

JOHN L. CINTINEO,^a MICHAEL J. PAVOLONIS,^b JUSTIN M. SIEGLAFF,^a DANIEL T. LINDSEY,^c
LEE CRONCE,^a JORDAN GERTH,^a BENJAMIN RODENKIRCH,^a JASON BRUNNER,^a AND
CHAD GRAVELLE^{a,d}

^a Cooperative Institute of Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin

^b NOAA/NESDIS/Center for Satellite Applications and Research/Advanced Satellite Products Team, Madison, Wisconsin

^c NOAA/NESDIS/Center for Satellite Applications and Research/Regional and Mesoscale Meteorology Branch,
Fort Collins, Colorado

^d NWS Operations Proving Ground, Kansas City, Missouri

(Manuscript received 12 July 2017, in final form 5 January 2018)

ABSTRACT

The empirical Probability of Severe (ProbSevere) model, developed by the National Oceanic and Atmospheric Administration (NOAA) and the Cooperative Institute for Meteorological Satellite Studies (CIMSS), automatically extracts information related to thunderstorm development from several data sources to produce timely, short-term, statistical forecasts of thunderstorm intensity. More specifically, ProbSevere utilizes short-term numerical weather prediction guidance (NWP), geostationary satellite, ground-based radar, and ground-based lightning data to determine the probability that convective storm cells will produce severe weather up to 90 min in the future. ProbSevere guidance, which updates approximately every 2 min, is available to National Weather Service (NWS) Weather Forecast Offices with very short latency. This paper focuses on the integration of ground-based lightning detection data into ProbSevere. In addition, a thorough validation analysis is presented. The validation analysis demonstrates that ProbSevere has slightly less skill compared to NWS severe weather warnings, but can offer greater lead time to initial hazards. Feedback from NWS users has been highly favorable, with most forecasters responding that ProbSevere increases confidence and lead time in numerous warning situations.

1. Introduction

Issuing severe weather warnings is a critical function of the National Weather Service (NWS). Real-time meteorological datasets are becoming more advanced and sophisticated, with greater spatial resolution, frequency, and content. The combination of high-resolution numerical weather prediction (NWP) models [e.g., High Resolution Rapid Refresh (HRRR); Benjamin et al. (2011)], next-generation Geostationary Observational Environmental Satellites (e.g., *GOES-16*; Schmit et al. 2015), spaceborne lightning mappers [e.g., Geostationary Lightning Mapper (GLM); Goodman et al. (2013)], terrestrial lightning arrays [e.g., Earth Networks Total Lightning Network (ENTLN), Vaisala National Lightning Detection Network (NLDN)], Multi-Radar Multi-Sensor products (MRMS; Smith et al.

2016), and other datasets means that forecasters have routine access to very large volumes of data. For short-fuse operational products such as severe thunderstorm and tornado warnings, forecasters must quickly analyze relevant data, identify threats, and issue warnings to the public in a timely manner. Given the very large data volume applicable to severe weather warning operations, the manual analysis techniques typically employed in operations will not always extract all of the pertinent information, especially when numerous storms are present.

The NWS is exploring a new paradigm to its advisory/watch/warning products, whereby severe weather watches and warnings (as well as other hazards) may be disseminated in a grid-based, frequently updating, probabilistic manner. This paradigm is part of NOAA's Forecasting a Continuum of Environmental Threats (FACETs) effort (C. D. Karstens et al. 2018, unpublished manuscript; Rothfus et al. 2014). Because the probabilistic FACETs

Corresponding author: John L. Cintineo, jlc248@gmail.com

DOI: 10.1175/WAF-D-17-0099.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

paradigm is starkly different from the current binary yes/no warning paradigm, this creates the need for probabilistic guidance to communicate a level of certainty of potential threats.

In response to the “big data” challenge, researchers at the NOAA/National Environmental Satellite, Data, and Information Service (NESDIS) and the Cooperative Institute of Meteorological Satellite Studies at the University of Wisconsin–Madison (UW-CIMSS) have developed the NOAA/CIMSS Probability of Severe (ProbSevere) model [Cintineo et al. (2013) and Cintineo et al. (2014), hereafter C13 and C14, respectively]. ProbSevere is a naïve Bayesian classifier that utilizes multiple meteorological datasets to compute the probability that any developing or present thunderstorm will produce severe weather in the near future (0–90 min), anywhere in the continental United States (CONUS). Using the NWS definition, severe weather is characterized by at least one of the following conditions: one or more hailstones with a diameter of at least 1 in., a convective wind gust measuring at least 50 kt (where 1 kt = 0.51 m s^{-1}) or producing significant damage to structures or trees, or the presence of a tornado (NOAA/NWS 2017). The goals of this paper are to describe the integration of ground-based lightning measurements into ProbSevere, present a rigorous validation analysis, and highlight the efforts to assess the performance of ProbSevere, with lightning, in operational settings.

2. ProbSevere – Introduction

a. ProbSevere overview

Data mining meteorological observations and model output within an object-based framework is not a new concept (e.g., Lakshmanan and Smith 2009). For instance, one can find recent data-driven work in the atmospheric sciences for severe wind prediction (Lagerquist et al. 2017), flash flooding (Gourley et al. 2017), offshore precipitation estimation (Veillette et al. 2016), hail prediction with storm-scale NWP models (Gagne et al. 2015), and severe weather climatological metrics (Smith et al. 2017), all of which use some elements of data mining, machine learning, and image processing while integrating multiple sources of observations and NWP model output. ProbSevere tackles the joint prediction of severe hail, wind, and tornadoes, uniquely integrating derived satellite observations with lightning, radar, and NWP output.

ProbSevere uses the naïve Bayesian classifier to determine the probability that a given thunderstorm is a member of the “severe” class (i.e., a member of the class of storms that will produce severe weather in the short term).

C14 gave an overview of the naïve Bayesian classifier and constituent model predictors from satellite, radar, and NWP sources. One unique aspect of ProbSevere is its multisensor storm identification and tracking capability. ProbSevere identifies and tracks clouds on a derived satellite field (C14; see Sieglaff et al. 2013 for tracking details) and on a derived radar field [C14; see Lakshmanan et al. (2007a,b) for details on the Warning Decision Support System–Integrated Information (WDSS-II)]. ProbSevere is designed to take advantage of geostationary satellite metrics that often effectively depict the evolution of a cumulus cloud into a cumulonimbus cloud as well as radar metrics that are most relevant after the cumulonimbus stage is reached. The combined use of satellite and radar metrics, which are time lagged, is achieved through the association of satellite cloud objects with radar objects, after the satellite data are corrected for parallax effects.¹ In this way, ProbSevere attempts to represent a more complete picture of a thunderstorm’s evolution, which better informs its future severe potential. Predictors in the classifier must have severe and nonsevere distributions with sufficiently different means and/or standard deviations in order to add skill. Please see C13 and C14 for an analysis of the separation of classes for the following predictors.

b. Predictors

ProbSevere incorporates five predictors (Table 1); one of which is derived from NOAA’s operational Rapid Refresh NWP model (RAP). The RAP predictor combines the most-unstable convective available potential energy (MUCAPE) and the effective bulk shear (EBS; Thompson et al. 2007) into a two-dimensional (2D) predictor. Two of the predictors are derived from Geostationary Operational Environmental Satellite (GOES) data. From GOES, the rate of change in the 11- μm top-of-the-troposphere emissivity $\Delta\epsilon_{\text{tot}}$ (C13, C14, Pavolonis 2010a) and the cloud-top glaciation rate ΔIce (C13, C14, Pavolonis 2010b) are utilized. In addition, the maximum expected size of hail (MESH; Witt et al. 1998) from the MRMS product suite is a predictor. Finally, the total lightning flash rate, the integration of which is detailed in section 3, is utilized.

The $\Delta\epsilon_{\text{tot}}$ is often referred to as the normalized vertical growth rate and is analogous to decreases in the minimum observed 11- μm brightness temperature in the cloud object. The 11- μm top-of-the-troposphere emissivity is the emissivity a cloud would have if it were at the tropopause (Pavolonis 2010a). Using the emissivity instead of brightness temperature helps

¹Parallax correction is performed using an assumption of a constant cloud height of 9 km.

TABLE 1. ProbSevere model predictors for 2014–16. This study reprocessed 2014 days with the 2016 model.

Year	Predictor name	Source
2014–15	MUCAPE vs EBS	RAP
	Max expected size of hail (MESH)	MRMS
	Max rate of change in the ϵ_{tot} in 2.5-h window ($\Delta\epsilon_{\text{tot}}$)	GOES
	Max rate of change in the cloud-top ice fraction in 2.5-h window (Δice)	GOES
2016	MUCAPE vs EBS	RAP
	MESH	MRMS
	$\Delta\epsilon_{\text{tot}}$	GOES
	Δice	GOES
	Total lightning flash rate vs EBS	ENTLN and RAP

reduce the impact of surface features being misconstrued as clouds and is less sensitive to the depth of the troposphere compared to the brightness temperature. The Δice uses the cloud-top phase field to capture how quickly cloud tops transition from liquid water to ice. Severe storms tend to exhibit stronger image-to-image increases in these satellite fields than nonsevere storms (C13). Satellite growth observable from geostationary imagery has long aided in thunderstorm intensity diagnoses (e.g., Adler and Fenn 1979, 1981; Reynolds 1980; Adler et al. 1985; Roberts and Rutledge 2003; Mecikalski and Bedka 2006; Sieglaff et al. 2011). ProbSevere uses the *maximum* image-to-image satellite growth rate in the latest 2.5-h window, that is, the greatest trends computed from either GOES-East or GOES-West observations.

The environmental shear and instability are generally good first-order fields for determining the severe potential in a given region. ProbSevere uses MUCAPE to better depict elevated hail and wind threats, compared to surface-based or mixed-layer CAPE. The EBS is used to better discern environmental shear for storms ranging from shallow to very tall, compared to 0–6-km bulk shear (Thompson et al. 2007). C14 explains how the a priori probability in the naïve Bayesian equation, or “first guess” probability, is derived as a function of these two NWP predictors, which are also spatially and temporally smoothed to mitigate phase and placement errors in the RAP.

The MRMS MESH is very useful for identifying storms capable of producing hail (e.g., Cintineo et al. 2012). It is empirically derived from a thermally weighted integration of the radar reflectivity from the melting layer to the storm top. MESH can also be considered a good proxy for updraft strength in a storm, which is related to storm severity. Thus, the MESH is a good predictor for damaging wind threats in certain situations, wherein greater MESH may be indicative of precipitation loading.

Note that other reflectivity-derived fields are not used as predictors. The “naïve” part of the Bayesian classifier is due to the fact that it assumes independence of predictors,

although it has been shown that a naïve Bayesian approach can still be skillful even when the independence assumption is clearly violated (e.g., Domingos and Pazzani 1997; Hand and Yu 2001; Kossin and Sitkowski 2009; Heidinger et al. 2012; Pavolonis et al. 2015). The MESH and other MRMS fields (e.g., vertically integrated liquid, reflectivity at -20°C , height of the 50-dBZ echo above 0°C) are highly correlated. Inclusion of several such predictors may create a probabilistic product that is poorly calibrated. While the ProbSevere predictors are not completely independent, they are derived from several independent measurement platforms (weather radar, satellite imagers, and lightning networks).

The total lightning flash rate (FR) in a storm is simply the sum of intra/intercloud flashes (IC) and the number of cloud-to-ground flashes (CG) in a storm per minute, with units of flashes per minute. The FR is directly related to ice production aloft and charge separation in the storm and can help infer updraft intensity (e.g., Deierling and Petersen 2008; Steiger et al. 2007). Flash data from ENTLN are proprietary and provided through a contract between NOAA and Earth Networks. Flash data are delivered in a comma-separated values format, with information provided for each flash including time, location, amplitude, polarity, height, and type of flash (IC or CG). The ProbSevere system currently uses a WDSS-II algorithm (w2ltg) to take the flash data and create a lightning density field with 2-min temporal resolution and $0.01^{\circ} \times 0.01^{\circ}$ spatial resolution. Flashes are smoothed with a radius of influence of 3 km. The total lightning flash density field has units of flashes per minute per square kilometer. When values from the lightning density field are aggregated inside a storm object, the FR is rounded to the nearest whole flash. Section 3 describes how the FR is incorporated into ProbSevere.

The ProbSevere model computes and updates probabilities for all CONUS storms at the MRMS frequency of 2 min, using the most recent MRMS, ENTLN-derived, RAP-derived, and GOES-derived data available. Model output can be displayed in the Advanced Weather Interactive Processing System II (AWIPS-II), which the NWS uses to view meteorological data and issue products such as severe weather warnings. Figure 1 shows an example of ProbSevere output in AWIPS-II, with polygons contoured around storms in the shaded MRMS MergedReflectivity field, colored by the final (i.e., posterior) probability of severe. With sampling enabled, forecasters can hover their mouse cursor over an object and get a text display of the probability of severe weather along with the constituent predictors. The text display helps users interpret changes in probability over time.

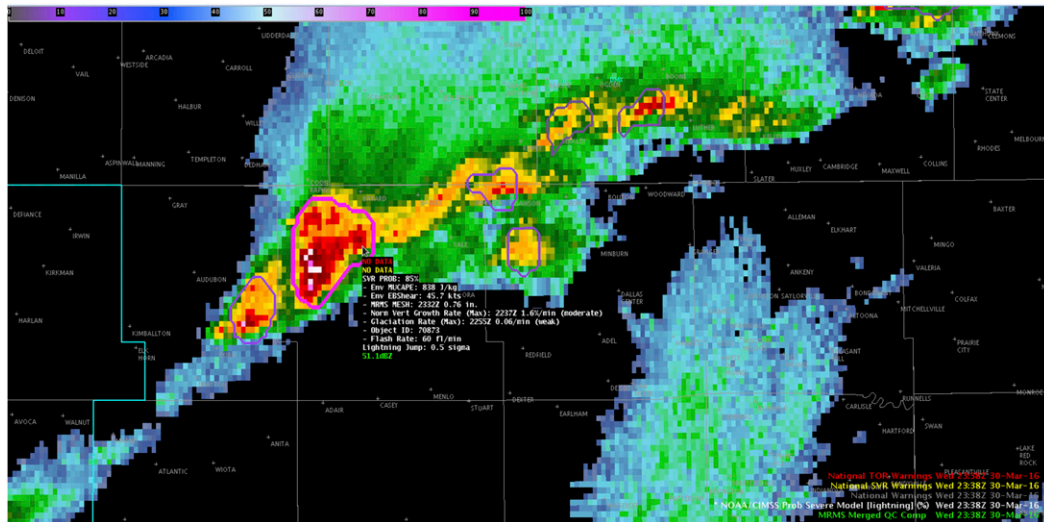


FIG. 1. NOAA/CIMSS ProbSevere model output visualized in AWIPS-II (image time is 2338 UTC; ProbSevere = 85%). Polygons represent storms identified and tracked by ProbSevere, colored by the computed probability of severe weather in the next 90 min. Pink shades denote probabilities in the 75%–100% range, whereas gray-to-purple shades denote probabilities in the 0%–15% range. When sampling is enabled in AWIPS-II, forecasters can scroll over polygons with their mouse cursors to see readouts of predictor values and the probability of severe weather. MRMS Merged Reflectivity is shaded from light blue (~15–20 dBZ) to green (~25–35 dBZ) to yellow and orange (~35–45 dBZ) to red (~50–55 dBZ) and then to white and magenta (60+ dBZ). This storm west of Des Moines, IA, on the evening of 30 March 2016 was cause for a severe thunderstorm warning by the NWS 20 min after this image time (2358 UTC). Five minutes after the warning was issued, multiple reports of 1-in.-diameter hail were recorded (at 0003 UTC).

3.

Prior to 2016, ProbSevere used the radar, satellite, and NWP predictors described previously (Table 1). The availability of ENTLN data to NOAA made it possible to test the potential benefit of using total lightning data in ProbSevere and gain insight into the potential benefits of spaceborne lightning measurements available in the GOES-R satellite series era (Goodman et al. 2013). When testing the FR as a univariate predictor in ProbSevere, it was found that it contributed to a large increase in the probability of detection and false alarm ratio at greater forecast probability thresholds (e.g., 80%–90%). The flash rate was therefore coupled with the EBS, which is closely tied to storm organization. The coupling of FR and EBS is somewhat analogous to using the MUCAPE/EBS 2D predictor for the first-guess probability (see C14), except that the flash rate is an *observed* quantity that is a realization of convective potential instability, which only yields a theoretical maximum updraft velocity.

With respect to total lightning incorporation, the severe and nonsevere storm classes for all of the predictors were extracted from 88 days in May, June, July, and August of 2015 (approximately the first 20–25 days of each month). Severe hail, wind, and tornado preliminary local storm reports (LSRs) from NOAA's Storm

Prediction Center (SPC 2016) were used to determine which storms were severe in an automatic fashion. Each LSR was assigned to the storm with the spatially closest centroid within 2 min of the report time. Nonsevere storms were defined as radar-identified objects that exhibited at least 40 dBZ at the -10°C isotherm at some point in their life cycle [convective initiation was defined in Kain et al. (2013) using this criterion], were tracked for at least 30 min, and were never associated with a severe LSR.

The joint distributions of FR/EBS for nonsevere and severe storms are shown in Figs. 2 and 3, respectively. These distributions were smoothed with 2D kernel density estimation, using optimally chosen bandwidths [i.e., bandwidths that smooth the sample distribution such that the mean integrated squared error is minimized; see Mielniczuk (1997)]. The shaded values and white contours denote the frequency of storms as a fraction of the nonsevere or severe storm population. Thus, the integration over each grid is 1. Note how the distribution for the severe class is more disperse and contains a more pronounced tail of data in the flash rate dimension. The gridded data in Fig. 4 show the ratio of the severe and nonsevere distributions. Ratios greater than 1 (see the white contours) indicate that the FR/EBS predictor favors the severe class (i.e., the larger the ratio,

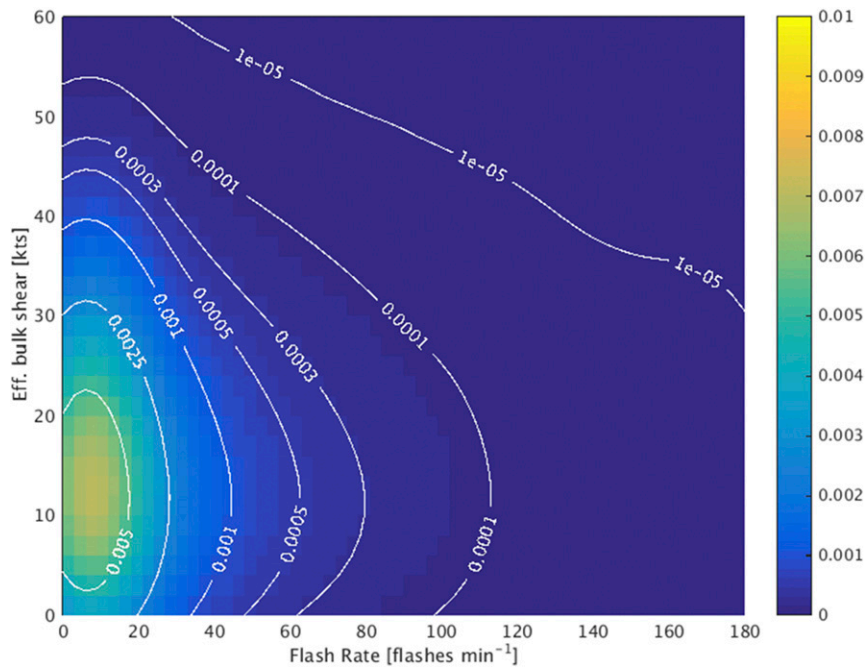


FIG. 2. The joint distribution, or frequency density, of total lightning FR and EBS for the nonsevere storms class. Shaded values and white contours denote the frequency of storms as a fraction of all nonsevere storms.

the more the severe class is favored). For low-EBS storms (e.g., 10 kt), a greater flash rate is required (~40 flashes per minute) to add to the *posterior* probability,

whereas for storms residing in relatively high-EBS environments (e.g., 40 kt), a relatively lesser flash rate (~10 flashes per minute) is necessary to increase the

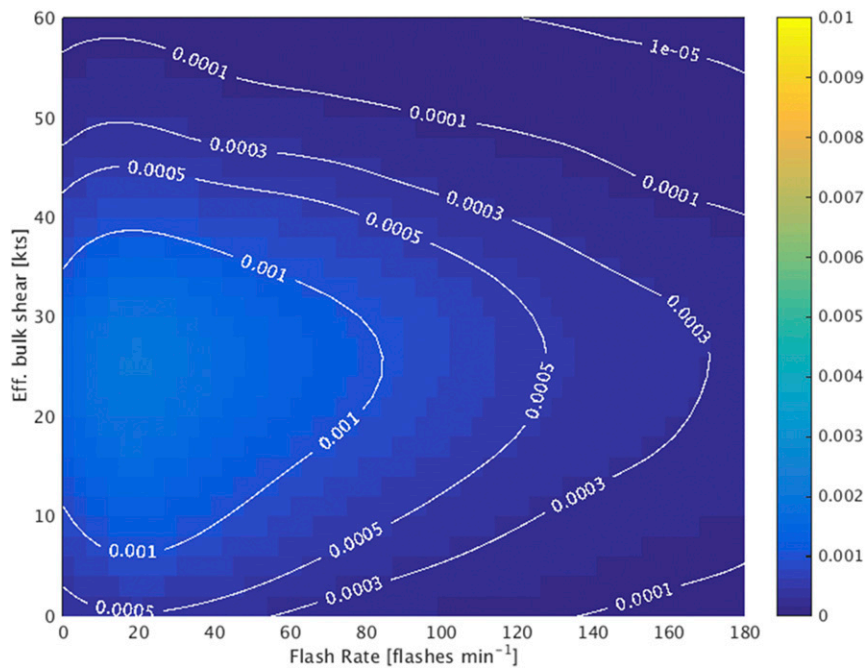


FIG. 3. As in Fig. 2, but for the severe storm class.

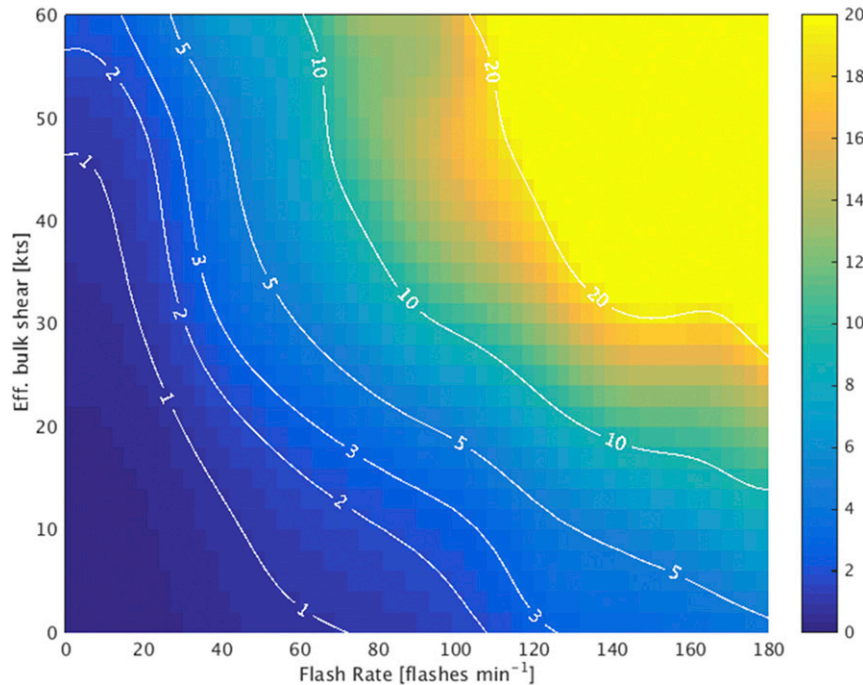


FIG. 4. The joint distribution of the total lightning FR and EBS for the severe class divided by the same joint distribution for the nonsevere class. Regions in the data space greater than 20 were assigned a value of 20 because of sparse sampling.

final probability. Both examples (10 kt, 40 flashes per minute; 40 kt, 10 flashes per minute) yield a ratio of approximately 1, which does not modify the a priori probability. The joint distributions of severe and nonsevere FR/EBS express a spatial character similar to the MUCAPE/EBS joint predictor for the a priori in ProbSevere (see Fig. 2 in C14) whereby moderate quantities in each predictor make severe storms more probable than a large value in one and a small value in the other.

Several aspects of total lightning interpretation in ProbSevere require additional consideration. First, supercell thunderstorms exhibit different FR patterns than multicells, single cells, or squall lines (e.g., Miller et al. 2015), suggesting that total lightning utility is tied to storm morphology. The FR/EBS distributions presented in this section do not discriminate by storm mode. Parsing the FR/EBS predictor (and other predictors) by a storm's general morphology can add value to users' interpretation of changes in FR in ProbSevere and has the potential to improve the model quantitatively.

The utility of total lightning FR is also strongly dependent on the accuracy of the storm object identification and tracking methodology. Abrupt changes in the definition of an object boundary may artificially produce a large increase or decrease in the FR, which will affect the final probability. For example, many quasi-linear convective

system (QLCS) storms are prolific lightning producers and sometimes suffer from abrupt object area changes as a result of segments of the QLCS merging or splitting, or simply because local maxima of composite reflectivity (which the identification algorithm relies on to create objects) can be difficult to discern from scan to scan. In this case, a large increase in area will correspond to a large increase in FR, which is unrelated to the storm's meteorology, yet still increases the final ProbSevere value. Thus, it is important that users visually inspect storms' evolution for abrupt object changes and for aid in storm mode diagnosis, which will allow for anticipation of possible inaccurate model forecasts. Model validation based on storm morphology and the use of spatial metrics to mitigate nonmeteorological changes in the ProbSevere values are both planned research activities.

4.

a. Methodology

1) PROBABILISTIC VALIDATION

The skill, lead time, and reliability (i.e., forecast probability calibration) of ProbSevere were assessed using an independent dataset of 119 days from 2014 and 2016 (Table 2), comprising nearly 3200 severe storms

TABLE 2. A list of the days for the validation of the ProbSevere model, independent from the days used to create the training dataset.

Year	Month	Day
2014	May	7, 8, 15, 17, 18, 19, 20, 21
	Jun	4, 7, 17
	Jul	17, 18, 21, 22, 25, 26, 27
2016	Mar	14, 15, 16
	Apr	6, 10, 11, 12, 13, 14, 15, 16
	May	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30
	Jun	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24
	Jul	1, 2, 3, 4, 5, 6
	Aug	1, 18, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31
	Sep	1, 2, 3, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30
	Oct	1, 2, 3, 4, 5, 6
	Nov	2, 4, 5, 18, 28, 29, 30
	Dec	17, 25, 28

and approximately 61 500 nonsevere storms. The skill of ProbSevere was measured with traditional metrics such as the probability of detection (POD; i.e., the ratio of correctly forecast events to all events), false alarm ratio (FAR; i.e., the ratio of false alarms to the sum of false alarms and hits), and critical success index (CSI; i.e., the ratio of hits to the sum of hits, misses, and false alarms) as a function of the posterior probability. CSI is a notably useful metric for rare-event forecasting (Wilks 2006). In the validation analysis, an event was defined as a thunderstorm (i.e., an identified object with a flash rate ≥ 2 flashes per minute at any point in its lifetime) meeting a certain probability threshold. In order for a storm to be severe, it must have been associated with at least one LSR of severe hail, severe wind, or a tornado at some point in its lifetime. An event was correctly forecast, or considered a “hit,” if the given probability threshold was first achieved for the storm at or prior to the LSR time. An event was a “miss” if the given probability threshold was first achieved after the LSR time, or if the probability threshold was never met. A “false alarm” event occurred when a probability threshold was met, but the storm was not associated with a severe LSR within the next 90 min. Thus, a storm can at most have either one hit or miss or false alarm at any given probability threshold. For example, a storm with a maximum ProbSevere value of 70% prior to the first LSR would count as hits at every probability threshold $\leq 70\%$ and misses at thresholds $> 70\%$. There would be no false alarms in this case. If the same storm failed to produce an LSR, it would then have false alarms at every probability threshold $\leq 70\%$, with no misses or hits.

2) NWS VERSUS PROBSEVERE VALIDATION

In addition to probabilistic validation of ProbSevere, a comparison between ProbSevere and NWS severe weather

warnings was performed. For this study, NWS verification practices were modified to enable a storm-by-storm validation. Therefore, NWS skill presented herein is unofficial. Manual analysis was used to identify coherent storm objects in MRMS composite reflectivity imagery over time, linking ProbSevere object IDs to individual storms. Manual analysis was chosen over an automated method to mitigate potential biases in automation and to more accurately link ProbSevere IDs together. For instance, a given storm could have multiple ProbSevere IDs if there were automated tracking anomalies, including splitting or merging storm cells and tracking algorithm errors in assigning the correct object ID. The manual analysis also associated storms with an *initial* LSR time and an *initial* NWS warning issuance time, if either occurred. NWS warnings without ProbSevere objects present were also recorded and scored.

In this way, we can directly compare ProbSevere model skill and lead time to the NWS for the *initial* LSR for severe storms. An example of this process is illustrated in Fig. 5. If an NWS severe thunderstorm or tornado warning was present at or before the first LSR for a storm, it was scored as one hit for the NWS, regardless of whether there were multiple warnings issued for the storm in its life, or if it produced multiple LSRs, or if the LSRs resided in the polygon (provided that they were easily attributable to the storm). If there were one or more LSRs produced by a storm that was never warned, or it was first warned after the initial LSR, it was scored as one miss only. If there were one or more warnings issued by the NWS for a storm that never produced LSRs, it was scored as one false alarm only. Thus, a warned storm can result in one hit or one false alarm, while an initial LSR can either validate a warned storm or cause the storm to be a miss. The storm in Fig. 5 was counted as one hit for ProbSevere at the 90% threshold with a lead time of 50 min to the initial LSR (the hail report). For the NWS, the storm would also be scored

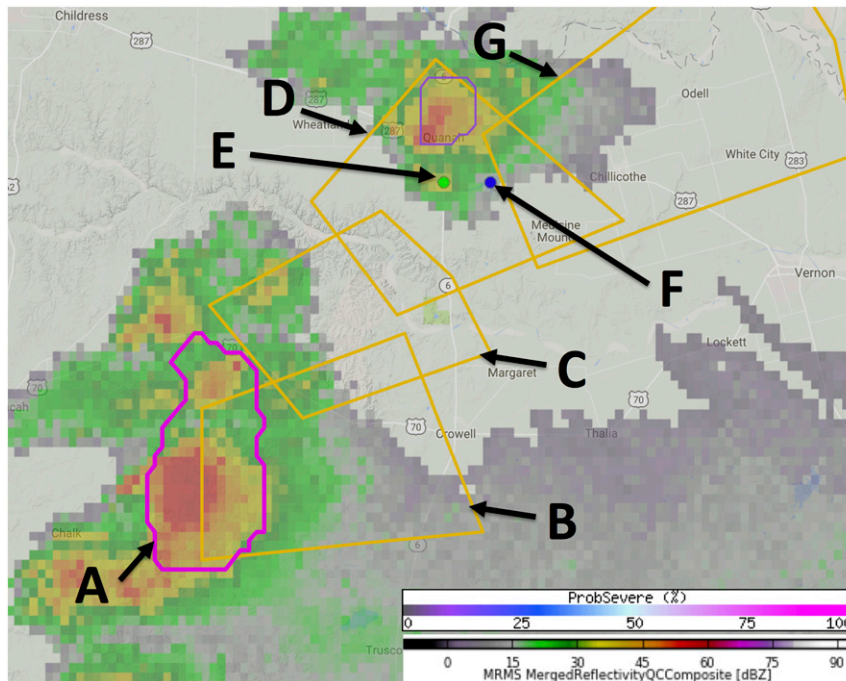


FIG. 5. An example of how storm-by-storm validation is performed in this study. The timeline for this example shows the following: A, ProbSevere object = 94% valid at 0012 UTC; B, NWS warning valid 0012–0100 UTC; C, NWS warning valid 0035–0100 UTC; D, NWS warning valid 0054–0145 UTC; E, 2.5-in.-hail report at 0102 UTC; F, 60 mi h^{-1} wind report at 0115 UTC; and G, NWS warning valid 0120–0200 UTC. The hail report verifies the ProbSevere object at the 90% threshold and verifies the set of four NWS warnings for this storm. The wind report is not considered, since it occurred after the hail report. This methodology yields one hit only for the NWS (E). This differs from official NWS verification methods, which would yield one verified warning (D), two warned events (E and F), and three false alarm warnings (B, C, and G).

as one hit with a lead time of 50 min to the initial LSR as well, based on the first warning issuance time at 0012 UTC and the first LSR time of 0102 UTC. The three subsequent NWS warnings and wind report are null points, since all occurred after the initial warning or LSR. Using official NWS scoring practices, the third NWS warning would be considered a “verified warning,” whereas the other three warnings would be false alarms. Both reports would be considered “warned events,” with a lead time of 8 min to the hail report and 21 min to the wind report. The modifications made to traditional verification practices help mitigate the effects of ProbSevere having a longer valid time and no explicit warning area for its probabilistic guidance compared to NWS warnings, creating a better “apples to apples” comparison.

Nonsevere storms were automatically extracted by identifying storms that were tracked by ProbSevere for at least 30 min and were clearly not associated with any LSRs during their history. The 30-min criterion was subjectively chosen because of some limitations in the automated tracking. For instance, one coherent storm

may, for a number of reasons, change its object ID on several occasions over its lifetime (e.g., mergers, splits). If every object ID from that storm were counted as a false alarm, the FAR could be artificially high. Thus, only more sustained storms were counted as nonsevere, helping to mitigate (but not eliminate) this problem. The ratio of severe to nonsevere storms is approximately 5% in this study, which is comparable to the U.S. climatology of severe thunderstorm frequency [$\sim 5\%$ – 10% ; NOAA/NWS (2010); FEMA (2007)].

3) LOCAL STORM REPORTS

Because LSRs serve to validate both ProbSevere and NWS warnings in this study, some limitations must be acknowledged, which are listed below with references:

- a secular increase in reporting over time (Doswell et al. 2005; Weiss et al. 2002);
- a dependence of report frequency on population density and time of day (Trapp et al. 2006; Doswell et al. 2005; Hales and Kelly 1985; Kelly et al. 1985);

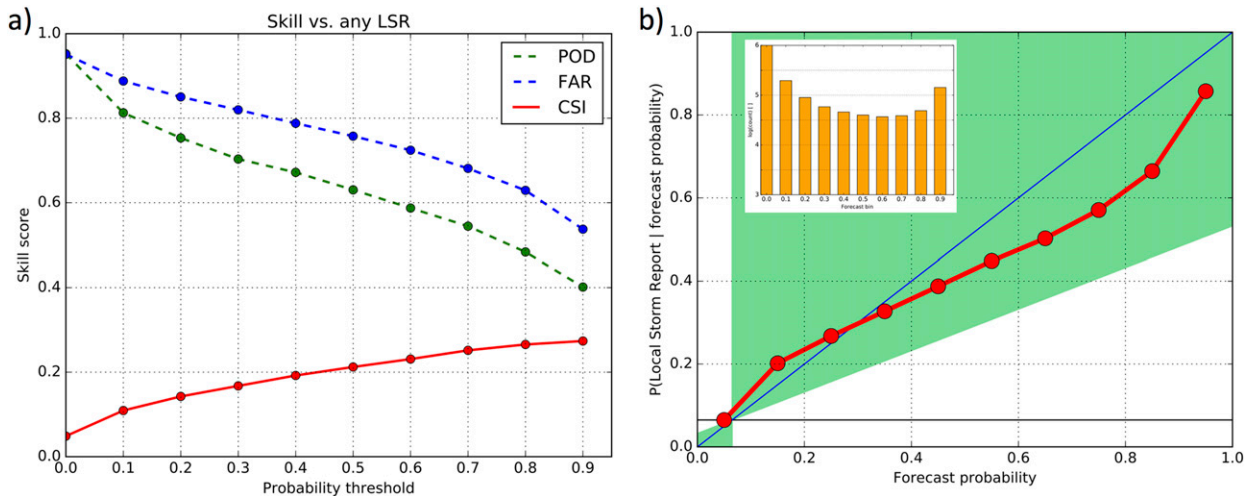


FIG. 6. (left) ProbSevere skill scores for the entire validation as a function of forecast probability threshold on a storm-by-storm basis and (right) a reliability diagram of ProbSevere skill, computed for the aggregation of every 2-min ProbSevere probability forecast. Note that the y axis of the inset graph in (b) is log scaled.

- overestimated wind speeds by observers (Doswell et al. 2005);
- quantized hail size and nonstandard hail-size reporting (Schaefer et al. 2004);
- demographic factors such as urbanization, highway distribution, and education level of reporters (Kelly et al. 1985);
- underreporting of “lesser” hazards (Morgan and Summers 1982; e.g., lack of hail reports with a tornadic storm); and
- inconsistent verification practices among NWS offices (Doswell et al. 2005).

Furthermore, limitations should be mitigated by the very large number of storms in the dataset and by negligible secular trends over the short verification duration. Readers are reminded that all NWS statistics presented here are unofficial. This is due to 1) this study’s use of preliminary LSRs from the SPC instead of the vetted official reports found in the *Storm Data* publication from the National Centers for Environmental Information (NCEI) and 2) the verification rules used for NWS warnings in this study (described in section 4a) are designed to facilitate storm-by-storm validation in concert with ProbSevere and therefore differ substantially from official NWS verification rules.

b. Aggregated validation

As a function of the ProbSevere forecast probability threshold, a maximum CSI value of 0.27 is achieved at the ProbSevere 90% probability threshold (Fig. 6a). CSI steadily increases to 80%, and then flattens somewhat in the 80%–90% range. The reliability diagram (Fig. 6b) shows

good correspondence between the forecast probability and the occurrence of an LSR event, although some over-forecasting is evident at probability thresholds of 40% and greater. The inset in Fig. 6b demonstrates good sharpness in the forecasts (note that the y axis is log scaled). Figure 7 shows ProbSevere (with and without the FR/EBS predictor) and NWS skill (POD, FAR, and CSI) and median lead time to the initial LSR for a storm. Note that the 80% probability threshold is shown here because the ProbSevere CSI was only slightly less compared to the 90% threshold, but the median lead time increased by 5 min. We see that ProbSevere has a smaller POD and greater FAR compared to NWS and a CSI of 0.27 compared to 0.35 for the NWS results. However, the median lead time of ProbSevere is 35 min compared to 17 min for the NWS version. Figure 7 also demonstrates that the inclusion of total lightning improves ProbSevere (0.1 increase in POD, 0.01 increase in CSI, and the median lead time to the first LSR increases by 5 min). Recall that a “valid window” of 90 min was chosen for ProbSevere forecasts. Given shortcomings in storm reports as discussed previously, a 90-min window is more inclusive than shorter windows for severe storms that may not have LSRs within 60 min because of nonmeteorological reasons (e.g., low population density, storms occurring at night). Using a 60-min valid window did not substantially diminish the skill (CSI decreased by ≤ 0.008 and median lead time decreased by ≤ 5 min at all probability thresholds $> 10\%$, compared to a 90-min window).

c. Quasi-seasonal validation

In a day-by-day analysis of ProbSevere (forecast probability threshold of 80%) and NWS skill (Fig. 8), we

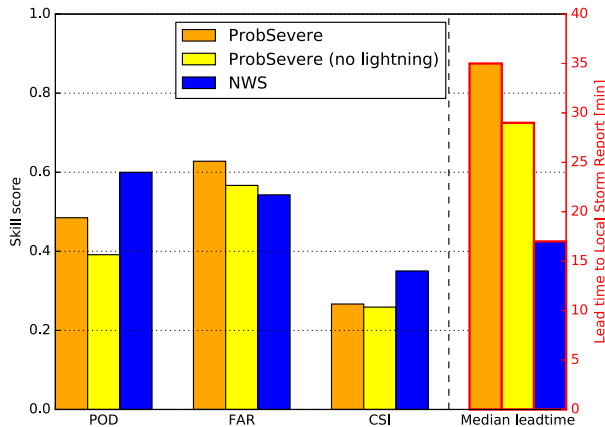


FIG. 7. ProbSevere, ProbSevere without the total lightning/EBS predictor, and NWS skill scores and median lead time to initial LSR. The ProbSevere metrics are measured from the 80% forecast probability threshold.

see that ProbSevere often underperforms the NWS, which is expected given the value of human expertise, but on some days ProbSevere can actually be more skillful than NWS warnings (e.g., 30 May, 12 June, 21 June, and 22 August 2016). In general, days where ProbSevere performs well with respect to the NWS tend to have storms that produced hail and wind reports, whereas days where ProbSevere skill is less than that of NWS tend to have storms that produce straight-line winds and/or weak tornadoes (i.e., no hail reports). For example, the dates of 6 April, 20 August, 18 November, and 28 November 2016 contained numerous linear or QLCS storms, which exhibited smaller MESH values and flash rates compared to cellular severe storms on other days. Within the Hazardous Weather Test Bed (HWT) and NWS Operations Proving Ground (OPG) experiments (Gravelle 2017), forecasters also noted that ProbSevere was more skilled at forecasting storms that only produced severe hail or produced severe hail along

with severe straight-line winds or tornadoes. As alluded to in section 3, this anecdotal evidence suggests the need for additional research addressing ProbSevere performance with different storm modes. Thompson et al. (2012) show that while EBS is a helpful predictor for severe versus nonsevere, it cannot (and thus ProbSevere cannot) readily discern storm mode. In the meantime, users of ProbSevere should be aware that these caveats exist and should therefore calibrate expectations accordingly.

d. Regional validation

A geographical assessment of ProbSevere was also performed on the 119 days noted in Table 2. Storms were assigned to NWS Weather Forecast Office (WFO) county warning areas (CWAs) based on their mean lifetime centroid latitudes and longitudes. To diminish the effects of a small sample size for some WFOs, skill statistics were aggregated for each CWA by including the CWAs of neighboring WFOs, that is, spatially adjacent CWAs. For instance, the skill calculated for the Nashville, Tennessee, WFO would aggregate storms for seven different CWAs: Nashville Tennessee; Knoxville, Tennessee; Huntsville, Alabama; Memphis, Tennessee; Paducah, Kentucky; Louisville, Kentucky; and Jackson, Kentucky. The aggregation procedure was repeated for each NWS WFO in the contiguous United States, except Key West, Florida, which was excluded from the analysis because no warnings were issued for the dates considered.

The CWA analysis shows that the probability threshold associated with the best CSI is a function of region (see Fig. 9). For instance, the most skillful probability threshold in the Ohio valley and Midwest is 60%–80%, depending on NWS CWA, while the most skillful threshold for the Great Plains is generally 90%. The southeast United States has maximum CSI scores

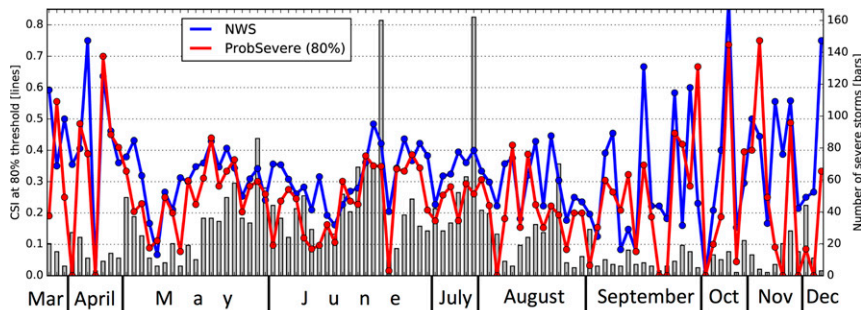


FIG. 8. A time series of ProbSevere CSI at the 80% forecast probability threshold and NWS CSI to initial LSRs (lines) and the number of severe storms analyzed on each day (bars). The time series covers a portion of the annual cycle for 2016 only. Please see Table 2 for a complete list of days in the validation.

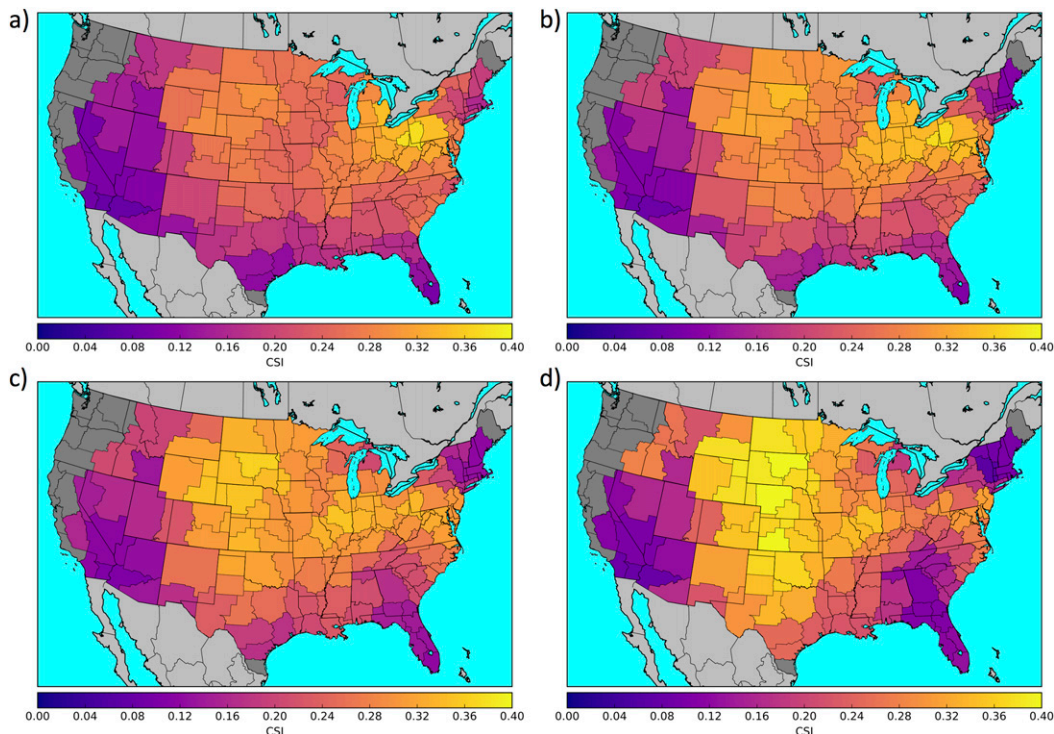


FIG. 9. ProbSevere model CSI by NWS region (see text for details) for the (a) 60%, (b) 70%, (c) 80%, and (d) 90% forecast probability thresholds. Dark gray regions are excluded because the sample size is too small. Please see the following link for more interactive maps detailing specific information such as POD, FAR, CSI, median lead time, number of storms, NWS offices included in the aggregation, and NWS skill (http://cimss.ssec.wisc.edu/severe_conv/training/validation.html).

generally achieved with the 60%–70% probability thresholds, ranging from 0.15 to 0.30. The Gulf Coast and Northeast U.S. regions have maximum CSI values of 0.1–0.25. The model tends to suffer from a large FAR at most thresholds in the Gulf Coast region, while suffering from a lower POD in the Northeast (not shown).

The western United States (i.e., west of the Rocky Mountains) also shows diminished CSI compared to the plains and Midwest, which may in part be due to sampling and possibly also due to other factors such as increased radar blockage, a relatively reduced lightning detection efficiency, difficulty in obtaining severe reports, and fewer supercell and convective-line storms in the region (Kolodziej et al. 2011). Readers should bear in mind that skill scores for the NWS are unofficial and hits are only counted for the *initial* LSR.

e. ENTLN influence on validation

There is much recent research showing that ENTLN has spatially and seasonally heterogeneous detection efficiency (DE) in North America when compared to the Lightning Imaging Sensor (LIS), a space-based optical detector similar to the GLM (e.g., Bitzer et al. 2016;

Rudlosky 2015; Thompson et al. 2014). Furthermore, ENTLN DE across the CONUS has increased over time (Bitzer et al. 2016), complicating validation statistics. Thompson et al. (2014) found that ENTLN had the greatest number of flashes coincident with LIS (80%–90%) in a semicircle that extended from central Oklahoma, through east Texas, along the northern Gulf of Mexico, across southern Florida, and along the U.S. Southeast coast (their Fig. 5). This greater DE along the Gulf Coast and the Southeast United States is one possible explanation for greater ProbSevere FAR in the region. Surrounding regions (i.e., Arizona, New Mexico, west Texas, inland parts of the U.S. Southeast) had between 40% and 80% DE. Rudlosky (2015) showed large daily variability of ENTLN DE in North America, along with seasonal decreases in 30-day average DE from March through September of 80%–40% in 2011, 80%–50% in 2012, and 80%–60% in 2013 (his Fig. 2). Rudlosky (2015) also noted that this could be a result of a number of factors, including ENTLN sensor placement and limited LIS sampling.

Recall that the total lightning predictor in ProbSevere was trained using storms across the CONUS from May

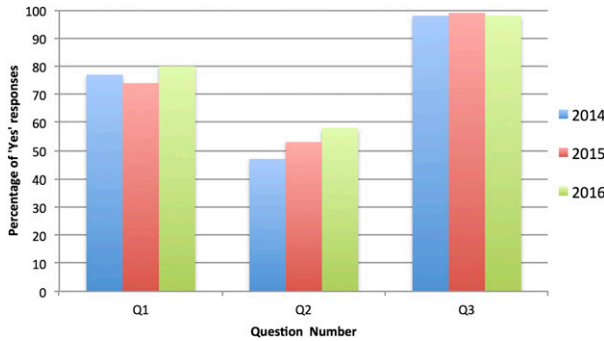


FIG. 10. Forecaster survey results for three questions, pertaining to the ProbSevere model, during the HWT from 2014 to 2016. Question 1 (Q1): In general, did the NOAA/CIMSS ProbSevere model output help increase your confidence in issuing severe thunderstorm (tornado) warnings? Q2: In general, did the NOAA/CIMSS ProbSevere model output help increase the lead time in which you were able to issue severe thunderstorm (tornado) warnings? Q3: Would you use the NOAA/CIMSS ProbSevere model output during warning operations at your WFO if available?

through August 2015. Training did not take into account the disparity in DE both regionally and seasonally. While ENTLN DE may be improving as time goes on (e.g., Bitzer et al. 2016), these heterogeneities underscore the need for renewed training of this predictor in ProbSevere every few years, utilizing the DE as a function of location and time. The GLM is expected to have a much more uniform and potentially greater DE across the CONUS, though the instrument will have limited spatial resolution (8 km at nadir) and will not be able to distinguish between IC and CG. Thus, terrestrial lightning networks and space-based sensors may be used together for improved severe hazard prediction.

5. *Forecasters' Feedback*

ProbSevere has been evaluated in real time by over 60 NWS and broadcast meteorologists at NOAA's HWT in Norman, Oklahoma, during the spring seasons of 2014, 2015, and 2016. Overall, the feedback was highly favorable, as demonstrated by participant blog posts ([Satellite Proving Ground at the Hazardous Weather Testbed 2016](#)) and forecaster survey responses.

Forecasters answered numerous questions regarding ProbSevere, with the results of three shown in Fig. 10. Over these three evaluation years, approximately 75%–80% of forecasters thought ProbSevere helped increase confidence in issuing (or not issuing) severe weather warnings, 50%–58% responded that ProbSevere increased lead time to severe hazards for their warnings, and 99% of forecasters would use ProbSevere during warning operations. With total lightning included in ProbSevere in 2016, forecasters gave the highest levels of favorable responses.

A wider evaluation of ProbSevere was performed from the spring through fall of 2016, with 52 NWS WFOs participating, 37 from the Central Region (CR) and 15 from the Eastern Region (ER). The ER/CR evaluation was conducted as an extension of the NWS OPG (Gravelle 2017). With 52 offices participating, it is likely that hundreds of NWS forecasters were able to use ProbSevere for multiple severe events. At the end of the evaluation period, the majority of forecasters that responded to questions indicated that ProbSevere model output was useful in providing confidence for warning decision-making and that they would use ProbSevere guidance again for upcoming convective events (Table 3). An additional 25 NWS forecasters served as the “point of contact” (POC) for their WFO and responded to a separate survey. These forecasters often (but not exclusively) held the managerial position of science operations officer (SOO). Based on the responses from these 25 POCs, ProbSevere aided in the warning decision process at least 500 times during the experiment [Table 4; Gravelle (2017)]. The feedback from the NWS OPG ER/CR evaluation corroborates the HWT forecaster feedback from 2014 to 2016: that ProbSevere contributes to forecaster confidence and potential lead time on severe weather warnings in many situations.

6. *Model Data Sources*

The NOAA/CIMSS ProbSevere model fuses data from the Rapid Refresh NWP model, geostationary weather satellites (GOES), the NWS weather radar network (MRMS), and, most recently, ground-based lightning measurements (from ENTLN) to determine

TABLE 3. Mean forecaster responses to questions following a 6-month-long experiment conducted by the NWS OPG. Forecasters could respond with a number from 1 to 10, with 1 being least useful or least likely and 10 being most useful or most likely.

Question	Mean forecaster response
Q1: How useful did you find the ProbSevere Model for providing confidence in issuing or not issuing convective warnings on thunderstorms?	7/10
Q2: How likely are you to use the ProbSevere Model for an upcoming convective event?	9/10

TABLE 4. NWS WFO POC responses after the experiment.

Question	No. of times	POC responses
Q1: Overall, during the evaluation, how many times would you estimate that your office used the ProbSevere Model to provide confidence in issuing or not issuing convective warnings on thunderstorms?	>30	9/25 (36%)
	21–30	6/25 (24%)
	11–20	9/25 (36%)
	≤10	1/25 (4%)

the probability that a thunderstorm will produce severe weather in the near term. Severe weather probabilities are computed on a storm object basis using a naïve Bayesian classifier. ProbSevere output has been experimentally provided in real time to the NWS since 2014 to aid in severe weather warning operations. Overall, ProbSevere has been shown to increase forecaster confidence and extend lead time to severe weather hazards. ProbSevere aids severe weather warning operations by distilling gigabytes of data into kilobytes of useful information directly relevant to decision-making.

The inclusion of a bivariate total lightning flash rate and effective bulk shear predictor improved the accuracy and lead time of ProbSevere (relative to the first report of severe weather). A robust validation analysis revealed that ProbSevere performance varies over the CONUS, performing best in the Ohio valley, Midwest United States, and Great Plains regions, and worst along the Gulf Coast, New England, and the Southwest United States. A frequent piece of feedback from forecasters is the request for probabilistic guidance based on hazard type (tornado, hail, or wind). To this end, work is under way in providing such guidance using the ProbSevere framework (i.e., multiscale storm tracking, data mining, naïve Bayesian classifier), but with different combinations of meteorological predictors appropriate to each hazard. This effort will result in more accurate and better-calibrated guidance for the wind and tornado hazards in particular. Work is also under way to retrain ProbSevere with on-orbit GOES-16 data. With much improved spatial, spectral, and temporal capabilities, the GOES-16 Advanced Baseline Imager (ABI) should help improve the accuracy and lead time of ProbSevere. In addition, the GLM, which is available on the GOES-R series of satellites, will allow for expanded use of lightning-based predictors. Tools like ProbSevere, which transform very large volumes of data into actionable information, enable users to better utilize advanced measurement and modeling capabilities without being overwhelmed by their information volume.

Acknowledgments. The authors acknowledge the NOAA GOES-R Risk Reduction Program (Grant NA15NES4320001) and NOAA/OAR (Grant NA15OAR4590188) for support of this research, as well as Dr. Kristin Calhoun

at the University of Oklahoma for aid in processing the total lightning data. The authors are also grateful for productive comments by Dr. Matthew Bunkers, Thomas Turnage, and one anonymous reviewer. The GOES-13 and GOES-15 data used in this study can be freely obtained from NOAA’s Comprehensive Large Array Data Stewardship System (CLASS; online at <https://www.class.ncdc.noaa.gov>). The Rapid Refresh NWP data can be freely obtained from NOAA’s National Centers for Environmental Information (NCEI; online at <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/rapid-refresh-rap>). The MRMS data in this study can be obtained from UW-CIMSS upon request of the corresponding author. The ENI total lightning data can be obtained from UW-CIMSS upon request of the corresponding author, pending confirmation of NOAA partnership or expressed written consent from Earth Networks. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

REFERENCES

Adler, R. F., and D. D. Fenn, 1979: Thunderstorm intensity as determined from satellite data. *J. Appl. Meteor.*, **18**, 502–517, [https://doi.org/10.1175/1520-0450\(1979\)018<0502:TIADFS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<0502:TIADFS>2.0.CO;2).

—, and —, 1981: Satellite-observed cloud-top height changes in tornadic thunderstorms. *J. Appl. Meteor.*, **20**, 1369–1375, [https://doi.org/10.1175/1520-0450\(1981\)020<1369:SOCTHC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1981)020<1369:SOCTHC>2.0.CO;2).

—, M. J. Markus, and D. D. Fenn, 1985: Detection of severe Midwest thunderstorms using geosynchronous satellite data. *Mon. Wea. Rev.*, **113**, 769–781, [https://doi.org/10.1175/1520-0493\(1985\)113<0769:DOSMTU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<0769:DOSMTU>2.0.CO;2).

Benjamin, S. G., S. S. Weygandt, C. R. Alexander, J. M. Brown, T. Smirnova, P. Hofmann, E. P. James, and G. Dimeg, 2011: NOAA’s hourly-updated 3km HRRR and RUC/Rapid Refresh—Recent (2010) and upcoming changes toward improving weather guidance for air-traffic management. *Special Symp. on Weather–Air Traffic Management Integration*, Seattle, WA, Amer. Meteor. Soc., 3.2, <https://ams.confex.com/ams/91Annual/webprogram/Paper185659.html>.

Bitzer, P. M., J. C. Burchfield, and H. J. Christian, 2016: A Bayesian approach to assess the performance of lightning detection systems. *J. Atmos. Oceanic Technol.*, **33**, 563–578, <https://doi.org/10.1175/JTECH-D-15-0032.1>.

Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail

- climatology of the contiguous United States. *Wea. Forecasting*, **2**, 1235–1248, <https://doi.org/10.1175/WAF-D-11-00151.1>.
- , M. J. Pavolonis, J. M. Sieglaff, and A. K. Heidinger, 2013: Evolution of severe and nonsevere convection inferred from GOES-derived cloud properties. *J. Appl. Meteor. Climatol.*, **52**, 2009–2023, <https://doi.org/10.1175/JAMC-D-12-0330.1>.
- , —, —, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Wea. Forecasting*, **2**, 639–653, <https://doi.org/10.1175/WAF-D-13-00113.1>.
- Deierling, W., and W. A. Petersen, 2008: Total lightning activity as an indicator of updraft characteristics. *J. Geophys. Res.*, **113**, D16210, <https://doi.org/10.1029/2007JD009598>.
- Domingos, P., and M. Pazzani, 1997: On the optimality of the simple Bayesian classifier under zero-one loss. *Mach. Learn.*, **2**, 103–130, <https://doi.org/10.1023/A:1007413511361>.
- Doswell, C. A., III, H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595, <https://doi.org/10.1175/WAF866.1>.
- FEMA, 2007: Thunderstorms fact sheet. FEMA Rep. 557, 2 pp., <https://www.fema.gov/media-library/assets/documents/12392>.
- Gagne, D. J., II, A. McGovern, J. Brotzge, M. Coniglio, J. Correia Jr., and M. Xue, 2015: Day-ahead hail prediction integrating machine learning with storm-scale numerical weather models. *Proc. 27th Conf. on Innovative Applications of Artificial Intelligence*, Association for the Advancement of Artificial Intelligence, 3954–3960, <https://www.aaai.org/ocs/index.php/IAAI/IAAI15/paper/viewFile/9724/9898>.
- Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.*, **125**–**126**, 34–49, <https://doi.org/10.1016/j.atmosres.2013.01.006>.
- Gourley, J. J., and Coauthors, 2017: The FLASH Project: Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Amer. Meteor. Soc.*, **98**, 361–372, <https://doi.org/10.1175/BAMS-D-15-00247.1>.
- Gravelle, C. M., 2017: Evaluating the ProbSevere Model at NWS Eastern and Central Region Weather Forecast Offices. *13th Symp. on New Generation Operational Environmental Satellite Systems*, Seattle, WA, Amer. Meteor. Soc., <https://ams.confex.com/ams/97Annual/webprogram/Paper315764.html>.
- Hales, J. E., Jr., and D. L. Kelly, 1985: The relationship between the collection of severe thunderstorm reports and warning verification. Preprints, *14th Conf. on Severe Local Storms*, Indianapolis, IN, Amer. Meteor. Soc., 13–16.
- Hand, D. J., and K. M. Yu, 2001: Idiot's Bayes—Not so stupid after all? *Int. Stat. Rev.*, **6**, 385–398.
- Heidinger, A. K., A. T. Evan, M. J. Foster, and A. Walther, 2012: A naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x. *J. Appl. Meteor. Climatol.*, **51**, 1129–1144, <https://doi.org/10.1175/JAMC-D-11-02.1>.
- Kain, J. S., and Coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bull. Amer. Meteor. Soc.*, **94**, 1213–1225, <https://doi.org/10.1175/BAMS-D-11-00264.1>.
- Kelly, D. L., J. T. Schaefer, and C. A. Doswell III, 1985: Climatology of nontornadic severe thunderstorm events in the United States. *Mon. Wea. Rev.*, **113**, 1997–2014, [https://doi.org/10.1175/1520-0493\(1985\)113<1997:CONSTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<1997:CONSTE>2.0.CO;2).
- Kolodziej, A., V. Lakshmanan, and T. Smith, 2011: The development of a storm type climatology using an automated storm classification system. *27th Conf. on Interactive Information Processing Systems*, Seattle, WA, Amer. Meteor. Soc., P.9, <https://ams.confex.com/ams/97Annual/webprogram/Paper180517.html>.
- Kossin, J. P., and M. Sitkowski, 2009: An objective model for identifying secondary eyewall formation in hurricanes. *Mon. Wea. Rev.*, **137**, 876–892, <https://doi.org/10.1175/2008MWR2701.1>.
- Lagerquist, R. A., A. McGovern, and T. Smith, 2017: Using machine learning to predict straight-line convective wind hazards throughout the continental United States. *27th Conf. on Interactive Information Processing Systems*, Seattle, WA, Amer. Meteor. Soc., 4.3, <https://ams.confex.com/ams/97Annual/webprogram/Paper316107.html>.
- Lakshmanan, V., and T. Smith, 2009: Data mining storm attributes from spatial grids. *J. Atmos. Oceanic Technol.*, **26**, 2353–2365, <https://doi.org/10.1175/2009JTECHA1257.1>.
- , A. Fritz, T. Smith, K. Hondl, and G. Stumpf, 2007a: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor. Climatol.*, **46**, 288–305, <https://doi.org/10.1175/JAM2460.1>.
- , T. Smith, G. Stumpf, and K. Hondl, 2007b: The Warning Decision Support System—Integrated Information. *Wea. Forecasting*, **22**, 596–612, <https://doi.org/10.1175/WAF1009.1>.
- Mecikalski, J. R., and K. M. Bedka, 2006: Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery. *Mon. Wea. Rev.*, **134**, 49–78, <https://doi.org/10.1175/MWR3062.1>.
- Mielniczuk, J., 1997: On the asymptotic mean integrated squared error of a kernel density estimator for dependent data. *Stat. Probab. Lett.*, **34**, 53–58, [https://doi.org/10.1016/S0167-7152\(96\)00165-4](https://doi.org/10.1016/S0167-7152(96)00165-4).
- Miller, P., A. W. Ellis, and S. Keighton, 2015: A preliminary assessment of using spatiotemporal lightning patterns for a binary classification of thunderstorm mode. *Wea. Forecasting*, **30**, 38–56, <https://doi.org/10.1175/WAF-D-14-00024.1>.
- Morgan, G. M., Jr., and P. W. Summers, 1982: Hailfall and hailstorm characteristics. *Thunderstorms: A Social, Scientific and Technological Documentary*, Vol. 2, *Thunderstorm Morphology and Dynamics*, E. Kessler, Ed., U.S. Government Printing Office, 363–408.
- NOAA/NWS, 2010: Thunderstorms, tornadoes, lightning. . . nature's most violent storms. NOAA/PA 201051, 18 pp., <https://www.weather.gov/media/owlie/ttl6-10.pdf>.
- , 2017: WFO severe weather products specification. National Weather Service Instruction 10-511, 35 pp., <http://www.nws.noaa.gov/directives/sym/pd01005011curr.pdf>.
- Pavolonis, M. J., 2010a: Advances in extracting cloud composition information from spaceborne infrared radiances—A robust alternative to brightness temperatures. Part I: Theory. *J. Appl. Meteor. Climatol.*, **49**, 1992–2012, <https://doi.org/10.1175/2010JAMC2433.1>.
- , 2010b: GOES-R Advanced Baseline Imager (ABI) algorithm theoretical basis document for cloud type and cloud phase, version 2. NOAA/NESDIS/Center for Satellite Applications and Research Rep., 86 pp., https://www.star.nesdis.noaa.gov/goesr/docs/ATBD/Cloud_Phase.pdf.
- , J. Sieglaff, and J. Cintineo, 2015: Spectrally enhanced cloud objects—A generalized framework for automated detection of volcanic ash and dust clouds using passive satellite measurements: 1. Multispectral analysis. *J. Geophys. Res. Atmos.*, **120**, 7813–7841, <https://doi.org/10.1002/2014JD022968>.
- Reynolds, D. W., 1980: Observations of damaging hailstorms from geosynchronous satellite digital data. *Mon. Wea. Rev.*, **108**, 337–348, [https://doi.org/10.1175/1520-0493\(1980\)108<0337:OODHFG>2.0.CO;2](https://doi.org/10.1175/1520-0493(1980)108<0337:OODHFG>2.0.CO;2).

- Roberts, R. D., and S. Rutledge, 2003: Nowcasting storm initiation and growth using *GOES-8* and WSR-88D data. *Wea. Forecasting*, **1**, 562–584, [https://doi.org/10.1175/1520-0434\(2003\)018<0562:NSIAGU>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0562:NSIAGU>2.0.CO;2).
- Rothfus, L. P., P. T. Schlatter, E. Jacks, and T. M. Smith, 2014: A future warning concept: Forecasting a Continuum of Environmental Threats (FACETs). *Second Symp. on Building a Weather-Ready Nation*, Atlanta, GA, Amer. Meteor. Soc., 2.1, <https://ams.confex.com/ams/94Annual/webprogram/Paper232407.html>.
- Rudlosky, S. D., 2015: Evaluating ENTLN performance relative to TRMM/LIS. *J. Oper. Meteor.*, **3**, 11–20, <https://doi.org/10.15191/nwajom.2015.0302>.
- Satellite Proving Ground at the Hazardous Weather Testbed, 2016: Showing posts sorted by relevance for query ProbSevere. Satellite Proving Ground at the Hazardous Weather Testbed blog, <http://goesrhwt.blogspot.com/search?q=ProbSevere>.
- Schaefer, J., J. J. Levit, S. J. Weiss, and D. W. McCarthy, 2004: The frequency of large hail over the contiguous United States. *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., 3.3, <https://ams.confex.com/ams/pdfpapers/69834.pdf>.
- Schmit, T. J., and Coauthors, 2015: Rapid Refresh information of geostationary operational satellites. *Bull. Amer. Meteor. Soc.*, **6**, 561–576, <https://doi.org/10.1175/BAMS-D-13-00210.1>.
- Sieglauff, J. M., L. M. Counce, W. F. Feltz, K. M. Bedka, M. J. Pavolonis, and A. K. Heidinger, 2011: Nowcasting convective storm initiation using satellite-based box-averaged cloud-top cooling and cloud-type trends. *J. Appl. Meteor. Climatol.*, **50**, 110–126, <https://doi.org/10.1175/2010JAMC2496.1>.
- , D. C. Hartung, W. F. Feltz, L. M. Counce, and V. Lakshmanan, 2013: A satellite-based convective cloud object tracking and multipurpose data fusion tool with application to developing convection. *J. Atmos. Oceanic Technol.*, **30**, 510–525, <https://doi.org/10.1175/JTECH-D-12-00114.1>.
- Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **1**, 1617–1630, <https://doi.org/10.1175/BAMS-D-14-00173.1>.
- , and Coauthors, 2017: Initial results from MYRORSS: A Multi-Radar/Multi-Sensor climatology of the United States. *Special Symp. on Severe Local Storms*, Seattle, WA, Amer. Meteor. Soc., P.949, <https://ams.confex.com/ams/97Annual/webprogram/Paper314791.html>.
- SPC, 2016: Severe weather event summaries. NOAA/Storm Prediction Center, <http://www.spc.noaa.gov/climo/online/>.
- Steiger, S. M., R. E. Orville, and L. D. Carey, 2007: Lightning signatures of thunderstorm intensity over north Texas. Part I: Supercells. *Mon. Wea. Rev.*, **135**, 3281–3302, <https://doi.org/10.1175/MWR3472.1>.
- Thompson, K. B., M. G. Bateman, and L. D. Carey, 2014: A comparison of two ground-based lightning detection networks against the satellite-based Lightning Imaging Sensor (LIS). *J. Atmos. Oceanic Technol.*, **31**, 2191–2205, <https://doi.org/10.1175/JTECH-D-13-00186.1>.
- Thompson, R. L., C. M. Mead, and R. Edwards, 2007: Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. *Wea. Forecasting*, **22**, 102–115, <https://doi.org/10.1175/WAF969.1>.
- , B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **2**, 1136–1154, <https://doi.org/10.1175/WAF-D-11-00116.1>.
- Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting*, **21**, 408–415, <https://doi.org/10.1175/WAF925.1>.
- Veillette, M., H. Iskenderian, M. Wolfson, C. Mattioli, E. Hassey, and P. Lamey, 2016: The offshore precipitation capability. MIT Lincoln Laboratory Project Rep. ATC-430, 24 pp., https://www.ll.mit.edu/mission/aviation/publications/publication-files/atc-reports/Veillette_2016_ATC-430.pdf.
- Weiss, S. J., J. A. Hart, and P. R. Janish, 2002: An examination of severe thunderstorm wind report climatology: 1970–1999. *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 11B.2, <https://ams.confex.com/ams/pdfpapers/47494.pdf> 446–449.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Elsevier, 627 pp.
- Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, [https://doi.org/10.1175/1520-0434\(1998\)013<0286:AEHDAF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2).