JGR Atmospheres

RESEARCH ARTICLE
10.1029/2021JD034603

Key Points:

- Focuses on short-term quantitative precipitation forecasts (QPFs) over the whole CONUS
- A chief emphasis was purposively placed over regions with poor radar coverage, where GLM lightning data is expected to offer the most benefits
- The simulations made use of an operational convection-allowing model configuration (the forecast model in HRRR)

Correspondence to:
A. Fierro,
Alex.Fierro@noaa.gov

Citation:
Hu, J., Fierro, A. O., Wang, Y., Gao, J., Clark, A. J., Jirak, I. L., et al. (2021). Assessment of storm-scale real time assimilation of GOES-16 GLM lightning-derived water vapor mass on short term precipitation forecasts during the 2020 Spring Forecast Experiment. *Journal of Geophysical Research: Atmospheres*, *126*, e2021JD034603. https://doi.org/10.1029/2021JD034603

Received 16 JAN 2021
Accepted 21 OCT 2021

Assessment of Storm-Scale Real Time Assimilation of GOES-16 GLM Lightning-Derived Water Vapor Mass on Short Term Precipitation Forecasts During the 2020 Spring Forecast Experiment

Junjun Hu[1,2], Alexandre O. Fierro[1,2,3], Yunheng Wang[1,2], Jidong Gao[2], Adam J. Clark[2], Israel L. Jirak[4], Edward R. Mansell[2], Brett Roberts[2,4], Eric P. James[5,6], and Ming Hu[6]

[1]Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, OK, USA, [2]NOAA/OAR/National Severe Storms Laboratory, Norman, OK, USA, [3]Department of Forecasting Models – ZAMG. Vienna, Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria, [4]NOAA/Storm Prediction Center, Norman, OK, USA, [5]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [6]NOAA/Global Systems Laboratory, Boulder, CO, USA

Abstract This study assesses the impact of assimilating pseudo-observations for water vapor mass derived from the Geostationary Operational Environmental Satellites GOES-16 lightning data on short-term quantitative precipitation forecasts (QPFs) over the contiguous United States (CONUS), with an emphasis given to regions characterized by an overall poor radar coverage. The GOES-16/17 provide nearly uniform and high temporal resolution total lightning observations over most of the Americas. To leverage this information for convective scale weather forecasts, a three-dimensional variational data assimilation (DA) package developed by the National Severe Storm Laboratory was employed to assimilate these data. To mimic operational settings, the Weather Research and Forecasting Model configuration used in NOAA Global Systems Laboratory's High-Resolution Rapid Refresh Model version 4 was utilized. During the 2020 NOAA Hazardous Weather Testbed Spring Forecasting Experiment, four experiments were run in real time during a 29-day period over CONUS and surrounding territories to assess the added value of GOES-16 lightning over conventional radar data. Overall, the lightning DA (LDA) showed benefit in improving QPFs up to 6 hr, with the best improvements seen during the first 3 hr. Owing to notably larger lightning activity over the eastern CONUS, the most noticeable impacts from the LDA were seen there. Case-by-case analysis revealed that positive impacts from the LDA were seen over areas characterized by both good (eastern two thirds of CONUS) and poor radar coverage (western one third of CONUS). Highlighting the inherent difficulties in developing an optimal observation operator based on moisture for LDA applications, the GLM-based DA experiments systematically overestimated rainfall over the eastern two thirds of CONUS while underestimating it over the western one third of CONUS.

1. Introduction

Timely, accurate and reliable convective-scale severe weather forecasts remain a critical challenge for forecasters confronted with the responsibility of warning the public and stakeholders of imminent, potentially life threatening high-impact weather threats. In the past few decades, notable forecast improvements spanning a wide range of scales have been achieved owing in large part to notable strides in data assimilation (DA) techniques, as well as the increasing volume of available observations from different observing platforms. In the context of convection-allowing (3–4 km) to cloud scale (<3 km) forecasting, Doppler radar data (i.e., reflectivity factor and radial velocity) have been extensively studied and proved to be a critical component in today's forecast operations. An inherent limitation of ground-based radar networks is that they are incapable of providing uniform observations across vast regions with inhomogeneous terrain (e.g., beam blockage; Maddox et al., 2002; Zhang et al., 2011) and over oceanic regions remaining outside the network's range.

The assimilation of spaceborne total lightning (i.e., cloud-to-ground plus intracloud) observations is expected to complement radar data over such vast regions because of its wide coverage and its insensitivity to terrain (e.g., Pessi & Businger, 2009). The Geostationary Lightning Mapper (GLM) aboard the Geostationary

Operational Environmental Satellites Series-R (GOES-R, operationally known as GOES-16 and GOES-17 for eastern and western sectors, respectively; Goodman et al., 2013; Gurka et al., 2006) provides a nearly uniform, 24-hr high spatial (~8–14 km) and temporal resolution (20 s) total lightning data over the Americas (Goodman et al., 2013; Peterson, 2019; Rudlosky et al., 2019), including typically data sparse regions such as mountainous terrain and vast oceanic regions. Thus, GLM total lightning data have great potential for cloud-scale DA applications to improve forecasts of thunderstorm events, particularly when combined with other storm scale datasets such as, for example, ground-based radar data.

Pioneering works aimed at assimilating lightning in NWP models were primarily designed for parameterized convection scales (~9–30 km) and employed cloud-to-ground (CG) flash data from ground-based networks (e.g., Alexander et al., 1999; Chang et al., 2001; Jones & Macpherson, 1997a, 1997b; Papadopoulos et al., 2005; Pessi & Businger, 2009) or combined CG flash with limited-area total lightning data (Mansell et al., 2007). The core concept behind their lightning data assimilation (LDA) method was to trigger the observed convection by boosting latent heating (and, hence, thermal buoyancy) within the columns or grid volumes where lightning was observed in the forecast domain.

The first attempt toward storm-scale (≤3 km) assimilation of total lightning was examined in Fierro et al. (2012, 2014). Instead of devising functional relationships between lightning and latent heating, Fierro et al. (2012) opted to make use of water vapor mass mixing ratio (q_v). Rationale for selecting q_v arises from its linkage with virtual potential temperature perturbation and hence thermal buoyancy. This simple LDA nudging method was tested over a 70-day sample period over the contiguous United States (CONUS) during the spring of 2013, and notable forecast improvements in the timing and placement of convection were seen (Fierro et al., 2015). Similar forecast improvements were also documented by parallel research efforts considering either sub-cloud warming (Marchand & Fuelberg, 2014) or adjusting latent heating derived from microphysical quantities/variables (graupel mass, radar reflectivity) to boost convection at observed lightning locations (Wang et al., 2017).

The lightning nudging scheme of Fierro et al. (2012) was later adapted for use in the three-dimensional variational (3DVAR) framework (Fierro et al., 2016) and within the Gridpoint Statistical Interpolation (GSI) framework (Fierro, Zhao, et al., 2018). More recently, Fierro et al. (2019) and Hu et al. (2020) adapted the 3DVAR LDA technique for operational GLM data and evaluated their impact on short-term forecasts of high-impact convective events that primarily occurred over the southern Great Plains of the Unites States.

Auxiliary research efforts explored ensemble Kalman filters (EnKF) to assimilate total lightning information (Allen et al., 2016; Mansell, 2014). A recent study (Kong et al., 2020) adopting the graupel-based forward operator of Allen et al. (2016) to assimilate GLM data in the GSI-EnKF framework revealed noteworthy forecast improvements for a retrospective high-impact mesoscale convective event.

Building upon the earlier works of Fierro et al. (2019) and Hu et al. (2020), this research aims to assess more systematically the assimilation of GLM total lightning data for forecasts conducted daily in real time settings over a 5-week period and covering a notably wider forecast domain (i.e., beyond regional scales) encompassing the entire CONUS and surrounding territories. These real time simulations were conducted as part of the 2020 NOAA Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE). The cloud-scale DA component employed a 3DVAR package developed at the National Severe Storm Laboratory (NSSL). A 3DVAR approach was used due to its lower computational burden and, therefore, its ability to provide operational forecasters with a faster delivery of forecast products.

The systematic evaluations herein place emphasis on short-term (≤6-hr) quantitative precipitation forecasts (QPFs) over the whole CONUS, including both areas characterized by good radar coverage (eastern two thirds of CONUS) and poor radar coverage (western one third of CONUS). For completeness, 6–12 hr forecasts are also evaluated as needed. For this assessment, four experiments were devised: a baseline "reference" simulation (labeled REF) complemented by experiments assimilating, individually GLM lightning data (GLM), radar observations (RAD), and both datasets combined (GLM + RAD)—all using REF as background. In other words, all cycled DA experiments begin with the same background data (from REF) and with the same boundary conditions when initiating a forecast. To the best of the authors' knowledge, this is a first effort aimed at systematically assessing the added value of operational GLM lightning data over the whole CONUS in real time settings using an operational convection-allowing model configuration.

2. The 3DVAR Analysis and Forecast System and Data Used for the DA

2.1. The Forecast Model

The forecasts utilized the Advanced Research Version of the Weather Research and Forecasting model version v3.9 (WRF-ARW, Skamarock & Klemp, 2008) adopted in the fourth generation high-resolution rapid refresh model (HRRRv4) by NOAA's Global Systems Laboratory (GSL). The SFE mimicked the operational settings of HRRRv4, using the same version of the WRF-ARW forecast model adopted in HRRRv4 together with the same physics parameterizations, namely: the Thompson "aerosol aware" bulk microphysics scheme (Thompson et al., 2008), the Mellor-Yamada-Nakanishi-Niino level 2.5 (MYNN2) planetary boundary layer (PBL) scheme (Nakanishi & Niino, 2006; Olson et al., 2019), the Rapid Update Cycle (RUC) land surface scheme (Benjamin et al., 2004; Smirnova et al., 2016), and the longwave (LW) and shortwave (SW) radiation schemes from Rapid Radiative Transfer Model for General Circulation Models (RRTMG; Iacono et al., 2008). No cumulus parameterization is employed.

2.2. The 3DVAR System

The 3DVAR DA framework used herein was originally developed for the Advanced Regional Prediction System at the Center for Analysis and Prediction of Storms and later upgraded at NSSL (Wang et al., 2019). Besides the capability of assimilating surface observations, radar radial velocity and its derived velocity–azimuth display (VAD) wind observations (Gao et al., 2004; Hu et al., 2006), this convective-scale DA scheme directly assimilates radar reflectivity with a forward operator classifying the hydrometer species based on background temperature (Gao & Stensrud, 2012; Gao et al., 2013). The 3DVAR DA system is under continuous development to assimilate a broader range of observations, such as satellite derived Cloud Water Path (CWP), ground based or spaceborne lightning data (Fierro et al., 2016; Hu et al., 2020), total precipitation water (Pan et al., 2018), as well as radar-derived pseudo-observations for q_v from vertical integrated liquid water (Lai et al., 2019).

2.3. The Data Set

The datasets used in this study include radar measurements from the US National Weather Service (NWS) operational Weather Surveillance Radar-1988 Doppler (WSR-88D) network and total lightning observations from the GLM aboard GOES-16.

The level II radar data (i.e., radar reflectivity factor and radial velocity) from the WSR-88D network (e.g., Gao et al., 2013; Zhang et al., 2011) and radar-derived VAD wind observations from a total of 143 radar sites covering the CONUS were used. Before being interpolated onto the model grid for the DA, the radar data were quality controlled (QC-ed) through a procedure outlined in Gao et al. (2013) that includes the removal of weak or nonmeteorological radar echoes and radial velocity dealiasing. For a given grid point wherein data from multiple radars overlap, the largest observed reflectivity value is selected for the DA. Radial velocity and VAD observations are also assimilated to adjust the wind components, and radar reflectivity data are used to adjust the mass mixing ratios for snow (q_s), rainwater (q_r), and hail (q_h) based on temperature thresholds (Gao & Stensrud, 2012).

The GLM instrument aboard the GOES-16 and GOES-17 satellites detects lightning activity 24/7 at a 2 ms frame rate across the Americas and adjacent oceanic regions, with a 24hr-average detection efficiency exceeding 70% (Edgington et al., 2019; Goodman et al., 2013; Mach, 2020; Rudlosky et al., 2019). The GLM provides the QC-ed and filtered "level 2" total lightning data in 20-s packets at a horizontal grid spacing ranging from about 8 km near the center of the field of view to about 14 km farther away near the edges (Goodman et al., 2013). The elemental variable provided by the GLM data directly related to lightning flashes is the pixel-level lightning optical energy "event". Based on temporal and spatial coincidence thresholds, events are combined into groups and further into flashes through a parent-to-child relationship (Mach et al., 2007). In this work, only GLM "flash" data are used to approximate total flash origin densities on the model grid. Following earlier studies (Fierro, Stevenson, & Rabin, 2018; Fierro et al., 2019), the areal footprint/extent of lightning flashes is purposively not considered in this study to reduce the impact of the LDA-induced q_v adjustments outside convection regions.

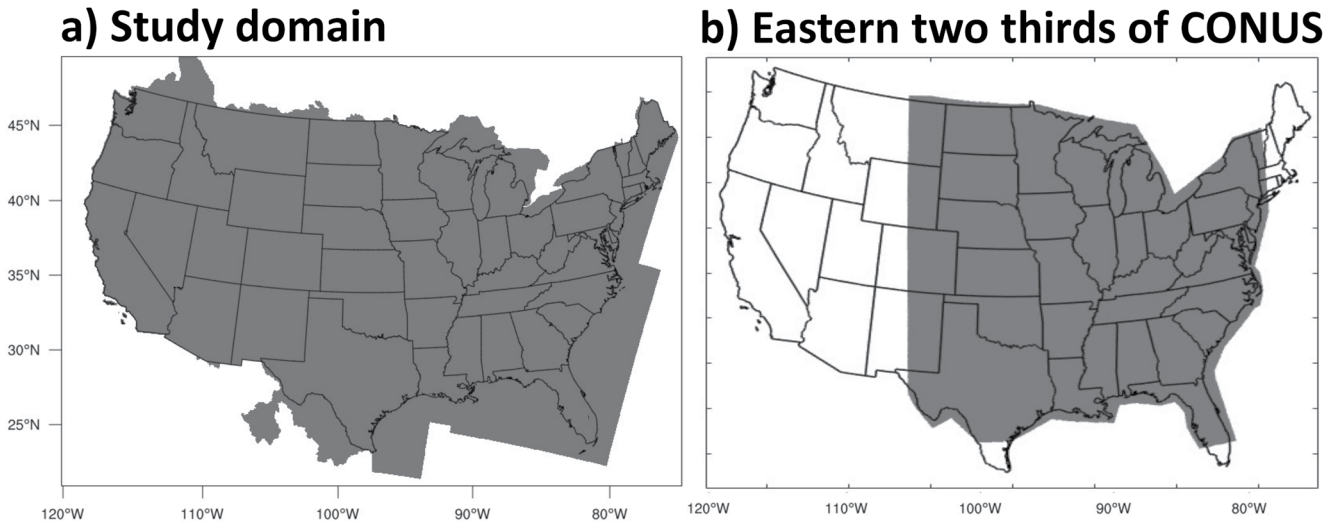

Figure 1. Simulation domain (3-km grid spacing) with (a) the gray shading indicating the area where the forecast evaluation was performed over the CONUS, and (b) the gray-shading delineating the area referred to as the eastern two thirds of CONUS in the text.

The GLM DA scheme and settings employed in this study are conceptually the same as those in Hu et al. (2020). Prior to each 3DVAR analysis, the 20-s GLM data packets recording the latitude-longitude coordinates of the flash pixel centroids were accumulated over a predetermined time interval (here, 10 min). At each observed GLM centroid horizontal location, pseudo-observations for q_v are created by adjusting the background q_v within a fixed depth (3 km) above the LCL toward near saturation conditions (set to a relative humidity of 95% herein, i.e., $q_v = q_{vsat} \times 95\%$). Afterward, these GLM-derived pseudo-observations for q_v are directly assimilated into the 3DVAR system. This DA procedure is independent of flash rate and is only applied to any qualifying grid points if the modeled (i.e., background) relative humidity is less than 95%. Areas with zero GLM-observed gridded hourly flash are treated as missing observations, and the pseudo-observations for q_v are not created. Additional details pertaining to the LDA scheme and the 3DVAR parameters used are provided in Hu et al. (2020).

Ongoing preliminary observational work revealed that, overall, the vast majority of lightning optical emissions detected by GOES-16 over CONUS overlaps with those from GOES-17 (Rudlosky & Virts, 2021). Because the current lightning DA scheme does not consider flash rates, only GOES-16 data were utilized herein. The inclusion of GOES-17 lightning data will be deferred to forthcoming studies especially in the advent that a lightning observation operator emphasizing flash density rate is considered during the DA (e.g., Fierro et al., 2012; Kong et al., 2020).

Although both the GLM and WSR-88D data can be publicly obtained from online NOAA databases such as those provided by the National Centers for Environmental Prediction (NCEP), these datasets were fed in real time into the experimental simulations from a large public data repository made available to researchers and collaborators on the Jet NOAA high performance computing cluster.

3. Experimental Design

3.1. General Set Up

To provide forecasters with daily forecast products consistent with other ongoing experiments during the SFE, the simulation domain follows the guidelines from the Community Leveraged Unified Ensemble (CLUE; Clark et al., 2018; Figure 1), which is characterized by horizontal dimensions in grid points of 1620 × 1120 and a uniform horizontal grid spacing of 3 km (or 4860 × 3360 km^2). The domain uses a stretched vertical grid with 51 levels extending from the surface up to 15 hPa. The integration time step is 20 s.

Figure 2. April–May 2020 calendar showing the sample of forecast days that were used in this work (gray shading, totaling 29 days) and, in white, the forecast days (in May) that were not part of the forecast sample owing to technical issues encountered during the real time experiment.

Owing to technical problems occasionally encountered during the 5-week duration (from 27 April to 29 May) of the SFE including system maintenance, node failure/network outages, unavailability of critical input data, complete runs are available for 29 days out of a target maximum of 33 days (Figure 2). Each day, the GLM and/or radar DA was performed between 2300 and 0000 UTC using a 15-min 3DVAR cycling frequency (Figure 3). At each analysis step, data accumulated within a 10-min time interval prior to the analysis time t (i.e., from t-10 min to t) are assimilated (following Hu et al., 2020). Following the 1-hr 3DVAR DA period, a 12-hr deterministic forecast was launched at 0000 UTC.

The background at the initial DA step (2300 UTC) was the 2-hr forecast from the fifth generation Rapid Refresh (RAPv5) initialized at 2100 UTC. When launching the forecast at 0000 UTC, boundary conditions are also based on RAPv5 forecasts initialized at 2100 UTC (available in real time on the Jet clusters). The forecast output data in the real time experiments were saved hourly throughout the 12-hr forecast period.

During the daily evaluation performed by operational NWS forecasters and participants, the SFE focused on two (out of the four) experiments, namely the RAD and GLM + RAD to focus on the added value of the GLM lightning relative to the more conventionally assimilated WSR-88D radar data. Each day, three on-demand custom evaluation domains were selected based on the observed total GLM lightning densities during the DA period (2300-0000 UTC). Furthermore, to better assess the added benefits of GLM lightning data and/or radar data over a deterministic forecast without assimilating GLM lightning or radar data, two additional experiments—namely REF and GLM—were conducted in real-time and examined offline.

3.2. Performance Evaluation: Methods and Data

Akin to Fierro et al. (2015), the performance evaluation is focused on short-term QPFs. Composite reflectivity fields were also examined but revealed, overall, very similar results in terms of forecast skill as for the QPFs and, hence, are not shown. The QPFs are evaluated against hourly rainfall quantitative precipitation estimates (QPEs) from NCEP's Stage IV (UCAR, 2020) product (Baldwin & Mitchell, 1997; Lin & Mitchell, 2005; Nelson et al., 2016), with a native polar stereographic grid spacing of 4 km across the CONUS. Independent studies showed that the fused Stage IV estimates generally outperform radar-only and radar-gauge products that did not undergo a separate manual quality control (Gourley et al., 2010; Westcott et al., 2008), which accounts for its wide usage in modeling studies to evaluate precipitation products (Smalley et al., 2014). Because of its more systematic usage for model verification over the years, the Stage IV QPE data were selected over the NSSL Multi-Radar/Multi-Sensor (MRMS) QPE product (Smith et al., 2016; Zhang et al., 2011). We expect the forecast verification results to be relatively similar using either product, especially given that forecasts metrics aggregated over a month period are emphasized herein.

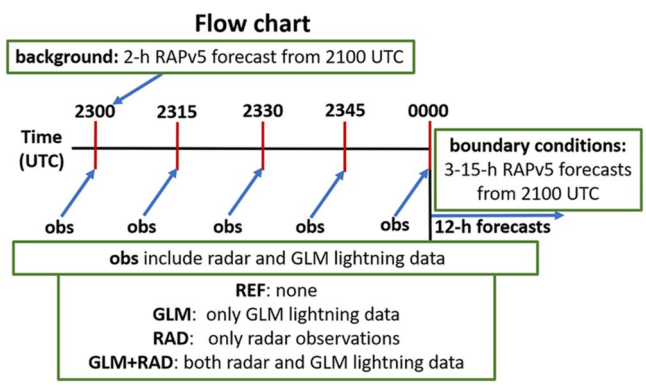

Figure 3. Data assimilation (DA) and forecast flow chart.

Before performing the comparison between observations and forecasts, the 4-km Stage IV QPFs were interpolated onto the 3-km grid of the simulation domain. Contingency-table measures including the probability of detection (POD), the false alarm ratio (FAR) together with the success ratio (SR = 1-FAR), the frequency bias (BIAS), the critical success index (CSI) as well as equitable threat score (ETS, e.g., Clark et al., 2010; Wilks, 2006) were calculated for hourly accumulated precipitation (HPRCP) forecasts relative to the Stage IV rainfall estimates. Their calculations are based on the standard 2 × 2 contingency table elements (Table 1) and are as follows:

$$\mathrm{POD} = N_{\_\mathrm{hits}} / \left(N_{\_\mathrm{hits}} + N_{\_\mathrm{misses}} \right) \quad (1)$$

Table 1
A 2 × 2 Contingency Table for Forecasting of Events, Where N_hits: The Number of Correct Forecasts of Events (Hits), N_falsealarms: The Number of Forecasts of Events That did not Occur (False Alarms), N_miss: The Number of Events That Occurred but Were not Forecasted (Misses), N_reject the Number of Correct Forecasts for the Events That did not Occur (the Number of Correct Rejections)

		Event observed	
		Yes	No
Event forecast	Yes	N_hits	N_falsealarms
	No	N_miss	N_reject

$$\mathrm{FAR} = N_{\_\mathrm{falsealarm}} / \left(N_{\_\mathrm{total}}\right) \tag{2}$$

$$\mathrm{SR} = 1 - \mathrm{FAR} = N_{\_\mathrm{hits}} / \left(N_{\_\mathrm{total}}\right) \tag{3}$$

$$\mathrm{CSI} = N_{\_\mathrm{hits}} / \left(N_{\_\mathrm{total}} + N_{\_\mathrm{misses}}\right) \tag{4}$$

$$\mathrm{BIAS} = \left(N_{\_\mathrm{total}}\right) / \left(N_{\_\mathrm{hits}} + N_{\_\mathrm{misses}}\right) \tag{5}$$

$$\mathrm{ETS} = \left(N_{\_\mathrm{hits}} - S\right) / \left(N_{\_\mathrm{total}} + N_{\_\mathrm{misses}} - S\right) \tag{6}$$

$$S = \left(N_{\_\mathrm{hits}} + N_{\_\mathrm{misses}}\right)\left(N_{\_\mathrm{total}}\right) / \left(N_{\_\mathrm{total}} + N_{\_\mathrm{misses}} + N_{\_\mathrm{reject}}\right),$$
$$N_{\_\mathrm{total}} = N_{\_\mathrm{hits}} + N_{\_\mathrm{falsealarm}}$$

where $N_{\_\mathrm{total}}$ is the number of forecasted events and the parameter S in 6 represents the number of hits forecasted at random $N_{\_\mathrm{total}}$ times. All other variables are described in Table 1. As formulated, the ETS ranges from −1/3 (no skill) to 1 (perfect skill).

The scores were calculated for a series of neighborhood radii ranging from 3 km (grid point based) to 39 km by an increment of 3 km for the following thresholds: 1, 2.5, 5, and 10 mm. By considering adjacent grid points within a specified radius of a given grid point, a neighborhood-based index relaxes the conditions required to qualify an event as a hit. That is, given a radius r and at grid point j, if an event is observed at j, it is a hit if the event is forecast at j or any grid point within radius r of grid point j, otherwise it is a miss.

Additionally, the performance evaluation made use of the normalized mean square error (NMSE) defined as follows:

$$\mathrm{NMSE} = E[\|x^f - H(y^o)\|^2] / E[\|x^f + H(y^o)\|^2] \tag{7}$$

where $E[.]$ is the expected value over a given domain area, $\boldsymbol{x^f}$ is the forecast, and $\boldsymbol{y^o}$ is the observation vector.

To investigate in more detail the forecast performance on a case-to-case basis, the composite reflectivity (CREF) forecasts were compared against CREF observations from the MRMS product (Smith et al., 2016; Zhang et al., 2011). The MRMS CREF data were retrieved from a local server (vMRMS) at NSSL with a horizontal grid spacing of 0.01° and are available in 5-min increments. Akin to the Stage IV data, the MRMS CREF fields were interpolated onto the 3-km model grid prior to performing the comparison.

4. Results

4.1. Statistical Evaluation

To gain an overall appreciation of the rainfall forecast skill throughout the 2020 SFE evaluation period, the forecast evaluation measures listed earlier were computed for hourly rainfall aggregated over all 29 forecast days. The POD, FAR, CSI, and BIAS were examined simultaneously using performance diagrams (Roebber, 2009).

For each of the four experiments, performance diagrams aggregated over all 29 forecast days over the entire CONUS (i.e., regions with valid stage IV data, Figure 1a) for hourly QPF at 1, 3, and 6-hr forecast are presented in Figure 4. Because, overall, the relative performance of each of the four experiments was similar for the neighborhood radii selected, only the results for a neighborhood radius of 18 km are shown.

An aggregate, domain-wide general improvement in forecast skill over REF can be seen for all DA runs (Figure 4), with the best results obtained for GLM + RAD. For all four experiments, the forecast skill for hourly QPF peaks at 1-hr forecast and progressively decreases afterward, as evidenced by decreasing POD,

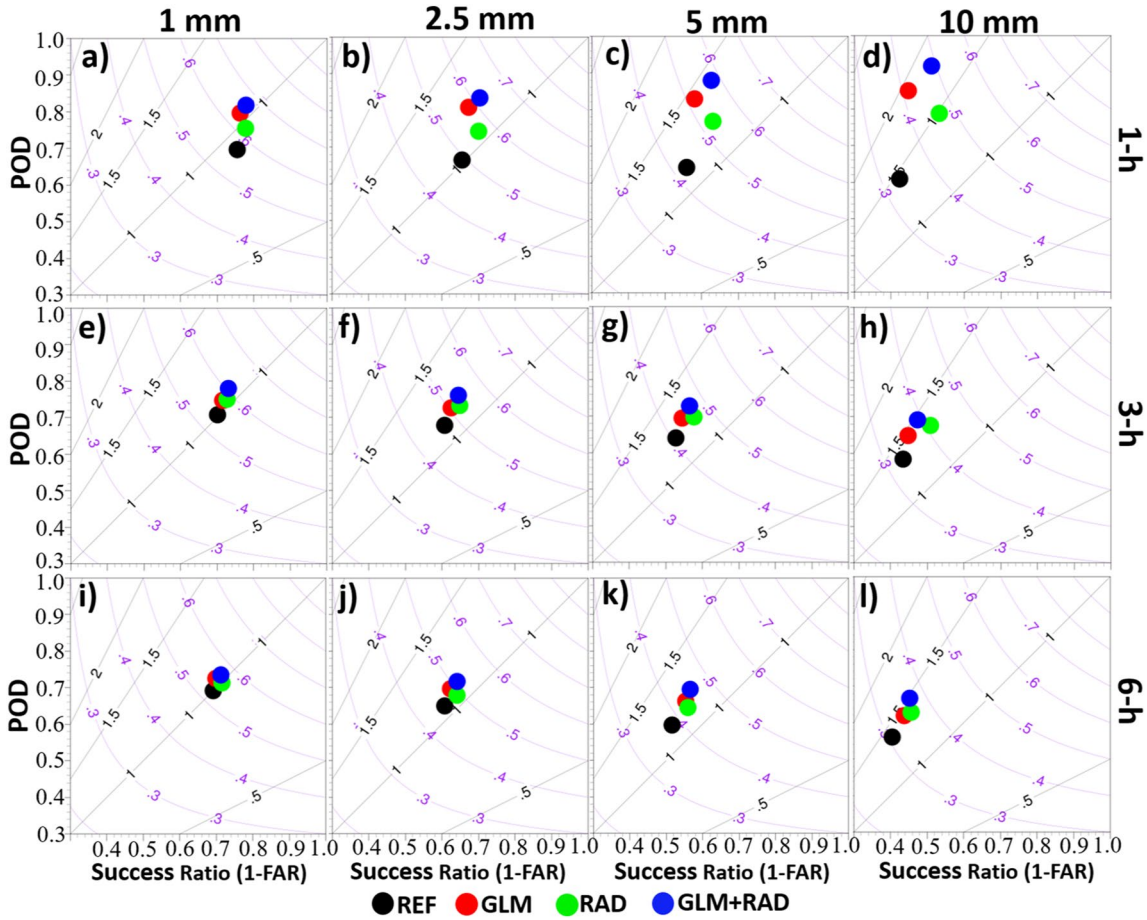

Figure 4. Performance diagrams for 1- (a–d), 3- (e–h), 6- (i–l) hourly accumulated precipitation (HPRCP) forecasts relative to Stage IV hourly rainfall data aggregated over all 29 cases (Figure 2) across the CONUS (the gray-shading area in Figure 1a) for HPRCP thresholds of 1 mm (a, e, i), 2.5 mm (b, f, j), 5 mm (c, g, k), and 10 mm (d, h, l). The results are presented for a neighborhood radius of 18 km. The upper-right corner of each plot represents perfect forecast skill and, conversely, the lower-right corner no skill. The thin purple curves show the critical success index (CSI), and the thin diagonal black lines the frequency bias such that dots located on the $y = x$ diagonal indicate a forecast devoid of bias (i.e., frequency bias of unity).

decreasing CSI and increasing FAR, especially at higher rainfall thresholds (i.e., 5 and 10 mm hr^{-1}). Despite these improvements, rainfall totals were generally overestimated by all DA experiments and, arguably, even by the REF run (BIAS > 1 in Figure 4). This overestimation worsens with increasing threshold and appears to be the largest for the two GLM-based experiments. Consistent with this result, statistics for CREF forecasts relative to MRMS CREF products shows an overall overprediction, especially for thresholds exceeding 30 dBZ (not shown). This result strongly suggests that the forecast model has a tendency to overpredict the strength and/or areal coverage of late afternoon convection, which might not only be partially attributed to bias and errors in initial and boundary conditions, but to errors in the physical processes associated with convective precipitation (e.g., by the Thompson microphysics, Fierro et al., 2015).

Comparing the relative aggregate forecast performances of GLM over REF and GLM + RAD over RAD reveals that the QPFs generally benefit from the GLM data during the assimilation (e.g., higher POD) at the expense, however, of exacerbating any existing wet biases, similar to previous q_v-based LDA studies (Fierro et al., 2015). Figure 4 shows that in fact, at higher rainfall thresholds, the REF run already exhibits increasingly wetter biases. Future research efforts could, for instance, adopt a q_v conservation approach (Fierro et al., 2019) or explore alternatives aimed at specifically targeting convective core areas prior to applying the lightning-induced moisture adjustments. To better assess the tendency of the present DA method to generate wet biases in the forecasts, additional, off-line, GLM DA experiments (not shown) were performed using the same 3DVAR DA flowchart as in Figure 3 but over a 24-hr period instead of 1-hr (prior to 00UTC).

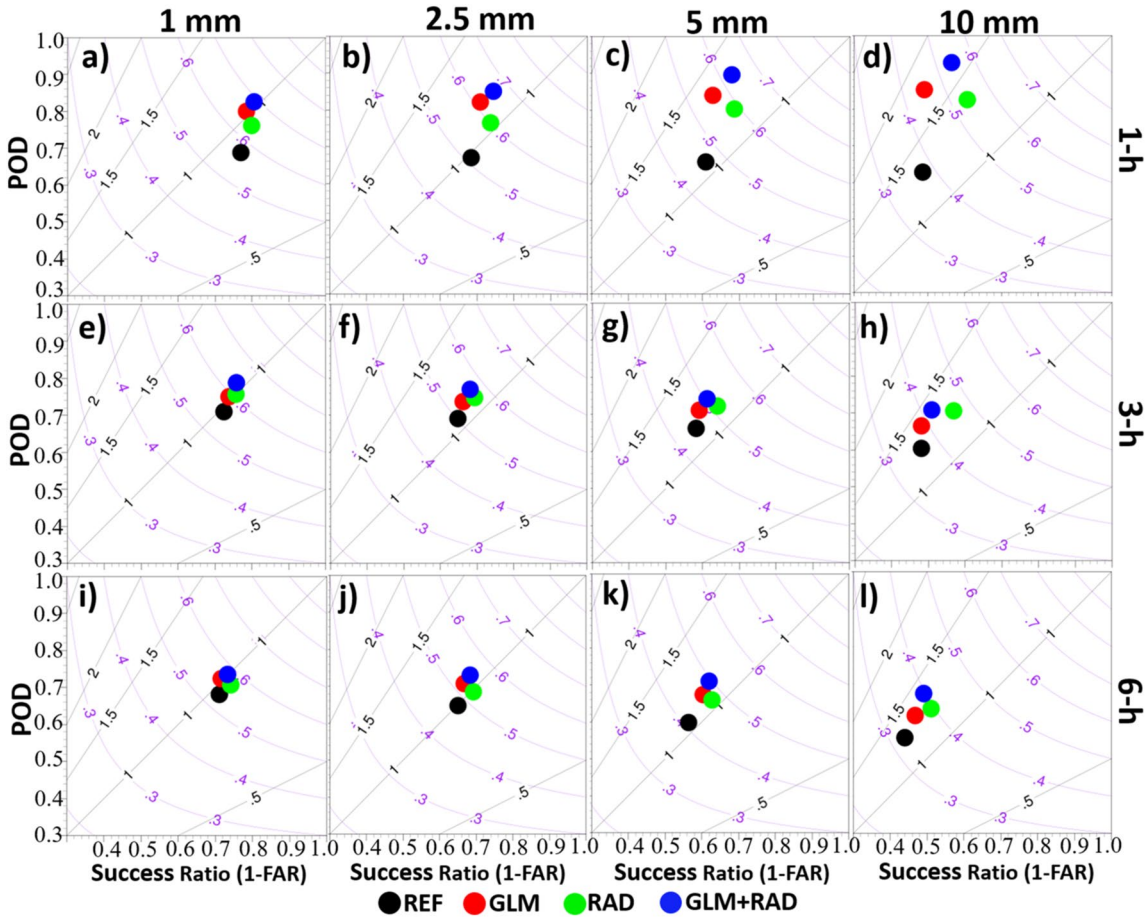

Figure 5. Same as Figure 4 but for precipitation data covering the eastern two thirds of CONUS only (cf. Figure 1b).

These experiments purposively targeted (here 3) days in the SFE samples containing the largest amount of GLM observed lightning. In addition to the notable increase in computational burden incurred by the 24-fold increase in 3DVAR analyses, the results revealed that not only the frequency biases remained relatively similar to that of the 1-hr cycling experiment but also that the forecast skill was generally degraded for both rainfall and CREF. This degradation in forecast skill with increasing number of 3DVAR cycles is in part likely attributed to the successive introduction in each DA cycles of errors already contained in the larger scale model, and reanalysis/forecast datasets used to derive the background fields for each successive 3DVAR analysis. Wet biases also arise from the systematic development of spurious convection, which motivates further refinements to the present DA method to address this drawback by assimilating, for example, zero lightning observations (Lai et al., 2019; Kong et al., 2020).

Though the Stage IV QPEs are regarded as one of the best observational rainfall datasets covering CONUS, QPEs of Stage IV are underestimated over the mountainous west due to beam blockage (e.g., Smalley et al., 2014). To partially take this into account, the performance diagrams over the eastern two thirds of CONUS (gray shading in Figure 1b) were examined separately but generally revealed similar aggregate performance improvements relative to the whole CONUS (Figure 3). This is because during the 1-hr DA period, the overwhelming majority of the lightning activity occurred over the eastern two thirds of CONUS (see later in the section). Slightly improved aggregate forecast skill is obtained over the eastern two thirds of CONUS as corroborated by higher POD, SR, and CSI (Figure 5), but lower BIAS values.

Statistics for forecast lead times beyond 6-hr (not shown) shows and confirms that any improvements/gains induced by the assimilation of either lightning or radar data are gradually lost. One relevant and atypical aspect of this evaluation is the overall good aggregate performance of the REF experiment (POD

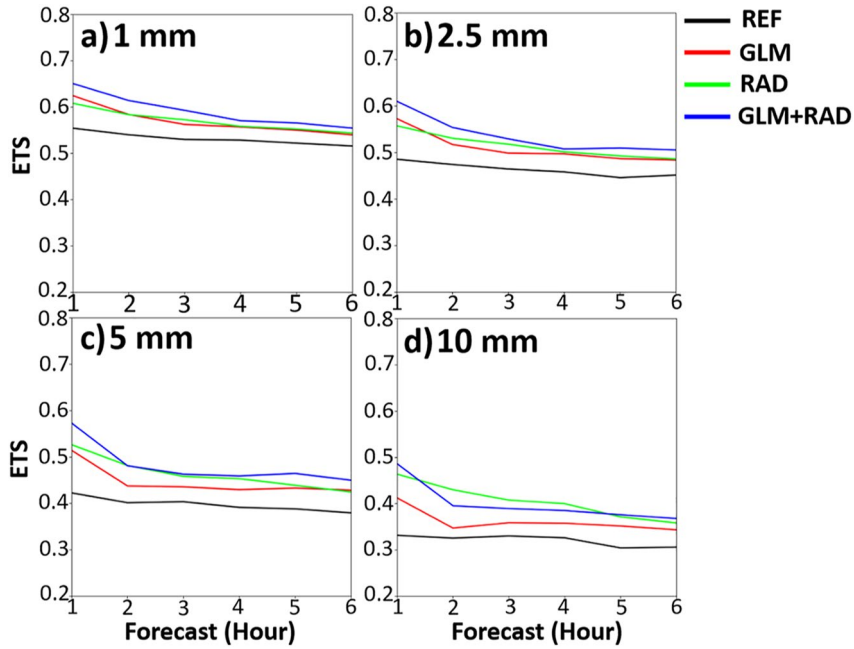

Figure 6. Equitable threat scores (ETS) for 1–6 hr HPRCP forecasts relative to Stage IV hourly rainfall estimates aggregated over all 29 cases (Figure 2) over the CONUS (gray-shading area in Figure 1a) for hourly thresholds of (a) 1 (b) 2.5, (c) 5, and (d) 10 mm. As in Figures 4 and 5, the results are shown for a neighborhood radius of 18 km.

ranging between 0.55 and 0.7, Figures 4 and 5). This is because, as mentioned earlier, the RAPv5 product operationally used by GSL already contains blended information from many observing platforms, including temperature, moisture and wind observations from rawinsonde, surface stations and aircrafts, WSR-88D radar reflectivity and radial wind, proxy reflectivity converted from National Lightning Detection Network (NLDN) lightning flash rate data, GOES atmospheric motion vectors (AMVs), among others (cf. Table 4 in Benjamin et al., 2016 for details). Although forecast improvements might be more challenging to achieve in this context, the RAPv5 was intentionally utilized herein to mimic GSL's operational settings with the HRRR. Daily evaluation with SFE participants revealed that the overall good performance of REF was most evident for strongly forced systems (higher predictability, not shown). Consequently, improvements induced by the GLM DA were more readily seen for convective events associated with weak forcing (cf. individual case analyses below).

The aggregate ETSs for hourly rainfall generally agree well with the performance diagrams. All DA experiments produce higher aggregate ETSs than REF (Figure 6) at 1–6 hr forecasts. The ETSs also decrease with time, but the relative differences in scores at 1-hr forecast still persist between experiments until 6-hr. Except for REF, all DA experiments exhibit a notable decrease in ETS during the first two hours of forecast. GLM + RAD produces overall the best forecast skill for thresholds of 1, 2.5 and arguably of 5 mm (Figures 6a–6c). This result does not hold for the 10 mm threshold and is partially attributed to fewer occurrences of strong convection in both the forecast and verification datasets.

4.2. Spatial Evaluation

To illustrate more concisely the differences in spatial distribution of QPFs between each of the four experiments, the 0–6 hr rainfall summed over all 29 forecast days is examined herein. For this analysis, the total flash density rate used during the DA period aggregated over all 29 days is conjointly analyzed with the rainfall to help better depict areas where the GLM DA produced the most noticeable impacts (Figure 7a).

From Figure 7, it becomes clear that two GLM runs, that is, GLM and GLM + RAD, produce notably more rainfall than either RAD or REF, especially offshore over southeastern Florida, southwestern Texas, extreme northeastern Mexico, the southern Great Plains, and western CONUS. Coincidentally, all these areas collo-

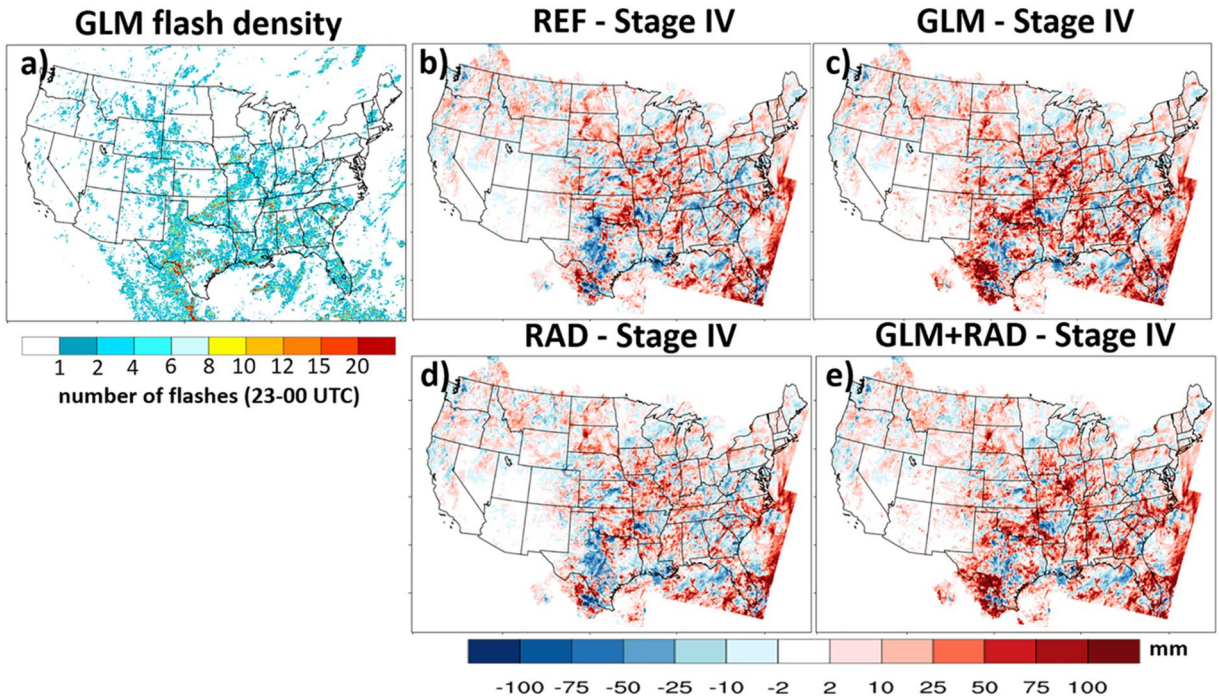

Figure 7. Total GLM flash density used during the DA between 2250 and 0000 UTC (a), presented side-by-side with differences of total 6-hr accumulated QPFs between (b) REF (c) GLM (d) RAD (e) GLM + RAD and the Stage IV QPE. The GLM flashes and QPF are accumulated over all 29 cases herein. The total flash density rate employed for the DA in (a) is defined as the sum of all 10-min GLM flash accumulation intervals that were used for each 3DVAR cycle between 2300 and 0000 UTC, namely: 2250–2300, 2305–2315, 2320–2330, 2335–2345, and 2350–0000 UTC.

cate well with the areas of largest aggregate GLM flash densities (Figure 7a). When directly comparing the rainfall maps for GLM and GLM + RAD against the Stage IV data, it becomes also evident that the rainfall totals over the above regions of the forecast domain are overestimated. This systematic overestimation in QPF for the GLM-based runs is consistent with the aggregate wet biases documented in Figures 4 and 5. The larger amounts of precipitation added in the GLM-based experiments relative to the RAD DA forecasts is undeniably partially attributed to the moisture-based LDA strategy, and this despite efforts to confine q_v adjustments vertically in a 3-km deep layer and spatially with a horizontal decorrelation length scale on the order of 1 grid point (3 km). The other two experiments–RAD and REF–generally exhibit smaller differences relative to the Stage IV QPE but tend to underestimate rainfall over the mountainous areas and radar sparse regions of the west. During the SFE, there were repeated occurrences of convective weather events over western/southwestern Texas, echoed by an abundance of GLM lightning there (Figure 7a). On the aggregate, REF severely underestimated the precipitation over western Texas with RAD offering some slight improvement that still remained well below the Stage IV estimates.

To get a quantitative assessment of quality of the QPFs from each experiment, the NMSE relative to Stage IV QPEs was computed together with the NMSE skill score relative to REF. Table 2 lists the NMSE values of

Table 2
NMSEs of Each Experiment Relative to the Stage IV Product, Averaged Over the Whole CONUS, the Western One Third and Eastern Two Thirds of CONUS

Domain	REF	GLM	RAD	GLM + RAD
CONUS	0.451	0.436 (0.033)	0.435 (0.035)	0.426 (0.055)
Western one third	0.481	0.466 (0.031)	0.487 (−0.012)	0.475 (0.012)
Eastern two thirds	0.439	0.425 (0.032)	0.414 (0.057)	0.407 (0.073)

Note. Calculations are for the 6-hr accumulated rainfall, the numbers in the parentheses are the NMSE skills scores with respect to the REF experiment, that is, NMSE skill of GLM = 1-NMSE(GLM)/NMSE(REF).

Table 3
Mean Bias (mm) of Each Experiment Relative to the Stage IV Product, Averaged Over the Whole CONUS, the Western One Third and Eastern Two Thirds of CONUS

Domain	REF	GLM	RAD	GLM + RAD
CONUS	0.152	0.328	0.154	0.321
Western one third	0.036	0.073	0.028	0.064
Eastern two thirds	0.183	0.437	0.188	0.429

Note. Calculations are for the 6-hr accumulated rainfall.

each experiment averaged over: the whole CONUS, the eastern two thirds of CONUS, and the western one third of CONUS. The NMSE skill score of experiment X relative to experiment REF is given by:

$$\text{NMSE skill core}\left(X\right)=1-\text{NMSE}\left(X\right)/\text{NMSE}\left(\text{REF}\right) \quad (8)$$

where X refers to either GLM, RAD, or GLM + RAD.

Overall, GLM and radar data help reduce the QP deviations from the Stage IV product, with GLM + RAD showing the smallest error over the whole CONUS and, by extension, the eastern two-third of CONUS. For the western one third of CONUS, due to the relatively smaller amount of reliable radar data, RAD produces similar forecasts as REF. This analysis reveals that GLM generally slightly outperforms RAD and GLM + RAD over the western one third of CONUS, highlighting the added value of GLM lightning data over radar-sparse areas. Complementing the NMSE statistics, Table 3 underscores that the forecast skill and NMSE improvements are accompanied by CONUS-wide aggregate wet biases on the order of 0.32 mm for the GLM-based experiments, which are largely attributed to the excess rainfall induced by the LDA in the lightning-active, eastern 2/3rd region of CONUS. Biases in this region are about an order of magnitude larger there compared to the western 1/3rd region (Table 3). Auxiliary analysis (not shown) revealed that the positive differences in 29-day aggregate rainfall in excess of 50 mm for the GLM-based runs (Figures 7c and 7e)—nearly all of which collocated within regions of GLM hourly flash rates exceeding ~2 (Figure 7a)—were generally associated with positive differences in z = 3–7 km or z = 1–5 km layer averaged q_v (29-day aggregate) on the order of 0.1–0.45 g kg^{-1} (not shown).

Hovmöller (time-longitude) diagrams of latitudinally averaged hourly precipitation fields aggregated over all 29 forecast days were constructed for each of the four experiments and the Stage IV QPE. For reference, the latitudinally averaged total GLM flash density used for the DA between 2250 and 0000 UTC is also examined. As anticipated, the Hovmöller diagrams corroborate the results from the spatial rainfall maps in Figure 7 relatively well. All four experiments exhibit a longitudinal distribution of precipitation forecasts that is overall consistent with the Stage IV estimates, with heavy precipitation dominating the eastern two thirds of the United States. The underestimation of rainfall by REF over western Texas discussed earlier (Figure 7b) can also be seen here (Figure 8b) together with the slight improvements offered by RAD (Figure 8d). Most importantly and corroborating the results in Table 3, this analysis highlights even further the tendency of the GLM DA to generate notably more precipitation during the first 2–3 h forecasts (Figures 8c and 8e), especially over the eastern two thirds of CONUS where the bulk of the observed lightning activity used for the DA occurred (Figures 8a and 8f).

In sharp contrast to the QPFs over the eastern two thirds of CONUS (Figure 8f), notably less precipitation and, thus, lightning activity occurred in the western one third of CONUS (Figure 9). A closer inspection of the Hovmöller diagrams over the western one third of CONUS (Figure 9) illustrates that—similar to the eastern two thirds of CONUS (Figure 8)—the GLM DA adds more precipitation during the first 2–3 hr forecasts. The Hovmöller plots over the west (Figure 9), however, illustrate that notably less rainfall was added overall by the GLM-based experiment (consistent with Table 3), owing to the weaker convective nature of the storms over the west (and, hence, a smaller areal coverage of the lightning density field). When comparing the rainfall amounts to the Stage IV QPE over the CONUS area west of 121°W, an underestimation is noted in all four experiments. Given that the Stage IV amounts themselves likely represent an underestimate of the true rainfall amounts owing to poor radar coverage (beam blockage), it is likely that the QPF underestimates might be even more pronounced than shown in Figure 9. This analysis illustrated that while the QPFs from the GLM-based experiments are generally overestimated over most of the lightning-active areas in the eastern two thirds of CONUS, the GLM-based runs underestimate the QPFs over the western one third of the country, especially areas west of 121°W–highlighting the inherent difficulties in developing an optimal observation operator.

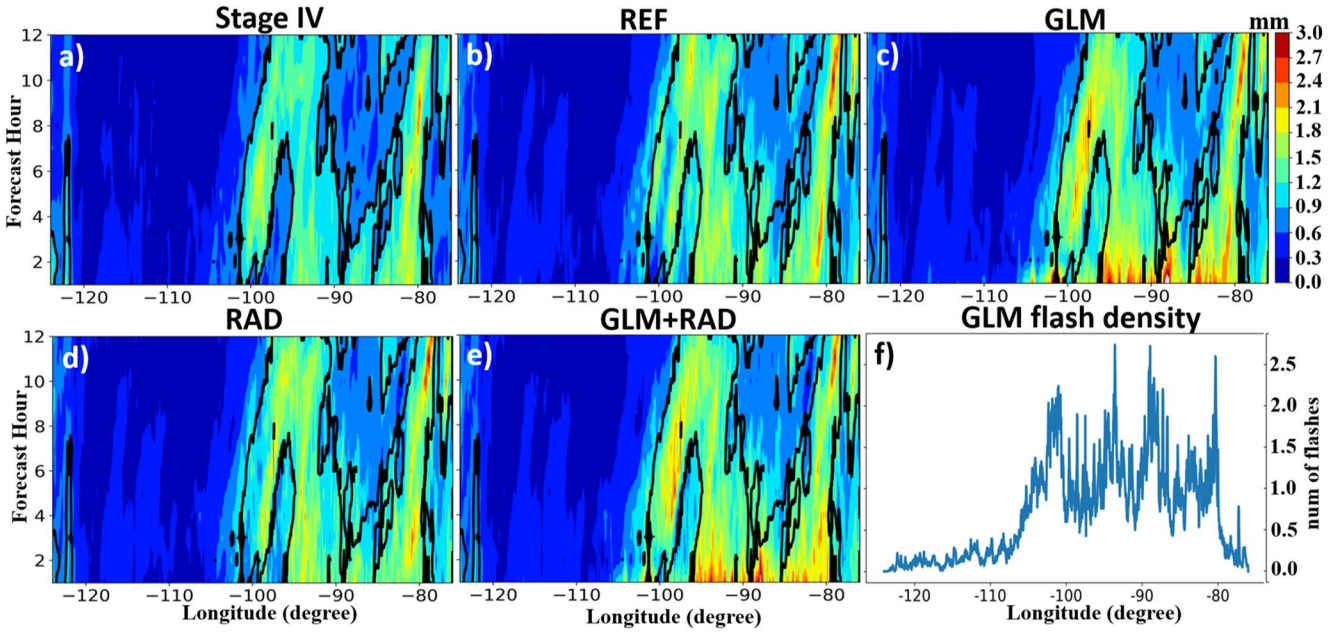

Figure 8. Time longitude (Hovmöller) diagrams of latitudinally-averaged precipitation averaged over all 29 cases along the longitudinal belt outlined by the gray-shading area (mask) in Figure 1a for (a) Stage IV rainfall estimates, (b) REF, (c) GLM, (d) RAD, and (e) GLM + RAD. The latitudinally averaged GLM flash density used during the DA between 2250 and 0000 UTC are provided in (f), for reference.

4.3. Individual Cases

Complementing the general evaluation and aggregate statistics in the previous sections, this analysis is aimed at examining in greater detail individual representative cases on the regional scale in both the radar-rich (eastern two thirds of CONUS) and radar-sparse regions (western one third of CONUS) of the CLUE domain. Emphasis is placed here on the CREF and QPFs evaluated against observations from the MRMS and Stage IV product, respectively. For rainfall, accumulated total GLM flash density maps up to the

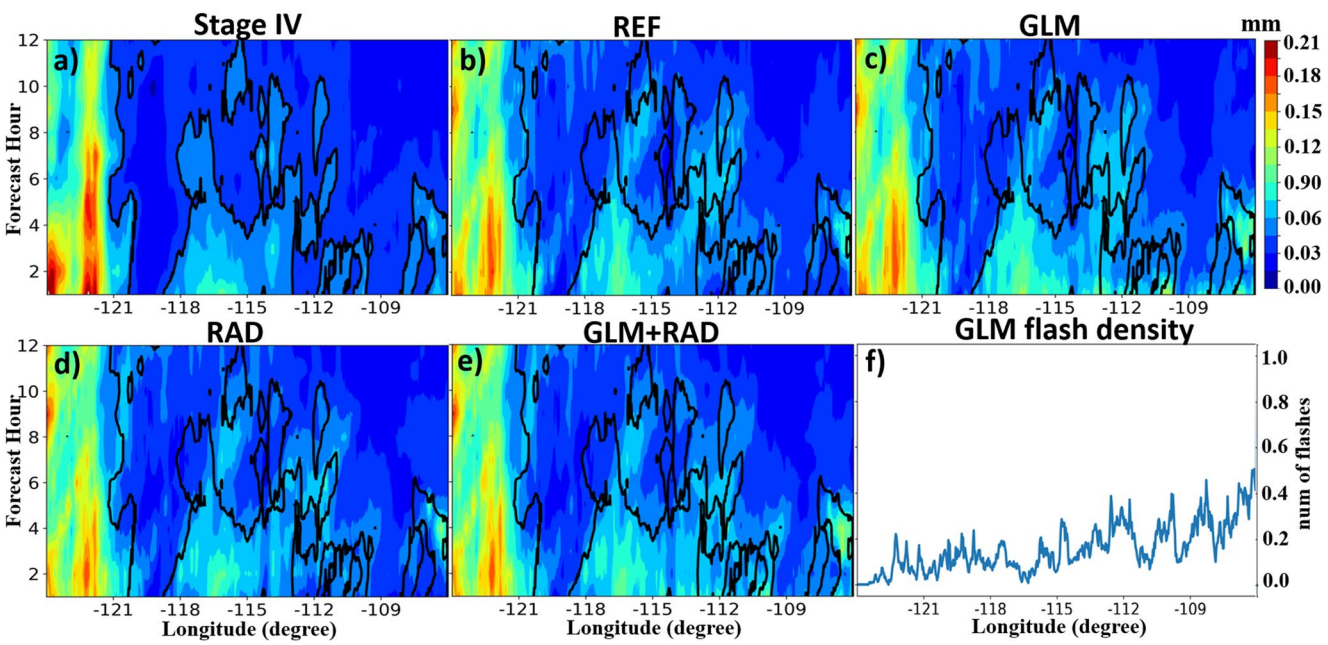

Figure 9. As in Figure 8, but aggregated over the western one third of the CONUS for the longitudinal belt between −124° and −106°.

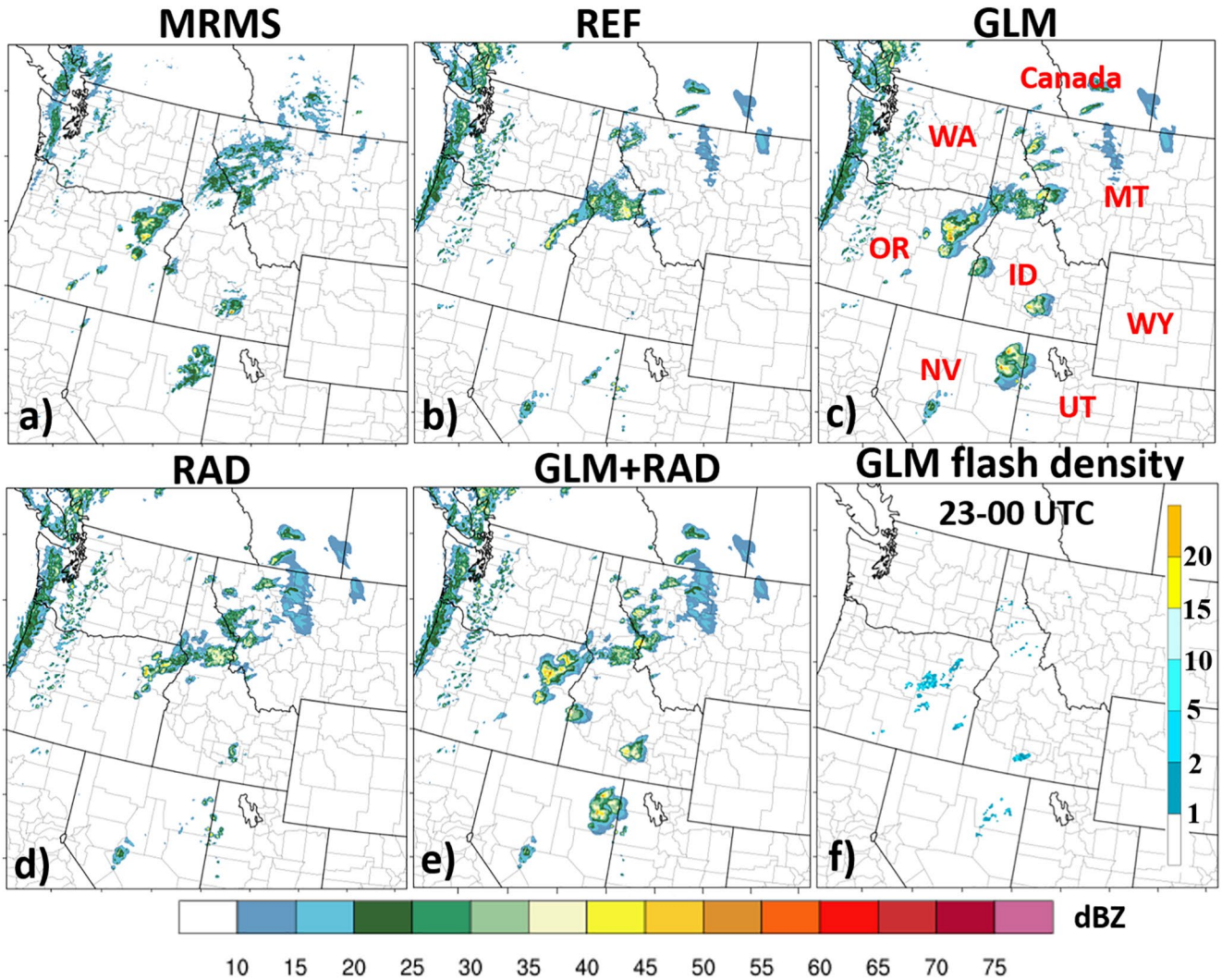

Figure 10. MRMS composite reflectivity (CREF) observations (a) and forecast CREF fields for (b) REF, (c) GLM, (d) RAD, and (e) GLM + RAD at 0100 UTC (1-hr forecast) for 30 April 2020. The GLM flash density used during the DA between 2250 UTC 29 April 2020 and 0000 UTC 30 April 2020 are also shown in (f), for reference. The standard US states abbreviations are also indicated in (c) for reference.

forecast time were used as reference to pinpoint swaths where the heaviest rainfall is expected to occur. To better identify areas where storms were most likely triggered by the LDA, the CREF analysis employs maps of the total GLM flash densities used during the DA (2250–0000 UTC)

4.3.1. 30 April 2020 Case Over the Northwestern United States

On 30 April 2020, widespread convection developed over the northwestern states including Oregon, Idaho, Montana, and Nevada (Figure 10a).

For this particular case, the REF run shows an overall poor forecast performance especially over south-central/southwestern Idaho and northeastern Nevada. Although some storm cells were forecasted by REF in northwestern Idaho and northeastern Oregon, their structure, orientation and areal coverage differ notably from the MRMS CREF observations (Figure 10b). Due to an overall sparser radar coverage over the northwestern U.S., RAD showed negligible impact in forecasting the observed convection relative to REF (Figure 10d) and this despite being able to partially capture some of the storms near the Idaho and Montana border as well as southcentral Idaho. For this radar-data sparse region, Figure 10f highlights the benefit of the GLM, which is able to monitor a reasonable portion of the lightning activity there (Rudlosky et al., 2019) and, hence, the presence of deeper mixed-phased convection. By adjusting the background moisture around

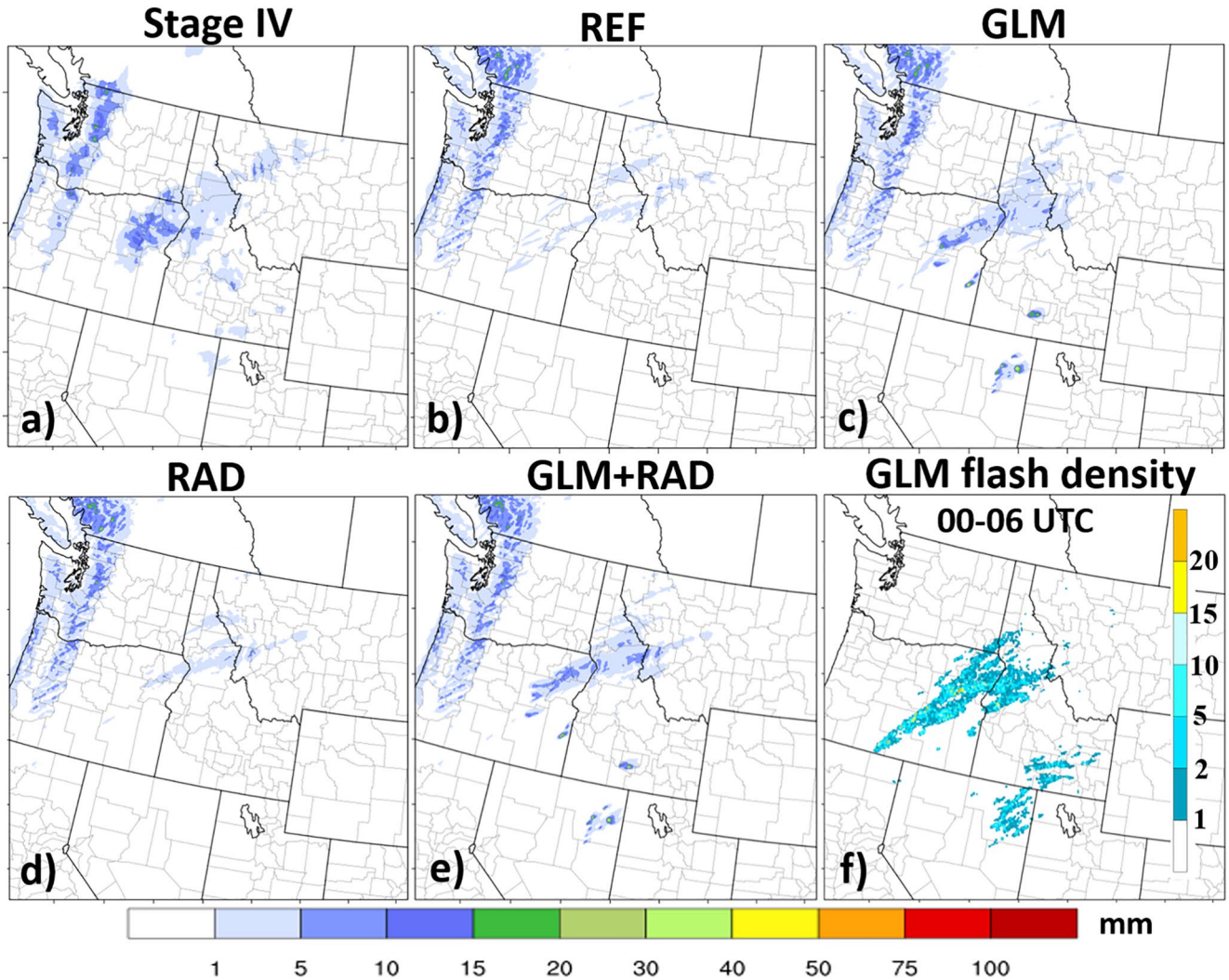

Figure 11. The total 6-hr (0000 UTC to 0600 UTC 30 April 2020) QPE for (a) Stage IV, and QPFs for (b) REF, (c) GLM, (d) RAD, and (e) GLM + RAD. The 6-hr accumulated (total) GLM flash densities during the 6-hr QP accumulation period are also shown to highlight areas of convectively active weather and, hence, heaviest rainfall potential.

lightning locations, the GLM-based DA experiments are able to notably improve the forecast (Figures 10c and 10e). The shape, location, orientation, and areal coverage agree overall remarkably well with the observed storms. When establishing a direct comparison with the MRMS CREF observations, one notices a relative overestimation in the CREF fields, which might partially be attributed to underestimates in the observed CREF fields due to beam blockage.

As also revealed by the aggregate statistics, the QPF performance is in accord with that of the CREF forecasts. Both REF and RAD severely underestimated the 6-hr accumulated precipitation over the northwestern CONUS (Figures 11b and 11d). In contrast, the two GLM-based experiments notably improved the forecast skill for 0–6 hr precipitation (Figures 11c and 11e), as evidenced by a relatively good agreement with the Stage IV estimates in terms of amounts and location. The location and areal coverage of the 6-hr QPF from both GLM-based runs agree well with the 0–6 hr accumulated GLM flash density fields, underlining further that, for this case, the improvements in 6-hr QPFs were primarily attributed to the inherent advantages of the GLM over mountainous terrain.

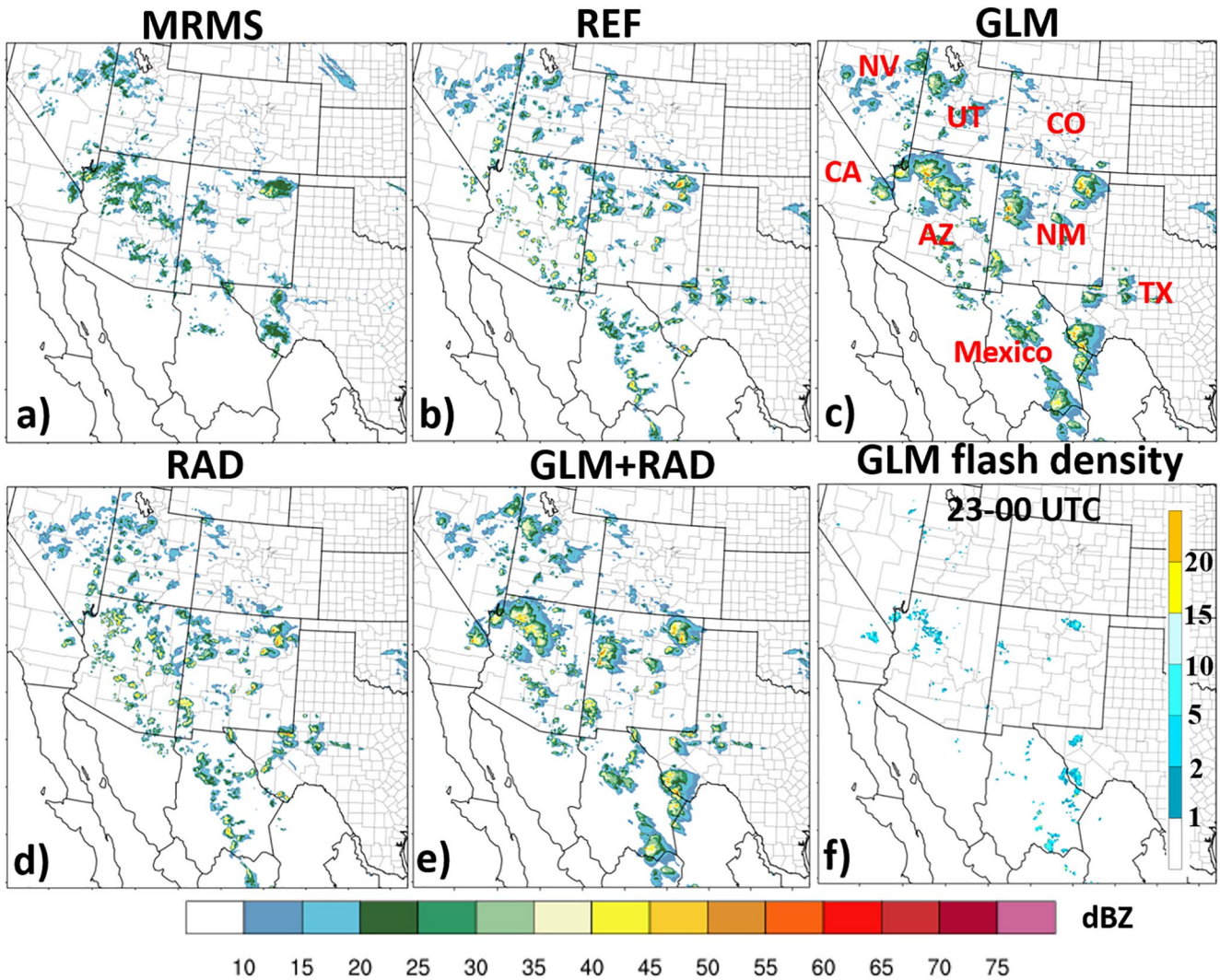

Figure 12. As in Figure 10 but for 11 May 2020.

4.3.2. 11 May 2020 Case Over the Southwestern United States

On 11 May 2020, scattered storms formed across the southwestern U.S., over portions of Nevada, southeastern California, Arizona, Utah, New Mexico and western Texas (Figure 12a). Over that broad area, REF and RAD only produce scattered weak radar echoes (Figures 12b and 12d). When assimilating the GLM observations (Figure 12f), the simulation is able to capture most of the observed storms (Figures 12c and 12e) despite a slight overestimation of areal coverage and intensity (Figures 12c and 12e). Similar to 30 April 2020, this overestimation might be attributed to an underestimation of the observed CREF values due to beam blockage

All DA experiments were generally able to capture the observed 6-hr rainfall over northeastern New Mexico. Over northwestern Arizona, however, REF and RAD underestimate their amounts (Figures 13b and 13d) while, in contrast, the GLM-based induced QPFs are slightly overpredicted (Figures 13c, 13e and 13f). Similar to CREF, it is reasonable to assume that—despite the lack of observations for verification—the QPFs over northeastern Mexico are more faithfully captured by the GLM-based experiments than either REF or RAD given the abundance of lightning detected by the GLM there and the lack of convective CREFs produced by either RAD or REF.

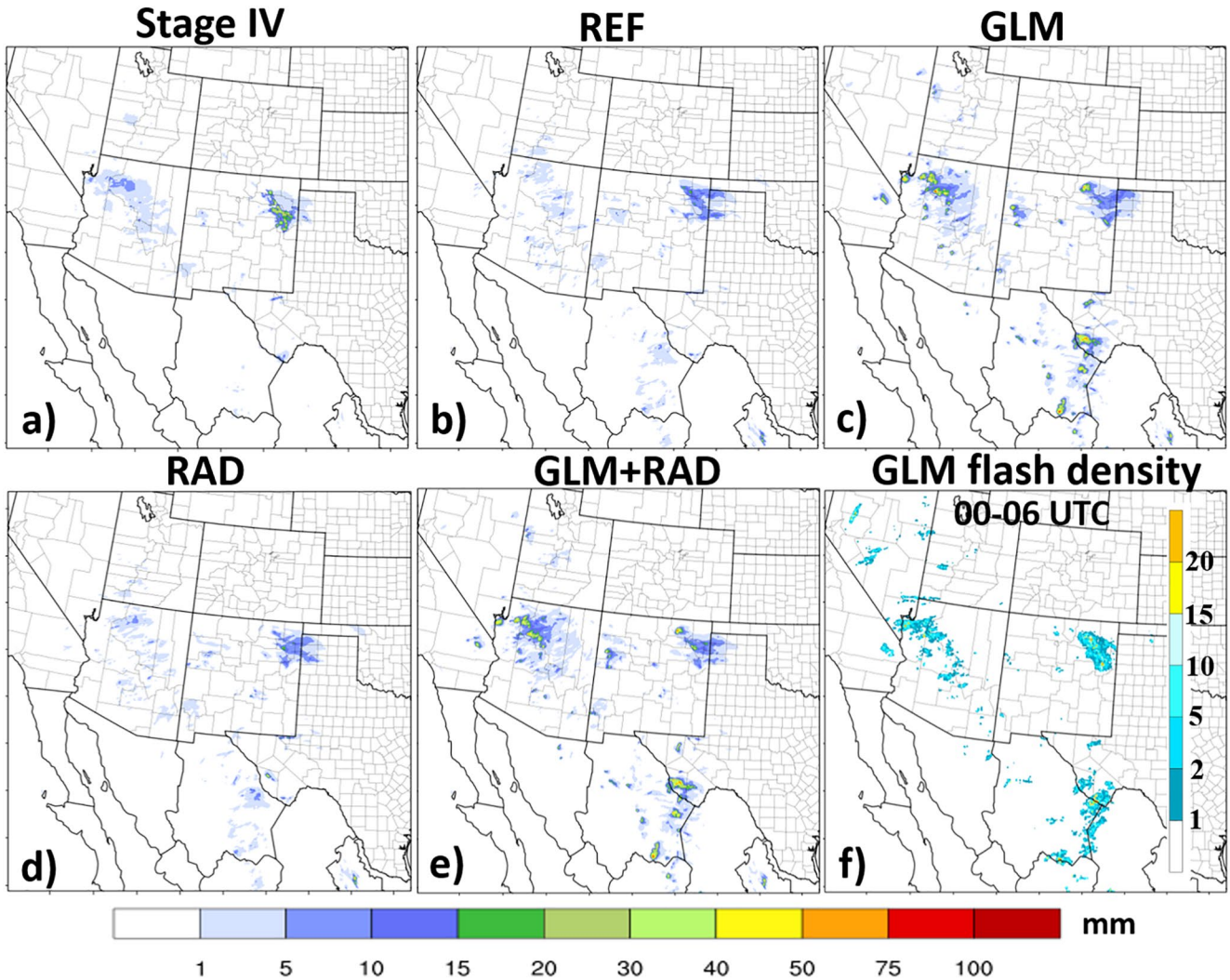

Figure 13. As in Figure 11 but for 11 May 2020.

4.3.3. 6 May 2020 Case Over Eastern CONUS

The above two cases provided illustrative examples of the added benefit of GLM total lightning data in radar sparse areas. Although relatively infrequent during this SFE, the following case will illustrate that even in areas characterized by good radar coverage (eastern two thirds of CONUS in Figure 1), the inclusion of the GLM was able on some rarer occasions to still prove beneficial.

During the early evening hours on 6 May 2020, strong storms affected most of the Carolinas. The 6-hr QPF and 3-hr CREF forecasts (Figures 14 and 15) highlight rather nicely the superiority of the GLM-based experiment for this case. The GLM DA is able to better resolve the upscale growth of the quasi-linear convective system (QLCS) that affected North Carolina by 0300 UTC, despite a slight northward displacement bias relative to the MRMS CREF observations. Both GLM experiments overestimate the maximum CREF (Figures 14c and 14e), which accounts for the 6-hr QPF overestimation (Figures 15c and 15e). Despite good radar coverage, the GLM DA is still able to add benefit to the QPF further by adjusting the moisture (Fierro et al., 2014), especially across the South Carolina/North Carolina border (Figures 15c and 15e).

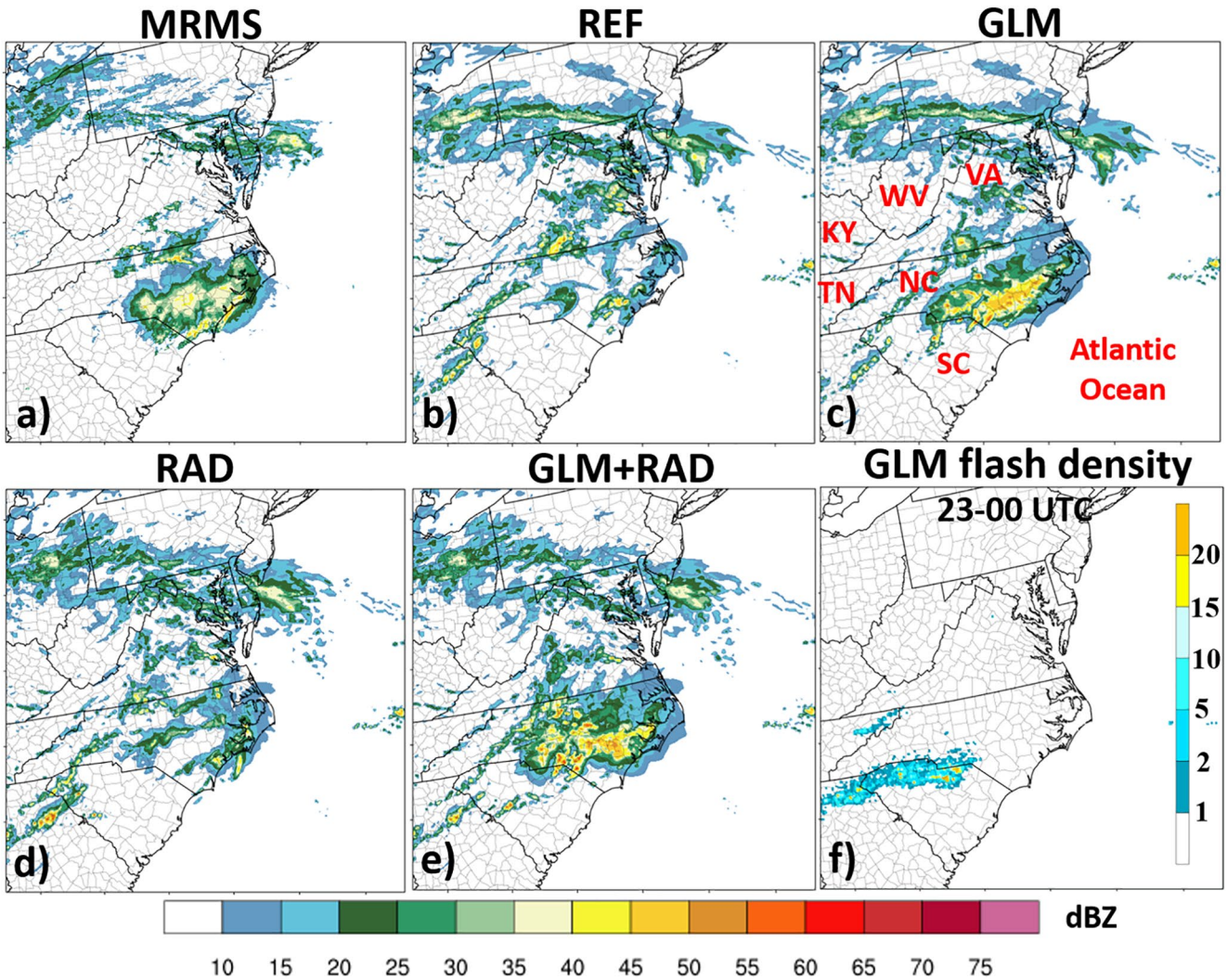

Figure 14. As in Figure 10 but for 6 May 2020 at 3-hr forecast.

5. Summary and Future Work

To better leverage the spatially uniform and high temporal frequency total lightning data provided by the GOES-16/17 GLM toward short term forecasting of convective events, this work analyses real time, convection-allowing, CONUS-scale DA simulations performed during the 5-week period of the 2020 SFE. These experiments were designed to mimic the operational settings of the HRRRv4 by using the same version of the WRF-ARW forecast model and same physics suite adopted in the HRRRv4. With the chief goal of gauging the added benefit of GLM lightning data over conventional WSR-88D radar data during the DA, this work assessed short-term (≤6h) rainfall forecast skill over the whole CONUS for a total of 29 forecast days, with a particular emphasis geared toward areas characterized by poorer radar coverage (i.e., western one third of the CONUS). Toward this goal, four experiments were devised: the GLM (assimilating GLM lightning data only), RAD (assimilating radar data only), GLM + RAD (assimilating both GLM lightning and radar data) and the no DA or "reference" experiment (REF). Because the experiments were intended to mimic operational settings of the HRRRv4, REF already contains, by design, blended information from a broad array of observing platforms (including lightning and radar) inherited from the RAPv5 background data. The QPF evaluation was performed against NCEP's Stage IV rainfall estimates.

During the 5-week SFE period, significantly larger amount of precipitation occurred over the radar-rich eastern two thirds of the CONUS compared to the radar-poor western one third region of the CONUS.

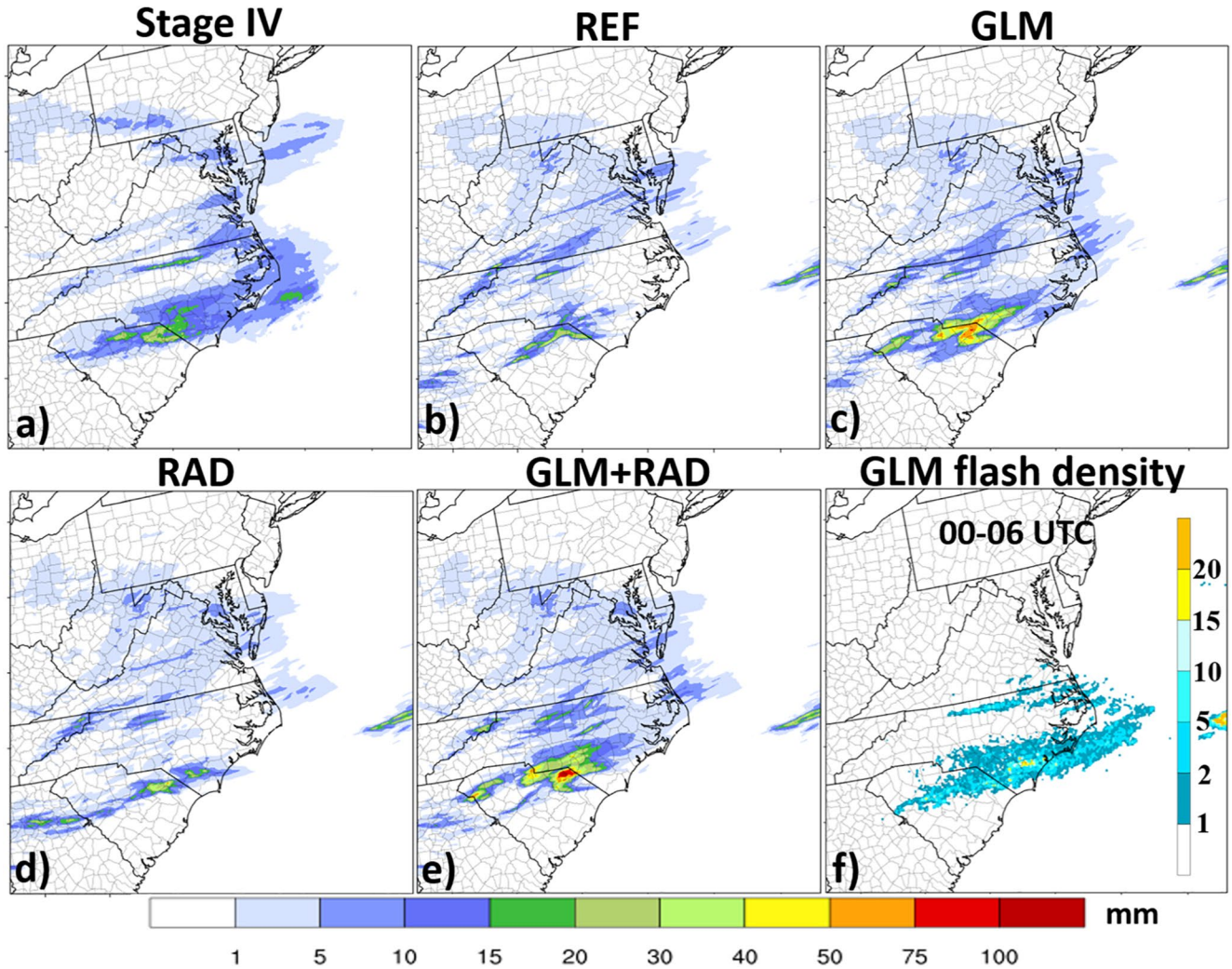

Figure 15. As in Figure 11 but for 6 May 2020.

Contingency elements and associated score metrics aggregated over all 29 forecast days show a general improvement in forecast skill over REF for all DA runs; with the best results obtained for GLM + RAD. These improvements were noted despite a generally good performance of REF, which is attributed to the usage of the RAPv5 dataset to derive its background fields. Despite this improvement in skill, the GLM-based DA runs produced overall larger-than-observed precipitation amounts over the eastern two thirds of CONUS, which was characterized by abundant lightning activity during the DA period (2300–0000 UTC) relative to the western one third of CONUS. Over the west, notably less rainfall was produced by the GLM experiments, with amounts generally remaining in closer agreement with the observations when taking into account underestimates in observed rainfall amounts due to beam blockage.

The performance of the GLM based runs was also assessed by means of Hovmöller diagrams for rainfall and through a more detailed examination of three illustrative individual cases: two in radar data sparse regions over the west (30 April and 11 May) and one over the radar-richer regions of the east (6 May). All demonstrated that, despite an overestimation of rainfall amounts over the eastern CONUS, the GLM helped alleviate some of the systematic underestimates in 1–3 hr QPFs over the western CONUS (i.e., west of 121°W) and produced the smallest NMSE overall. The simultaneous overestimation of the QPFs over the eastern two thirds and underestimation over the western CONUS, however, illustrates the inherent challenges in deriving an optimal observation operator for lightning assimilation when solely considering adjustments to the background water vapor mass mixing ratio.

As a potential remedy, future research efforts should consider resorting to another LDA strategy and/or alternative lightning observations operators with focus on those able to assimilate zero lightning observations (e.g., Kong et al., 2020). Additionally, these research efforts should take into consideration the areal extent of flashes by assimilating the flash extent density, and/or combine the GLM with additional datasets to help better address the development of spurious convection during the DA.

Data Availability Statement

For this work, the GLM data were retrieved from tape archives available on high performance computing resources (HPCs) at NOAA but could also be obtained publicly from multiple public sources (e.g., Comprehensive Large Array-data Stewardship System or CLASS, NOAA, 2020). Similarly, The WSR-88D Level-II data (reflectivity factor and radial velocity) used during the assimilation were made available in real time on NOAA HPCs but are made freely accessible to the public by the National Center of Environmental information at http://www.ncdc.noaa.gov/. The Stage IV QPF data are publicly available at https://www.eol.ucar.edu. All data and source codes generated in the work are stored locally on NOAA (NSSL) and OU computers for at least 3 years after the project has concluded, as require by NOAA. The WRF source code (Skamarock & Klemp, 2008; Skamarock et al., 2008) is made publicly available by NCAR/UCAR (https://github.com/wrf-model/WRF) and the developmental NEWS3DVAR software (Gao et al., 2013; Wang et al., 2019; Hu et al., 2020) can be made available to interested collaborators by contacting Yunheng Wang (Yunheng.Wang@noaa.gov) or obtained directly at https://bitbucket.org/newe3dvar/newe3dvar/src/develop. Owing to the large size (~65Tb) of all the simulation output files used, a subset of these data is freely provided to any interested researchers and collaborators as netCDF files through the following ftp link: ftp.nssl.noaa.gov:/users/yhwang/GLM/.

Acknowledgments
Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce. This work was further supported by the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce under Grants NOAA-NWS-NWSPO-2018-2005317 (Award No. NA18NWS4680063). The simulations were conducted on the NOAA HPC 'Jet' resources housed in Boulder, CO. Auxiliary computer resources were provided by the Oklahoma Supercomputing Center for Education and Research (OSCER) hosted at the University of Oklahoma. Thanks also go out to Thomas Jones for his valuable comments and suggestions on an earlier version of the manuscript.

References

Alexander, G. D., Weinman, J. A., Karyampudi, V. M., Olson, W. S., & Lee, A. C. L. (1999). The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 superstorm. *Monthly Weather Review*, *127*(7), 1433–1457. https://doi.org/10.1175/1520-0493(1999)127<1433:teoarr>2.0.co;2

Allen, B. J., Mansell, E. R., Dowell, D. C., & Deierling, W. (2016). Assimilation of pseudo-GLM data using the ensemble Kalman filter. *Monthly Weather Review*, *144*(9), 3465–3486. https://doi.org/10.1175/mwr-d-16-0117.1

Baldwin, M. E., & Mitchell, K. E. (1997). The NCEP hourly multisensor US precipitation analysis for operations and GCIP research. *Preprints, 13th Conference On Hydrology, Long Beach, CA* (Vol. *5455*), American Meteorlogical Society.

Benjamin, S. G., Dévényi, D., Weygandt, S. S., Brundage, K. J., Brown, J. M., Grell, G. A., et al. (2004). An hourly assimilation–forecast cycle: The RUC. *Monthly Weather Review*, *132*(2), 495–518. https://doi.org/10.1175/1520-0493(2004)132<0495:ahactr>2.0.co;2

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Monthly Weather Review*, *144*(4), 1669–1694. https://doi.org/10.1175/mwr-d-15-0242.1

Chang, D. E., Weinman, J. A., Morales, C. A., & Olson, W. S. (2001). The effect of spaceborne microwave and ground-based continuous lightning measurements on forecasts of the 1998 Groundhog Day storm. *Monthly Weather Review*, *129*(8), 1809–1833. https://doi.org/10.1175/1520-0493(2001)129<1809:teosma>2.0.co;2

Clark, A., Jirak, I., Gallo, B. T., Roberts, B., Dean, A., Knopfmeier, K., et al. (2020). *Spring forecasting experiment 2020: Preliminary results and findings* (p. 77). NOAA/NSSL/SPC. Retrieved from https://hwt.nssl.noaa.gov/sfe/2020/docs/HWT_SFE_2020_Prelim_Findings_FINAL.pdf

Clark, A. J., Gallus, W. A., Jr, & Weisman, M. L. (2010). Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Weather and Forecasting*, *25*(5), 1495–1509. https://doi.org/10.1175/2010waf2222404.1

Clark, A. J., Jirak, I. L., Dembek, S. R., Creager, G. J., Kong, F., Thomas, K. W., et al. (2018). The community leveraged unified ensemble (CLUE) in the 2016 NOAA/hazardous weather testbed spring forecasting experiment. *Bulletin of the American Meteorological Society*, *99*(7), 1433–1448. https://doi.org/10.1175/bams-d-16-0309.1

Edgington, S., Tillier, C., & Anderson, M. (2019). Design, calibration, and on-orbit testing of the geostationary lightning mapper on the GOES-R series weather satellite. In: *International Conference on Space Optics—ICSO 2018* (Vol. *11180*, p. 1118040), International Society for Optics and Photonics. https://doi.org/10.1117/12.2536063

Fierro, A. O., Clark, A. J., Mansell, E. R., MacGorman, D. R., Dembek, S. R., & Ziegler, C. L. (2015). Impact of storm-scale lightning data assimilation on WRF-ARW precipitation forecasts during the 2013 warm season over the contiguous United States. *Monthly Weather Review*, *143*(3), 757–777. https://doi.org/10.1175/mwr-d-14-00183.1

Fierro, A. O., Gao, J., Ziegler, C. L., Calhoun, K. M., Mansell, E. R., & MacGorman, D. R. (2016). Assimilation of flash extent data in the variational framework at convection-allowing scales: Proof-of-concept and evaluation for the short-term forecast of the 24 May 2011 tornado outbreak. *Monthly Weather Review*, *144*(11), 4373–4393. https://doi.org/10.1175/mwr-d-16-0053.1

Fierro, A. O., Gao, J., Ziegler, C. L., Mansell, E. R., MacGorman, D. R., & Dembek, S. R. (2014). Evaluation of a cloud-scale lightning data assimilation technique and a 3DVAR method for the analysis and short-term forecast of the 29 June 2012 derecho event. *Monthly Weather Review*, *142*(1), 183–202. https://doi.org/10.1175/mwr-d-13-00142.1

Fierro, A. O., Mansell, E. R., Ziegler, C. L., & MacGorman, D. R. (2012). Application of a lightning data assimilation technique in the WRF-ARW model at cloud-resolving scales for the tornado outbreak of 24 May 2011. *Monthly Weather Review*, *140*(8), 2609–2627. https://doi.org/10.1175/mwr-d-11-00299.1

Fierro, A. O., Stevenson, S. N., & Rabin, R. M. (2018). Evolution of GLM-observed total lightning in Hurricane Maria (2017) during the period of maximum intensity. *Monthly Weather Review*, *146*(6), 1641–1666. https://doi.org/10.1175/mwr-d-18-0066.1

Fierro, A. O., Wang, Y., Gao, J., & Mansell, E. R. (2019). Variational assimilation of radar data and GLM lightning-derived water vapor for the short-term forecasts of high-impact convective events. *Monthly Weather Review*, *147*(11), 4045–4069. https://doi.org/10.1175/mwr-d-18-0421.1

Fierro, A. O., Zhao, G., Liu, S., Wang, Y. J., Gao, K. C., Calhoun, K., et al. (2018). Assimilation of total lightning with GSI and NEWS3DVAR to improve short-term forecasts of high impact weather events at cloud resolving scales. *JCSDA Quarterly Newsletter*, *58*, 5–12.

Gao, J., Smith, T. M., Stensrud, D. J., Fu, C., Calhoun, K., Manross, K. L., et al. (2013). A real-time weather-adaptive 3DVAR analysis system for severe weather detections and warnings. *Weather and Forecasting*, *28*(3), 727–745. https://doi.org/10.1175/waf-d-12-00093.1

Gao, J., & Stensrud, D. J. (2012). Assimilation of reflectivity data in a convective-scale, cycled 3DVAR framework with hydrometeor classification. *Journal of the Atmospheric Sciences*, *69*(3), 1054–1065. https://doi.org/10.1175/jas-d-11-0162.1

Gao, J., Xue, M., Brewster, K., & Droegemeier, K. K. (2004). A three-dimensional variational data analysis method with recursive filter for Doppler radars. *Journal of Atmospheric and Oceanic Technology*, *21*(3), 457–469. https://doi.org/10.1175/1520-0426(2004)021<0457:atvdam>2.0.co;2

Goodman, S. J., Blakeslee, R. J., Koshak, W. J., Mach, D., Bailey, J., Buechler, D., et al. (2013). The GOES-R geostationary lightning mapper (GLM). *Atmospheric Research*, *125*, 34–49. https://doi.org/10.1016/j.atmosres.2013.01.006

Gourley, J. J., Hong, Y., Flamig, Z. L., Li, L., & Wang, J. (2010). Intercomparison of rainfall estimates from radar, satellite, gauge, and combinations for a season of record rainfall. *Journal of Applied Meteorology and Climatology*, *49*(3), 437–452. https://doi.org/10.1175/2009jamc2302.1

Gurka, J. J., Schmit, T. J., Renkevens, T. M., Gunshor, M. M., & Li, J. (2006). 2006 update on baseline instruments for the GOES-R series. In *Atmospheric and environmental remote sensing data processing and utilization II: Perspective on calibration/validation initiatives and strategies* (Vol. *6301*, p. 63010H), International Society for Optics and Photonics. https://doi.org/10.1117/12.683701

Hu, J., Fierro, A. O., Wang, Y., Gao, J., & Mansell, E. R. (2020). Exploring the assimilation of GLM-derived water vapor mass in a cycled 3DVAR framework for the short-term forecasts of high-impact convective events. *Monthly Weather Review*, *148*(3), 1005–1028. https://doi.org/10.1175/mwr-d-19-0198.1

Hu, M., Xue, M., & Brewster, K. (2006). 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part I: Cloud analysis and its impact. *Monthly Weather Review*, *134*(2), 675–698. https://doi.org/10.1175/mwr3092.1

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*(D13). https://doi.org/10.1029/2008jd009944

Jones, C. D., & Macpherson, B. (1997a). A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, *4*(3), 269–277. https://doi.org/10.1017/s1350482797000522

Jones, C. D., & Macpherson, B. (1997b). Sensitivity of the limited area model to the assimilation of precipitation estimates derived from lightning data. *UKMO Forecasting Research Technical Report*, *212*(11).

Kong, R., Xue, M., Fierro, A. O., Jung, Y., Liu, C., Mansell, E. R., & MacGorman, D. R. (2020). Assimilation of GOES-R geostationary lightning mapper flash extent density data in GSI EnKF for the analysis and short-term forecast of a mesoscale convective system. *Monthly Weather Review*, *148*(5), 2111–2133. https://doi.org/10.1175/mwr-d-19-0192.1

Lai, A., Gao, J., Koch, S. E., Wang, Y., Pan, S., Fierro, A. O., et al. (2019). Assimilation of radar radial velocity, reflectivity, and pseudo–water vapor for convective-scale NWP in a variational framework. *Monthly Weather Review*, *147*(8), 2877–2900. https://doi.org/10.1175/mwr-d-18-0403.1

Lin, Y., & Mitchell, K. E. (2005). the NCEP stage II/IV hourly precipitation analyses: Development and applications. In *Proceedings of the 19th Conference Hydrology* (Vol. *10*). American Meteorological Society.

Mach, D. M. (2020). Geostationary lightning mapper clustering algorithm stability. *Journal of Geophysical Research: Atmospheres*, *125*(5), e2019JD031900. https://doi.org/10.1029/2019jd031900

Mach, D. M., Christian, H. J., Blakeslee, R. J., Boccippio, D. J., Goodman, S. J., & Boeck, W. L. (2007). Performance assessment of the optical transient detector and lightning imaging sensor. *Journal of Geophysical Research*, *112*(D9). https://doi.org/10.1029/2006jd007787

Maddox, R. A., Zhang, J., Gourley, J. J., & Howard, K. W. (2002). Weather radar coverage over the contiguous United States. *Weather and Forecasting*, *17*(4), 927–934. https://doi.org/10.1175/1520-0434(2002)017<0927:wrcotc>2.0.co;2

Mansell, E. R. (2014). Storm-scale ensemble Kalman filter assimilation of total lightning flash-extent data. *Monthly Weather Review*, *142*(10), 3683–3695. https://doi.org/10.1175/mwr-d-14-00061.1

Mansell, E. R., Ziegler, C. L., & MacGorman, D. R. (2007). A lightning data assimilation technique for mesoscale forecast models. *Monthly Weather Review*, *135*(5), 1732–1748. https://doi.org/10.1175/mwr3387.1

Marchand, M. R., & Fuelberg, H. E. (2014). Assimilation of lightning data using a nudging method involving low-level warming. *Monthly Weather Review*, *142*(12), 4850–4871. https://doi.org/10.1175/mwr-d-14-00076.1

Nakanishi, M., & Niino, H. (2006). An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, *119*(2), 397–407. https://doi.org/10.1007/s10546-005-9030-8

Nelson, B. R., Prat, O. P., Seo, D. J., & Habib, E. (2016). Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. *Weather and Forecasting*, *31*(2), 371–394. https://doi.org/10.1175/waf-d-14-00112.1

Olson, J. B., Kenyon, J. S., Angevine, W., Brown, J. M., Pagowski, M., & Sušelj, K. (2019). *A description of the MYNN-EDMF scheme and the coupling to other components in WRF–ARW*.

Pan, S., Gao, J., Stensrud, D. J., Wang, X., & Jones, T. A. (2018). Assimilation of radar radial velocity and reflectivity, satellite cloud water path, and total precipitable water for convective-scale NWP in OSSEs. *Journal of Atmospheric and Oceanic Technology*, *35*(1), 67–89. https://doi.org/10.1175/jtech-d-17-0081.1

Papadopoulos, A., Chronis, T. G., & Anagnostou, E. N. (2005). Improving convective precipitation forecasting through assimilation of regional lightning measurements in a mesoscale model. *Monthly Weather Review*, *133*(7), 1961–1977. https://doi.org/10.1175/mwr2957.1

Pessi, A. T., & Businger, S. (2009). The impact of lightning data assimilation on a winter storm simulation over the North Pacific Ocean. *Monthly Weather Review*, *137*(10), 3177–3195. https://doi.org/10.1175/2009mwr2765.1

Peterson, M. (2019). Research applications for the geostationary lightning mapper operational lightning flash data product. *Journal of Geophysical Research: Atmospheres*, *124*(17–18), 10205–10231. https://doi.org/10.1029/2019jd031054

Roebber, P. J. (2009). Visualizing multiple measures of forecast quality. *Weather and Forecasting*, *24*(2), 601–608. https://doi.org/10.1175/2008waf2222159.1

Rudlosky, S. D., Goodman, S. J., Virts, K. S., & Bruning, E. C. (2019). Initial geostationary lightning mapper observations. *Geophysical Research Letters*, *46*(2), 1097–1104. https://doi.org/10.1029/2018gl081052

Rudlosky, S. D., & Virts, K. S. (2021). Dual geostationary lightning mapper observations. *Monthly Weather Review*, *149*(4), 979–998. https://doi.org/10.1175/mwr-d-20-0242.1

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, *227*(7), 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., et al., (2008). *A description of the advanced research WRF Version 3* (pp. 113), NCAR Technical Note NCAR/TN-475+STR, https://doi.org/10.5065/D68S4MVH

Smalley, M., L'Ecuyer, T., Lebsock, M., & Haynes, J. (2014). A comparison of precipitation occurrence from the NCEP Stage IV QPE product and the CloudSat Cloud Profiling Radar. *Journal of Hydrometeorology*, *15*(1), 444–458. https://doi.org/10.1175/jhm-d-13-048.1

Smirnova, T. G., Brown, J. M., Benjamin, S. G., & Kenyon, J. S. (2016). Modifications to the rapid update cycle land surface model (RUC LSM) available in the weather research and forecasting (WRF) model. *Monthly Weather Review*, *144*(5), 1851–1865. https://doi.org/10.1175/mwr-d-15-0198.1

Smith, T. M., Lakshmanan, V., Stumpf, G. J., Ortega, K. L., Hondl, K., Cooper, K., et al. (2016). Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(9), 1617–1630. https://doi.org/10.1175/bams-d-14-00173.1

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*(12), 5095–5115. https://doi.org/10.1175/2008mwr2387.1

UCAR. (2020). *GCIP/EOP surface: Precipitation NCEP/EMC 4KM gridded data (GRIB) stage IV data*. Retrieved from https://data.eol.ucar.edu/dataset?21.093 Accessed 10 February 2020.

Wang, H., Liu, Y., Cheng, W. Y., Zhao, T., Xu, M., Liu, Y., et al. (2017). Improving lightning and precipitation prediction of severe convection using lightning data assimilation with NCAR WRF-RTFDDA. *Journal of Geophysical Research: Atmospheres*, *122*(22), 12–296. https://doi.org/10.1002/2017jd027340

Wang, Y., Gao, J., Skinner, P. S., Knopfmeier, K., Jones, T., Creager, G., et al. (2019). Test of a Weather-Adaptive Dual-Resolution Hybrid Warn-on-Forecast Analysis and Forecast System for Several Severe Weather Events. *Weather and Forecasting*, *34*(6), 1807–1827. https://doi.org/10.1175/waf-d-19-0071.1

Westcott, N. E., Knapp, H. V., & Hilberg, S. D. (2008). Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United States. *Journal of Hydrology*, *351*(1–2), 1–12. https://doi.org/10.1016/j.jhydrol.2007.10.057

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (Vol. *59*). Academic press.

Zhang, J., Howard, K., Langston, C., Vasiloff, S., Kaney, B., Arthur, A., et al. (2011). National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bulletin of the American Meteorological Society*, *92*(10), 1321–1338. https://doi.org/10.1175/2011bams-d-11-00047.1