Exploring the Assimilation of GLM-Derived Water Vapor Mass in a Cycled 3DVAR Framework for the Short-Term Forecasts of High-Impact Convective Events

Junjun Hu, Alexandre O. Fierro, and Yunheng Wang

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

Jidong Gao

NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

Edward R. Mansell

NOAA/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 14 June 2019, in final form 9 December 2019)

ABSTRACT

The recent successful deployment of the Geostationary Lightning Mapper (GLM) on board the Geostationary Operational Environmental Satellite R series (*GOES-16/17*) provides nearly uniform spatiotemporal measurements of total lightning (intracloud plus cloud to ground) over the Americas and adjacent vast oceanic regions. This study evaluates the potential value of assimilating GLM-derived water vapor mixing ratio on short-term (≤ 6 h), cloud-scale ($dx = 1.5$ km) forecasts of five severe weather events over the Great Plains of the United States using a three-dimensional variational (3DVAR) data assimilation (DA) system. Toward a more systematic assimilation of real GLM data, this study conducted sensitivity tests aimed at evaluating the impact of the horizontal decorrelation length scale, DA cycling frequency, and the time window size for accumulating GLM lightning observations prior to the DA. Forecast statistics aggregated over all five cases suggested that an optimal forecast performance is obtained when lightning measurements are accumulated over a 10-min interval and GLM-derived water vapor mixing ratio values are assimilated every 15 min with a horizontal decorrelation length scale of 3 km. This suggested configuration for the GLM DA together with companion experiments (i) not assimilating any data, (ii) assimilating radar data only, and (iii) assimilating both GLM and radar data were evaluated for the same five cases. Overall, GLM data have shown potential to help improve the short-term (<3 h) forecast skill of composite reflectivity fields and individual storm tracks. While this result also held for accumulated rainfall, longer-term (≥ 3 h) forecasts were generally characterized by noteworthy wet biases.

1. Introduction

Timely and accurate short-term convection-allowing (<3 km) numerical weather prediction (NWP) is critical for making informed decisions to warn the public and stakeholders of high-impact, and occasionally life-threatening, weather events. Increasing the lead times of such warnings is challenging given the complex, nonlinear physical processes governing the evolution of atmospheric weather phenomena across a large range of scales. Pioneering work by Lilly (1990) outlined the principal scientific challenges of the development of convective-scale NWPs and underlined that increasing computer power combined with advances in NWP parameterizations and data assimilation (DA) methods are expected to play critical roles in improving forecasts in the shorter and longer term (Stensrud et al. 2009, 2013).

In addition to NWP-related strides, the increasing availability of detailed datasets spanning a large range of scales played a significant role toward recent improvements of convective-scale severe weather forecasting. Various observational datasets at the meso and convective scales from in situ to remote sensing instruments have been used in convective-scale DA applications. These include, for instance, traditional observations

Corresponding author: Junjun Hu, junjun.hu@noaa.gov

(e.g., pressure, moisture, temperature, and wind) from surface instruments like mesonets and automated surface observing systems (e.g., Fujita et al. 2007; Knopfmeier and Stensrud 2013; Sobash and Stensrud 2015) and upper-air observation platforms such as rawinsonde, wind profiler, aircraft communications addressing and reporting systems (ACARS; e.g., Hitchcock et al. 2016; Coniglio et al. 2016), radial velocity and reflectivity from Weather Surveillance Radar-1988 Doppler (WSR-88D) (e.g., Snyder and Zhang 2003; Gao and Stensrud 2012; Yussouf et al. 2015; Johnson et al. 2015; Wang and Wang 2017), satellite measurements (e.g., Jones et al. 2016, 2018; Zhang et al. 2018), and combinations of these observations (e.g., Benjamin et al. 2016; Hu et al. 2017). Among these, Doppler radar data (i.e., radial velocity and reflectivity factor) were proven to be critical toward convective-scale forecast improvements. It is well established, however, that ground-based radar networks suffer from poor coverage in mountainous terrains due to beam blockage (Zhang et al. 2011; Maddox et al. 2002) and oceanic regions lying beyond the network's range. Ancillary detection of deep convection is critical for these regions, given that mountainous areas are susceptible to forest fires while some oceanic regions are prone to hurricane and tropical systems.

Technical advances in spaceborne instruments such as the Geostationary Operational Environmental Satellite R series (GOES-16/17; Goodman et al. 2013) provide an unprecedented opportunity for researchers and operational forecasters to monitor the atmosphere over vast regions devoid of conventional data at nearly uniform spatiotemporal resolution (e.g., Jones et al. 2018; Minamide and Zhang 2018; Zhang et al. 2018). The Geostationary Lightning Mapper (GLM) on board each of the GOES-16/17 satellites provides continuous detection of total lightning over most of the Western Hemisphere with high temporal (20 s) and spatial (8–12 km) resolution (Goodman et al. 2013). Many studies have shown that lightning activity, in general, is an unambiguous indicator of electrified mixed phase convection (Petersen and Rutledge 1998; Rudlosky and Fuelberg 2013; Makowski et al. 2013), especially total lightning (cloud to ground plus intracloud) (e.g., MacGorman et al. 1989; Carey and Rutledge 1998; Wiens et al. 2005; Kuhlman et al. 2006; Fierro et al. 2006; Deierling and Petersen 2008; MacGorman et al. 2011; Weiss et al. 2012; Medici et al. 2017).

Previous studies focusing on lightning DA have generally demonstrated positive impacts on NWP forecast primarily through functional relationships based on proxies exhibiting a well-established linkage with lightning, such as latent heating (Alexander et al. 1999;

Chang et al. 2001; Pessi and Businger 2009), specific humidity (Papadopoulos et al. 2005, 2009), the trigger function in convection parameterization schemes (Mansell et al. 2007; Lagouvardos et al. 2013; Giannaros et al. 2016; Heath et al. 2016), the water vapor mass mixing ratio (Fierro et al. 2012, 2016, 2019; Zhang et al. 2017), or radar reflectivity (Wang et al. 2014). In addition to simpler and computationally efficient nudging DA methods (Fierro et al. 2012; Marchand and Fuelberg 2014), variational (e.g., Fierro et al. 2016, 2019; Zhang et al. 2017) and ensemble-based approaches (e.g., Mansell 2014; Allen et al. 2016; H. Wang et al. 2018) have been investigated to assimilate lightning. To the best of our knowledge, there have not yet been any studies evaluating more systematically the impact of assimilating GLM-derived water vapor on short-term forecasts of high-impact convective events in the variational framework at the cloud-scale (~1.5 km) with high-frequency cycling (≤15 min).

Toward the goal of investigating how to best use GLM lightning measurements in a cycled variational (3DVAR) framework, five cases that occurred over the Great Plains of the United States in May 2018 were examined. The 3DVAR DA system that was initially developed at the Center for Analysis and Prediction of Storms (CAPS; Gao et al. 2004), and further updated at the National Severe Storms Laboratory (NSSL) is employed in this study (Gao and Stensrud 2012; Gao et al. 2013). Fierro et al. (2019) adapted the previous 3DVAR method for lightning (ground-based network) to use GLM data. The current study builds upon Fierro et al.'s (2019) proof-of-concept adaptation by performing a detailed sensitivity analysis of the impact of assimilating GLM-derived water vapor mixing ratio on short-term forecasts of storm-scale weather events by varying (i) the length of the accumulation interval of the GLM data prior to the DA, (ii) the frequency of DA cycling, and (iii) the scale of horizontal decorrelation. In light of the significant improvements in short-term forecasts benefiting from the assimilation of radar data (reflectivity factor and radial wind), GLM DA experiments are further evaluated against radar DA experiments.

2. Brief description of GLM-observed lightning data

As mentioned in the introduction, the GLM instrument mounted on board the GOES-16/17 satellites provides a unique opportunity to monitor total lightning over the Western Hemisphere with nearly uniform spatiotemporal resolution. The GLM detects optical pulses emitted by lightning discharges night and day,

without being able to differentiate between flash type and polarity. The GLM is expected to detect ~90% of all flashes during the night and ~70% during the day with a daily average exceeding 70% (Goodman et al. 2013). The spatial resolution of the individual pixels varies between 8 and 12 km over the contiguous Unites States. Data processing provides a hierarchy of geo-located and time-stamped lightning data products (level 2 lightning data products), which include the events, groups, and flashes (Goodman et al. 2013; Rudlosky et al. 2019). At each pixel, the GLM tracks the average background energy. During each 2-ms sampling frame, an event is defined if a single pixel exceeds the background threshold. The GLM lightning cluster-filter algorithm (LCFA) then assembles one or more simultaneous GLM events in adjacent pixels into groups, and finally combines one or more sequential groups separated by less than 330 ms and 16.5 km into flashes (Mach et al. 2007). Similar to Fierro et al. (2019), only the GLM flash product is used in this study. The level 2 GLM data products also contain a built-in adjustment for parallax due to the viewing angle of cloud tops from the satellite. That correction is based on satellite position, latitude and longitude of each pixel, and an assumed cloud-top height, which is dependent on latitude. The parallax errors over the Great Plains of the United States is estimated to be on the order of 10 km or less (Fierro et al. 2018a) and, thus, is expected to have negligible impacts on shorter-term, storm-scale forecasts.

3. Methodology/model and data assimilation configuration

a. Forecast model

This work employs version 3.7.1 of the nonhydrostatic Weather Research and Forecasting (WRF) Model with the Advanced Research WRF dynamic core (WRF-ARW; Skamarock et al. 2008). The selected physical parameterizations are the NSSL two-moment 4-ice category bulk microphysics scheme (Ziegler 1985; Mansell et al. 2010; Mansell and Ziegler 2013), the Dudhia shortwave (SW) radiation scheme (Dudhia 1989), the Rapid Radiative Transfer Model (RRTM) longwave (LW) radiation scheme (Mlawer et al. 1997), the Yonsei University (YSU) planetary boundary layer (PBL) scheme (Hong et al. 2006), and the Rapid Update Cycle (RUC) land surface scheme (Benjamin et al. 2004). The DA experiments are performed on one single domain with a uniform horizontal grid spacing of 1.5 km and horizontal dimensions of 500 × 500 grid points (i.e., 750 km × 750 km). There are 51 vertical levels with a top set at 50-hPa corresponding to an

altitude of about 20–22 km. The computational time step for the integration is 6 s.

b. Data assimilation procedures

1) 3DVAR SYSTEM

This study employs the NSSL 3DVAR system (Gao et al. 2013; Gao and Stensrud 2014), which is an upgrade from the Advanced Regional Prediction System (ARPS) 3DVAR code (Gao et al. 2004; Hu et al. 2006a,b). This earlier version was only able to assimilate surface and radar radial velocity observations and update the following six state variables: the pressure p, potential temperature θ, water vapor mixing ratio q_v, horizontal wind components u and v, and vertical wind component w. The adjustments of water vapor and hydrometer variables were originally performed by means of a complex cloud analysis scheme (Hu et al. 2006a), appearing to be too aggressive due to the introduction of relatively large amount of ice and water mass into the model domain (Schenkman et al. 2011; Fierro et al. 2014; Gao et al. 2016). To mitigate this drawback, NSSL 3DVAR was upgraded to include the capability of assimilating radar reflectivity factor (Gao and Stensrud 2012) to adjust the mass mixing ratios for rainwater (q_r), snow (q_s), and hail (q_h) in the same variational framework in which radar radial velocity data and other traditional datasets are assimilated. The reflectivity forward operator uses classification of hydrometer types so that updates are physically consistent by preventing, for example, snow mass from being introduced at temperatures below the freezing level. Another feature of NSSL 3DVAR not used in this study is the incorporation of ensemble-estimated covariance in a variational approach (Gao and Stensrud 2014; Gao et al. 2016). In addition, NSSL 3DVAR includes a direct interface for WRF-ARW (Zhuang et al. 2016; Y. Wang et al. 2018).

The data assimilated in this study include WSR-88D level II radar data and pseudo water vapor mixing ratio observations derived from GLM-observed total lightning data, both of which are described in more detail in the subsections below.

2) RADAR

This work utilizes WSR-88D level II data (i.e., reflectivity factor and radial velocity) from multiple radars covering the forecast domain. Prior to the DA, necessary quality control (QC) procedures including radial velocity dealiasing and removal of weak radar returns or nonmeteorological scatters are performed on each radar. For thinning purpose, QC-ed radar data from multiple sites are then interpolated onto the

WRF model grid. If there are multiple radars over-lapping at a given grid point, the largest radar reflec-tivity value is chosen. The QC and processing of radar data follow the same procedures as in Gao et al. (2013) and Fierro et al. (2016, 2019). Additionally, the radar-derived velocity–azimuth display (VAD) wind obser-vations are assimilated into NSSL 3DVAR to adjust the wind components of the storm environment.

3) GLM TOTAL LIGHTNING

To provide guidance for utilizing GLM lightning observations in real-time DA applications, lightning observations first are accumulated up to the analysis time within a fixed time window. That is, for a given window size Δt and analysis time t, the GLM total flash data (from 20-s operational packets) accumulated between $t - \Delta t$ and t are used in the assimilation. It is worth underlining that this approach differs from previous works where, instead, observations falling within a time window centered at the analysis time t (i.e., observations from $t - \Delta t/2$ to $t + \Delta t/2$) are as-similated. In other words, this implementation of the GLM DA does not consider future observations (relative to the analysis time, i.e., smoothing); which topic will be deferred to future research. One chief task of this study is to determine an optimal Δt for the GLM data, which were varied from 10 to 60 min in the present study. The lightning data assimilation (LDA) procedure follows, overall, a similar philosophy as Fierro et al. (2019) wherein the model background q_v values are adjusted around the locations where lightning occurs, regardless of flash rate. In other words, the creation of pseudo-q_v observations is per-formed on a binary basis and is, thus, only contingent on whether or not a flash is observed at a given grid point.

Figure 1 illustrates a difference of strategy adopted in this work with respect to Fierro et al. (2019) to inter-polate the observed GLM data onto the model grid. Each GLM flash has a latitude–longitude coordinate denoting its pixel centroid with an average 8–12 km pixel footprint resolution (Goodman et al. 2013). To ac-count for this coarser resolution relative to the typical convection-allowing grid spacing used ($dx \leq 3$ km), Fierro et al. (2019) assumed that, given a GLM flash centroid, all the grid cells contained within an assumed, fixed (e.g., $\sim 9 \times 9$ km^2) area around the centroid on the model grid (blue region in Fig. 1) were assigned the same lightning rate values as the centroid observations. In this study, however, only the (1.5×1.5 km^2) grid cell con-taining the flash centroid (referred to as lightning col-umn) is assigned the actual observed flash rate (i.e., the cell at the center of the grid in Fig. 1). The reduction of
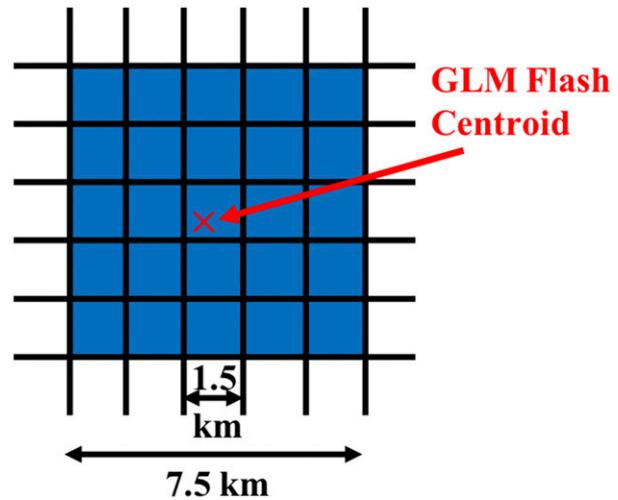

FIG. 1. Schematic illustration of the spreading strategy for GLM total lightning data used in Fierro et al. (2019) vs the one used in this study. In this example, the horizontal dimension of individual grid cell is of 1.5 km. A red "×" symbol indicates the centroid of GLM flash observation, which does not necessarily coincide with the geometrical center of the grid cell in the model.

the lightning footprint on the model grid is motivated by the systematic trade-off seen between shorter-term forecast improvements and increases in wet biases in the forecast when precipitation- or moisture-related variables are adjusted (increased) during the assimilation (Fierro et al. 2015, 2016). The decorrelation length from the 3DVAR is then used to determine a distance-weighted factor such that grid points located farther away from the centroid bear lower weight (i.e., smaller effect) reminis-cent of the Gaspari and Cohn (1999) correlation function with compact support used in EnKF methods. At each lightning column, a pseudo q_v field is created wherein the background q_v within a fixed layer (here 3 km) above the lifted condensation level (LCL) is adjusted to near satu-ration (set RH to 95%, unless the background RH al-ready exceeds 95%). Different layer depths above the LCL ranging from 2 to 10 km were evaluated and yielded overall similar results, consistent with Fierro et al. (2016). The GLM pseudo-q_v observations are produced using the saturation water vapor mixing ratio $q_{v\mathrm{sat}}$ and the adjusted background RH of 95% through $q_v = q_{v\mathrm{sat}} \times$ RH and then assimilated into the 3DVAR system. Different sat-uration RH values ranging from 80% to 95% were tested and yielded overall qualitatively very similar results.

Similar to Fierro et al. (2016, 2019), the background and observation error for q_v is set to 1×10^{-2} and 3×10^{-3} kg kg^{-1}, respectively, with a smaller observation error aiming to impose more weight on observations during the 3DVAR minimization. These values were obtained by trial and error in a manner that would ensure comparable

Simulation domains (blue square) and locations of radar sites
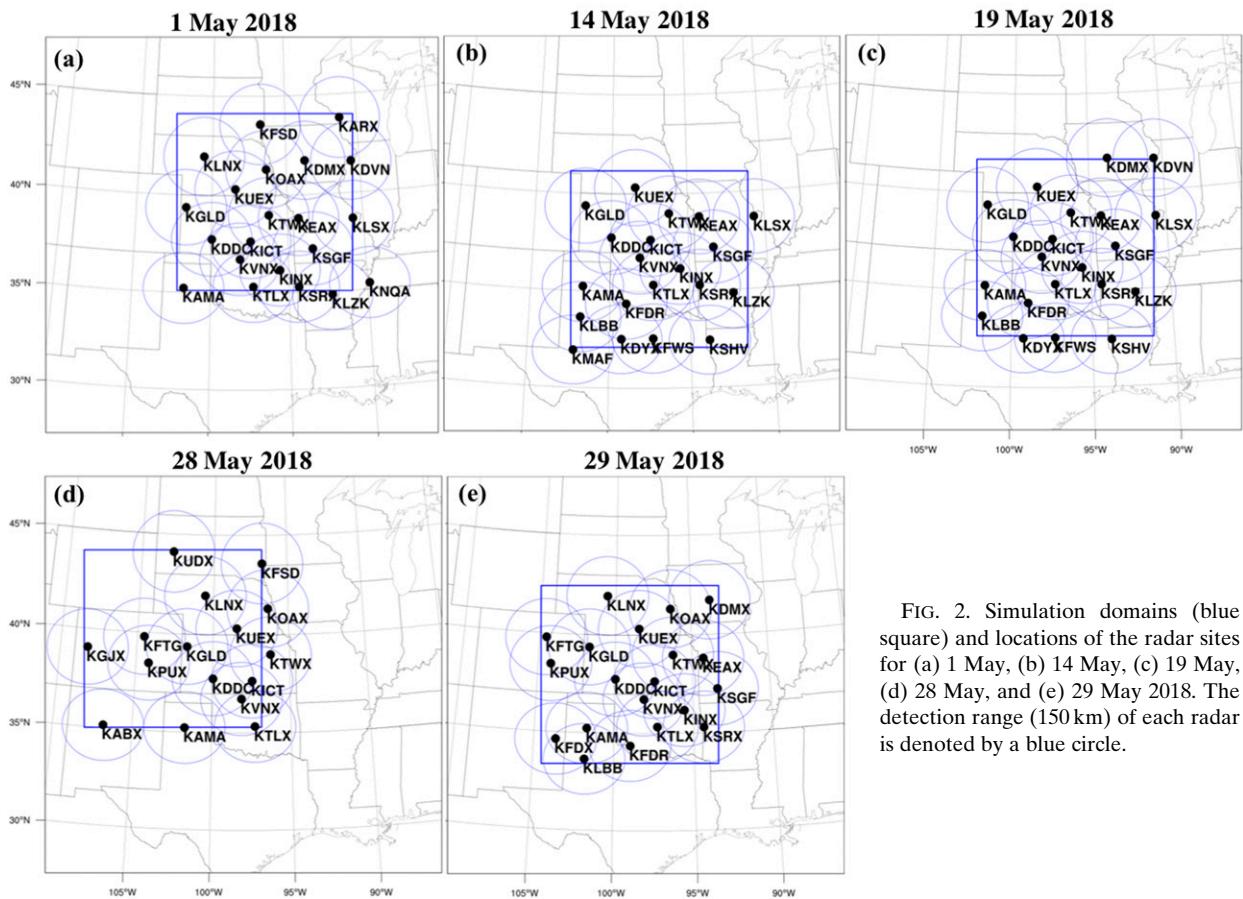

FIG. 2. Simulation domains (blue square) and locations of the radar sites for (a) 1 May, (b) 14 May, (c) 19 May, (d) 28 May, and (e) 29 May 2018. The detection range (150 km) of each radar is denoted by a blue circle.

weights assigned to lightning observations with respect to other storm-scale dataset (in this case radar data) that are used for the real-time Spring Forecast Experiment (SFE).

As indicated by Xie et al. (2011) and Li et al. (2010), a multipass recursive filter (from larger to finer length scales) yields generally superior forecast performance over single-pass 3DVAR methods. This study adopts a two-pass recursive filter, each with a prescribed, fixed horizontal decorrelation length scale (labeled L1 and L2, respectively) and vertical decorrelation length scale. GLM lightning data are only used in the second pass. As in Fierro et al. (2019), L1 = 24 km was chosen for the first pass. Different values of L2 for the second pass were tested for the GLM-based LDA as will be discussed in sections 4 and 5. The vertical decorrelation lengths adopted in this study are 4 grid points in the first pass and 2 grid points in the second pass.

4. Experimental design

To perform a more systematic evaluation of the impact of assimilating GLM lightning-derived water vapor on short-term forecasts, five high-impact weather events from the 2018 SFE hosted by the NOAA Hazardous Weather Testbed (HWT) are examined, namely, 1, 14, 19, 28, and 29 May. The NSSL Experimental Warn-on-Forecast System (WoFS) developed by the Warn-on-Forecast (WoF) project has been run in real time for SFE since 2016. The study domain is selected on a daily basis based on the ''day 1'' convective outlook product from the Storm Prediction Center (SPC). The same simulation domains as those from WoFS real-time SFE runs are utilized in this study and presented in Fig. 2 with the locations of radar sites used overlaid for all five cases.

The flowchart of the cycled DA and forecast system is illustrated in Fig. 3. For each case, the model is cold started at 1800 UTC and the 3DVAR is cycled at 15-min (or 60-min) intervals until 0300 UTC the following day. Starting from 1900 to 0300 UTC, a 3-h forecast is launched every hour. The initial and lateral boundary conditions are derived from the 3-km High-Resolution Rapid Refresh (HRRR) forecasts initialized at 1800 UTC and downscaled onto the 1.5-km simulation domain (Fig. 2). To better gauge the
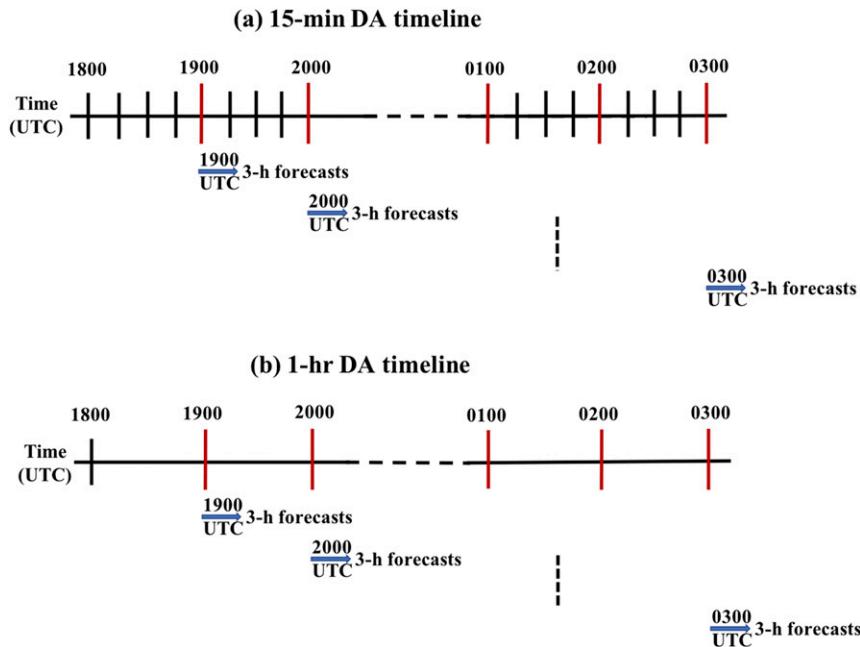
FIG. 3. Illustration of data assimilation and forecast cycle workflow for two scenarios: (a) observations being assimilated every 15 min and (b) 60 min. In both (a) and (b), a 3-h forecast is launched every hour from 1900 to 0300 UTC (viz., eight separate forecasts).

respective impacts of the radar and/or GLM DA, a control experiment (denoted as CTRL) is conducted first by advancing the model forward without assimilating any observations.

The observed composite reflectivity fields (CREF) from NSSL's Multi-Radar Multi-Sensor (MRMS) product (Smith et al. 2016) and the Stage IV hourly rainfall estimates from the National Centers for Environmental Prediction (Baldwin and Mitchell 1997) were used to evaluate the simulations. To facilitate the comparison between observations and model simulations, all the observations were interpolated onto the 1.5-km model grid (Fig. 2).

Contingency elements including the probability of detection (POD), the false alarm rate (FAR), and critical success index (CSI) along with the frequency bias and neighborhood-based approach such as the equitable threat scores (ETS; Clark et al. 2010) and the fractions skill scores (FSS; Roberts and Lean 2008) were used to quantitatively evaluate the DA impact on the analysis and short-term forecasts for each experiment individually, and in the aggregate.

To determine more optimal settings for assimilating GLM lightning-derived water vapor mixing ratio in convective-scale NWP models, this study first performed several sensitivity tests related to the DA configuration. Focus was directed on the horizontal decorrelation length scale L2 [ranging from 3 to 12 km

to limit the overspreading of q_v described in Fierro et al. (2016, 2018b)], the frequency of 3DVAR cycling (15–60 min) and the length of the time window Δt for accumulating GLM lightning data up to the analysis time t (10–60 min). Additional sensitivity tests using GLM accumulation periods centered at time t were conducted and showed overall similar analysis/forecast performance. The main objective is to explore the potential usage of GLM-derived water mixing ratio in a 3DVAR data assimilation system for real-time forecasts of high-impact weather events. Therefore, at each analysis time, it is reasonable to assimilate observations that have been collected up to the current analysis time to minimize latency. The design of the sensitivity test experiments and associated nomenclature are listed in Table 1. For example, the experiment denoted as "GLM_15min_10win_12km" assimilates water vapor mixing ratio derived from 10-min accumulated GLM total lightning data every 15 min using L2 = 12 km. The flow diagrams for the 15- and 60-min 3DVAR cycling experiments are shown in Figs. 3a and 3b, respectively. In addition to the three aforementioned DA parameters, these sensitivity tests evaluated how the GLM DA strategy herein (Fig. 1) performed against the original method from Fierro et al. (2019).

Based on aggregate forecast evaluation metrics for all five cases, the GLM experiment in Table 1 yielding the best results will be selected for evaluation against

TABLE 1. Key parameters used for the GLM-based DA sensitivity experiments. From left to right, the column lists the nomenclature for each experiment, the 3DVAR cycling frequency, the accumulation interval used for the GLM data prior to the DA and the horizontal decorrelation length scale used in the second 3DVAR pass "L2" (recalling that the GLM DA is not activated in the first pass). The suffix "Spread" refers to the original LDA method from Fierro et al. (2019) illustrated in Fig. 1.

Experiment	DA cycle (min)	Time window Δt (min)	Horizontal decorrelation length L2 (km)
GLM_15min_10win_12km	15	10	12
GLM_15min_10win_6km	15	10	6
GLM_15min_10win_3km	15	10	3
GLM_15min_10win_3km_Spread	15	10	3
GLM_60min_60win_3km	60	60	3
GLM_60min_30win_3km	60	30	3
GLM_60min_15win_3km	60	15	3

(i) CTRL, (ii) a DA experiment only assimilating WSR-88D radar data (RAD), and (iii) a DA experiment assimilating GLM total lightning data in conjunction with WSR-88D radar data (referred to as GLM + RAD).

5. Results and discussion

a. Sensitivity tests for the GLM-based DA experiments

Prior to describing the impact of the GLM DA on the analysis and short-term forecast, it is useful to first provide a brief quantitative depiction of the q_v adjustments made by the 3DVAR. For the sake of brevity, the 1 May 2018 case is used for illustration.

On 1 May 2018, environmental conditions favored the development of severe thunderstorms across Kansas and southeastern portions of Nebraska. Between 1930 and 2000 UTC, the first convective cells initiated along a front across western Kansas and southeastern Nebraska. In the following hours, more storms formed along the front to eventually merge into a quasi-linear mesoscale convective system (QLCS) as they moved northeastward. Based on National Weather Service (NWS) local storm reports, this QLCS produced several large hail (up to baseball size) and tornado events. Between 2300 and 0200 UTC 2 May alone, a total of 13 tornadoes and several large hail events were reported in Kansas, Nebraska, and Oklahoma.

The flash origin density rate data (i.e., the number of flashes in each grid cell) derived from the accumulated GLM flash centroid observations at 2300 UTC in each experiment (Table 1) for the 1 May 2018 case are given in Fig. 4. Comparing Figs. 4a and 4b, it becomes clearer that the "spreading" strategy used in Fierro et al. (2019) leads to a larger areal coverage of nonzero flash density rates, indicating that the q_v adjustments from the 3DVAR will be performed on a larger number of grid columns (or grid volume). Because the probability of a given GLM flash centroid to fall within an individual $9 \times 9 \, km^2$ box will be larger than within a finer

$1.5 \times 1.5 \, km^2$ box, the spreading method of Fierro et al.(2019) will also lead to higher flash density rates on the model grid (Figs. 4a,b). Although the current study imposes q_v adjustments regardless of lightning rates, Fig. 4 illustrates that consideration must be taken if the q_v adjustments are assumed to depend on flash density as in Fierro et al. (2016). Figures 4c–e highlight and contrast for this case how the number of nonzero lightning cells decreases as the accumulation interval for the lightning decreases (smaller rates).

To assess how the GLM DA impacts the q_v fields for each of the sensitivity experiments listed in Table 1, the 3–7 km (MSL) layer-averaged q_v fields of the analysis are examined for 1 May 2018. Note that due to the cycling, each experiment will evolve a different background over time. Therefore, in this study, the q_v differences relative to CTRL instead of the q_v increments (analysis background) are analyzed at 2300 UTC (Fig. 5). It is worth noting that because 2300 UTC is not the first 3DVAR cycle (Fig. 3a), the q_v differences can be negative owing to successive model adjustments to the background fields from the prior 3DVAR cycles. Rationale for selecting 2300 UTC was because this time produced a copious amount of lightning, better highlighting the differences and sensitivities between the experiments.

As expected, the largest positive q_v differences are collocated with areas of nonzero flash density (Fig. 4). As L2 decreases, the q_v differences are imposed on progressively smaller areas (Figs. 5a–c). For a given value of L2 (here at 3 km), Figs. 5b and 5d highlight how the spreading strategy of Fierro et al. (2019) yields approximately comparable q_v differences as those produced by the nonspreading strategy herein using a larger decorrelation length (L2 = 6 km, Fig. 5b). Comparison of Fig. 5e against Fig. 5c reveals that 15-min DA (five cycles) produces noticeably larger q_v differences than 60-min DA (1 cycle). This is because, over a 1-h period, DA experiment GLM_15min_10win_3km accumulates the impact (q_v adjustments) of five individual 3DVAR cycles, compared to only one 3DVAR cycle
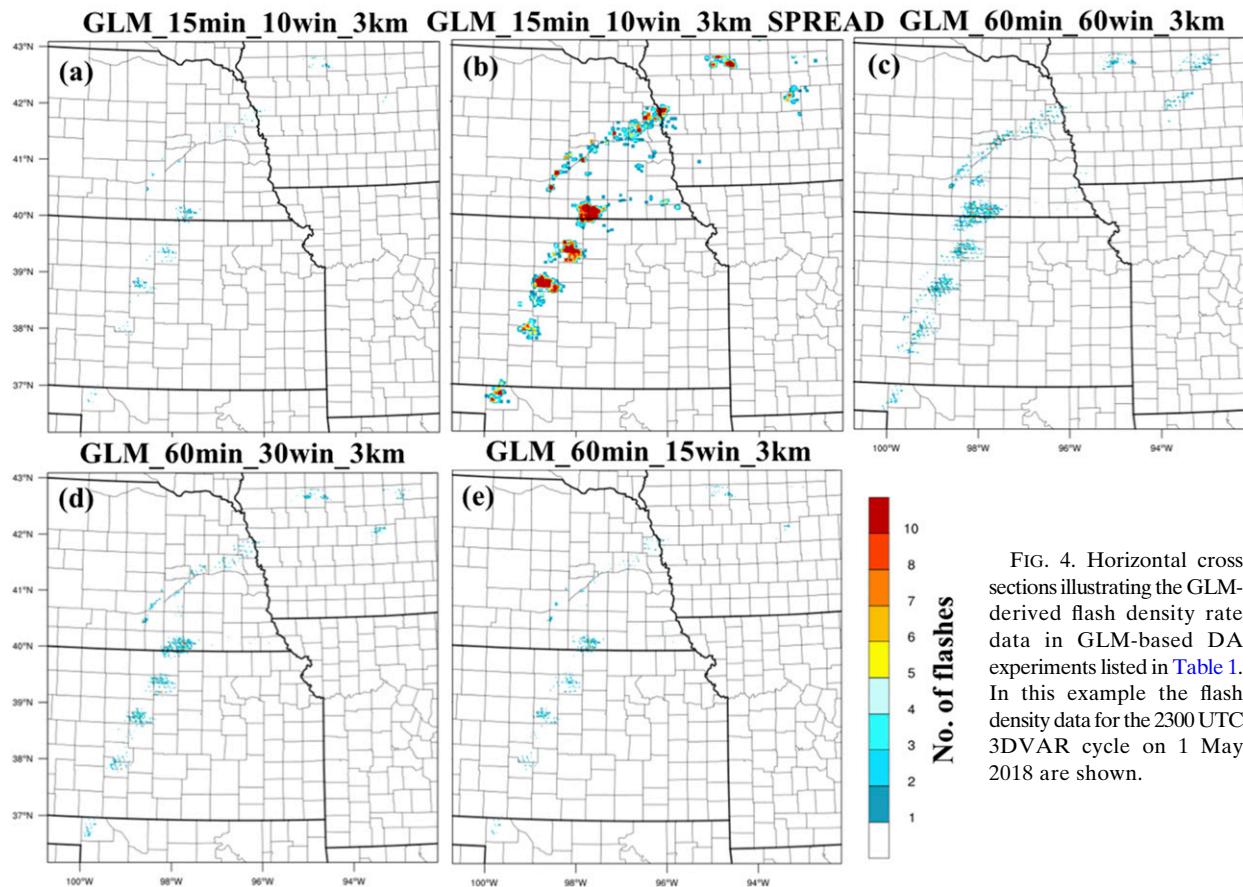
GLM lightning observations at 2300 UTC, 1 May 2018

FIG. 4. Horizontal cross sections illustrating the GLM-derived flash density rate data in GLM-based DA experiments listed in Table 1. In this example the flash density data for the 2300 UTC 3DVAR cycle on 1 May 2018 are shown.

in GLM_60min_60win_3km. The 60-min DA experiments show that a slightly larger areal coverage of positive q_v differences will ensue if a longer accumulation interval is used (Figs. 5e–g).

The analyzed CREF fields at 2300 UTC are shown in Fig. 6 for each of the GLM-based experiments in Table 1 together with CTRL and the MRMS observations. To better differentiate and analyze the comparisons, the QLCS is divided into four main regions, labeled A to D from south to north (Fig. 6a). CTRL fails to simulate the tornadic storm in region A (Figs. 6a,b) while all 15-min cycling DA experiments successfully captured it (Figs. 6c–f). Highlighting the benefit of using more frequent 3DVAR cycles, Figs. 6g–i reveal that all 60-min cycling DA experiments fail to analyze the storm despite the positive q_v adjustments incurred there (Figs. 5e–g). The observed storms in region B produced several tornadoes and large hail events. CTRL misses some of the cells in this region and the intensity is overall weaker than observed (Fig. 6b). Although, both 15- and 60-min DA experiments

improved the forecast of these isolated cells, the forecasted CREF values are notably larger than observed. The advantage of using higher-frequency 15-min cycling over 60 min is illustrated in Figs. 6c–i: none of the 60-min cycling experiments forecasts the individual storms well enough (Figs. 6g–i), while, in contrast, the 15-min DA experiments are able to produce storm cells that are more isolated in nature and, in turn, more concordant with the observations despite a slight northward bias (Fig. 6a). In region C, all simulations appear to reproduce the largely stratiform, weaker convection there, with the GLM-based DA experiments generally overestimating reflectivity values. In region D, all 15-min DA experiments exhibit reasonable storm coverage with generally larger-than-observed reflectivity values (Figs. 6c–f). In contrast, the 60-min DA experiments produce slightly smaller storm coverage but with reflectivity values generally more aligned with the observations (Figs. 6g–i). Generally, all GLM-based DA experiments show noteworthy improvements over CTRL with the 15-min DA experiments producing the best

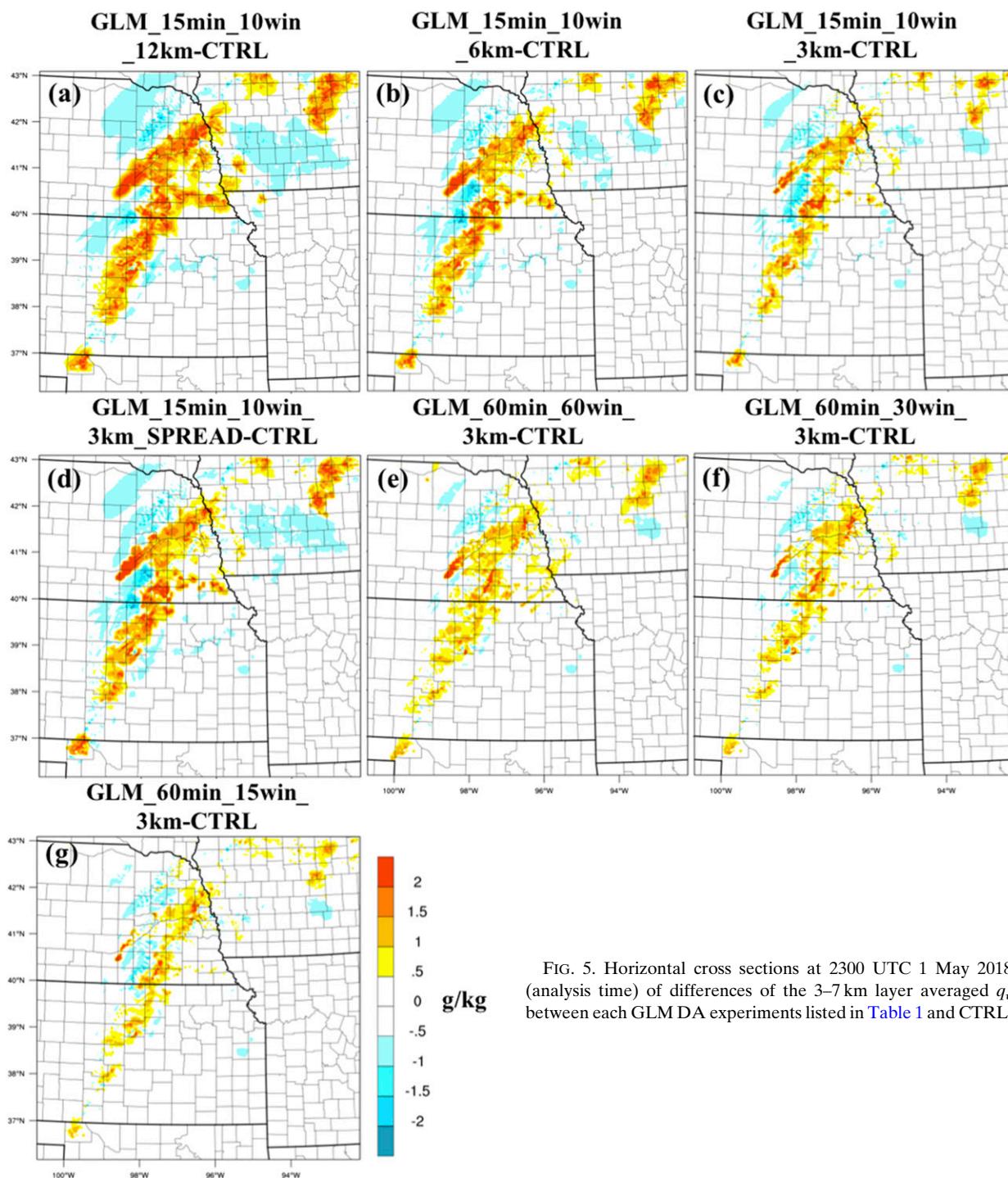## 3–7 km layer averaged q_v from analysis at 2300 UTC, 1 May 2018

FIG. 5. Horizontal cross sections at 2300 UTC 1 May 2018 (analysis time) of differences of the 3–7 km layer averaged q_v between each GLM DA experiments listed in Table 1 and CTRL.

results, in terms of intensity, structure, location, and areal coverage. In line with the q_v differences highlighted in Fig. 5, the usage of a larger decorrelation length scale yields larger values and areal coverage of simulated

reflectivity (Figs. 6c–e). Although the difference is arguably small, storm coverage generally decreases as the lightning accumulation interval decreases (Figs. 6g–i). Consistent with earlier findings, the spreading strategy
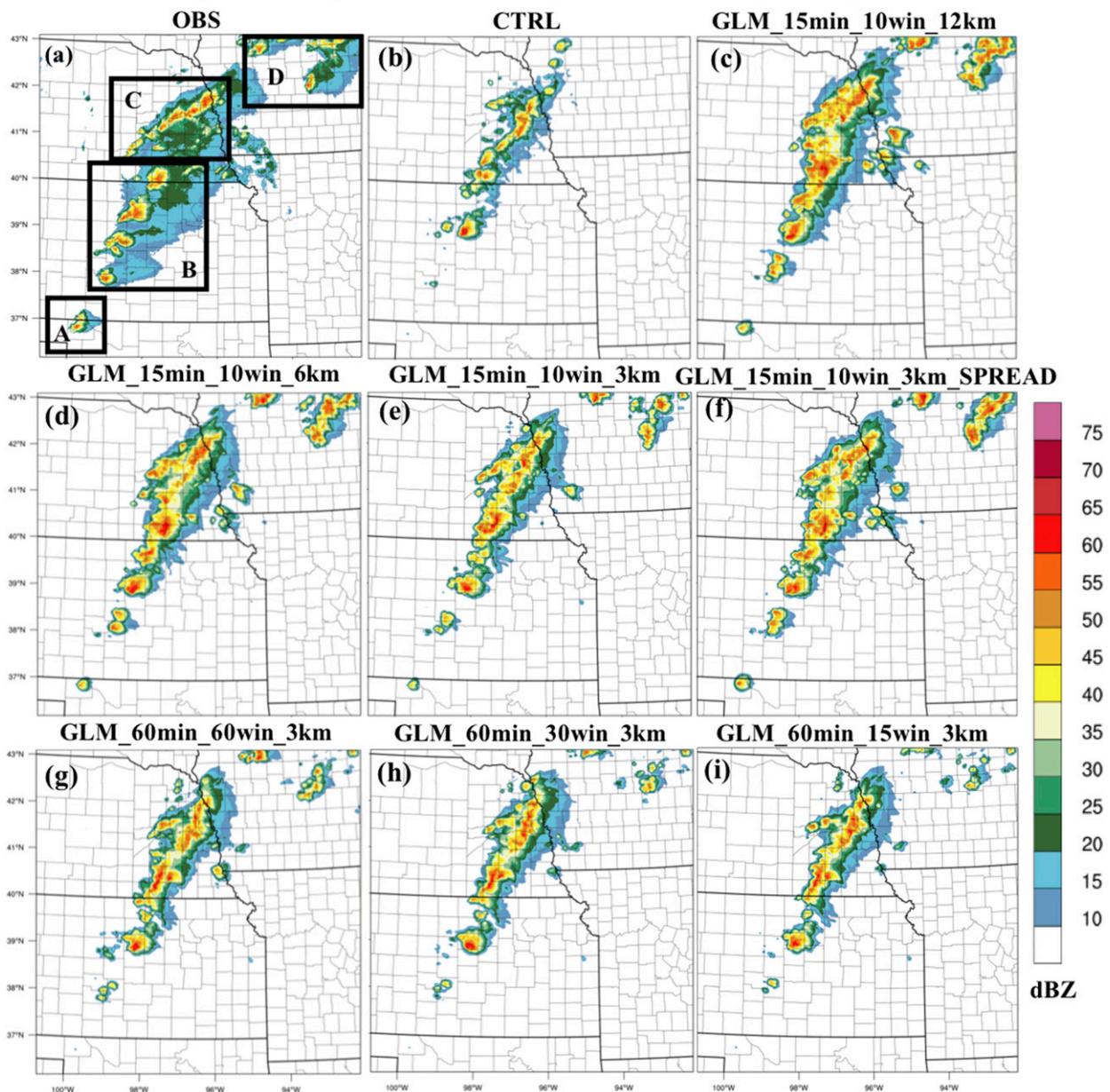
FIG. 6. Horizontal cross sections of composite reflectivity fields at 2300 UTC 1 May 2018 (analysis time) for (a) MRMS observations interpolated onto the 1.5 km simulation domain, (b) CTRL, and (c)–(i) the GLM DA experiments in Table 1.

from Fierro et al. (2019) is effectively equivalent to increasing the decorrelation length scale (Figs. 6d,f).

To analyze the overall performance of each GLM DA experiment, aggregate forecast metrics for composite reflectivity fields and accumulated rainfall were computed for 19 separate 3-h forecasts across all five cases (viz. 2100–0000 UTC 1 May 2018, 1900–2100 UTC 19 May 2018, and 0000–0300 UTC 14, 28, and 29 May 2018). These metrics are: POD, CSI, FAR, success ratio

(SR = 1 − FAR), the frequency bias (referred to as bias) and the neighborhood-based ETS. The FSS values were also computed showing, overall, very similar results to the ETS (not shown). The neighborhood radius shown for the ETS is 12 km (8 grid points). To provide a more concise and clearer depiction of forecast performance, this work makes use of the categorical performance diagram described in Roebber (2009), which conveniently merge bias, POD, FAR (through SR), and CSI

Performance of 0-3h forecast composite reflectivity relative to MRMS observations
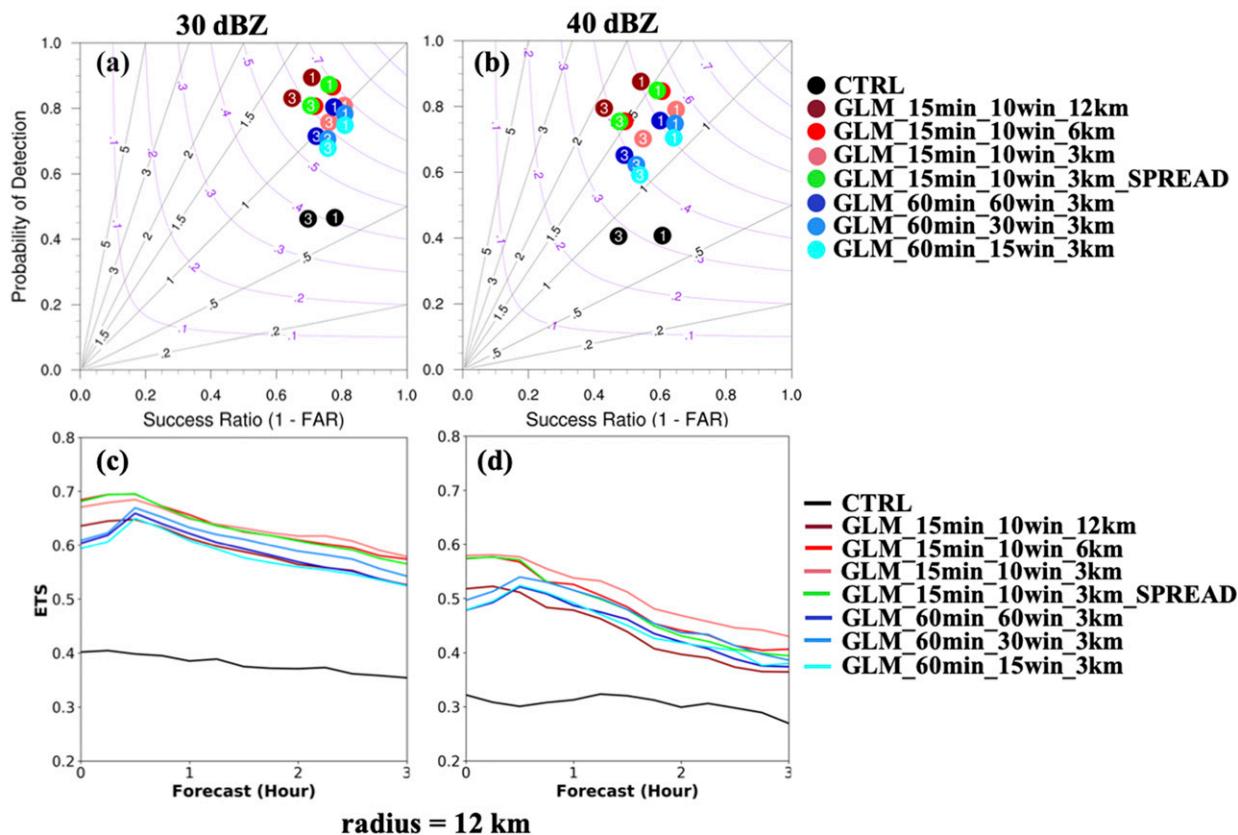

FIG. 7. Aggregate score metrics of composite reflectivity fields relative to MRMS observations for 19 three-hour forecasts over all five cases for each of the experiments listed in Table 1. (a),(b) Performance diagram and (c),(d) equitable threat score (ETS) for (a),(c) 30 dB*Z* and (b),(d) 40 dB*Z* thresholds, respectively. Results are shown for a neighborhood radius of 12 km. The numbers within the colored dots denote the forecast hour (i.e., 1- and 3-h forecasts).

information into one graph. Forecast skill increases as the POD, SR, bias, and CSI all approach unity, meaning that a perfect forecast lies on the upper-right corner of the diagram.

The improvements from the GLM-based DA experiments relative to CTRL become more evident when visualized in this diagram (Figs. 7a,b). Consistent with Fierro et al. (2016), the 15-min DA experiments indicate that when the decorrelation length scale L2 increases, so does the bias along with progressively larger POD and smaller SR and CSI. The 60-min DA experiments produce smaller CSI values than their 15-min counterparts. The nearly identical placement of the dots on the diagram for both GLM_15min_10win_3km_SPREAD and GLM_15min_10win_6km (Table 1), confirms that the spreading strategy used in Fierro et al. (2019) is effectively equivalent to increasing the decorrelation length scale in the nonspreading strategy. When increasing the CREF threshold from 30 dBZ (Fig. 7a) to 40 dBZ

(Fig. 7b), the POD, SR, and CSI values decrease in all experiments. Overall, the experiment labeled GLM_15min_10win_3km produces the best forecast as evidenced by higher POD and CSI values together with a lower bias. Generally speaking, when inspecting the respective aggregate ETS aggregated over 19 forecasts, it also becomes clear that all GLM-based DA experiments outperform CTRL, and that the 15-min cycling experiments produce overall larger scores than with 60-min cycling (Figs. 7c,d). Metrics for hourly precipitation using thresholds of 2.5, 5, and 10 mm relative to the Stage IV estimates have similar forecast behavior and performance, though the relative differences in scores between the experiments are more pronounced than for CREF (Figs. 8a–f). For precipitation, it is also worth noting that between 1 and 3 h forecast, (i) the biases exhibit a larger relative increase than for CREF, and (ii) the SRs at 3 h forecasts are generally smaller than for CREF (owing to larger FAR). This
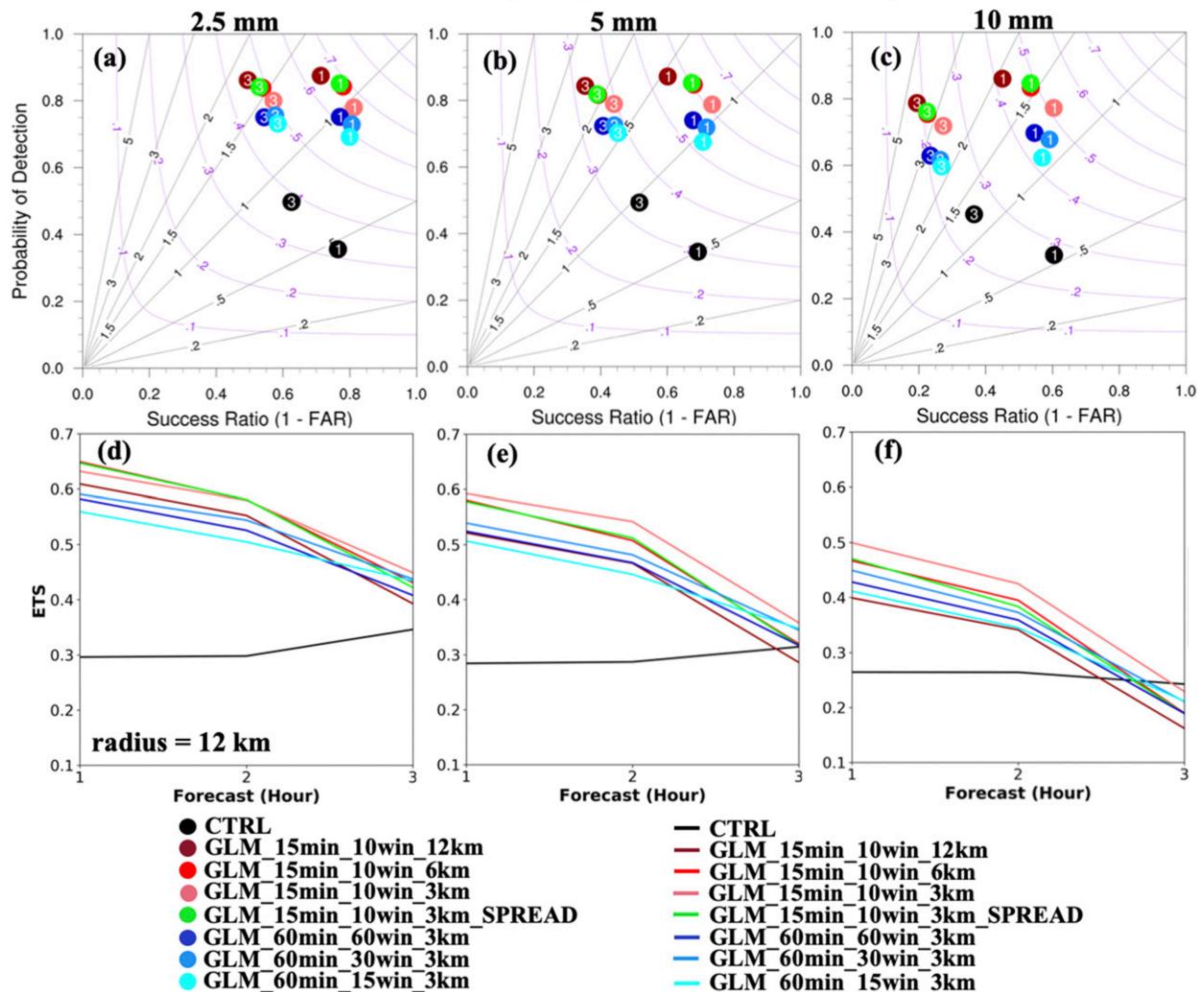
FIG. 8. As in Fig. 7, but for hourly accumulated precipitation relative to Stage IV multisensor rainfall estimates calculated using the following thresholds: (a),(d) 2.5, (b),(e) 5, and (c),(f) 10 mm.

suggests a larger sensitivity of precipitation predictions to some of the factors contributing to forecast errors such as spurious convection, bias in the microphysics (Fierro et al. 2015) and/or phase errors.

The aggregate performance diagrams and ETS time series for rainfall and CREF all indicate that the GLM_15min_10win_3km experiment—referred to as "GLM" in the remainder of the manuscript—generally produces the best forecast.

b. Overall evaluation

To better gauge the impact and added value of the assimilation of GLM data on short-term, cloud-scale forecasts, the GLM DA experiment (i.e., GLM_15min_10win_3km) will be evaluated against simulations assimilating

(level II) radar data (RAD) and assimilating both GLM and radar data (GLM + RAD). To maintain consistency with the original (most optimal) GLM experiment, both radar DA experiments (i.e., RAD and GLM + RAD) make use of 15-min cycling frequency as well (Fig. 3a). In the following sections, the results from two case studies representative of the best and worst forecast according to the performance diagram and ETS/FSS values will be first discussed followed by an analysis of the aggregate forecast statistics over all five cases.

1) BEST FORECAST SCENARIO: 1 MAY 2018 CASE

Although CTRL is able to produce some of the observed storms at 1-h forecast (Figs. 9a,b), all three DA

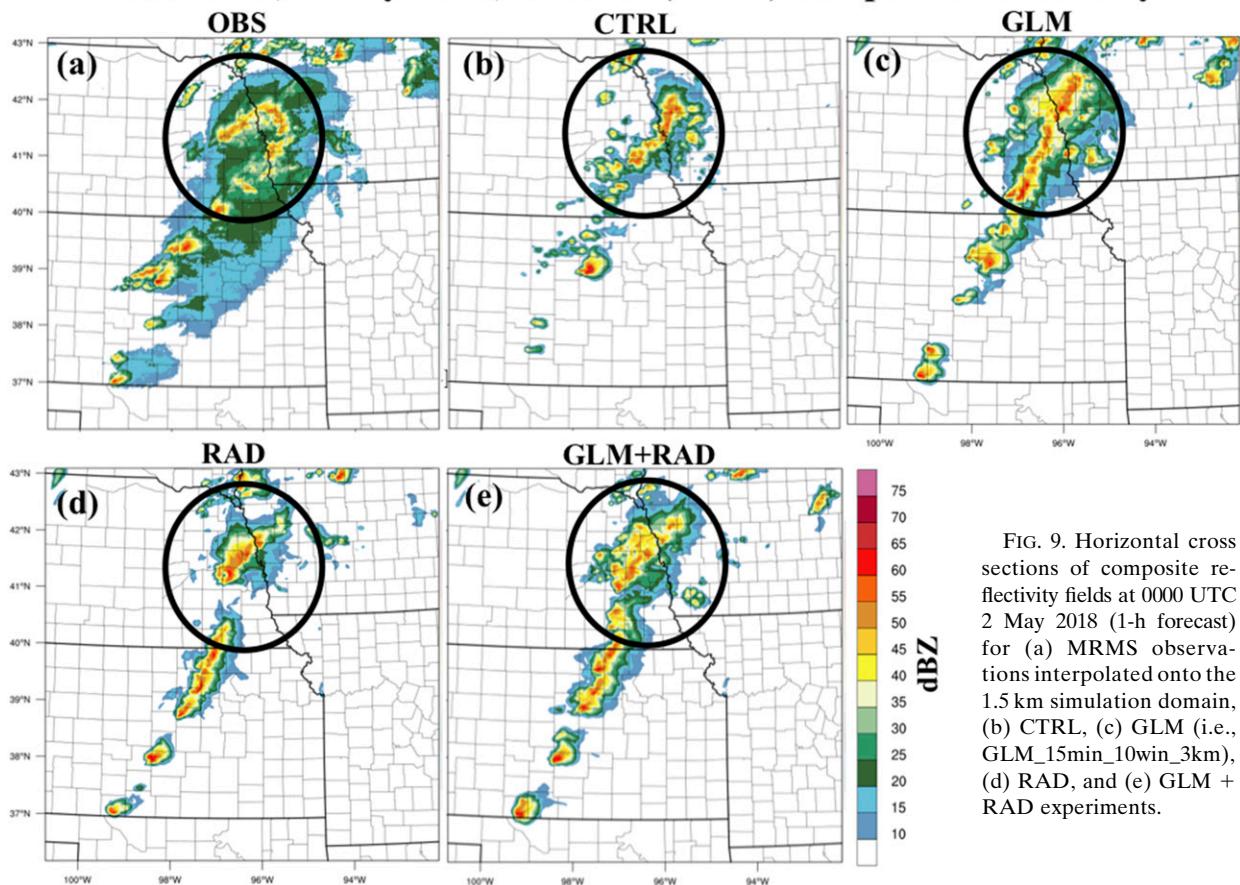## 2300 UTC, 1 May 2018, forecast (t = 1h) composite reflectivity

FIG. 9. Horizontal cross sections of composite reflectivity fields at 0000 UTC 2 May 2018 (1-h forecast) for (a) MRMS observations interpolated onto the 1.5 km simulation domain, (b) CTRL, (c) GLM (i.e., GLM_15min_10win_3km), (d) RAD, and (e) GLM + RAD experiments.

experiments GLM, RAD, and GLM + RAD improve the forecast further, as evidenced by the two storms just north of the Oklahoma border and one cell on the border between Kansas and Nebraska. Among the three DA experiments, only GLM is able to produce the two isolated supercellular storms just north of the Oklahoma border (Fig. 9c). Additionally, all DA experiments capture the QLCS structure better than CTRL. Both experiments assimilating radar data, that is, RAD and GLM + RAD, tend to produce a more linear convective mode than in the observations. It is relevant to note, however, that GLM also produces a more linear convective mode but in southeast Nebraska and western Iowa (Figs. 9a,c,d,e).

Generally speaking, all DA experiments improve the areal coverage of 3-h APCP < 40 mm further (Fig. 10). For amounts exceeding 50 mm, GLM produces a fairly reasonable areal coverage and amount prediction. However, all experiments appear to overpredict the areal coverage as well as the precipitation amount, which is most evident for the cluster of storms that crossed from Nebraska into Iowa (Figs. 10c–e).

To provide a more thorough assessment of the impact of the GLM DA, the 2–5 km updraft helicity (UH) tracks computed from 15-min model output are depicted in Fig. 11 and were overlaid with the NWS local storm reports. Although CTRL predicts the storm tracks associated with most tornadoes and hail events, it fails in predicting the two strong tracks in south-central Kansas and one additional track to the north of it (Fig. 11a). GLM predicts most of the tracks and their locations aligned generally well with the NWS local storm reports. The GLM storm tracks, however, exhibit a northward displacement bias for storms near the border of Nebraska and Kansas and a weaker intensity for one of the storms that produced two hail events in south-central Kansas (Fig. 11b). The tracks for all storms associated with NWS local storm reports are well forecasted by RAD, with the exception of weaker tracks in north-central Kansas associated with tornadoes and hail reports (Fig. 11c). When assimilating both GLM and radar data, the rotation tracks become more consistent with the NWS local storm reports (Fig. 11d). A similar analysis was conducted for
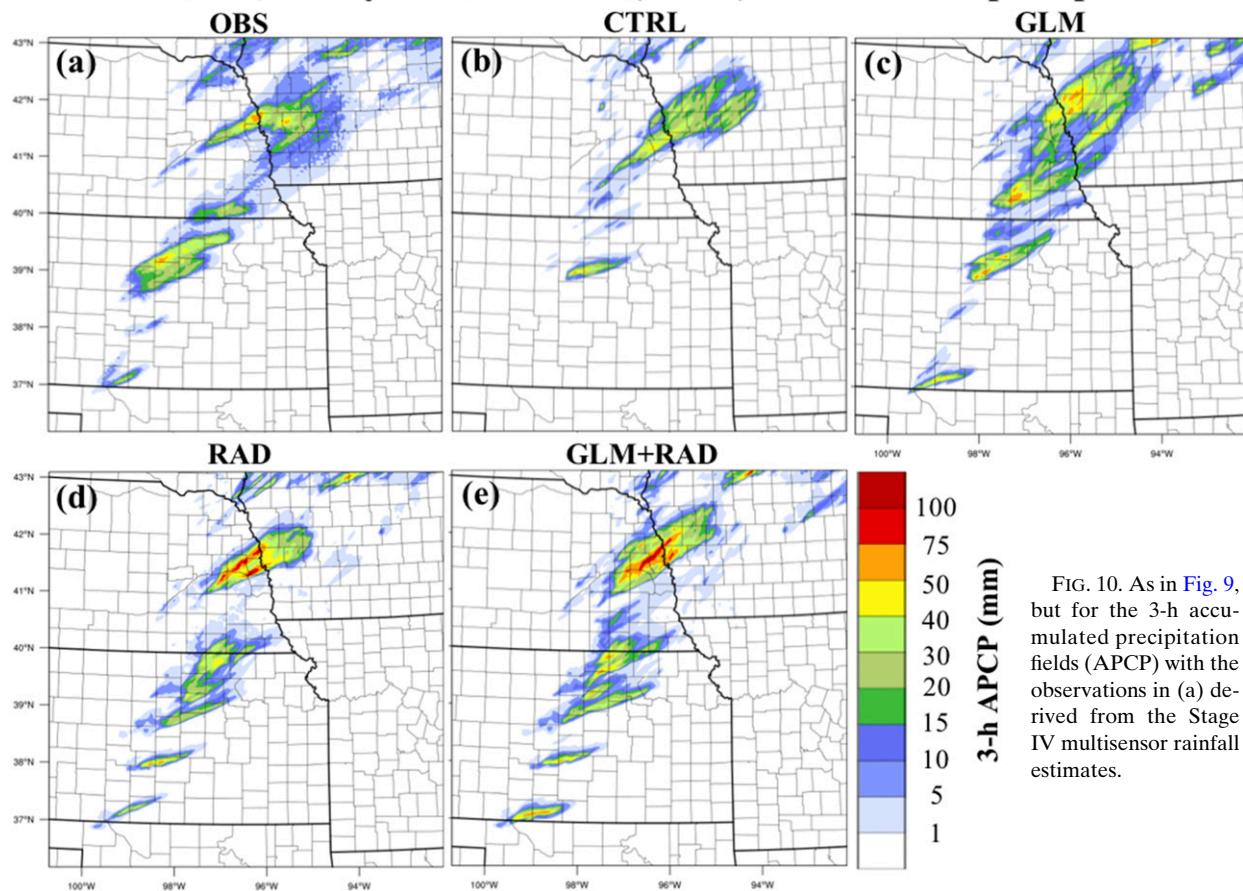
FIG. 10. As in Fig. 9, but for the 3-h accumulated precipitation fields (APCP) with the observations in (a) derived from the Stage IV multisensor rainfall estimates.

30+ dBZ CREF tracks and revealed overall very similar qualitative behaviors (not shown).

2) WORST FORECAST SCENARIO: 14 MAY 2018 CASE

On 14 May 2018, severe thunderstorms developed along and ahead of a quasi-stationary front moving through the central Great Plains. Four tornadoes were reported in Kansas—three of them in Cowley County and the fourth one in Chautauqua County (Fig. 12a). Hail up to baseball size (7.0 cm) was reported in western Kansas (near Scott city) and as large as ping pong balls (3.8 cm) in southeast Kansas. These severe thunderstorms also produced flooding rains over Chanute in Neosho County setting a 24-h rainfall record of 2.96 in., and Iola in Allen County of 2.34 in. Additionally, several thunderstorms initiated along a dryline in northwest Texas and the Oklahoma panhandle, and moved northeastward producing several large hail and strong wind reports there.

For the analysis of the CREF forecast, the study domain is, as before, divided into four key regions, labeled from A to D (Fig. 12a). For brevity of illustration, this

analysis focuses on 1-h forecasts initialized at 0200 UTC 15 May 2018. In A, all DA experiments reasonably predict most of the storms, with GLM generally producing the best results (Fig. 12c). RAD and GLM + RAD systematically displace the storms southward and with larger-than-observed CREF (Figs. 12d,e). All DA experiments reasonably reproduce most of the storms in B in terms of location and intensity (Figs. 12c–e). The CTRL run generates fewer storms and generally suffers from a notable southward bias (Fig. 12b), while GLM + RAD produce the largest overestimate for their areal coverage. For the storms in C, GLM generally overpredicts their intensity (Fig. 12c) while RAD does not distinctly improve over CTRL in this area (Fig. 12d). In region C, however, radar data are helpful in suppressing some of the spurious radar echoes over Missouri when combined with GLM lightning data (GLM + RAD, Fig. 12e). In region D, all the experiments underestimate CREF and fail to simulate the two main storm cells there.

When examining the 3-h precipitation forecasts, CTRL produces a distinctive southward bias in the APCP maximum observed in southeastern Kansas

2-5 km updraft helicity tracks from 0-3h forecasts initialized at 2300 UTC, 1 May 2018
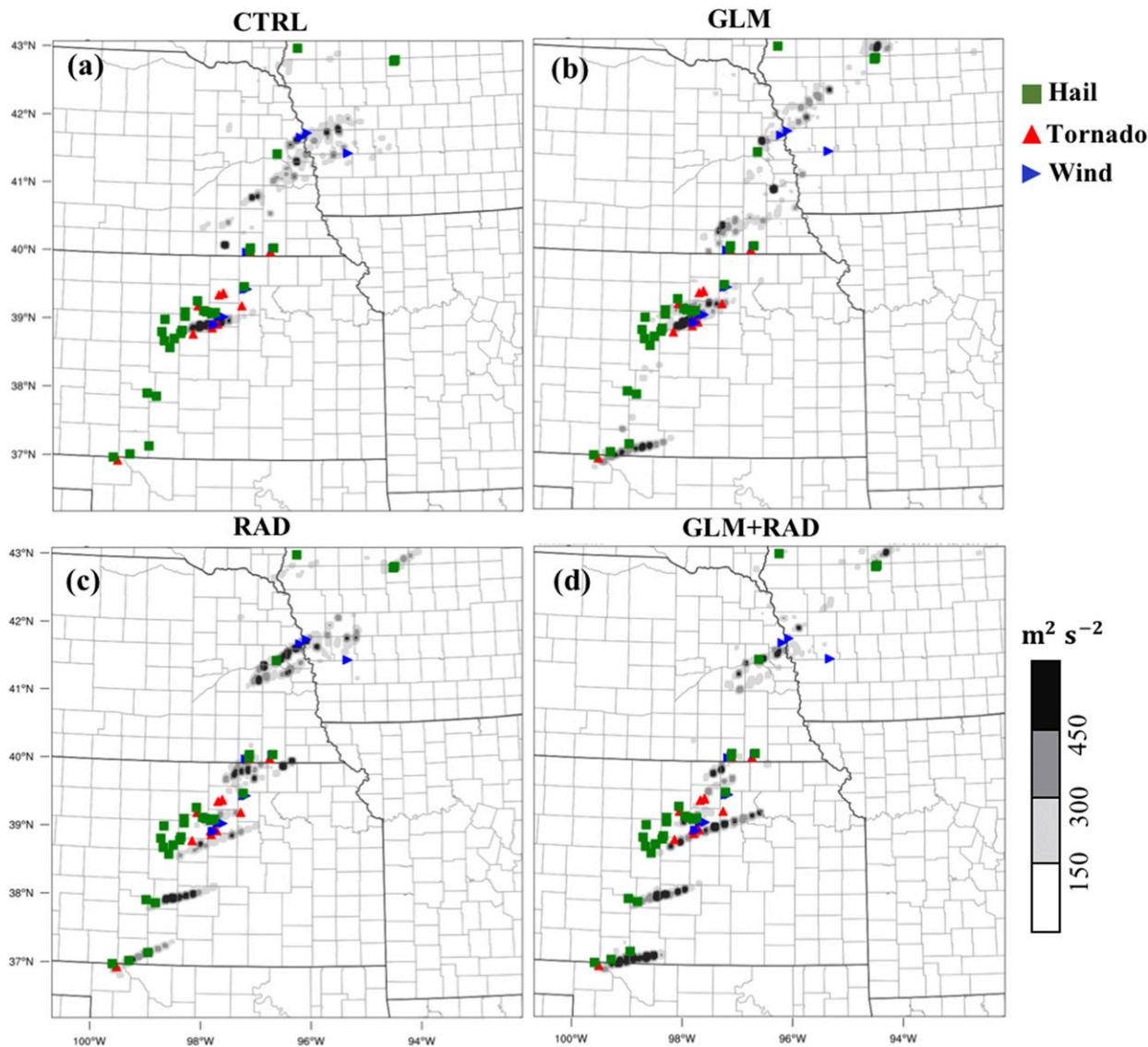

FIG. 11. 2–5 km updraft helicity tracks for 0–3 h forecast beginning at 2300 UTC 1 May 2018 for (a) CTRL, (b) GLM, (c) RAD, and (d) GLM + RAD experiments. The red triangles, green squares, and blue triangles represent the observed tornadoes, hail, and severe wind events, respectively.

(Figs. 13a,b). While GLM, RAD, and GLM + RAD alleviate this displacement error somewhat (Figs. 13c–e), GLM overestimates the rainfall amounts there (Figs. 13a,c), with GLM + RAD exacerbating rainfall overprediction further in eastern Kansas (Fig. 13e). Consistent with the CREF forecasts, RAD and GLM + RAD both overpredict precipitation amounts in southwestern Oklahoma (Figs. 13d,e).

The CREF tracks throughout the 3-h forecast and UH tracks (not shown) exhibit overall similar behavior

to the rainfall and CREF forecasts. In contrast to the southwest bias in CTRL, all DA experiments predict the CREF tracks in southern Kansas and northern Oklahoma reasonably well, demonstrating their superiority relative to CTRL. In this region, the CREF tracks from RAD are the most consistent with the NWS local storm reports, while both RAD and GLM + RAD generally overestimate their areal coverage there, agreeing well with CREF and rainfall forecasts in Figs. 12 and 13. As discussed before, CTRL as well as

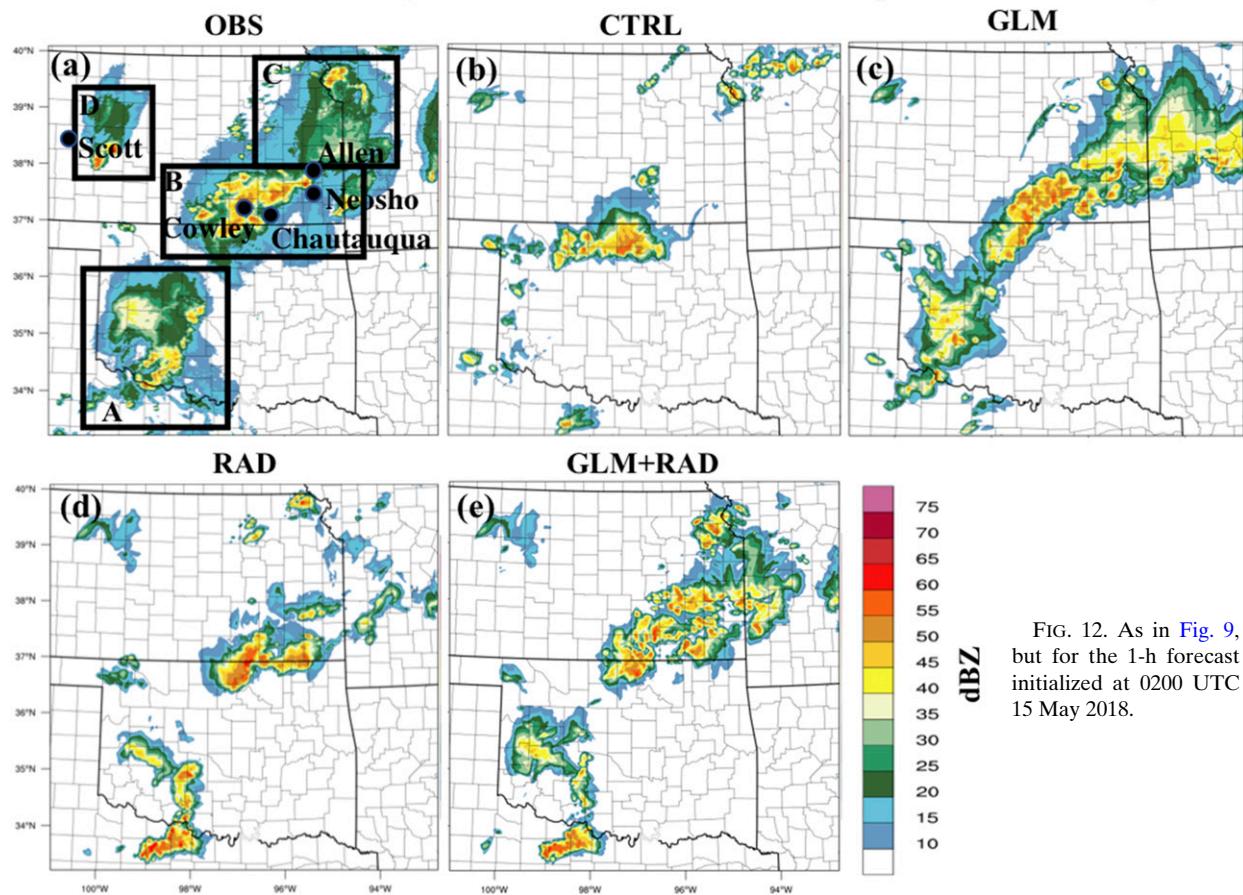## 0200 UTC, 15 May 2018 forecast (t = 1h) composite reflectivity

FIG. 12. As in Fig. 9, but for the 1-h forecast initialized at 0200 UTC 15 May 2018.

all DA experiments fail to predict the storms in western of Kansas and, consequently, the hail storm tracks over that region.

3) AGGREGATE PERFORMANCE

Similar to the GLM sensitivity experiments discussed earlier, aggregate forecast statistics for 19 separate 3-h forecasts from all five cases were computed to provide a more comprehensive view of the performance of GLM, RAD, and GLM + RAD. The aggregate metrics include POD, CSI, SR, bias, and ETS (radius of 12 km). These were calculated, as before, for CREF and 1-h accumulated precipitation with thresholds of 30 and 40 dBZ and 2.5, 5, and 10 mm, respectively. These statistics were also computed for the FSS and revealed overall similar behavior to the ETS (not shown).

In terms of CREF, all DA experiments outperform CTRL and produce biases ranging between 0.7 and 1.4 (Fig. 14). GLM produces the highest POD but largest bias, RAD has the smallest POD but smallest bias, while GLM + RAD lies in between (Figs. 14a,b). GLM + RAD exhibits slightly higher ETS than RAD for 0–1 h

CREF forecast at the cost of adding more bias due to the overprediction and spurious cells (Figs. 14c,d). For the 30-dBZ threshold, the improvement of ETS and CSI by GLM + RAD over RAD persists until the 3-h forecast. At 40 dBZ, however, RAD produces higher ETS and CSI for the 1.5–3 h forecast. GLM has lower ETS for the first 1.5 h of forecast but then surpasses GLM + RAD. The difference between RAD and GLM is more notable during the first half hour, where RAD produces much higher ETS than GLM. After that, GLM outperforms RAD in terms of ETS, POD, and CSI forecast at 30 dBZ but produced similar ETS scores at 40 dBZ. The ETSs for all DA experiments undergo a sharp decrease with time after 1 h forecast. Such a marked drop is especially evident for RAD and GLM + RAD, a result that is consistent with past cloud-scale lightning DA works highlighting that longer-term solutions tend to progressively become bounded by bias and errors contained in the large-scale environment (e.g., Fierro et al. 2015).

The hourly rainfall forecasts (Fig. 15) exhibit broadly consistent results with those from CREF. The 1-h forecasts from GLM + RAD and RAD are relatively

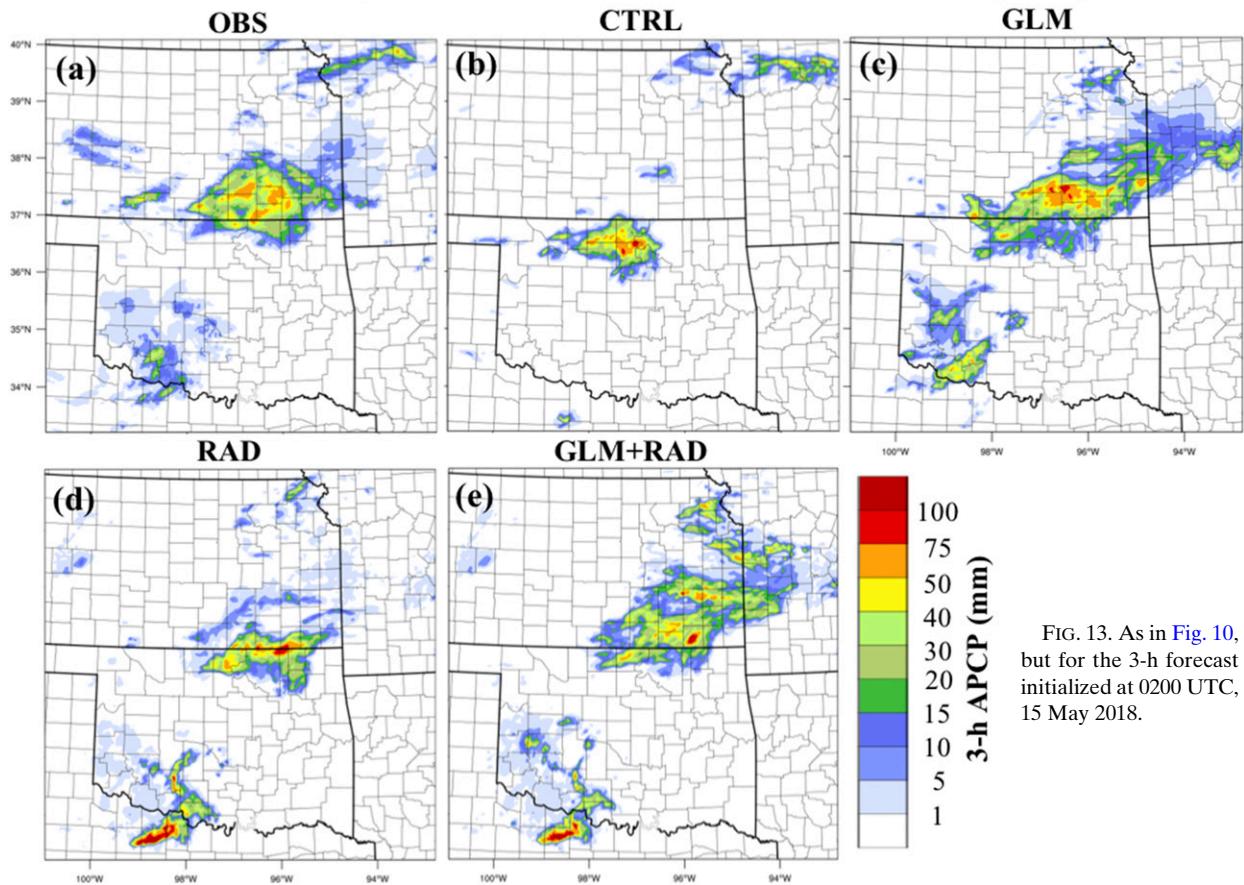## 0200 UTC , 15 May 2018 forecast (t = 3h) accumulated precipitation

FIG. 13. As in Fig. 10, but for the 3-h forecast initialized at 0200 UTC, 15 May 2018.

similar, with higher POD and SR than GLM (Figs. 15a–c). At 3-h forecast, however, GLM produces the largest POD (at the cost of producing the largest bias and lower SR). RAD shows overall the largest value of SR and CSI but smallest value of POD (Figs. 15a–c). Similar to CREF, both RAD and GLM + RAD produce the best ETS at 1 h, whose values experience a marked drop afterward (Figs. 15d–f). At 2–3 h forecasts, the ETS for GLM and GLM + RAD remain similar. For the 2-h forecast, GLM produces the highest ETS at 2.5 and 5.0 mm while RAD produces the best ETS at 10 mm. For all thresholds, RAD generally shows the largest ETS at 3-h forecasts although being small (e.g., <0.4 at 10 mm) and not significantly different from CTRL (∼0.25, Fig. 15f).

Overall, the assimilation of GLM lightning data over radar has a neutral to positive impact on the short-term forecasts of convective-scale severe weather events. In line with Fierro et al. (2019), the assimilation of GLM lightning data alone is able to notably improve short-term (1–2 h) forecast skill over CTRL. The sharp, gradual loss in model skill only a few hours (∼3 h)

following the analysis time has also been reported in prior cloud-scale 3DVAR DA studies assimilating radar (e.g., Schenkman et al. 2011; Gao et al. 2018) and/or lightning data (Fierro et al. 2015, 2019). When simultaneously assimilating both GLM and radar data, forecast skill is improved further relative to when either only radar or only lightning data are assimilated, especially for 0–1 h forecasts. Similar to previous DA studies involving adjustments of moisture- or precipitation-related proxy variables for lightning (or hydrometeors through a cloud analysis), however, a higher POD is generally accompanied by an increase in wet bias. Fierro et al. (2015) indicated that in the context of q_v-based LDA, there generally was a systematic trade-off between wet biases in the forecast and short-term forecast improvements. Such wet biases earlier on in the forecast can have profound detrimental impacts on the longer-term solution through faster saturation of errors at the smaller scales, and thus should be addressed in future endeavors either by building on the q_v conservation concept proposed in Fierro et al. (2019) or by exploring alternative means to confine

Performance of 0–3h forecast composite reflectivity relative to MRMS observations
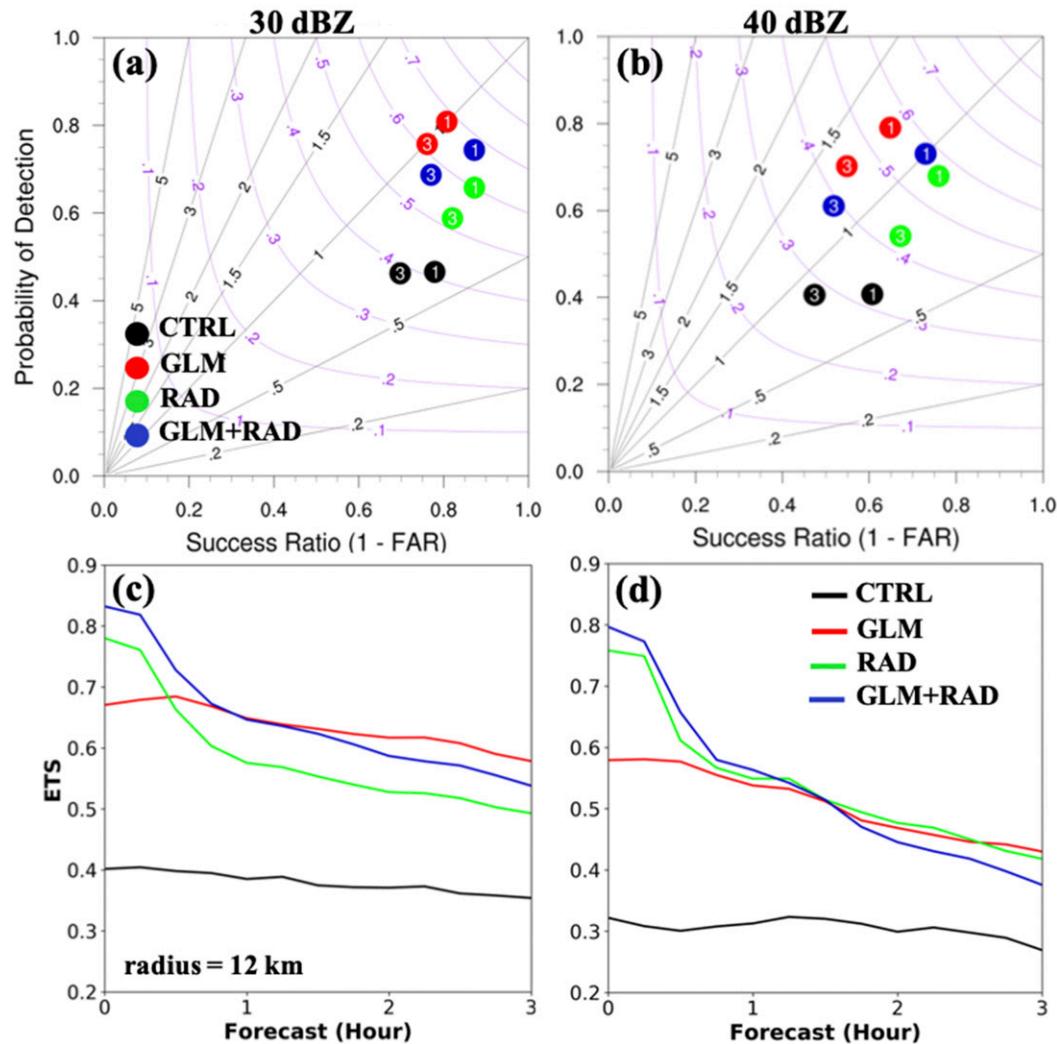

FIG. 14. Aggregate score metrics of composite reflectivity fields relative to MRMS observations for 19 three-hour forecasts over all five cases for experiments GLM, RAD, and GLM + RAD. (a),(b) The performance diagram and (c),(d) the equitable threat score (ETS) for (a),(c) 30 dBZ and (b),(d) 40 dBZ thresholds, respectively. Results are shown for a neighborhood radius of 12-km. The numbers in the circle in (a) and (b) denote the forecast length (i.e., 1- and 3-h forecasts).

the impact of moisture adjustments inferred by lightning or other types of observations near convective cores.

c. DA statistics

Following the line of work of Fierro et al. (2019), the variation of total cost function together with DA statistics including the mean and root-mean-square (RMS) of the innovation and analysis residual during DA cycles are analyzed to provide an evaluation of the quality and performance of the 3DVAR analyses. Because the results for each case were qualitatively similar, only one

case is presented for brevity, namely, 1 May 2018. The cost function minimized iteratively by the 3DVAR algorithm during each analysis is defined, for instance, by Eq. (1) in Gao et al. (2013). For a converging analysis, the cost function is anticipated to decrease with iterations in each minimization step.

The innovation defines the difference between the observation and the background, and the analysis residual measures the difference between the observation and the analysis. The mean innovation (referred to as MEAN$_{innov}$) and MEAN analysis residual (referred to

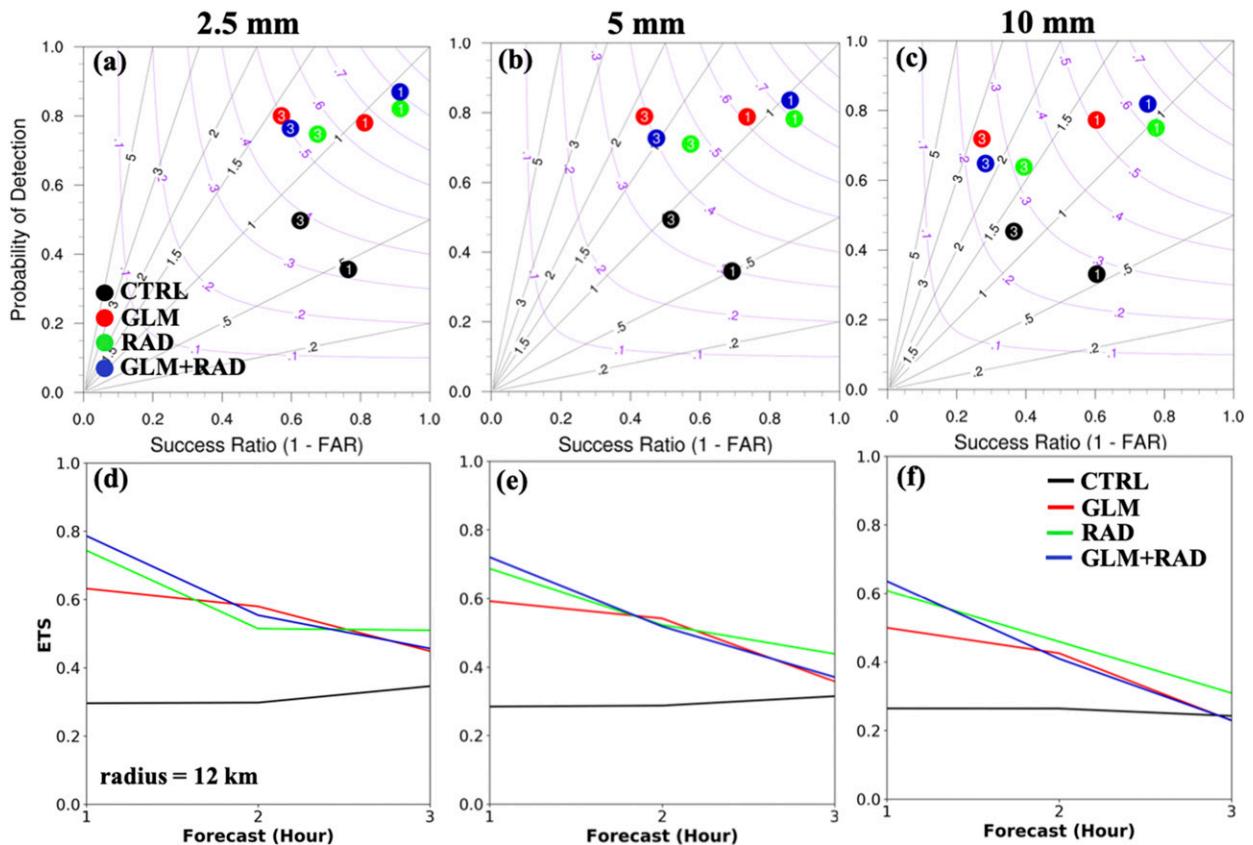## Performance of 0-3h forecast precipitation relative to Stage IV observations

FIG. 15. As in Fig. 14, but for hourly accumulated precipitation relative to Stage IV multisensor rainfall estimates calculated using the following thresholds: (a),(d) 2.5, (b),(e) 5, and (c),(f) 10 mm.

as MEAN$_{res}$) are calculated to estimate a measure of bias between the observation and model. The RMS statistics for the innovation (referred to as RMS$_{innov}$) and analysis residual (referred to as RMS$_{res}$) provide a measure of the Euclidian distance between the observation and the background and between the observation and the analysis, respectively. The reader is invited to consult Fierro et al. (2019) for details behind their formulations. Because statistics from both 3DVAR passes have similar behavior, only those for the second pass are discussed.

In general, the magnitude of the (normalized) total cost function decreases with iterations, with GLM producing the steepest drop (about 90%) during the first few iterations (Fig. 16a). In RAD and GLM + RAD, significantly more radar observations are being assimilated (volumetric scans) and because of the nonlinearity relationship between reflectivity and hydrometer variables, the total cost function decreases less rapidly (Figs. 16b,c). The lower values of RMS$_{res}$ compared to corresponding values of RMS$_{innov}$ for all

variables in each experiment indicate that the analysis is closer to the observations than the background, which confirms that the 3DVAR system generally ingests all the available observations reasonably well (Figs. 16d–f). The slight decrease of RMS$_{innov}$ for the GLM-derived q_v and radial velocity v_r with DA cycles in all experiments indicates particularly good assimilation of GLM-derived q_v observations and radar radial velocity observations by the system. The nonlinearity of the radar reflectivity operator (Gao et al. 2016) coupled with phase errors and the development of spurious convection all account for the gradual increases seen for RMS$_{innov}$ with cycling before 0000 UTC (Figs. 16e,f).

Positive MEAN$_{innov}$ and MEAN$_{res}$ values (close to zero) for q_v in (Figs. 16g,i) indicate that q_v fields from both the background and the analysis are smaller than the GLM-derived pseudo observations for q_v (which assumes near saturation w.r.t. water). This indicates that the 3DVAR analysis properly adjust the background q_v toward saturation within lightning columns.
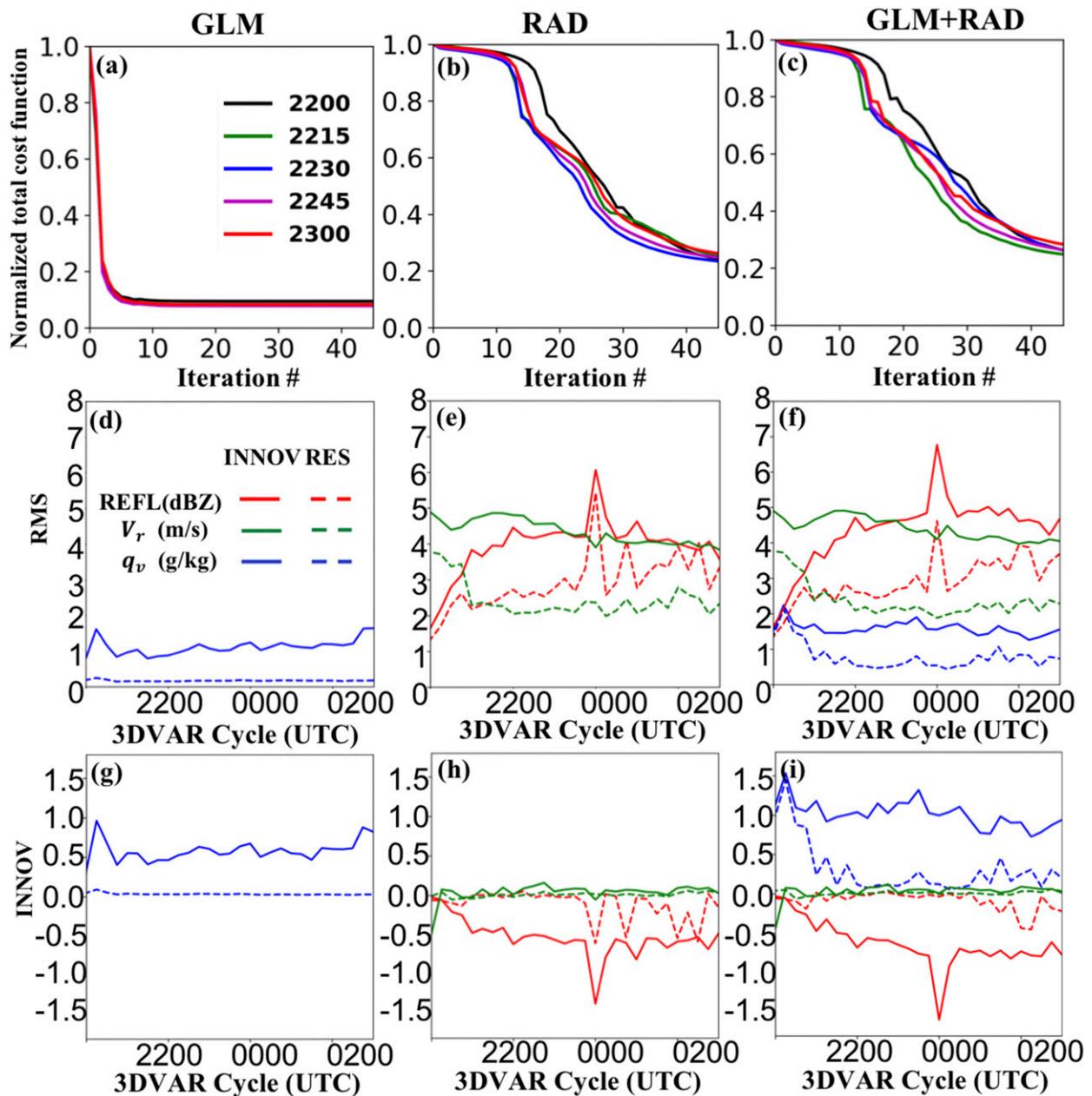
FIG. 16. (a)–(c) Change of the normalized total cost function with iteration number for (a) GLM, (b) RAD, and (c) GLM + RAD experiments during each cycle between 2200 and 2300 UTC 1 May 2018. The curve for each cycle is colored following the legends given in (a). (d)–(f) Time series of root-mean-square (RMS) innovation (solid curve) and analysis residual (dashed curve) for reflectivity factor (REFL in dBZ, red), radial velocity (V_r in m s^{-1}, blue), and water vapor mixing ratio (q_v in g kg^{-1}, green) for (d) GLM, (e) RAD, and (f) GLM + RAD experiments. (g)–(i) As in (d)–(f), but for the mean statistics.

MEAN$_{innov}$ and MEAN$_{res}$ values for v_r are quite small (Figs. 16h,i), indicating little overall bias. Both background and analysis have higher reflectivity values than the radar observations, as revealed by the negative MEAN$_{innov}$ and MEAN$_{res}$ values for reflectivity in both RAD-based experiments (Figs. 16h,i). This systematic overestimation of reflectivity by the 3DVAR

analysis accounts for the overprediction of rainfall as seen, for instance, in Figs. 10d and 10e.

6. Summary and conclusions

This study evaluated the impact of assimilating pseudo-q_v observations derived from spaceborne GLM

total lightning data into convective-scale NWP for short-term severe weather forecasts. The NSSL 3DVAR system was utilized to assimilate GLM lightning-derived pseudo-q_v observations by building on initial concepts put forth in Fierro et al. (2016, 2019) wherein water vapor mass mixing ratio is adjusted to near saturation within a confined layer at and around observed lightning locations.

To provide a more systematic assessment of the performance of GLM lightning DA, five high-impact weather events that occurred in May 2018 over the Great Plains of the United States were examined. For each of the five cases, the study domain was chosen based on the SPC "day 1" convective outlook product. To determine an optimal configuration for the GLM DA, sensitivity tests were conducted first, which focused on evaluating the impact of the (i) horizontal decorrelation length scale, (ii) 3DVAR cycling frequency, and (iii) accumulation period for the GLM data prior to each 3DVAR analysis. Qualitative and quantitative assessments of composite reflectivity fields and accumulated precipitation analysis/forecast against observations using aggregate statistics revealed that assimilating 10-min accumulated GLM lightning data every 15 min with a decorrelation length scale of 3 km for q_v yielded the best results.

With the above settings for the GLM DA, two companion experiments assimilating level II radar data (VAD-derived winds, radial velocity, and radar reflectivity) with or without GLM total lightning-derived pseudo-q_v observations were conducted. Overall, improvements on short-term (0–3 h) forecasts in the aggregate were revealed by all assimilation experiments including either radar and/or GLM lightning data. Similar to prior cloud-scale DA studies, forecast skill was characterized by a marked drop after about 1 h. For these five cases, the forecast skill for the experiment assimilating only GLM lightning-derived q_v data was comparable to that of the radar-only DA experiment (Fierro et al. 2016). Similar to previous cloud-scale lightning DA studies, the short-term improvements in precipitation forecasts were generally accompanied by a noteworthy increase in wet bias, especially later in the simulation ($\geq$3 h, e.g., Fierro et al. 2015). Overall, higher 0–1 h forecast skill was obtained when radar data were assimilated. The combination of GLM and RAD helps improve POD and ETS forecast but at the cost of exacerbating any existing wet biases (Fierro et al. 2016).

One of the critical limitations of the present LDA scheme lies in its inability to suppress spurious convection even when combined with current radar DA approach. Current parallel work is investigating modifications to the

radar DA scheme wherein radar reflectivity fields from background and observed fields are used in concert to identify areas of spurious convection where negative q_v increments should be imposed. Besides, available traditional observations such as pressure, temperature, moisture, and wind observations from surface observing platforms such as mesonets, as well as upper-air observing instruments such as radiosondes and wind profiles, could also be effectively assimilated in future work to help better represent the observed near-storm environment. It is relevant to also point out that this study only considered GLM flash centroids as an effort to primarily facilitate comparisons between previous studies that assimilated typical point-flash data from ground-based sensors (e.g., Fierro et al. 2012; Marchand and Fuelberg 2014; Fierro et al. 2019). Further research will be devoted at leveraging the GLM group and event locations, which better depict the spatial extent of the GLM flashes and, hence, could help better represent convective and stratiform areas in the analysis (Peterson 2019). Auxiliary spaceborne datasets such as the GOES-R Series Advanced Baseline Imager (ABI) product should also be used in tandem with the GLM to better depict the nature of clouds beyond electrically active regions.

REFERENCES

Alexander, G. D., J. A. Weinman, V. Karyampudi, W. S. Olson, and A. C. L. Lee, 1999: The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 Superstorm. *Mon. Wea. Rev.*, **127**, 1433–1457, https://doi.org/10.1175/1520-0493(1999)127<1433:TEOARR>2.0.CO;2.

Allen, B. J., E. R. Mansell, D. C. Dowell, and W. Deierling, 2016: Assimilation of pseudo-GLM data using the ensemble Kalman filter. *Mon. Wea. Rev.*, **144**, 3465–3486, https://doi.org/10.1175/MWR-D-16-0117.1.

Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly mul-tisensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.

Benjamin, S. G., G. A. Grell, J. M. Brown, and T. G. Smirnova, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494, https://doi.org/10.1175/1520-0493(2004) 132<0473:MWPWTR>2.0.CO;2.

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

Carey, L. D., and S. A. Rutledge, 1998: Electrical and multipa-rameter radar observations of a severe hailstorm. *J. Geophys. Res.*, **103**, 13 979–14 000, https://doi.org/10.1029/97JD02626.

Chang, D. E., J. A. Weinman, C. A. Morales, and W. S. Olson, 2001: The effect of spaceborne microwave and ground-based continu-ous lightning measurements on forecasts of the 1998 Groundhog Day storm. *Mon. Wea. Rev.*, **129**, 1809–1833, https://doi.org/ 10.1175/1520-0493(2001)129<1809:TEOSMA>2.0.CO;2.

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

Coniglio, M. C., S. M. Hitchcock, and K. H. Knopfmeier, 2016: Impact of assimilating preconvective upsonde observations on short-term forecasts of convection observed during MPEX. *Mon. Wea. Rev.*, **144**, 4301–4325, https://doi.org/ 10.1175/MWR-D-16-0091.1.

Deierling, W., and W. A. Petersen, 2008: Total lightning activity as an indicator of updraft characteristics. *J. Geophys. Res.*, **113**, D16210, https://doi.org/10.1029/2007JD009598.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two–dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, https://doi.org/ 10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Fierro, A. O., M. S. Gilmore, E. R. Mansell, L. J. Wicker, and J. M. Straka, 2006: Electrification and lightning in an ideal-ized boundary-crossing supercell simulation of 2 June 1995. *Mon. Wea. Rev.*, **134**, 3149–3172, https://doi.org/10.1175/ MWR3231.1.

——, E. R. Mansell, C. L. Ziegler, and D. R. MacGorman, 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. *Mon. Wea. Rev.*, **140**, 2609–2627, https://doi.org/10.1175/MWR-D-11-00299.1.

——, J. Gao, C. L. Ziegler, E. R. Mansell, D. R. MacGorman, and S. R. Dembek, 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. *Mon. Wea. Rev.*, **142**, 183–202, https://doi.org/10.1175/MWR-D-13-00142.1.

——, A. J. Clark, E. R. Mansell, D. R. MacGorman, S. Dembek, and C. Ziegler, 2015: Impact of storm-scale lightning data assimilation on WRF-ARW precipitation forecasts during the 2013 warm season over the contiguous United States. *Mon. Wea. Rev.*, **143**, 757–777, https://doi.org/10.1175/ MWR-D-14-00183.1.

——, J. Gao, C. L. Ziegler, K. M. Calhoun, E. R. Mansell, and D. R. MacGorman, 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. *Mon. Wea. Rev.*, **144**, 4373–4393, https://doi.org/10.1175/MWR-D-16-0053.1.

——, S. Stevenson, and R. Rabin, 2018a: Evolution of GLM-observed total lightning in Hurricane Maria (2017) during the period of maximum intensity. *Mon. Wea. Rev.*, **146**, 1641–1666, https://doi.org/10.1175/MWR-D-18-0066.1.

——, and Coauthors, 2018b: 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*, No. 58, Joint Center for Satellite Data Assimilation, Boulder, CO, 5–12, https:// doi.org/10.7289/V5CJ8BR2.

——, Y. Wang, J. Gao, and E. R. Mansell, 2019: Variational as-similation of radar data and GLM lightning-derived water vapor for the short-term forecasts of high-impact convective events. *Mon. Wea. Rev.*, **147**, 4045–4069, https://doi.org/ 10.1175/MWR-D-18-0421.1.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2007: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135**, 1846–1868, https://doi.org/10.1175/MWR3391.1.

Gao, J., and D. J. Stensrud, 2012: Assimilation of reflectivity data in a convective-scale, cycled 3DVAR framework with hy-drometeor classification. *J. Atmos. Sci.*, **69**, 1054–1065, https:// doi.org/10.1175/JAS-D-11-0162.1.

——, and ——, 2014: Some observing system simulation experi-ments with a hybrid 3DEnVAR system for storm-scale radar data assimilation. *Mon. Wea. Rev.*, **142**, 3326–3346, https:// doi.org/10.1175/MWR-D-14-00025.1.

——, M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, **21**, 457–469, https://doi.org/10.1175/1520-0426(2004) 021<0457:ATVDAM>2.0.CO;2.

——, and Coauthors, 2013: A real-time weather-adaptive 3DVAR analysis system for severe weather detections and warnings. *Wea. Forecasting*, **28**, 727–745, https://doi.org/10.1175/WAF-D-12-00093.1.

——, C. Fu, J. Kain, and D. J. Stensrud, 2016: OSSEs for an en-semble 3DVAR data assimilation system with radar obser-vations of convective storms. *J. Atmos. Sci.*, **73**, 2403–2426, https://doi.org/10.1175/JAS-D-15-0311.1.

Gao, S., J. Sun, J. Min, Y. Zhang, and Z. Ying, 2018: A scheme to assimilate "no rain" 575 observations from Doppler radar. *Wea. Forecasting*, **33**, 71–88, https://doi.org/10.1175/WAF-D-17-0108.1.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

Giannaros, T. M., V. Kotroni, and K. Lagouvardos, 2016: A lightning data assimilation technique implemented in the WRF model for improving precipitation forecasts. *Environ. Modell. Software*, **76**, 54–68, https://doi.org/10.1016/j.envsoft. 2015.11.017.

Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.*, **125–126**, 34–49, https://doi.org/10.1016/j.atmosres.2013.01.006.

Heath, N. K., J. E. Pleim, R. C. Gilliam, and D. Kang, 2016: A simple lightning assimilation technique for improving ret-rospective WRF simulations. *J. Adv. Model. Earth Syst.*, **8**, 1806–1824, https://doi.org/10.1002/2016MS000735.

Hitchcock, S. M., M. C. Coniglio, and K. H. Knopfmeier, 2016: Impact of MPEX upsonde observations on ensemble analyses

and forecasts of the 31 May 2013 convective event over Oklahoma. *Mon. Wea. Rev.*, **144**, 2889–2913, https://doi.org/10.1175/MWR-D-15-0344.1.

Hong, S.-Y., 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.

Hu, M., M. Xue, and K. Brewster, 2006a: 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. *Mon. Wea. Rev.*, **134**, 675–698, https://doi.org/10.1175/MWR3092.1.

——, ——, J. Gao, and K. Brewster, 2006b: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134**, 699–721, https://doi.org/10.1175/MWR3093.1.

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

Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, https://doi.org/10.1175/MWR-D-14-00345.1.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

——, X. Wang, P. Skinner, A. Johnson, and Y. Wang, 2018: Assimilation of *GOES-13* imager clear-sky water vapor (6.5 μm) radiances into a Warn-on-Forecast system. *Mon. Wea. Rev.*, **146**, 1077–1107, https://doi.org/10.1175/MWR-D-17-0280.1.

Knopfmeier, K. H., and D. J. Stensrud, 2013: Influence of mesonet observations on the accuracy of surface analyses generated by an ensemble Kalman filter. *Wea. Forecasting*, **28**, 815–841, https://doi.org/10.1175/WAF-D-12-00078.1.

Kuhlman, K. M., C. L. Zielger, E. R. Mansell, D. R. MacGorman, and J. M. Straka, 2006: Numerically simulated electrification and lightning of the 29 June 2000 STEPS supercell storm. *Mon. Wea. Rev.*, **134**, 2734–2757, https://doi.org/10.1175/MWR3217.1.

Lagouvardos, K., V. Kotroni, E. Defer, and O. Bousquet, 2013: Study of a heavy precipitation event over southern France, in the frame of HYMEX project: Observational analysis and model results using assimilation of lightning. *Atmos. Res.*, **134**, 45–55, https://doi.org/10.1016/j.atmosres.2013.07.003.

Li, W., Y. F. Xie, S. M. Deng, and Q. Wang, 2010: Application of the multigrid method to the two-dimensional Doppler radar radial velocity data assimilation. *J. Atmos. Oceanic Technol.*, **27**, 319–332, https://doi.org/10.1175/2009JTECHA1271.1.

Lilly, D. K., 1990: Numerical prediction of thunderstorms—Has its time come? *Quart. J. Roy. Meteor. Soc.*, **116**, 779–798, https://doi.org/10.1002/qj.49711649402.

MacGorman, D. R., D. W. Burgess, V. Mazur, W. D. Rust, W. L. Taylor, and B. C. Johnson, 1989: Lightning rates relative to

tornadic storm evolution on 22 May 1981. *J. Atmos. Sci.*, **46**, 221–251, https://doi.org/10.1175/1520-0469(1989)046<0221:LRRTTS>2.0.CO;2.

——, I. R. Apostolakopoulos, N. R. Lund, N. W. S. Demetriades, M. J. Murphy, and P. R. Krehbiel, 2011: The timing of cloud-to-ground lightning relative to total lightning activity. *Mon. Wea. Rev.*, **139**, 3871–3886, https://doi.org/10.1175/MWR-D-11-00047.1.

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

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

Makowski, J. A., D. R. Macgorman, M. I. Biggerstaff, and W. H. Beasley, 2013: Total lightning characteristics relative to radar and satellite observations of Oklahoma mesoscale convective systems. *Mon. Wea. Rev.*, **141**, 1593–1611, https://doi.org/10.1175/MWR-D-11-00268.1.

Mansell, E. R., 2014: Storm-scale ensemble Kalman filter assimilation of total lightning-extent data. *Mon. Wea. Rev.*, **142**, 3683–3695, https://doi.org/10.1175/MWR-D-14-00061.1.

——, and C. L. Ziegler, 2013: Aerosol effects on simulated storm electrification and precipitation in a two-moment bulk microphysics model. *J. Atmos. Sci.*, **70**, 2032–2050, https://doi.org/10.1175/JAS-D-12-0264.1.

——, ——, and D. R. MacGorman, 2007: A lightning data assimilation technique for mesoscale forecast models. *Mon. Wea. Rev.*, **135**, 1732–1748, https://doi.org/10.1175/MWR3387.1.

——, ——, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, https://doi.org/10.1175/2009JAS2965.1.

Marchand, M. R., and H. E. Fuelberg, 2014: Assimilation of lightning data using a nudging method involving low-level warming. *Mon. Wea. Rev.*, **142**, 4850–4871, https://doi.org/10.1175/MWR-D-14-00076.1.

Medici, G., K. L. Cummins, W. J. Koshak, S. D. Rudlosky, R. J. Blakeslee, S. L. Goodman, and D. J. Cecil, 2017: The intra-cloud lightning fraction in the contiguous United States. *Mon. Wea. Rev.*, **145**, 4481–4499, https://doi.org/10.1175/MWR-D-16-0426.1.

Minamide, M., and F. Zhang, 2018: Assimilation of all-sky infrared radiances from *Himawari-8* and impacts of moisture and hydrometer initialization on convection-permitting tropical cyclone prediction. *Mon. Wea. Rev.*, **146**, 3241–3258, https://doi.org/10.1175/MWR-D-17-0367.1.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Papadopoulos, A., T. G. Chronis, and E. N. Anagnostou, 2005: Improving convective precipitation forecasting through assimilation of regional lightning measurements in a mesoscale model. *Mon. Wea. Rev.*, **133**, 1961–1977, https://doi.org/10.1175/MWR2957.1.

——, E. Serpetzoglou, and E. N. Anagnostou, 2009: Evaluating the impact of lightning data assimilation on mesoscale model simulations of a flash flood inducing storm. *Atmos. Res.*, **94**, 715–725, https://doi.org/10.1016/j.atmosres.2009.05.008.

Pessi, A. T., and S. Businger, 2009: The impact of lightning data assimilation on a winter storm simulation over the North Pacific Ocean. *Mon. Wea. Rev.*, **137**, 3177–3195, https://doi.org/10.1175/2009MWR2765.1.

Petersen, W. A., and S. A. Rutledge, 1998: On the relationship between cloud-to-ground lightning and convective rainfall. *J. Geophys. Res.*, **103**, 14 025–14 040, https://doi.org/10.1029/97JD02064.

Peterson, M., 2019: Research applications for the Geostationary Lightning Mapper operational lightning flash data product. *J. Geophys. Res. Atmos.*, **124**, 10 205–10 231, https://doi.org/10.1029/2019JD031054.

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

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Rudlosky, S. D., and H. E. Fuelberg, 2013: Documenting storm severity in the Mid-Atlantic region using lightning and radar information. *Mon. Wea. Rev.*, **141**, 3186–3202, https://doi.org/10.1175/MWR-D-12-00287.1.

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

Schenkman, A. D., X. Ming, A. Shapiro, K. Brewster, and J. Gao, 2011: The analysis and prediction of the 8–9 May 2007 Oklahoma tornadic mesoscale convective system by assimilating WSR-88D and CASA radar data using 3DVAR. *Mon. Wea. Rev.*, **139**, 224–246, https://doi.org/10.1175/2010MWR3336.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Smith, T., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677, https://doi.org/10.1175/2555.1.

Sobash, R. A., and D. J. Stensrud, 2015: Assimilating surface mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 May 2012. *Mon. Wea. Rev.*, **143**, 3700–3725, https://doi.org/10.1175/MWR-D-14-00126.1.

Stensrud, D. J., N. Yussouf, D. C. Dowell, and M. C. Coniglio, 2009: Assimilating surface data into a mesoscale model ensemble: Cold pool analyses from spring 2007. *Atmos. Res.*, **93**, 207–220, https://doi.org/10.1016/j.atmosres.2008.10.009.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Wang, H., and Coauthors, 2018: Continuous assimilation of lightning data using time-lagged ensembles for a convection-allowing numerical weather prediction model. *J. Geophys. Res. Atmos.*, **123**, 9652–9673, https://doi.org/10.1029/2018JD028494.

Wang, Y., and X. Wang, 2017: Direct assimilation of radar reflectivity without tangent linear and adjoint of the nonlinear observation operator in the GSI-based EnVar system: Methodology and experiment with the 8 May 2003 Oklahoma City tornadic supercell. *Mon. Wea. Rev.*, **145**, 1447–1471, https://doi.org/10.1175/MWR-D-16-0231.1.

——, Y. Yang, and C. Wang, 2014: Improving forecasting of strong convection by assimilating cloud-to-ground lightning data using the physical initialization method. *Atmos. Res.*, **150**, 31–41, https://doi.org/10.1016/j.atmosres.2014.06.017.

——, J. Gao, P. S. Skinner, K. Knopfmeier, T. Jones, P. L. Heiselman, and L. J. Wicker, 2018: Test of a weather-adaptive dual-resolution hybrid 3DEnVAR and WRF-DART analysis and forecast system for severe weather events. *22nd Conf. on Integrated Observing and Assimilation Systems for the Atmosphere, Ocean, and Land Surface (IOAS-AOLS)*, Austin, TX, Amer. Meteor. Soc., 168, https://ams.confex.com/ams/98Annual/webprogram/Manuscript/Paper330920/98AMSConf_paper_LWJ.pdf.

Weiss, S. A., D. R. MacGorman, and K. M. Calhoun, 2012: Lightning in the anvils of supercell thunderstorms. *Mon. Wea. Rev.*, **140**, 2064–2079, https://doi.org/10.1175/MWR-D-11-00312.1.

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.

Xie, Y., S. E. Koch, J. A. McGinley, S. Albers, P. Bieringer, M. Wolfson, and M. Chan, 2011: A space and time multiscale analysis system: A sequential variational analysis approach. *Mon. Wea. Rev.*, **139**, 1224–1240, https://doi.org/10.1175/2010MWR3338.1.

Yussouf, N., D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, https://doi.org/10.1175/MWR-D-14-00268.1.

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, https://doi.org/10.1175/2011BAMS-D-11-00047.1.

Zhang, R., Y. Zhang, L. Xu, D. Zheng, and W. Yao, 2017: Assimilation of total lightning data using the three-dimensional variational method at convection-allowing resolution. *J. Meteor. Res.*, **31**, 731–746, https://doi.org/10.1007/s13351-017-6133-3.

Zhang, Y., F. Zhang, and D. J. Stensrud, 2018: Assimilating all-sky infrared radiances from *GOES-16* ABI using an ensemble Kalman filter for convection-allowing severe thunderstorms prediction. *Mon. Wea. Rev.*, **146**, 3363–3381, https://doi.org/10.1175/MWR-D-18-0062.1.

Zhuang, Z., N. Yussouf, and J. Gao, 2016: Analyses and forecasts of a tornadic supercell outbreak using a 3DVAR system ensemble. *Adv. Atmos. Sci.*, **33**, 544–558, https://doi.org/10.1007/s00376-015-5072-0.

Ziegler, C. L., 1985: Retrieval of thermal and microphysical variables in observed convective storms. Part I: Model development and preliminary testing. *J. Atmos. Sci.*, **42**, 1487–1509, https://doi.org/10.1175/1520-0469(1985)042<1487:ROTAMV>2.0.CO;2.