May 2016

The Influence of Assimilated Targeted Observations Upon Ensemble Forecasts of Convection Initiation

Alexandra Marie Keclik
University of Wisconsin-Milwaukee

THE INFLUENCE OF ASSIMILATED TARGETED OBSERVATIONS UPON ENSEMBLE FORECASTS OF CONVECTION INITIATION

by

Alexandra Marie Keclik

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Mathematics

at

The University of Wisconsin-Milwaukee

May 2016

ABSTRACT

THE INFLUENCE OF ASSIMILATED TARGETED OBSERVATIONS
UPON ENSEMBLE FORECASTS OF CONVECTION INITIATION

by

Alexandra Marie Keclik

The University of Wisconsin-Milwaukee, 2016
Under the Supervision of Professor Clark Evans and Professor Paul Roebber

The influence of assimilating targeted meso-α- to synoptic-scale observations collected in the upstream, pre-convective environment upon subsequent short-range ensemble forecasts of convection initiation (CI) across the central United States for the fifteen aircraft missions conducted by the Mesoscale Predictability Experiment (MPEX) in May and June 2013 is evaluated in this study. Utilizing the ensemble Kalman filter implementation within the Data Assimilation Research Testbed software package as coupled to version 3.4.1 of the Advanced Research version of the Weather Research and Forecasting model, two nearly-identical thirty-member ensembles of short-range forecasts are conducted for each mission. Initial conditions for one ensemble are generated through a cycled data assimilation process that incorporates the targeted MPEX dropsonde observations from that day's mission, and initial conditions for the other ensemble are generated through a cycled data assimilation process that excludes the targeted MPEX dropsonde observations. All forecasts for a given mission begin at 1500 UTC, extend forward 15 h, and are conducted on a domain encompassing the conterminous United States with 3 km horizontal grid spacing and 40 vertical levels. Verification is conducted over

spatiotemporal thresholds of 50 km/0.5 h, 100 km/1 h, and 200 km/2 h of an observed CI event to assess the skill of probabilistic forecasts and quantify the influence that assimilating targeted observations has upon forecast skill for the events considered. Forecasts without the targeted observations have high probabilities of detection but also greatly overproduce CI, and the inclusion of targeted observations minimally improves some forecasts and minimally degrades other forecasts. Within the 100 km/1 h threshold, the targeted observations on average reduce distance errors between matched modeled and observed objects by 0.22 km while adding a time bias of 0.24 minutes. The forecast performance of specific cases as well as implications for CI predictability are discussed.

To

my mom

and abuela

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank my advisors Clark Evans and Paul Roebber for their guidance and for sharing their wealth of knowledge. Thank you to Sergey Kravtsov for being on my thesis committee. Many thanks to Bryan Burlingame for his help in all things WRF and WDSS-II related. Thank you to Glen Romine of NCAR and Ryan Torn of the State University of New York at Albany for providing the EnKF-derived initial conditions for the numerical simulations. This work has been supported by the National Science Foundation Grant No. AGS-1347545 and the UWM Advanced Opportunity Fellowship. Computing support has been provided by the NCAR *yellowstone* supercomputer.

1. Introduction

Convection initiation (CI) or the formation of deep, moist convection is a sequence of events in which air parcels accelerate beyond their level of free convection to create a visible cloud top (Kain et al. 2013). CI is triggered by a convergence mechanism; common examples include drylines, elevated convergence zones, frontal boundaries, gust fronts, horizontal convective rolls, orographic circulations, sea breezes, and undular bores (Jorgensen and Weckwerth 2003; Weckwerth and Parsons 2006; Burghardt et al. 2014). Further, CI is a classic scale interaction problem, requiring a favorable interaction between phenomena from multiple scales. The synoptic and meso-α scales establish the thermodynamic and kinematic environment favorable for CI (Weisman et al. 2008). The meso-β scale contributes to horizontal variability in the large-scale environment in which CI occurs, and meso-γ to microscale phenomena determine local planetary boundary layer (PBL) lifting, moistening, and environmental variability crucial to CI timing and location (Markowski et al 2006; Weckwerth et al. 2008; Burghardt et al. 2014).

In a favorable atmospheric environment with sufficiently large vertical wind shear, CI can lead to the development of severe thunderstorms capable of producing damaging surface winds, flash flooding, large hail and tornadoes. Severe storms annually cause substantial loss of life and are responsible for the largest amount of U.S. billion-dollar natural disaster events during 1980–2014, with an average billion-dollar severe storm event cost of approximately $2.2 billion (NOAA 2015). Despite the significant societal impacts that can result, accurately predicting the initiation, intensity, and evolution of deep, moist convection remains a significant challenge for numerical weather prediction models and human forecasters. Contributions to forecast error include the stochastic nature of the atmospheric system, the dependence of CI upon processes on multiple scales, shortcomings in physical parameterization packages employed within

convection-permitting numerical simulations, and data quality and availability (Weckwerth and Parsons 2006).

Characteristics of convective storms are strongly tied to the environment in which they develop, so it is important to accurately represent the initial environment when forecasting such events (Benjamin et al. 2010; Wandishin et al. 2010; Weisman et al. 2015). Several efforts in operational and experimental convection-permitting NWP (CP; horizontal grid spacing of 4 km or less) have been made in recent years to explicitly forecast CI. Duda and Gallus (2013) utilized the Advanced Research Weather Research and Forecasting (ARW-WRF; Skamarock et al. 2008) with 3-km horizontal grid spacing and Kain et al. (2013) utilized the ARW-WRF Model with 4-km horizontal grid spacing, and simulations in both studies showed no discernible timing error in forecasted CI. Large-scale forcing features were well-simulated and small-scale features that influence the timing and location of CI, such as storm-scale outflow boundaries and horizontal convective rolls in the boundary layer, were at least partially resolved. Burghardt et al. (2014) utilized the ARW-WRF model with a 429 m horizontal resolution on 25 cases in the west-central high plains and found a mean model distance error of 48.0 km and mean model early bias of about 3 min for the predictability of CI within a 100 km/1 h spatiotemporal threshold. In addition, an overproduction of modeled CI events caused high false alarm and bias ratios (Burghardt et al. 2014).

A key finding from NCAR's experimental forecasts with the WRF Model is that initial conditions (ICs) often have a larger influence on short-term (0–36 h) convective forecast skill than does model configuration (Weisman et al. 2008; Romine et al. 2013). Targeted observations, or the augmentation of the regular observation network with additional, specifically chosen observations into numerical weather prediction models, is thought to improve

model ICs by providing a better synoptic to sub-synoptic atmospheric representation. Motivated by studies demonstrating the influence of the synoptic- and meso- scales upon deep, moist convection, the Mesoscale Predictability Experiment (MPEX; Weisman et al. 2015) hypothesized that the collection of non-routine synoptic- and meso-α-scale observations in the upstream, pre-convective environment across the Intermountain West and their subsequent assimilation into convection-permitting numerical forecasts significantly improves forecasts of the timing, location, and mode of CI and its subsequent downstream evolution.

Prior to MPEX, observations have mostly been targeted at synoptic-scale systems for the purposes of improving global model 1–3 day forecasts (Majumdar et al. 2011). Targeted dropwindsonde observations from field projects such as the NOAA Synoptic Surveillance program (Aberson 2010) and Dropwindsonde Observations for Typhoon Surveillance near the Taiwan Region (DOTSTAR; Wu et al. 2007, Chou et al. 2011) have been found to be statistically beneficial for forecasting the track of tropical cyclones. Targeted observations have, on average, showed smaller yet still positive impacts upon extratropical, synoptic-scale forecasts. A campaign that found that targeted observations made a mix of improvements and degradations to forecasts is the Atlantic-THORPEX Regional Campaign (A-TReC) in which various forms of targeted observations, i.e. dropwindsondes, rawinsondes, drifting buoys, AMDAR, Doppler Wind Lidar (DWL), and rapid-scan atmospheric motion vectors, were assimilated to evaluate the impact of targeted observations on 1–3 day forecasts of high-impact weather over Europe (Rabier et al. 2008). Fourrié et al. (2006) suggested that the small number of targeted observations and their proximity to conventional observations rather than within areas that were deemed most sensitive to additional observations by various methods (i.e. Ensemble Kalman Filter, Hessian vectors, and singular vectors) and the lack of high-impact weather cases that were

difficult to predict played a role in mostly neutral results (Fourrié et al. 2006; Majumdar et al. 2011).

The International H_2O Project (IHOP_2002) sampled the three-dimensional time-varying moisture field via in situ and remote sensing techniques to better understand convective processes (Weckwerth et al. 2008). Liu and Xue (2008) numerically simulated the IHOP_2002 12–13 June 2002 convection initiation event over the Central U.S. Great Plains using the 3-km horizontal resolution Advanced Regional Prediction System (ARPS). Sensitivity experiments assess how data assimilation intervals and additional, nonstandard observations influence the prediction of CI. The simulation that assimilates the most data produces the best forecast as the additional observations removes the resolution-related delay of CI and overly moist initial conditions at the low-levels (Liu and Xue 2008). However, other experiments excluding the nonstandard observations did better in the timing and location of the initiation of some cell groups.

Observational tools for MPEX included the NCAR G-V aircraft, which featured the Airborne Vertical Atmospheric Profiling System mini-dropsonde system and the JPL Microwave Temperature Profiling (MTP) system, as well as several ground-based mobile upsonde systems. Operations involved two missions per active program day: (1) an early morning mission with the NCAR G-V, well upstream of anticipated convective storms, and (2) an afternoon and early evening mission with the mobile sounding units, to sample the environment surrounding CI and the influence that convection has on its local and larger-scale environment (Weisman et al. 2015). This study utilizes the MPEX dropsonde observations that were collected across the Intermountain West of the United States (Fig. 1) during the course of fifteen research flights (RFs) between 15 May and 15 June 2013. An average of 28 dropsondes were released during a

RF with a maximum of 33 dropsondes and a minimum of 17 dropsondes during an RF. By using ensembles of initial and lateral boundary conditions, one with and one without the MPEX dropsonde observations, we test the hypothesis that the assimilation of targeted meso-α to synoptic-scale observations are sufficient to increase 0–18 hour CI forecast skill. Section 2 describes the methodology, including ensemble analysis and simulation configuration, the CI identification process, and forecast verification methods. Results from this work will be presented in section 3, and a summary and possible future work will be discussed in section 4.

2. Methodology

a. *Experimental design*

Two parallel cycles of Ensemble Kalman filter-initialized (EnKF), convection-permitting real-data numerical simulations, each containing thirty ensemble members, are identically configured and utilized to generate ensemble estimates of the initial, pre-convective atmospheric state for each of the fifteen RFs. The first, termed the Control ensemble, does not incorporate any MPEX observations. The second, termed the Perturbed ensemble, incorporates the MPEX dropsonde observations for that particular RF.

The ensembles are initialized using analyses generated using the ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003) from the Data Assimilation Research Testbed (DART; Anderson et al. 2009). Cycling of the data assimilation system is conducted using version 3.4.1 of the WRF-ARW (Skamarock et al. 2008) numerical forecast model. The analysis domain upon which all data assimilation is conducted has a horizontal grid spacing of 15 km (415 x 325 grid points) and covers the conterminous United States, the Gulf of Mexico and portions of Canada, Mexico, the eastern North Pacific Ocean, and the western North Atlantic Ocean. The domain contains 40 terrain-following vertical levels between the surface and 50 hPa with approximately eight levels being in the planetary boundary layer (PBL). The Mellor-

Yamada-Janjic (MYJ; Janjic 1994) PBL, Thompson et al. (2008) hybrid double-moment microphysical, RRTM longwave radiation (Mlawer et al. 1997), Goddard shortwave radiation (Chou and Suarez 1994), NOAH (Chen and Dudhia 2001) land-surface process parameterization, and Tiedtke (1989) cumulus convection schemes are utilized by the cycled analysis system (Table 1).

The initial 15 km ensemble is produced by adding Gaussian random samples with zero mean and covariance provided by the WRFDA package (Barker et al. 2012) to the 1800 UTC 30 April 2013 Global Forecast System (GFS) analysis interpolated to the 15-km WRF model domain (Schwartz et al. 2015). The lateral boundary conditions are generated utilizing the fixed covariance perturbation technique of Torn et al. (2006). Sampling errors are minimized by sampling error correction as well as horizontal and vertical observation localization applied via a Gaspari and Cohn (1999) weighting function, and ensemble spread is preserved with adaptive inflation (Anderson 2009).

Analysis fields are updated by WRF-DART included horizontal and vertical velocity, perturbation potential temperature, geopotential height, water vapor, diabatic heating rate, and the mixing ratios and number concentrations for all carried microphysical species. Routine observations assimilated by WRF-DART included mandatory-level rawinsonde data; surface-based METAR, buoy, and ship observations; Aircraft Meteorological Data Relay (AMDAR) reports; satellite-derived atmospheric horizontal motion vectors (Velden et al. 2005); and Global Positioning System (GPS)-derived radio occultation (Kursinski et al. 1997) observations (e.g., as in Romine et al. 2013; Schwartz et al. 2015; Torn and Romine 2015). AMDAR observations are averaged over boxes with dimensions of 30 km in the horizontal and 25 hPa in the vertical as in Torn (2010) and AMVs are averaged over 60 km in the horizontal and are excluded over land.

Both surface observations with model terrain and station height differing by more than 300 m or within three grid lengths of the lateral boundary conditions are excluded. In addition, radio-occultation profiles are thinned to 15 levels and observations within three grid lengths of the lateral boundary conditions are excluded. The characteristics of all assimilated observations are compared to the pre-assimilation atmospheric state space provided by the ensemble, and observations whose squared difference from the ensemble mean estimate exceeded three times the sum of the prior ensemble and observational error variances are rejected by the assimilation system (Romine et al. 2013). The assumed errors for each observation type are the same as in Romine et al. (2013) and are listed in Table 3.

Data is continuously cycled every 6 hours through 0000 UTC 15 June. However, hourly data assimilation begins at 0000 UTC the day of a given RF and continues for 15 hours, and the ensemble for a RF does not incorporate the previous RF's hourly cyclic analysis. At 0900 UTC the day of an RF, the targeted observations are assimilated into one of the ensembles. The ensemble without the targeted observations is termed the Control ensemble and the ensemble with the targeted observations is termed the Perturbed ensemble. Data assimilation experiments for an RF do not include MPEX observations from a previous RF. The MPEX data are quality-controlled by the staff of NCAR's Earth Observing Laboratory prior to assimilation.

Numerical simulations initialized from the analyses provided by the cycled EnKF-based data assimilation system are conducted upon a two-way-interacting 15 km/3 km nested domain. The outer 15 km domain is identical to the WRF-DART analysis domain. The inner 3 km domain (1046 x 871 grid points) extends from the Intermountain West of the United States to the Appalachian Mountains and from Baja California to the Canadian border (Fig. 2). The model configuration is identical to that utilized by the cycled EnKF data assimilation system with the

exception that cumulus convection is treated explicitly on the 3 km domain. The WRF simulations for the Control and Perturbed ensembles are initialized at 1500 UTC on the day of an MPEX RF and extend forward 15 h.

b. CI event identification

Following Burghardt et al. (2014), Kain et al. (2013), and references therein, CI objects are defined as radar reflectivity ≥ 35 dBZ at the –10°C isotherm for a minimum of half an hour. Radar reflectivity at the –10°C isotherm is selected to prevent brightbanding effects and the half hour time minimum is selected to prevent the inclusion of CI false alarms. Observed radar reflectivity from the Level-II Next Generation Weather Radar System (NEXRAD) is obtained for 42 radars extending eastward from the Continental Divide to the Mississippi River for the duration of each RF simulation (Fig. 3). The Warning Decision Support System – Integrated Information (WDSS-II; Lakshmanan et al. 2007) spatial analysis tools (Lakshamanan 2012) are used to generate a uniform 0.03° latitude × 0.03° longitude gridded domain and interpolate the NEXRAD data to the –10°C isotherm extracted from 13 km Rapid Update Cycle (RUC; Benjamin et al. 2004) 0-h analyses. The WDSS-II w2merger is used to merge all the data from the 42 radar locations (Lakshmanan et al. 2006; Lakshmanan and Humphrey 2013), and the w2segmotionll tool from WDSS-II is used to identify individual storm cells of a minimum of 4 conjoined grid points (Lakshmanan et al. 2009) and track the cells forward in time (Lakshmanan and Smith 2010). Motion estimates and advection qualities are extracted and attributed to each CI object using w2segmotionll (Lakshmanan and Smith 2010; Lakshmanan et al. 2013). Forecast reflectivity at the –10°C level of each ensemble member is computed in-line by the WRF model and extracted every 5 min across the domain from each simulation and bilinearly interpolated to

the same grid as the observed reflectivity at –10°C RUC data. The resulting data are then provided to WDSS-II, which then identify and track CI events from the reflectivity field.

c. *Verification metrics*

Each ensemble member forecast is verified deterministically with ensemble performance evaluated over the range of ensemble members. Sets of modeled and observed CI objects within the area of NEXRAD radar coverage are listed and compared for each RF forecast. Following Burghardt et al. (2014), a flow-dependent error metric is utilized to quantify the proximity between the modeled and observed objects in both space and time dimensions. The error metric takes the form:

$$C^2 = (\text{Error}_d)^2 + (\text{Velocity}_c \times \text{Error}_t)^2. \quad (1)$$

where C is the spatiotemporal error (km), Error_d is the spatial difference between the forecasted and observed CI (km) and Error_t is the temporal error (h). Velocity_c is the speed of the observed CI object (km h^{-1}) calculated using the time taken and distance traveled from when the object first reached 35 dBZ until it no longer met the CI criteria. The Velocity_c used herein differs from that of Burghardt et al. (2014) which considers a velocity based off of the layer-mean wind field rather than object motion. The flow dependence allows for consistency in dimensional units of spatial and temporal errors and characterizes timing errors associated with differences in storm motion (Burghardt et al. 2014). Thereafter, spatiotemporal thresholds of maximum Error_d and maximum Error_t 50 km/0.5 h, 100 km/1 h, and 200 km/2 h are applied. There is a focus of a maximum Error_d being 100 km and a maximum Error_t of 1 hour as it represents a balance of space and time errors in the mesoscale and it allows for the comparison to the model performance found within Burghardt et al. 2014.

Verification of CI is performed through the use of a contingency table for dichotomous yes-no forecasts (Wilks 2011; Fowle and Roebber 2003; Kain et al. 2013; Burghardt et al. 2014; Table 2). In this study, true positives (a), false positives (b), and false negatives (c) are identified and used to compute four quality measures (Roebber 2009). True negatives (d) are not evaluated due to the object-based verification method selected. The probability of detection (POD; Eq. 2) is the ratio of correctly forecasted CI objects to the total number of observed CI objects and ranges from 0 to 1. The false alarm ratio (FAR; Eq. 3) is the ratio of unobserved forecast CI objects to the total number of forecast CI objects and also ranges from 0 to 1. The bias (Eq. 4) is the ratio of total forecast events to total observed objects. The Critical Success Index (CSI; Eq. 5), sometimes referred to as the Threat Score, is the ratio of correctly forecast objects to the total number of observed and forecast CI objects with a perfect score being 1.

$$\mathrm{POD} = \frac{a}{b + c} \tag{2}$$

$$\mathrm{FAR} = \frac{b}{a + b} \tag{3}$$

$$\mathrm{bias} = \frac{a + b}{a + c} \tag{4}$$

$$\mathrm{CSI} = \frac{a}{a + b + c} \tag{5}$$

3. Results

The control and perturbed ensembles are compared to each other and to observations to assess the impact of the MPEX targeted observations upon CI forecasts. The verification of CI occurrences is performed through the use of the contingency table. In general, false alarms and misses decrease and hits increase with increasing spatiotemporal thresholds, but there still remains a model overproduction of CI even at the most lenient spatiotemporal threshold, 200

km/2 h. Some cases show minimal improvement and others show nominal degradation of CI forecast skill with the addition of the MPEX observations.

Within a 100 km/1 h threshold, the PODs averaged over all 15 RFs for the Control and Perturbed ensembles are 0.566 and 0.575, and the FARs are 0.695 and 0.691, respectively. There is also slight improvement regarding the mean C error as the Control has a value of 51.30 km and the Perturbed C error is 51.0 km, but the C errors remain slightly larger than the matches found from the matches from the deterministic forecasts of Burghardt et al. (2014) which had a mean C error of 48.0 km within the 100 km/1 h threshold. The average distance error for the Control is 45.68 km and the average distance error for the Perturbed is 45.46 km. The average time bias for the Control is 0.87 minutes (positive denotes later than observed) with a mean absolute time error of 26.66 minutes, while the average time bias for the Perturbed is 1.11 minutes with a mean absolute time error of 26.55 minutes. The mean CSI scores for the Control (0.243) and the Perturbed (0.249) show better performance than the CSI score of 0.191 found by the deterministic forecasts of Burghardt et al. (2014). The maximum improvement of average CSI with the inclusion of MPEX dropsonde observations is 0.028 for RF12 and the most CSI degradation due to the targeted observations occurs for RF6 with 0.016 (Fig. 8; Table 4).

When making the spatiotemporal thresholds stricter to 50 km/0.5 h, the decrease in matches causes mean PODs to decrease to 0.247 for the Control and 0.254 for the Perturbed. The mean false alarm ratio increases to 0.867 for the Control and 0.864 for the Perturbed ensembles due to the increase in false alarms. CSI scores also degrade to 0.093 and 0.096 for the Control and Perturbed ensembles, respectively, but are still slightly more skillful than the CSI of 0.088 at 50 km/0.5 h found by Burghardt et al. (2014). However, the lower number of matches contains lower space and time errors and lower C errors. Mean C error values are 31.44 km for the

Control ensembles and 31.56 km for the Perturbed ensembles which are still slightly larger than the 30.2 km C error of Burghardt et al. (2014).

Relaxing the spatiotemporal-matching thresholds to 200 km/2 h improves CI occurrence skill scores. The average PODs increase to 0.814 (Control) and 0.819 (Perturbed) while mean FARs decrease to 0.557 (Control) and 0.554 (Perturbed). Thus, CSI scores show improvement, increasing to 0.397 and 0.401 for the Control and Perturbed ensembles, respectively. Despite the less strict threshold yielding more matches between observed and modeled CI objects, the matches are more distant in time and space, thus increasing average object C errors for CI cases. Mean C errors become 73.78 km (Control) and 72.9 km (Perturbed). Next, we examine the forecast performances of individual cases, namely RF6 and RF12.

a. RF6: May 23

The focus of RF6 was to sample the upstream environment of a surface front in northwest Texas and weak upper tropospheric potential vorticity (PV) features in New Mexico and Arizona that contributed to severe convection in west central Texas. The G-V flight sampled the environment in northwestern Texas, New Mexico, Arizona, and central Colorado (Fig. 6a). There was a severe, mesoscale convective system (MCS) over Oklahoma during the flight but convection did not interrupt the flight (UCAR 2013). The dropsonde soundings revealed a cyclonic wind shift over New Mexico in the 300-400 hPa layer around 09 UTC, and a lower tropopause and cooler, upper troposphere in accordance with this cyclonic anomaly. Severe convection occurred along a weak surface front in the Texas Panhandle by 19 UTC between Lubbock and Childress, TX (Fig. 10) which later merged with convection along the dry line in southeast New Mexico, forming an MCS over west Texas. In addition to the lifting mechanisms, low-level southerly flow from the Gulf of Mexico allowed for sufficient moisture with 2 m dew

point temperatures around 17–21°C over much of west central Texas. There was steep mid-level lapse rates as well as CAPE values over 3000 J/kg, making the environment favorable for CI.

The POD and CSI for the RF6 forecasts of convection initiation have the worst overall performance among all the cases for the 50 km/0.5 h and 100 km/1 h spatiotemporal thresholds. The 100 km/1 h spatiotemporal threshold verification skill scores for RF6 also show the most degradation of forecast skill as the CSI decrease from 0.145 to 0.129 with the inclusion of MPEX observations. The Control members have a mean POD of 0.375 with a FAR of 0.810 whereas the Perturbed members have a mean POD of 0.326 and FAR of 0.824 (Fig. 7a).

When the spatiotemporal threshold increases to 200 km/2 h, there is still a negative impact on forecast skill caused by targeted observations as the CSI increases to 0.364 (Control) and 0.306 (Perturbed). There are more matches at this threshold, but the mean distance at which the hits occur increases from 51.2 km (Control) and 53.25 km (Perturbed) at the 100 km/1 h threshold to 82.67 km (Control) and 86.77 km (Perturbed) at the 200 km/2 h threshold. This increase in threshold also results in an increase in time bias from 0.3 minute to 1.8 minutes in the Control and from –0.34 minute to 2.49 minutes for the Perturbed. The mean time error of 29.37 minutes for the Control increases to 55.14 minutes, and the mean time error of 30.49 minutes for the Perturbed increases to 56.54 minutes with the larger spatiotemporal threshold. Due to these increases in spatiotemporal error, the C error also increases from 54.6 km (Control) and 56.57 km (Perturbed) to 90.69 km (Control) and 94.75 km (Perturbed).

Both the Perturbed and Control ensembles show evidence of gravity waves beginning at 1500 UTC 23 May but by 1800 UTC, the models overcome the implied imbalance in the initial conditions. The models immediately overproduce CI as they show convection over western Texas as well as northeastern Mexico when convection does not occur there until after 1900

UTC (Fig. 10). The simulations show the ongoing convection in northwestern Oklahoma, albeit slightly displaced and intensified with a more widespread area of greater than 35 dBZ reflectivity. At 1900 UTC, convection in the Texas panhandle initiates along the stationary surface front enhanced by the outflow boundary caused by the convection in Oklahoma and the members begin to show CI except that this occurrence is displaced to the north. More members within both ensembles begin predicting convection in the Texas panhandle as observed convection continues to the south. Members also show more widespread convection over the western Texas panhandle. At 2100 UTC, more Control ensemble members predict convection in the area of observed CI, over western Texas, whereas the Perturbed ensemble keeps the bulk of the reflectivity to the southwest and northeast (Fig. 10). In addition, both ensembles show more widespread convection over eastern Texas.

The Perturbed ensemble produces lower heights over western Texas, New Mexico and Mexico at 300 hPa indicating an amplification of the upper level trough, and there are also higher Perturbed heights at lower levels prior to and during convective initiation over the Texas panhandle (Fig. 11). In addition, there is slightly lower SLP and warmer 2 m temperatures indicating a minimal eastward shift in the area of convergence within the Perturbed ensemble. Perturbed dewpoint temperatures along the New Mexico and Texas borders show variation from the Control on the order of +/–4°C with the Perturbed dew points being higher in central and northern New Mexico as well as northern Mexico and lower in southeastern New Mexico. This shift in dew point temperatures indicate a shift in the dryline which is a possible cause for the Perturbed creating more convectively active events to the east of where the observed convection occurs in western central Texas.

b. RF12: June 8

A trough embedded within strong, upper-tropospheric northwesterly flow over Wyoming and Colorado contributes to widespread convection from the Oklahoma Panhandle northeastward into Iowa during RF12. The GV flight sampled the environment in Northern Colorado, Eastern Nebraska, southwestern South Dakota, Wyoming, and eastern Idaho (Fig. 6b). The flight winds show a sharp trough over the Northeastern Colorado and a large contrast between the air within and ahead of the trough. The sounding in the trough is cooler throughout the tropospheric layer and has a lower tropopause (UCAR 2013). CI began in north central Kansas around 2000 UTC along a surface cold front with dew points near 21°C and 1500–2000 J/kg CAPE.

In general, RF12 shows modest forecast skill relative to all the other RFs and is the case that has the most improvement of skill with the inclusion of targeted observations at the 100 km/1 h and 200 km/2 h spatiotemporal thresholds. The CSI of the Control ensemble is 0.218 and the CSI of the Perturbed ensemble is 0.246 in the 100 km/1 h spatiotemporal threshold (Fig. 8b, 7b) as the addition of the observations increases hits and reduces false alarms. In the Perturbed ensemble, radar reflectivity of greater than 35 dBZ is reduced over North Dakota and South Dakota where there is ongoing convection of less than 35 dBZ (Fig. 10). Both ensembles are delayed in developing the bulk of the CI which occurs along the cold front in central Kansas beginning around 2000 UTC and have the CI displaced further to the northwest. This is a possible reason for why RF12 has a late Control time bias of 8.46 minutes and a Perturbed time bias of 7.46 minutes. The added observations help to delay the early production of CI over western Texas as the Control has the majority of the members predicting CI in that area by 1900 UTC when it does not begin until about 2100 UTC. Both ensembles indicate the CI along a weak

frontal boundary in southeastern Texas with each member having slightly different locations while the CI in Northern Mexico is overproduced by both the Control and Perturbed.

When the spatiotemporal threshold is increased to 200 km/2 h, CSI increases to 0.376 for the Control and 0.409 for the Perturbed. The POD increases from 0.556 to 0.848 for the Control and 0.600 to 0.881 for the Perturbed, but there is an increase in mean time and distance errors. On average, the time errors become more biased as the ensemble CI objects match with the observed objects 16.2 (Control) and 16.21 (Perturbed) minutes late. The mean time errors increase from 27.26 minutes to 46.8 minutes for the Control and from 27.1 minutes to 45.07 for the Perturbed. Distance error increases from 48.41 to 67.93 (Control) and from 46.95 to 64.7 (Perturbed).

Differences between the Perturbed and Control height fields, primarily over Kansas and Nebraska, increase with decreasing pressure levels which indicates a deepening of the trough with increasing height (Fig. 14). The height differences correspond well to the decrease in surface temperature and minimal increase in mean sea level pressure from North Dakota southward through northern Texas (Fig. 15). In addition, there is a slight decrease in 2 m dew point temperatures from that of the Control in the vicinity of the observed cold front, supplying a possible reason for the decrease in CI overproduction by the Perturbed ensemble.

4. Summary and discussion

Two thirty-member WRF-DART simulations, one incorporating MPEX observations and one not, are run for fifteen cases to test the hypothesis that meso-α- to synoptic-scale targeted observations improve forecasts of convection initiation. All forecasts for a given mission begin at 1500 UTC, extend forward 15 h, and are conducted on a domain encompassing the conterminous United States with 3 km horizontal grid spacing and 40 vertical levels. Convection initiation

events across the central United States are identified from objects in the convectively-active field with observed or simulated radar reflectivity of 35 dBZ or greater at the –10°C level for at least thirty minutes. Verification is conducted over multiple spatiotemporal thresholds: within 50 km/0.5 h, 100 km/1 h, and 200 km/2 h of an observed convection initiation event. Each ensemble member forecast is verified deterministically with ensemble performance evaluated over the range of ensemble members.

The inclusion of the MPEX targeted observations slightly improves some forecasts and marginally degrades some forecasts, but on average, there is minor improvement in the prediction of CI. Within the 100 km/1 h threshold, the targeted observations on average reduce distance errors between matched modeled and observed objects by 0.22 km while adding a time bias of 0.24 minutes and reducing the mean time error by 0.11 minutes. The average POD within the 100 km/1 h threshold is 0.566 for the Control and 0.575 for the Perturbed. However, due to a large overproduction of CI, FAR is 0.695 for the Control and 0.691 for the Perturbed. The overproduction of CI by convective-allowing, high-resolution simulations also occurs in the 25 deterministic cases considered by Burghardt et al. (2014) as a mean FAR of 0.765 within the 100 km/1 h threshold was found in that study. In addition, mean CSI values are 0.243 (Control) and 0.249 (Perturbed), which are slightly higher than the CSI of 0.191 found by Burghardt et al. (2014). When the spatiotemporal threshold is loosened to 200 km/2 h, the forecasts become more skillful but the distance and time error between the observed and modeled matches increase. The inverse is true for when the spatiotemporal thresholds are made stricter; there are less matches and more false alarms with matches closer in time and space for 50 km/0.5 h.

RF6 is a case in which a surface front and upper tropospheric potential vorticity contribute to severe convection over west central Texas. Utilizing the 50 km/0.5 h and 100 km/1

h spatiotemporal threshold, the simulations for this RF are the worst performing of the 15 RFs and show the most degradation in forecast skill by the addition of MPEX observations as CSI values decrease from 0.145 in the Control to 0.129 in the Perturbed within the 100 km/1 h threshold. The observed CI that occurs in north central is displaced to the north by the simulations, and the ensembles suggest CI over Northern Mexico that does not occur until later in the simulations. Difference fields between the Control and Perturbed ensemble mean states indicate minimal changes in height fields over northern Mexico, New Mexico, and Western Texas that contribute to the slight degradation in forecast skill while the remainder of the area height fields remain constant between the Perturbed and Control ensembles.

RF12 shows modest forecast skill relative to all the other RFs and is the case that has the most improvement of skill with the inclusion of targeted observations at the 100 km/1 h and 200 km/2 h spatiotemporal thresholds. The CSI of the Control is 0.218 and the CSI of the Perturbed is 0.246 when utilizing the 100 km/1 h spatiotemporal threshold. The bulk of CI that occurs within this case occurs over central Kansas and is largely forced by a trough and associated cold front. Difference fields between the Control and Perturbed reveal more influence by the targeted observations within this case than for RF6. Decreases in height fields aloft and cooler conditions at the surface within the areal extent of the upper trough are probable causes for the decrease in overproduction of CI in North Dakota, South Dakota, and Nebraska.

The simulation performances show similarities to previous works on high-resolution, convection-allowing simulations (e.g. Duda and Gallus 2013, Kain et al. 2013, Burghardt et al. 2014), but are only valid for the model configuration and the object-oriented method utilized within this study. The results are limited by initial condition uncertainty and model physics error. Changes in the model configuration and data assimilation methods may possibly provide

different results. Also, the previous study of the 12–13 June IHOP_2002 case has a similar finding in that the inclusion of targeted observations does not always improve the skill of CI prediction. In that case study, it is found that the 3-hourly data assimilation cycle provide better skill in CI prediction than the hourly data assimilation (Liu and Xue 2007). Although the targeted observations utilized for the IHOP_2002 case are from a variety of sources (e.g. radiosondes and mesonet networks) whereas dropsonde data are the only targeted observations utilized within this study and the Advanced Regional Prediction System Data Analysis System rather than DART is utilized by Liu and Xue (2007), perhaps changing the data assimilation cycling rate of the targeted MPEX observations may add value to the forecasts of CI prediction and is subject to future work.

Table 1. WRF options

Parameter	Cycled analysis domain	High-resolution forecast domain
Horizontal grid	415 x 325, Δx = 15 km	1046 x 871, Δx = 3 km
Vertical grid	40 levels, p_{top} = 50 hPa	Same
PBL scheme	Mellor-Yamada-Janjic	Same
Cumulus scheme	Tiedtke	None
Explicit microphysics	Thompson	Same
Radiation (shortwave)	Goddard	Same
Radiation (longwave)	RRTM	Same
Land surface scheme	NOAH	Same

Table 2. 2 × 2 Contingency Table

Contingency Table		**Event Observed**	
		Yes	No
Event Forecast	Yes	a	b
	No	c	d

Table 3. Assimilated observation types and assumed observation errors.

Platform	Variable	Observation Error
Radiosonde	Temperature E–W, N–S winds Dewpoint	NCEP statistics NCEP statistics Lin and Hubbard (2004)
AMDAR	Temperature E–W, N–S winds	NCEP statistics NCEP statistics
METAR	Temperature E–W, N–S winds Altimeter Dewpoint	2 K 1.75 m s^{-1} 1 hPa Lin and Hubbard (2004)
Buoy and ship reports	Temperature E–W, N–S winds Altimeter Dewpoint	2 K 1.75 m s^{-1} 1 hPa Lin and Hubbard (2004)
AMV	E–W, N–S winds	50% NCEP statistics
GPS (thinned to 15 levels)	RO refractivity	Kuo et al. (2004)

Table 4. RF cases and averaged ensemble performance for the 100 km/1 h threshold.

RF	Date (2013)	Dropsondes	Control CSI	Perturbed CSI	Control C error (km)	Perturbed C error (km)
1	15 May	27	0.2605	0.2880	53.88	53.61
2	16 May	30	0.2466	0.2471	50.14	48.63
3	18 May	17	0.2854	0.2827	52.45	52.46
4	19 May	29	0.3259	0.3353	51.54	50.04
5	21 May	27	0.2426	0.2644	53.39	53.69
6	23 May	29	0.1445	0.1290	54.60	56.57
7	27 May	29	0.1686	0.1644	52.59	52.59
8	28 May	21	0.2675	0.2766	52.82	52.19
9	30 May	26	0.2593	0.2622	52.42	52.68
10	31 May	28	0.2991	0.2867	49.18	50.34
11	03 June	32	0.2339	0.2341	46.00	43.68
12	08 June	31	0.2176	0.2459	54.71	53.33
13	11 June	33	0.2556	0.2691	47.86	48.21
14	12 June	33	0.1536	0.1549	52.33	51.79
15	14 June	33	0.2916	0.2934	45.58	45.19

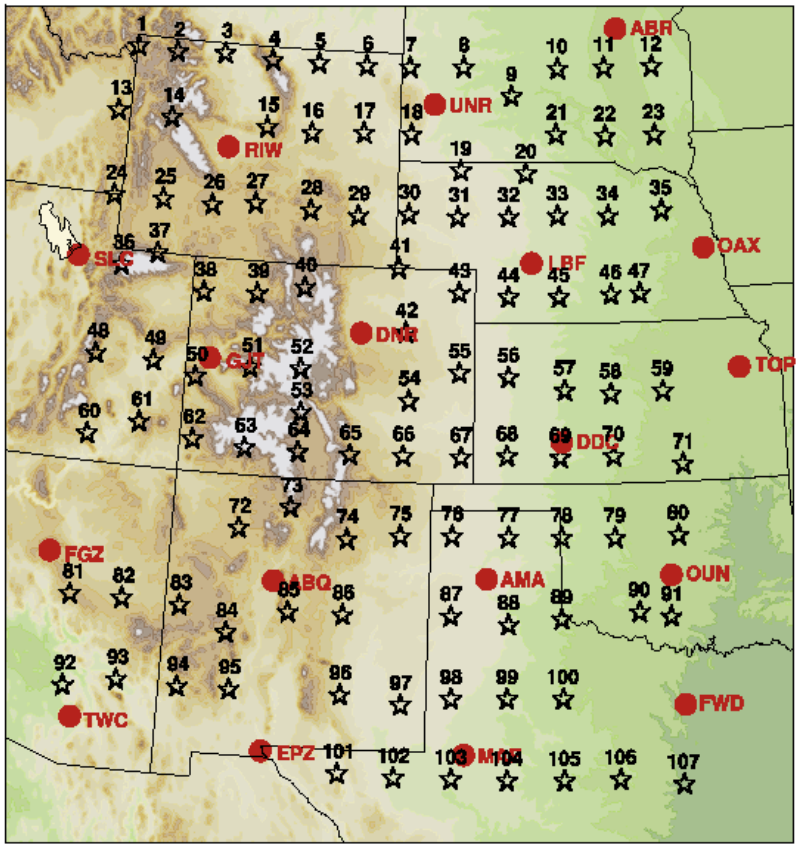


Figure 1. The full domain of interest for MPEX morning dropsonde operations, along with **a** pre-vetted set of dropsonde sites (numbered stars). Operational National Weather Service (NWS) sounding sites are indicated by the red dots (Weisman et al. 2015).

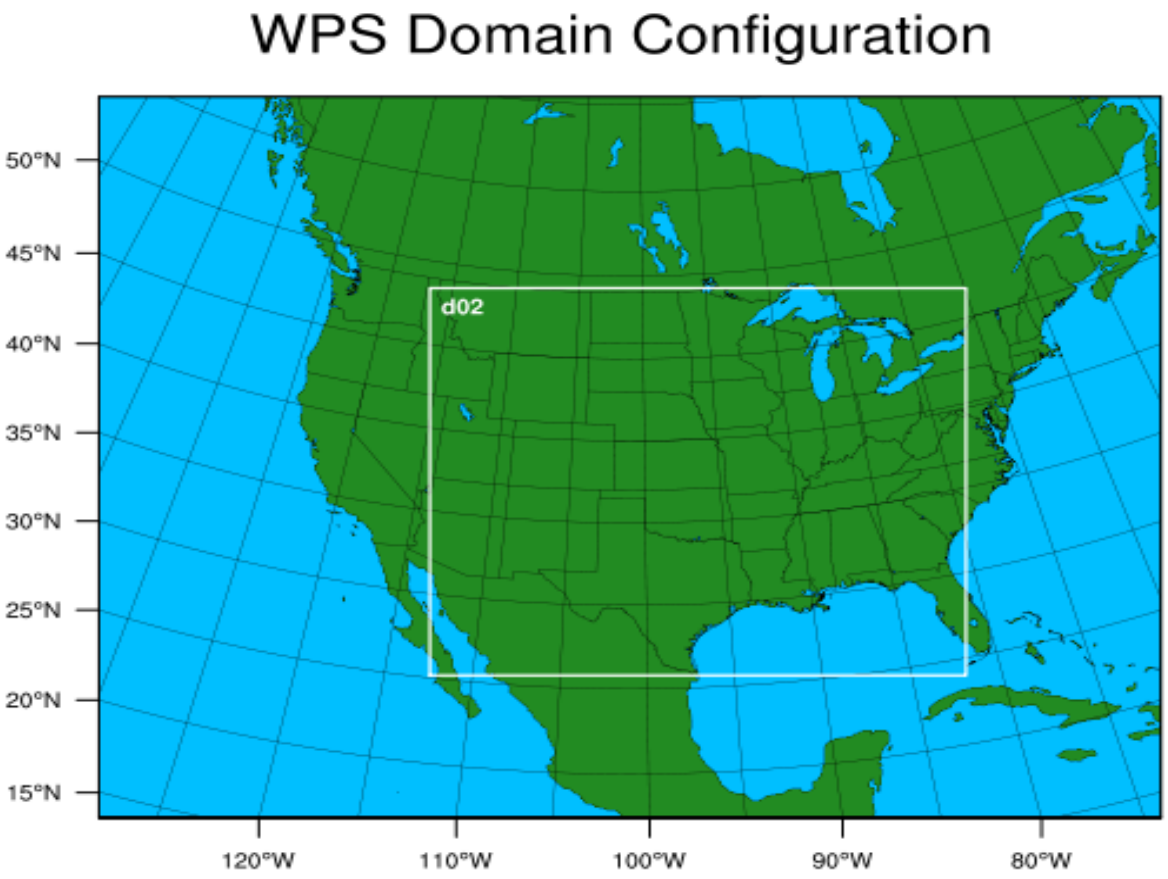


Figure 2. The computational domain.

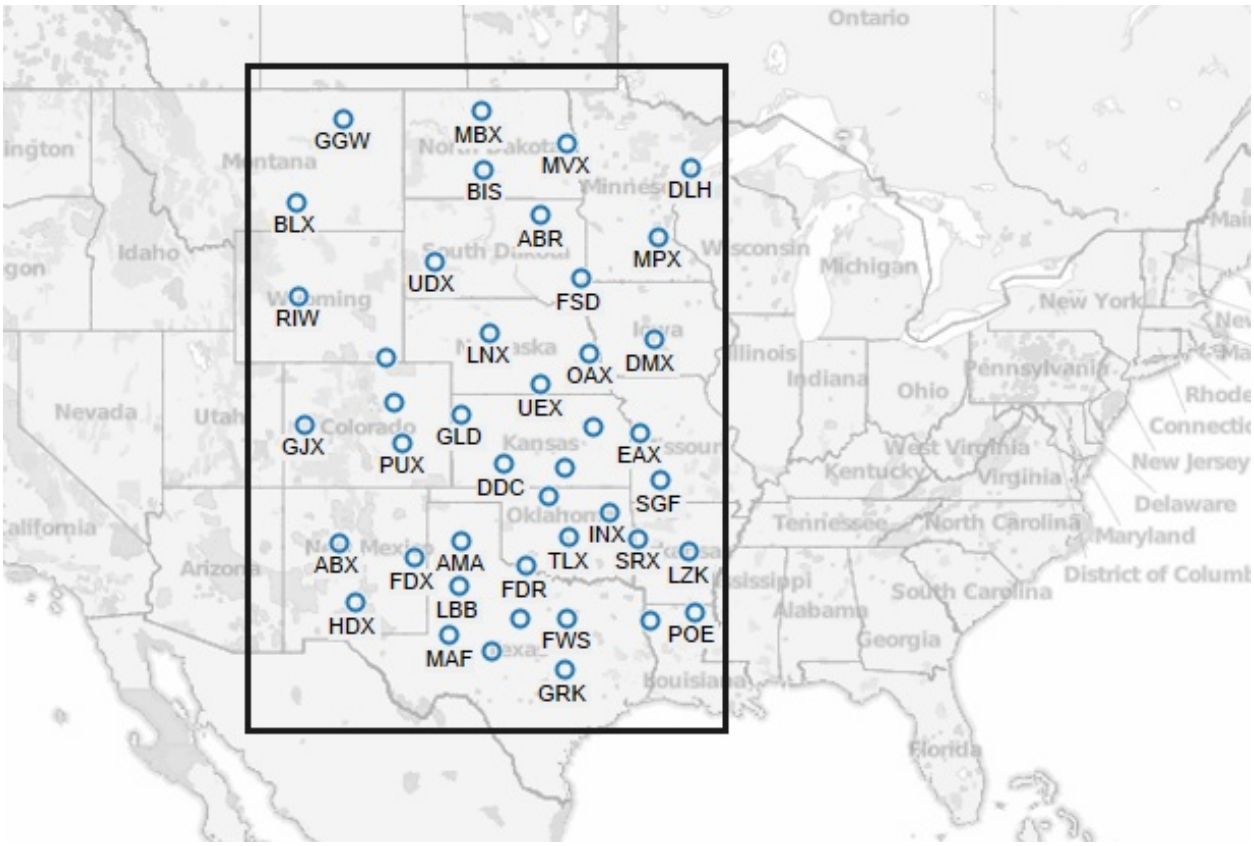


Figure 3. Domain covered by radars.

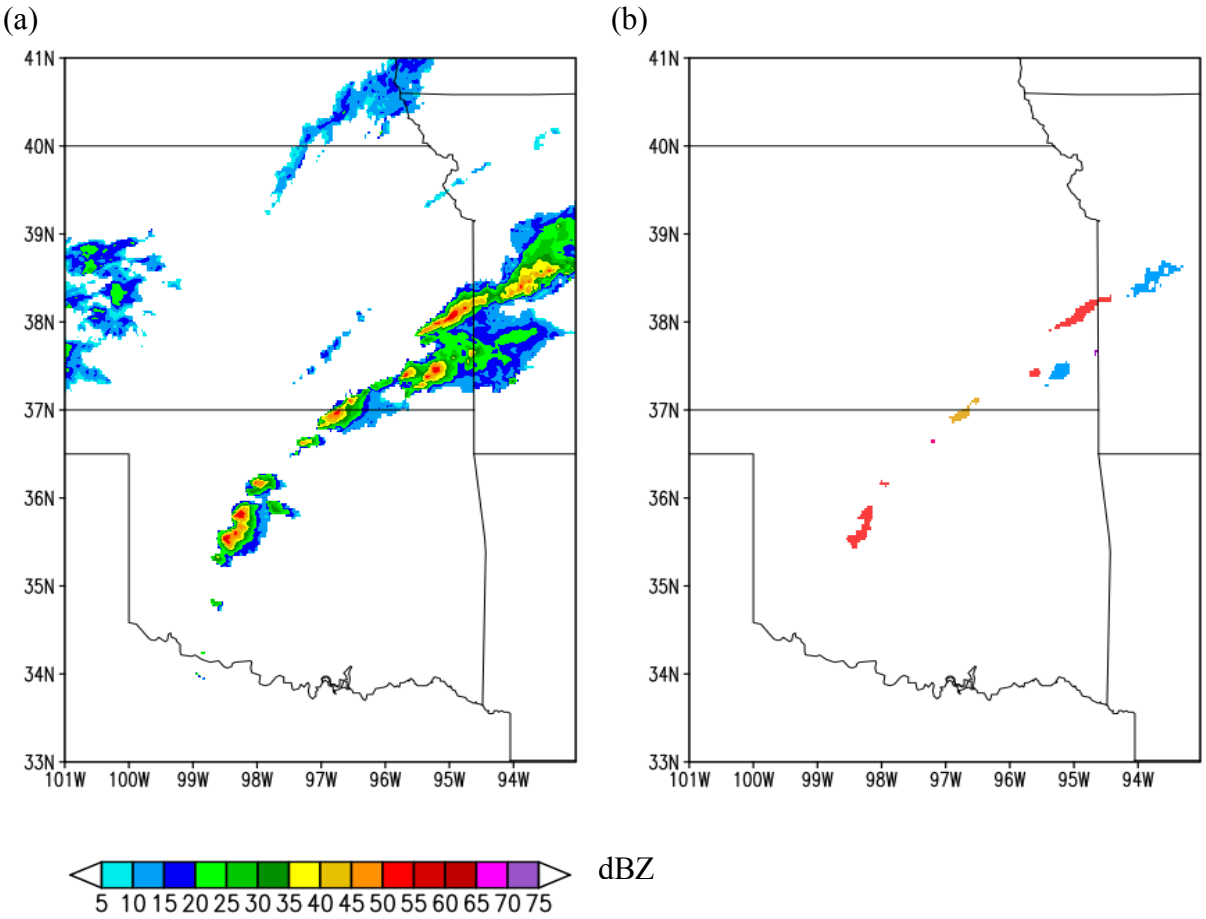


Figure 4. An example of the CA algorithm output from 2200 UTC 31 May 2013 showing the (a) interpolated observed reflectivity and (b) identified convectively active objects.

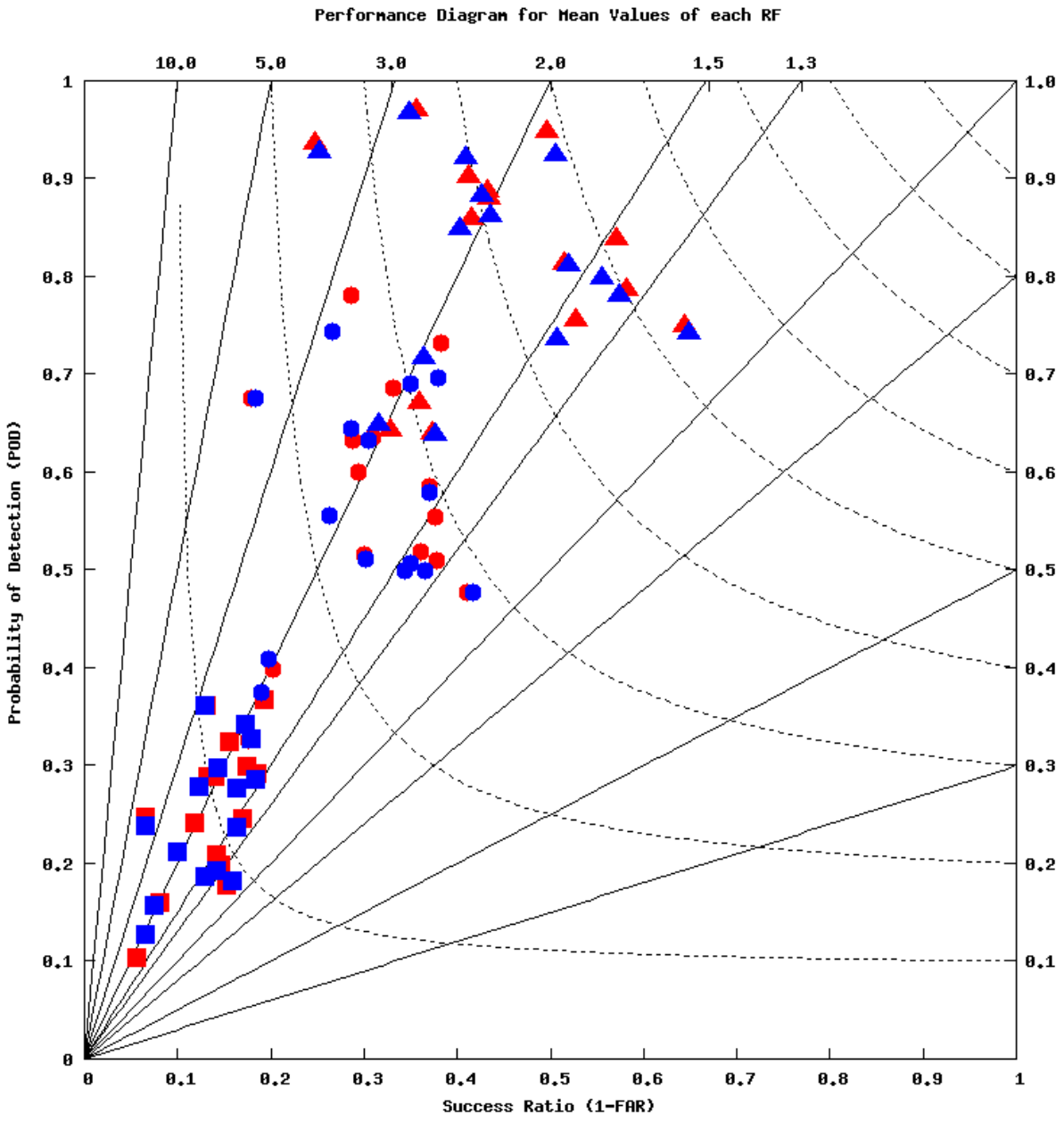
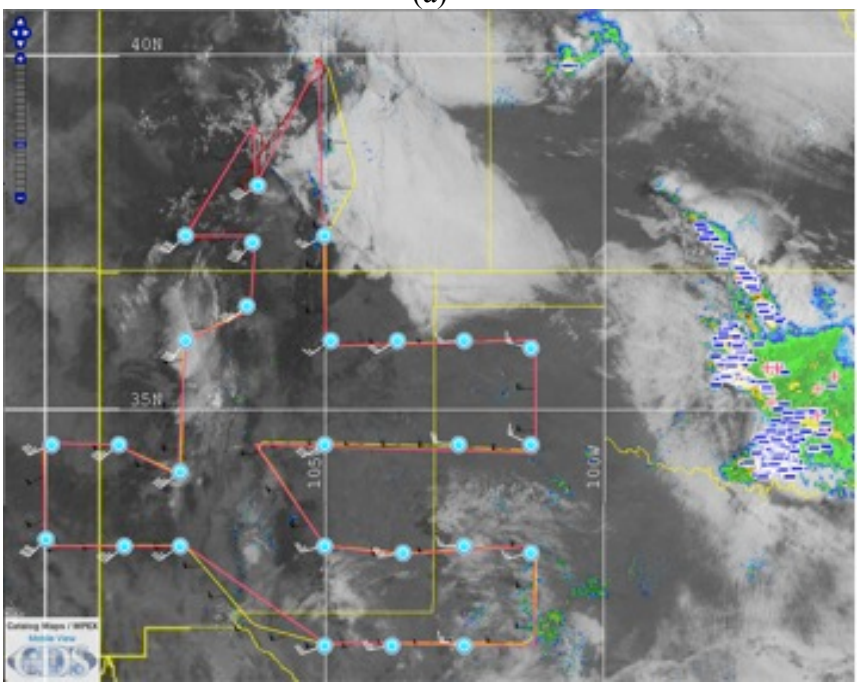


Figure 5. Performance (Roebber) Diagram for the mean values of all 15 IOP With MPEX and Without MPEX ensembles for the three spatiotemporal thresholds: 50km/30 min (squares), 100 km/60 min (circles), and 200 km/120 min (triangles). Red indicates the Perturbed ensemble and Blue indicates the Control ensemble. Filled shapes are the ensemble mean values and the hollow shapes are individual ensemble members. Bias is shown through the solid lines and CSI is represented by the dashed, curved lines.

(a)



(b)

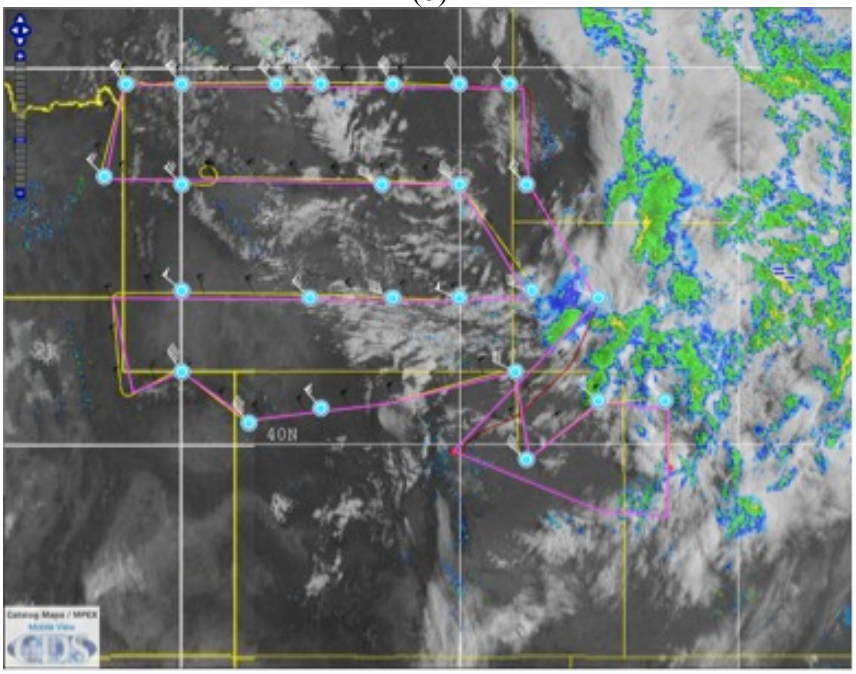
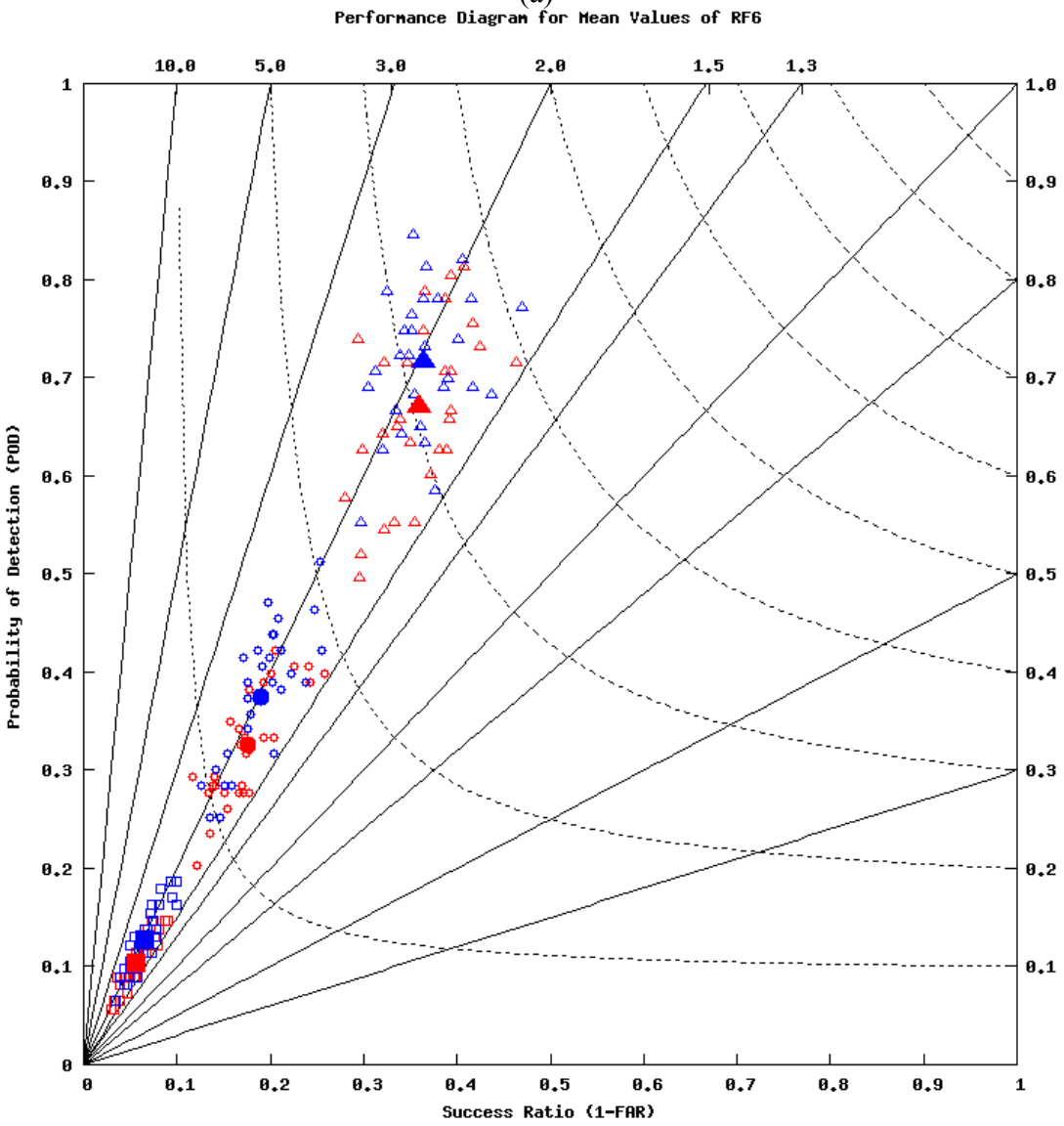


Figure 6. The GV flight tracks for (a) RF6 and (b) RF12.

(a)


Performance Diagram for Mean Values of RF6
10.0
5.0
3.0
2.0
1.5
1.3
Probability of Detection (POD)
Success Ratio (1-FAR)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

(b)

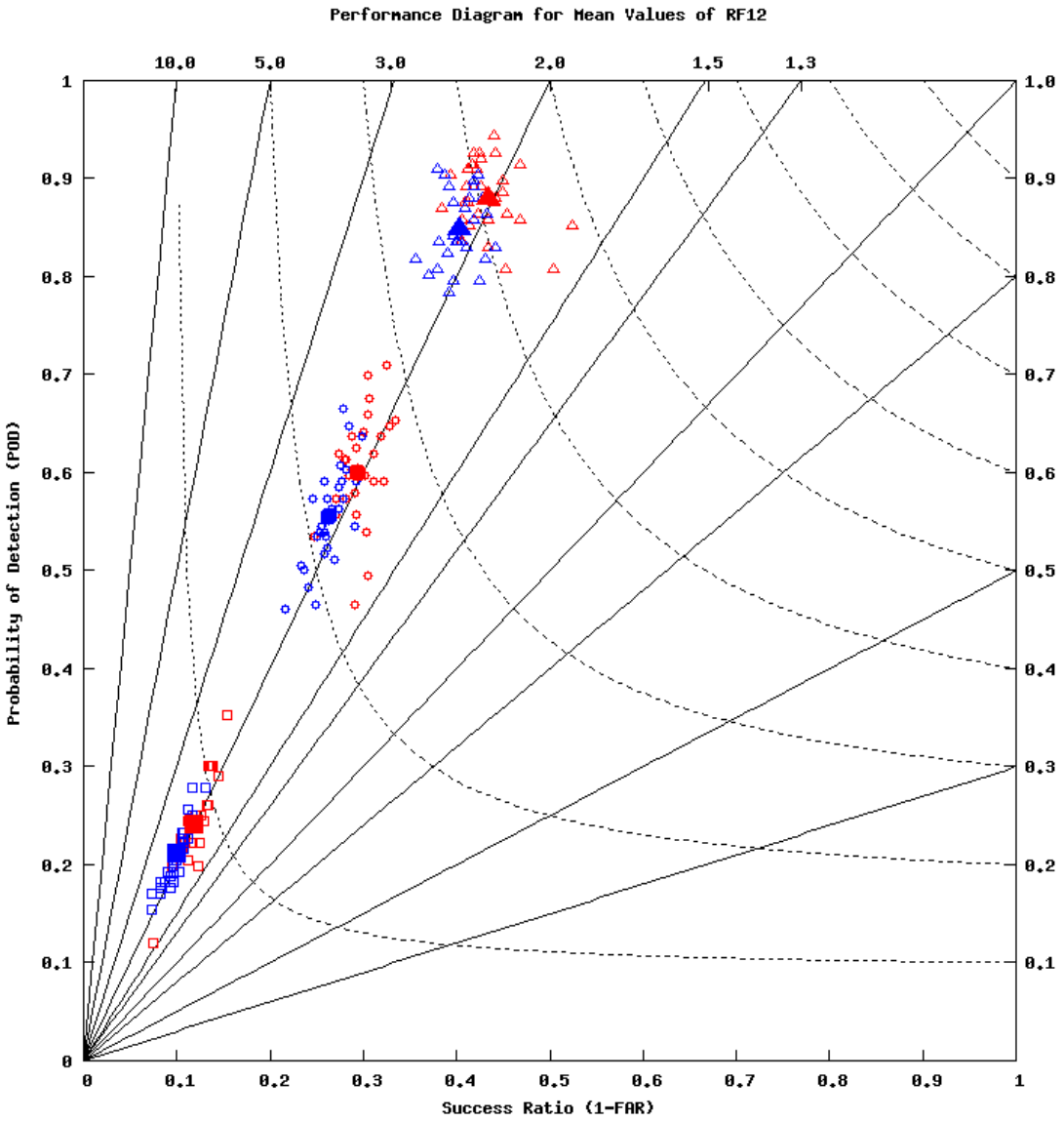
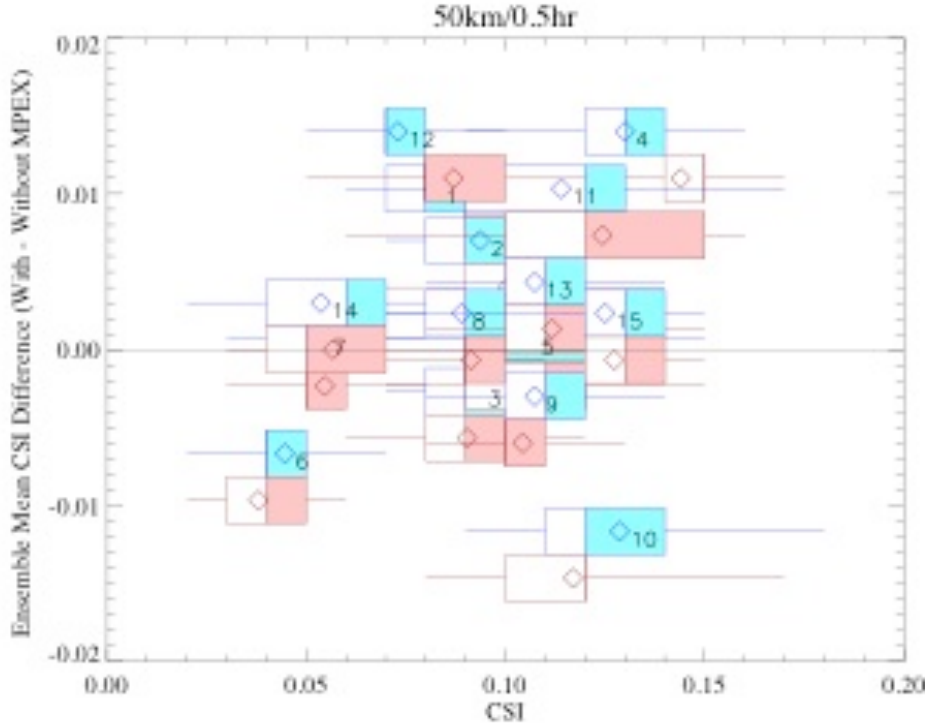
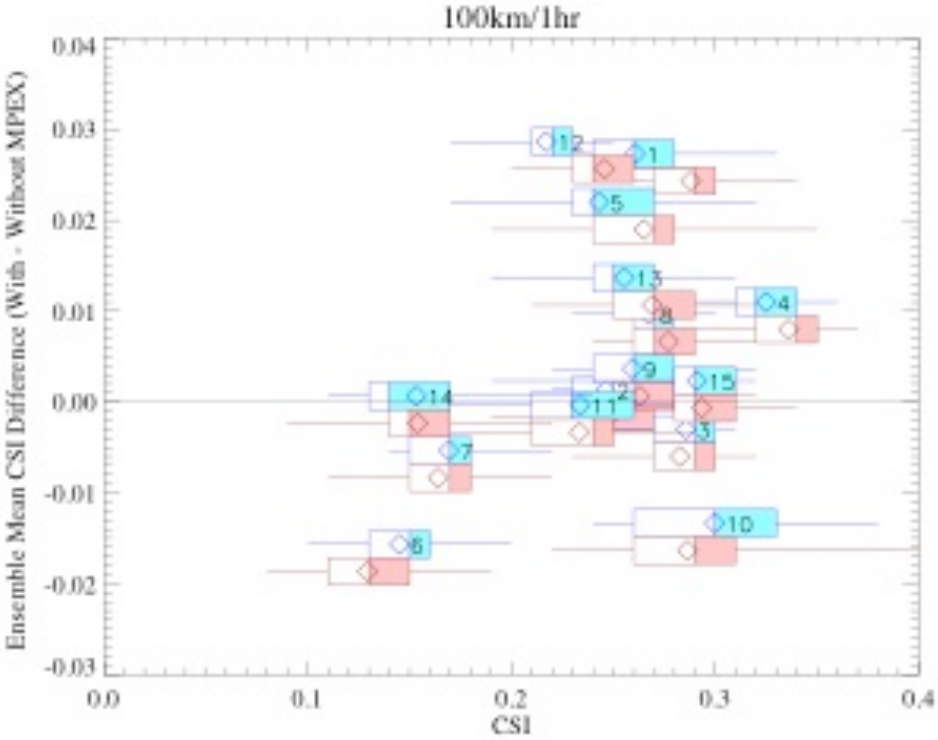


Figure 7. Performance (Roebber) Diagram for (a) RF6 and (b) RF12 for the 50km/30 min (squares), 100 km/60 min (circles), and 200 km/120 min (triangles) spatiotemporal thresholds. Shaded shapes are the ensemble mean and hollow shapes are individual ensemble members. Red indicates the Perturbed ensemble and blue indicates the Control ensemble.

(a)


50km/0.5hr
Ensemble Mean CSI Difference (With - Without MPEX)
CSI
0.02
0.01
0.00
-0.01
-0.02
0.00
0.05
0.10
0.15
0.20
12
4
1
11
2
13
14
8
15
7
5
3
9
6
10

(b)


100km/1hr
Ensemble Mean CSI Difference (With - Without MPEX)
CSI
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
0.0
0.1
0.2
0.3
0.4
12
1
5
13
4
8
9
15
14
2
11
3
7
10
6

(c)

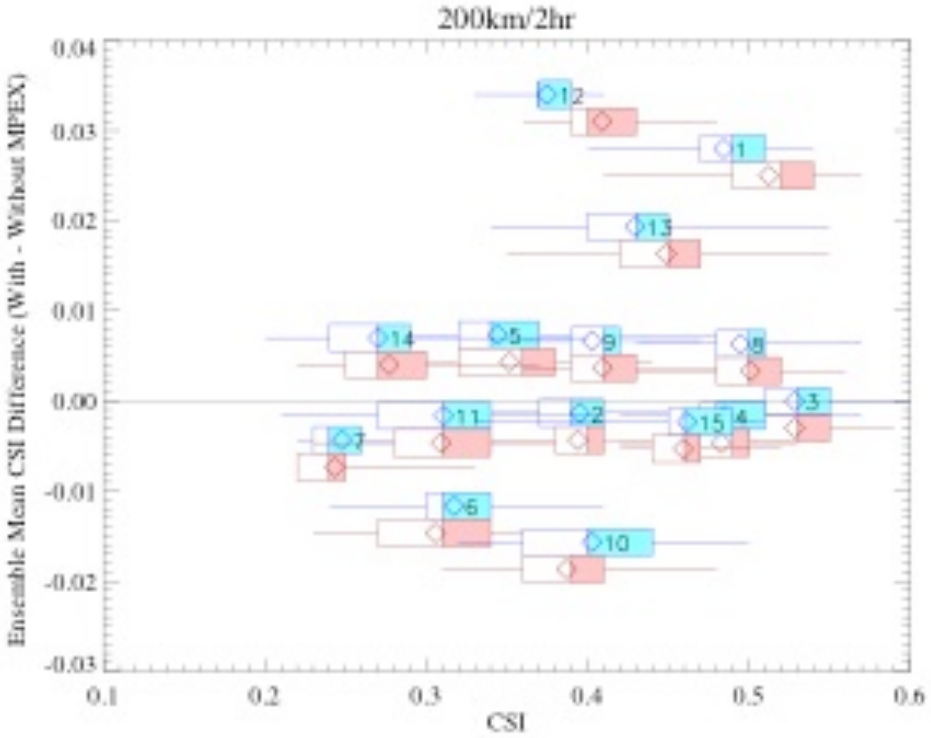
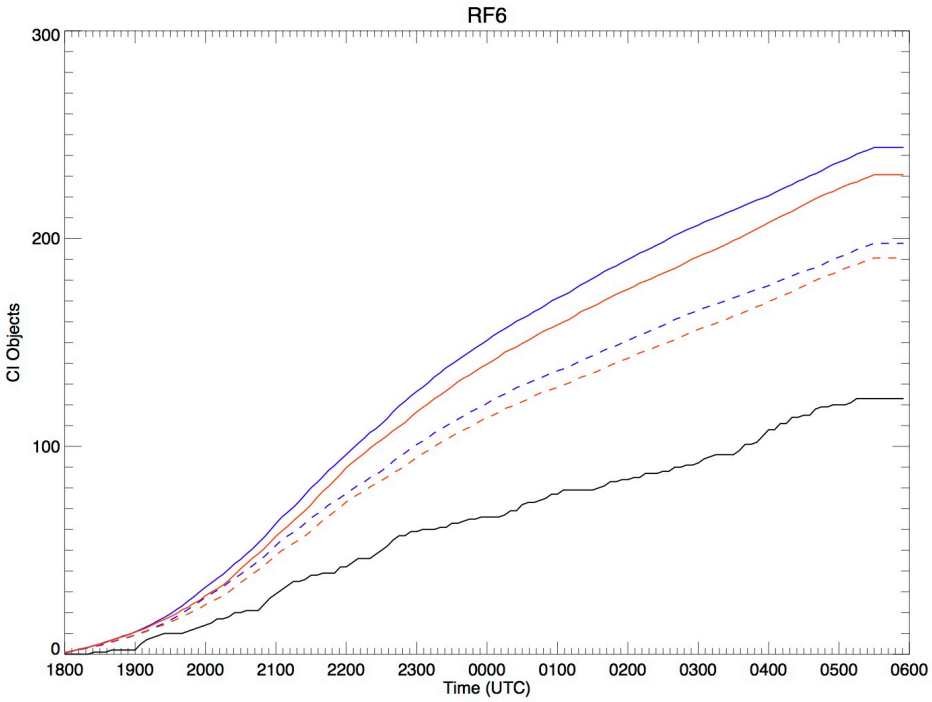


Figure 8. CSI Ensemble CSI mean (diamond), minimum (left whisker), maximum (right whisker), lower quartile, median and upper quartile values are plotted for the 15 Control (blue) and 15 Perturbed (red) cases.

(a)



(b)

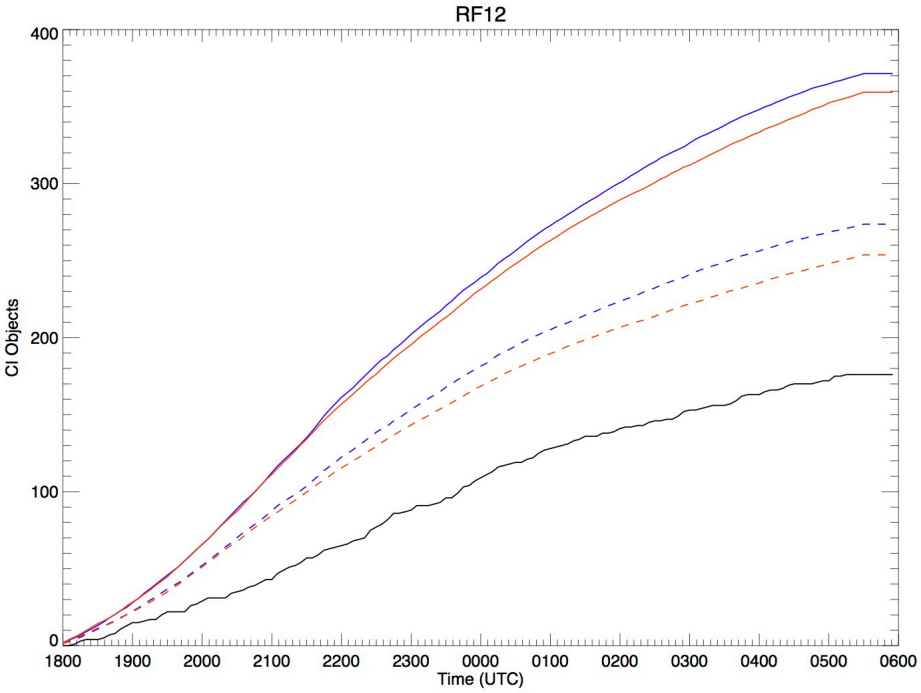


Figure 9. The count of observed CI events (black line), the mean Control CI objects (solid blue) and false alarms (dashed blue) as well as the mean Perturbed CI objects (solid red) and false alarms (dashed red) over the radar specified domain for (a) RF6 and (b) RF12.

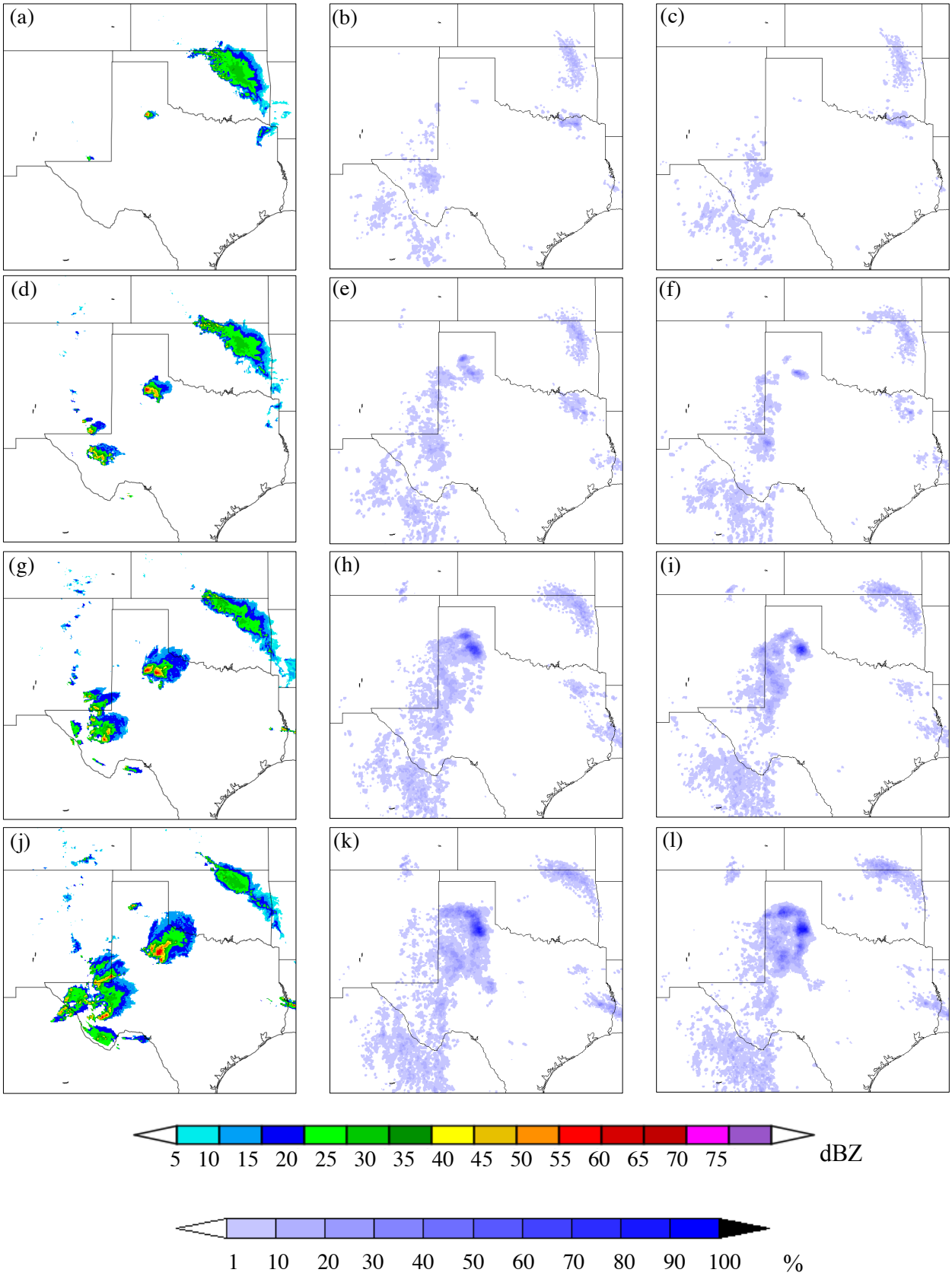
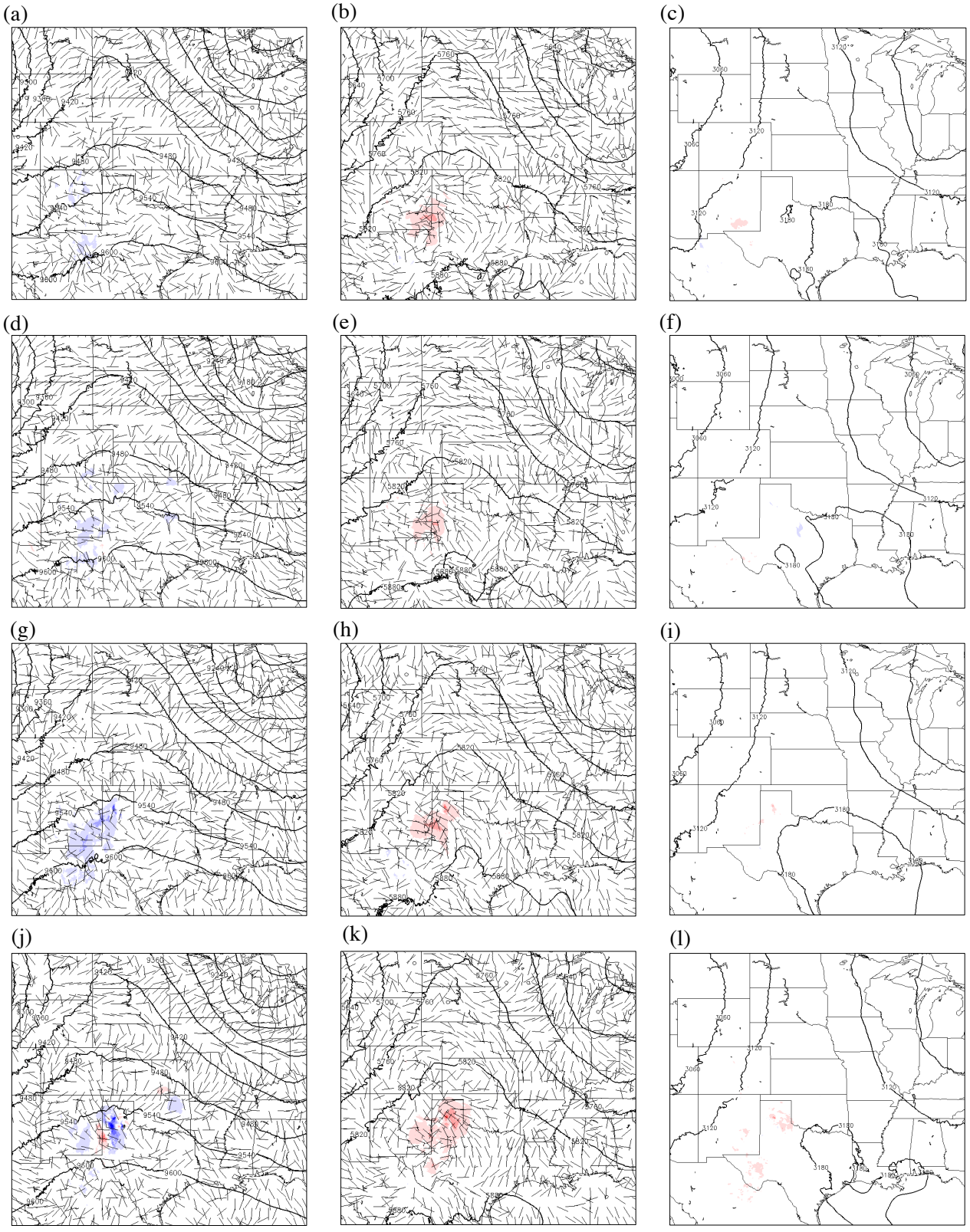

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
dBZ
1 10 20 30 40 50 60 70 80 90 100
%

Figure 10. The observed radar reflectivity at –10°C isotherm, probability of Control reflectivity greater than 35 dBZ at the –10°C isotherm, probability of Perturbed reflectivity greater than 35 dBZ at the –10°C isotherm for RF6, respectively at (a-c) 1900 UTC, (d-f) 2000 UTC, (g-i) 2100 UTC, and (j-l) 2200 UTC 23 May 2013.


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

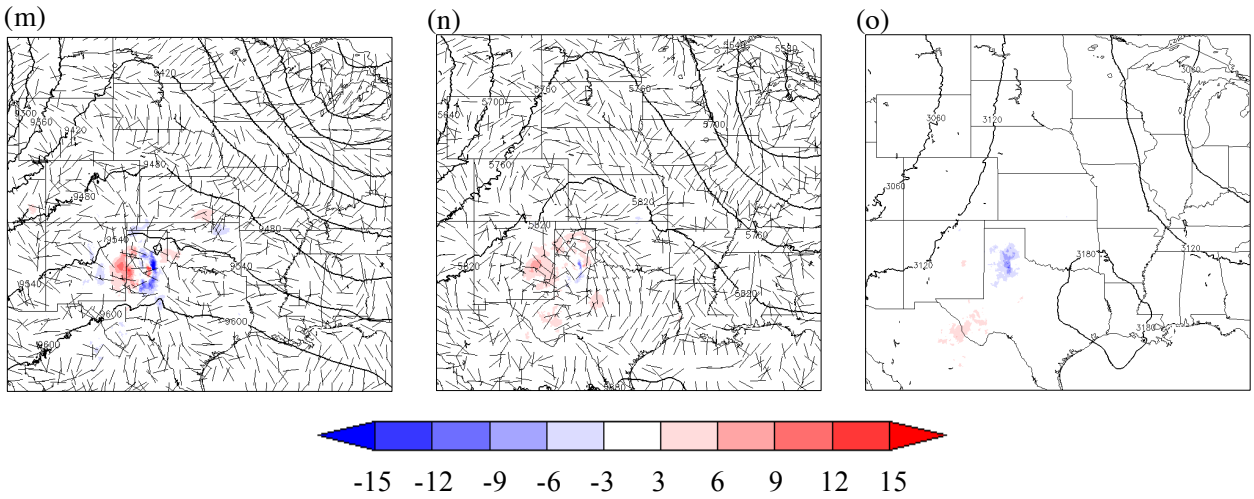
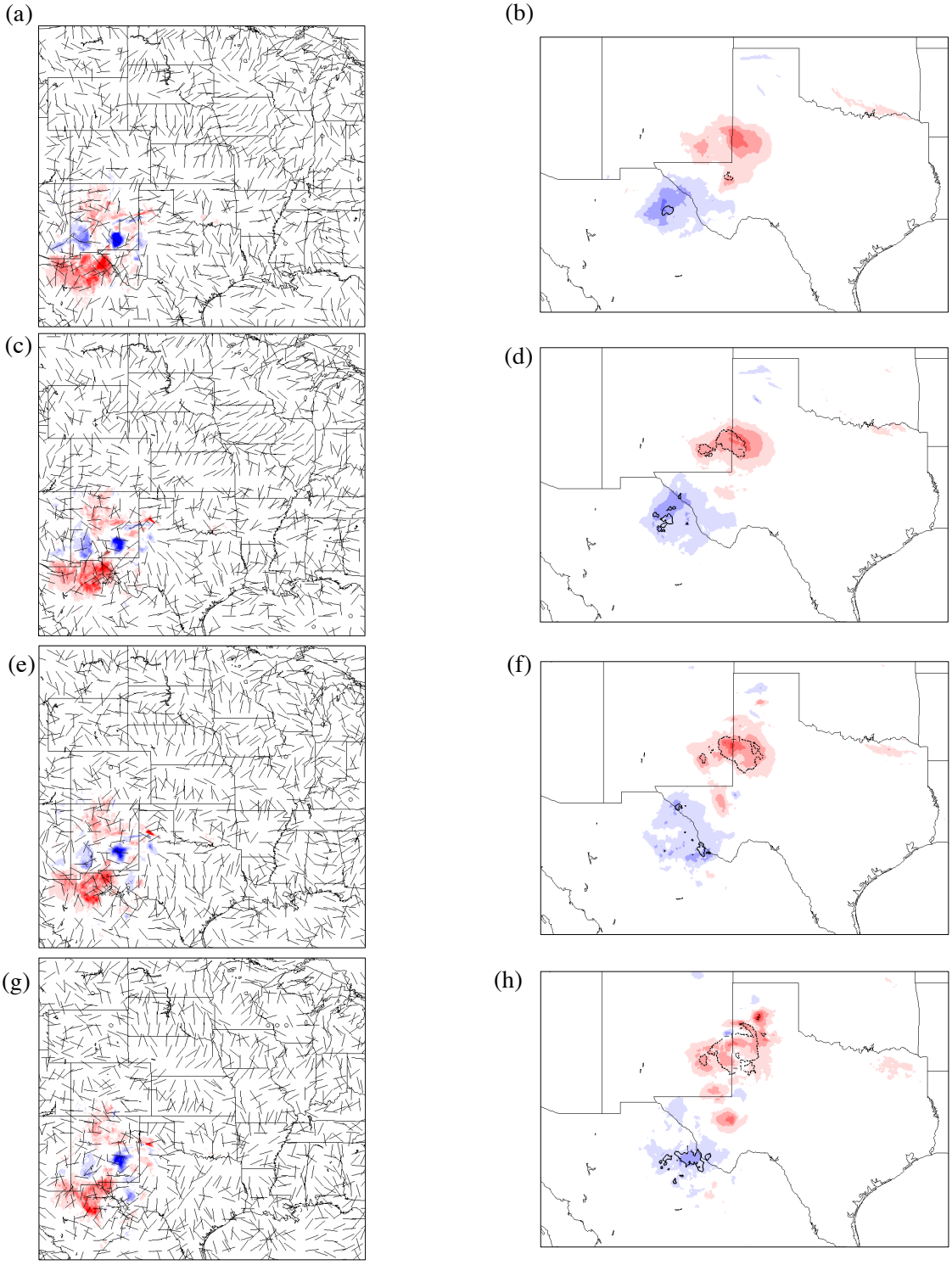


Figure 11. 300 hPa Control height (m; contour), 300 hPa Perturbed – Control height (m; shaded), 300 hPa Perturbed – Control winds (kt; barb); 500 hPa Control height (m; contour), 500 hPa Perturbed – Control height (m; shaded), 500 hPa Perturbed – Control winds (kt; barb); 700 hPa Control height (m; contour), 700 hPa Perturbed – Control height (m; shaded), respectively for (a-c) 1800 UTC, (d-f) 1900 UTC, (g-i) 2000 UTC, (j-l) 2100 UTC, (m-o) 2200 UTC 23 May 2013.


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)

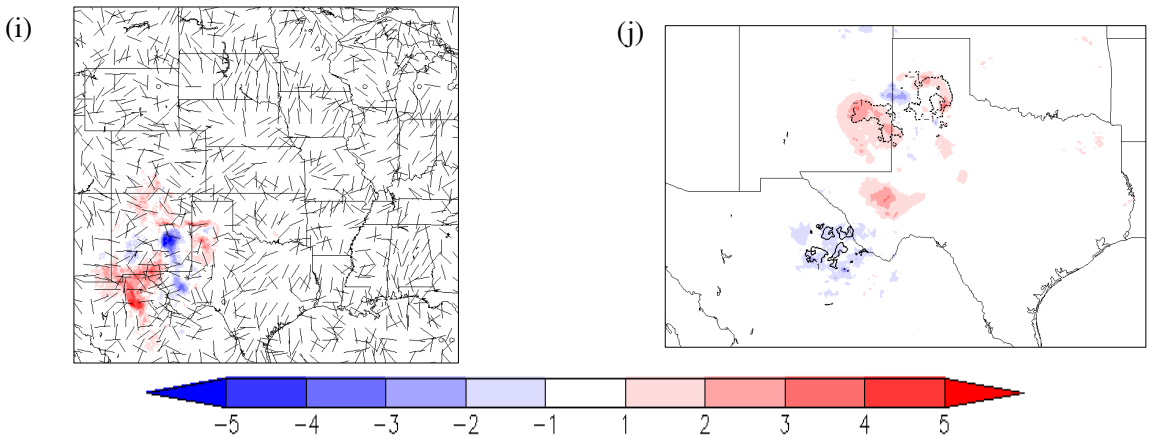


Figure 12. 2 meter Perturbed – Control dewpoint temperature (°C; shaded) and 10 meter Perturbed – Control winds (kt; barb); Perturbed – Control SLP (hPa; contour) and 2 meter Perturbed – Control temperature (°C; shaded) for (a-b) 1800 UTC, (c-d) 1900 UTC, (e-f) 2000 UTC, (g-h) 2100 UTC, (i-j) 2200 UTC 23 May 2013.

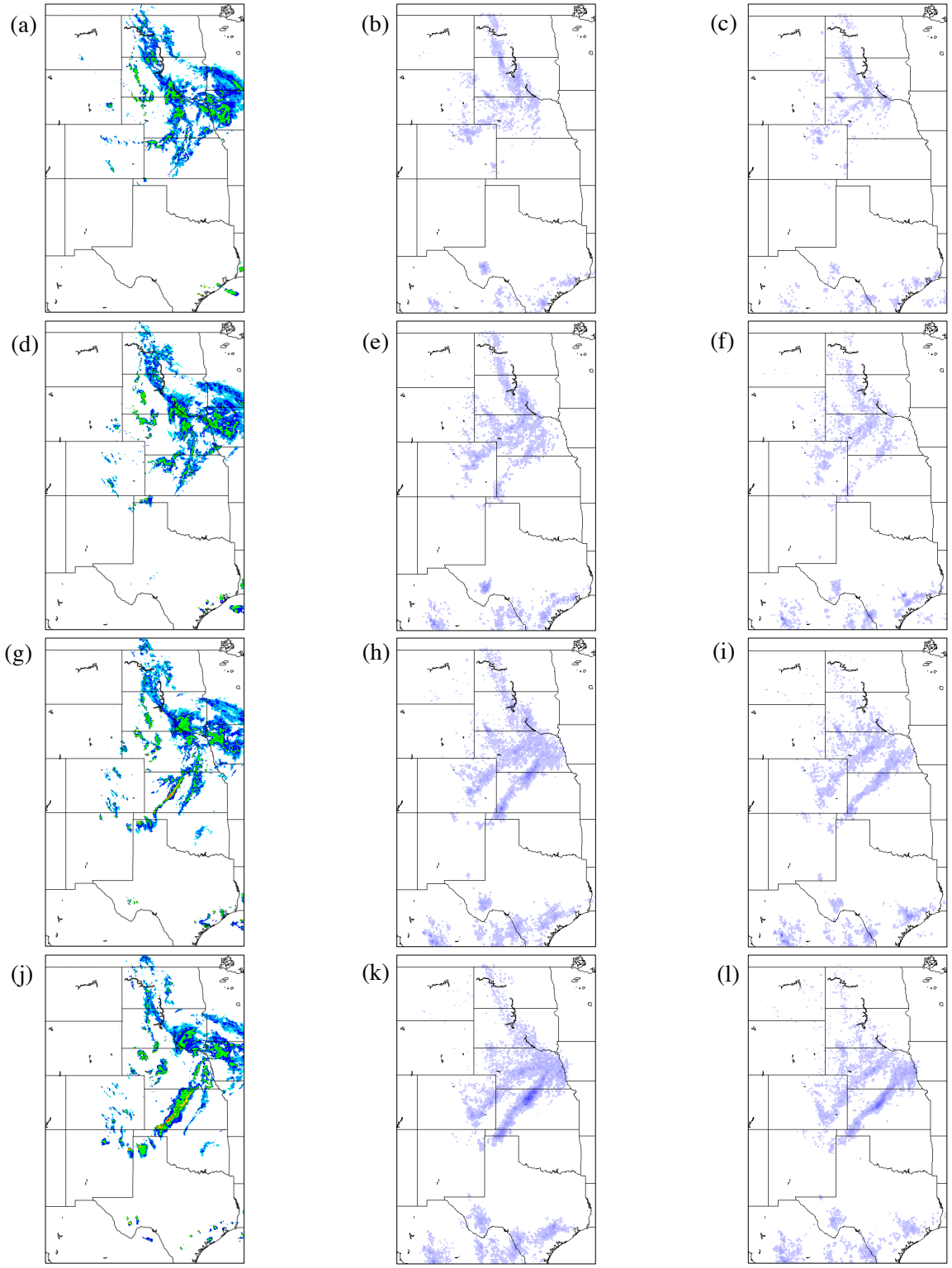
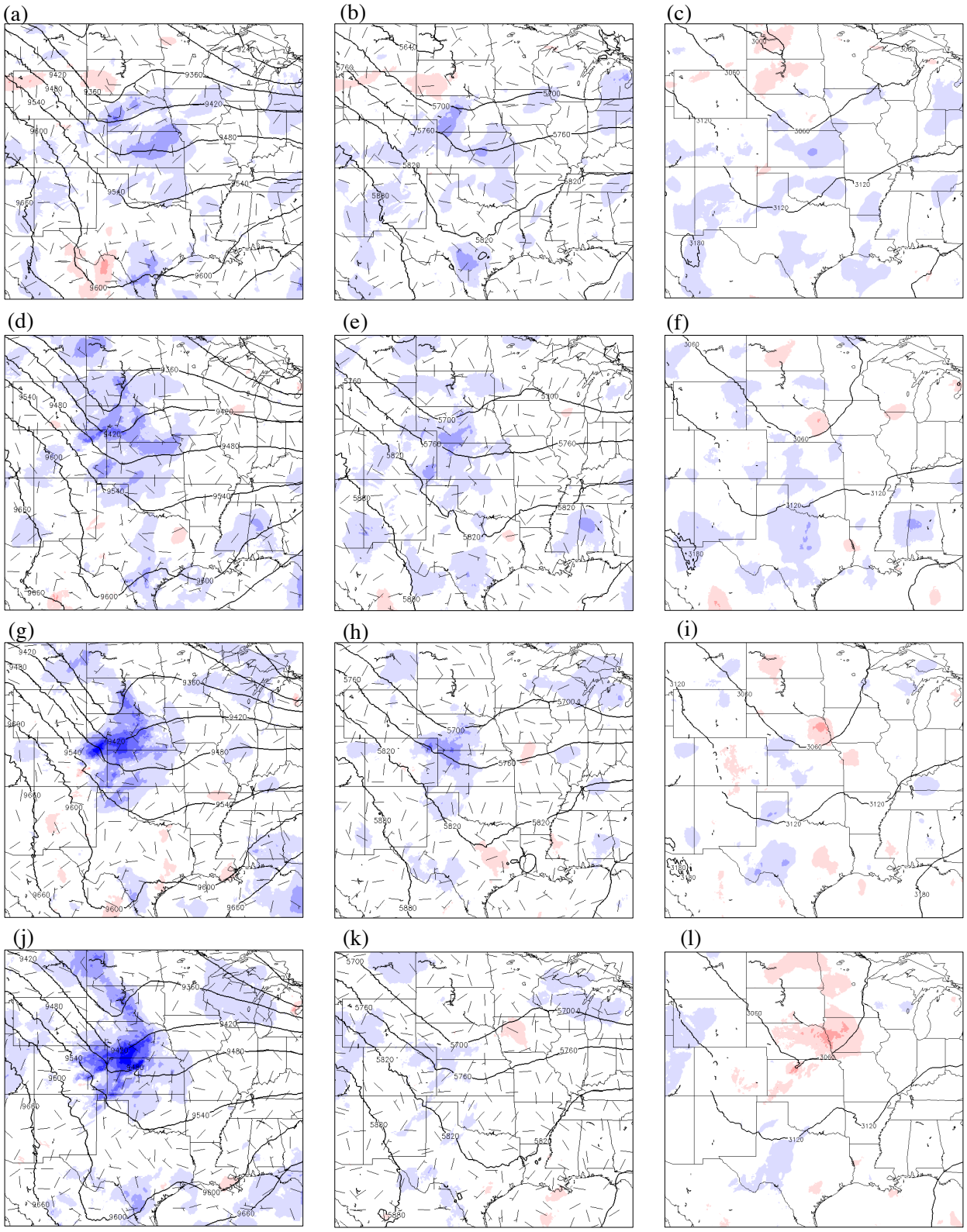


Figure 13. The observed radar reflectivity at the –10°C isotherm, probability of Control reflectivity greater than 35 dBZ at the –10°C isotherm, probability of Perturbed reflectivity greater than 35 dBZ at the –10°C isotherm for RF12, respectively at (a-c) 1900 UTC, (d-f) 2000

UTC, (g-i) 2100 UTC, and (j-l) 2200 UTC 08 June 2013 (shading legend is the same as Figure 10).


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

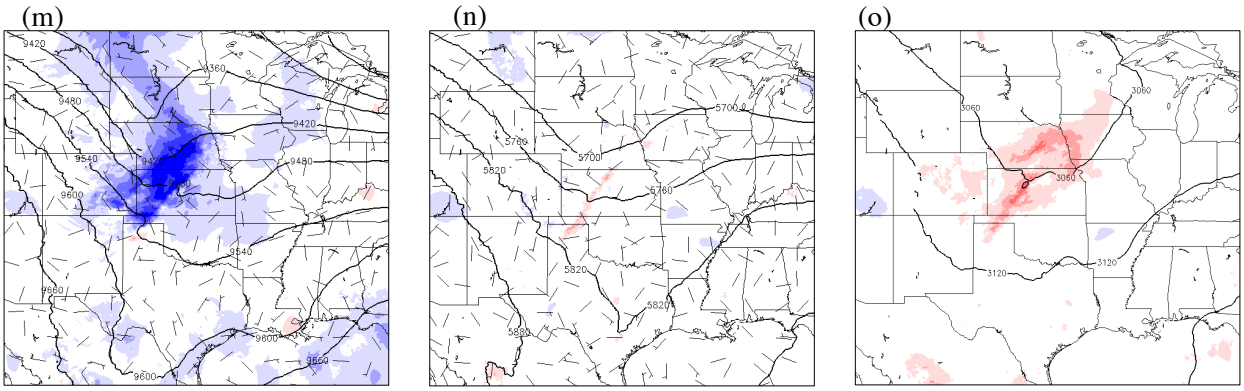
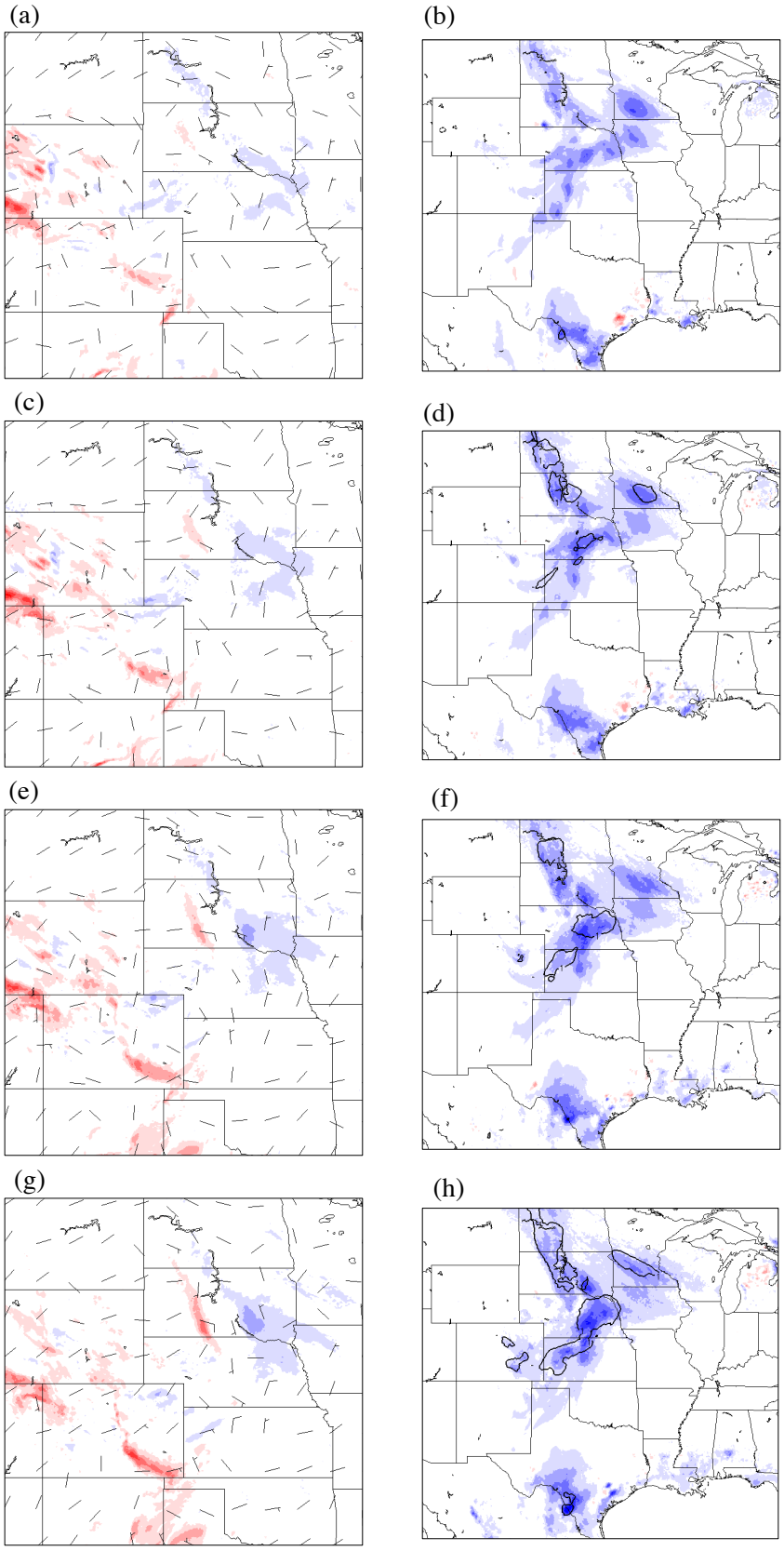


Figure 14. 300 hPa Control height (m; contour), 300 hPa Perturbed – Control height (m; shaded), 300 hPa Perturbed – Control winds (kt; barb); 500 hPa Control height (m; contour), 500 hPa Perturbed – Control height (m; shaded), 500 hPa Perturbed – Control winds (kt; barb); 700 hPa Control height (m; contour), 700 hPa Perturbed – Control height (m; shaded), respectively for (a-c) 1900 UTC, (d-f) 2000 UTC, (g-i) 2100 UTC, (j-l) 2200 UTC, (m-o) 2300 UTC 08 June 2013 (shading legend the same as Figure 11).


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)

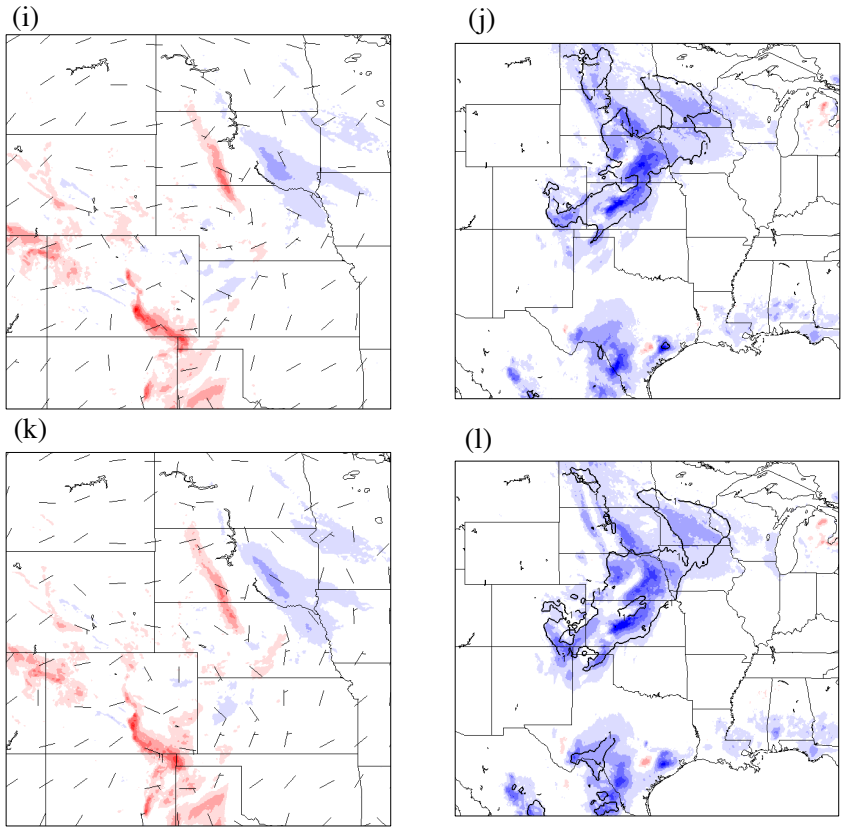


Figure 15. 2 meter Perturbed – Control dewpoint temperature (°C; shaded) and 10 meter Perturbed – Control winds (kt; barb); Perturbed – Control SLP (hPa; contour) and 2 meter Perturbed – Control temperature (°C; shaded) for (a-b) 1800 UTC, (c-d) 1900 UTC, (e-f) 2000 UTC, (g-h) 2100 UTC, (i-j) 2200, and (k-l) 2300 UTC 08 June 2013 (shading legend is the same as Figure 12).

References

Aberson, S.D., 2010: 10 Years of Hurricane Synoptic Surveillance. *Mon Wea. Rev.*, **138**, 1536–1549.

Anderson, J.L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon Wea. Rev.*, **129**, 2884–2903.

——, 2003: A local least squares framework for ensemble filtering. *Mon Wea. Rev.*, **131**, 634–642.

——, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus.*, **61A**, 72–83.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296.

Barker, D. M., and Coauthors, 2012: The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843.

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation-forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518.

——, B. D. Jamison, W. R. Moninger, S. R. Sahm, B. E. Schwartz, and T. W. Schiatter, 2010: Relative short-range forecast impact from aircraft, profiler, rawinsonde, VAD, GPS-PW, METAR, and mesonet observations via the RUC hourly assimilation. *Mon. Wea. Rev.*, **138**, 1319–1343.

Burghardt B.J., C. Evans, and P.J. Roebber, 2014: Assessing the Predictability of Convection

Initiation in the High Plains Using an Object-Based Approach. *Wea. Forecasting*, **29**, 403–418.

Chen, F. and J. Dudhia, 2001: Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.*, **129**, 569–585.

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

Chou, K.-H., C.-C. Wu, P.-H. Lin, S. D. Aberson, M. Weissmann, F. Harnisch and T. Nakazawa, 2011: The Impact of Dropwindsonde Observations on Typhoon Track Forecasts in DOTSTAR and T-PARC, *Mon Wea. Rev.*, **139**, 1728–1743.

Fourrié, N. D., M. F. Rabier, B. Chapnik, and G. Desroziers, 2006: Impact study of the 2003 North Atlantic THORPEX Regional Campaign. *Quart. J. Roy. Meteor. Soc.*, **132**, 275–295.

Fowle, M.A., and P.J. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode and location. *Wea. Forecasting*, **18**, 782–794.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Janjić, Z.I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Jorgensen, D. P. and T.M. Weckwerth, 2003: Forcing and organization of convective storms. Radar and Atmospheric Science: A Collection of Essays in Honor of David Atlas.

Meteor. Monogr., **52**, 75–103.

Kain, J.S., S.J. Weiss, D.R. Bright, M.E. Baldwin, J.J. Levit, G.W. Carbin, C.S. Schwartz, M. Weisman, K.K. Droegemeier, D. Weber, and K.W. Thomas, 2008: Some practical considerations for the first generation of operational convection allowing NWP: How much resolution is enough? *Wea. and Forecasting*, **23**, 931–952.

Kursinski, E.R., G. A. Hajj, J. T. Shofield, R. P. Linfield, and K. R. Hardy, 1997: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning *System. J. Geophys. Res.*, **102** (D19), 23 429–423 465.

Lakshmanan V., T. Smith, G. J. Stumpf, and K. Hondl, 2007: The warning decision support system - integrated information. *Wea. Forecasting*, **22**, 596–612.

——, 2012: Automating the analysis of spatial grids: A practical guide to data mining geospatial images for human and environmental applications. Springer

——, and T. W. Humphrey, 2013: A mapreduce technique to mosaic continental-scale weather radar data in real-time. *IEEE J. of Select Topics in Appl. Earth Obs. and Remote Sensing*, in Review.

——, T. Smith, K. Hondl, G. J. Stumpf, and A. Witt, 2006: A real-time, three dimensional, rapidly updating, heterogeneous radar merger technique for reflectivity, velocity and derived products. *Wea. Forecasting*, **21**, 802–823.

——, K. Hondl, and R. Rabin, 2009: An efficient, general-purpose technique for identifying storm cells in geospatial images. *J. Ocean. Atmos. Tech.*, **26**, 523–537.

——, and T. Smith, 2010: An objective method of evaluating and devising storm tracking algorithms. *Wea. and Forecasting*, **29**, 721–729.

——, and T. Smith, 2009: Data mining storm attributes from spatial grids. *J. Ocea. and Atmos. Tech.*, **26**, 2353–2365.

Liu, H., M. Xue, 2008: Prediction of Convective Initiation and Storm Evolution on 12 June 2002 during IHOP_2002. Part I: Control Simulation and Sensitivity Experiments. *Wea. Rev.*, **136**, 2261–2282.

Majumdar, S.J., and Coauthors, 2011: Targeted observations for improving numerical weather prediction: An overview. *A World Research Programme/THORPEX*, **15**.

Markowski, P., C. Hannon, and E. Rasmussen, 2006: Observations of convection initiation "failure" from the 12 June 2002 IHOP deployment. *Mon. Wea. Rev.*, **134**, 375–405.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102**, 16 663–16 682.

Rabier, F., P. Gauthier, C. Cardinali, R.H. Langland, M. Tsyrulnikov, A. Lorenc, P. Steinle, R. Gelaro, K. Koizumi, 2008: An update on THORPEX-related research in data assimilation and observing strategies. *Nonlinear Processes Geophys.*, **15**, 81–94.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608.

Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilated system and its influence of convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284.

Schwartz, C. S., G. S. Romine, M. L. Weisman, R. A. Sobash, K. R. Fossell, K. W. Manning, and S.B. Trier, 2015: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. and Forecasting*, **30**,

1158–1181.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN–475+STR, 113 pp. [Available online at www2.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]

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

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502.

——, 2010: Performance of a mesoscale ensemble Kalman filter (EnKF) during the NOAA high-resolution hurricane test. *Mon. Wea. Rev.*, **138**, 4375–4392.

——, and G.S. Romine, 2015: Sensitivity of central Oklahoma convection forecasts to upstream potential vorticity anomalies during two strongly forced cases during MPEX. *Mon. Wea. Rev.*, **143**, 4064–4087.

Wandishin, M. S., D. J. Stensrud, S. L. Mullen, and L. J. Wicker, 2010: On the predictability of mesoscale convective systems: Three-dimensional simulations. *Mon. Wea. Rev.*, **138**, 863–885.

Weckwerth T. M., and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP_2002. *Mon. Wea. Rev.*, **134**, 5–22.

——, H. V. Murphey, C. Flamant, J. Goldstein, and C. R. Pettet, 2008: An observational study of convection initiation on 12 June 2012 during IHOP_2002. *Mon. Wea. Rev.*, **136**,

2283–2304.

Weisman, M. L., C. Davis, W. Wang, K.W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437.

——, and Coauthors, 2015 The Mesoscale Predictability Experiment (MPEX). *Bull. Amer. Meteor. Soc.* **96**. 2127–2150.

Wilks, D. S., 1995: Statistical Methods in Atmospheric Sciences: An Introduction. *Academic Press*, 704 pp.

Wu, C.-C., K.-H. Chou, P-H. Lin, S. D. Aberson, M. S. Peng, and T. Nakazawa, 2007: The impact of dropwindsonde data on typhoon track forecasts in DOTSTAR. *Wea. Forecasting*, **22**, 1157–1176.

University Corporation for Atmospheric Research (UCAR), 2013: MPEX Field Catalog. [Available online at http://catalog.eol.ucar.edu/mpex]

University of Oklahoma, 2010: Warning Decision Support System – Integrated Information. [Available online at http://www.cimms.ou.edu/~lakshman/WDSSII/index.shtml].

U.S. Department of Commerce (USDC), National Climate Data Center (NCDC), (2015): Billion-dollar weather and climate disasters: Summary stats. [Available online at http://www.ncdc.noaa.gov/billions/summary-stats].

Velden, C. and Coauthors, 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.*, **86**, 205–223.