

Evaluating Forecast Performance and Sensitivity to the GSI EnKF Data Assimilation Configuration for the 28 – 29 May 2017 Mesoscale Convective System Case

Jonathan Labriola^{1,2}, Youngsun Jung², Chengsi, Liu², and Ming Xue^{1,2}

School of Meteorology¹ and Center for Analysis and Prediction of Storms²
University of Oklahoma, Norman, OK 73072

Submitted to Weather and Forecasting
May 2020

Revised and accepted November 2020

Corresponding author address:

Jonathan Labriola

Center for Analysis and Prediction of Storms

University of Oklahoma,

120 David Boren Blvd, Room 2500, Norman OK 73072

jonathan.labriola@noaa.gov

Abstract

In an effort to improve radar data assimilation configurations for potential operational implementation, GSI EnKF data assimilation experiments based on the operational system employed by the Center for Analysis and Prediction of Storms (CAPS) realtime Spring Forecast Experiments are performed. These experiments are followed by 6-hour forecasts for an MCS on 28 – 29 May 2017. Configurations examined include data thinning, covariance localization radii and inflation, observation error settings, and data assimilation frequency for radar observations.

The results show experiments that assimilate radar observations more frequently (i.e., 5 – 10 minutes) are initially better at suppressing spurious convection. However, assimilating observations every 5 minutes causes spurious convection to become more widespread with time, and modestly degrades forecast skill through the remainder of the forecast window. Ensembles that assimilate more observations with less thinning of data or use a larger horizontal covariance localization radius for radar data predict fewer spurious storms and better predict the location of observed storms. Optimized data thinning and horizontal covariance localization radius have positive impacts on forecast skill during the first forecast hour that are quickly lost due to the growth of forecast error. Forecast skill is less sensitive to the ensemble spread inflation factors and observation errors tested during this study. These results provide guidance towards optimizing the configuration of the GSI EnKF system. Among DA the configurations tested, the one employed by the CAPS Spring Forecast Experiment produces the most skilled forecasts while remaining computationally efficient for realtime use.

Significance Statement

High-resolution ensemble forecasts that can skillfully predict thunderstorms provide an opportunity for warning severe weather further in advance. However, forecast accuracy is dependent upon many factors including the data assimilation system used to create the forecasts. This study optimizes the design of a realtime ensemble forecast and data assimilation system for a

severe weather event where a line of thunderstorms produced hail, wind, and tornadoes in the Southern United States. 0 – 6 hour forecasts predict the thunderstorms with moderate skill. Forecast accuracy is most sensitive to how frequently the data assimilation system assimilates radar observations, the degree to which radar observations are thinned, and the covariance localization radius. Results of this study can be used to design future realtime forecast systems for severe weather events.

1. Introduction

Thunderstorms have caused approximately \$200 billion in property damage (NOAA 2019) since the beginning of the millennium. Most severe weather hazards occur with only tens of minutes of advance warning (average tornado warning lead time ≈ 17 minutes) because the National Weather Service issues warnings based upon the detection of a hazard, either from surface reports, radar detections, or an imminent threat as determined by the forecaster (Stensrud et al. 2013). Based on the current warning paradigm (i.e., warn-on-detection), the average warning lead time for detected tornadoes has remained relatively constant since 1986 (Stensrud et al. 2013) and is unlikely to substantially increase without degrading warning skill (e.g., increasing the number of false alarms) (Brooks 2004). To extend warning lead time, the NWS is working on adopting the Warn-on-Forecast paradigm (Stensrud et al. 2009, 2013) where frequently updated, convection-allowing model (CAM) ensemble forecast guidance is incorporated into the warning process. Due to the rapid error growth of convective scale forecasts, in addition to limiting sources of forecast model error, one key component of a skilled forecast system is to limit initial condition errors by developing a skilled data assimilation (DA) system.

Various Warn-on-Forecast prototype systems (e.g., Yussouf et al. 2013; Wheatley et al. 2015; Jones et al. 2016; Snook et al. 2016; Lawson et al. 2018; Labriola et al. 2019; Stratman et al. 2020) use an ensemble Kalman filter (EnKF; Evensen 1994, 2003) to assimilate observations and initialize 0 – 6 hour CAM forecasts. An EnKF is a Monte Carlo implementation of Kalman Filter (Kalman 1960), which is itself based on optimal estimation theory. EnKFs derive forecast background error covariance statistics from an ensemble of short-term forecasts instead of trying to evolve it in time using expensive prediction equations (Evensen 1994, 2003). This technique is particularly attractive when compared to static covariance models, which often assume the background error covariance is spatially homogenous and isotropic (Parrish and Derber 1992; Purser et al. 2003a). Error covariances at the convective scale are inhomogenous and anisotropic, and are thus better represented by flow-dependent covariances derived from the forecast ensemble. The filter can update unobserved model state variables using often a limited number of observed parameters. This is the case with radar observations where radar reflectivity (Z) and radial velocity (V_r) are primarily available at the convective scale. Many observation system simulation experiments (OSSEs) and real-data experiments have demonstrated that assimilating Z and V_r observations improves estimated thermodynamic, dynamic, and microphysical state variables (e.g, Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Caya et al. 2005; Tong and Xue 2005; Xue et al. 2010; Dawson et al. 2012; Jung et al. 2012; Johnson et al. 2015; Snook et al. 2015; Supinie et al. 2016; Wang and Wang 2017). Assimilating other convective-scale observations such as polarimetric radar observations and satellite radiances may also decrease initial conditions errors (Jung et al. 2008; Jones et al. 2018b; Zhang et al. 2019; Putnam et al. 2019). While EnKF systems have demonstrated their potential to assimilate radar observations, performance of the DA system and skill of subsequent forecasts is contingent upon the optimization of the DA

configurations. Each DA system contains many tuning parameters and further, the optimization is sensitive to the model configurations and observations assimilated (e.g., Dowell et al. 2004; Tong and Xue 2005; Sobash and Stensrud 2013; Johnson and Wang 2017).

Previous studies (e.g., Zhang et al. 2004; Dowell and Wicker 2009; Johnson and Wang 2017) have found DA system configurations can have a substantial impact on the subsequent forecast skill. For example, EnKF experiments employ covariance inflation and localization to mitigate poor estimates of the background error statistics. Additional considerations when assimilating radar observations include observation error specification, data thinning interval, and assimilation frequency. The remainder of this introduction discusses these commonly used techniques used to improve EnKF initial condition estimates.

EnKF systems are often under-dispersive due to the limited ensemble size and unaccounted model errors (e.g., Houtekamer and Mitchell 1998; Romine et al. 2014), which makes the model over-confident and reduces the influence of observations. Repeated assimilation of dense observations can reduce ensemble spread so much that the EnKF can no longer effectively assimilate observations because it becomes over-confident about the ensemble forecast, leading to filter divergence (Jazwinski 1970; Anderson and Anderson 1999). Spread maintenance algorithms artificially increase ensemble variance to better represent the true uncertainty of the atmospheric state during DA (Anderson 2001; Hamill et al. 2001). Some spread inflation algorithms add noise to the posterior ensemble (e.g., Dowell and Wicker 2009; Sobash and Wicker 2015) while some, such as the relaxation-to-prior-spread (RTPS) algorithm (Whitaker and Hamill 2012), increase the spread of posterior ensemble to a fraction of the spread of the prior ensemble. All methods are met with varying degrees of success between experiments, and often a combination of these methods is used to optimize EnKF performance (Jung et al. 2012).

The quality of background error covariance is often poor because it is computed from an insufficient number of ensemble members (e.g., Houtekamer and Mitchell 1998; Hamill et al. 2001). Such sampling errors cause grid points to become spuriously correlated with distant observations, and consequently the filter erroneously updates the background forecasts far from the assimilated observation. A distance-based Gaussian weighting function (Gaspari and Cohn 1999) is often used to localize the influence of observations; the optimal cutoff radius is contingent upon many factors including ensemble size, observation type, density, location, computational cost, model resolution, and correlation length scale from model dynamics (Sobash and Stensrud 2013; Ying et al. 2018). Most studies (e.g., Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Caya et al. 2005; Tong and Xue 2005; Aksoy et al. 2009; Jung et al. 2012; Sobash and Stensrud 2013; Johnson et al. 2015; Wheatley et al. 2015) use a small cutoff radius for radar observations (4 – 18 km horizontal, 4 – 8 km vertical) because they are dense and the convective scale flows have small spatial correlation length scales.

Radar data require pre-processing prior to assimilation because observations are provided on radar polar coordinates at resolutions finer than most forecast models. Most EnKF experiments assimilate coarsened radar data that are interpolated to a regular grid or the model grid itself. For example, Xue et al. (2006) assimilates radar observations that are interpolated to the horizontal model grid but kept on the radar elevation levels in the vertical. After interpolation, some experiments assimilate a fraction of the available radar observations to decrease the computational expense (Gao and Xue 2008) or mitigate the effects of spatially correlated observation errors by removing neighboring observations (Chang et al. 2014). Thinning assimilated observations can also decrease the ensemble spread reduction during EnKF DA, though this does not necessarily

improve forecast skill (Aksoy et al. 2012). Radar data are most commonly thinned at regular user-specified intervals throughout the model domain. Despite potential benefits, thinning observations can remove important in-storm observational information during DA and limit the analysis skill.

Modeling errors on the convective scale in particular are poorly understood and thus represent a challenge when assimilating radar observations. While unrealistic, most EnKF-based studies assume radar observation errors are Gaussian in nature, constant in standard deviation, and spatially uncorrelated. Well-calibrated WSR-88D radars have observational errors of approximately 1 dBZ (Z) and 1 m s^{-1} (V_r) (Doviak and Zrnić 1993; Ryzhkov et al. 2005), but DA experiments typically assume larger errors to account for representativeness errors and other uncertainties (such as those with observation operators). Increasing observation errors also alleviates ensemble under-dispersion by decreasing spread reduction during DA and is shown to improve performance of ensemble analyses (e.g., Dowell et al. 2004; Snook et al. 2013).

To obtain accurate initial conditions for convective storms, radar observations are often assimilated at high frequencies. Real-time systems often assimilate radar observations every 15 minutes (e.g., Wheatley et al. 2015; Johnson et al. 2017; Snook et al. 2019), although improvements to computational infrastructure and new weather radar technology such as phased-array radars (Weber et al. 2007; Zrnic et al. 2007; Heinselman and Torres 2011; Curtis and Torres 2011) provide the opportunity to assimilate observations more frequently in the future. OSSEs that assimilate rapid-scanning radar information note a decrease in the spin-up time for convection and a reduction in errors in both observed and unobserved variables (Zhang et al. 2004; Xue et al. 2006; Yussouf and Stensrud 2010). Real-data cases have successfully assimilated radar information at relatively high frequencies (≤ 5 minutes) (Snook et al. 2011, 2016; Jung et al. 2012; Supinie et al. 2017; Labriola et al. 2017; Stratman et al. 2020); however, rapid updates can introduce imbalances that are unable to adjust to the model before the next DA cycle and can degrade both analysis and forecast skill (Wang et al. 2013; Johnson and Wang 2017) unless extra care is taken.

Most studies designed within the Warn-on-Forecast framework evaluate forecast and analysis sensitivity to a limited number of EnKF parameters for radar DA including covariance inflation (Dowell and Wicker 2009), covariance localization (Sobash and Stensrud 2013), observation data thinning (Gao and Xue 2008), DA frequency (Yussouf and Stensrud 2010; Stratman et al. 2020), and prescribed observation errors (Gao and Xue 2008; Dowell et al. 2011; Snook et al. 2013). Evaluating only a few parameters allows these studies to analyze forecast sensitivities in greater detail; however, the forecast skill remains suboptimal given the other parameters are untuned. Further, these experiments do not consider the computational limitations when running a realtime CAM forecast system over a large-area domain such as the contiguous United States (CONUS). This is the first study to tune the EnKF radar DA parameters for a realtime short-term CAM forecast system that can be deployed over the full CONUS domain. Comparing forecasts initialized using many different EnKF configurations demonstrates how each parameter impacts forecast skill, and which configuration optimizes forecast skill.

This study evaluates short-term (0 - 6 hour) forecasts for a mesoscale convective system (MCS) event on 28 – 29 May 2017 that are initialized using an EnKF system based on the NCEP operational gridpoint statistical interpolation (GSI) framework, using a forecast ensemble that is based upon the Center for Analysis and Prediction of Storms (CAPS) Storm Scale Ensemble Forecast (SSEF) run during the 2017 Hazardous Weather Testbed Spring Forecast Experiment (CAPS 2017). The rest of the paper is organized as follows: section 2 provides a brief overview of

the case study, a description of the experiments, and verification procedures. The control experiment results including objective and subjective forecast evaluations are discussed in section 3, while section 4 discusses the radar DA parameter sensitivity experiment results. Section 5 summarizes the results of this study and discusses potential future directions for research.

2. Event overview, experiment configuration, and verification methodology

a.) Case overview

The focus of this study is to forecast the evolution of an MCS and nearby isolated convective storms that produced strong winds, hail, and tornadoes in Texas, Louisiana, and Mississippi on 28 – 29 May 2017 (Fig. 1). Upper-level wind patterns were favorable for convective development during this event; a trough was located over the Central United States, and Texas was to the right of a jet entrance region. At approximately 2000 UTC on 28 May multiple thunderstorms initiated along a frontal boundary extending from Arkansas towards the Texas-Mexico border. Thunderstorms that initiated over eastern Texas quickly grew in scale between 2200 and 0000 UTC as they ingested unstable air ($CAPE > 2200 \text{ J kg}^{-1}$). The storms eventually merged to form a squall line that impacted Louisiana and Mississippi between 0000 and 0600 UTC on 29 May (Fig. 2). Storms embedded within the line produced several weak tornadoes and multiple wind reports extending from Shreveport to Jackson (Fig. 1). Between 0000 and 0200 UTC isolated thunderstorms in Texas (Fig. 2a-b) produced multiple hail and wind reports (Fig. 1), but by 0300 UTC (Fig. 2c) many of these storms weakened and formed a large region of mostly stratiform precipitation. With the exception of convection near Houston and the MCS located east of Jackson, precipitation remains stratiform through 0600 UTC (Fig. 2d).

The presence of active MCS convection within the DA period before the forecast initial condition time, and during several hours of forecast make this case suitable for investigating the impact of DA and its configurations on subsequent storm forecasts. Forecasts focus on the Texas, Louisiana, and Mississippi regions. Experiment conclusions are limited because this is a single case study; however, analyzing a single case allows this study to test multiple EnKF parameters and determine an optimal radar DA configuration that can be employed in the future.

b.) Control experiment DA System Settings

This subsection describes the control experiment (hereafter referred to as CTRL) DA configuration (**Error! Reference source not found.**), modified DA configuration parameters are discussed in section 2d. This study uses the GSI EnKF system for DA. The GSI system performs observation quality control (QC) and applies forward observation operators to the model background to generate observation priors. A version of the ensemble square root filter (Whitaker and Hamill 2002) that pre-calculates observation priors then updates them together with the state variables within the filter (Anderson and Collins 2007) is used. Recently, the GSI EnKF system was updated to assimilate radar observations; radar forward operators were added to the GSI suite and the EnKF system was given the capability to update hydrometeor variables (e.g., Johnson et al. 2015; Jones et al. 2018a; Tong et al. 2020). CAPS recently added radar reflectivity forward operator consistent with the Thompson microphysics scheme. The new operator follows Jung et al. (2008a), which uses the T-matrix method (Vivekanandan et al. 1991; Bringi and Chandrasekar 2001) for raindrops and Rayleigh scattering approximation for ice hydrometeor species (i.e., snow, hail, graupel) to calculate scattering amplitudes. In this study, radar observations undergo automatic QC including velocity dealiasing, using a procedure developed by CAPS (Brewster et al. 2005). These enhancements allow the GSI EnKF system to be used in several storm-scale modeling studies (e.g., Johnson et al. 2015; Johnson and Wang 2017; Johnson et al. 2017; Jones et

al. 2018b; Jung et al. 2018a,b; Chipilski et al. 2020).

CTRL follows the 2017 CAPS storm-scale ensemble configuration (CAPS 2017), except the DA parameters (**Error! Reference source not found.**), which are inherited from the GSI EnKF system used to generate the 2019 CAPS storm-scale ensemble initial conditions (Clark et al. 2019). CTRL is initialized from the 1800 UTC North American Mesoscale (NAM) analysis plus perturbations derived from 3-hour forecasts of the 1500 UTC cycle Short-Range Ensemble Forecast (SREF) (Table 1). Perturbations are constructed by taking the difference between 2 selected members among the 24 SREF members. The differences and their negative counterparts, which are the same in magnitude but opposite in direction, are added to the NAM analysis to produce 2 perturbed members. This is repeated to create 39 perturbed members, the first ensemble member remains unperturbed. SREF and NAM forecasts provide lateral boundary condition information. For more details on the initial and lateral boundary conditions, see Table 3 of the “CAPS Spring Forecast Experiment Program Plan” (CAPS 2017).

Conventional observations (e.g., surface stations, buoys, soundings) are assimilated hourly between 1900 – 0000 UTC, and radar (Z and V_r) observations are assimilated every 15 minutes in the last hour. A schematic on the setup of DA cycles is shown in Fig. 3. Covariance localization uses the Gaspari and Cohn (1999) weighting function with user-specified cutoff radii. In this study the horizontal cutoff radius is 300 km for conventional observations and 12 km for radar observations. The vertical cutoff radius for all observations is 0.7 scale height (natural log of pressure). The RTPS covariance inflation algorithm is used to restore the spread of the analysis to 99% of the background spread. At 0000 UTC an ensemble of 10 free forecasts are initialized from the final analyses and run to 0600 UTC. The ensemble members are initialized following the same procedure as the CAPS realtime system; eight forecasts are initialized from the ensemble member analyses selected considering physics and graupel density diversity and two are initialized from the ensemble mean analysis (Table 4 of CAPS [2017]). The 0000 UTC NAM forecast and 2100 UTC SREF forecasts provide lateral boundary conditions. Ensemble members selected to initialize forecasts employ a diverse suite of model physics options, and are chosen to enhance forecast diversity. The forecast configuration is discussed in the following subsection.

For radar DA, Z observations are from the Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) products. The MRMS system performs QC on 140 WSR-88D radars and generates a mosaic of the observations on a three-dimensional grid with a horizontal resolution of 0.01° latitude \times 0.01° longitude and 33 vertical levels. V_r observations are processed using CAPS’s software package that includes QC. The data are interpolated to the model grid horizontally and kept on radar elevation levels in the vertical for each radar site (Xue et al. 2006). Z and V_r observations are thinned horizontally to every 6 km during assimilation. Z observations are also thinned vertically to every 1 km in radar echoes ($Z > -90$ dBZ) and to every 2 km in clear air regions ($Z < -90$ dBZ). Such thinning is done to fit the GSI EnKF analyses into available computer memory, which is necessary when a large continent-size domain is used during, e.g., the CLUE experiment (the GSI memory usage is inefficient). The V_r and Z observation errors are assumed to be 3 m s^{-1} and 6 dBZ, respectively, during DA.

c.) Prediction model settings

Aside from the smaller forecast domain focused on the MCS impacted region, the grid specifications and model physics largely follow the CAPS GSI EnKF-initialized ensemble (Jung et al. 2018b) that is part of the CLUE (Clark et al. 2018). Forecasts are run using the advanced research Weather Research and Forecast Model (WRF-ARW; Skamarock et al. 2008) version

3.8.1. The forecast domain spans much of Texas, Louisiana, and Mississippi (Fig. 1) with 433×241 grid points in the horizontal and 51 vertical levels in sigma-pressure coordinates. The horizontal grid spacing is 3 km. The vertical grid follows the 2017 CLUE configuration where the finest vertical grid spacing is located near the surface. All ensemble members use the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) for general circulation model (RRTMG; Iacono et al. 2008) to represent short- and longwave radiation. During DA (1800 – 0000 UTC), forecasts are run with the aerosol aware Thompson microphysics scheme (Thompson et al. 2008; Thompson and Eidhammer 2014) but with varying graupel density across the members. 0000 UTC EnKF-initialized forecasts vary microphysics schemes (Thompson [Thompson et al. 2008], Morrison [Morrison and Grabowski 2008], Milbrandt and Yau [MY; Milbrandt and Yau 2005], National Severe Storms Laboratory [NSSL; Mansell et al. 2010], and Predicted Particle Properties [P3; Morrison and Milbrandt 2015]) between ensemble members. Ensemble members also employ different planetary boundary layer physics parameterizations (Mellor-Yamada-Janjić [MYJ; Janjić 1990, 1996, 2001], Yonsei University [YSU; Hong et al. 2006], and Mellor-Yamada-Nakanishi-Niino [MYNN; Nakanishi and Niino 2009]) and the Noah land surface model (Chen and Dudhia 2001) both during the DA and forecast periods. An in-depth overview of ensemble member physics options is provided in CAPS (2017).

d.) DA configuration experiments

Sensitivity experiments are designed to evaluate the impact of EnKF DA configurations on the forecast skill. Each experiment repeats the CTRL DA procedure (**Error! Reference source not found.**) except one aspect of the DA configuration is modified when assimilating radar observations. Experiment results analyze forecast sensitivity to data thinning, covariance localization, covariance inflation, observation errors, and assimilation frequency for radar observations. These parameters are tested over a range of values employed by previous studies (e.g., Gao and Xue 2008; Dowell et al. 2011; Sobash and Stensrud 2013; Snook et al. 2013; Wheatley et al. 2015; Supinie et al. 2017; Stratman et al. 2020). The remainder of the subsection defines the range of tested parameter values.

Data thinning experiments assimilate radar observations thinned over increasingly large intervals either horizontally (3 km, 6 km, 9 km) or vertically (500m, 1 km, 2 km) and are labeled by the direction observations are thinned in the horizontal and vertical directions (ThinH3V1, ThinH6V1, ThinH9V1, ThinH3V0.5, ThinH3V2). Following CTRL, all experiments thin Z observations over twice as large of a vertical interval in clear air regions. V_r observations are not further thinned during this study because the data are preserved on about a dozen radar elevations and thus already sparser than MRMS Z . The covariance localization experiments vary the localization radius for radar observations either horizontally (6 km, 12 km, 18 km) or vertically (0.4 scalar height, 0.7 scalar height, 1.0 scalar height) during DA and are referred to by the length of the applied radius in the horizontal and vertical directions (CovH6V0.7, CovH12V0.7, CovH18V0.7, CovH12V0.4, CovH18V1.0). The covariance inflation experiments vary the percentage (80%, 90%, 99%, 110%) that posterior ensemble spread is relaxed to that of the prior ensemble via the RTPS algorithm when radar observations are assimilated (2300 – 0000 UTC). Experiments are referred to by the inflation factor applied (Inf80, Inf90, Inf99, Inf110). Observation error experiments vary observation errors for V_r (3 m s^{-1} , 6 m s^{-1}) and Z (6 dBZ, 9dBZ) and are referred to by the assumed errors (3ms6dBZ, 6ms9dBZ).

Experiments that assimilate radar data at higher frequencies have demonstrated mixed success in previous studies. Stratman et al. (2020) suggest that assimilating observed Z more

frequently can cause predicted storms to spin up more quickly and better suppress spurious convection. Frequently assimilating observations can also introduce imbalances into the ensemble that propagate with time to degrade forecast skill (Hu and Xue 2007; Johnson and Wang 2017). To determine the impact of radar DA frequency, Z and V_r observations are assimilated at 5, 10 and 15 minute intervals during the final hour of DA (2300 – 0000 UTC). These intervals roughly correspond to the frequency that WSR-88D radars sample the atmosphere (~ 5 minutes) and the frequency some current real-time systems assimilate radar observations (e.g., Wheatley et al. 2015; Jung et al. 2018a). DA frequency experiments are referred to by how frequently observations are assimilated (5Min, 10Min, 15Min).

e.) Forecast evaluation

The 0000 – 0600 UTC forecast Z is subjectively and objectively verified against observations to evaluate the predicted evolution of storm structure. This forecast evaluation period corresponds to when the impact of assimilated radar observations is most prominent (Kain et al. 2010), and when the Warn-on-Forecast paradigm is expected to offer the most benefit to operational forecasters (e.g., Stensrud et al. 2009).

During forecast evaluation small errors in storm placement can substantially degrade objective performance by double penalizing a forecast i.e., adjoining grid points may receive a false negative and false positive. This problem is exacerbated when verifying localized events (e.g., convective storms) but can be ameliorated by verifying the occurrence of an event within a prescribed radius. The neighborhood maximum ensemble probability (NMEP; Schwartz et al. 2010; Schwartz and Sobash 2017) method is used to generate probabilistic forecasts. This study verifies the probability of Z exceeding 40 dBZ ($P[Z > 40 \text{ dBZ}]$) within a 12-km neighborhood. This neighborhood radius reduces the impact of small forecast displacement errors but ensures the short-term forecast remains precise enough to detect storm structures related to localized severe weather impacts. A Gaussian filter with a smoothing length scale of 12-km is applied to the output to smooth probabilistic forecasts. It is noted that this study verifies a relatively large Z threshold; this is done to verify the location of predicted storm cores and to eliminate regions of stratiform precipitation from the statistics.

Probabilistic forecast skill is objectively evaluated using the Brier Skill Score (BSS; Brier 1950), which can be decomposed into three distinct components (Murphy 1973): reliability, resolution, and uncertainty (Table 2 **Error! Reference source not found.**). Reliability is the difference between predicted probability and observed frequency, forecast skill improves when this difference is minimized. Resolution, which should be maximized to improve ensemble performance, is the difference between the climatological probability and the observed relative frequency for a given probability threshold. Unlike the other two components, forecast uncertainty cannot be changed through calibration and is a function of the climatological probability. Reliability diagrams provide a visual representation of the BSS by plotting forecast probability against observed frequency over increasingly large thresholds. Forecast probability is equal to observed frequency in an unbiased system, and the reliability curve falls along the one-to-one line. If forecast probability is larger (smaller) than observed frequency then the curve falls below (above) the one-to-one line and the forecast is high (low) biased. Reliability diagrams also provide the frequency occurrence of each probability to evaluate forecast sharpness.

3. CTRL results

Shortly after the start of the forecast at 0015 UTC (Fig. 4a) CTRL predicts the $P(Z > 40 \text{ dBZ})$ to be relatively high (> 0.90) within the confines of observed storms including the MCS located near Shreveport, Louisiana. Although CTRL predicts the $P(Z > 40 \text{ dBZ})$ to exceed 0.7 for isolated storms located along the Mexico-Texas border at 0015 UTC (Fig. 4a), probabilities diminish by 0100 UTC (Fig. 4b). This region is cooler and drier than locations near the coast, suggesting that the environment is less conducive for convective development. Several forecasts in the CTRL ensemble (Fig. 4a) also predict spurious storms to initiate to the north and east of San Antonio. Reliability diagrams show probabilistic 0015 UTC forecasts exhibit a slight overprediction bias for both low and moderate probability threshold events ($P[Z > 40 \text{ dBZ}] < 0.6$) (Fig. 5) primarily due to the spurious convection in Texas. Despite this bias, the BSS at 0015 UTC (0.52) is relatively high compared to later forecast times because CTRL is unbiased for high-probability events ($P[Z > 40 \text{ dBZ}] > 0.8$).

As the MCS moves eastward through Louisiana (0100 – 0300 UTC), CTRL predicts the $P(Z > 40 \text{ dBZ})$ to remain relatively high (> 0.7) near the observed storms (Fig. 4b-c). Some CTRL members predict the isolated convection near San Antonio to grow in scale and form a line of storms that increases moderate forecast probabilities ($P[Z > 40 \text{ dBZ}] > 0.4$) in Central Texas by 0100 UTC (Fig. 4b). This line of storms is spurious; the observed storms south and west of Houston at 0100 UTC are isolated (Fig. 2b) and the observed MCS is further east. By 0300 UTC, most of the observed storms in Central Texas weaken (Fig. 2c) and leave behind a large swath of stratiform precipitation and some weak storms. CTRL also predicts the storms in Central Texas to weaken around 0300 UTC, which causes forecast probabilities near San Antonio to decrease (Fig. 4c). The BSS at 0100 UTC is lower than at 0300 UTC (Fig. 5) because more spurious storms are predicted at the earlier time. Spurious storms also cause the 0100 UTC reliability curve (Fig. 5) to become high-biased for moderate and high probability events ($P[Z > 40 \text{ dBZ}] > 0.4$). By 0300 UTC, when the spurious storms begin to weaken, the reliability curve (Fig. 5) becomes less biased and more closely follows the one-to-one line.

CTRL predicts storms to move too quickly, and by 0600 UTC the predicted MCS (Fig. 4d) is east of the observed storms. Storm motion biases, such as this, are commonly observed in CAM forecasts and are often a consequence of model errors (e.g., Yussouf et al. 2016). Displacement errors cause the BSS at 0600 UTC to become negative and the reliability curve to become high-biased (Fig. 5), suggesting CTRL exhibits no objective skill at this time. It is noted that the $P(Z > 40 \text{ dBZ})$ exceeds 0.7 on the southern edge of the observed MCS (Fig. 4d). Increasing the neighborhood radius when calculating the NMEP could improve forecast performance but would make forecasts less precise, thus this is not performed. Despite displacement errors, CTRL demonstrates some qualitative skill and predicts the MCS to become less organized between 0500 – 0600 UTC, which is approximately the same time as observations (Fig. 2d).

4. Radar DA parameter results

Many studies show radar observations provide important storm scale information, and when assimilated by an optimally configured EnKF, improve analyzed storm structure (e.g., Snyder and Zhang 2003; Dowell et al. 2004; Tong and Xue 2005) and subsequent forecasts (e.g., Snook et al. 2012, 2015). Modifying parameters directly related to radar DA, including data thinning, covariance localization radius and inflation, observation errors, and assimilation frequency initially impacts storm structure and the near-storm environment. With the exception of the DA frequency experiments (Fig. 6g), which impact Z forecast skill several hours after DA, the

differences between experiments are most prominent during the first forecast hour because large- and small-scale errors grow quickly (e.g., Melhauser and Zhang 2012) and degrade the initial benefits of radar DA. At 0015 UTC reliability curves (and BSSs) are already similar between all experiments (Fig. 6). Despite considerable overlap between the reliability curves, experiments that assimilate more radar observations (i.e., smaller data thinning intervals; Fig. 6a-b) or employ a larger horizontal covariance localization radius (Fig. 6c) are more skilled. Impacts of radar DA are localized and not readily obvious in objective verification metrics that consider the full model domain and multiple storm systems. Subjective forecast evaluations performed in the remainder of this section show these DA parameters impact Z forecasts for individual storm systems (i.e., isolated convection and the MCS).

a.) Data thinning

Thinning radar observations potentially removes important fine-scale details of the storms. For small storms, such as those near San Antonio, thinning radar observations removes many if not all available in-storm observations. ThinH3V1 (Fig. 7b) predicts the $P(Z > 40 \text{ dBZ})$ to be larger than ThinH9V1 (Fig. 7c), particularly for small storms located near Mexico, because the experiment assimilates more in-storm observations. Assimilating clear air observations also improves forecast skill by suppressing spurious convection. ThinH6V2, which assimilates few clear air observations (every 4 km vertically), predicts more spurious convection, causing the $P(Z > 40 \text{ dBZ})$ to exceed 0.7 outside of an observed Z core near San Antonio (Fig. 7e). These spurious storms contribute to a modest overprediction bias observed in the ThinH6V2 reliability curve for moderate to high probability thresholds ($P[Z > 40 \text{ dBZ}] > 0.6$) (Fig. 6b).

BSSs are calculated over forecast subdomains that encompass either the isolated convection in Texas (black square, Fig. 7) or the MCS (red square, Fig. 7) to determine how the Z forecast skill varies between the storm regions (Table 3). Fifteen-minute forecasts predict the MCS with more skill ($BSS \geq 0.51$) than the isolated convection ($BSS \geq 0.35$). Enhanced Z forecast skill is in part because the MCS is mature throughout the radar DA period (Fig. 8), while some isolated storms, particularly those northeast of San Antonio, rapidly grow and merge during the final 15 minutes of the DA window (Fig. 8d-e). Studies often assimilate observations of mature storms over many DA cycles to reduce initial conditions errors and improve forecast skill (e.g., Snyder and Zhang 2003; Dawson et al. 2012; Stratman et al. 2020). Thus, isolated convection BSSs are lower in part because fewer DA cycles capture the evolution of storms. The MCS forecasts are also less sensitive to the data thinning experiments (Fig. 7), MCS BSSs decrease less (Table 3) when experiments assimilate fewer radar observations (i.e., ThinH9V1, ThinH6V2).

b.) Covariance localization

Increasing the horizontal covariance localization radius for radar observations causes ensembles to become more confident in high Z values located near observed storms for short-term forecasts. CovH6V0.7 (Fig. 9b) predicts the $P(Z > 40 \text{ dBZ})$ to be smaller than either CovH12V0.7 (Fig. 9a) or CovH18V0.7 (Fig. 9c) for the isolated storms located near the Mexico-Texas border. CovH18V0.7 also predicts fewer spurious storms to be located between San Antonio and Houston (Fig. 9c), which decreases the $P(Z > 40 \text{ dBZ})$ outside of observed storm cores. Because CovH18V0.7 and CovH12V0.7 predict fewer spurious storms in Texas and have improved resolution, the ensembles score a larger BSS than COVH6V0.7 at 0015 UTC (Fig. 6c). Increasing vertical localization radii modestly increases the BSS (Fig. 6d) but probabilistic forecasts appear to be insensitive to this EnKF parameter (Fig. 9d-e). Although Z forecast skill is relatively

insensitive to the vertical covariance localization radius in this study, Sobash and Stensrud (2013) suggest ensembles that employ a smaller vertical localization radius produce analyses with smaller root-mean square errors. Result differences are due to many factors such as changes in the vertical distribution of radar observations, how the vertical covariance localization radii are applied, and that the referred study is an OSSE that assumes perfect-model conditions.

CovH6V0.7 predicts more spurious convection to develop because the ensemble is initialized at 0000 UTC (Fig. 10a) with more spurious radar echoes than CovH12V0.7 (Fig. 10b) or CovH18V0.7 (Fig. 10c). The smaller localization radius employed by COVH6V0.7 reduces the number of surrounding grid points updated by an observation and causes the EnKF to remove less precipitation from clear air regions. Regions of spurious precipitation intensify to form new convection, thus degrading COVH6V0.7 skill during the first forecast hour relative to the other covariance localization experiments (Fig. 6c).

Modifying horizontal covariance localization radii also alters the near-surface environment, which can have a detrimental impact on the predicted thunderstorm evolution. The CovH6V0.7 ensemble mean analysis (Fig. 11a) predicts storms to produce warmer cold pools than CovH12V0.7 (Fig. 11b) or CovH18V0.7 (Fig. 11c). Since cold pools alter the evolution of convection (e.g., Droegemeier and Wilhelmson 1985; Rotunno et al. 1988), changes in the near surface environment impact the forecasts. Thirty-minute forecasts initialized from the ensemble mean analysis (Fig. 11d-f) show CovH12V0.7 (Fig. 11e) and CovH18V0.7 (Fig. 11f) predict storm updrafts to form along the cold pool boundaries. CovH6V0.7 predicts storm updrafts to be smaller in area and more fragmented (Fig. 11d) in part because of the initially weak cold pool (Fig. 11a). Although the impacts on Z forecast skill are relatively small for this case study, altering the near-storm environment, such as cold pool intensity, can potentially degrade forecast skill for other cases.

Improved initial condition estimates cause CovH12v0.7 and CovH18v0.7 to predict the $P(Z > 40 \text{ dBZ})$ with more skill than CovH6v0.7 during the first forecast hour (Fig. 6c). These horizontal covariance localization radii are similar to the optimal radius determined by Sobash and Stensrud (2013) (18 km) but larger than what was used in Tong and Xue (2005) (6 km). The optimal localization radius employed by Tong and Xue (2005) is smaller in part because the experiment assimilates radar observations at every model grid point. Localization radii are sensitive to observation density (e.g., Gao and Xue 2008), and thus can change with the number of assimilated observations.

c.) Covariance inflation

Although BSSs do not substantially change between the covariance inflation experiments conducted (Fig. 6e), increasing the RTPS inflation factor causes the reliability curves to have an increased high bias. When the RTPS inflation factor increases the areal coverage of low and moderate forecast probabilities ($P[Z > 40 \text{ dBZ}] \leq 0.5$) increases modestly within the vicinity of observed convection (Fig. 12). Experiments run with the three smallest inflation factors (i.e., Inf80, Inf90, Inf99) do not substantially alter Z forecast performance (Fig. 6e); however, the Inf110 reliability curve has the largest overprediction bias and lowest BSS. This is because Inf110 predicts the $P(Z > 40 \text{ dBZ})$ to increase most ahead of observed convection near San Antonio (Fig. 12d). Results suggest no forecast skill is gained when inflating the posterior ensemble spread beyond that of the prior ensemble for this case study. Although Z forecast skill is relatively

insensitive to the RTPS inflation factor, it remains important to perform ensemble covariance inflation to prevent collapse of ensemble spread and potential filter divergence when assimilating a dense network of observations.

d.) Observation errors

With some minor differences, 3ms6dBZ and 6ms9dBZ reliability curves are quite similar at 0015 UTC (Fig. 6f) and consequently both ensembles have similar BSSs. Because there are only minor differences between probabilistic forecasts (not pictured), both sets of experiments are not further discussed.

e.) DA frequency

Z forecast skill is moderately sensitive to how frequently radar observations are assimilated by the GSI EnKF system. 5Min (Fig. 13a) and 10Min (Fig. 13b) predict fewer spurious storms than 15Min (Fig. 13c), which decreases forecast probabilities outside of the observed storm cores. Assimilating radar observations more frequently suppresses the coverage of spurious convection (Stratman et al. 2020); however, the benefit of more frequent DA cycling is short-lived. 5Min forecasts at 0015 UTC predict the $P(Z > 40 \text{ dBZ})$ to increase outside of the observed storm cores (Fig. 13d) because spurious convection becomes more widespread in coverage.

Experiments that assimilate observations less frequently (e.g., 10Min, 15Min) exhibit the most skill throughout the remainder of the forecast period. Z forecasts differ the most for the Texas-based convection, but are relatively similar for the MCS. All DA frequency experiments predict forecast probabilities in eastern Texas to be high ($P[Z > 40 \text{ dBZ}] > 0.7$) at 0300 UTC (Fig. 13g-i); however, 5Min (Fig. 13g) predicts the largest probabilities to be displaced from the observed storm cores. 5Min also predicts low and moderate forecast probabilities ($P[Z > 40 \text{ dBZ}] < 0.5$) to be more widespread in coverage throughout the domain because the ensemble predicts more storms to occur. Errors in storm placement and the enhanced coverage of spurious storms modestly reduces the 5Min BSS throughout much of the forecast period.

Experiments that assimilate observations more frequently (i.e., 5Min, 10Min) are the most skilled at 0005 UTC ($\text{BSS} \geq 0.57$) because the forecasts predict fewer spurious storms. This causes subsequent reliability curves to have a smaller high bias (Fig. 14a). Reliability curves for all DA frequency experiments become more similar by 0015 UTC (Fig. 6g) because spurious convection is predicted to grow in coverage. 10Min is most skilled at 0015 UTC ($\text{BSS} = 0.56$) because the reliability curve is closest to the one-to-one line for most moderate probability thresholds ($0.3 < P[Z > 40 \text{ dBZ}] < 0.7$). After 0100 UTC, the 5Min BSS is modestly lower than the other DA frequency experiments. For example, the 5Min 0300 UTC reliability curve (Fig. 14b) exhibits an increased high bias and a decreased BSS because the predicted storms are displaced from regions of high observed Z , and because the ensemble predicts low and moderate forecast probabilities to be more widespread in coverage (Fig. 13g). While 10Min and 15Min predict the $P(Z > 40 \text{ dBZ})$ with more skill at 0300 UTC ($\text{BSS} \geq 0.28$), both reliability curves (Fig. 14b) have a similar, albeit a modestly smaller, high bias.

Despite assimilating more radar observations, 5Min predicts the $P(Z > 40 \text{ dBZ})$ with similar or less skill than experiments that assimilate observations less frequently. To determine why Z forecast skill does not improve with increased DA frequency, the balance of model states for each DA frequency experiment is evaluated using the absolute mean surface pressure tendency (N) (Lynch and Huang 1992; Hu and Xue 2007; Pan and Wang 2019). N , which is averaged over

the domain and ensemble members, is defined by:

$$N = \left(\frac{1}{n_x \times n_u \times n_{mem}} \right) \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{k=1}^{n_{mem}} \left| \frac{\partial p_s}{\partial t} \right|_{i,j} \quad (1)$$

where n_x and n_y are the dimensions of the domain, n_{mem} is the number of ensemble members, and $\left| \frac{\partial p_s}{\partial t} \right|$ is the absolute change in air pressure at the lowest grid point above the surface. N values are calculated every 5 minutes during the 6-hour forecast period. For all experiments, N values are initially large ($\sim 40 - 50 \text{ pa min}^{-1}$) but decrease with time to approximately 5 pa min^{-1} as the model adjusts to any imbalances that originate from DA (Fig. 15). To a large extent this is done through acoustic waves and hydrostatic adjustment. Many of the storms also weaken during the forecast period, further minimizing changes in pressure. 5Min produces the largest N values throughout the forecast period (Fig. 15), while 10Min and 15Min N values are smaller and more similar in magnitude. Large pressure tendencies suggest that the more frequent DA cycling causes 5Min forecasts to become less balanced. Since 10Min and 15Min N values are more similar in magnitude, small increases in the DA frequency (i.e., 10Min) do not appear to cause the ensemble to become as unbalanced and can improve forecast skill.

Results of the DA frequency experiments concur with previous studies (e.g., Wang et al. 2013; Johnson et al. 2017; Pan and Wang 2019). Johnson and Wang (2017) hypothesize that imbalances introduced during DA do not have enough time to adjust to the model before the next assimilation cycle. It is noted that many other studies (e.g., Aksoy et al. 2009; Jung et al. 2012; Dawson et al. 2012) have successfully assimilated radar observations much more frequently (≤ 5 minute intervals). Stratman et al. (2020), who assimilates phased array radar observations every minute, suggests frequent DA cycling can quickly spin up thunderstorms and suppress spurious convection. Given that most of the adjustment occurs within the first few forecast minutes so that all experiments have similar levels of surface pressure noise after 5 minutes (Fig. 15), the negative impact of assimilating 5 minute data is likely insignificant given the potential benefit quickly spinning up observed storms within the model and suppressing spurious convection. It is likely 5Min skill can be further improved by more thoroughly tuning the GSI EnKF system. For realtime forecasting systems where computational cost is a significant issue, the fact that assimilating radar observations at 10 to 15 min intervals yields good results is encouraging.

5. Summary and discussion

This study evaluates short-term (0 – 6 hour) convection-allowing model (CAM) forecasts initialized using the GSI EnKF system enhanced with radar data assimilation (DA) capabilities. Forecasts are run for a mesoscale convective system (MCS) event that occurred in the Southern United States on 28 – 29 May 2017 and produced tornadoes, strong wind gusts, and hail. The control configuration of the GSI EnKF system resembles the Center for the Analysis and Prediction of Storms (CAPS) storm scale ensemble forecast run during the 2019 Hazardous Weather Testbed Spring Forecast Experiment as part of the CLUE. In addition to verifying forecasts run using this configuration, sensitivity experiments are run to evaluate the impact of GSI EnKF configurations for assimilating radar observations including data thinning, covariance localization and inflation, observation error specification, and DA frequency. Ten-member multi-physics ensemble forecasts and deterministic forecasts from the final EnKF ensemble and mean analyses, respectively, are run for the assessment of DA impacts on forecasts.

The CTRL DA configuration creates the most skilled forecasts in this study while

remaining computationally efficient for realtime use. The initial ensemble of CTRL is centered on the 1800 UTC NAM analysis with perturbations derived from 3 hour SREF forecasts. The DA system assimilates thinned radar observations (6 km horizontally, 1 km vertically for in-storm regions, and 2 km vertically for clear air regions) every 15 minutes to remain computationally efficient for experiments run over the CONUS domain. Although forecast skill can be further improved when assimilating more observations (i.e., less data thinning, increased DA frequency) the experiments require additional computing resources and become impractical for realtime use over a large domain such as the CONUS. The covariance localization radius for Z and V_r is set to 12 km in the horizontal and 0.7 scalar height in the vertical to update unobserved regions. CTRL skillfully predicts the evolution of the MCS during the first three forecast hours; however, the predicted storm becomes displaced from observations at later times because it moves too quickly. Despite displacement errors that limit objective forecast skill, CTRL predicts the MCS to weaken during the final forecast hour, which is approximately the same time as observations. CTRL predicts isolated convection with less skill. The ensemble predicts small storms located near the domain boundary with less confidence and predicts nearby spurious convection.

Modifying radar DA parameters, including data thinning, covariance localization radii and inflation impacts forecast skill most during the first forecast hour; large- and small-scale errors grow rapidly to degrade the benefits of radar DA that mainly improve storm-scale structures. Results show increasing the horizontal covariance localization radius and assimilating more radar observations (i.e., less data thinning) increases forecast confidence in high reflectivity values near observed storms and also decreases the coverage of spurious convection. Small, isolated convection is most sensitive to changes in these parameters. In contrast, the predicted MCS is less sensitive to the covariance localization radius or data thinning because the storm system is mature throughout the radar DA window and because the MCS is larger in scale so many more in-storm observations are available for assimilation. Changing the assumed radial velocity and reflectivity observation errors (3 - 6 m s^{-1} and 6 - 9 dBZ, respectively) does not substantially impact forecast skill in this study. Covariance inflation is an important aspect of EnKF DA and prevents the collapse of ensemble spread when assimilating a dense network of observations (e.g., radar observations). Forecast skill is relatively insensitive to the range of relaxation-to-prior-spread inflation factors (80% - 110%) employed in this study; however, when the posterior ensemble spread is inflated beyond that of the prior ensemble (i.e., 110% RTPS) forecast skill decreases because the ensemble overpredicts storm coverage.

DA frequency experiments demonstrate that assimilating radar observations more frequently (i.e., every 5 minutes) initially improves forecast skill by suppressing spurious storms; however, after 15 minutes forecasts predict spurious convection to initiate and forecast skill decreases. The enhanced coverage of spurious convection and storm position errors cause the experiment that assimilates radar observations every 5 minutes to have modestly less skill than experiments that assimilate observations less frequently (i.e., 10 minutes, 15 minutes) throughout much of the forecast period. Results of this study concur with Johnson and Wang (2017), which hypothesizes assimilating observations too frequently degrades forecast skill because the DA-induced imbalances are unable to adjust to the model before the next cycle. Although assimilating observations every 10 minutes improves objective forecast skill in this study, Stratman et al. (2020) note that even more frequent DA cycling can reduce the spin up time for storms and suppress spurious convection.

Although this study evaluates several key aspects of the experimental design, many other factors that impact forecast performance should be investigated. For example, future studies should compare forecast differences between mixed physics and single physics ensembles. CTRL is a mixed physics ensemble which can enhance forecast diversity (Snook et al. 2012; Johnson and Wang 2017) but can also introduce more systematic biases into the forecast system than a single physics ensemble (e.g., Romine et al. 2013). Future studies should also evaluate forecast sensitivity to conventional observation (ground-based observations, soundings) DA parameters because these observations modify atmospheric environmental conditions and can substantially impact storm evolution (e.g., Sobash and Stensrud 2015; Snook et al. 2015). It is noted that the conclusions of this work are limited to a single case study, and that results are sensitive to many factors (e.g, storm morphology, environment, ensemble and domain configuration, observation availability, DA system). Additional experiments are required to ensure results are robust for a variety of cases. Even with these limitations, results provide an initial EnKF configuration that can be deployed in future studies, and the insights gained from this work contribute to the development of future ensemble DA systems.

Acknowledgements: This work was supported by NOAA Joint Technology Transfer Initiative (JTII) grant NA18OAR4590385. Computing was performed on the XSEDE Stampede2 supercomputer at the University of Texas Advanced Computing Center (TACC). Ensemble background and initial conditions were generated under the support of NOAA CSTAR grant NA10NWS4680001.

References

- Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicas e comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Weather Rev.*, **137**, 1805–1824, <https://doi.org/10.1175/2008MWR2691.1>.
- , S. Lorusolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airborne doppler radar observations in an OSSE. *Mon. Weather Rev.*, **140**, 1843–1862, <https://doi.org/10.1175/MWR-D-11-00212.1>.
- Anderson, J. L., 2001: An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Weather Rev.*, **129**, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2).
- , and S. L. Anderson, 1999: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Mon. Weather Rev.*, **127**, 2741–2758.
- , and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Ocean. Technol.*, **24**, 1452–1463, <https://doi.org/10.1175/JTECH2049.1>.
- Brewster, K. A., M. Hu, M. Xue, and J. Gao, 2005: Efficient Assimilation of Radar Data at High Resolution for Short-Range Numerical Weather Prediction. *Int. Symp. on Nowcasting Very Short Range Forecasting*, Toulouse, France, WMO/World Weather Research Program, 3.06.

- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.*, **78**, 1–3, <https://doi.org/10.1126/science.27.693.594>.
- Bringi, V. N., and V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar*. Cambridge, 636 pp.
- Brooks, H. E., 2004: Tornado-warning performance in the past and future. *Bull. Am. Meteorol. Soc.*, **85**, 837–843, <https://doi.org/10.1175/BAMS-85-6-837>.
- CAPS, 2017: 2017 CAPS Spring Forecast Experiment Program Plan. Center for The Analysis and Prediction of Storms. http://forecast.caps.ou.edu/SpringProgram2017_Plan-CAPS.pdf.
- Caya, A., J. Sun, and C. Snyder, 2005: A Comparison between the 4DVAR and the Ensemble Kalman Filter Techniques for. *Mon. Weather Rev.*, **133**, 3081–3094.
- Chang, W., K.-S. Chung, L. Fillion, and S.-J. Baek, 2014: Radar data assimilation in the Canadian high-resolution ensemble Kalman filter system: Performance and verification with real summer cases. *Mon. Weather Rev.*, **142**, 2118–2138, <https://doi.org/10.1175/MWR-D-13-00291.1>.
- Chen, F., and J. Dudhia, 2001: Coupling an Advanced Land Surface- Hydrology Model with the Penn State NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. *Mon. Weather Rev.*, **129**, 569–585.
- Chipilski, H. G., X. Wang, and D. B. Parsons, 2020: Impact of Assimilating PECAN Profilers on the Prediction of Bore-Driven Nocturnal Convection: A Multiscale Forecast Evaluation for the 6 July 2015 Case Study. *Mon. Weather Rev.*, **148**, 1147–1175, <https://doi.org/10.1175/mwr-d-19-0171.1>.
- Clark, A. J., and Coauthors, 2018: The community leveraged unified ensemble (CLUE) in the 2016 NOAA/hazardous weather testbed spring forecasting experiment. *Bull. Am. Meteorol. Soc.*, **99**, 1433–1448, <https://doi.org/10.1175/BAMS-D-16-0309.1>.
- , and Coauthors, 2019: *SPRING FORECASTING EXPERIMENT 2019*. 1–39 pp. https://hwt.nssl.noaa.gov/sfe/2019/docs/HWT_SFE2019_operations_plan.pdf.
- Curtis, C. D., and S. M. Torres, 2011: Adaptive range oversampling to achieve faster scanning on the National Weather Radar Testbed phased-array radar. *J. Atmos. Ocean. Technol.*, **28**, 1581–1597, <https://doi.org/10.1175/JTECH-D-10-05042.1>.
- Dawson, D. T. I., L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2012: Impact of the Environmental Low-Level Wind Profile on Ensemble Forecasts of the 4 May 2007 Greensburg, Kansas, Tornadic Storm and Associated Mesocyclones. *Mon. Weather Rev.*, **140**, 696–716, <https://doi.org/10.1175/MWR-D-11-00008.1>.
- Doviak, R., and D. S. Zrnić, 1993: *No Doppler Radar and Weather Observations*. 2nd ed. Academic Press, 562 pp.
- Dowell, D., L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman Filter Assimilation of Radar Observations of the 8 May 2003 Oklahoma City Supercell : Influences of Reflectivity Observations on Storm-Scale Analyses. *Mon. Weather Rev.*, **139**, 272–294, <https://doi.org/10.1175/2010MWR3438.1>.

- Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Ocean. Technol.*, **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. *Mon. Weather Rev.*, **132**, 1982–2005, [https://doi.org/10.1175/1520-0493\(2004\)132<1982:WATRIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2).
- Droegemeier, K. K., and R. B. Wilhelmson, 1985: Three-dimensional numerical modeling of convection produced by interacting thunderstorm outflows. Part I: control simulation and low-level moisture variations. *J. Atmos. Sci.*, **42**, 2381–2403, [https://doi.org/10.1175/1520-0469\(1985\)042<2381:TDNMOC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<2381:TDNMOC>2.0.CO;2).
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 143–162.
- , 2003: The Ensemble Kalman Filter : theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Gao, J., and M. Xue, 2008: An efficient dual-resolution approach for ensemble data assimilation and tests with simulated Doppler radar data. *Mon. Weather Rev.*, **136**, 945–963, <https://doi.org/10.1175/2007MWR2120.1>.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, 723–757.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.*, **129**, 2776–2790, [https://doi.org/10.1175/1520-0493\(2001\)129<2776:DDFOBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2).
- Heinselman, P. L., and S. M. Torres, 2011: High-temporal-resolution capabilities of the National Weather Radar Testbed Phased-Array Radar. *J. Appl. Meteorol. Climatol.*, **50**, 579–593, <https://doi.org/10.1175/2010JAMC2588.1>.
- Hong, S. Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Weather Rev.*, **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. *Mon. Weather Rev.*, **126**, 796–811, [https://doi.org/10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2).
- Hu, M., and M. Xue, 2007: Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. *Mon. Weather Rev.*, **135**, 507–525, <https://doi.org/10.1175/MWR3313.1>.
- 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. *J. Geophys. Res. Atmos.*, **113**, 1–9, <https://doi.org/10.1029/2008JD009944>.

- Janjić, Z. I., 1990: The step-mountain coordinate: physical package. *Mon. Weather Rev.*, **118**, 1429–1443, [https://doi.org/10.1175/1520-0493\(1990\)118<1429:TSMCPP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<1429:TSMCPP>2.0.CO;2).
- , 1996: The surface layer in the NCEP Eta Model. *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 354–355.
- , 2001: *Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso Model*. 61 pp.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Johnson, A., and X. Wang, 2017: Design and Implementation of a GSI-Based Convection-Allowing Ensemble Data Assimilation and Forecast System for the PECAN Field Experiment. Part I: Optimal Configurations for Nocturnal Convection Prediction Using Retrospective Cases. *Weather Forecast.*, **32**, 289–315, <https://doi.org/10.1175/waf-d-16-0102.1>.
- , ———, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Weather Rev.*, **143**, 3087–3108, <https://doi.org/10.1175/MWR-D-14-00345.1>.
- , ———, and S. Degelia, 2017: Design and implementation of a GSI-based convection-allowing ensemble-based data assimilation and forecast system for the PECAN field experiment. Part II: Overview and evaluation of a real-time system. *Weather Forecast.*, **32**, 1227–1251, <https://doi.org/10.1175/WAF-D-16-0201.1>.
- Jones, T. A., K. H. Knopfmeier, D. M. Wheatley, G. J. 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 Forecast.*, **31**, 297–327, <https://doi.org/10.1175/WAF-D-15-0107.1>.
- , P. S. Skinner, K. H. Knopfmeier, E. R. Mansell, P. Minnis, R. Palikonda, and W. J. Smith, 2018a: Comparison of Cloud Microphysics Schemes in a Warn-on-Forecast System Using Synthetic Satellite Objects. *Weather*, **33**, 1681–1708, <https://doi.org/10.1175/WAF-D-18-0112.1>.
- , X. Wang, P. Skinner, A. Johnson, and Y. Wang, 2018b: Assimilation of GOES-13 imager clear-sky water vapor (6.5 μm) radiances into a warn-on-forecast system. *Mon. Weather Rev.*, **146**, 1077–1107, <https://doi.org/10.1175/MWR-D-17-0280.1>.
- Jung, Y., M. Xue, G. Zhang, and J. M. Straka, 2008: Assimilation of Simulated Polarimetric Radar Data for a Convective Storm Using the Ensemble Kalman Filter. Part I: Observation Operators for Reflectivity and Polarimetric Variables. *Mon. Weather Rev.*, **136**, 2228–2245, <https://doi.org/10.1175/2007MWR2288.1>.
- , ———, and M. Tong, 2012: Ensemble Kalman Filter Analyses of the 29–30 May 2004 Oklahoma Tornadoic Thunderstorm Using One- and Two-Moment Bulk Microphysics Schemes, with Verification against Polarimetric Radar Data. *Mon. Weather Rev.*, **140**, 1457–1475, <https://doi.org/10.1175/MWR-D-11-00032.1>.
- , and Coauthors, 2018a: Development of GSI-Based EnKF and Hybrid EnVar Data

- Assimilation Capabilities for Continental-Scale 3-km Convection-Permitting Ensemble Forecasting and Testing via NOAA Hazardous Weather Testbed Spring Forecasting Experiments. *29th Conference on Weather Analysis and Forecasting/25th Conference on Numerical Weather Prediction*, Denver, CO.
- , R. Kong, C. Liu, J. Luo, M. Xue, and G. Zhao, 2018b: CAPS Real-time Storm-Scale EnKF Data Assimilation and Forecasts for the NOAA Hazardous Weather Testbed Spring Forecasting Experiments: Towards the Goal of Operational Ensemble-Variational Data Assimilation for Convection Permitting Models. *The 8th EnKF Data Assimilation Workshop*, Sainte-Adèle, Québec, Canada.
- Kain, J. S., and Coauthors, 2010: Assessing advances in the assimilation of radar data and other mesoscale observations within a collaborative forecasting-research environment. *Weather Forecast.*, **25**, 1510–1521, <https://doi.org/10.1175/2010WAF2222405.1>.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Fluids Eng. Trans. ASME*, **82**, 35–45, <https://doi.org/10.1115/1.3662552>.
- Labriola, J., N. A. Snook, Y. Jung, B. J. Putnam, and M. Xue, 2017: Ensemble Hail Prediction for the Storms of 10 May 2010 in South-Central Oklahoma Using Single- and Double-Moment Microphysical Schemes. *Mon. Weather Rev.*, <https://doi.org/10.1175/MWR-D-17-0039.1>.
- , N. Snook, Y. Jung, and M. Xue, 2019: Explicit Ensemble Prediction of Hail in 19 May 2013 Oklahoma City Thunderstorms and Analysis of Hail Growth Processes with Several Multi-Moment Microphysics Schemes. *Mon. Weather Rev.*, **147**, 1193–1213, <https://doi.org/10.1175/MWR-D-18-0266.1>.
- Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: evidence of progress. *Weather Forecast.*, **33**, 599–607, <https://doi.org/10.1175/waf-d-17-0145.1>.
- Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Weather Rev.*, **120**, 1019–1034, [https://doi.org/10.1175/1520-0493\(1992\)120<1019:IOTHMU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2).
- Mansell, E. R., C. Ziegler, and E. Bruning, 2010: Simulated Electrification of a Small Thunderstorm with Two-Moment Bulk Microphysics. *J. Atmos. Sci.*, **67**, 171–194, <https://doi.org/10.1175/2009JAS2965.1>.
- Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, <https://doi.org/10.1175/JAS-D-11-0315.1>.
- Milbrandt, J. A., and M. K. Yau, 2005: A Multimoment Bulk Microphysics Parameterization. Part II: A Proposed Three-Moment Closure and Scheme Description. *J. Atmos. Sci.*, **62**, 3065–3081, <https://doi.org/10.1175/JAS3535.1>.
- 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. *J. Geophys. Res. D Atmos.*, **102**, 16663–16682, <https://doi.org/10.1029/97jd00237>.

- Morrison, H., and W. W. Grabowski, 2008: A Novel Approach for Representing Ice Microphysics in Models: Description and Tests Using a Kinematic Framework. *J. Atmos. Sci.*, **65**, 1528–1548, <https://doi.org/10.1175/2007JAS2491.1>.
- , and J. A. Milbrandt, 2015: Parameterization of Cloud Microphysics Based on the Prediction of Bulk Ice Particle Properties. Part I: Scheme Description and Idealized Tests. *J. Atmos. Sci.*, **72**, 287–311, <https://doi.org/10.1175/JAS-D-14-0065.1>.
- Murphy, A. H., 1973: A New Vector Partition of the Probability Score. *J. Appl. Meteorol.*, **12**, 595–600, [https://doi.org/doi.org/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2](https://doi.org/doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2).
- Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteorol. Soc. Japan*, **87**, 895–912, <https://doi.org/10.2151/jmsj.87.895>.
- NOAA, 2019: National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters. <https://www.ncdc.noaa.gov/billions/> (Accessed July 11, 2019).
- Pan, Y., and M. Wang, 2019: Impact of the Assimilation Frequency of Radar Data with the ARPS 3DVar and Cloud Analysis System on Forecasts of a Squall Line in Southern China. *Adv. Atmos. Sci.*, **36**, 160–172, <https://doi.org/10.1007/s00376-018-8087-5>.
- Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2019: Ensemble Kalman Filter Assimilation of Polarimetric Radar Observations for the 20 May 2013 Oklahoma Tornadoic Supercell Case. *Mon. Weather Rev.*, **147**, 2511–2533, <https://doi.org/10.1175/mwr-d-18-0251.1>.
- Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Weather Rev.*, **141**, 1263–1284, <https://doi.org/10.1175/MWR-D-12-00112.1>.
- , ———, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Weather Rev.*, **142**, 4519–4541, <https://doi.org/10.1175/MWR-D-14-00100.1>.
- Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A Theory for Strong, Long-Lived Squall Lines. *J. Atmos. Sci.*, **45**, 463–485.
- Ryzhkov, A. V., S. E. Giangrande, V. M. Melnikov, and T. J. Schuur, 2005: Calibration issues of dual-polarization radar measurements. *J. Atmos. Ocean. Technol.*, **22**, 1138–1155, <https://doi.org/10.1175/JTECH1772.1>.
- Schwartz, C. S., and R. A. Sobash, 2017: Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations. *Mon. Weather Rev.*, **145**, 3397–3418, <https://doi.org/10.1175/MWR-D-16-0400.1>.
- , and Coauthors, 2010: Toward Improved Convection-Allowing Ensembles: Model Physics Sensitivities and Optimizing Probabilistic Guidance with Small Ensemble Membership. *Weather Forecast.*, **25**, 263–280, <https://doi.org/10.1175/2009WAF2222267.1>.

- Skamarock, W. C., and Coauthors, 2008: A Description of the Advanced Research WRF Version 3. *NCAR Tech. Note NCAR/TN-4751STR*, 113, <https://doi.org/10.5065/D6DZ069T>.
- Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial Operating Capabilities. *Bull. Am. Meteorol. Soc.*, **97**, 1617–1630, <https://doi.org/10.1175/BAMS-D-14-00173.1>.
- Snook, N., M. Xue, and Y. Jung, 2012: Ensemble Probabilistic Forecasts of a Tornadoic Mesoscale Convective System from Ensemble Kalman Filter Analyses Using WSR-88D and CASA Radar Data. *Mon. Weather Rev.*, **140**, 2126–2146, <https://doi.org/10.1175/MWR-D-11-00117.1>.
- , ———, and ———, 2013: Impacts of assumed observation errors in EnKF analyses and ensemble forecasts of a tornadoic mesoscale convective system. *17th Conf. on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, Austin, TX
<https://ams.confex.com/ams/93Annual/webprogram/Paper219933.html>.
- , ———, and ———, 2015: Multi-Scale EnKF Assimilation of Radar and Conventional Observations and Ensemble Forecasting for a Tornadoic Mesoscale Convective System. *Mon. Weather Rev.*, 1035–1058, <https://doi.org/10.1175/MWR-D-13-00262.1>.
- , Y. Jung, J. Brotzge, B. J. Putnam, and M. Xue, 2016: Prediction and Ensemble Forecast Verification of Hail in the Supercell Storms of 20 May 2013. *Weather Forecast.*, **4**, 811–825, <https://doi.org/10.1175/WAF-D-15-0152.1>.
- , F. Kong, K. A. Brewster, M. Xue, K. W. Thomas, T. A. Supinie, S. Perfater, and B. Albright, 2019: Evaluation of Convection-Permitting Precipitation Forecast Products Using WRF, NMMB, and FV3 for the 2016–17 NOAA Hydrometeorology Testbed Flash Flood and Intense Rainfall Experiments. *Weather Forecast.*, **34**, 781–804, <https://doi.org/10.1175/waf-d-18-0155.1>.
- Snook, N. A., M. Xue, and Y. Jung, 2011: Analysis of a Tornadoic Mesoscale Convective Vortex Based on Ensemble Kalman Filter Assimilation of CASA X-Band and WSR-88D Radar Data. *Mon. Weather Rev.*, **139**, 3446–3468, <https://doi.org/10.1175/MWR-D-10-05053.1>.
- Snyder, C., and F. Zhang, 2003: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter. *Mon. Weather Rev.*, **131**, 1663–1677.
- Sobash, R. A., and D. J. Stensrud, 2013: The impact of covariance localization for radar data on EnKF analyses of a developing MCS: Observing system simulation experiments. *Mon. Weather Rev.*, **141**, 3691–3709, <https://doi.org/10.1175/MWR-D-12-00203.1>.
- , and ———, 2015: Assimilating surface mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 May 2012. *Mon. Weather Rev.*, **143**, 3700–3725, <https://doi.org/10.1175/MWR-D-14-00126.1>.
- , and L. J. Wicker, 2015: On the impact of additive noise in storm-scale EnKF experiments. *Mon. Weather Rev.*, **143**, 3067–3086, <https://doi.org/10.1175/MWR-D-14-00323.1>.
- Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Am. Meteorol. Soc.*, **90**, 1487–1499, <https://doi.org/10.1175/2009BAMS2795.1>.

- , and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, <https://doi.org/10.1016/j.atmosres.2012.04.004>.
- Stratman, D., N. Yussouf, Y. Jung, T. Supinie, M. Xue, P. S. Skinner, and B. J. Putnam, 2020: Optimal Temporal Frequency of NSSL Phased-Array Radar Observations for an Experimental Warn-on-Forecast System. *Weather Forecast.*, accepted, <https://doi.org/10.1175/WAF-D-19-0165.1>.
- Supinie, T. A., Y. Jung, M. Xue, D. J. Stensrud, M. M. French, and H. B. Bluestein, 2016: Impact of VORTEX2 Observations on Analyses and Forecasts of the 5 June 2009 Goshen County, Wyoming, Supercell. *Mon. Weather Rev.*, **44**, 429–449, <https://doi.org/10.1175/MWR-D-15-0171.1>.
- , N. Yussouf, Y. Jung, M. Xue, J. Cheng, and S. Wang, 2017: Comparison of the Analyses and Forecasts of a Tornadoic Supercell Storm from Assimilating Phased-Array Radar and WSR-88D Observations. *Weather Forecast.*, **32**, 1379–1401, <https://doi.org/10.1175/WAF-D-16-0159.1>.
- Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, <https://doi.org/10.1175/JAS-D-13-0305.1>.
- , P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Mon. Weather Rev.*, **136**, 5095–5115, [https://doi.org/10.1175/1520-0493\(2004\)132<0519:EFOWPU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2).
- Tong, C.-C., Y. Jung, C. Liu, and M. Xue, 2020: Implementation of Radar Data Assimilation Capabilities within Ensemble-Variational Hybrid GSI for the Stand-Alone Regional FV3-Based Convection-Allowing Forecasting System. *30th Conference on Weather Analysis and Forecasting (WAF)/26th Conference on Numerical Weather Prediction (NWP)*, Boston, MA, Amer. Meteor. Soc.
- Tong, M., and M. Xue, 2005: Ensemble Kalman Filter Assimilation of Doppler Radar Data with a Compressible Nonhydrostatic Model: OSS Experiments. *Mon. Weather Rev.*, **133**, 1789–1807, <https://doi.org/10.1175/MWR2898.1>.
- Vivekanandan, J., W. M. Adams, and V. N. Bringi, 1991: Rigorous Approach to Polarimetric Radar Modeling of Hydrometeor Orientation Distributions. *J. Appl. Meteorol.*, **30**, 1053–1063, [https://doi.org/10.1175/1520-0450\(1991\)030<1053:RATPRM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1991)030<1053:RATPRM>2.0.CO;2).
- Wang, S., M. Xue, and J. Min, 2013: A four-dimensional asynchronous ensemble square-root filter (4DEnSRF) algorithm and tests with simulated radar data. *Q. J. R. Meteorol. Soc.*, **139**, 805–819, <https://doi.org/10.1002/qj.1987>.
- 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 tornadoic supercell. *Mon. Weather Rev.*, **145**, 1447–1471, <https://doi.org/10.1175/MWR-D-16-0231.1>.
- Weber, M. E., J. Y. N. Cho, J. S. Herd, J. M. Flavin, W. E. Benner, and G. S. Torok, 2007: The

- next-generation multimission U.S. surveillance radar network. *Bull. Am. Meteorol. Soc.*, **88**, 1739–1751, <https://doi.org/10.1175/BAMS-88-11-1739>.
- Wheatley, D. M., 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 Forecast.*, **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. *Mon. Weather Rev.*, **130**, 1913–1924, [https://doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- , and ———, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Mon. Weather Rev.*, **140**, 3078–3089, <https://doi.org/10.1175/MWR-D-11-00276.1>.
- Xue, M., M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square root Kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting. *J. Atmos. Ocean. Technol.*, **23**, 46–66, <https://doi.org/10.1175/JTECH1835.1>.
- , Y. Jung, and G. Zhang, 2010: State estimation of convective storms with a two-moment microphysics scheme and an ensemble Kalman filter: Experiments with simulated radar data. *Q. J. R. Meteorol. Soc.*, **136**, 685–700, <https://doi.org/10.1002/qj.593>.
- Ying, Y., F. Zhang, and J. L. Anderson, 2018: On the selection of localization radius in ensemble filtering for multiscale quasigeostrophic dynamics. *Mon. Weather Rev.*, **146**, 543–560, <https://doi.org/10.1175/MWR-D-17-0336.1>.
- Yussouf, N., and D. J. Stensrud, 2010: Impact of phased-array radar observations over a short assimilation period: Observing system simulation experiments using an ensemble Kalman filter. *Mon. Weather Rev.*, **138**, 517–538, <https://doi.org/10.1175/2009MWR2925.1>.
- , E. R. Mansell, L. J. Wicker, D. Wheatley, and D. Stensrud, 2013: The Ensemble Kalman Filter Analyses and Forecasts of the 8 May 2003 Oklahoma City Tornadoic Supercell Storm Using Single- and Double-Moment Microphysics Schemes. *Mon. Weather Rev.*, **141**, 3388–3412, <https://doi.org/10.1175/MWR-D-12-00237.1>.
- , J. S. Kain, and A. J. Clark, 2016: Short-Term Probabilistic Forecasts of the 31 May 2013 Oklahoma Tornado and Flash Flood Event Using a Continuous-Update-Cycle Storm-Scale Ensemble System. *Weather Forecast.*, **31**, 957–983, <https://doi.org/10.1175/waf-d-15-0160.1>.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Mon. Weather Rev.*, **132**, 1238–1253.
- Zhang, Y., D. J. Stensrud, and F. Zhang, 2019: Simultaneous Assimilation of Radar and All-Sky Satellite Infrared Radiance Observations for Convection-Allowing Ensemble Analysis and Prediction of Severe Thunderstorms. *Mon. Weather Rev.*, **147**, 4389–4409, <https://doi.org/10.1175/mwr-d-19-0163.1>.

Zrnic, D. S., and Coauthors, 2007: Agile-beam phased array radar for weather observations. *Bull. Am. Meteorol. Soc.*, **88**, 1753–1766, <https://doi.org/10.1175/BAMS-88-11-1753>.

Table 1: CTRL experiment configuration

Initial conditions	1800 UTC NAM analysis with 1500 UTC SREF perturbations
Assimilation frequency	60 Minutes 1800 – 2300 UTC (conventional only) 15 Minutes 2300 – 0000 UTC (radar and conventional when available)
Covariance localization radii - radar observations (horizontal / vertical)	12 km / 0.7 scale height
Covariance localization radii – conventional observations (horizontal / vertical)	300 km / 0.7 scale height
Observation errors (reflectivity / radial velocity)	6 dBZ / 3 m s ⁻¹
Ensemble spread inflation	99% RTPS
Radar data thinning interval (horizontal / vertical in-storm / vertical clear air)	6 km / 1 km / 2 km

Table 2: Scores used to evaluate forecasts in the study. The number of forecasts is N , the number of forecasts for a given k probability threshold is n_k , \bar{o} is the observed climatology, and \bar{o}_k is the observed relative frequency, and p_k is the forecast probability.

Metric	Formula
Reliability	$\frac{1}{N} \sum_{k=1}^K n_k (p_k - \bar{o}_k)^2$
Resolution	$\frac{1}{N} \sum_{k=1}^K n_k (\bar{o}_k - \bar{o})^2$
Uncertainty	$\bar{o}(1 - \bar{o})$
Brier skill score	$\frac{\text{Resolution} - \text{Reliability}}{\text{Uncertainty}}$

Table 3: Data thinning experiment BSSs at 0015 UTC verifying the $P(Z > 40 \text{ dBZ})$. Skill scores are computed over subdomains (Fig. 7) that encompass either the isolated convection (black square) or the MCS (red square).

	Isolated Convection	MCS
ThinH6V1	0.45	0.57
ThinH3V1	0.45	0.55
ThinH9V1	0.40	0.53
ThinH6V0.5	0.45	0.56
ThinH6V2	0.35	0.51

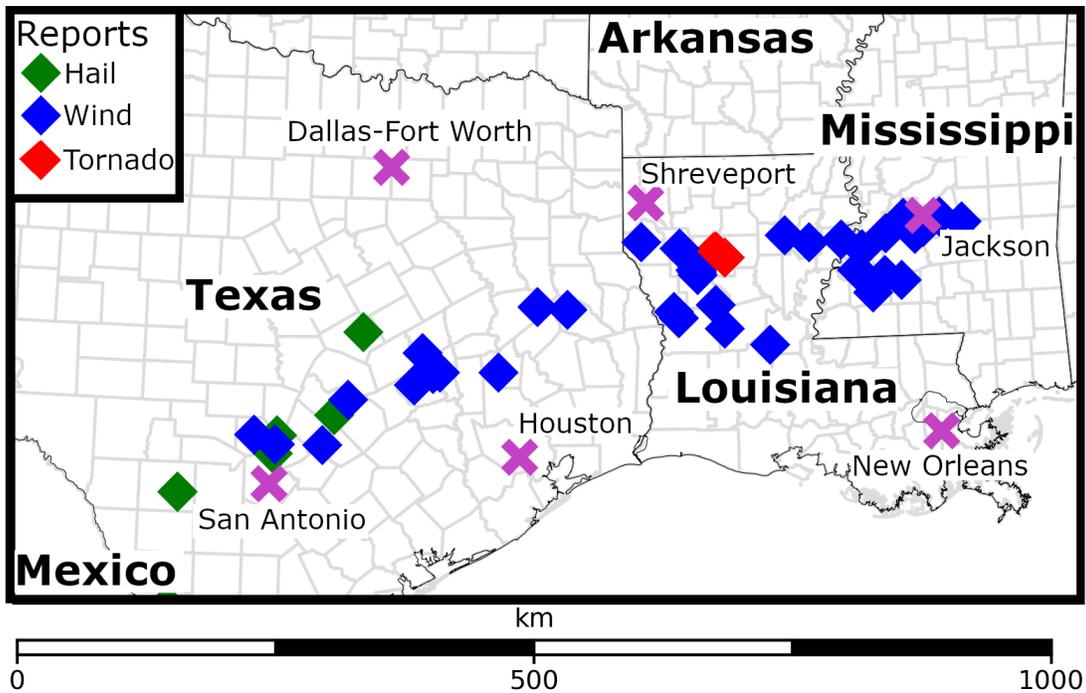


Fig. 1. A diagram of the 28-29 May 2017 forecast domain, states and countries are labeled in bold. A legend for hail, wind, and tornado SPC storm reports is provided in the upper left-hand corner. Cities referred to during the study are marked with a fuchsia “X”.

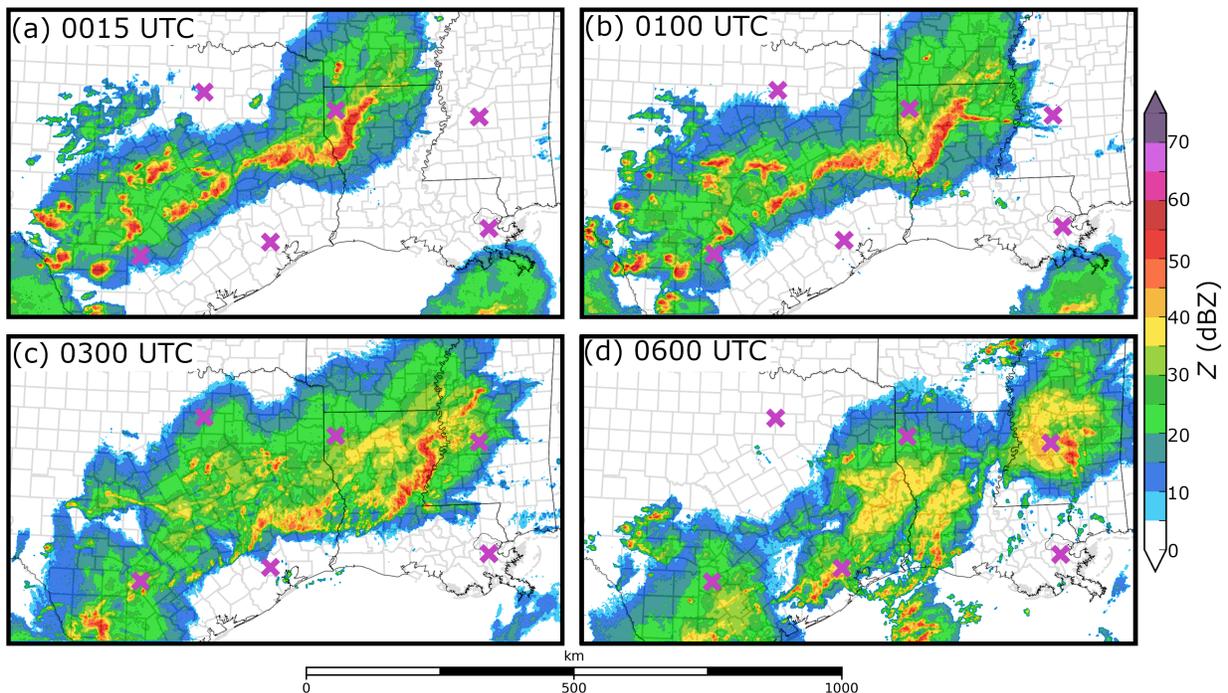


Fig. 2. MRMS observed column maximum Z valid at (a) 0015 UTC, (b) 0100 UTC, (c) 0300 UTC, and (d) 0600 UTC. Cities referred to during the study are marked with a fuchsia “X” and are the same as Fig. 1.

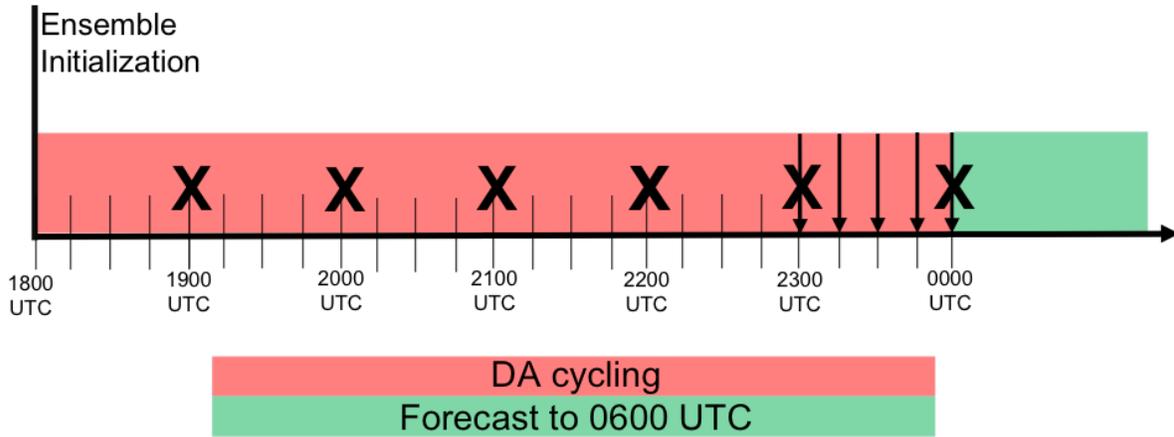


Fig. 3. Flow diagram detailing the CTRL DA experiment configuration. A bold vertical line at 1800 UTC marks when the ensemble of forecasts is first initialized. The “X” symbols mark when surface observations are assimilated and the downward pointing arrows mark when radar observations are assimilated. The final free forecasts start at 0000 UTC, after 6 hours of DA cycles.

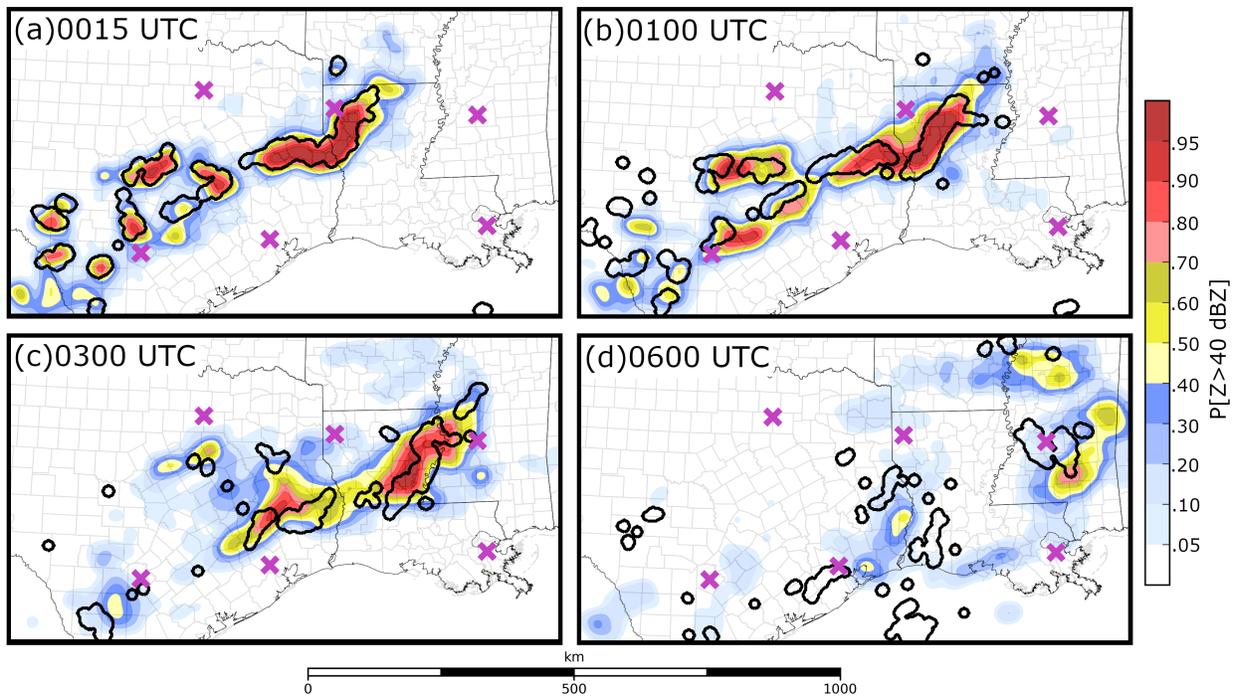


Fig. 4. The $P(Z > 40 \text{ dBZ})$ predicted by CTRL valid at the labeled times. Thick black contours represent locations where observed $Z > 40 \text{ dBZ}$. Background maps are the same as Fig. 2.

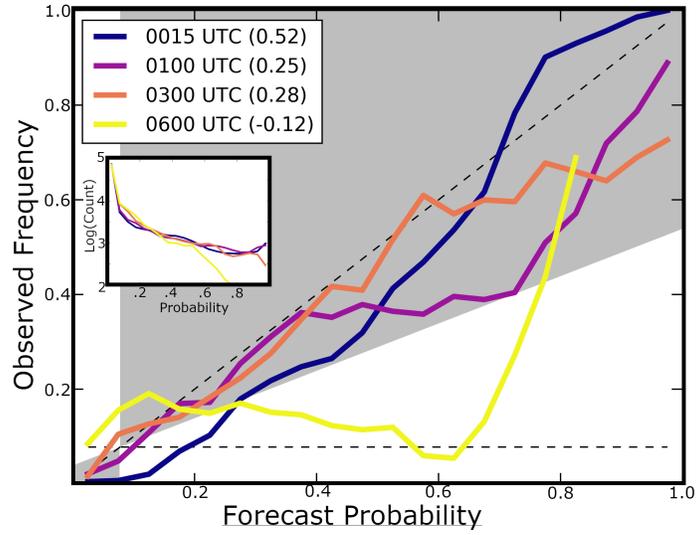


Fig. 5. Reliability diagrams for the CTRL probabilistic forecasts shown in Fig. 4. Line colors correspond to when the forecasts are valid, the BSS for each forecast time is included in the legend.

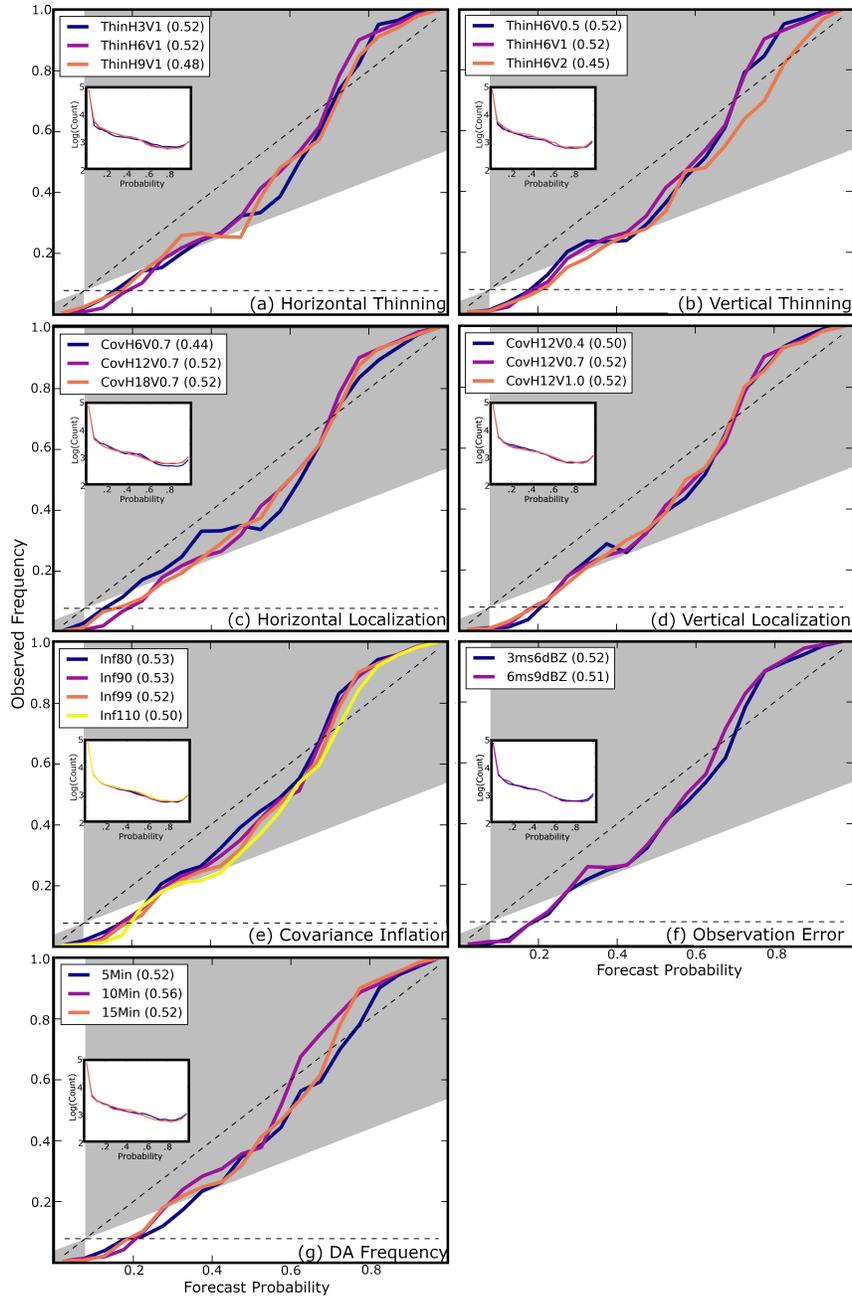


Fig. 6. Reliability diagrams for the forecast $P(Z > 40 \text{ dBZ})$ valid at 0015 UTC for the labeled experiments including (a-b) radar data thinning, (c-d) covariance localization radius, (e) covariance inflation, (f) observation errors, and (g) DA frequency. Experiments correspond to different colors, the BSS for each experiment is included in the legend.

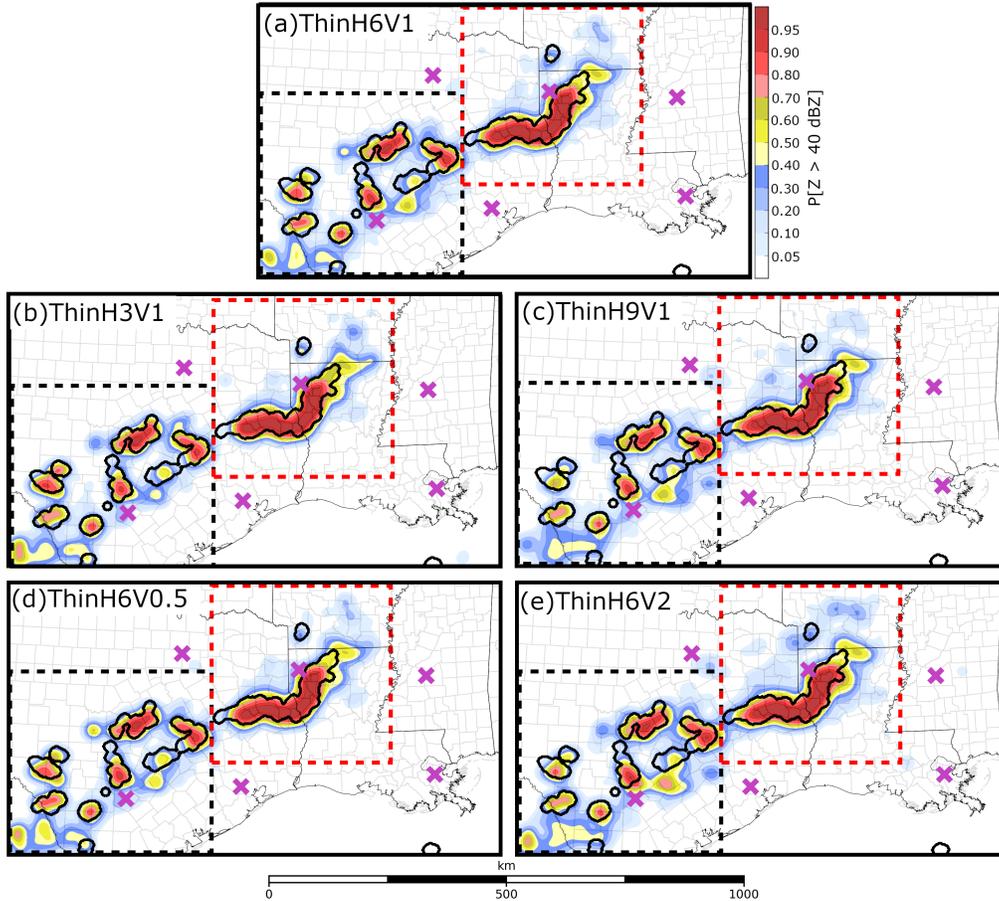


Fig. 7. The $P(Z > 40 \text{ dBZ})$ predicted by (a) CTRL and the (b-c) horizontal and (d-e) vertical radar data thinning experiments valid at 0015 UTC. Thick black contours represent locations where observed $Z > 40 \text{ dBZ}$. Dashed black and red squares mark the verification subdomains discussed in Table 3. Background maps are the same as Fig. 2.

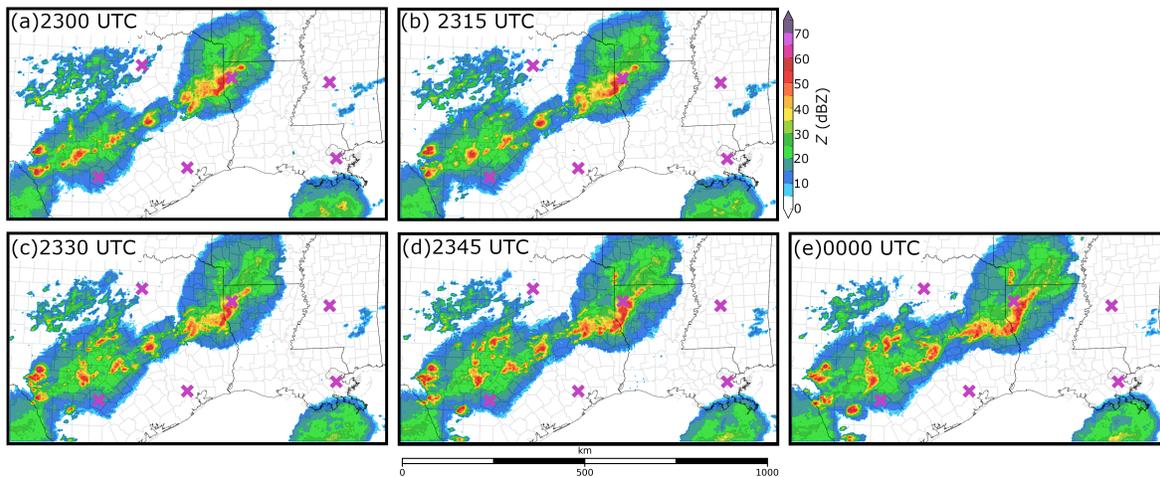


Fig. 8. MRMS observed column maximum Z between 2300 – 0000 UTC, when radar observations are assimilated

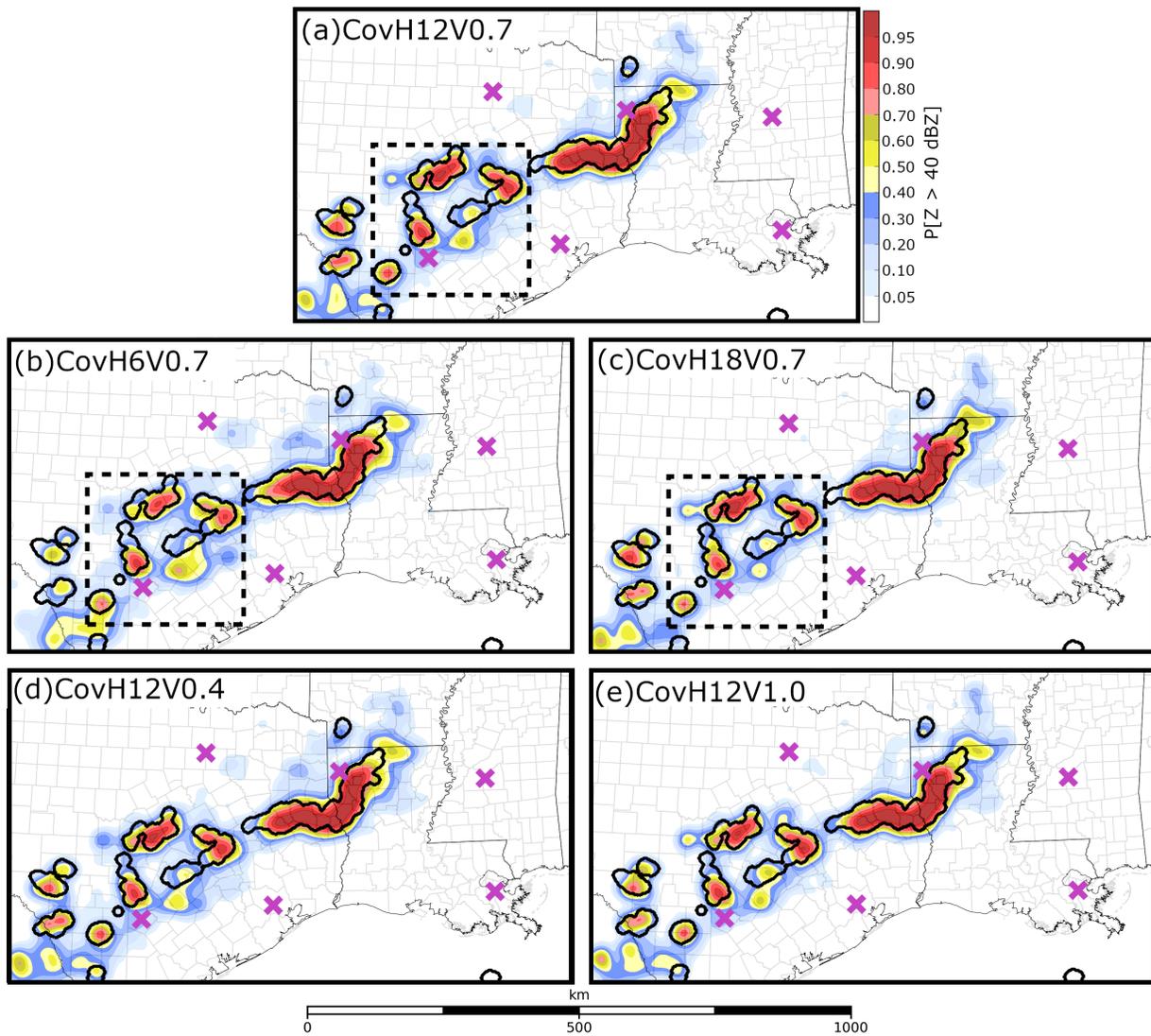


Fig. 9. The $P(Z > 40 \text{ dBZ})$ predicted by (a) CTRL and the (b-c) horizontal and (d-e) vertical covariance localization experiments valid at 0015 UTC. Thick black contours represent locations where observed $Z > 40 \text{ dBZ}$. Dashed squares in (a-c) mark the boundaries of the Fig. 11 domain. Background maps are the same as Fig. 2.

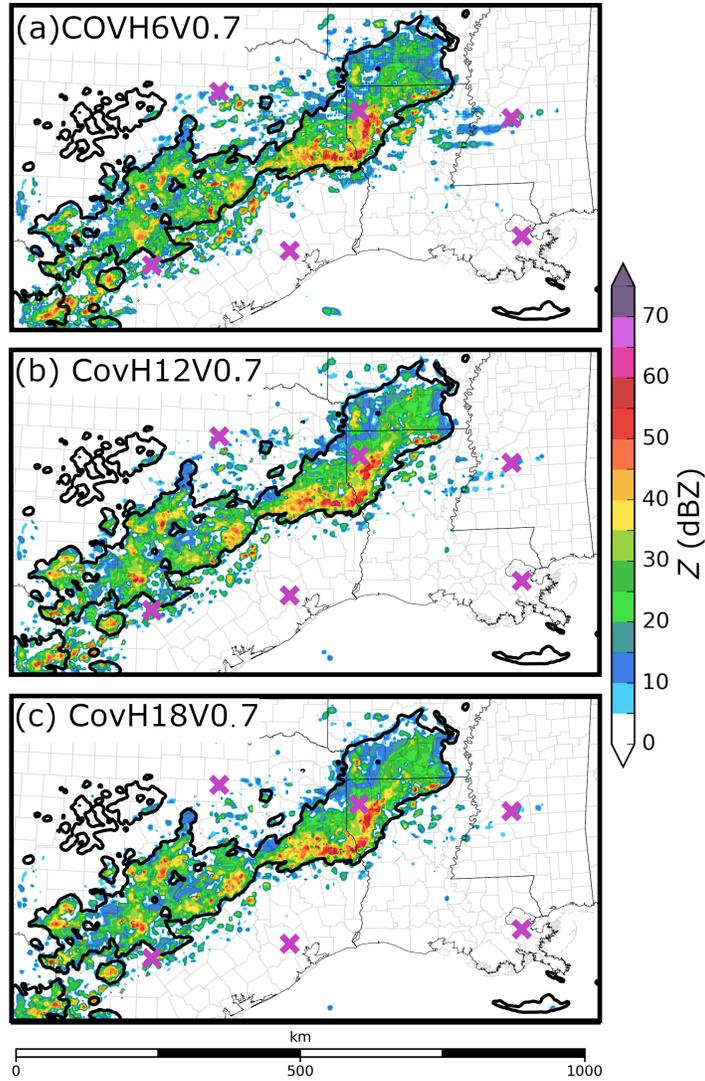


Fig. 10. 0000 UTC ensemble mean analysis simulated Z at 3 km above ground level (AGL). Thick black contours mark where the observed Z at 3 km AGL exceeds 5 dBZ, and the background maps are the same as Fig. 2.

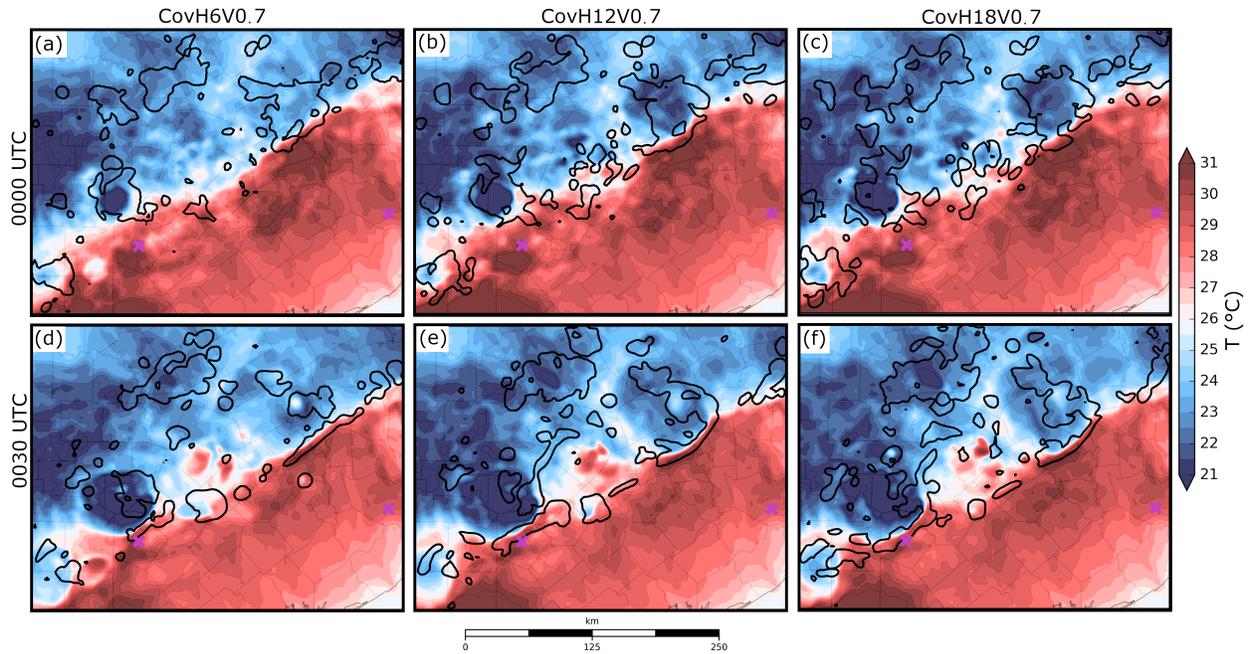


Fig. 11. Air temperature (T) at the lowest model level above the surface for the horizontal covariance localization radius experiments. Thick black contours mark where the predicted updrafts exceed 1 m s^{-1} , the fuschia “x” marks the location of San Antonio.

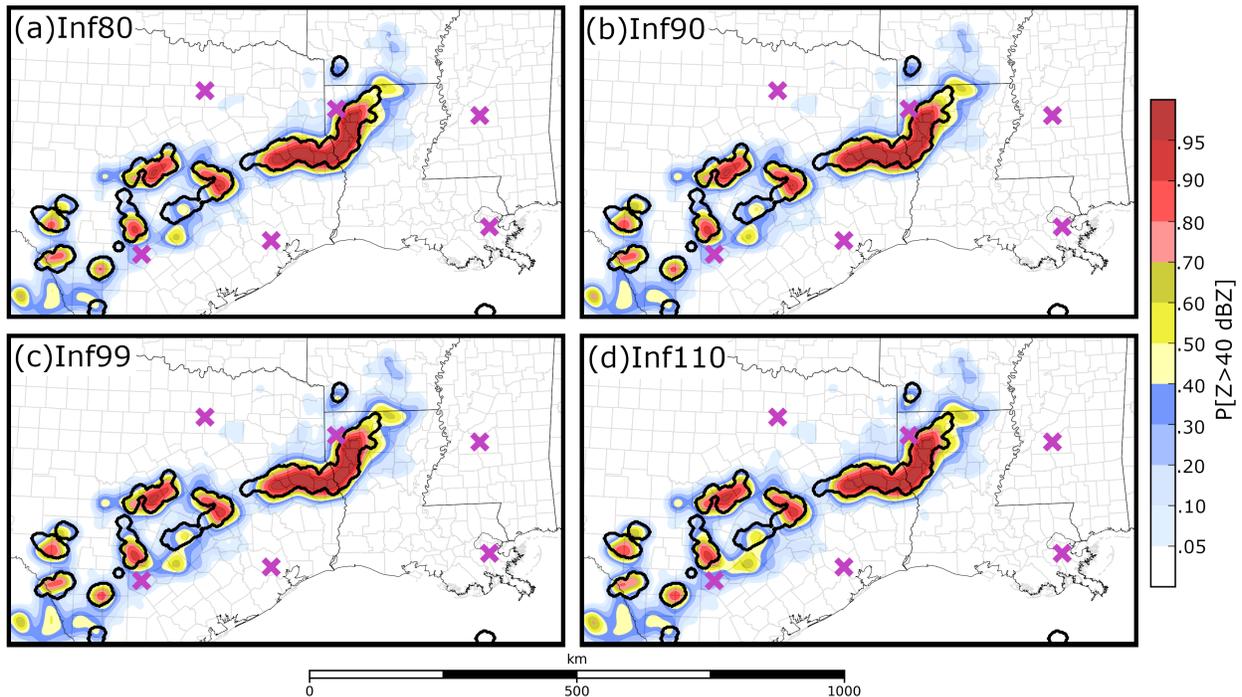


Fig. 12. The $P(Z > 40 \text{ dBZ})$ predicted by the covariance inflation experiments valid at 0015 UTC. Thick black contours represent locations where observed $Z > 40 \text{ dBZ}$. Background maps are the same as Fig. 2.

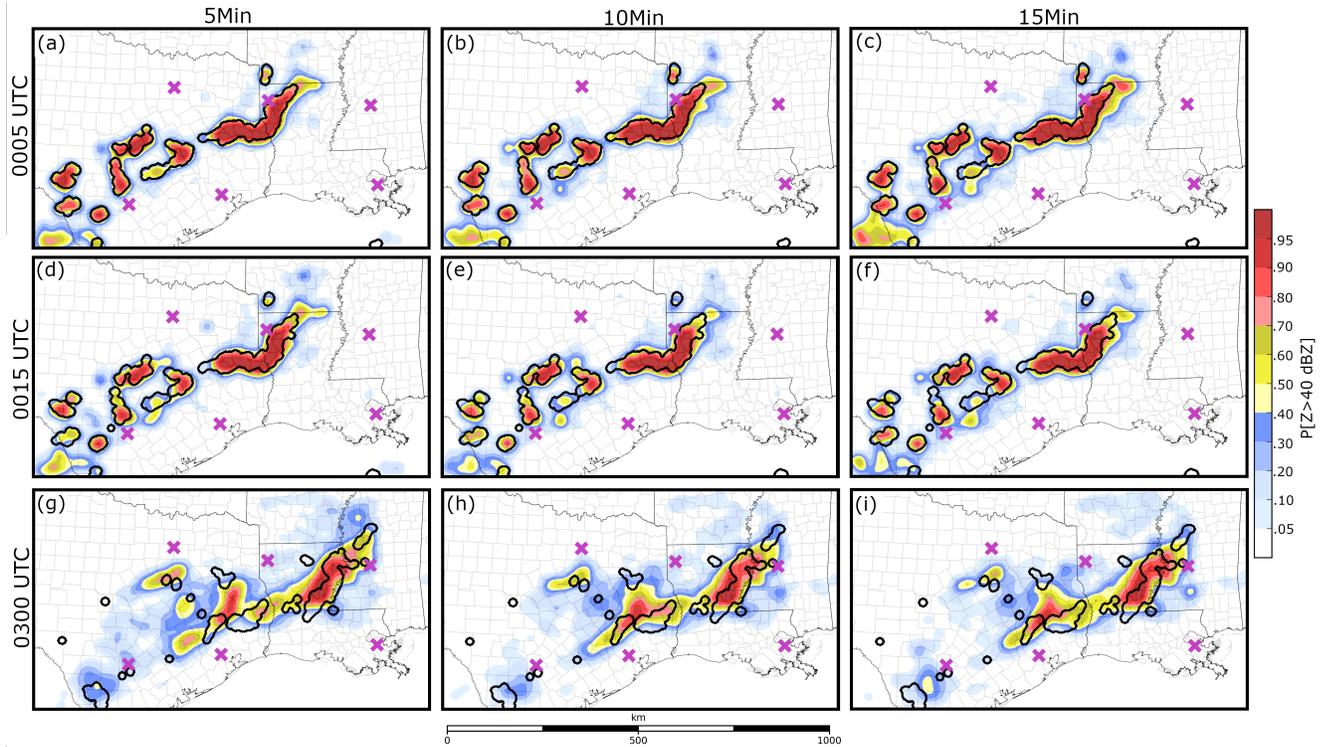


Fig. 13. The $P(Z > 40 \text{ dBZ})$ predicted by (a, d, g) 5Min, (b, e, h) 10Min, and (c, f, i) 15Min valid at the labeled times. Thick black contours represent locations where observed $Z > 40 \text{ dBZ}$. Background maps are the same as Fig. 2.

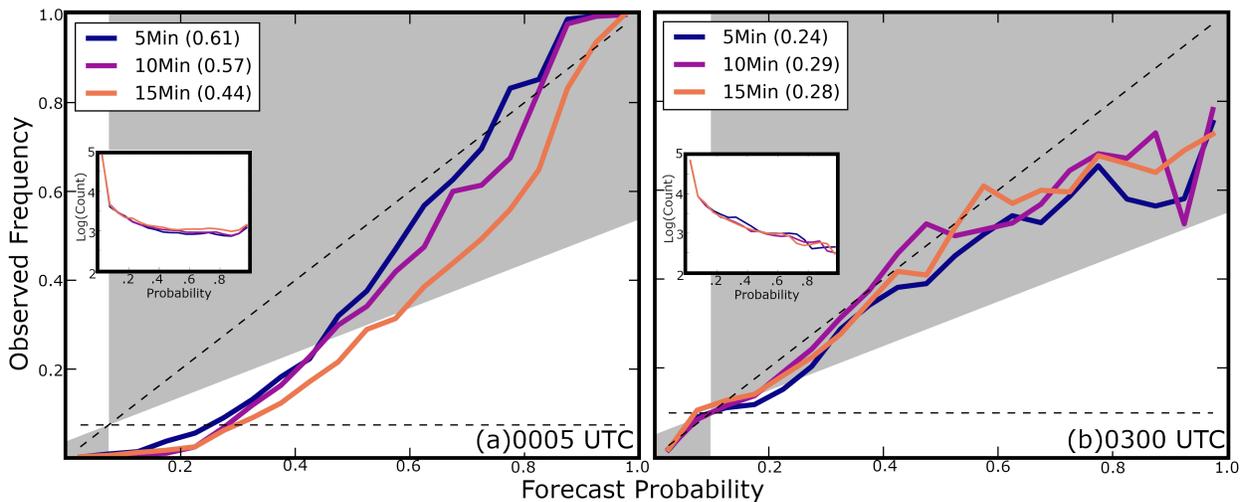


Fig. 14. Reliability diagrams for the probabilistic forecasts shown in (a) Fig. 13a-c and (b) Fig. 13g-i. DA frequency experiments correspond to different colors, the BSS for each experiment is included in the legend.

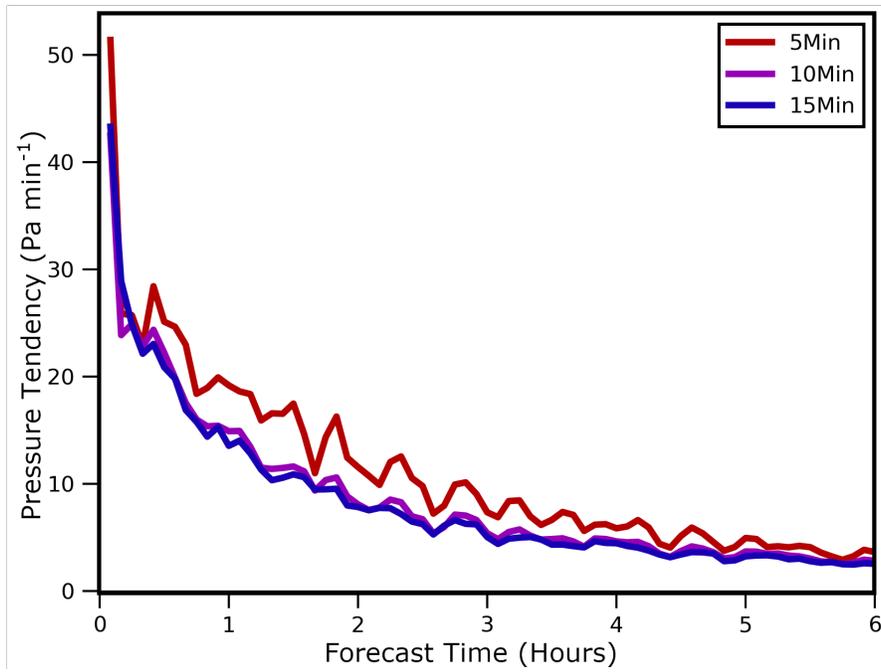


Fig. 15. The ensemble mean absolute pressure tendency during the first six forecast hours for the three DA frequency experiments.