Comparison of Lightning Forecasts from the High-Resolution Rapid Refresh Model to Geostationary Lightning Mapper Observations

BRIAN K. BLAYLOCK AND JOHN D. HOREL

Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah

ABSTRACT

The ability of the High-Resolution Rapid Refresh (HRRR) model to forecast the location of convective storms is of interest for a variety of applications. Since lightning is often present with intense convection, lightning observations from the Geostationary Lightning Mapper (GLM) on GOES-East are used to evaluate the performance of the HRRR lightning forecasts from May through September during 2018 and 2019. Model skill is presented in terms of the fractions skill score (FSS) evaluated within circular neighborhoods with radial distances from 30 to 240 km. Case studies of individual events illustrate that the HRRR lightning forecasts FSS varies from storm to storm. Mean FSS is summarized for the months with peak lightning activity (June–August) for the west, central, and east United States. Our results suggest that forecasters should use HRRR lightning forecasts to indicate general tendencies for the occurrence, region, and timing of thunderstorms in a broad region rather than expect high forecast accuracy for lightning locally. For example, when FSS is evaluated within small neighborhoods (30-km radius), mean FSS drops sharply after the first two hours of model integration in all regions and during all hours of the day. However, when evaluated within larger neighborhoods (60-km radius and larger), FSS in the western United States and northern Mexico remains high for all lead times in the late afternoon and early evening. This result is likely due to the model capturing the tendency for convection to break out over higher terrain during those hours.

1. Introduction

Convective weather—with its erratic outflows, precipitation, and lightning—constitute threats to life and property with implications for public safety, aviation, floods, and wildland fires. For wildland fires, convective outflows may cause shifts in wind direction and intensity that lead to changes in fire behavior and put firefighters and property at risk (Yarnell Hill Fire 2013; Johnson et al. 2014). Thus, the ability to forecast the location and extent of convective weather is vital for many operational forecasting applications.

Guidance from convection-allowing numerical weather models is available operationally today. Since 2014, the operational High-Resolution Rapid Refresh (HRRR) model has provided hourly updating, storm-scale forecasts out to 18 h (Benjamin et al. 2016). National Weather Service Weather Forecast Office staff, including hydrologists and incident meteorologists, relies on the HRRR as one of many tools to subjectively assess the potential for convective activity, intense precipitation, hazardous lightning, and abrupt changes in near-surface winds.

Hourly HRRR analyses and forecasts are especially useful for the wildland fire managers who work in remote locations where there might be little information from observational networks. Evaluation of HRRR convection forecasts in these remote locations, however, can be difficult for the same reason—lack of surface wind observations and limited radar coverage in the remote, mountainous areas (Radford et al. 2019). Nevertheless, HRRR output is relied on for a variety of postprocessing tools for situational awareness in the vicinity of wildfires. For instance, the WindNinja software developed by Wagenbrenner et al. (2016) and used by fire managers can be configured to use HRRR model forecasts to produce highly detailed wind fields in complex terrain near a fire. Large errors in the downscaled wind field, however, occur if the HRRR inaccurately forecasts the timing and location of convective systems. Additional research under rapid development involves complex, coupled fire–atmosphere models that often rely on the HRRR for initial and lateral boundary conditions (Kochanski et al. 2013; Prichard et al. 2019). Those

Corresponding author: Brian K. Blaylock, brian.blaylock@utah.edu

subsequent forecasts may also be impacted by imperfect convective and boundary layer conditions specified by the HRRR.

We evaluate here the ability of the HRRR model to forecast the correct locations of convective systems as a function of forecast lead time and time of day for two summer seasons across the contiguous United States. Total lightning events observed from the Geostationary Lightning Mapper (GLM) on board Geostationary Operational Environmental Satellites (GOES-East) are used as proxies for identifying the locations of thunderstorms. The extent of lightning activity measured by the GLM is used to quantify the accuracy of HRRR thunderstorm forecasts. Knowledge of the HRRR lightning threat product accuracy can benefit operational forecasters who interpret HRRR forecasts, especially in situations when the forecasted impacts are sensitive to the location of those convective events.

Characteristics of the GLM lightning data and HRRR forecasted lightning threat fields are reviewed in section 2, followed by a description of the methods used to evaluate model skill in section 3. Section 4 discusses skill metrics during four case events from 2018. Section 5 reviews model skill for the hour between 1700 and 1800 LT for all days during May–September 2018, while section 6 summarizes observed lightning and model verification statistics for the May–September 2018 and 2019 summer seasons.

2. Data

a. Geostationary lightning mapper

The GLM instrument onboard GOES-East and GOES-West (previously known as *GOES-16* and *GOES-17*, respectively) provide the first opportunity to observe lightning from geostationary orbit (see Fig. 1; Goodman et al. 2013; Rudlosky et al. 2019). GOES-East and GOES-West became operational on 18 December 2018 and 12 February 2019, respectively. The GLM instrument can provide the location and extent of thunderstorm activity and augments the current climatology record for the contiguous United States provided by proprietary ground-based lightning networks (e.g., Goodman et al. 2013; Rudlosky et al. 2019). Although the GLM cannot distinguish between cloud-to-ground and cloud-to-cloud flashes, total lightning is a useful indicator of strong convection (Schultz et al. 2011). The GLM lightning observations are a useful independent dataset to evaluate the current HRRR lightning forecasts because these observations were not used in the HRRR model's routine data assimilation during the 2018 and 2019 study period. GLM lightning data, however, are expected to

be assimilated after future upgrades to the HRRR model (E. James 2019, personal communication).

Lightning events measured by the GLM are identified by a charge-coupled device (CCD) sensitive to the 777.4-nm wavelength—an emission sensitive to lightning activity associated with the dissociation of molecular oxygen (Christian et al. 1989; Nikitovic et al. 2008; Goodman et al. 2013). The detector continuously monitors the scene for changes to the background field at a frequency of 500 Hz (2 ms). The CCD detection resolution is 8 km at nadir and 14 km near the boundaries of the field of view. CCD pixels illuminated by lightning within 2-ms frames are defined as lightning events. Since intracloud and cloud-ground lightning can often illuminate one or more pixels, the GLM level 2 Lightning Cluster Filter Algorithm is used to define the location of GLM groups and flashes (Mach et al. 2007; Goodman et al. 2013; Zhang et al. 2019). A group is the radiance-weighted centroid of multiple adjacent events that occur in the same 2-ms frame. GLM flashes, as shown in Fig. 1 for a specific 5-min period, are the radiance-weighted centroid of one or more groups that occur within 330 ms and no more than 16.5 km from each other (i.e., within one GLM pixel).

The ongoing evaluation of the GLM instrument by others is encouraging, although its accuracy varies spatially, diurnally, and seasonally. Rudlosky et al. (2019) showed that the first nine months of GLM data between December 2017 and August 2018 agreed well with lightning patterns documented in previous studies during summer including higher frequencies of lightning over Florida, the Great Plains, and the Sierra Madre mountains of northern Mexico. They also point out artifacts of false event detection along the boundaries of the detector subarrays (Rudlosky et al. 2019). For example, a prominent false detection artifact occurs east of the Bahama Islands in the GOES-East GLM data. Since this area lies for the most part outside the HRRR domain, this artifact is not expected to impact our results.

Furthermore, Koshak et al. (2018) showed that the January–August 2018 GLM detection efficiency from GOES-East is lower in the northwestern United States than other parts of the GLM field-of-view. For much of this area over land, GLM detection efficiency west and north of Colorado was below 50% when compared to the Earth Networks Total Lightning Network and below 70% when compared to Vaisala's Global Lightning Detection Network (Koshak et al. 2018). For the purpose of evaluating HRRR forecasts, our focus is on the area where lightning occurs during hourly periods and not the total number of events. Since occasional undetected lightning events might be expected to overlap
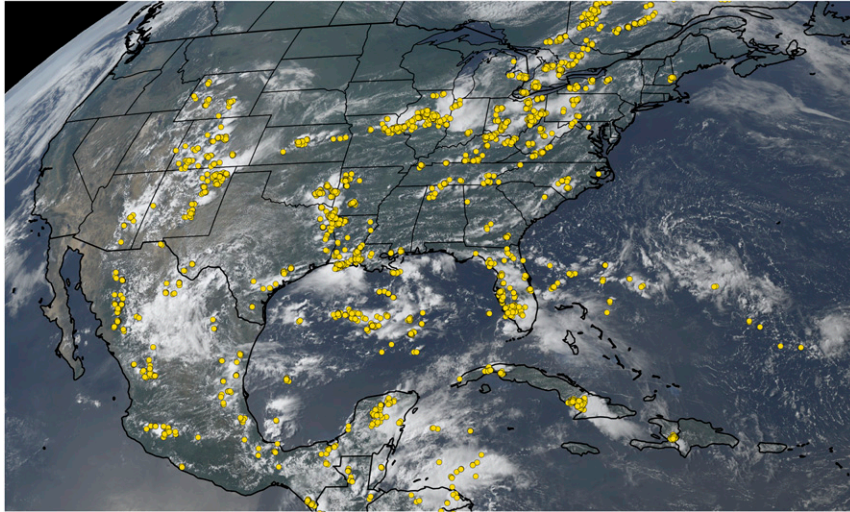
FIG. 1. GOES-East Advanced Baseline Imager natural color image and GLM flashes (yellow dots) for the 5-min period beginning at 2057 UTC 5 Jul 2018. (The Lake Christine fire, discussed in section 4b, was burning in western Colorado on this day.)

with other events during an hour, the detection efficiency errors are likely mitigated. Future algorithm updates and additional instrument calibrations are expected to reduce these types of errors. Other known types of false detections (e.g., solar glint) did not impact the cases studies shown here and likely have minimal impact on the average verification scores.

GLM Level 2 data were obtained from Amazon Web Services (https://registry.opendata.aws/noaa-goes/) for the May–September study period during 2018 and 2019. Data from the instrument are packaged in files each representing observations for a duration of 20 s. The locations of observed events and flashes within each hour of the study period are retrieved from the 180 files corresponding to each hour. These aggregated lightning records match the same hour interval for which the HRRR lightning forecasts are valid. Hourly GLM flash accumulations provide how much lightning activity is occurring over each hour. The GLM event data provide the spatial extent of this lightning activity using the native pixel resolution of the GLM.

Estimates of the occurrence of observed lightning based on the spatial extent of GLM events were generated for each hour interval on the HRRR model grid (~3-km horizontal grid spacing). Although others prefer to convert the GLM Level 2 data to the GOES Advanced Baseline Imager grid to collocate imagery features and the GLM data, we have elected to use the Level 2 geolocated GLM data that have been corrected for parallax. Since the parallax correction assumes lightning height varies only with latitude, large deviations between the assumed and actual height of the lightning

may lead to inaccurate locations. From the Level 2 GLM data, we define the footprint of each GLM event as the 21 HRRR grid points within the 5 × 5 square surrounding the event location excluding the four corner grid points—an area roughly the size of a GLM pixel. It is reasonable to expect that the HRRR grid points within the footprint of a GLM event would also experience convective conditions at that time. To reduce occasional false detections, the GLM binary fields only include those grid points where at least two GLM events occurred within the hour.

We will focus initially on case studies during the 2018 summer season followed by examination of variations in HRRR forecast skill from day-to-day during that season. Cumulative statistics will be presented using GOES-East GLM data during both the 2018 and 2019 summer seasons, which provides a sample size of over a hundred million GLM flashes across the contiguous United States. We have now examined a number of cases using both GOES-East and GOES-West GLM data during the 2019 summer season and found the differences in event and flash data obtained from the two satellites do not measurably impact the types of statistics presented here (not shown).

b. HRRR forecasted lightning threat

HRRR model output was retrieved from the University of Utah's HRRR archive (Blaylock et al. 2017; Blaylock et al. 2018). The HRRR model is a convection-allowing, short-range forecast system with a horizontal grid spacing of 3 km that runs operationally every hour at the National Centers for Environmental Prediction's Environmental

Modeling Center (Benjamin et al. 2016). The short-range 0–18-h forecasts benefit from an advanced data assimilation system that includes the incorporation of radar reflectivity data and lightning data from the Vaisala Global Lightning Detection Network during a 1-h preforecast (James and Benjamin 2017; Hu et al. 2017). The HRRR model has been upgraded regularly with steady improvements in convective forecasts (Bytheway et al. 2017) and underwent another operational upgrade on 12 July 2018 from HRRR version 2 (HRRRv2) to HRRR version 3 (HRRRv3). This latest upgrade included improved boundary layer and cloud physics, improved radar reflectivity data assimilation that reduces overforecasting convection and convective outflows, and introduced the use of additional observational data including radar radial velocities and Tropospheric Airborne Meteorological Data Reporting observations (E. James, personal communication; https://rapidrefresh.noaa.gov/hrrr/). Although 36-h forecasts are available every 6 h after the implementation of HRRRv3, this study only focuses on the 1–18-h forecasts (hereafter F01–F18).

The HRRR model forecasts provide a parameterized lightning threat product that is based on two criteria: 1) the vertical flux of graupel at the $-15°C$ level and 2) the vertically integrated hydrometeor content (McCaul et al. 2009; Kain et al. 2010; Wong et al. 2013; Goodman et al. 2013; Tippett and Koshak 2018). The resultant hourly output expresses lightning threat as the maximum flash density in 5 min during the hour before the valid time. Since the output is given by a 5-min maximum, it is not possible to estimate total number of flashes in the hour. Of course, we do not expect the forecasts to resolve lightning events or flashes obtained from the GLM explicitly. Rather, the HRRR lightning output field is used as a measure of lightning threat potential in an area derived from relevant model-based cloud microphysical processes.

The lightning threat forecasts at F01–F18 were converted to a binary grid of forecasted lightning occurrence for each hour using the threshold > 0.04 flashes km^{-2}. The resultant binary field of forecasted lightning occurrence is an estimate of the location of forecasted thunderstorm activity. We tried other thresholds (including >0.0 flashes km^{-2}) but found that frequent, very weak model convection led to many locations with lightning between 0 and 0.04 flashes km^{-2} that led to a high model detection bias. We do not expect the overall results to differ substantively if a somewhat smaller or higher threshold had been used.

3. Method

Although deterministic high-resolution models produce realistic convective forecasts, they are not necessarily more accurate because errors at small scales proliferate and are highly dependent on environmental conditions (Mass et al. 2002; Lynn et al. 2012; Mittermaier and Csima 2017). Forecasted storms may be displaced in time and space with varying levels of intensity than what is observed. Forecasts may also be missed entirely. Rather than rely on individual deterministic forecasts, it is recognized that forecasts of convective processes may be more accurately estimated probabilistically from a suite of ensemble members (Mass et al. 2002; Schwartz et al. 2019; Frogner et al. 2019). For example, convection-allowing ensemble experiments by Frogner et al. (2019) showed that even with limited predictability at small scales, ensemble forecasts provided more valuable forecasts than deterministic forecasts, especially in the summer months when predictability is typically lower. However, no operational high-resolution ensemble forecast system guidance is available across the United States at this time, which necessitates most operational forecasters to rely in part on subjective comparisons of the output from many different high-resolution deterministic forecast systems.

The objective verification of convective-scale forecasts of precipitation and lightning has been undertaken in many studies (e.g., Casati and Wilson 2007; Clark et al. 2010; Mittermaier et al. 2013; Nachamkin and Schmidt 2015; Wilkinson 2017). Contingency table approaches are common that count the number of hits, misses, false alarms, and correct rejections at each model grid point and computing basic statistics [e.g., probability of detection (POD) and false alarm ratio (FAR)] (Jolliffe and Stephenson 2011; Wilks 2011). Those approaches, however, are less appropriate for convection-scale features in convection-allowing models because even small displacement errors for features of interest are subject to "double penalty" (i.e., both where the feature is and where it is not). A better model approach for evaluating convection-scale features is to evaluate conditions within a neighborhood of points around specific grid points (Clark et al. 2010; Lynn et al. 2012; Wilkinson 2017).

Many studies have used neighborhood and object-based verification methods to provide meaningful measures of convective-scale forecast performance (Ebert 2009; Clark et al. 2010; Wolff et al. 2014; Nachamkin and Schmidt 2015; Xu et al. 2019; Radford et al. 2019). Davis et al. (2006) used object-based verification to assess the displacement of simulated and observed areas of rainfall—a method now referred to as Method for Object-based Diagnostic Evaluation (MODE; see also Bytheway and Kummerow 2015 and Radford et al. 2019). Roberts and Lean (2008) introduced the fractions skill score (FSS) now frequently used by many verification studies to evaluate skill as a function of spatial scale (Roberts 2008; Wolff et al. 2014; Sobash et al. 2019). Clark et al. (2010) and

Lynn et al. (2015) used the equitable threat score within a neighborhood for the evaluation of precipitation and lightning forecasts made by convection-allowing models. Wilkinson (2017) introduced an alternative technique for evaluating lightning forecasts with a new contingency table based on lightning coverage, distance, and intensity.

We rely on the FSS introduced by Roberts and Lean (2008) to assess location errors of HRRR lightning forecasts relative to the observed GLM lightning event area. FSS is defined as one minus the fractional area mean square error at the grid points within a neighborhood divided by the maximum possible mean square error in that neighborhood (Roberts and Lean 2008; Wolff et al. 2014). FSS provides no information on the spatial structure or intensity of the features within each neighborhood, but it does compare the area of the condition observed and forecasted within the neighborhood. Evaluating the conditions within a neighborhood helps mitigate potential errors in the position of the observed lightning introduced by the assumptions used to correct the GLM data for parallax. A FSS value of 1 is a perfect forecast indicating that the areal coverage of the condition within that specific neighborhood is identical between observations and forecasts while a value of zero indicates no correspondence between the two. Nonzero FSS implies the phenomenon in the observed and forecast grid overlap to some extent within the neighborhood being evaluated.

FSS depends strongly on the size of the neighborhood evaluated and asymptotes to 1 with increasing neighborhood size for an unbiased forecast. As the forecast bias increases in either direction (over or underforecasting), the maximum FSS possible is reduced (Mittermaier and Roberts 2010). An indicator of a ''useful'' forecast independent of the neighborhood size was defined by Roberts and Lean (2008) as FSS $\geq$ FSS$_{uniform}$. The threshold FSS$_{uniform} = (1.0 + f_0)/2$ is the average of a skillful forecast and that of a random unbiased forecast, f_0, whose probability of occurrence is the same as that of the phenomenon's observed probability throughout the domain of interest at that time. For example, if the phenomenon is observed over 10% of the entire region in a specific hour, then a useful forecast is defined here as one with FSS $\geq$ 0.55. FSS is commonly computed independently over a range of neighborhood sizes to help assess the spatial scales for which forecasts are useful (i.e., when FSS $\geq$ FSS$_{uniform}$).

For each hour of the day from May through September in 2018 and 2019, FSS was computed between the GLM observed binary field based on the GOES-East observations and each F01–F18 HRRR forecast binary field. As shown in Fig. 2, FSS was computed for the entire HRRR

FIG. 2. The HRRR model domain, shaded by topography, with boundaries of west, central, and east subregions. Center circles represent footprints used in FSS calculations with a radius of 30, 60, 120, and 240 km.

domain and three subregions (west, central, and east) that each span 16° longitude × 25.8° latitude (each with an areal extent of ~4 million km^2). The east and central regions primarily cover sections of the United States, but the west region includes the convectively active southwest summer monsoon region of northern Mexico (Nesbitt et al. 2008; Carlaw et al. 2017). Computing FSS within each subregion is useful since forecast skill varies across the entire HRRR domain, and regional successes or limitations may be masked when evaluating FSS over the entire domain (Mittermaier and Roberts 2010). FSS was also computed within specific state boundaries—Utah, Colorado, and Texas—for a few case studies presented later.

FSS was calculated for spatial scales with circular neighborhoods with radii ranging from 30 to 240 km. Although past studies using FSS have often relied on square neighborhoods for computational simplicity, using square neighborhoods has been shown to influence FSS when the displacement of a feature is oriented near the corner of a square (Skok 2016; Skok and Roberts 2016). Figure 2 shows the relative size of circular footprints used in this work with radii of 30, 60, 120, and 240 km. For reference, a circle with 30-km radius has an area of ~2800 km^2, which would encompass the environment surrounding large wildfires likely of interest to a fire manager. The larger radii might be more useful to evaluate model performance or be applicable in cases when forecasters issue warnings for large areas, like those issued by the Storm Prediction Center. However, as the neighborhood radius increases, the number of independent samples used to compute the statistics within a region of interest drops. For example, although there are ~450 000 grid points within each of the three regional domains, there are ~1400 (~90) independent 30-km (120-km) neighborhoods within those domains.

4. Case studies

a. Mallard fire—16 May 2018

The Mallard fire burned an estimated $306\,km^2$ (75 530 acres) in northern Texas between 8 May 2018 and 28 May 2018. At 0200 UTC (2000 CDT) 16 May 2018, convection along a frontal boundary approached the fire location (Fig. 3a). Gusts near the fire observed by West Texas Mesonet stations peaked at $22.2\,m\,s^{-1}$ at the Claude 12SW station and $26.4\,m\,s^{-1}$ at the Vigo Park station during the previous hour (black crosses in Figs. 3a–c). The HRRR analysis (F00) of simulated composite radar reflectivity valid at 0200 UTC (Fig. 3a) shows the extent of the storm near the fire at the top of the hour. The HRRR simulated reflectivity agrees very well with the observed radar reflectivity in this case (not shown). We find that the HRRR reflectivity analyses provide useful simulated radar coverage in otherwise data void regions between radars and over mountainous terrain affected by beam blockage.

Figure 3b shows the areal coverage where the forecasted lightning threat exceeded at least 0.04 flashes km^{-2} $(5\,min)^{-1}$ during the 0100–0200 UTC period 16 May at F01, F06, F12, and F18 lead times. The areal extent of GLM-observed lightning events during this period is denoted by the dark blue areas in Fig. 3c with the centroid locations of GLM flashes in yellow superimposed on top. All HRRR forecasts reproduced the general structure of the organized convection when compared to the instantaneous simulated radar reflectivity and the observed GLM events and flashes during that hour. As might be expected based on the northeasterly movement and redevelopment of thunderstorms during this hour, the observed and forecasted lightning areas tend to extend from the southwest (corresponding to earlier in the hour) to the northeast (later in the hour). Subjectively, there is a more direct correspondence between the areas of GLM events and F01 forecasts than at longer lead times, particularly in the vicinity of the Mallard fire where the longer lead-time forecasts extended the lightning farther downstream than observed.

Score statistics for the conditions in Texas at all forecast lead times are shown in Figs. 4a and 4b. Note that the statistics for forecast lead times from F18 to F01 are displayed from left to right along the x axes to correspond to the order the forecasts become available to operational users (i.e., each successive value to the right represents new forecast guidance closer to the valid time). The POD (hit rate) at F18–F06 is far below the FAR (or probability of false alarm) when evaluated at individual grid points (Fig. 4a). These values become comparable at shorter lead times (F06–F02). The very low FAR at F01 reflects that the model under forecasted lightning at that short lead time.

FSS at each lead time is shown in Fig. 4b using radial neighborhoods of 30 (FSS_{30}), 60 (FSS_{60}), 120 (FSS_{120}), and 240 km (FSS_{240}). As defined in section 3, useful forecasts in this situation are those where FSS > $FSS_{uniform}$ = 0.57 (above the dashed line in the Fig. 4b). Considering the accuracy of the lightning forecasts with a small neighborhood (FSS_{30}), the forecasts would be judged to be useful only at lead times of 6 h or less. If the forecasts are evaluated within a larger neighborhood (FSS_{60}), then F08–F01 forecasts might be judged useful. Comparing the areas of lightning forecast at 6-h lead time in Fig. 3b (blue shaded regions) to that observed tends to confirm that result subjectively.

On even larger scales (120- and 240-km neighborhoods), the forecasts are judged "useful" by the FSS metric at all lead times, but they have limited practical significance. That is, knowing the HRRR model has similar amounts of lightning within 120- or 240-km radii of that observed is likely of limited value for most operational forecasting applications, especially if the forecaster is interested in specific locations like a wildland fire or airport.

b. Lake Christine fire—5 July 2018

The Lake Christine fire burned $50\,km^2$ (12 588 acres) between 3 July 2018 and 17 September 2018. Based on the HRRR F00 simulated reflectivity and GLM observations in Figs. 3d and 3f, convective activity in Colorado between 2000 and 2100 UTC (1400–1500 MDT) 5 July 2018 was tied primarily to areas of locally higher terrain. Less of the state is forecast to have lightning for F01 than observed while F06, F12, and F18 show a mix of scattered and isolated thunderstorms over the high terrain combined with more organized convection drifting to the east of the Rockies during the hour. These features east of the Rockies were not observed in the HRRR analyzed reflectivity or GLM data. Over the next several hours, convection over the higher terrain weakened and organized convective storms did develop in the plains to the east of the Rockies (not shown).

Based on the subjective comparisons possible from Figs. 3e–f, it is not surprising that POD is low and FAR is high at all forecast lead times (Fig. 4c). It is typical for forecasts of scattered convection to have low POD and high FAR in point-by-point comparisons. However, POD did increase to over 0.5 at a 6-h lead time during the next hour as greater amounts of organized convection shifted east of the Rockies (not shown).

FSS values computed within 30-km radii for all forecasts valid at 2100 UTC shown in Fig. 4d are generally

F00 Simulated Reflectivity
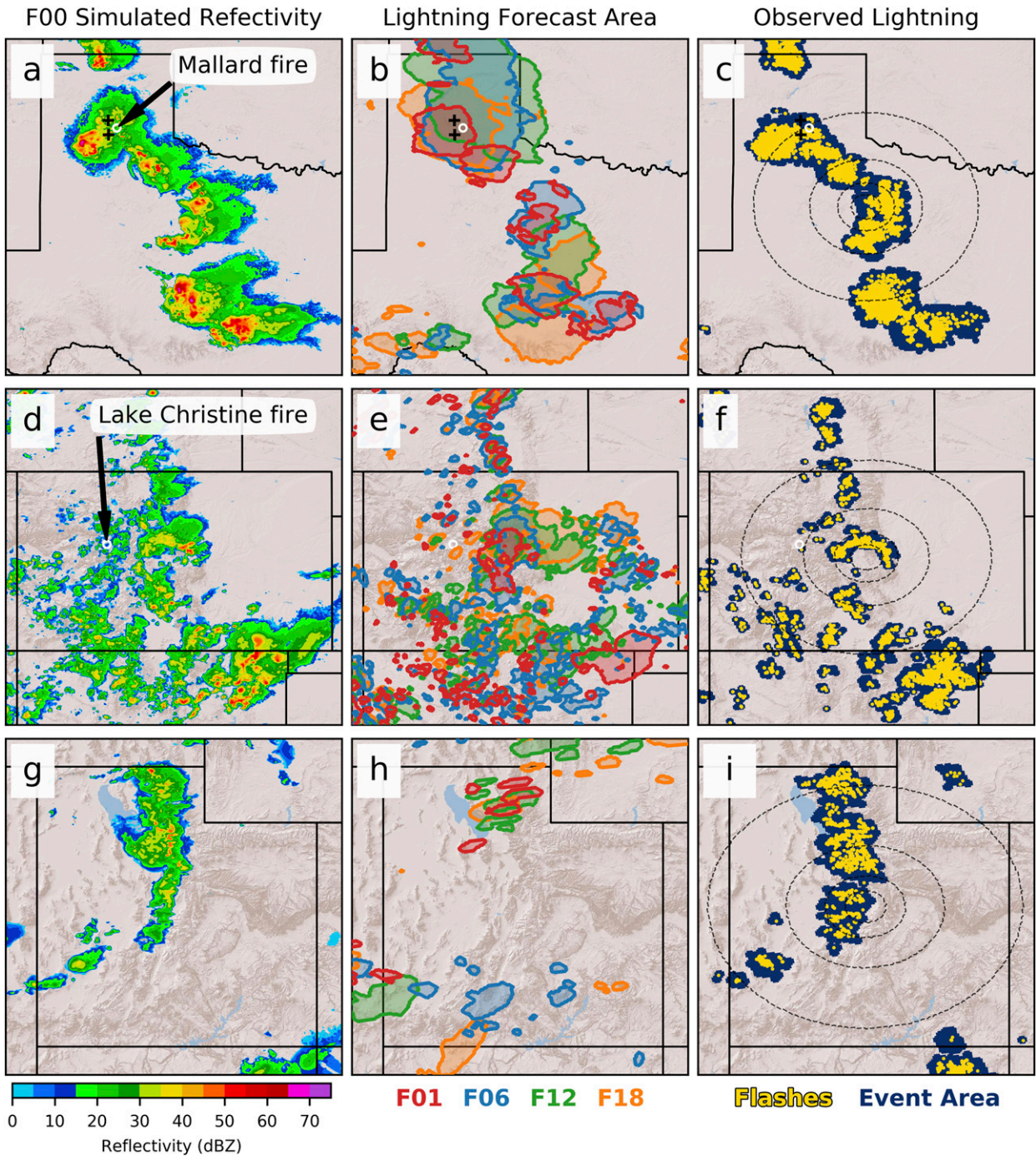
Lightning Forecast Area

Observed Lightning

FIG. 3. (a) HRRR analysis (F00) simulated composite reflectivity, according to scale, valid at 0200 UTC 16 May 2018 during the Mallard Fire in Texas. The north and south black crosses denote the locations of the Claude and Vigo Park stations, respectively. (b) Shading and contours enclosing HRRR-forecasted lightning threat during 0100–0200 UTC 16 May 2018 greater than 0.04 flashes km^{-2} (5 min)$^{-1}$ for forecast lead times of F01 (red), F06 (blue), F12 (green), and F18 (orange). (c) GLM events (dark blue) beneath GLM flashes (yellow dots) during 0100–0200 UTC 16 May 2018. Circular rings illustrate neighborhood sizes for 30, 60, 120, and 240-km radius used to calculate the FSS. (d)–(f) As in (a)–(c), but for 2000–2100 UTC 5 Jul 2018 during the Lake Christine fire in Colorado. (g)–(i) As in (a)–(c), but for 0500–0600 UTC 17 Jul 2018 during a thunderstorm in Utah.
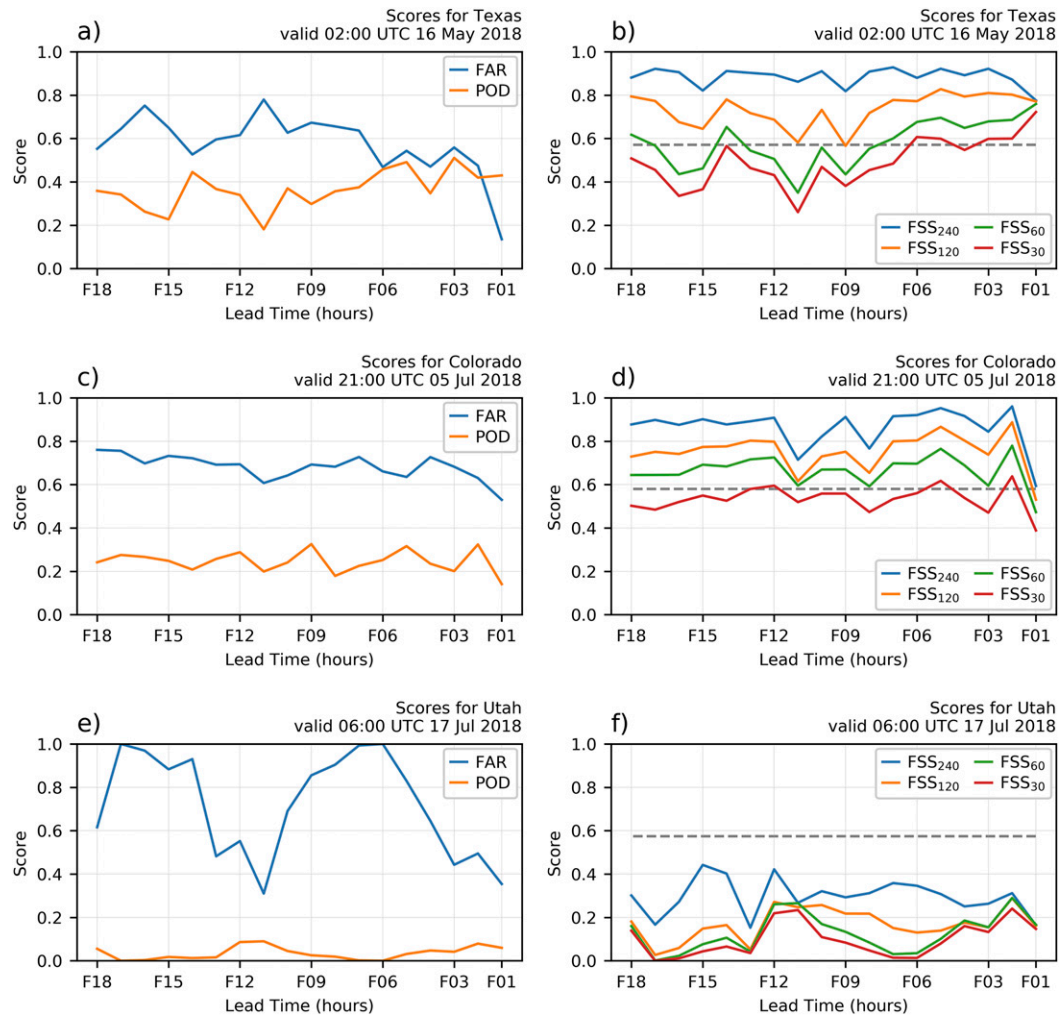
FIG. 4. (a) FAR and POD valid at 0200 UTC 16 May 2018 for F18–F01. Statistics are calculated with grid points within the Texas state boundary. (b) As in (a), but for FSS with 30-, 60-, 120-, and 240-km radial neighborhoods and FSS$_{uniform}$ (gray dashed line). (c), (d) As in (a) and (b), but valid at 2100 UTC 5 Jul 2018 for grid points within the Colorado state boundary. (e), (f) As in (a) and (b), but valid at 0600 UTC 17 Jul 2018 for grid points within the Utah state boundary.

below the useful forecast threshold, FSS$_{useful}$, defined by Roberts and Lean (2008). The FSS values from F18 to F02 for larger neighborhood sizes are above the threshold, which in part reflects the tendency for the HRRR to place lightning across higher terrain correctly.

The high FSS values computed within the 120- and 240-km radii neighborhoods in Fig. 4d are not necessarily relevant for a detailed examination of the model forecast skill of lightning. Nearly all of Colorado lies within a radius of 240 km (see Fig. 3f)—the neighborhood size approaches the size of the domain being evaluated and, in this instance, little information is gained for a specific location by suggesting the estimate of the lightning forecast over nearly the entire state of Colorado is similar to that observed. The large-scale metrics, however,

may still be useful in determining timing of general initiation and dissipation of convection for the region.

c. Utah thunderstorm—17 July 2018

A line of thunderstorms stretched from central to northern Utah at 0600 UTC 17 July 2018. This case provides an example of a poor forecast made by the HRRR model (Figs. 3g–i). Minimal lightning threat was forecasted in northern Utah with no lightning forecasted during F06 in northern Utah (lack of blue contours in Fig. 3h) while lightning was forecasted in southern Utah at F06 that never occurred. It is perhaps most surprising that even at short lead times (e.g., F01) the HRRR did not capture the extent of convection that then developed. This might be a result of the model environment conditions not

1 hr Forecasted Lightning Area
valid 00:00 UTC 23 Aug 2018

F02
F12

1 hr Observed Lightning Area
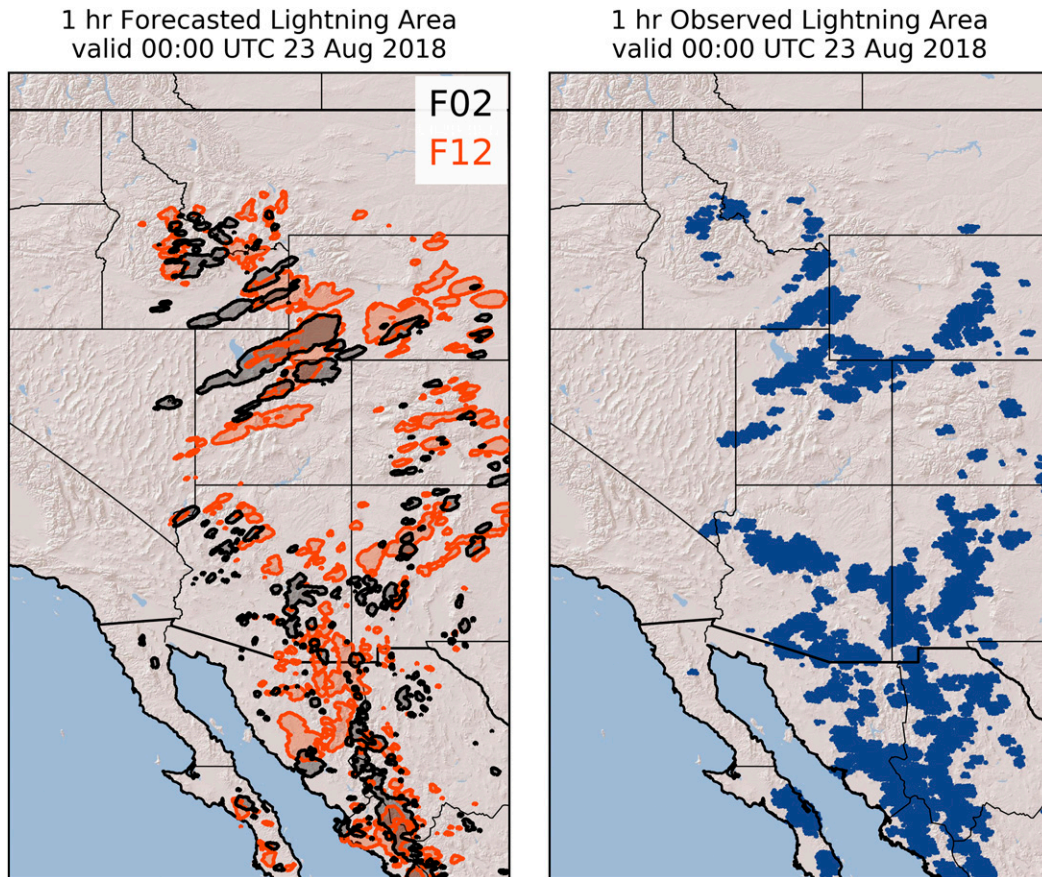## valid 00:00 UTC 23 Aug 2018

FIG. 5. Area in the west subregion with forecasted and observed lightning valid 2300–0000 UTC 23 Aug 2018. (left) Area with HRRR-forecasted lightning threat for lead times F02 (black) and F12 (orange). (right) Area with observed GLM lightning events (blue).

being suitable to sustain the storm shown in the model analysis. The poor forecasts resulted in large FAR, small POD, and small FSS values well below the useful forecast threshold at all spatial scales and all forecast lead times. This example illustrates that physically plausible scenarios and run-to-run consistency of HRRR forecasts as the forecast lead time shortens may not necessarily correspond to situations of high forecast skill.

d. West region—23 August 2018

To aid transitioning to the cumulative results to be presented in sections 5 and 6, Fig. 5 shows the area of GLM lightning events (blue shading in right panel) in the west subregion during the hour ending at 0000 UTC 23 August 2018. Nearly 29 000 distinct GLM flashes were observed in the region during this hour, which was the highest amount during this hour of the day for the west region during May–September 2018 (Fig. 6a). Extensive lightning was observed across the Sierra Madre Mountains and Sonoran deserts of northern Mexico as well as the Mogollon Rim and other mountain and desert regions of Arizona and New Mexico.

Storm reports indicated some wind damage in central Arizona near this time. In addition, thunderstorm bands developed and then extended downstream across Utah and Idaho into Wyoming, Colorado, and Montana.

F02 and F12 HRRR forecasts of the lightning threat reveal the shift in the forecast track across Utah, Wyoming, and Colorado from downstream of the terrain to closer to the observed GLM lightning locations with the shorter forecast time period (Fig. 5). The FSS_{60} values for the F02 and F12 forecasts are nearly the same—0.44 and 0.49, respectively—but both are below $FSS_{uniform} = 0.53$. These forecasts, however, may still be considered useful (FSS > $FSS_{uniform}$) if a larger neighborhood is relevant for the users' application. However, the benefit of the forecasts for a small area, such as the immediate surroundings of a large wildfire, is likely low.

5. 1700–1800 LT May–September 2018

The 1-h accumulations of GLM flashes between 1700 and 1800 LT during May–September 2018 are shown in
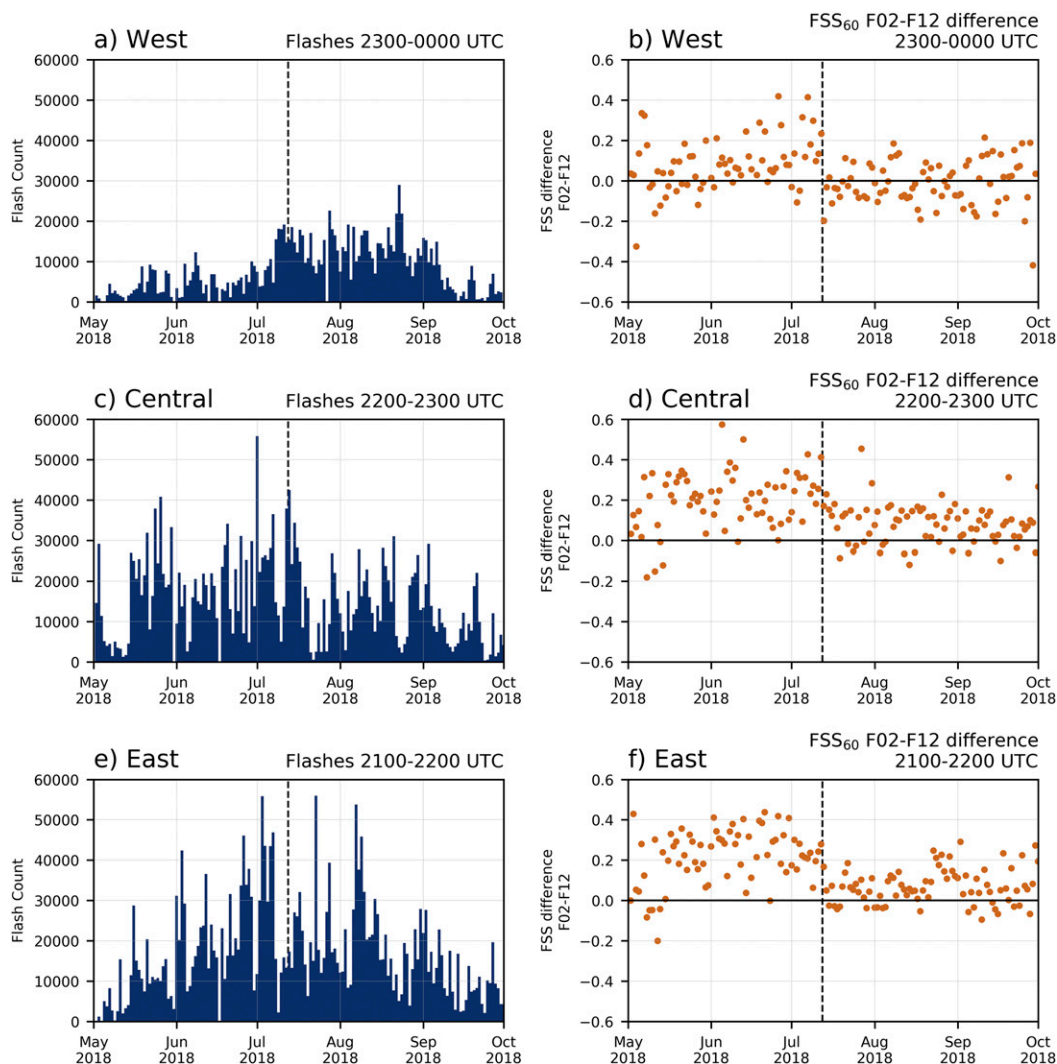
FIG. 6. (a) Total GLM flashes for 2300–0000 UTC (1700–1800 LT) in the west subregion from May through September 2018. The vertical black dashed line is the date of the HRRRv3 upgrade. (b) Differences in FSS_{60} between F02 and F12 forecasts valid 2300–0000 UTC. (c), (d) As in (a) and (b), but for the central subregion for 2200–2300 UTC. (e), (f) As in (a) and (b), but for the east subregion for 2100–2200 UTC.

Figs. 6a, 6c, and 6e for the west, central, and east subregions, respectively. As a reminder, GLM flash data are only used as an estimate of how much lightning occurred and not used in calculating FSS (FSS is based on GLM event data). The west, central, and east regions had 5, 3, and 1 h missing, respectively, for these summary statistics. The amount of lightning during this late-afternoon hour varies from day to day with overall fewer flashes occurring in the west region and more in the central and east regions. The peak lightning at this hour in the west region discussed in section 4d is quite prominent in Fig. 6a with some distinct peaks in lightning occurrence in the other regions as well. For example, the central region on 1 July 2018 had nearly 60 000 GLM flashes

accompanied by numerous wind damage reports across the northern plains.

The difference between F02 and F12 FSS_{60} scores derived for the subregions during May through September 2018 during this hour are shown in Figs. 6b, 6d, and 6f. A positive (negative) difference indicates that FSS_{60} at F02 was higher (lower) than F12. Our choice to focus on this neighborhood size and lead times will become apparent when we present the mean statistics for the entire season (section 6).

Although the HRRR model was upgraded on 12 July 2018 that included changes in the microphysics and data assimilation (https://rapidrefresh.noaa.gov/hrrr/), there are no obvious abrupt changes in skill after the upgrade
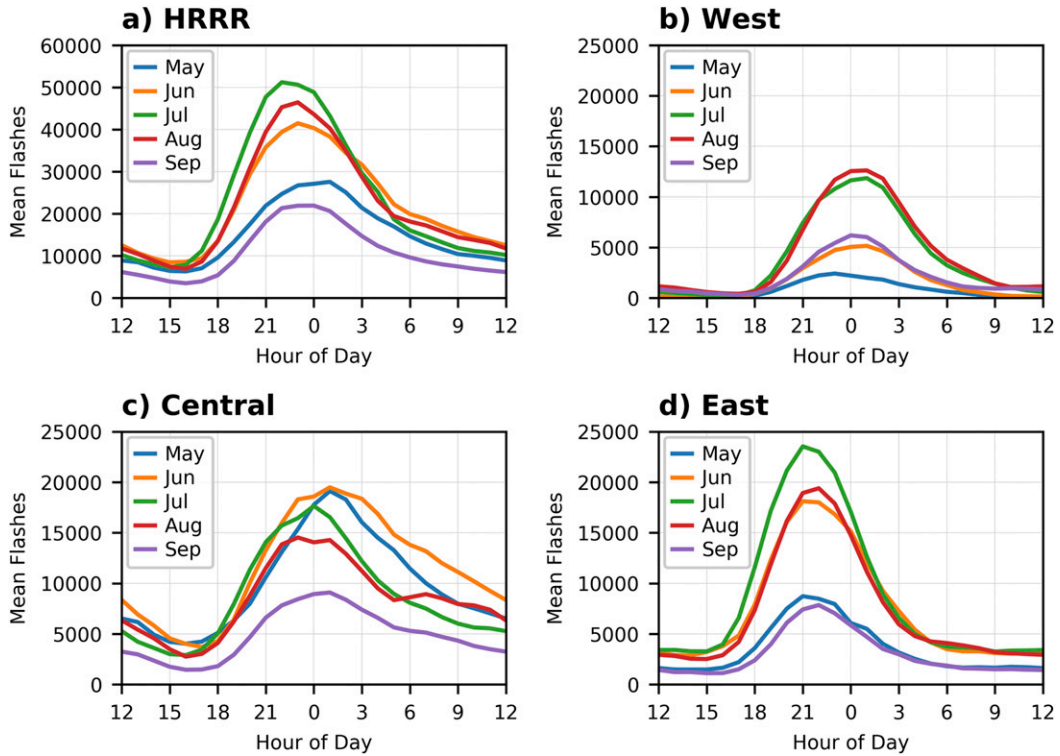
FIG. 7. Mean GLM flashes, as observed by GOES-East, during each hour of the day (1200 UTC–1200 UTC) from May through September during 2018 and 2019 for the (a) full HRRR domain, (b) west, (c) central, and (d) east subregions. The number of GLM flashes was summed within hourly periods (e.g., flashes at hour 0 were accumulated between 2300 and 0000 UTC).

that can be easily identified in Fig. 6 particularly in the central subregion. The decrease in skill in the west after mid-July is most likely due to the sharp increase in lightning caused by intraseasonal variations in lightning occurrence and not related to the model changes (Fig. 6a).

For the central and east regions, the FSS_{60} scores at 2-h lead time are more often higher than FSS_{60} at F12 (Figs. 6d,f). Over the west region, there are many days when the F12 forecasts have higher FSS_{60} values than F02 forecasts (Fig. 6b), as illustrated in the case discussed in section 4d. Hence, the morning HRRR forecasts of lightning potential for later that afternoon on those days have some skill at identifying the conditions favorable for lightning outbreaks over the mountain regions of the west.

6. Mean FSS June–August 2018 and 2019

Figure 7 shows the mean number of GLM flashes in 2018 and 2019 from GOES-East for each hour during each May–September month for the full HRRR domain and the three subregions. Over one hundred million flashes were estimated to have taken place during these two summer seasons across the contiguous

United States. July had more flashes on average than any other month across the entire HRRR domain (Fig. 7a). GLM flashes peak during the late afternoon during the hours ending at 2200–0000 UTC and decrease overnight reaching a minimum in the morning between 1400 and 1600 UTC. The times of peak solar heating shift the maxima for the east and west regions by 1–2 h during each month. Lightning activity in the west region is markedly suppressed late at night and during the early morning. Lightning storms in the central region, and to a lesser extent eastern region, often linger throughout the night due to mesoscale convective systems with more flashes occurring later into the evening, particularly during June in the central region. Overall, the three most active months for lightning are June, July, and August.

We next present mean HRRR lightning forecast FSS scores based on the GLM event observations from GOES-East during June–August 2018 and 2019. FSS computed over the entire HRRR domain for circular neighborhoods with radius 30 (FSS_{30}), 60 (FSS_{60}), and 120 km (FSS_{120}) are shown in Figs. 8a–c, respectively. For all neighborhood sizes, FSS is lowest around 1800 UTC when convection and lightning activity is beginning to develop and increases sharply as lightning
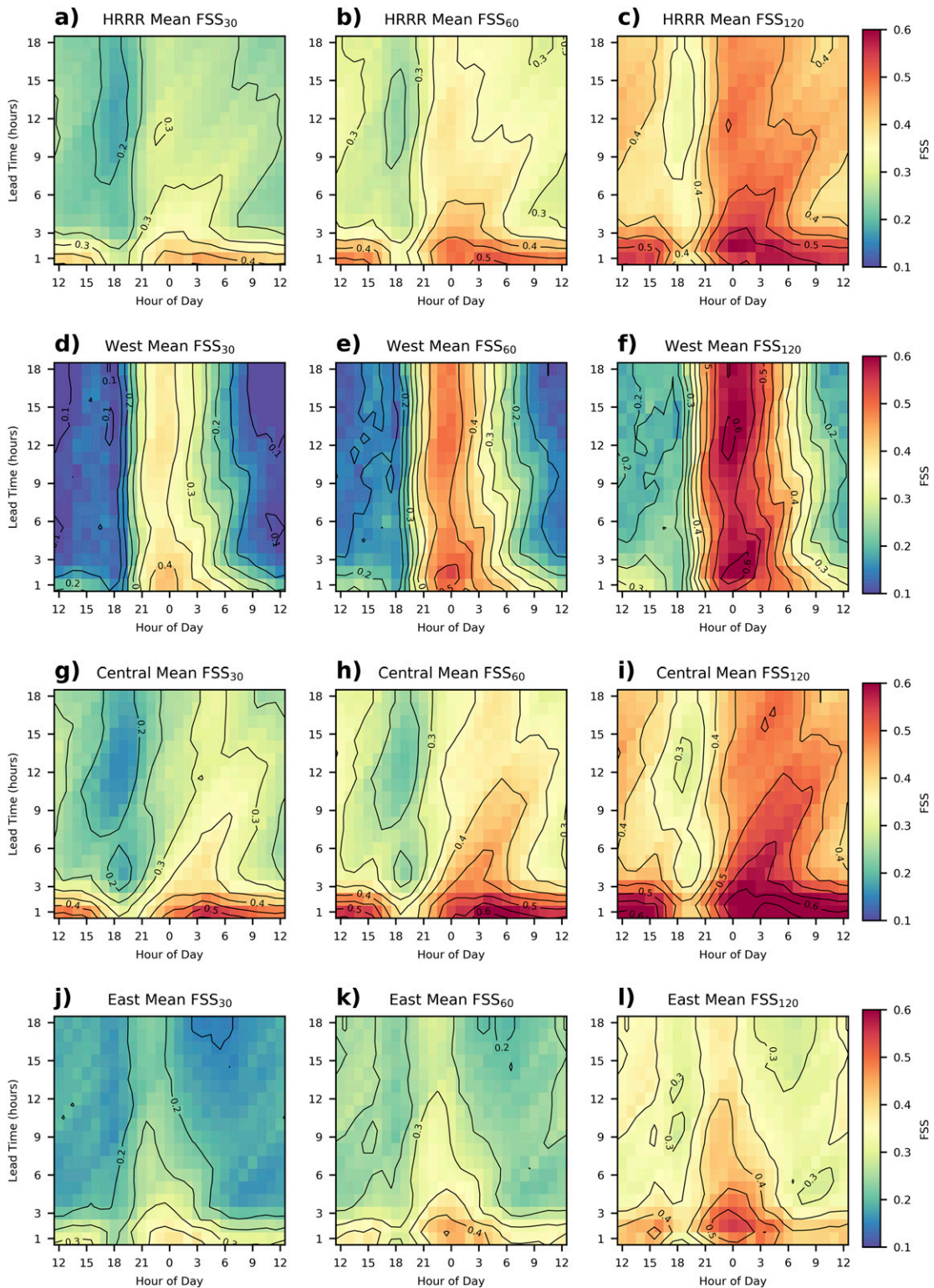
FIG. 8. (a) Mean FSS_{30} of lightning occurrence by hour of day (1200 UTC–1200 UTC) and forecast lead time (F01–F18) for the entire HRRR domain for the period 1 June–31 August during 2018 and 2019. Shaded according to the scale and contoured at an interval of 0.05. (b) As in (a), but for FSS_{60}. (c) As in (a), but for FSS_{120}. (d)–(f) As in (a)–(c), but for the west subregion. (g)–(i) As in (a)–(c), but for the central subregion. (j)–(l) As in (a)–(c), but for the east subregion.

activity peaks later in the afternoon when convection tends to become more organized and covers larger areas. Higher FSS values continue through the evening (2100–0600 UTC) with a tendency for the skill to increase slightly with decreasing lead time. FSS is highest at very short forecast lead times (F01–F02) and increases with neighborhood size.

The FSS$_{uniform}$ threshold that defines a "useful" forecast (Roberts and Lean 2008) is approximately ~0.51 for the HRRR domain. This threshold is not exceeded on average on the 30-km neighborhood scale and only at lead times of 1–2 h for the 60-km scale between 0300 and 0600 UTC. Useful forecasts at the 120-km scale exist for short lead times at all hours except 1600–2000 UTC when there is minimal lightning activity. The useful threshold is exceeded at longer lead times, up to F07, between 2200 and 0400 UTC. As explained before, FSS evaluated at the 120-km radial neighborhood scale is useful for understanding model tendencies but has little practical utility for operational forecasting applications at specific locations.

We suggest the overall tendencies exhibited by FSS as a function of lead time and time of day over the entire HRRR domain result from two factors. First, the continuously updating assimilation cycle fosters the short-range higher accuracy to predict lightning locations. However, that accuracy of the forecasted positions of those convective storms is lost after a couple of hours. Second, skill during the evening hours decreases much slower with lead time than morning and early afternoon hours. The HRRR has better skill at handling the progression of larger-scale organized convective storm complexes during the evening hours (e.g., 16 May 2018 in Texas in section 4a), and the widespread thunderstorm complexes common during the southwest monsoon across northern Mexico extending into the southwestern United States (e.g., 23 August 2018 in section 4d).

For the west subregion (Figs. 8d–f), FSS values increase very rapidly from 1800 to 2100 UTC, followed by a slower decrease through the evening and night. The FSS is low in the morning hours because there are much fewer lightning events observed between 0900 and 1800 UTC. The higher FSS at all forecast lead times during the time thunderstorms are active in the west is particularly noteworthy. The tendency for higher FSS values during the late afternoon at long lead times suggests that the model's propensity to generate afternoon convection over higher terrain features in the west is captured to some extent (see section 4d). While evaluation of the forecasts locally (FSS$_{30}$) indicates marginal skill, there are general tendencies at all neighborhood scales for similarly high skill at ~F12 and F02 between 2200 and 0000 UTC (see also Fig. 6b). The slightly lower values at F01 (i.e., lower FSS$_{120}$ at 0000 UTC from F01

to F02) suggest that the HRRR model has to spin up convection in this region that may not be initially present due to gaps in the radar coverage in the western United States and limited radar data from northern Mexico.

For the central region (Figs. 8g–i), FSS is high between 0000 and 0900 UTC for all neighborhood sizes, but the skill is lower for longer lead time as opposed to the skill remaining relatively constant with lead time as in the west. Forecast skill at F03–F09 lead times tends to be higher at 0600 UTC than 0000 UTC. FSS in the eastern region (Figs. 8j–l), however, tend to be lower than the west and central subregions. We hypothesize there are implications related to the lack of radar coverage over the large area of ocean in the east region relative to the others.

7. Summary and discussion

A wide variety of forecasting applications are sensitive to the location of thunderstorms such as tactical operations for fighting wildland fires, flood management, aviation operations, severe weather forecasting, and public safety. Forecasters rely in part on short-term forecasts provided by the HRRR model to communicate the potential of hazardous weather conditions associated with the location of these thunderstorms. This research evaluated the ability of the HRRR model to forecast thunderstorm locations relative to the location of observed lightning from the GLM sensor on GOES-East. Skill was evaluated with the FSS computed from circular neighborhoods within 30–240-km radii. Model evaluation with FSS has advantages over traditional point-by-point metrics since FSS considers a neighborhood of grid points and is less sensitive to differences in the position or areal extent of the observed or predicted lightning within that neighborhood. We computed these FSS statistics for each hour of the day from May through September during 2018 and 2019 for forecast lead times from 1 to 18 h and within different regions of the United States. Thus, the HRRR lightning threat forecasts are compared to a very large sample of GLM-derived lightning events and evaluated over many locations and mesoscale/synoptic situations.

The GLM sensor system onboard GOES-East provides valuable delineation of lightning. We have also contrasted GLM data obtained from GOES-East and GOES-West for a number of cases during the 2019 summer season (not shown). There are several caveats related to using GLM data: 1) it is not possible to discriminate between intracloud and cloud-to-ground lightning; 2) the CCD detection resolution of ~8–14 km and the processing involved to define events, groups, and flashes obfuscates the specific locations of lightning larger than the

HRRR grid resolution; and 3) the differing viewing angles of the two satellites (e.g., the western border of Utah is ~39° and ~23° off nadir for GOES-East and GOES-West, respectively) lead to slight offsets in the positions of lightning events and flashes. Many other comparisons of GLM data from GOES-East and GOES-West are becoming available with recommendations to rely on GOES-East (GOES-West) data for regions east (west) of 106°W (Bill Line, personal communication, https://satelliteliaisonblog.com/2019/07/09/goes-east-west-glm-compare-and-glm-parallax/).

We first illustrated the methods to evaluate the skill of the HRRR forecasts for thunderstorms with differing distributions of observed lightning. During 16 May 2018, organized convective complexes developed along a frontal boundary in Texas where FSS_{30} was above the useful $FSS_{uniform}$ threshold, indicating a skillful forecast, for 6 h and shorter lead times. FSS evaluated for larger neighborhoods, FSS_{120} and FSS_{240}, exceeded the useful threshold for all lead times. Scattered thunderstorms forecasted over the mountains of Colorado during the afternoon of 5 July 2018 were not generally collocated with GLM lightning events, but FSS values indicated useful forecasts when considered over relatively broad neighborhoods (i.e., radii greater than 60 km). Thunderstorms across Utah on 18 July 2018 highlighted a situation where little lightning was forecasted consistently from run-to-run although lightning did occur over a large area. Hence, the run-to-run consistency of HRRR model forecasts in this situation was a poor predictor of forecast accuracy.

The extensive outbreak of thunderstorms in the west region on 23 August 2018 with nearly 30 000 GLM lightning flashes in one hour illustrated the complex mix of types of convective storms underway at any one time when viewed regionally. Large thunderstorm complexes were evident within the summer monsoon across northern Mexico and nearby states combined with widespread fast-moving thunderstorm complexes across states farther north. This case helps explain the higher FSS at all lead times in the west evident in the mean FSS statistics. FSS in the west during the afternoon and early evening (Fig. 8e) is greater than 0.45 for FSS_{60} and greater than 0.55 for FSS_{120} for all F01–F18 lead times. FSS is much lower in the central and east subregions than in the west at the longer lead times.

It was somewhat surprising to us that FSS values for the west region were on average higher during the afternoon than those for the other two regions and less dependent on forecast lead time and neighborhood scale. One possible explanation is that the HRRR model is simply capturing the climatological tendency for thunderstorms to develop over higher terrain after extensive periods of surface heating during the afternoon

during summer. Whether that sort of information is useful depends on the application. The HRRR lightning forecasts are not sufficiently accurate on smaller scales (e.g., 30-km radial neighborhoods) to provide specific guidance on lightning location. However, when considered over larger neighborhoods (60 km radially and larger), the FSS values > 0.5 reflect a loose association between where lightning is forecast and later observed that might be useful for situational awareness over larger regions. Recognizing that HRRR forecasts are better when evaluated over large neighborhoods might help focus attention on those situations when the 12-h forecast guidance in the morning highlights greater widespread risk for afternoon lightning than has been evident recently.

Our findings are consistent with other studies that evaluated HRRR forecasts of convective storms that relying on earlier versions of the HRRR model (Bytheway and Kummerow 2015; Bytheway et al. 2017). Bytheway and Kummerow (2015) found that for short lead times, the centroid of larger storms in the central United States is often within 50 km of the observed storm centroid position. Those results are similar to the FSS_{60} shown here for the central subregion (Fig. 8h) that indicate higher FSS when the location of convective storms are evaluated with respect to larger neighborhoods and not at specific points.

Interpretation of results obtained from nearly all numerical weather models and new technologies are potentially impacted by ongoing work by the science teams to improve those models and technologies. Hence, it is important to extend these types of evaluation over time. For example, the GLM science team is continuing to enhance lightning detection efficiency and limit false detections. Also, the HRRR model is slated to undergo continual improvements in the physics and data assimilation, including assimilation of GLM and other data assets. Future evaluations of the HRRR based on GLM observations should consider the impact the assimilation of the GLM data has on the forecasts when it is included in the assimilation process.

Perhaps the most anticipated feature is the ensemble version of the HRRR model and transformation to the Rapid Refresh Forecast System using the NOAA's unified model based on the Finite-Volume Cubed-Sphere (FV3) dynamical core. The implementation of an operational high-resolution ensemble modeling system will provide forecasters with probabilistic estimates of lightning occurrence. An ensemble convection-allowing modeling system will also benefit other forecast objectives such as forecasting precipitation and snowbands (Radford et al. 2019). Our work serves as a baseline that will aid in evaluating the future improvements made to the GLM resource and HRRR modeling systems.

Additional advances in convective forecasting are happening. Ensemble-based methods developed for the High-Resolution Ensemble Forecast system (HREF) and experimental NCAR ensemble are particularly relevant (Roberts et al. 2019; Schwartz et al. 2019). In addition, coupled fire–atmosphere ensemble modeling approaches are in development and those will likely provide critical assessments on the potential for convective activity near wildfires and the impacts of fire behavior on the development of convection in the vicinity of fire crews (Kochanski et al. 2013; Prichard et al. 2019). Postprocessing techniques combined with recent observations and artificial intelligence may enhance warning of convective activity and extend the hours that forecasts are useful. Enhanced postprocessed products are already being used, such as ProbSevere (Cintineo et al. 2018) and the Localized Aviation MOS Program (Charba et al. 2019).

REFERENCES

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Blaylock, B. K., J. D. Horel, and S. T. Liston, 2017: Cloud archiving and data mining of High-Resolution Rapid Refresh forecast model output. *Comput. Geosci.*, **109**, 43–50, https://doi.org/10.1016/j.cageo.2017.08.005.

——, ——, and C. Galli, 2018: High-Resolution Rapid Refresh model data analytics derived on the open science grid to assist wildland fire weather assessment. *J. Atmos. Oceanic Technol.*, **35**, 2213–2227, https://doi.org/10.1175/JTECH-D-18-0073.1.

Bytheway, J. L., and C. D. Kummerow, 2015: Toward an object-based assessment of high-resolution forecasts of long-lived convective precipitation in the central US. *J. Adv. Model. Earth Syst.*, **7**, 1248–1264, https://doi.org/10.1002/2015MS000497.

——, ——, and C. Alexander, 2017: A features-based assessment of the evolution of warm season precipitation forecasts from the HRRR model over three years of development. *Wea. Forecasting*, **32**, 1841–1856, https://doi.org/10.1175/WAF-D-17-0050.1.

Carlaw, L. B., A. E. Cohen, and J. W. Rogers, 2017: Synoptic and mesoscale environment of convection during the North American monsoon across central and southern Arizona. *Wea. Forecasting*, **32**, 361–375, https://doi.org/10.1175/WAF-D-15-0098.1.

Casati, B., and L. J. Wilson, 2007: A new spatial-scale decomposition of the Brier score: Application to the verification of lightning probability forecasts. *Mon. Wea. Rev.*, **135**, 3052–3069, https://doi.org/10.1175/MWR3442.1.

Charba, J. P., F. G. Samplatsky, A. J. Kochenash, P. E. Shafer, J. E. Ghirardelli, and C. Huang, 2019: LAMP upgraded convection and total lightning probability and "potential" guidance for the conterminous United States. *Wea. Forecasting*, **34**, 1519–1545, https://doi.org/10.1175/WAF-D-19-0015.1.

Christian, H., R. Blakeslee, and S. Goodman, 1989: The detection of lightning from geostationary orbit. *J. Geophys. Res.*, **94**, 13 329–13 337, https://doi.org/10.1029/JD094iD11p13329.

Cintineo, J. L., and Coauthors, 2018: The NOAA/CIMSS ProbSevere model: Incorporation of total lightning and validation. *Wea. Forecasting*, **33**, 331–345, https://doi.org/10.1175/WAF-D-17-0099.1.

Clark, A. J., W. A. Gallus, and M. L. Weisman, 2010: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Wea. Forecasting*, **25**, 1495–1509, https://doi.org/10.1175/2010WAF2222404.1.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting*, **24**, 1498–1510, https://doi.org/10.1175/2009WAF2222251.1.

Frogner, I.-L., A. Singleton, M. Køltzow, and U. Andrae, 2019: Convection-permitting ensembles: Challenges related to their design and use. *Quart. J. Roy. Meteor. Soc.*, **145**, 90–106, https://doi.org/10.1002/QJ.3525.

Goodman, S., 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.

Hu, M., S. G. Benjamin, T. T. Ladwig, D. C. Dowell, S. S. Weygandt, C. R. Alexander, and J. S. Whitaker, 2017: GSI three-dimensional ensemble–variational hybrid data assimilation using a global ensemble for the regional Rapid Refresh model. *Mon. Wea. Rev.*, **145**, 4205–4225, https://doi.org/10.1175/MWR-D-16-0418.1.

James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble–variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, https://doi.org/10.1175/MWR-D-16-0398.1.

Johnson, R. H., R. S. Schumacher, J. H. Ruppert, D. T. Lindsey, J. E. Ruthford, and L. Kriederman, 2014: The role of convective outflow in the Waldo Canyon fire. *Mon. Wea. Rev.*, **142**, 3061–3080, https://doi.org/10.1175/MWR-D-13-00361.1.

Jolliffe, I. T., and D. B. Stephenson, Eds., 2011: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, 292 pp.

Kain, J. S., 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.

Kochanski, A. K., M. A. Jenkins, J. Mandel, J. D. Beezley, and S. K. Krueger, 2013: Evaluation of WRF-SFIRE performance with

field observations from the FireFlux experiment. *Geosci. Model Dev.*, **6**, 1109–1126, https://doi.org/10.5194/gmd-6-1109-2013.

Koshak, W., D. Mach, M. Bateman, P. Armstrong, and K. Virts, 2018: GOES-16 GLM level 2 data full validation data quality: Product performance guide for data users. NOAA, 16 pp., https://www.ncdc.noaa.gov/sites/default/files/attachments/GOES16_GLM_FullValidation_ProductPerformanceGuide.pdf.

Lynn, B. H., Y. Yair, C. Price, G. Kelman, and A. J. Clark, 2012: Predicting cloud-to-ground and intracloud lightning in weather forecast models. *Wea. Forecasting*, **27**, 1470–1488, https://doi.org/10.1175/WAF-D-11-00144.1.

——, G. Kelman, and G. Ellrod, 2015: An evaluation of the efficacy of using observed lightning to improve convective lightning forecasts. *Wea. Forecasting*, **30**, 405–423, https://doi.org/10.1175/WAF-D-13-00028.1.

Mach, D., H. Christian, R. Blakeslee, D. Boccippio, S. Goodman, and W. Boeck, 2007: Performance assessment of the optical transient detector and lightning imaging sensor. *J. Geophys. Res.*, **112**, D09210, https://doi.org/10.1029/2006JD007787.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skill forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430, https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2.

McCaul, E. W., S. J. Goodman, K. M. LaCasse, and D. J. Cecil, 2009: Forecasting lightning threat using cloud-resolving model simulations. *Wea. Forecasting*, **24**, 709–729, https://doi.org/10.1175/2008WAF2222152.1.

Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, https://doi.org/10.1175/2009WAF2222260.1.

——, and G. Csima, 2017: Ensemble versus deterministic performance at the kilometer scale. *Wea. Forecasting*, **32**, 1697–1709, https://doi.org/10.1175/WAF-D-16-0164.1.

——, N. Roberts, and S. Thompson, 2013: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteor. Appl.*, **20**, 176–186, https://doi.org/10.1002/met.296.

Nachamkin, J. E., and J. Schmidt, 2015: Applying a neighborhood fractions sampling approach as a diagnostic tool. *Mon. Wea. Rev.*, **143**, 4736–4749, https://doi.org/10.1175/MWR-D-14-00411.1.

Nesbitt, S. W., D. J. Gochis, and T. J. Lang, 2008: The diurnal cycle of clouds and precipitation along the Sierra Madre Occidental observed during NAME-2004: Implications for warm season precipitation estimation in complex terrain. *J. Hydrometeor.*, **9**, 728–743, https://doi.org/10.1175/2008JHM939.1.

Nikitovic, Z., A. Strinkic, G. Malovic, V. Stojanovic, and Z. Petrovic, 2008: Measurements and analysis of spatial profiles of 777.4 nm line in a townsend discharge in oxygen. *Acta Chim. Slov.*, **55**, 219–222.

Prichard, S., and Coauthors, 2019: The fire and smoke model evaluation experiment—A plan for integrated, large fire-atmosphere field campaigns. *Atmosphere*, **10**, 66, https://doi.org/10.3390/atmos10020066.

Radford, J. T., G. M. Lackmann, and M. A. Baxter, 2019: An evaluation of snowband predictability in the High-Resolution Rapid Refresh. *Wea. Forecasting*, **34**, 1477–1494, https://doi.org/10.1175/WAF-D-19-0089.1.

Roberts, B., I. L. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Postprocessing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, https://doi.org/10.1175/BAMS-D-18-0041.1.

Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169, https://doi.org/10.1002/met.57.

——, and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Rudlosky, S. D., S. J. Goodman, K. S. Virts, and E. C. Bruning, 2019: Initial geostationary lightning mapper observations. *Geophys. Res. Lett.*, **46**, 1097–1104, https://doi.org/10.1029/2018GL081052.

Schultz, C. J., W. A. Petersen, and L. D. Carey, 2011: Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends. *Wea. Forecasting*, **26**, 744–755, https://doi.org/10.1175/WAF-D-10-05026.1.

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2019: NCAR's real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, https://doi.org/10.1175/BAMS-D-17-0297.1.

Skok, G., 2016: Analysis of fraction skill score properties for a displaced rainy grid point in a rectangular domain. *Atmos. Res.*, **169**, 556–565, https://doi.org/10.1016/j.atmosres.2015.04.012.

——, and N. Roberts, 2016: Analysis of fractions skill score properties for random precipitation fields and ECMWF forecasts. *Quart. J. Roy. Meteor. Soc.*, **142**, 2599–2610, https://doi.org/10.1002/qj.2849.

Sobash, R. A., C. S. Schwartz, G. S. Romine, and M. L. Weisman, 2019: Next-day prediction of tornadoes using convection-allowing models with 1-km horizontal grid spacing. *Wea. Forecasting*, **34**, 1117–1135, https://doi.org/10.1175/WAF-D-19-0044.1.

Tippett, M. K., and W. J. Koshak, 2018: A baseline for the predictability of U.S. cloud-to-ground lightning. *Geophys. Res. Lett.*, **45**, 10 719–10 728, https://doi.org/10.1029/2018GL079750.

Wagenbrenner, N. S., J. M. Forthofer, B. K. Lamb, K. S. Shannon, and B. W. Butler, 2016: Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja. *Atmos. Chem. Phys.*, **16**, 5229–5241, https://doi.org/10.5194/acp-16-5229-2016.

Wilkinson, J. M., 2017: A technique for verification of convection-permitting NWP model deterministic forecasts of lightning activity. *Wea. Forecasting*, **32**, 97–115, https://doi.org/10.1175/WAF-D-16-0106.1.

Wilks, D., 2011: *Statistical Methods in Atmospheric Sciences*. Vol. 100. Academic Press, 704 pp.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, https://doi.org/10.1175/WAF-D-13-00135.1.

Wong, J., M. C. Barth, and D. Noone, 2013: Evaluating a lightning parameterization based on cloud-top height for mesoscale numerical model simulations. *Geosci. Model Dev.*, **6**, 429–443, https://doi.org/10.5194/gmd-6-429-2013.

Xu, M., G. Thompson, D. R. Adriaansen, and S. D. Landolt, 2019: On the value of time-lag-ensemble averaging to improve numerical model predictions of aircraft icing conditions. *Wea. Forecasting*, **34**, 507–519, https://doi.org/10.1175/WAF-D-18-0087.1.

Yarnell Hill Fire, 2013: Yarnell Hill fire serious accident investigation report. Yarnell, AZ, 122 pp., https://docs.google.com/file/d/0B36DIycSgbzWSUtjNkl1Z2ROT0k/edit.

Zhang, D., K. L. Cummins, P. Bitzer, and W. J. Koshak, 2019: Evaluation of the performance characteristics of the lightning imaging sensor. *J. Atmos. Oceanic Technol.*, **36**, 1015–1031, https://doi.org/10.1175/JTECH-D-18-0173.1.