Impact of Assimilating GOES-R Geostationary Lightning Mapper Flash Extent Density Data on Severe Convection Forecasts in a Warn-on-Forecast System

YAPING WANG,[a,b] NUSRAT YUSSOUF,[a,b,c] EDWARD R. MANSELL,[b] BRIAN C. MATILLA,[a,b] RONG KONG,[d] MING XUE,[d,c] AND VANNA C. CHMIELEWSKI[a,b]

[a] *Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma*
[b] *NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma*
[c] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[d] *Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 17 December 2020, in final form 26 June 2021)

ABSTRACT: The Geostationary Operational Environmental Satellite-R (GOES-R) Geostationary Lightning Mapper (GLM) instrument detects total lightning rate at high temporal and spatial resolution over the Americas and adjacent oceanic regions. The GLM observations provide detection and monitoring of deep electrified convection. This study explores the impact of assimilating the GLM-derived flash extent density (FED) on the analyses and short-term forecasts of two severe weather events into an experimental Warn-on-Forecast system (WoFS) using the ensemble Kalman filter data assimilation technique. Sensitivity experiments are conducted using two tornadic severe storm events: one with a line of individual supercells and the other one with both isolated cells and a severe convective line. The control experiment (CTRL) assimilates conventional surface observations and geostationary satellite cloud water path into WoFS. Additional experiments also assimilate either GLM FED or radar data (RAD), or a combination of both (RAD+GLM). It is found that assimilating GLM data in the absence of radar data into the WoFS improves the short-term forecast skill over CTRL in one case, while in the other case it degrades the forecast skill by generating weaker cold pools and overly suppressing convection, mainly owing to assimilating zero FED values in the trailing stratiform regions. Assimilating unexpectedly low FED values in some regions due to low GLM detection efficiency also accounts for the poorer forecasts. Although RAD provides superior forecasts over GLM, the combination RAD+GLM shows further gains in both cases. Additional observation operators should consider different storm types and GLM detection efficiency.

KEYWORDS: Lightning; Numerical analysis/modeling; Numerical weather prediction/forecasting; Data assimilation; Ensembles

1. Introduction

Lightning activity, especially total lightning flash rate (cloud-to-ground and intracloud), has been recognized as a good indicator of deep electrified convection. Previous studies have shown that total flash rate is well correlated with measures of graupel and ice content, updraft intensity, and rainfall amount in single cell and multicell storms (e.g., Goodman et al. 1988; MacGorman et al. 1989; Carey and Rutledge 1996; Wiens et al. 2005; Deierling and Petersen 2008; Kuhlman et al. 2006; Fierro et al. 2006), as well as in tropical cyclones (Fierro et al. 2015b; Fierro and Mansell 2017, 2018).

The recently operational Geostationary Lightning Mapper (GLM) instrument on board the Geostationary Operational Environmental Satellite-R (GOES-R) series of weather satellites greatly extends the ability of total lightning detection. The GLM continuously detects total lightning over the Americas and adjacent oceanic regions in the western hemisphere, providing lightning products with high temporal (2 ms) and good spatial (8–12 km) resolution (Goodman et al. 2013). The GLM lightning observations are a valuable source of convective activity information, which are a potential complement to ground-based networks especially in mountainous and oceanic regions, where both radar coverage and the detection efficiency (DE) of other lightning networks are poor.

Earlier studies have demonstrated the positive impact of lightning data assimilation (LDA) in mesoscale and convective-scale models. Through the simple and effective nudging methods, previous studies used lightning data to adjust the latent heating rate (e.g., Alexander et al. 1999; Chang et al. 2001; Pessi and Businger 2009b) or the humidity profiles (Papadopoulos et al. 2005, 2009) in a convection parameterization scheme (CPS), to force the CPS directly (Mansell et al. 2007), to adjust the water vapor mixing ratio at observed lightning locations (Fierro et al. 2012, 2014, 2015a, 2016), to warm the updraft source layer at lightning locations to promote convection (Marchand and Fuelberg 2014), and to adjust the graupel mixing ratio and latent heat (Wang et al. 2017). Fierro et al. (2016) evaluated the added value of LDA approach from Fierro et al. (2012) in tandem with radar data in a three-dimensional variational (3DVAR) framework. In an ensemble Kalman filter (EnKF) framework, Hakim et al. (2008) was the first one to assimilate the lightning data using the EnKF method by converting the lightning data to convective rainfall amount through the lightning-rainfall relationship following Pessi et al. (2006). The correlation between lightning and precipitation were explored by analyzing available lightning and rainfall observations (e.g., Pessi et al. 2004; Pessi and Businger 2009a). Mansell (2014)

Corresponding author: Yaping Wang, yaping.wang@noaa.gov

DOI: 10.1175/MWR-D-20-0406.1

directly assimilated the flash extent density (FED) derived from simulated total lightning data using the EnKF in an observing system simulation experiment. Allen et al. (2016) applied the LDA technique in Mansell (2014) to real cases using pseudo-GLM data derived from ground-based Lightning Mapping Array (LMA) data.

Most of the previous LDA studies used lightning data detected from ground-based platforms, such as the LMAs (Mansell et al. 2007; Allen et al. 2016), the Earth Networks Total Lightning Network (ENTLN, e.g., Fierro et al. 2012, 2015a; Marchand and Fuelberg 2014), the U.S. National Lightning Detection Network (NLDN, e.g., Mansell et al. 2007; Alexander et al. 1999), and the Pacific Lightning Detection Network/Long-Range Lightning Detection Network (PacNet/LLDN, e.g., Pessi and Businger 2009b). As GLM lightning data became available, Fierro et al. (2019) and Hu et al. (2020) converted GLM flash centroid locations to pseudowater vapor observations, which were assimilated in a 3DVAR system and demonstrated the added value of lightning with radar data also being assimilated. Kong et al. (2020) for the first time assimilated GLM-derived FED into the EnKF framework.

Building upon Allen et al. (2016) and Kong et al. (2020), the present study assesses the impact of assimilating GLM-derived FED using the EnKF technique into the National Severe Storms Laboratory (NSSL) Experimental Warn-on-Forecast System (WoFS) with all available observations including radar data also being assimilated. The WoFS is a convection-allowing ensemble analyses and forecast system designed to provide short-range (0–6 h) probabilistic guidance for thunderstorm hazards (e.g., tornadoes, severe hail, damaging winds, and flash flooding). The WoFS has been run in real time for NOAA Hazardous Weather Testbed (HWT) since 2016 (Skinner et al. 2018), and has demonstrated the ability to provide accurate short-term probabilistic forecast for thunderstorm hazards (Wheatley et al. 2015; Yussouf et al. 2013, 2015; Jones et al. 2016) and heavy rainfall events (Lawson et al. 2018; Yussouf et al. 2016, 2020; Yussouf and Knopfmeier 2019; Jones et al. 2019).The current experimental WoFS assimilates conventional observations, reflectivity and radial velocity from Weather Surveillance Radar-1988 Doppler (WSR-88D), and satellite retrievals.

While Kong et al. (2020) evaluated the performance of GLM data assimilation for one MCS case in the absence of radar data, this study further explores the potential value of GLM LDA in tandem with radar DA for two tornadic severe storm events with different convection patterns using WoFS. As experimental WoFS system output is increasingly used in the forecasting community, the results of GLM LDA into the system could be very impactful. The following questions are addressed in this work: How does GLM LDA perform compared to radar DA? Is there a forecast benefit from additional GLM LDA in combination with radar DA? How does GLM detection efficiency impact DA performance? Following the introduction, section 2 describes the DA experiments setups, including the DA system, the processing of GLM data and the lightning observation operator. Section 3a discusses the optimal configuration for GLM DA, and section 3b compares the impacts of radar DA and lightning DA. Summary and conclusions are given in section 4.

2. The data assimilation experiments

a. GSI-EnKF data assimilation system

The WoFS is an on-demand, ensemble DA and prediction system. The initial and boundary conditions for the WoFS are provided by the experimental High-Resolution Rapid Refresh Ensemble (HRRRE; Dowell et al. 2016). The configuration used in this work is similar to the ones used in a prototype real-time WoFS run during the 2018 and 2019 NOAA HWT Spring Forecast Experiment (SFE; Kain et al. 2003; Gallo et al. 2017; Clark et al. 2020). The WoFS uses the Advanced Research Version of the Weather Research and Forecasting (WRF-ARW) Model, version 3.8.1 (Skamarock et al. 2008) and comprises an ensemble of 36 members. The multiphysics ensemble is configured with different sets of boundary layer and radiation schemes (e.g., Stensrud et al. 2000; Wheatley et al. 2014) and initial and boundary conditions. The cloud microphysics parameterization scheme used in all ensemble members is the NSSL two-moment scheme (Mansell et al. 2010). A domain was selected to cover the primary severe weather event each day based on the Storm Prediction Center's Day 1 convective outlook. The domain contains 250 × 250 grid points (approximately 750 km × 750 km) for the 2018 version and 300 × 300 grid points (approximately 900 km ×900 km) for the 2019 version, both with a 3-km horizontal grid spacing. Conventional observations, radar reflectivity and radial velocity from the WSR-88D, and satellite-derived cloud water path (CWP) are assimilated at 15-min intervals using the Community Gridpoint Statistical Interpolation (GSI; Kleist et al. 2009) based ensemble Kalman filter (EnKF; Houtekamer and Mitchell 2005) (GSI-EnKF) DA system. Conventional observations, which include pressure, temperature, moisture, and wind from surface and marine stations, Aircraft Communications Addressing and Reporting Systems (ACARS), air report (AIREP)/pilot weather report (PIREP), rawinsondes, Next Generation Weather Radar (NEXRAD) velocity azimuth display (VAD) wind reports, and profilers from NCEP's PrepBUFR files, and Oklahoma mesonet observations (McPherson et al. 2007) are assimilated within the WoFS domain. Assimilated WSR-88D reflectivity data derived from the 1-km Multi-Radar Multi-Sensor (MRMS) products are objectively analyzed to 5 km spacing using a Cressman interpolation scheme (Cressman 1959; Smith et al. 2016). Radial velocites from WSR-88D data are also objectively analyzed to the 5-km grid using the same Cressman scheme. Cloud water path (CWP) retrievals from the *Geostationary Operational Environmental Satellite-16* (*GOES-16*) imager data are also assimilated. A complete description of the WoFS is available in Yussouf et al. (2015), Wheatley et al. (2015), Jones et al. (2016, 2018), and Skinner et al. (2018).

b. GLM FED data processing

The publicly available GLM 20-s Level-2 data files contain the three-level parent–child hierarchy metrics, which are flashes, groups, and events, at the pixel resolution of approximately

8 km at nadir and less than 12 km over the continental United States (CONUS). An event is the most elementary product, defined as the occurrence of a single pixel exceeding the background threshold during a single 2-ms frame. A group consists of adjacent events within the same time frame. A set of groups that are sequentially separated by 330 ms or less in time and no more than 16.5 km in space are aggregated as a flash (Goodman et al. 2013). Each event, group or flash includes a centroid latitude and longitude, time and number identification (ID). Each event or group also includes a parent ID for the next level up the hierarchy. That is, events belonging to the same group share a same parent group and flash ID; groups belonging to the same flash share the same parent flash ID. More details regarding the GLM data structure could be found in Goodman et al. (2013), Bruning et al. (2019), and Kong et al. (2020).

The FED derived from the GLM data are assimilated in this study. Events, groups, and flashes in the GLM dataset are reported as geolocated points, and FED is defined as the count of flashes that propagate through a grid column during some period of time. In this study, the GLM FED is obtained by utilizing a Python package called "glmtools" (Bruning et al. 2019). In this algorithm, the GLM grids are first produced in the GOES fixed grid coordinates, which is the standard coordinate reference frame for the Advanced Baseline Imager (ABI) observations and other geostationary satellite observations. The algorithm restores the spatial footprint of GLM events on the fixed grid coordinates, and then connects the spatial footprint of events to flashes through the parent–child relationship information (flash/group IDs stored in each event). The events belonging to different flashes in a fixed grid pixel within a certain time period (e.g., 1 and 5 min in this work) are counted as the number of flashes passing through that pixel, which is the calculated FED. The fixed grid coordinates are then converted to a longitude/latitude array by intersecting the fixed grid angle with Earth's surface before DA. The algorithm in glmtools provides a universal solution to the gap and double-counting problems caused by resampling the GLM data to grids with a smaller or larger grid spacing. In this study, the pixel width at nadir is set to 8 km for the fixed grids. After the fixed grid projecting to Earth's surface, the pixel width of GLM data on ground varies from ~8 km at nadir to ~14 km at corners, similar to the GLM charge-coupled device (CCD) detection pixel size.

c. Observation operator

A FED observation operator based on graupel mass (FEDM) from Allen et al. (2016) and then tuned by Kong et al. (2020) is employed in this study. In Allen et al. (2016), FEDM is a linear best-fit operator between model flash rate and graupel mass derived from a simulation with radar radial velocity assimilated. Their work also showed that the operator performs well when using 1-min pseudo-GLM FED data, which was derived from LMA data. To evaluate the feasibility of this operator using real GLM FED, Kong et al. (2020) conducted several sensitivity experiments and found that applying a scaling factor of 0.5 to the operator is more suitable for the real 1-min GLM FED data, which is shown below:

$$\mathrm{FEDM} = 0.5 \times 2.088 \times 10^{-8}(\mathrm{GM}). \quad (1)$$

In this study, Eq. (1) is used as the observation operator and GM is the vertically integrated graupel mass in a volume covering a $15 \times 15\,\mathrm{km}^2$ area in the horizontal dimensions centered on model grids before being interpolated to the GLM FED locations. The horizontal coverage of the graupel mass is slightly larger than the pixel size of the GLM data, in order to account for lightning occurring on the edge of a GLM pixel (Mansell 2014; Allen et al. 2016). Zero FED observations are also assimilated in order to limit spurious deep convection (Mansell 2014). Following Allen et al. (2016) and Kong et al. (2020), the 1-min accumulated FED observations are used and are assigned to a height of 6.5 km in this study. The observation error is set to $0.5\,\mathrm{min}^{-1}$ per pixel (Kong et al. 2020). Also similar to Allen et al. (2016) and Kong et al. (2020), the horizontal localization radius for FED is set to 15 km and the vertical localization is set to 4.0 in units of scale height, which corresponds to a very large radius of influence. A large vertical localization allows the two-dimensional GLM-FED observations impact the entire domain depth. In addition, the outlier check in GSI-EnKF is turned off for FED observations in order to maximize the impact of FED DA.

d. Case description and experimental design

In this study, two high-impact weather events that occurred on 1–2 May 2018 and 20 May 2019 are examined. The 1–2 May 2018 severe weather outbreak occurred across Kansas and southeastern regions of Nebraska, where convection first initiated along a front across western Kansas and southeastern Nebraska between 1900 and 2000 UTC. In the following hours, several thunderstorms including several isolated tornadic supercells, further developed and later merged into a quasi-linear mesoscale convective system. There was a total of 18 tornadoes as well as several damaging wind and large hail reports in Kansas, Nebraska, and Oklahoma for this case. The other case examined is the severe weather outbreak on 20–21 May 2019 that occurred across northern Texas, Oklahoma, southeastern Kansas, and western Missouri. High instability and deep-layer shear favored a number of supercells to develop and to merge into a squall line, with a broad trailing stratiform region. A total of 37 tornadoes and many severe wind and damaging hail reports were recorded in Texas, Oklahoma, Kansas, and Missouri. The 20 May 2019 event was a challenging case to forecasters as well. The National Weather Service (NWS) Storm Prediction Center (SPC) issued a high-risk outlook for tornadoes and severe thunderstorms from northwest Texas into western and central Oklahoma. While tornadoes and other severe weather did happen over the high-risk region, they were not as intense or as widespread as anticipated, especially in central Oklahoma.

The simulation domains and the locations of radar sites for the two cases are shown in Fig. 1. The continuous 15-min DA cycling starts at 1800 UTC and ends at 0300 UTC for the 1 May 2018 event, and from 1700 to 0100 UTC for the 20 May 2019 event. The 3-h forecasts are launched at 1-h intervals when thunderstorms were most active in each case, between 2100 and 0300 UTC, and between 1900 and 0100 UTC, respectively.

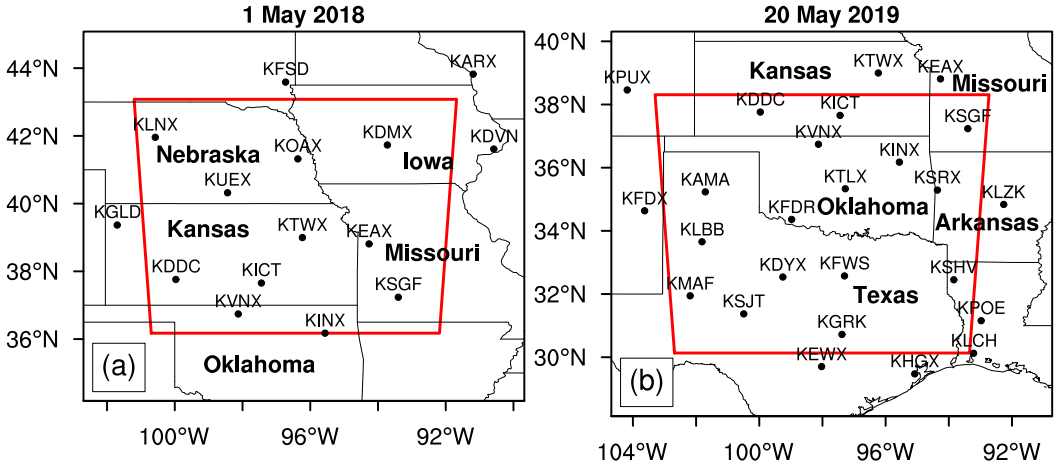

FIG. 1. Simulation domains (red polygon) and the locations of radar sites for two cases.

A set of sensitivity experiments are performed to evaluate the added value of assimilating GLM FED into WoFS on short-term forecasts (Table 1). A control experiment (CTRL) with only the conventional observations and the CWP data assimilated, and an experiment (GLM_1min) with additional 1-min GLM FED data assimilated into CTRL are first conducted to show the effect of GLM assimilation. In WoFS, smooth and random additive noise is introduced to the u, v, w, θ, and T_d variables of each ensemble member in and near high-reflectivity regions to help initiate convection in the model and to maintain ensemble spread (Caya et al. 2005; Dowell and Wicker 2009). In GLM_1min experiment, similar three-dimensional fields of additive noise are added in and near the regions where the FED is greater than 1 min^{-1} per pixel, vertically centered on the assumed altitude of GLM observations, which is 6.5 km according to previous studies (Allen et al. 2016; Kong et al. 2020). The horizontal and vertical length scales for smoothing perturbations are set to 9 and 3 km, respectively. An experiment (GLM_1min_NPT) identical to the GLM_1min experiment but without FED-based additive noise (Table 1) is conducted to assess the impact of additive noise. Additionally, considering the wide zero-FED regions when aggregating lightning in a short time period, an experiment (GLM_1min_NZR) using only nonzero GLM FED data is conducted. Another experiment (GLM_5min) using GLM FED temporally smoothed (i.e., averaged) in a 5-min window is conducted. Figure 2 shows examples of 1-min GLM FED and GLM FED averaged in the 5-min window prior to the analysis time. Generally, the smoothed FED has slightly lower maximum values but a larger nonzero FED region compared to the nonsmoothed FED. To further assess the potential value of assimilating GLM data with radar data assimilated, two more experiments are conducted, one with radar data assimilated into CTRL (RAD), and another experiment in which both radar and 5-min average accumulated GLM data with zero values are assimilated into CTRL (RAD+GLM). Experiments with radar data assimilated (RAD, RAD+GLM) used high-reflectivity regions to determine where additive noise is placed. The RAD+GLM experiment also used the FED-based additive noise (Table 1). The GLM FED and radar observations are both assimilated at 15-min intervals over the entire DA window. Although withholding radar data from CTRL and GLM-based experiments is not practical in operational WoFS, these experiments simulate a radar dropout scenario. Assimilating GLM observations alone provides insights on how much GLM FED could improve the forecasts compared to radar data and suggestions on the potential benefit of GLM LDA in radar sparse regions.

e. *Evaluation and verification methods*

To assess the accuracy of ensemble forecasts in each experiment and each case, a combination of subjective evaluation and objective verification methods is used in this study.

TABLE 1. Description of sensitivity experiments.

Expt	Data being assimilated	Additive noise
CTRL	Conventional observations, CWP	None
GLM_1min	As in CTRL, but additional 1-min GLM FED	FED-based additive noise
GLM_1min_NPT	As in GLM_1min	None
GLM_1min_NZR	As in CTRL, but additional 1-min nonzero GLM FED	FED-based additive noise
GLM_5min	As in CTRL, but additional 1-min GLM FED averaged in a 5-min window	FED-based additive noise
RAD	As in CTRL, but additional radar reflectivity and radial velocity data	Radar-based additive noise
RAD+GLM	As in RAD, but additional 1-min GLM FED as used in GLM_5min	Both FED-based and radar-based additive noise

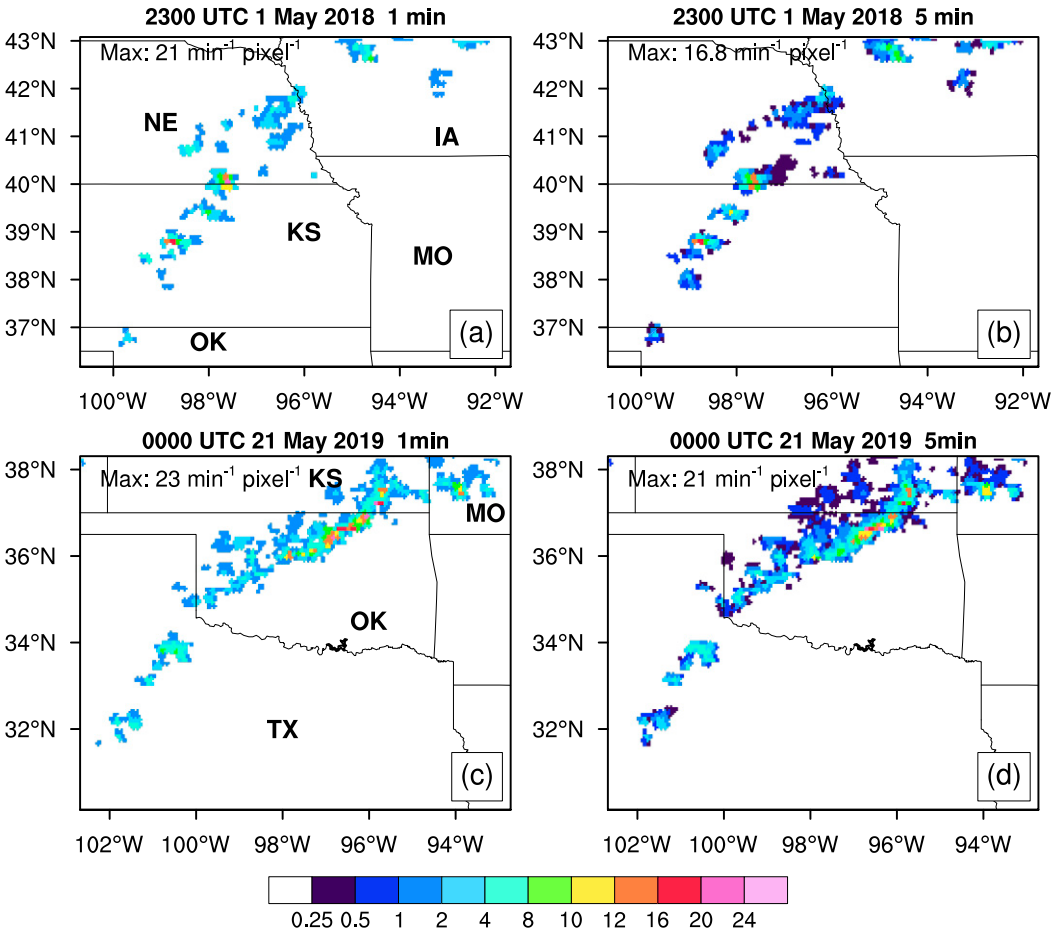

FIG. 2. Plan views of the1-min GLM FED (flashes per minute per pixel) (a),(c) without temporal smoothing and (b),(d) smoothed in a 5-min time window at (top) 2300 UTC 1 May 2018 and (bottom) 0000 UTC 21 May 2019. The FED are aggregated over the prior 1- or 5-min intervals from the listed times.

The observation-space diagnostic statistics (Dowell et al. 2004; Dowell and Wicker 2009; Dowell et al. 2011; Yussouf et al. 2013; Wheatley et al. 2015; Jones et al. 2018) of the mean innovation, the root-mean-square innovation (RMSI), total spread and consistency ratio (CR) for the assimilated FED are calculated to evaluate the performance of the system in assimilating GLM observations. Here, the difference between the observations and the model prior (forecast) or posterior (analysis) fields represent the innovation. The mean innovation and RMSI are the corresponding statistics between the observations and the model variables. Compared to the prior, a smaller absolute innovation and RMSI from the posterior is expected, indicating a successful DA cycle. A decreasing RMSI as a function of time is desirable, representing a decreasing forecast error. The total spread is the summation of the prior ensemble variance and the observation error variance, which measures the degree of the ensemble spread. A comparable magnitude of prior total spread and RMSI indicates that the forecast error is representative of the ensemble spread. The CR is defined as the ratio of the square of total spread to the variance of prior innovation (square of RMSI). A CR value of ~1.0 indicates that the ensemble variance is an optimal approximation for the forecast error variance for the assumed observation error. Details for these metrics can be found in the above-cited references.

To evaluate the reflectivity and the precipitation from the forecasts, the observed MRMS reflectivity and the Stage IV multisensor hourly rainfall estimates from the National Centers for Environmental Prediction (Baldwin and Mitchell 1997) are employed in this study. To compare with the model-simulated variables, the observations are all interpolated onto the WoFS domain with a 3-km grid spacing. Subjective comparisons between the observations and the model variables directly depict how the model reproduces the storm features (e.g., convection distribution and intensity). The neighborhood-based equitable threat score (ETS; Clark et al. 2010) and the performance diagram (Roebber 2009) depicting contingency elements including the probability of detection (POD), the false alarm ratio (FAR), the success ratio (SR = 1 − FAR), the frequency bias, and the critical success index (CSI), are used to quantitatively evaluate the short-term forecasts for each experiment and each case. These statistics

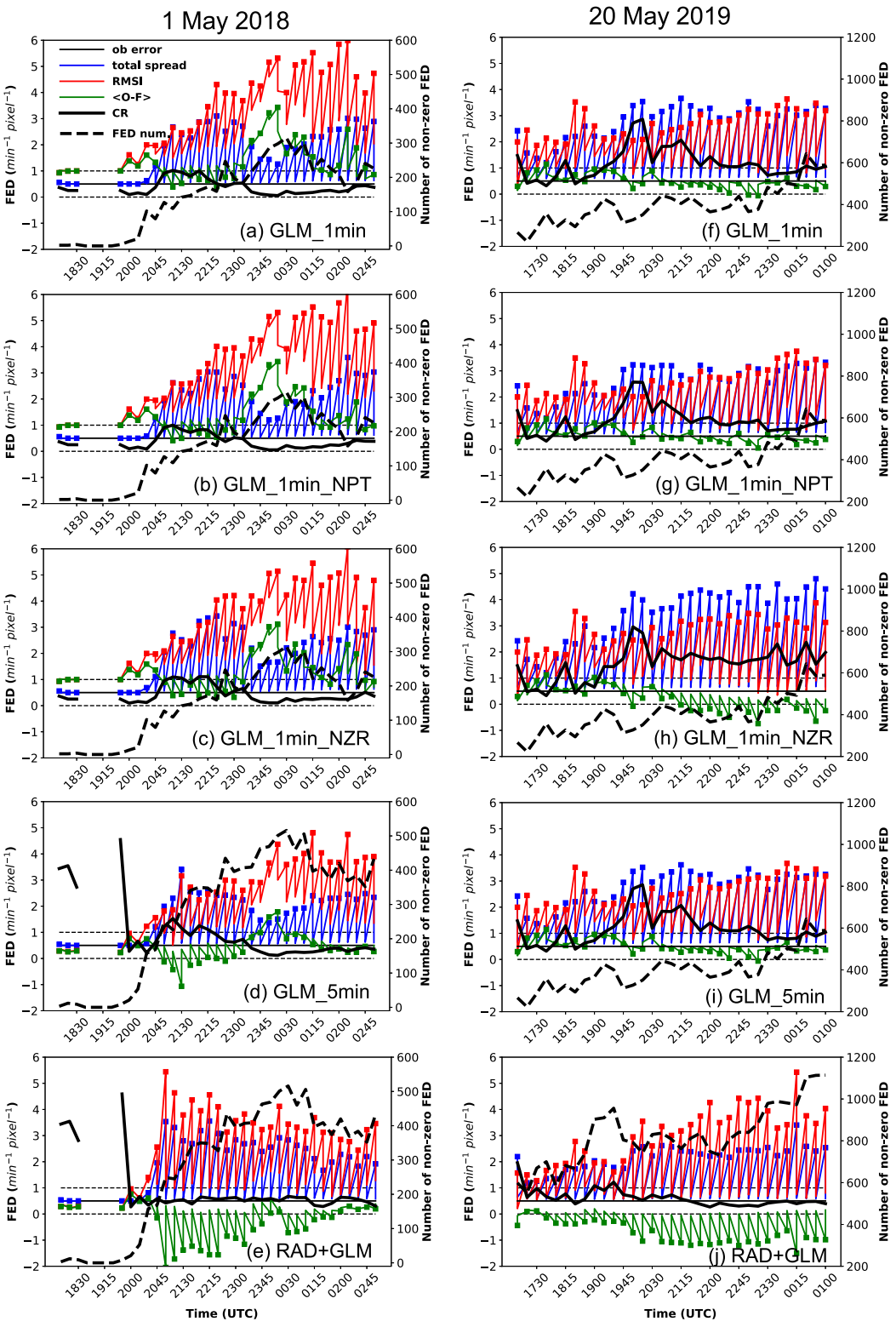

FIG. 3. Mean innovation (⟨O-F⟩), root-mean-square innovation (RMSI), observation error (ob error), total spread, consistency ratio (CR) of FED (left axis), and number of assimilated nonzero GLM FED observations (right axis) from all experiments listed in Table 1 for cases (a)–(e) 1 May 2018 and (f)–(j) 20 May 2019. Two thin dashed lines denote 0 and 1 values on the left axis. The FED statistics are computed only where the assimilated observed FED is >0 min^{-1} per pixel. The sawtooth patterns of ⟨O-F⟩, RMSI, and total spread are due to the plotted prior and posterior statistics. The square-marked end of the sawtooth lines denotes the prior values, and the other end denotes the posterior values.

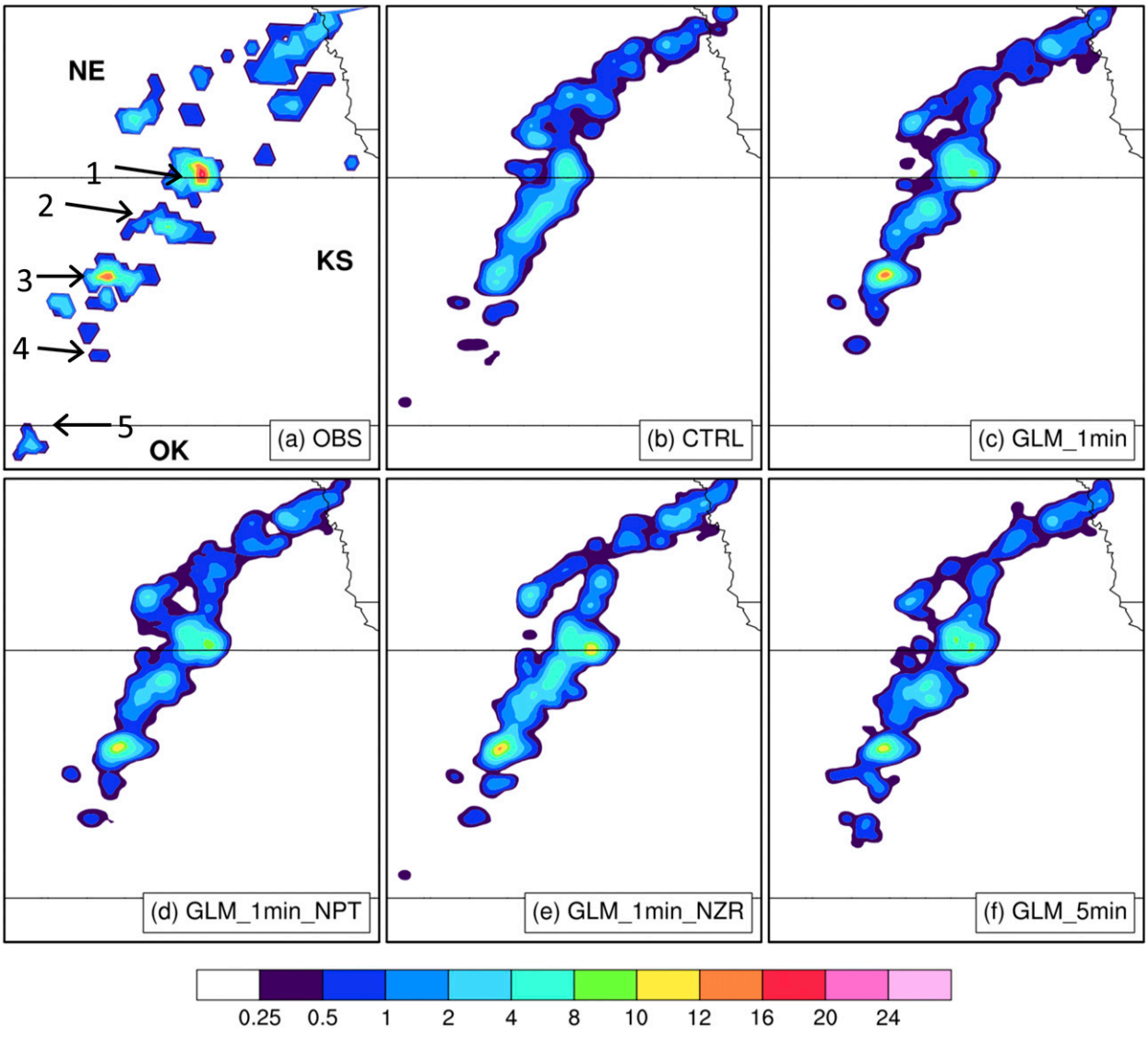

FIG. 4. Observed 1-min GLM FED (flashes per minute per pixel) and ensemble mean FED analyses at 2300 UTC 1 May 2018 from GLM-based experiments.

are calculated using all the ensemble members of each case grouped together.

3. Results and discussion

a. Sensitivity tests on GLM data assimilation

1) GLM DA DIAGNOSTIC STATISTICS

The observation-space diagnostic statistics of FED and the number of assimilated nonzero FED throughout the entire DA window for the two cases are presented in Fig. 3. In each experiment, the statistics are only calculated in areas where observed FED is greater than $0\,\text{min}^{-1}$ per pixel. Note that the posterior FED in Fig. 3 is not recalculated from the posterior state, but rather is the prior operator value that has been updated by the DA itself, and thus will have difference from a "true" posterior FED. For 1 May 2018, the RMSI for the experiments increases at ~2000 UTC (Figs. 3a–e) as convective cells develop and lightning occurs. The RMSI for all experiments is smaller than $6\,\text{min}^{-1}$ per pixel throughout the entire DA period and the posterior RMSI decreases toward $0\,\text{min}^{-1}$ per pixel. The prior mean innovation ranges from -1 to $3.5\,\text{min}^{-1}$ per pixel and the posterior mean innovation differs from it after each DA cycle. Due to the observed increasing flashes in isolated supercells, the underpredicted FED and the displacement error (not shown), the RMSI and mean innovation sharply increase at ~2330 UTC except experiment RAD+GLM, followed by the increase of the total spread and the ensemble spread. Correspondingly, the CR increases rapidly with convection initiation, then slowly decreases, and eventually remains lower than 0.5. The small CR and the low total spread are mainly due to the choice of a small

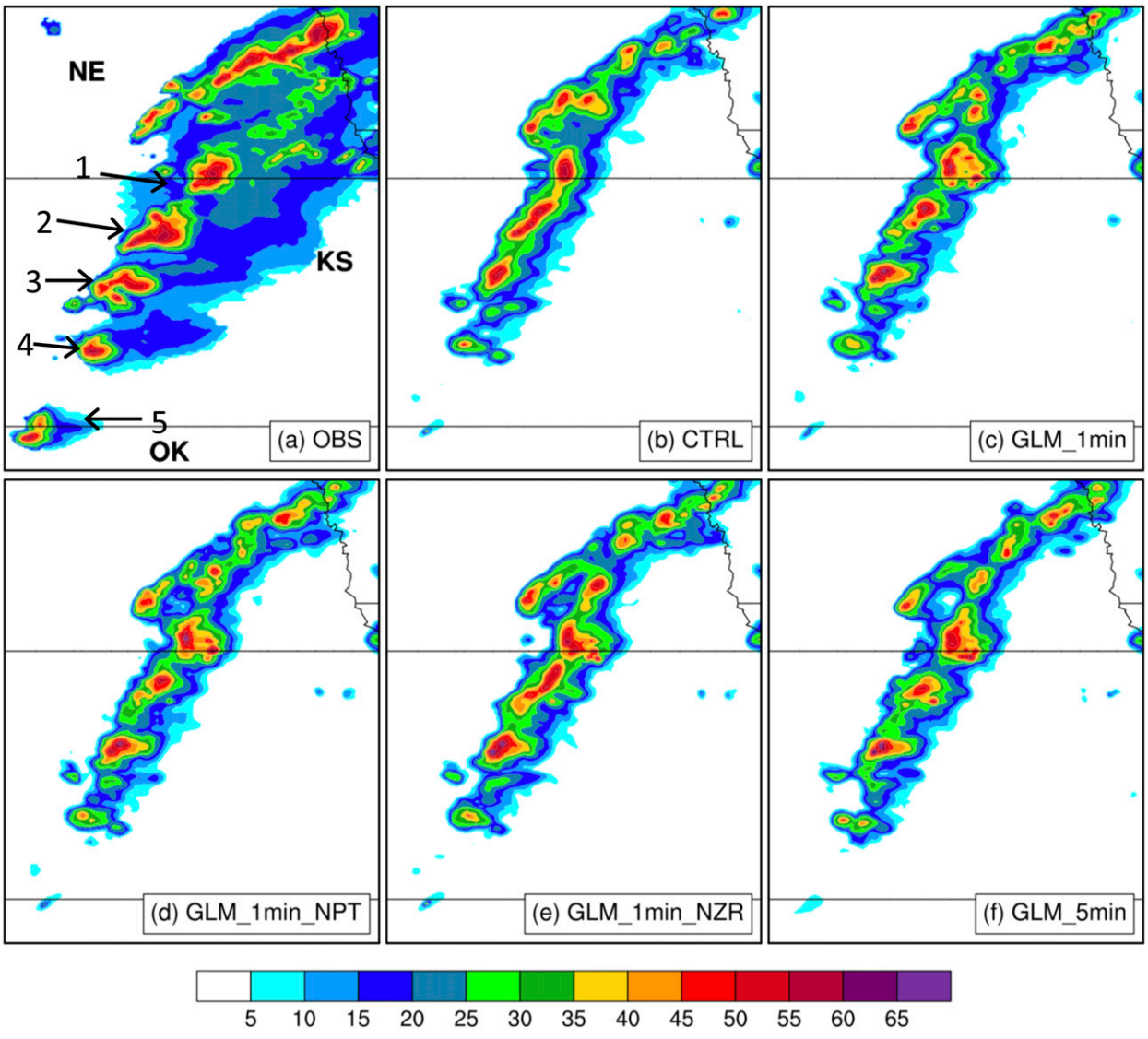

FIG. 5. Observed MRMS composite reflectivity (dBZ) and simulated probability matched mean composite reflectivity analyses at 2300 UTC 1 May 2018 from GLM-based experiments.

observation error, which increases the influence of low FED observations.

The CR in all the GLM experiments for case 20 May 2019 increases rapidly, then gradually decreases, and eventually remains around 1.0 or slightly greater (Figs. 3f–i). A CR of ~1.0 indicates that the ensemble variance is a good approximation of the forecast error variance for an assumed observation error (e.g., Dowell and Wicker 2009; Yussouf et al. 2013). That is, the observation error of 0.5 min^{-1} per pixel is a reasonable estimate for this case. The RMSI are all lower than 4 min^{-1} per pixel while GLM_5min shows smaller RMSI among all experiments for this case (Fig. 3i).

The statistics from GLM_1min and GLM_1min_NPT for the both events are very similar except for slightly larger total spread or ensemble spread from GLM_1min, which is expected with additive noise applied. Also, experiment GLM_5min for both events have larger CR and smaller RMSI compared to experiment GLM_1min. Smaller mean innovation or RMSI in GLM_5min is because in addition to FED greater than 1 min^{-1} per pixel, fractional FED observations ($0 < \text{FED} < 1\ \text{min}^{-1}$ per pixel) due to smoothing are also assimilated and included in calculating statistics. Overall, the saw-tooth pattern of innovation and RMSI indicate that assimilating GLM FED does reduce the model simulated FED error in all experiments. The statistics from RAD+GLM are discussed in section 3b.

2) ANALYSES OF FED AND COMPOSITE REFLECTIVITY

The performances of each experiment are evaluated by comparing with the observations and the CTRL experiment. Severe weather was relatively active on 1 May 2018 between 2300 and 0200 UTC, with a total of 12 tornadoes within this time period and a number of severe wind and large hail reported

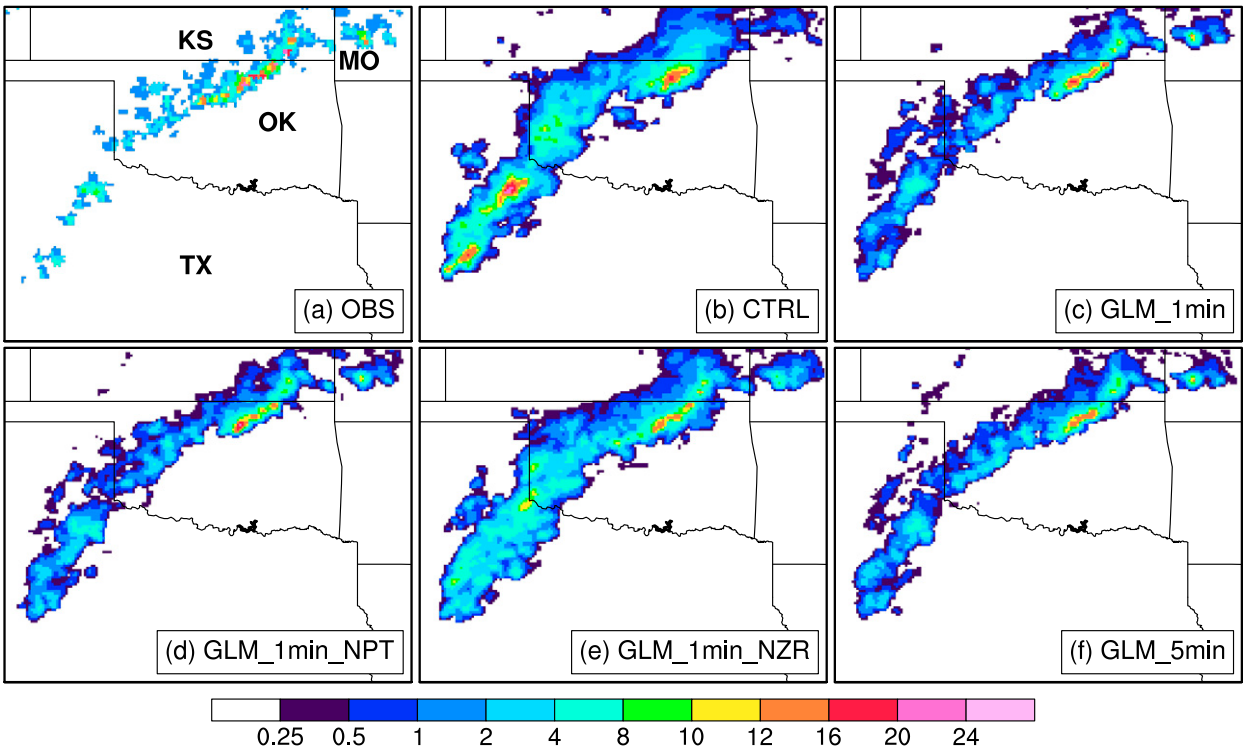

FIG. 6. As in Fig. 4, but at 0000 UTC 21 May 2019.

in four tornadic supercells (denoted 1, 2, 3, 5 in Fig. 5a) and one nontornadic supercell (denoted 4 in Fig. 5a) in Kansas, Nebraska and Oklahoma. As such, the analyses and the 3-h forecast initialized at 2300 UTC are evaluated. The FED analyses from all GLM-based experiments are shown in Fig. 4. The five FED cells corresponding to five supercells in Fig. 5a are denoted as 1–5. All five sensitivity experiments are able to produce flashes in most of the supercells except supercell 5, but the FED values in all the experiments are weaker than the observations (Figs. 4b–f). Compared to GLM_1min_NZR, all the other GLM-based experiments produce more isolated supercells similar to the observed ones, indicating that zero FED helps to better resolve individual supercells. Experiments GLM_1min_NPT, GLM_1min, and GLM_5min show very similar FED features.

The model-simulated probability matched mean (PMM) composite reflectivity (CREF) at the analysis time (2300 UTC) for case 1 May 2018 are shown in Fig. 5. The probability matching assumes that the most likely spatial representation is given by the ensemble mean and restores the amplitude characteristics of the full ensemble to the ensemble mean (Ebert 2001). Specifically, the values from ensemble mean and full ensemble are sorted from highest to lowest. Then the ensemble mean field is substituted by the values from the full ensemble, which are thinned by the ensemble size. Similar to the FED, the five experiments generally reproduce the four supercells (denoted 1, 2, 3 4) but miss the supercell 5. They overall have very similar reflectivity fields. Compared to CTRL and GLM_1min_NZR, experiments GLM_1min, GLM_1min_NPT and GLM_5min reproduce better convection mode, and the shapes and the isolation features of the tornadic convective cells (denoted 1, 2, 3) are more like the observation.

For the case 20 May 2019, FED at 0000 UTC 21 May 2019 shows a quasi-linear mode of deep electrified convection that extends from northern Texas into southeastern Kansas and eastern Missouri through Oklahoma (Fig. 6a). Compared to the observation, CTRL produces a broader intense lightning band, with slightly lower FED near the border between Oklahoma and Kansas but much higher FED in Texas. Assimilating both zero and nonzero GLM data significantly narrows the lightning band and reduces the FED in Texas (Figs. 6c,d,f). In GLM_1min_NZR (Fig. 6e), the maximum lightning intensity in Texas decreases but the lightning band is still broad.

The MRMS CREF at 0000 UTC 21 May 2019 shows that the MCS consists of a quasi-linear structure of severe convection and a broad trailing stratiform region (Fig. 7a). The lightning flashes were mainly within the convective region (Figs. 6a and 7a). Generally, CTRL overpredicts convection in intensity and areal coverage. Assimilating GLM data including zero values yields more reasonable convection intensity in northern Oklahoma and Kansas (Figs. 7c,d,f). The convective cells in Texas and southwestern Oklahoma, however, are substantially suppressed. The stratiform region to the north is already poorly simulated in CTRL, and is further reduced by assimilation of zero FED. With only nonzero GLM FED assimilated, GLM_1min_NZR produces a broader stratiform rainband near the border between Oklahoma and Kansas compared

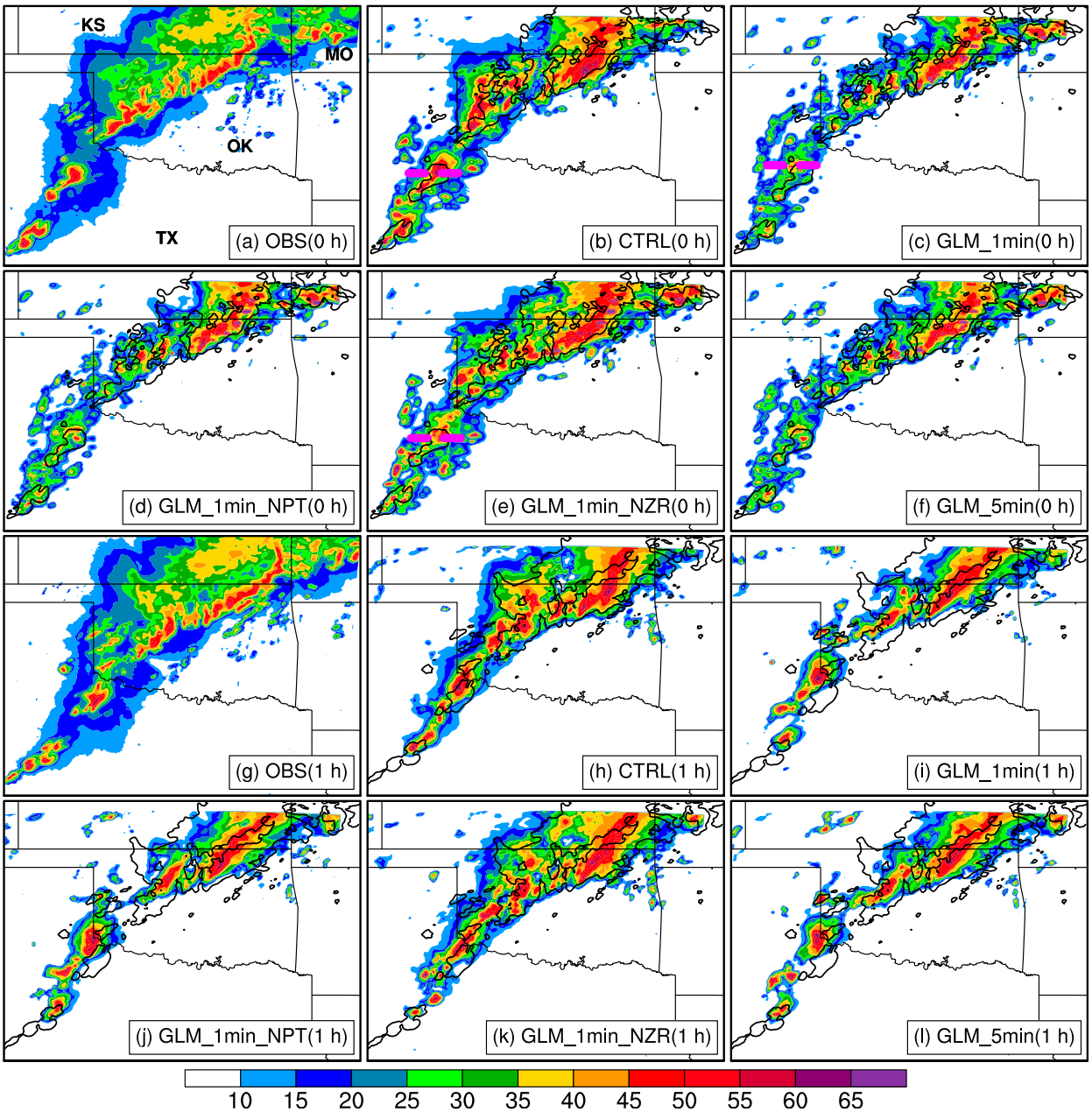

FIG. 7. Observed MRMS composite reflectivity (dBZ) and simulated probability matched mean composite reflectivity from analyses and 1-h forecasts initialized at 0000 UTC 21 May 2019 from GLM-based experiments. Black contours represent the 30-dBZ value from MRMS.

to GLM_1min and GLM_5min (Figs. 7b,c,e,f). Given that GLM FED values are nearly 0 flashes per minute in the trailing stratiform region (Figs. 2c,d), the stratiform rainband in GLM_1min and GLM_5min are likely to be weakened by assimilating zero FED values. Also, GLM_1min_NZR outperforms GLM_1min or GLM_5min by not suppressing the convection in southwestern Oklahoma and northern Texas as much. Nevertheless, the convective cells in northern Texas in GLM_1min_NZR still are neither as well organized, nor as intense as those in observation and CTRL. The PMM CREF from the 1-h forecasts (Figs. 7h–l) show that the GLM-based experiments, especially GLM_1min and GLM_5min, perform worse in terms of larger convection displacement errors in southwestern Oklahoma and Texas. Experiment GLM_1min_NZR performs better in reproducing the stratiform region of the MCS than GLM_1min and GLM_5min. However, GLM_1min_NZR is still slightly worse than CTRL by displacing the convection in Texas too far north.

The comparison among these sensitivity experiments suggests that assimilating zero GLM FED over stratiform region is

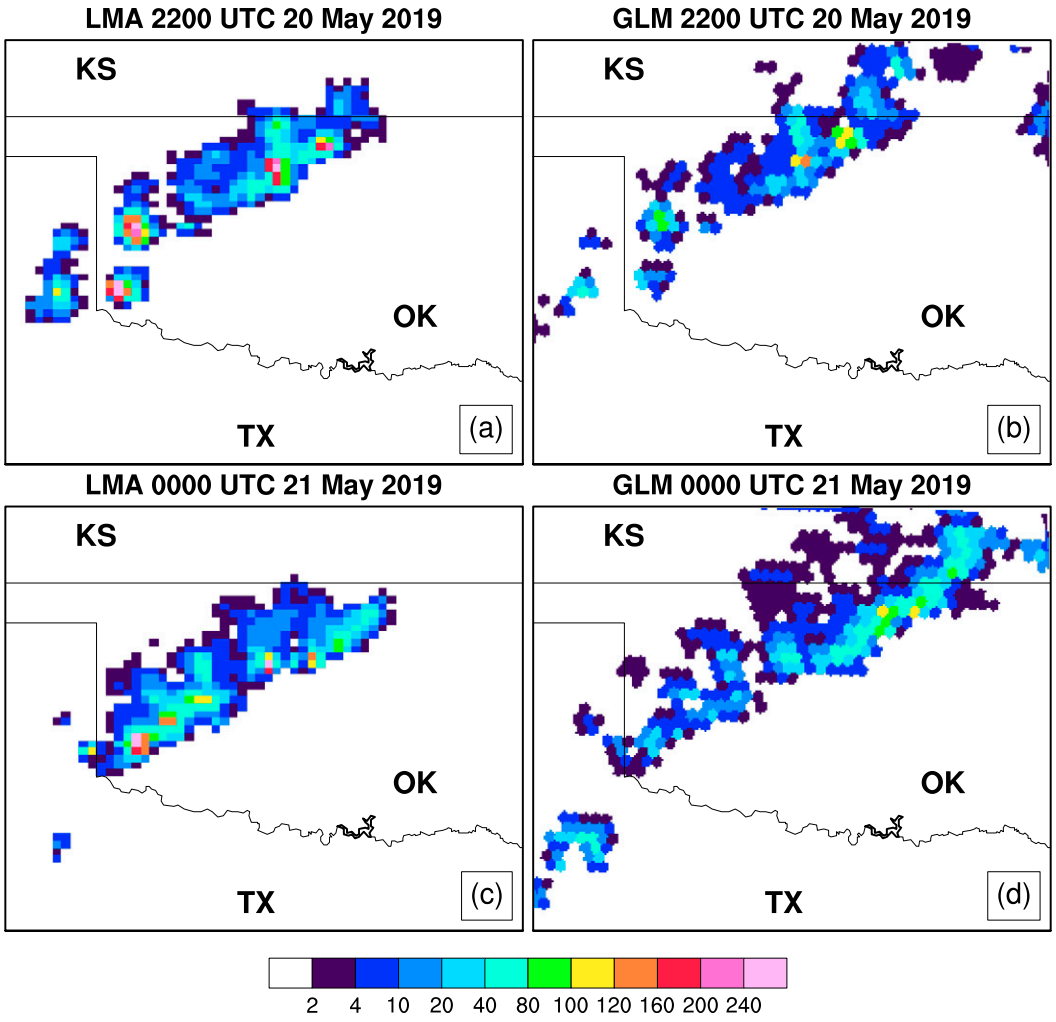

FIG. 8. Horizontal cross sections illustrating the 5-min accumulated (a),(c) Oklahoma LMA FED (flashes per 5 minutes per pixel) and (b),(d) GLM FED on a ~10 × 10 km^2 grid box at (top) 2200 UTC 20 May 2019 and (bottom) 0000 UTC 21 May 2019. LMA FED is generated using "lmatools" (Fuchs et al. 2016) with at least 10 points per flash.

detrimental for this MCS case. However, assimilating nonzero GLM FED tends to degrade the analyses in the southwestern part of the domain. Note that the observed reflectivity intensities of the convective cells in Texas and southwestern Oklahoma are comparable to those in northern Oklahoma (Fig. 7a), but the GLM flash rates in southwestern domain are much lower (Fig. 6a). One possibility is that flashes from those storms are less easily detected by GLM. Further comparing GLM data with Oklahoma Lightning Mapping Array (OKLMA) data (for information on the network operation, accuracy and the covered domain, see Thomas et al. 2004; MacGorman et al. 2008; Chmielewski and Bruning 2016; Chmielewski et al. 2020) shows that the GLM FED is much lower than the OKLMA FED in southwestern and central Oklahoma (Fig. 8), indicating lower GLM DE in these regions. For instance, in the tornadic storm near the southwestern corner of Oklahoma around 2200 UTC 20 May 2019, the maximum of the 5-min accumulated OKLMA FED was above 400 per pixel (Fig. 8a) while that of GLM FED is less than 40 per pixel (Fig. 8b). The maximum OKLMA FED from only the flashes with durations of at least 0.3 s and areas of at least 10 km^2, which would be expected to have GLM DE > 50% (Zhang and Cummins 2020), still reached over 120 per pixel in the southwestern storm (not shown). It is unclear why the GLM flash rate was so much lower in this event. The convective cells in Texas had slightly warmer cloud tops than those in northern Oklahoma during the DA window according to the *GOES-16* infrared images (not shown). Some recent studies have shown that GLM DE could be particularly low in storms with large cloud water and cloud ice content, and compact flashes at lower altitudes (Rutledge et al. 2020), with very high midaltitude reflectivity (Murphy and Said 2020), and with small, short duration flashes (Zhang and Cummins 2020). DE is also sensitive to the vertical cloud inhomogeneity (Brunner and Bitzer 2020) and the location of small parent flashes relative to the pixels (Zhang et al. 2019). Studies comparing GLM

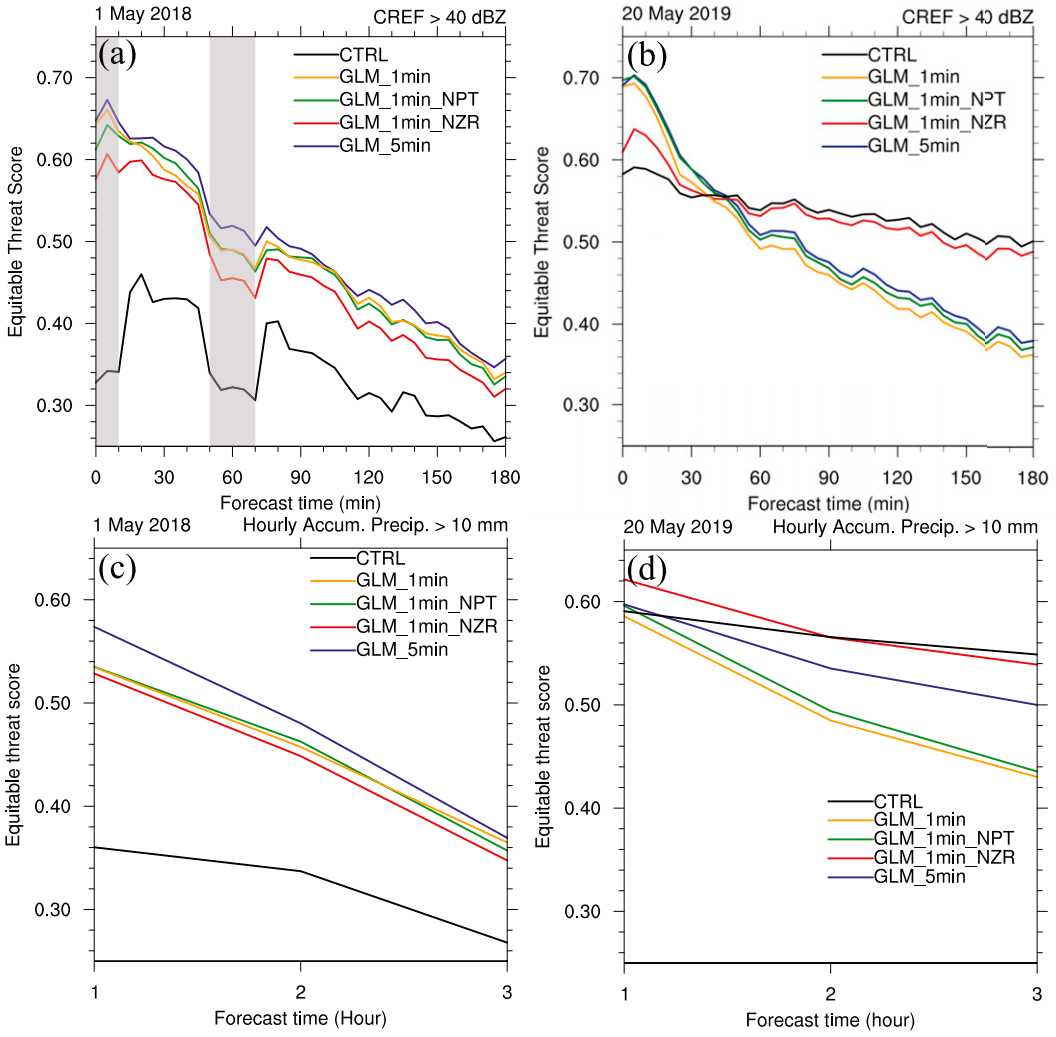

FIG. 9. Aggregate equitable threat score (ETS) of (a),(b) composite reflectivity greater than 40 dBZ relative to the MRMS observations, and (c),(d) hourly accumulated precipitation greater than 10 mm relative to the Stage IV multisensor precipitation analyses over seven 3-h forecast periods of each GLM-based experiment for (left) 1 May 2018 and (right) 20 May 2019. The neighborhood radius is 15 km. The shaded areas denote the ETS affected by missing MRMS observations.

to ground-based lightning networks have also found lower DEs in western Great Plains (Marchand et al. 2019; Rutledge et al. 2020). The FED forward operator function may also overestimate the flash rates for this convection, which would tend cause erroneous weakening.

3) SHORT-TERM FORECAST EVALUATION

To quantitatively evaluate the performance of each experiment, the aggregated neighborhood-based ETS are computed using a neighborhood radius of 15 km for the CREF using a threshold of 40 dBZ and for the hourly accumulated precipitation using a threshold of 10 mm from the 3-h forecasts initialized at the top of the hour for each of the experiments for the two events (Fig. 9). Different thresholds (e.g., 35 and 45 dBZ for CREF; 5 mm for hourly precipitation) yield similar results. For the 1 May 2018 event, experiment CTRL produces forecasts with a relatively low ETS time series for CREF, generally below 0.5 throughout the 3-h forecast period (Fig. 9a). The sharp increase or decrease of ETS shown during the forecast period is due to the missing MRMS CREF data between 2150 and 2210 UTC, which affects evaluation of the two forecasts at 2100 and 2200 UTC at varying lead times. Overall, assimilating GLM FED data greatly improves the forecast skill over the entire 3-h forecast compared to CTRL. The experiment GLM_5min outperforms other GLM-based experiments, by improving the ETS about 0.15 relative to CTRL. The GLM_1min and GLM_1min_NPT have comparable ETS with GLM_1min presenting a slightly better analyses than GLM_1min_NPT. GLM_1min_NZR also has lower ETS compared to GLM_1min, indicating the positive impact of zero FED. The performance of the precipitation forecasts from all experiments are consistent with those of CREF (Fig. 9c).

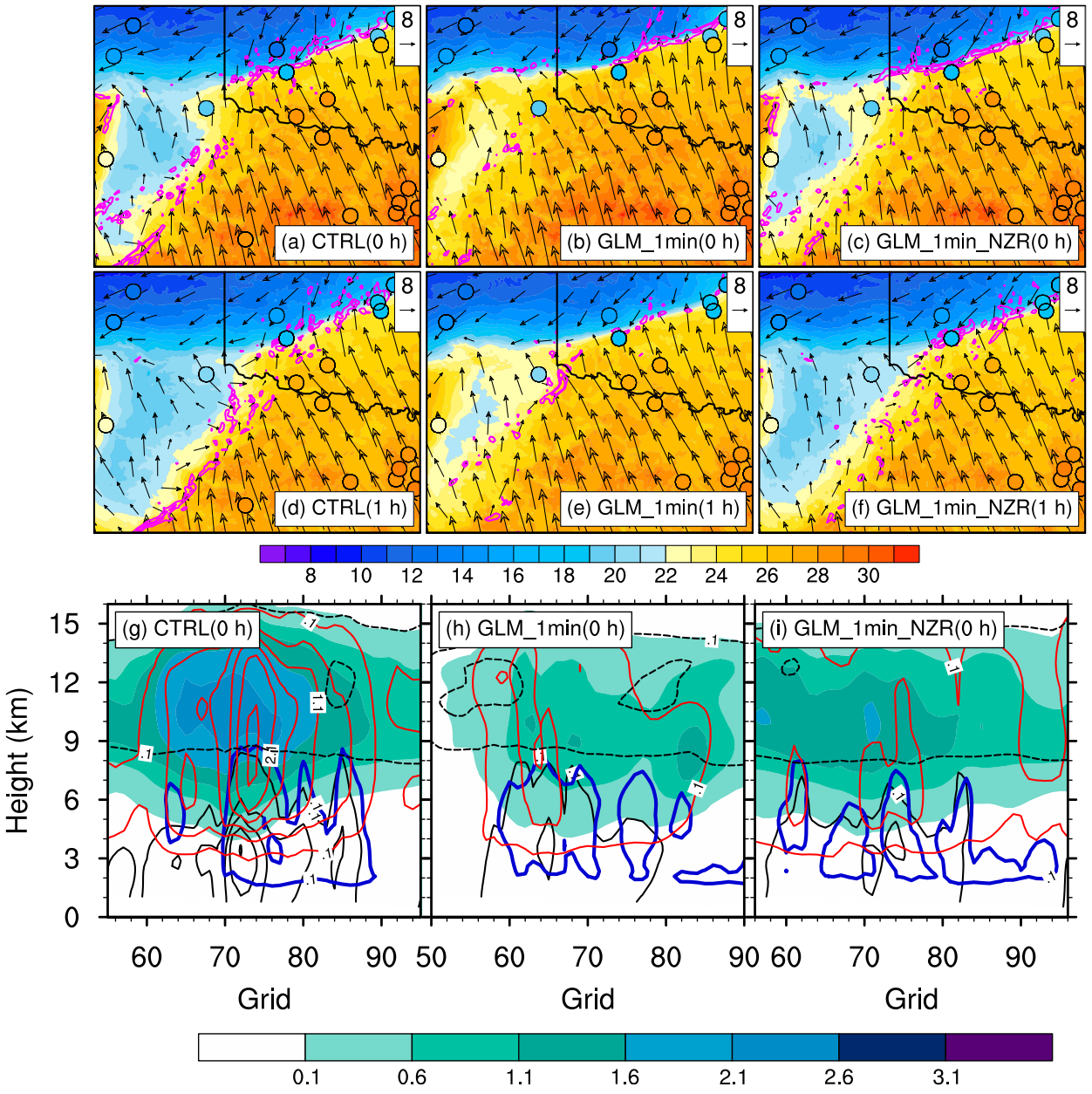

FIG. 10. (a)–(f) The ensemble mean 2-m temperature (colors), 10-m wind vectors (m s^{-1}), and 500-m vertical velocity (magenta contours at 0.6 and 1 m s^{-1}) from experiments CTRL, GLM_1min, and GLM_1min_NZR from analyses and 1-h forecast initialized at 0000 UTC 21 May 2019. The color-filled dots are the corresponding 2-m temperature from Automated Surface Observing System (ASOS); (g)–(i) vertical cross sections of ensemble mean mixing ratio (10^{-3} g g^{-1}) of cloud water (q_c; blue contours), rainwater (q_r; black solid contours), snow (q_s; color shaded), graupel (q_g; red contours), and ice (q_i; black dashed contours) from the analyses at 0000 UTC 21 May 2019 along the magenta dashed lines in Figs. 7b, 7c, and 7e. The x axis gives the indices of zonal grid points.

For the 20 May 2019 event, all GLM-based experiments improve the analyses and the 0–40-min forecast of CREF (Fig. 9b). After 40 min, however, ETS for all GLM experiments except GLM_1min_NZR decreases and eventually becomes ~0.12 lower than CTRL. GLM_1min_NZR shows less improvement to the performance of the first 40-min forecast, but it significantly outperforms GLM_1min and only is slightly worse than CTRL in the following 2.5-h forecast. This result suggests that, despite giving a better ETS in the analyses, assimilating zero FED value is the dominant source of the forecast error in GLM_1min for this case, which is consistent with the subjective analyses in Fig. 7, which shows correct suppression of some spurious cells, but also incorrect weakening of some real storms. GLM_5min appears to slightly outperform GLM_1min, probably owing to GLM_5min having fewer zero-FED values after smoothing. Also, GLM_1min_NPT performs

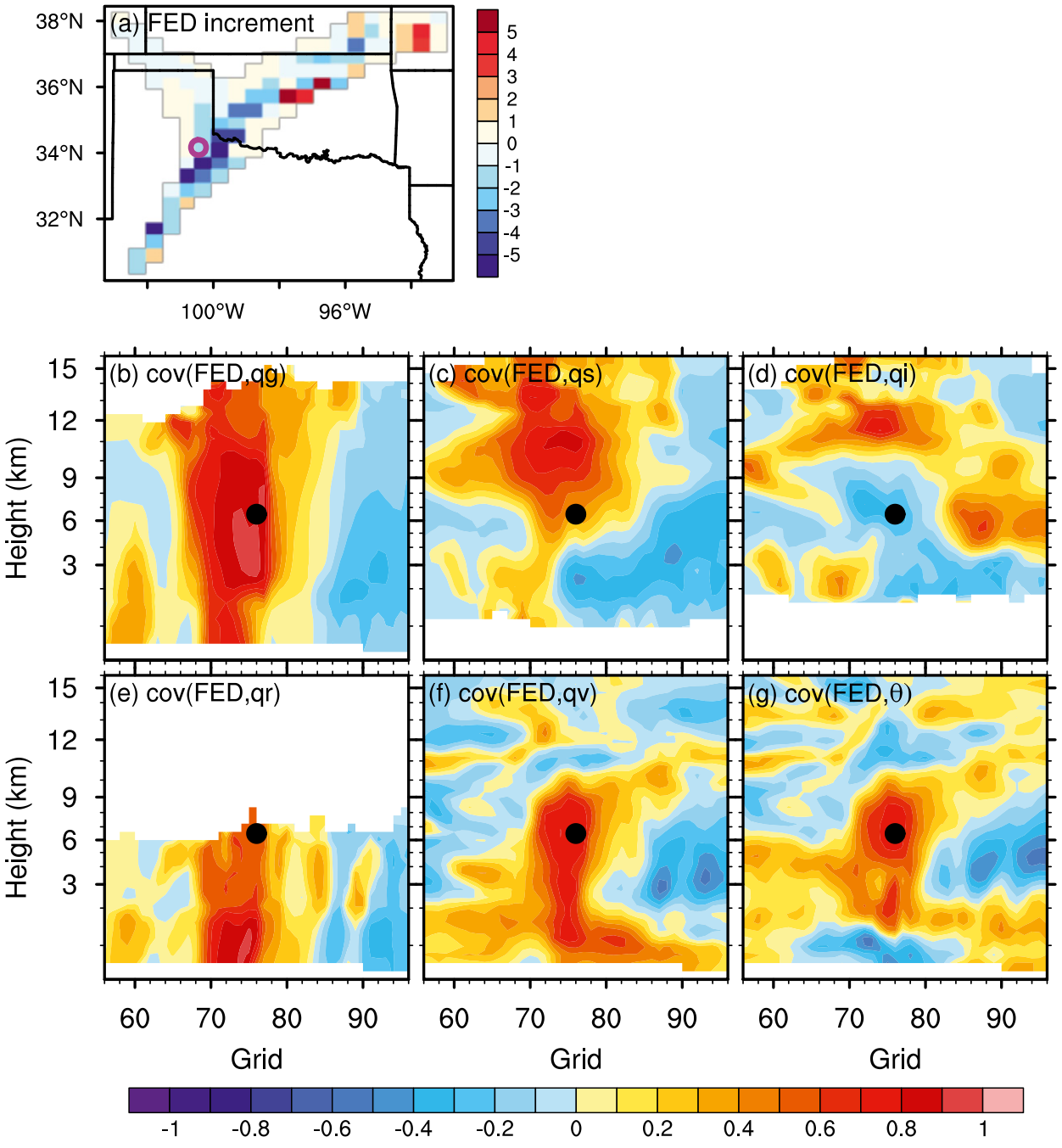

FIG. 11. (a) The increment of FED (posterior − prior; flashes per minute per pixel) after the DA at 0000 UTC 21 May 2019 from GLM_1min_NZR; (b)–(g) background error correlations between a simulated FED located at the magenta circle location and model state variables q_g, q_s, q_i, q_r, q_v, and potential temperature θ. The x axis in (b)–(g) gives the indices of zonal grid points.

slightly better than GLM_1min. This is likely because GLM_1min_NPT has a smaller ensemble spread than GLM_1min (Figs. 3f,g), and thus gives more weight to the prior forecasts over the observation, which in turn mitigates the negative impact caused by GLM observations.

The precipitation forecasts for 20 May 2019 event also are quite consistent with the CREF forecasts (Fig. 9d). Overall, GLM_1min_NZR appears to outperform other experiments by notably improving the forecast of the first-hour precipitation. Another experiment assimilating nonzero values from the 1-min accumulated GLM FED data and zero values from 10-min accumulated GLM FED data into CTRL was conducted for this case such that fewer zero FED observations are used over the stratiform or nonintense convective region. Results show

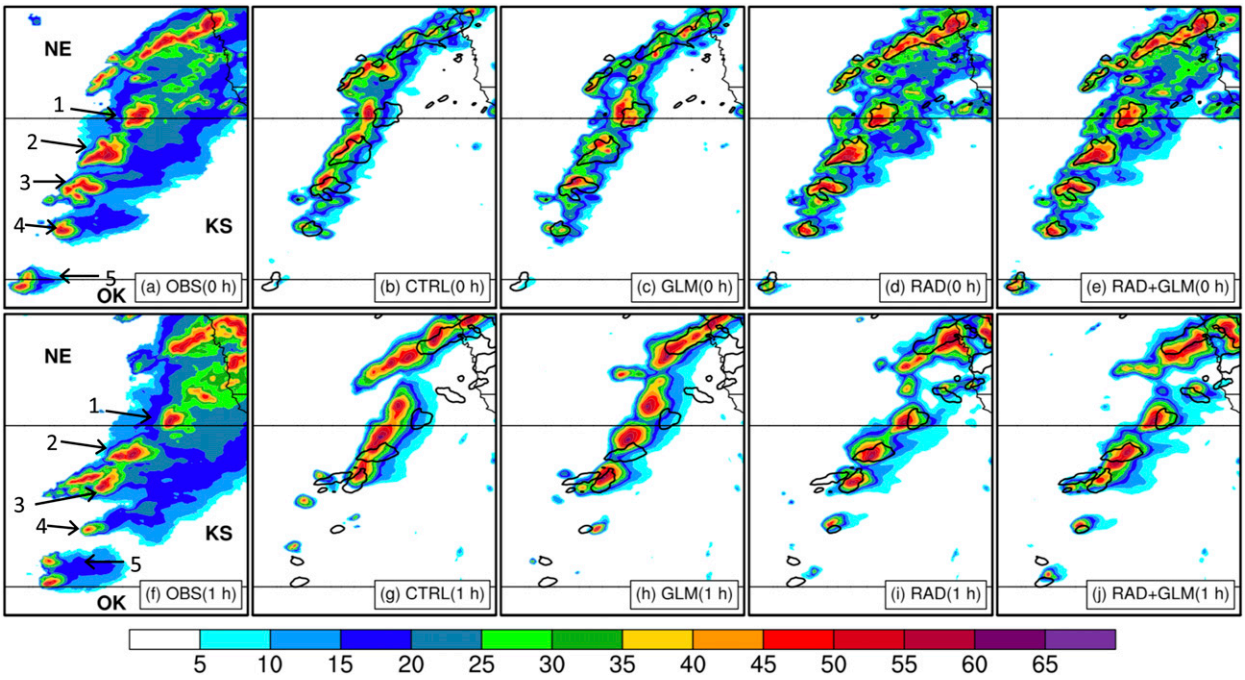

FIG. 12. Observed MRMS composite reflectivity (dBZ) and simulated probability matched mean composite reflectivity from analyses and 1-h forecasts initiated at 2300 UTC 1 May 2018 from experiments CTRL, GLM, RAD, and RAD+GLM. Black contours represent the 35-dBZ value from MRMS.

that this GLM experiment still performs better than CTRL in the first 40 min and then performs worse (not shown).

Figures 10a–f shows that GLM-related experiments, especially GLM_1min, have larger warm bias near surface in Texas, compared to the observation and that from CTRL, though all the experiments are able to reproduce the cold front across Oklahoma. The poorer forecasts in the southwestern domain from the GLM-based experiments are likely due to the weakened cold pools. In CTRL, the eastward outflow at the edge of the strong cold pool in Texas interacts with the southeast flow, stimulating strong updrafts along the cold pool east edge (Figs. 10a,d). However, in GLM_1min and GLM_1min_NZR, the weaker cold pool-driven outflows are mainly northward, resulting in fewer and weaker low-level updrafts in Texas (Figs. 10b,c,e,f). GLM_1min_NZR has a better cold pool because it allows spurious but close storms to persist while the real storms are kept weak (incorrectly) by low GLM FED.

The aforementioned zero- and low-FED observations likely result in negative FED increment after DA (difference between posterior and prior FED operator values; Fig. 11a). Via the overall positive background error correlations between FED and hydrometeor mixing ratios (e.g., q_g, q_s, q_i, q_r, q_v, Figs. 11b–f), hydrometeor contents in GLM_1min and GLM_1min_NZR decrease (Figs. 10g–i) as a result of the negative FED increment, which was also discussed by Kong et al. (2020). The weaker cold pools are partly owing to the associated weaker evaporation, melting, and sublimation of precipitation. Negative FED increments could also warm the near surface air via a negative correlation between FED and low-level potential temperature (Fig. 11g), which could also suppress cold pools. Significant warmer cold pools in GLM_1min are likely due to many more zero-FED values being assimilated and suppressing neighboring (yet spurious) storms.

In summary, assimilating total GLM data (a combination of zero and nonzero values) could better capture the features of the individual supercells and substantially improve the forecast skill. However, for the MCS case consisting of deep electrified convection, nonintense convection and trailing stratiform regions, assimilating GLM data without radar data appears to harm the forecast despite the improvement in the analyzed FED and CREF field. Assimilating only nonzero GLM FED values mitigates the negative impact (partially via compensating spurious cells) but does not eliminate it owing to unexpected low GLM FED values, which appears to be the root source of degradation. The current observation operator, which is a linear fit using model output variables, does not fit for the GLM observations at some regions of the MCS, especially when the GLM data suffer from the detection efficiency issue. Lower GLM DE, particularly in severe storms with lots of small, intracloud flashes at low altitudes over the western great plain, should be considered for a more advanced observation operator.

b. Data assimilation of GLM data with radar data

Based on the results in section 3a, experiment GLM_5min is selected to compare against additional two experiments: one that assimilates radar data (RAD), and another experiment that uses the same setup as in GLM_5min to assimilate both GLM data and radar data (RAD+GLM). Experiment GLM_5min is renamed to ''GLM'' for convenience in

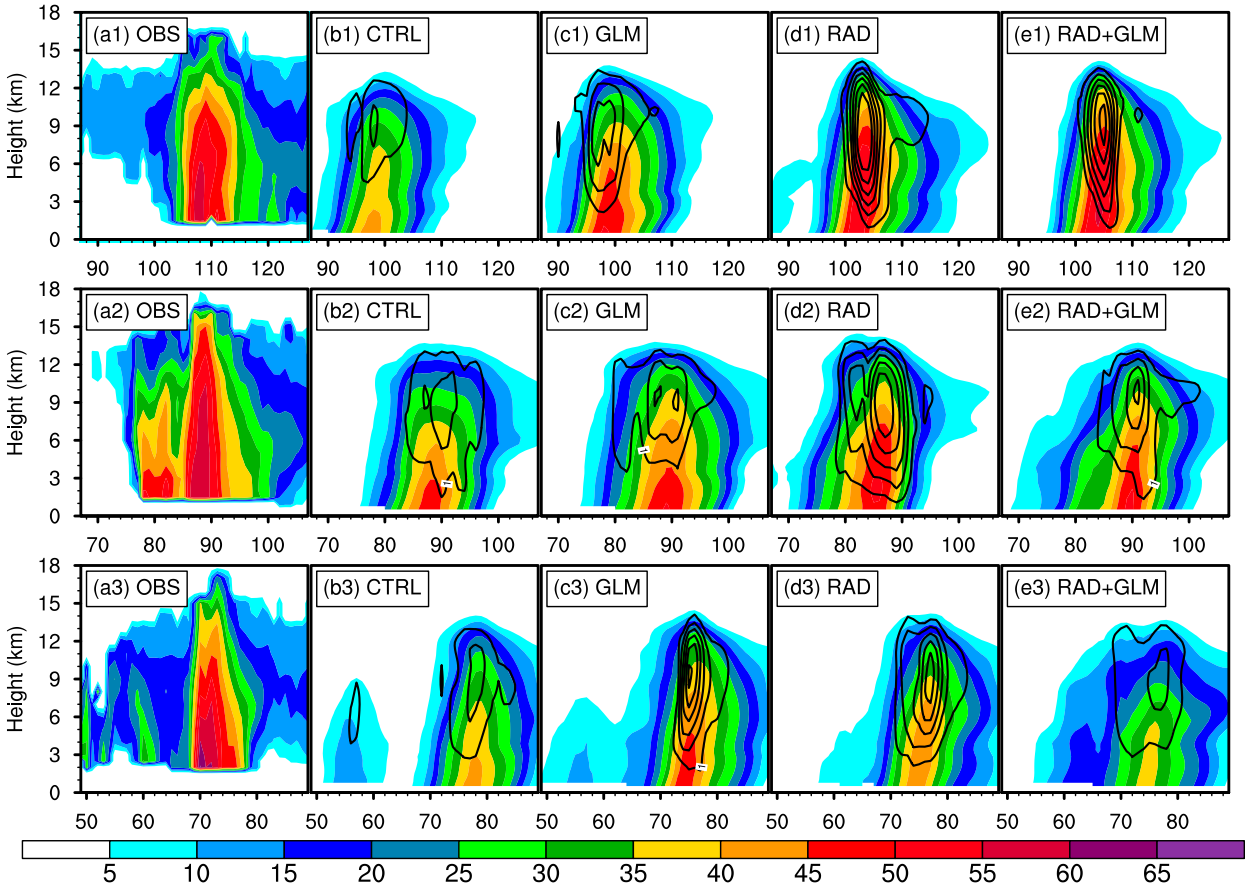

FIG. 13. Vertical cross sections of the ensemble mean reflectivity (shaded; dBZ) and vertical velocity (contours; $m\,s^{-1}$) of (a1)–(e1) convection 1, (a2)–(e2) convection 2, and (a3)–(e3) convection 3 from 1-h forecasts in Figs. 10f–j along the west–east direction through the maximum reflectivity center. The *x* axis gives the indices of zonal grid points.

this section. The forecast skill in predicting CREF, precipitation, and storm tracks will be evaluated for the two selected cases.

1) CASE 1 MAY 2018

Figure 3e shows the diagnostic statistics of FED for RAD+GLM. The magnitudes of RMSI and total spread match each other better than those in experiments with no radar DA, resulting in more constant CR. The faster growth in RMSI from RAD+GLM than GLM indicates faster storm spinup with radar DA. In general, RAD and RAD+GLM better capture the five main convective cells across Kansas compared to CTRL and GLM (Figs. 12a–e). RAD and RAD+GLM perform very similarly except that RAD+GLM has a slightly more intense convective cell at the Kansas–Oklahoma border, which is more consistent with the observation. In the 1-h forecasts, all experiments present some larger reflectivity cores than the analyses (Figs. 12f–j). GLM performs slightly better than CTRL with somewhat weaker convection on the east border of Nebraska (Figs. 12f–h). The two experiments with radar DA still outperform CTRL and GLM with less displacement error. RAD+GLM performs best, and is still able to forecast the southernmost tornadic convective cell, although it is weaker than the observation (Fig. 12j).

To further examine the impact of radar and GLM DA, the vertical cross sections of the ensemble mean reflectivity of the three tornadic convective cells (denoted as 1, 2, and 3 in Fig. 12f) from CTRL, GLM, RAD, and RAD+GLM are shown in Fig. 13. The west–east cross sections are through the maximum reflectivity center in each experiment. Overall, all the experiments with radar or lightning data assimilated (GLM, RAD, RAD+GLM) enhance the convection columns similar to the observation with notably stronger reflectivity and updraft than those in CTRL. Convection cores in experiments with radar data assimilated (RAD, RAD+GLM) are generally stronger than those in GLM (Figs. 13c1–e1,c2–e2) except the third convective cell (Figs. 13c3–e3). The low ensemble-mean reflectivity and the high PMM CREF (Figs. 12g–j) associated with convection 3 indicates a large amplification between ensemble members or a large ensemble spread in predicting this cell. Experiment GLM performs best in reproducing the third storm cell with strongest updraft and reflectivity (Fig. 13c3). Assimilating GLM FED with radar data creates slightly weaker convective cells (Figs. 13e1,e2,e3). Among all the experiments, the displacement error between the simulated storms in experiment RAD+GLM and observations, at least along the zonal direction, are generally the smallest.

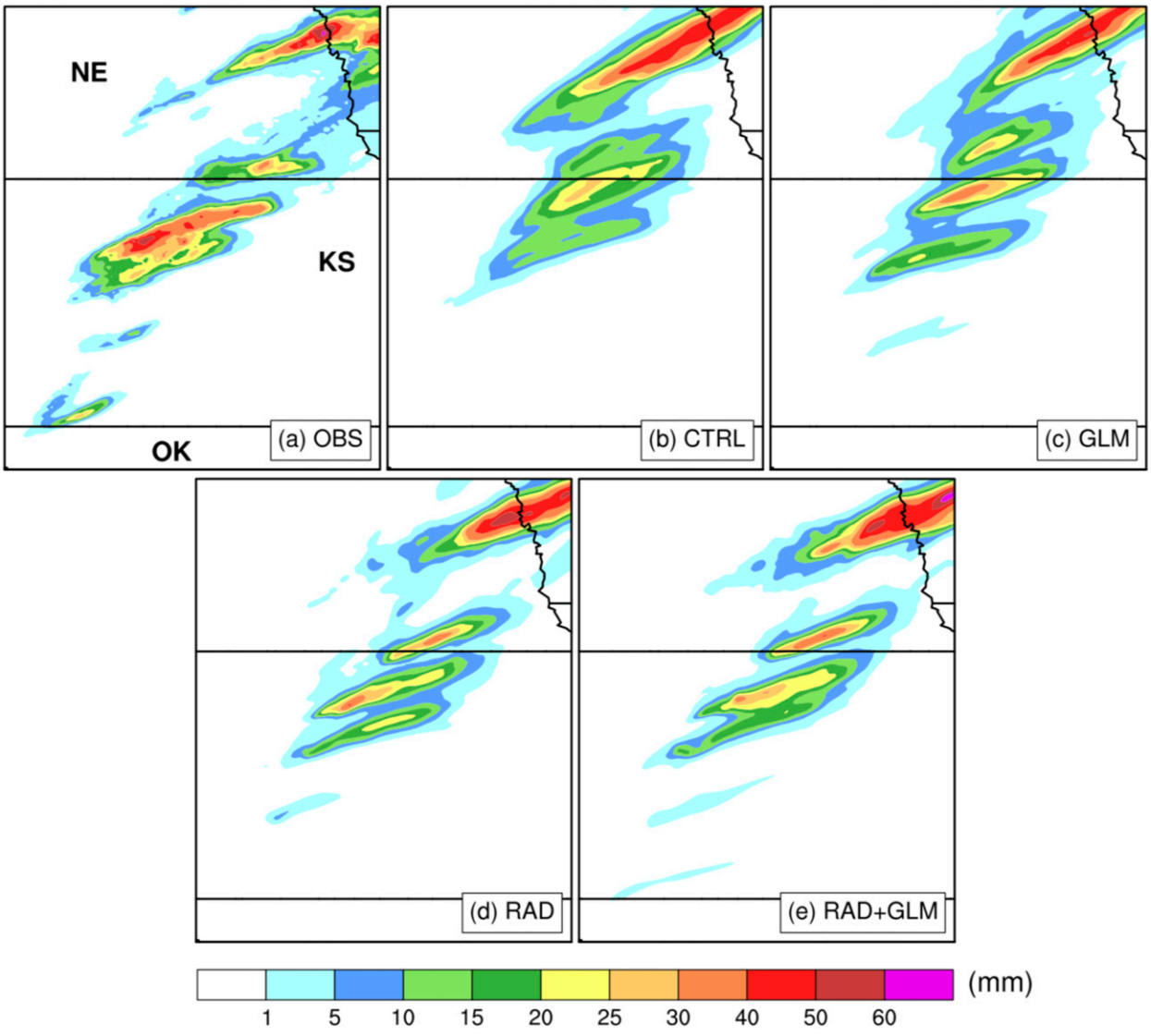

FIG. 14. The 3-h accumulated precipitation (mm) from Stage IV multisensor precipitation products and from forecasts beginning at 2300 UTC 1 May 2018 for experiments CTRL, GLM, RAD, and RAD+GLM.

The 3-h accumulated precipitation from the forecast initiated at 2300 UTC is compared with the Stage IV precipitation products in Fig. 14. Generally, all the DA experiments overpredict the coverage of the heavy precipitation region crossing the east border of Nebraska while underpredicting the precipitation amount in northern Kansas. Assimilating radar data suppresses the northern false precipitation region exhibited in CTRL and reduces the northern displacement of the heavy precipitation center in Kansas (Figs. 14a,b,d,e). Assimilating GLM FED into CTRL produces a slightly narrower rainband over the east border of Nebraska and larger rainfall amount in the south, which are more consistent with the observation (Figs. 14a–c).

There were 12 tornado events as well as a number of severe wind and large hail reports during 2300 UTC 1 May to 0200 UTC 2 May 2018. Figure 15 shows the probability swaths of the model-simulated 2–5-km updraft helicity (UH) greater than $60\,m^2\,s^{-2}$ in the 3-h forecast beginning at 2300 UTC for four experiments, overlaid by the severe weather reports from NWS during that period. This UH value is similar to the intensity thresholds used in previous studies (Skinner et al. 2018; Jones et al. 2018). With conventional observations and CWP data assimilated, CTRL is able to predict storm tracks overlapping with most of the severe weather events (Fig. 15a). With additional GLM data assimilated (GLM), three high-probability swaths, associated with convective cells 1, 2, and 3 become clearer with probability exceeding 90% (Fig. 15b). However, both CTRL and GLM miss the tornado over the border between Kansas and Oklahoma and several hail events to north of that tornado. There is also notable spatial displacement error between the high-probability swaths and the observed events in these two experiments. As expected, assimilating radar data

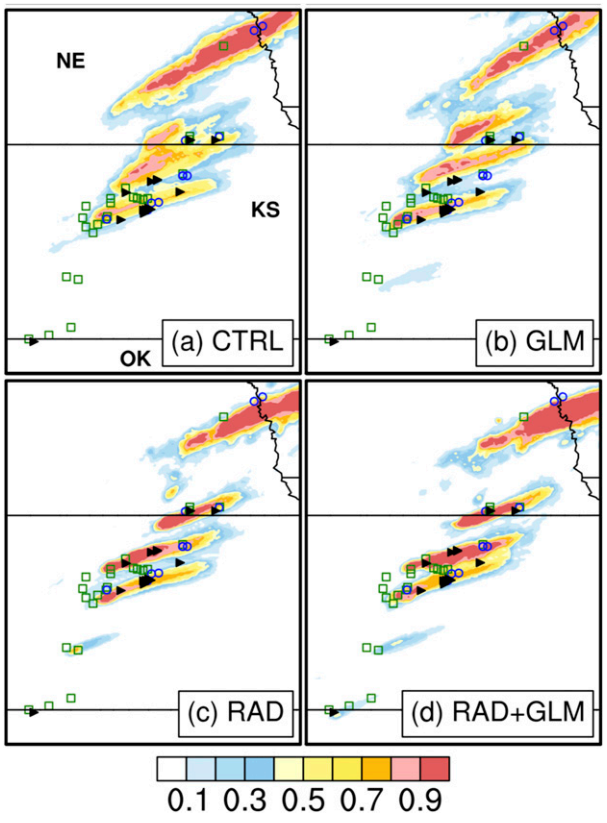

FIG. 15. Probability of simulated 2–5-km updraft helicity greater than 60 $m^2 s^{-2}$ within the 3-h forecast period beginning at 2300 UTC 1 May 2018 for experiments CTRL, GLM, RAD, and RAD+GLM. Blue circles represent severe wind reports, green squares represent large hail reports, and black triangles represent tornado reports during the 3-h forecast period.

(RAD and RAD+GLM) significantly improves the forecast skill of tornadic potential (Figs. 15c,d). Specifically, the three high-probability swaths overlay the observed tornadoes over the northern portions of Kansas quite well. The RAD+GLM experiment slightly outperforms RAD experiment by exhibiting a short low-probability (>0.3) swath over the border of Kansas and Oklahoma, where one tornado and several hail events occurred.

The impact of GLM FED assimilation on the short-term forecast is further evaluated by objective methods to supplement the above subjective verification. In additional to ETS, aggregate forecast metrics, including POD, CSI, FAR, SR, and the frequency bias are computed for the composite reflectivity and hourly accumulated precipitation relative to observations over seven separate 3-h forecasts in each experiment. These additional metrics are displayed on a performance diagram. Overall, assimilating GLM FED or radar data greatly improves the ETS for CREF compared to CTRL (Fig. 16a). The GLM and RAD experiments consistently improve the ETS for CREF by about 0.1 and 0.25 relative to CTRL, respectively (Fig. 16a). The RAD+GLM generates a slightly but consistently better performance than RAD. Correspondingly, the performance diagram indicates higher POD, lower FAR and the resultant higher CSI in GLM, RAD, and RAD+GLM relative to CTRL throughout the entire 3-h forecast period (Fig. 16c). Specifically, CTRL has a CSI value lower than 0.4 and a slightly low (<1) frequency bias for the first forecast hour. GLM increases the CSI up to ~0.6. Assimilating GLM FED into CTRL has little effect on the bias, which is a result of approximately equally increased POD and SR. Overall, CTRL and GLM both have quite good bias. Assimilating radar data (RAD, RAD+GLM) substantially outperforms CTRL and GLM. In addition to notably increasing the values of POD, SR and CSI (~0.8 for the first hour), radar DA is likely to change the bias for the first forecast hour from under forecasting to over forecasting. This indicates that radar DA tends to increase POD more than increase SR, which is probably due to the convection introduced by assimilating a large number of radar observations. Meanwhile, RAD+GLM creates a 3-h forecast with the highest CSI, which is contributed by a highest POD and a lowest FAR.

Hourly accumulated precipitation results (with a threshold of 10 mm) are presented in Figs. 16b and 16d. Overall, the results are consistent with those from CREF. RAD+GLM performs only slightly better than RAD, in terms of a slightly higher ETS, a higher POD, SR, and CSI. GLM produces an ETS about 0.1 higher relative to CTRL while RAD produces an ETS about 0.3 higher. The performance diagram shows that the precipitation from CTRL has a dry bias in the first forecast hour, then the frequency bias gradually increases to ~1.0 in the following two hours (Fig. 16d). GLM produces higher POD, SR, and CSI compared to CTRL, and generally maintains the same bias as CTRL. RAD produces nearly unbiased precipitation with higher POD, SR and CSI. Once again, RAD+GLM slightly outperforms RAD.

2) CASE 20 MAY 2019

The observation-space diagnostic statistics of FED for 20 May show that the CR decreases and becomes constant faster in RAD+GLM than in the GLM experiments for this case (Figs. 3f,j). Radar DA again spins the storms up faster so that RMSI in RAD+GLM grows more rapidly and is larger than GLM. The severe weather events were most active during 2100 to 0000 UTC as reported by NWS. Due to the incomplete MRMS CREF observation between 2130 and 2330 UTC, however, the PMM CREF from the analysis at 0000 UTC 21 May 2019 and the corresponding 1-h forecast in each experiment are compared to the MRMS observation in Fig. 17. Experiments RAD and RAD+GLM both well capture the entire MCS structure including the severe leading line and the associated stratiform area (Figs. 17d,e). Specifically, radar DA is able to enhance the analyzed reflectivity in both the stratiform and convective areas, while additional GLM DA modulates the reflectivity field to be slightly more consistent with the observation in intensity (Figs. 17a,d,e). For the 1-h forecast, severe supercells in both CTRL and GLM are less organized as a line through Oklahoma, and convection in CTRL is still more intense than the observation while GLM produces slightly weaker convection than CTRL (Figs. 17f–h). Correspondingly, CTRL slightly overpredicts the precipitation in central

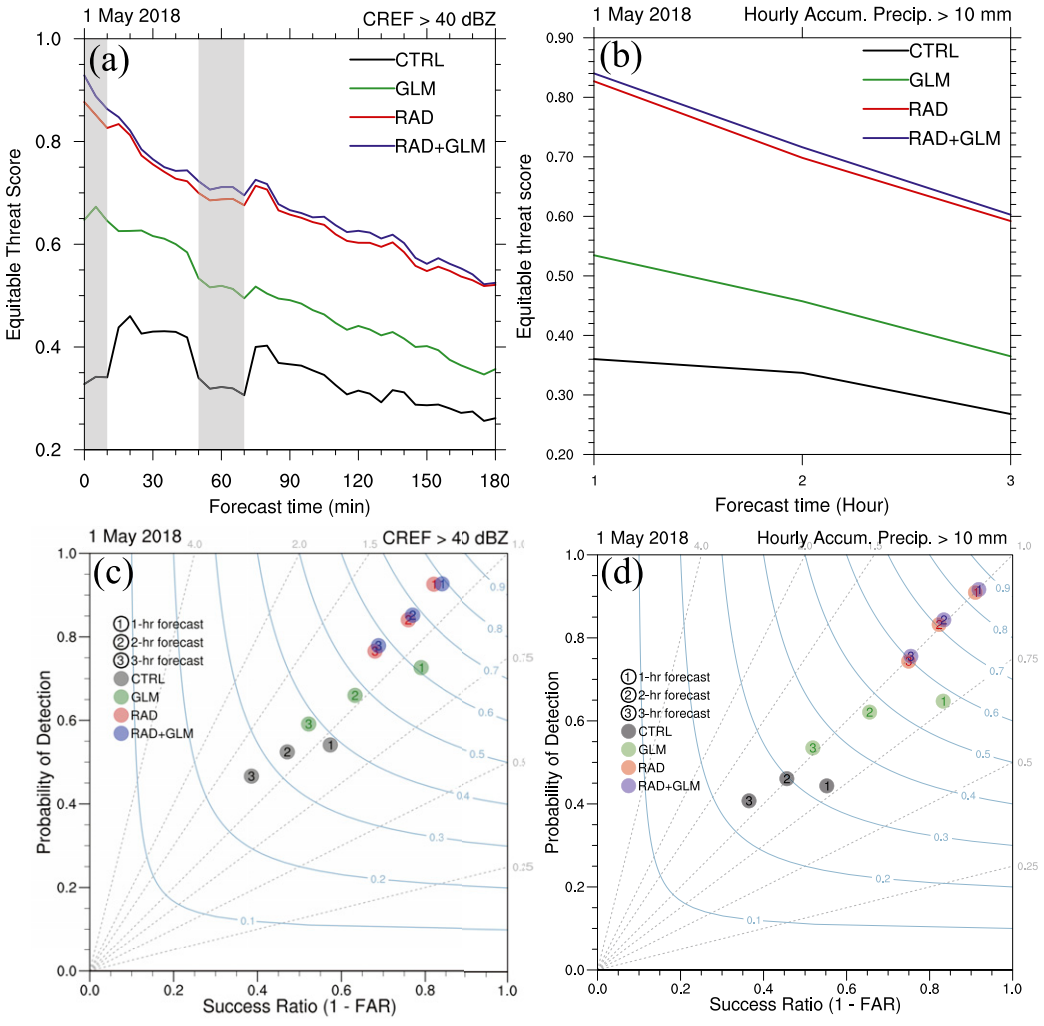

FIG. 16. (a),(b) Aggregate equitable threat score (ETS) and (c),(d) performance diagram of composite reflectivity greater than 40 dBZ relative to (left) the MRMS observations and (right) hourly accumulated precipitation greater than 10 mm relative to the Stage IV multisensor precipitation analyses over seven 3-h forecast periods of each experiment for 1 May 2018. The neighborhood radius is 15 km. Score metrics in the performance diagrams are averaged in the first, second, and third forecast hours. The shaded areas denote the ETS affected by missing MRMS observations.

Oklahoma while GLM underpredicts it (Figs. 18a–c). The size and the detailed structure of the MCS in RAD and RAD+GLM from the analyses are more like the observations and generally outperform the no-radar experiments (Figs. 17a–e). However, as the forecast continues, RAD and RAD+GLM produce excessive convection (Figs. 17f–j), resulting in overpredicted 3-h accumulated precipitation (Figs. 18d,e).

The probability of the model-simulated strong midlevel UH in the 3-h forecast initiated at 2100 UTC for this case is presented in Fig. 19, overlaid by a number of reported severe weather events including 16 tornadoes. The four experiments all capture the overall high-probability swaths of severe weather extending from the southwest to the northeast of the domain. The severe weather events in western Texas are the most poorly predicted with a large displacement error. In general, CTRL performs well but misses the tornado reports near the border between Kansas and Missouri and the tornadoes in Texas (Fig. 19a). Compared to CTRL, GLM is able to produce a moderate-probability (about 0.4) swath near the tornado reports on the border between Kansas and Missouri (Fig. 19b). As expected, assimilating radar data (RAD) remarkably increases the probability in regions overlaid by severe weather events (Fig. 19c). RAD also adds a high-probability swath near to, but south of the tornado in southwestern Oklahoma. A primary improvement of RAD+GLM compared to RAD is a substantial increase in the UH probability crossing the Kansas–Missouri border.

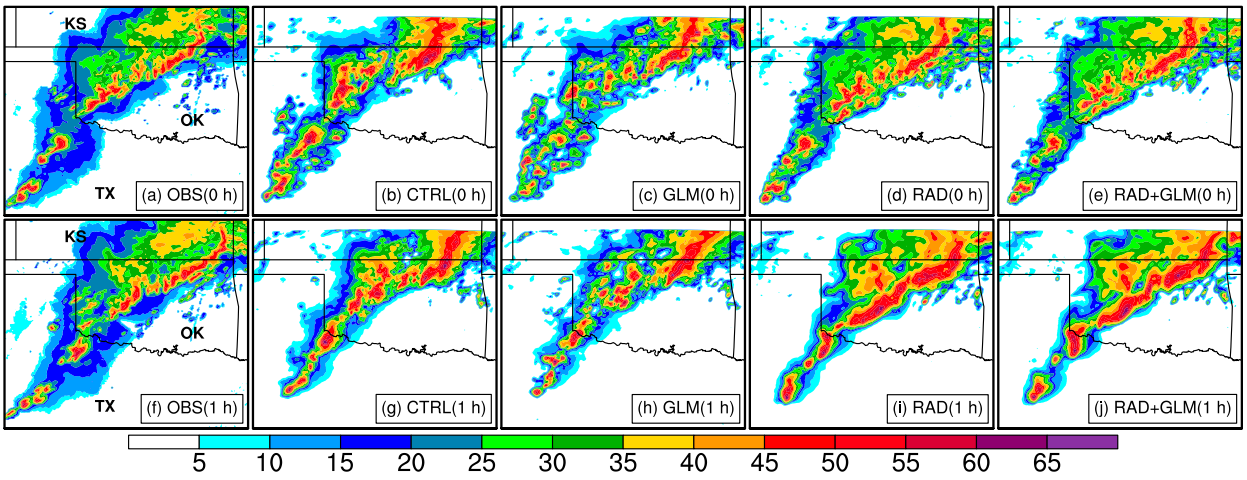

FIG. 17. As in Fig. 12, but for the analyses and 1-h forecasts initiated at 0000 UTC 21 May 2019.

The ETS time series and the performance diagram for CREF and precipitation prediction for 20 May 2019 are presented in Fig. 20. For CREF, both the ETS and the performance diagram show that CTRL has very similar performance in the first-, the second- and the third-forecast hour (Figs. 20a,c). GLM shares comparable CSI with CTRL in the first forecast hour, but with a lower POD, a higher SR and smaller bias (Fig. 20c). In the following two hours, lower POD and SR make GLM perform worse than CTRL. The experiments with radar assimilated (RAD, RAD+GLM) once again have a higher ETS than both CTRL and GLM throughout the 3 forecast hours, but the ETS values decrease faster than in CTRL (particularly in the first hour). Specifically, the difference between the ETS of RAD and CTRL gradually decreases from 0.2 at the beginning to 0.05 at the end. In terms of the contingency metrics, RAD and RAD+GLM have notably

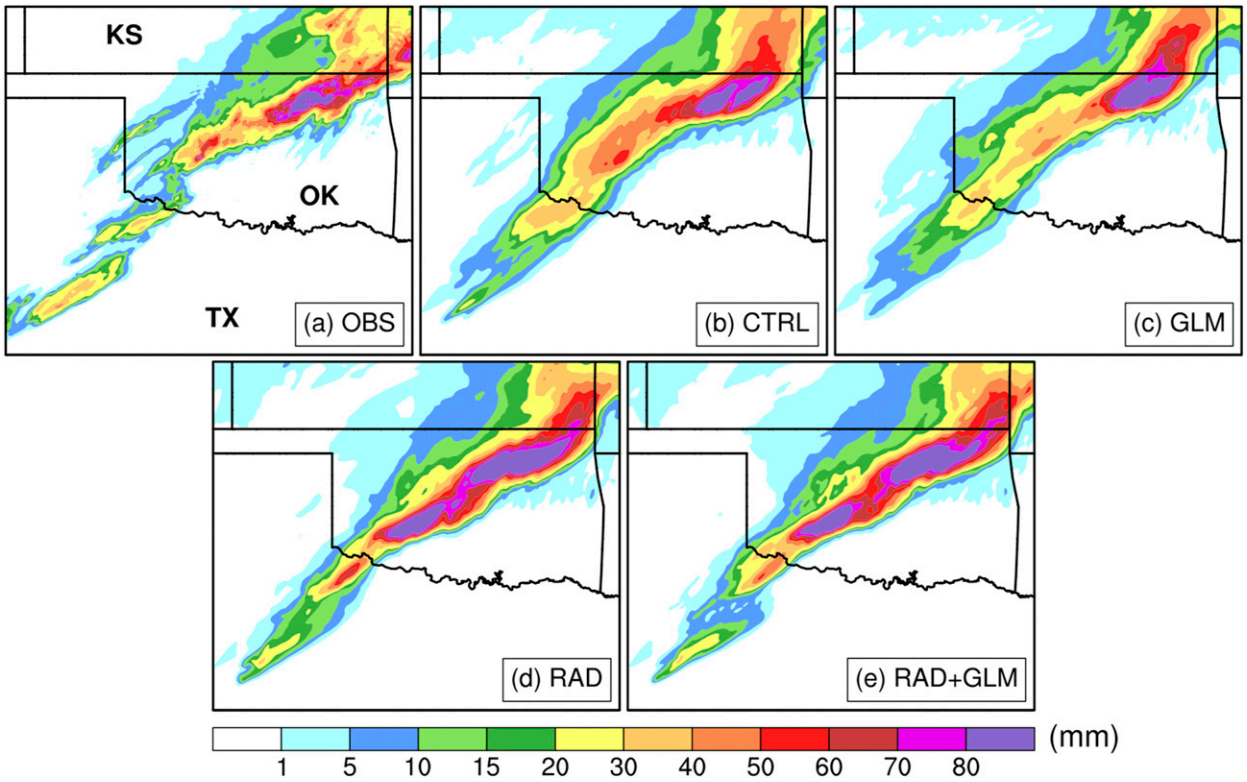

FIG. 18. As in Fig. 14 (accumulated precipitation), but for the forecasts initiated at 0000 UTC 21 May 2019.

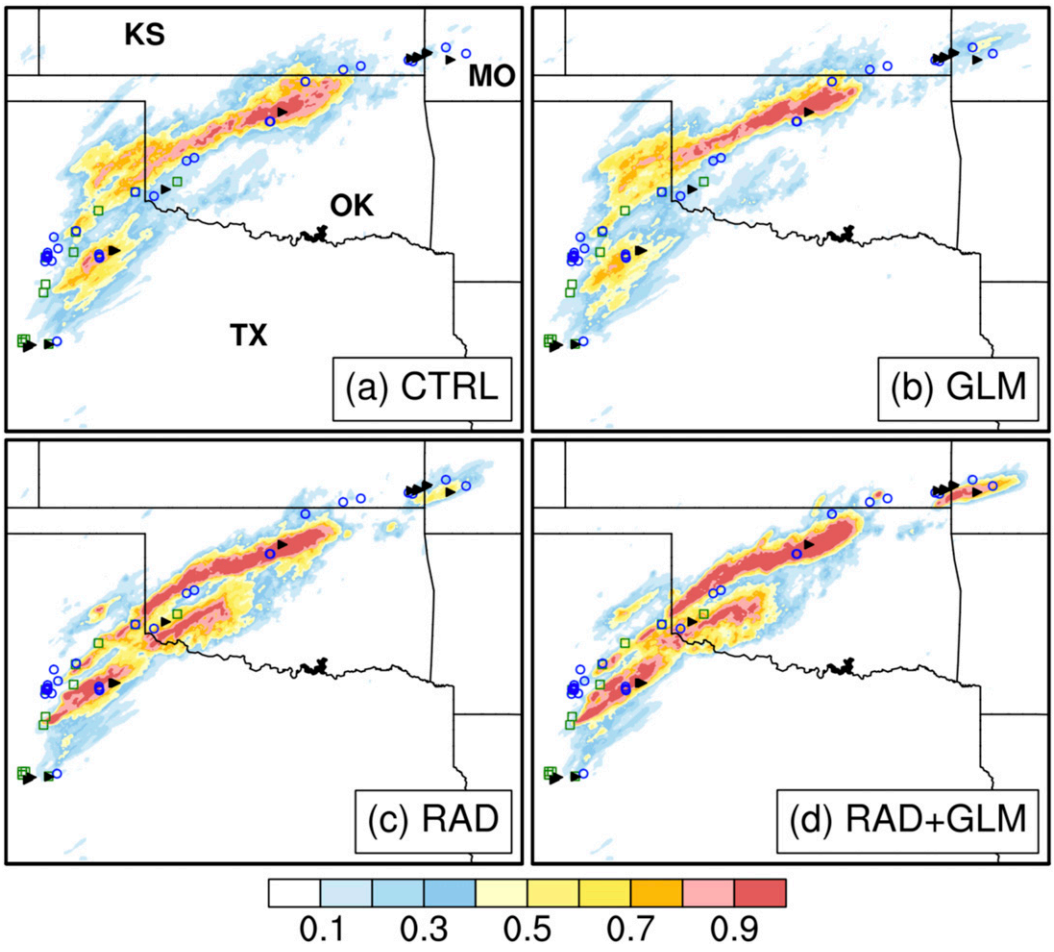

FIG. 19. As in Fig. 15 (updraft helicity probability), but for the forecasts initiated at 2100 UTC 20 May 2019.

higher POD, SR and CSI than GLM and CTRL. Again, RAD+GLM performs slightly better than RAD.

For the precipitation forecast in the case 20 May 2019, experiment GLM performs worst with the lowest ETS for the second and the third hour forecast while RAD and RAD+GLM perform much better with ETS at least 0.2 higher than that of CTRL (Fig. 20b). Both the ETS and the performance diagram show that CTRL has a very stable forecast skill through the 3 forecast hours (Figs. 20b,d). Though GLM has a similar CSI with CTRL in the first hour forecast, it had a larger dry bias with bias frequency value of ~0.8. In the following two hours, experiment GLM shows lower POD, SR, and CSI relative to CTRL. RAD and RAD+GLM performs very similar and both better than CTRL and GLM. RAD+GLM keeps slightly better performance than RAD in the three forecast hours. The positive impact of GLM DA in RAD+GLM is likely because the radar observations overwhelm the negative impact of the anomalously low FED observations.

In summary, compared to GLM FED data, assimilating radar data improves the forecast skill more substantially for both selected storms by reducing storm displacement error, better capturing and predicting the intensity and structure of thunderstorms, the associated accumulated precipitation, and the high-rotation tracks. This is an expected result given the higher density of radar observations, which provide direct wind and hydrometeor information. The assimilation of GLM data only, however, could obtain about 50% of the improvement of radar data in one case based on the ETS values (Figs. 16a,b), which is encouraging, given that the assimilated 2D GLM data are much sparser than radar data. With radar data assimilated, adding GLM data could also slightly improve the forecasts of both the supercell case and the MCS case.

4. Summary and conclusions

The GOES-R GLM instrument detects total lightning rate at high temporal and spatial resolution (8–12 km) over the Americas and adjacent oceanic regions. The GLM observations provide detection and monitoring of deep electrified convection, which has the potential to fill the gaps of the ground-based network in mountainous and ocean regions. This study assimilates FED derived from GLM flash data into WoFS using the EnKF DA technique for two severe convection events, one with a line of individual supercells and the other one with severe convective and trailing stratiform areas.

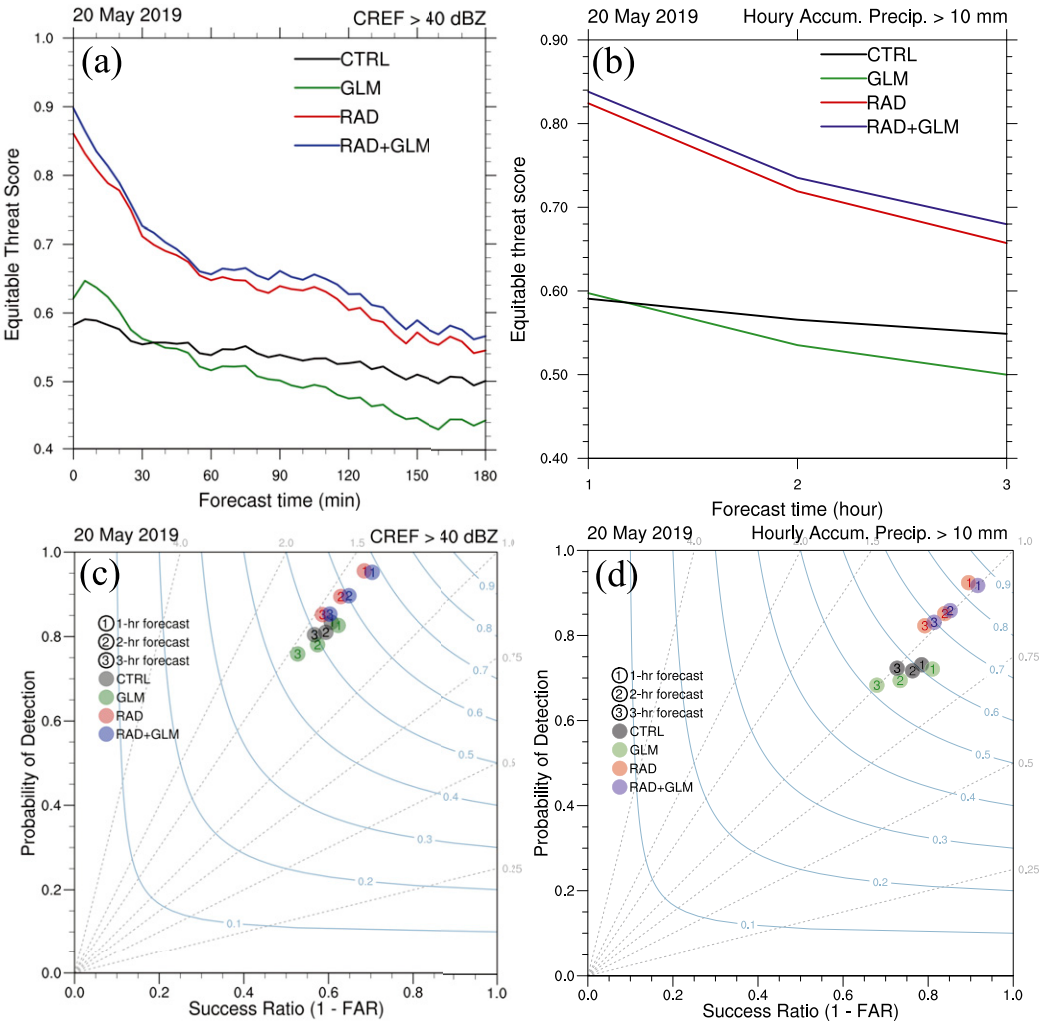

FIG. 20. As in Fig. 16, but for 20 May 2019 event.

A CTRL experiment solely assimilates conventional surface observations and geostationary satellite CWP. Additionally, a set of experiments assimilating GLM FED into CTRL are conducted to evaluate different DA configurations. In general, assimilating GLM FED produces FED analyses more consistent with the observation. However, the impact on the 3-h forecasts of composite reflectivity and precipitation are different between the two severe storm events. For the case with multiple individual supercells, experiment GLM_5min, which assimilates 5-min average GLM FED and applies additive noise to high-FED regions, substantially improves the forecast skill by better capturing the size, the intensity, and the tracks of supercells. For the MCS case consisting of deep electrified convection, nonintense convection and trailing stratiform regions, the GLM FED is unexpectedly low in the southwestern parts of the domain, apparently due to low GLM DE. While assimilating zero GLM FED values over stratiform regions without radar data suppresses the development of the MCS and greatly degrades the forecast despite the improvement in the analyzed FED and CREF field, assimilating the abnormally low GLM FED further harms the forecasts. The cold pools are weakened from assimilating low FED values and are less able to maintain reasonable convection. Nevertheless, UH probabilities are still enhanced by GLM where severe events occurred in the northeastern part of the domain.

The potential value of assimilating GLM FED observations together with radar data is also evaluated with two more experiments, one assimilating radar data and the other one assimilating a combination of radar data and GLM data. Assimilating GLM FED data in addition to radar data slightly outperforms the radar-only experiment in predicting composite reflectivity, accumulated precipitation, and rotation tracks in both two cases. Improvement in the analyzed fields and the subsequent short-term forecasts contributed by radar DA is more notable than assimilating GLM data alone. This result is expected from the higher information content of 3D radar reflectivity and radial velocity data compared to 2D GLM.

While the results in this paper indicate a positive impact of assimilating lightning observations together with radar data, more cases should be evaluated in the future to find an optimal

way to combine the lightning data with the radar data. The opposite impact of GLM FED DA in one case with and without assimilating dense radar data indicates a need to better understand and account for variations in GLM detection efficiency. Given that only two severe storm events in the locations with dense radar coverage are tested in this work, future work should expand case studies to regions with sparse radar coverage, e.g., the western CONUS and over sea, and over a variety of storm types, e.g., southwest monsoon thunderstorms and landfalling tropical cyclones. Additional future work could explore advanced observation operators considering different storm types by building statistics between GLM observations and WoFS forecasts. Calibration and quality control of GLM data before assimilation would also be necessary given the detection efficiency issues.

Acknowledgments. Local computer assistance is provided by Gerry Creager, Brett Morrow, Steven Fletcher, and Robert Coggins. The authors thank Dr. Eric Bruning for his assistance in using the glmtools package and Dr. Alexandre Fierro. The lead author thanks Junjun Hu and Derek Stratman for providing plotting scripts and Yongjie Huang for helpful discussions. The authors thank the three anonymous reviewers for their comments and suggestions, which improved the manuscript. This work is supported by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreements NA11OAR4320072, NA18OAR4590361, and NA16OAR4320072 (U.S. Department of Commerce).

Data availability statement. The GOES-R GLM data and radar data are freely available online in archives hosted by the National Centers for Environmental Information (NCEI). The national Stage IV QPE product can be found at https://www.emc.ncep.noaa.gov/mmb/SREF/pcpanl/stage4/. Due to confidentiality agreements, assimilation experiments data cannot be made available.

REFERENCES

Alexander, G. D., J. A. Weinman, V. M. 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 multisensory U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.

Bruning, E., and Coauthors, 2019: Meteorological imagery for the Geostationary Lightning Mapper. *J. Geophys. Res. Atmos.*, **124**, 14 285–14 309, https://doi.org/10.1029/2019JD030874.

Brunner, K., and P. Bitzer, 2020: A first look at cloud inhomogeneity and its effect on lightning optical emission. *Geophys. Res. Lett.*, **47**, e2020GL087094, https://doi.org/10.1029/2020GL087094.

Carey, L. D., and S. A. Rutledge, 1996: A multiparameter radar case study of the microphysical and kinematic evolution of a lightning producing storm. *Meteor. Atmos. Phys.*, **59**, 33–64, https://doi.org/10.1007/BF01032000.

Caya, A., J. Sun, and C. Snyder, 2005: A Comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094, https://doi.org/10.1175/MWR3021.1.

Chang, D., J. A. Weinman, C. A. Morales, and W. S. Olson, 2001: The effect of spaceborne microwave and ground-based continuous 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.

Chmielewski, V. C., and E. C. Bruning, 2016: Lightning Mapping Array flash detection performance with variable receiver thresholds. *J. Geophys. Res. Atmos.*, **121**, 8600–8614, https://doi.org/10.1002/2016JD025159.

——, D. MacGorman, D. Kennedy, and J. Blair, 2020: Surprisingly different outcomes from two different processing methodologies for the Oklahoma Lightning Mapping Array. *2020 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract AE007-06.

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

——, and Coauthors, 2020: A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bull. Amer. Meteor. Soc.*, **101**, E2022–E2024, https://doi.org/10.1175/BAMS-D-19-0298.1.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374, https://doi.org/10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.

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.

Dowell, D., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927, https://doi.org/10.1175/2008JTECHA1156.1.

——, ——, and D. J. Stensrud, 2004: High resolution analyses of the 8 May 2003 Oklahoma City storm. Part II: EnKF data assimilation and forecast experiments. Preprints, *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 12.5, http://ams.confex.com/ams/pdfpapers/81393.pdf.

——, ——, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Wea. Rev.*, **139**, 272–294, https://doi.org/10.1175/2010MWR3438.1.

——, and Coauthors, 2016: Development of a High-Resolution Rapid Refresh Ensemble (HRRRE) for severe weather forecasting. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 8B.2, https://ams.confex.com/ams/28SLS/webprogram/Paper301555.html.

Ebert, E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2.

Fierro, A. O., and E. R. Mansell, 2017: Electrification and lightning in idealized simulations of a hurricane-like vortex subject to wind shear and sea surface temperature cooling. *J. Atmos. Sci.*, **74**, 2023–2041, https://doi.org/10.1175/JAS-D-16-0270.1.

——, and ——, 2018: Relationships between electrification and storm-scale properties based on idealized simulations of an

intensifying hurricane-like vortex. *J. Atmos. Sci.*, **75**, 657–674, https://doi.org/10.1175/JAS-D-17-0202.1.

——, M. S. Gilmore, E. R. Mansell, L. J. Wicker, and J. M. Straka, 2006: Electrification and lightning in an idealized 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. R. Dembek, and C. L. Ziegler, 2015a: 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.

——, E. R. Mansell, C. L. Ziegler, and D. R. MacGorman, 2015b: Explicitly simulated electrification and lightning within a tropical cyclone based on the environment of Hurricane Isaac (2012). *J. Atmos. Sci.*, **72**, 4167–4193, https://doi.org/10.1175/JAS-D-14-0374.1.

——, 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.

——, Y. Wang, J. Gao, and E. R. Mansell, 2019: Variational assimilation 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.

Fuchs, B. R., E. C. Bruning, S. A. Rutledge, L. D. Carey, P. R. Krehbiel, and W. Rison, 2016: Climatological analyses of LMA data with an open-source lightning flash-clustering algorithm. *J. Geophys. Res. Atmos.*, **121**, 8625–8648, https://doi.org/10.1002/2015JD024663.

Gallo, B. T., and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting*, **32**, 1541–1568, https://doi.org/10.1175/WAF-D-16-0178.1.

Goodman, S. J., D. E. Buechler, P. D. Wright, and W. D. Rust, 1988: Lightning and precipitation history of a microburst-producing storm. *Geophys. Res. Lett.*, **15**, 1185–1188, https://doi.org/10.1029/GL015i011p01185.

——, 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.

Hakim, G. J., P. Regulski, C. Mass, and R. Torn, 2008: Lightning data assimilation using an ensemble Kalman filter. *20th Int. Lightning Detection Conf.*, Tucson, AZ, Vaisala, 6 pp., https://www.vaisala.com/sites/default/files/documents/Lightning%20Data%20Assimilation%20using%20an%20Ensemble%20Kalman%20Filter.PDF.

Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289, https://doi.org/10.1256/qj.05.135.

Hu, J., A. O. Fierro, Y. Wang, J. Gao, and E. R. Mansell, 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. *Mon. Wea. Rev.*, **148**, 1005–1028, https://doi.org/10.1175/MWR-D-19-0198.1.

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.

——, P. S. Skinner, N. Yussouf, K. Knopfmeier, A. E. Reinhart, and D. C. Dowell, 2019: Forecasting high-impact weather in landfalling tropical cyclones using a Warn-on-Forecast system. *Bull. Amer. Meteor. Soc.*, **100**, 1405–1417, https://doi.org/10.1175/BAMS-D-18-0203.1.

Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research Scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806, https://doi.org/10.1175/BAMS-84-12-1797.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP global data assimilation system. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Kong, R., and Coauthors, 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. *Mon. Wea. Rev.*, **148**, 2111–2133, https://doi.org/10.1175/MWR-D-19-0192.1.

Kuhlman, K. M., C. L. Ziegler, 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.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, https://doi.org/10.1175/WAF-D-17-0145.1.

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.

——, and Coauthors, 2008: TELEX: The Thunderstorm Electrification and Lightning Experiment. *Bull. Amer. Meteor. Soc.*, **89**, 997–1014, https://doi.org/10.1175/2007BAMS2352.1.

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

——, C. L. Ziegler, 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.

——, K. Hillburn, and S. D. Miller, 2019: Geostationary Lightning Mapper and Earth Networks lightning detection over the contiguous United States and dependence on flash characteristics. *J. Geophys. Res. Atmos.*, **124**, 11 552–11 567, https://doi.org/10.1029/2019JD031039.

McPherson, R. A., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma mesonet. *J. Atmos. Oceanic Technol.*, **24**, 301–321, https://doi.org/10.1175/JTECH1976.1.

Murphy, M., and R. Said, 2020: Comparisons of lightning rates and properties from the U.S. National Lightning Detection Network (NLDN) and GLD360 with GOES-16 Geostationary Lightning Mapper and Advanced Baseline Imager data. *J. Geophys. Res. Atmos.*, **125**, e2019JD031172, https://doi.org/10.1029/2019JD031172.

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, 2009a: Relationships among lightning, precipitation, and hydrometeor characteristics over the North Pacific Ocean. *J. Appl. Meteor. Climatol.*, **48**, 833–848, https://doi.org/10.1175/2008JAMC1817.1.

——, and ——, 2009b: 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.

——, ——, K. L. Cummins, and T. Turner,2004: On the relationship between lightning and convective rainfall over the central Pacific Ocean. *18th Int. Lightning Detection Conf.*, Helsinki, Finland, Vaisala, Ref. 21, http://www.soest.hawaii.edu/met/Faculty/businger/personnel/pessi/ILDC2004_Pessi.pdf.

——, ——, and T. Cherubini, 2006: Comparison of two methods for assimilation of lightning data into NWP models. Preprints, *First Int. Lightning Meteorology Conf.*, Tucson, AZ, Vaisala, CD-ROM.

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

Rutledge, S. A., K. A. Hilburn, A. Clayton, B. Fuchs and S. D. Miller, 2020: Evaluating Geostationary Lightning Mapper flash rates within intense convective storms. *J. Geophys. Res. Atmos.*, **125**, e2020JD032827, https://doi.org/10.1029/2020JD032827.

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.

Skinner, P. S., and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast System. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smith, T. M., 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.

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2.

Thomas, R. J., P. R. Krehbiel, W. Rison, S. J. Hunyady, W. P. Winn, T. Hamlin, and J. Harlin, 2004: Accuracy of the lightning mapping array. *J. Geophys. Res.*, **109**, D14207, https://doi.org/10.1029/2004JD004549.

Wang, H., and Coauthors, 2017: Improving lightning and precipitation prediction of severe convection using lightning data assimilation with NCAR WRF-RTFDDA. *J. Geophys. Res. Atmos.*, **122**, 12 296–12 316, https://doi.org/10.1002/2017JD027340.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman filter analyses and forecasts of a severe mesoscale convective system using different choices of microphysics schemes. *Mon. Wea. Rev.*, **142**, 3243–3263, https://doi.org/10.1175/MWR-D-13-00260.1.

——, K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

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.

Yussouf, N., and K. H. Knopfmeier, 2019: Application of Warn-on-Forecast system for flash-flood producing heavy convective rainfall events. *Quart. J. Roy. Meteor. Soc.*, **145**, 2385–2403, https://doi.org/10.1002/qj.3568.

——, E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, https://doi.org/10.1175/MWR-D-12-00237.1.

——, 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.

——, J. S. Kain, and A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Wea. Forecasting*, **31**, 957–983, https://doi.org/10.1175/WAF-D-15-0160.1.

——, T. A. Jones, and P. S. Skinner, 2020: Probabilistic high-impact rainfall forecasts from landfalling tropical cyclones using Warn-on-Forecast system. *Quart. J. Roy. Meteor. Soc.*, **146**, 2050–2065, https://doi.org/10.1002/qj.3779.

Zhang, D. K., and K. Cummins, 2020: Time evolution of satellite-based optical properties of lightning flashes, and its impact on GLM flash detection. *J. Geophys. Res. Atmos.*, **125**, e2019JD032024, https://doi.org/10.1029/2019JD032024.

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