. The experiments, Hyb_retro and Hyb_WD are indicated by blue and orange lines, respectively. Dots indicate the statistical significance of changes in Hyb_WD as compared to Hyb_retro. Scores are provided for a neighborhood radius of 12 km.

As pointed out in Table 5.1, the quantitative evaluation of the 10 Dec 2021 event is carried out using hourly forecasts from 2100 to 0300 UTC, as significant losses occurred after 2300 UTC.

The aggregated ETSs for both 20 and 40 dBZ thresholds are characterized by an almost linear decrease throughout the forecast period. The ETSs for Hyb_WD are slightly below those values for Hyb_retro for the 1- to 3-hour forecast with the 20 dBZ threshold (Fig. 5.15 a). In contrast, for the 0- to 4-hour forecast with the 40 dBZ threshold, the ETSs for Hyb_WD are slightly higher than those for Hyb_retro (Fig. 5.15 b). Similar improvement of Hyb_WD is also observed in ETSs of 1- to 4-hour midlevel rotation track forecasts regardless of the threshold used in the verification (Figs. 5.15 c, d). Further analysis of the contingency table reveals that the changes in ETSs between two experiments are due to the increased hits as well as decreased false alarms and misses compared to those of Hyb_retro (not shown). It is worth noting that the number of misses is an order of magnitude greater than that of hits, and two orders of magnitude greater than that of false alarms. This results in POD, which includes misses in the denominator, being dominated by the count of misses, displaying unusually low values in the performance diagram, and is barely changed in Hyb_WD compared to Hyb_retro by improved hits and false alarms (The symbols in purple color in Figs. 5.4 and 5.5). This is the reason why the benefit of using weather dependent error covariances in WoF-Hybrid is small for this particular case.

5.4 Summary and Discussion

In this chapter, the weather-dependent static error covariances are implemented into the WoF-Hybrid. The goal is to examine if the updated WoF-Hybrid can improve the convective-scale analyses and forecasts. The performance of the system is evaluated with five severe weather events that occurred in 2020 and 2021 using statistical evaluation metrics, such as POD, FAR, BIAS, CSI, and ETS. Furthermore, this study discusses the results of three specific events that took place on 5 May 2020, 22 May 2020, and 10 Dec 2021 using both qualitative and quantitative verification methods. MRMS CREF products, MRMS AzShear products and NWS/SPC local storm reports were used for these verifications.

The aggregate metrics over forecasts from all 5 events suggest an overall improvement in the quality of 0- to 6-hour forecasts for CREF and midlevel rotation track with the inclusion of the weather-dependent static error covariances in WoF-Hybrid. The improvement has a slightly greater margin in early CREF and later rotation track forecasts. Furthermore, the forecast skills are found to exhibit more variability at longer forecast times (3-hour and 6-hour) and less variability at 1-hour forecasts. For CREF, most cases show an underpredicted bias for the 20 dBZ threshold for 3-hour and 6-hour forecasts, and an overpredicted bias for the 40 dBZ threshold for all forecasts. For rotation forecasts, underpredicted bias exists for all cases.

The effectiveness of the weather-dependent static error covariances in WoF-Hybrid is clearly seen in the first detailed examined event, i.e., 5 May 2020. The improvement of Hyb_WD compared to Hyb_retro is observed in several aspects of the forecast, including CREF and rotation track, and is reflected in the quantitative verification metrics. The CREF forecasts from Hyb_WD exhibit a better match with the observed CREF from MRMS, particularly for the deep convection at the leading edge of the system and the developing convection within the trailing areas of the quasi-linear convective complex. Moreover, the UH track forecasts from Hyb_WD are more aligned with the storm reports for this case. The results for the 22 May 2020 event are more complicated. The improvement is typically neutral to slightly better for CREF, while it is mixed for midlevel rotation tracks. The mixed results of the rotation track forecast are due to some spurious cells existing in later hour forecasts. Despite this, a slight correction of midlevel rotation track and its associated CREF is found in Hyb_WD. The 10 Dec 2021 event has the worst quantitative results among 5 events. The reason behind it is that both Hyb_retro and Hyb_WD have significant failures in predicting two quasi-linear convective systems. Only Hyb_WD can predict the famous long-lived supercell, which produced significant damages over four states including Arkansas, Missouri, Kentucky, and Tennessee. Beside this, the quantitative analysis of Hyb_WD reveals an improvement in ETSs for high reflectivity and midlevel rotation tracks. The validity of this improvement is confirmed by visually comparing the NWS/SPC storm reports with the UH track forecasts initialized prior to the formation of the tornadoes.

In summary, this study exhibits the potential of improving the accuracy of severe weather forecasts by implementing weather-dependent static error covariances in the WoF-Hybrid system. The system is evaluated using several severe weather events, and it was found that the storm mode and orientations are more skillfully predicted with the new implementation.

Chapter 6 Summary and Future Plans

6.1 General Summary

A real-time, deterministic, weather adaptive, dual-resolution hybrid EnVAR DA system, i.e., WoF-Hybrid, is further developed for the purpose of better forecasting convective-scale DA and NWP. The updates allow the system (i) to assimilate satellite-derived CWP and LPW observations, (ii) to adopt newly developed weather-dependent static error covariances instead of the homogeneous static error covariances, which is not suitable for convective DA, (iii) to perform analyses using either pure 3DVAR (i.e., WoF-Var) or 3DEnVAR DA (i.e., WoF-Hybrid), and finally (iv) to be evaluated using operational configurations and observations as in HWT SFE. The description of the early version of WoF-Hybrid system is first described in Chapter 2 of this dissertation. The updated DA system with newly developed components, and a series of experiments and performance evaluations are conducted in the remaining part of this dissertation.

First, the satellite derived CWP and LPW observation forward operators are built into the WoF-Hybrid system. To understand if the CWP and LPW are appropriately assimilated into the system and if the satellite-derived CWP and LPW observations in addition to the radar observations can help improve the analysis and prediction of convective systems, experiments are conducted with three severe weather events in 2017 in Chapter 3. The comparisons and evaluations of the results from the three types of experiments indicates that assimilating CWP and/or LPW improves model initial conditions and short-range forecasts for the three severe weather events:

i) For the 10 May 2017 event, assimilating CWP improves the cloud property analysis, downward shortwave radiation over cloudy areas, surface temperature, and atmospheric instability. Assimilating the LPW product yields positive results for CI by saturating the moisture fields surrounding the ongoing convection. Spurious cells are mostly removed in later forecasts due to this improvement. Improved analysis generates high UH swaths that are more appropriately placed when compared against the SPC damage reports.

ii) For the 16 and 26 May 2017 events, the experiment with assimilating both satellite derived CWP and LPW as well as radar data leads to intensified moisture gradient and low-level vertical wind shear at the surface and resulting in better analyzed and predicted storms compared to the experiment with assimilating only CWP and LPW data, and the experiment with assimilating only radar data.

As discussed in Chapter 1 and Chapter 4, simultaneous assimilation of observations with different spatiotemporal scales into the NWP model is still challenging in the early version of our DA system. Homogenous static error covariances were generally applied in our early version of the DA system which typically provided similar error structure inside and outside the convective core when the analysis was done. To solve this problem, weather-dependent background error structure is constructed based on the intensity of radar composite reflectivities calculated from the WoFS ensemble forecast product. The procedure includes classifying ensemble perturbations according to the strength of convection, filtering outliers, and applying Gaussian smoothing in the vertical direction. Then, the derived weather-dependent background error structure is integrated into the WoF-Hybrid DA system.

In Chapter 4, the impact of the weather-dependent background error structure on the WoF-Var (i.e., the variational component of WoF-Hybrid) analyses is first assessed with a single observation experiment. Without hybridization, the heterogeneity of the analysis increment is also displayed in the single observation experiment. Following it, two severe weather events, the 5 May 2020 linear complex and the 10 Dec 2021 long-lived supercell storm, are used to evaluate the updated WoF-Var system. It is found that:

i) For the 5 May 2020 event, the temperature at 500 hPa along the front is greatly increased when using weather-dependent error covariances. This increase may lead to robust buoyancy above the level of free convection, and hence the forecast improvement with composite reflectivity and UH in the primary convective complex. In this event, it also suggests relatively low MLCAPE in areas outside observed convection, causing the alleviation of the spurious convective cells in Tennessee and North Carolina.

ii) For the 10 Dec 2021 event, Improved reflectivity pattern and stronger convergence ahead of the rear-flank gust front, along with the presence of a strong cold pool in the forward-flank downdraft area and better coordination among these factors, are observed in the experiment that utilizes weather-dependent error covariances. These enhancements facilitated the development of a complete and well-formed supercell structure.

Even in a pure 3DVAR framework, the study suggests that the use of weather-dependent error covariances can significantly improve the accuracy of severe weather forecasts, especially in predicting the formation of tornadic supercells. This study is promising and attractive because the heterogeneity of the analysis increment can be achieved even without using flow-dependent error covariances from the ensemble. Without using ensemble forecasts, the deterministic DA and forecasts cycle with pure 3DVAR could be much more efficient. As we understand, severe convective weather and high-impact events, which are characterized by local, sudden, and short life cycles, usually require fast and timely forecasts, early warnings, and quick delivery of these forecasts and warnings to the public. Therefore, we believe that pure 3DVAR is also quite suitable for the severe weather forecasts by providing a highly efficient and rapid updating analysis and forecast cycle if needed. Certainly Chapter 5 demonstrated adding flow-dependent ensemble covariances can further improve the short-term severe weather forecasts.

Further investigations related to the evaluation of the benefit of newly updated WoF-Hybrid to convective-scale short-term forecasts are conducted in Chapter 5. The performance of the system is assessed using five severe weather and high-impact weather events that occurred in 2020 and 2021, which most related to HWT activities. Some statistical metrics, such as POD, FAR, BIAS, CSI, and ETS are used. Three specific events are examined in detail, and the results are assessed both qualitatively and quantitatively.

The aggregate metrics over forecasts from all five severe weather and high-impact weather events suggest an overall improvement in the quality of 0 – 6-h forecasts for CREF and midlevel rotation track with the inclusion of the weather-dependent static error covariances in WoF-Hybrid. However, forecast skills exhibit more variability at longer forecast times and less variability at 1-hour forecasts. The analysis also reveals an underpredicted bias for the 20 dBZ threshold for 3- and 6-hour forecasts, and an overpredicted bias for the 40 dBZ threshold for all forecasts. Detailed evaluation of the 5 May 2020 event show that the weather-dependent static error covariances are effective in improving CREF and midlevel rotation track forecasts. However, mixed results are seen for rotation track forecasts in the 22 May 2020 event due to spurious convection. The worst quantitative results are obtained for the 10 Dec 2021 event, where the system has significant difficulty in predicting two quasi-linear convective systems. Nonetheless, the use of the weather-dependent static error covariances allows for the prediction of a long-lived supercell that caused significant damage across four states for 10 Dec 2021 event. This result is significantly better that the quasi-operational run of WoFS for this particular case (Carpenter, 2022). Although these findings are based on only five of the worst-performed events in HWT SFE 2020 and 2021, the study suggests that the updated WoF-Hybrid system with the weather-dependent static error covariances has a generally positive impact on convective-scale analyses and forecasts, although there is still room for further improvement, particularly in predicting quasi-linear convective systems. Obviously, testing the newly updated WoF-Hybrid with more severe and high-impact weather events is needed.

6.2 Future Plans

The further development of the WoF-Hybrid provides a valuable platform for exploring the scientific challenges associated with improving severe weather forecasting. Although numerous experiments have been conducted in this dissertation, further studies are necessary. For instance, (1) the assimilation of CWP and LPW observations still results in unwanted artifacts, such as the emergence of spurious convection during the free forecast; (2) the detrimental impact of satellite derived high layer precipitable water observations remains unresolved; (3) the reliability of the upgraded WoF-Hybrid system needs to be further tested in real-time situations like HWT SFE; (4) how the improved WoF-Hybrid system can benefit the performance of the WoFS ensemble system, is not well studied. Our future work will focus on investigating these problems to achieve our ultimate objective of improving prediction of severe and high-impact weather events for society, particularly by reducing fatalities and property damages, one of long-time goals of NSSL WoF project.

References

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses. *Monthly Weather Review*, **137**, 1805–1824, https://doi.org/10.1175/2008MWR2691.1.

Bannister, R. N., 2008: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1951–1970, https://doi.org/10.1002/qj.339.

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale Weather Prediction with the RUC Hybrid Isentropic–Terrain-Following Coordinate Model. *Monthly Weather Review*, **132**, 473–494, https://doi.org/10.1175/1520-0493(2004)132<0473:MWPWTR>2.0.CO;2.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The Oklahoma Mesonet: A Technical Overview. *Journal of Atmospheric and Oceanic Technology*, **12**, 5–19, https://doi.org/10.1175/1520-0426(1995)012<0005:TOMATO>2.0.CO;2.

Brousseau, P., L. Berre, F. Bouttier, and G. Desroziers, 2011: Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var. *Quarterly Journal of the Royal Meteorological Society*, **137**, 409–422, https://doi.org/10.1002/qj.750.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, **131**, 1013–1043, https://doi.org/10.1256/qj.04.15.

——, P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *Monthly Weather Review*, **138**, 1550–1566, https://doi.org/10.1175/2009MWR3157.1.

——, ——, ——, ——, and ——, 2010b: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations. *Monthly Weather Review*, **138**, 1567–1586, https://doi.org/10.1175/2009MWR3158.1.

——, and Coauthors, 2015: Implementation of Deterministic Weather Forecasting Systems Based on Ensemble–Variational Data Assimilation at Environment Canada. Part I: The Global System. *Monthly Weather Review*, **143**, 2532–2559, https://doi.org/10.1175/MWR-D-14-00354.1.

Calhoun, K. M., T. M. Smith, D. M. Kingfield, J. Gao, and D. J. Stensrud, 2014: Forecaster Use and Evaluation of Real-Time 3DVAR Analyses during Severe Thunderstorm and

Tornado Warning Operations in the Hazardous Weather Testbed. *Weather and Forecasting*, **29**, 601–613, https://doi.org/10.1175/WAF-D-13-00107.1.

Caron, J.-F., and L. Fillion, 2010: An Examination of Background Error Correlations between Mass and Rotational Wind over Precipitation Regions. *Monthly Weather Review*, **138**, 563–578, https://doi.org/10.1175/2009MWR2998.1.

Carpenter, N., 2022: Evaluation of the experimental warn-on-forecast system and WoF-hybrid 3DEnVar system on short-term forecasts for 2021 real-time cases.

Caya, A., J. Sun, and C. Snyder, 2005: A Comparison between the 4DVAR and the Ensemble Kalman Filter Techniques for Radar Data Assimilation. *Monthly Weather Review*, **133**, 3081–3094, https://doi.org/10.1175/MWR3021.1.

Chou, M.-D., K.-T. Lee, S.-C. Tsay, and Q. Fu, 1999: Parameterization for Cloud Longwave Scattering for Use in Atmospheric Models. *Journal of Climate*, **12**, 159–169, https://doi.org/10.1175/1520-0442(1999)012<0159:PFCLSF>2.0.CO;2.

Clark, A. J., W. A. Gallus, and M. L. Weisman, 2010: Neighborhood-Based Verification of Precipitation Forecasts from Convection-Allowing NCAR WRF Model Simulations and the Operational NAM. *Weather and Forecasting*, **25**, 1495–1509, https://doi.org/10.1175/2010WAF2222404.1.

——, and Coauthors, 2011: Probabilistic Precipitation Forecast Skill as a Function of Ensemble Size and Spatial Scale in a Convection-Allowing Ensemble. *Monthly Weather Review*, **139**, 1410–1418, https://doi.org/10.1175/2010MWR3624.1.

Crook, N. A., and J. Sun, 2002: Assimilating Radar, Surface, and Profiler Data for the Sydney 2000 Forecast Demonstration Project. *Journal of Atmospheric and Oceanic Technology*, **19**, 888–898, https://doi.org/10.1175/1520-0426(2002)019<0888:ARSAPD>2.0.CO;2.

Davies, E. R., 2005: Machine Vision: Theory, Algorithms and Practicalities. E.R. Davies, Ed., Morgan Kaufmann, 47–101.

Dowell, D. C., and L. J. Wicker, 2009: Additive Noise for Storm-Scale Ensemble Data Assimilation. *Journal of Atmospheric and Oceanic Technology*, **26**, 911–927, https://doi.org/10.1175/2008JTECHA1156.1.

——, F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and Temperature Retrievals in the 17 May 1981 Arcadia, Oklahoma, Supercell: Ensemble Kalman Filter Experiments. *Monthly Weather Review*, **132**, 1982–2005, https://doi.org/10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2.

——, L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman Filter Assimilation of Radar Observations of the 8 May 2003 Oklahoma City Supercell: Influences of Reflectivity Observations on Storm-Scale Analyses. *Monthly Weather Review*, **139**, 272–294, https://doi.org/10.1175/2010MWR3438.1.

——, and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description. *Weather and Forecasting*, **37**, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1.

Du, J., 2011: NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0. 291407 data files, 2 ancillary/documentation files, 17 GiB, https://doi.org/10.5065/D6PG1QDD.

Dudhia, J., 1989: Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model. *Journal of the Atmospheric Sciences*, **46**, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, **99**, 10143–10162, https://doi.org/10.1029/94JC00572.

——, 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

Fierro, A. O., J. Gao, C. L. Ziegler, K. M. Calhoun, E. R. Mansell, and D. R. MacGorman, 2016: Assimilation of Flash Extent Data in the Variational Framework at Convection-Allowing Scales: Proof-of-Concept and Evaluation for the Short-Term Forecast of the 24 May 2011 Tornado Outbreak. *Monthly Weather Review*, **144**, 4373–4393, https://doi.org/10.1175/MWR-D-16-0053.1.

——, Y. Wang, J. Gao, and E. R. Mansell, 2019: Variational Assimilation of Radar Data and GLM Lightning-Derived Water Vapor for the Short-Term Forecasts of High-Impact Convective Events. *Monthly Weather Review*, **147**, 4045–4069, https://doi.org/10.1175/MWR-D-18-0421.1.

Fisher, M., 2007: The sensitivity of analysis errors to the specification of background error covariances.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface Data Assimilation Using an Ensemble Kalman Filter Approach with Initial Condition and Model Physics Uncertainties. *Monthly Weather Review*, **135**, 1846–1868, https://doi.org/10.1175/MWR3391.1.

Gallo, B. T., and Coauthors, 2017: Breaking New Ground in Severe Weather Prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Weather and Forecasting*, **32**, 1541–1568, https://doi.org/10.1175/WAF-D-16-0178.1.

Gao, J., and D. J. Stensrud, 2012: Assimilation of Reflectivity Data in a Convective-Scale, Cycled 3DVAR Framework with Hydrometeor Classification. *Journal of the Atmospheric Sciences*, **69**, 1054–1065, https://doi.org/10.1175/JAS-D-11-0162.1.

——, and ——, 2014: Some Observing System Simulation Experiments with a Hybrid 3DEnVAR System for Storm-Scale Radar Data Assimilation. *Monthly Weather Review*, **142**, 3326–3346, https://doi.org/10.1175/MWR-D-14-00025.1.

——, and Coauthors, 2021: Testing of the Warn-on-Forecast (WoF) Hybrid Data Assimilation and Forecasting System during the HWT Spring Experiment in 2020. *AMS - 101st American Meteorological Society Annual Meeting*,. https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Program/1513 (Accessed March 22, 2023).

——, and ——, 2022: The Performance of the Warn-on-Forecast Hybrid Data Assimilation and Forecasting System During the HWT Spring Forecast Experiment in 2021. *AMS - 30th Conference on Severe Local Storms*,. https://ams.confex.com/ams/30SLS/meetingapp.cgi/Home/0 (Accessed March 22, 2023).

——, M. Xue, A. Shapiro, and K. K. Droegemeier, 1999: A Variational Method for the Analysis of Three-Dimensional Wind Fields from Two Doppler Radars. *Monthly Weather Review*, **127**, 2128–2142, https://doi.org/10.1175/1520-0493(1999)127<2128:AVMFTA>2.0.CO;2.

——, ——, K. Brewster, and K. K. Droegemeier, 2004: A Three-Dimensional Variational Data Analysis Method with Recursive Filter for Doppler Radars. *Journal of Atmospheric and Oceanic Technology*, **21**, 457–469, https://doi.org/10.1175/1520-0426(2004)021<0457:ATVDAM>2.0.CO;2.

——, and Coauthors, 2013: A Real-Time Weather-Adaptive 3DVAR Analysis System for Severe Weather Detections and Warnings. *Weather and Forecasting*, **28**, 727–745, https://doi.org/10.1175/WAF-D-12-00093.1.

——, C. Fu, D. J. Stensrud, and J. S. Kain, 2016: OSSEs for an Ensemble 3DVAR Data Assimilation System with Radar Observations of Convective Storms. *Journal of the Atmospheric Sciences*, **73**, 2403–2426, https://doi.org/10.1175/JAS-D-15-0311.1.

Ge, G., 2011: On the Further Studies of Suitable Storm-Scale 3DVar Data Assimilation for the Prediction of Tornadic Thunderstorms. School of Meteorology, University of Oklahoma, 191 pp.

——, J. Gao, and M. Xue, 2012: Diagnostic Pressure Equation as a Weak Constraint in a Storm-Scale Three-Dimensional Variational Radar Data Assimilation System. *Journal of Atmospheric and Oceanic Technology*, **29**, 1075–1092, https://doi.org/10.1175/JTECH-D-11-00201.1.

Guerra, J. E., P. S. Skinner, A. Clark, M. Flora, B. Matilla, K. Knopfmeier, and A. E. Reinhart, 2022: Quantification of NSSL Warn-on-Forecast System Accuracy by Storm Age Using Object-Based Verification. *Weather and Forecasting*, **37**, 1973–1983, https://doi.org/10.1175/WAF-D-22-0043.1.

Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1218–1256, https://doi.org/10.1002/qj.3179.

Hamill, T. M., 1999: Hypothesis Tests for Evaluating Numerical Precipitation Forecasts. *Weather and Forecasting*, **14**, 155–167, https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

——, and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Monthly Weather Review*, **128**, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

Hamrud, M., M. Bonavita, and L. Isaksen, 2015: EnKF and Hybrid Gain Ensemble Data Assimilation. Part I: EnKF Implementation. *Monthly Weather Review*, **143**, 4847–4864, https://doi.org/10.1175/MWR-D-14-00333.1.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Monthly Weather Review*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

——, and ——, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **129**, 123–137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Hu, J., A. O. Fierro, Y. Wang, J. Gao, and E. R. Mansell, 2020: Exploring the Assimilation of GLM-Derived Water Vapor Mass in a Cycled 3DVAR Framework for the Short-Term Forecasts of High-Impact Convective Events. *Monthly Weather Review*, **148**, 1005–1028, https://doi.org/10.1175/MWR-D-19-0198.1.

——, and Coauthors, 2021: Evaluation of an experimental Warn-on-Forecast 3DVAR analysis and forecast system on quasi-real-time short-term forecasts of high-impact weather events. *Quarterly Journal of the Royal Meteorological Society*, **147**, 4063–4082, https://doi.org/10.1002/qj.4168.

Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and Cloud Analysis with WSR-88D Level-II Data for the Prediction of the Fort Worth, Texas, Tornadic Thunderstorms. Part I: Cloud Analysis and Its Impact. *Monthly Weather Review*, **134**, 675–698, https://doi.org/10.1175/MWR3092.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, **113**, https://doi.org/10.1029/2008JD009944.

Isaksen, L., M. Fisher, and J. Berner, 2007: Use of analysis ensembles in estimating flow-dependent background error variance. 22.

Janjić, Z. I., 1994: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Monthly Weather Review*, **122**, 927–945, https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A Comparison of Multiscale GSI-Based EnKF and 3DVar Data Assimilation Using Radar and Conventional Observations for Midlatitude Convective-Scale Precipitation Forecasts. *Monthly Weather Review*, **143**, 3087–3108, https://doi.org/10.1175/MWR-D-14-00345.1.

Jones, T. A., J. A. Otkin, D. J. Stensrud, and K. Knopfmeier, 2013: Assimilation of Satellite Infrared Radiances and Doppler Radar Observations during a Cool Season Observing System Simulation Experiment. *Monthly Weather Review*, **141**, 3273–3299, https://doi.org/10.1175/MWR-D-12-00267.1.

——, D. Stensrud, L. Wicker, P. Minnis, and R. Palikonda, 2015: Simultaneous Radar and Satellite Data Storm-Scale Assimilation Using an Ensemble Kalman Filter Approach for 24 May 2011. *Monthly Weather Review*, **143**, 165–194, https://doi.org/10.1175/MWR-D-14-00180.1.

——, K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-Scale Data Assimilation and Ensemble Forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined Radar and Satellite Data Experiments. *Weather and Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

——, X. Wang, P. Skinner, A. Johnson, and Y. Wang, 2018: Assimilation of GOES-13 Imager Clear-Sky Water Vapor (6.5 μm) Radiances into a Warn-on-Forecast System. *Monthly Weather Review*, **146**, 1077–1107, https://doi.org/10.1175/MWR-D-17-0280.1.

——, and Coauthors, 2020: Assimilation of GOES-16 Radiances and Retrievals into the Warn-on-Forecast System. *Monthly Weather Review*, **148**, 1829–1859, https://doi.org/10.1175/MWR-D-19-0379.1.

Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman Filter Analyses of the 29–30 May 2004 Oklahoma Tornadic Thunderstorm Using One- and Two-Moment Bulk Microphysics Schemes, with Verification against Polarimetric Radar Data. *Monthly Weather Review*, **140**, 1457–1475, https://doi.org/10.1175/MWR-D-11-00032.1.

Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between Forecasters and Research Scientists at the NSSL and SPC: The Spring Program: The Spring Program. *Bulletin of the American Meteorological Society*, **84**, 1797–1806, https://doi.org/10.1175/BAMS-84-12-1797.

——, and Coauthors, 2008: Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP. *Weather and Forecasting*, **23**, 931–952, https://doi.org/10.1175/WAF2007106.1.

——, and Coauthors, 2010: Assessing Advances in the Assimilation of Radar Data and Other Mesoscale Observations within a Collaborative Forecasting–Research Environment. *Weather and Forecasting*, **25**, 1510–1521, https://doi.org/10.1175/2010WAF2222405.1.

Kalina, E. A., I. Jankov, T. Alcott, J. Olson, J. Beck, J. Berner, D. Dowell, and C. Alexander, 2021: A Progress Report on the Development of the High-Resolution Rapid Refresh Ensemble. *Weather and Forecasting*, **36**, 791–804, https://doi.org/10.1175/WAF-D-20-0098.1.

Kalnay, E., and S.-C. Yang, 2010: Accelerating the spin-up of Ensemble Kalman Filtering. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1644–1651, https://doi.org/10.1002/qj.652.

Khairoutdinov, M., and D. Randall, 2006: High-Resolution Simulation of Shallow-to-Deep Convection Transition over Land. *Journal of the Atmospheric Sciences*, **63**, 3421–3436, https://doi.org/10.1175/JAS3810.1.

Kong, R., M. Xue, C. Liu, and Y. Jung, 2020: Comparisons of Hybrid En3DVar with 3DVar and EnKF for Radar Data Assimilation: Tests with the 10 May 2010 Oklahoma Tornado Outbreak. *Monthly Weather Review*, **149**, 21–40, https://doi.org/10.1175/MWR-D-20-0053.1.

Kucukkaraca, E., and M. Fisher, 2006: Use of analysis ensembles in estimating flow-dependent background error variances. 16, https://doi.org/10.21957/36n2z0p1p.

Lai, A., J. Gao, S. E. Koch, Y. Wang, S. Pan, A. O. Fierro, C. Cui, and J. Min, 2019: Assimilation of Radar Radial Velocity, Reflectivity, and Pseudo–Water Vapor for Convective-Scale NWP in a Variational Framework. *Monthly Weather Review*, **147**, 2877–2900, https://doi.org/10.1175/MWR-D-18-0403.1.

Li, H., C. Liu, M. Xue, J. Park, L. Chen, Y. Jung, R. Kong, and C.-C. Tong, 2022: Use of Power Transform Total Number Concentration as Control Variable for Direct Assimilation of Radar Reflectivity in GSI En3DVar and Tests with Six Convective Storms Cases. *Monthly Weather Review*, **150**, 821–842, https://doi.org/10.1175/MWR-D-21-0041.1.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **129**, 3183–3203, https://doi.org/10.1256/qj.02.132.

Lu, J., T. Feng, J. Li, Z. Cai, X. Xu, L. Li, and J. Li, 2019: Impact of Assimilating Himawari-8-Derived Layered Precipitable Water With Varying Cumulus and Microphysics Parameterization Schemes on the Simulation of Typhoon Hato. *Journal of Geophysical Research: Atmospheres*, **124**, 3050–3071, https://doi.org/10.1029/2018JD029364.

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated Electrification of a Small Thunderstorm with Two-Moment Bulk Microphysics. *Journal of the Atmospheric Sciences*, **67**, 171–194, https://doi.org/10.1175/2009JAS2965.1.

McPherson, R. A., and Coauthors, 2007: Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, **24**, 301–321, https://doi.org/10.1175/JTECH1976.1.

Michel, Y., T. Auligné, and T. Montmerle, 2011: Heterogeneous Convective-Scale Background Error Covariances with the Inclusion of Hydrometeor Variables. *Monthly Weather Review*, **139**, 2994–3015, https://doi.org/10.1175/2011MWR3632.1.

Miller, W. J. S., and Coauthors, 2022: Exploring the Usefulness of Downscaling Free Forecasts from the Warn-on-Forecast System. *Weather and Forecasting*, **37**, 181–203, https://doi.org/10.1175/WAF-D-21-0079.1.

Minnis, P., and Coauthors, 2011: CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data—Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, **49**, 4374–4400, https://doi.org/10.1109/TGRS.2011.2144601.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, **102**, 16663–16682, https://doi.org/10.1029/97JD00237.

Monteiro, M., and L. Berre, 2010: A diagnostic study of time variations of regionally averaged background error covariances. *Journal of Geophysical Research: Atmospheres*, **115**, https://doi.org/10.1029/2010JD014095.

Montmerle, T., and L. Berre, 2010: Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1408–1420, https://doi.org/10.1002/qj.655.

Nakanishi, M., and H. Niino, 2004: An Improved Mellor–Yamada Level-3 Model with Condensation Physics: Its Design and Verification. *Boundary-Layer Meteorology*, **112**, 1–31, https://doi.org/10.1023/B:BOUN.0000020164.04146.98.

NOAA National Centers for Environmental Information (NCEI), 2021: Billion-Dollar Weather and Climate Disasters. https://doi.org/10.25921/stkw-7w73.

Pan, S., and J. Gao, 2022: A Method for Assimilating Pseudo Dewpoint Temperature as a Function of GLM Flash Extent Density in GSI-Based EnKF Data Assimilation System – A Proof of Concept study. *Earth and Space Science*, **9**, e2022EA002378, https://doi.org/10.1029/2022EA002378.

——, ——, D. J. Stensrud, X. Wang, and T. A. Jones, 2018: Assimilation of Radar Radial Velocity and Reflectivity, Satellite Cloud Water Path, and Total Precipitable Water for Convective-Scale NWP in OSSEs. *Journal of Atmospheric and Oceanic Technology*, **35**, 67–89, https://doi.org/10.1175/JTECH-D-17-0081.1.

——, ——, T. A. Jones, Y. Wang, X. Wang, and J. Li, 2021: The Impact of Assimilating Satellite-Derived Layered Precipitable Water, Cloud Water Path, and Radar Data on Short-Range Thunderstorm Forecasts. *Monthly Weather Review*, **149**, 1359–1380, https://doi.org/10.1175/MWR-D-20-0040.1.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Monthly Weather Review*, **120**, 1747–1763, https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

Potvin, C. K., C. Broyles, P. S. Skinner, H. E. Brooks, and E. Rasmussen, 2019: A Bayesian Hierarchical Modeling Framework for Correcting Reporting Bias in the U.S. Tornado Database. *Weather and Forecasting*, **34**, 15–30, https://doi.org/10.1175/WAF-D-18-0137.1.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003a: Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part I: Spatially Homogeneous and Isotropic Gaussian Covariances. *Monthly Weather Review*, **131**, 1524–1535, https://doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2.

——, ——, ——, and ——, 2003b: Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially Inhomogeneous and Anisotropic General Covariances. *Monthly Weather Review*, **131**, 1536–1548, https://doi.org/10.1175//2543.1.

Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, **15**, 163–169, https://doi.org/10.1002/met.57.

Roberts, N. M., and H. W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Roebber, P. J., 2009: Visualizing Multiple Measures of Forecast Quality. *Weather and Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Schmit, T. J., M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier, 2005: INTRODUCING THE NEXT-GENERATION ADVANCED BASELINE IMAGER ON GOES-R. *Bulletin of the American Meteorological Society*, **86**, 1079–1096, https://doi.org/10.1175/BAMS-86-8-1079.

——, P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebair, 2017: A Closer Look at the ABI on the GOES-R Series. *Bulletin of the American Meteorological Society*, **98**, 681–698, https://doi.org/10.1175/BAMS-D-15-00230.1.

——, and Coauthors, 2019: Legacy Atmospheric Profiles and Derived Products From GOES-16: Validation and Applications. *Earth and Space Science*, **6**, 1730–1748, https://doi.org/10.1029/2019EA000729.

Skamarock, W., J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, X.-Y. Huang, and M. Duda, 2008: A Description of the Advanced Research WRF Version 3 (No. NCAR/TN-475+STR). *University Corporation for Atmospheric Research*, http://dx.doi.org/10.5065/D68S4MVH.

Skinner, P. S., and Coauthors, 2018: Object-Based Verification of a Prototype Warn-on-Forecast System. *Weather and Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smith, T. M., and Coauthors, 2014: Examination of a Real-Time 3DVAR Analysis System in the Hazardous Weather Testbed. *Weather and Forecasting*, **29**, 63–77, https://doi.org/10.1175/WAF-D-13-00044.1.

——, and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) Severe Weather and Aviation Products: Initial Operating Capabilities. *Bulletin of the American Meteorological Society*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Snyder, C., and F. Zhang, 2003: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter. *Monthly Weather Review*, **131**, 1663–1677, https://doi.org/10.1175//2555.1.

Stensrud, D. J., and J. Gao, 2010: Importance of Horizontally Inhomogeneous Environmental Initial Conditions to Ensemble Storm-Scale Radar Data Assimilation and Very Short-Range Forecasts. *Monthly Weather Review*, **138**, 1250–1272, https://doi.org/10.1175/2009MWR3027.1.

——, J.-W. Bao, and T. T. Warner, 2000: Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems. *Monthly Weather Review*, **128**, 2077–2107, https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.

——, and Coauthors, 2009: Convective-Scale Warn-on-Forecast System: A Vision for 2020. *Bulletin of the American Meteorological Society*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmospheric Research*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Stratman, D. R., and K. A. Brewster, 2015: Verification of 24 May 2011 Simulated Mesocyclones Using Various Microphysics Schemes at 1-km Grid Resolution. 27th Conference On Weather Analysis And Forecasting/23rd Conference On Numerical Weather Prediction, AMS https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273795.html (Accessed March 23, 2023).

Sun, J., 2005: Convective-scale assimilation of radar data: progress and challenges. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3439–3463, https://doi.org/10.1256/qj.05.149.

——, and N. A. Crook, 1997: Dynamical and Microphysical Retrieval from Doppler Radar Observations Using a Cloud Model and Its Adjoint. Part I: Model Development and Simulated Data Experiments. *Journal of the Atmospheric Sciences*, **54**, 1642–1661, https://doi.org/10.1175/1520-0469(1997)054<1642:DAMRFD>2.0.CO;2.

——, and ——, 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. *Journal of the Atmospheric Sciences*, **55**, 835–852, https://doi.org/10.1175/1520-0469(1998)055<0835:DAMRFD>2.0.CO;2.

——, and ——, 2001: Real-Time Low-Level Wind and Temperature Analysis Using Single WSR-88D Data. *Weather and Forecasting*, **16**, 117–132, https://doi.org/10.1175/1520-0434(2001)016<0117:RTLLWA>2.0.CO;2.

——, H. Wang, W. Tong, Y. Zhang, C.-Y. Lin, and D. Xu, 2016: Comparison of the Impacts of Momentum Control Variables on High-Resolution Variational Data Assimilation and Precipitation Forecasting. *Monthly Weather Review*, **144**, 149–169, https://doi.org/10.1175/MWR-D-14-00205.1.

——, Y. Zhang, J. Ban, J.-S. Hong, and C.-Y. Lin, 2020: Impact of Combined Assimilation of Radar and Rainfall Data on Short-Term Heavy Rainfall Prediction: A Case Study. *Monthly Weather Review*, **148**, 2211–2232, https://doi.org/10.1175/MWR-D-19-0337.1.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Monthly Weather Review*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Tong, M., and M. Xue, 2005: Ensemble Kalman Filter Assimilation of Doppler Radar Data with a Compressible Nonhydrostatic Model: OSS Experiments. *Monthly Weather Review*, **133**, 1789–1807, https://doi.org/10.1175/MWR2898.1.

Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer Beware: Some Words of Caution on the Use of Severe Wind Reports in Postevent Assessment and Research. *Weather and Forecasting*, **21**, 408–415, https://doi.org/10.1175/WAF925.1.

Wang, H., J. Sun, X. Zhang, X.-Y. Huang, and T. Auligné, 2013: Radar Data Assimilation with WRF 4D-Var. Part I: System Development and Preliminary Testing. *Monthly Weather Review*, **141**, 2224–2244, https://doi.org/10.1175/MWR-D-12-00168.1.

Wang, P., J. Li, B. Lu, T. J. Schmit, J. Lu, Y.-K. Lee, J. Li, and Z. Liu, 2018: Impact of Moisture Information From Advanced Himawari Imager Measurements on Heavy Precipitation Forecasts in a Regional NWP Model. *Journal of Geophysical Research: Atmospheres*, **123**, 6022–6038, https://doi.org/10.1029/2017JD028012.

Wang, X., C. Snyder, and T. M. Hamill, 2007: On the Theoretical Equivalence of Differently Proposed Ensemble–3DVAR Hybrid Analysis Schemes. *Monthly Weather Review*, **135**, 222–227, https://doi.org/10.1175/MWR3282.1.

——, D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment. *Monthly Weather Review*, **136**, 5116–5131, https://doi.org/10.1175/2008MWR2444.1.

——, ——, ——, and ——, 2008b: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments. *Monthly Weather Review*, **136**, 5132–5147, https://doi.org/10.1175/2008MWR2445.1.

——, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2009: A Comparison of the Hybrid and EnSRF Analysis Schemes in the Presence of Model Errors due to Unresolved Scales. *Monthly Weather Review*, **137**, 3219–3232, https://doi.org/10.1175/2009MWR2923.1.

Wang, Y., and X. Wang, 2017: Direct Assimilation of Radar Reflectivity without Tangent Linear and Adjoint of the Nonlinear Observation Operator in the GSI-Based EnVar System: Methodology and Experiment with the 8 May 2003 Oklahoma City Tornadic Supercell. *Monthly Weather Review*, **145**, 1447–1471, https://doi.org/10.1175/MWR-D-16-0231.1.

——, and ——, 2021: Development of Convective-Scale Static Background Error Covariance within GSI-Based Hybrid EnVar System for Direct Radar Reflectivity Data Assimilation. *Monthly Weather Review*, **149**, 2713–2736, https://doi.org/10.1175/MWR-D-20-0215.1.

Wang, Y., J. Gao, P. S. Skinner, K. Knopfmeier, T. Jones, G. Creager, P. L. Heiselman, and L. J. Wicker, 2019: Test of a Weather-Adaptive Dual-Resolution Hybrid Warn-on-Forecast Analysis and Forecast System for Several Severe Weather Events. *Weather and Forecasting*, **34**, 1807–1827, https://doi.org/10.1175/WAF-D-19-0071.1.

Wattrelot, E., O. Caumont, and J.-F. Mahfouf, 2014: Operational Implementation of the 1D+3D-Var Assimilation Method of Radar Reflectivity Data in the AROME Model. *Monthly Weather Review*, **142**, 1852–1873, https://doi.org/10.1175/MWR-D-13-00230.1.

Weisman, M. L., and J. B. Klemp, 1982: The Dependence of Numerically Simulated Convective Storms on Vertical Wind Shear and Buoyancy. *Monthly Weather Review*, **110**, 504–520, https://doi.org/10.1175/1520-0493(1982)110<0504:TDONSC>2.0.CO;2.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman Filter Analyses and Forecasts of a Severe Mesoscale Convective System Using Different Choices of Microphysics Schemes. *Monthly Weather Review*, **142**, 3243–3263, https://doi.org/10.1175/MWR-D-13-00260.1.

——, K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-Scale Data Assimilation and Ensemble Forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar Data Experiments. *Weather and Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Wilson, K. A., and Coauthors, 2019: Exploring Applications of Storm-Scale Probabilistic Warn-on-Forecast Guidance in Weather Forecasting. *Virtual, Augmented and Mixed Reality. Applications and Case Studies*, J.Y.C. Chen and G. Fragomeni, Eds., *Lecture Notes in Computer Science*, Cham, Springer International Publishing, 557–572.

Xiao, Q., and J. Sun, 2007: Multiple-Radar Data Assimilation and Short-Range Quantitative Precipitation Forecasting of a Squall Line Observed during IHOP_2002. *Monthly Weather Review*, **135**, 3381–3404, https://doi.org/10.1175/MWR3471.1.

——, Y.-H. 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. *Journal of Applied Meteorology and Climatology*, **44**, 768–788, https://doi.org/10.1175/JAM2248.1.

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) – A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification. *Meteorol Atmos Phys*, **75**, 161–193, https://doi.org/10.1007/s007030070003.

——, and Coauthors, 2001: The Advanced Regional Prediction System (ARPS) – A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteorol Atmos Phys*, **76**, 143–165, https://doi.org/10.1007/s007030170027.

Yussouf, N., and D. J. Stensrud, 2010: Impact of Phased-Array Radar Observations over a Short Assimilation Period: Observing System Simulation Experiments Using an Ensemble Kalman Filter. *Monthly Weather Review*, **138**, 517–538, https://doi.org/10.1175/2009MWR2925.1.

——, and K. H. Knopfmeier, 2019: Application of the Warn-on-Forecast system for flash-flood-producing heavy convective rainfall events. *Quarterly Journal of the Royal Meteorological Society*, **145**, 2385–2403, https://doi.org/10.1002/qj.3568.

——, T. A. Jones, and P. S. Skinner, 2020a: Probabilistic high-impact rainfall forecasts from landfalling tropical cyclones using Warn-on-Forecast system. *Quarterly Journal of the Royal Meteorological Society*, **146**, 2050–2065, https://doi.org/10.1002/qj.3779.

——, K. A. Wilson, S. M. Martinaitis, H. Vergara, P. L. Heinselman, and J. J. Gourley, 2020b: The Coupling of NSSL Warn-on-Forecast and FLASH Systems for Probabilistic Flash Flood Prediction. *Journal of Hydrometeorology*, **21**, 123–141, https://doi.org/10.1175/JHM-D-19-0131.1.

Zhang, F., M. Minamide, R. G. Nystrom, X. Chen, S.-J. Lin, and L. M. Harris, 2019a: Improving Harvey Forecasts with Next-Generation Weather Satellites: Advanced Hurricane

Analysis and Prediction with Assimilation of GOES-R All-Sky Radiances. *Bulletin of the American Meteorological Society*, **100**, 1217–1222, https://doi.org/10.1175/BAMS-D-18-0149.1.

Zhang, Y., F. Zhang, and D. J. Stensrud, 2018: Assimilating All-Sky Infrared Radiances from GOES-16 ABI Using an Ensemble Kalman Filter for Convection-Allowing Severe Thunderstorms Prediction. *Monthly Weather Review*, **146**, 3363–3381, https://doi.org/10.1175/MWR-D-18-0062.1.

——, D. J. Stensrud, and F. Zhang, 2019b: Simultaneous Assimilation of Radar and All-Sky Satellite Infrared Radiance Observations for Convection-Allowing Ensemble Analysis and Prediction of Severe Thunderstorms. *Monthly Weather Review*, **147**, 4389–4409, https://doi.org/10.1175/MWR-D-19-0163.1.

Zhao, J., J. Gao, T. A. Jones, and J. Hu, 2021a: Impact of Assimilating High-Resolution Atmospheric Motion Vectors on Convective Scale Short-Term Forecasts: 1. Observing System Simulation Experiment (OSSE). *Journal of Advances in Modeling Earth Systems*, **13**, e2021MS002484, https://doi.org/10.1029/2021MS002484.

——, ——, ——, and ——, 2021b: Impact of Assimilating High-Resolution Atmospheric Motion Vectors on Convective Scale Short-Term Forecasts: 2. Assimilation Experiments of GOES-16 Satellite Derived Winds. *Journal of Advances in Modeling Earth Systems*, **13**, e2021MS002486, https://doi.org/10.1029/2021MS002486.

——, ——, T. Jones, and J. Hu, 2022: Impact of Assimilating High-Resolution Atmospheric Motion Vectors on Convective Scale Short-Term Forecasts: 3. Experiments With Radar Reflectivity and Radial Velocity. *Journal of Advances in Modeling Earth Systems*, **14**, e2022MS003246, https://doi.org/10.1029/2022MS003246.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the New NCEP Global Ensemble Forecast System in a Parallel Experiment. *Weather and Forecasting*, **32**, 1989–2004, https://doi.org/10.1175/WAF-D-17-0023.1.