

Incorporating UH Occurrence Time to Ensemble-Derived Tornado Probabilities

BURKELY T. GALLO

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

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

(Manuscript received 25 June 2018, in final form 29 November 2018)

ABSTRACT

Probabilistic ensemble-derived tornado forecasts generated from convection-allowing models often use hourly maximum updraft helicity (UH) alone or in combination with environmental parameters as a proxy for right-moving (RM) supercells. However, when UH occurrence is a condition for tornado probability generation, false alarm areas can occur from UH swaths associated with nocturnal mesoscale convective systems, which climatologically produce fewer tornadoes than RM supercells. This study incorporates UH timing information with the forecast near-storm significant tornado parameter (STP) to calibrate the forecast tornado probability. To generate the probabilistic forecasts, three sets of observed climatological tornado frequencies given an RM supercell and STP value are incorporated with the model output, two of which use UH timing information. One method uses the observed climatological tornado frequency for a given 3-h window to generate the probabilities. Another normalizes the observed climatological tornado frequency by the number of hail, wind, and tornado reports observed in that 3-h window compared to the maximum number of reports in any 3-h window. The final method is independent of when UH occurs and uses the observed climatological tornado frequency encompassing all hours. The normalized probabilities reduce the false alarm area compared to the other methods but have a smaller area under the ROC curve and require a much higher percentile of the STP distribution to be used in probability generation to become reliable. Case studies demonstrate that the normalized probabilities highlight the most likely area for evening RM supercellular tornadoes, decreasing the nocturnal false alarm by assuming a linear convective mode.

1. Introduction

The addition of convection-allowing model (CAM) ensembles to the suite of available numerical guidance provides severe weather forecasters information regarding convective mode when generating forecasts (Kain et al. 2003; Clark et al. 2012). Indeed, as computing power increases, ever more guidance is becoming

available to forecasters (Gallo et al. 2017). As such, summary products for severe weather forecasters have been developed using storm-scale metrics alone and in combination with environmental information (Sobash et al. 2011; Gallo et al. 2016; Loken et al. 2017; Gagne et al. 2017). Many of the products include the hourly maximum updraft helicity (UH; Kain et al. 2010), a storm-scale rotation metric that indicates a forecasted midlevel mesocyclone and is often used as a proxy for a right-moving (RM) supercell (Naylor et al. 2012). Since supercells produce many severe convective storm reports,

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

DOI: 10.1175/WAF-D-18-0108.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](https://www.ametsoc.org/PUBSReuseLicenses).

UH has been widely used to forecast severe thunderstorms (Sobash et al. 2011; Sobash et al. 2016b; Loken et al. 2017). Efforts have recently expanded from forecasting any type of severe convection to specific hazards (Gallo et al. 2016; Gagne et al. 2017) by including environmental parameters. However, these individual hazard forecasts often suffer from overforecasting, particularly in the case of tornado forecasts (Gallo et al. 2016, 2018; Sobash et al. 2016a). Reducing the overforecasting of tornadoes from ensemble-based methods would greatly benefit forecasters by ensuring that they have reliable first-guess information from the CAMs, and that is the aim of this work.

The significant tornado parameter (STP) is a common environmental parameter in numerical weather prediction forecasts and is often considered by forecasters when characterizing the severe convective environment. The STP was developed by Thompson et al. (2003) and adapted by Thompson et al. (2012) to highlight favorable environmental conditions for significant (EF2+) tornado occurrence. Smith et al. (2015) and Thompson et al. (2017) developed climatologies of tornado occurrence given an RM supercell and a corresponding STP value. These climatologies were employed by Gallo et al. (2018) to generate probabilistic tornado forecasts using a 10-member CAM ensemble. The ensemble was based on a 4-km experimental version of the Weather Research and Forecasting (WRF; Skamarock et al. 2008) Model run at the National Severe Storms Laboratory (NSSL), known as the NSSL-WRF ensemble (Gallo et al. 2016; Clark 2017). Member variety came from different initial and lateral boundary conditions; otherwise, members had the same configuration. Ensemble probabilities were calibrated by empirical climatological frequencies, resulting in skillful forecasts of tornadoes from RM supercells (RM tornadoes) that slightly overforecasted tornado occurrence (Gallo et al. 2018).

Utilizing the climatological frequencies created more reliable and skillful forecasts than other methods of probabilistic forecast generation that consider both the UH value and the STP value as thresholds to be exceeded. The binary paradigm stemming from a threshold exceedance method only assigned a “yes” or a “no” to each point and member; an STP value either did or did not exceed the threshold. The climatological frequencies, however, treated each grid point probabilistically by assigning a climatological frequency at each point based on the specific STP value. While in the threshold exceedance paradigm an STP of 1.5 and an STP of 6.3 were assigned the same value (yes or 1), in the probabilistic paradigm the higher STP would be assigned a higher probability according to the climatology. Despite the improved reliability of Gallo et al.’s (2018) climatologically based

forecasts, they still occasionally generated large false alarm areas that were often linked to nocturnal mesoscale convective systems (MCSs). MCSs are less likely to produce tornadoes than supercells (Smith et al. 2012), so false alarms associated with MCSs could provide incorrect information to forecasters. This work attempts to use the observed climatology and timing of UH occurrence to reduce the false alarm areas from nocturnal UH occurrence, assuming that much of that UH is associated with linear convective modes.

Section 2a briefly describes how this study adapts the methodology of Gallo et al. (2018) using normalization and time-dependent techniques, followed by a description of the NSSL-WRF ensemble forecasts, Storm Prediction Center (SPC) forecasts, and verification metrics used in section 2b. Aggregated statistical results from all cases encompassed by this study are presented in section 3a, and section 3b hones in on three example case studies to demonstrate how the forecaster would see the different methodologies on a given day. Finally, section 4 presents conclusions and ideas for future work.

2. Data and methodology

a. Probabilistic forecast generation

Probabilistic forecasts were generated across 244 days in 2014 and 2015 during spring and early summer (1 April–31 July) using the technique of Gallo et al. (2018), key components of which will be described here. This technique incorporates empirical environmental frequencies of a tornado given an RM supercell associated with a severe report and a modified STP value. The modified STP utilizes capping functions from the effective-layer STP [e.g., if $\text{SHR6} < 12.5 \text{ m s}^{-1}$, that term is set to 0; Thompson et al. (2012)] while recognizing the inability to efficiently calculate the effective inflow layer at every point within the CAM ensemble. It is defined by

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

where SBCAPE, SBCIN, and SBLCL are the convective available potential energy (CAPE), convective inhibition (CIN), and lifted condensation level (LCL) of a surface-based parcel, respectively. SHR6 is the 0–6-km bulk shear, and SRH1 is the 0–1-km storm-relative helicity. The climatological tornado occurrence frequencies utilize 1202 tornado reports and 5422 hail or wind reports occurring from January 2014 to

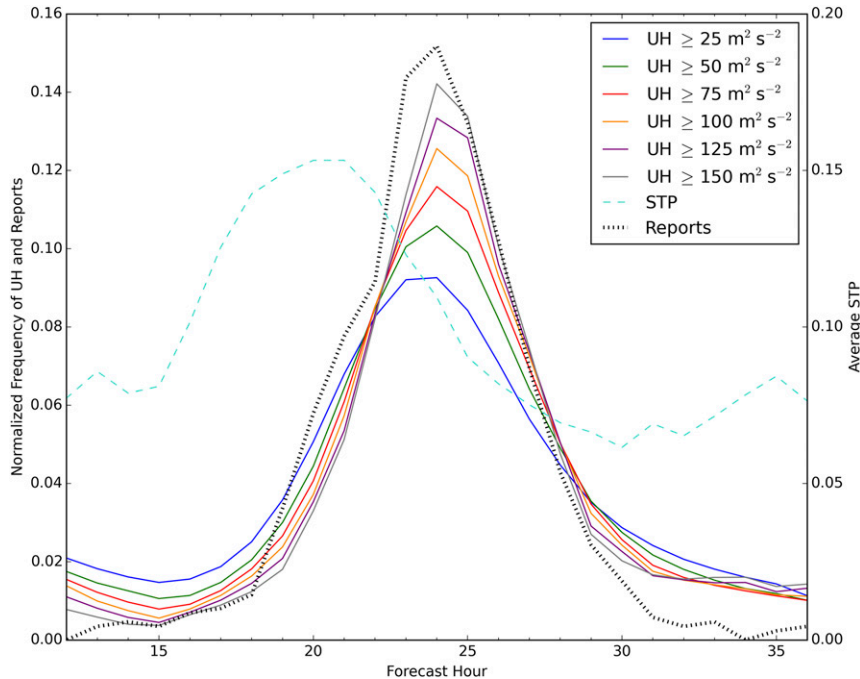


FIG. 1. Reports from RM supercells, UH, and STP diurnal distributions. Plots begin at forecast hour 13, corresponding to 1300 UTC on the day of the forecasts, and end at forecast hour 36, corresponding to 1200 UTC on the following day.

December 2015 (Thompson et al. 2017; Gallo et al. 2018). Frequencies are calculated by dividing the number of tornado reports from RM supercells [using Smith et al.'s (2015) RM supercell definition] by the number of hail, wind, and tornado reports from RM supercells in each STP bin. This calculation provides the frequency of tornado occurrence given a RM supercell that produces a report.

These frequencies were combined with NSSL-WRF ensemble forecasts initialized at 0000 UTC and extending to 36 h to generate 24-h probabilistic tornado forecasts valid from 1200 UTC to 1200 UTC the following day. Following Gallo et al. (2018), hourly forecast values of UH and STP were extracted from the ensemble. For each ensemble member and forecast hour between 1200 and 1200 UTC the following day (spanning forecast hours 13–36), each grid point was checked for UH exceeding $25 \text{ m}^2 \text{ s}^{-2}$ anywhere within a 40-km radius as a proxy for an RM supercell. If UH exceeded this threshold, the point STP value from the previous hour was added to a distribution of STP at that grid point such that every member, hour, and grid point maintained a separate STP distribution. Creating the distribution of STP accounts for the multiple potential inflow STP values the storm may ingest, since the STP is from the hour prior to UH occurrence. Next, a percentile of the distribution was selected as the representative

STP value for that forecast hour. Percentiles tested in this work were the 10th, 25th, 50th, 75th, 90th, and 100th. After representative STP values were assigned to each member, the daily maximum STP value from this process was calculated. That daily maximum STP was then associated with a frequency using the climatology, and the climatological frequency became the probability assigned to that point and member. Finally, an average of the probabilities at each grid point was taken across all members and smoothed using a Gaussian kernel, creating a final probabilistic field similar to probabilistic forecasts issued by the SPC. This method will henceforth be called the daylong method and is based on climatological frequencies generated using all available reports, no matter their time of occurrence.

When this methodology used the observed climatological frequencies generated independent of time, UH swaths associated with nocturnal MCSs often produced false alarm areas (Gallo et al. 2018). While RM tornado reports show a steep peak during the afternoon hours, overnight hours contain only a small fraction of reports from RM supercells (Fig. 1). However, the diurnal UH cycle maintains UH throughout the evening hours at even high thresholds (e.g., $\text{UH} \geq 150 \text{ m}^2 \text{ s}^{-2}$). Thus, to reduce false alarm, two methods of incorporating timing information through the climatological frequencies were applied. While the daylong methodology described

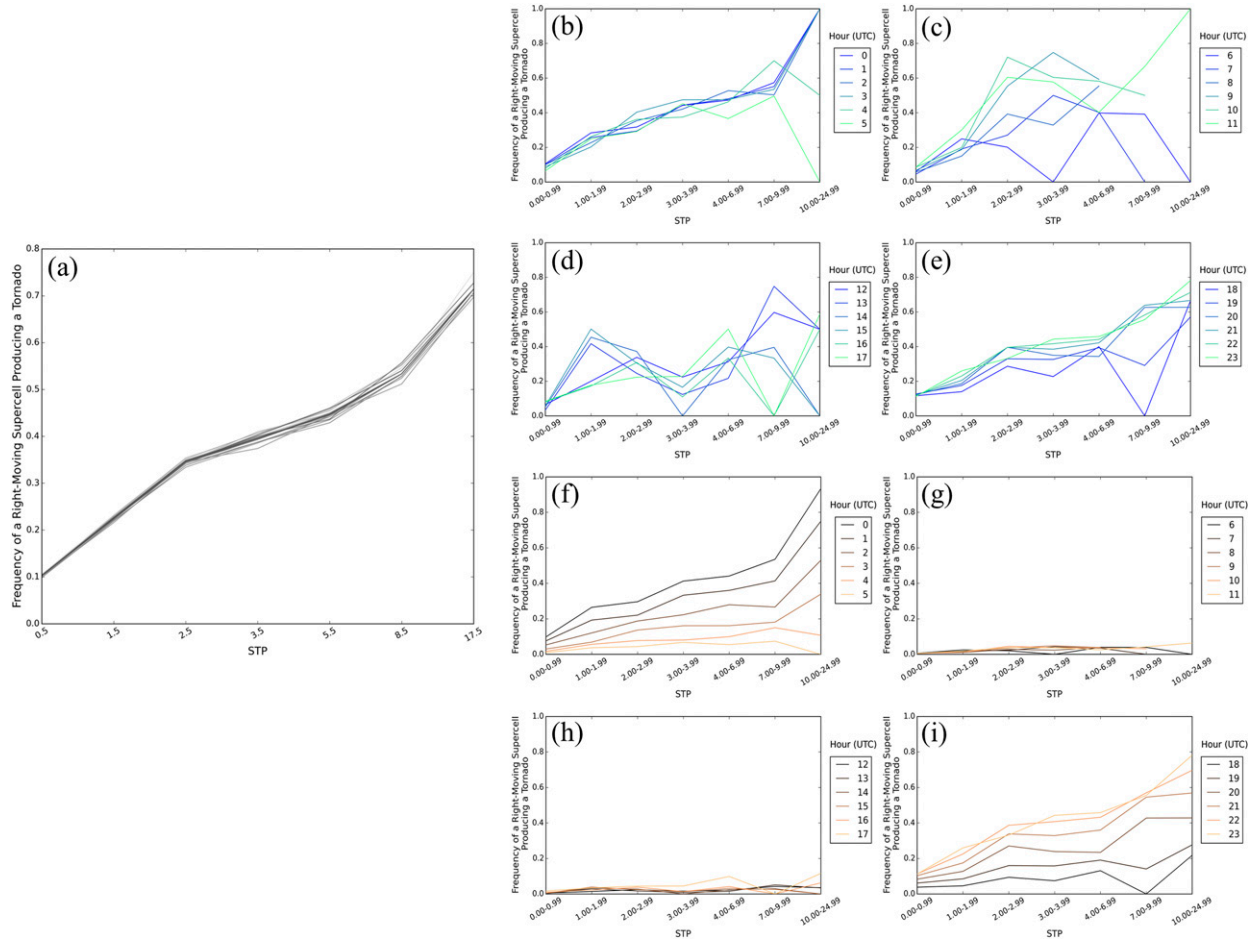


FIG. 2. (a) The climatological frequencies given an RM supercell and STP value used to calculate the daylong probabilities. Each line represents a different week's frequency calculated based on the cross-validation technique. (b)–(e) Climatological frequency of tornado occurrence given an RM supercell, time of day, and STP value. Each color represents the center of a 3-h time window, averaged over the cross-validated weeks in the April–July 2014 and 2015 verification period. (f)–(i) As in (b)–(e), but for the climatological frequencies of tornado occurrence normalized by the maximum number of hail, wind, and tornado reports during a given 3-h window.

previously calculated the probabilities independent of the report occurrence time (Fig. 2a), this study also tested methodologies that parse the climatological frequencies using a moving 3-h time window centered on the hour of interest (Figs. 2b–i). The first approach incorporating time will be called the nonnormalized time-dependent method, and has been used to explore how the diurnal variation in the relationship between STP and tornado frequency would affect the subsequent probabilities (Figs. 2b–e). The second method incorporated the frequencies from the 3-h windows *and* the number of hail, wind, and tornado reports associated with RM supercells occurring in each window. Each 3-h window's climatological frequencies were weighted by the number of reports occurring therein, with the 3-h window containing the most reports (2300 UTC) assigned a weight of one (Figs. 2f–i). This approach will be known as

the normalized time-dependent method and tested the effect of calibrating the probabilities with both timing information and a supercell report climatology. Additionally, while the daylong probabilities were interpolated between STP bins, the time-dependent probabilities were not due to a smaller sample size for the 3-h windows.

Separate climatological frequencies were calculated for each week during the study period, and the reports for a given week were excluded from the climatology used to calculate the probabilities for that week. This cross-validation technique (Elsner and Schmertmann 1994) follows Gallo et al. (2018) and has previously been applied to surrogate severe probabilities (Sobash and Kain 2017), preventing the reports that the probabilities are verified on from informing the probabilities. Variability was highest from week to week when

the sample sizes were small, as at high STP values (Fig. 2a).

b. Verification metrics and data

Ensemble-generated forecasts were verified alongside the 0600 UTC forecasts from the SPC, since the ensemble probabilities are designed as operational first-guess tornado probability forecasts and ideally would behave comparably to the SPC forecasts, which have a high POD and have been shown to be reliable when forecasting RM supercells (Hitchens and Brooks 2017; Gallo et al. 2018). The 0600 UTC convective outlooks from the SPC were chosen for comparison because they are the first product issuance where forecasters have information from 0000 UTC runs of CAM ensembles. Verification metrics used include the area under the receiver operating curve (ROC area; Mason 1982), reliability diagrams, and performance diagrams (Roebber 2009). The ROC area describes how forecasts discriminate areas of event occurrence from areas of event nonoccurrence by plotting the probability of detection (POD) versus the probability of false detection (POFD) but contains no bias information. Reliability diagrams plot the observed frequency versus the forecast probability, complementing the ROC areas. Performance diagrams visualize four different contingency-table-based metrics, including the bias, the success ratio (SR), the POD, and the critical success index (CSI), which is often used as a rare-event score (Wilks 2011). Statistics were generated at each of the probability thresholds forecast by the SPC: 2%, 5%, 10%, 15%, 30%, 45%, and 60%. Verification statistics were computed over approximately the eastern $\frac{2}{3}$ of the contiguous United States for all methods and for the SPC forecasts, which were regridded onto the 4-km NSSL-WRF ensemble grid. Observed tornado path data were also regridded onto the 4-km NSSL-WRF ensemble grid prior to verification, and treated as yes/no events. A yes event occurred if a tornado from a RM supercell passed within 40 km of a point, consistent with the SPC's spatial forecast definition.

3. Results

a. Aggregated performance statistics

The aim of incorporating the time of UH occurrence is to reduce the nocturnal false alarm by leveraging the daily cycle of severe storm reports. To determine the impact of the timing information, average probabilities for each hour across the entire domain and study period were created using each forecast method. The diurnal cycle of the daylong and the nonnormalized time-dependent probabilities maintained areas of

probability throughout the nocturnal hours, while the normalized time-dependent probabilities showed a sharp decrease from the afternoon peak that resulted in nearly zero probability overnight (Fig. 3). The non-normalized time-dependent probabilities were greater throughout the diurnal cycle compared to the daylong probabilities, and the normalized time-dependent probabilities had the highest afternoon probabilities, likely due to the different percentiles of STP used in their generation (i.e., 90th percentile vs 10th or 25th percentile).

The percentile differences in the average forecast probabilities were the result of selecting the most reliable percentile with sufficiently high ROC areas (ROC area ≥ 0.7) for comparison between each method. Reliability diagrams were used to subjectively determine which percentile was the most reliable for each method: the 10th percentile for the daylong probabilities (Fig. 4a), the 25th percentile for the nonnormalized time-dependent probabilities (Fig. 4b), and the 90th percentile for the normalized time-dependent probabilities (Fig. 4c). Optimal reliability for this study included values that fall on the “underforecast” side of the reliability diagram, due to operational constraints in forecast probabilities that the SPC may issue. For example, even if a forecaster thinks that there is an area of 14% probability, they will issue a 10% contour, as the next highest contour they can issue is 15%. Therefore, SPC forecasts by nature underforecast, and so the most reliable percentile chosen for each method is also occasionally an underforecast. Reliability varied greatly between percentile of STP used and between probability generation methods because of the differences in the climatological frequency equations used to calibrate the probabilities. For example, at 1600 UTC, an STP of 1.5 would produce a frequency of ~ 0.22 in the daylong method, a frequency of ~ 0.08 in the non-normalized time-dependent method, and a frequency of ~ 0.009 in the normalized time-dependent method.

In contrast to the reliability, ROC areas differed little between the percentiles of STP chosen for a given method (not shown). All percentiles of the daylong and the nonnormalized time-dependent probabilities had higher ROC areas than the SPC forecasts, while all of the normalized time-dependent probabilities had lower ROC areas than the SPC (in most cases because the POD was lower than the SPC with a very similar POFD; not shown). Between methods, the normalized time-dependent probabilities had a lower ROC area than the other methods largely due to decreased POD (Fig. 4d), while the other methods were similar to one another. The largest difference between the methods was at the 2% threshold, showing the sensitivity to this lowest forecast

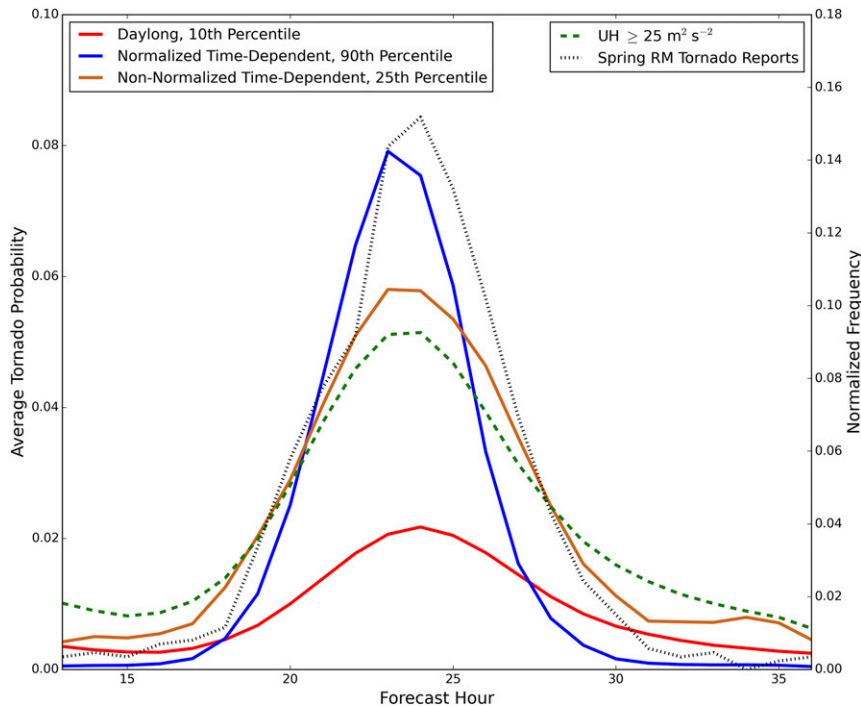


FIG. 3. The diurnal cycle of report frequency, UH frequency, and average probability over the verification domain for each forecast method.

probability. If the model-derived forecasts were calculated at smaller and smaller probabilities, approaching 0.01%, the curves would likely look very similar, demonstrating that the inclusion of the time-dependent observational frequencies largely does not change the underlying ability of the forecasts to discriminate tornado events from nonevents. Rather, it is the shift in where the 2% point falls on the curve that results in differing ROC areas. When the percentiles of the best reliability for each method were selected, the reliabilities were similar, as expected. All but the daylong method reliably forecasted RM tornadoes up to the 10% threshold (Fig. 4e).

The CSI of the normalized time-dependent probabilities is higher than the other methods at 5% and 10%, despite this method having a lower ROC area than the other forecasts (Fig. 5). At lower thresholds, such as the 2% threshold, its POD is much lower than the other methods, with a slight increase in SR comparable to the increase relative to the other methods in the SPC forecasts. At the 5% threshold, the SPC has the highest POD of any forecast, but also has a lower SR than the normalized time-dependent forecast probabilities. At the higher-impact 10% probability, all methods have similar or higher PODs than SPC forecasts, but the two time-dependent methods have less false alarm (higher SR) than the SPC forecasts.

The 10% forecast threshold also has the highest CSIs of any forecast threshold overall. At the 15% probability, the daylong probabilities have a similar false alarm area to the SPC forecasts while having a higher POD, while the time-dependent probabilities have a much smaller false alarm and a POD between the SPC forecasts and the daylong forecasts. The 30% threshold was only reached by the daylong probabilities and has a very low POD and high SR. Overall, the differences in the methods were slight, with larger differences in the POD and SR than the CSI, particularly at the higher forecast thresholds. In aggregate, the incorporation of UH timing does not appear to have much impact on overall skill; therefore, we examined case studies to illustrate the day-to-day impact of the different methodologies.

b. Case studies

To demonstrate the strengths and weaknesses of these techniques and their appearance to forecasters, three case studies are presented. The first is a late spring case where the guidance produced mixed modes, with 6 of 10 ensemble members showing a linear convective mode in the reflectivity. These cases are where we expect the best performance from the normalized, time-dependent probability. The second and third cases illustrate the behavior of the probabilities

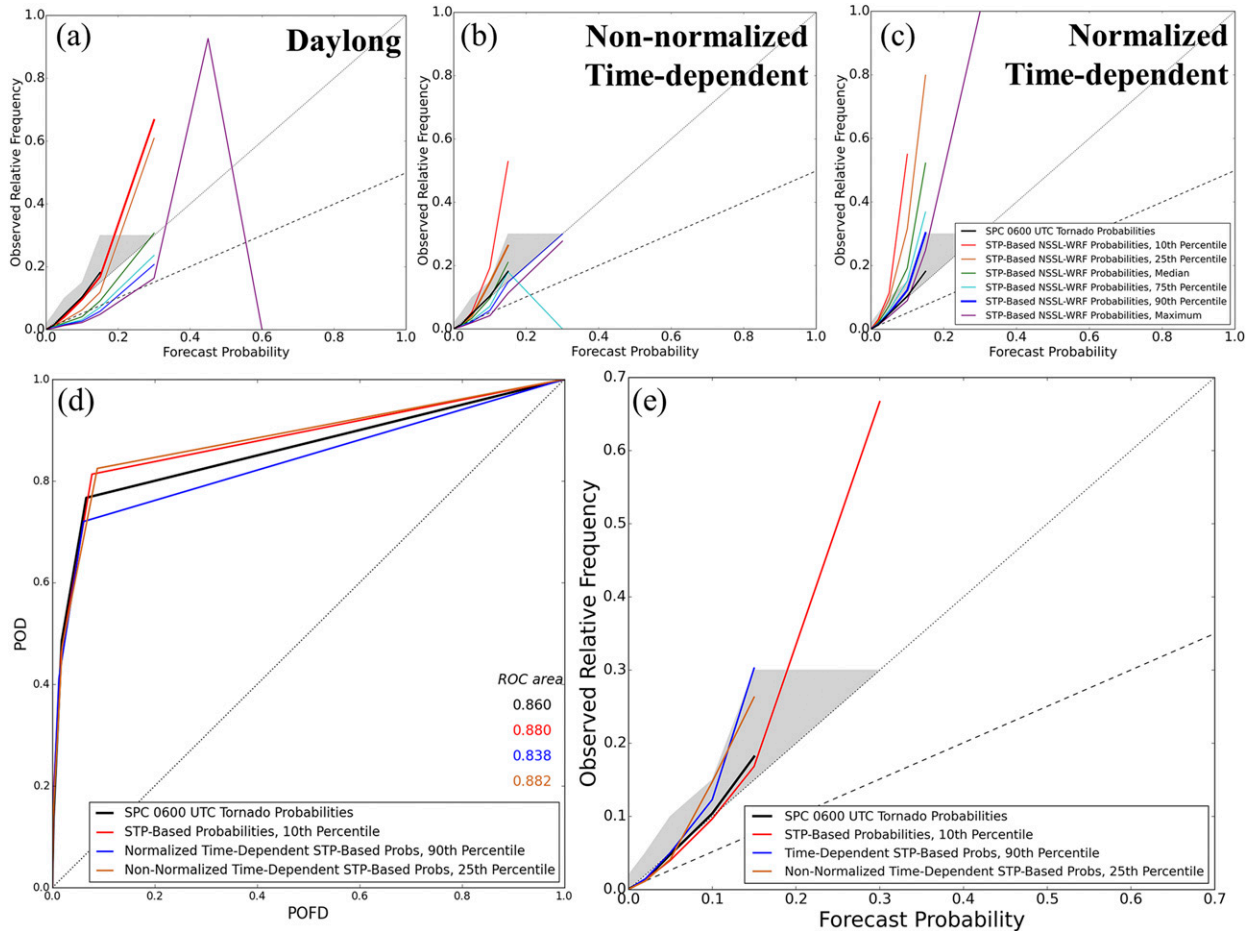


FIG. 4. Attributes diagrams for different percentiles of STP used to formulate (a) the daylong probabilities, (b) nonnormalized time-dependent probabilities, and (c) the normalized time-dependent probabilities. The diagonal represents perfect reliability, the dashed line is the no-skill line, and the shaded area shows where SPC forecasts can be considered reliable and credible. The most reliable forecasts are set in boldface. (d) ROC curve, with the diagonal representing a forecast with no skill and (e) attributes diagram for the SPC and selected percentiles of each probabilistic forecast generation method, with the shading and diagonal as in (a)–(c).

outside of the peak spring convective season, when the diurnal cycle of severe convection can be less pronounced and damping of the nocturnal probabilities could incorrectly suppress a signal for supercellular tornadoes.

1) 29 JUNE 2014

The first case study illustrates the forecast improvement provided by including the time of UH occurrence in the probability generation during a warm-season case, when the threat from nocturnal UH was associated with a linear convective mode in a majority of the ensemble members. On 29 June 2014 a surface low pressure center evolved across the south-central High Plains, with ample low-level moisture ahead of the main low. The 0600 UTC convective outlook from the SPC mentioned appreciable uncertainty in the storm coverage and timing, making

this a case where forecasters could use first-guess tornado guidance that reduces the false alarm from less favorable convective modes (i.e., MCSs). The SPC highlighted a 10% tornado threat across the Iowa–Missouri border, with a broad 5% area extending north through Wisconsin and west to the middle of Nebraska (Fig. 6a). A few initial supercells developed near a residual outflow boundary, but a complex pattern of storm evolution with multiple mergers ensued and a MCS developed around 0300 UTC. The tornado threat was primarily associated with the supercellular storms; 12 RM tornadoes occurred out of 14 total tornado reports. All ensemble-generated probabilities had the same magnitude as the 0600 UTC SPC forecasts: 10% (Figs. 6b–d). However, the placement and extent of the 10% probabilities differed. The daylong probabilities and the nonnormalized time-dependent probabilities both had a broad swath of probabilities

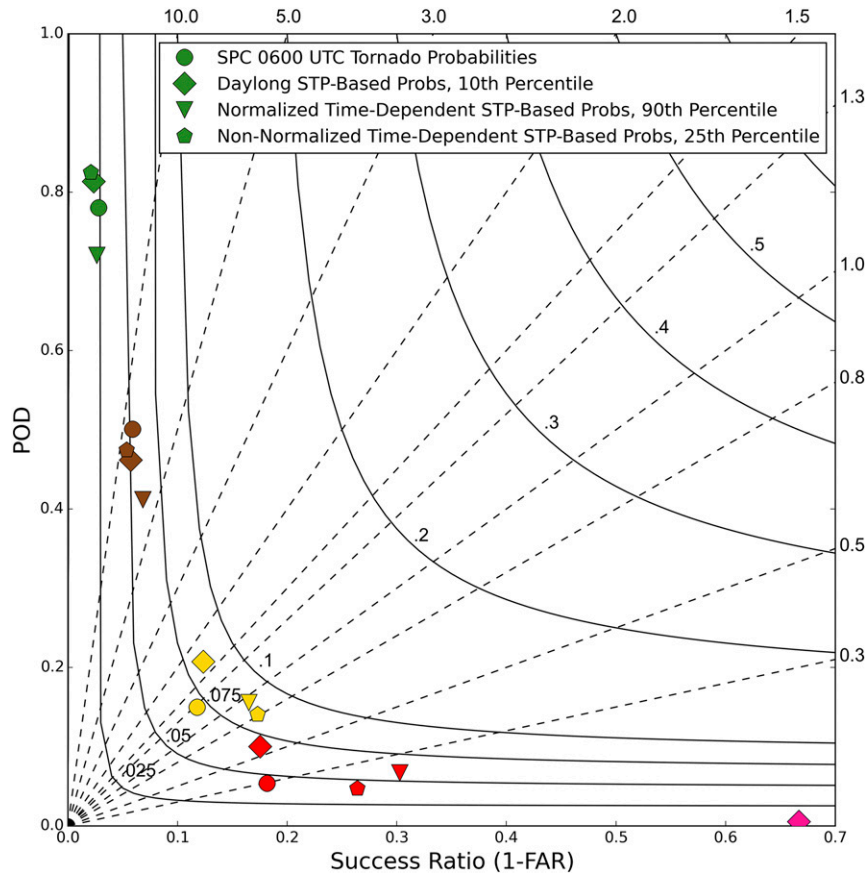


FIG. 5. Performance diagram for the three different methods of probability generation and the SPC. Green, brown, yellow, red, and pink shapes represent the 2%, 5%, 10%, 15%, and 30% forecast threshold, respectively. Dashed lines are of constant bias, and solid curved lines are lines of constant CSI.

extending into Illinois and a secondary area of probabilities across Kentucky, whereas the normalized time-dependent probabilities largely eliminated this area because the UH was occurring at 0300–0600 UTC (forecast hours 27–30; Fig. 6e). The UH in this case was also characteristic of MCSs, which tend to have broader UH that is less consistent than UH from supercells. Besides eliminating the false alarm area, the normalized time-dependent probabilities also maintained high probabilities where tornadoes occurred, resulting in a better forecast. This case's diurnal cycle of the probabilities showed that the nonnormalized time-dependent probabilities and the daylong probabilities maintained relatively high values through the night, while the nonnormalized time-dependent probabilities had much lower overnight average probabilities (Fig. 6f).

2) 16 NOVEMBER 2014

An obvious concern regarding the dampening of nocturnal UH is that the signal from UH associated with

nocturnal supercells will be incorrectly eliminated. This poses a particular problem in the southeastern United States where the annual and diurnal distributions of tornadoes exhibit much less of a peak compared to distributions in the Great Plains, which tend to have a sharp peak during the afternoon and early evening of the spring months (Krocak and Brooks 2018). The following case study demonstrates a failure mode for the time-dependent probabilities, when supercellular storms occurred overnight in the southeastern United States. Eleven tornadoes were reported from Louisiana through the Florida panhandle, with seven of those occurring between the hours of 0700 and 1200 UTC.

The SPC's 0600 UTC outlook explicitly mentioned the risk of nocturnal severe weather, including the possibility of quasi-discrete storms and tornadoes. The initial outlook was located west of the area of main tornado occurrence (Fig. 7a) and was focused around a maritime warm front advancing northward. The daylong (Fig. 7b) and nonnormalized time-dependent (Fig. 7c) tornado

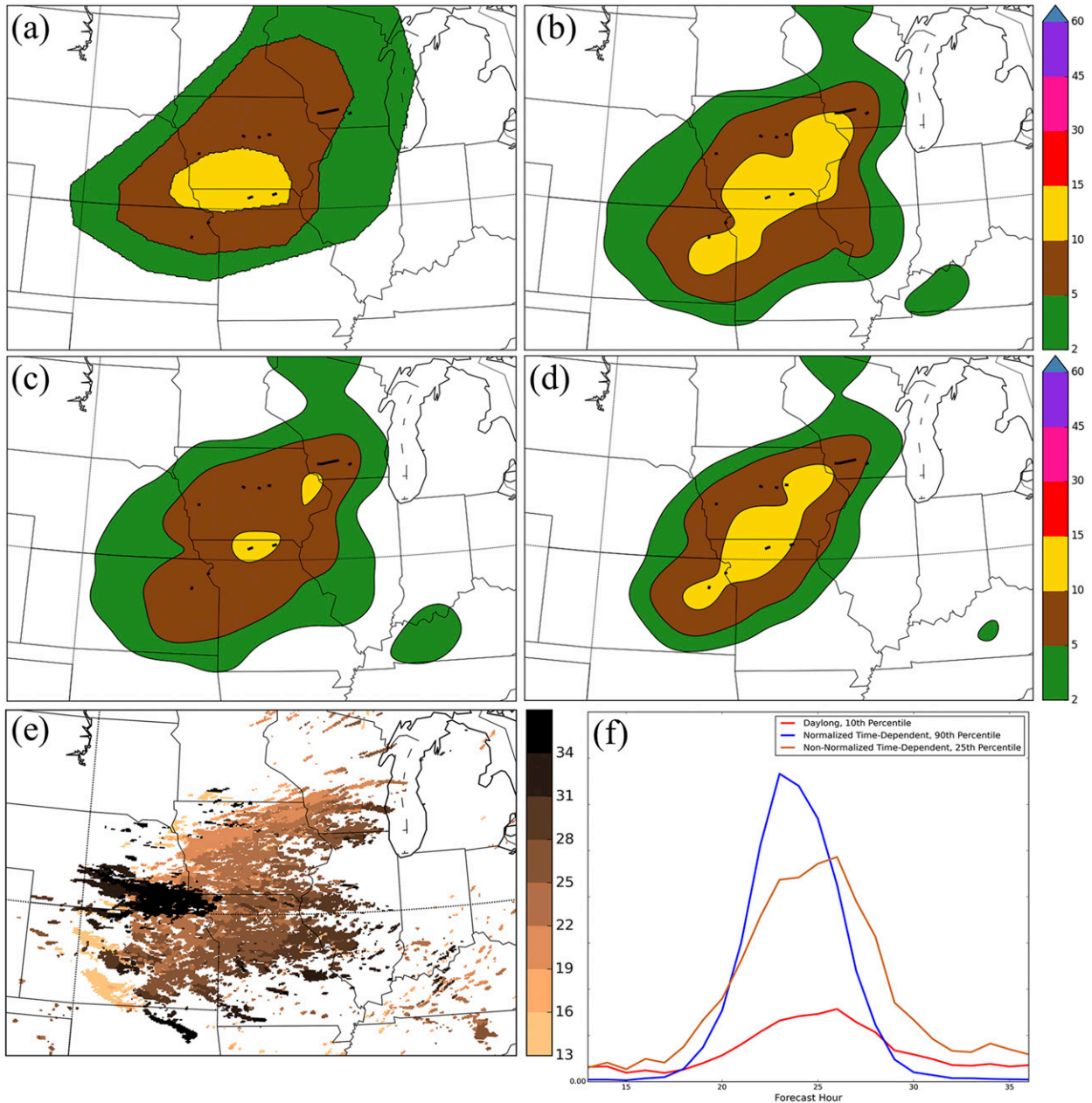


FIG. 6. Tornado forecasts for 29 Jun 2014 from (a) the SPC, (b) the daylong probabilities using the 10th percentile of the STP, (c) the nonnormalized time-dependent probabilities using the 25th percentile of the STP, and (d) the normalized time-dependent probabilities using the 90th percentile of the STP. Black lines show the tracks of RM tornadoes. (e) Ensemble 2–5-km UH $\geq 25 \text{ m}^2 \text{ s}^{-2}$, color coded by hour of UH occurrence. (f) The diurnal cycle of report frequency, UH frequency, and average probability for 29 Jun 2014.

probabilities extended slightly farther northeast than the SPC’s initial outlook, encompassing most of the long-track tornado that occurred in Alabama, but still excluded the easternmost tornadoes. Overall, the magnitude and extent of both sets of probabilities were very similar to the SPC’s forecast. The normalized time-dependent probabilities, however, decreased the probabilities across Alabama and Mississippi enough that

only the two tornadoes in Louisiana were captured by the forecast. Although discrete, long UH tracks occurred across Alabama and Mississippi, indicating rotating supercells (Fig. 7e), they were incorrectly dampened by the time-dependent probabilities. The diurnal histograms of the probabilities highlighted that the daylong and nonnormalized time-dependent probabilities both maintained a higher domain-averaged probability during

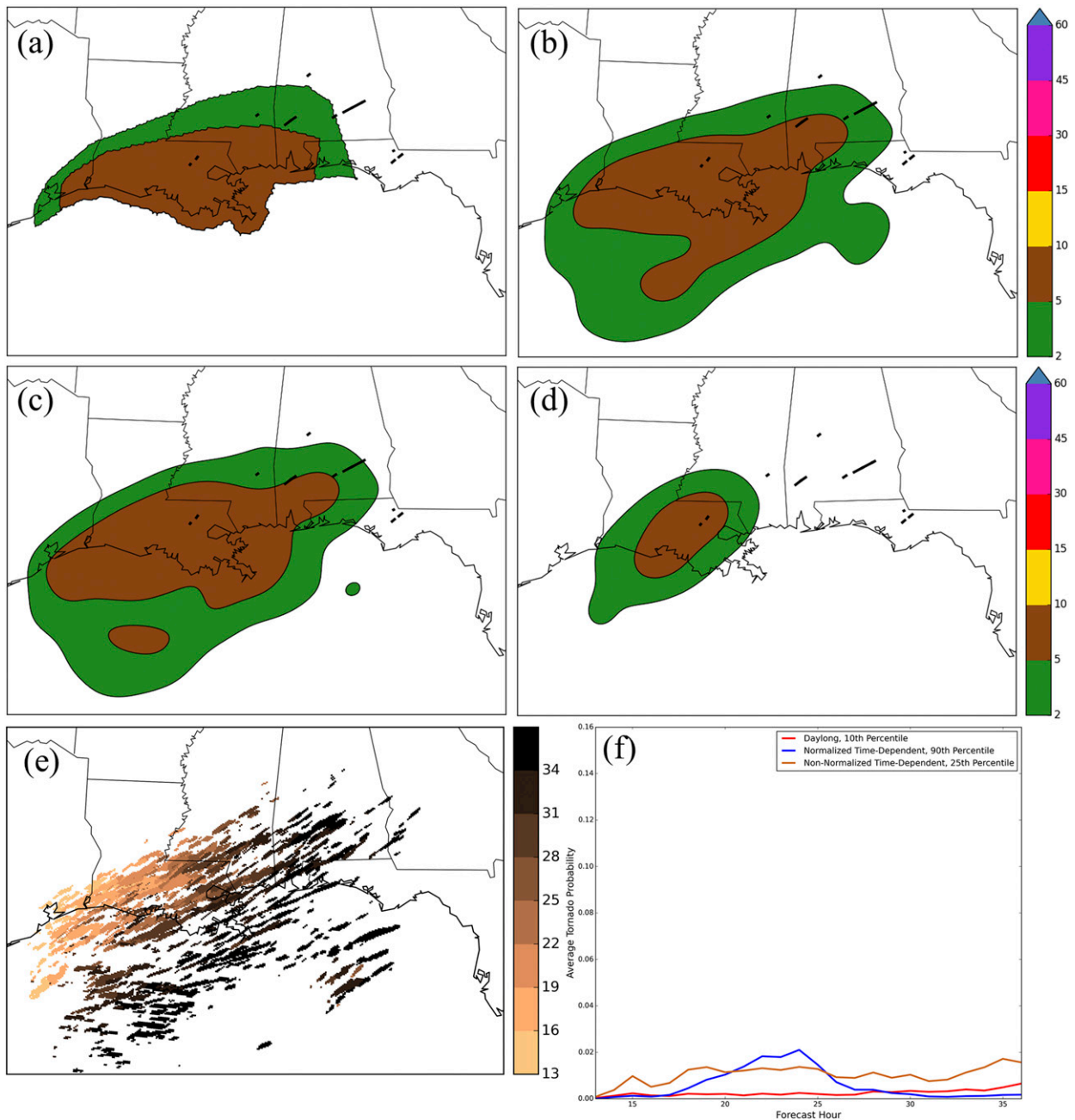


FIG. 7. Tornado forecasts for 16 Nov 2014 from (a) the SPC, (b) the daylong probabilities using the 10th percentile of the STP, (c) the nonnormalized time-dependent probabilities using the 25th percentile of the STP, and (d) the normalized time-dependent probabilities using the 90th percentile of the STP. Black lines show the tracks of RM tornadoes. (e) Ensemble 2–5-km UH $\geq 25 \text{ m}^2 \text{ s}^{-2}$, color coded by hour of UH occurrence. (f) The diurnal cycle of report frequency, UH frequency, and average probability for 16 Nov 2014.

the overnight hours (~forecast hours 29–35) than the time-dependent probabilities (Fig. 7f).

3) 16 NOVEMBER 2015

The final case study examines another November case, but one in which the normalized time-dependent probabilities worked as intended, dampening the tornado threat

from nocturnal UH that corresponded to a linear MCS in the ensemble. On 16 November 2015, 47 tornadoes were reported from southern Nebraska to Texas, with the greatest concentration of tornadoes occurring from western Kansas to the Texas panhandle. Discrete supercells initiated around 2000 UTC, with the first tornadoes occurring just before 2200 UTC. By 0500 UTC, the

discrete storms had grown upscale into a large squall line spanning from Nebraska south to Texas. This line maintained its intensity overnight through Oklahoma and Kansas, producing multiple high wind reports.

The SPC's 0600 day 1 outlook indicated the potential for tornadoes, but the initial forecast was focused too far south. A main forecast challenge on this day was the maintenance of a low-level capping inversion and weaker instability than what actually occurred along the dryline. Since the storms occurring along the dryline then moved into a cooler, yet still moist, environment with low-level shear, the tornado risk continued for several more hours than initially expected. The forecast discussion associated with the 0600 day 1 outlook discussed the potential for fast upscale growth of the initial storms, so the highest SPC-issued tornado probabilities were confined to an area where discrete storms were expected (Fig. 8a). All first-guess ensemble probabilities were also too far south but exhibited large differences in the east–west extent of the tornado forecasts, depending on the method of the probability generation. The daylong (Fig. 8b) and nonnormalized time-dependent (Fig. 8c) probabilities both had large areas of at least 15% probability, one category higher than the SPC's official forecast. However, the forecasts were focused on a broad area across southern Oklahoma and central Texas, east of where the initial storms were forecast. This area was associated with large UH swaths (Fig. 8e), but many of these broad swaths occurred overnight. The normalized time-dependent probabilities (Fig. 8d) eliminated this area, focusing instead on the initial storms. Only small 10% and 15% probabilities were generated, and the false alarm area was greatly decreased while capturing most of the tornadoes within the probabilities. The diurnal histograms for this case (Fig. 8f) showed the peak of the daylong and nonnormalized time-dependent probabilities overnight, while the nonnormalized time-dependent probabilities tapered off overnight without being eliminated entirely. Despite this case occurring during the cool season, using the diurnal cycle to dampen the nocturnal probabilities resulted in a forecast with much a much lower false alarm than forecasts that did not consider the timing of the UH occurrence or the diurnal cycle of the report occurrence.

4. Summary and discussion

Probabilities were developed that consider the time of UH occurrence within an ensemble and the climatological frequency of a tornado given the existence of a right-moving supercell. These probabilities address a shortcoming of prior first-guess forecasts, which often

had false alarms associated with nocturnal UH produced by unfavorable convective modes. While the relationship between the STP and frequency of right-moving supercells producing tornadoes varies somewhat by time of day, the nonnormalized time-dependent probabilities maintained a nocturnal false alarm signal. However, weighting the timing information by the overall number of reports during a given 3-h window lessened the nocturnal false alarm, because the most heavily weighted time occurs in the same window as the majority of reports: around 0000 UTC.

The most reliable percentile for each set of probabilities was compared, since the aim of this work was to reduce false alarms. The diurnal cycle of the normalized time-dependent probabilities more accurately reflected the diurnal report cycle than the other probabilities did, decreasing the nocturnal false alarm area compared to the UH occurrence. The normalized time-dependent probabilities had lower ROC areas than any other method, likely because the reduction in area covered by the probabilities decreased the POD at the 2% threshold. The sensitivity of the ROC area to the 2% threshold also likely plays a role in the difference between the ROC areas, and if the probabilities were calculated at lower and lower decision thresholds, the differences between the probabilities would likely be small. Additionally, since tornadoes are rare events, missed events greatly affect the statistical scores such as the POD. The CSI of the normalized time-dependent probabilities was less affected by missed events than the ROC area and performed similarly to the other methodologies. All methods performed well, particularly at high probabilistic thresholds, which often have larger potential impacts than the lower, more common thresholds. At these higher thresholds, the normalized time-dependent probabilities maintained PODs as high or higher than for other forecast methods, while also maintaining high SRs. Incorporating the timing of UH occurrence to reduce the false alarm area focused the forecast on areas at risk for supercellular tornadoes, remaining true to the underlying climatological frequencies while providing forecasters with a skillful and reliable first-guess tornado forecast. However, the differences in statistical scores over the April–July season were relatively small, suggesting that the addition of timing to the probabilities does not improve the overall skill by much when looked at in aggregate.

Three case studies demonstrated the advantages and limitations of the normalization methodology from a forecaster standpoint; the maps generated in the case studies show how the probabilities would appear day to day. Though the normalized time-dependent probabilities are most applicable to warm season cases such as

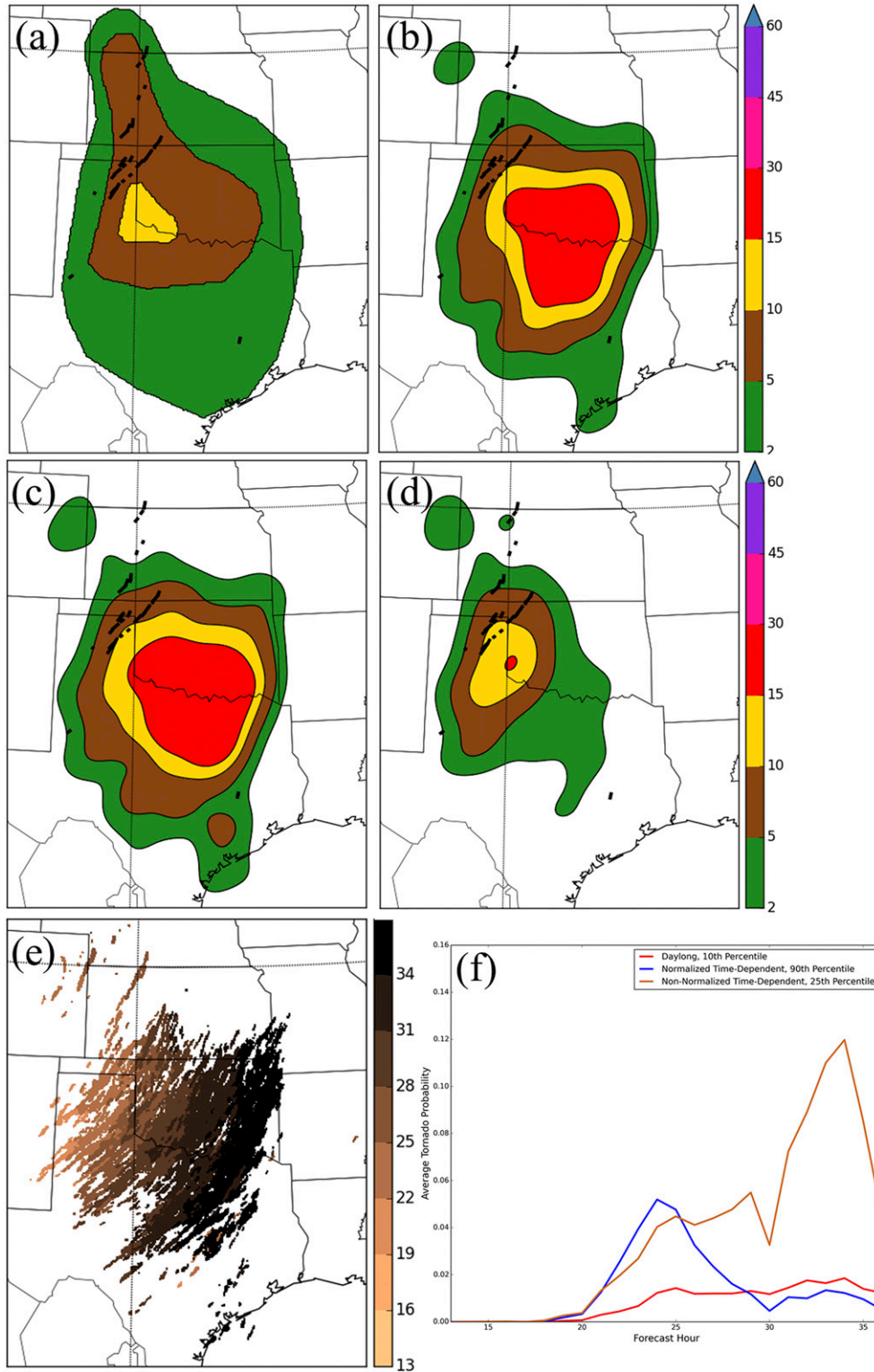


FIG. 8. Tornado forecasts for 16 November 2015 from (a) the SPC, (b) the daylong probabilities using the 10th percentile of the STP, (c) the nonnormalized time-dependent probabilities using the 25th percentile of the STP, and (d) the normalized time-dependent probabilities using the 90th percentile of the STP. Black lines show the tracks of RM tornadoes. (e) Ensemble 2–5-km UH $\geq 25 \text{ m}^2 \text{ s}^{-2}$, color coded by hour of UH occurrence. (f) The diurnal cycle of report frequency, UH frequency, and average probability for 16 Nov 2015.

29 June 2014, whenever nocturnal UH is associated with a linear convective mode this methodology will be useful. A cold season case that benefited from generating probabilities and using the normalization occurred on 16 November 2015, when the daylong probabilities and the nonnormalized time-dependent probabilities both had large areas of false alarm where no tornadoes occurred. The normalized probabilities are less appropriate when the nocturnal UH is actually associated with RM supercells; an example of this occurred on 16 November 2014. In this scenario, if forecasters know that discrete supercells may occur overnight, we would recommend the daylong and nonnormalized probabilities over the normalized probabilities. Conversely, using any of these methodologies when an MCS is present during the afternoon is likely to result in too many false alarms, as the normalization process will not dampen the probabilities. Making it clear through training and annotation that the probabilities are normalized based on the report distribution should indicate to forecasters that these probabilities are best used in the afternoon and during supercellular events.

Throughout the case studies and the aggregate statistics, the daylong and the nonnormalized time-dependent probabilities were quite similar, suggesting two main conclusions. First, the relationship between tornado occurrence and STP is quite consistent throughout the diurnal cycle—more favorable environments are more likely to be associated with a tornadic supercell, no matter the time of day. Based on that initial finding, our second conclusion is that the addition of timing information alone does not sufficiently increase the statistical scores to a level many users would find to be worthwhile.

Future iterations of the probabilities should explicitly take into account the convective mode, rather than using the time of UH occurrence as a coarse proxy for mode. While the ideal probabilities would incorporate the mode of the convection via simulated reflectivity, to do so is beyond the scope of this work. This work instead determines whether the time of UH occurrence can be used as a coarse proxy for the mode, and in which scenarios such a proxy can be used to reduce excessive false alarms from nocturnal UH occurrence.

Future work could also improve the probabilities by sensitivity testing involving the UH values. While we use a relatively low UH threshold to catch a majority of rotating updrafts, it is possible that a higher UH threshold would not have as many problems with false alarms. Additionally, we have noticed that UH associated with supercells tends to form long, coherent tracks of high UH values, while linear systems tend to have shorter, less coherent tracks with weaker UH values overall. A future convective mode filter could leverage these tendencies to eliminate UH that is unlikely to be associated with a right-moving supercell.

For forecasters, these probabilities summarize numerical output and calibrate the output based on observed climatological tornado frequencies given an STP value. In addition to the climatological tornado frequencies, the normalized probabilities also bring in the diurnal report distribution. By providing a summary of parameters output by CAM ensembles and basing the summary on observations, we hope that these probabilities can help forecasters protect life and property.

Acknowledgments. The authors thank Chris Melick and Robert Hepper of the SPC for providing regridded SPC forecasts and Andrew Dean of the SPC for obtaining the data for the climatological frequency calculations. We also thank three anonymous reviewers who contributed thoughtful and insightful suggestions during the review process. This material is based upon work supported by an NSF Graduate Research Fellowship under Grant DGE-1102691, Project A00-4125. BTG, 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.

REFERENCES

- 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, 2012: 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>.
- Elsner, J. B., and C. P. Schertmann, 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).
- Gagne, D. J., A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, <https://doi.org/10.1175/WAF-D-17-0010.1>.
- 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>.
- , A. J. Clark, B. T. Smith, R. L. Thompson, I. Jirak, and S. R. Dembek, 2018: Blended probabilistic tornado forecasts: Combining climatological frequencies with NSSL-WRF ensemble forecasts. *Wea. Forecasting*, **33**, 443–460, <https://doi.org/10.1175/WAF-D-17-0132.1>.
- Hitchens, N. M., and H. E. Brooks, 2017: Determining criteria for missed events to evaluate significant severe convective out-

- looks. *Wea. Forecasting*, **32**, 1321–1328, <https://doi.org/10.1175/WAF-D-16-0170.1>.
- Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806, <https://doi.org/10.1175/BAMS-84-12-1797>.
- , 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>.
- Krocak, M. J., and H. E. Brooks, 2018: Climatological estimates of hourly tornado probability for the United States. *Wea. Forecasting*, **33**, 59–69, <https://doi.org/10.1175/WAF-D-17-0123.1>.
- Loken, E. D., A. J. Clark, M. Xue, and F. Kong, 2017: Comparison of next-day probabilistic severe weather forecasts from coarse- and fine-resolution CAMs and a convection-allowing ensemble. *Wea. Forecasting*, **32**, 1403–1421, <https://doi.org/10.1175/WAF-D-16-0200.1>.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Naylor, J., M. S. Gilmore, R. L. Thompson, R. Edwards, and R. B. Wilhelmson, 2012: Comparison of objective supercell identification techniques using an idealized cloud model. *Mon. Wea. Rev.*, **140**, 2090–2102, <https://doi.org/10.1175/MWR-D-11-00209.1>.
- 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., <https://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).
- , 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>.
- , 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>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.