AN OSSE FRAMEWORK FOR ASSESSING THE UTILITY OF STORM-SCALE ENSEMBLE FORECASTS TO TORNADO WARNING OPERATIONS

Corey K. Potvin[1,2] and Louis J. Wicker[2]

[1]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK
[2]NOAA/OAR National Severe Storms Laboratory, Norman, OK

1. INTRODUCTION

National Weather Service forecasters' ability to provide advanced warning of supercell tornadoes currently relies heavily upon the detection (by radar or human observers) of strong low-level storm rotation. This "warn-on-detection" paradigm hinders the tornado warning process in three important ways. First, tornado warning lead times are significantly limited in cases where the onset of strong low-level rotation precedes tornadogenesis by only several minutes. It is therefore not surprising that the average warning lead time for events in which the warning precedes the tornado, about 17 min, did not increase from 1986 to 2006 (Stensrud et al. 2012). Second, the existence of low-level mesocyclone-scale rotation, however intense, does not guarantee tornadogenesis. Achieving a satisfactory probability of detection (POD; currently ~80 %) while maintaining sufficiently long warning lead times therefore results in a high false alarm rate (FAR; currently ~75 %). Third, much of the lowest 1-3 km of the atmosphere lies below the Weather Surveillance Radar-1988 Doppler (WSR-88D) domain (Maddox et al. 2002), precluding low-level radar observations of many storms. The tradeoff between the POD and FAR is sharpened in those cases.

These problems are mitigated under the envisioned "warn-on-forecast" paradigm (Stensrud et al. 2009), in which forecasters would utilize short-term meso- and storm-scale ensemble numerical weather prediction (NWP) models to increase tornado (as well as severe thunderstorm and flash flood) warning lead times and possibly reduce FARs while maintaining high PODs. Available computational resources will presumably constrain the horizontal grid spacing of initial operational warn-on-forecast systems to 1 km or larger, thus precluding direct simulation of tornadoes (thus, the second warn-on-detection weakness listed above may not be addressed by early warn-on-forecast systems). Fortunately, 1 km horizontal grid spacing is sufficiently fine to permit simulation of low-level mesocyclones (LLMs). Using a large and geographically diverse dataset, Trapp et al. (2005) determined that 40 % of mesocyclones with bases below 1 km AGL were tornadic, versus only 15 % of mesocyclones with bases 3-5 km AGL. The potential utility of ensemble systems to LLM forecasting and thereby to tornado warning operations therefore merits serious consideration. Toward that end, this study adopts an observing system simulation experiment (OSSE) framework to estimate the maximum accuracy with which near-term-realizable ensemble forecast systems can predict LLM path, timing and intensity.

Due largely to the nonlinearities associated with convective instability and cloud microphysics, moist convection is chaotic (i.e., initial condition errors grow rapidly, especially as finer scales are simulated; e.g., Zhang et al. 2003; Hohenegger and Schar 2007). Numerical forecasts of supercell thunderstorms, therefore, are sensitive to errors in the initial state estimate provided by the data assimilation procedure. Such errors inevitably arise from deficiencies in (1) the assimilated observations (e.g., data gaps, measurement errors), (2) the assimilation system (e.g., simplified forward operators) and, in four-dimensional variational and ensemble Kalman filter methods, (3) the NWP model (e.g., discretization and physical parameterization errors). Additional errors occur as the model is integrated forward from the initial condition. In this study, we introduce model error into our experiments by using a finer horizontal grid for the truth simulation than for the ensemble analysis-forecasting system.

In deducing implications of our idealized forecasts for near-future warn-on-forecast ensemble systems, it is important to consider how the predictability of the supercells simulated in our experiments compares to the predictability of real supercells. The grid resolution in our "truth" simulations (described in Section 2a) is coarse relative to the inertial subrange of cumulus convection (Bryan et al. 2003). The resulting absence in the simulations of the smallest supercell scales of motion artificially reduces upscale error growth (Lorenz 1969) in forecasts. This suggests the *intrinsic* predictability (i.e., the predictability that would be achieved given a perfect forecast model and very small initial condition uncertainty) of our simulated supercells is greater than that of atmospheric supercells. Moreover, since errors arise in current convection-permitting models from a multitude of sources, our use of a model that, apart from its coarsened resolution, is identical to the model used to generate "truth" presumably leads to our simulated supercells having greater *practical* predictability (i.e., the predictability given constraints in the observational network, data assimilation techniques, and NWP model) than atmospheric supercells. This is true even if, unexpectedly, the intrinsic predictability of our simulated supercells is similar to that of atmospheric supercells. The results of this study are therefore best viewed as sampling the *upper limit* of the accuracy with which early warn-on-forecast systems will forecast LLMs. Establishing such a baseline is critical to assessing the feasibility of operationally useful, long-term tornado warnings.

Studies of the 4-5 May 2007 Greensburg, Kansas, tornadic thunderstorm by Stensrud and Gao (2010) and Dawson et al. (2012) demonstrate that operationally useful ensemble forecasts of low-level vorticity can be achieved in at

* Corresponding Author Address: Dr. Corey K. Potvin, 120 David L. Boren Blvd, Rm 4355, Norman, OK 73072; Email: corey.potvin@noaa.gov.

least some cases despite current observational, model and data assimilation method limitations. The experiments of Snook et al. (2012) support a similar hypothesis for 0-3 h forecasts of mesoscale convective system mesovortices. Assessing the generality of these results, and identifying scenarios that pose a particularly significant challenge to ensemble low-level rotation forecasts, requires that such forecasts be performed and evaluated for a large number of cases spanning a range of radar-storm geometries and atmospheric environments.

The OSSE framework provides a powerful complement to real case studies of such problems, for two primary reasons. First, since the "truth" is known, analysis errors and their sensitivity to important experimental parameters (e.g., radar cross-beam angle) can be precisely determined. Second, there is no need to collect and quality-control numerous observation sets meeting specific desired criteria. In this paper, we simulate scenarios where a supercell is observed by two WSR-88D radars and evaluate forecasts of LLMs for different radar-storm distances, radar cross-beam angles, assimilation period lengths, and low-level rotational intensities. The simulated supercell used as "truth" in the majority of our experiments rapidly develops an intense LLM-like vortex (hereafter, simply "LLM") that undergoes cyclic mesocyclogenesis through the remainder of the simulation. Evaluations of those forecasts focus upon the timing of the development of the initial LLM, as well as the path and rotational intensity of the LLM "family". To test the ability of the ensemble system to distinguish between supercells that develop very different magnitudes of low-level rotation, we also perform experiments in which the "true" storm develops a weaker LLM, or fails to sustain strong low-level rotation at all (null case).

The rest of the paper is organized as follows. Section 2 describes the model configuration for our supercell simulations, the procedure for emulating radar observations of the simulated storms, the ensemble data assimilation and forecast system, and our verification methods. The results of the LLM forecasts are described in Section 3. Implications of the forecast results for the proposed warn-on-forecast paradigm are discussed in Section 4.

2. METHODS

2.1. "Truth" simulations

The two "truth" simulations for our experiments were generated by the National Severe Storms Laboratory Collaborative Model for Multiscale Atmospheric Simulation (NCOMMAS; Wicker and Skamarock 2002; Coniglio et al. 2006). The NCOMMAS is a nonhydrostatic, compressible cloud model designed to simulate convective storms in a simplified setting (e.g., flat surface, no surface fluxes nor radiative transfer, and horizontally uniform base state). The prognostic variables in NCOMMAS are the wind components u, v and w, Exner function π, turbulent mixing coefficient K_m, potential temperature θ, water vapor mixing ratio q_v, and the microphysical parameterization (MP) scheme variables (listed below). The supercell simulations proceeded on a stationary $200 \times 200 \times 20$ km domain with horizontal grid spacing $\Delta_H =$ 1/3 km and vertical spacing increasing from 200 m over the lowest 1 km to 600 m above z = 5.2 km. Both simulations were integrated for 2 h using large and small time steps of 4 s and 2/3 s, respectively.

The sounding (Fig. 1a) that provided the model base state for the default supercell simulation (used in most of our experiments) is a composite of the wind profile from the 1200 UTC 3 April 1974 Covington, Kentucky rawinsonde, modified to yield a storm motion slow enough to permit use of a stationary model grid, and a thermodynamic profile similar to that of Weisman and Klemp (1982; 1984) with some modifications to increase the low-level stability below 800 mb to introduce a weak capping inversion more indicative of supercell environments (G. Bryan 2011, personal communication). The sounding used in the second simulation has the same thermodynamic profile as that used in the default simulation, but the wind profile is modified to reduce the wind shear by 1/3 (Fig. 1b). In both simulations, the storm was initiated with an ellipsoidal 4-K thermal bubble with horizontal and vertical radii of 10 km and 1.4 km, respectively. A fully dual-moment version of the Ziegler et al. (1985) MP scheme (Mansell et al. 2010) was used. The scheme predicts mixing ratio and number concentration for distributions of cloud droplets, rain, cloud ice crystals, snow, graupel and hail, as well as bulk concentration of cloud condensation nuclei, average bulk densities of graupel and hail, and the melted fractional diameters of graupel and hail.

In both supercell simulations, the initial supercell splits several times during the course of the model integration, consistent with the straight-line hodographs above z = 1 km. We restrict our attention to the initial supercell pair in the default simulation and to the initial right-moving supercell in the weaker-shear simulation. Cursory inspection of the simulated low-level reflectivity fields reveals marked differences in the evolution of the three supercells (Fig. 2). Time-height plots of the horizontal-domain-maximum-amplitude cyclonic (anticyclonic) vorticity, ζ_{max}, are shown for the right-moving supercells (left-moving supercell) in Figs. 3a-c (description of Figs. 3d-f is differed to Section 2d). Cyclic mesocyclogenesis occurs in both default-simulation supercells; this process is reflected in temporal oscillations in ζ_{max} (Figs. 3a, b). Series of LLMs from the same storm are considered a single object for the purpose of the verification. The right-moving supercell in the default simulation, "supA", develops a fairly intense LLM just after t = 60 min; the LLM weakens after t = 65 min, rapidly reintensifies after t = 75 min, and remains strong through the end of the simulation (Fig. 3a). The evolution of the LLM of the left-moving supercell in the default simulation, "supB", bears qualitative similarities to that of the supA LLM (Fig. 3b). However, low-level mesocyclogenesis is delayed relative to supA, and ζ_{max} is generally lower above z = 1 km. The latter difference is at least partly attributable to the fact that both storms experience positive storm-relative environmental helicity, the tilting of which enhances the (cyclonic) LLM in supA but weakens the (anticyclonic) LLM in supB. The supercell in the weaker-shear simulation, "supC", fails to sustain strong low-level rotation (Fig. 2c). This is the intended consequence of reducing the environmental vertical wind shear, and likely results from some combination of the following effects: (1) lesser low-level horizontal vorticity available to be tilted into the vertical; (2) lesser mid-level horizontal vorticity, which tends to promote a weaker mid-level mesocyclone and, due to the resultingly reduced perturbation vertical pressure gradient force, weaker low-level updraft and thus vertical vorticity stretching; and (3) reduced storm-relative flow, which tends to cause precipitation to fall too close to the updraft, which in turn is more likely to become undercut by strong outflow (supC indeed becomes outflow-dominant early in the simulation). The large differences in evolution between the three supercells allow us to pose a more varied, meaningful challenge to the ensemble analysis-forecasting system.

As with any OSSE study, the relevance of our results to real-data applications is largely determined by the physical realism of the model used to generate our "truth" simulations. Of particular importance is the representativeness of the low-level vertical vorticity generation processes in the simulations. Given the important contributions of microphysical processes to the magnitudes and locations of thermodynamic and wind gradients in supercells, gross MP scheme errors can lead to unphysical simulation of baroclinic and barotropic vorticity processes. Fortunately, such concerns are mitigated in the present study for three reasons. First, our use of a double-moment MP scheme affords greater flexibility in hydrometeor size distributions than single-moment schemes, presumably improving representation of size sorting and other processes (e.g., Milbrandt and Yau 2006; Dawson et al. 2010). Second, the surface gradients and maximum deficits of perturbation virtual potential temperature in our simulations (not shown) are consistent with the real supercell observations described in Shabbot and Markowski (2006) and Markowski et al. (2002). Third, the storm morphology and evolution (not shown) comport with observations in many important ways, including the bowing of the RFD gust front by surging outflow, and attendant horseshoe-shaped updraft; the existence of a vertical vorticity dipole straddling the hook echo; and the diffuse nature of the forward flank downdraft "gust front". These considerations suggest that our simulations reasonably represent the processes most important to LLM evolution in real supercells.

2.2. Radar emulation and experiments

Pseudo-observations of reflectivity Z^{obs} and Doppler velocity V^{obs} are generated from the model reflectivity Z, wind components u, v, and w, and hydrometeor fall speeds w_t using a slightly modified version of the Wood et al. (2009) radar emulator. This technique simulates the power-weighted averaging of radial velocities and reflectivities of scatterers within a Gaussian radar beam, and accounts for earth curvature and standard atmospheric beam refraction in computing the beam path. The same hydrometeor fall speed formula is used in the calculation of the V^{obs}, the V^{obs} forward operator in the EnKF, and the model: $w_t = -2.6Z^{.107}\left(1.2/\rho_{sim}\right)^{0.4}$, where ρ_{sim} (kg m^{-3}) is the height-varying base state air density in the simulation and Z is given in mm^6 mm^{-3} (Joss and Waldvogel 1970). Reflectivity observations < 0 dBZ are set to 0 dBZ to imitate the common practice of treating missing or very low reflectivities as "no-precipitation" observations to suppress spurious convection in the ensemble (Dowell et al. 2004; Tong and Xue 2005; Aksoy et al. 2009). To emulate the lack of radial velocity data in regions of low signal-to-noise ratio, V^{obs} are only computed in regions with Z^{obs} > 5 dBZ. Random errors having 2 m s^{-1} (3 dBZ) standard deviation are added to the V^{obs} (Z^{obs}).

In all of our experiments, the simulated supercell is observed by two stationary radars having characteristics consistent with the WSR-88D network. Volume coverage pattern 11 (VCP-11) is used, with successively higher groups of (2-3) sweeps computed from model fields valid at successively later (in 1-min increments) simulation times. At the lowest two sweeps, super-resolution data (currently available in the WSR-88D network for visualization only) are emulated; legacy-resolution observations are generated at steeper elevation angles (Table 1). The positions of the emulated radars and the approximate paths of the main low-level updraft of each of the three supercells during the data assimilation period are depicted in Fig. 4. One radar is fixed at the same location, ~130 km east-southeast of the supA low-level updraft at t = 40 min, in all of the experiments. The second radar is repositioned from experiment to experiment to investigate the sensitivity of the LLM forecasts to the radar-storm geometry. Experiments are labeled according to the supercell being forecast, the approximate distance of the second radar from the low-level updraft at t = 40 min, and the approximate cross-beam angle of the two radars at the same time and location (e.g., '130km_CBA70_supA'). Forecasts are initialized at t = 50 min (after 30 min of data assimilation) or t = 70 min (after 50 min of data assimilation) and labeled accordingly (e.g., '130km_CBA70_supA_50min'). The t = 70 min forecasts ideally benefit from the additional 20 min of radar data assimilated, but at the cost of reduced forecast lead time.

2.3. Ensemble analysis-forecasting system

Initial conditions for the ensemble forecasts are obtained from assimilating the emulated radar observations using the NCOMMAS ensemble square root filter (based on the filter of Whitaker and Hamill 2002). Eighty ensemble members are used; repeating the 130km_CBA70_supA_50min forecast (Section 3a) with a 120-member ensemble did not substantially improve the results. The NWP model used to advance the ensemble members to each successive assimilation time is equivalent to that used to generate the truth simulations, except that Δ_H is increased from 1/3 km to 1 km. The covariance localization factor is calculated using the Gaspari and Cohn (1999) correlation function with covariance estimation cutoff radii of 6 km and 3 km in the horizontal and vertical directions, respectively. To imitate the practice of accounting for uncertainty in the sounding, and to mitigate ensemble underdispersion at higher altitudes, perturbations are added to the base-state u and v of each ensemble member following the procedure of Potvin et al. (2012). The perturbations are computed by generating random sinusoidal perturbations of the form used in Aksoy et al. (2009), then scaling them such that their standard deviation at each level is a fraction (= .025 in this study) of the base-state wind speed multiplied by exp(z/22), where z is the model level height (km). Ellipsoidal thermal bubbles having random sizes and magnitudes are inserted in each member at t = 0 to initiate storms. The bubbles are randomly positioned within a 40 × 40 × 1.5 km box centered on the location of the initiation bubble in the truth simulation. The ensemble members are then integrated 20 mins forward to the beginning of the data assimilation period (t = 20 min). This allows physically realistic covariances to develop in the ensemble, thus maximizing the utility of radar data early in the assimilation period (e.g., Snyder and Zhang 2003; Dowell et al. 2004).

Prior to assimilation, observations are analyzed to a quasi-horizontal grid on each conical scan surface (e.g., Dowell et al. 2004; Dowell and Wicker 2009) using Cressman interpolation. Observations from the radars further from the storms ($x \geq 145$ km; Fig. 4) are interpolated to 2 km grids using a Cressman radius of 1.5 km. Observations from the radars closer to the storms are interpolated to 1 km grids using a Cressman radius of 1.0 km. To account for storm motion between the times at which observations are valid and the times at which they are analyzed, the interpolated observations are shifted to locations determined by the estimated storm translational velocity components U and V. The U and V (= 13 m s^{-1} and 4 m s^{-1}, respectively) are treated as constants in space and time and were determined by visually tracking features in the Z^{obs} field. Observations are assimilated every two minutes using a two-minute window centered on t. As in many EnKF radar data assimilation studies, to reduce computational cost, the observation operator H trilinearly interpolates model fields to observational locations, and thus makes no provision for the shape of, nor inhomogeneous reflectivity distribution within, the radar beam (Thompson et al. 2012 showed these simplifications do not severely degrade EnKF analyses and subsequent forecasts). Following Dowell and Wicker (2009), to save computational time, observations are not used to update π and K_m since the impact of the observations on these variables is negligible. Observational error standard deviations of 2 m s^{-1} and 5 dBZ are assumed in the filter. As in Potvin and Wicker (2012), we used larger filter-assumed Z^{obs} errors (5 dBZ) than were actually added to the Z^{obs} (3 dBZ) to reduce the impact of errors in the forecast hydrometeor fields and Z^{obs} operator. In experiments with both radars located roughly equidistantly from the storm, Z^{obs} are assimilated only from radar # 1 (Fig. 4) since Z^{obs} from the second radar would contain little independent information.

A procedure similar to the additive noise method (Dowell and Wicker 2009; based on the ensemble initialization procedure of Caya et al. 2005) is used to maintain ensemble spread consistent with the ensemble forecast error variance. Smoothed perturbations having horizontal and vertical length scales of 4 km and 2 km, respectively, are added to u, v, θ, and dewpoint temperature T_d below z = 10 km wherever Z^{obs} > 20 dBZ during the data assimilation. Prior to being smoothed, the u, v, θ and T_d perturbations have standard deviations of 2 m s^{-1}, 2 m s^{-1}, 1 K and 1 K, respectively. Time-height plots of V^{obs} consistency ratio and mean forecast innovation valid where Z^{obs} > 10 dBZ (not shown) suggest sufficient ensemble spread was obtained in all of our experiments.

2.4. Forecast verification

Our evaluations of the LLM ensemble forecasts focus on estimates of the peak azimuthal-mean vortex-maximum tangential velocity averaged over the z = 0.5-1.5 km layer. This parameter, V_T, is computed at each time within the evaluation period (t = 50-120 min) for both the truth simulations and the ensemble member forecasts using the following procedure:

1) Vertically average ζ over z = 0.5-1.5 km, yielding $\overline{\zeta}$ (x, y).

2) Determine the location of the maximum $\overline{\zeta}$, (x_{max}, y_{max}).

3) For each grid coordinate (x_0, y_0) within 3 km of (x_{max}, y_{max}), compute the circulation, $\Gamma \equiv \sum_C \mathbf{V} \bullet d\mathbf{l}$, for a series of circles C centered on (x_0, y_0) with radius R alternately set to 1.0 km, 1.5 km, 2.0 km, 2.5 km and 3.0 km, where $d\mathbf{l}$ is the line element vector tangent to C at a given point, and $\mathbf{V}$ is the vertically averaged horizontal wind field valid over the same layer as $\overline{\zeta}$ (z = 0.5 km to ~1.5 km).

4) For each Γ, compute $v_t = |\Gamma / 2\pi R|$.

5) $V_T = max\{v_t\}$.

The degree to which each forecast replicates the timing and intensity of the true low-level rotation is assessed by comparing time series of the ensemble probability of V_T exceeding prescribed thresholds in the forecast supercell to time series of V_T in the "true" supercell (Figs. 3d-f). The skill with which each forecast replicates the path of the maximum low-level rotation is evaluated by comparing 3 × 3 point neighborhood ensemble probabilities (Schwarz et al. 2010) of the forecast-period-maximum V_T exceeding a threshold to the region where the forecast-period-maximum true V_T exceeds the same threshold. The temporal and spatial ensemble probabilities are labeled P_t and P_{xy}, respectively.

3. RESULTS

3.1. Both radars distant from storm

Within the WSR-88D network, supercells are often > 100 km from the nearest radars. In such cases, the planetary boundary layer, which is dynamically critical to supercell evolution, is largely unobserved. In addition, the resolution of radar observations is substantially reduced at such long ranges. To explore whether useful ensemble forecasts of potentially tornadic supercells can be achieved in such suboptimal circumstances, we performed experiments (130km_CBA70_supA, 130km_CBA65_supB, and 140km_CBA65_supC) with both emulated radars positioned > 100 km from the low-level updraft throughout the data assimilation period (Fig. 4). At these distances, the lowest (0.5°) beam from each radar is centered > 1.5 km above the ground. Thus, the development of accurate ensemble covariances between the model state variables above and below the data cutoff is critical to retrieving the low-level storm fields during the data assimilation. Favorable CBAs of roughly 60° to 90° obtain during the assimilation period for all three supercells; the impact of poor CBAs on forecasts of supA is examined in Section 3c.

The P_t and P_{xy} provide mixed, but overall positive, support to the ability of a warn-on-forecast system to predict low-level rotation in a supercell located within a gap in the low-level domain of the WSR-88D network. We first evaluate the supA forecast initialized at t = 50 min (130km_CBA70_supA_50min). Consistent with the brief LLM that occurs just after t = 60 min in the truth simulation (Fig. 3d), roughly 40 % of the ensemble members contain V_T > 10 m s^{-1} shortly after initialization (Fig. 5a). The forecast LLMs, however, generally form 5-10 mins too early (a possible reason for this is given in the next subsection) and, due partly to their premature development, are displaced southwest of the true LLM (Fig. 6a). In addition, there is only a weak signal in the forecast for the subsequent weakening of the initial LLM (Figs. 5a, 6a). The timing of the onset of sustained strong low-level rotation, on the other hand, is well forecast, with $P_t(V_T > 10 \text{ m s}^{-1})$ rapidly increasing beginning near t = 75 min, roughly coincident with the true V_T exceeding the same threshold. During the period of maximum true V_T, $t \approx$ 90-100 min, $P_t(V_T > 10 \text{ m s}^{-1})$ and $P_t(V_T > 15 \text{ m s}^{-1})$ average near 75 % and 30 %, respectively. The P_t for both thresholds decrease after $t \approx$ 105 min, consistent with, though slightly delayed from, the decline in true V_T after $t \approx$ 100 min. The $P_t(V_T > 10 \text{ m s}^{-1})$ remains high, generally > 60 %, through the end of the forecast period, consistent with the maintenance of $V_T > 10 \text{ m s}^{-1}$ in the true supercell.

The peak $P_{xy}(V_T > 10 \text{ m s}^{-1})$ are generally displaced several kilometers south of the true LLM track, but these errors are quite acceptable given the long forecast lead times. The swath of peak P_{xy} is generally centered within the envelope of LLM tracks, which is ≤ 30 km wide (along the direction perpendicular to the storm motion) through the forecast period. Given that current tornado warning boxes are generally ~20-30 km wide at 30 min lead times, a tornado warning polygon constructed to encompass the envelope of $P_{xy} > 0$ (a conservative approach) in this case would comfortably include the true LLM (and potential tornado) track without being unduly large. Computations of $P_{xy}(V_T > 10 \text{ m s}^{-1})$ for 10-min subintervals of the forecast period (Fig. 7a) show that the forecast LLM trajectory is reasonably accurate in time as well as in space. This suggests that warn-on-forecast ensembles will ultimately permit greater temporal resolution in tornado warnings.

As implied by the large width of the $P_{xy} > 0$ envelope relative to the true path of $V_T > 10 \text{ m s}^{-1}$, large variance exists among the individual low-level rotation forecasts (Fig. 8a). The differences between the forecasts are striking given the qualitative similarities between the member initial conditions (shown at low levels in Fig. 8b), and serve to underscore the chaotic nature of the phenomena being predicted. The large errors that occur in many of the individual member forecasts highlight the advantage of using an ensemble, rather than deterministic, forecast approach.

The 130km_CBA70_supA_70min forecast is superior to the 130km_CBA70_supA_50min forecast, an expected result of the larger number of radar volumes assimilated and the shorter forecast lead times. The timing of the onset of $V_T > 15 \text{ m s}^{-1}$ is better captured, as is the decrease in V_T after t = 100 min (Fig. 5a). After t = 75 min (when the true V_T exceeds 10 m s^{-1}), the $P_t(V_T > 10 \text{ m s}^{-1})$ is also better (Fig. 5a). The P_{xy} swatch is considerably narrower at later times than for 130km_CBA70_supA_50min, and the maximum P_{xy} are substantially larger (cf. Figs. 6a,b). These results support the expectation that warn-on-forecast ensemble output will be valuable not just to issuing tornado warnings, but also to refining existing warnings as newer forecasts become available.

We now turn to evaluating the 130km_CBA65_supB forecasts (recall that supB is the left-moving counterpart to supA; Section 2a). Consistent with the lower V_T in supB (cf. Figs. 3a, b), the P_t are much smaller than for the supA forecasts (cf. Figs. 5a, b; cf. Figs. 5c, d). The rapid increase in the true V_T to above 5 m s^{-1} is reasonably well captured by the $P_t(V_T > 5 \text{ m s}^{-1})$ in both 130km_CBA65_supB_50min and (especially) 130km_CBA65_supB_70min. The decrease in V_T after t = 100 min, however, is not reflected in either forecast.

Plots of $P_{xy}(V_T > 5 \text{ m s}^{-1})$ indicate that, as in the supA forecasts, the supB forecast LLM tracks are displaced only several kilometers from the true LLM track (Figs. 9a,b). Moreover, the envelope of $P_{xy} > 0$ again encompasses the true LLM path. While the $P_{xy} > 0$ envelope is substantially wider than in the 130km_CBA70_supA forecasts, using a slightly less conservative criterion, such as $P_{xy} > 0.1$, defines a much narrower tornado risk area that still includes the true LLM path. Thus, both the supA and supB forecasts effectively outline the region of greatest tornado risk. Moreover, as with 130km_CBA70_supA_50min, 130km_CBA65_supB_50min accurately predicts the timing in addition to the path of the LLM (Fig. 7b), further suggesting that warn-on-forecast ensemble output may permit enhanced temporal information in warnings. Also consistent with the 130km_CBA70_supA forecasts, 130km_CBA65_supB_70min is substantially better than 130km_CBA65_supB_50min (Fig. 5b; cf. Figs. 9a,b).

In the supC simulation, intense low-level rotation is absent for most of the forecast period, with V_T generally remaining well below 10 m s^{-1} (except near t = 70 min; Figs. 3c,f). This is therefore a suitable null case test for our ensemble system. Unfortunately, the peak $P_t(V_T > 10 \text{ m s}^{-1})$ is substantially higher in 140km_CBA65_supC_50min than in 130km_CBA65_supB_50min (cf. Figs. 5a,c) despite supB exhibiting stronger low-level rotation and supC never actually exceeding the $V_T = 10 \text{ m s}^{-1}$ threshold (Fig. 3f). In addition, while the true V_T generally remains below 5 m s^{-1} after t = 80 min, the 140km_CBA65_supC_50min $P_t(V_T > 5 \text{ m s}^{-1})$ ranges between 60 % and 80 % during the same period. Thus, based solely on the P_t, supC would be regarded as a greater tornado threat than supB, despite supC never developing a distinct LLM. The overprediction of the low-level rotation in supC is also starkly reflected in the $P_{xy}(V_T > 5 \text{ m s}^{-1})$ and $P_{xy}(V_T > 10 \text{ m s}^{-1})$ plots (Figs. 10a, e). These results raise concerns about the reliability of warn-on-forecast low-level rotation guidance in null cases, particularly with respect to false alarms. These concerns are enhanced by the fact that the results of our idealized forecasts are likely more accurate than would typically be obtained in practice for similar cases (Section 1). Fortunately, the overestimation of V_T is subtantially mitigated in the t = 70 min forecast, particularly with respect to $V_T > 10 \text{ m s}^{-1}$ (Fig. 5c; cf. Figs. 10a,b; cf. Figs. 10e,f).

3.2. One radar close to storm, one distant from storm

The analyses and subsequent forecasts in the above experiments are hindered by the absence of radar data over the lowest 1.5 km of the storms and the relatively coarse resolution of the assimilated observations, both of which result from the large distances between the supercells and both radars. While that scenario is common, the WSR-88D network is sufficiently dense that storms are often located relatively close to one radar. A set of experiments was therefore performed (40km_CBA70_supA, 35km_CBA60_supB, and 50km_CBA65_supC) in which radar #2 was relocated closer to the supercells (Fig. 4).

In the supA forecasts, much of the southward bias in the LLM track disappears, and the peak P_{xy} increases relative to the original forecasts (cf. Figs. 6a, c and Figs. 6b, d). In addition, the $P_t(V_T > 10\ \mathrm{m\ s^{-1}})$ and $P_t(V_T > 15\ \mathrm{m\ s^{-1}})$ are larger than in the original forecasts during the peak in the true V_T, and the subsequent decline in V_T is better captured (Figs. 3d, 11a,d). Comparisons of surface θ', surface divergence and 1 km AGL reflectivity fields from the EnKF mean analyses and individual ensemble members (not shown) from 130km_CBA70_supA and 40km_CBA70_supA reveal that all the fields are generally slightly better retrieved in the latter analysis (Fig. 12a). Perhaps the most important difference between the 130km_CBA70_supA and 40km_CBA70_supA analyses is that the analyzed surface RFD gust front (RFDGF), and thus the leading edge of the storm cold pool, is generally too far east and too meridionally oriented in many of the 130km_CBA70_supA member analyses. This bias was also found 1 km and 2 km AGL, but not at higher altitudes where observations were available (not shown). We speculate that the premature development of low-level rotation in many of the 130km_CBA70_supA_50min member forecasts (Section 3a) resulted from the analyzed RFDGF and associated regions of baroclinic (horizontal) vorticity generation and barotropic (vertical) vorticity generation and stretching having advanced too close to the low-level updraft (within which horizontal vorticity is tilted into the vertical and vertical vorticity is stretched) by the initialization time.

Moving radar #2 closer to the storms has a more varied impact on the supB forecasts than on the supA forecasts (Figs. 11b, e; cf. Figs. 9a,c; cf. Figs. 9b,d). On one hand, the P_t and P_{xy} (fortunately) increase when and where the true V_T is highest. On the other hand, the rapid increase of P_t is delayed by ~5-10 min relative to the original forecast and to the truth simulation, and in 35km_CBA60_supB_50min, the maximum P_{xy} is generally displaced eastward of the maximum true V_T (the maximum P_{xy} in 130km_CBA65_supB_50min was roughly collocated with the maximum true V_T). As a result, the $P_{xy}(V_T > 5\ \mathrm{m\ s^{-1}})$ envelope in 35km_CBA60_supB_50min excludes part of the region of true $V_T > 5\ \mathrm{m\ s^{-1}}$. These results are somewhat surprising given that the 35km_CBA60_supB_50min initialization appears mildly better than the 130km_CBA65_supB_50min initialization (Fig. 12b). Fortunately, both P_t and (especially) P_{xy} are improved in 35km_CBA60_supB_70min relative to 35km_CBA60_supB_50min (cf. Figs. 11c,f; cf. Figs. 10e,f; cf. Figs. 10g,h). Overall, however, neither the t = 50 min nor t = 70 min forecasts benefit substantially from the greater proximity of radar #2.

In the case of supC, the decreased distance to radar #2 substantially *degrades* the t = 50 min forecast (Fig. 11c; cf. Figs. 10a,c; cf. Figs. 10b,d; cf. Figs. 10e,g; cf. Figs. 10f,h). While the $P_t(V_T > 5\ \mathrm{m\ s^{-1}})$ and $P_{xy}(V_T > 5\ \mathrm{m\ s^{-1}})$ are now much larger within the spatiotemporal window of true $V_T > 5\ \mathrm{m\ s^{-1}}$, they are also much larger at subsequent times/locations along the storm path. Moreover, the $P_t(V_T > 10\ \mathrm{m\ s^{-1}})$ and $P_{xy}(V_T > 10\ \mathrm{m\ s^{-1}})$ are substantially increased, despite the fact that the true $V_T < 10\ \mathrm{m\ s^{-1}}$ at all times.

Visual comparison of the EnKF means and individual member fields at t = 50 min (not shown) reveals the RFDGF is analyzed slightly too far east in 140km_CBA65_supC, and slightly too far west in 50km_CBA65_supC. Perhaps as a consequence of this, many of the 50km_CBA65_supC_50min member forecasts, but not the 140km_CBA65_supC_50min forecasts, erroneously delay the undercutting of the updraft by the cold pool (not shown). As a result, the period of vertical vorticity generation is erroneously prolonged in the 50km_CBA65_supC_50min forecast, which presumably explains the overprediction of V_T.

Fortunately, the t = 70 min supC forecast is not degraded overall by the greater proximity of radar #2 (cf. Fig. 11f; cf. Figs. 10b,d; cf. Figs. 10f,h). As a result, 50km_CBA65_supC_70min improves substantially upon 50km_CBA65_supC_50min. As was the case with supB, however, forecasts of supC do not appear to generally benefit from decreasing the distance to radar #2, despite the additional information content of the assimilated observations.

3.3. Impact of poor radar cross-beam angles

The 130km_CBA70_supA and 40km_CBA70_supA experiments (Section 3a) were repeated with the second radar relocated so as to maintain roughly the same distance from the storm while effecting much poorer CBAs (Fig. 4). In 130km_CBA25_supA, the CBA over the low-level updraft varies between ~20° and ~30° during the t = 20-70 min period. In 40km_CBA0_supA, the CBAs are particularly poor, varying from 30° to as low as 0° (in which case the wind component perpendicular to the radar baseline is totally unsampled). The P_t and P_{xy} for the ensemble forecasts initialized at t = 50 min and t = 70 min are presented in Figs. 13 and 14, respectively.

The impact of the poorer CBAs, rather than being consistently undesirable as might be expected, is mixed. Only minor differences occur between 130km_CBA25_supA_50min and 130km_CBA70_supA_50min (Fig. 13a; cf. Figs. 14a, 6a). From t = 80 to t = 95, the 130km_CBA25_supA_70min P_t are higher than in 130km_CBA70_supA_70min (Fig. 14b), a desirable result given the true $V_T > 15\ \mathrm{m\ s^{-1}}$ during that period (Fig. 3d). On the other hand, the subsequent rapid decrease in the 130km_CBA25_supA_70min $P_t(V_T > 15\ \mathrm{m\ s^{-1}})$ occurs too early, while that of 130km_CBA70_supA_70min comports well with the true V_T falling below 15 m s^{-1} around t = 100 min.

Turning to the forecasts with radar #2 positioned closer to the storm, the maintenance of true $V_T > 10\ \mathrm{m\ s^{-1}}$ after t = 100 min is much better signalled in 40km_CBA0_supA_50min than in 40km_CBA70_supA_50min (Fig. 13c; cf. Figs. 14c, 6c), as is the timing of the onset of $V_T > 15\ \mathrm{m\ s^{-1}}$ (Fig. 13c). The decline of V_T below 15 m s^{-1}, however, is better reflected in 40km_CBA70_supA_50min (Fig. 13c). The 40km_CBA0_supA_70min forecast is the only one that is substantially degraded by the poor CBAs. The $P_t(V_T > 10\ \mathrm{m\ s^{-1}})$ and $P_{xy}(V_T > 10\ \mathrm{m\ s^{-1}})$ are substantially lower in 40km_CBA0_supA_70min than in 40km_CBA70_supA_70min after t = 80 min (Fig. 13d; cf. Figs. 14d, 6d), during which the true $V_T > 10\ \mathrm{m\ s^{-1}}$. The $P_t(V_T > 15\ \mathrm{m\ s^{-1}})$ is also greatly reduced during this period, which is a desirable result after t = 105 min (when V_T falls below 15 m s^{-1}; Fig. 3d), but is inconsistent with the true $V_T > 15\ \mathrm{m\ s^{-1}}$ during 80 min < t < 105 min.

Despite the varied impacts of reducing the CBAs, two tentative conclusions can be drawn from the results. As exemplified in 40km_CBA0_supA_70min, very poor radar

CBAs can substantially limit the accuracy of the LLM forecasts. On the other hand, the results of the remaining three forecasts suggest slightly less narrow CBAs (20°-30°) do not necessarily introduce large errors. The latter conclusion is encouraging given that such CBAs are common within the WSR-88D network.

4. SUMMARY AND CONCLUSIONS

The OSSEs presented above provide tentative support to one of the primary hypotheses of the warn-on-forecast vision (Stensrud et al. 2009): that storm-scale ensemble forecast systems achievable in the near future will enable mean tornado warning lead times of 30 minutes or more. The most encouraging results were obtained in experiments where both emulated WSR-88D radars were positioned > 100 km (during the data assimilation period) from a pair of supercells (supA and supB) that later developed distinct LLMs. Despite the relatively coarse radar resolution and the absence of observations over the lowest 1.5 km of the atmosphere, the EnKF data assimilation system retrieved the boundary layer well enough for ensuing ensemble forecasts to effectively predict the development and evolution of the LLMs. SupA was correctly forecast to develop stronger low-level rotation than supB, and the timing of the onset of significant low-level rotation was predicted fairly well in both cases. In addition, the trajectories of both storms' LLMs were captured reasonably well. These results suggest that even in the common scenario where a supercell exists within a low-level gap in the WSR-88D domain, operationally useful probabilistic guidance can be obtained on the timing, path and magnitude of tornado risk. Moreover, additional experiments with supA indicated that while extremely narrow radar cross-beam angles may substantially degrade LLM forecasts (though not enough to render them useless), cross-beam angles as low as 20-30º may not be unduly detrimental. The latter result is broadly consistent with the OSSEs of Potvin and Wicker (2012), in which decreasing radar cross-beam angles from ~90º to ~30º had a relatively minor impact on EnKF analyses of a supercell wind field, but assimilating single-radar data introduced severe errors into dynamical analyses. The results of the present study suggest the frequently suboptimal radar-storm geometry within the WSR-88D network does not preclude useful numerical prediction of LLMs.

Forecasts of a supercell that failed to develop a distinct LLM (supC) were less successful than forecasts of supA and supB. The magnitude of low-level rotation was overpredicted during much of the forecast period; in fact, more ensemble members predicted the development of strong low-level rotation in supC than in supB. This result implies that mitigating the tornado warning false alarm rate may continue to be a significant challenge under the warn-on-forecast paradigm.

In experiments in which one of the radars was relocated to within 30-50 km of the storms, supA forecasts generally benefited from the higher-resolution observations and data availability nearer the ground. This result is consistent with previous studies (e.g., Dong et al. 2011; Schenkman et al. 2011; Snook et al. 2012) and motivates the installation of gap-filling (e.g., Collaborative Adaptive Sensing of the Atmosphere; McLaughlin et al. 2009) radars within the current WSR-88D network. The supB forecasts, however, generally did not improve when the radar was moved closer to the storm. Worse, the supC forecast initialized at t = 50 min was substantially *degraded*, with low-level rotation being even more overpredicted than in the forecast with both radars > 100 km away.

Fortunately, forecasts of all three supercells were substantially improved in all but once case (40km_CBA0_supA) when initialized at t = 70 min rather than t = 50 min, presumably due to both the better initial conditions (owing to the additional 20 mins of data assimilation) and the shorter forecast lead times. In instances where forecasts initialized at t = 50 min were degraded from moving one of the radars closer to the storm, initializing the forecasts at t = 70 min significantly mitigated those errors. To the extent that the improvements in the t = 70 min forecasts resulted from improved initial conditions, it is possible that methods for reducing the ensemble spinup time (e.g., "Running-In-Place", or RIP; Kalnay and Yang 2010) could substantially improve the t = 50 min forecasts. This would soften the tradeoff between increased forecast accuracy (later initialization) and increased forecast lead time (earlier initialization).

As explained in the introduction, the forecast results presented in these idealized experiments provide an estimate of the best-case scenario achievable in practice. It is plausible that if the above experiments with supA and supB were repeated for real storms in similar scenarios (i.e., comparable radar-storm geometries and storm environments), larger model errors could increase the spread and/or bias in the LLM path forecasts enough that a reasonably-sized tornado warning polygon based on that guidance would fail to encompass the true LLM path. Perhaps more concerning is the possibility that supercells that do not develop strong low-level rotation (as with supC) may pose an even greater false alarm risk than our idealized experiments suggest. To explore these possibilities, the OSSE framework presented herein could be extended to examine the impact of various model errors (including in the initial storm environment) on LLM forecasts. The results of those experiments would further clarify expectations for the performance of near-future warn-on-forecast systems, and potentially identify additional scenarios where numerical forecasts of low-level rotation may fare poorly. It would also be valuable to explore how much the forecast accuracy suffers under more unpredictable scenarios, such as when the supercell is strongly interacting with nearby storms or traversing a highly heterogeneous, poorly-sampled environment. Finally, forecast improvements from recent EnKF innovations, including asynchronous filters (Sakov et al. 2010; Wang et al. 2012) and the RIP method mentioned above, should be examined. The authors plan to pursue at least some of these lines of future work.

Acknowledgments. The first author was primarily supported by a National Research Council Research Associateship Award at the NOAA/National Severe Storms Laboratory. Additional funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA08OAR4320904, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the U.S.

Department of Commerce. We thank Thomas Jones for his comments on an earlier version of this paper.

REFERENCES

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824.

Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394–2416.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094.

Coniglio, M. C., D. J. Stensrud, and L. J. Wicker, 2006: Effects of upper-level shear on the structure and maintenance of strong quasi-linear mesoscale convective systems. *J. Atmos. Sci.*, **63**, 1231–1252.

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.

Dawson II, D. T., L. J. Wicker, E. R. Mansell, 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.

Dong, J., M. Xue, and K. Droegemeier, 2011: The analysis and impact of simulated high-resolution surface observations in addition to radar data for convective storms with an ensemble Kalman filter. *Meteorol. Atmos. Phys.*, **112**, 41-61.

Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005.

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

Hohenegger, C., and C. Schär, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478.

Joss, J., and A. Waldvogel, 1970: Raindrop size distribution and Doppler velocities. Preprints, 14th Conf. on Radar Meteorology, Tucson, AZ, Amer. Meteor. Soc.

Kalnay, E. and S.-C. Yang, 2010: Accelerating the spin-up of Ensemble Kalman Filtering. *Quart. J. Roy. Meteor. Soc*., **136**, 1644-1651.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21,** 289–307.

Maddox, R. A., J. Zhang, J. J. Gourley, and K. W. Howard, 2002: Weather radar coverage over the contiguous United States. *Wea. Forecasting*, **17**, 927-934.

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194.

Markowski, P. M., J. M. Straka, and E. N. Rasmussen, 2002: Direct surface thermodynamic observations within the rear-flank downdrafts of nontornadic and tornadic supercells. *Mon. Wea. Rev.*, **130**, 1692–1721.

McLaughlin, D., and Coauthors, 2009: Short-wavelength technology and the potential for distributed networks of small radar systems. *Bull. Amer. Meteor. Soc.*, **90**, 1797–1817.

Milbrandt, J. A., and M. K. Yau, 2006: A multimoment bulk microphysics parameterization. Part IV: Sensitivity experiments. *J. Atmos. Sci.*, **63**, 3137–3159.

Potvin, C. K., and L. J. Wicker, 2012: Comparison between dual-Doppler and EnKF storm-scale wind analyses: Observing system simulation experiments with a supercell thunderstorm. *Mon. Wea. Rev.*, in press.

Potvin, C. K., L. J. Wicker, D. Betten, M. I. Biggerstaff, and A. Shapiro, 2012: Comparison between dual-Doppler and EnKF storm-scale wind analyses: The 29-30 May 2004 Geary, Oklahoma, supercell thunderstorm. *Mon. Wea. Rev.*, conditionally accepted.

Sakov, P., G. Evensen, and L. Bertino, 2010: Asynchronous data assimilation with the EnKF. *Tellus*, **62**, 24–29.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, J. Gao, 2011: Impact of CASA radar and Oklahoma mesonet data assimilation on the analysis and prediction of tornadic mesovortices in an MCS. *Mon. Wea. Rev.*, **139**, 3422–3445.

Shabbott, C. J., and P. M. Markowski, 2006: Surface in situ observations within the outflow of forward-flank downdrafts of supercell thunderstorms. *Mon. Wea. Rev.*, **134**, 1422–1441.

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.

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

Stensrud, D. J., and J. Gao, 2010: Importance of horizontally inhomogeneous environmental initial conditions to ensemble storm-scale radar data assimilation and very short-range forecasts. *Mon. Wea. Rev.*, **138**, 1250–1272.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

Stensrud, D. J., and Co-authors, 2012: Progress and Challenges with Warn-on-Forecast. *Atmos. Res*., in press.

Thompson, T. E., L. J. Wicker, and X. Wang, 2012: Impact from a volumetric radar-sampling operator for radial velocity observations within EnKF supercell assimilation. *J. Atmos. Oceanic Technol*., **29**, 1417-1427.

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.

Trapp, R. J., G. J. Stumpf, and K. L. Manross, 2005: A reassessment of the percentage of tornadic mesocyclones. *Wea. Forecasting*, **20**, 680–687.

Wang, S., M. Xue, and J. Min, 2012: A four-dimensional asynchronous ensemble square-root filter (4DEnSRF) algorithm and tests with simulated radar data. *Quart J. Roy. Meteor. Soc*., in press.

Weisman, M. L., and J. B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504–520.

Weisman, M. L., and J. B. Klemp, 1984: The structure and classification of numerically simulated convective storms in directionally varying wind shears. *Mon. Wea. Rev.*, **112**, 2479–2498.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Wicker, L. J., and W. C. Skamarock, 2002: Time-splitting methods for elastic models using forward time schemes. *Mon. Wea. Rev.*, **130**, 2088–2097.

Wood, V. T., R. A. Brown, and D. C. Dowell, 2009: Simulated WSR-88D velocity and reflectivity signatures of numerically modeled tornadoes. *J. Atmos. Oceanic Technol.*, **26**, 876–893.

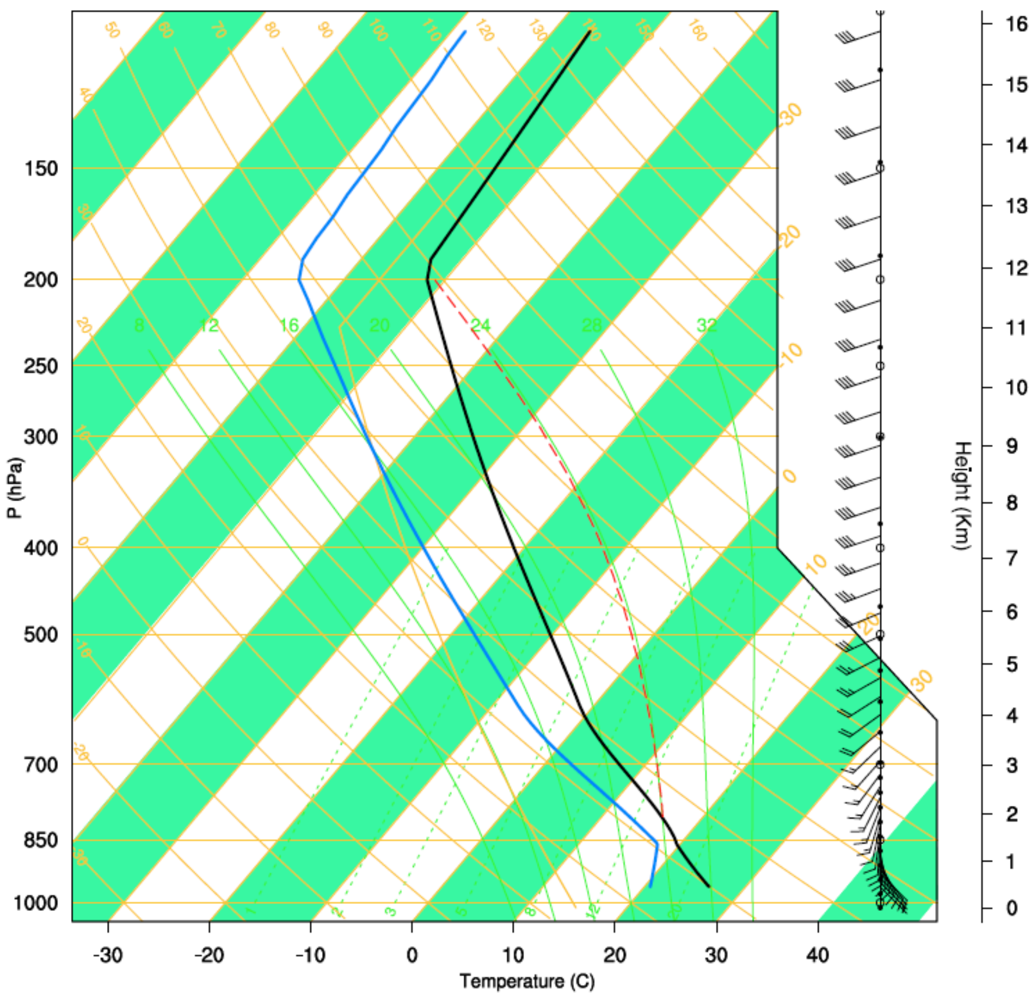
Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.,* **60**, 1173-1185.

Ziegler, C. L., 1985: Retrieval of thermal and microphysical variables in observed convective storms. Part 1: Model development and preliminary testing. *J. Atmos. Sci.*, **42**, 1487–1509.

Elevation angles (°)	**Effective half-power beamwidth**	V^{obs} **sampling intervals**	Z^{obs} **sampling intervals**
0.5, 1.5	1.03°	0.25 km × 0.5°	0.25 km × 0.5°
2.4, 3.4, 4.3, 5.3, 6.2, 7.5, 8.7, 10, 12, 14, 16.7, 19.5	1.39°	0.25 km × 1°	1 km × 1°

Table 1. Radar sampling characteristics for lowest two sweeps (super-resolution) and higher sweeps (legacy resolution).

(a)



(b)

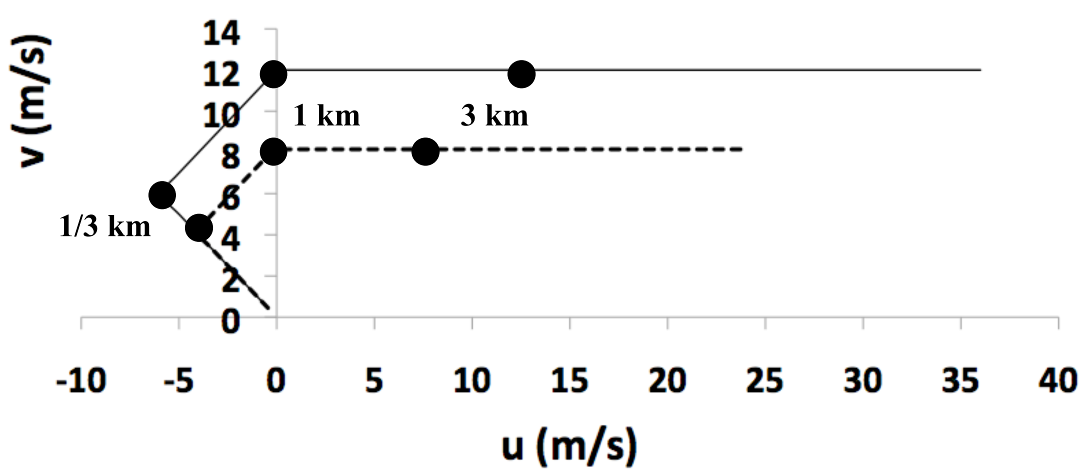


Fig 1. Model base state (a) thermodynamic profile used in both simulations, and (b) hodographs used in default (solid) and lower-shear (dashed) simulations. Heights (km) are indicated for three points on each hodograph.

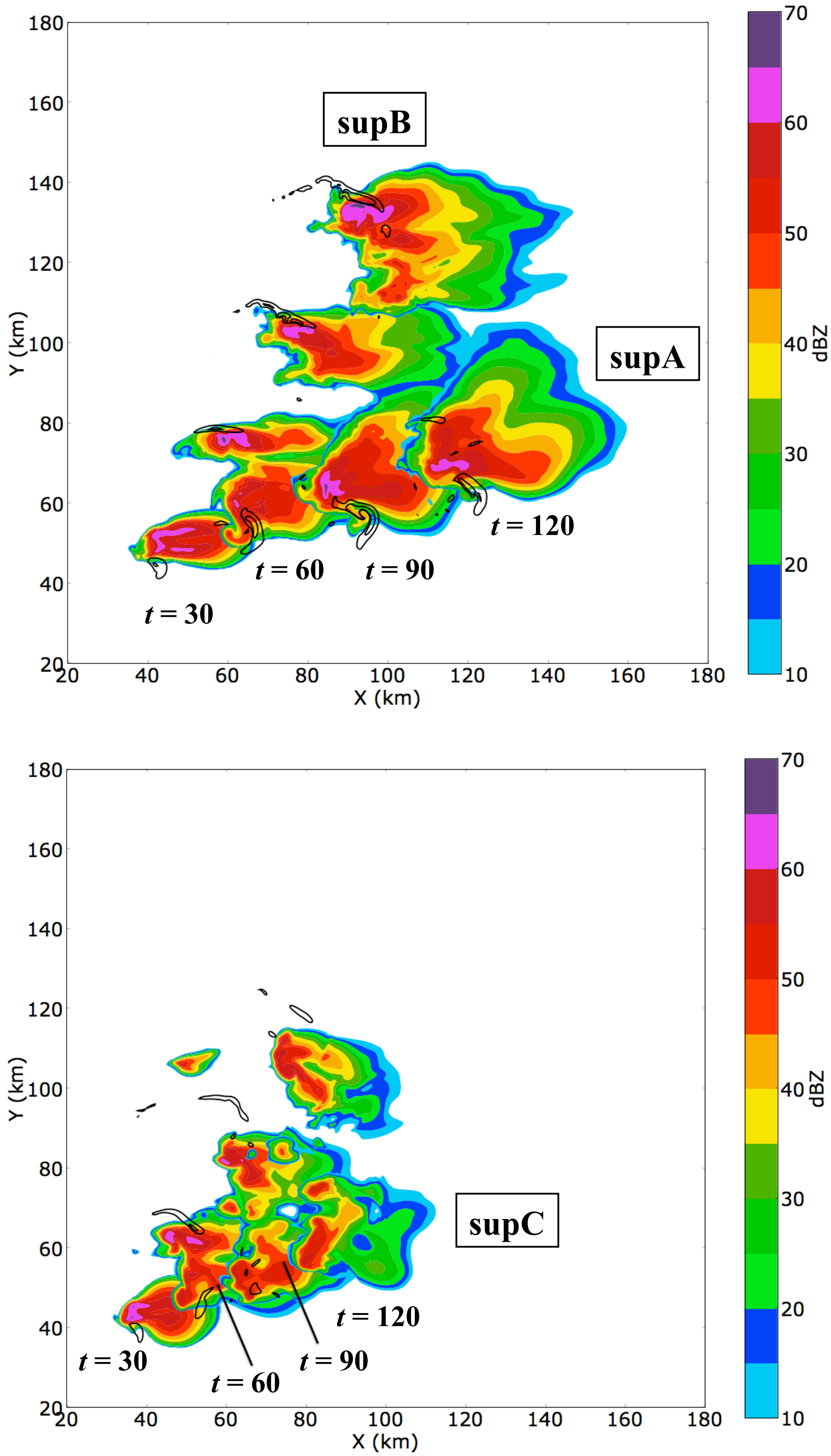


Fig. 2. Horizontal cross-sections of z = 1 km reflectivity (shading; dBZ) and w (contoured at 5, 10 m s^{-1}) at t = 30, 60, 90 and 120 min: (a) default simulation; (b) lower-shear simulation.

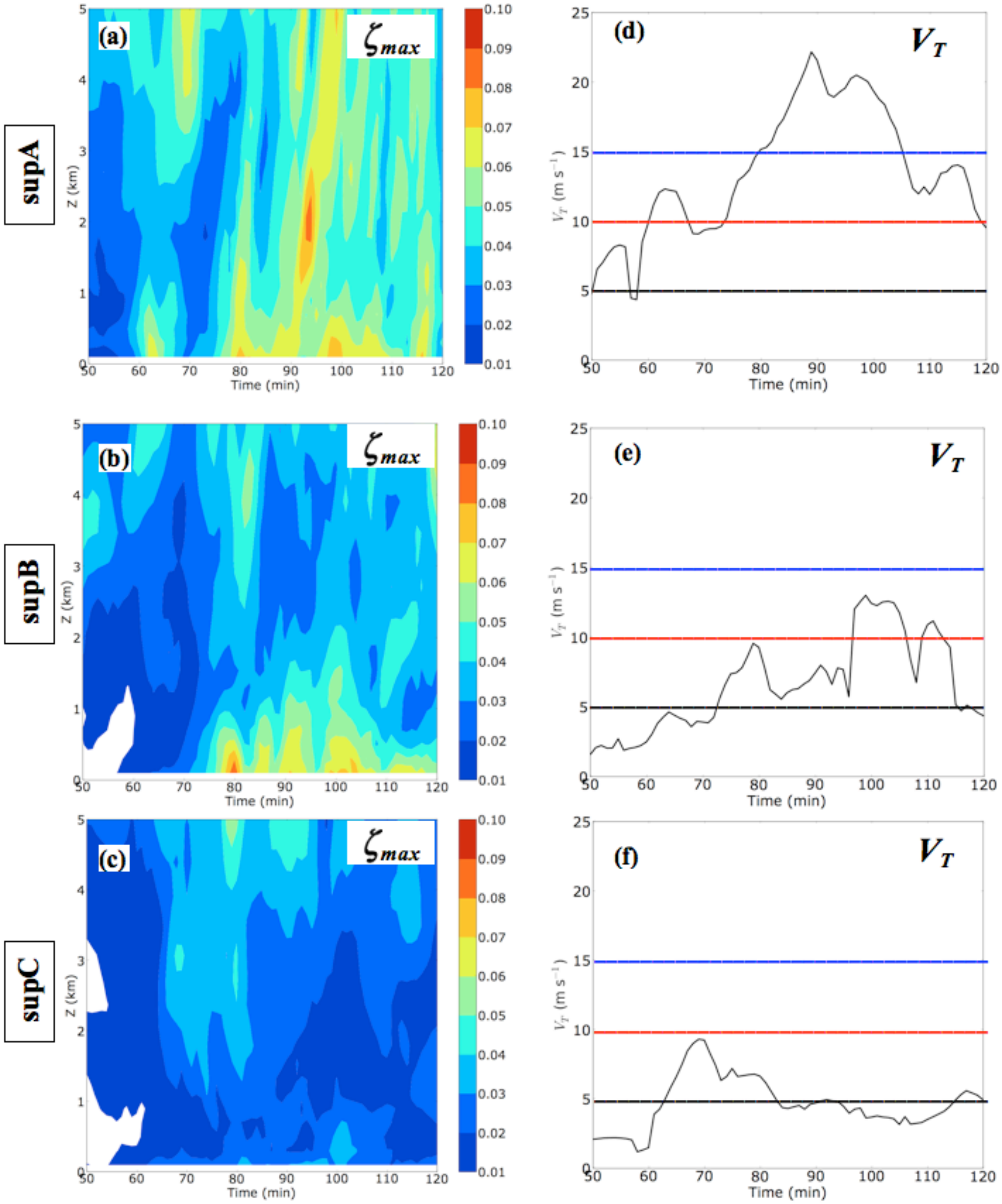


Fig. 3. Left panels: time-height plots of maximum-amplitude (a) cyclonic vertical vorticity in supA, (b) anticyclonic vertical vorticity in supB, and (c) cyclonic vertical vorticity in supC. Right panels: time series of V_T for (d) supA, (e) supB and (f) supC.

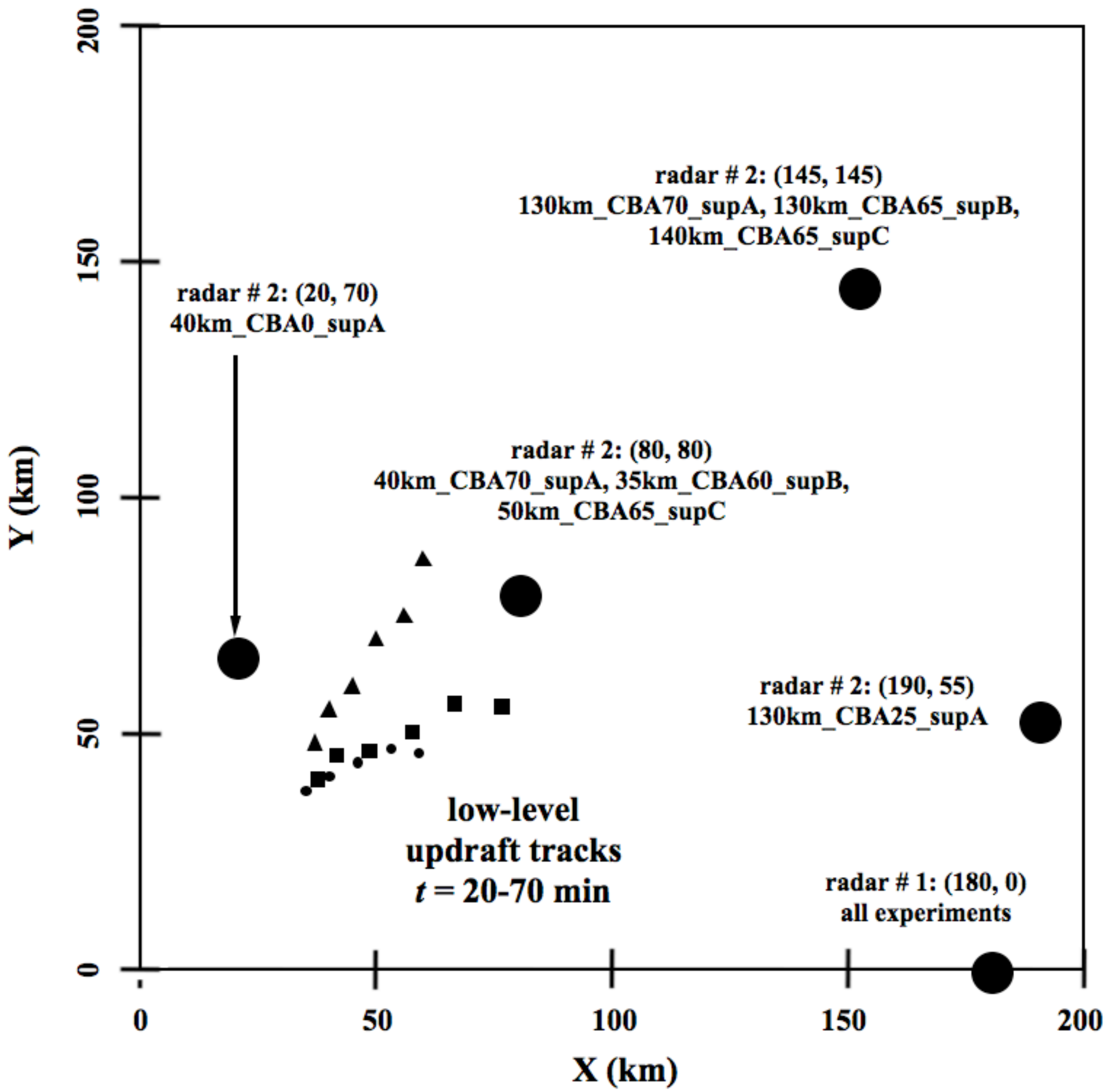


Fig. 4. Model domain used in truth simulation and EnKF experiments. The locations of the emulated radars are indicated by large dots. The *x*-*y* coordinates of each radar site relative to the southwest corner of the domain are listed in parentheses. The experiments in which each radar site is used are listed below the radar coordinates. Also shown are the paths of the low-level updrafts of supA (squares), supB (triangles) and supC (dots) during the data assimilation period (t = 20-70 min).

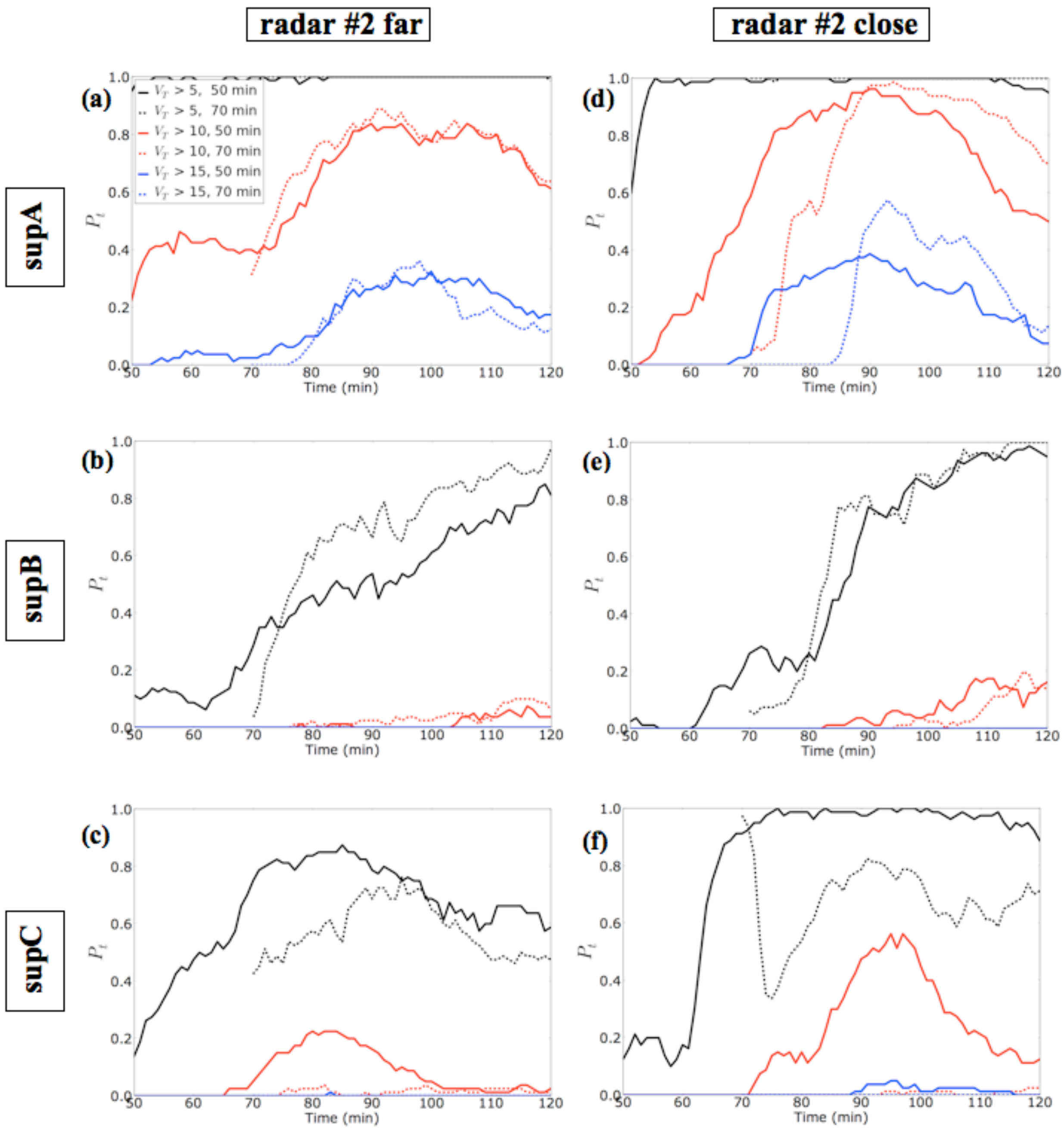


Fig. 5. P_t of $V_T > 5$ m s^{-1} (black), 10 m s^{-1} (red) and 15 m s^{-1} (blue) for forecasts initialized at t = 50 min (solid) and t = 70 min (dashed): (a) 130km_CBA70_supA, (b) 130km_CBA65_supB, (c) 140km_CBA65_supC, (d) 40km_CBA70_supA, (e) 35km_CBA60_supB, and (f) 50km_CBA65_supC. Note that $P_t(V_T > 15$ m s$^{-1}) = 0$ at all times in (b) and (e).

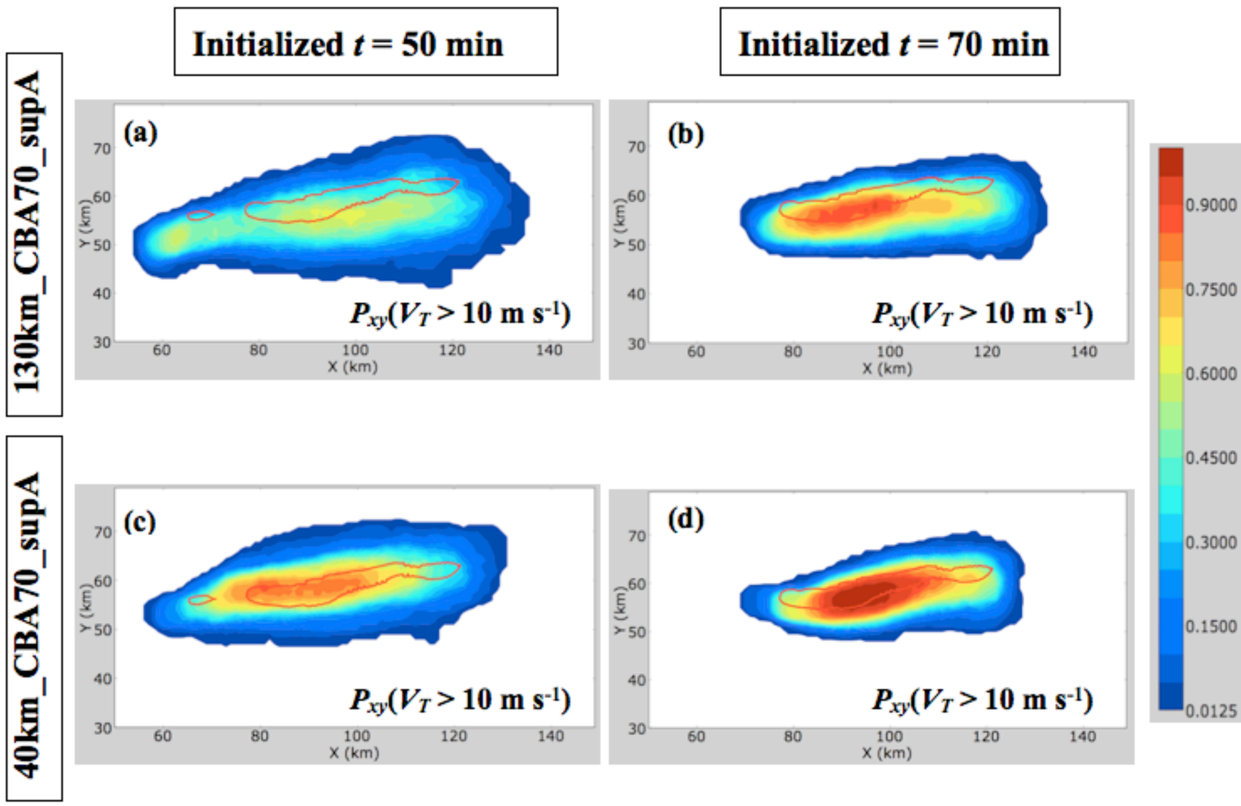


Fig. 6. $P_{xy}(V_T > 10\ \mathrm{m\ s^{-1}}$; shading) for forecasts initialized at t = 50 min (left panels) and t = 70 min (right panels): (a, b) 130km_CBA70_supA and (c, d) 40km_CBA70_supA. The red contours enclose the regions where the true $V_T > 10\ \mathrm{m\ s^{-1}}$ during the forecast period.

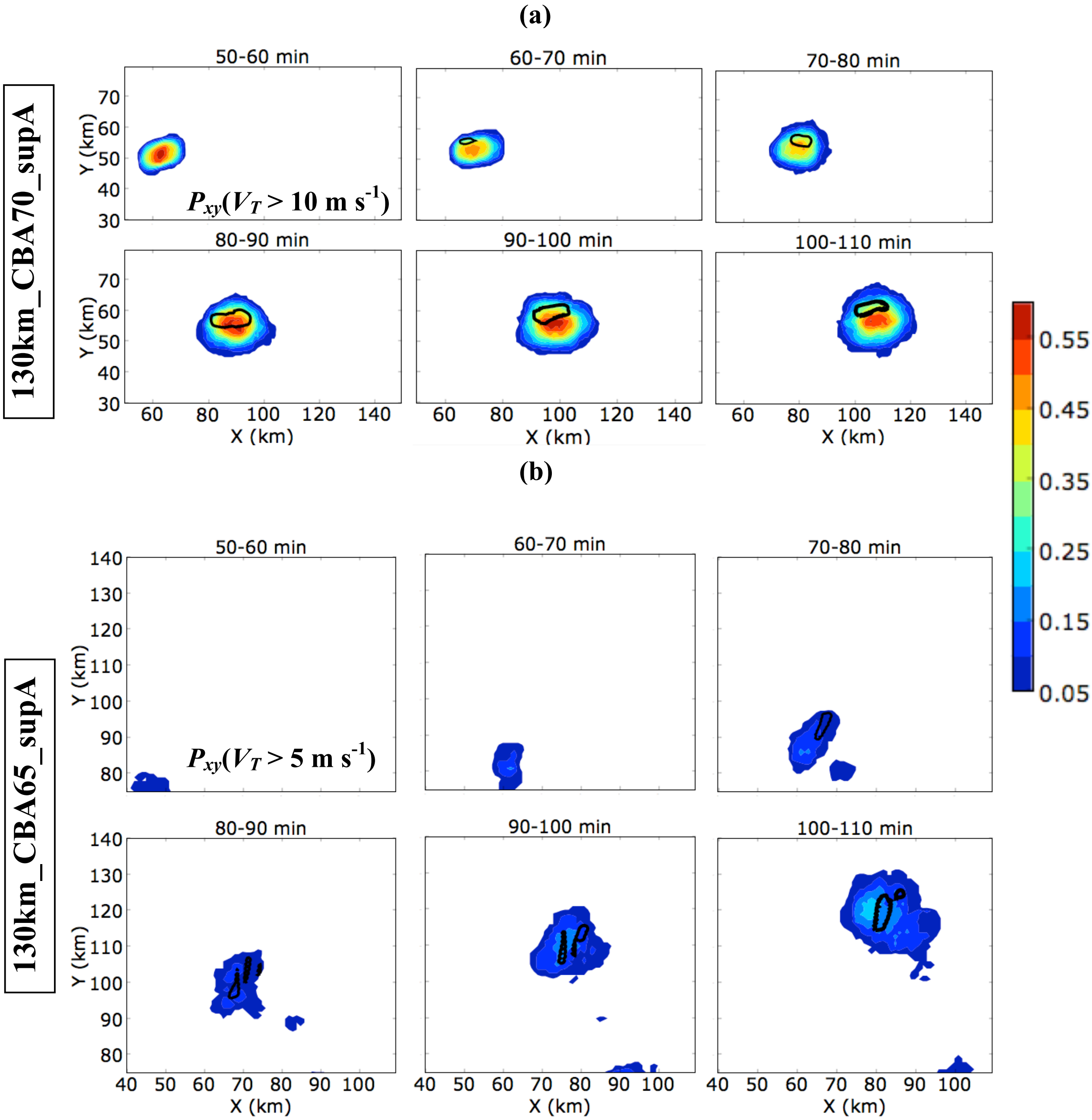


Fig. 7. P_{xy} valid over ten minute subintervals (shading) of (a) V_T > 10 m s^{-1} in 130km_CBA70_supA_50min and (b) V_T > 5 m s^{-1} in 130km_CBA65_supB. The black contours enclose regions of (a) true V_T > 10 m s^{-1} and (b) true V_T > 5 m s^{-1} during each subinterval.

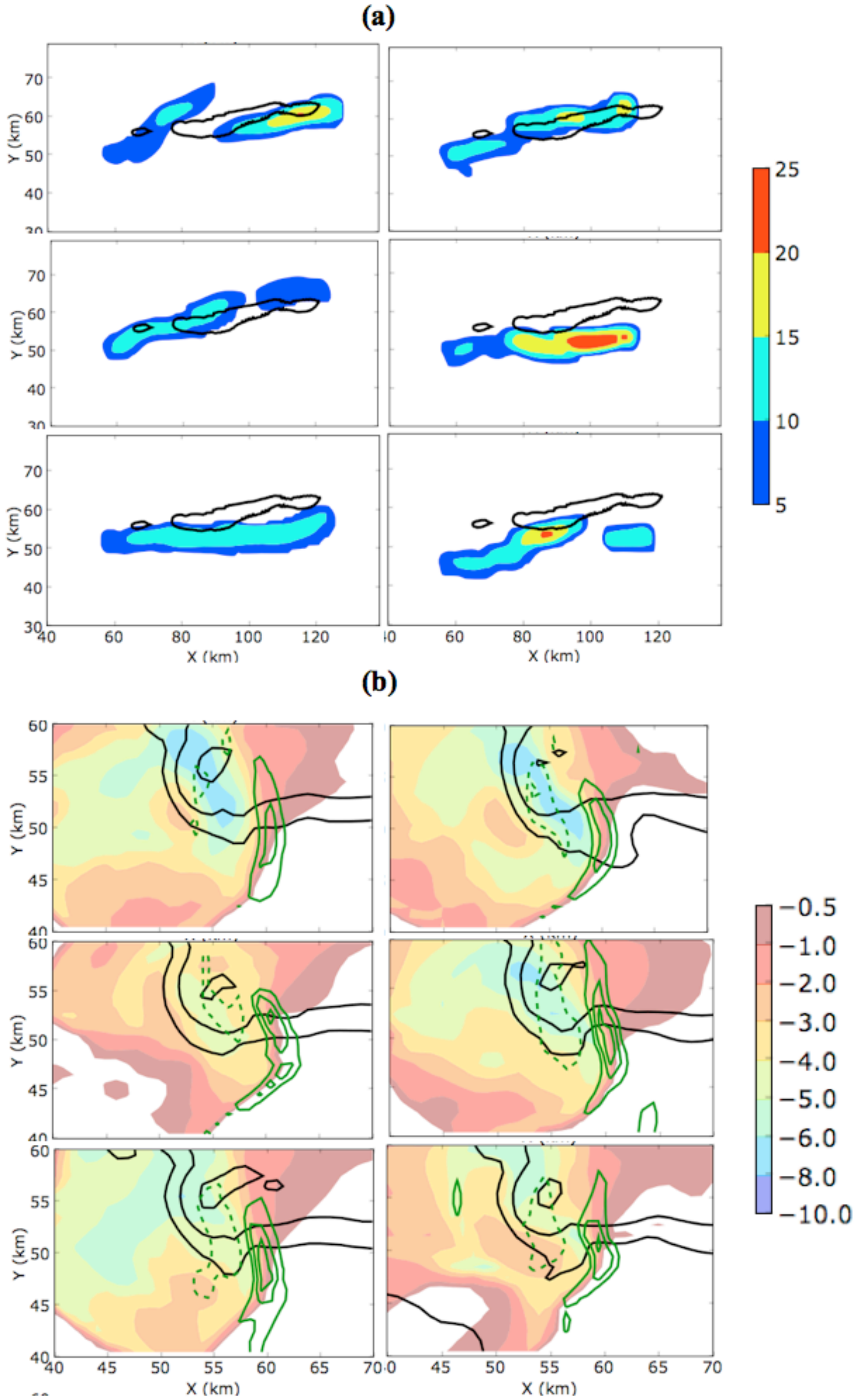


Fig. 8. (a) Forecast-period-maximum V_T (shading) for a representative subset of the 130km_CBA70_supA_50min member forecasts. The black contours enclose the regions of true V_T > 10 m s^{-1} during the forecast period. (b) Initial conditions of the member forecasts in (a): surface θ' (shading), surface convergence (green contours: -.015, -.010, -.005, .005, .01, and .015 s^{-1}), and 1 km AGL reflectivity (black contours: 20, 40 and 60 dB*Z*).

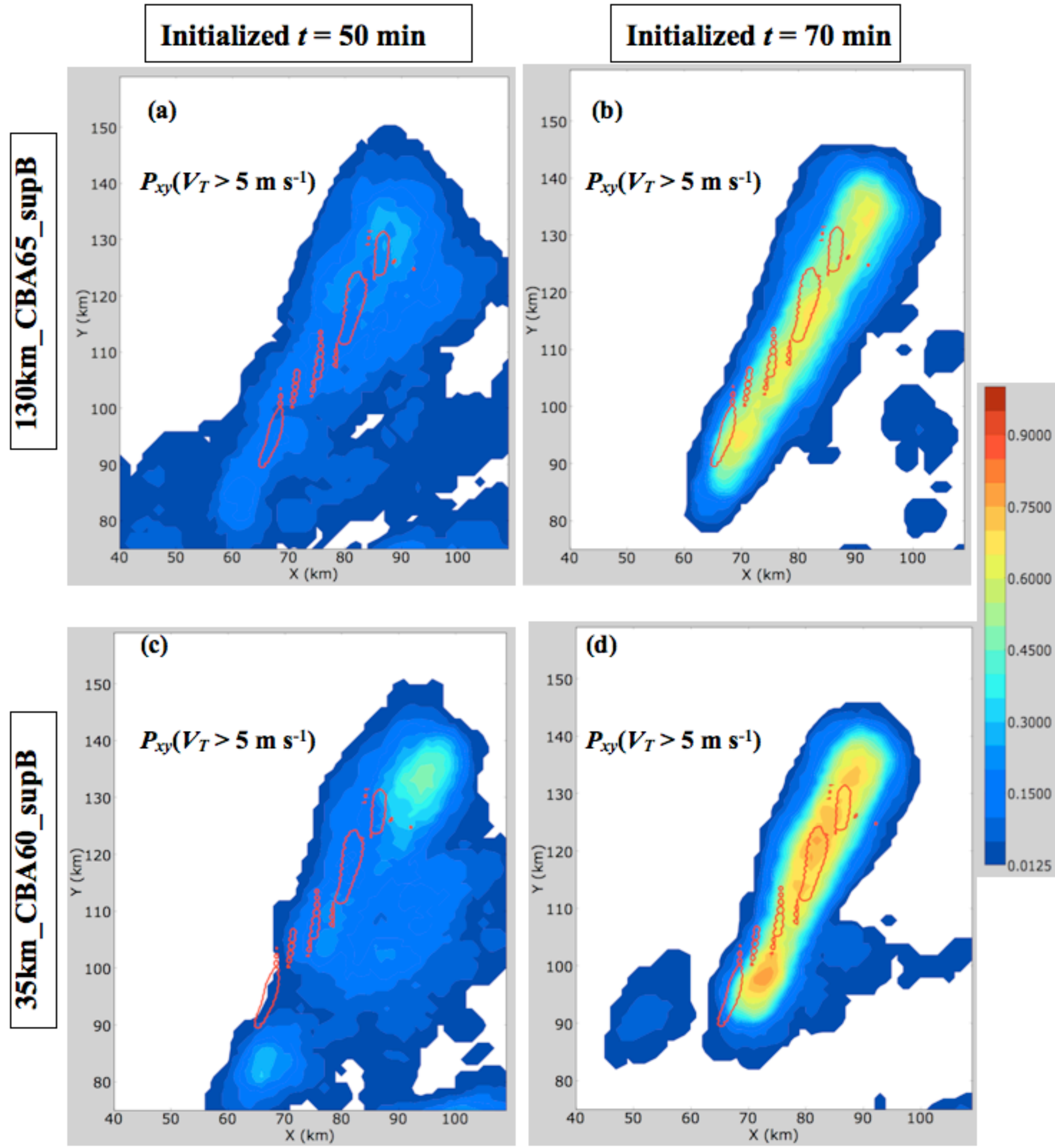


Fig. 9. $P_{xy}(V_T > 5$ m s^{-1}; shading) for forecasts initialized at t = 50 min (left panels) and t = 70 min (right panels): (a, b) 130km_CBA65_supB and (c, d) 35km_CBA60_supB. The red contours enclose the regions where the true V_T > 5 m s^{-1} during the forecast period.

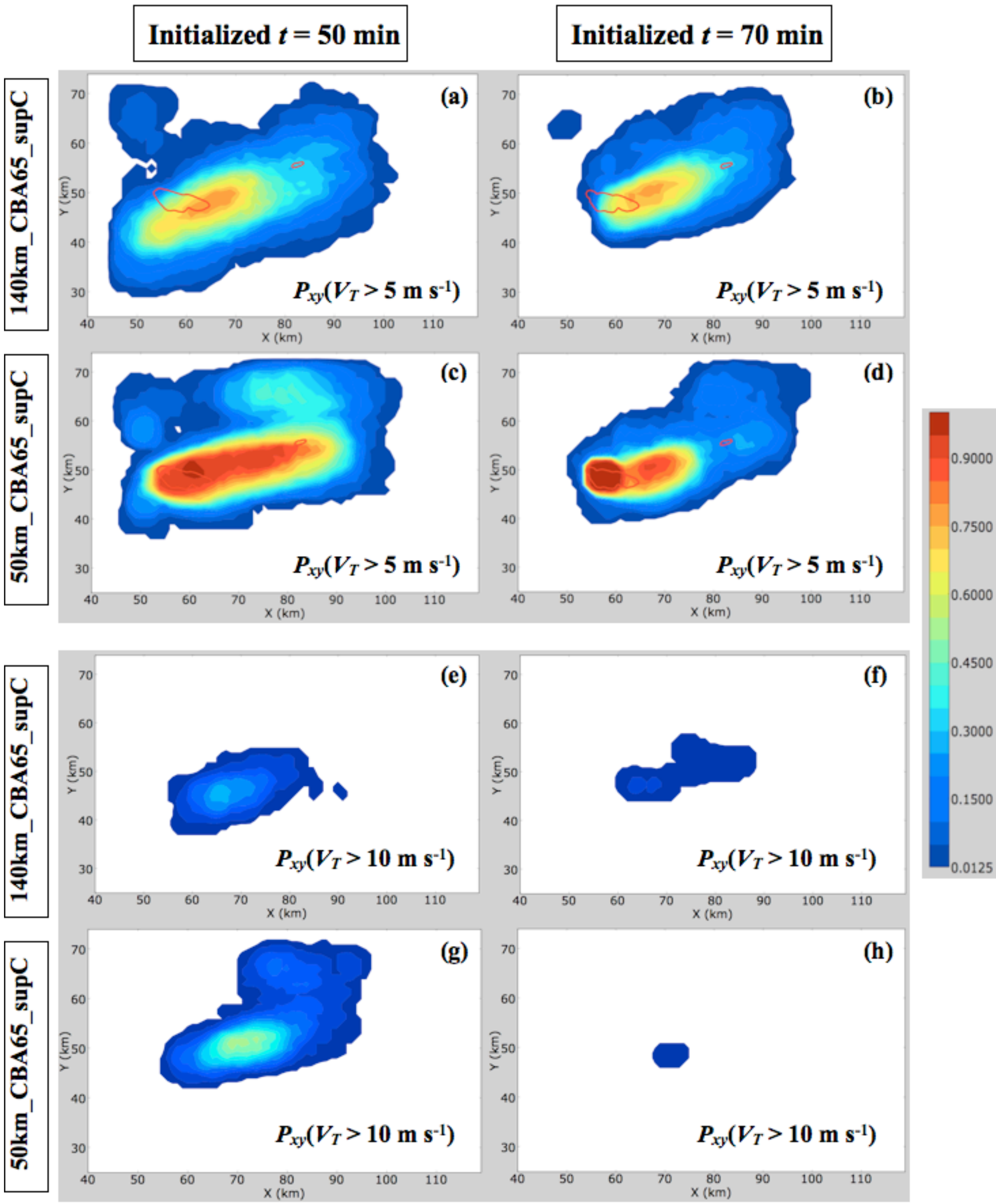


Fig. 10. P_{xy} (V_T > 5 m s^{-1}; shading) for forecasts initialized at t = 50 min (left panels) and t = 70 min (right panels): (a, b) 140km_CBA65_supC and (c, d) 50km_CBA65_supC. The red contours enclose the regions where the true V_T > 5 m s^{-1} during the forecast period. (e-h): Same as (a-d) but for V_T > 10 m s^{-1} (note that the true V_T never exceeds 10 m s^{-1}).

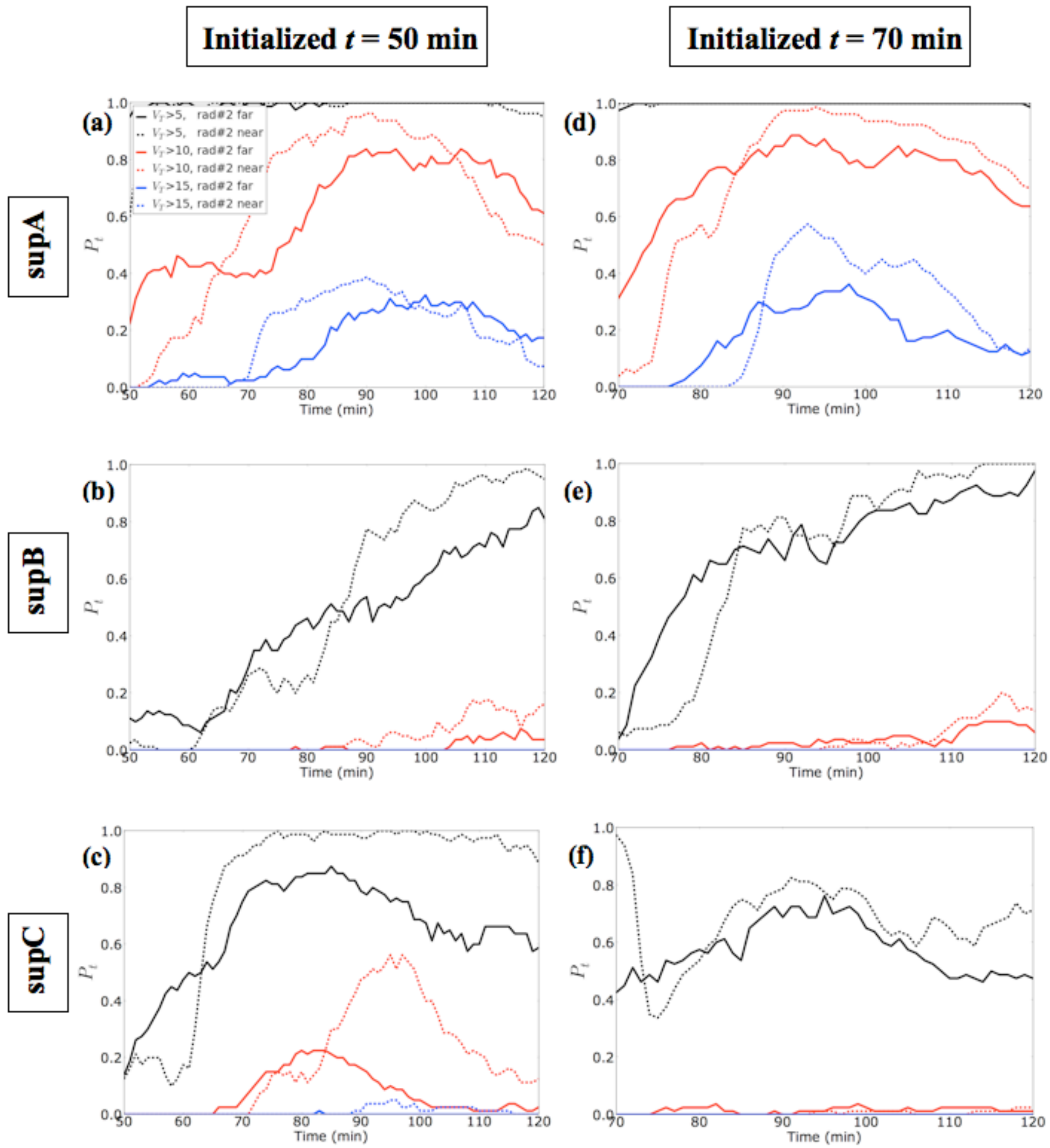


Fig. 11. P_t of $V_T > 5$ m s^{-1} (black), 10 m s^{-1} (red) and 15 m s^{-1} (blue) for forecasts initialized at t = 50 min (left panels) and t = 70 min (right panels): (a, d) 130km_CBA70_supA (solid) and 40km_CBA70_supA (dashed), (b, e) 130km_CBA65_supB (solid) and 35km_CBA60_supB (dashed), (c, f) 140km_CBA65_supC (solid) and 50km_CBA65_supC (dashed).

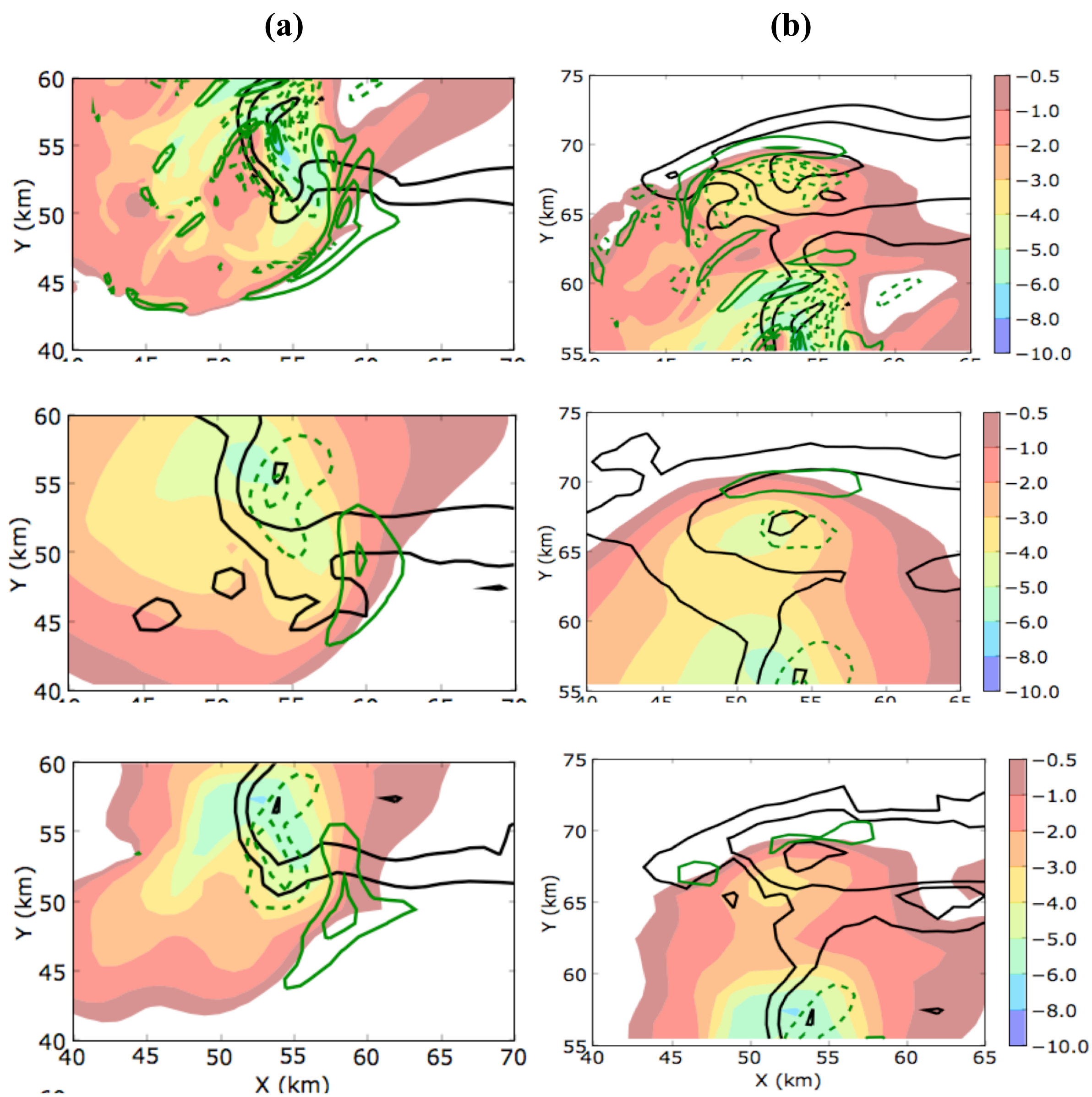


Fig. 12. Surface θ' (shading), surface convergence (green contours: -.015, -.010, -.005, .005, .01, and .015 s^{-1}), and 1 km AGL reflectivity (black contours: 20, 40 and 60 dBZ) for (a) supA and (b) supB analyses. (Top panels): truth; (middle panels): 130km_CBA70_supA_50min & 130km_CBA65_supB_50min analyses; (bottom panels): 40km_CBA70_supA_50min & 35km_CBA60_supB_50min analyses.

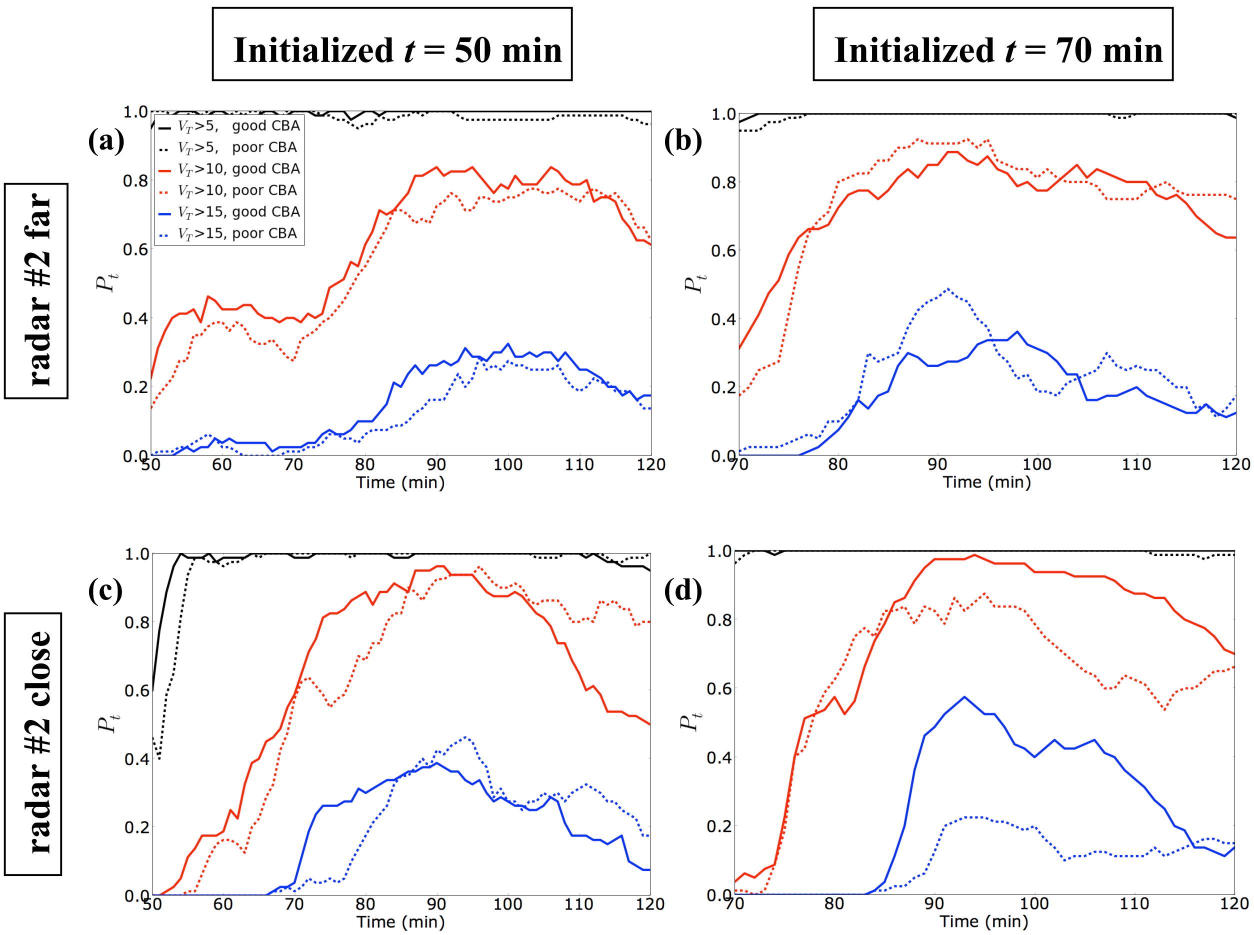


Fig. 13. P_t of (black) V_T > 5 m s^{-1}, (red) V_T > 10 m s^{-1} and (blue) V_T > 15 m s^{-1} for forecasts initialized at t = 50 min (left panels) and t = 70 min (right panels): (a, b) 130km_CBA70_supA (solid) and 130km_CBA25_supA (dashed); (c, d) 40km_CBA70_supA (solid) and 40km_CBA0_supA (dashed).

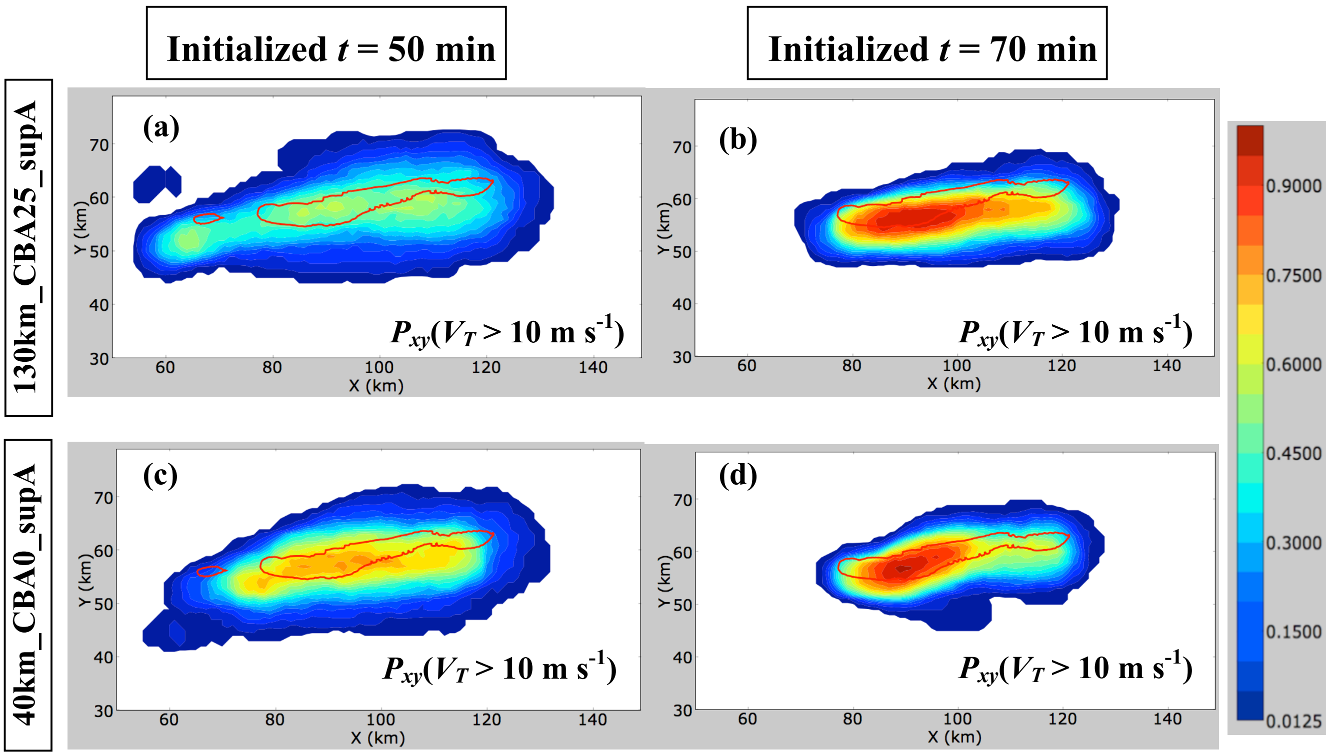


Fig. 14. $P_{xy}(V_T > 10$ m s^{-1}; shading) for forecasts initialized at t = 50 min (left panels) and t = 70 min (right panels): (a, b) 130km_CBA25_supA and (c, d) 40km_CBA0_supA.