

Impact of VORTEX2 Observations on Analyses and Forecasts of the 5 June 2009 Goshen County, Wyoming, Supercell

TIMOTHY A. SUPINIE^{1,2}, YOUNGSUN JUNG², MING XUE^{1,2}, DAVID J. STENSRUD³, MICHAEL M. FRENCH⁴,
AND HOWARD B. BLUESTEIN²

¹Center for Analysis and Prediction of Storms and ²School of Meteorology, University of Oklahoma, Norman, Oklahoma

³Department of Meteorology, The Pennsylvania State University; University Park, Pennsylvania

⁴School of Marine and Atmospheric Sciences, Stony Brook University; Stony Brook, New York

Submitted to Monthly Weather Review
April 2015

Revised October 2015

Corresponding author address:

Ming Xue, University of Oklahoma School of Meteorology
120 David L. Boren Blvd. Ste. 2500
E-mail: mxue@ou.edu

ABSTRACT

Several data assimilation and forecast experiments are undertaken to determine the impact of special observations taken during the second Verification of the Origins of Rotation in Tornadoes Experiment (VORTEX2) on forecasts of the 5 June 2009 Goshen County, Wyoming, supercell. The data used in these experiments are those from the Mobile Weather Radar, 2005 X-band, Phased Array (MWR-05XP), two mobile mesonets (MM), and several mobile sounding units. Data sources are divided into “routine,” including those from operational Weather Surveillance Radars-1988 Doppler (WSR-88Ds) and the Automated Surface Observing System (ASOS) network, and “special” observations from the VORTEX2 project.

VORTEX2 data sources are denied individually from a total of 6 ensemble square root filter (EnSRF) data assimilation and forecasting experiments. The EnSRF data assimilation uses 40 ensemble members on a 1 km grid nested inside a 3 km grid. Each experiment assimilates data every 5 minutes for 1 hour, followed by a 1-hour forecast. All experiments are able to reproduce the basic evolution of the supercell, though the impact of the V2 observations was mixed. The V2 sounding data decreased the mesocyclone intensity in the latter stages of the forecast, consistent with observations. The MWR-05XP data increased the forecast vorticity above approximately 1 km AGL in all experiments and had little impact on forecast vorticity below 1 km AGL. The MM data had negative impacts on the intensity of the low-level mesocyclone, by decreasing the vertical vorticity and indirectly by decreasing the buoyancy of the inflow.

1. Introduction

The ultimate goal of the Warn-on-Forecast project (Stensrud et al. 2009) is to issue reliable probabilistic tornado warnings based upon explicit numerical ensemble predictions of tornadoes, rather than based on radar detection. Achieving this goal faces many challenges, including the needs ranging from understanding the behavior of tornadoes and their parent supercell thunderstorms to understanding the behavior of data assimilation and numerical prediction systems used. Given these difficulties and knowing that the forecast will never be perfect, probabilistic information can be useful in the decision-making process. Recent results using ensemble Kalman filter (EnKF) methods (Evensen 2003) to initialize convective-scale ensembles for the prediction of low-level vertical vorticity swaths yield favorable comparisons to observed tornado damage tracks (Dawson et al. 2012; Yussouf et al. 2013), providing hope that the goal of Warn-on-Forecast is attainable. Snook et al. (2012; 2015) also demonstrated probabilistic forecasting skills for low-level mesovortices associated with observed tornadoes within a mesoscale convective system, when EnKF was used for data assimilation.

In addition to a rigorously-developed data assimilation algorithm, however, we require data that resolve the features of interest. Radar data provide the best opportunity for resolving supercells. However, one problem for detecting tornadoes is that because of the curvature of Earth, a single radar can only observe the lowest levels of the atmosphere over a very short range. Low-level radar observations have already proven useful in EnKF analyses of a supercell (Tanamachi et al. 2013) and in EnKF analyses and forecasts of a tornadic mesoscale convective system (Snook et al. 2011), suggesting that it is important to have low-level radar observations.

One source of low-level radar observations is the second Verification of the Origins of Rotation in Tornadoes Experiment (VORTEX2; Wurman et al. 2012) conducted during the springs of 2009 and 2010. The purpose of the experiment was to build integrated data sets of supercell thunderstorms using many observation platforms, including multiple mobile Doppler radars. Approximately 50 thunderstorms were sampled during the VORTEX2 campaign, many of which were non-tornadic or weakly tornadic, and observations from several mobile radars often are available.

One tornadic storm in particular from the 2009 campaign has already received much attention in the literature: the Goshen County, Wyoming, storm of 5 June 2009. Most of the work thus far has focused on physical understanding of supercell and tornado dynamics (e.g., Markowski et al. 2012a,b; Wakimoto et al. 2011; Atkins et al. 2012; Marquis et al. 2014; Kosiba et al. 2013). Work by Marquis et al. (2014) is unique in that it used EnKF analyses of the storm in order to close gaps in data coverage of the storm, though the study focused on physical understanding. Some preliminary forecasting work has been done using this supercell, notably by Dowell et al. (2010). Their experiments used a CONUS-scale outer ensemble at 15 km grid spacing and a nest along the Front Range of the Rocky Mountains at 3 km grid spacing. They did not assimilate observations from VORTEX2, but did assimilate observations from the Meteorological Assimilation Data Ingest System (MADIS) and 6 WSR-88D's. Their experiments found that assimilation of the WSR-88D's improved the correspondence between observed storm reports and areas of high probability of updraft helicity generated by their ensembles.

After a forecast, either deterministic or ensemble, has been obtained, one must decide the best way to evaluate the performance of that forecast. Several methods have been used in the literature. For deterministic forecasts, various observation-space statistics can be generated, as in Aksoy et al. (2010), including mean difference (bias), root mean square difference and Equitable Threat Score (ETS). Direct comparisons of model predicted fields with radar observations in the radar coordinates had also been performed (e.g., Xue et al. 2014). On fine-scale grids, positioning error becomes more prominent than on coarse-scale grids, and double-penalization, that is penalizing the model for having a feature where it did not occur *and* not having a feature where it did occur, becomes an issue. One way to alleviate this problem is to use “neighborhood” scores, such as the Fractions Skill Score (FSS; Roberts and Lean 2008), which evaluate not only at a given grid point, but all grid points within a certain radius. For ensemble forecasts, one could use probabilistic skill scores, such as relative operating characteristic (ROC) score, and ensemble probability. For example, Snook et al. (2012; 2015) used a neighborhood ensemble probability for reflectivity and low-level vertical vorticity exceeding certain thresholds to alleviate double-penalty, and ROC was used in Snook et al. (2015) for evaluating probabilistic forecasting skills of precipitation. Object-based methods can also be used for tracking discrete convective-scale features in numerical predictions (e.g., Johnson et al. 2013; Clark et al. 2014). One subjective method used by Dawson et al. (2012) is to use visual comparisons of the probability of time-maximum vorticity at a given level being greater than a threshold against an actual tornado track.

In this study, data denial experiments are conducted to determine the impact of special VORTEX2 observations on the analysis and prediction of the Goshen County supercell storm. As the study is part of the larger Warn on Forecast goal, there is a focus on the utility of these observations in an operational NWP setting. This study is similar in design to Dowell et al. (2010), but deals specifically with special observations taken during VORTEX2 and evaluating their importance to creating a realistic forecast of supercell behavior. Similar experiments have also been conducted by Marquis et al. (2014). Tanamachi et al. (2013) are the closest in spirit to the current study, as they investigated the impact of University of Massachusetts X-band, polarimetric radar data on analyses of the Greensburg, Kansas, supercell of 4 May 2007. However, they did not examine storm-scale (1-hour) forecasts resulting from these analyses. Snook et al. (2012) examined the impact of data from limited-range X-band radars on 3-hour forecasts of a mesoscale convective system. This study is unique in examining the relative impact of special VORTEX2 observations on convective-scale analyses and forecasts of a tornadic supercell. Studies looking at convective-scale forecasts, but not investigating the impact of different data sources based on ensemble data assimilation methods include Dawson et al. (2012), Yussouf et al. (2013), and Jung et al. (2012).

The rest of this paper is organized as follows. Section 2 gives the information about the observations used and the design of the experiments. General evolution of the simulated supercell and comparisons to observed data are discussed in Section 3. Section 4 focuses on the impact of the observations on the analyses forecasts. Conclusions are given in section 5.

2. Experiment Design

a. Prediction model configurations

The Advanced Regional Prediction System (ARPS; Xue et al. 2000; 2003) version 5.3 is used for the forward prediction model, and the Ensemble Square Root Filter (EnSRF; Whitaker and Hamill 2002) is used as the data assimilation method. The data assimilation and forecast experiments are performed on two domains: an inner domain with 1 km grid spacing one-way nested inside an outer domain with 3 km grid spacing (Fig. 1), each with 40 ensemble members. Each member on the inner domain receives boundary conditions from the corresponding member on the outer domain. The location for the inner domain is centered on one of the VORTEX2 soundings launched in the inflow region to the supercell, and it is large enough to contain the storm throughout its life cycle (Fig. 1). Full physics and terrain are used for all experiments, including Lin et al. (1983) microphysics, with the rain intercept parameter reduced by a factor of 20 from the default value, as suggested by Snook and Xue (2008) and Dawson et al. (2010). The Lin scheme was used instead of a theoretically more accurate, but more computationally expensive, double- or triple-moment scheme in order to mimic a potential operational configuration. A planetary boundary layer scheme based on Sun and Chang (1986) is used for all experiments. More details on the physics options can be found in Xue et al. (2001). Selected configuration settings for the experiments for both domains are given in Table 1. These parameter settings were chosen after numerous experiments. For example, smaller observation error standard deviations (OESD) often led to filter divergence because of the large volume of data. Using larger OESDs helped prevent filter divergence and maintain ensemble spreads at desirable levels.

b. Ensemble filter configurations

The EnKF data assimilation system based on the EnSRF algorithm is the one developed for the ARPS modeling system, as documented in a number of papers including Xue et al. (2007), Tong and Xue (2008), and Wang et al. (2013b). The initial ensemble on the outer domain was generated by adding initial perturbations to the North American Mesoscale forecast system (NAM) analysis at 1800 UTC on 5 June 2009. Similar to Jung et al. (2012), the added initial perturbations were spatially correlated and generated by applying a recursive filter (Lorenc 1992) to random Gaussian perturbations. This procedure is a computationally more efficient alternative to the perturbation smoothing procedure introduced into the ARPS EnKF data assimilation system by Tong and Xue (2008) and used in many later studies (e.g., Xue et al. 2009, 2010). In addition, perturbations with a smaller spatial correlation scale were added to the 1-km inner domain ensemble at the start of this nested grid ensemble to introduce additional convective perturbations. On the outer and inner domains, the horizontal decorrelation length scales of the perturbations were 12 km and 6 km, respectively, and the vertical decorrelation length scale was 3 km for both domains. The perturbation magnitudes in terms of standard deviations were 2 m s^{-1} for the zonal and meridional components of wind (u and v , respectively), 1 K for potential temperature (θ), 0.6 g kg^{-1} for water vapor mixing ratio (q_v) on the outer domain, and 0.4 g kg^{-1} for q_v on the inner domain. These magnitudes were chosen by trial and error to give as much spread as possible in the initial conditions without initiating too much spurious convection.

Deep convection initiated along the Laramie Mountains in southeastern Wyoming after 1930 UTC on 5 June 2009, with the convection showing supercellular characteristics after 2100 UTC (Markowski et al. 2012a). Thus, on the outer domain (c.f. Fig. 1), routine radar and surface data are assimilated every 30 minutes between 1800 and 2100 UTC, while sounding

and profiler data are assimilated every hour during the same period. Between 2100 and 2200 UTC, radar and surface data are assimilated every 5 minutes, with profiler and sounding data still assimilated every hour. Clear-air reflectivity observations from all three WSR-88D radars are assimilated in order to suppress spurious convection (Tong and Xue 2005). Data from all WSR-88Ds are assimilated on the outer domain, while MWR-05XP data are not. On the inner domain, radar and surface data are assimilated every 5 minutes between 2100 and 2200 UTC, with sounding and profiler data assimilated every hour. Data from the KRIW radar are not assimilated on the inner domain because of its distance from the domain. See Fig. 1c for a timeline of the experiment and radar data assimilation configurations.

Spread during the assimilation period is maintained using a combination of static multiplicative (Anderson 2001) and adaptive relaxation to prior spread (RTPS; Whitaker and Hamill 2012) covariance inflation. For the first two cycles, multiplicative inflation with inflation factor $\alpha_{mult} = 1.03$ is used over the entire domain. From there (1930 UTC) on, a combination of multiplicative inflation with $\alpha_{mult} = 1.20$ in regions where the reflectivity of the ensemble mean is greater than 15 dBZ and RTPS everywhere in the domain with $\alpha_{adapt} = 0.9$ is used. Other filter parameter configurations, including localization radii and assumed OESDs for the various observation types, are found in Table 1.

c. Observations

The data used in the experiments are divided into two groups: routine observations and special observations (i.e., those taken as part of the VORTEX2 experiment). Routine observations include surface temperature, moisture, and horizontal wind from the ASOS network, horizontal wind from the NOAA Profiler Network, and Level-II reflectivity and radial velocity from three WSR-88D radars: Cheyenne, WY (KCYS), Denver, CO (KFTG), and Riverton, WY (KRIW). As mentioned earlier, the radar data assimilation includes clear-air reflectivity observations to help suppress convection where precipitation not observed. As the experiments take place between synoptic times, no routine rawinsonde data are available. Manual quality control was performed on data from KCYS and KFTG, including velocity unfolding and ground clutter removal. Data from KRIW were not manually quality controlled, as the radar was out of range of the supercell; however those data were subjected to automated quality control routines, such as velocity unfolding and automated ground clutter removal [see Brewster et al. (2005) for details].

Special observations include surface temperature, moisture, and horizontal wind from the NSSL mobile mesonet (NSSLMM; Straka et al. 1996) and the Texas Tech University Sticknet (Weiss and Schroeder 2008), and upper air temperature, moisture, and horizontal wind from mobile sounding units. The locations of the assimilated conventional data can be found in Fig. 2. Because of limitations in the implementation of the data assimilation algorithm, we retain the synoptic-scale assumption that each sounding profile is representative of the column above its release point at the beginning of the hour following its release. These are represented by the dashed lines in Fig. 2.

The NSSLMM probes were mounted on cars, labeled “P1” through “P8”, excluding “P7”, and thus could be mobile while taking data. The Sticknet probes were stationary once deployed. Both data sets included several quality control flags for suspicious readings of the observed quantities. The high temporal resolution of these data sets, at 1 second or less

intervals, allowed for a strict quality control procedure while keeping temporal errors in the observations used for assimilation low; thus, observations were discarded if any of the quality control flags were set. The NSSLMM cars also have two temperature sensors: a “slow” and “fast” response sensor, and a flag is included in the data file that specifies whether the sensors are well-aspirated. The slow-response sensor is used wherever the sensors were well-aspirated; the fast-response sensor is used otherwise. The time constant for the slow-response sensor is 1 minute, and errors because of finite instrument response time are considered negligible.

In addition, “superobs” were created from the MM data with 11 closest observations to the target time. To create the superobs, the minimum and maximum observation for each quantity over the 11-second interval were removed and the remaining observations were averaged together. If the wind direction observations fell over three or more quadrants, the winds were set to missing. None of the raw observations that were averaged to create a given superob were more than 180 seconds apart in time or 1 km apart in space.

An additional source of special observations is the Mobile Weather Radar, 2005 X-band, Phased Array (MWR-05XP; Bluestein et al. 2010; French et al. 2013), an X-band phased-array radar. For the 5 June 2009 deployment, the volumetric update interval is 6-9 s, and volumes are scanned up to 20° in elevation. A nice property of the phased array radar for data assimilation is that elevation sweeps are taken nearly simultaneously in the context of the scan. This reduces or eliminates position errors between elevation angles associated with the scanning strategies of dish radars. Additionally, MWR-05XP data begin at 2143 UTC, 9 minutes before tornadogenesis. Furthermore, while the MWR-05XP scanned up to 20° in elevation, DOW radars scanned up to 16° in elevation (Kosiba et al. 2013), NOXP scanned up to 11° (Schwartz and Burgess 2010), and the U-Mass X-Pol scanned up to 14.4° (Jeffrey Snyder 2013, personal communication). Thus, using MWR-05XP allows for better sampling of the mid-levels of the storm than the other mobile radars scanning the storm.

The angular resolution of the MWR-05XP is fairly coarse: the half-power beamwidth is 1.8° and 2.0° in azimuth and elevation, respectively, and the sampling intervals are 1.5° in both directions. This is permissible for this study because the radar is located no more than 25 km, and sometimes as close as 10 km, away from the updraft region of the storm, during the assimilation period (French et al. 2013). The angular sampling interval results in a maximum azimuthal gate spacing of ~650 m at a range of 25 km, while the radial gate spacing is 150 m. Since these are both less than the grid spacing used for the inner domain, even in the worst case, the data density is greater than the grid density. Additionally, the focus for this study is on the low-level mesocyclone, which is well-resolved by the MWR-05XP data and mostly unobserved by the nearest WSR-88D. In addition to the relatively coarse angular resolution, another drawback is that the MWR-05XP truck is not leveled, which can introduce position errors in the data. Fortunately, for the 5 June 2009 deployment, the road on which the radar truck was sited was relatively level (see Fig. 1a of French et al. 2014). Thus, for this data set, the position errors are likely less than 500 m and considered to be negligible for purposes of this study.

The MWR-05XP Doppler velocity data were de-aliased and data contaminated by ground clutter were removed (French et al. 2013). Because a radar beam at X-band wavelength attenuates strongly in heavy precipitation, reflectivity data are not used directly in the assimilation, as this would require attenuation correction in the forward operator and increase the complexity of the experiments. Because of the large resolution volumes in the

MWR-05XP data, the velocity data are prone to sidelobe contamination in regions of strong reflectivity gradients, such as along the forward flank. To mitigate this contamination, radial velocity data are used only in areas where reflectivity from MWR-05XP is greater than 30 dBZ.

All radar data are interpolated linearly to model grid point locations (see the next section) in the horizontal. In the vertical direction, the data are left on the original conical sweep surfaces, and all elevation angles from all radars are used. WSR-88D volumes are available approximately every 4-5 minutes, and data are assimilated by volume. For each assimilation time, the closest volume prior to the assimilation time is used, meaning the nominal time for the assimilation always occurs during the scan. The relatively slow forward speed of the storm ($\sim 10 \text{ m s}^{-1}$) allows for volume-by-volume use of data in this manner without incurring large errors.

The total number of observations is heavily skewed towards the radar data. At 2200 UTC on the inner domain, the number of observations of radial velocity from KCYS is 63,618, those from KFTG are 9495, and from the MWR-05XP there are 1059 radial velocity observations. Also at this time, there are 638,138 reflectivity observations from KCYS and 284,715 reflectivity observations from KFTG (including clear-air reflectivity observations). For each radar and each observed quantity, the number of radar observations is the same order of magnitude across time steps. By comparison, there are just 4 soundings with 224 sounding observations of each quantity at 2200 UTC, 3 ASOS observations, and 9 MM observations.

d. Data-denial experiments

The control (CTRL) run assimilates all routine observations, including those from the WSR-88D radars, as well as MWR-05XP, MM, and sounding data. To investigate the impact of each VORTEX2 data source individually, MWR-05XP data, MM data, and sounding data were removed from the assimilation in turn. These experiments will be referred to as NO_MWR, NO_MM, and NO_SND, respectively. Additionally, MWR-05XP and MM data were removed together in one experiment (NO_MM_MWR), and all three were removed in another experiment (NO_V2). VORTEX2 data were only denied on the inner domain in order to have a consistent set of boundary conditions and thereby reduce the number of potential sources of variability.

3. Analyses and evolution of the supercell

First, the model reflectivity is examined to determine the basic evolution of the supercell in the model. Reflectivity is computed using a beam-weighting technique on radar scan surfaces to facilitate direct comparison to radar observations. The technique is the same as that used in Xue et al. (2006), and this is hereafter referred to as “simulated reflectivity.” Instead of examining the ensemble mean, which has a tendency to “smear” the storm as placement errors increase, we examine the ensemble member closest to the mean, which is identified in the same manner as Yussouf et al. (2013). The storm motion of the member closest to the ensemble mean mirrors that of the observed storm (Fig. 3), though the observed storm moved more slowly. The major difference is that the reflectivity in all five experiments shown weakens significantly by 2220 UTC in comparison to that observed. This behavior is observed in a few previous studies of high-resolution supercell simulations

(Yussouf et al. 2013; Lei et al. 2008). Lei et al. (2008) suspected that smaller scale structures, such as the low-level rotation could be under-predicted at 1 km grid spacing. Additionally, in the forecasts, by 2240 UTC, a secondary storm has developed to the southwest and has become the dominant storm by 2300 UTC. The development of the secondary storm occurred in the observed atmosphere between 2300 and 0000 UTC (not shown), and the secondary storm produced a second weak tornado at 2349 UTC. The storm in NO_V2 experiment had the strongest updraft at the end of the forecast period. The storm in NO_MWR_MM evolved in largely the same fashion as that in NO_V2, and is therefore not shown.

In terms of the observation-space statistics [Fig. 4; summarized below, see Dowell and Wicker (2009) for a full description of the parameters], the experiments are largely the same. The consistency ratios are similar across all experiments as RTPS covariance inflation and multiplicative inflation were applied to help maintain the ensemble spread. Even with the covariance inflation, the observation-averaged ensemble spread is less than the root-mean-square innovation throughout the assimilation period and forecast, a typical sign of under-dispersion. The under- or over-dispersion of an ensemble can be quantified by the consistency ratio (Fig. 4b); values near or above 1, as seen in these experiments, typically mean the ensemble is properly dispersed. This is mostly due to a large observation error standard deviation, 7 dBZ, which appears in the numerator of the consistency ratio equation [eq. (3.4) of Dowell and Wicker (2009)] and is likely inflating the statistic. Thus, we consider the ensemble to be under-dispersed. However, as noted earlier, this under-dispersion did not necessarily degrade the filter performance, consistent with previous studies (Tanamachi et al. 2013; Dawson et al. 2012; Snook et al. 2011).

In the case of both MM and MWR-05XP data, neither involve direct assimilation of reflectivity observations, so it is through cross-covariances among model state variables as well as interactions among the variable during the prediction that establish the storm reflectivity in the model. The behavior of the observation-space statistics for radial velocity (v_r) are similar.

To quantitatively evaluate the performance of the ensemble, the Area Under ROC Curves (AUC; Mason 1982; Wilks 2006) are computed and presented in Fig 5. The score is derived from Relative Operating Characteristic (ROC) curves and AUC has the ability to discriminate between events and non-events. The range of values for AUC is 0.0 to 1.0, with 1.0 representing a perfect probabilistic forecast and 0.5 representing probabilistic forecasts no better than random. To compute AUC, simulated reflectivity is computed for each ensemble member for all 14 tilts from KCYS. The verifying observations were taken from KCYS only, which covered the majority of the inner domain, save the far northeastern corner (see Fig. 1).

The AUCs for all experiments start near 0.95 for reflectivity ≥ 25 dBZ and near 0.9 for reflectivity ≥ 45 dBZ (Fig. 5). For the 25 dBZ threshold, which corresponds to areas of any precipitation, the score for each experiment falls steadily until the skill reaches 0.5 by the 1-hour forecast. The primary reason for the decrease in performance with increasing lead time is the increase in false alarms, i.e., regions where reflectivity ≥ 25 dBZ was forecast but not observed. The false alarms can be seen in Fig. 3, particularly in regions northwest of the supercell, where convection was forecast, but not observed. For the 45 dBZ threshold (Fig. 5b), which corresponds to areas of heavy precipitation, the AUC score remains higher than that for the 25 dBZ threshold until approximately the 45-minute forecast. At this point, the

performance starts to decrease and by 2255 UTC drops below the 0.7 value suggested by Kong et al. (2011) as the threshold for a useful forecast. The relatively steady AUC in the early part of the forecast is likely because most areas of spurious convection in the forecast have reflectivity less than 45 dBZ. Thus, they do not show up as false alarms for that threshold.

Additionally, the NO_V2, and NO_MWR_MM experiments have the highest AUC for both thresholds throughout each experiment, though statistically insignificant for the 25 dBZ threshold. However, a close look at the reflectivity patterns reveals that all experiments but CTRL show splitting storms that do not match observations (see Fig 3). AUCs for high thresholds are less confident because of displacement errors of reflectivity cores and underestimation of areal coverage of those cores in CTRL, in addition to the small sample size.

The performance of the ensemble, in terms of ensemble mean forecast, is also evaluated against MM data (Fig. 6). Three mobile mesonet instruments are used in the comparison: one Sticknet site (“102B” in Fig. 6) and two NSSL MM cars (“P2” and “P4” in Fig. 6). Experiments assimilating MM data generally better fit the analyses and forecasts at all three sites. The 102B site starts on the forward flank gust front at 2200 UTC and transitions to the rear-flank cold pool by 2215 UTC. The corresponding increase in u occurs 10 minutes too soon in the ensembles, suggesting that the cold pool is advancing faster in the model than in the real atmosphere. The same behavior is noted for P4, which starts in the inflow at 2130 and transitions to the rear-flank cold pool by 2215 in the observed atmosphere. This transition occurs 5 minutes too early in the ensembles, again suggesting that the ensemble cold pool is too fast. P2 starts in the inflow as well and transitions to the rear-flank cold pool very near the mesocyclone around 2150 UTC. The P2 car also moves eastward and transitions back to the inflow by 2240. These transitions are captured in the temperature fields of the ensembles, but less so in the u wind fields. The early transitions from inflow to rear-flank cold pool suggest that, more so than the observed storm, the modeled storm is being forced by the cold pool, which helps explain the faster propagation of the ensemble storm when compared to the observed storm.

This can be seen in Fig. 7. In the CTRL experiment, P1 and P4 in the inflow are much colder than the ensemble mean, meaning the ensemble mean forecast is too warm. However, the temperature at P2 just behind the rear-flank gust front is near the ensemble mean. The same can be seen from the difference plots for other experiments; the differences between the forecasts are approximately the same between the inflow and outflow, but the P2 is closer to the CTRL experiment than P1 and P4, meaning the gradient is stronger in the ensemble. In addition, the assimilation of MM v observations at the time of Fig. 7 (2150 UTC) cooled the region just west of the gust front, and assimilation of MM potential temperature observations had little effect on the cold pool temperature. The effect of these is that the propensity of the gust front to propagate like a density current is enhanced compared to the real atmosphere. Overall, however, the ensembles are able to capture the general trend of the observations well, even though the ensemble is generally warmer and drier than observed.

Finally, in Fig. 6, clustering of experiments with and without MM observations is evident. This is likely a product of the near-surface nature of the observations, which is subject to the land-surface model and boundary layer scheme in the forward prediction model. These both are constant across all members and experiments, meaning that the evolution of the near-surface environment has little variability. This issue could be alleviated using a multi-physics

ensemble (Berner et al. 2011). In any case, these results suggest the ensembles are capturing the general evolution of the supercell reasonably well.

4. Forecast mesocyclone tracks

In examining the forecast mesocyclone tracks in these experiments, we use the Updraft Helicity (UH) metric (Fig. 8). UH is used as defined in Kain et al. (2008), except herein, the limits of integration are from the surface to 3 km AGL so as to pick out low-level rotating updrafts. Additionally, instead of a 9-point smoother, a neighborhood ensemble probability has been computed using a circular neighborhood 2.5 km in radius. Here, the probability of UH represents a mesocyclone-scale circulation as a proxy for a tornado track. The results presented here use a UH threshold of $75 \text{ m}^2 \text{ s}^{-2}$, but they are relatively insensitive to the threshold used.

The UH swaths are compared to the Warning Decision Support System - Integrated Information (WDSS-II) rotation swath product (Smith and Elmore 2004). The peak probabilities in the forecast mesocyclone tracks from all six experiments generally follow the WDSS-II rotation track in space, as expected for having initialized the 1-hour forecasts after tornadogenesis. This again supports the conclusion that the ensembles are capturing the evolution of the mesocyclone in the forecast. However, some differences are apparent between the individual experiments, discussed in the next sections.

a. Impact of mobile sounding observations

Omitting the sounding data results in a larger area of UH probabilities greater than 50% in the latter half of the forecast than assimilating sounding data (Fig. 8). This suggests higher probabilities in the NO_SND and NO_V2 experiments than the other experiments, while the WDSS-II rotation track suggests a weaker mesocyclone during this period of the forecast.

To examine the reason for the higher UH probabilities in the experiments without soundings, we examine the difference between CTRL and NO_SND experiments and between the NO_MM_MWR and NO_V2 experiments (Fig. 9). The source of the decreased UH in the CTRL and NO_MM_MWR experiments appears to be from the sounding launched farthest southeast. That sounding produces southeastward increments in the velocity, which corresponds to a decrease in wind speed in the inflow to the storm. The information from the soundings is advected northwestward with the wind in the inflow, producing the observed changes in the hodograph at 2240 UTC. The effect on the storm is that the soundings decrease vertical velocity at 500 m and therefore decrease low-level vorticity stretching. Both of these have a direct impact on the UH.

b. Impact of mobile radar observations

The MWR-05XP data appear to have no systematic effects on the UH probability swaths. However, additional information can be extracted by examining how the low-level vorticity field (Fig. 10) evolves with time. The CTRL and NO_MM experiments have initialized the most intense mesocyclone in terms of vertical vorticity (ζ) in the ensemble mean. The next most intense in the mean is the NO_MWR_MM experiment, followed by the NO_MWR experiment, which is the least intense in the ensemble mean. As expected, the experiments assimilating MWR-05XP data have a more intense mean low-level mesocyclone than those experiments not assimilating them. This is consistent with the findings of Tanamachi et al. (2013).

Without MWR-05XP data, the members of the NO_MM_MWR experiment show larger variability in the placement of the $\zeta = 0.015 \text{ s}^{-1}$ contours, whereas the individual members of the CTRL and NO_MM experiments are relatively consistent on the placement of those contours. This suggests that the high-spatial-resolution MWR-05XP data reduce the uncertainty in the placement of the vortex. Larger variability in the placement of the vorticity maxima in the ensemble is partially responsible for the lessened ensemble mean ζ in experiments without MWR-05XP data.

By the 5-minute forecast (Fig. 11), the mesocyclones in all four experiments have weakened considerably. The most dramatic weakening is in the CTRL experiment, which has one of the strongest mesocyclones in the analysis and one of the weakest mesocyclones in the 5-minute forecast. The decrease in intensity of the ensemble mean ζ in this experiment appears to be primarily the result of the decreased intensity in the individual members. The NO_MM and NO_MM_MWR members also exhibit a reduction in intensity, but the magnitude of reduction is smaller than the CTRL and NO_MWR experiments. A rapid reduction in mesocyclone intensity has also been observed in the 5-minute forecast of a Mesoscale Convective System using an EnKF (Snook et al. 2011). Some of the weakening in NO_MM and NO_MWR_MM appears to be associated with differences in placement of the vortex. The vertical vorticity analyses of Atkins et al. (2012) suggest that the observed low-level mesocyclone undergoes a general intensification between 2200 and 2210 UTC, but this is not observed in any of the ensembles here, as they are likely still in the adjustment period of the initial forecast (Putnam et al. 2014).

Furthermore, during the assimilation period, MWR-05XP data increased the ensemble mean of the maximum ζ over the domain, particularly between approximately 2.5 to 4 km MSL (1 to 2.5 km AGL; Fig. 12). However, there is no discernible difference between the 5-minute ζ forecasts at 300 m AGL from the NO_MM and NO_MM_MWR experiments in Fig 11. We view this as evidence that the inclusion of MWR-05XP data do not always increase near-surface ζ in a forecast, even though the influence is apparent in the analysis.

c. Impact of mobile mesonet observations

An examination of Fig. 8b shows that the NO_MWR experiment has the least area of 50% UH probability out of all the experiments in the first part of the forecast, implying that adding MM observations without the presence of the MWR-05XP data decreased the intensity of the mesocyclone in the ensemble. In addition, ζ at 300 m AGL decreases drastically in the first 5 minutes of the 1-hour forecast (Figs. 10 and 11) in experiments that assimilate MM data (CTRL and NO_MWR). Furthermore, two side experiments are performed that are identical in configuration to the CTRL experiment, except that in one, MM wind observations were omitted, and in the other MM thermodynamic observations were omitted. In both these experiments, the ensemble mean ζ at 300 m AGL is larger, though this appears to be mainly due to better agreement among the members in placement of the mesocyclone. Further, ζ decreases in the 5-minute forecast as quickly in both side experiments as in the CTRL experiment. The MM observations were anticipated to provide better information on surface conditions around the storm, so a decrease in mesocyclone intensity and longevity resulting from MM assimilation was unexpected during an intense period in the observed storm.

A possible reason for the unexpected behavior of the mesocyclones can be found in examining the forecast ζ during the analysis period (Fig. 12). Whereas the last 10 minutes of

the assimilation period in the NO_MM and NO_MWR_MM experiments feature a general increase in or maintenance of forecast low-level ζ , the same period in the CTRL and NO_MWR experiments features a decrease in forecast low-level ζ . Atkins et al. (2012) and Kosiba et al. (2013) found that 2150 UTC begins a period of intensification in the observed low-level mesocyclone, so the low-level mesocyclones in the CTRL and NO_MWR experiments do not mimic the observed. Additionally, the experiments assimilating MM data decrease ζ in the 2150 UTC analyses, where the forecast ζ at 2150 UTC is already weaker in those experiments than in the experiments not assimilating MM data.

To explain the decrease in ζ at the 2150 UTC analysis, we examine the analysis increment of ζ produced by wind observations from individual MM cars (Fig. 13). It is apparent that the v wind observation from probe 2 (P2) contributes the overwhelming majority of the ζ decrease, with most of the contributions from the other observations being neutral or slightly positive. This is partially the result of the large deviation of the ensemble mean v wind from the observed v , 5.6 m s^{-1} , with the ensemble mean wind being much stronger than observed. This is consistent with the cold pool being too strong, as noted above.

In addition to the cold pool being too cool, the MM observations decrease the temperature of the near inflow to the storm (Fig. 7). This affects the buoyancy of the parcels lifted in the updraft; less buoyant parcels will not be lifted as quickly. This is reflected in the ensemble mean of the domain-maximum forecast $\frac{\partial w}{\partial z}$ (Fig. 14), particularly in the 2155 and 2200 UTC forecasts. The quantity $\frac{\partial w}{\partial z}$ was chosen to evaluate the stretching of ζ , while removing the effect of the ζ already present. It also reflects the vertical acceleration, which is proportional to the buoyancy of the inflow air. One would expect low-level $\frac{\partial w}{\partial z}$ to be higher in cases with larger low-level buoyancy (Markowski and Richardson 2014), and this is consistent with Fig. 14. The CTRL and NO_MWR experiments both have lower buoyancy in the inflow, and they both exhibit lower $\frac{\partial w}{\partial z}$ than the NO_MM and NO_MM_MWR experiments. Thus, pre-existing ζ undergoes less stretching in experiments that assimilate MM observations, which accounts for the weaker low-level mesocyclone in those experiments. In addition to the forecast $\frac{\partial w}{\partial z}$, which is at least partially a reflection of the low-level buoyancy in the model, the increases in $\frac{\partial w}{\partial z}$ by the data assimilation are greater in the NO_MM experiment than the CTRL experiment. The same is true to a lesser extent in the NO_MM_MWR experiment versus the NO_MWR experiment. Thus, the assimilation of MM data is also directly decreasing vertical acceleration, and therefore the stretching of ζ .

5. Conclusions

We have presented data assimilation and forecast experiments using data from the 5 June 2009 Goshen County, Wyoming, supercell. This storm was well-observed by the VORTEX2 project, including by the MWR-05XP mobile X-band phased array radar, two mobile mesonets (MMs), and several sounding teams. Data were divided into “routine” and “VORTEX2” (V2) observations; groups of V2 observations were removed one by one or all together from the assimilation in each experiment to determine the impact of special, high-resolution observations on forecasts of the supercell. Data were assimilated at 5-minute intervals for an hour into an outer mesh at 3 km resolution and an inner mesh at 1 km

resolution, which produced a realistic initial state for the storm. The subsequent 1-hour forecasts from all experiments captured the evolution of the supercell structure reasonably well and captured some observed trends in surface observations.

While most V2 data sources were found to be beneficial to the analyses and subsequent forecasts, one data set exhibits a negative impact on the forecasts. Special soundings were found to generally have a positive impact on the latter stages of the 1-hour forecast. Specifically, the sounding data from the upstream inflow region resulted in a negative velocity increment in the inflow, which decreased the strength of the updraft in the storm. Both the increase in updraft and subsequent increase in low-level vorticity stretching account for the increased UH seen in the latter stages of the forecast.

Furthermore, the MWR-05XP radial velocity data were found to have the effect of increasing the mean low-level vorticity in the analyses, consistent with that found by Tanamachi et al. (2013). The increase in vorticity was at least partially a result of reducing vortex placement differences in the members. Additionally, while the MWR-05XP data are useful in increasing vertical vorticity (ζ) aloft in a forecast, it is less useful in increasing forecast ζ near the surface. Collectively, these results are consistent with the assimilation of radar data, which provide kinematic information about the mesocyclone, but convey little directly about the storm environment, and therefore, the conditions that at least partially contribute to the maintenance of the mesocyclone.

The MM data were found to have a negative impact, through both direct and indirect means. Directly, the MM observations decrease vertical vorticity in the 2150 UTC analysis through covariances between observed v and ensemble u and v . This is likely because outflow in the model background is too strong, which in turn is because the contrast between the cold outflow and warm inflow is greater than observed. Indirectly, the MM obs also decrease the temperature of the inflow to the storm, which also decreases the low-level stretching of parcels from the inflow. We applied several forms of quality control to the MM data, including creating super-obs and removing data that did not meet supplied quality control metrics. Markowski et al. (2012a,b) and Marquis et al. (2014) used similar quality control procedures for the same MM data in their studies. Additionally, Marquis et al. (2014) assimilated thermodynamic MM data and found it had a positive impact on the analyses; however, they did not examine the forecasts from those analyses. Despite extensive efforts in quality controlling the data, and in tuning the assimilation configurations, we were not able to obtain a positive impact of the MM data. This is not necessarily because the analyses were not fitting the observations (see Fig. 6), but because fitting the observations moved the ensemble state into one less favorable for sustaining a mesocyclone. While such results are not expected in a statistical sense, it can happen with individual cases, when model errors, background state errors, and errors from other observations can work in complex way against the realization of the benefit of one particular data type. We consider the current case one of such examples.

The computational requirements for these experiments were quite steep; each forecast step took approximately 10 minutes, and the EnSRF program took about 15 minutes. Thus an optimal configuration would take 25 minutes for a 5-minute cycle, much too slow for operational implementation. While the radar data comprised the bulk of the assimilated data volume, the efficient parallelization for radar data (Wang et al. 2013b) meant that the conventional data assimilation required a disproportionate amount of time. Unfortunately from this perspective, sounding data had the greatest positive impact on the 1-hour forecast,

and the MWR-05XP data had negligible impact after the first 10 minutes of the 1-hour forecast. Additionally, we note that the sounding data contain mostly information on the environment, while the MWR-05XP data contain mostly information on the storm itself. These suggest that, for operational forecasting at 1 km grid spacing, it is important to invest computational time in assimilating data that describe the environment, rather than the storm itself.

In addition to the data used in the assimilation, the microphysics scheme used in the prediction model has been shown to have a strong influence on the structure and evolution of simulated storms (Dawson et al. 2012). Whereas a single-moment microphysics scheme such as the Lin scheme keeps a constant intercept parameter on the Marshall-Palmer distribution, double-moment schemes let those intercept parameters vary. Several studies show that reflectivity structure, cold pool intensity and polarimetric signatures simulated with multi-moment schemes are more consistent with observations (e.g. Dawson et al. 2010, 2014; Jung et al. 2012; Putnam et al. 2014). Thus, a future avenue for work will be to investigate any structure and evolution changes to the storm with different microphysics and whether the MM data have a more positive impact given a more complex and hopefully more accurate microphysics scheme. This could also lead to using polarimetric observations in the data assimilation, such as in Jung et al. (2008).

Finally, the phased array MWR-05XP produced 6- to 9-second volume scans, which were not fully utilized in these experiments. Future work could assimilate data more frequently than 5 minutes, which would allow for more MWR-05XP volumes to be utilized. The 4DEnSRF implementation by Wang et al. (2013a) would be suitable, which assimilates frequent data collected over a time span in a single filter step. This allows one to relax the assumption that all observations are valid at the same time in each assimilation cycle, which reduces the timing errors of observations and reduces model adjustments at the beginning of each cycle.

Acknowledgments. Thanks go to Kevin Manross, Brandon Smith, and Lamont Bain for quality controlling the WSR-88D data used in these experiments. Paul Markowski and Yvette Richardson of Penn State University and Jerry Straka of the University of Oklahoma collected the NSSL mobile mesonet data, and Chris Weiss of Texas Tech University collected the Sticknet data used in these experiments. Josh Wurman provided the observed tornado track data as part of the NSF-sponsored Doppler On Wheels (DOW) project. Additionally, Robin Tanamachi and Yunheng Wang helped in troubleshooting problems with the experiments, and discussions with Nate Snook helped increase the quality of the research. Jim Marquis and one anonymous reviewer also increased the quality of the manuscript. This research has been primarily supported by the NOAA Warn-on-Forecast project under grant number DOC-NOAA NA080AR4320904 and by NSF grant AGS-0802888. The MWR-05XP has been supported by NSF Grants ATM-0637148 and AGS-0934307. Finally, thanks to the National Institute for Computational Sciences (NICS) and the Texas Advanced Computing Center (TACC) for the use of their supercomputers.

References

- Aksoy, A., D. C. Dowell, and C. Snyder, 2010: A multicas e comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts. *Mon. Wea. Rev.*, **138**, 1273–1292.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Atkins, N. T., A. McGee, R. Ducharme, R. M. Wakimoto, and J. Wurman, 2012: The LaGrange tornado during VORTEX2. Part II: Photogrammetric analysis of the tornado combined with dual-Doppler radar data. *Mon. Wea. Rev.*, **140**, 2939–2958.
- Atkins, N. T., E. M. Glidden, and T. M. Nicholson, 2014: Observations of wall cloud formation in supercell thunderstorms during VORTEX2. *Mon. Wea. Rev.*, **142**, 4823–4838
- Berner, J., S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995.
- Bluestein, H. B., M. M. French, I. PopStefanija, R. T. Bluth, and J. B. Knorr, 2010: A mobile, phased-array Doppler radar for the study of severe convective storms. *Bull. Amer. Meteor. Soc.*, **91**, 579–600.
- Brewster, K., M. Hu, M. Xue, and J. Gao, 2005: Efficient assimilation of radar data at high resolution for short-range numerical weather prediction. *Preprints, WWRP Int. Symp. on Now-casting and Very Short Range Forecasting*.
- Chapra, S. C. and R. P. Canale, 2006: *Numerical Methods for Engineers*. McGraw-Hill.
- Chou, M.-D., 1992: A solar radiation model for use in climate studies. *J. Atmos. Sci.*, **49**, 762–772.
- Chou, M.-D. and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. Tech. Rep. 104606, National Aeronautics and Space Administration.
- Clark, A. J., R. G. Bullock, T. L. Jensen, M. Xue, and F. Kong, 2014: Application of object-based time-domain diagnostics for tracking precipitation systems in convection-allowing models. *Wea. Forecasting*, **29**, 517–542.
- Dawson, D. T., E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2014: Low-level Z_{DR} signatures in supercell forward flanks: The role of size sorting and melting of hail. *J. Atmos. Sci.*, **71**, 276–299.
- Dawson, D. T., L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2012: Impact of the environmental low-level wind profile on ensemble forecasts of the 4 May 2007 Greensburg, Kansas, tornadic storm and associated mesocyclones. *Mon. Wea. Rev.*, **140**, 696–716.
- Dawson, D. T., M. Xue, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms. *Mon. Wea. Rev.*, **138**, 1152–1171.
- Dowell, D., G. Romine, and C. Snyder, 2010: Ensemble storm-scale data assimilation and prediction for severe convective storms. Preprints, *25th Conf. on Severe Local Storms*, Denver, CO. Amer. Meteor. Soc., 9.5. [Available online at <http://ams.confex.com/ams/papers/176121.pdf>]
- Dowell, D. C. and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Ocean. Tech.*, **26**, 911–927.

- Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.
- French, M. M., H. B. Bluestein, I. PopStefanija, C. A. Baldi, and R. T. Bluth, 2013: Reexamining the vertical development of tornadic vortex signatures in supercells. *Mon. Wea. Rev.*, **141**, 4576–4601.
- French, M. M., H. B. Bluestein, I. PopStefanija, C. A. Baldi, and R. T. Bluth, 2014: Mobile, phased-array, Doppler radar observations of tornadoes at X band. *Mon. Wea. Rev.*, **142**, 1010–1036.
- Johnson, A., X. Wang, F. Kong, and M. Xue, 2013: Object-based evaluation of the impact of horizontal grid spacing on storm-scale forecasts. *Mon. Wea. Rev.*, **141**, 3413–3425.
- Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 29–30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric radar data. *Mon. Wea. Rev.*, **140**, 1457–1475.
- 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 II: Impact of polarimetric data on storm analysis. *Mon. Wea. Rev.*, **136**, 2246–2260.
- Kain, J. S., et al., 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952.
- Kong, F., et al., 2011: Evaluation of CAPS multi-model storm-scale ensemble forecast for the NOAA HWT 2010 spring experiment. *Preprints: 24th Conf. Wea. Forecasting/20th Conf. Num. Wea. Prediction. AMS, Seattle, WA, January 23–27, 2011*.
- Kosiba, K., J. Wurman, Y. Richardson, P. Markowski, P. Robinson, and J. Marquis, 2013: Genesis of the Goshen County, Wyoming, tornado on 5 June 2009 during VORTEX2. *Mon. Wea. Rev.*, **141**, 1157–1181.
- Lei, T., M. Xue, and T. Yu, 2008: Multi-scale analysis and prediction of the 8 May 2003 Oklahoma City tornadic supercell storm assimilating radar and surface network data using EnKF. *Preprints: 13th Conf. IOAS-AOLS. AMS, New Orleans, LA, January 20–24, 2008*.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Clim. Appl. Meteor.*, **22**, 1065–1092.
- Lorenc, A., 1992: Iterative analysis using covariance functions and filters. *Q. J. R. Meteorol. Soc.*, **118**, 569–591.
- Markowski, P., et al., 2012a: The pretornadic phase of the Goshen County, Wyoming, supercell of 5 June 2009 intercepted by VORTEX2. Part I: Evolution of kinematic and surface thermodynamic fields. *Mon. Wea. Rev.*, **140**, 2887–2915.
- Markowski, P., et al., 2012b: The pretornadic phase of the Goshen County, Wyoming, supercell of 5 June 2009 intercepted by VORTEX2. Part II: Intensification of low-level rotation. *Mon. Wea. Rev.*, **140**, 2916–2938.
- Markowski, P. M. and Y. P. Richardson, 2014: The influence of environmental low-level shear and cold pools on tornadogenesis: Insights from idealized simulations. *J. Atmos. Sci.*, **71**, 243–275.
- Marquis, J., Y. Richardson, P. Markowski, D. Dowell, J. Wurman, K. Kosiba, P. Robinson, and G. Romine, 2014: An investigation of the Goshen County, Wyoming, tornadic supercell of 5 June 2009 using EnKF assimilation of mobile mesonet and radar observations collected during VORTEX2. Part I: Experiment design and verification of the EnKF analyses. *Mon. Wea. Rev.*, **142**, 530–554.

- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteorol. Mag.*, **30**, 291–303.
- Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2014: The analysis and prediction of microphysical states and polarimetric radar variables in a mesoscale convective system using double-moment microphysics, multinet radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 141–162.
- Roberts, N. M. and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97.
- Schwartz, C. M. and D. W. Burgess, 2010: Verification of the origins of rotation in tornadoes experiment, part 2 (VORTEX2): Data from the NOAA (NSSL) X-band dual-polarized radar. *Preprints, 25th Conf. Severe Local Storms*.
- Smith, T. M. and K. L. Elmore, 2004: The use of radial velocity derivatives to diagnose rotation and divergence. *Preprints, 11th Conf. on Aviation, Range, Aerospace, Hyannis, MA*.
- Snook, N. and M. Xue, 2008: Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophys. Res. Lett.*, **35**, L24 803.
- Snook, N. J., M. Xue, and Y. Jung, 2011: Analysis of a tornadic mesoscale convective vortex based on ensemble Kalman filter assimilation of CASA X-band and WSR-88D radar data. *Mon. Wea. Rev.*, **139**, 3446–3468.
- Snook, N., M. Xue, and Y. Jung, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140**, 2126–2146.
- Snook, N. A., M. Xue, and Y. Jung, 2015: Multi-scale EnKF assimilation of radar and conventional observations and ensemble forecasting for a tornadic mesoscale convective system. *Mon. Wea. Rev.*, **143**, 1035–1057.
- Stensrud, D. J., M. Xue, L. J. Wicker, K. E. Kelleher, M. P. Foster, J. T. Schaefer, R. S. Schneider, S. G. Benjamin, S. S. Weygandt, J. T. Ferree, and J. P. Tuell, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.
- Straka, J. M., E. N. Rasmussen, and S. E. Fredrickson, 1996: A mobile mesonet for finescale meteorological observations. *J. Atmos. Ocean. Tech.*, **13**, 921–936.
- Tanamachi, R. L., L. J. Wicker, D. C. Dowell, H. B. Bluestein, D. T. Dawson, and M. Xue, 2013: EnKF assimilation of high-resolution, mobile Doppler radar data of the 4 May 2007 Greensburg, Kansas, supercell into a numerical cloud model. *Mon. Wea. Rev.*, **141**, 625–648.
- Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807.
- Tong, M. and M. Xue, 2008: Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble Kalman filter. Part I: Sensitivity analysis and parameter identifiability. *Mon. Wea. Rev.*, **136**, 1630–1648.
- Wakimoto, R. M., N. T. Atkins, and J. Wurman, 2011: The LaGrange tornado during VORTEX2. Part I: Photogrammetric analysis of the tornado combined with single-Doppler radar data. *Mon. Wea. Rev.*, **138**, 2233–2258.
- Wakimoto, R. M., P. Stauffer, W.-C. Lee, N. T. Atkins, and J. Wurman, 2012: Finescale structure of the LaGrange, Wyoming, tornado during VORTEX2: GBVTD and photogrammetric analyses. *Mon. Wea. Rev.*, **140**, 3397–3418.

- Wang, S., M. Xue, and J. Min, 2013a: A four-dimensional asynchronous ensemble square-root filter (4DEnSRF) algorithm and tests with simulated radar data. *Q. J. R. Meteor. Soc.*, **139**, 805–819.
- Wang, Y., Y. Jung, T. A. Supinie, and M. Xue, 2013b: A hybrid MPI/OpenMP parallel algorithm and performance analysis for an ensemble square root filter suitable for dense observations. *J. Atmos. Ocean. Tech.*, **30**, 1382-1397.
- Weiss, C. C. and J. L. Schroeder, 2008: Sticknet - a new portable, rapidly-deployable, surface observing system. *Preprints: 24th Conf. IIPS. AMS, New Orleans, LA, January 20-24, 2008*.
- Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Whitaker, J. S. and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed., Academic Press.
- Wurman, J., D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The second verification of the origins of rotation in tornadoes experiment: VORTEX2. *Bull. Amer. Meteor. Soc.*, **93**, 1147–1170.
- Wurman, J., K. Kosiba, and P. Robinson, 2013: In situ, Doppler radar, and video observations of the interior structure of a tornado and the wind-damage relationship. *Bull. Amer. Meteor. Soc.*, **94**, 835-846.
- 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. *Meteor. Atmos. Phys.*, **75**, 161–193.
- Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D. Wang, 2001: The Advanced Regional Prediction System (ARPS) - A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143-166.
- Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82**, 139-170.
- Xue, M., M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square root Kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting. *J. Atmos. Ocean. Tech.*, **23**, 46–66.
- Xue, M., M. Tong, and G. Zhang, 2009: Simultaneous state estimation and attenuation correction for thunderstorms with radar data using an ensemble Kalman filter: Tests with simulated data. *Quart. J. Royal Meteor. Soc.*, **135**, 1409-1423.
- Xue, M., 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. Roy. Meteor. Soc.*, **136**, 685-700.
- Xue, M., M. Hu, and A. D. Schenkman, 2014: Numerical prediction of 8 May 2003 Oklahoma City tornadic supercell and embedded tornado using ARPS with assimilation of WSR-88D radar data. *Wea. Forecasting*, **29**, 39-62.
- Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single and double moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412.

Table 1: Forward prediction model and filter configurations for the inner domain and outer domain experiments.

Parameter	Outer Domain	Inner Domain
Grid Size	411 × 411 × 51	255 × 255 × 51
Horizontal Grid Spacing	3 km	1 km
Vertical Grid Spacing	Hyperbolic Tangent Stretching, 50 m minimum	
Integration Time Step	6s	1s
Initial Conditions	5 June 2009 1200 UTC NAM + 3DVAR	2100 UTC Outer Domain
Lateral Boundary Conditions	NAM	Outer Domain Analyses
Microphysics	Lin et al. (1983)	
Rain Intercept Parameter (N0r)	$4 \times 10^5 \text{ m}^{-4}$	
Snow Intercept Parameter (N0s)	$3 \times 10^6 \text{ m}^{-4}$	
Hail Intercept Parameter (N0h)	$4 \times 10^4 \text{ m}^{-4}$	
Radiation	SW: Chou (1992); LW: Chou and Suarez (1994)	
Planetary Boundary Layer	1.5-order TKE closure (Sun and Chang 1986)	
Radar Localization Radius	6 km	6 km
Radar Error Standard Deviation	5 dBZ, 3 m s^{-1}	7 dBZ, 5 m s^{-1}
Sounding Localization Radius	150 km	25 km
Sounding Error Standard Deviation	2 K, 2.5 m s^{-1}	
Surface Localization Radius	300 km (Mesonet: 50 km)	
Surface Error Standard Deviation	1 K, 1.5 m s^{-1} (Mesonet: 1.11 K, 1.5 m s^{-1})	
Profiler Localization Radius	150 km	
Profiler Error Standard Deviation	2.5 m s^{-1}	

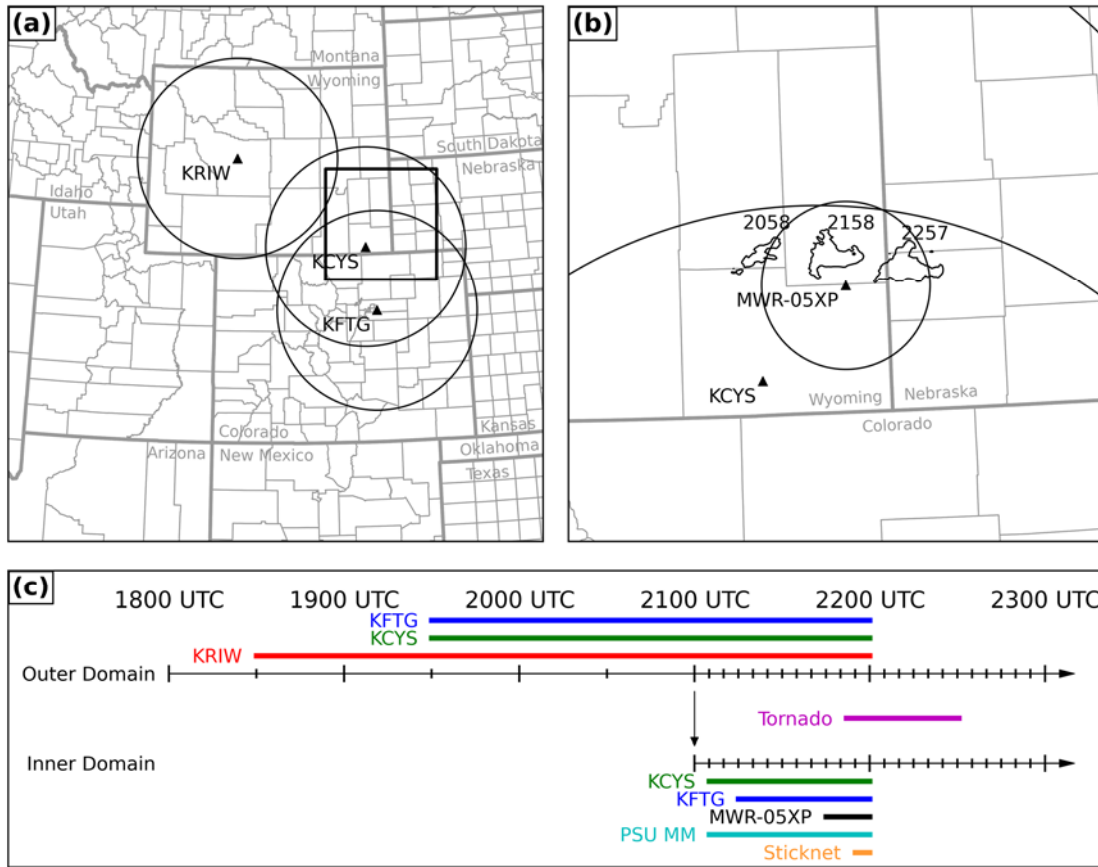


Figure 1: The locations and range rings of the radars used for the experiments in the outer domain (a) and inner domain (b). The timeline and availability of radar and mobile mesonet data for the experiments are given in (c). The location of the inner domain is plotted as a thick black square in (a), observed 40 dBZ contours from KCYS are plotted in (b), and the time of the tornado is given in (c). Times are given in UTC.

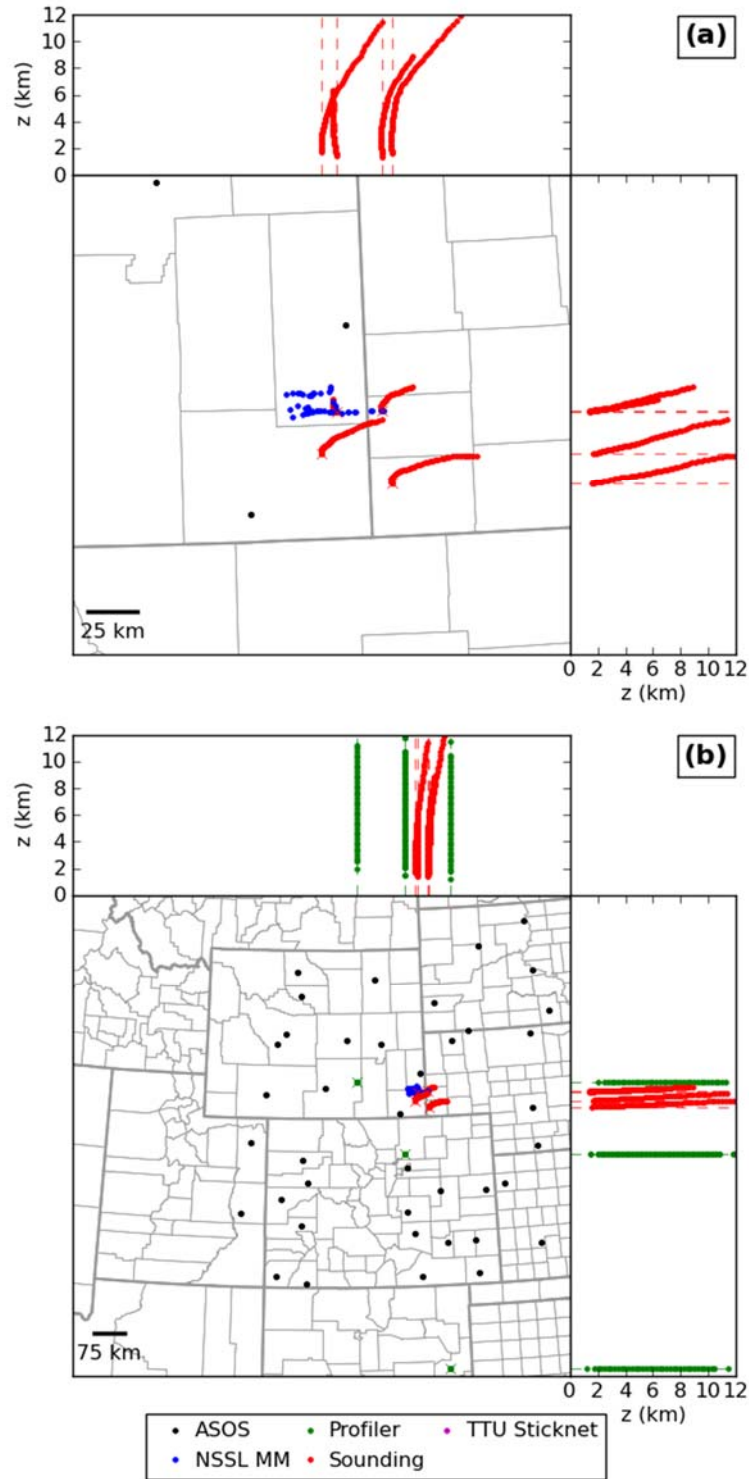


Figure 2: Locations of all conventional observations assimilated between 2105 and 2200 UTC on the inner domain (a) and between 1830 and 2200 UTC on the outer domain (b). See Fig. 1 for more information about the domains. The dashed lines on the x - z and y - z plots indicate the assumed horizontal location for sounding (red) and profiler (green) observations.

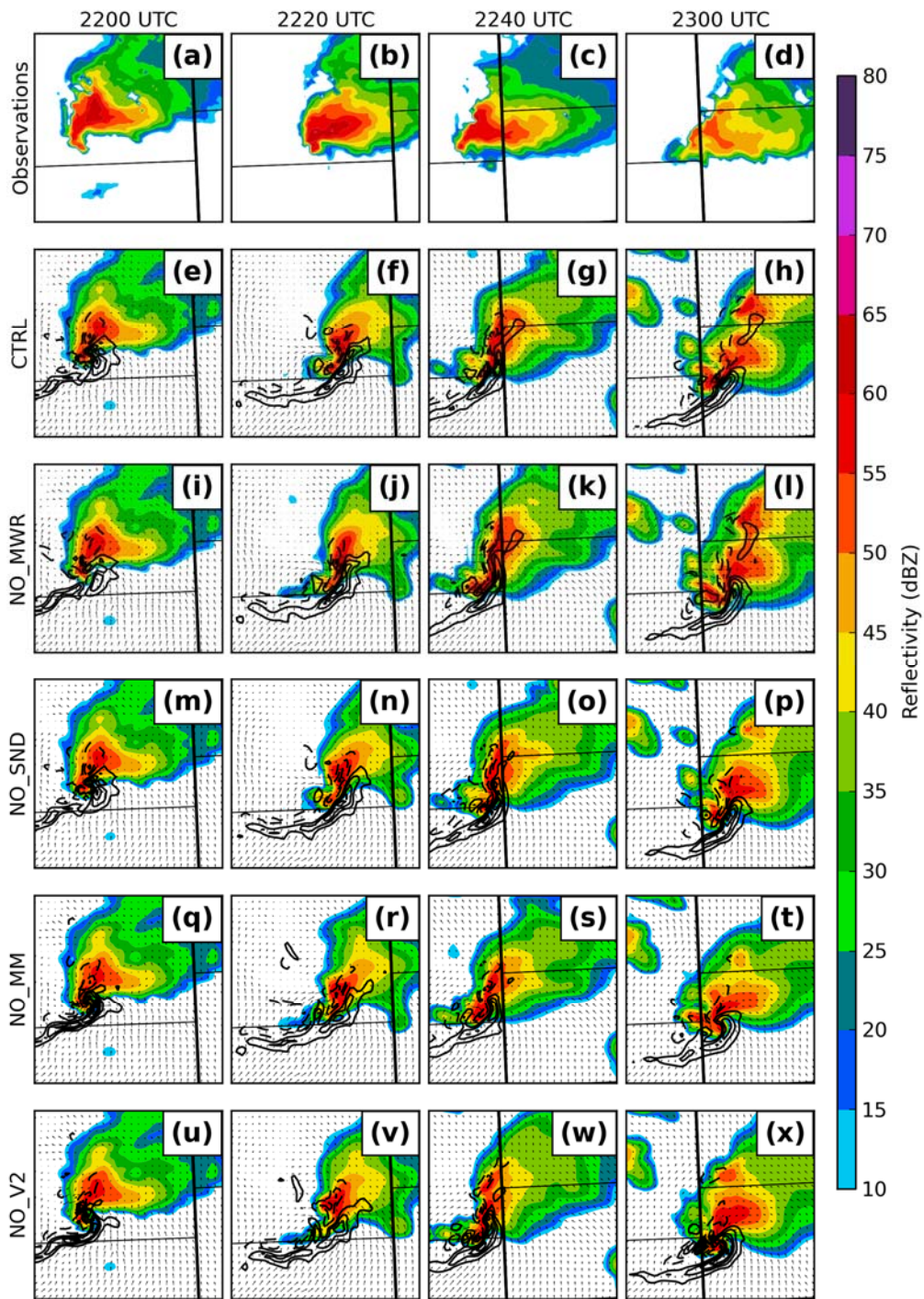


Figure 3: Simulated base reflectivity from KCYS and 1 km AGL vertical velocity and horizontal wind vectors from the ensemble member closest to the mean. Vertical velocity is contoured every 2 m s^{-1} , with negative values dashed, and simulated reflectivity is color filled. Additionally, observed base reflectivity from KCYS is shown in color fill on the top row.

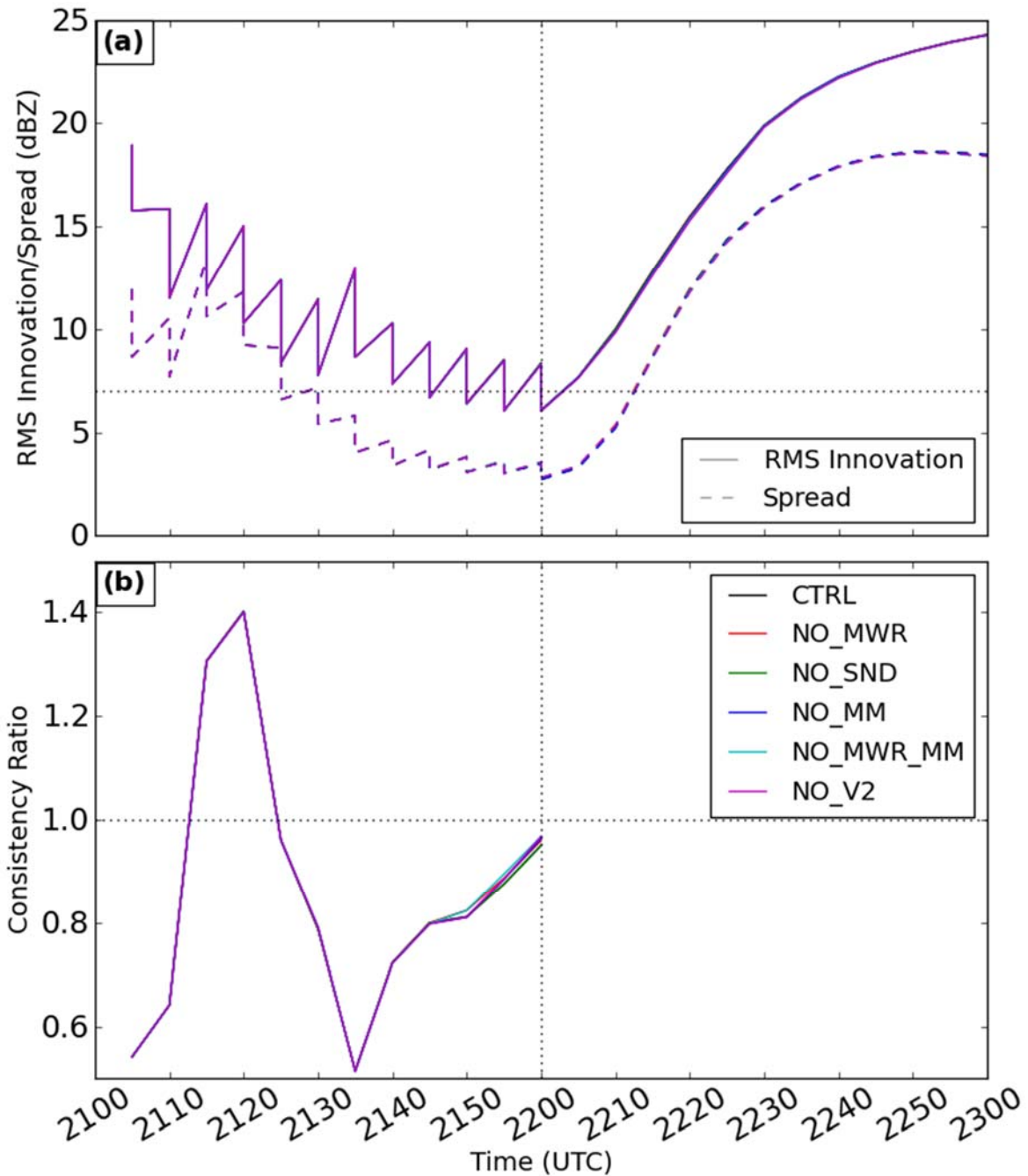


Figure 4: The RMS innovation and domain-averaged spread (a) and consistency ratio (b) for reflectivity from the KCYS radar. In both (a) and (b), the vertical dotted line denotes the time of the last analysis. In (a), the horizontal dotted line denotes the observation error standard deviation (7 dBZ), and in (b), the horizontal dotted line denotes the ideal consistency ratio (1). Statistics are only computed where either observed or model reflectivity is greater than or equal to 15 dBZ.

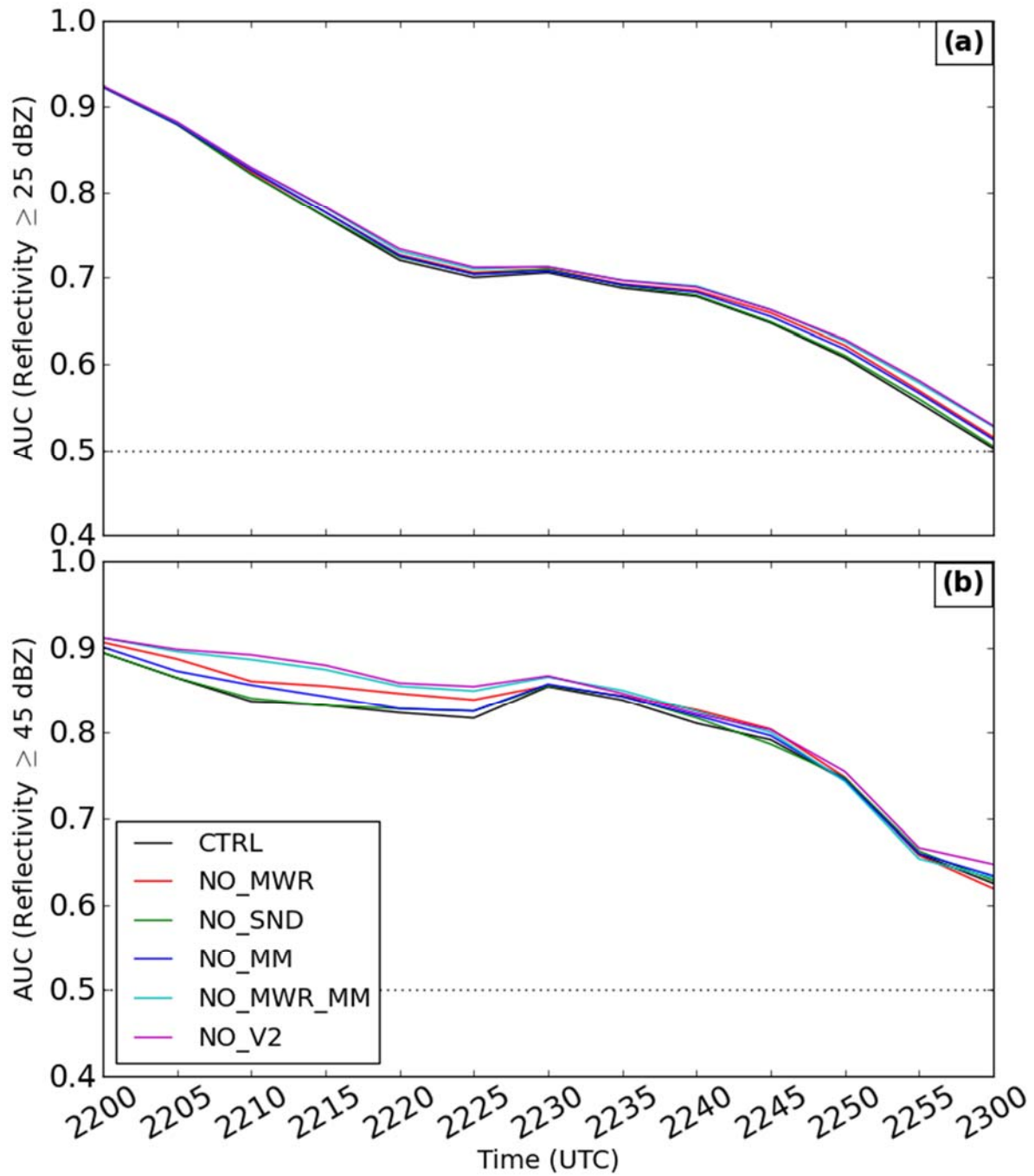


Figure 5: ROC area for KCYS base reflectivity ≥ 25 dBZ (a) and reflectivity ≥ 45 dBZ (b). The dotted black line denotes an AUC of 0.5, at which point the forecast is no better than a random forecast.

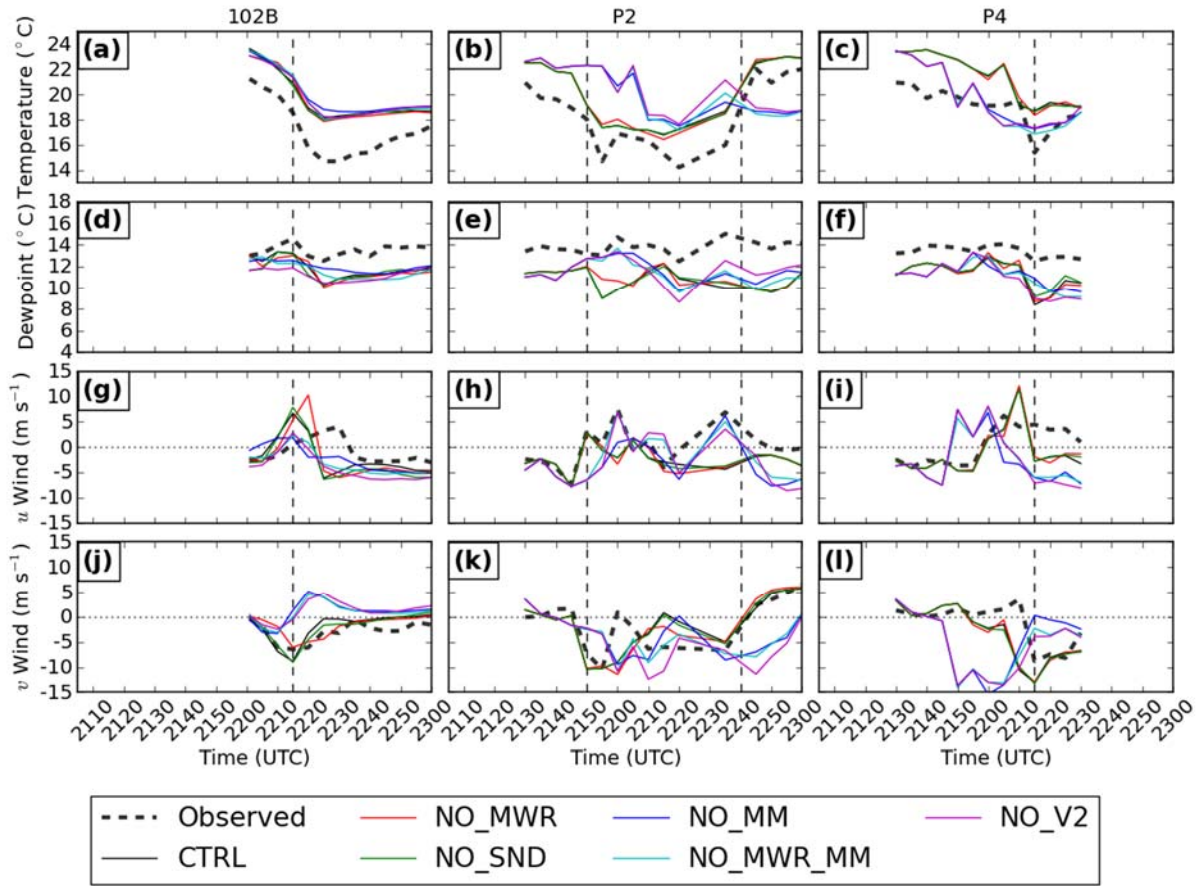


Figure 6: Time series graphs comparing mobile mesonet observations (thick dashed black) and the ensemble mean forecasts from each experiment (solid). Vertical dashed lines represent times of transitions between air masses.

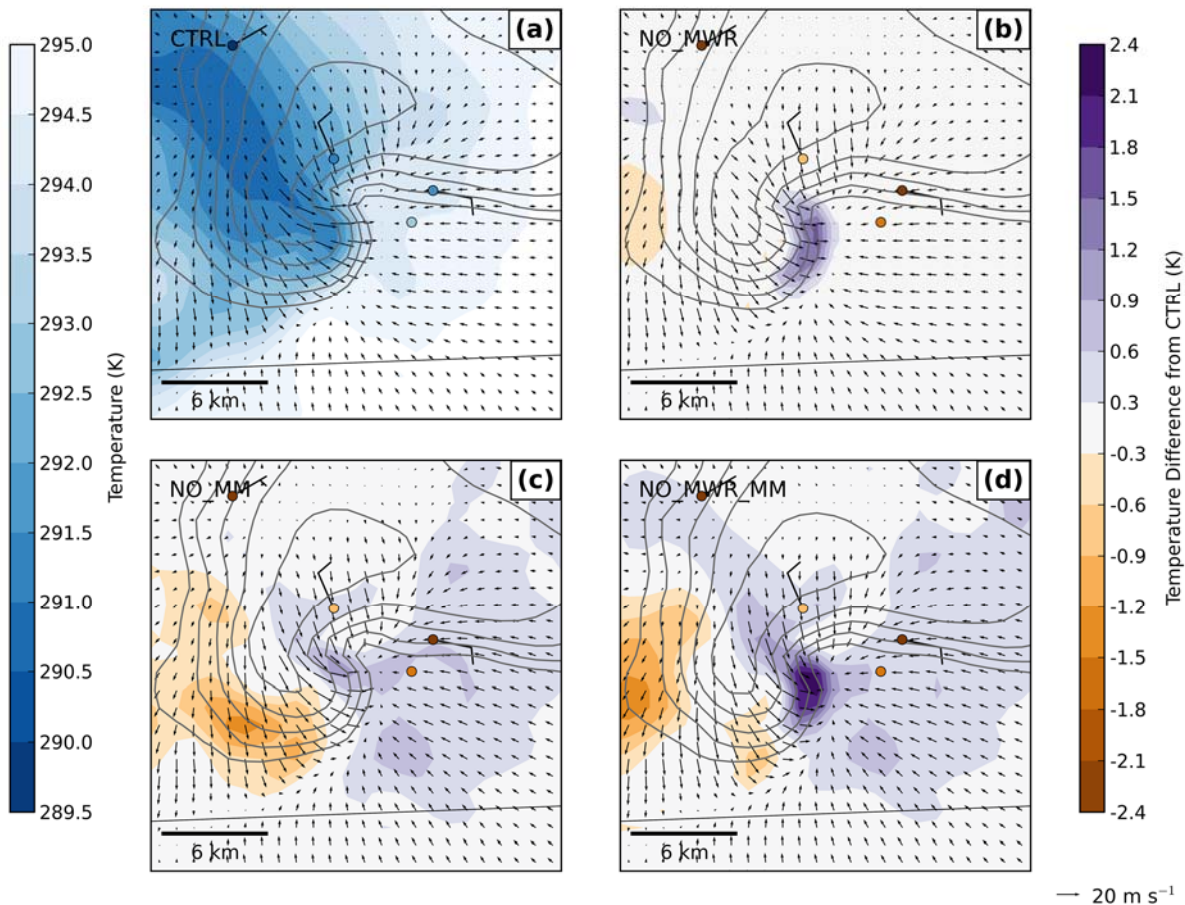


Figure 7: Temperature of the CTRL experiment (a) and differences from the CTRL experiment (b, c, and d) and winds from a 5-minute forecast valid at 2150 UTC at the lowest model level above ground (25 m AGL). Dots in (a) are colored according to observed temperature, and dots in (b), (c) and (d) are colored according to the observed temperature difference from the CTRL forecast. Forecast simulated reflectivity from KCYS is contoured in gray every 10 dBZ. In addition, observed wind barbs (m s^{-1}) are plotted.

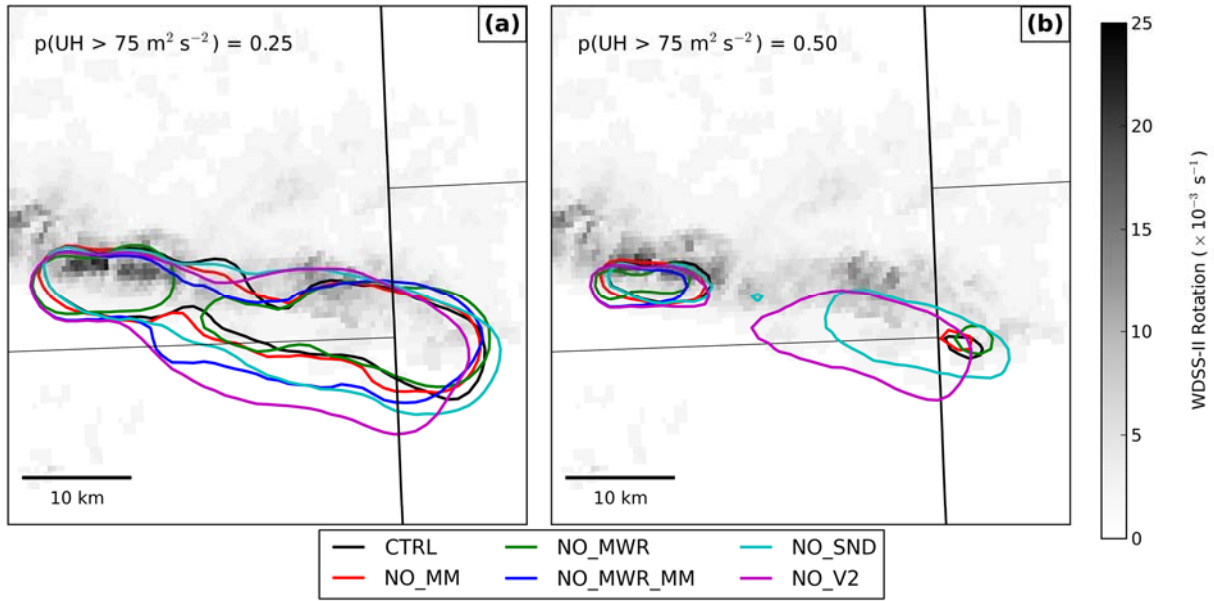


Figure 8: Neighborhood ensemble probability of updraft helicity (UH) $\geq 75 \text{ m}^2 \text{ s}^{-2}$ over the duration of the 1-hour forecast. UH is integrated over the 0-3 km AGL layer, and the neighborhood is 2.5 km in radius. The 0.25 probability contour is given in (a) and 0.50 probability contour is given in (b). The Warning Decision Support System - Integrated Information (WDSS-II) rotation swath product is in grayscale.

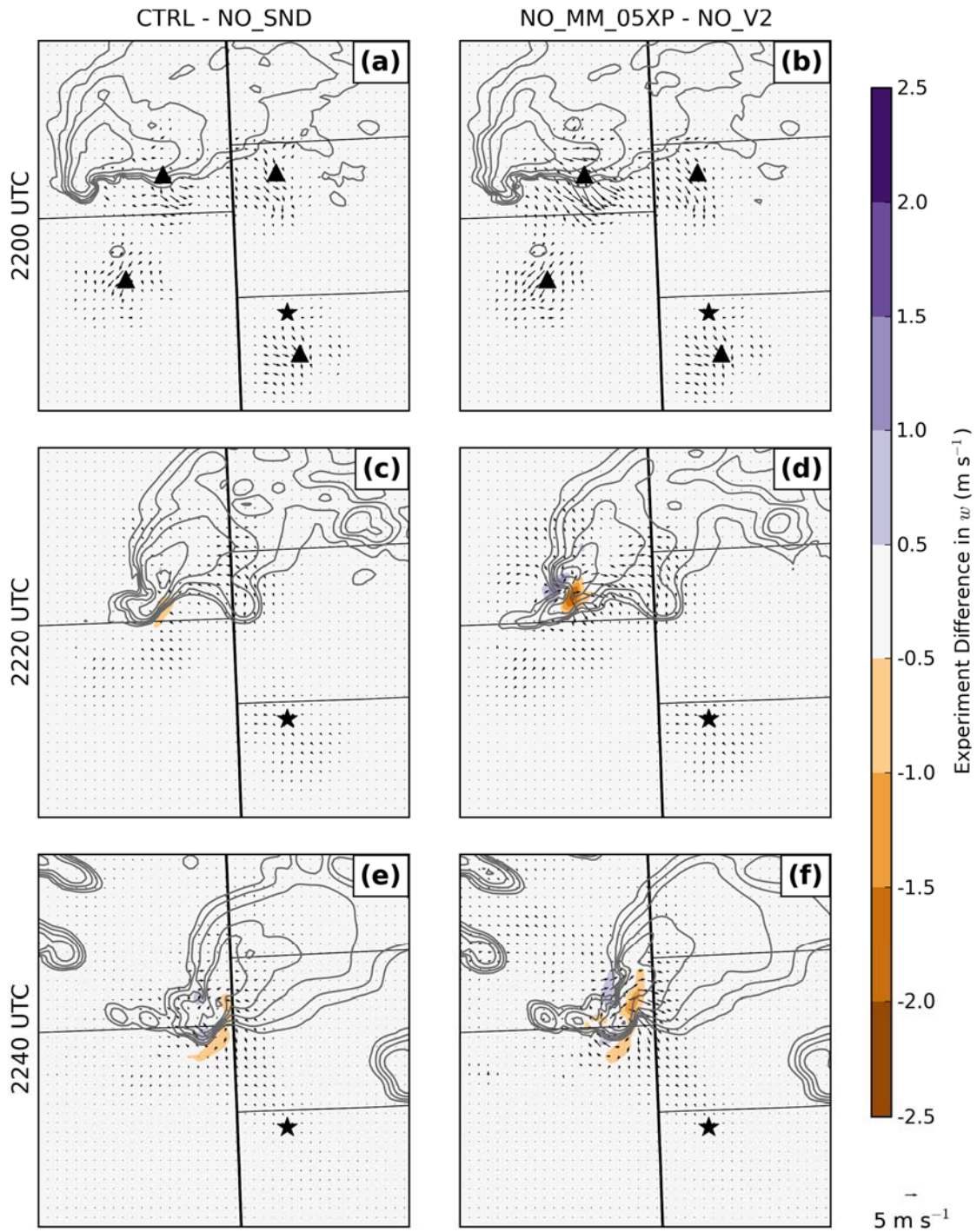


Figure 9: Difference in ensemble mean vertical velocity w (color fills) and horizontal wind (vectors) at 500 m AGL between experiments. Simulated base reflectivity from the ensemble member closest to the mean is contoured in gray in 10 dBZ increments, starting at 10 dBZ. Triangles denote locations of assimilated soundings, and the star denotes the location of the hodograph retrieved from the model.

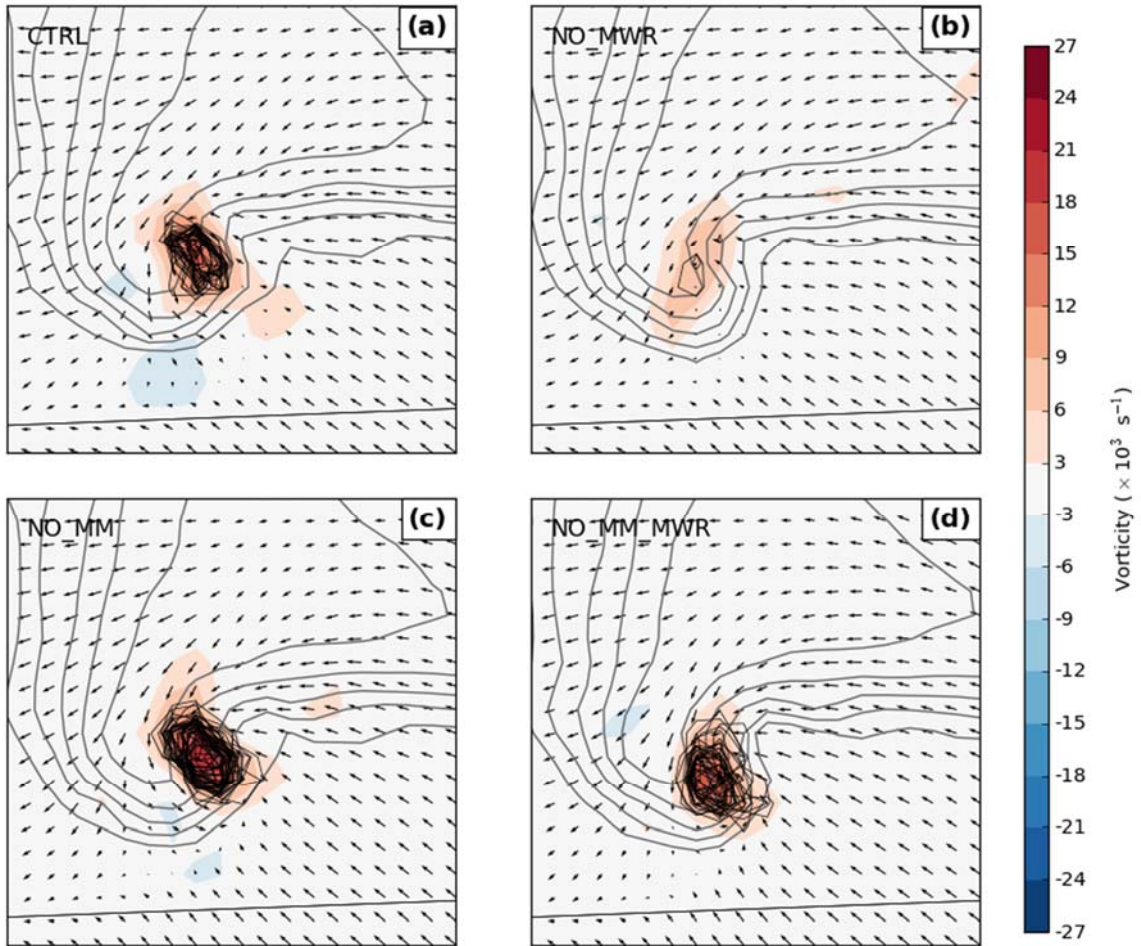


Figure 10: Ensemble mean vertical vorticity (ζ) at 300 m AGL at the final analysis (2200 UTC). Storm-relative wind vectors at 300 m AGL are also plotted, and simulated base reflectivity from KCYS is contoured every 20 dBZ, starting at 20 dBZ, in gray. Individual member $\zeta \geq 0.015 \text{ s}^{-1}$ are contoured in thin black lines.

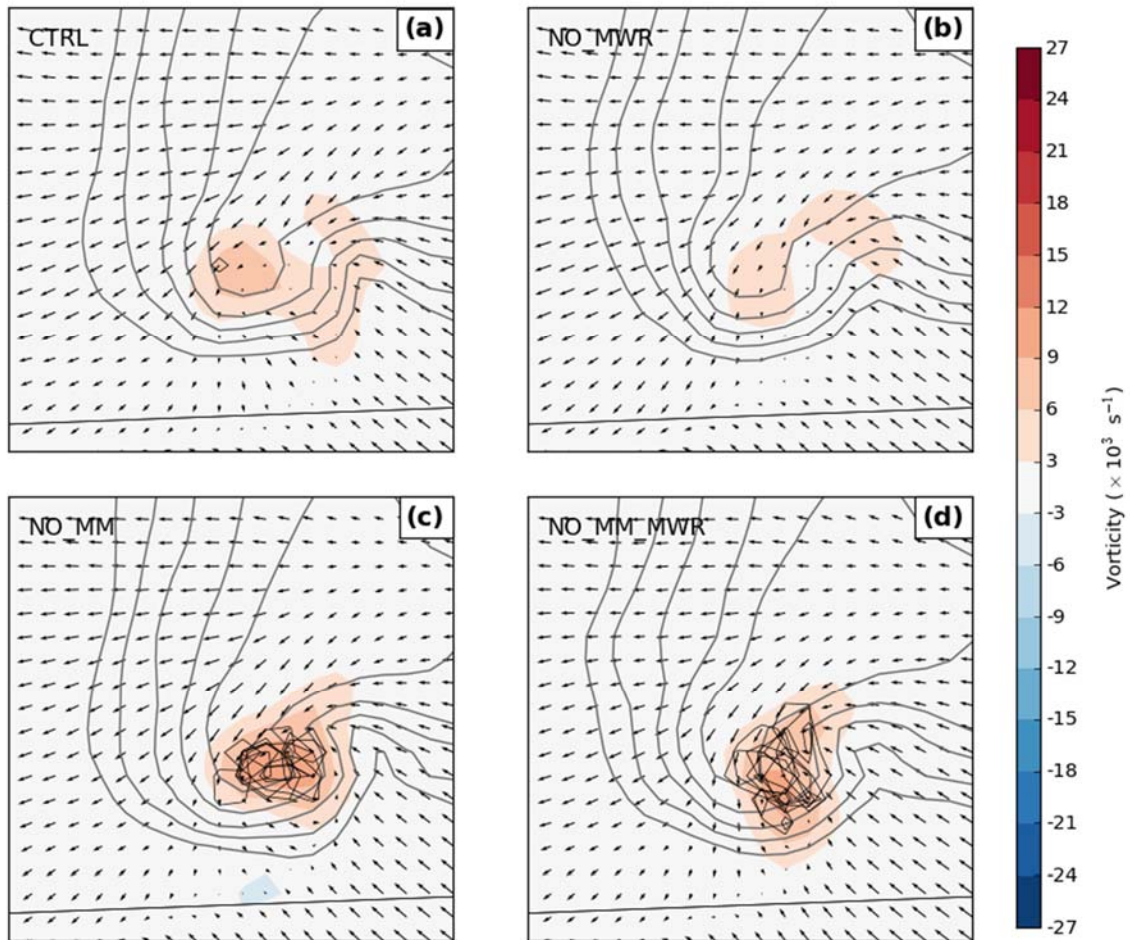


Figure 11: As in Fig. 10, but for a 5-minute forecast valid 2205 UTC.

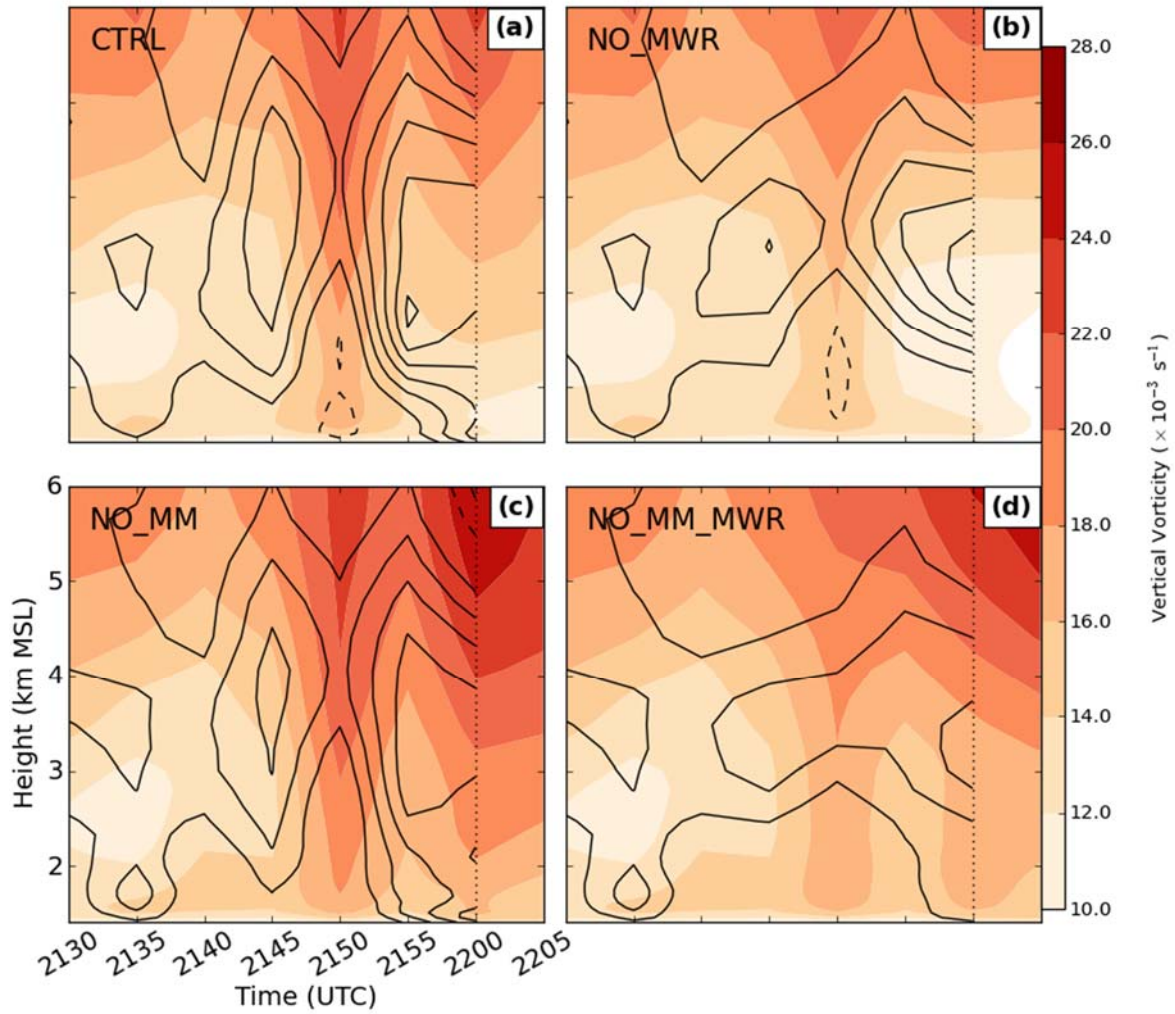


Figure 12: Time-height plot of the ensemble mean of the maximum vertical vorticity (color fills) from the ensemble forecasts during the analysis period and first 5 minutes of the 1-hour forecast. The maximum is taken over the entire domain. The difference between ensemble mean maximum vertical vorticity in the analyses and that in the forecasts is given in contours, every $2 \times 10^{-3} \text{ s}^{-1}$. Solid contours are positive (analysis greater than forecast) and dashed contours are negative (analysis less than forecast). The dotted line represents the time of the final analysis.

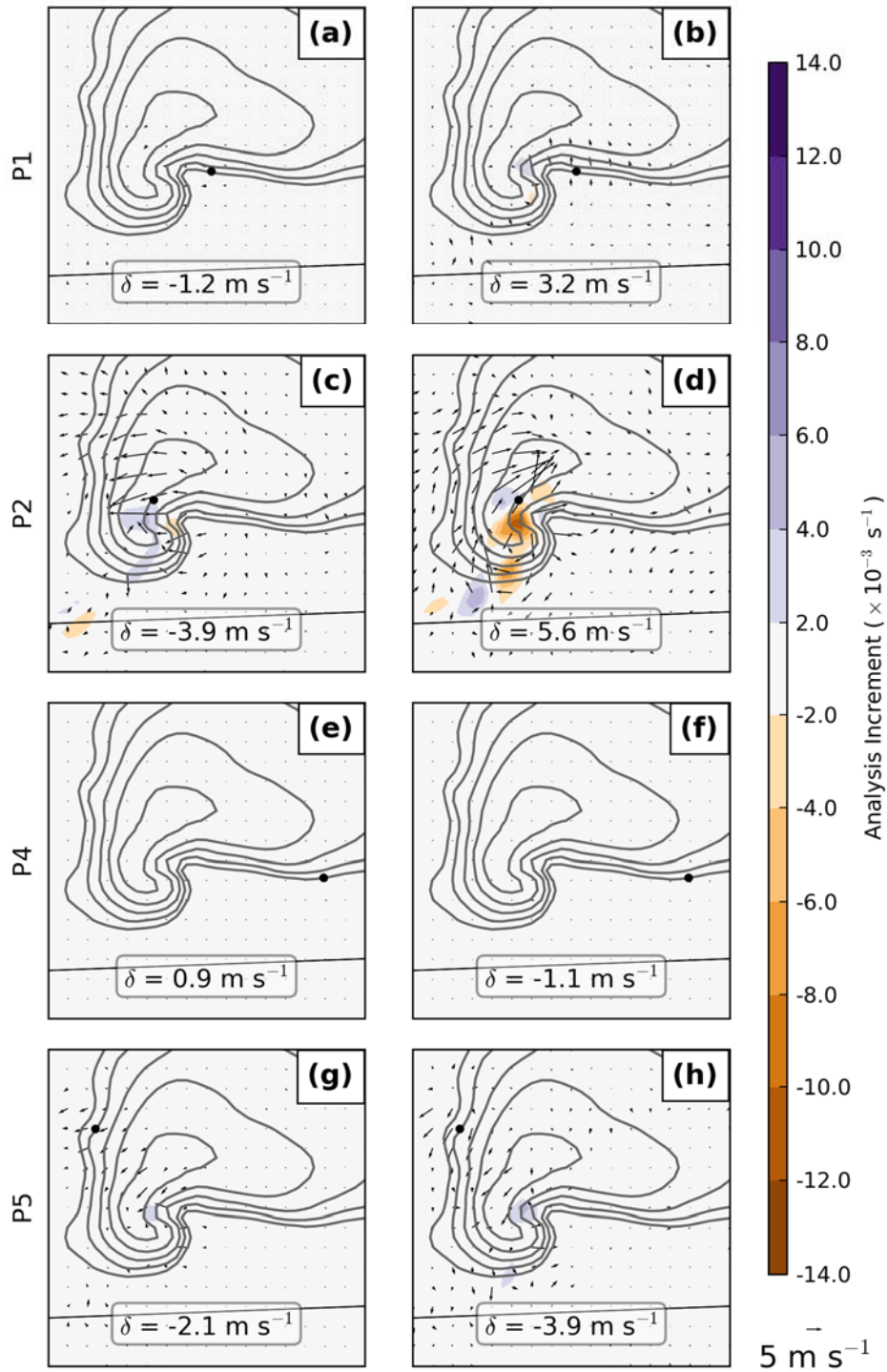


Figure 13: Analysis increments of vorticity from the (color fills) and horizontal wind (vectors) from the CTRL experiment by u (left column) and v (right column) observations from each mobile mesonet site (P1, P2, P4, and P5). Forecast simulated base reflectivity from KCYS is contoured in gray every 10 dBZ. δ represents the ensemble mean deviation from the observation $y^o - H(\bar{x})$.

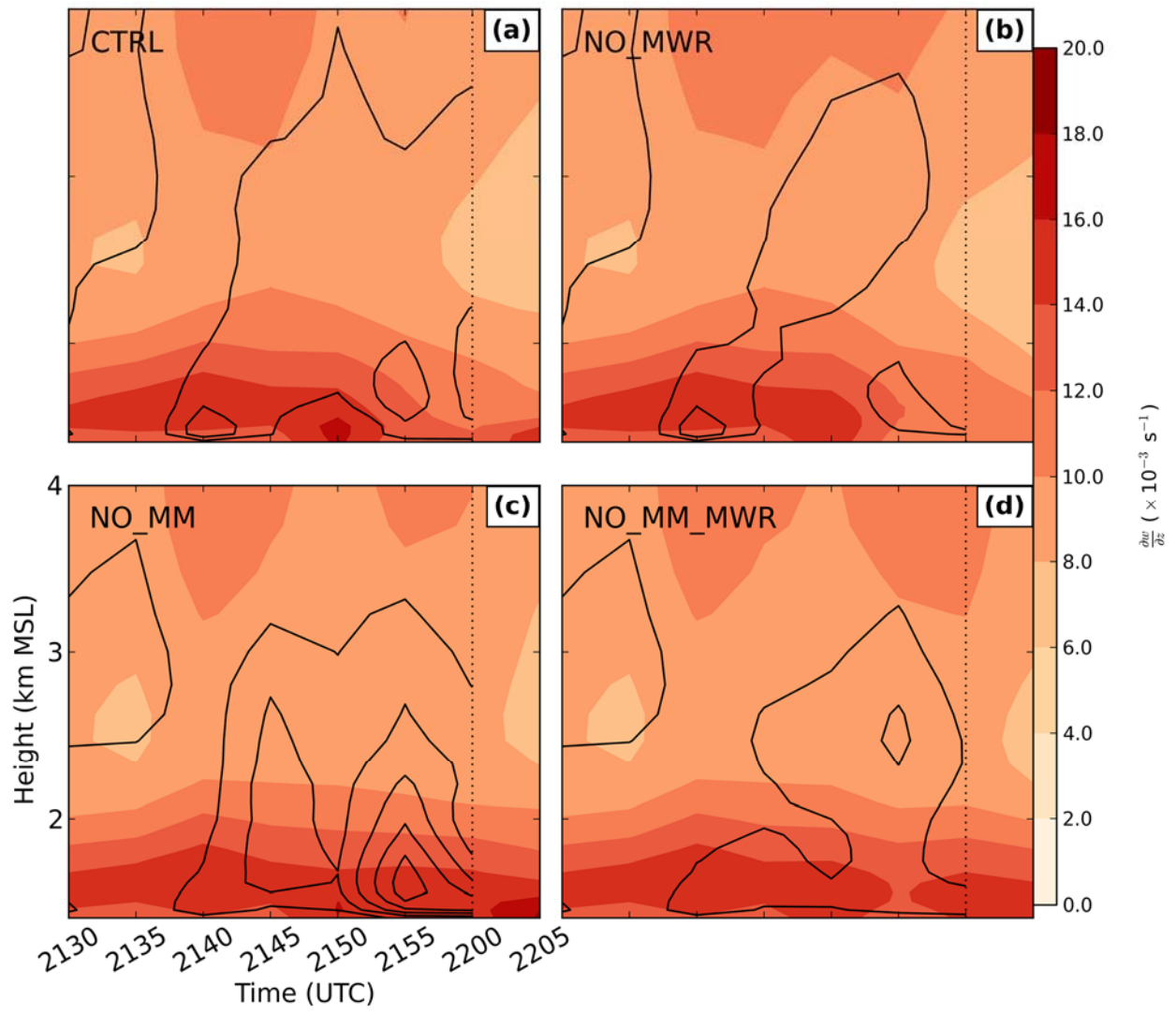


Figure 14: As in Fig. 12, but for $\frac{\partial w}{\partial z}$.