

## Sensitivities of the WRF Lightning Forecasting Algorithm to Parameterized Microphysics and Boundary Layer Schemes

EUGENE W. MCCAUL JR.,<sup>a</sup> GEORGIOS PRIFTIS,<sup>a,g</sup> JONATHAN L. CASE,<sup>b</sup> THEMIS CHRONIS,<sup>c</sup>  
PATRICK N. GATLIN,<sup>d</sup> STEVEN J. GOODMAN,<sup>e,h</sup> AND FANYOU KONG<sup>f</sup>

<sup>a</sup> *Universities Space Research Association, Huntsville, Alabama*

<sup>b</sup> *ENSCO, Inc., Huntsville, Alabama*

<sup>c</sup> *Earth System Science Center, University of Alabama in Huntsville, Huntsville, Alabama*

<sup>d</sup> *NASA Marshall Space Flight Center, Huntsville, Alabama*

<sup>e</sup> *NOAA/NESDIS, Silver Spring, Maryland*

<sup>f</sup> *University of Oklahoma, Norman, Oklahoma*

(Manuscript received 14 May 2019, in final form 6 March 2020)

### ABSTRACT

The Lightning Forecasting Algorithm (LFA), a simple empirical procedure that transforms kinematic and microphysical fields from explicit-convection numerical models into mapped fields of estimated total lightning flash origin density, has been incorporated into operational forecast models in recent years. While several changes designed to improve LFA accuracy and reliability have been implemented, the basic linear relationship between model proxy amplitudes and diagnosed total lightning flash rate densities remains unchanged. The LFA has also been added to many models configured with microphysics and boundary layer parameterizations different from those used in the original study, suggesting the need for checks of the LFA calibration factors. To assist users, quantitative comparisons of LFA output for some commonly used model physics choices are performed. Results are reported here from a 12-member ensemble that combines four microphysics with three boundary layer schemes, to provide insight into the extent of LFA output variability. Data from spring 2018 in Nepal–Bangladesh–India show that across the ensemble of forecasts in the entire three-month period, the LFA peak flash rate densities all fell within a factor of 1.21 of well-calibrated LFA-equipped codes, with most schemes failing to show differences that are statistically significant. Sensitivities of threat areal coverage are, however, larger, suggesting substantial variation in the amounts of ice species produced in storm anvils by the various microphysics schemes. Current explicit-convection operational models in the United States employ schemes that are among those exhibiting the larger biases. For users seeking optimum performance, we present recommended methods for recalibrating the LFA.

### 1. Introduction

Lightning from thunderstorms is recognized as posing a significant threat to human life, with annual death tolls in the United States alone comparable to typical death tolls from tornadoes (Curran et al. 2000). Other parts of the world experience even larger lightning casualty rates (Holle 2008). The threat from lightning is particularly

large in the realms of outdoor labor and recreation, where cloud-to-ground (CG) lightning flashes cause almost all the casualties. In the spring of 2018, for example, some 215 lightning-strike fatalities and 60 other injuries were reported just in Bangladesh (NIRAPAD 2018), with no guarantee that those reported casualty numbers are complete. Some investigators have suggested that worldwide annual lightning deaths may exceed 20000 (Holle 2016). Aside from the risk to human life, CG lightning also causes problems with the agriculture industry, including casualties to farm animals, and to the power generation and transmission industry (Holle et al. 2005; Zhang et al. 2011). Although CG flashes naturally pose the greatest hazard, even intra-cloud (IC) lightning is capable of causing problems for the aviation sector (see Laroche et al. 2015). In addition,

---

<sup>g</sup> Current affiliation: Earth System Science Center, University of Alabama in Huntsville, Huntsville, Alabama.

<sup>h</sup> Current affiliation: Thunderbolt Global Analytics, Huntsville, Alabama.

---

*Corresponding author:* Eugene W. McCaul Jr., emccaul@usra.edu

IC lightning is usually the dominant contributor to total lightning activity, which is often closely related to storm severity, with additional threats of tornadoes, high winds, and hail (Williams et al. 1999; Gatlin and Goodman 2010; Schultz et al. 2017). Because of these concerns, interest remains high in having reliable forecasts of the amounts of total lightning expected from storms.

Early approaches to forecasting lightning were based on analytical studies relating storm lightning rates to storm electrical power and, often, to storm cloud top height (Price and Rind 1992). These studies were consistent with earlier speculations (Vonnegut 1963) and were echoed by later observational research (Williams 1985). Later, other investigators based forecast methods on statistical use of lightning climatologies (Bothwell 2005). Other methods were also implemented using measures of predicted buoyant instability aloft, as derived from numerical simulations (Bright et al. 2004). These methods, however, tended to produce overly broad forecast areas of lightning threat compared to observations of thunderstorm coverage, and had difficulty in providing quantitative flash rate guidance. Recently, Lopez (2016) devised a calibrated lightning flash scheme for use in global-scale models such as those run by the European Centre for Medium-Range Weather Forecasts (ECMWF), which uses gridded fields from a parameterized convection model to build lightning flash rate fields designed to match total regional flash rate observations when integrated on time scales of months or longer.

Meanwhile, at finer space and time scales, more elaborate full electrification schemes were developed for inclusion in explicit-convection numerical forecast models (e.g., Mansell et al. 2002; Fierro et al. 2007; Lynn et al. 2012). These latter methods offer considerable detail and insight into storm electrical behavior, and provide quantitative forecasts of lightning flash rates, flash locations, and in some schemes, even flash type and structure in simulated storms. However, even after simplification of the complexities of modeling the lightning discharge process (see, e.g., Fierro et al. 2007; Lynn et al. 2012; Fierro et al. 2013), most full electrification schemes remain computationally intensive, and are still subject to errors in their quantitative forecasts of lightning event flash rates, owing to the complexity of hydrometeor charging processes (see, e.g., Takahashi 1978; Emersic and Saunders 2010), and the intrinsic low predictability of deep convection in the parent explicit-convection model.

An important point is that many existing lightning prediction schemes are designed for specific purposes based on specific assumptions and are thus best judged by their own preferred metrics. Lynn (2017), for example, designed a physics-based Dynamic Lightning System

(DLS) and compared forecasts of lightning from the DLS with those from other published schemes. The primary metric used to evaluate the DLS was simply the occurrence or nonoccurrence of lightning, but this metric emphasizes scheme performance in marginal lightning threat situations and is not necessarily the metric best suited for evaluating the other schemes.

McCaul et al. (2009) proposed a simple, empirical, physics-based lightning forecast tool, called the Lightning Forecasting Algorithm (LFA), which converts selected proxy fields from convection-allowing models (CAMs; see Kain et al. 2010) into time-dependent horizontal two-dimensional fields of estimated total lightning flash origin density. The lightning threat proxy fields used in the LFA are the upward graupel flux (GFX) in the mixed phase layer, approximated by the product of updraft speed and graupel mixing ratio in the layer with temperature of  $-15^{\circ}\text{C}$  and the total vertical ice integral (VII). Both proxies have been found in global observational studies (Petersen et al. 2005; Cecil et al. 2005; Deierling and Petersen 2008) to be strongly related to storm total flash rates, and are believed to have general validity, independent of geography, season, or weather regime; the observational study by Wiens et al. (2005) found both updraft and graupel volumes to be important.

Since its introduction, interest in the forecasting community has led to the incorporation of the LFA into a number of explicit-convection forecast models, allowing for more complete testing and analysis of its performance in varying seasons and locations. To date, the LFA has been added to the Weather Research and Forecasting (WRF; Skamarock et al. 2008) Model, and has been run daily at the National Severe Storms Laboratory (NSSL), in the WRF ensembles run each Spring by the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma (Xue et al. 2007), and in the operational hourly High-Resolution Rapid Refresh (HRRR; see Alexander et al. 2014) WRF runs disseminated by the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP). More recently, the LFA was also used in a 10-member ensemble of daily 48-h HRRR-like WRF forecasts featuring fixed physics but diverse initial conditions, run through late 2017 by the National Center for Atmospheric Research (NCAR; Schwartz et al. 2015).

The LFA's goal is to provide quantitatively calibrated estimates of gridded total flash rate density (FRD) that should, for any given regional convective outbreak event, yield peak FRDs in the strongest simulated storms that statistically match those found in observations of the strongest actual storms in the event, as measured by trusted, well-validated lightning observing systems,

such as the North Alabama Lightning Mapping Array (NALMA; Krehbiel et al. 2000; Goodman et al. 2005). The LFA emphasizes the strongest storms in simulated and observed outbreaks, not only because they tend to produce the most damaging weather, but also because experience indicates that the explicit convection models tend to predict the peak severity of the strongest storms more faithfully than they predict storm numbers, sizes and intensities at specific times and locations. For example, when designing the LFA, McCaul et al. (2009) found that cross correlations between peak FRDs in observed storm outbreaks and peak values of the LFA proxies in simulations of those outbreaks ranged up to approximately 0.8. On the other hand, instantaneous correlations between fields of observed and simulated storm signatures are often so negligible that nontraditional methods must be used (see Clark et al. 2014) to gauge forecast usefulness. The LFA's basic strategy is thus to ensure accuracy of lightning forecasts for the strongest storms predicted by contemporary CAMs, with the expectation that future models will eventually depict accurately all aspects of convective storms, once data assimilation methods and other modeling improvements allow.

The LFA also seeks to depict realistically the areal footprint of total lightning threat aloft in storm anvils by using the proxy fields in conjunction with lower-bound truncation thresholds that constrain the estimate of maximum threat areal coverage in an event, so that it matches statistically the observed maximum threat areal coverage. As a simple proxy-based diagnostic algorithm, the LFA can in theory be applied to any suitable explicit-convection model output that contains the needed kinematic, thermal, and microphysical fields.

In terms of spatial pattern, the LFA's GFX-based FRD fields resemble the VII-based fields, but GFX shows more compact spatial footprints and more FRD temporal variability. The temporal smoothness of VII is likely due to the effect of the integrations used in its computation. Thus, GFX and VII could both be properly calibrated, but GFX was more sensitive to storm updraft time variation. Such temporal updraft variability is important, because analyses of lightning "jumps" have exhibited utility in nowcasts of severe weather (Schultz et al. 2017), and to the extent the forecast models can accurately depict the time variations in storm intensity, it is desirable that products like the LFA also be able to reflect these properties. McCaul et al. (2016) found preliminary evidence that the LFA could accurately reproduce the amplitude and approximate frequency of occurrence of lightning jumps in WRF simulations of a major severe weather outbreak.

GFX thus offers some advantages over VII with respect to its ability to depict storm time variability. VII,

however, when properly calibrated, has the advantage of better depicting the commonly seen spread of total lightning threat into storm anvil regions, where explicit occurrences of flash origins are rare, and thus not often represented by GFX. To achieve an optimum overall lightning threat product, having both a realistic time variability but with a more realistic threat spatial footprint, McCaul et al. (2009) constructed a blended threat field, based on a simple weighted average of the co-calibrated GFX and VII fields, followed by application of a lower-bound threshold for diagnosed FRD to control the threat footprint. Assignment of weights of 0.95 to GFX and 0.05 to VII was found to preserve most of the desired temporal variability of lightning threat, while also producing a peak net spatial footprint roughly consistent with NALMA observations of the peak areal coverage of total lightning threat, as inferred from flash extent density (FXD; Murphy and Demetriades 2005; Lojou and Cummins 2005).

Although the LFA is based on a statistical treatment of lightning using known lightning physics, there are reasons to suspect that explicit full-electrification lightning schemes might ultimately prove to be more useful. Such schemes could, if configured for their full capabilities, yield information about lightning flash origin points, flash structure, polarity, peak current, energy and charge transfer, things that the LFA cannot easily do. However, it remains the case at the time of this writing, that rigorous explicit lightning discharge schemes remain costly, although E-WRF's use of a simplified lightning discharge geometry allows for operational use at only a 10%–15% run-time penalty (Fierro et al. 2013) compared to the LFA. Even with simplifications, however, the explicit schemes, despite greater precision, completeness, and detail, remain tied to the parent CAM and its susceptibility to general forecast errors associated with low predictability on the convective scale, just like the LFA.

The LFA uses two FRD calibration parameters, one for GFX (critically important) and another for VII (less important, as shown in the appendix), plus two FRD thresholds, one a lower-bound threshold for the GFX-based threat, to limit false alarms, and another for the final blended threat, to ensure good matching of threat areal coverage to FXD-based observations of actual lightning threat coverage. As discussed herein, users of the LFA should be aware of the possible need to check and perhaps modify two of these parameters, the GFX threat calibration factor that controls peak FRD, and the blended threat FRD threshold that controls blended threat areal coverage. Thus, the cost of the LFA's portability and simplicity is that users must be aware of the LFA's sensitivity to model physics, and be prepared to examine the need to modify those two scheme

parameters when applying the LFA in a customized model system differing from that used in McCaul et al. (2009). Use of the LFA in models having model meshes finer than 2 km may also require calibration checks, as only meshes sized from 2 to 4 km have been tested thus far. Although there is some sensitivity to model mesh size even in the 2–4-km range, anecdotal evidence thus far suggests that such sensitivities are much smaller than those associated with model physics. We are not familiar with any LFA model applications on meshes coarser than 4 km, but such use would not be recommended. The LFA, of course, is only to be used with explicit convection models, and not with larger-scale models using parameterized convection.

Comparisons have been attempted by some authors of the performance of the LFA versus operational versions of full-electrification lightning forecast schemes such as E-WRF (see, e.g., Fierro et al. 2013; Dafis et al. 2018), who found evidence of E-WRF's better skill scores with only a small penalty in computation time. Some of the results, however, should be viewed with caution, as only the original version of the LFA (as in McCaul et al. 2009) was used, and that version was found by McCaul et al. (2011, 2012) to be susceptible to false alarms, especially in cold-season systems. Use of the modified LFA proposed by McCaul et al. (2011, 2012), likely would curtail false alarms and improve the LFA skill score statistics. In fact, McCaul et al. (2011, 2012) found that the addition of the lower-bound GFX threshold of  $1.5 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$  greatly reduced the number of cold-season false alarms, and also led to consistent improvements of approximately 0.2 in Heidke skill scores across cold, warm, and transitional seasons alike.

The original LFA, as published in McCaul et al. (2009), was trained on a small but diverse set of observed and simulated convective storm events occurring near Huntsville, AL, carefully chosen to represent a wide range of storm intensities and types so that the statistical relations between simulated model proxies and observations of lightning could be reliably determined. All simulations in the original paper used identical model physics and microphysics configurations, but, in recognition of expected sensitivities of the results to changes in those parameterizations, it was recommended then that LFA calibrations be rechecked whenever significant changes are made to model physics parameterizations or grid mesh spacings. The purpose of this research is to explore the quantitative sensitivity of the LFA output to a set of commonly used model physics parameterizations, by examining the output of WRF Model forecasts using those varied parameterizations against output from a trusted "reference run." However, no attempt is made here to perform

intercomparisons of LFA performance versus any other lightning forecast scheme, a task that is beyond the scope of the present research. The data and methods used in studying the LFA physics scheme sensitivities are presented in section 2, with results given in section 3. In section 4, there is discussion of the findings and their limitations and implications. Section 5 contains a summary and prospects for the future. Additional details regarding the several basic improvements made to the LFA since its original publication in 2009 are given in the appendix.

## 2. Data and methodology

Although some insight might be gained from study of LFA output variability among the several competing regularly published LFA-equipped forecast model runs, such study is complicated by the diversity of many other details of the model configurations used. To study more systematically the sensitivity of the LFA to choices of commonly used model physics schemes, a daily ensemble of 12 WRF CAM forecasts generated in Spring 2018 under the auspices of a National Aeronautics and Space Administration (NASA) SERVIR Applied Sciences Team Project by personnel in Huntsville, Alabama, was examined. These ensemble runs covered the convectively active regions of Nepal, Bangladesh, and northeast India from March to May, during the peak premonsoon severe weather season. This underserved region of southern Asia is susceptible to intense lightning activity, which the SERVIR ensemble was designed to address, using a probability approach based on output from the physics-diverse ensemble members. To handle lightning threats, a simple but reliable and flexible lightning scheme was desired, and the LFA was the most portable scheme, and thus the most appropriate choice.

The SERVIR ensemble consisted of 12 WRF configurations, derived from four choices of microphysics schemes paired with three planetary boundary layer (PBL) schemes. All other aspects of the ensemble member simulations were held constant, except for the use of randomized initial and boundary conditions derived from members of the Global Ensemble Forecast System (GEFS) ensemble, to promote enhanced diversity of solutions. Although changing the model initialization method can introduce additional variability to the forecast solutions, it is our belief that the effects of these randomized initializations tend to fade into secondary importance compared to the more enduring and specific impacts of the microphysics and PBL choices, when averaged over an entire season (3 months of forecasts), as is done here.

One of the 12 members of the ensemble featured a model configuration using the WRF single-moment 6-class microphysics scheme (WSM6; Hong and Lim 2006) microphysics scheme, as in McCaul et al. (2009), paired with the Mellor–Yamada–Jancić (MYJ; Janjić 1994) PBL scheme. This combination is similar to that used regularly in the 4-km NSSL WRF forecasts and has gained widespread familiarity in the forecast community. It is thus considered to serve as a “reference” or benchmark result against which the LFA FRD estimates from the other 11 scheme configurations can be compared. The use of such a “reference” run was necessitated by the absence of total lightning observational data of LMA quality over the geographical study region. Our original project plan was to study the LFA’s performance in a version of WRF that used HRRR two-moment microphysics in quasi-operational forecasts across the continental United States (CONUS), employing high-quality ground-truth lightning data for validation, but this had to be scrapped and redesigned when unexpected and irreparable problems were found in some of the LFA-related model fields. We then decided to study the NASA/SERVIR ensemble results, which, although lacking in high-quality ground-truth lightning observations, provided 12 WRF Model configurations for intercomparison, including one run similar to the HRRR, the one run similar to the NSSL runs that served as the “reference,” and 10 other runs as a bonus.

In this paper we compare the LFA peak FRD metrics from the reference run with those from each of the other 11 ensemble members, and estimate multiplicative recalibration factors appropriate for those 11 other nonreference WRF configurations, the use of which would make their peak FRD results match most closely those of the benchmark MYJ-WSM6 reference configuration. We make no claim here of any inherent superiority of WSM6—most microphysics schemes exhibit idiosyncratic tendencies—but it happens that WSM6 was a reasonable choice to use when the original LFA was being designed.

We also compare the lightning threat areal coverages of the ensemble members and form analogous ratios of reference areal coverage to that of each other scheme in order to estimate recalibrated lower-bound threat thresholds for each alternate scheme. In this case, the ratio must be used as a divisor; however, because a ratio in excess of unity implies a need to boost threat areal coverage, which is accomplished by reducing the blended threat threshold for that scheme. The details of these recalibration procedures are described below.

Although it is not feasible to make an exhaustive study of all possible configurations of choices of the

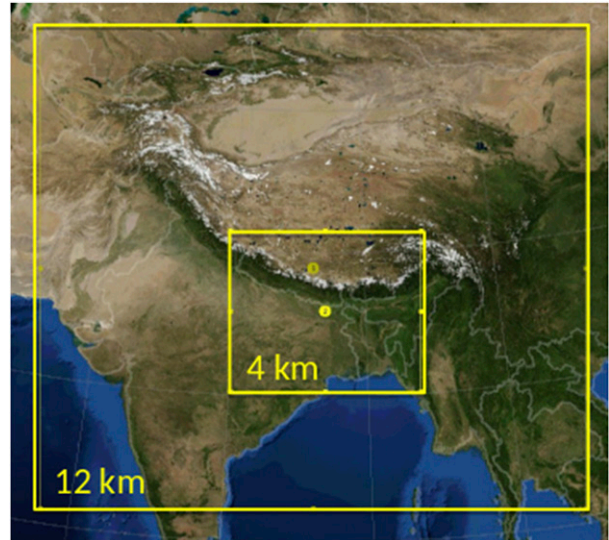


FIG. 1. Map of forecast simulation domains used for this study. The outer domain consisted of a  $12\text{ km} \times 12\text{ km}$  mesh, while the inner nested domain used a  $4\text{ km} \times 4\text{ km}$  mesh.

ever-evolving model physics schemes and their impact on LFA performance, the present findings will at least offer insight into the LFA’s overall sensitivities, provide understanding of the possible range of errors on both a seasonal-term and also a day-by-day basis, and show how basic recalibrations might be performed on any prospective model configuration.

The 12-member ensembles of WRF simulations were executed over a domain covering Nepal, Bangladesh, northeastern India, and the Himalayas (see Fig. 1), for the spring and summer of 2018. Of the 92 days studied here between 1 March and 31 May 2018, convective storm events provided useful data on 73 days; the 19 excluded days either produced no convection, or convection that was too weak to trigger LFA forecasts of lightning threat, for one or more physics schemes.

The model was run on a  $4\text{ km} \times 4\text{ km}$  nested mesh, with 42 terrain-following levels in the vertical (Case et al. 2018). Forecasts began at 1800 UTC each day and covered 48-h periods using a 20-s time step on the nested mesh. As mentioned earlier, four separate microphysics schemes were paired with three separate PBL schemes to give 12-member physics-diversity ensembles for each day. The schemes used for PBL were

- 1) Yonsei University (YSU; Hong et al. 2006);
- 2) Mellor–Yamada–Jancić (MYJ; Janjić 1994); and
- 3) Mellor–Yamada–Nakanishi–Niino; (MYNN; Nakanishi and Niino 2006).

The microphysics schemes were

- 1) Morrison (Morr; Morrison et al. 2009).

- 2) WRF single moment 6-class microphysics scheme (WSM6; [Hong and Lim 2006](#));
- 3) Thompson (Thom; [Thompson et al. 2008](#)); and
- 4) Goddard (Godd; [Lang et al. 2007](#); [Lang et al. 2011](#)).

Initialization was accomplished in cold-start mode, using an arbitrary but consistent set of 12 members of the randomized Global Ensemble Forecast System (GEFS; [Zhou et al. 2017](#)) data. To eliminate model spinup effects, and to focus unambiguously and without redundancy on the local daily convective cycle, analyses were limited to forecast times from 6 to 30 h inclusive. Model output fields were saved every 1 h, including the hourly fields of 1-h maxima ([Kain et al. 2010](#)), where available. Of the 73 active storm days sampled, some of the storm events were quite intense, with updrafts as strong as  $52 \text{ m s}^{-1}$  noted on the  $4 \text{ km} \times 4 \text{ km}$  mesh.

Lightning event metrics were determined as follows. For each convective event  $n$  ranging from 1 to 73, there were three PBL options, with  $i$  ranging from 1 to 3, and four microphysics options, with  $j$  ranging from 1 to 4, as mentioned earlier. For each  $n$  and each physics option  $i$  and  $j$ , the LFA FRD hourly maxima were found for each hour from 6 to 30, and the largest of these values was assigned as  $F_{ijn}$ . LFA FRD output from the MYJ-WSM6 run was taken as reference run data for each convective event. Since MYJ data were assigned as  $i = 2$ , and WSM6 data as  $j = 3$ , the reference FRD value for the  $n$ th event was taken as  $F_{23n}$ . Scatterplots were then made for each  $ij$  physics combination, each containing all 73 useful events, with  $F_{23n}$  plotted along the  $y$  axis and  $F_{ijn}$  along the  $x$  axis. Best-fit linear relationships passing through the origin (as in [McCaul et al. 2009](#)), were then evaluated and drawn on the scatterplots, in the functional form  $F_{ij}(x) = a_{ij}x$ , where the slopes  $a_{ij}$  are determined by least squares fitting of origin-based straight lines through the 73 available data points having FRD values  $x = F_{ijn}$ . Note that for each datum, there is an error (actually a discrepancy)  $e_{ijn} = F_{23n} - a_{ij}F_{ijn}$ . If none of the  $ij$  experiments showed any bias over the 73 data points relative to  $F_{23n}$ , then all slopes  $a_{ij}$  would be close to unity. Physics scheme combinations that produce, say, smaller LFA flash rate densities  $F_{ijn}$  than the desired  $F_{23n}$ , would feature slopes greater than unity. The slopes shown on each of the  $ij$  scatterplots thus may be taken as recalibration factors that should be applied multiplicatively to each  $ij$  scheme combination's  $F_{ijn}$  FRD data so as to yield results that are a statistical match to the "reference" value  $F_{23n}$ .

Of course, the natural variability of simulated convection produces scatterplots with most points showing differences from the best-fit linear function. To estimate quantitatively how much these variations differ from the

reference run, we calculated the rms errors using the  $e_{ij}$  of all the 73 points in each  $ij$  physics plot, and also the standard deviations of the off-diagonal differences  $F_{23n} - F_{ijn}$  and then of the  $F_{ijn}$  alone. Armed with this information, the magnitudes of the differences in slope "SL\_SIG" needed to allow conclusion that those slope differences met 95% confidence criteria in a  $t$  test were tallied (see, e.g., [Wilks 2011](#)). The slopes  $a_{ij}$  and the slope differences SL\_SIG needed for statistical significance are printed for easy inspection on each scatterplot. Note that we have included the  $ij = 23$  plot for completeness, even though as expected, it shows a slope  $a_{23}$  of unity and zero SL\_SIG difference.

While the best-fit slopes  $a_{ij}$  exhibit variation from panel to panel, it is also worthwhile examining the intrapanel slopes of individual daily data points from the origin. This reveals the range of variability of LFA calibration accuracy from one storm day to another, resulting from the effects of day-to-day fidelity of the overall forecasts. To investigate this, we have calculated the slopes of every daily datum on each of the  $ij$  experiment scatterplots, and mention in the results section below some of the more salient findings from that exercise.

The recalibration of a scheme to improve its accuracy within a specific model configuration requires consideration of not only peak flash rate density, but also lightning threat peak areal coverage. This areal coverage assessment has many similarities to that described above for recalibration of peak FRD, but also some differences. In the specific case of areal coverage, we first build scatterplots of areal fractions from the reference scheme versus those of the alternate schemes and determine best-fit slopes  $b_{ij}$ . In this case, however, the slopes  $b_{ij}$  are not used multiplicatively to boost or trim the scheme's areal fraction back to that of the reference scheme, since in the LFA the ultimate areal coverage is determined by an FRD threshold, which must be lowered (raised) to achieve an increase (decrease) in threat area. Thus, to arrive at a recalibrated FRD threshold that provides the desired threat areal coverage, the reference scheme's threshold,  $b_{23} = 0.072 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$ , is divided by the factor  $b_{ij}$ . By design, the LFA threat areal footprint is strongly influenced by VII and is thus quite sensitive to cloud microphysics. This fact, in concert with the specific topography of the VII field in each storm anvil, makes it unlikely that the revised areal coverage will increase by the exact desired factor  $b_{ij}$ . The magnitude of the discrepancy is, however, small, as discussed below, so that our procedure is actually quite effective.

The amplitude and pattern of variations in the behavior of the peak FRD slopes  $a_{ij}$  and peak areal coverage slopes

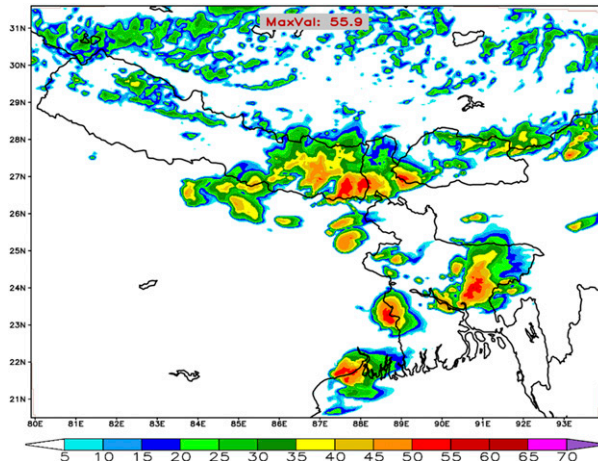


FIG. 2. Instantaneous forecast map at 1100 UTC 30 Mar 2018 of simulated composite radar reflectivity (dBZ). Calculated peak reflectivity at this time was nearly 57 dBZ, while peak updrafts simulated at 1100 UTC for this storm complex exceeded  $39 \text{ m s}^{-1}$ .

$b_{ij}$ , and their associated critical slope differences, are thus the main subjects of this paper, and are described and discussed below.

### 3. Results

For context, we provide Fig. 2, which illustrates the simulated composite reflectivity field for an intense convective storm system that was successfully forecast to occur on 30 March 2018. Figure 3 provides the corresponding LFA FRD field for this event at the same time. This storm, fed by convective available potential energy of at least  $4000 \text{ J kg}^{-1}$ , achieved simulated peak updraft speeds (not shown) at 1100 UTC 30 March 2018 of  $39 \text{ m s}^{-1}$ , and also produced large hail, damaging winds, and lightning that caused seven fatalities and additional casualties. Peak composite reflectivities simulated at 1100 UTC were close to 57 dBZ. The peak LFA-prognosed FRD values in Fig. 3 exceeded  $14 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$ , which, based on experience, is a value large enough to suggest a high potential for severe weather.

In Fig. 4, we present the 12 panels of scatterplots of the 73 stormy days for each of the 12 model physics configurations. Each of the three columns features one PBL scheme, and each row one microphysics scheme, as marked. In what follows, our subscripts for PBL schemes feature index  $i$  increasing to the right, while for microphysics, the index  $j$  increases downward. From Fig. 4 we observe that, while many of the best-fit slopes are close to unity, some achieve values greater than 1.10, reaching an extreme of 1.21 for  $a_{ij} = a_{33}$ . It is seen that only two experiments exhibit FRD slopes differing from

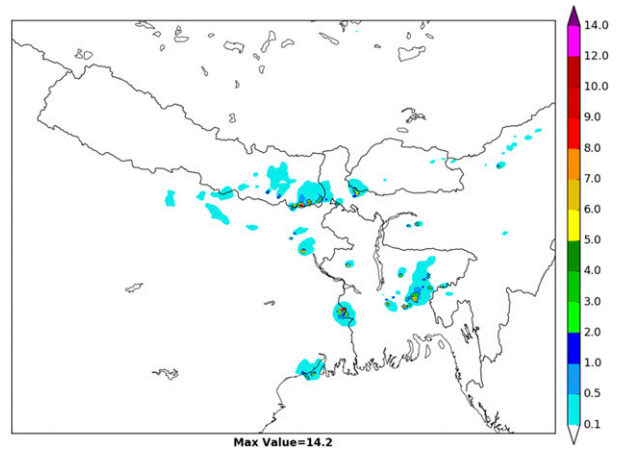


FIG. 3. Instantaneous forecast map at 1100 UTC 30 March 2018 of LFA-derived flash rate density [ $\text{flashes km}^{-2} (5 \text{ min})^{-1}$ ]. Peak flash rate densities exceeded  $14 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$ , suggestive of a high probability of severe weather.

the benchmark value of unity by amounts greater than their SL\_SIG quantity, and thus can be said to have calibration factors significantly different than the benchmark run. Of particular interest is the fact that all slopes  $a_{3j}$  are above 1.10, showing that the MYNN PBL scheme is vulnerable to larger bias errors than the other PBL schemes. In fact, both of the experiments featuring statistically significant differences from the benchmark run use MYNN PBL physics. Inspection of the slope values for the microphysics schemes indicates that the Thompson scheme tends to have slightly larger slopes  $a_{i2}$  than the others (with the exception of WSM6 itself), but even these do not differ from the  $a_{23}$  benchmark value of unity by more than the SL\_SIG values needed to qualify as statistically significant differences. Note that both the Thompson and MYNN schemes are employed in the operational HRRR model.

There is also a subtle pattern in the scatter for the Thompson microphysics scheme, which consists of enhanced scatter at both low and high values of  $F_{ijn}$ . This scatter shows a tendency toward reduced values of  $F_{ijn}$  for weak lightning events, and enhanced values for strong lightning events. This possible bias tendency of the Thompson scheme at low flash rates also seems to promote the largest day-to-day “recalibration” errors seen in the data, with one point’s slope value reaching a magnitude of 11.0, indicating an approximate order-of-magnitude underforecast of FRD versus the reference scheme. This sort of occasional error is perhaps not surprising, in view of the inherently low predictability in forecasts of convective-scale phenomena.

By analogy with Fig. 4, we present in Fig. 5 the scatterplots of LFA threat areal coverage and their best-fit

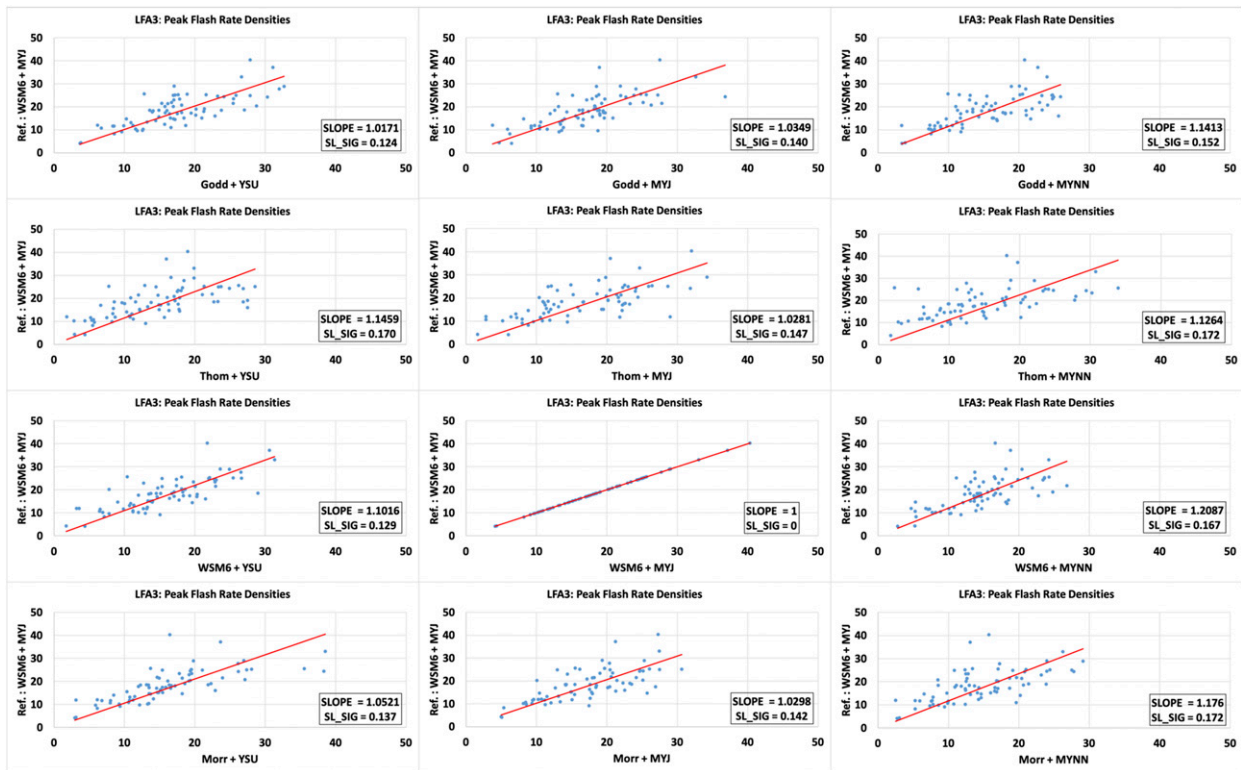


FIG. 4. Scatterplots of values of peak FRD from LFA blended lightning threat (LFA3) within “reference” MYJ-WSM6 run (plotted on y axis in all 12 panels), vs the same threat from the other 11 model configurations (plotted on x axis in each panel). Each column features one of three PBL schemes, as noted in the panel labels; each row features one of four microphysics schemes, also shown in the panel labels. Scheme abbreviated titles are as listed in the text. On each panel, the best-fit slope and critical deviation from unity (SL\_SIG) needed to infer statistically significant differences from the reference scheme with 95% confidence are marked for convenience. Note that panel  $(i, j) = (2, 3)$  is a plot of MYJ-WSM6 results against itself, so that the slope is unity and SL\_SIG is zero. Units are flashes  $\text{km}^{-2} (5 \text{ min})^{-1}$ . Note the tendency for larger best-fit slopes for the Thompson “Thom” microphysics and MYNN PBL schemes, suggesting they produce less graupel than the reference schemes.

curves for each of the ensemble members. As in Fig. 4, the slopes for each scheme configuration indicate the factor by which the diagnosed areal threat coverage fraction needs to be boosted to match that of the reference run. Also shown is the fractional factor SL\_SIG by which a given scheme’s diagnosed threat area slope needs to exceed or fall short of the reference run’s unity value to qualify as having a slope difference that is statistically significant.

In Fig. 5, the findings reveal different patterns than those seen in Fig. 4. In particular, all WSM6 schemes are associated with slopes near unity, but the other schemes show evidence of sizeable deviations from that standard. The largest deviations occur with the Thompson microphysics scheme, where the slopes are consistently near 2.0, with only relatively small modulations by the PBL schemes. This indicates a strong tendency for the Thompson scheme to produce less ice in storm anvils than the WSM6-based reference model, leading to a corresponding deficit of about half in VII-based threat

areal coverage. On the other hand, for both the Morrison and Goddard microphysics schemes, the best-fit area recalibration factor slopes range from about 0.5 to 0.7, indicating the need to shrink areal coverages inferred from those microphysics schemes, which evidently produce more anvil ice than WSM6. In general, Fig. 5 shows that areal coverage is much more sensitive to microphysics than PBL physics.

If recalibration of peak FRD output from one of the 11 studied schemes involves simple multiplication of output FRD amplitudes by the slopes seen in Fig. 4, the recalibration procedure for adjustment of peak areal coverage back to reference values is almost as simple but less precise. For any given best-fit slope diagnosed for one of the 11 scheme configurations in Fig. 5, the task is to boost (shrink) the diagnosed threat areal coverage by a factor equal to that slope. To do this requires a decrease (boost) in the lower-bound critical FRD threshold, and the simplest approach is to divide (rather than multiply) the basic critical threshold—given in the appendix



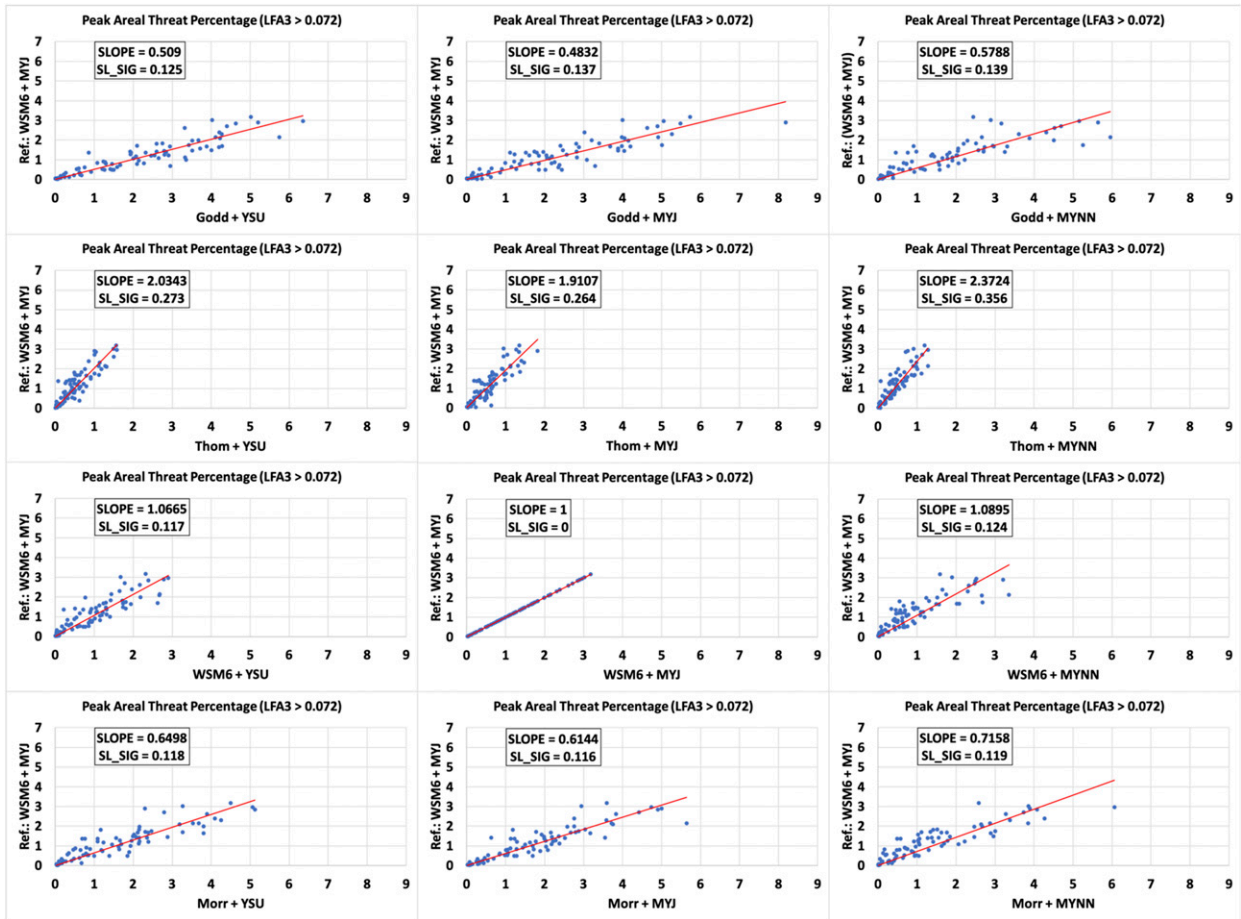


FIG. 5. As in Fig. 4, but for values of LFA peak areal coverage fraction (see text) of blended lightning threat (LFA3) from “reference” MYJ-WSM6 run (plotted on y axis in all 12 panels), vs the same threat from the other 11 model configurations (plotted on x axis in each panel). Scheme abbreviated titles are as listed in the text. On each panel, the best-fit slope and critical deviation from unity (SL\_SIG) needed to infer statistically significant differences from the reference scheme with 95% confidence are marked for convenience. As in Fig. 4, panel (i, j) = (2, 3) is a plot of MYJ-WSM6 results against itself, so that the slope is unity and SL\_SIG is zero. Units are percentages. Note the tendency for much larger best-fit slopes for the three experiments using the Thompson “Thom” microphysics scheme, suggesting it produces less anvil ice than the reference schemes.

as  $0.072 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$ —by the best-fit slope from Fig. 5. Such a decrease (boost) in critical threshold will produce a larger (smaller) diagnosed threat areal footprint.

While we propose using the slopes from areal coverage scatterplots in Fig. 5 as divisors into the baseline critical FRD in order to perform recalibration for areal coverage, it is acknowledged that such division does not always guarantee new output having precisely correct revised areal coverages. This is because the precise fall-off of anvil ice hydrometeor amounts and associated VII with distance from storm updraft cores may not always be simple and well behaved. Tests have been performed, however, on the most sensitive scheme (Thompson), using the proposed method of dividing the threshold critical FRD by the slopes in Fig. 5, and

these tests show that the resulting new threat areas are, on average, generally within a few percent of the values suggested as being needed by the slopes in the figure. We thus conclude that our simple proposed recalibration method for areal coverage does indeed yield satisfactory results.

The large low bias and also enhanced variability of peak threat area in the Thompson microphysics environment shown in Fig. 5 are seen mainly when examining the full peak threat area defined by the LFA lower-bound critical FRD threshold. However, storm-core threat areas reveal less scheme-to-scheme variability when a larger FRD threshold of, say,  $5.0 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$ , is examined (not shown). This suggests that most microphysics schemes yield a blend of large ice hydrometeors in storm cores that is more consistent, at

least for quantities used in the LFA, than the ice species in outer anvils.

#### 4. Discussion

The results show that both the MYNN PBL and Thompson microphysics schemes tend to produce LFA FRD results that differ from the reference MYJ-WSM6 model configuration more than most model scheme configurations tested, although the size of the bias errors is only about 10%–20% when averaged over a 3-month convective season. All 11 alternate model configurations gave smaller peak FRD values than the original MYJ-WSM6 scheme, but only two experiments, both using MYNN PBL physics, showed differences from the reference scheme that were large enough to be judged statistically significant.

MYNN and Thompson both underforecast peak FRD relative to the reference experiment, but Thompson had the additional problem of showing larger threat areal coverage deficits. The threat areal coverage differences among the tested microphysics schemes were found to be statistically significant, with the differences likely attributable to variations in the amounts of anvil ice produced by the various schemes. The reference scheme involving MYJ-WSM6 occupies middle ground in terms of handling threat areal coverage, while Thompson produced threat areas too small, with Morrison and Goddard too large. The effects of the PBL schemes on areal coverage were less important than microphysics.

In summary, recalibration of peak FRD can be accomplished by multiplying the various schemes' FRD outputs by the slopes shown in Fig. 4, although the fact that so few scheme combinations showed significant differences from the reference run suggests there is no urgent need to implement the FRD recalibration. To recalibrate peak areal coverages, the LFA default threshold of 0.072 must be divided by the slopes shown in Fig. 5. The areal coverage sensitivities, unlike those for peak FRD, were large enough to warrant recalibration in forecast applications that require accurate forecasts of lightning threat areal coverage. Although our suggested areal threat recalibration method does not guarantee precise statistical matching of areal coverages, any resulting areal coverage errors appear to be small enough to be ignored in practice. As with all lightning forecast schemes, error budgets are generally dominated by the overall day-to-day performance of the parent convection model.

The Thompson microphysics scheme also tended to underforecast lightning in weak storms and overforecast it in a few very strong storms. This pattern could signal the existence of a possible nonlinearity in the relationship

between actual observed peak FRD and the FRD proxy fields used by the LFA, or the need to add an offset to the regression line in the scatterplot. The latter option implies in turn the possibility of a need to reduce the lower GFX threshold of  $1.5 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$  (see the appendix), in the Thompson scheme, in order to boost GFX-based threat at the very low end of the storm intensity spectrum. Further investigation is needed to determine if nonlinear calibration of the LFA is actually required when the Thompson scheme is used.

The large sensitivity of threat areal coverage to model microphysics offers useful insight into the specific partitioning of hydrometeor types by each scheme, and thus poses challenges for diagnosis of lightning threat extension into storm anvils using simulated ice hydrometeor fields alone. The sensitivities seen herein are not unknown; they are in fact reminiscent of those discussed in Gilmore et al. (2004) and Liu and Moncrieff (2007). Ultimately, the need for recalibration of threat areas back to a credible reference seems to be unavoidable for statistically based lightning schemes such as the LFA.

There is at present insufficient information to say whether the Thompson scheme's peak FRD performance idiosyncrasies might match LMA-quality lightning ground-truth observations better than the reference WSM6 scheme. As shown in the appendix, one major change made to the LFA in 2011 was to revise the algorithm so as to let the sensitive field of graupel flux GFX dictate the quantitative peak FRD estimate (to whose peak amplitude both proxies are forced by rescaling to conform), rather than the smoother field of vertical ice integral VII. The main reasons this was done included making the LFA more responsive to very high flash rate storm events and providing more realistic temporal response of the LFA to storm updraft intensity changes, with both of which VII had difficulty because of its integral nature. It is not easy to validate the LFA on storm events having very high flash rates, not only because of the rarity of such events, but also because of the inherent observational difficulty of distinguishing individual flashes in storms producing large numbers of flashes that overlap one another in close time and space proximity.

#### 5. Summary and future prospects

The results shown here indicate that the LFA needs only relatively modest adjustment of its basic linear proportionality factor relating the GFX proxy to total lightning FRD—by no more than about 20% or so—in order to achieve a good match to the peak FRD output from existing well-calibrated model versions. Most of the tested model configurations did not exhibit calibration

differences relative to the reference scheme that could be called statistically significant. We leave it to the forecast community to decide whether the relatively small improvements in LFA FRD output described here warrant LFA code modifications. Even if FRD recalibrations are made, however, LFA users can expect that the quantitative accuracy of peak FRD results will still vary on a day-to-day basis, owing to the limitations of model physical parameterizations, initialization data and procedures, and model numerics, all of which produce inherent uncertainty in model forecasts of small-scale convective phenomena. Particular findings of interest here are that the MYNN PBL scheme and the Thompson microphysics scheme—both employed in the operational HRRR forecast model—feature relatively large FRD bias errors relative to the reference MYJ-WSM6 scheme; in addition, the HRRR combination occasionally underpredicts for very weak storms and overpredicts for very strong ones.

The LFA lightning threat areal coverages turned out to be more sensitive to model physics than did the FRD output. The areal coverages varied in a statistically significant manner for each of the microphysics schemes, with lesser impact from PBL physics. Some microphysics schemes yielded threat areal coverages larger than the default WSM6 choice, while others yielded less. It was found that the Thompson scheme showed the largest discrepancies, underpredicting lightning threat areal coverage by about half for all tested PBL schemes, suggesting a need for an approximate 50% reduction in the LFA FRD threshold that controls threat areal coverage, from 0.072 down to about 0.036 flashes  $\text{km}^{-2}$  ( $5 \text{ min}^{-1}$ ), for contemporary versions of that microphysics scheme only. Increases of the threshold by somewhat smaller factors were indicated for the Morrison and Goddard schemes.

Additional observational studies that allow direct comparison of HRRR LFA results with reliable lightning observations are needed to see whether the HRRR physics configuration actually performs better than the original WSM6-based design; new space-borne lightning observing systems such as the Geostationary Lightning Mapper (GLM; Goodman et al. 2013), along with expansion of high-quality ground-based lightning detection systems, should ultimately provide many new opportunities for such validation work.

*Acknowledgments.* This research was funded by a Risk Reduction grant from the National Oceanic and Atmospheric Administration to Project 68, in preparation for the launch of the Geostationary Lightning Mapper (GLM) on the GOES-R (now *GOES-16*) satellite platform, with extended funding from NASA.

Additional support for TC and GP was provided by NASA Cooperative Agreement NNM11AA01A, and for JC, by NASA/SERVIR Applied Sciences Team funding. We would also like to acknowledge NASA's SERVIR program for funding the ensemble modeling simulations used in this study. The authors also thank Drs. William Koshak and Richard Blakeslee of NASA and Dan Lindsey of NOAA for their ongoing interest and support. SG's contributions to the review and refinement of the manuscript were supported by NASA Grant 80NSSC18K1689. We also wish to thank the anonymous AMS reviewers for their insightful and constructive critiques, which helped us improve the paper substantially. Datasets used in this research are, because of their nature, voluminous, but may be obtained by making arrangements with the corresponding author.

## APPENDIX

### Operational Changes to the LFA

Since its original publication (see McCaul et al. 2009), the increased use of the LFA in various research and operational models has offered opportunities to examine its performance in real-world circumstances, and to uncover ways to make it more robust. These real-world tests have indeed led to a few changes in the details of the LFA, as described below.

Results compiled from the 2010–11 daily WRF runs by NSSL and 2011 runs by CAPS were examined and used to make revisions to the LFA in 2012. Emphasis shifted to the performance of the LFA over larger regions and on year-round time scales, especially on previously unexamined cold-season nonconvective weather scenarios and on rare but important very high-flash rate convective storm events not available for earlier study. Preliminary inspection of the 2010–11 LFA output suggested weaknesses in the original LFA in these latter two situations (see McCaul et al. 2011).

The use of the LFA on large CONUS-scale forecast grids dictated a simple change in the handling of FRD values on a latitude–longitude mesh. In the original design study, only a small region surrounding the Tennessee Valley was simulated, and for such small domains, only a single center-of-grid latitude with its own map factor was used in the conversion of gridcell values of lightning flash rates to flash rate densities. However, to ensure the LFA FRD calculations were independent of latitude on CONUS-scale grids and could be used anywhere, latitude-dependent map-scale factors were introduced starting in 2011 in all LFA versions distributed to operational and research centers such as NSSL, CAPS, NCAR, NOAA, and NASA. This step

guaranteed that FRD values would always be given in standardized units [i.e., flashes  $\text{km}^{-2}$  ( $5 \text{ min})^{-1}$ ].

A more significant change to the LFA following operational experience dealt with prioritizing the two main proxy fields, GFX and VII, before the application of weighted averaging in the blending step. This aspect was reexamined when results from the original LFA were found to underestimate FRD peaks in very high-flash rate storms, because reliance on VII to control the FRD values failed to handle very large FRD situations. An overriding concern here is that, when performing a weighted average of two correlated fields that are both designed statistically to have similar peak magnitudes, it is necessary to ensure both fields do in fact have the same peak magnitudes in order to guarantee that the blended field resulting from the weighted averaging also inherits that same peak magnitude. Although both GFX and VII are calibrated independently to produce statistically similar peak values, local statistical variations of fields sometimes allow mismatched peak values. In such cases, it is necessary to rescale one of the fields prior to the averaging, so that the final blended field has the desired peak amplitude. The only question is: Which proxy field should have priority in the definition of the diagnosed peak FRD amplitude?

Initially there were concerns that the coarser 4-km grid mesh used in the NSSL WRF runs would reduce the GFX-based threat values, owing to the likelihood of smaller updraft values being simulated compared to what might have been found on a 2-km mesh. On the other hand, it was suspected that the VII field would be less sensitive to this change of grid mesh. Thus, the original version of the LFA used in the NSSL WRF runs included a provision to forcibly rescale the GFX field by the ratio of the peak VII value to the peak GFX value, in the hopes that this would mitigate any loss in amplitude that might be suffered by GFX in simulations on a coarser mesh. While this approach was used in the early applications to the NSSL WRF, other factors relating to LFA performance in the quasi-operational environment were ultimately found to be of overriding importance, as discussed below. The end result is that the most recent configurations of the LFA now require a forced rescaling of the VII field so that it matches the GFX field at the time-dependent location of the latter's peak value. This rescaling is performed every hour during a typical WRF Model run. The LFA thus modified was added to both the NSSL and CAPS models starting with the 2010–11 period, without any other customization.

In all NSSL runs, to ensure the actual peak values of NSSL WRF LFA-derived FRD were recorded for analysis and comparison with LMA 5-min data, hourly fields of the hourly maximum values of the simulated

GFX, VII and blended lightning FRD estimates were saved and examined, as in [Kain et al. \(2010\)](#). Thus although the peak FRD data from the WRF LFA and the LMA networks did not have exactly the same space or time resolutions, the coarser space resolution (2–4 km versus 1 km) but finer time resolution (24 s versus 5 min) of the LFA data tended to counteract each other to yield LFA values that compared favorably to those from LMA. The calibration process of comparing simulated LFA proxy data to observed LMA data, each on their own time and space grids, was allowed to handle the final step of quantitatively relating the simulated and observed data at resolutions conveniently obtainable from each datastream. The end result was that the NSSL WRF LFA provided quantitative guidance based on the model 4-km gridded data that compared well statistically with LMA FRD observations of the same events on an imposed 1-km grid.

Operational use of the LFA also revealed a tendency for the algorithm to overestimate lightning FRD in cold-season storm systems. This bias was also found to come mostly from the VII field, which often exhibited significant magnitude even in the absence of actual convection. The GFX field was less prone to show this bias, although some susceptibility occurred in forecast simulations that erroneously assigned too much convective character to winter precipitation systems.

As NSSL WRF output were available daily on an ongoing basis, LFA statistics were accordingly studied year-round to examine its robustness of performance across the seasons. The seasonal results appeared to be most clearly stratified in terms of three categories: cool-season, transitional season, and warm season. In each of the wide-ranging weather patterns encountered within these categories, some of which were nonconvective, distinct LFA bias and skill patterns emerged. These error patterns were not evident in the original LFA study of [McCaul et al. \(2009\)](#), which was restricted only to a relatively small set of weather events dominated by deep convection of varying types and intensities, and lacked null events and events strongly contaminated with cold stratiform clouds.

In the cool-season forecasts from the NSSL WRF, a sizable number of false alarm lightning threat events (days) were diagnosed from the original LFA scheme, which employed a lower-bound GFX FRD threshold of only  $0.01 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$  to eliminate spurious noise. A contingency table based on LFA output from the two winter periods January–March 2010 along with December 2010–March 2011 (not shown, see Fig. A2 in [McCaul et al. 2011](#)) revealed that 40 days of false alarm events were found in those cool-season months. In studying these cool-season cases, care was taken to

exclude the occasional anomalous warm convective events. The preponderance of cases was “true nulls,” although there was a single “hit” consisting of a rare winter thundersnow event in the Tennessee Valley area on 10 January 2011, which was successfully prognosed by the WRF and LFA. There were no “miss” events in these two winters. Based on these cool-season results, it was suspected that most of the false alarms were caused by too-large amounts of VII and occasionally GFX within typical cool-season stratiform cloud systems, the latter of which were able to override the small GFX thresholds in such systems during the weighted averaging step that creates the blended lightning threat. Thus it was thought that it might be helpful to boost the FRD threshold in GFX to a larger value that would eliminate most of the false alarms without sacrificing the thundersnow hit, or other true convective hits in other seasons.

To study this issue further, the contingency table of LFA daily event output for the warm-season months of June–August in both 2010 and 2011 was then consulted. For the LFA configured with the original small lower-bound GFX FRD of  $0.01 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$ , the results were found to be dominated by hits (169), with only 11 false alarms. There was one true null event, and, as in the cool-season study, no misses. The fact that there are no misses in both the cool season and warm season periods suggests that WRF is somewhat over aggressive in its forecasts of deep convection.

Based on study of the FRD values in these two years of daily events, it was apparent that the GFX lower-bound FRD threshold could be raised to  $1.50 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$  without losing true convective lightning threats, while retaining the correct prognosis for the single winter thundersnow event. Reassessment of the cool-season contingency table using this higher threshold showed that the number of false alarm cases was reduced 85% down to 6, without diminishing the numbers of hits and true nulls. Meanwhile, application of the higher FRD threshold to the warm-season data yielded an additional reduction of false alarm cases from 11 down to 9. There was an insufficient number of true winter thundersnow events to be able to determine if a revised FRD threshold other than  $1.50 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$  might have improved these contingency statistics further, while still identifying thundersnow events satisfactorily. No attempt was made to try to reduce the false alarm rate by more than the 85% already achieved, as it was believed that the remaining false alarms were likely the result of the inevitable occasional subpar model forecasts, and that further attempts to reduce false alarms might compromise the existing high probability of detection. In fact, many of those remaining

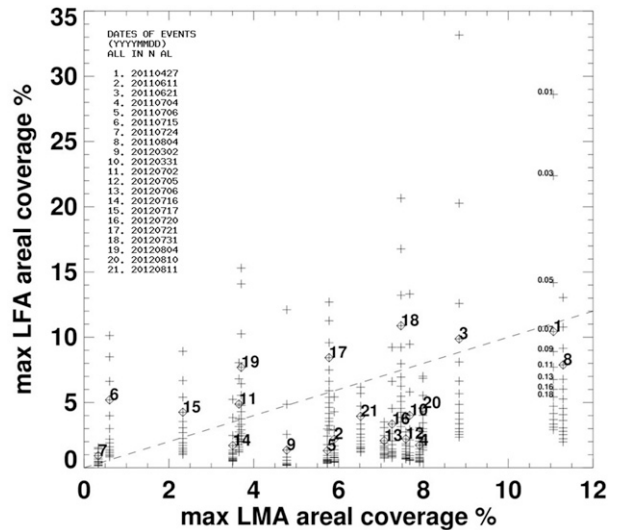


FIG. A1. Scatterplot of simulated (y axis) peak areal footprint fractions of LFA-diagnosed lightning threat, for a series of assumed lower-bound FRD threat thresholds (denoted by columns of + symbols), vs actual LMA-observed peak areal footprints (x axis) derived from flash extent density data. The magnitudes of threat FRD threshold are shown for a sample event along the right side of the plot. Dates of the storm events studied and their assigned event numbers are provided on the figure; all events were studied using observations from the North Alabama Lightning Mapping Array. Dashed line indicates the true diagonal, which intercepts the various threshold columns at a best-fit value of  $0.072 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$ .

winter false alarms appeared to be WRF forecasts of sleet, most of which constituted localized erroneous forecasts of precipitation hydrometeor type.

Note that a lower-bound threshold of  $0.4 \text{ flashes km}^{-2} (5 \text{ min})^{-1}$  has always been applied to the VII field. However, as shown below, this limit is found to play only a minor role, as the small threat values it allows in the blending step are later eliminated by another overall FRD threshold applied during the blending step, which is described in the last paragraph below and in Fig. A1. Thus, no further discussion of this early VII threshold is needed.

Standard skill statistics scores for the low-FRD and higher-FRD threshold versions of the LFA have been provided in McCaul et al. (2012). For both cool season and warm season events, improvements in both totals skill scores (TSS) and Heidke skill scores (HSS), and reductions in bias errors were evident. HSS values generally improved by approximately 0.20 using the new higher GFX lower-bound threshold to curtail false alarms.

To estimate the optimum blended threat FRD threshold needed to produce the peak LFA threat areal coverage that best matches LMA observations of peak

TABLE A1. Equations and history of LFA parameter values. Note that  $F_1 = a_1 \text{GFX}$ ;  $F_1 > G_1$ , otherwise 0;  $F_2 = a_2 \text{VII}$ ;  $F_2 > G_2$ , otherwise 0; and  $F_3 = w_1 F_1 + (1 - w_1) F_2$ ;  $F_3 > G_3$ , otherwise 0. The units of FRD are flashes  $\text{km}^{-2} (5 \text{ min})^{-1}$ . M09 is McCaul et al. (2009), M11 is McCaul et al. (2011), and M14 is McCaul et al. (2014).

Parameter (units)	Value	Date	Reference
$a_1$ {FRD [(g kg <sup>-1</sup> ) (m s <sup>-1</sup> )] <sup>-1</sup> }	0.0420	2009	M09
$a_2$ {FRD (kg m <sup>-2</sup> ) <sup>-1</sup> }	0.2000	2009	M09
$G_1$ (FRD)	0.0100	2009	M09
$G_2$ (FRD)	0.4000	2009	M09
$w_1$ (dimensionless)	0.9500	2009	M09
$a_1$ {FRD [(g kg <sup>-1</sup> ) (m s <sup>-1</sup> )] <sup>-1</sup> }	0.0512	2011	M11
$a_2$ {FRD (kg m <sup>-2</sup> ) <sup>-1</sup> }	0.2440	2011	M11
$G_1$ (FRD)	1.5000	2011	M11
$G_3$ (FRD)	0.0800	2014	M14
$G_3$ (FRD)	0.0720	2015	This study

FXD areal coverage, we collected output from a set of NSSL WRF forecasts of 21 diverse convective events, and stratified them by peak LMA FXD areal coverage (Fig. A1), depicted on the abscissa, and then plotted a series of points along the ordinate, one for each of the LFA threat areas resulting from each of a series of specified increasing values of assumed LFA FRD threshold. The value of LFA FRD threshold that fit the diagonal of this plot in a least squares sense was then determined and used as the optimum threshold value in the revised algorithm. As Fig. A1 shows, this optimum value turned out to be 0.072 flashes  $\text{km}^{-2} (5 \text{ min})^{-1}$ . This value is slightly lower than the value obtained in McCaul et al. (2014) based on cursory visual inspection of an early version of Fig. A1. Note that the figure shows that, while reasonable values of threshold often straddle the diagonal of the plot, there are some storm events for which the LFA underestimates the threat areal coverage, and others for which it overestimates it. Because of the lack of LMA data in India–Bangladesh, it was not possible to develop a plot like Fig. A1 for the cases that are the subject of this paper.

The equations used in assigning LFA lightning threats, and the history of all relevant parameter values used in the evolving algorithm are summarized in Table A1.

## REFERENCES

- Alexander, C., and Coauthors, 2014: The High-Resolution Rapid Refresh: Recent model and data assimilation development towards an operational implementation in 2014. *26th Conf. on Weather Analysis and Forecasting/22nd Conf. on Numerical Weather Prediction*, Atlanta, GA, Amer. Meteor. Soc., J1.2, <https://ams.confex.com/ams/94Annual/webprogram/Paper239964.html>.
- Bothwell, P., 2005: Development of an operational statistical scheme to predict the location and intensity of lightning. *First Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 4.2, [https://ams.confex.com/ams/Annual2005/techprogram/paper\\_85013.htm](https://ams.confex.com/ams/Annual2005/techprogram/paper_85013.htm).
- Bright, D. R., R. E. Jewell, M. S. Wandishin, and S. J. Weiss, 2004: A physically based parameter for lightning prediction and its calibration in ensemble forecasts. *First Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 4.3, [https://ams.confex.com/ams/Annual2005/techprogram/paper\\_84173.htm](https://ams.confex.com/ams/Annual2005/techprogram/paper_84173.htm).
- Case, J. L., and Coauthors, 2018: Forecasting and monitoring intense thunderstorms in the Hindu Kush Himalayan region: Spring 2018 forecasting experiment. *25th Conf. on Numerical Weather Forecasting*, Denver, CO, Amer. Meteor. Soc., 15B.3, <https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345576.html>.
- , P. N. Gatlin, J. Srikishen, and E. W. McCaul Jr., 2020: Verification of convection-allowing ensemble system over the Hindu Kush Himalaya region during the 2018 and 2019 pre-monsoon severe thunderstorm seasons. *30th Conf. on Weather and Forecasting(WAF)/26th Conf. on Numerical Weather Prediction (NWP)*, Boston, MA, Amer. Meteor. Soc., P152, <https://ams.confex.com/ams/2020Annual/webprogram/Paper366029.html>.
- Cecil, D. J., S. J. Goodman, D. J. Boccippio, E. J. Zipser, and S. W. Nesbitt, 2005: Three years of TRMM precipitation features. Part I: Radar, radiometric, and lightning characteristics. *Mon. Wea. Rev.*, **133**, 543–566, <https://doi.org/10.1175/MWR-2876.1>.
- Clark, A. J., R. G. Bullock, T. L. Jensen, M. Xue, and F. Kong, 2014: Application of object-based time domain diagnostic for tracking precipitation systems in convection-allowing models. *Wea. Forecasting*, **29**, 517–542, <https://doi.org/10.1175/WAF-D-13-00098.1>.
- Curran, E. B., R. L. Holle, and R. E. Lopez, 2000: Lightning casualties in the United States from 1959 to 1994. *J. Climate*, **13**, 3448–3464, [https://doi.org/10.1175/1520-0442\(2000\)013<3448:LCADIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3448:LCADIT>2.0.CO;2).
- Dafis, S., A. O. Fierro, T. M. Giannaros, V. Kotroni, K. Lagouvardos, and E. Mansell, 2018: Performance evaluation of an explicit lightning forecast system. *J. Geophys. Res. Atmos.*, **123**, 5130–5148, <https://doi.org/10.1029/2017JD027930>.
- 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>.
- Emersic, C., and C. P. R. Saunders, 2010: Further laboratory investigations into the relative diffusional growth rate theory of thunderstorm electrification. *Atmos. Res.*, **98**, 327–340, <https://doi.org/10.1016/j.atmosres.2010.07.011>.
- Fierro, A. O., L. M. Leslie, E. R. Mansell, J. M. Straka, D. R. MacGorman, and C. Ziegler, 2007: A high-resolution simulation of the microphysics and electrification in an idealized hurricane-like vortex. *Meteor. Atmos. Phys.*, **98**, 13–33, <https://doi.org/10.1007/s00703-006-0237-0>.
- , E. R. Mansell, C. Ziegler, and D. R. MacGorman, 2013: The implementation of an explicit charging and discharge lightning scheme within the WRF-ARW Model: Benchmark simulations of a continental squall line, a tropical cyclone, and a winter storm. *Mon. Wea. Rev.*, **141**, 2390–2415, <https://doi.org/10.1175/MWR-D-12-00278.1>.
- Gatlin, P. N., and S. J. Goodman, 2010: A total lightning trending algorithm to identify severe thunderstorms. *J. Atmos. Oceanic Technol.*, **27**, 3–22, <https://doi.org/10.1175/2009JTECHA1286.1>.
- Gilmore, M., J. M. Straka, and E. N. Rasmussen, 2004: Precipitation and evolution sensitivity in simulated deep convective storms: Comparisons between liquid-only and simple ice and liquid

- phase microphysics. *Mon. Wea. Rev.*, **132**, 1897–1916, [https://doi.org/10.1175/1520-0493\(2004\)132<1897:PAESIS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1897:PAESIS>2.0.CO;2).
- Goodman, S. J., and Coauthors, 2005: The North Alabama Lightning Mapping Array: Recent severe storm observations and future prospects. *Atmos. Res.*, **76**, 423–437, <https://doi.org/10.1016/j.atmosres.2004.11.035>.
- , 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>.
- Holle, R., 2008: Annual rates of lightning fatalities by country. Preprints, *20th Int. Lightning Detection Conf.*, Tucson, AZ, Vaisala, 1–14, [https://www.vaisala.com/sites/default/files/documents/Annual\\_rates\\_of\\_lightning\\_fatalities\\_by\\_country.pdf](https://www.vaisala.com/sites/default/files/documents/Annual_rates_of_lightning_fatalities_by_country.pdf).
- , 2016: A summary of recent national-scale lightning fatality studies. *Wea. Climate Soc.*, **8**, 35–42, <https://doi.org/10.1175/WCAS-D-15-0032.1>.
- , R. E. Lopez, and B. C. Navarro, 2005: Deaths, injuries, and damages from lightning in the United States in the 1890s in comparison with the 1990s. *J. Appl. Meteor.*, **44**, 1563–1573, <https://doi.org/10.1175/JAM2287.1>.
- Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- , Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, <https://doi.org/10.1175/MWR3199.1>.
- Janjić, Z. I. 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, [https://doi.org/10.1175/1520-0493\(1994\)122<0927:TSMECM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2).
- Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, <https://doi.org/10.1175/WAF2007106.1>.
- , S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levitt, 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>.
- Krehbiel, P. R., R. J. Thomas, W. Rison, T. Hamlin, J. Harlin, and M. Davis, 2000: GPS-based mapping system reveals lightning inside storms. *Eos, Trans. Amer. Geophys. Union*, **81**, 21–25, <https://doi.org/10.1029/00EO00014>.
- Lang, S., W.-K. Tao, R. Cifelli, W. Olson, J. Halverson, S. Rutledge, and J. Simpson, 2007: Improving simulations of convective system from TRMM LBA: Easterly and westerly regimes. *J. Atmos. Sci.*, **64**, 1141–1164, <https://doi.org/10.1175/JAS3879.1>.
- , —, X. Zeng, and Y. Li, 2011: Reducing the biases in simulated radar reflectivities from a bulk microphysics scheme: Tropical convective systems. *J. Atmos. Sci.*, **68**, 2306–2320, <https://doi.org/10.1175/JAS-D-10-05000.1>.
- Laroche, P., P. Blanchet, A. Delannoy, and F. Issac, 2015: Experimental studies of lightning strikes to aircraft. AerospaceLab, Tech. Doc. hal-01184400f, 14 pp., <https://hal.archives-ouvertes.fr/hal-01184400/document>.
- Liu, C., and M. W. Moncrieff, 2007: Sensitivity of cloud-resolving simulations of warm-season convection to cloud microphysics parameterizations. *Mon. Wea. Rev.*, **135**, 2854–2868, <https://doi.org/10.1175/MWR3437.1>.
- Lojou, J.-Y., and K. L. Cummins, 2005: On the representation of two- and three-dimensional total lightning information. *First Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 2.4, [https://ams.confex.com/ams/Annual2005/techprogram/paper\\_86442.htm](https://ams.confex.com/ams/Annual2005/techprogram/paper_86442.htm).
- Lopez, P., 2016: A lightning parameterization for the ECMWF integrated forecasting system. *Mon. Wea. Rev.*, **144**, 3057–3075, <https://doi.org/10.1175/MWR-D-16-0026.1>.
- Lynn, B. H., 2017: The usefulness and economic value of total lightning forecasts made with a dynamic lightning scheme coupled with lightning data assimilation. *Wea. Forecasting*, **32**, 645–663, <https://doi.org/10.1175/WAF-D-16-0031.1>.
- , Y. Yair, C. Price, G. Kelman, and A. 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>.
- Mansell, E. R., D. R. MacGorman, C. L. Ziegler, and J. M. Straka, 2002: Simulated three-dimensional branched lightning in a numerical thunderstorm model. *J. Geophys. Res.*, **107**, 4075, <https://doi.org/10.1029/2000JD000244>.
- McCauley, E. W., Jr., 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>.
- , J. L. Case, S. J. Goodman, S. R. Dembek, F. Kong, and S. Weiss, 2011: Optimizing the lightning forecast algorithm within the Weather Research and Forecasting Model. *Seventh GOES Users Conf.*, Birmingham, AL, NOAA-NWA, <https://www.goes-r.gov/downloads/GUC-7/day1/session04/04-mccauley.pdf>.
- , —, S. R. Dembek, F. Kong, S. J. Goodman, and S. Weiss, 2012: The WRF Lightning Forecast Algorithm: Recent updates and results from convective ensemble forecasts. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., P7.139, <https://ams.confex.com/ams/26SLS/webprogram/Paper211819.html>.
- , T. Chronis, J. L. Case, S. R. Dembek, F. Kong, S. J. Goodman, and S. Weiss, 2014: The WRF Lightning Forecast Algorithm: Recent updates and extension to forecasts of CG lightning. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 84, <https://ams.confex.com/ams/27SLS/webprogram/Paper255198.html>.
- , J. L. Case, and T. Chronis, and S. J. Goodman 2016: The WRF Lightning Forecast Algorithm: Characterization of lightning jumps in simulated storms. *41st Annual NWA Meeting*, Norfolk, VA, National Weather Association, A.52, <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160011362.pdf>.
- Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, <https://doi.org/10.1175/2008MWR2556.1>.
- Murphy, M. J., and N. Demetriades, 2005: An analysis of lightning holes in a DFW supercell storm using total lightning and radar information. *First Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., 2.3, [https://ams.confex.com/ams/Annual2005/techprogram/paper\\_84505.htm](https://ams.confex.com/ams/Annual2005/techprogram/paper_84505.htm).
- Nakanishi, M., and H. Niino, 2006: An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, <https://doi.org/10.1007/s10546-005-9030-8>.
- NIRAPAD, 2018: Monthly Hazard Incident Report, March, April, May (2018). Accessed 1 July 2020, [www.nirapad.org.bd/home/resources/monthlyHazard](http://www.nirapad.org.bd/home/resources/monthlyHazard).
- Petersen, W. A., H. J. Christian, and S. A. Rutledge, 2005: TRMM observations of the global relationship between ice water

- content and lightning. *Geophys. Res. Lett.*, **32**, L14819, <https://doi.org/10.1029/2005GL023236>.
- Price, C., and D. Rind, 1992: A simple lightning parameterization for calculating global lightning distributions. *J. Geophys. Res.*, **97**, 9919–9933, <https://doi.org/10.1029/92JD00719>.
- Schultz, C. J., L. D. Carey, E. V. Schultz, and R. J. Blakeslee, 2017: Kinematic and microphysical significance of lightning jumps versus nonjump increases in flash rate. *Wea. Forecasting*, **32**, 275–288, <https://doi.org/10.1175/WAF-D-15-0175.1>.
- Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, <https://doi.org/10.1175/WAF-D-15-0103.1>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Takahashi, T., 1978: Riming electrification as a charge generation mechanism in thunderstorms. *J. Atmos. Sci.*, **35**, 1536–1548, [https://doi.org/10.1175/1520-0469\(1978\)035<1536:REAACG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<1536:REAACG>2.0.CO;2).
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- Vonnegut, B., 1963: Some facts and speculation concerning the origin and role of thunderstorm electricity. *Severe Local Storms*, Meteor. Monogr., No. 27, Amer. Meteor. Soc., 224–241.
- Wiens, K. C., S. A. Rutledge, and S. A. Tessendorf, 2005: The 29 June 2000 supercell observed during STEPS. Part II: Lightning and charge structure. *J. Atmos. Sci.*, **62**, 4151–4177, <https://doi.org/10.1175/JAS3615.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Williams, E. R., 1985: Large-scale charge separation in thunderclouds. *J. Geophys. Res.*, **90**, 6013–6025, <https://doi.org/10.1029/JD090iD04p06013>.
- Williams, E., and Coauthors, 1999: The behavior of total lightning activity in severe Florida thunderstorms. *Atmos. Res.*, **51**, 245–265, [https://doi.org/10.1016/S0169-8095\(99\)00011-3](https://doi.org/10.1016/S0169-8095(99)00011-3).
- Xue, M., and Coauthors, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 3B.1, <https://ams.confex.com/ams/pdfpapers/124587.pdf>.
- Zhang, W., Q. Meng, M. Ma, and Y. Zhang, 2011: Lightning casualties and damages in China from 1997 to 2009. *Nat. Hazards*, **57**, 465–476, <https://doi.org/10.1007/s11069-010-9628-0>.
- Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP Global Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, <https://doi.org/10.1175/WAF-D-17-0023.1>.