Assimilating All-Sky Infrared Radiances from *GOES-16* ABI Using an Ensemble Kalman Filter for Convection-Allowing Severe Thunderstorms Prediction

YUNJI ZHANG, FUQING ZHANG, AND DAVID J. STENSRUD

Center for Advanced Data Assimilation and Predictability Techniques, and Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 19 February 2018, in final form 19 July 2018)

ABSTRACT

This article presents the first practice of assimilating real-world all-sky *GOES-16* ABI infrared brightness temperature (BT) observations using an ensemble-based data assimilation system coupled with the Weather Research and Forecasting (WRF) Model at a convection-allowing (1 km) horizontal resolution, focusing on the tornadic thunderstorm event across Wyoming and Nebraska on 12 June 2017. It is found that spurious clouds created before observed convection initiation are rapidly removed, and the analysis and forecasts of thunderstorms are significantly improved, when all-sky BT observations are assimilated with the adaptive observation error inflation (AOEI) and adaptive background error inflation (ABEI) techniques. Better forecasts of the timing and location of convection initiation can be achieved after only 30 min of assimilating BT observations; both deterministic and probabilistic WRF forecasts of midlevel mesocyclones and low-level vortices, started from the final analysis with 100 min of BT assimilation, closely coincide with the tornado reports. These improvements result not only from the effective suppression of spurious clouds, but also from the better estimations of hydrometeors owing to the frequent assimilation of all-sky BT observations that yield a more accurate analysis of the storm. Results show that BT observations generally have a greater impact on ice particles than liquid water species, which might provide guidance on how to better constrain modeled clouds using these spaceborne observations.

1. Introduction

Severe thunderstorms that produce damaging winds, flash floods, hail, and tornadoes have long been one of the major threats to human life and property. When examining billion-dollar weather and climate disasters in the United States during 1980–2017, it is shown that severe thunderstorms are responsible for over 40% of the total number of these events, almost 1/7 the total economic losses, and about 1/6 of the deaths (NCDC 2018). Monitoring and predicting severe thunderstorms is one of the most important and most difficult parts of operational weather forecast and warning operations. The average warning lead time for tornadoes has increased from 3 min in 1978 to 14 min in 2011 (Wurman et al. 2012; Stensrud et al. 2013), thanks in large part to the establishment of the nationwide Weather Surveillance Radar-1988 Doppler (WSR-88D) network.

Doppler weather radars play a critical role in the current severe weather warning paradigm, either via direct utilization of their reflectivity and radial velocity observations or in combination with numerical weather prediction models through data assimilation techniques (e.g., Polger et al. 1994; Clark et al. 2012). The next-generation geostationary weather satellites, including *Himawari-8/9* of the Japan Meteorological Agency (Bessho et al. 2016), the GOES-R series of NASA (Schmit et al. 2005), the FY-4 series of the China Meteorological Administration (Yang et al. 2017), and the Meteosat Third-Generation (MTG) series of the EUMETSAT (Stuhlmann et al. 2005), will provide a similarly important opportunity to enhance the monitoring and prediction of severe weather events with more spectral bands and higher spatial and temporal resolutions than their predecessors. *Himawari-8*, *GOES-16*, and *FY-4A* launched on 7 October 2014, 9 November 2016, and 10 December 2016, respectively, and are already providing high-quality, continuous images of the atmosphere.

Corresponding author: Fuqing Zhang, fzhang@psu.edu

Infrared imagers on board geostationary weather satellites, such as the Advanced Baseline Imager (ABI) on board *GOES-16*, have some unique advantages over ground-based Doppler weather radars. Cloud development and extent is observed earlier by satellite imagers that focus on radiance near Earth's surface than by Doppler weather radars where precipitating hydrometeors must be present for detection. Unlike Doppler radars that suffer from limited horizontal coverage and beam blocking from built structures, satellite imagers can provide nephograms without "gaps." Imagers like ABI can provide scans of the continental United States (CONUS) every 5 min, comparable to the interval of each WSR-88D volume scan in precipitation mode (5–6 min), with a subpoint horizontal resolution of 2 km for infrared channels. This 2-km resolution is sufficient to monitor the initiation and development of convective clouds (Schmit et al. 2017). Because of these potential benefits, it is desirable to directly assimilate infrared brightness temperature (BT; will be used interchangeably with "radiance" in this article) observations from geostationary satellites into numerical weather prediction (NWP) models with the goal to improve severe weather forecasts.

Previous studies have assimilated synthetic infrared BT observations from ABI using ensemble-based techniques like the ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Zhang 2016). The EnKF has the advantage of providing flow-dependent time-varying estimation of background error covariances, compared with variational-based assimilation techniques like 3DVar (Zhang et al. 2011), and thus is widely used in data assimilation applications for severe thunderstorms at convection-allowing storm scales (e.g., Snyder and Zhang 2003; Aksoy et al. 2009; Dowell et al. 2011; Wheatley et al. 2015; Yussouf et al. 2015; Yokota et al. 2016). Observing system simulation experiments (OSSEs) of the direct assimilation of synthetic ABI BT observations using EnKF have mostly focused on extratropical cyclones (Otkin 2010, 2012; Zupanski et al. 2011; Jones et al. 2013), mesoscale convective systems (Jones et al. 2014; Cintineo et al. 2016), or tropical cyclones (F. Zhang et al. 2016; Minamide and Zhang 2017, 2018a); recently, Honda et al. (2018a,b) assimilated real-data all-sky radiance observations from the Advanced Himawari Imager (AHI) on board the *Himawari-8* satellite, which has similar infrared channels as the ABI onboard the *GOES-16* satellites, to improve predictions of tropical cyclones and associated torrential precipitation and floods. However, storm-scale data assimilation studies using geostationary satellite observations only assimilated temperature and moisture profile retrievals (Jones et al. 2017),

cloud-top temperature (Kerr et al. 2015), water paths of different hydrometeor species (Jones and Stensrud 2015; Jones et al. 2015, 2016), and *GOES-13* clear-sky infrared radiance (Jones et al. 2018), rather than all-sky (clear sky and cloud affected) infrared radiance observations from high spatiotemporal imagers like the ABI. This study is the first attempt to directly assimilate real-world all-sky ABI infrared BT observations using an EnKF approach with a numerical model running at a convection-allowing 1-km horizontal resolution (finer than all previous studies) to improve model forecasts of a tornadic thunderstorm event.

The tornadic thunderstorm event from 12 June 2017 in Wyoming and Nebraska is briefly summarized in section 2. The observations, numerical model, and experiment design are presented in section 3. Verification of the EnKF is discussed in section 4, followed in section 5 by an evaluation of the performance of the deterministic and ensemble forecasts initialized at various times from the EnKF analysis. Results are summarized in section 6.

2. Overview of the 12 June 2017 severe weather outbreak

During the morning of 12 June 2017, an upper-level low was moving eastward from northern Nevada into Utah, with an upper-level jet located to its southeast and stretching from southern Nevada and northern New Mexico into Wyoming and Nebraska. Southeasterly winds in the low levels to the east of the Rocky Mountains transported abundant moisture into eastern Colorado and southeastern Wyoming during the previous day. Strong instability with surface-based CAPE above $3000 \, \text{J} \, \text{kg}^{-1}$ and little to no CIN was present in this region by early afternoon of 12 June. The 0–6-km wind shear exceeded $10 \, \text{m} \, \text{s}^{-1}$ in northeastern Colorado, southeastern Wyoming, and the Nebraska panhandle, accompanied by a very large supercell composite parameter (Thompson et al. 2003) greater than 8 as well as a very large significant tornado parameter (Thompson et al. 2003) greater than 2 in the adjacent regions of these three states at local noon [mountain standard time (MST); MST = UTC − 0700]. The NOAA Storm Prediction Center (SPC) predicted a moderate risk of severe thunderstorms with tornado probability greater than 15% (with possible significant tornadoes) and large hail probability greater than 45% in eastern Wyoming and surrounding areas in its 1630 UTC day 1 convective outlook issued on 12 June 2017. SPC later issued a tornado watch at 1910 UTC for northeast Colorado, western Nebraska, and southeast Wyoming effective until 0200 UTC 13 June, indicating a particularly dangerous situation (PDS) for possible intense tornadoes,
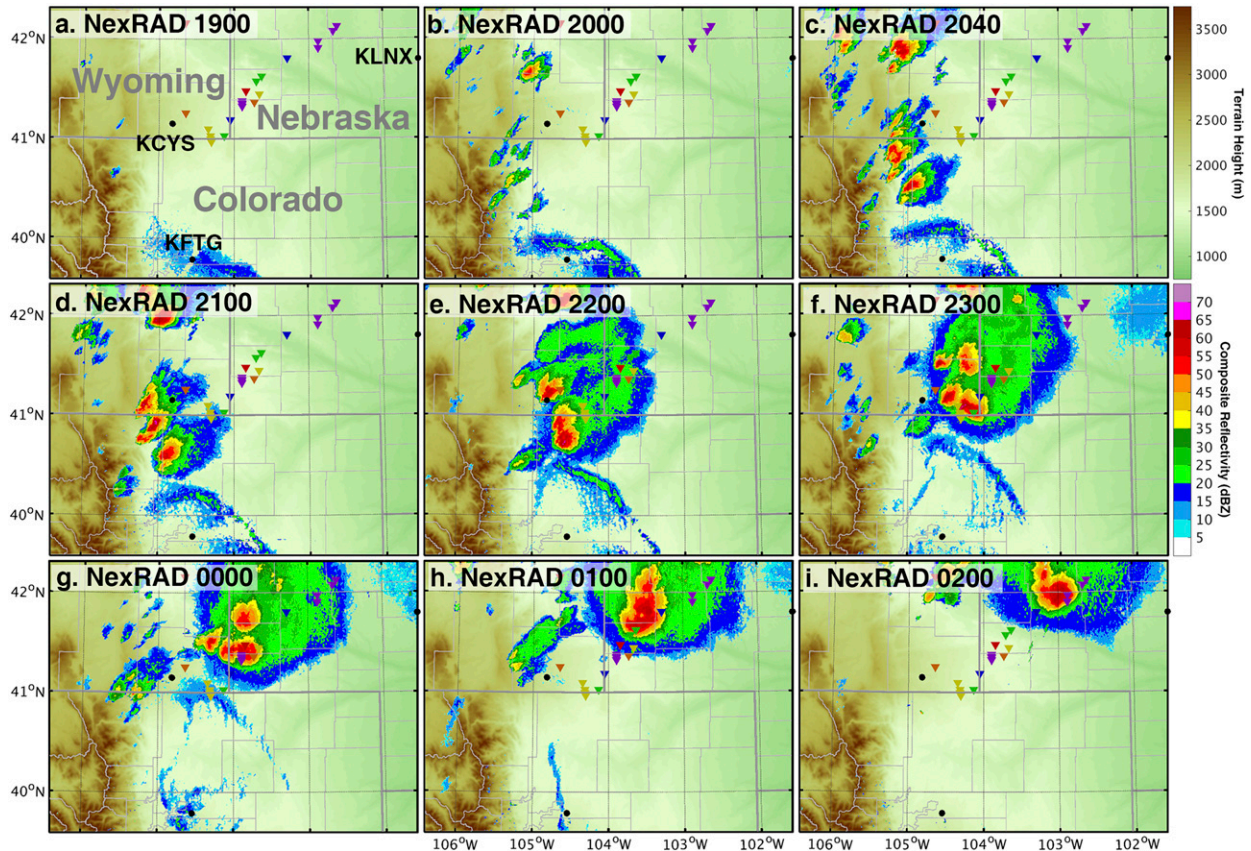
FIG. 1. Composite reflectivity at (a) 1900, (b) 2000, (c) 2040, (d) 2100, (e) 2200, and (f) 2300 UTC 12 Jun, and (g) 0000, (h) 0100, and (i) 0200 UTC 13 Jun 2017. Shading of background indicates elevations. Filled triangles in all panels are tornado reports from SPC, with different colors indicating reports from different tornadoes. Locations of WY, NE, and CO and the three WSR-88Ds that were used to generate composite reflectivity are also marked in (a).

large hail, and damaging winds—the first PDS tornado watch ever issued for Wyoming.

Remotely sensed observations of this severe weather event from before convection initiation until after the final tornado report are shown in Figs. 1 and 2. The composite reflectivity images of this event (Fig. 1) are generated by interpolating the level 2 reflectivity of three WSR-88Ds located at Denver, Colorado (KFTG), Cheyenne, Wyoming (KCYS), and North Platte, Nebraska (KLNX), onto a domain that is used for numerical experiments with a horizontal resolution of 1 km (see section 3 for the model settings) and taking the maximum values vertically as well as 2.5 min before and after each plotting time. The BT images from channel 10 (Fig. 2), the lower-level tropospheric water vapor channel of ABI with a central wavelength of 7.3 μm, are interpolated onto same plotting grid as in Fig. 1.

Widespread convection initiated along the slopes of the Rocky Mountains in north-central Colorado near the border between Colorado and Wyoming, as well as farther north in central Wyoming during early afternoon

around 2000 UTC (Figs. 1b, 2b). Several storms developed parallel to the mountain slopes (Figs. 1c,d), and their overshooting tops can be clearly identified in satellite BT images (Figs. 2c,d). The storms moved northeastward into Wyoming and Nebraska (Figs. 1d–f, 2d–f), with the southernmost storm generating the first EF-2 tornado of the day from 2258 to 2308 UTC with a track across the Colorado–Wyoming border. There was also another EF-2 tornado, which formed at 2309 UTC from a separate storm farther to the north and lasted until 0001 UTC 13 June. Between 2300 and 0100 UTC, some of the storms merged (Fig. 1f), and new storms also initiated alongside the existing storms (Figs. 1f–h), generating multiple EF-0 and EF-1 tornadoes; several studies imply that there might be relationships between storm mergers and subsequent tornadogenesis (e.g., Wurman et al. 2007; Hastings and Richardson 2016; Honda and Kawano 2016). All these storms eventually combined to form a single very intense supercell thunderstorm (Figs. 1h, 2h). This supercell produced a long-lived EF-2 tornado from 0119 to 0225 UTC that tracked
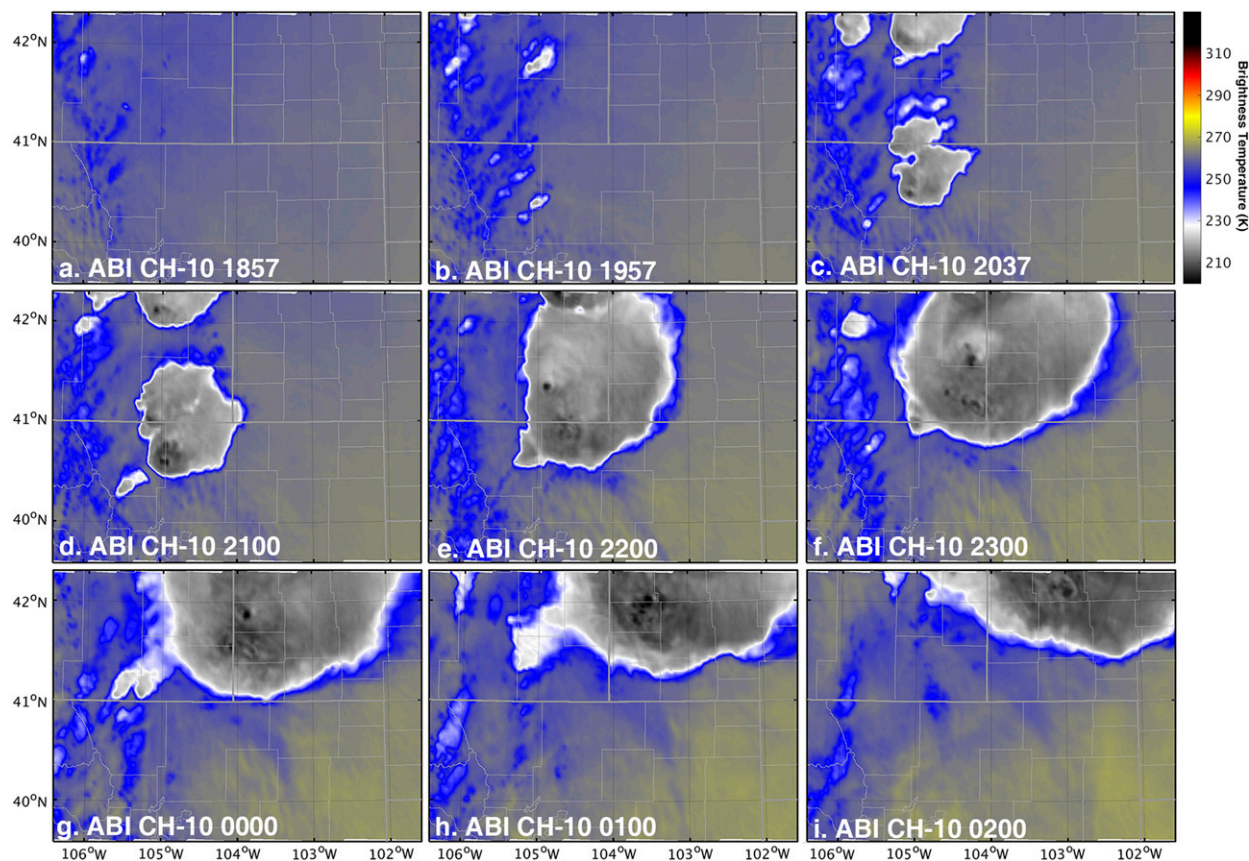
FIG. 2. Brightness temperature of channel 10 of *GOES-16* ABI at (a) 1857, (b) 1957, (c) 2037, (d) 2100, (e) 2200, and (f) 2300 UTC 12 Jun, and (g) 0000, (h) 0100, and (l) 0200 UTC 13 Jun 2017.

nearly 40 mi. This storm continued moving northeastward (Figs. 1h,i)—weakening and becoming disorganized, producing widespread anvil clouds (Figs. 2h,i)—and eventually dissipated in South Dakota.

While convection initiation (CI) processes and storm overshooting tops can be clearly observed in the satellite images, it is hard to determine the precise storm evolution and storm interactions from satellite observations. There is also no clear relationship between the height of the overshooting tops and the severity of the storms. For example, the storm associated with the highest overshooting top during storm development from 2200 to 0000 UTC (Figs. 2e–g) was comparably weak in radar reflectivity (Figs. 1e–g) and only spawned an EF-0 tornado with 4-min lifespan to the northeast of Cheyenne, Wyoming.

3. Data, model, and experiments

a. GOES-16, *ABI, and the observations*

GOES-16, the first satellite of the GOES-R series, was launched by an Atlas V rocket at Cape Canaveral, Florida, on 19 November 2016. It underwent multiple examinations, calibrations, and validations for a year while holding at its checkout location at 89.5°W above the equator, which is the location of the satellite during the event of this study. It was shifted to the GOES-East operational location at 75.2°W on 11 December 2017 and was announced as the operational GOES-East satellite on 18 December 2017, succeeding *GOES-13*. The successor of current GOES-West (*GOES-15*), *GOES-17*, was launched on 1 March 2018, and is expected to become operational in late 2018.

Among the multiple instruments on board *GOES-16* that sense the Earth, the sun, and space environment, the ABI is the primary instrument for imaging Earth's weather and climate. It has two visible, four near-infrared, and 10 infrared channels with subpoint resolutions of 0.5–2 km, up to 4 times finer than its predecessors. The ABI is able to produce a full disk scan every 15 min, a CONUS scan (5000 km × 3000 km) every 5 min, and a mesoscale scan (1000 km × 1000 km) every 30 s using a "flex" scan mode, or it can produce a full disk scan every 5 min using a "continuous full disk" scan mode. Previous GOES imagers took at least 25 min to finish a full disk scan. Further details

on the ABI and characteristics of its channels can be found in Schmit et al. (2005, 2017).

Observations used in this study are retrieved from NOAA's Comprehensive Large Array-Data Stewardship System (CLASS). Channel 10 of the ABI has a weighting function that peaks at roughly 620 hPa in clear-sky conditions (Schmit et al. 2017), and observations from this channel are assimilated. The ABI measured level 1b radiances are converted to BT observations as the Cloud and Moisture Imagery product (CMIP). The BT observations have been calibrated and geolocated, and raw BT observations with a horizontal resolution of ~2.5 km in Colorado, Wyoming, and Nebraska taken during the event will be used without data thinning. It should be pointed out that observational errors of high-resolution observations like BT from *GOES-16* ABI usually contain spatial correlations; however, in order to take the most advantage of the higher resolutions of ABI compared with its predecessors, especially the high-resolution details of the convective-scale structures of the developing thunderstorms, we choose to use raw BT observations without data thinning, while using a relatively large observational error together with the adaptive observation error inflation (AOEI) technique [Minamide and Zhang (2017); will be explained in more detail in section 3b] to implicitly take spatially correlated observational errors into account. We do not perform any additional quality control procedures on the raw observations.

b. Numerical model and data assimilation systems

The Pennsylvania State University (PSU) WRF–EnKF cycling data assimilation system (Zhang et al. 2009; Weng and Zhang 2016)—evolved from an early development of Meng and Zhang (2008a,b) based on the fully compressible, nonhydrostatic Advanced Research core of the Weather Research and Forecasting (WRF-ARW) numerical model (Skamarock et al. 2008), version 3.8.1—is used in this study. A single model domain consisting of $401 \times 301 \times 61$ grid points with a horizontal grid spacing of 1 km, 61 vertical layers with 19 in the lowest 1 km above ground level (AGL) and the uppermost level at 50 hPa, is used. This domain covers northern Colorado, southeastern Wyoming, and southwest Nebraska, where the tornadic thunderstorm formed and developed (Fig. 1). Physical parameterization schemes are chosen based on sensitivity experiments and are similar to the suite used in the High-Resolution Rapid Refresh (HRRR) model, including the Thompson et al. (2008) microphysics scheme with mixing ratios of water vapor (Qv), cloud water (Qc), cloud ice (Qi), rainwater (Qr), snow (Qs), and graupel (Qg) and number concentration of ice and rainwater; unified Noah land surface model (Ek et al. 2003); Monin–Obukhov–Janjić Eta scheme

(Janjić 1996) for surface layer parameterization; Mellor–Yamada–Janjić TKE scheme (Janjić 1994) for PBL processes; and the Rapid Radiative Transfer Model for general circulation models (RRTMG) schemes (Iacono et al. 2008) for longwave and shortwave radiation. Simulated radar reflectivity is calculated using the built-in module of the Thompson microphysics scheme.

The Community Radiative Transfer Model (CRTM; Han et al. 2006) is a rapid radiative transfer model that is widely used in the satellite data assimilation community as the observation operator to calculate simulated brightness temperature from numerical model variables. This study uses CRTM to generate simulated satellite infrared BT from the WRF-ARW variables with the successive order of interaction (SOI) forward solver (Heidinger et al. 2006) using the OPTRAN code. The default standard tropical region profile in CRTM was used above the model top of 50 hPa; this profile choice has a minor impact on the calculations, given that atmosphere above 50 hPa contributes little to the simulated BT according to the weighting function of channel 10.

The data assimilation part of the cycling PSU WRF–EnKF system uses the ensemble square root filter (EnSRF; Whitaker and Hamill 2002) variation of EnKF. The AOEI technique (Minamide and Zhang 2017) is used to limit potentially erroneous innovations for all-sky radiance assimilation and is shown to provide better analysis and forecasts than a fixed observation error; it adaptively inflates observation error for BT observations if the increment is large, while keeping a minimum 3-K error when the increment is small. The adaptive background error inflation (ABEI) technique (Minamide and Zhang 2018b, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*) is used to help initiate clouds and is proven to be effective; it provides inflation factors proportional to the discrepancies between simulated and observed brightness temperatures where observed clouds are not simulated, while truncated at a maximum value of 1.4 corresponding to a discrepancy of 44.4 K for channel 10 of ABI, same as in Minamide and Zhang (2018b, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*). The covariance relaxation method (Zhang et al. 2004) is used to maintain ensemble spread, combining 80% of prior perturbation and 20% of posterior perturbation in the EnKF analysis. Using a criterion similar to F. Zhang et al. (2016), every channel 10 BT observation is first assessed to determine whether it is in a clear-sky region, as defined when channel 14 (window channel) BT are higher than 285 K, or in a cloudy region, as defined when channel 14 BT are lower than 285 K. Each observation height is then assigned to be either 620 hPa, the peak of weighting function of channel 10 in clear-sky

condition, or 250 hPa, the approximate height of cloud top, depending upon whether the observation is in a clear-sky or cloudy region, respectively. A broad radius of influence (ROI) for vertical localization of 5 times the altitude AGL of each observation is used during the EnKF; considering that BT is an accumulated nonlocal type of observation, the length scale of vertical ROI is selected so that influences from observations will reduce by about 25% at both the model bottom and top under clear-sky conditions. The horizontal ROI is fixed to 30 km for all BT observations, which is slightly narrower than the 40–60-km horizontal ROI that has been used in previous simulated and real ABI radiance assimilation studies with horizontal model grid spacing of 3–6 km (Jones et al. 2015; Cintineo et al. 2016; Honda et al. 2018b); this study uses a much higher 1-km resolution for the numerical model together with raw ABI radiance observations.

c. Experiment design

Two sets of simulations without data assimilation are first carried out as references. One deterministic forecast is run from 1800 UTC 12 June to 0000 UTC 13 June using the hourly HRRR analysis as initial and lateral boundary conditions (referred to as "DETER" hereafter). Another simulation is a set of ensemble forecasts generated by first running 6- and 12-h 20-member ensemble forecasts from 1200 and 0600 UTC 12 June using corresponding GEFS analysis and forecasts; then, 40 ensemble initial conditions are generated by subtracting the mean of these 40 GEFS simulations at 1800 UTC from each simulation and adding these 40 "perturbations" to the 1800 UTC HRRR analysis. The perturbations are downscaled before adding them to the HRRR analysis to generate the ensemble initial conditions and have a domain-averaged standard deviation of 0.55 K, 1.05 g kg^{-1}, 1.14 m s^{-1}, and 1.12 m s^{-1} for temperature, water vapor mixing ratio, and the two horizontal components of wind, respectively. The calculations of moisture are performed on dewpoint temperature instead of water vapor mixing ratio to avoid negative mixing ratio values, assuming unchanged pressure of dry air during conversion between water vapor mixing ratio and dewpoint temperature. Last, a 40-member ensemble forecast is carried out from 1800 until 0000 UTC 13 June using these 40 initial conditions and corresponding updated boundary conditions (referred to as "NODA" hereafter). These two sets of simulations are briefly described in section 3d.

Cycling EnKF is started at 1900 UTC from the NODA ensemble forecast valid at the same time, an hour before observed CI around 2000 UTC. The EnKF is performed every 5 min, consistent with the 5-min interval of BT observations, and only BT observations are assimilated.
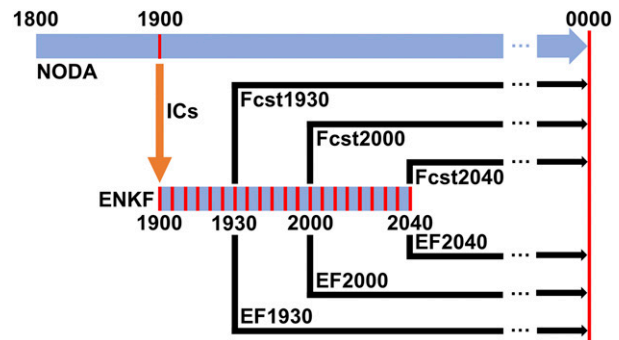

FIG. 3. Schematic diagram of the experiment design for cycling EnKF assimilating *GOES-16* ABI BT observations and associated deterministic and ensemble forecasts.

There is a change from "flex" scan mode to "continuous full disk" scan mode for the ABI data from 2040 to 2100 UTC, during which observations were unavailable; thus, the cycling EnKF ends at 2040 UTC, providing 100 min of EnKF cycles (referred to as "ENKF" hereafter). Deterministic forecasts are initialized from the 1930, 2000, and 2040 UTC EnKF analysis mean (referred to as "Fcst1930," "Fcst2000," and "Fcst2040," respectively, hereafter), and ensemble forecasts are also initialized at these times using EnKF analysis ensembles (referred to as "EF1930," "EF2000," and "EF2040," respectively, hereafter). A schematic diagram showing the experiment design is provided in Fig. 3.

d. Simulations without data assimilation

The reference deterministic and ensemble simulations without data assimilation are examined in this subsection. Figure 4 shows composite reflectivity from the DETER forecast. Storms are initiated along the slopes of the Rocky Mountains shortly after initialization at 1800 UTC, more than 1 h earlier compared with the observed CI (Fig. 1b), and form into several strong storms as early as 1900 UTC (Fig. 4a). These storms undergo a merging process similar to the observed storms (Figs. 4c,d). Although large discrepancies are present in the timing of CI in the DETER forecast compared with observations, the general location of CI and movement of the storms are close to the actual storms, as these characteristics are primarily regulated by mesoscale to synoptic-scale environmental conditions, which are generally well simulated by rapid-cycling convection-allowing HRRR system. The DETER forecast also simulates a well-defined track of 2–5-km updraft helicity (UH; Fig. 5a) in close proximity to the tornado reports, indicating the existence of an intense mesocyclone. However, there are also several UH tracks alongside the longest one in the eastern half of the model domain. Furthermore, a low-level vortex (represented by maximum
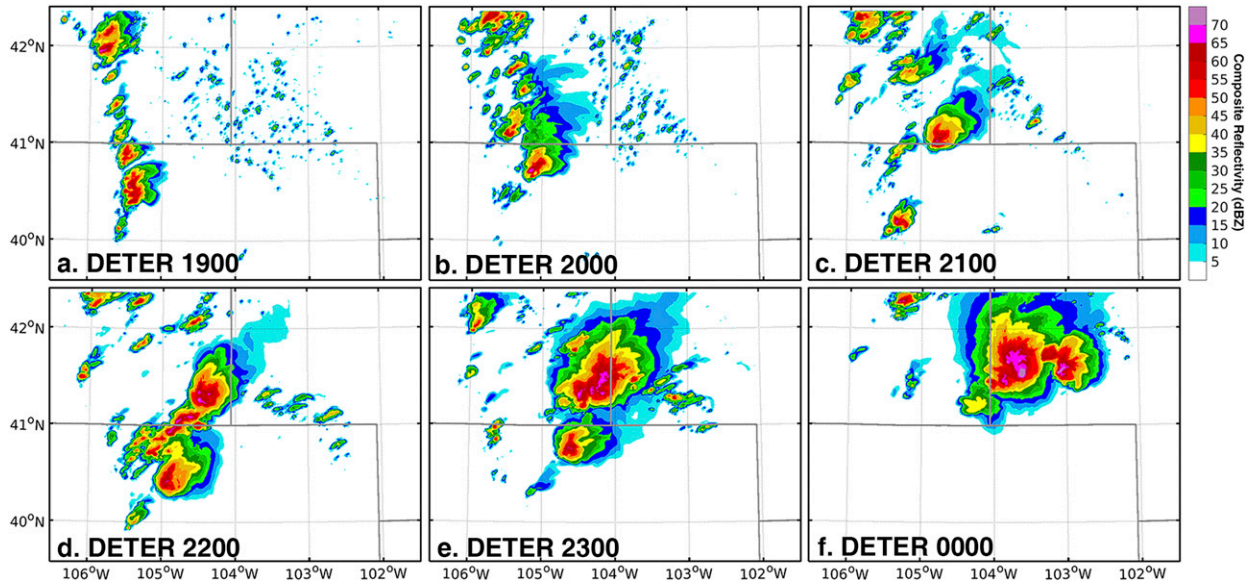
FIG. 4. Simulated composite reflectivity of DETER forecast at (a) 1900, (b) 2000, (c) 2100, (d) 2200, and (e) 2300 UTC 12 Jun, and (f) 0000 UTC 13 Jun 2017.

vertical vorticity within 1 km AGL) associated with these UH tracks is intermittent, and strong near-surface rotation only existed for a short period of time (Fig. 5c).

Similar to the DETER forecast, there also are earlier storm initiations along the slopes of the Rockies in the NODA ensemble forecast (Fig. 6a). Furthermore, the ensemble has very large uncertainties at later times, especially after 2200 UTC (Figs. 6d–f), with previously higher probabilities (greater than 60%) gradually diminishing to values lower than 50%, indicating a loss of confidence in the storm prediction. As can be expected from the large spread of composite reflectivity, ensemble
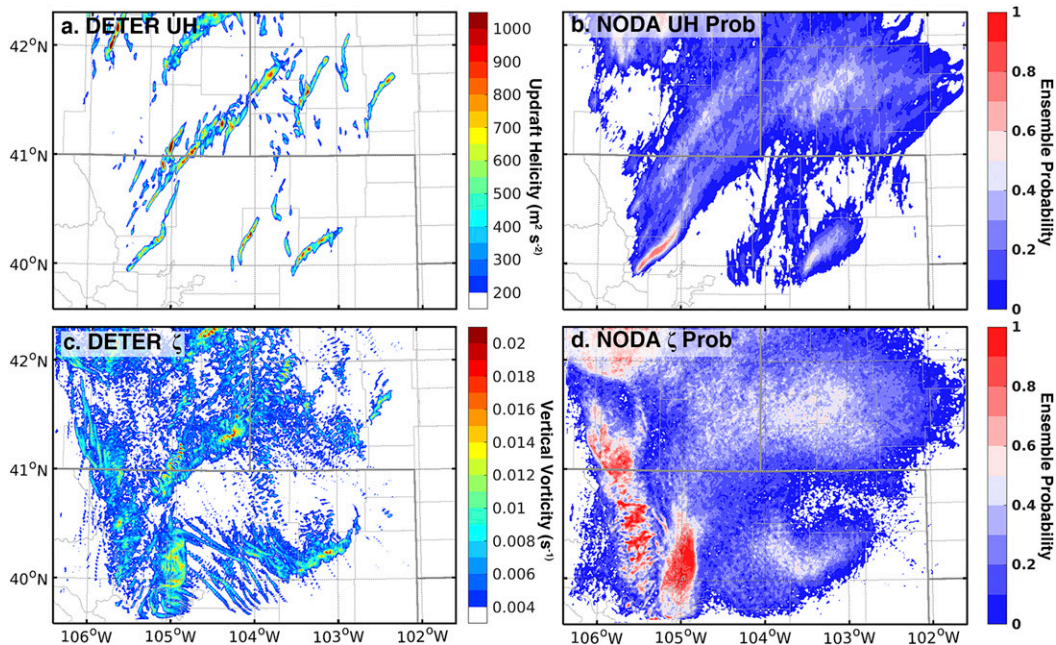

FIG. 5. Deterministic forecasts of (a) maximum of 2–5-km UH and (c) maximum of 0–1-km maximum vertical vorticity of DETER forecast. Probabilistic forecasts of (b) $180 \, m^2 \, s^{-2}$ 2–5-km UH and (d) $0.004 \, s^{-1}$ 0–1-km maximum vertical vorticity of NODA ensemble forecast. Temporal maximum values throughout the entire simulations were used.
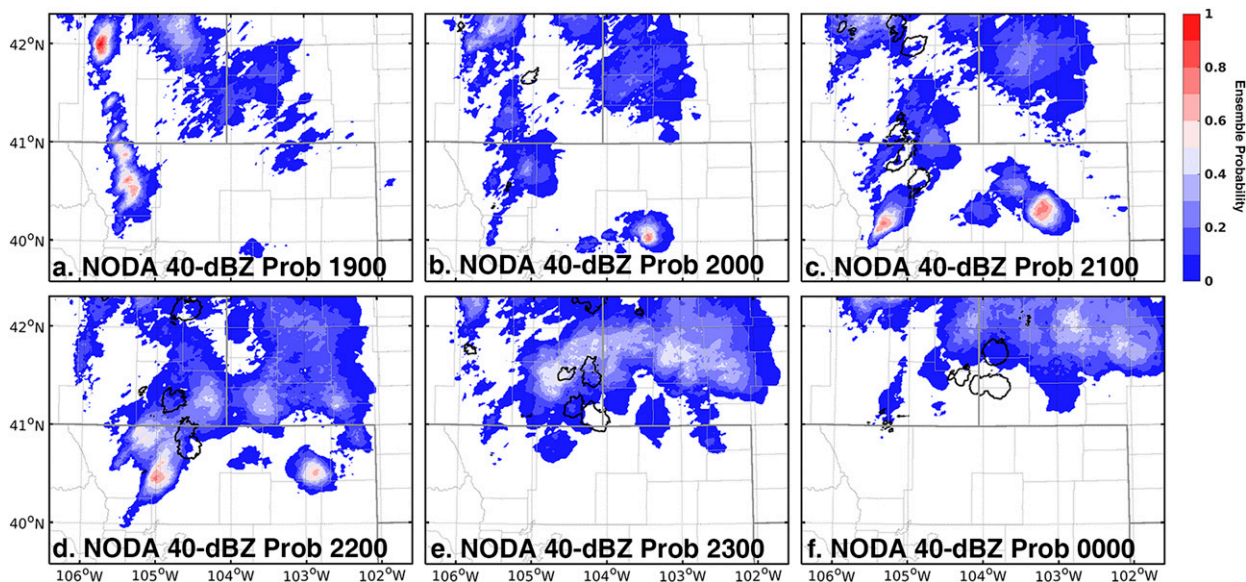
FIG. 6. Ensemble probability of 40-dBZ composite reflectivity of NODA ensemble forecast at (a) 1900, (b) 2000, (c) 2100, (d) 2200, and (e) 2300 UTC 12 Jun, and (f) 0000 UTC 13 Jun 2017. Black contours in all panels are observed 40-dBZ composite reflectivity.

probability of midlevel mesocyclone in NODA is widespread and generally weak (lower than 50%; Fig. 5b), although there is a region with slightly higher probability in southeast Wyoming, coincident with the tornado reports. However, consistent with that of DETER forecast, there are no organized low-level vortex tracks in high-UH-probability regions (Fig. 5d), indicating large uncertainties in the prediction of near-surface rotation in the NODA ensemble forecast.

4. The accuracy of EnKF analysis

a. Quantitative verifications

The performance of the WRF–EnKF system is first investigated using observation–space diagnostic metrics, including root-mean-square innovation/fit (RMSI/RMSF), bias, and ensemble spread. RMSI/RMSF is calculated as

$$ \text{RMSI/RMSF} = \sqrt{\left\langle (\mathbf{d} - \mathbf{d})^2 \right\rangle}, $$

where $\mathbf{d} = \mathbf{y}^o - \overline{H(\mathbf{x}^b)}$ or $\mathbf{d} = \mathbf{y}^o - \overline{H(\mathbf{x}^a)}$ represents the innovation of background mean $[\overline{H(\mathbf{x}^b)}]$ or fit of analysis mean $[\overline{H(\mathbf{x}^a)}]$, compared with observations $(\mathbf{y}^o)$ in observation space, respectively; H is the observation operator (which is the CRTM model here); and $\mathbf{x}$ is the model state vector with superscripts b and a indicating the background (prior) and analysis (posterior) estimates. Bias is calculated as

$$ \text{Bias} = \langle -\mathbf{d} \rangle, $$

where $\mathbf{d}$ is the same as defined in RMSI/RMSF calculations. Ensemble spread is calculated as

$$ \text{Spread} = \sqrt{\left\langle \frac{1}{N-1} \sum_{n=1}^{N} \left[H(\mathbf{x}_n) - \overline{H(\mathbf{x}_n)} \right]^2 \right\rangle}, $$

where N is the ensemble size (which is 40 here), and n is the index of each ensemble member.

These verification metrics during EnKF cycles are shown in Fig. 7a. The BT RMSI/RMSF experiences a very steep drop from 23 to 7 K at the first (1900 UTC) EnKF cycle, and a similar reduction of bias from 16 to 2 K occurs simultaneously. RMSI/RMSF remains at a relatively small magnitude afterward and shows a very slight increase after 2000 UTC due to storm initiation. The persistent decrease of the magnitude of bias throughout the entire EnKF cycles indicates a reduction of spurious convection (will also be examined later). Ensemble spread also drops rapidly during the first several EnKF cycles and remains at almost constant values around 2 K after 1945 UTC, while root-mean-square observational errors $(\sqrt{\langle \sigma_o^2 \rangle})$ generally follow the trend of RMSI/RMSF, consistent with its formulation. The quasi-steady state of all these metrics after several EnKF cycles indicates that the EnKF system is working properly and maintains a close representation of the observations.

We further divide mean bias of ensemble mean into averages of positive bias and negative bias. A positive bias resulted from higher brightness temperature in the
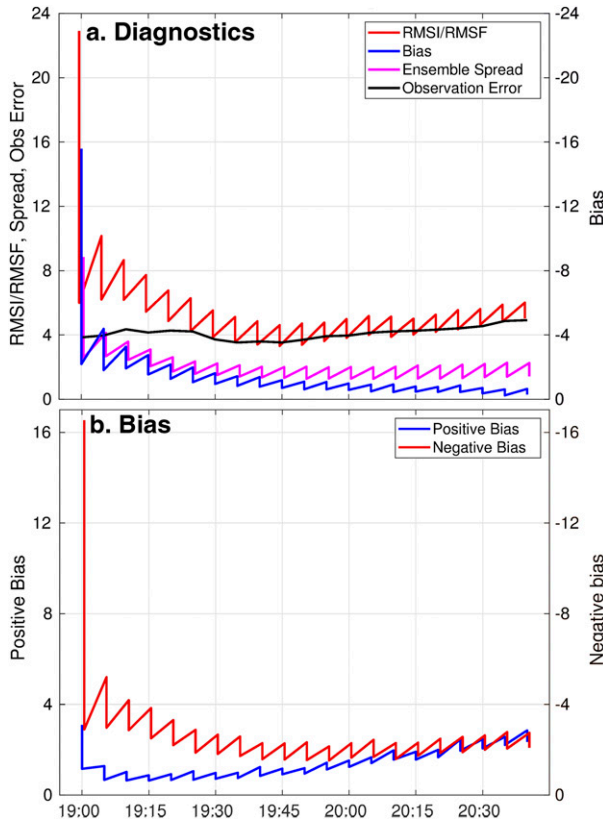
FIG. 7. Evolutions of (a) observation–space diagnostics of RMSI/ RMSF (red) and mean bias (blue) of ensemble mean, ensemble spread (purple), and RMS observation error (black), and (b) mean positive bias (blue) and mean negative bias (red) of ensemble mean throughout the EnKF cycles.

observation than prior or posterior ensemble mean indicates spuriously simulated clouds, while a negative bias indicates insufficient coverage of clouds or lower cloud tops in the simulation, compared with observations. It can be seen in Fig. 7b that the negative bias persistently remained at a relatively small value of ~3 K after the first several cycles, indicating an effective removal of spuriously simulated clouds during EnKF cycles and subsequent short-term ensemble forecasts between EnKF cycles. On the other hand, although the positive bias also remained below 3 K throughout all EnKF cycles, there is a slight increase in the magnitude of positive bias after 1945 UTC, resulted from a comparably slower initiation of clouds in the EnKF cycles, compared with the observations. However, magnitudes of both positive and negative biases experience notable reduction during every EnKF cycle (became closer to 0), proving that our technique can both suppress and initiate clouds, and the increase of positive bias could potentially be reduced by tuning ABEI parameters.

The rank histogram (Anderson 1996; Hamill and Colucci 1996, 1997; Harrison et al. 1995; Talagrand et al. 1997; Hamill 2001) is used to examine whether the ensemble spread is reasonable and representative to background uncertainties (Fig. 8). It plots the distribution of the ranks of each observation among corresponding predicted background or analysis values in the observation space. A flatter histogram indicates a sufficient variance, a U-shaped histogram indicates an insufficient variance, and a left- or right-skewed histogram indicates a positive or negative ensemble bias, respectively. At the beginning of the EnKF cycles, the ensemble has insufficient variance and negative bias (observation values larger than simulated values; Fig. 8a), consistent with Fig. 7a. The bias is greatly reduced after the first EnKF cycle, with the number of observations ranked at the upper end reduced by 70% (from 11 851 to 3442). The insufficient variance is also partly reduced, although still skewed rightward. Several cycles later at 2000 UTC, the shape of the rank histogram is much flatter than the first cycle at 1900 UTC, and there is even a little overdispersion at this time when the leftmost and rightmost biases are excluded (Fig. 8b). At the end of the EnKF cycles (Fig. 8c), the shape of the rank histogram is considerably flatter than earlier cycles, indicating that EnKF helps to maintain a reasonably good ensemble spread, although there are some amounts of both right bias (resulted from spurious clouds) and left bias (resulted from insufficient coverage of observed clouds).

Ensemble equitable threat scores (ETSs; Wilks 2011) of composite reflectivity are used to examine the improvement of assimilating BT observations. The ETS is calculated as

$$ETS = \frac{H - R}{H + M + F - R},$$

$$R = \frac{(H + M)(H + F)}{n},$$

where H, M, and F are the numbers of total hits, misses, and false alarms of all (40) ensemble members combined, and n is the total number of grid points of all ensemble members combined, which equals $40 \times 400 \times 300 = 4\,800\,000$ here. ETS of deterministic forecasts are calculated similarly using one forecast instead of the 40-member ensemble to determine hits, false alarms, misses, and the total number of grid points. Two thresholds of 20 and 40 dBZ for composite reflectivity are evaluated, representing regions of general precipitation and convective precipitation, respectively. From Fig. 9, it is clear that before 2000 UTC, because there is no organized storm, ETS for 40-dBZ reflectivity remains near
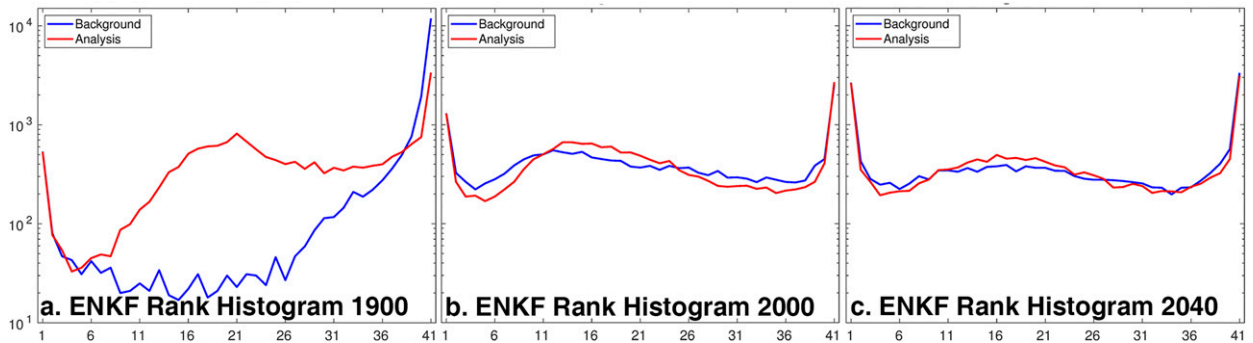
FIG. 8. Rank histograms of ENKF prior ensemble (blue) and posterior ensemble (red) at (a) 1900, (b) 2000, and (c) 2040 UTC.

0 (Fig. 9b) for EnKF, NODA, and DETER, while the ETS values for 20-dBZ reflectivity show fluctuations (Fig. 9a). After 2000 UTC, a persistent increase occurs for both scores, indicating increasing skill in the analysis of hydrometeors, although neither these variables nor composite reflectivity is directly assimilated. Meanwhile, it is clear that EnKF ETS values are constantly higher than those of the NODA ensemble at both reflectivity thresholds throughout the entire EnKF cycles, indicating an improved analysis of the initiation and early development of the storms, resulting from the efficient removal of spurious clouds as well as triggering of new clouds, which is examined next by comparing EnKF analysis with observations.

b. Comparisons with observations

Simulated ABI channel 10 BT of prior (background) and posterior (analysis) mean of the EnKF experiment at 1900, 2000, and 2040 UTC is shown in Fig. 10, while their respective differences with assimilated observations ($O - B$ and $O - A$, which have opposite signs

compared with bias defined in previous subsection) are shown in Fig. 11. Because of much earlier CI in the 1-h ensemble forecast before the first EnKF cycle at 1900 UTC (Fig. 6a), there is a vast region within the model domain that is covered with clouds (Fig. 10a) with large positive innovations (Fig. 11a), while there is no cloud in observations at all (Fig. 2a). However, it is apparent that the spurious clouds are significantly reduced after the first EnKF cycle (Fig. 10b): the large regions of positive innovations in prior mean (Fig. 11a) are greatly reduced, and only localized regions of clouds with positive innovations along the mountain slopes exist after EnKF in posterior mean (Fig. 11b). There are a few new clouds that form during the 5-min ensemble forecasts between each cycle, but assimilating BT observations effectively reduces the number of clouds that are inconsistent with the observations, and the spurious clouds are almost completely removed 1 h later at 2000 UTC (Fig. 10d). New clouds also are effectively generated by assimilating BT observations as the coverage of clouds at the last EnKF cycle (Figs. 10e,f) is close to the
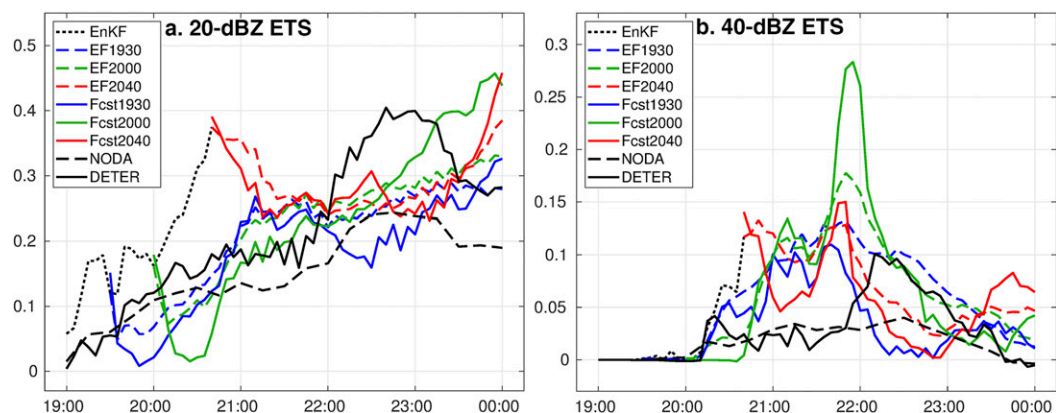

FIG. 9. ETS scores for (a) 20- and (b) 40-dBZ composite reflectivity regions of DETER (black solid), NODA (black dashed), ENKF (black dotted), deterministic forecasts from EnKF analysis at 1930 (blue solid), 2000 (green solid), and 2040 (red solid) UTC, and ensemble forecasts at 1930 (blue dashed), 2000 (green dashed), and 2040 (red dashed) UTC.
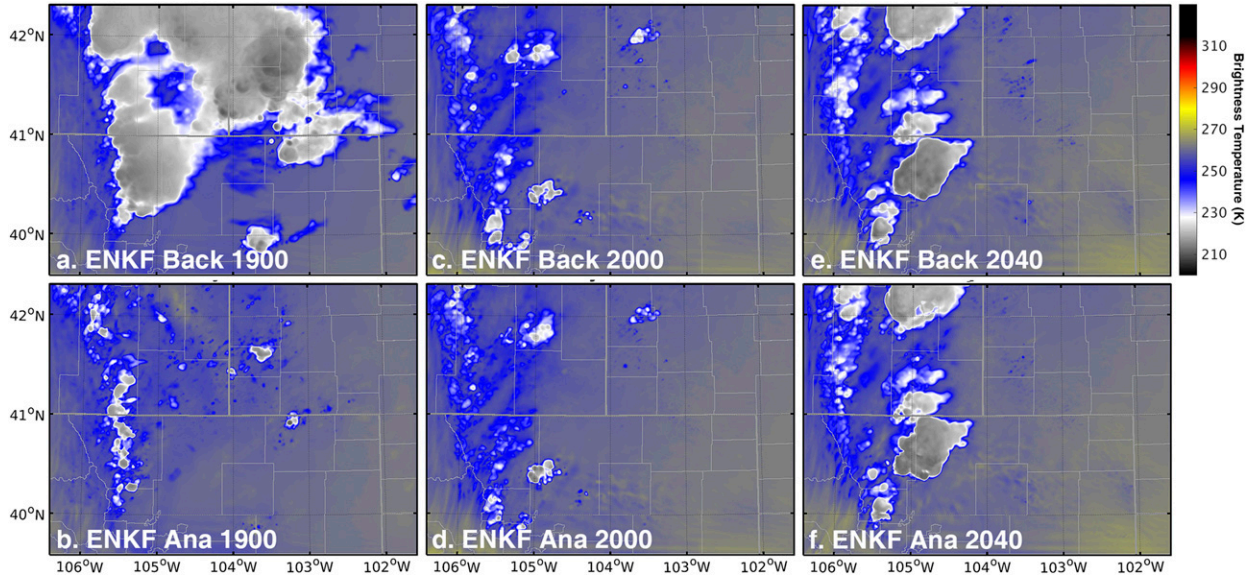
FIG. 10. Simulated brightness temperature of ABI channel 10 of (a),(c),(e) prior mean and (b),(d),(f) posterior mean of EnKF at (left) 1900, (middle) 2000, and (right) 2040 UTC.

observed clouds (Fig. 2c). The difference $O - A$ at this time (Fig. 11f) also indicates a good representation of clouds in EnKF analysis mean compared to the observed clouds, with only a slightly smaller coverage and warmer temperature, and the reduction of negative difference in $O - A$ (Fig. 11f) compared with $O - B$ (Fig. 11e), proves that our techniques are initiating clouds using BT observations, consistent with evolution of biases in Fig. 7b.

The influence of assimilating BT on hydrometeors is further investigated by examining the evolution of domain-averaged liquid water path (LWP; mass-weighted vertical integration of cloud water and rainwater), ice water path (IWP; integration of cloud ice, snow and graupel), and total water path (TWP; combining both LWP and IWP) in Fig. 12. At the first EnKF cycle at 1900 UTC, the domain-averaged LWP is reduced by about 60%, while ice water path is reduced by
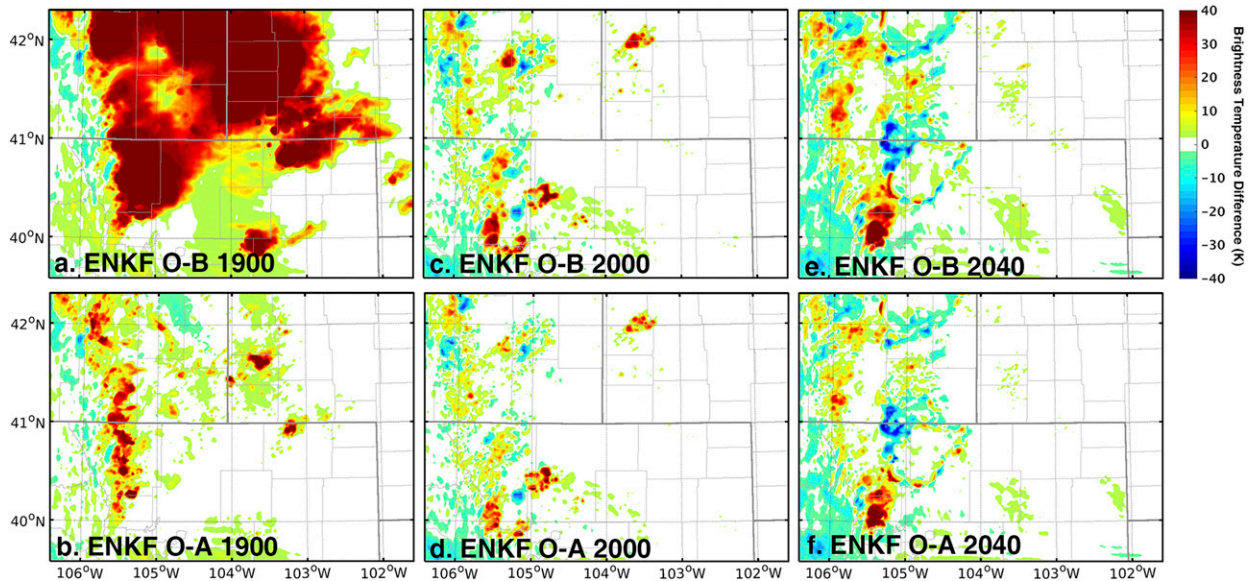

FIG. 11. Differences between observations and simulated BT of ABI channel 10 (observation minus EnKF) of (a),(c),(e) prior mean and (b),(d),(f) posterior mean of EnKF at (left) 1900, (middle) 2000, and (right) 2040 UTC.
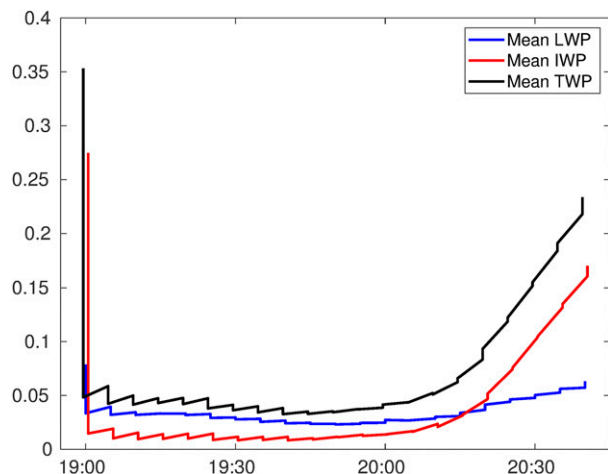
FIG. 12. Evolution of domain-averaged LWP (blue), IWP (red), and TWP (black) during EnKF.

almost 95%. With the attempted formation of new clouds or recovery of the EnKF removed clouds, there are some slight increases in both water and ice contents during the subsequent 5-min free forecast period, but the reduction with the BT assimilation at the next cycle is bigger than such increases due to cloud formation or recovery. Such increase in LWP and IWP became little in the free forecast periods after 1930 UTC, indicating continuous cycling BT assimilating cannot only remove spurious clouds, but also make them harder to form. Subsequently, after 2000 UTC, both IWP and LWP show persistent increases during both ensemble forecasts and EnKF, but the increase of LWP is much slower than that of IWP. It can also be seen from changes of IWP and LWP during each EnKF cycle throughout the entire experiment that assimilating BT observations might have a greater impact on ice hydrometeors than liquid hydrometeors, which is consistent with previous studies of Zupanski et al. (2011) that updating ice particles when assimilating BT generally have greater improvements than updating liquid water particles. The different impact of BT on liquid versus ice water particles might suggest ways to further improve the performance of BT assimilation in the future.

Assimilating infrared BT observations using EnKF can influence mesoscale environmental conditions as well. However, due to the generally weak correlations between BT and atmospheric state variables, the influence is limited for this event and does not pose statistically significant improvement compared with NODA (figure not shown). The combined impact of simultaneous assimilation of satellite and conventional observations will be explored in future studies.

5. Improvements in forecasts

a. Deterministic forecasts

Deterministic forecasts are carried out from the 1930, 2000, and 2040 UTC EnKF analysis mean (referred to as Fcst1930, Fcst2000, and Fcst2040, respectively). Simulated composite reflectivity and ABI channel 10 BT of these three experiments are shown in Figs. 13 and 14. All three deterministic forecasts share some similar characteristics, yet also are different from each other in the detailed development of the storms. For example, they all have storm initiation along the Rocky Mountain slopes and are located similarly as of 2100 UTC (Figs. 13a,e,i); however, Fcst1930 has two strong individual storms (Fig. 13a), while in Fcst2000, the southern storm just to the north of 40°N is significantly weaker than the other storm closer to the Wyoming–Colorado border (Fig. 13e), and in Fcst2040 multiple storms are initiated (Fig. 13i). The strength and juxtaposition of early storms also influences later development, as detailed by Y. Zhang et al. (2016): at 2200 UTC, there is an intense and large storm just across the Wyoming–Colorado border in Fcst2040 accompanied by several weaker ones to the northwest (Fig. 13j), while in Fcst2000, a cluster of strong storms appeared (Fig. 13f), and for Fcst1930, the major storm is located farther southwest with a lower translation speed (Fig. 13b). The major storm in Fcst1930 later triggered new storms ahead (Fig. 13c) and dissipated gradually (Fig. 13d). For Fcst2000, the storm cluster also triggered new storms ahead (Fig. 13g), which were maintained (Fig. 13h), while for Fcst2040, the strong storm was maintained with fewer new storms initiated (Figs. 13k,l). It can also be seen that simulated BT of Fcst2040 at 2100 UTC (Fig. 14i) is closest to the observed storms (Fig. 2d), compared with the other two; however, it is hard to infer storm development and organization beneath cloud tops from satellite images.

The ETS of these three deterministic forecasts are also calculated and presented in Fig. 9. A sharp decrease of ETS occurs after the initialization of each forecast, resulting from the model adjustment of imbalances in EnKF analysis, as well as subsequent adjustment of hydrometeors according to the thermodynamic conditions. The ETS of the general precipitative region (20 dBZ) show no distinct differences among the three forecasts (Fig. 9a). For the convective region (40 dBZ; Fig. 9b), Fcst2000 has the largest values around 2200 UTC, and Fcst1930 has the smallest values during this period. This results from a better collocation of the 40-dBZ regions in Fcst2000, as compared with observations (Fig. 13f), than the other two forecasts (Figs. 13b,j). Although the ETS of all three forecasts for 40-dBZ reflectivity decreases after 2200 UTC, ETS of Fcst2040 increases again after 2300 UTC, owing to a slightly
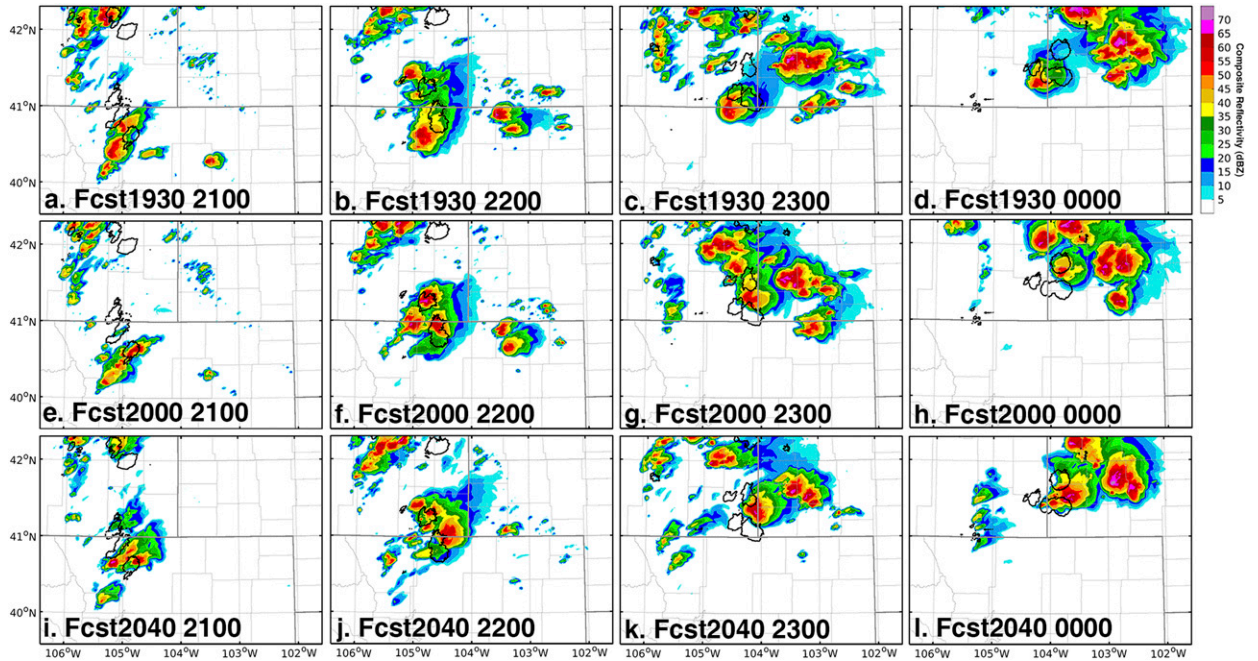
FIG. 13. Simulated composite reflectivity of (a)–(d) Fcst1930, (e)–(h) Fcst2000, and (i)–(l) Fcst2040 at (first column) 2100, (second column) 2200, and (third column) 2300 UTC 12 Jun, and (fourth column) 0000 UTC 13 Jun 2017. Black contours in all panels are observed 40-dBZ composite reflectivity.

better forecast of the storm interactions. However, it should be noted that 40-dBZ ETS values are almost constantly smaller than 0.3 for all forecasts, and none of them can be regarded as skillful, although all three

experiments show much better scores than DETER before 2200 UTC. It is also worth noting that although the ETS for the general precipitative region of all three forecasts show persistent increases during the entire
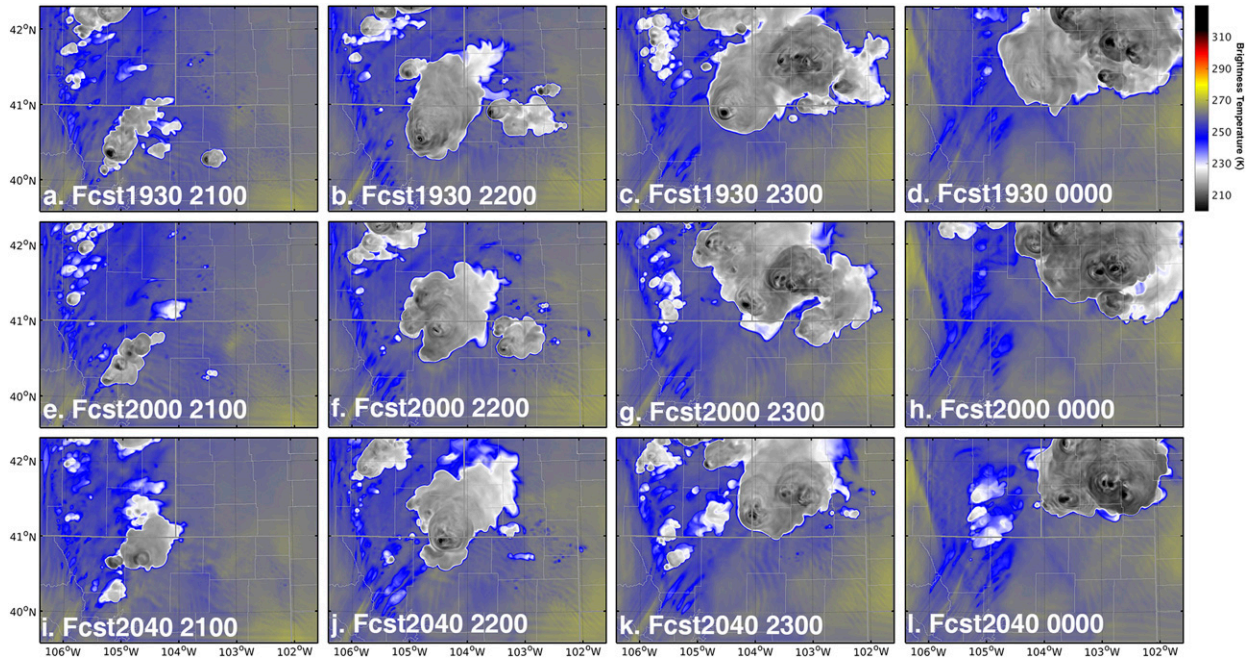

FIG. 14. Simulated BT of ABI channel 10 of (a)–(d) Fcst1930, (e)–(h) Fcst2000, and (i)–(l) Fcst2040 at (first column) 2100, (second column) 2200, and (third column) 2300 UTC 12 Jun, and (fourth column) 0000 UTC 13 Jun 2017.
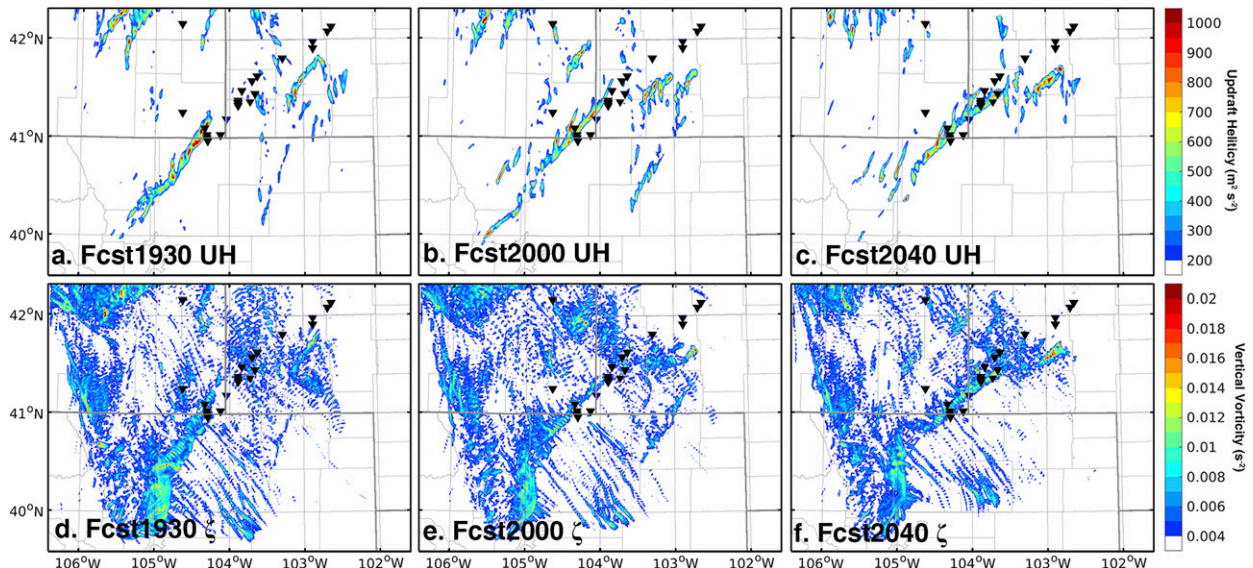
FIG. 15. (a)–(c) Maximum of 2–5-km UH and (d)–(f) maximum of 0–1-km maximum vertical vorticity of (left) Fcst1930, (middle) Fcst2000, and (right) Fcst2040. Black triangles indicate tornado reports (as in Fig. 1).

forecast period (Fig. 9a), in the convective regions the ETS values peak around 2200 UTC and decrease afterward (except for Fcst2040, as stated earlier; Fig. 9b), indicating different predictability characteristics for the two kinds of precipitation structures.

Although composite reflectivity of the three forecasts is similar both qualitatively and quantitatively, their predictions of mesocyclone and low-level tornado vortex signatures were significantly different, with a clear improvement seen when more satellite observations are assimilated. The major UH tracks in Fcst1930 (Fig. 15a) are located to the southwest of the observed tornado reports, and there is only a slight signature of near-surface rotation along the UH tracks (Fig. 15b). Some improvement in the location of the UH tracks occurs in Fcst2000 (Fig. 15c), although there are still no continuous low-level vortex tracks in this deterministic forecast (Fig. 15d). The UH tracks are further improved in Fcst2040 (Fig. 15e), becoming stronger and wider and tracking farther into Nebraska as observed for the tornadoes, and there is a clear track of low-level vortices overlaid on top of the same UH track (Fig. 15f), indicating the existence of a strong near-surface vortex signature beneath the midlevel mesocyclone in this specific prediction. These results indicate that a better prediction of the tornadoes could be achieved with more cycles of infrared BT assimilation for this event.

b. Ensemble forecasts

The probabilistic forecasts of this event from the EnKF analysis are also examined. The ensemble probabilities of

ensemble forecasts show some similarity to the deterministic forecasts from the EnKF analysis at the same time. For EF1930, high probabilities at 2100 UTC (Fig. 16a) show a similar shape as the storms in Fcst1930 at the same time (Fig. 13a), but become dislocated from observed storms at 2300 UTC (Fig. 16c). The dislocation of higher probability becomes smaller in EF2000 at both 2100 (Fig. 16e) and 2200 UTC (Fig. 16f), and the southern high probability region in EF2000 at 2100 UTC further shrinks in EF2040 (Fig. 16i). The spurious storms in northeast Colorado in EF1930 and EF2000 are also much weaker in EF2040 both at 2100 (Fig. 16i) and 2200 UTC (Fig. 16j). Furthermore, EF1930 and EF2000 trigger new storms ahead of the old ones at 2300 UTC (Figs. 16c,g), and the old storms dissipate gradually (Figs. 16d,h); although EF2040 experiences similar processes at 2300 UTC (Fig. 16k), old storms in EF2040 did not undergo rapid dissipation, better matching with the observations.

Similar to the results from the deterministic forecasts, the ETS of composite reflectivity for EF1900 and EF2000 also decrease significantly after initialization (Fig. 9a), but the magnitudes are smaller. For the general precipitation region (Fig. 9a), EF1930 and EF2000 generally have similar values throughout the ensemble forecasts, while EF2040 has higher values before about 2130 UTC. On the other hand, for the convective precipitation region (Fig. 9b), although ETS for the three ensemble forecasts are similar before 2100 UTC, EF2040 has smaller ETS values than the other two ensembles after 2100 UTC, owing to its smaller spread in the prediction of storm locations (Figs. 16k,l). Given that
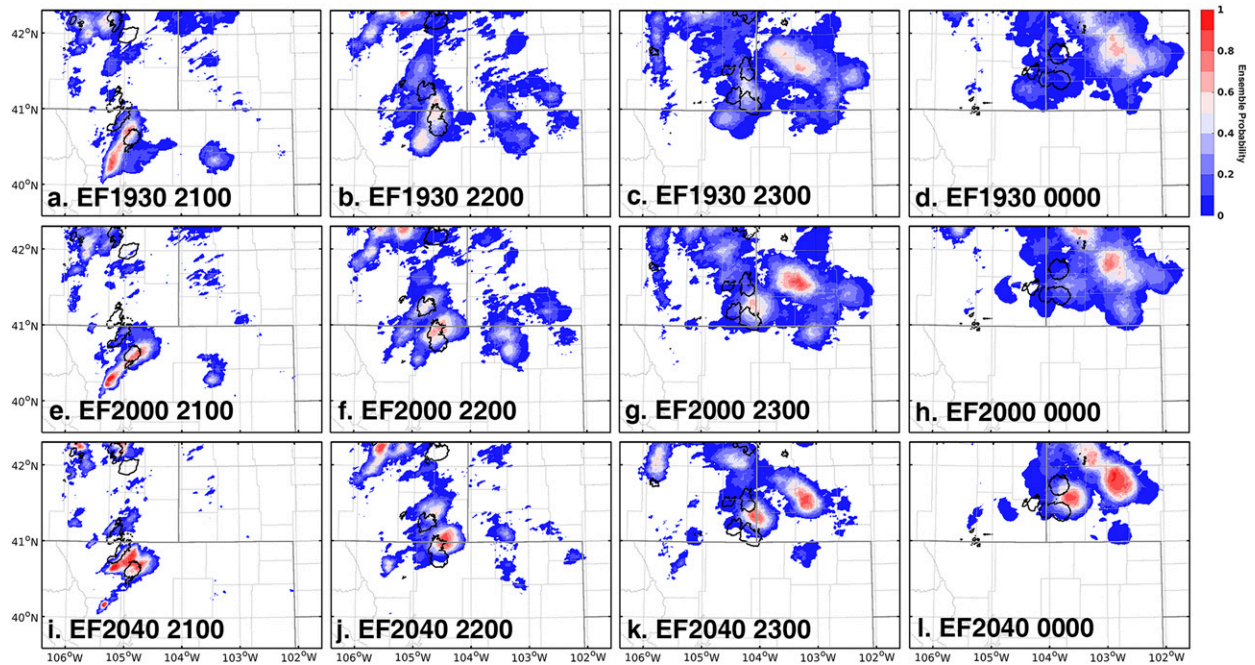
FIG. 16. Ensemble probability of 40-dBZ simulated composite reflectivity of (a)–(d) EF1930, (e)–(h) EF2000, and (i)–(l) EF2040 at (first column) 2100, (second column) 2200, and (third column) 2300 UTC 12 Jun, and (fourth column) 0000 UTC 13 Jun 2017. Black contours in all panels are observed 40-dBZ composite reflectivity.

satellite infrared imagers provide information only above cloud tops and are unable to provide information on storm structures beneath the cloud tops, these results are not surprising. All three ensemble forecasts show higher ETS scores than NODA in precipitative regions, although similar to ETS of deterministic forecasts, and 40-dBZ ETS of all ensemble forecasts remain below 0.25.

Despite similar quantitative scores for the simulated composite reflectivity of the three ensemble forecasts, a persistent improvement on the probabilistic forecasts of mesocyclone and low-level vortex signature can be seen when more BT observations are assimilated. Probabilistic forecast of UH of EF1930 (Fig. 17a) is primarily located in Colorado. Probabilistic forecast of UH of EF2040 (Fig. 17c) has a similar track to that of EF2000 (Fig. 17b), but EF2040 has consistently higher probabilities, and the track is better defined and more compact. Probabilistic track of vertical vorticity of EF2040 (Fig. 17f) is also much longer than the other two ensemble forecasts (Figs. 17b,d), extending farther into Nebraska. These improvements, especially the higher and more compact probabilities, indicate a higher confidence in the probabilistic predictions of midlevel mesocyclones and near-surface rotation in EF2040 than EF1930 and EF2000.

In summary, both deterministic forecasts and ensemble forecasts from EnKF analysis show that prediction of

the midlevel mesocyclone and low-level rotation can be improved when satellite infrared radiance observations are assimilated using EnKF, although the quantitative measurements of simulated composite reflectivity of forecasts from the EnKF analysis at different times are similar.

6. Summary

Using an EnKF data assimilation system coupled with the WRF-ARW Model running at a convection-allowing resolution, this study presents the first attempt to assimilate real-world all-sky infrared BT observations of the ABI onboard the *GOES-16* satellite for a tornadic thunderstorm event. The analysis and prediction of the severe convective storms and associated severe weather signatures in Colorado, Wyoming, and Nebraska on 12 June 2017, with and without BT data assimilation, are examined.

Observation–space diagnostics including RMS innovation and fit, bias, ensemble spread, and rank histogram indicate that the EnKF system works properly and maintains a reasonable variance among ensemble members. Errors at the beginning of the EnKF cycles at 1900 UTC result from spreading spurious clouds that are rapidly reduced within the first several cycles when hydrometeors, especially ice-phased particles, are effectively removed and spurious clouds suppressed.
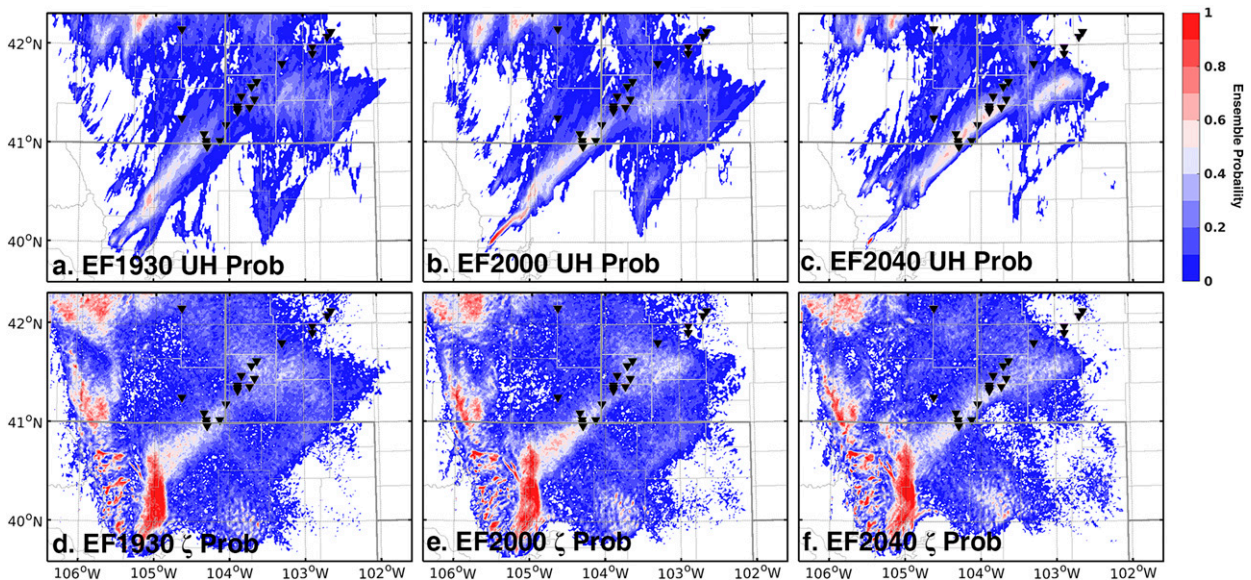
FIG. 17. Ensemble probability of (a)–(c) 180 m^2 s^{-2} 2–5-km UH and (d)–(f) 0.004 s^{-1} 0–1-km maximum vertical vorticity of (left) EF1930, (middle) EF2000, and (right) EF2040. Black triangles indicate tornado reports (as in Fig. 1).

The magnitude of errors, as well as the amount of water condensates, remains at relatively low values before observed convection initiation at approximately 2000 UTC. After the storms initiate, errors show a slight increase but still remain below 6 K, indicating the effectiveness of both removing and generating clouds of our techniques. Nonetheless, ensemble ETS of stratiform and convective precipitation regions indicate that the EnKF assimilation of ABI BT observations provides a better estimation of the storms than the NODA ensemble forecast that did not include data assimilation.

Both deterministic and ensemble forecasts are carried out from the 1930, 2000, and 2040 UTC EnKF analysis. ETS values indicate that both the general precipitative region (composite reflectivity greater than 20 dBZ) and the convective precipitation region (composite reflectivity greater than 40 dBZ) forecasts from different EnKF analyses generally have similar skill. It is worth noting that none of the forecasts have 40-dBZ ETS above 0.3. Although more EnKF cycles do not improve deterministic or probabilistic forecast skill of the storms in terms of composite reflectivity, there is a substantial improvement of the prediction of midlevel mesocyclone and low-level vortex signatures when more BT observations are assimilated. Deterministic forecasts from 1930 and 2000 UTC EnKF analysis produce curvy UH tracks and unorganized short low-level vertical vorticity tracks at the wrong locations, and corresponding ensemble probabilistic forecasts of UH and vertical vorticity are also short, ending before the storms enter Nebraska. In contrast, a well-defined low-level vertical vorticity track accompanying a straight and long-lasting UH track is generated by the deterministic forecast from 2040 UTC EnKF analysis, and the probabilistic forecasts of both UH and low-level vertical vorticity from 2040 UTC EnKF analysis are also better defined and longer lasting than ensemble forecasts from the other two times, indicating improvement on the structure and the development of the storms. Generally speaking, greater confidence on the prediction of low-level rotation as well as other severe weather signatures associated with thunderstorms can be gained when more BT observations are assimilated.

The results presented in this study indicate that assimilating ABI BT observations using EnKF techniques has the potential to improve the prediction of severe thunderstorms that may lead to improvements in the operational warning processes for severe weather. Spaceborne instruments provide unique observations that complement those from ground-based Doppler weather radars and might be especially useful during convection initiation when assimilated in combination with radar observations; the simultaneous assimilation of radar, satellite, and conventional observations for severe thunderstorm events will be explored in the future. Besides, how to better utilize infrared BT observations to extract more information—including but not limited to bias correction, vertical and horizontal localization, simultaneous assimilation of multiple channels, and treatment of spatially correlated observational errors—will also need further exploration.

REFERENCES

Aksoy, A., D. C. Dowell, and C. A. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, https://doi.org/10.1175/2008MWR2691.1.

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530, https://doi.org/10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.

Bessho, K., and Coauthors, 2016: An introduction to *Himawari-8/9*—Japan's new-generation geostationary meteorological satellites. *J. Meteor. Soc. Japan*, **94**, 151–183, https://doi.org/10.2151/jmsj.2016-009.

Cintineo, R. M., J. A. Otkin, T. A. Jones, S. Koch, and D. J. Stensrud, 2016: Assimilation of synthetic GOES-R ABI infrared brightness temperatures and WSR-88D radar observations in a high-resolution OSSE. *Mon. Wea. Rev.*, **144**, 3159–3180, https://doi.org/10.1175/MWR-D-15-0366.1.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, https://doi.org/10.1175/BAMS-D-11-00040.1.

Dowell, D. C., L. J. Wicker, 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.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, https://doi.org/10.1029/2002JD003296.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

——, and S. J. Colucci, 1996: Random and systematic error in NMC's short-range Eta ensembles. Preprints, *13th Conf. on Probability and Statistics in the Atmospheric Sciences*, San Francisco, CA, Amer. Meteor. Soc., 51–56.

——, and ——, 1997: Verification of Eta–RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2.

Han, Y., P. van Delst, Q. Liu, F. Weng, B. Yan, R. Treadon, and J. Derber, 2006: JCSDA Community Radiative Transfer Model (CRTM)—version 1. NOAA Tech. Rep. NESDIS 122, 40 pp.

Harrison, M. S. J., D. S. Richardson, K. Robertson, and A. Woodcock, 1995: Medium-range ensembles using both the ECMWF T63 and unified models—An initial report. UKMO Tech. Rep. 153, 25 pp.

Hastings, R., and Y. Richardson, 2016: Long-term morphological changes in simulated supercells following mergers with nascent supercells in directionally varying shear. *Mon. Wea. Rev.*, **144**, 471–499, https://doi.org/10.1175/MWR-D-15-0193.1.

Heidinger, A., C. O'Dell, R. Bennartz, and T. Greenwald, 2006: The successive-order-of-interaction radiative transfer model. Part I: Model development. *J. Appl. Meteor. Climatol.*, **45**, 1388–1402, https://doi.org/10.1175/JAM2387.1.

Honda, T., and T. Kawano, 2016: A possible mechanism of tornadogenesis associated with the interaction between a supercell and an outflow boundary without horizontal shear. *J. Atmos. Sci.*, **73**, 1273–1292, https://doi.org/10.1175/JAS-D-14-0347.1.

——, and Coauthors, 2018a: Assimilating all-sky *Himawari-8* infrared radiances: A case of Typhoon Soudelor (2015). *Mon. Wea. Rev.*, **146**, 213–229, https://doi.org/10.1175/MWR-D-16-0357.1.

——, S. Kotsuki, G.-Y. Lien, Y. Maejima, K. Okamoto, and T. Miyoshi, 2018b: Assimilation of *Himawari-8* all-sky radiances every 10 minutes: Impact on precipitation and flood risk prediction. *J. Geophys. Res. Atmos.*, **123**, 965–976, https://doi.org/10.1002/2017JD027096.

Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

Iacono, M., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

——, 1996: The surface layer in the NCEP Eta model. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 354–355.

Jones, T. A., and D. J. Stensrud, 2015: Assimilating cloud water path as a function of model cloud microphysics in an idealized simulation. *Mon. Wea. Rev.*, **143**, 2052–2081, https://doi.org/10.1175/MWR-D-14-00266.1.

——, J. A. Otkin, D. J. Stensrud, and K. Knopfmeier, 2013: Assimilation of satellite infrared radiances and Doppler radar observations during a cool season observing system simulation experiment. *Mon. Wea. Rev.*, **141**, 3273–3299, https://doi.org/10.1175/MWR-D-12-00267.1.

——, ——, ——, and ——, 2014: Forecast evaluation of an observing system simulation experiment assimilating both radar and satellite data. *Mon. Wea. Rev.*, **142**, 107–124, https://doi.org/10.1175/MWR-D-13-00151.1.

——, D. Stensrud, L. Wicker, P. Minnis, and R. Palikonda, 2015: Simultaneous radar and satellite data storm-scale assimilation using an ensemble Kalman filter approach for 24 May 2011. *Mon. Wea. Rev.*, **143**, 165–194, https://doi.org/10.1175/MWR-D-14-00180.1.

——, 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.

——, S. Koch, and Z. Li, 2017: Assimilating synthetic hyperspectral sounder temperature and humidity retrievals to improve severe weather forecasts. *Atmos. Res.*, **186**, 9–25, https://doi.org/10.1016/j.atmosres.2016.11.004.

——, 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.

Kerr, C. A., D. J. Stensrud, and X. Wang, 2015: Assimilation of cloud-top temperature and radar observations of an idealized splitting supercell using an observing system simulation experiment. *Mon. Wea. Rev.*, **143**, 1018–1034, https://doi.org/10.1175/MWR-D-14-00146.1.

Meng, Z., and F. Zhang, 2008a: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Mon. Wea. Rev.*, **136**, 522–540, https://doi.org/10.1175/2007MWR2106.1.

——, and ——, 2008b: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Wea. Rev.*, **136**, 3671–3682, https://doi.org/10.1175/2008MWR2270.1.

Minamide, M., and F. Zhang, 2017: Adaptive observation error inflation for assimilating all-sky satellite radiance. *Mon. Wea. Rev.*, **145**, 1063–1081, https://doi.org/10.1175/MWR-D-16-0257.1.

——, and ——, 2018a: 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.

NCDC, 2018: Billion-dollar weather and climate disasters: Summary stats. NOAA/National Centers for Environmental Information, accessed January 2018, https://www.ncdc.noaa.gov/billions/summary-stats.

Otkin, J. A., 2010: Clear and cloudy sky infrared brightness temperature assimilation using an ensemble Kalman filter. *J. Geophys. Res.*, **115**, D19207, https://doi.org/10.1029/2009JD013759.

——, 2012: Assimilation of water vapor sensitive infrared brightness temperature observations during a high impact weather event. *J. Geophys. Res.*, **117**, D19203, https://doi.org/10.1029/2012JD017568.

Polger, P. D., B. S. Goldsmith, R. C. Przywarty, and J. R. Bocchieri, 1994: National Weather Service warning performance based on the WSR-88D. *Bull. Amer. Meteor. Soc.*, **75**, 203–214, https://doi.org/10.1175/1520-0477(1994)075<0203:NWSWPB>2.0.CO;2.

Schmit, T. J., M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier, 2005: Introducing the next-generation Advanced Baseline Imager on GOES-R. *Bull. Amer. Meteor. Soc.*, **86**, 1079–1096, https://doi.org/10.1175/BAMS-86-8-1079.

——, P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebair, 2017: A closer look at the ABI on the GOES-R series. *Bull. Amer. Meteor. Soc.*, **98**, 681–698, https://doi.org/10.1175/BAMS-D-15-00230.1.

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.

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.

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

Stuhlmann, R., A. Rodriguez, S. Tjemkes, J. Grandell, A. Arriaga, J.-L. Bézy, D. Aminou, and P. Bensi, 2005: Plans for EUMETSAT's Third Generation Meteosat geostationary satellite programme. *Adv. Space Res.*, **36**, 975–981, https://doi.org/10.1016/j.asr.2005.03.091.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability*, Shinfield Park, Reading, United Kingdom, ECMWF, 1–25.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

Weng, Y., and F. Zhang, 2016: Advances in convection-permitting tropical cyclone analysis and prediction through EnKF assimilation of reconnaissance aircraft observations. *J. Meteor. Soc. Japan*, **94**, 345–358, https://doi.org/10.2151/jmsj.2016-018.

Wheatley, D. M., 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.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.

Wurman, J. M., Y. P. Richardson, C. Alexander, S. Weygandt, and P. F. Zhang, 2007: Dual-Doppler and single-Doppler analysis of a tornadic storm undergoing mergers and repeated tornadogenesis. *Mon. Wea. Rev.*, **135**, 736–758, https://doi.org/10.1175/MWR3276.1.

——, D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The second Verification of the Origins of Rotation in Tornadoes Experiment: VORTEX2. *Bull. Amer. Meteor. Soc.*, **93**, 1147–1170, https://doi.org/10.1175/BAMS-D-11-00010.1.

Yang, J., Z. Zhang, C. Wei, F. Lu, and Q. Guo, 2017: Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4. *Bull. Amer. Meteor. Soc.*, **98**, 1637–1658, https://doi.org/10.1175/BAMS-D-16-0065.1.

Yokota, S., H. Seko, M. Kunii, H. Yamauchi, and H. Niino, 2016: The tornadic supercell on the Kanto Plain on 6 May 2012: Polarimetric radar and surface data assimilation with EnKF and ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **144**, 3133–3157, https://doi.org/10.1175/MWR-D-15-0365.1.

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, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

——, Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, https://doi.org/10.1175/2009MWR2645.1.

——, M. Minamide, and E. E. Clothiaux, 2016: Potential impacts of assimilating all-sky infrared satellite radiance from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophys. Res. Lett.*, **43**, 2954–2963, https://doi.org/10.1002/2016GL068468.

Zhang, M., F. Zhang, X.-Y. Huang, and X. Zhang, 2011: Intercomparison of an ensemble Kalman filter with three- and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003. *Mon. Wea. Rev.*, **139**, 566–572, https://doi.org/10.1175/2010MWR3610.1.

Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2016: Intrinsic predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma at storm scales. *Mon. Wea. Rev.*, **144**, 1273–1298, https://doi.org/10.1175/MWR-D-15-0105.1.

Zupanski, D., M. Zupanski, L. D. Grasso, R. Brummer, I. Jankov, D. Lindsey, M. Sengupta, and M. Demaria, 2011: Assimilating synthetic GOES-R radiances in cloudy conditions using an ensemble-based method. *Int. J. Remote Sens.*, **32**, 9637–9659, https://doi.org/10.1080/01431161.2011.572094.