

# Blended Probabilistic Tornado Forecasts: Combining Climatological Frequencies with NSSL–WRF Ensemble Forecasts

BURKELY T. GALLO

*School of Meteorology, and NOAA/OAR/National Severe Storms Laboratory, and Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma*

ADAM J. CLARK

*NOAA/OAR/National Severe Storms Laboratory, and School of Meteorology, University of Oklahoma, Norman, Oklahoma*

BRYAN T. SMITH, RICHARD L. THOMPSON, AND ISRAEL JIRAK

*NOAA/NWS/NCEP/Storm Prediction Center, Norman, Oklahoma*

SCOTT R. DEMBEK

*NOAA/OAR/National Severe Storms Laboratory, and Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 5 September 2017, in final form 19 January 2018)

## ABSTRACT

Attempts at probabilistic tornado forecasting using convection-allowing models (CAMs) have thus far used CAM attribute [e.g., hourly maximum 2–5-km updraft helicity (UH)] thresholds, treating them as binary events—either a grid point exceeds a given threshold or it does not. This study approaches these attributes probabilistically, using empirical observations of storm environment attributes and the subsequent climatological tornado occurrence frequency to assign a probability that a point will be within 40 km of a tornado, given the model-derived storm environment attributes. Combining empirical frequencies and forecast attributes produces better forecasts than solely using mid- or low-level UH, even if the UH is filtered using environmental parameter thresholds. Empirical tornado frequencies were derived using severe right-moving supercellular storms associated with a local storm report (LSR) of a tornado, severe wind, or severe hail for a given significant tornado parameter (STP) value from Storm Prediction Center (SPC) mesoanalysis grids in 2014–15. The NSSL–WRF ensemble produced the forecast STP values and simulated right-moving supercells, which were identified using a UH exceedance threshold. Model-derived probabilities are verified using tornado segment data from just right-moving supercells and from all tornadoes, as are the SPC-issued 0600 UTC tornado probabilities from the initial day 1 forecast valid 1200–1159 UTC the following day. The STP-based probabilistic forecasts perform comparably to SPC tornado probability forecasts in many skill metrics (e.g., reliability) and thus could be used as first-guess forecasts. Comparison with prior methodologies shows that probabilistic environmental information improves CAM-based tornado forecasts.

## 1. Introduction

Discriminating a tornado threat from an overall severe convective threat poses a unique forecast challenge. Forecasters incorporate knowledge of internal storm dynamics and environments conducive to tornadogenesis, a thorough understanding of current observations, and numerical

weather prediction (NWP) guidance to forecast tornadoes. Until very recently, NWP guidance has been too coarse to depict specific storm modes, but recent expansion of computational resources has enabled models that explicitly depict convection and can thus provide specific information on mode, initiation, and evolution (Kain et al. 2008; Clark et al. 2012a; Gallo et al. 2017).

Several parameters have been associated with environmental conditions supportive of supercells, which can produce tornadoes. Supercell environments require

---

*Corresponding author:* Burkely T. Gallo, burkely.twiest@noaa.gov

enough convective available potential energy (CAPE) to maintain convection and strong deep-layer shear to create midlevel rotation (Weisman and Klemp 1982, 1984, 1986; Weisman and Rotunno 2000). Storms exhibiting at least some marginal supercell characteristics produce all types of severe convective weather (defined herein as hail  $\geq 2.54$  cm in diameter, thunderstorm wind gusts  $\geq 25$  m s<sup>-1</sup>, and tornadoes). However, distinguishing which storms in an environment will become tornadic is more difficult than determining if environmental conditions could support supercells and remains a large forecast challenge (Anderson-Frey et al.

2016). Environments conducive to supercell-based tornadogenesis typically have low lifted condensation levels (LCLs) and high 0–1-km storm-relative helicity (SRH; Rasmussen 2003; Craven and Brooks 2004; Thompson et al. 2012). Thompson et al. (2003) combined these parameters into the fixed-layer significant tornado parameter (STP), which attempts to distinguish significantly tornadic (EF2+) environments from nontornadic environments. The formulation was then updated by Thompson et al. (2012) to incorporate convective inhibition (CIN) and effective shear terms:

$$\text{STP} = (\text{MLCAPE}/1500 \text{ J kg}^{-1}) \times (\text{EBWD}/20 \text{ m s}^{-1}) \times (\text{ESRH1}/150 \text{ m}^2 \text{ s}^{-2}) \\ \times [(2000 \text{ m} - \text{MLLCL})/1000 \text{ m}] \times [(200 + \text{MLCIN})/150 \text{ J kg}^{-1}], \quad (1)$$

where mixed-layer CAPE (MLCAPE), mixed-layer CIN (MLCIN), and mixed-layer LCL (MLLCL) are the CAPE, CIN, and LCL calculated using the lowest 100-hPa mean parcel; EBWD is the effective bulk wind difference; and ESRH1 is the effective storm-relative helicity [calculated using the Bunkers et al. (2000) storm motion estimate]. If the STP is  $\geq 1.0$ , the environment is more supportive of significant tornadoes.

STP as a composite parameter also better discriminates between weak and significant right-moving supercellular (RM) tornadoes than individual thermodynamic or kinematic parameters (Thompson et al. 2013). Smith et al. (2015) examined tornadic storms from 2009 to 2013 within 101 mi of a WSR-88D, creating conditional probabilities of maximum hourly tornado intensity based on the maximum STP within 80 km of each tornadic storm. Larger STPs yielded generally stronger tornadoes in a gridpoint hour, further extending the application of STP as a discriminatory parameter.

While potential storm environment evolutions depicted by convection-parameterizing NWP help forecasters understand large-scale environmental conditions, key storm characteristics depend on smaller-scale features such as boundaries (Markowski et al. 1998; Boustead et al. 2013) and storm-to-storm interactions (e.g., Klees et al. 2016). These finescale details, which convection-allowing models (CAMs) can depict, often determine how the convective mode and subsequent hazards evolve (Fowle and Roebber 2003). CAMs also supply storm-scale metrics such as hourly maximum updraft helicity (UH; Kain et al. 2010), which has been successfully used as a midlevel (Kain et al. 2008; Clark et al. 2012b) and low-level (Sobash et al. 2016a) mesocyclone-scale rotation diagnostic. Swaths of

positive UH typically indicate simulated right-moving supercells (similarly, swaths of negative UH typically depict simulated left-moving supercells). Since supercells often generate severe weather reports, UH can indicate severe storm occurrence within both deterministic (Sobash et al. 2011) and ensemble frameworks (Sobash et al. 2016b).

Extending UH application from severe convective forecasting to tornado forecasting has begun in recent years. Taking a countrywide perspective, daily accumulated UH swaths positively correlate with total tornado pathlength over the continental United States (CONUS; Clark et al. 2013). On an individual storm level, Sobash et al. (2016a) argue that 0–3-km UH can serve as a tornado proxy by showing that simulated storms with strong low-level mesocyclone-scale rotation occur in simulated environments with STP and individual kinematic and thermodynamic parameters similar to observed proxy soundings from tornadic storm environments. Combining UH and environmental information can also help parse the tornado threat from the overall severe convective threat (Jirak et al. 2014; Gallo et al. 2016). Since simulated mesocyclones often occur in environments unfavorable to tornadogenesis (Clark et al. 2012b), environmental criteria can reduce false alarms by limiting probabilistic tornado forecasts to favorable environments (Rasmussen and Blanchard 1998; Rasmussen 2003; Thompson et al. 2003; Grünwald and Brooks 2011; Grams et al. 2012; Thompson et al. 2012; Thompson et al. 2013). Indeed, both coarse-scale (Jirak et al. 2014) and finescale (Gallo et al. 2016) environmental information demonstratively improves tornado guidance skill beyond forecasts generated solely using UH.

This work blends CAM environmental and storm-scale output with observed, empirical frequencies of a tornado of any intensity given environmental characteristics from right-moving supercells. [Smith et al. \(2015\)](#) developed initial frequencies from environmental tornado climatologies, which [Thompson et al. \(2017\)](#) improved upon by determining the frequency of a tornado given a right-moving supercell [as defined by [Smith et al. \(2012\)](#)] with a local storm report (LSR) using data from 2014 and 2015. By applying these observed frequencies to the NWP output, this study creates forecasts resembling Storm Prediction Center (SPC) convective outlooks using a paradigm that represents each point as having a probability of tornado occurrence rather than assuming a tornado if deterministic attribute thresholds are exceeded. This process was also designed to reduce the overforecasting seen in prior probabilistic tornado forecasts ([Jirak et al. 2014](#); [Gallo et al. 2016](#); [Sobash et al. 2016a](#)) by constraining the magnitude of the probabilities to the observed frequencies roughly based on the environmental probabilities from [Thompson et al. \(2017\)](#). The forecasts produced by this methodology are also compared to other methods of probability generation described in the literature, including using 2–5-km UH or 0–3-km UH as a tornado proxy sans environmental information [as in [Sobash et al. \(2016a\)](#)], or by

requiring that 2–5-km UH exist in an environment exceeding a threshold of STP [as in [Gallo et al. \(2016\)](#)].

[Section 2a](#) of this paper describes the modified STP used throughout this study, which is a surface-based parcel and fixed-layer shear version of the effective-layer STP ([Thompson et al. 2012](#)). [Section 2b](#) describes the empirical climatological frequency generation, while [section 2c](#) outlines the ensemble system and probabilistic forecast generation algorithm. [Sections 2d](#) and [2e](#) specify SPC forecasts and objective verification metrics used in this study, respectively. Determination of the optimum STP percentile composes [section 3a](#), while [section 3b](#) compares four probability generation methods and the 0600 UTC SPC forecasts. Case studies in [section 3c](#) illustrate the daily tornado probabilities on two high-end days and a more marginal day. Finally, [section 4](#) summarizes and discusses the results and future research directions.

## 2. Data and methodology

### a. STP formulation

The STP calculation herein uses surface-based parcels and fixed-layer calculations within the effective-layer STP equation ([Thompson et al. 2012](#)):

$$\text{STP} = (\text{SBCAPE}/1500 \text{ J kg}^{-1}) \times (\text{SHR6}/20 \text{ m s}^{-1}) \times (\text{SRH1}/150 \text{ m}^2 \text{ s}^{-2}) \\ \times [(2000 \text{ m} - \text{SBLCL})/1000 \text{ m}] \times [(200 \text{ J kg}^{-1} + \text{SBCIN})/150 \text{ J kg}^{-1}], \quad (2)$$

where the SBCAPE, SBLCL, and SBCIN are the surface-based CAPE, LCL, and CIN. As in the fixed-layer STP, the CAPE and LCL height are calculated from surface-based parcels because of availability constraints within the National Severe Storms Laboratory (NSSL)–Weather Research and Forecasting Model (WRF; [Skamarock et al. 2008](#)) ensemble, and the shear and SRH are computed from fixed layers. Similar to the effective-layer STP, the modified STP includes CIN, albeit calculated from the surface-based parcel rather than the 100-mb (1 mb = 1 hPa) mixed-layer parcel. Additionally, the capping terms [e.g., if  $\text{SHR6} < 12.5 \text{ kt}$  (1 kt =  $0.51 \text{ m s}^{-1}$ ), the SHR6 term is set to zero] are taken from the effective-layer STP. This STP formulation utilizes improvements within the effective-layer STP while balancing the computational expense of running a CONUS-wide CAM ensemble (i.e., the inability to calculate the effective-layer inflow for each grid point and time on a 4-km grid efficiently).

### b. Tornado frequency calculation

The climatological frequency of tornado occurrence was calculated following [Thompson et al. \(2017\)](#), but using the modified STP formulation described in [section 2a](#). LSRs from 1 January 2014 to 31 December 2015 were filtered in three ways: 1) all tornado reports were filtered by maximum EF scale per 40-km grid hour,<sup>1</sup> 2) all hail/wind reports were required to meet effective bulk wind difference ([Thompson et al. 2007](#)) criteria ( $>20 \text{ kt}$  for 2014,  $>40 \text{ kt}$  for 2015<sup>2</sup>), and 3) a convective mode filter ensured that only right-moving supercells

<sup>1</sup> This study does not use intensity information; this step was performed such that the most intense tornado supported by each environment was used.

<sup>2</sup> The more strict effective bulk wind difference criteria for 2015 were estimated to reduce the number of potential 40-km grid hour events by ~35% for 2015 based on 2014 data, thereby reducing the workload while capturing a majority of the low-level circulations within the sample. For further details, see [Thompson et al. \(2017\)](#).

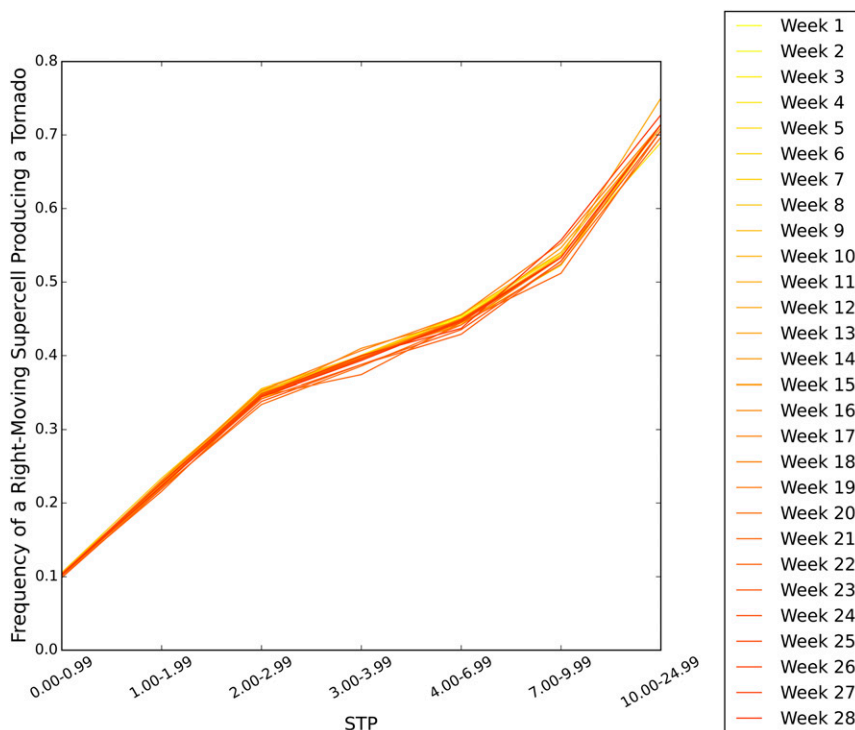


FIG. 1. The climatological frequency of tornadoes given a right-moving supercellular storm associated with an LSR and a given modified fixed-layer STP using all data from 1 Feb 2014 to 31 Dec 2015 except the week indicated in the legend. Week 1 begins on 30 Mar 2014, week 14 begins on 29 Jun 2014, week 15 begins on 29 Mar 2015, and week 28 begins on 28 Jun 2015.

and right-moving marginal supercells were included. The supercell definition required an azimuthal velocity difference of  $\geq 10 \text{ m s}^{-1}$  across less than  $\sim 7 \text{ km}$  throughout more than one-quarter of the storm's depth for at least 10–15 min (Smith et al. 2012). After filtering, 1202 tornadic cases and 5422 nontornadic cases were used to generate the climatological frequencies. To ensure separation of the training and testing datasets, weekly frequencies were generated withholding the reports for that week. Each week's frequencies were then used in probability generation. This cross-validation technique (Elsner and Schmertmann 1994) has previously been applied to surrogate severe probabilities (Sobash and Kain 2017). Hourly SPC objective analyses (Bothwell et al. 2002) provided the nearest 40-km gridpoint-modified STP assigned to each event. The weekly climatological tornado frequency in each STP bin equaled the tornadic storm count divided by the total number of storms in that bin (Fig. 1). Variability in the equations was largest at high STP values, which have more limited sample sizes than lower STP values.

### c. Probabilistic forecast generation

Probabilistic tornado forecasts were generated using output from a 4-km horizontal grid-spacing ensemble

based around an experimental version of the NSSL–WRF ensemble using the Advanced Research version of WRF (WRF-ARW; Kain et al. 2010). The NSSL–WRF ensemble contains the NSSL–WRF and nine additional members with varied initial conditions (ICs) and lateral boundary conditions (LBCs) (Gallo et al. 2016; Clark 2017; Table 1). Ensemble runs began in February 2014 and produce forecasts to 36 h beginning at 0000 UTC. Probabilistic tornado forecasts were generated for the spring seasons (defined as 1 April–30 June) of 2014 and 2015; seasonal statistics are aggregated over that time. The probabilistic forecasts herein are intended as automated first-guess tornado forecasts for 12–36-h lead time covering the day 1 period defined by the SPC.

Ensemble membership shifted slightly between June 2014 and April 2015, exchanging two members initialized from Eulerian mass (EM) Short-Range Ensemble Forecast (SREF) members for two members initialized from Non-hydrostatic Multiscale Model on the B grid (NMMB) SREF members. This change occurred when SPC forecasters noticed tight clustering within the EM SREF members compared to other subsets. The ensemble membership shift has minimal impact on subsequent tornado forecasts (Gallo et al. 2016), and therefore the 2014 and 2015 spring seasons are combined.

TABLE 1. A summary of the NSSL–WRF ensemble configurations with differing LBCs and ICs. All members use WRF single-moment 6-class (WSM-6; Hong and Lim 2006) microphysics, the Dudhia shortwave radiation scheme (Dudhia 1989), the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) longwave radiation, the Noah land surface model (Chen and Dudhia 2001), and the Mellor–Yamada–Janjić (MYJ; Mellor and Yamada 1982; Janjić 2002) PBL scheme. Members with years in parentheses by the ensemble member were only part of the ensemble for that year. Aside from the control NSSL–WRF member and GFS member, members are initialized using 3-h SREF member forecasts initialized at 2100 UTC for the ICs and LBCs.

Ensemble member	ICs/LBCs	Microphysics	PBL	Radiation	Land surface
1	00Z NAM	WSM6	MYJ	RRTM/Dudhia	Noah
2	0000 UTC GFS	WSM6	MYJ	RRTM/Dudhia	Noah
3	2100 UTC em_ctl	WSM6	MYJ	RRTM/Dudhia	Noah
4	2100 UTC nmb_ctl	WSM6	MYJ	RRTM/Dudhia	Noah
5	2100 UTC nmb_p1	WSM6	MYJ	RRTM/Dudhia	Noah
6	2100 UTC nmm_ctl	WSM6	MYJ	RRTM/Dudhia	Noah
7	2100 UTC nmm_n1	WSM6	MYJ	RRTM/Dudhia	Noah
8	2100 UTC nmm_p1	WSM6	MYJ	RRTM/Dudhia	Noah
9 (2015)	2100 UTC nmb_n1	WSM6	MYJ	RRTM/Dudhia	Noah
10 (2015)	2100 UTC nmb_p2	WSM6	MYJ	RRTM/Dudhia	Noah
11 (2014)	2100 UTC em_n1	WSM6	MYJ	RRTM/Dudhia	Noah
12 (2014)	2100 UTC em_p1	WSM6	MYJ	RRTM/Dudhia	Noah

This work compares four methods of probabilistic forecast generation. Method 1 uses 2–5-km  $UH \geq 75 \text{ m}^2 \text{ s}^{-2}$  as a coarse proxy for tornado occurrence from the daily maximum UH field of each member, as in Gallo et al. (2016) and following the Hamill and Colucci (1998) method for calculating probabilities. Each member has a distribution of UH values from the daily maximum UH within a 40-km radius of a point, and probabilities are generated by determining where  $75 \text{ m}^2 \text{ s}^{-2}$  occurs within the distribution. Methods 2 and 3 are similar but use 2–5-km  $UH \geq 75 \text{ m}^2 \text{ s}^{-2}$  only at points where the preceding hour had  $STP \geq 1$  or use 0–3-km  $UH \geq 33 \text{ m}^2 \text{ s}^{-2}$ , respectively. The 0–3-km threshold was chosen by determining the percentile of 2–5-km UH corresponding to  $75 \text{ m}^2 \text{ s}^{-2}$  during the study period and the subsequent value of 0–3-km UH at that percentile. These three methods are derived from those previously explored in the literature and solely use output from CAM ensembles.

The final probabilistic tornado forecast method (i.e., method 4) combines ensemble information and the observed climatological frequencies described in section 2b (Fig. 2). First, forecast hours 12–36 of each ensemble member are checked for 2–5-km  $UH \geq 25 \text{ m}^2 \text{ s}^{-2}$ , indicating a right-moving supercell (Clark et al. 2013; Gallo et al. 2016; Sobash et al. 2016b). If a grid point exceeds the UH threshold, the STP from the prior hour is collected from every point where the threshold is exceeded within a 40-km radius, creating an STP distribution at each grid point and for each hour. From these STP distributions, a percentile value is extracted and assigned to the grid point and hour. The percentiles examined herein are the 10th, 25th, 50th, 75th, 90th, and 100th (maximum value). Once each grid point and hour has an STP value, the daily maximum STP is assigned to

the point, representing the most favorable environment over a 24-h period. The climatological frequency values are then used to assign an STP-based tornado probability at that grid point. The calculated climatological frequency values (Fig. 1) represent the centerpoint of their bins, and linear interpolations between the bin centers assign frequencies between centerpoints.

The final step averages the individual member probabilities and smooths the resultant field using a Gaussian kernel density weighting function with weights determined by

$$f(x, y) = \frac{1}{2\pi(\sigma/\Delta x)^2} \times \exp\left[\frac{-(x^2 + y^2)}{2(\sigma/\Delta x)^2}\right], \quad (3)$$

where  $\sigma$  is the user-defined standard deviation (km), and  $\Delta x$  is the grid spacing. Varying values of  $\sigma$  were tested (not shown), and  $\sigma = 50 \text{ km}$  creates a field of comparable resolution to SPC tornado probabilities.

#### d. SPC forecasts

All ensemble probabilities were verified in conjunction with the initial SPC day 1 tornado probabilities issued at 0600 UTC (valid 1200–1159 UTC the following day) to compare the skill of the first-guess probabilities and initial SPC tornado forecasts. For these probabilities to become a useful first-guess forecast, the resolution and accuracy should resemble the SPC forecasts. The SPC issues 0600 UTC tornado forecasts using information from 0000 UTC, making them the most applicable comparison to the first-guess forecasts since the ensemble initializes at 0000 UTC. The outlooks herein were largely independent of the NSSL–WRF ensemble probabilities, as the ensemble fields were unavailable to

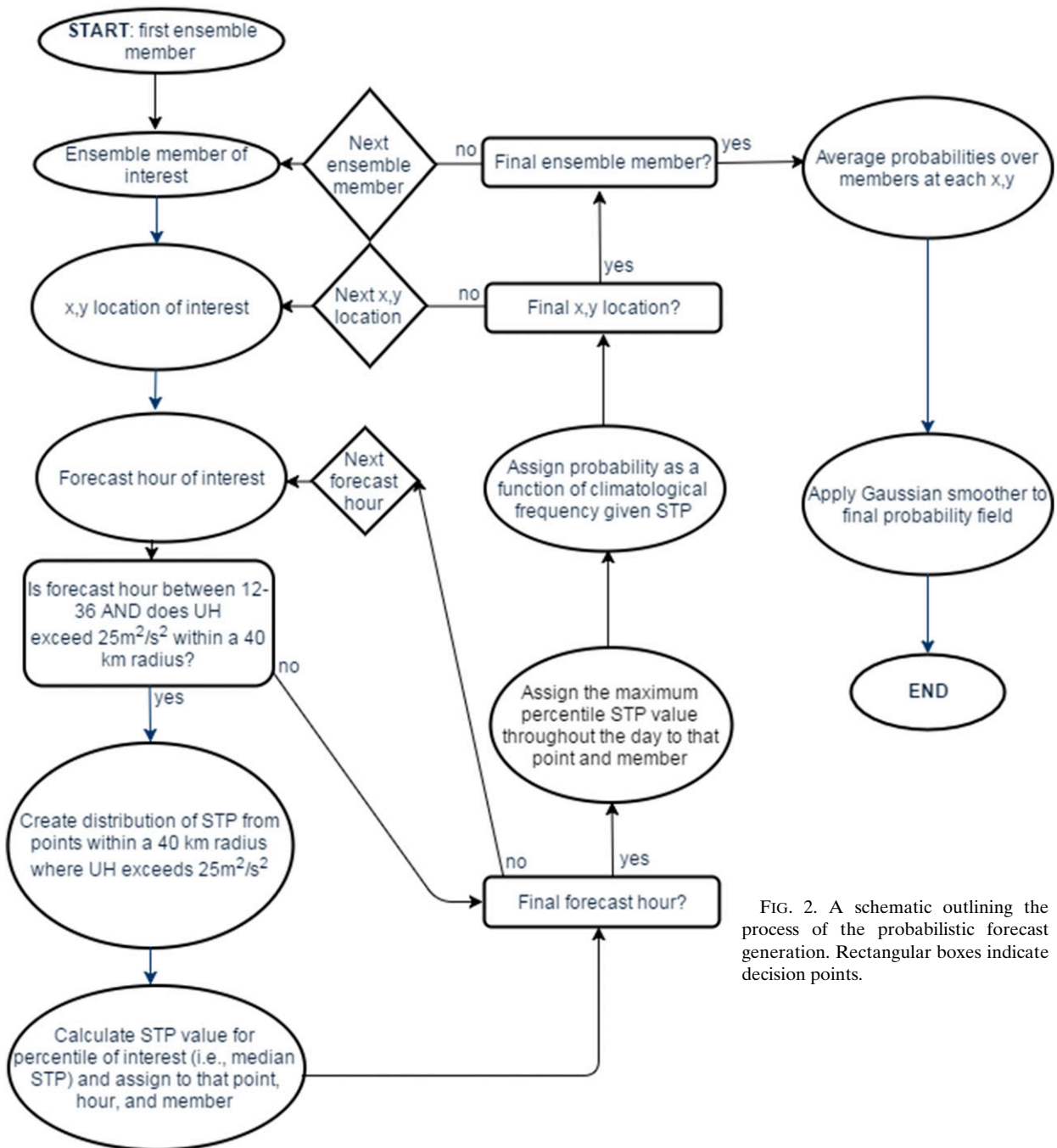


FIG. 2. A schematic outlining the process of the probabilistic forecast generation. Rectangular boxes indicate decision points.

forecasters producing the 0600 UTC outlooks. The SPC probabilities were regridded to the NSSL–WRF grid before verification, ensuring consistency between the ensemble and SPC forecasts.

#### e. Verification

Verification occurred across approximately the eastern two-thirds of the CONUS (Fig. 3). All probabilities (NSSL–WRF and SPC) were considered only within this

domain and over the 182 days of April–June 2014 and 2015. Tornado path data were georeferenced to the 4-km grid of the NSSL–WRF ensemble and treated as binary yes/no events. Yes events occurred if a tornado passed within 40 km of a point. Though the severe report database has documented shortcomings regarding tornado reports (Doswell and Burgess 1988; Brooks et al. 2003; Verbout et al. 2006; Doswell et al. 2009) and hail reports (Blair et al. 2017), more low-magnitude tornadoes

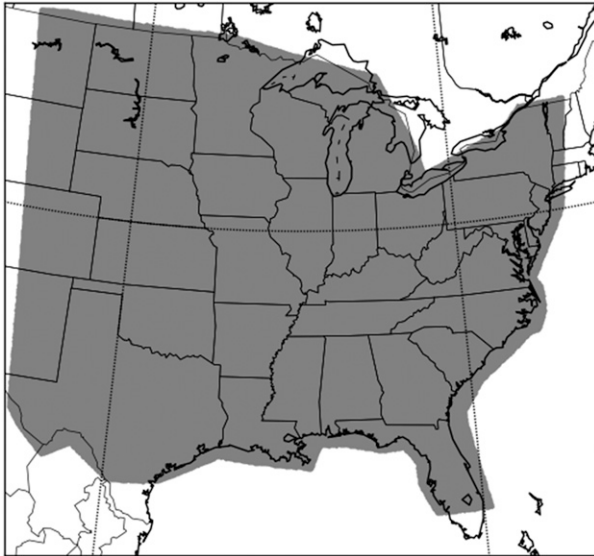


FIG. 3. A subset of the model domain for the NSSL–WRF ensemble showing where objective verification measures were computed (shaded region)

have been reported in recent decades (Brooks and Doswell 2001).

Two subsets of the tornado database were considered for this project. The first subset included tornado path data from all modes of parent convection. The second subset solely included tornadoes produced by either right-moving supercells or marginal right-moving supercells (RM tornadoes). Since the new methodology derives probabilities from observed climatological frequencies of RM tornadoes, applying the forecasts to the second subset is truer to the underlying data than using them as forecasts of all tornadoes. Comparing the verification methods may help determine whether the probabilities are appropriate as tornado forecasts or should solely be considered a forecast of RM tornadoes. The other methods previously documented in the literature were also verified with both datasets.

Forecasts were verified using reliability diagrams (Wilks 2011), performance diagrams (Roebber 2009), and the area under the receiver operating characteristic (ROC) curve, which measures the ability of a forecast to discern an event from a nonevent by plotting the probability of detection (POD) against the probability of false detection (POFD) at different thresholds. POD and POFD were generated using a standard  $2 \times 2$  contingency table and are defined as

$$POD = \frac{\text{hits}}{\text{hits} + \text{misses}} \tag{4}$$

and

$$POFD = \frac{\text{false alarms}}{\text{false alarms} + \text{correct negatives}}. \tag{5}$$

One POD and one POFD were defined for each probabilistic tornado forecast threshold that the SPC issues: 2%, 5%, 10%, 15%, 30%, 45%, and 60%. The model forecast verification occurred at these thresholds to enable comparisons. The area under the curve was then computed using the trapezoidal method (Wandishin et al. 2001). ROC areas range from 0.0 to 1.0, where 1.0 indicates a perfect forecast, and 0.5 is the skill of a random forecast. Generally, a score of 0.7 or higher is considered skillful (Buizza et al. 1999).

The ROC area difference between the SPC forecasts and ensemble forecasts was tested for statistical significance using resampling, following Hamill (1999). All cases were randomly assigned to one of the two forecasts, seasonally aggregated ROC areas were calculated for the two groups, and the difference was computed 1000 times to create a ROC area difference distribution. Significant ROC area differences between the SPC forecasts and the NSSL–WRF ensemble forecasts fell outside of the 95% confidence interval of this subsequent distribution.

Reliability diagrams plot the observed relative frequency against the forecast probability, providing information about bias to supplement the ROC areas, which are insensitive to bias. A perfect forecast follows the 45° diagonal: when there is a 40% probability of a tornado, a tornado observation occurs in 4 out of 10 forecasts. The SPC’s forecasts largely occur at low probabilities and are only issued at specific thresholds: forecasters typically assume some higher probabilities exist within the contours that do not exceed the following threshold. For example, the 15% contour may contain probabilities as high as 29.99%, since 30% is the next probabilistic contour issued. Thus, SPC forecasts by design underforecast according to the reliability diagram, resulting in values that are above the diagonal. Conversely, overforecasting results in values beneath the diagonal.

Performance diagrams visualize four different statistical metrics including the critical success index, defined as

$$CSI = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{false alarms}}. \tag{6}$$

This is typically a rare-event score (Wilks 2011) and has verified prior tornado forecasts (Gallo et al. 2016; Sobash et al. 2016a). It ranges from 0.0 to 1.0, with 1.0

TABLE 2. Area under the ROC curve statistics for ensemble-generated forecasts based on differing percentiles of STP. Bold-faced seasonally aggregated areas under the ROC curve are statistically significantly different from the SPC area under the ROC curve at  $\alpha = 0.05$ . Numbers outside the parentheses were verified using all tornadoes; values within the parentheses used solely RM tornadoes.

STP percentile	Seasonally aggregated ROC area
10th	0.845 (0.879)
25th	0.855 (0.889)
Median	<b>0.868 (0.902)</b>
75th	<b>0.878 (0.911)</b>
90th	<b>0.884 (0.916)</b>
Max	<b>0.890 (0.921)</b>
SPC	0.824 (0.865)

being a perfect score. Performance diagrams plot POD versus success ratio (SR), defined as

$$SR = 1 - \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}, \quad (7)$$

with lines of constant CSI and bias to aid in interpretation. Reliability information at each threshold can also be extracted (i.e., ideally an SR of 15% would occur at the 15% forecast threshold).

### 3. Results

#### a. STP percentile sensitivity

The seasonally aggregated SPC 0600 UTC tornado forecasts had ROC areas of 0.824 for all tornadoes and 0.865 for RM tornadoes, respectively (Table 2), showing that the SPC is more skillful at forecasting RM tornadoes than tornadoes from other convective modes. However, both subsets easily exceed the 0.7 criteria determining a skillful forecast. Similarly, the ensemble-based probabilities achieved skillful ROC areas for all tested percentiles, ranging from a low score of 0.845 for probabilities using the 10th percentile of STP and verified on all tornadoes to a high score of 0.921 for probabilities using the maximum STP and verified on RM tornadoes (Table 2). Across all percentiles, verification on RM tornadoes scored higher than verification on all tornadoes, indicating that the forecasts were more adept at discerning areas of RM tornadoes. Given the underlying climatological frequencies and the strong correlation between UH and supercells, the probabilities were expected to particularly highlight areas where RM tornadoes may occur. Higher STP percentiles attained significantly higher ROC areas than the SPC, likely because of their broader coverage as a harsh penalty is imposed by the ROC area when missing a tornado report (Gallo et al. 2016).

ROC curves for all STP percentiles had higher POD and POFD values than the SPC forecasts, particularly at lower forecast thresholds such as 2% (Figs. 4a,d). The curves also showed that the increase in ROC area at higher STP percentiles comes mostly from increased POD at the 2% and the 5% thresholds. Above the 5% threshold, the POD and the POFD values were nearly indistinguishable from the SPC's forecasts. Thus, STP-based ensemble forecasts could provide forecasters with objectively skillful first-guess tornado probabilities, particularly for RM tornadoes, with the understanding that at low thresholds the improvement in POD is accompanied by a slightly higher POFD. The largest difference between verifying with all tornadoes and RM tornadoes stemmed from the POD difference at low-probability thresholds, with all forecasts having a higher POD for RM tornadoes than for all tornadoes.

Since the ROC area solely distinguishes events from nonevents, forecast reliability is key in determining the practical usefulness of the probabilities. Reliability diagrams showed that the ensemble-based probabilities closely resembled the SPC forecasts when they were generated using the 10th percentile of STP (Figs. 4b,e). Higher percentiles overforecasted all tornadoes, especially at low probabilities (Fig. 4b); only the 10th percentile forecast was nearly reliable until the 30% forecast probability. When forecasting RM tornadoes, overforecasting increased, and the 10th percentile remained most reliable (Fig. 4e). The increase in overforecasting when looking at RM tornadoes compared to all tornadoes was expected, since the RM constraint ensures fewer tornadoes in the verification dataset.

Performance diagrams allow a closer examination of individual probabilistic forecast thresholds. Since tornadoes rarely occur, the ideal forecast would contain a majority of tornadoes with limited false alarms, leading to an SR equal to the probability at each probability threshold. At nearly all percentiles and probability thresholds, the ensemble forecasts had a higher POD and a lower SR than the SPC probabilities (Figs. 4c,f). An exception occurred with the probabilities generated using the 10th percentile of STP for the 10% or 15% threshold, when the ensemble forecasts had higher PODs and higher SRs than the SPC forecasts. SPC forecasts of 10% and 15% are reserved for high-impact days, and so these thresholds warrant special attention.

Performance diagram results were consistent between all tornadoes (Fig. 4c) and RM tornadoes (Fig. 4f), but the RM tornadoes generally had a lower CSI despite having an increased ROC area. Since RM tornadoes are a subset of all tornadoes, when verifying solely on RM tornadoes, the false alarm and correct negatives will increase, the misses will decrease, and at best the number of hits will remain the same (if the probabilities



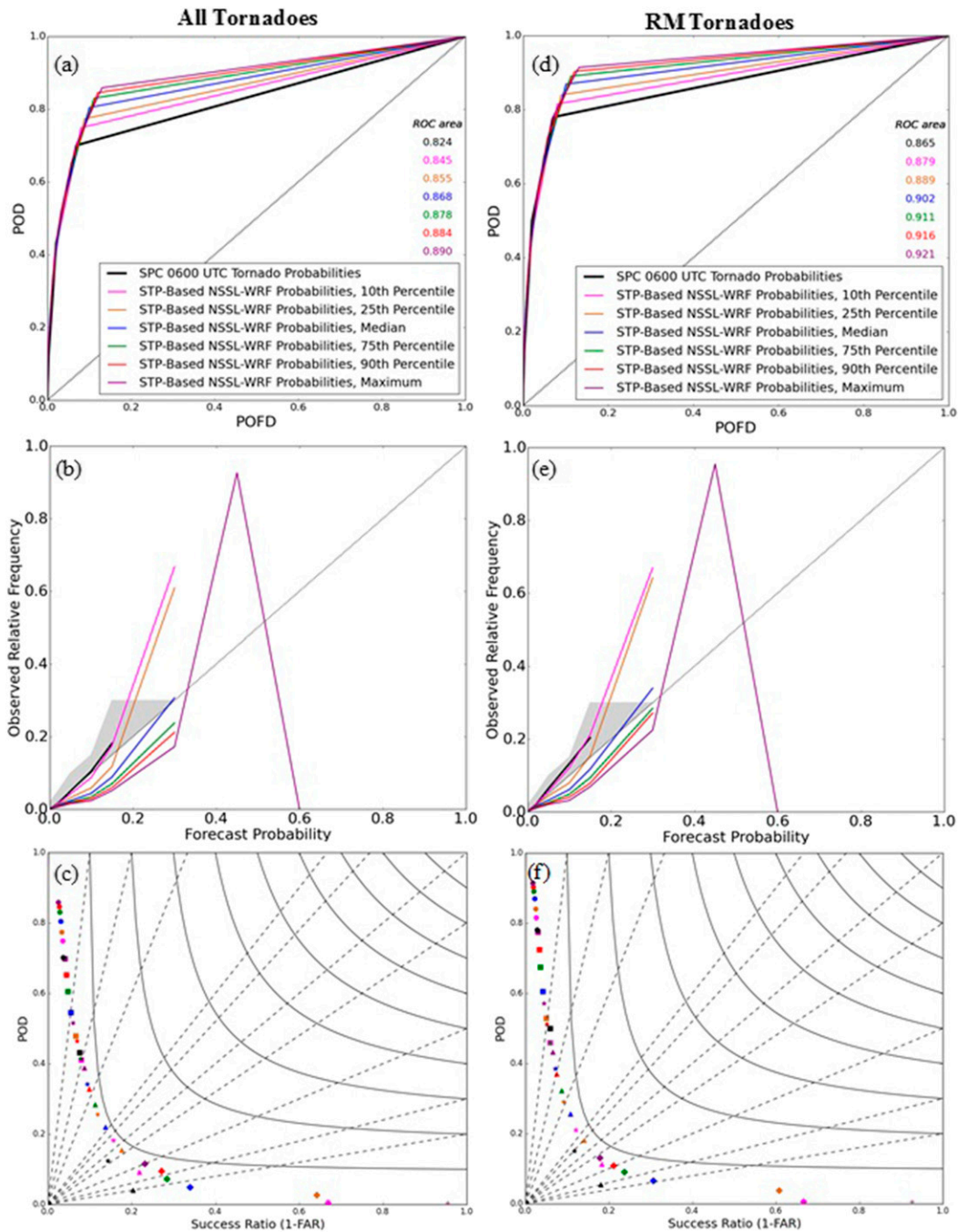


FIG. 4. Summary statistics for different percentiles of STP used to calculate the STP-based NSSL-WRF ensemble probabilities: (a),(d) seasonally aggregated ROC curves annotated with the areas under the ROC curve, (b), (e) reliability diagrams, and (c),(f) performance diagrams for (left) all tornadoes and (right) RM tornadoes. Colors represent percentiles of STP used in probability generation. Black lines and symbols represent the SPC 0600 UTC forecasts. In (a) and (d), the thin black line indicates the performance of a random forecast, while in (b) and (e), it represents perfect reliability. In (c) and (f), the different symbols represent the different probability thresholds: circles, squares, stars, triangles, and diamonds represent 2%, 5%, 10%, 15%, and 30%, respectively. Black dashed lines indicate constant bias, while solid black lines are for constant CSI.

TABLE 3. Area under the ROC curve statistics for different methods of generating ensemble-based probabilities. Boldfaced seasonally aggregated areas under the ROC curve are statistically significantly different from the SPC area under the ROC curve at  $\alpha = 0.05$ . Numbers outside the parentheses were verified using all tornadoes; values within the parentheses used solely RM tornadoes.

Method	Seasonally aggregated ROC area
2–5-km UH, unfiltered	<b>0.867 (0.900)</b>
0–3-km UH, unfiltered	<b>0.889 (0.919)</b>
2–5-km UH, filtered by STP $\geq 1$	0.810 (0.848)
STP based, 10th percentile	0.845 (0.879)
SPC	0.824 (0.865)

are encompassing all RM tornadoes) or decrease. In a rare-event scenario, false alarms are often the largest term in the CSI (compared to hits and misses), and the increased false alarm of verifying on RM tornadoes decreases the CSI. False alarms affect CSI more than the ROC area because the CSI does not incorporate correct negatives. False alarms are incorporated in the ROC area through the POFD, which is overwhelmingly dominated by correct negatives in the rare-event scenario. The ROC area is instead sensitive to the POD and increases because of the decreased misses.

#### b. Probability generation method comparison

The probabilities generated using the 10th percentile of STP were the most reliable while maintaining high skill, so those forecasts were compared with other methodologies of probability generation (Gallo et al. 2016; Sobash et al. 2016a). From this point, the STP-based probabilities denote the probabilities computed using the 10th percentile of STP. Seasonally aggregated ROC areas between the 0–3-km UH-only, 2–5-km UH-only, and STP-based probabilities were similar, while the filtered 2–5-km UH had a much lower ROC area. However, neither the filtered 2–5-km nor the STP-based method was statistically significantly different from the SPC forecasts for either verification dataset (Table 3). Across both verifications, ROC curves of the UH-only methods had higher POD and POFD values at low-probability thresholds than methods incorporating the STP (Fig. 5). The filtered 2–5-km UH method had lower POFDs than the other methods, as well as a much lower POD than the other methods and the SPC forecasts. The STP-based probabilities had a slightly lower POD than the UH-only methods, but also had a lower POFD that more closely resembles the SPC forecasts. The most obvious difference between the RM tornado verification and the all-tornado verification was that the RM tornadoes produced higher ROC areas than the all-

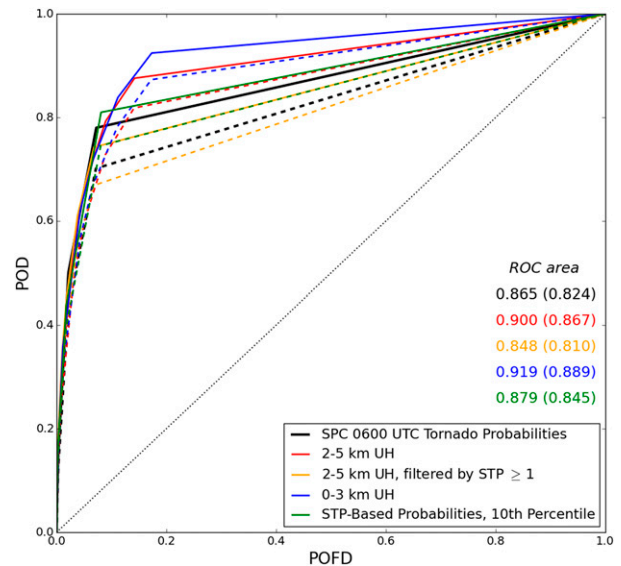


FIG. 5. ROC curves for different probabilistic tornado forecasting methods, annotated with the area under the ROC curve for RM tornadoes (all tornadoes). Different colors represent the different methods. Solid lines are verified using only RM tornadoes, while dashed lines are verified using all tornadoes. The dotted black line indicates the ROC area of a random forecast.

tornado dataset across all methods, mostly as a result of an increase in POD at low thresholds. Otherwise, the results were consistent between verifications.

The methods differed immensely in their reliability (Fig. 6). High SPC forecast probabilities are rare, and unnecessarily high first-guess ensemble probabilities can mislead forecasters trying to anticipate the severity of a day (Gallo et al. 2016). Vast overforecasting occurred in the methods solely using UH despite their high ROC areas, and verification using only the RM tornado dataset exacerbated this signal. Filtering the 2–5-km UH probabilities by requiring STP  $\geq 1$  improved reliability, but still overforecasted. The STP-based probabilities, however, were remarkably reliable, particularly when forecasting RM tornadoes. The SPC was also extremely reliable for both verification methods. Indeed, the SPC forecasts achieved nearly perfect reliability up to 15% when forecasting RM tornadoes, while the STP-based probabilities overforecasted at 10% and below. Clearly, using empirical observations as a basis for the probabilistic tornado forecasts improved reliability over the other methods, which solely rely on an ensemble and Gaussian smoother to moderate the probabilities.

A performance diagram illustrates the verification statistics at SPC forecast thresholds (Fig. 7). At the 2% level the UH-only and STP-based methods have similar SRs, although the STP-based method had higher CSI and lower POD than the UH-only methods. However,

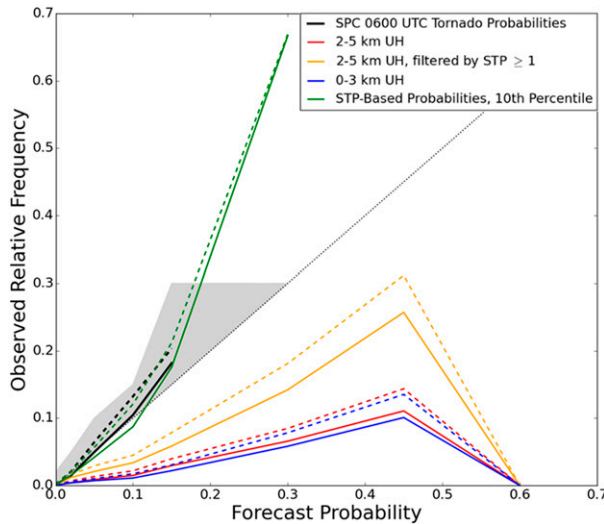


FIG. 6. Reliability diagrams for different probabilistic tornado forecast methods. Different colors represent the different methods. Dashed lines are verified on all tornadoes, and solid lines are verified solely on RM tornadoes. The dotted black line indicates perfect reliability. The shaded region represents where categorical forecasts currently issued by the SPC are reliable (e.g., the 2% forecast encompasses areas from 2% to 4.99%).

beginning at the 5% level, all methods except the STP-based probabilities have much higher PODs and lower SRs than the SPC forecasts. At the 10% and 15% thresholds, the STP-based probabilities have higher CSI, POD, and SR values than the SPC forecasts for all tornadoes and for RM tornadoes, although the increase in SR was larger for all tornadoes than for RM tornadoes. As the probability threshold increases, so do the discrepancies between the methods, with the UH-based methods having much higher PODs and much lower SRs than the SPC and the STP-based method and corresponding to their high bias. Therefore, the STP-based method performs better than all other first-guess methods for a given threshold and even scores higher than the SPC at high-impact thresholds.

c. Case studies

To demonstrate how the probabilities appear to a forecaster, three case studies are now presented. The first illustrates a high-impact day, with high probabilities and multiple tornadoes. The second highlights an area where forecast upscale growth contained embedded supercells, emphasizing that these probabilities are intended as a tool for forecasting supercellular tornadoes. The final case occurred on a more marginal day and had a relatively large false alarm area.

1) 28 APRIL 2014

Late April 2014 saw a multiday outbreak spanning from the Great Plains to the East Coast, with the most

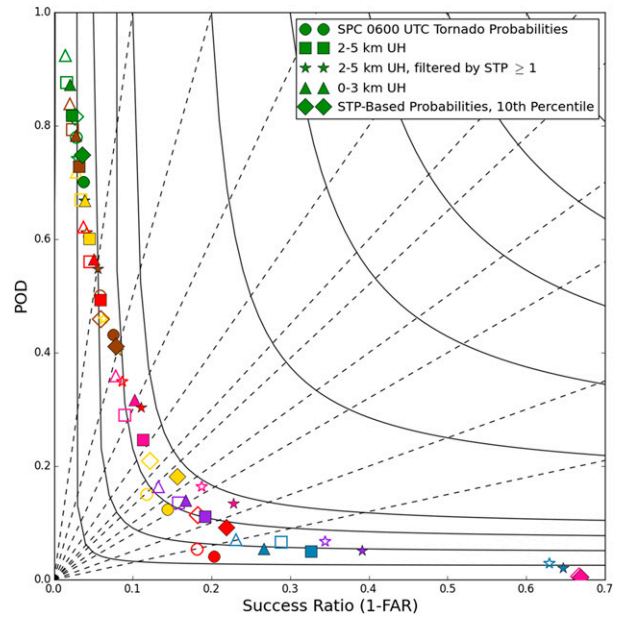


FIG. 7. Performance diagrams for the forecast tornado probabilities. Different colors indicate different probability thresholds. Green, brown, yellow, red, pink, purple, and blue represent 2%, 5%, 10%, 15%, 30%, 45%, and 60%, respectively. Filled shapes are verified on all tornadoes; hollow shapes are verified on RM tornadoes. Black dashed lines are lines of constant bias, while solid black lines are lines of constant CSI. SR is defined in terms of the false alarm ratio (FAR).

tornadoes occurring on 28 April. In fact, this day had the largest number of tornadoes (121) of any day in our dataset. Four of these tornadoes caused 15 deaths across Mississippi, Alabama, and Tennessee. On 28 April, a 500-mb closed low was located over Nebraska and a negatively tilted short-wave trough stretched from the central Great Plains into eastern Oklahoma and Louisiana. At the base of this trough, a 500-hPa jet streak with wind speeds exceeding 80 kt existed over Arkansas and moved eastward throughout the day. Thermodynamic parameters were also favorable, with MLCAPE exceeding  $2000 \text{ J kg}^{-1}$  where tornadoes would later occur. Objectively analyzed STP ranged from 3.0 to 6.0 in the area of interest (not shown).

The SPC forecasted this event well in advance, issuing a day 3 moderate risk. The SPC’s 0600 UTC tornado probabilities (Fig. 8a) had a broad area of 15% probability, corresponding to a “moderate” categorical risk. The 2000 UTC update to this forecast increased the tornado probabilities to 30% (not shown), leading to a categorical upgrade to high risk. The 0600 UTC SPC-issued probabilities successfully captured the largely RM tornado reports for that day, and most of the tornadoes occurred in the upper-tier probabilities. The NSSL–WRF ensemble also highlighted the Southeast,

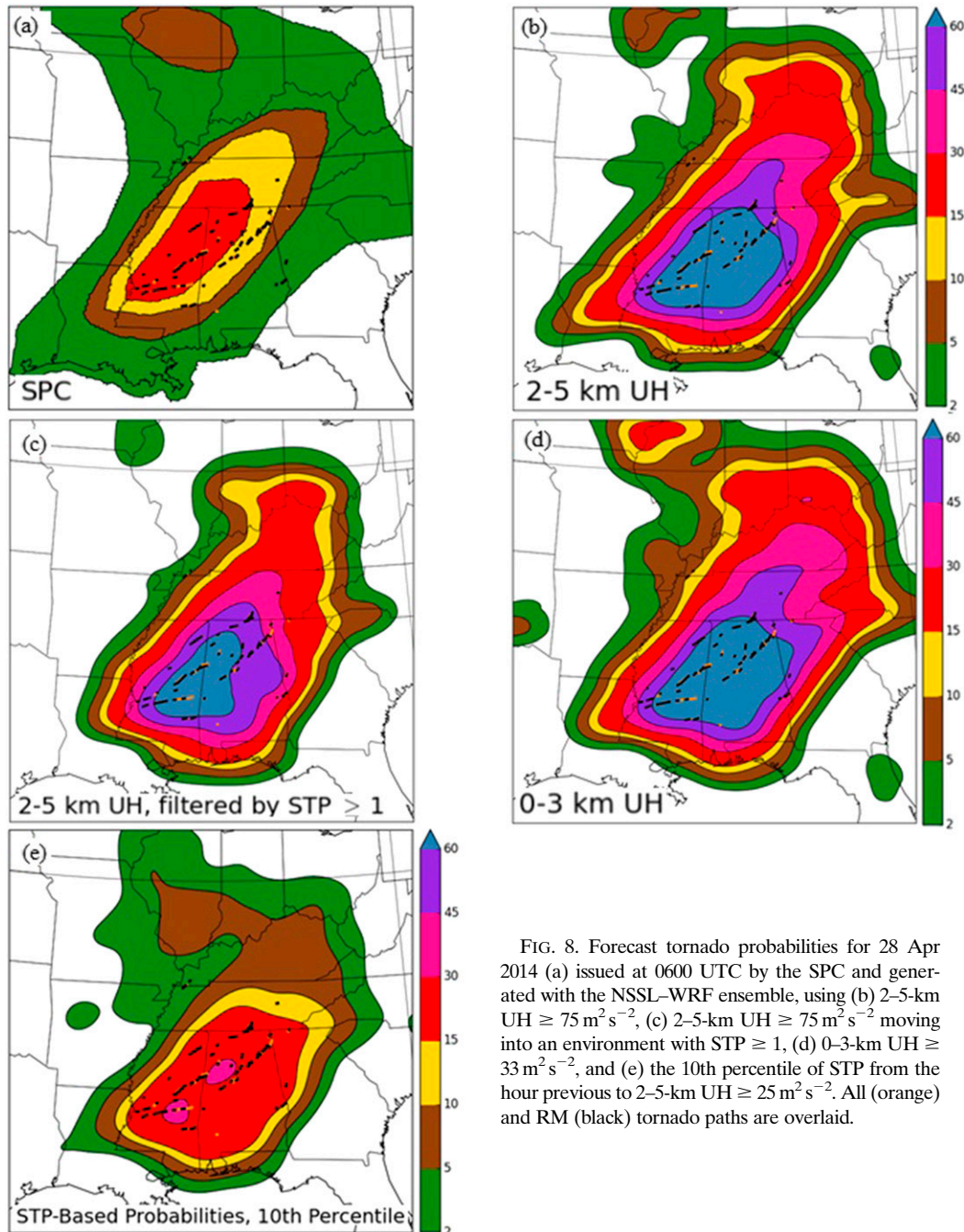


FIG. 8. Forecast tornado probabilities for 28 Apr 2014 (a) issued at 0600 UTC by the SPC and generated with the NSSL-WRF ensemble, using (b) 2–5-km UH  $\geq 75 \text{ m}^2 \text{ s}^{-2}$ , (c) 2–5-km UH  $\geq 75 \text{ m}^2 \text{ s}^{-2}$  moving into an environment with STP  $\geq 1$ , (d) 0–3-km UH  $\geq 33 \text{ m}^2 \text{ s}^{-2}$ , and (e) the 10th percentile of STP from the hour previous to 2–5-km UH  $\geq 25 \text{ m}^2 \text{ s}^{-2}$ . All (orange) and RM (black) tornado paths are overlaid.

with high ensemble STP and abundant UH, creating high probabilities for all methods (Figs. 8b–e).

This case demonstrates the value of restricting the maximum probability using observed frequencies. Initially, using midlevel rotation (Fig. 8b) or low-level rotation (Fig. 8d) alone created extremely high probabilities both within and well outside the region with numerous tornadoes. The overforecasting of the 2–5-km

UH probabilities (Fig. 8b) was not tempered much by requiring STP  $\geq 1$  (Fig. 8c), since high STP was abundant. However, the STP-based probabilities (Fig. 8e) had a maximum magnitude equivalent to the SPC's updated forecast, 30%, which is categorically equivalent to a high risk, although they had lower probabilities than the other methods within the region containing numerous tornadoes.

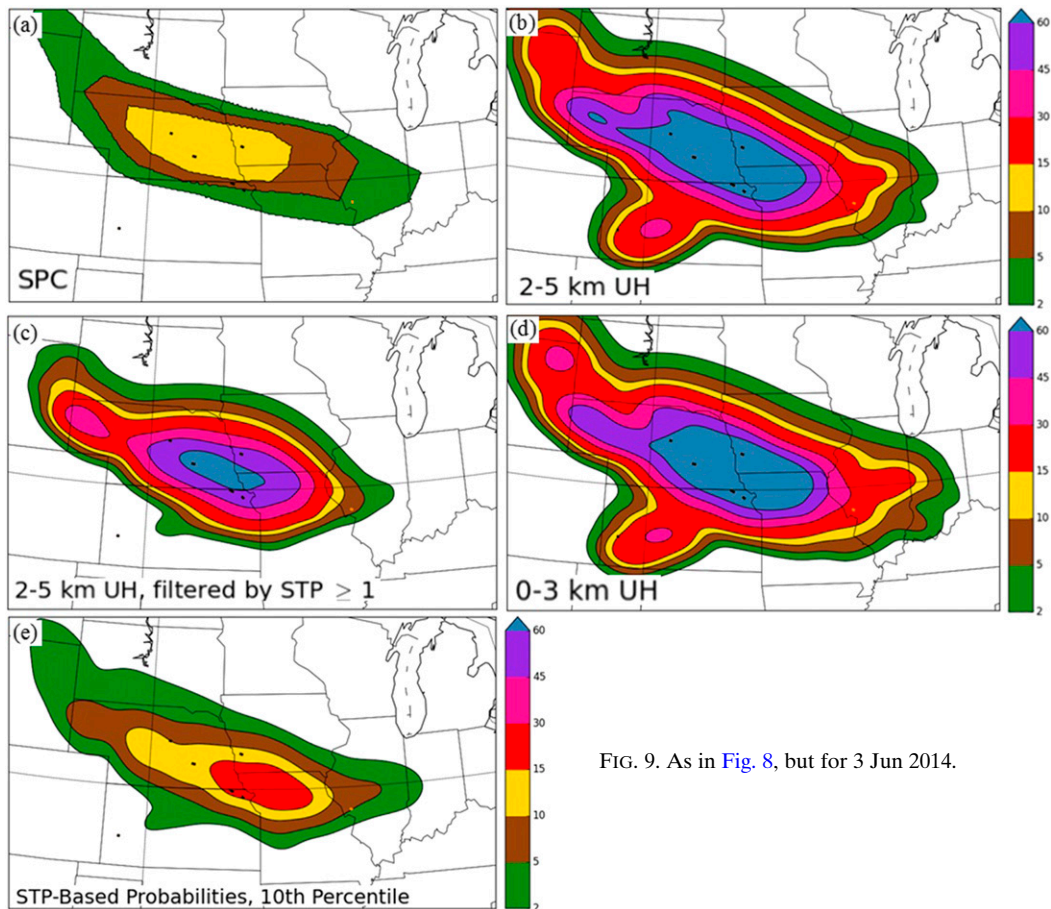


FIG. 9. As in Fig. 8, but for 3 Jun 2014.

## 2) 3 JUNE 2014

The second case contained mixed modes, where clusters of supercells produced most of the tornadoes. A vigorous short-wave trough was initially located across the north-central plains, with strong 250-hPa wind speeds (not shown). According to the 0600 UTC convective outlook, severe convection was expected to occur near a warm front. The forecast environment had ample shear and sufficient CAPE to support rotating storms. Isolated, high-based storms were anticipated initially, but much NWP guidance showed fast upscale growth into one or more mesoscale convective systems (MCSs). As a result, a 10% tornado threat was highlighted by the 0600 UTC SPC convective outlook (Fig. 9a), along with a 45% damaging wind threat (not shown). Although upscale growth occurred, many of the storms retained supercellular characteristics early in their convective life cycle. Six RM tornadoes and one nonsupercellular tornado resulted.

As in the previous case, the 2–5-km UH (Fig. 9b) and the 0–3-km UH (Fig. 9d) had vast swaths of probability

exceeding 60% (the highest possible tornado probability contour issued by the SPC), including in areas outside of the region with several tornadoes. However, the probabilities captured the tornado in western Kansas, which was missed by the 0600 UTC outlook (the 1630 UTC outlook extended the 2% probabilities into western Kansas). Forecasters might have excessive difficulty in determining the appropriate magnitude of the probabilities given this overforecasting, as was seen by Gallo et al. (2016). Incorporating environmental information by requiring an exceedance of STP reduced the probabilities somewhat (Fig. 9c), but the peak magnitude remained above 60% (which still far exceeds a typical SPC forecast and therefore does not produce useful first-guess guidance), and the Kansas tornado was now outside the 2% contour. The STP-based probabilities (Fig. 9e), however, handled the magnitude of the event the best of any of the automated probabilities, although the highest probabilities occurred east of the area with the most tornadoes. The highest probability contour was only one category higher than the official SPC forecast on this day, making it the most useful first guess of any

ensemble probabilities in creating a prediction consistent with the SPC forecast as the forecaster would not have to mentally calibrate the probabilities to typical operational values. This case also demonstrates the struggle the probabilities have with mode, in that UH swaths associated with MCSs can produce areas of false alarm, as seen across Illinois in all ensemble-generated methods.

### 3) 5 MAY 2015

The third case examined herein demonstrates how these probabilities are best used for forecasting RM tornadoes and shows the difficulties they may have on more weakly forced days. According to the SPC 0600 UTC convective outlook, a short-wave trough was forecast to evolve across the CONUS throughout the period of interest. Ongoing thunderstorms were expected to limit the instability across the central High Plains. A sharpening dryline and remnant boundaries from the morning convection were anticipated as the focus of the subsequent severe convection. Such meso-scale detail poses a forecasting challenge to humans and NWP alike, making this a difficult day to forecast. Effective bulk shear was noted by the SPC as being sufficient for supercells with a tornado threat east of the dryline, leading to an area of 5% tornado probability across the Texas Panhandle and a broader area of 2% stretching southward, where the shear was weaker (Fig. 10a). Subsequent outlooks reduced the area of 5% and eventually shifted it southward (not shown). While the UH-only methods had lower probabilities than in the prior two cases, they still showed areas of 10% (2–5-km UH only; Fig. 10b) and 15% (0–3-km UH only; Fig. 10d), which are typically used by the SPC on high-end days. These probabilities encompassed all of the tornadoes that occurred on 5 May, with the exception of the non-RM tornado in Oklahoma. Filtering the UH by requiring  $STP \geq 1$  decreased the area of false alarm in Oklahoma, but just excluded the tornadoes that occurred in central Texas and maintained the high-magnitude false alarm in southern Texas (Fig. 10c). Using the STP-based probabilities decreased the false alarm overall, and the maximum probability magnitude matched that of the SPC: 5%. Probabilities across southern Texas were especially reduced. However, the area highlighted by the 5% was in southwestern Oklahoma, which had no tornadoes, and some of the southern tornadoes were excluded.

## 4. Summary and discussion

Forecast probabilities generated using combined ensemble output and observed climatological tornado

frequencies performed comparably to the SPC 0600 UTC forecasts for all tornadoes and solely RM tornadoes. These model forecasts are designed for quick forecaster interpretation by summarizing relevant environmental and convective ensemble parameters into one graphic. Additionally, the ensemble forecasts currently become available for the 1300 UTC forecast updates, allowing forecasters to adjust the magnitude and location of the 0600 UTC tornado probabilities if they think the ensemble forecast probabilities add value. Incorporating this method into other ensembles would even allow the probabilities to be available in time for the initial day 1 forecast at 0600 UTC and is the subject of ongoing work.

These probabilities are the first to incorporate observed climatological frequencies given environmental parameters, unlike other ensemble-based tornado forecast techniques to date. The climatological frequencies calibrate the tornado probability given model-based storm environments and attributes, improving upon the idea of using thresholds of simulated environmental values, as is seen in Gallo et al. (2016). Calibrating on the STP magnitude presumes that tornado occurrence in a high-STP environment when a supercell is present is more probable, all else being equal. By calculating the probability using the value of environmental STP, the newly proposed methodology provides more information than a simple threshold exceedance paradigm. To construct the probabilities and ensure that the environmental STP remains free of storm influences, each point and time has a unique STP distribution. The probabilities are calculated by taking different percentiles of this distribution, finding the maximum resultant STP throughout the day, and assigning the probability based on the climatological frequency to that point and ensemble member. Once all ensemble members have a probability field, a Gaussian-smoothed member average yields the final values.

Of the different percentiles of STP used for probability generation, the 10th percentile had the highest reliability while maintaining high ROC areas and was compared to other probabilistic forecast generation methods. The methods tested herein produced vastly different statistics. Using solely 2–5-km UH or 0–3-km UH as proxies for tornado occurrence produced large ROC areas, as seen in previous studies (Jirak et al. 2014; Gallo et al. 2016; Sobash et al. 2016a), capturing many tornado events but overforecasting. While the exact probability calculation method using the 0–3-km UH differed from Sobash et al. (2016a), using a UH threshold that produced the most reliable forecasts also misses many tornado events as evidenced by the relatively low ROC areas in Sobash et al. (2016a). Since these probabilities are to be operational forecasting

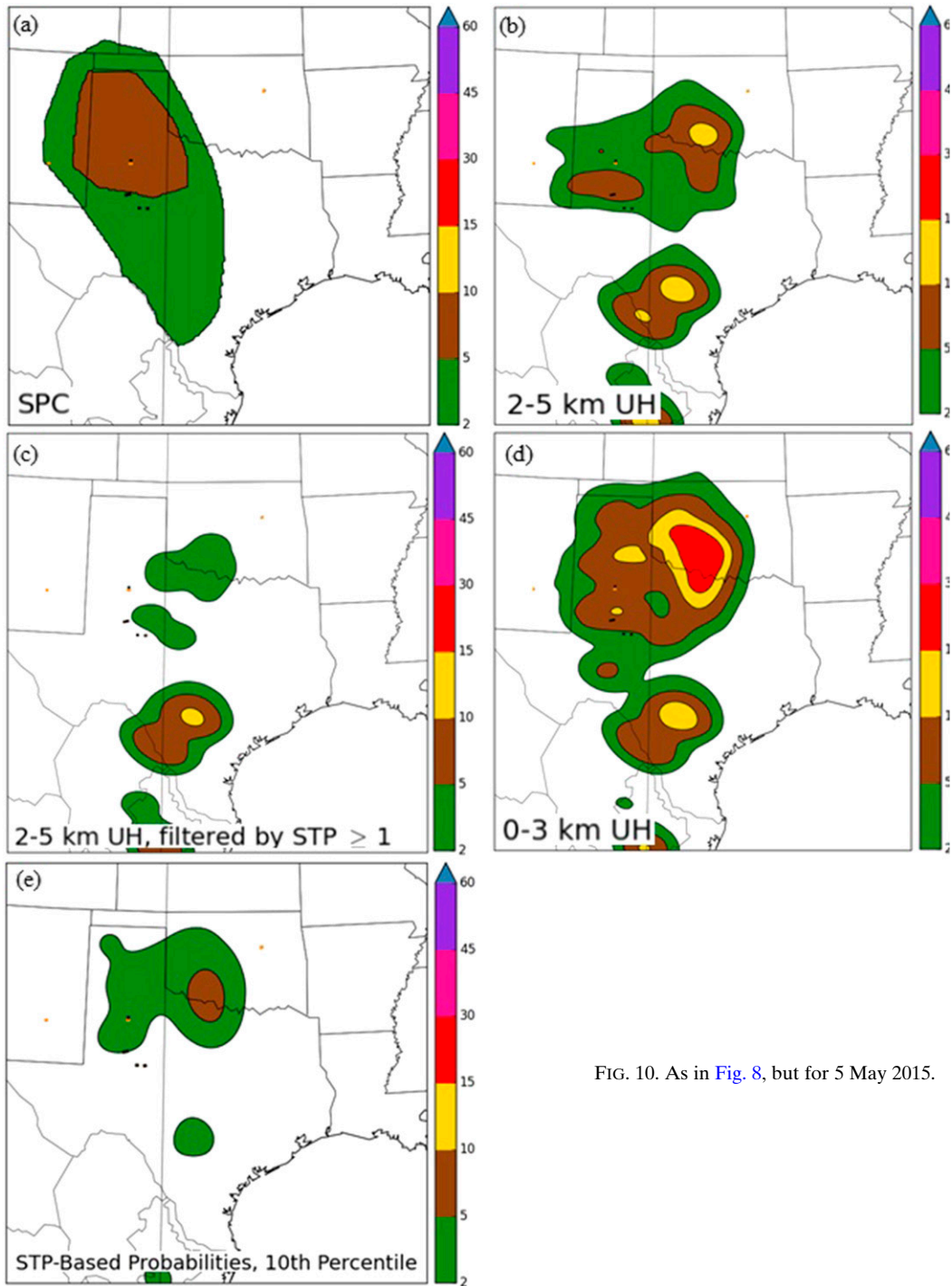


FIG. 10. As in Fig. 8, but for 5 May 2015.

tools, the 0–3-km UH threshold selected herein minimized missed events at the expense of perfect reliability.

Statistically, the STP-based probabilities resembled the 0600 UTC tornado forecasts issued by the SPC more than any other method, when verified by all tornadoes or solely by RM tornadoes. While the UH-only methods

captured more tornado events than the STP-based probabilities (i.e., higher ROC areas), both the low- and midlevel UH methods overforecasted the threat areas and magnitude. Incorporating environmental information by requiring  $STP \geq 1$  increased reliability compared to solely using UH, but excluded some

tornadoes, lowering the ROC area and still over-forecasting. The STP-based probabilities scored high ROC areas by increasing the POD with a slight increase in the POFD at the low forecast thresholds that compose most of the SPC's forecasts. They also drastically reduced overforecasting, with relatively reliable forecasts at most probabilistic thresholds, especially when considering all tornadoes. Until NWP models can directly resolve tornado-like vortices with finer grid spacing, environmental information still adds value to tornado forecasts at ~3–4-km grid spacing.

On a day-to-day basis, the STP-based probabilities often appeared comparable to the SPC forecasts, while the opposite was true for probabilities determined using a threshold of STP. The STP-based probabilities resulted in lower-probability magnitudes, as shown in the case studies, while maintaining a higher ROC area. Since these forecasts are designed to be available and can be considered a first guess for operational forecasters (with caveats of the ensemble correctly forecasting the convective mode and environment), magnitudes that are more accurate save forecasters from trying to mentally calibrate unrealistically high probabilities. For example, forecasters on 3 June 2014 could have seen the potential for supercellular tornadoes, despite the forecasted upscale growth into linear convective modes. With this guidance, it may have been easier to determine that embedded supercells were a threat within the large storm clusters, although the UH generated by linear MCSs would lend caution to the veracity of the underlying tornado probabilities. Indeed, only one non-RM tornado occurred after the line grew upscale.

The case studies also demonstrate the limitations of using environmental parameter thresholds. On 28 April 2014, STP was abundant throughout the domain of concern, so limiting the probabilities by requiring that STP exceed one still created widespread high probabilities. On 3 June 2014, high STP occurred even after the storms grew upscale, leading to high probabilities east of where most tornadoes occurred. However, using the STP-based method, the probabilities were lowered and somewhat constrained. This method also decreased the magnitudes of the probabilities in less severe cases such as on 5 May 2015 and focused the probabilities on the RM tornadoes, although weakly forced cases remain challenging.

The probabilistic paradigm discussed herein generates a probabilistic forecast from each ensemble member before averaging those forecasts. Therefore, this methodology is applicable to deterministic forecasts and ensembles of multiple sizes and implementation in such ensembles is the subject of future work. Future work will also extend these forecasts to differing modes and tornado intensities, perhaps developing similar probabilities for tornadoes with

quasi-linear convective systems or forecasting the probability of a significant tornado. Further work also remains in isolating mode: a great improvement to these probabilities would eliminate the false alarm produced by UH from MCSs, which are far less likely to produce significant tornadoes than supercellular modes. Additionally, the data examined herein covered only spring seasons; in order for these probabilities to be increasingly validated by forecasters, applicability across seasons must be tested. While a version of these probabilities is running twice daily in the HREFv2 ensemble ([available at www.spc.noaa.gov/exper/href/](http://www.spc.noaa.gov/exper/href/)) and anecdotally appear to be useful outside of the peak convective season, formal operational evaluation has yet to occur.

*Acknowledgments.* The authors thank Chris Melick and Robert Hepper of the SPC for providing regridded SPC forecasts, as well as Andrew Dean of the SPC for obtaining the environmental and radar data used in the climatological frequency calculation. Thanks also go to Ryan Lagerquist for insight to the cross-validation technique performed herein. This material is based upon work supported by a NSF Graduate Research Fellowship under Grant DGE-1102691, Project A00-4125. BTG and SRD were provided support by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA16OAR4320115, U.S. Department of Commerce. AJC also received support from a Presidential Early Career Award for Scientists and Engineers. We would also like to thank three anonymous reviewers for their comments, which improved the content and clarity of the manuscript.

## REFERENCES

- Anderson-Frey, A., Y. P. Richardson, A. R. Dean, R. L. Thompson, and B. T. Smith, 2016: Investigation of near-storm environments for tornado events and warnings. *Wea. Forecasting*, **31**, 1771–1790, <https://doi.org/10.1175/WAF-D-16-0046.1>.
- Blair, S. F., and Coauthors, 2017: High-resolution hail observations: Implications for NWS warning operations. *Wea. Forecasting*, **32**, 1101–1119, <https://doi.org/10.1175/WAF-D-16-0203.1>.
- Bothwell, P. D., J. A. Hart, and R. L. Thompson, 2002: An integrated three-dimensional objective analysis scheme. Preprints, *21st Conf. on Severe Local Storms/19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., JP3.1, <https://ams.confex.com/ams/pdfpapers/47482.pdf>.
- Boustead, J. M., B. E. Mayes, W. Gargan, J. L. Leighton, G. Phillips, and P. N. Schumacher, 2013: Discriminating environmental conditions for significant warm sector and boundary tornadoes in parts of the Great Plains. *Wea. Forecasting*, **28**, 1498–1523, <https://doi.org/10.1175/WAF-D-12-00102.1>.



- Brooks, H. E., and C. A. Doswell III, 2001: Some aspects of the international climatology of tornadoes by damage classification. *Atmos. Res.*, **56**, 191–201, [https://doi.org/10.1016/S0169-8095\(00\)00098-3](https://doi.org/10.1016/S0169-8095(00)00098-3).
- , —, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640, [https://doi.org/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. Lalauette, and A. Ghelli, 1999: Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Wea. Forecasting*, **14**, 168–189, [https://doi.org/10.1175/1520-0434\(1999\)014<0168:PPOPUT>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0168:PPOPUT>2.0.CO;2).
- Bunkers, M. J., B. A. Klimowski, J. W. Zeitler, R. L. Thompson, and M. L. Weisman, 2000: Predicting supercell motion using a new hodograph technique. *Wea. Forecasting*, **15**, 61–79, [https://doi.org/10.1175/1520-0434\(2000\)015<0061:PSMUAN>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0061:PSMUAN>2.0.CO;2).
- Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.*, **129**, 569–585, [https://doi.org/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., 2017: Generation of ensemble mean precipitation forecasts from convection-allowing ensembles. *Wea. Forecasting*, **32**, 1569–1583, <https://doi.org/10.1175/WAF-D-16-0199.1>.
- , and Coauthors, 2012a: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, <https://doi.org/10.1175/BAMS-D-11-00040.1>.
- , J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012b: Forecasting tornado pathlengths using a three-dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113, <https://doi.org/10.1175/WAF-D-11-00147.1>.
- , J. Gao, P. Marsh, T. Smith, J. Kain, J. Correia, M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387–407, <https://doi.org/10.1175/WAF-D-12-00038.1>.
- Craven, J. P., and H. E. Brooks, 2004: Baseline climatology of sounding derived parameters associated with deep, moist convection. *Natl. Wea. Dig.*, **28**, 13–24.
- Doswell, C. A., III, and D. W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.*, **116**, 495–501, [https://doi.org/10.1175/1520-0493\(1988\)116<0495:OSIOUS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<0495:OSIOUS>2.0.CO;2).
- , H. E. Brooks, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563, <https://doi.org/10.1016/j.atmosres.2008.11.003>.
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, [https://doi.org/10.1175/1520-0469\(1989\)046<3077:NSOCOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2).
- Elsner, J. B., and C. P. Schmertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619–624, [https://doi.org/10.1175/1520-0434\(1994\)009<0619:AFSTCV>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0619:AFSTCV>2.0.CO;2).
- Fowle, M., and P. J. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782–794, [https://doi.org/10.1175/1520-0434\(2003\)018<0782:SHNPOC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0782:SHNPOC>2.0.CO;2).
- Gallo, B. T., A. J. Clark, and S. R. Dembek, 2016: Forecasting tornadoes using convection-permitting ensembles. *Wea. Forecasting*, **31**, 273–295, <https://doi.org/10.1175/WAF-D-15-0134.1>.
- , and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting*, **32**, 1541–1568, <https://doi.org/10.1175/WAF-D-16-0178.1>.
- Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2012: A climatology and comparison of parameters for significant tornado events in the United States. *Wea. Forecasting*, **27**, 106–123, <https://doi.org/10.1175/WAF-D-11-00008.1>.
- Grünwald, S., and H. E. Brooks, 2011: Relationship between sounding derived parameters and the strength of tornadoes in Europe and the USA from reanalysis data. *Atmos. Res.*, **100**, 479–488, <https://doi.org/10.1016/j.atmosres.2010.11.011>.
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, [https://doi.org/10.1175/1520-0434\(1999\)014<0155:HTFENP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2).
- , and S. J. Colucci, 1998: Evaluation of Eta–RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.*, **126**, 711–724, [https://doi.org/10.1175/1520-0493\(1998\)126<0711:EOEREP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0711:EOEREP>2.0.CO;2).
- Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- Janjić, Z. I., 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, NOAA/NWS, 61 pp., <http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf>.
- Jirak, I. L., C. J. Melick, and S. J. Weiss, 2014: Combining probabilistic ensemble information from the environment with simulated storm attributes to generate calibrated probabilities of severe weather hazards. Preprints, *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 2.5, <https://ams.confex.com/ams/27SLS/webprogram/Paper254649.html>.
- 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, <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, <https://doi.org/10.1175/2010WAF2222430.1>.
- Klees, A. M., Y. P. Richardson, P. M. Markowski, C. Weiss, J. M. Wurman, and K. K. Kosiba, 2016: Comparison of the tornadic and nontornadic supercells intercepted by VORTEX2 on 10 June 2010. *Mon. Wea. Rev.*, **144**, 3201–3231, <https://doi.org/10.1175/MWR-D-15-0345.1>.
- Markowski, P. M., E. N. Rasmussen, and J. M. Straka, 1998: The occurrence of tornadoes in supercells interacting with boundaries during VORTEX-95. *Wea. Forecasting*, **13**, 852–859, [https://doi.org/10.1175/1520-0434\(1998\)013<0852:TOOTIS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0852:TOOTIS>2.0.CO;2).
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875, <https://doi.org/10.1029/RG020i004p00851>.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, <https://doi.org/10.1029/97JD00237>.

- Rasmussen, E. N., 2003: Refined supercell and tornado forecast parameters. *Wea. Forecasting*, **18**, 530–535, [https://doi.org/10.1175/1520-0434\(2003\)18<530:RSATFP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)18<530:RSATFP>2.0.CO;2).
- , and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164, [https://doi.org/10.1175/1520-0434\(1998\)013<1148:ABCOSD>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<1148:ABCOSD>2.0.CO;2).
- Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, <https://doi.org/10.1175/2008WAF2222159.1>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <http://dx.doi.org/10.5065/D68S4MVH>.
- Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. *Wea. Forecasting*, **27**, 1114–1135, <https://doi.org/10.1175/WAF-D-11-00115.1>.
- , —, A. R. Dean, and P. T. Marsh, 2015: Diagnosing the conditional probability of tornado damage rating using environmental and radar attributes. *Wea. Forecasting*, **30**, 914–932, <https://doi.org/10.1175/WAF-D-14-00122.1>.
- Sobash, R. A., and J. S. Kain, 2017: Seasonal variations in severe weather forecast skill in an experimental convection-allowing model. *Wea. Forecasting*, **32**, 1885–1902, <https://doi.org/10.1175/WAF-D-17-0043.1>.
- , —, 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, <https://doi.org/10.1175/WAF-D-10-05046.1>.
- , G. S. Romine, C. S. Schwartz, D. J. Gagne II, and M. L. Weisman, 2016a: Explicit forecasts of low-level rotation from convection-allowing models for next-day tornado prediction. *Wea. Forecasting*, **31**, 1591–1614, <https://doi.org/10.1175/WAF-D-16-0073.1>.
- , C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016b: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, <https://doi.org/10.1175/WAF-D-15-0138.1>.
- Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, [https://doi.org/10.1175/1520-0434\(2003\)018<1243:CPSWSE>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2).
- , C. M. Mead, and R. Edwards, 2007: Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. *Wea. Forecasting*, **22**, 102–115, <https://doi.org/10.1175/WAF969.1>.
- , B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **27**, 1136–1154, <https://doi.org/10.1175/WAF-D-11-00116.1>.
- , —, A. R. Dean, and P. T. Marsh, 2013: Spatial distributions of tornadic near-storm environments by convective mode. *Electron. J. Severe Storms Meteor.*, **8** (5), <http://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/125>.
- , and Coauthors, 2017: Tornado damage rating probabilities derived from WSR-88D data. *Wea. Forecasting*, **32**, 1509–1528, <https://doi.org/10.1175/WAF-D-17-0004.1>.
- 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, <https://doi.org/10.1175/WAF910.1>.
- Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev.*, **129**, 729–747, [https://doi.org/10.1175/1520-0493\(2001\)129<0729:EOASRM>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0729:EOASRM>2.0.CO;2).
- 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, [https://doi.org/10.1175/1520-0493\(1982\)110<0504:TDonSC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<0504:TDonSC>2.0.CO;2).
- , and —, 1984: The structure and classification of numerically simulated convective storms in directionally varying wind shears. *Mon. Wea. Rev.*, **112**, 2479–2498, [https://doi.org/10.1175/1520-0493\(1984\)112<2479:TSACON>2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112<2479:TSACON>2.0.CO;2).
- , and —, 1986: Characteristics of isolated convective storms. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Amer. Meteor. Soc., 331–358.
- , and R. Rotunno, 2000: The use of vertical wind shear versus helicity in interpreting supercell dynamics. *J. Atmos. Sci.*, **57**, 1452–1472, [https://doi.org/10.1175/1520-0469\(2000\)057<1452:TUOVWS>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<1452:TUOVWS>2.0.CO;2).
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.