

## Verifying Supercellular Rotation in a Convection-Permitting Ensemble Forecasting System with Radar-Derived Rotation Track Data

LOGAN C. DAWSON

*Purdue University, West Lafayette, Indiana*

GLEN S. ROMINE

*National Center for Atmospheric Research, Boulder, Colorado*

ROBERT J. TRAPP

*University of Illinois at Urbana-Champaign, Urbana, Illinois*

MICHAEL E. BALDWIN

*Purdue University, West Lafayette, Indiana*

(Manuscript received 23 June 2016, in final form 17 January 2017)

### ABSTRACT

The utility of radar-derived rotation track data for the verification of supercell thunderstorm forecasts was quantified through this study. The forecasts were generated using a convection-permitting model ensemble, and supercell occurrence was diagnosed via updraft helicity and low-level vertical vorticity. Forecasts of four severe convective weather events were considered. Probability fields were computed from the model data, and forecast skill was quantified using rotation track data, storm report data, and a neighborhood-based verification approach. The ability to adjust the rotation track threshold for verification purposes was shown to be an advantage of the rotation track data over the storms reports, because the reports are inherently binary observations whereas the rotation tracks are based on values of Doppler velocity shear. These results encourage further pursuit of incorporating observed rotation track data in the forecasting and verification of severe weather events.

### 1. Introduction

Convection-permitting modeling has become increasingly prevalent in research and operational numerical weather prediction (NWP) applications. With horizontal grid spacings on the order of a few kilometers, NWP models, like the Weather Research and Forecasting (WRF; Skamarock et al. 2008) Model, are capable of nominally resolving cloud-scale processes without using a cumulus parameterization (CP) scheme (e.g., Done et al. 2004; Weisman et al. 2008). Explicit simulation of mesoscale and convective processes provides qualitative information about the expected convective mode, which aids forecasters in anticipating predominant severe weather hazards (Kain et al. 2008; Weisman et al. 2008; Kain et al. 2010).

---

Corresponding author e-mail: Logan C. Dawson, ldawson@purdue.edu

Even with horizontal grid spacings of 1–4 km, convection-permitting forecasts are still too coarse to fully resolve smaller-scale convective features and severe weather phenomena such as mesocyclones and tornadoes. In an effort to overcome this limitation of grid resolution, various surrogates of these features and phenomena have been developed using grid-resolved model output fields. For example, a widely used surrogate for mesocyclones in supercell thunderstorms is a diagnostic quantity known as updraft helicity (UH; e.g., Kain et al. 2008; Carley et al. 2011; Sobash et al. 2011; Clark et al. 2012, 2013). Considering that supercell thunderstorms account for a large portion of severe weather events, especially large hail and tornadoes (e.g., Duda and Gallus 2010; Kain et al. 2008), a means of identifying simulated supercells in NWP forecasts has proven to be especially valuable.

Objective verification of supercell forecasts in convection-permitting models has two particularly

challenging issues. The first relates to the relatively small scale of predictands, like convective storms, and how even small errors in predicted location and timing severely degrade forecast skill in traditional measures of grid-scale model performance such as root-mean-square error. Discussed at length in Ebert (2008), fuzzy or neighborhood verification approaches have been developed to mitigate this verification dilemma. For example, such methods have been used to objectively verify quantitative precipitation forecasts produced from convection-permitting ensembles (e.g., Schwartz et al. 2010; Clark et al. 2010), as well as to verify storm surrogate forecasts against storm reports (Schwartz et al. 2015; Sobash et al. 2016). The second issue regards the lack of optimal observations. Severe thunderstorm occurrence is documented mostly through voluntarily reported observations of tornadoes, damaging winds, and large hail. These storm reports are known to possess numerous nonmeteorological biases (e.g., Brooks et al. 2003; Doswell et al. 2005). In addition to a known dependence on population density, reporting practices create uncertainty in estimating the intensity of these phenomena (Trapp et al. 2006; Verbout et al. 2006). Also, these discontinuous point observations fail to capture the full spatial extent of the severe weather hazards. Indeed, Sobash et al. (2011) conducted neighborhood-based verification of severe storm surrogates determined from UH forecasts using storm reports but noted “significant concerns” with using these observations.

A proposed alternative to storm report observations is the rotation tracks product developed by the National Severe Storms Laboratory (NSSL), which has recently been employed in the verification of low-level vertical vorticity forecasts by Skinner et al. (2016). Formally, the rotation track data used in this study are a measure of maximum azimuthal shear in the lowest 3 km of the atmosphere (Smith and Elmore 2004). These radar-derived data have been used to identify tracks of potentially severe storms (Miller et al. 2013). While these data are subject to the shortcomings of Doppler radar sampling (e.g., Wood and Brown 1997), the rotation tracks provide a dataset appropriate for verification that is not prone to the aforementioned nonmeteorological biases found in the storm reports database. Additionally, the rotation tracks are more continuous than storm reports and are more closely related to the severe storm surrogates for rotation in the model.

The increasing use of model-predicted severe storm surrogates in research and operational forecasting practices suggests the need to identify the most appropriate methods and datasets for surrogate verification. This study aims to show that rotation track data are comparable, and in many ways advantageous, to storm reports for objectively verifying supercell surrogates as predicted

via model diagnostics such as UH and maximum 0–3-km vertical vorticity. Because of the inherent predictability challenges of convective-scale NWP, a probabilistic approach to forecasting and verification is taken. An ensemble forecasting system is employed to produce retrospective forecasts of the severe weather that occurred on 19, 20, 30, and 31 May 2013. Neighborhood-based verification methods are applied to quantify and compare forecast skill when rotation track data and storm report data are used as alternate sets of verifying observations. Quantifying the skill in probabilistic supercell surrogate forecasts provides a means of assessing the practical predictability of severe weather phenomena in convection-permitting ensembles. This type of exercise is especially valuable as NWP advances toward approaches like Warn-on-Forecast (Stensrud et al. 2009).

An overview of the modeling framework and observation data is provided in the following section. Section 3 describes methodology for producing probabilistic guidance and the approaches for quantifying forecast skill. The results are presented in section 4, followed by a summary in section 5.

## 2. Model and data

### a. Ensemble design

Ensembles of explicit convection forecasts were generated using the Advanced Research version of the WRF (WRF-ARW, version 3.5.1; Skamarock et al. 2008). For each event, a five-member ensemble was produced by varying the initial conditions (ICs) and lateral boundary conditions (LBCs). ICs for the five-member ensemble were retrospectively drawn at random from the 30-member National Center for Atmospheric Research (NCAR) real-time WRF Data Assimilation Research Testbed (WRF-DART; Anderson et al. 2009) mesoscale ensemble analyses (Schwartz et al. 2015; Weisman et al. 2015) at 1200 UTC on the date of each event. Unique LBCs were generated for each ensemble member forecast using the fixed covariance perturbation method of Torn et al. (2006) centered on forecasts from the operational GFS model. The forecasts were initialized at 1200 UTC and were integrated for 18 h. The outermost computational domain covered the continental United States (CONUS) in addition to parts of Canada and Mexico with 15-km horizontal grid spacing (Fig. 1). An inner, nested domain spanned roughly the eastern two-thirds of the CONUS and had a horizontal grid spacing of 3 km. Both domains used a vertical grid with 40 levels.

Physical parameterizations included the Morrison double-moment microphysics scheme (Morrison et al. 2009), the Mellor–Yamada–Janjić (MYJ; Janjić 1994) planetary boundary layer (PBL) scheme, and the Noah

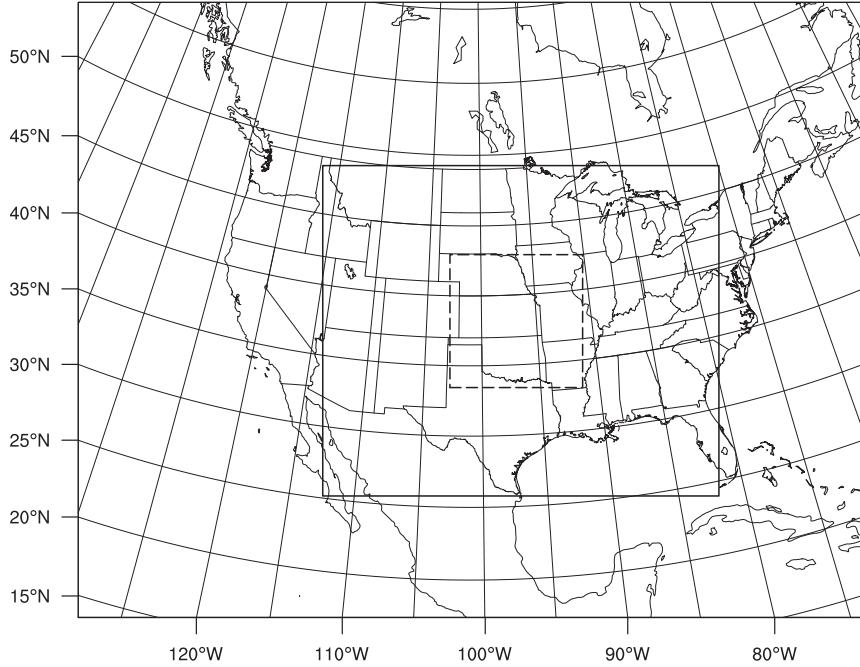


FIG. 1. Mesoscale outer domain, convection-permitting inner domain, and verification domain (dashed).

land surface model (Chen and Dudhia 2001). Longwave and shortwave radiation were parameterized with the Rapid Radiative Transfer Model for GCMs (RRTMG; Mlawer et al. 1997). Cumulus convection was parameterized on the outer domain using the Tiedtke scheme (Tiedtke 1989; Zhang et al. 2011); convective processes were treated explicitly on the inner domain.

### b. Supercell surrogates

To combat the challenges of spatially and temporally resolving severe weather phenomena in convection-permitting NWP, Kain et al. (2010) recommended the computation of relevant model diagnostics at each time step and saving the maximum value in every grid column between model output times. These resultant surrogates are superior to those from traditional hourly data as the surrogates exhibit swaths revealing the storm evolution rather than instantaneous snapshots of storm characteristics. One such supercell surrogate is UH, which as noted above is widely used to identify mesocyclone surrogates in model output. Following the definition outlined by Kain et al. (2008), UH was computed in the model as

$$UH = \int_{z=2000m}^{z=5000m} w\zeta dz, \quad (1)$$

where  $w$  is the vertical velocity ( $m s^{-1}$ ),  $\zeta$  is the vertical component of vorticity ( $s^{-1}$ ), and  $z$  is the height above

ground level (AGL). Because UH incorporates updraft strength, which is not available directly from radar data, maximum vertical vorticity in the 0–3-km layer of the model (hereafter Z3) was also diagnosed to simulate a surrogate more analogous to the rotation track data. Low-level maximum vertical vorticity has previously been used as a diagnostic in works focused on explicit prediction of tornadic supercells (e.g., Yussouf et al. 2013; Yussouf et al. 2015; Skinner et al. 2016). In this present work, analysis of both diagnostics was conducted using the hourly maximum fields as opposed to the instantaneous fields at model output times.

Locations of supercell surrogates were determined when and where UH or Z3 forecasts exceeded specified thresholds within a  $900\text{ km} \times 900\text{ km}$  verification domain located over the central Great Plains (see Fig. 1). Two thresholds were chosen for each diagnostic, with one threshold corresponding to the 99.5th percentile and the other at the more restrictive 99.9th percentile. The 99.5th and 99.9th percentile values for UH and Z3 were determined using the distribution of all values within the verification domain at each forecast hour. For each of the four cases, a 12-h verification period was considered from 1800 to 0600 UTC (i.e., forecast hours 7–18). In total, 48 hourly percentile values were computed and averaged to determine the respective thresholds for UH and Z3. These percentile values corresponded to 40 and

$100 \text{ m}^2 \text{s}^{-2}$ , respectively, for UH (i.e.,  $\text{UH}_{\text{thresh}}$ ) and  $0.006$  and  $0.008 \text{ s}^{-1}$ , respectively, for Z3. The UH thresholds are within the range of values that have been used for various research and forecasting applications (e.g., Kain et al. 2008; Trapp et al. 2011; Carley et al. 2011; Sobash et al. 2011; Clark et al. 2012). The Z3 thresholds are slightly lower than the nominal value often used to define mesocyclones in observation-based and idealized model simulation studies of supercells [ $0.01 \text{ s}^{-1}$ ; e.g., Brandes et al. (1988); Trapp (2013)].

### c. Observational data

Two sets of observational data, storm reports and rotation tracks, were employed to verify the supercell surrogate forecasts. Rotation track data were produced using the Warning Decision Support System–Integrated Information (WDSS-II; Lakshmanan et al. 2007) software package. Level II data (Crum and Alberty 1993) for the entire 12-h verification period of each case were obtained from the National Climatic Data Center (now known as NCEI) for 19 Weather Surveillance Radar-1988 Doppler (WSR-88D) stations located within the verification domain: Amarillo, Texas (KAMA); Dodge City, Kansas (KDDC); Des Moines, Iowa (KDMX); Kansas City, Missouri (KEAX); Frederick, Oklahoma (KFDR); Dallas–Fort Worth, Texas (KFWS); Goodland, Kansas (KGKD); Wichita, Kansas (KICT); Tulsa, Oklahoma (KINX); North Platte, Nebraska (KLNX); St. Louis, Missouri (KLSX); Little Rock, Arkansas (KLZK); Omaha, Nebraska (KOAX); Springfield, Missouri (KSGF); Fort Smith, Arkansas (KSRX); Oklahoma City, Oklahoma (KTLX); Topeka, Kansas (KTWX); Hastings, Nebraska (KUEX); and Vance Air Force Base, Oklahoma (KVNX). The reflectivity data were quality controlled, and radial velocities were dealiased before calculating the azimuthal shear with the linear least squares derivative (LLSD) algorithm (Smith and Elmore 2004; Newman et al. 2013; Lakshmanan et al. 2014) for all available tilts and scans for each radar. The single-radar azimuthal shear radial data were mapped onto a three-dimensional latitude–longitude–height grid, and the vertical maximum was recorded in a two-dimensional latitude–longitude grid. The single-radar vertical maximum azimuthal shear values were merged using data from all 19 radars. When merging the data, the maximum  $0\text{--}3\text{-km}$  azimuthal shear value was again recorded if two or more radars observed this layer at the same 2D grid box. The merged rotation tracks were accumulated to create hourly rotation tracks for the entire verification domain and time period, and the hourly data were interpolated onto the model grid for verification.

The azimuthal shear recorded in these rotation track data (hereafter ROT) provides observations of shearing radial

velocities that were potentially attributable to mesocyclonic circulations. To further constrain the data, thresholds were applied to the azimuthal shear magnitudes to discriminate severe storms in the ROT data. Following the approach used for the model diagnostics, the 99.5th and 99.9th percentile thresholds were determined by calculating the percentile values from hourly distributions of ROT within the verification domain and averaging the 48 hourly percentile values. For ROT, the 99.5th and 99.9th percentile thresholds corresponded to  $0.009$  and  $0.013 \text{ s}^{-1}$ , respectively.

Local storm reports (hereafter LSRs) were obtained from the Storm Prediction Center website (<http://www.spc.noaa.gov/wcm/>). Because supercell thunderstorms can produce a broad spectrum of hazards and account for roughly half of LSRs (Duda and Gallus 2010), reports of all severe hazard types (i.e., tornado, wind, and hail) were included in this study. The LSR data were mapped to the 3-km model grid using a nearest-neighbor approach. This process was conducted in hourly intervals over the 12-h verification period. Each 3-km grid box containing one LSR was treated the same as a grid box containing more than one LSR; therefore, the number of LSRs in each grid box had no influence on the verification scores. Because the LSR data essentially were treated as binary data (yes/no of a report) in each grid box, thresholds could not be applied to these data in the same way as were applied to the model diagnostics and ROT data. Therefore, the probability of spatial occurrence (i.e., fractional spatial coverage) was used to ensure that the ROT and LSR data used for verification were consistent for the appropriate supercell surrogates. The probability of spatial occurrence of LSRs was computed for each hour of the verification period for each case, and the 48 values were averaged. The average spatial coverage of gridded LSRs within the verification domain was approximately  $0.01\%$ , which corresponded to the average spatial coverage of the ROT exceedances identified with the 99.9th percentile threshold. [Note that although a 99.9th percentile would suggest that the spatial coverage of exceedances should be  $0.1\%$ , the spatial pattern of exceedances (e.g., clustering) likely caused a lower average spatial coverage.] Therefore, the grid-scale LSR data were used (in contrast with the 99.9th percentile ROT exceedances) to verify the simulated supercell surrogates determined with the 99.9th percentile thresholds (Fig. 2a). To verify the 99.5th percentile supercell surrogates, the spatial coverage of LSRs was artificially increased to approximate that of the ROT exceedances identified with the 99.5th percentile threshold (Fig. 2b). This process was achieved by adding an LSR to a previously empty grid box if at least one LSR was recorded within  $3\text{ km}$  from the center of the empty grid box (thus, expanding LSRs to

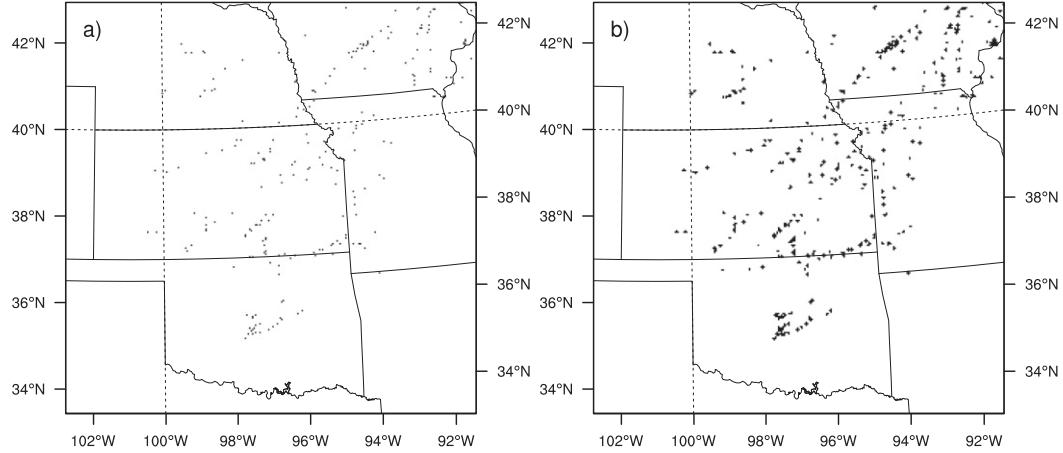


FIG. 2. (a) Storm reports from 1800 UTC 19 May to 0600 UTC 20 May 2013 plotted on the model grid. (b) Enhanced storm reports using the 3-km artificial areal coverage expansion radius.

include “hits” in adjacent grid boxes). All grid boxes that contained LSRs in the grid-scale data remained unchanged. After applying this procedure, the average spatial coverage of the expanded LSRs within the verification domain was approximately 0.075%, which roughly corresponded to the average spatial coverage (0.082%) of the ROT exceedances identified with the 99.5th percentile threshold.

### 3. Verification approach

#### a. Probabilistic guidance

Ensemble forecasting systems inherently possess probabilistic information that can aid in quantifying the uncertainty in a forecast. This implicit probabilistic information can be extracted from the ensemble system by way of the ensemble probability (EP). Using the specified threshold for each supercell surrogate, the forecasts from each ensemble member were converted into a dichotomous field of binary probabilities (BPs), such as

$$\text{BP}_{ki} = \begin{cases} 1 & \text{if } \text{UH}_{ki} \geq \text{UHthresh} \\ 0 & \text{if } \text{UH}_{ki} < \text{UHthresh} \end{cases}, \quad (2)$$

for the  $k$ th ensemble member, at the  $i$ th grid point; an analogous procedure was followed to determine the BPs based on the Z3 threshold. Using BPs, EP was then computed at the  $i$ th grid point as

$$\text{EP}_i = \frac{1}{n} \sum_{k=1}^n \text{BP}_{ki}, \quad (3)$$

where  $n$  is the number of ensemble members. Within the framework of this study, the EP estimates the likelihood of midlevel or low-level rotation occurring within each grid column based on the number of

ensemble members that predict UH or Z3 to exceed its specified threshold.

In an effort to account for the previously mentioned challenge of verifying convective-scale predictands, neighborhood-based probabilistic guidance was incorporated. Following [Roberts and Lean \(2008\)](#), a neighborhood was established for each grid box by specifying a radius of influence (ROI). The neighborhood was then defined to include all of the grid boxes whose centers fell within the ROI. Using additional information from these surrounding grid boxes, neighborhood probabilities (NPs) were computed to allow for spatial uncertainty on the high-resolution model domain. NPs were computed as

$$\text{NP}_{ki} = \frac{1}{N_b} \sum_{m=1}^{N_b} \text{BP}_{km}, \quad (4)$$

where  $m$  is the grid point within the neighborhood and  $N_b$  is the number of grid boxes within the neighborhood. The NP can be thought of as the relative frequency of events forecasted or observed to occur within a given radius of a grid box. Figure 11 in [Schwartz et al. \(2010\)](#) illustrated how the NP for the center grid box would be equal for the forecast and observations, but the point-to-point comparison indicates an incorrect forecast.

In [Schwartz et al. \(2010\)](#), the ensemble and neighborhood approaches to forecast probabilities were combined to create the neighborhood ensemble probability (NEP). In this study, the NEP was computed as

$$\text{NEP}_i = \frac{1}{n} \sum_{k=1}^n \text{NP}_{ki}. \quad (5)$$

[Schwartz et al. \(2010\)](#) posited that merging the neighborhood and ensemble approaches to produce forecast

TABLE 1. A  $2 \times 2$  contingency table for dichotomous events.

|                | Observed | Not observed     |
|----------------|----------|------------------|
| Forecasted     | Hit      | False alarm      |
| Not forecasted | Miss     | Correct negative |

probabilities serves to better represent the true probability density function of the atmospheric state. Moreover, for forecasts with fine grid spacing, guidance can still be of considerable value even if the grid-scale prediction is not exact. If the mean forecast probabilities within a specified ROI are equivalent to the areal coverage of the event exceeding the threshold over the same ROI, the forecast is considered perfect. Within this framework, the NP and NEP depend on the specified ROI, but the EP is independent of this specification.

### b. Quantification of forecast skill

To calculate forecast skill, the fractions skill score (FSS; [Roberts 2005](#); [Roberts and Lean 2008](#)) and area under the relative operating characteristic curve (ROC; [Mason 1982](#)) were computed for the entire verification domain, as these performance metrics are appropriate for a probabilistic forecasting and verification approach. FSS quantifies the skill of a forecast relative to a worst-case reference forecast, and is defined as

$$\text{FSS} = 1 - \frac{\text{FBS}}{\text{FBS}_{\text{worst}}}, \quad (6)$$

where the fractions Brier score (FBS; [Roberts 2005](#)) of the model forecast and the worst-case reference forecast are defined as

$$\text{FBS} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (\text{FP}_{ij} - \text{OP}_{ij})^2 \quad (7)$$

and

$$\text{FBS}_{\text{worst}} = \frac{1}{N_x N_y} \left( \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \text{FP}_{ij}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \text{OP}_{ij}^2 \right), \quad (8)$$

where  $N_x$  ( $N_y$ ) is the number of grid points in the east-west (north-south) direction and  $\text{FP}_{ij}$  and  $\text{OP}_{ij}$  are the respective forecast and observation probabilities at the  $ij$ th grid point. FSS ranges from 0 to 1. A value of 1 indicates a perfect forecast, and 0 indicates a forecast with no skill.

While FSS only grades the forecasts against the observations in a holistic sense, ROC curves assess how well

TABLE 2. Forecast and observation probability pairs. NEP was computed from model forecasts of UH and Z3. Neighborhood probabilities were computed from observed ROT and LSR data. The appended numerals 5 and 9 denote probabilities computed using the 99.5th and 99.9th percentiles, respectively.

|   | Forecast probability | Supercell surrogate | Observed data |
|---|----------------------|---------------------|---------------|
| 1 | NEP5                 | Z3                  | ROT5          |
| 2 | NEP5                 | UH                  | ROT5          |
| 3 | NEP5                 | Z3                  | LSR5          |
| 4 | NEP5                 | UH                  | LSR5          |
| 5 | NEP9                 | Z3                  | ROT9          |
| 6 | NEP9                 | UH                  | ROT9          |
| 7 | NEP9                 | Z3                  | LSR9          |
| 8 | NEP9                 | UH                  | LSR9          |

the forecasts discriminate between observed events and nonevents ([Mason 1982](#)). ROC curves were developed by selecting a range of probability thresholds (0%–100%) to convert the continuous probabilistic forecast and observation fields into discrete dichotomous events. At each probability threshold, the elements of the  $2 \times 2$  contingency table for dichotomous events ([Table 1](#); [Wilks 1995](#)) were then used to compute the probability of detection [POD = hits/(hits + misses)] and probability of false detection [POFD = false alarms/(false alarms + correct negatives)]. POD and POFD were plotted against one another to produce the ROC curves.

In this study, the area under the curve (AUC) was computed using a trapezoidal approximation to quantify the degree of discrimination shown in the forecasts. AUC ranges from 0 to 1, with an area of 1.0 representing a perfect probabilistic forecast. An AUC below 0.5 indicates a forecast possesses no discriminatory skill (i.e., POFD is greater than POD), while [Buizza et al. \(1999\)](#) determined that AUC must be greater than 0.7 in order to classify a forecast as useful.

As previously noted, two thresholds (99.5th and 99.9th percentiles) were applied to UH and Z3 to determine supercell surrogates, which indicate simulated storms capable of producing storm damage. This process was conducted for each forecast hour using the hourly maximum fields, and the hourly surrogates were aggregated over a 12-h period (1800–0600 UTC, i.e., forecast hours 7–18) before computing NEP. This occurrence determination and aggregation over the same 12-h period was also conducted for the ROT and LSR data. For simplicity, the NEPs computed with the 99.5th and 99.9th percentile thresholds will be referred to as NEP5s and NEP9s. In a similar fashion, NPs were computed using exceedances determined from the gridded ROTs using the 99.5th and 99.9th percentile thresholds (hereafter, ROT5s and ROT9s, respectively). NPs were also

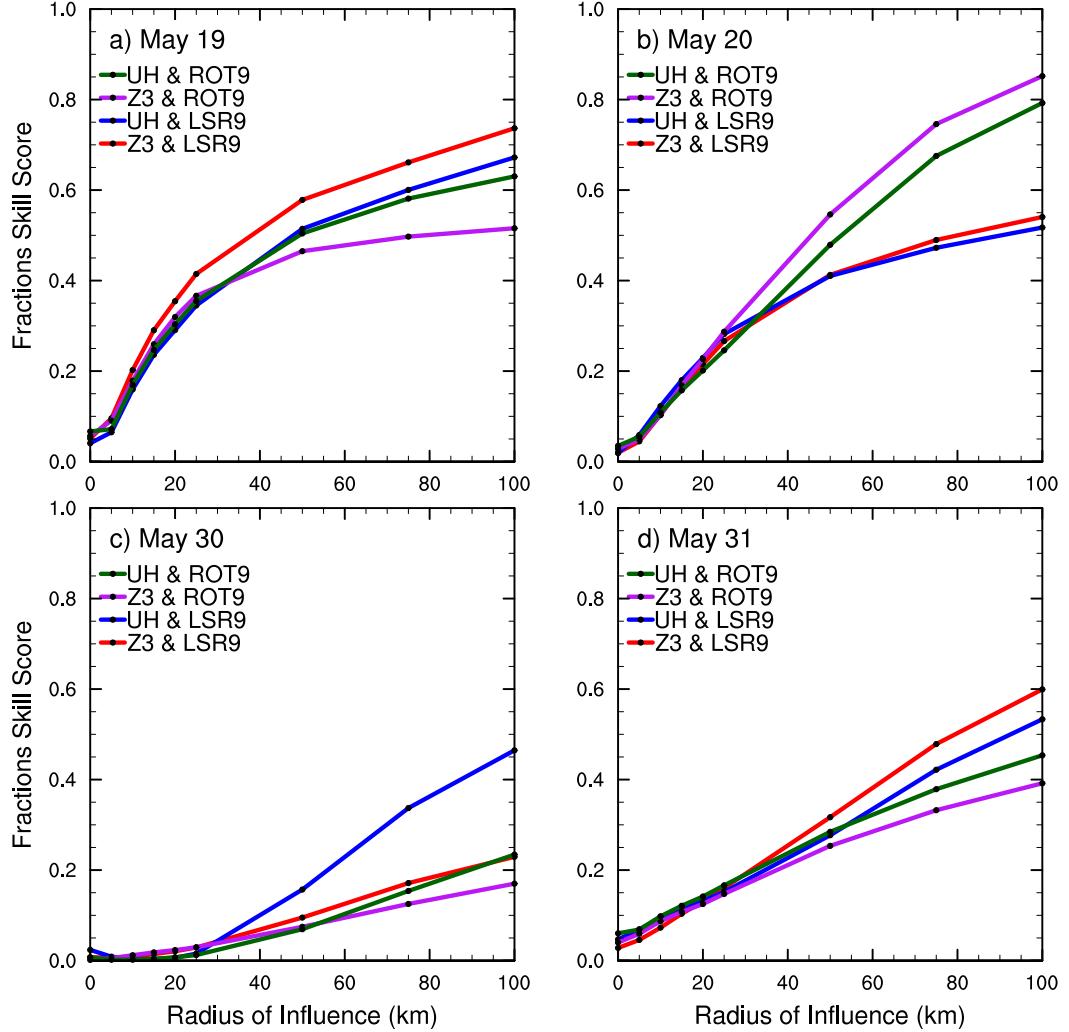


FIG. 3. FSS as a function of ROI for the (a) 19, (b) 20, (c) 30, and (d) 31 May 2013 NEP9 forecasts. NEPs compared against ROT9s (LSR9s) are shown in green and purple (red and blue).

computed using the grid-scale and artificially expanded areal coverage LSRs approximating the 99.9th and 99.5th percentiles (hereafter, LSR9s and LSR5s, respectively). Following this procedure, four forecast probabilities and four observation fractions fields (see Table 2) were computed for the entire verification domain and used for verifying the four cases of interest.

## 4. Results

### a. Low- and high-threshold results

In all four cases, FSSs computed from the UH and Z3 NEP forecasts displayed a clear dependency on ROI (Figs. 3 and 4). Beginning with a 0-km ROI, little skill was found in the NEPs, regardless of whether

LSRs or ROTs were used for verification. The NEPs calculated with this 0-km ROI were essentially EPs and, thus, served as probabilistic forecast information verified at the grid scale. Grid-scale forecasts showed a dependency on the specified threshold level. The FSS for all the NEP5s was approximately 0.1, while the NEP9 FSSs averaged 0.03. For each respective forecast–observation pair, the FSS for the NEP9s never exceeded the FSS calculated from the less restrictive NEP5s. When comparing forecast–observation pairs at the 99.9th or 99.5th percentile threshold levels, FSS did not show a consistent difference based on the model surrogate or observation type.

As the ROI was increased, the FSSs increased for both the NEP9s (Fig. 3) and NEP5s (Fig. 4). This was expected because both the forecast probabilities and

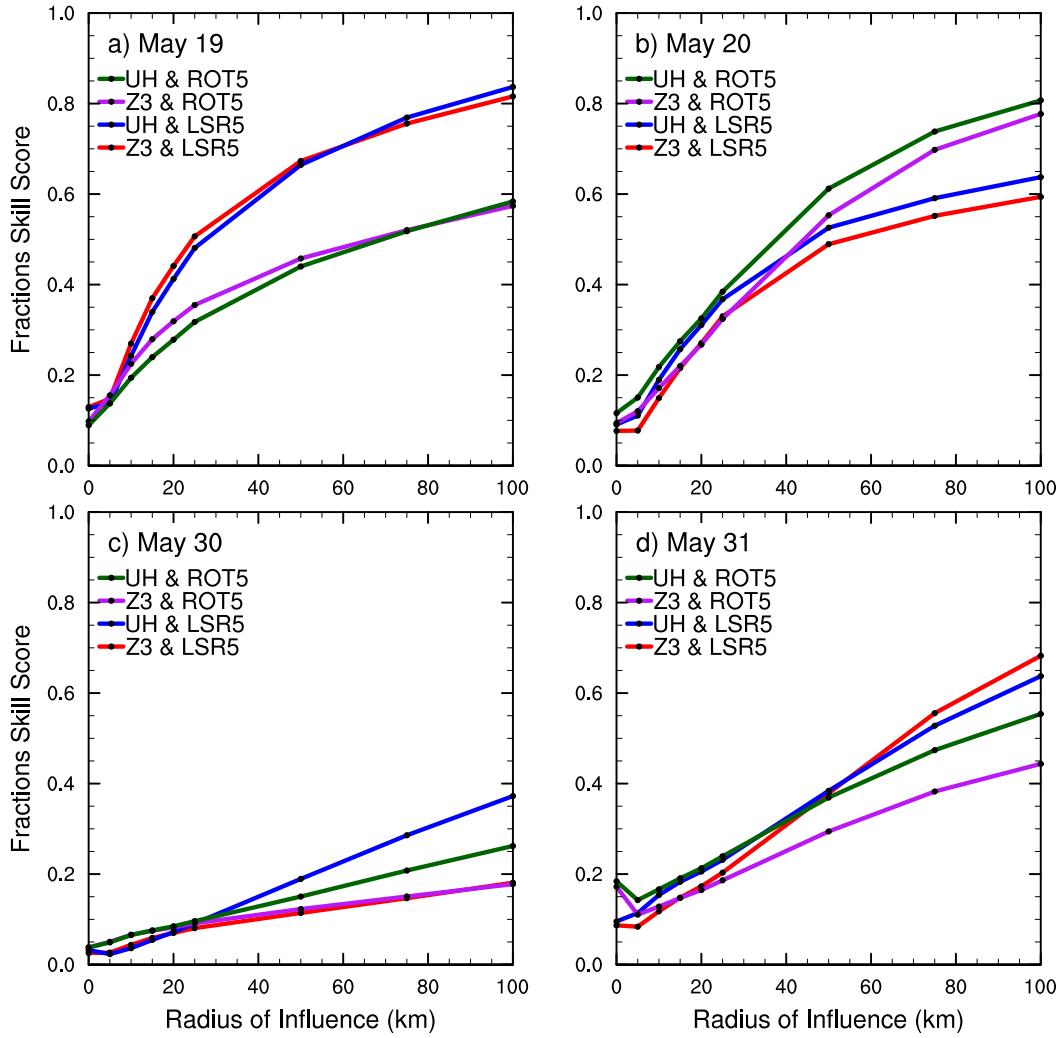


FIG. 4. FSS as a function of ROI for the (a) 19, (b) 20, (c) 30, and (d) 31 May 2013 NEP5 forecasts. NEP5s compared against ROT5s (LSR5s) are shown in green and purple (red and blue).

observed fractions fields were smoothed (see Fig. 5) as the neighborhoods grew in size (Roberts and Lean 2008). At ROIs of 50 km and greater, FSS was found to approach or exceed 0.5 for some forecast–observation probability pairs. Again, no consistent distinction was found between the FSS determined with the LSRs and that determined with the ROTs. This was true for the forecasts and observations at both the 99.5th and 99.9th percentiles.

While these FSS trends indicated an apparent degree of skill in the NEP forecasts, examination of ROC curves showed that these forecasts possessed very little discriminatory skill. For the NEP5s, the AUC was generally between 0.5 and 0.7 for all forecast–observation pairs at all ROIs (Fig. 6). The AUC was always highest for the grid-scale NEPs. In fact, the AUC computed from the NEP5s only approached the 0.7 “useful”

threshold determined by Buizza et al. (1999) for the grid-scale verification of the 31 May case (Fig. 6d). The lack of discriminatory skill displayed by the NEP forecasts was likely due to a large discrepancy between the forecast and observation probability distributions at low ROIs. With a 10-km ROI, the maximum NEP5s found within the verification domain were approximately 20% for UH and Z3 (Table 3). The corresponding maximum NPs computed from the ROT5s and LSR5s ranged from approximately 62% to 92%.

As the ROI was increased above 50 km, the AUC decreased and approached a constant 0.5 for all forecast–observation probability pairs. The NEP9s showed even less discriminatory skill, with AUC rarely exceeding 0.5 for any ROI (Fig. 7). This constant 0.5 AUC was attributable to the smoothing of the forecast probabilities and observation fractions fields toward

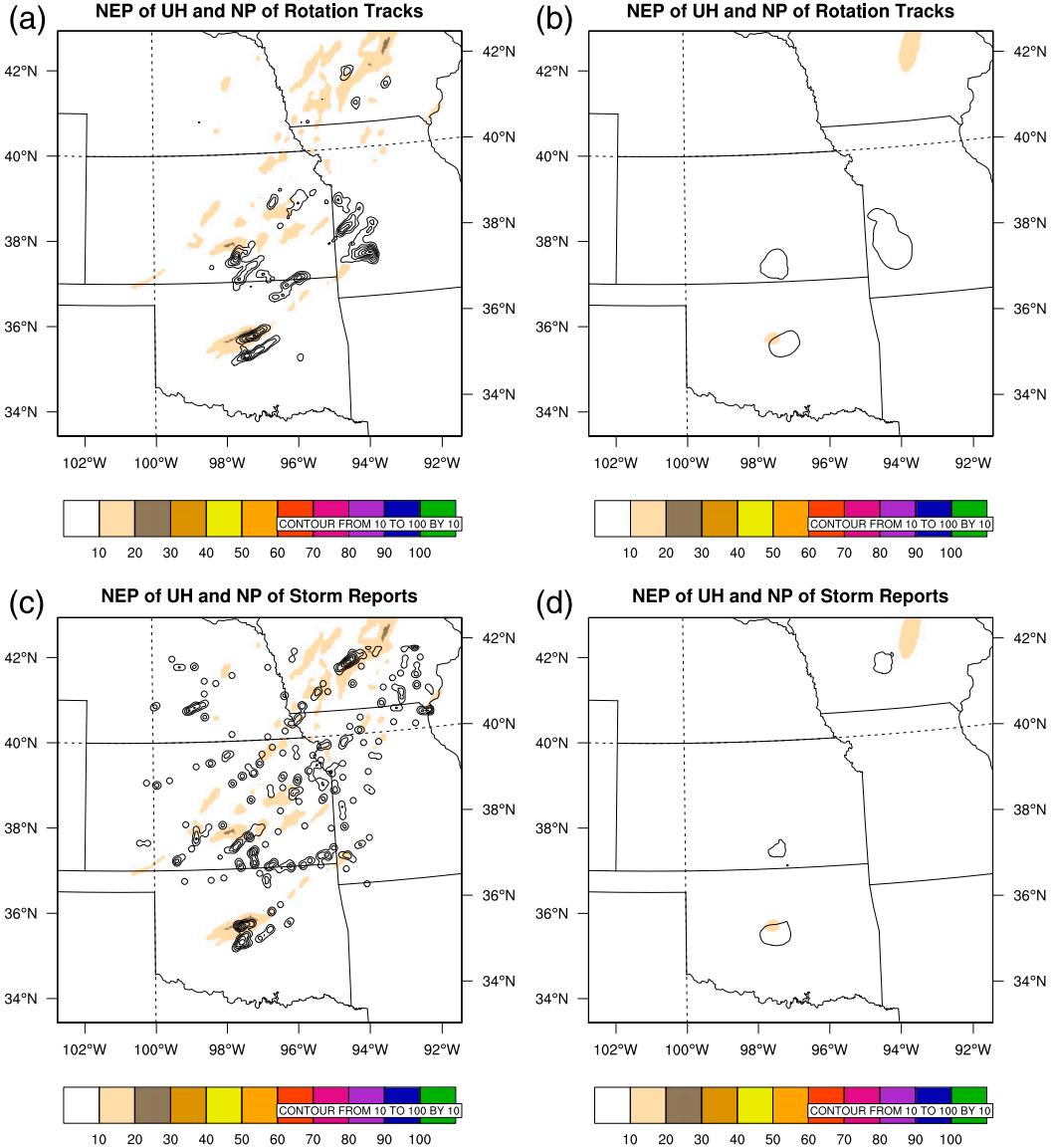


FIG. 5. UH NEP5 forecasts (color fill) and observation fractions fields (contours) for the 19 May 2013 event. Observation fractions fields shown on the left were computed with a 10-km ROI. Those shown on the right were computed with a 50-km ROI.

zero as the neighborhood sizes were increased with larger ROI (Table 3). In this scenario, when POD and POFD are calculated at each probability threshold, POD and POFD equal 1 for the 0% probability threshold, and both equal 0 for all other probability thresholds. This returned the constant ROC AUC of 0.5 that was found at large ROIs.

#### b. Mixed threshold results

In the previously discussed results, the LSR and ROT data were applied such that their respective spatial coverage percentages were roughly equal. Within the

context of the cases included in this study, the reporting of severe weather phenomena was assumed to be adequate given the high impact of these events and the proximity to a populated metropolitan area. Even if this assumption was invalid, the LSR expansion increased the neighborhood where storm damage was reported to have occurred and theoretically accounted for the potential underreporting of severe weather phenomena. These hazards occur over continuous swaths rather than discrete point locations along the paths of observed storms. Rotation tracks convey storm damage potential aligned with the storm motion, and these more

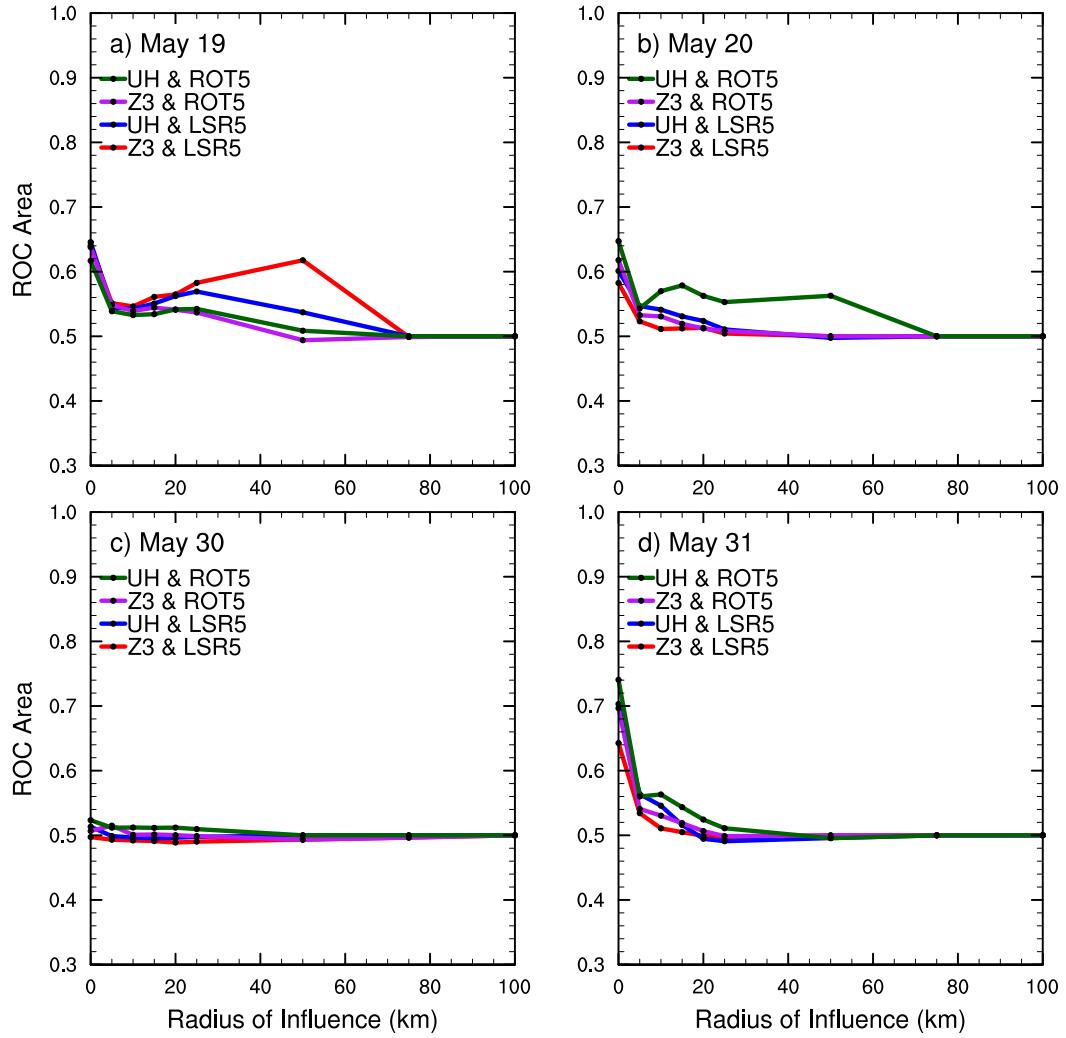


FIG. 6. AUC as a function of ROI for the (a) 19, (b) 20, (c) 30, and (d) 31 May 2013 NEP5 forecasts. NEP5s compared against ROT5s (LSR5s) are shown in green and purple (red and blue).

continuous swaths provide verification data in areas where storm reports might have been difficult to obtain as a result of a lack of infrastructure. Achieving LSR5 areal coverage meant altering the local number of LSRs in an isotropic fashion, thus introducing an additional degree of freedom into the analysis. While we do not advocate this practice, it was demonstrated herein to maintain parity between the observation datasets for verifying the 99.5th percentile surrogates. To eliminate this procedure and demonstrate the advantage of the rotation track data relative to the storm report data, mixed threshold verification was conducted using the LSR9s and the 99.5th percentile forecast surrogates and rotation track exceedances.

FSS values for the NEP5s were always higher when ROT5s were employed as opposed to the LSR9s

(Fig. 8). This was especially evident in the 19 and 20 May results. Using the LSR9s as observational truth yielded much lower FSSs because the NEP5s were biased forecasts within this framework (Mittermaier and Roberts 2010). This frequency bias effect was removed when the NEP5s were verified with the ROT5s. While the disparity in the spatial coverage of the observations was intentional, these results demonstrate an advantage for the rotation tracks over storm report data. The rotation track data are a continuous distribution akin to the model diagnostics, which enables specification of thresholds to prevent the frequency bias from affecting the FSS results. This characteristic arguably adds increased adaptability to the list of advantages the rotation track data possess relative to the storm report data.

TABLE 3. Domain-maximum neighborhood probabilities computed using low-threshold surrogates.

|         |        | ROI (km) |        |        |        |        |        |        |        |        |
|---------|--------|----------|--------|--------|--------|--------|--------|--------|--------|--------|
|         |        | 0        | 5      | 10     | 15     | 20     | 25     | 50     | 75     | 100    |
| NEP5 UH | 19 May | 100      | 22.497 | 20.549 | 20.218 | 19.653 | 19.048 | 11.985 | 8.978  | 7.905  |
|         | 20 May | 100      | 22.487 | 20.511 | 19.244 | 16.715 | 14.616 | 13.031 | 9.154  | 6.992  |
|         | 30 May | 100      | 22.454 | 20.029 | 20.003 | 19.274 | 17.401 | 8.532  | 4.865  | 3.361  |
|         | 31 May | 100      | 22.5   | 20.517 | 20.08  | 19.59  | 18.571 | 12.804 | 10.183 | 7.439  |
| NEP5 Z3 | 19 May | 100      | 22.5   | 20.547 | 19.952 | 19.945 | 19.517 | 15.133 | 11.092 | 9.028  |
|         | 20 May | 100      | 22.415 | 20.283 | 17.891 | 15.824 | 14.602 | 9.565  | 7.628  | 5.542  |
|         | 30 May | 100      | 22.222 | 20.002 | 20     | 18.102 | 17.376 | 14.504 | 11.371 | 9.159  |
|         | 31 May | 100      | 22.469 | 20.44  | 20.012 | 18.408 | 15.409 | 10.746 | 8.499  | 5.807  |
| NP ROT5 | 19 May | 100      | 100    | 62.162 | 39.506 | 29.197 | 22.179 | 15.621 | 8.108  | 6.305  |
|         | 20 May | 100      | 100    | 62.162 | 49.383 | 37.956 | 30.769 | 15.621 | 10.827 | 8.402  |
|         | 30 May | 100      | 100    | 56.757 | 48.148 | 34.307 | 28.959 | 11.631 | 6.68   | 4.165  |
|         | 31 May | 100      | 100    | 91.892 | 79.012 | 77.372 | 68.326 | 35.918 | 23.773 | 15.407 |
| NP LSR5 | 19 May | 100      | 100    | 70.27  | 50.617 | 40.146 | 29.864 | 14.823 | 8.822  | 6.298  |
|         | 20 May | 100      | 100    | 54.054 | 32.099 | 26.277 | 20.814 | 9.806  | 8.006  | 6.476  |
|         | 30 May | 100      | 100    | 72.973 | 48.148 | 39.416 | 25.339 | 8.78   | 6.295  | 4.43   |
|         | 31 May | 100      | 100    | 81.081 | 60.493 | 48.175 | 41.176 | 15.735 | 9.007  | 5.849  |

### c. Results within the context of previous studies

NEP forecasts of UH were found to have comparable degrees of forecast skill when verified against ROT and LSR data. Because the ROT data were shown to have a distinct advantage over the LSR data within this framework, the results of the previous section would encourage using ROT data to verify supercell surrogate forecasts. As previously mentioned, updraft strength is included in the UH calculation but is not directly available in the radar data. In addition, UH diagnoses rotation in the midlevels between 2 and 5 km AGL, while the ROT data used herein observed rotation in the lowest 3 km of the atmosphere. Therefore, Z3 was also diagnosed and verified as a surrogate that is more analogous to the ROT data. The FSS and ROC AUC calculated for the Z3 forecasts displayed similar trends to those found for the UH forecasts. In fact, the verification metrics were not consistently higher or lower for UH or Z3 (see Figs. 3, 4, and 6–8). This finding is reasonable as both surrogates diagnose mesocyclone rotation and overlap between 2 and 3 km. Given the similarity in the UH and Z3 results and the correspondence between the Z3 and the ROT data, this study suggests that diagnosing Z3 in addition to UH is reasonable for severe convective forecasting and for forecast verification against rotation track data.

Computing NEP as stated in section 3a led to an upper limit of about 20% for the chosen supercell surrogates considered for the four cases in this study. The NEP maxima were significantly lower than NEP values shown in work such as Yussouf et al. (2015). While this current study used thresholds of 0.006 and 0.008 s<sup>-1</sup> for 0–3-km maximum vertical vorticity, Yussouf et al. (2015) used

thresholds of 0.002 and 0.004 s<sup>-1</sup> for 0–1-km maximum vertical vorticity. In addition, Yussouf et al. (2015) was a storm-scale data assimilation study in which reflectivity and Doppler velocities were assimilated into a storm-scale ensemble every 5 min before generating 0–1-h forecasts. It is reasonable to expect that such storm-scale data assimilation would dramatically improve the location of forecast objects that would affect forecast probabilities in a strongly positive manner. The differences in thresholds, vertical vorticity layers, and the forecast integration periods could all be attributing to the disparities in maximum NEPs in these two studies.

This work aimed at determining the best way to utilize rotation track data via a probabilistic approach using percentile thresholds of select severe storm surrogates against two types of verification data. In a similar study, Skinner et al. (2016) also conducted objective verification of low-level vorticity forecasts against radar-derived rotation tracks. This study differs in several important ways. First, Skinner et al. (2016) utilized an object-based method with distance and amplitude score to quantify skill in 0–1-h forecasts. This study verified 0–18-h forecasts using a neighborhood-based probabilistic verification approach. Additionally, both updraft helicity and low-level vertical vorticity were verified, and storm reports were evaluated as verification observations in addition to rotation tracks.

### 5. Summary and conclusions

Ensembles of convection-permitting forecasts were produced for four severe weather events during May 2013, including high-impact events on 20 and 31 May 2013. The five ensemble members were initialized by

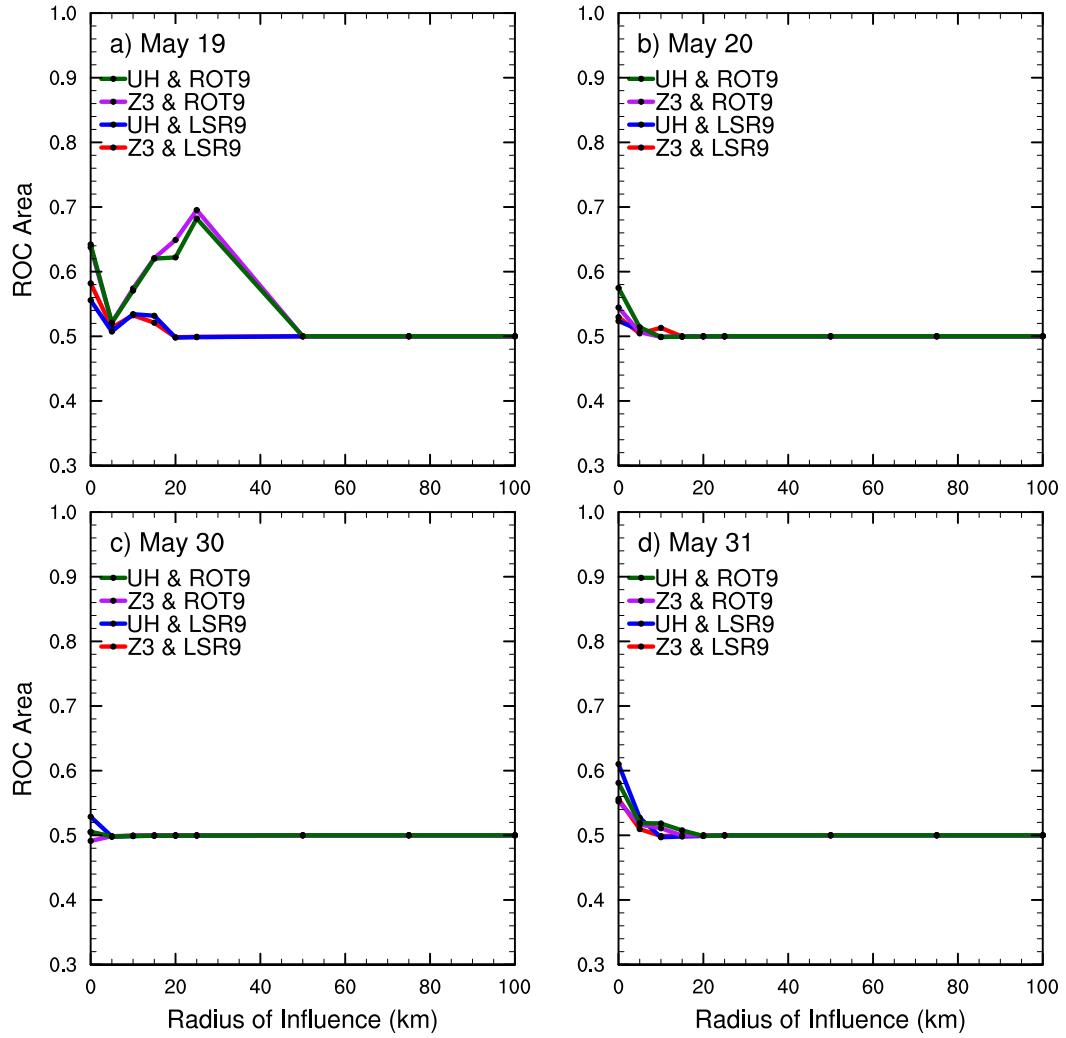


FIG. 7. AUC as a function of ROI for the (a) 19, (b) 20, (c) 30, and (d) 31 May 2013 NEP9 forecasts. NEP9s compared against ROT9s (LSR9s) are shown in green and purple (red and blue).

randomly drawing 1200 UTC mesoscale analyses from the 30-member NCAR real-time WRF-DART ensemble data assimilation system (Schwartz et al. 2015). Using WRF-ARW, 18-h convection-permitting forecasts were produced on a nested domain covering the central portions of the CONUS with 3-km horizontal grid spacing. Probabilistic guidance for model-simulated rotation was produced using forecasts of UH and maximum 0–3-km vertical vorticity. These probabilistic forecasts were then verified against storm reports and radar-derived rotation tracks to investigate the practicality of using these observations to objectively verify forecasts of supercell thunderstorm occurrences. Using the 99.5th and 99.9th percentiles as thresholds to identify supercell surrogates, the levels of model skill for the UH and Z3 forecasts were comparable for all four cases of interest. These positively correlated model diagnostics

appear to be interchangeable in this exercise, thus encouraging the further proliferation of diagnosing maximum low-level vorticity in addition to UH in severe weather forecasting across a broader range of forecast scenarios.

Finally, the selection of ROTs or LSRs for verification data did not have a consistent impact on the level of skill determined from the NEP forecasts, given consistent spatial coverage between the ROT exceedances and the gridded LSRs. It was found in the cases considered in this study that radar-derived rotation tracks were largely equivalent to storm reports when the areal coverages were equal. Yet, greater skill was evident when a lower rotation track threshold was used for verification, an option not reasonably obtainable with LSRs. Thus, the rotation track data have a distinct advantage over the storm reports in this application, as

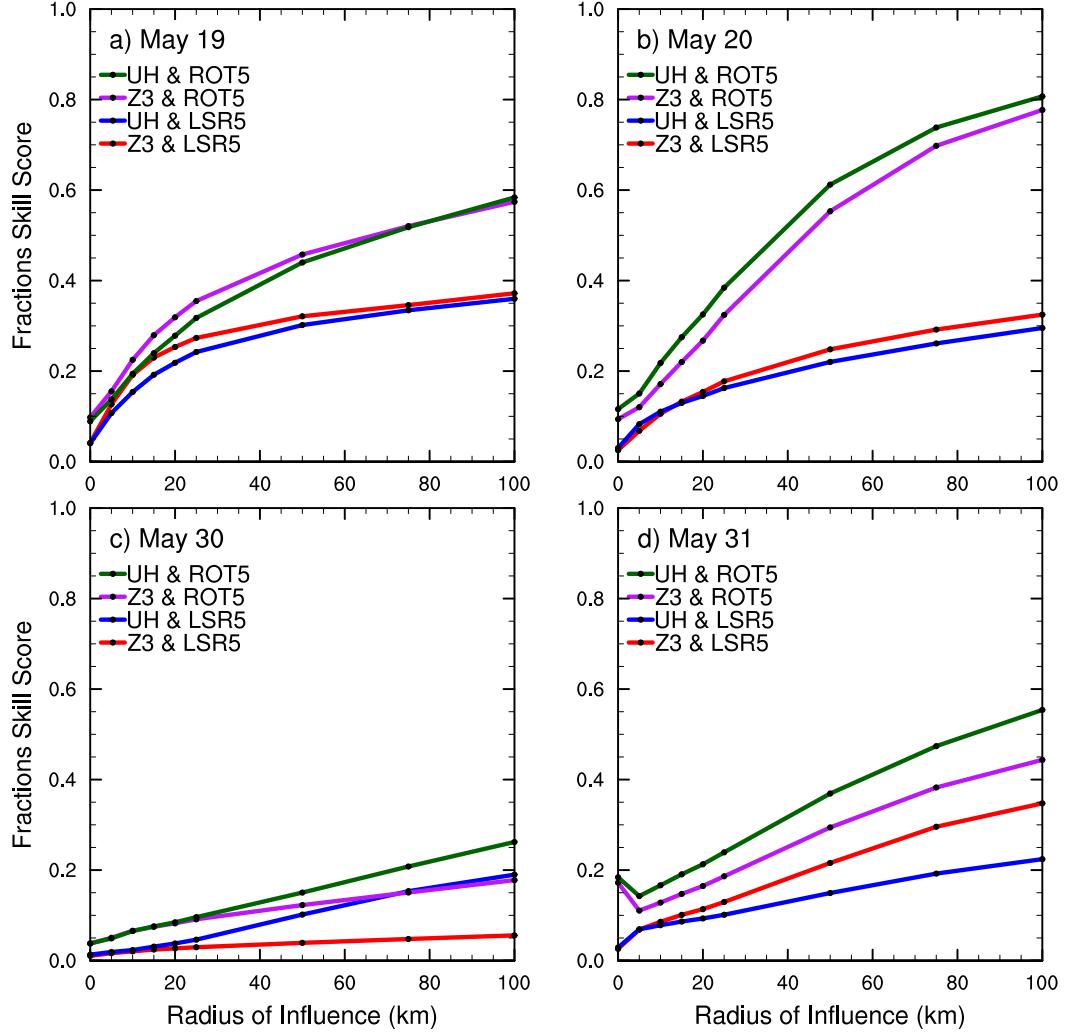


FIG. 8. FSS as a function of ROI for the (a) 19, (b) 20, (c) 30, and (d) 31 May 2013 NEP5 forecasts. NEP5s compared against the ROT5s (grid-scale LSR9s) are shown in green and purple (red and blue).

it is trivial to adjust the rotation track verification threshold. Here, specified thresholds for the ROT, UH, and Z3 distributions were modified to change the frequency of threshold exceedances for both model and verification data. We also demonstrate altering the spatial coverage of the storm reports to achieve equal areal coverage at lower forecast thresholds, thus requiring an undesirable artificial expansion of LSR areal coverage. Identifying model-predicted surrogates and radar-observed exceedances using frequencies from continuous distributions allows for the straightforward adjustment of thresholds, which can be easily modified to fit forecasting applications for varying severe weather types, intensities, and locations. Because of this ease in fitting the rotation track data and the absence of the nonmeteorological biases present in the

storm report database, the rotation track data warrant increased use in developing and calibrating severe storm surrogates used to identify thunderstorms capable of producing severe weather in convection-permitting forecasts.

**Acknowledgments.** We thank the National Science Foundation (NSF) for support of this work (AGS-1230085) and the National Severe Storms Laboratory for supplying the WDSS-II application that was used to generate the rotation tracks. This research began while the lead author participated in the Significant Opportunities in Atmospheric Research and Science (SOARS) program. Kiel Ortega, Craig Schwartz, and Sarah Tessendorf all provided helpful assistance and discussions during SOARS and/or thereafter. The authors

would also like to thank the three anonymous reviewers for their feedback.

## REFERENCES

- Anderson, J. L., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community data assimilation facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi:[10.1175/2009BAMS2618.1](https://doi.org/10.1175/2009BAMS2618.1).
- Brandes, E. A., R. P. Davies-Jones, and B. C. Johnson, 1988: Streamwise vorticity effects on supercell morphology and persistence. *J. Atmos. Sci.*, **45**, 947–963, doi:[10.1175/1520-0469\(1988\)045<0947:SVEOSM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<0947:SVEOSM>2.0.CO;2).
- Brooks, H. E., C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability. *Wea. Forecasting*, **18**, 626–640, doi:[10.1175/1520-0434\(2003\)018<0626:CEOLDT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0626:CEOLDT>2.0.CO;2).
- Buizza, R., A. Hollingsworth, F. Lalaurie, and A. Ghelli, 1999: Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Wea. Forecasting*, **14**, 168–189, doi:[10.1175/1520-0434\(1999\)014<0168:PPPOPUP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0168:PPPOPUP>2.0.CO;2).
- Carley, J. R., B. R. J. Schwedler, M. E. Baldwin, R. J. Trapp, J. Kwiatkowski, J. Logsdon, and S. J. Weiss, 2011: A model-based methodology for feature-specific prediction for high-impact weather. *Wea. Forecasting*, **26**, 243–249, doi:[10.1175/WAF-D-10-05008.1](https://doi.org/10.1175/WAF-D-10-05008.1).
- Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, doi:[10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2).
- Clark, A. J., W. A. Gallus Jr., and M. L. Weisman, 2010: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF simulations and the operational NAM. *Wea. Forecasting*, **25**, 1495–1509, doi:[10.1175/2010WAF2222404.1](https://doi.org/10.1175/2010WAF2222404.1).
- , J. S. Kain, P. T. Marsh, J. Corrieria Jr., M. Xue, and F. Kong, 2012: Forecasting tornado pathlengths using a three-dimensional object algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113, doi:[10.1175/WAF-D-11-00147.1](https://doi.org/10.1175/WAF-D-11-00147.1).
- , J. Gao, P. Marsh, T. Smith, J. Kain, J. Corrieria, M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387–407, doi:[10.1175/WAF-D-12-00038.1](https://doi.org/10.1175/WAF-D-12-00038.1).
- Crum, T. D., and R. K. Alberty, 1993: The WSR-88D and the WSR-88D Operational Support Facility. *Bull. Amer. Meteor. Soc.*, **74**, 1669–1687, doi:[10.1175/1520-0477\(1993\)074<1669:TWATWO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<1669:TWATWO>2.0.CO;2).
- Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecast (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, doi:[10.1002/asl.72](https://doi.org/10.1002/asl.72).
- Doswell, C. A., III, H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595, doi:[10.1175/WAF866.1](https://doi.org/10.1175/WAF866.1).
- Duda, J. D., and W. A. Gallus Jr., 2010: Spring and summer mid-western severe weather reports in supercells compared to other morphologies. *Wea. Forecasting*, **25**, 190–206, doi:[10.1175/2009WAF2222338.1](https://doi.org/10.1175/2009WAF2222338.1).
- Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64, doi:[10.1002/met.25](https://doi.org/10.1002/met.25).
- 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, doi:[10.1175/1520-0493\(1994\)122<0927:TSMECM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2).
- Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, doi:[10.1175/WAF2007106.1](https://doi.org/10.1175/WAF2007106.1).
- , S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542, doi:[10.1175/2010WAF2222430.1](https://doi.org/10.1175/2010WAF2222430.1).
- Lakshmanan, V., T. Smith, G. J. Stumpf, and K. Hondl, 2007: The Warning Decision Support System–Integrated Information. *Wea. Forecasting*, **22**, 596–612, doi:[10.1175/WAF1009.1](https://doi.org/10.1175/WAF1009.1).
- , C. Karstens, J. Krause, and L. Tang, 2014: Quality control of weather radar data using polarimetric variables. *J. Atmos. Oceanic Technol.*, **31**, 1234–1249, doi:[10.1175/JTECH-D-13-00073.1](https://doi.org/10.1175/JTECH-D-13-00073.1).
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Miller, M. L., V. Lakshmanan, and T. M. Smith, 2013: An automated method for depicting mesocyclone paths and intensities. *Wea. Forecasting*, **28**, 570–585, doi:[10.1175/WAF-D-12-00065.1](https://doi.org/10.1175/WAF-D-12-00065.1).
- Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, doi:[10.1175/2009WAF2222260.1](https://doi.org/10.1175/2009WAF2222260.1).
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:[10.1029/97JD00237](https://doi.org/10.1029/97JD00237).
- Morrison, H., G. Thompson, and V. Tatarkii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, doi:[10.1175/2008MWR2556.1](https://doi.org/10.1175/2008MWR2556.1).
- Newman, J. F., V. Lakshmanan, P. L. Heinzelman, M. B. Richman, and T. M. Smith, 2013: Range-correcting azimuthal shear in Doppler radar data. *Wea. Forecasting*, **28**, 194–211, doi:[10.1175/WAF-D-11-00154.1](https://doi.org/10.1175/WAF-D-11-00154.1).
- Roberts, N. M., 2005: An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall. Met Office Tech. Rep. 455, 80 pp.
- , and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:[10.1175/2007MWR2123.1](https://doi.org/10.1175/2007MWR2123.1).
- Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, doi:[10.1175/2009WAF222267.1](https://doi.org/10.1175/2009WAF222267.1).
- , 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. Forecasting*, **30**, 1158–1181, doi:[10.1175/WAF-D-15-0013.1](https://doi.org/10.1175/WAF-D-15-0013.1).
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:[10.5065/D68S4MVH](https://doi.org/10.5065/D68S4MVH).

- Skinner, P. S., L. J. Wicker, D. M. Wheatley, and K. H. Knopfmeier, 2016: Application of two spatial verification methods to ensemble forecasts of low-level rotation. *Wea. Forecasting*, **31**, 713–735, doi:[10.1175/WAF-D-15-0129.1](https://doi.org/10.1175/WAF-D-15-0129.1).
- Smith, T. M., and K. L. Elmore, 2004: The use of radial velocity derivatives to diagnose rotation and divergence. *11th Conf. on Aviation, Range, and Aerospace*, Hyannis, MA, Amer. Meteor. Soc., 5.6. [Available online at <https://ams.confex.com/ams/pdfpapers/81827.pdf>.]
- Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728, doi:[10.1175/WAF-D-10-05046.1](https://doi.org/10.1175/WAF-D-10-05046.1).
- , C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, doi:[10.1175/WAF-D-15-0138.1](https://doi.org/10.1175/WAF-D-15-0138.1).
- Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, doi:[10.1175/2009BAMS2795.1](https://doi.org/10.1175/2009BAMS2795.1).
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, doi:[10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2).
- Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502, doi:[10.1175/MWR3187.1](https://doi.org/10.1175/MWR3187.1).
- Trapp, R. J., 2013: *Mesoscale-Convective Processes in the Atmosphere*. Cambridge University Press, 346 pp.
- , D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting*, **21**, 408–415, doi:[10.1175/WAF925.1](https://doi.org/10.1175/WAF925.1).
- , E. D. Robinson, M. E. Baldwin, N. S. Diffenbaugh, and B. R. J. Schwedler, 2011: Regional climate of hazardous convective weather through high-resolution dynamical downscaling. *Climate Dyn.*, **37**, 667–688, doi:[10.1007/s00382-010-0826-y](https://doi.org/10.1007/s00382-010-0826-y).
- Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954–2003. *Wea. Forecasting*, **21**, 86–93, doi:[10.1175/WAF910.1](https://doi.org/10.1175/WAF910.1).
- 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, doi:[10.1175/2007WAF2007005.1](https://doi.org/10.1175/2007WAF2007005.1).
- , and Coauthors, 2015: The Mesoscale Predictability Experiment (MPEX). *Bull. Amer. Meteor. Soc.*, **96**, 2127–2149, doi:[10.1175/BAMS-D-13-00281.1](https://doi.org/10.1175/BAMS-D-13-00281.1).
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.
- Wood, V. T., and R. A. Brown, 1997: Effects of radar sampling on single-Doppler velocity signatures of mesocyclones and tornadoes. *Wea. Forecasting*, **12**, 928–938, doi:[10.1175/1520-0434\(1997\)012<0928:EORSOS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0928:EORSOS>2.0.CO;2).
- Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, doi:[10.1175/MWR-D-12-00237.1](https://doi.org/10.1175/MWR-D-12-00237.1).
- , D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, doi:[10.1175/MWR-D-14-00268.1](https://doi.org/10.1175/MWR-D-14-00268.1).
- Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Mon. Wea. Rev.*, **139**, 3489–3513, doi:[10.1175/MWR-D-10-05091.1](https://doi.org/10.1175/MWR-D-10-05091.1).