Forecasting Tornadoes Using Convection-Permitting Ensembles

BURKELY T. GALLO

School of Meteorology, University of Oklahoma, Norman, Oklahoma

ADAM J. CLARK

NOAA/OAR/National Severe Storms Laboratory, 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 6 October 2015, in final form 18 December 2015)

ABSTRACT

Hourly maximum fields of simulated storm diagnostics from experimental versions of convection-permitting models (CPMs) provide valuable information regarding severe weather potential. While past studies have focused on predicting any type of severe weather, this study uses a CPM-based Weather Research and Forecasting (WRF) Model ensemble initialized daily at the National Severe Storms Laboratory (NSSL) to derive tornado probabilities using a combination of simulated storm diagnostics and environmental parameters. Daily probabilistic tornado forecasts are developed from the NSSL-WRF ensemble using updraft helicity (UH) as a tornado proxy. The UH fields are combined with simulated environmental fields such as lifted condensation level (LCL) height, most unstable and surface-based CAPE (MUCAPE and SBCAPE, respectively), and multifield severe weather parameters such as the significant tornado parameter (STP). Varying thresholds of 2–5-km updraft helicity were tested with differing values of σ in the Gaussian smoother that was used to derive forecast probabilities, as well as different environmental information, with the aim of maximizing both forecast skill and reliability. The addition of environmental information improved the reliability and the critical success index (CSI) while slightly degrading the area under the receiver operating characteristic (ROC) curve across all UH thresholds and σ values. The probabilities accurately reflected the location of tornado reports, and three case studies demonstrate value to forecasters. Based on initial tests, four sets of tornado probabilities were chosen for evaluation by participants in the 2015 National Oceanic and Atmospheric Administration's Hazardous Weather Testbed Spring Forecasting Experiment from 4 May to 5 June 2015. Participants found the probabilities useful and noted an overforecasting tendency.

1. Introduction

High-resolution convective-permitting models (CPMs) are increasingly part of an operational forecaster's severe weather toolbox (Fowle and Roebber 2003; Weiss et al. 2006; Coniglio et al. 2010; Sobash et al. 2011; Clark et al. 2012a; Schwartz et al. 2015). These CPMs generally have grid spacing of 4 km or less, allowing them to represent bulk properties of convective circulations, skillfully differentiate convective modes (Fowle and Roebber 2003; Done et al. 2004; Weisman et al. 2008), and provide unique guidance using hourly maximum fields of simulated storm diagnostics (Kain et al. 2010). Spring forecasting experiments (SFEs) taking place in the National Oceanic and Atmospheric Administration's (NOAA) Hazardous Weather Testbed (HWT) examine how well experimental CPMs can provide guidance to forecasters (Clark et al. 2012a). At the SFEs, researchers and forecasters discuss forecaster needs and current capabilities of CPMs, fostering greater understanding between the research and operational communities. Input from forecasters to the research community allows for subjective information about perceived guidance value, rather than relying solely on objective measures of verification.

Corresponding author address: Burkely T. Gallo, National Weather Center, NSSL/FRDD, 120 David L. Boren Blvd., Norman, OK 73072. E-mail: burkely.twiest@noaa.gov

DOI: 10.1175/WAF-D-15-0134.1

Murphy (1993) discusses three types of "goodness" that a forecast can possess: 1) the agreement between the forecast and the forecaster's conceptual model ("consistency"), 2) the correspondence between the forecast and observations ("quality"), and 3) the usefulness of the forecast to the end user ("value"). While objective measures assess the quality of the probabilities, feedback from SFE participants helps to improve the consistency of the probabilities, as well as the value of the probabilities to the forecaster. As tools for the forecaster, guidance should be consistent and valuable; working with SFE participants allows for modifying the probabilities to achieve these objectives while maintaining forecast quality.

Forecasters already use ensembles of coarser-resolution models, such as the Short-Range Ensemble Forecast (SREF; Du et al. 2014), to assess forecast uncertainty. Computing capabilities continue to improve, to the point where NOAA's Environmental Modeling Center plans to implement an operational, CPM-based ensemble in the near future (University Corporation for Atmospheric Research 2015). Compared to convection-parameterizing ensembles, CPM-based ensembles have been shown to provide better guidance in terms of precipitation forecast skill. Clark et al. (2009) found that the skill gained by upgrading ensembles to convection-permitting resolutions more than made up for the skill lost by decreasing the number of ensemble members. However, exploring the effectiveness of CPMs at forecasting severe hazards is a relatively new endeavor.

Updraft helicity (UH), a product of vertical vorticity and updraft speed, is described by

$$\mathrm{UH} = \int_{z_0}^{z_1} \zeta w \, dz, \tag{1}$$

where z_0 and z_1 are the user-defined layer of the atmosphere, w is the updraft speed, and ζ is the vertical vorticity (Kain et al. 2010). UH has been used to create probabilistic hazard guidance for any type of severe weather and skillfully distinguished severe weather events from nonsevere weather events (Sobash et al. 2011). This skill is likely due to the detection of persistent midlevel mesocyclones, a characteristic of supercells, which cause a large percentage of severe weather reports (Duda and Gallus 2010). Indeed, hourly maximum UH correlates well with observations of mesocyclones (Kain et al. 2010).

While UH is a good predictor for severe hazards, it is not necessarily a good proxy for tornadoes when used alone. Like in reality, simulated mesocyclones often form in environments unfavorable for tornadogenesis (Clark et al. 2012b). Therefore, if generating tornado probabilities from UH alone, large areas of false alarms will occur in regions with unfavorable environments. However, adding environmental criteria for probability generation could reduce the false alarm area, increasing the precision of the tornado probabilities by combining the existence of simulated mesocyclones with environmental information conducive to tornadogenesis. This study focuses on combining model-generated rotation in the form of UH with environmental parameters conducive to tornadogenesis as identified by numerous previous studies (Rasmussen and Blanchard 1998; Thompson et al. 2004; Grünwald and Brooks 2011; Grams et al. 2012) to generate probabilistic forecasts of tornadoes.

Previously, high-resolution UH has been combined with coarser-resolution environmental information to separate the tornado threat from the hail and wind threat. Jirak et al. (2014) used the Storm-Scale Ensemble of Opportunity (SSEO; Jirak et al. 2012), a CPM ensemble produced by the Storm Prediction Center (SPC), for UH fields and the 40-km SREF for environmental parameters, to extract individual hazard probabilities. The combination of large-scale environmental information with the small-scale UH diagnostic is shown to provide skillful tornado guidance, with some overprediction of the hail and wind threats (Jirak et al. 2014). This study aims to investigate the benefits of combining UH with environmental parameters taken from the same model in generating probabilistic tornado forecasts. Probabilistic forecasts reflect both uncertainty in the exact location of the storms as well as whether or not an individual storm will produce a tornado. Several objective verification metrics assess the quality of the forecast probabilities, as well as examination of subjective comments provided by participants in the 2015 SFE.

Section 2a of this paper will describe the ensemble system and the parameters used to generate the tornado probabilities. Section 2b will elaborate upon the probability generation methodology, and section 2c will explain both the objective and subjective verification methods. Section 3a evaluates the quality of the tornado probabilities through objective verification metrics. Differences in probability generation methods will be highlighted by three case studies in section 3b. Section 3c will describe the subjective evaluation that took place, including common themes noted by the SFE 2015 participants. Finally, a summary and discussion of the results along with conclusions and suggestions for further research are provided in section 4.

2. Data and methodology

a. *The NSSL-WRF ensemble configuration*

Since fall 2006, SPC forecasters have used output from an experimental, 4-km version of the Weather Research

TABLE 1. A summary of the NSSL-WRF ensemble configurations with differing lateral boundary conditions (LBCs) and initial conditions (ICs). All members use WSM6 microphysics, Dudhia shortwave radiation, RRTM longwave radiation, the Noah land surface model, and the MYJ boundary layer. 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	0000 UTC 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

and Forecasting (WRF) Model (Skamarock et al. 2008) generated by the National Severe Storms Laboratory (NSSL) using the Advanced Research version of the WRF (ARW), known hereafter as the NSSL-WRF (Kain et al. 2010). This model runs twice daily, at both 0000 and 1200 UTC. Nine additional 4-km ARW members with varying initial conditions are run at 0000 UTC, composing an ensemble of 10 members known as the NSSL-WRF ensemble. Eight of the members are initialized at 0000 UTC using 3-h SREF forecasts initialized at 2100 UTC for initial conditions and corresponding SREF member forecasts as lateral boundary conditions. The remaining member uses the 0000 UTC National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) analysis for initial conditions and the corresponding NCEP GFS forecast as lateral boundary conditions. The physics parameterizations among all members are identical, using the Mellor–Yamada–Janjić (MYJ; Mellor and Yamada 1982; Janjić 2002) planetary boundary layer scheme, WRF single-moment 6-class (WSM-6; Hong and Lim 2006) microphysics, the Noah (Chen and Dudhia 2001) land surface model, and the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) longwave radiation and Dudhia (Dudhia 1989) shortwave radiation scheme (Table 1). The NSSL-WRF ensemble began running in February 2014. Each ensemble run includes 35 vertical levels and was integrated 36 h over the continental United States (CONUS), starting at 0000 UTC. For this study, the period from 1200 to 1200 UTC the following day is considered (forecast hours 12–36).

Two spring seasons are examined herein: 1 April–30 June 2014 and 1 April–30 June 2015. Ensemble membership changed slightly during that time period, with two members initialized from Eulerian mass (EM) SREF members switched for two members initialized from Nonhydrostatic Multiscale Model on the B grid (NMMB) SREF members. This change occurred because SPC forecasters noticed that the EM SREF members were much less dispersive than the other sets of SREF model cores, resulting in a clustering of members and a subsequent decrease in the ensemble variability. Thus, it was hoped that a switch to more NMMB SREF members would increase the spread and improve the reliability. The change in ensemble membership was tested by comparing reliability diagrams for each year using the consistent members and reliability diagrams for each year using all of the members. Reliability diagrams plot the forecast probability versus the observed relative frequency, and only small differences occurred through the addition of the varying members to the constant members. This change did not have significant effects on the composition of the generated probabilities. Thus, the change in two members does not significantly affect the overall forecast probabilities, and the years are combined throughout the following verification.

b. *Probability generation*

Probabilities based on the NSSL-WRF ensemble were generated using the 2–5-km hourly maximum UH (Kain et al. 2010), defined by integrating the vertical vorticity times the updraft velocity at 2–5 km AGL (e.g., Kain et al. 2008). These hourly maximum variables contain the maximum value of UH at a given point for each hour, providing insight into trends in storm intensity and movement hour by hour. Hereafter, UH will refer to the hourly maximum quantity. Probabilities were generated following Hamill and Colucci (1998). For each case, the daily maximum value of UH is found at each grid point

for each member. Next, for each grid point a distribution of UH values is created using the value of maximum UH within a 40-km radius for each member. Probabilities are found by determining where the chosen threshold of UH (e.g., $25\,m^2\,s^{-2}$) is within this distribution. If the threshold is greater than all members forming the distribution, the Gumbel distribution (Wilks 2011) is used. The resulting probabilities are smoothed using a Gaussian kernel density weighting function, whose weights are calculated by

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

where σ is the user-defined standard deviation in units of kilometers and Δx is the grid spacing. Varying σ results in different levels of smoothness in the resultant probability fields: the higher the σ, the smoother the probability fields.

The first aim of this study is to determine the optimal σ for the Gaussian kernel and the optimal UH value. Previous studies have found that UH greater than or equal to $40\,m^2\,s^{-2}$ generated reliable probabilities of any severe report (Sobash et al. 2011). To focus on the tornado problem rather than on the any severe problem, five thresholds of hourly maximum UH were examined, beginning at $25\,m^2\,s^{-2}$ and increasing to $125\,m^2\,s^{-2}$ in $25\,m^2\,s^{-2}$ intervals. While Sobash et al. (2011) found a relatively large smoothing radius of 200 km to best discriminate severe events from nonevents, it is expected that in the current study usage of an ensemble framework allows for a smaller optimum σ because the ensemble members will account for much of the spatial uncertainty. This differs from the results of the Sobash et al. (2011) study, in which the Gaussian kernel accounted for all spatial uncertainty.

The first set of verification statistics for probabilities with varying UH and σ without environmental information provided a baseline against which probabilities incorporating environmental information were compared. While the UH is an hourly maximum variable, the environmental variables were instantaneous and assumed to be representative of the environment into which the storm was moving. To assign values of environmental parameters to values of maximum UH at each grid point for each member, the hour of the maximum UH during the period of interest was determined. Then, the environmental information for the previous hour was used for that point. If the environmental information was below certain thresholds, the UH was not included in the probability generation (i.e., UH was set to zero). The environmental variables from the previous hour of the maximum UH were used in three different combinations. One combination, designed to eliminate elevated storms [where the inflow is drawn from an above-surface unstable layer; Colman (1990)], as well as high-based storms, required the ratio of surface-based convective available potential energy (SBCAPE) to most unstable convective available potential energy (MUCAPE) to be at least 0.75, and the lifted condensation level (LCL) height to be below 1500 m AGL. These requirements helped ensure that the storm inflow originated in the near-surface layer and that cloud bases would be relatively low. The values of 0.75 and 1500 m were chosen based on Clark et al. (2012b), where these values were found to successfully identify UH in environments supportive of elevated and high-based storms. In another combination, the fixed-layer significant tornado parameter (STP; Thompson et al. 2003) was required to be greater than 1. Thompson et al. (2003) designed the STP to discriminate significant from nonsignificant or nontornadic environments (Thompson et al. 2003), utilizing SBCAPE, 0–6-km bulk shear (SHR6), 0–1-km storm relative helicity (SRH1), and the surface-based lifting condensation level (SBLCL):

$$\text{STP} = (\text{SBCAPE}/1500\,\text{J kg}^{-1})(\text{SHR6}/20\,\text{m s}^{-1})(\text{SRH1}/150\,\text{m}^2\,\text{s}^{-2})[(2000\,\text{m} - \text{SBLCL})/1000\,\text{m}]. \qquad (3)$$

Since a value of 1 or greater indicates an environment supportive of significant tornadoes, it was selected as the threshold for this study. Because this study is verifying all tornadoes, both significant and nonsignificant, requiring STP to be at or greater than 1 may seem too stringent. However, based on the results (shown later), it still slightly overpredicts tornado occurrence. The final combination of environmental parameters used both prior combinations of environmental parameters: SBCAPE to MUCAPE ratio greater than 0.75, LCL heights below 1500 m, and STP greater than 1. Each UH threshold and smoothing radius was tested for these three sets of environmental parameters.

c. Verification

Objective verification of the forecasts was conducted using reliability diagrams (Wilks 2011), receiver operating characteristic (ROC) curves, the area beneath the ROC curves, and the critical success index (CSI). The area under the ROC curve measures the ability of a

forecast to discern the outcome of a binary event, and is computed by plotting the probability of detection (POD), defined as

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

against the probability of false detection (POFD), defined as

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

at specified levels of probability: 0.5%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, and 95%. Computing these statistics for smaller increments at lower probabilities than at higher probabilities follows SPC tornado probability forecasts and accounts for large differences in area between low-probability thresholds. The area under this curve is computed using the trapezoidal method (Wandishin et al. 2001) and ranges from 0 to 1. A value of 1 is a perfect forecast, a value above 0.5 is considered to have positive skill, and a ROC area of 0.7 is considered to be the lower limit of a useful forecast (Buizza et al. 1999). To test the statistical significance of the difference between ROC areas from two forecasts, resampling was done following Hamill (1999). Cases were randomly assigned to one of the two forecast methods (i.e., UH only versus UH and STP) 1000 times to create a distribution of ROC area differences. If the ROC area differences calculated using the two forecasting techniques lies outside the 95% confidence interval, they were deemed significant.

While ROC curves determine the discriminating ability between events and nonevents, the shape of the ROC curves is unaffected by probability magnitude and therefore is not impacted by biased probability forecasts. To visualize the bias in the forecasts, reliability diagrams were generated by plotting the forecast probability against the observed relative frequency. A diagonal line represents a forecast probability equal to the observed relative frequency (i.e., perfect reliability). Values above (below) the diagonal represent underforecasting (overforecasting), where the observed relative frequency is higher (lower) than the forecast probability.

The final metric considered, CSI, is the number of correct ''yes'' forecasts divided by the total number of hits, misses, and false alarms:

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

It is a score often used in rare events (Wilks 2011) and is, therefore, an appropriate score to consider in tornado forecasting. Scores range from 0 to 1, with 1 being a perfect score. Visualization of CSI is viewed through performance diagrams (Roebber 2009). Performance diagrams plot the POD versus the success ratio, which is defined as

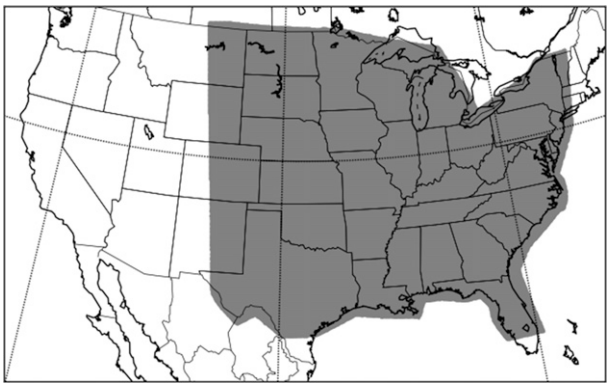

FIG. 1. The model domain for the NSSL–WRF ensemble. The shaded region shows where objective verification measures were computed.

$$\text{success ratio} = 1 - \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}. \tag{7}$$

Lines of constant reliability are plotted as dashed lines, and lines of constant CSI are plotted as solid, curved lines.

These measures were applied across the eastern two-thirds of the CONUS (Fig. 1). Verification was based on the Local Storm Reports (LSR) database for each day, as generated by the SPC. Reports filtered by the SPC were used to attempt to remove duplicate reports of the same tornado. While the tornado report database is flawed (Verbout et al. 2006; Doswell et al. 2009), underreporting has been reduced in recent decades (Brooks and Doswell 2002) and utilizing the location of reported tornadoes for verification emphasizes the utility of CPM ensembles in highlighting spatial areas of concern. Only the starting points of tornado paths are used to assign locations of the reports, and tornado pathlength is not considered. Verification was performed on the 4-km grid of the NSSL-WRF and observed reports were mapped onto the 4-km grid and treated as yes/no binary events, where a yes occurred if a tornado report was within a 40-km radius.

Subjective verification of the forecasts took place during the Experimental Forecast Program of the Spring Forecasting Experiment at the Hazardous Weather Testbed from 4 May to 5 June 2015. During this experiment, participants were presented with forecast probabilities and LSR tornadoes from the forecast period overlaid. The forecasters were then asked to assign

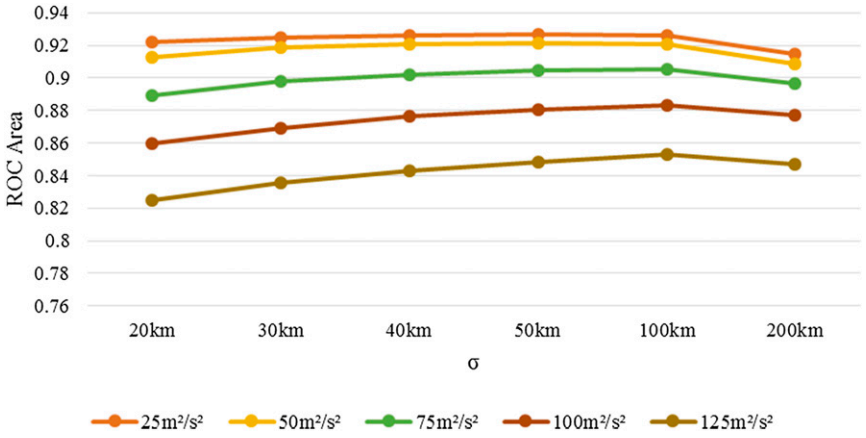

FIG. 2. ROC areas for tornado probabilities formed using differing σ values and UH thresholds. Different UH thresholds are shown in different colors. All ROC areas are for probabilities formed without incorporation of environmental information.

ratings to the forecasts on a scale from 1 (very poor) to 10 (very good), and to provide specific comments about the forecasts and the methods of incorporating environmental parameters into the probabilities. They could also explain why they assigned the ratings they chose for each forecast.

3. Results

a. *Objective verification*

Objective verification of the probabilities utilized the ROC curve, the area under the ROC curve, and reliability diagrams. ROC areas were first computed for the probabilities that solely incorporated UH (Fig. 2). The impact of changing the UH threshold and the σ value of the Gaussian kernel were tested. As the threshold of UH was increased, the ROC areas decreased at all σ levels, likely as a result of the POD decreasing more quickly than the POFD as the UH threshold increases. However, the ROC areas remained above 0.7 for all thresholds and σ values. Decreases in ROC area were greater above the UH threshold of $50\,m^2\,s^{-2}$, decreasing by ~0.01 or less from 25 to $50\,m^2\,s^{-2}$ and from 0.01 to 0.03 for each $25\,m^2\,s^{-2}$ of UH added to the threshold past $50\,m^2\,s^{-2}$. Differences in ROC areas between thresholds separated by $25\,m^2\,s^{-2}$ were not statistically significant, but differences in ROC areas between thresholds separated by $50\,m^2\,s^{-2}$ were significant. Increases in the σ value had smaller effects on the ROC area than increases in the UH thresholds. In general, as the σ value increased, (more smoothing; example given in Fig. 3), the ROC area increased slightly. However, at low UH thresholds, increasing σ beyond 50 km decreased the ROC area. The same effect at high UH thresholds was seen at σ beyond 100 km, suggesting a less skillful forecast. ROC area changes caused by σ variation were one to two orders of magnitude smaller than the changes caused by adjusting the UH threshold. In fact, differences in the ROC area between σ of 20 km and σ of 200 km show no statistically significant difference at any UH threshold. Thus, the variation in the UH threshold has a larger influence on the ROC area than does the smoothing level. This ROC area behavior is similar to the results of Sobash et al. (2011), who also found that the ROC area decreased with increased UH threshold and generally increased with increasing σ.

The same pattern occurs when the probabilities incorporate environmental information. ROC areas for varying levels of σ and UH (Fig. 4) show that environmental filtering decreases the ROC area in most instances, but the ROC area remains above 0.8 for all cases except for UH $\geq 125\,m^2\,s^{-2}$, the highest UH threshold tested. The ROC area decrease depends on the filtering method, UH threshold, and σ value. The LCL/CAPE ratio method shows the smallest difference from the UH-only probabilities, with an average difference across all σ values and all UH thresholds of −0.005. Indeed, in two cases (UH $\geq 25\,m^2\,s^{-2}$, $\sigma = 50$ km and UH $\geq 50\,m^2\,s^{-2}$, $\sigma = 100$ km) the environmental information increases the ROC area compared to UH only. However, neither of these differences were statistically significant, nor were other differences between the LCL/CAPE ratio method and the UH-only method across the σ and UH producing the largest average differences. The differences become larger and statistically significant for the STP method when compared to the UH-only method, with an average difference across all σ values and UH thresholds of −0.035.

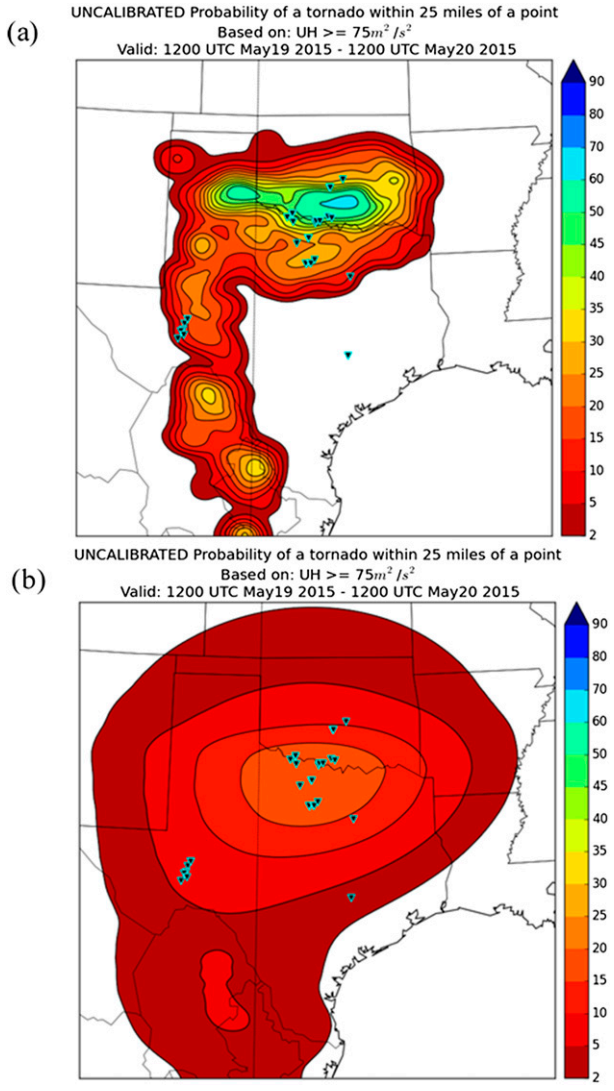

FIG. 3. Tornado probability maps valid from 1200 UTC 19 May to 1200 UTC 20 May 2015 for a UH threshold of 75 $m^2 s^{-2}$ and a Gaussian kernel of σ = (a) 20 and (b) 200 km. Probabilities are shaded contours, and tornado reports are overlaid black inverted triangles with cyan borders.

The difference from the UH-only method widens for the LCL/CAPE ratio/STP method at −0.039, while the difference between the LCL/CAPE ratio/STP method and the STP method is quite small, reflecting the large dependence on STP. The differences are larger across all methods for larger UH thresholds, often at the 0.01 order of magnitude. Therefore, the environmental information incorporated generally has as much of an impact on the ROC area metric as the selection of the UH threshold, and more of an impact than the selection of σ.

Figure 5 visualizes example ROC curves for five varying UH thresholds using four methods of probability generation. The σ of the Gaussian kernel is fixed at 50 km for Figs. 5a–e, as 50–100 km is the range above which most ROC areas began to decrease for a given threshold. Generally, POD and POFD decrease as the UH threshold increases, because more events are being missed at higher UH thresholds. While it may seem counterintuitive that environmental information causes lower ROC areas, the curves show that most of the information loss occurs at low probabilities (i.e., less than 0.5%). Since events are rare, missing one event causes a large decrease in POD at very low thresholds. At operational probability thresholds (i.e., 2%+), the environmental information causes slight improvement in the POFD, which is then offset by the decrease in POD at low-probability levels.

While the ROC areas are highest for low thresholds of UH, they are heavily influenced by correct negatives, which compose a large portion of the data for tornadoes on a high-resolution grid. Thus, CSI was examined to provide a metric that excludes correct negatives. Performance diagrams (Fig. 6) show that the addition of environmental information increases the CSI at ranges used by the SPC operationally: 2%, 5%, 10%, 15%, 30%, and 45%. CSI also improves as UH thresholds increase. While all values considered are far from a perfect forecast of 1, they are similar to the results of Sobash et al. (2011), and roughly what is expected from high-resolution verification of very rare events such as tornadoes. Finally, CSI shows improvement with additional environmental information, with the LCL/CAPE ratio/STP method often having the highest CSI at a given probability level.

The effect of changing σ is pronounced when considering the reliability diagrams for the UH-only probabilities (Fig. 7). For small values of σ, all forecast probabilities are much larger than the observed relative frequency, indicating overforecasting. This overforecasting persists as σ is increased, but the degree of overforecasting lessens with increased σ. For larger σ (Figs. 7e,f) the overforecasting is minimized for most levels, but sample size begins to limit the number of higher forecast probabilities, starting around a UH threshold of 75 $m^2 s^{-2}$. Since the ROC areas of each σ level were statistically indistinguishable and a limited sample size occurred at high-probability thresholds, the UH and σ combination used in SFE2015 was selected as the more computationally efficient σ of 50 km and a UH threshold of 75 $m^2 s^{-2}$ to maintain reliable high probabilities. Though these high probabilities are larger than what is currently operationally forecasted, the reliability of these probabilities combined with the relatively high ROC areas suggests skillful forecasts.

This chosen threshold (σ of 50 km and UH of 75 $m^2 s^{-2}$) is compared among all methods of probability

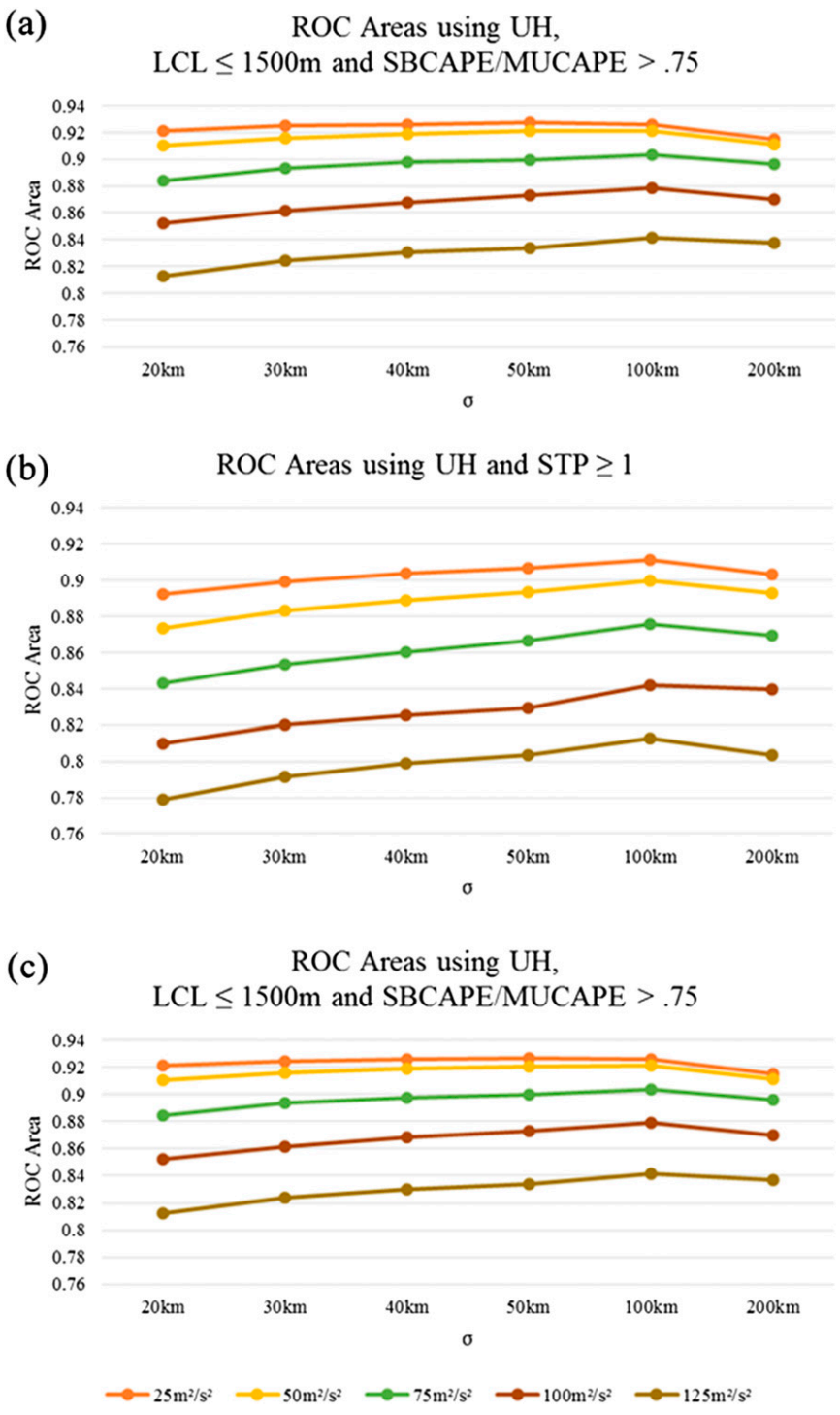

FIG. 4. ROC areas for tornado probabilities formed using differing σ values and UH thresholds. Different colors represent different UH thresholds. ROC areas are from probabilities incorporating (a) LCL ≤ 1500 m and SBCAPE/MUCAPE > 0.75, (b) STP ≥ 1, and (c) LCL ≤ 1500 m, SBCAPE/MUCAPE > 0.75, and STP ≥ 1.

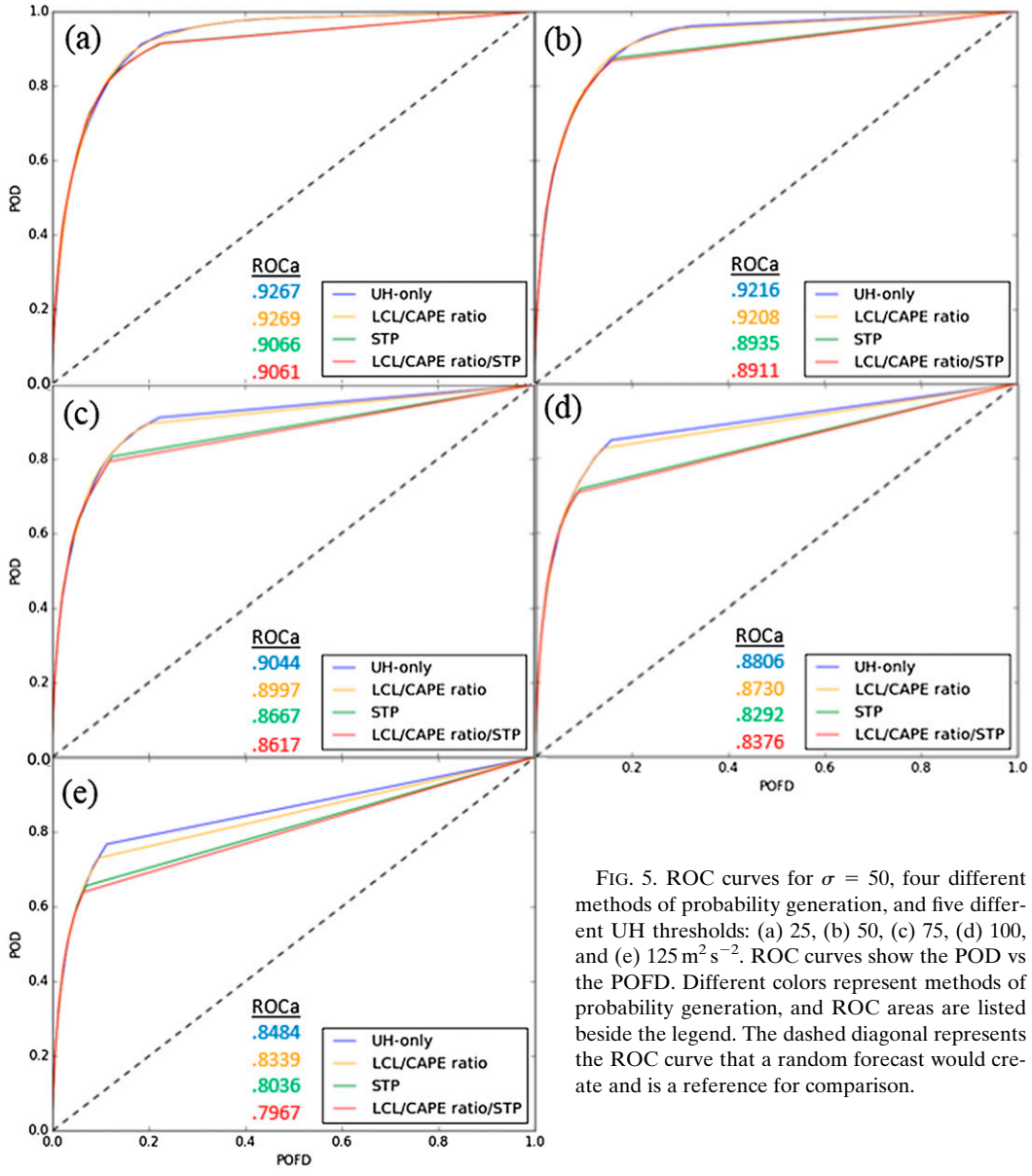

FIG. 5. ROC curves for $\sigma = 50$, four different methods of probability generation, and five different UH thresholds: (a) 25, (b) 50, (c) 75, (d) 100, and (e) 125 $m^2 s^{-2}$. ROC curves show the POD vs the POFD. Different colors represent methods of probability generation, and ROC areas are listed beside the legend. The dashed diagonal represents the ROC curve that a random forecast would create and is a reference for comparison.

generation (Fig. 8). Incorporation of the environmental information greatly increases the reliability, particularly at higher probability values. As it is harder for the probabilities to meet all environmental criteria (recall that the STP consists of four separate parameters), fewer ensemble members will meet the criteria in a given neighborhood. This dampens the magnitude of the probabilities and leads to a reduction in overforecasting. The environmental criteria also reduce the spatial area encompassed by the probabilities. The reduction in spatial area will be more fully illustrated in the case study examples given in section 2b. As can be seen in Fig. 8b, incorporating LCL height and the CAPE ratio increases the reliability at high forecast probabilities. However, only a slight increase occurs at the lower magnitudes. When the STP is considered, as in Figs. 8c and 8d, the reliability increases for all magnitudes of probability, and the overforecasting is more uniform than in both Figs. 8a and 8b.

When these results are compared to probabilities generated without UH, instead requiring that STP ≥ 1, vast overforecasting occurs at all levels, and large swaths of very high probabilities occur (Fig. 9). These results emphasize the need for multiple methods of evaluating the probabilities, as the ROC area from both spring seasons is 0.90, similar to that found with solely using UH. However, the large swaths of high probability seen on individual days (Fig. 9a) demonstrate how difficult it would be to use these probabilities as a first-guess

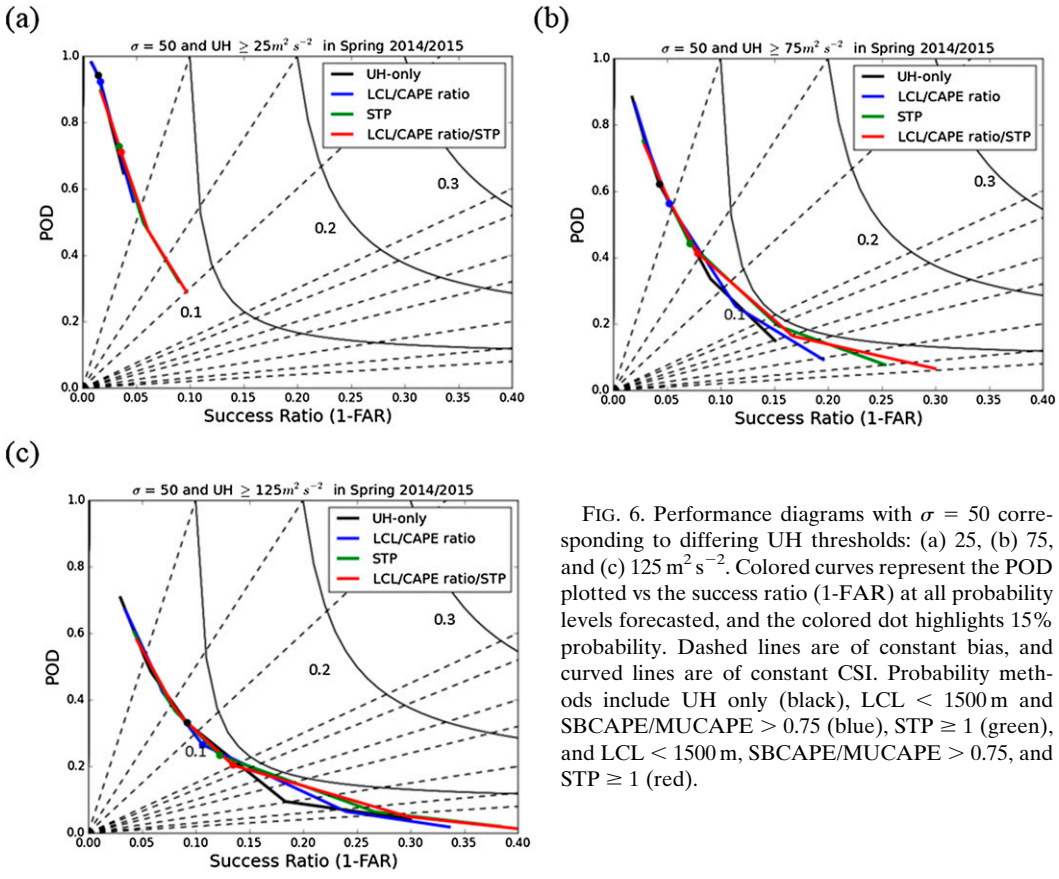

FIG. 6. Performance diagrams with $\sigma = 50$ corresponding to differing UH thresholds: (a) 25, (b) 75, and (c) 125 $m^2 s^{-2}$. Colored curves represent the POD plotted vs the success ratio (1-FAR) at all probability levels forecasted, and the colored dot highlights 15% probability. Dashed lines are of constant bias, and curved lines are of constant CSI. Probability methods include UH only (black), LCL < 1500 m and SBCAPE/MUCAPE > 0.75 (blue), STP ≥ 1 (green), and LCL < 1500 m, SBCAPE/MUCAPE > 0.75, and STP ≥ 1 (red).

forecast, as extremely high probabilities encompass much of Texas and Oklahoma. There is also a very sharp gradient in probabilities, reflecting the larger overforecasting problem identified in the reliability diagram (Fig. 9b).

Forecasters develop intuition about various models and products; these statistics may help calibrate forecasters' understanding of the probabilities. The high probabilities in all cases involving environmental parameters demonstrate high observed relative frequency, occasionally even underforecasting high-probability events. While the sample size at high probabilities is fairly small, when high probabilities occur, a tornado is relatively likely, and forecasters can proceed with heightened awareness. The high values of ROC areas found across all probabilities also indicate that these forecasts can successfully distinguish areas of tornado occurrence from areas without tornado occurrence.

b. *Example cases*

Three example cases are discussed in this section. The first case was a typical synoptic setup for spring in the southern plains, with ample CAPE, strong shear, and relatively little convective inhibition, spawning multiple tornadoes across southern Oklahoma and northern Texas. These tornadoes were well depicted by the probabilities. The second case is a late spring case, taking place in the northern plains with a secondary area of focus across the mid-Atlantic. This case had more tornadoes than the first case and demonstrates the performance of the probabilities in less climatologically favored regions for tornadoes. The final case demonstrates a day where the probabilities had difficulty pinpointing the area of highest tornado risk, instead portraying a broad area of false alarm, with the tornado reports occurring away from the highest magnitude of probabilities. While it is unwise to judge the quality of probabilistic forecasts based on individual days, these probabilities are meant to be tools for forecasters. As such, the potentially operational end products are presented here. These case studies further emphasize the operational potential of these forecasts.

1) 19 MAY 2015

At 1200 UTC 19 May 2015, a 500-hPa short-wave trough progressed across the Great Basin area, with a 500-hPa speed maximum of 55–60 kt (where 1 kt = 0.51 m s^{-1}) located over Arizona and New Mexico (Fig. 10). At 850 hPa (not shown), moist air was

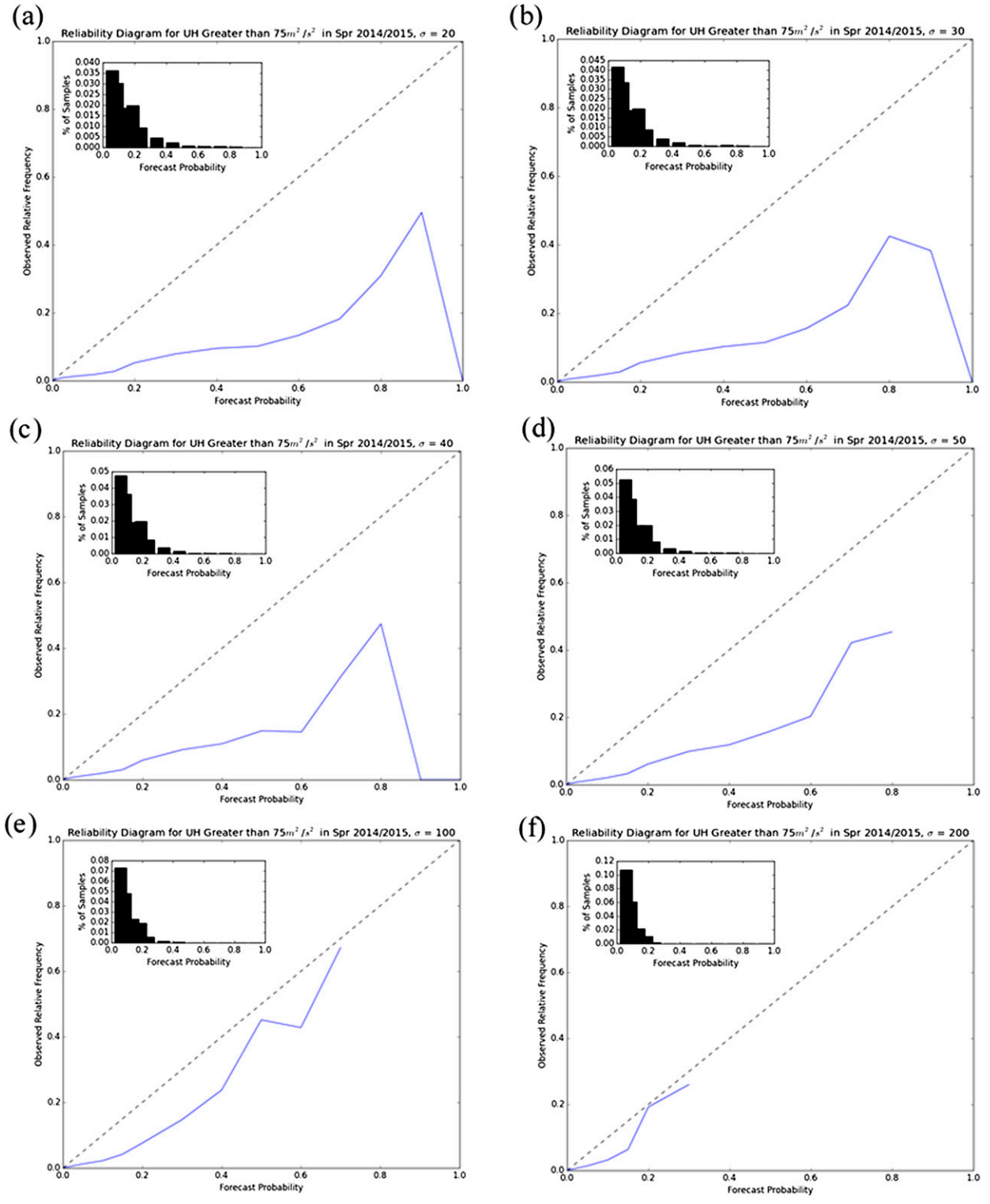

FIG. 7. Reliability diagrams for tornado probabilities solely incorporating UH $> 75\ \mathrm{m^2\ s^{-2}}$ and values of σ = (a) 20, (b) 30, (c) 40, (d) 50, (e) 100, and (f) 200 km. The dashed black line indicates perfect reliability, the area above the line indicates underforecasting, and the area below the line indicates overforecasting. Histograms in the corner show the percentage of samples in each forecast probability bin, with the 0% bin excluded for clarity because of its overwhelming majority of samples.

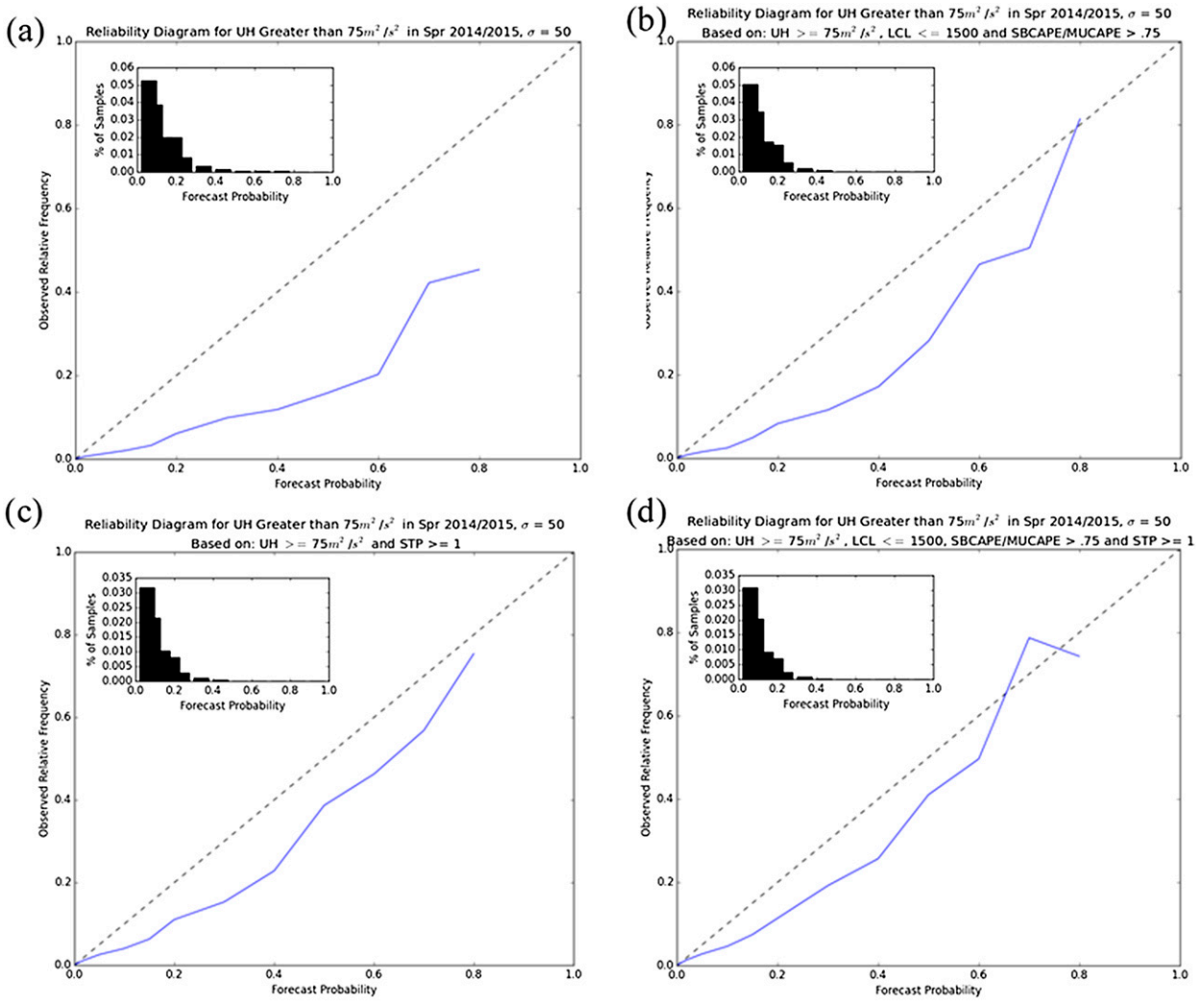

FIG. 8. Reliability diagrams for tornado probabilities with a UH threshold of 75 $m^2 s^{-2}$ and values of σ = 50 km for (a) no additional environmental information, (b) LCL < 1500 m and SBCAPE/MUCAPE > 0.75, (c) STP ≥ 1, and (d) LCL < 1500 m, SBCAPE/MUCAPE > 0.75, and STP ≥ 1. The dashed black line indicates perfect reliability.

advected northwestward from the Gulf of Mexico, and dewpoints across Oklahoma and northern Texas reached 10°–14°C. While this setup is often associated with outbreaks of severe weather across the southern plains (Mercer et al. 2012; Corfidi et al. 2010), this case was complicated by the presence of ongoing convection across the Texas Panhandle. On this day, a slight risk was issued by the SPC despite the high values of shear and potential for large CAPE, largely because of the morning convection and subsequent cloud cover, and a lack of an elevated mixed layer as discussed in the 1630 UTC day 1 convective outlook. This case took place during SFE 2015, and both experiment leaders and participants agreed that the convective mode, evolution, and timing were particularly difficult to forecast because of the ongoing storms and mixed numerical guidance regarding the convective mode. Many models showed multiple mesoscale convective systems (MCSs) moving across the region of interest during the day, but some suggested that supercellular storms would form in the warm sector ahead of the ongoing convection and south of an east–west-oriented surface stationary front.

This front progressed slowly northward throughout the day, and tornadic supercells formed after the passage of the weak MCS generated by the morning convection. These supercells grew upscale into a second MCS that stretched across Oklahoma and into northern Texas. Behind these supercells, surface heating was able to initiate a third MCS over the Texas Panhandle late in the day, which eventually caught up to and merged with the second MCS , forming an east–west-oriented MCS located along the stationary front. A few supercells also

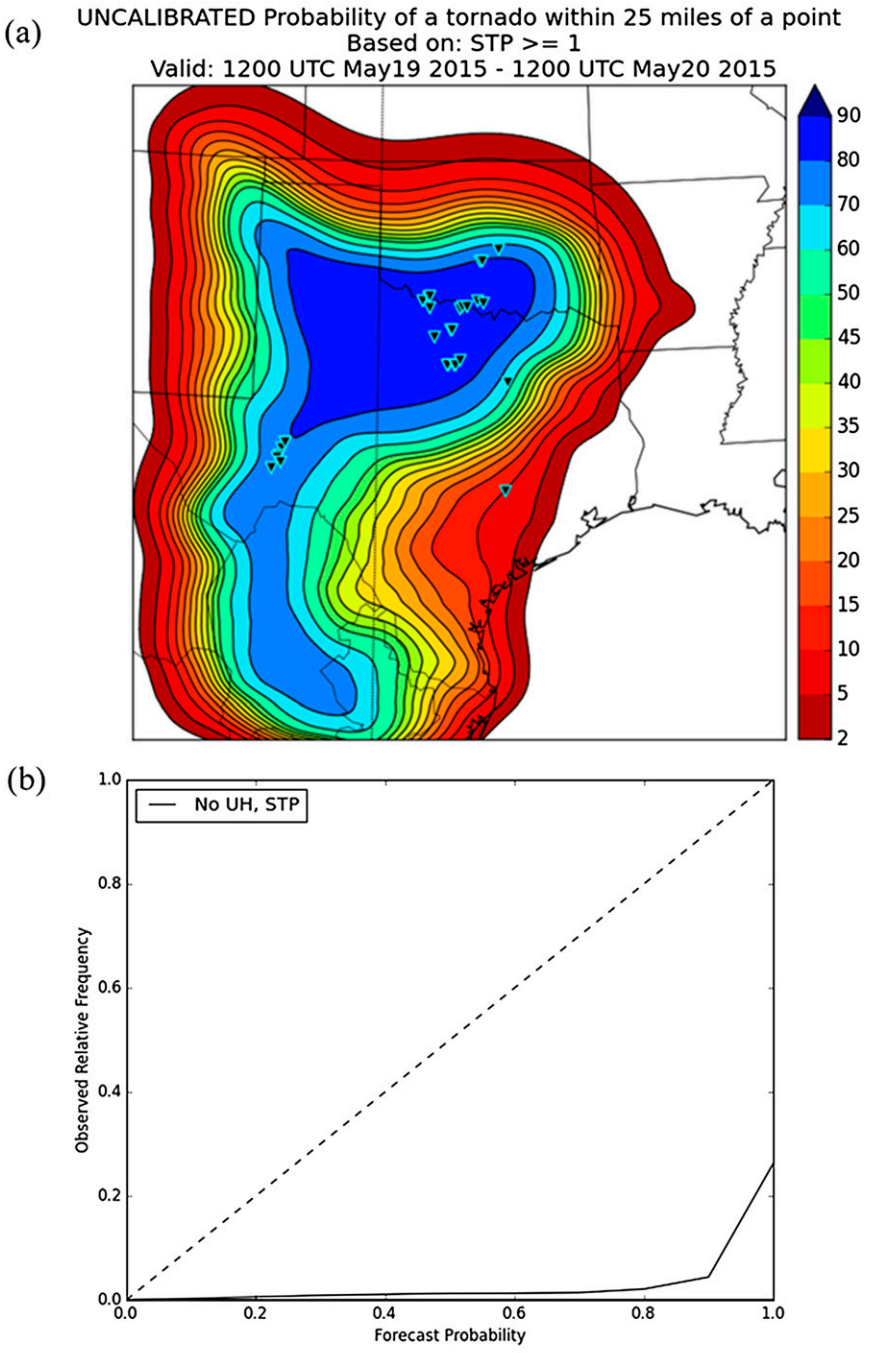

FIG. 9. (a) Tornado probability map valid from 1200 UTC 19 May to 1200 UTC 20 May 2015 generated solely using STP ≥ 1 and $\sigma = 50$ km, with tornado reports as overlaid black inverted triangles with cyan borders and (b) the reliability diagram for the springs of 2014 and 2015 for probabilities using solely STP ≥ 1. The dashed black line indicates perfect reliability.

initiated off of the Davis Mountains in southern Texas, far from the morning convection. At the end of the day, 29 tornadoes were reported across Oklahoma and Texas.

Clearly, this was a difficult day to forecast specific hazards. The mixed-mode signal suggested that wind, hail, and tornado threats were possible. The tornado probabilities provided an excellent first guess for the locations of the tornado reports (Fig. 11). The UH-only probabilities (Fig. 11a) broadly highlight northern Texas and southern Oklahoma, as well as a secondary area of

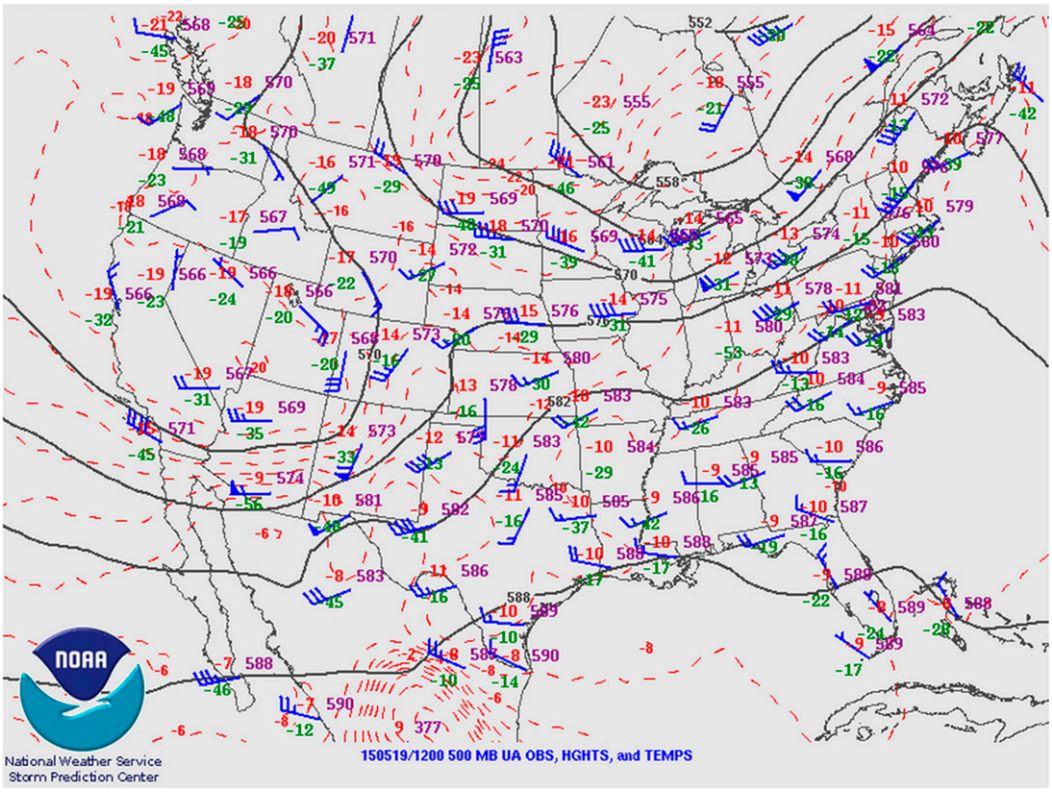

FIG. 10. A 500-hPa map valid at 1200 UTC 19 May 2015. Solid black lines are isobars, dashed red lines are isotherms, and blue barbs are 500-hPa wind speed and direction. Pressures (purple), temperatures (red), and dewpoints (green) at observation points are also shown. [Obtained from the SPC website (www.spc.noaa.gov/exper/archive/event.php?date=20150519).]

concern associated with the Davis Mountains. The highest probabilities were centered along the Red River, which forms the border between southern Oklahoma and northern Texas, and the highest magnitudes were ~45%. This bull's-eye was where the highest concentration of tornado reports occurred. The LCL and SBCAPE/MUCAPE ratio method (Fig. 11b) maintained the high magnitude of probabilities around the Red River, but correctly diminished the high probabilities of the UH-only method across the Texas Panhandle. The low-end probabilities generally encompassed the same area as the UH-only probabilities, but the magnitudes decreased (Fig. 11b). The STP method (Fig. 11c) greatly reduces the magnitude of probabilities far from the bull's-eye while maintaining high probabilities in the bull's-eye, although the magnitude of the probability reduction is less than with the SBCAPE/MUCAPE ratio method. The area of false alarm initially present across the Texas Panhandle (Figs. 11a,b) is also greatly reduced by the STP method. Finally, the method with LCL height, CAPE ratio, and STP decreases the probabilities the most and provides the greatest correspondence of the probabilities with the locations of the tornado reports (Fig. 11d). The secondary bull's-eye of higher probability across the Texas Panhandle is greatly diminished, while the area of higher probability remains present across the Davis Mountains.

The high magnitude of the probabilities along the Red River is maintained in all methods of tornado forecast generation, suggesting that incorporating environmental information maintains the high risk of tornadoes across this area. This contrasts with the area of relatively high probabilities across the Texas Panhandle, which was greatly reduced by using the environmental information. While the broad area encompassed by the probabilities remained consistent, the highest risk was shifted toward the observations through the addition of the environmental information, and highlighted the area of highest tornado risk despite mixed signals regarding the convective mode, evolution, and timing of the day's storms.

2) 27 JUNE 2015

At 1200 UTC 27 June 2015, a 500-hPa trough spread across the Mississippi valley (Fig. 12). The 250-hPa wind speeds (not shown) were high considering the location

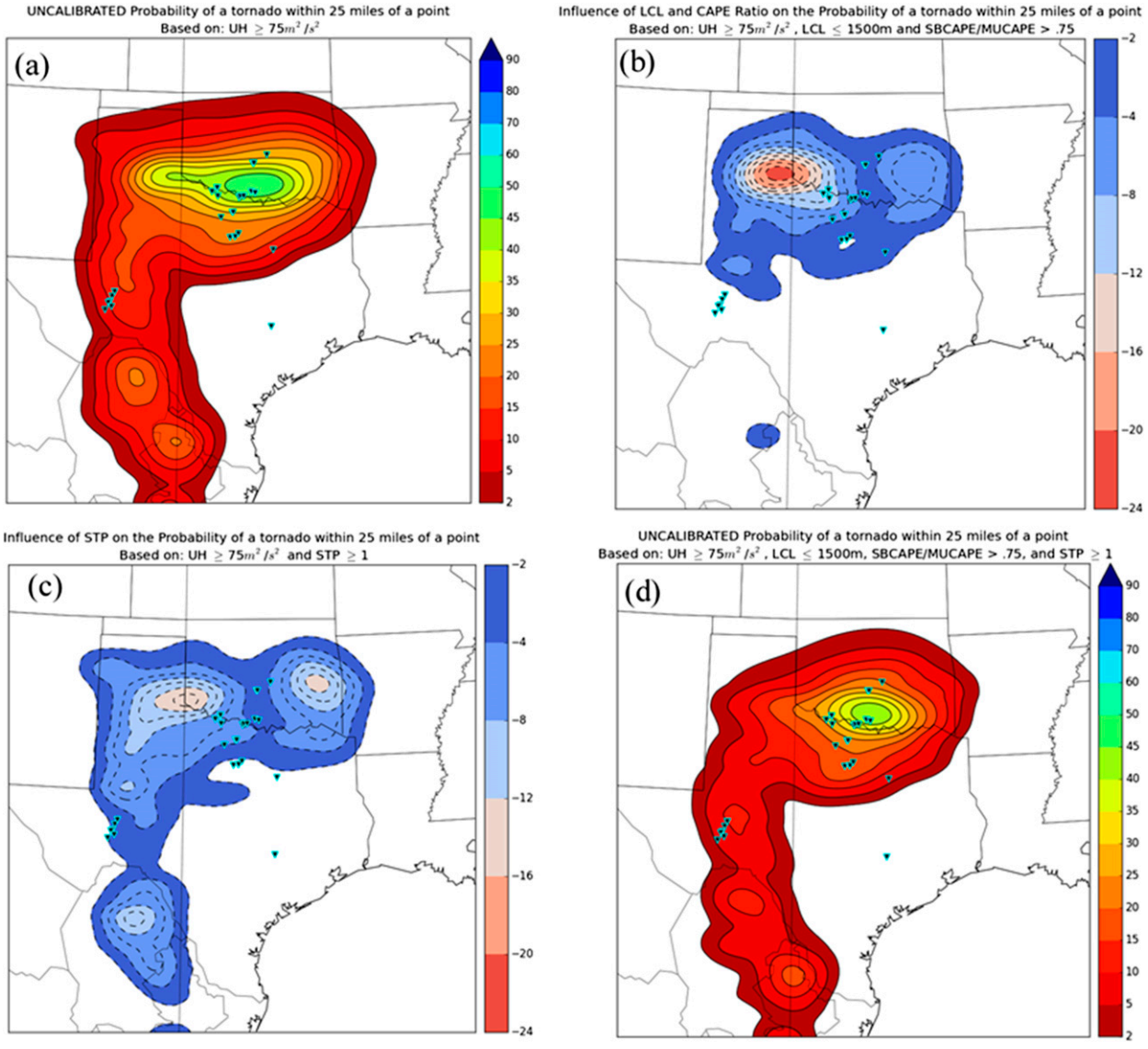

FIG. 11. (a) Tornado probability map valid from 1200 UTC 19 May to 1200 UTC 20 May 2015 for a UH threshold of 75 $m^2 s^{-2}$ and σ = 50 km generated using solely UH and (d) including environmental information. Probabilities are shaded contours, and tornado reports are overlaid black inverted triangles with cyan borders. Difference maps between probabilities generated solely using UH and (b) requiring LCL < 1500 m and SBCAPE/MUCAPE > 0.75 and (c) requiring STP ≥ 1. Dashed contours are drawn every 2%, starting at 0%. Negative numbers indicate a reduction in probability compared to (a).

and time of year, with wind speeds of over 100 kt ahead of the main trough axis. Two separate areas of tornadic storms formed: one across the eastern Dakotas and one across the mid-Atlantic and the Carolinas. The SPC issued an enhanced risk across both areas, and most tornado reports were encompassed within either the enhanced or the slight risk area. The SPC noted in their 1630 UTC convective outlook that the warm front in the east provided backed wind profiles capable of supporting rotating storms, as well as steep lapse rates and strong upper-level winds associated with the short-wave trough evolving from Canada into the Dakotas. However, the weak anticipated low-level wind shear caused uncertainty with regard to the tornado risk. By the end of the event, 35 tornadoes were reported, with a majority of the tornadoes occurring across North Dakota and into Minnesota.

On this day, the probabilities highlighted the northern system (Figs. 13a–d). The probabilities emphasized tornadic risk across the Dakotas, while maintaining low risk across the mid-Atlantic. The orientation of the probabilities in both cases also closely matched the orientation of the reports, suggesting that the synoptic

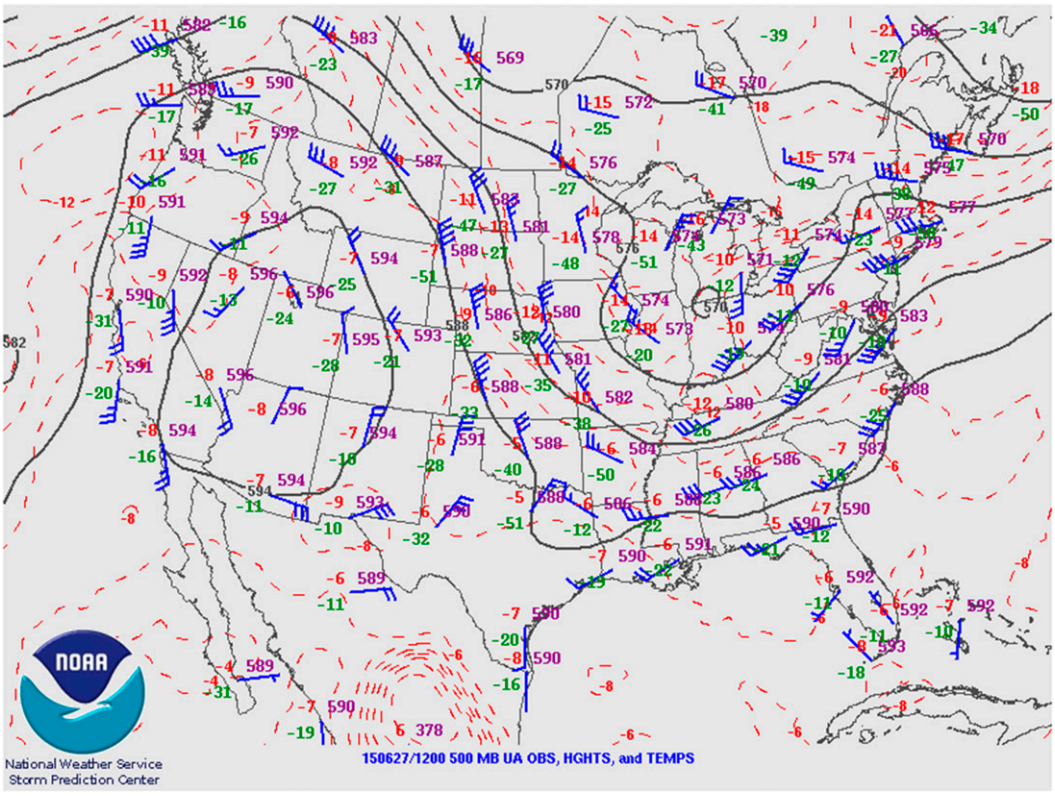

FIG. 12. As in Fig. 10, but for 1200 UTC 27 Jun 2015. [Obtained from the SPC website (www.spc.noaa.gov/exper/archive/event.php?date=20150627).]

setup was accurately portrayed. Comparing Figs. 13a and 13d demonstrates the reduction in both magnitude and areal coverage of probabilities provided by adding environmental information. The difference plots shown in Figs. 13b and 13c show that the STP method in this case caused a much larger reduction than did the LCL/CAPE ratio method. The STP method also eliminates the area of false alarm in Alabama. While all of the northern tornado reports remain within the envelope of probabilities with all environmental criteria (Fig. 13d), the focus of the tornado probabilities in the mid-Atlantic is much more northerly than in the reports. While the mid-Atlantic probabilities encompassed two of the tornado reports, on this day many of the North Carolina tornadoes were missed.

This case is discussed to demonstrate that the probabilities are useful across the United States; wherever the environmental conditions are favorable for tornadogenesis and UH is present within the ensemble, probabilities will be issued.

3) 28 MAY 2015

On 28 May 2015, a short-wave trough was located across the Rocky Mountains, with several smaller short-wave impulses along the larger trough axis. One such short-wave impulse was ejected from northern Oklahoma, with another impulse set to be ejected northeastward over Texas throughout the day (Fig. 14). Upper-level wind speeds at the trough's base were approximately 55–65 kt at 250 hPa (not shown) and low-level moisture was abundant. Prior convection left remnant outflow boundaries across Kansas, Oklahoma, and Texas, and the Storm Prediction Center's 1300 and 1630 UTC convective outlooks noted their potential as foci for convective initiation later in the day. Despite morning convection, storms initiated along the outflow boundaries and produced one tornado before they quickly grew upscale into a large MCS that spanned Texas, while farther northward isolated supercells across western Kansas also grew upscale into clusters. The supercells in Kansas produced a string of tornado reports, as did supercells near the Oklahoma–Colorado border.

Probabilities on this day suggested a widespread region of risk from southern Nebraska south to the Texas–Mexico border (Fig. 15a). These probabilities exceeded 30% across most of Texas. While one report did occur in this area, the majority of reports took place away from the area of highest probabilities. In addition, false alarms were present across most of Oklahoma and Texas. Again, usage of the environmental information

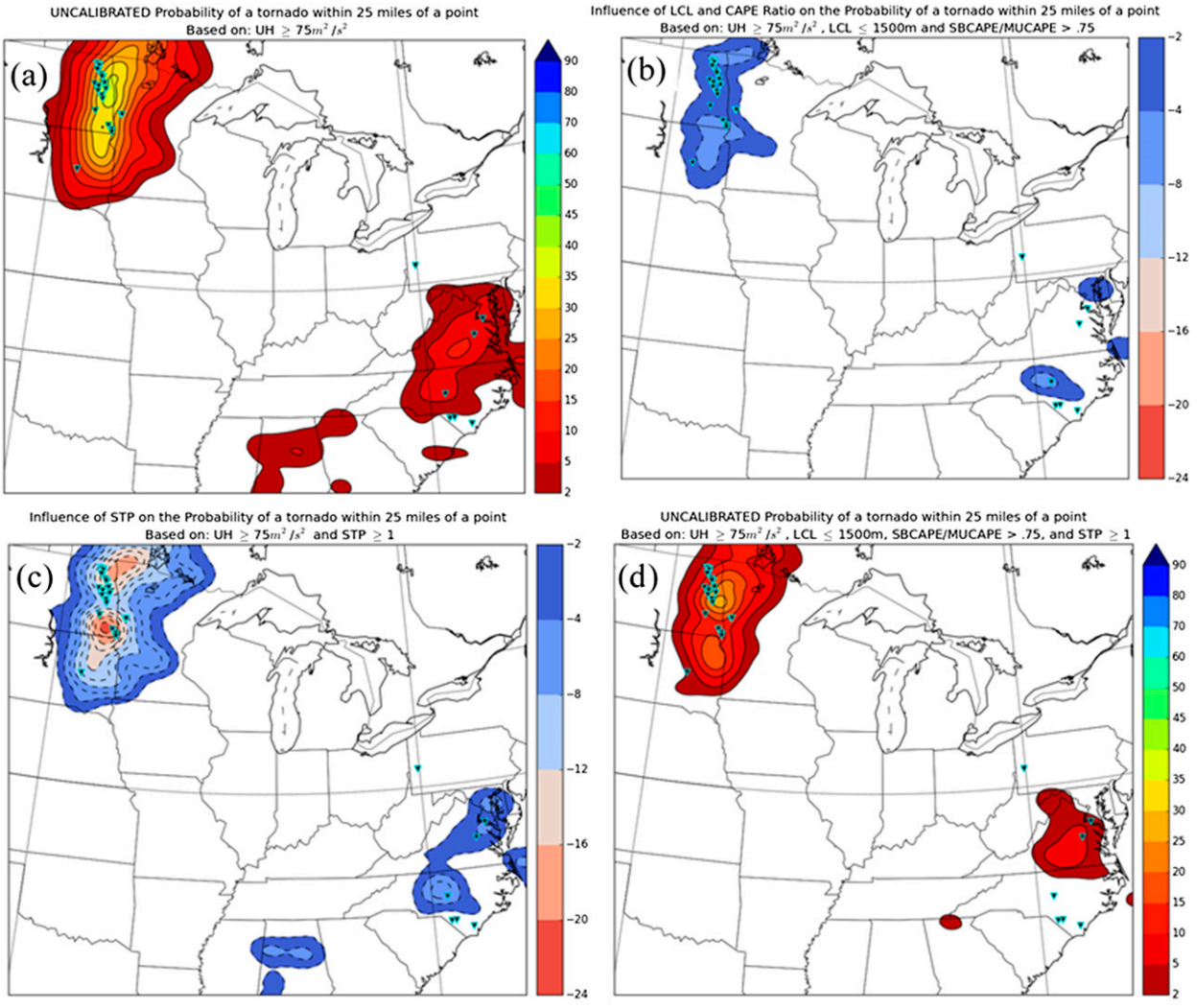

FIG. 13. As in Fig. 11 but from 1200 UTC 27 Jun to 1200 UTC 28 Jun 2015.

decreased the probabilities (Figs. 15b,c). The decrease in probabilities using the LCL/CAPE ratio method (Fig. 15b) was fairly uniform across Texas and Oklahoma, but lowered the probabilities where most of the tornado reports occurred. The STP method (Fig. 15c) reduced the probabilities much more than the LCL/CAPE ratio method did, but again the highest-magnitude reductions were near the actual string of reports. The large area of false alarms remained over Oklahoma and eastern Texas and was not reduced much by the inclusion of environmental information. When all of the environmental information is included in the probabilities (Fig. 15d), a large area of false alarms persists, particularly in south-central Texas, far from the majority of the reports. In addition, one of the tornado reports included in the UH-only method (Fig. 15a) now is outside the envelope of probabilities.

This case highlights the difficulties encountered when calculating probabilities in MCS situations. While the mode is often easily discernable when looking at simulated reflectivity, the presence of UH within the squall lines and the presence of ingredients conducive to tornadogenesis in systems present a difficult problem. Further, MCSs occasionally do produce tornadoes, and ideally probabilities would reflect this potential. It is beyond the scope of this work to lower the probabilities when the expected mode is linear in nature, while maintaining probabilities that reflect the MCS tornado threat.

c. *Subjective verification*

Subjective verification of the tornado probabilities took place during SFE 2015, from 5 May to 4 June 2015. Each participant was asked on a daily basis to rate the four probabilities from the previous day generated using a

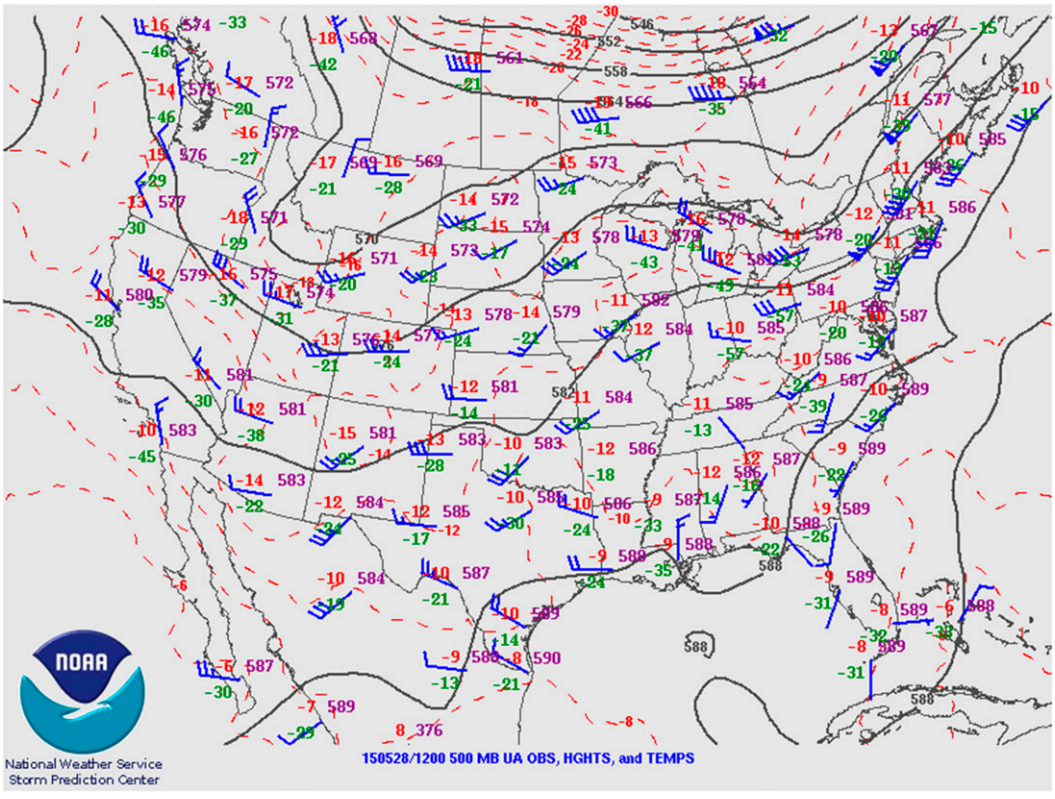

FIG. 14. As in Fig. 10, but for 1200 UTC 28 May 2015. [Obtained from the SPC website (www.spc.noaa.gov/exper/archive/event.php?date=20150528).]

UH threshold of 75 $m^2 s^{-2}$ and a σ of 50 km. In the case of Monday, the most active day from the previous weekend was considered. These ratings ranged from 1 (very poor) to 10 (very good), in response to the following question:

> "Subjectively rate the NSSL-WRF 24 h tornado probabilities using a rating scale of Very Poor (1) to Very Good (10). We are testing the use of updraft helicity as forecast by the NSSL-WRF ensemble to derive tornado probabilities at time and space scale consistent with SPC outlooks. UH $\geq$ 75 is used as a proxy for tornadoes and various methods are tested to only consider UH in environments typically supportive of tornadoes."

Incorporation of environmental information produced higher mean subjective ratings (Fig. 16) over the UH-only method for the 24-h probabilities. Of the 22 days of evaluation, the LCL/CAPE ratio method had or was tied for the highest average rating on 9 days, the STP method and the LCL/CAPE ratio/STP method had or were tied for the highest average rating on 8 days, and the UH-only method had or was tied for the highest average rating on 6 days. The UH-only method and LCL/CAPE ratio were rated the same on four days, and the STP and LCL/CAPE ratio/STP method were rated the same on seven days. Thus, many participants saw a strong similarity between the STP and the LCL/CAPE ratio/STP method.

Overall, the participants' comments described some common themes. Most of the participants found the guidance to be useful, and noted that the incorporation of environmental information focused the area of interest and reduced false alarms as per the aim of this study, with multiple comments such as

> "All products capture the area, axis, and grouping of the tornado reports very well. The naive UH probabilities show too much false alarm area in SW Oklahoma, but the additional filters correct that area very well."

These comments suggested that forecasters would like to have the probabilities available when they are forecasting, and that they would glean information at a glance, rather than mentally integrating all of the ensemble data upon which these probabilities are based.

The participants' main concerns were the high magnitude of probabilities on multiple days and displacement of the bull's-eye of high probabilities from eventual tornado reports on multiple days. The high magnitude of the probabilities correspond to relatively high-risk categories, as assigned by the SPC, resulting in comments such as

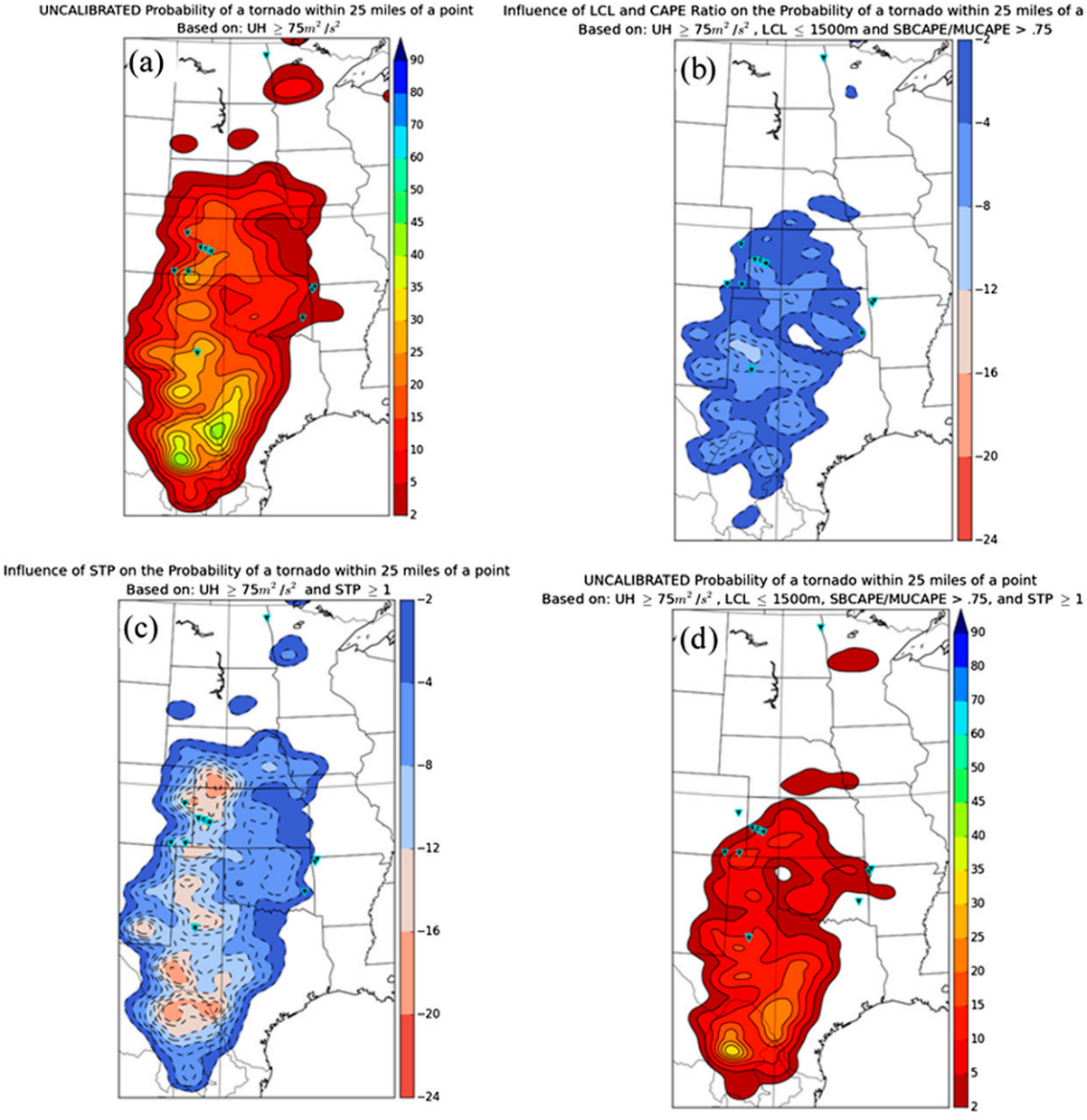

FIG. 15. As in Fig. 11 but from 1200 UTC 28 May to 1200 UTC 29 May 2015.

> "Several reports occurred outside of the bullseye of tornado probs, and there were only a few tornadoes in the area in Oklahoma that had probs over 30%, even with the most discriminating filters. 30% is high risk, and the reports did not seem like a high risk day to me."

However, from the objective verification discussed previously, high magnitudes are only slightly overforecast according to the reliability diagrams.

Adding environmental information did occasionally have a downside, as was noted in the 27 June 2015 case study; the STP-inclusive probabilities were occasionally noted by the participants as being too limiting, and excluding tornado reports that the less restrictive methods maintained within low probabilities:

> "Large false alarm areas. However the two tornado reports were near the high probability areas. After filtering, the Wyoming tornado was missed although the false alarm area was greatly reduced."

Forecasters have different opinions about whether it is more important to not miss events or to reduce the number of false alarms, and through the SFE the probabilities were rated by forecasters with a mix of these views.

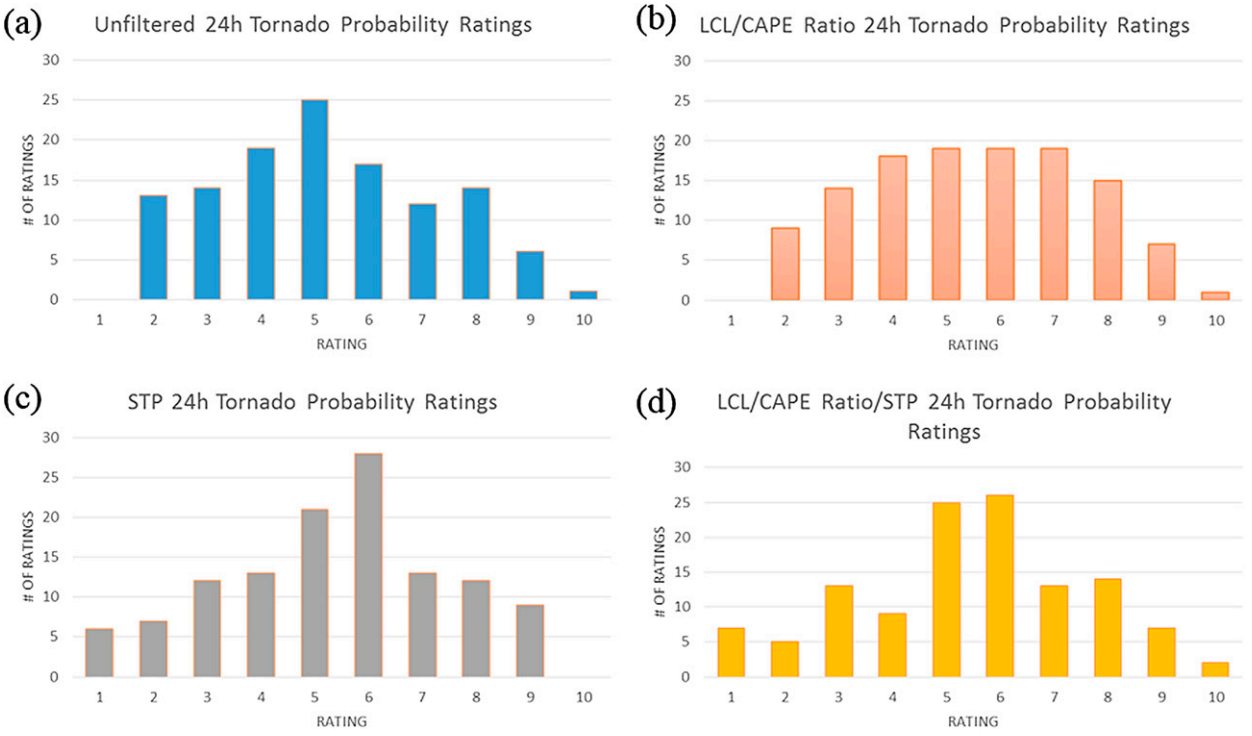

FIG. 16. Subjective ratings of the tornado probabilities by participants in SFE 2015 for (a) UH only, (b) requiring LCL < 1500 m and SBCAPE/MUCAPE > 0.75, (c) requiring STP ≥ 1, and (d) requiring LCL < 1500 m, SBCAPE/MUCAPE > 0.75, and STP ≥ 1. Ratings encompassed 24 cases.

Finally, the participants noted the difference between days with few tornadoes and days with more tornadoes, namely, that marginal days posed more difficulties because of the weaker environmental parameters naturally present on those days:

> "Some displacement from the area where reports occurred. Max probability value ~4X greater than the density of storm reports - so the parameter is running quite hot. Missed event, which probably is more a miss of the underlying forecast than any aspect of the parameter space shown. The filters that included STP reduced the max values, which for this event moved closer to observed report density."

Since the probabilities are mostly ingredients based, it is to be expected that the days with less favorable environments would produce fewer tornadoes, and that the probabilities would have difficulty pinpointing exactly where these tornadoes would occur.

Overall, the participant comments were positive and reinforced the results produced through objective analysis while providing insight into how a forecaster might utilize these probabilities operationally. They also highlighted areas of potential improvement and concerns, which will be taken into account in future work.

4. Summary and discussion

High-resolution models are a very useful resource for forecasters, but the amount of information available from these models continues to grow while the amount of time a forecaster has often is fixed. This work attempts to provide a "first guess" forecast of tornadoes from the high-resolution NSSL-WRF ensemble. Information output by the ensemble, such as UH, STP, LCL height, and SBCAPE/MUCAPE ratio are synthesized into probabilities. The first question addressed by this study asks which UH threshold and σ value maximized both reliability and skill in forecasting tornadoes. Utilizing the area under the ROC curve, CSI, and reliability diagrams, this study suggests a UH threshold of $75\,m^2\,s^{-2}$ maximizes reliability, while producing graphics of similar smoothness to those already issued operationally and maintaining a high ROC area. Lower thresholds of UH were also considered, but produced large areas of overforecasting. However, all thresholds of UH produced less overforecasting than what was found when considering environmental information, such as STP, without considering UH. Small smoothing radii greatly overforecasted and produced noisy graphics; using a larger σ ensures that the

probabilities are not tied to specific UH tracks within the model.

When our results are compared to the calibrated tornado forecasts of Jirak et al. (2014), they demonstrate higher CSI values at UH thresholds above $25\,m^2\,s^{-2}$. As Jirak et al. (2014) used calibrated probabilities based on historical relative frequencies, these probabilities have the advantage of higher CSIs while not requiring historical report information. Reliabilities between the two studies were comparable, and both performed more poorly than did the SPC day 1 outlooks reported by Jirak et al. (2014). However, the addition of higher-resolution ensemble data appears to improve the CSI of these uncalibrated probabilities beyond the calibrated probabilities using coarser-resolution environmental information, suggesting that the higher-resolution environmental information benefits the probabilities.

The second question of this study asked whether the incorporation of environmental information to UH information would improve the probabilities. While ROC areas decreased slightly with the addition of environmental information across all UH and σ thresholds, CSI increased. ROC area reduction is thought to be due to lower skill at very low probability thresholds and the large influence of correct negatives, as supported by the CSI. However, the inclusion of environmental information reduced the area of false alarms in many individual cases—STP generally more so than LCL height and CAPE ratio. The inclusion of environmental information also led to an improvement in reliability across all cases.

Subjectively, this finding was supported by participants during SFE 2015, in their comments and their ratings, which favored the probabilities incorporating environmental information over the UH-only probabilities. Both verifications suggest that high-resolution environmental information helps distinguish tornadoes from other severe convective hazards. Subjective evaluation also suggests that these probabilities are useful to forecasters, particularly from SFE 2015 participant comments. The integration of environmental parameters with UH values into one map of probabilities saves forecasters time and effort. To that end, three case studies are presented in which the probabilities could give forecasters an idea of the tornado threat. An overwhelmingly mixed-mode day and a day with the potential for tornadoes in a climatologically less-favored area for tornadoes than the central Great Plains show the ability of the probabilities to handle multiple-tornadic scenarios. A third case demonstrates weaknesses of the probabilities and provides the focus of the future work.

Future work includes ongoing collaboration with SPC forecasters on using UH and STP to generate empirically calibrated probabilities. Preliminary results suggest that these probabilities could provide very different guidance from the method described in this study. Future work will also focus on exploring the relationship between model-generated STP and STP obtained from the ROC reanalysis of tornado events, as well as the relationship between model-generated UH and the radar-observed rotational velocity of storms. Future probabilities will be tested in upcoming SFEs and objectively analyzed, to provide the best possible first guess tool for forecasters in their pursuit of an accurate tornado forecast.

Acknowledgments. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant DGE-1102691, Project A00-4125. AJC and SRD were provided support by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, 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

Brooks, H. E., and C. A. Doswell III, 2002: Deaths in the 3 May 1999 Oklahoma City tornado from a historical perspective. *Wea. Forecasting*, **17**, 354–361, doi:10.1175/1520-0434(2002)017<0354:DITMOC>2.0.CO;2.

Buizza, R., A. Hollingsworth, F. Lalaurette, 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:PPOPUT>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, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, doi:10.1175/2009WAF2222222.1.

——, and Coauthors, 2012a: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, doi: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, doi:10.1175/WAF-D-11-00147.1.

Colman, B. R., 1990: Thunderstorms above frontal surfaces in environments without positive CAPE. Part I: A climatology. *Mon. Wea. Rev.*, **118**, 1103–1122, doi:10.1175/1520-0493(1990)118<1103:TAFSIE>2.0.CO;2.

Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF model output for severe weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*, **25**, 408–427, doi:10.1175/2009WAF2222258.1.

Corfidi, S. F., S. J. Weiss, J. S. Kain, S. J. Corfidi, R. M. Rabin, and J. J. Levit, 2010: Revisiting the 3–4 April 1974 super outbreak of tornadoes. *Wea. Forecasting*, **25**, 465–510, doi:10.1175/2009WAF2222297.1.

Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) Model. *Atmos. Sci. Lett.*, **5**, 110–117, doi:10.1002/asl.72.

Doswell, C. A., III, H. E. Brooks, and N. Dotzek, 2009: On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563, doi:10.1016/j.atmosres.2008.11.003.

Du, J., and Coauthors, 2014: NCEP regional ensemble update: Current systems and planned storm-scale ensembles. *Proc. 26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction*, Atlanta, GA, Amer. Meteor. Soc., J1.4. [Available online at https://ams.confex.com/ams/94Annual/webprogram/Paper239030.html.]

Duda, J. D., and W. A. Gallus Jr., 2010: Spring and summer midwestern severe weather reports in supercells compared to other morphologies. *Wea. Forecasting*, **25**, 190–206, doi:10.1175/2009WAF2222338.1.

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, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>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, doi:10.1175/1520-0434(2003)018<0782:SHNPOC>2.0.CO;2.

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, doi: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, doi:10.1016/j.atmosres.2010.11.011.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, doi: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, doi: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.

Jirak, I. L., S. J. Weiss, and C. J. Melick, 2012: The SPC storm-scale ensemble of opportunity: Overview and results from the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *Proc. 26th Conf. Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 137. [Available online at https://ams.confex.com/ams/26SLS/webprogram/Paper211729.html.]

——, 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. *Proc. 27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 2.5. [Available online at 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, doi: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.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875, doi:10.1029/RG020i004p00851.

Mercer, A. E., C. M. Shafer, C. A. Doswell III, L. M. Leslie, and M. B. Richman, 2012: Synoptic composites of tornadic and nontornadic outbreaks. *Mon. Wea. Rev.*, **140**, 2590–2608, doi:10.1175/MWR-D-12-00029.1.

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, doi:10.1029/97JD00237.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, doi:10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164, doi: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, doi:10.1175/2008WAF2222159.1.

Schwartz, C., G. Romine, M. Weisman, R. Sobash, K. Fossell, K. Manning, and S. 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.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 2. NCAR Tech Note. NCAR/TN–475+STR, 113 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.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.

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, doi:10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

——, ——, and C. M. Mead, 2004: An update to the supercell composite and significant tornado parameters. Preprints, *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor.

Soc., P8.1. [Available online at https://ams.confex.com/ams/pdfpapers/82100.pdf.]

University Corporation for Atmospheric Research, 2015: Report of the UCACN Model Advisory Committee. UCAR, 72 pp. [Available online at http://www.ncep.noaa.gov/director/ucar_reports/ucacn_20151207/UMAC_Final_Report_20151207-v14.pdf.]

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.

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, doi:10.1175/1520-0493(2001)129<0729:EOASRM>2.0.CO;2.

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.

Weiss, S. J., D. R. Bright, J. S. Kain, J. J. Levit, M. E. Pyle, Z. I. Janjic, B. S. Ferrier, and J. Du, 2006: Complementary use of short-range ensemble and 4.5 KM WRF-NMM model guidance for severe weather forecasting at the Storm Prediction Center. Preprints, *23rd Conf. Severe Local Storms*, St. Louis MO, Amer. Meteor. Soc., 8.5. [Available online at http://www.spc.noaa.gov/publications/weiss/spc-sref.pdf.]

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences: An Introduction*. 3rd ed. Elsevier, 676 pp.

CORRIGENDUM

BURKELY T. GALLO

School of Meteorology, University of Oklahoma, Norman, Oklahoma

ADAM J. CLARK

NOAA/OAR/National Severe Storms Laboratory, 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 and in final form 18 April 2016)

After publication of our paper "Forecasting tornadoes using convection-permitting ensembles" (Gallo et al. 2016, hereafter G16), we noticed an error in the computation of the performance diagrams in G16's Fig. 6. Though this error has no effect on G16's conclusions, we would like to present the corrected Fig.6 by means of this corrigendum.

Rather than showing the probability levels of 2%, 5%, 10%, 15%, 30%, and 45%, with the 15% level highlighted via a point, Fig. 6 of G16 shows probability levels of 1%, 4%, 5%, 10%, 25%, and 40%, with the 10% level highlighted via a point. While the points composing the previous line were accurately computed, they simply depicted the incorrect range of probabilities. The corrected Fig. 6 gives the performance diagrams over the correct range of probabilistic tornado forecasts issued operationally by the Storm Prediction Center.

REFERENCE

Gallo, B. T., A. J. Clark, and S. R. Dembek, 2016: Forecasting tornadoes using convection-permitting ensembles. *Wea. Forecasting*, **31**, 273–295, doi:10.1175/WAF-D-15-0134.1.

Corresponding author address: Burkely T. Gallo, National Weather Center, NSSL/FRDD, 120 David L. Boren Blvd., Norman, OK 73072.
E-mail: burkely.twiest@noaa.gov

DOI: 10.1175/WAF-D-16-0076.1

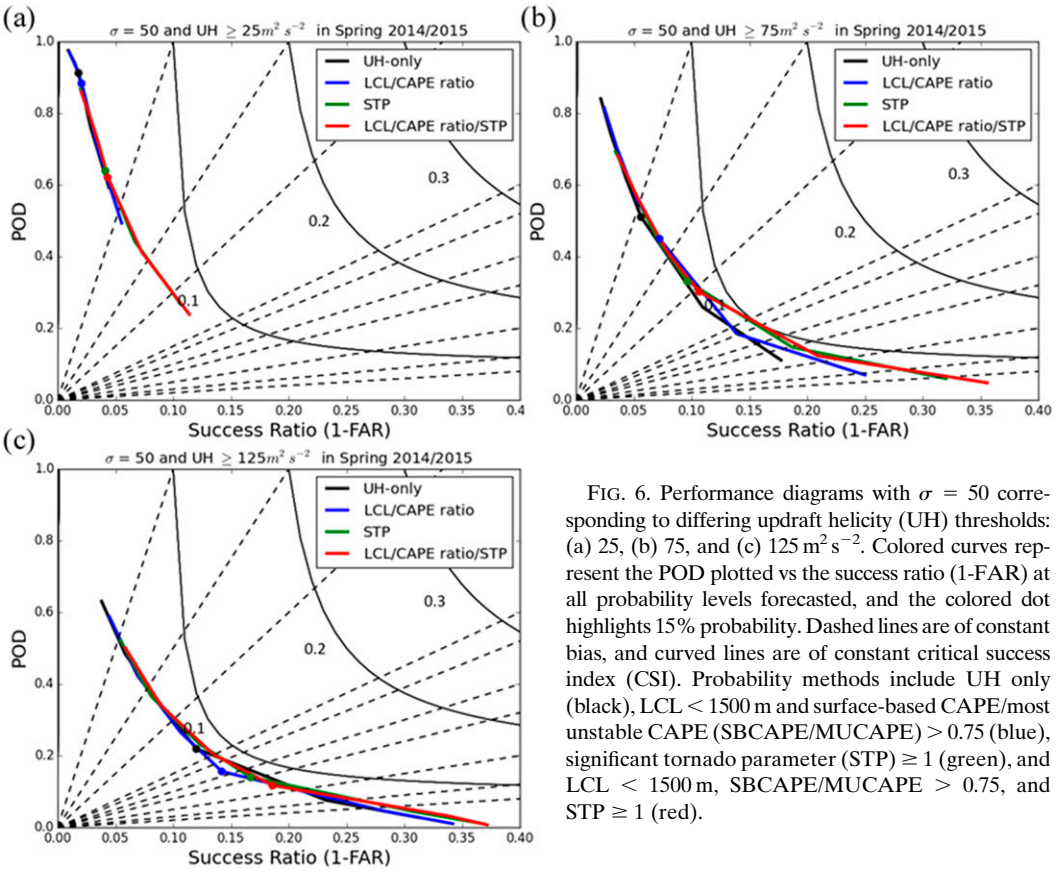

FIG. 6. Performance diagrams with $\sigma = 50$ corresponding to differing updraft helicity (UH) thresholds: (a) 25, (b) 75, and (c) 125 $m^2 s^{-2}$. Colored curves represent the POD plotted vs the success ratio (1-FAR) at all probability levels forecasted, and the colored dot highlights 15% probability. Dashed lines are of constant bias, and curved lines are of constant critical success index (CSI). Probability methods include UH only (black), LCL < 1500 m and surface-based CAPE/most unstable CAPE (SBCAPE/MUCAPE) > 0.75 (blue), significant tornado parameter (STP) ≥ 1 (green), and LCL < 1500 m, SBCAPE/MUCAPE > 0.75, and STP ≥ 1 (red).