Methods for Validating HRRR Simulated Cloud Properties for Different Weather Phenomena Using Satellite and Radar Observations

SARAH M. GRIFFIN,[a] JASON A. OTKIN,[a] AND WILLIAM E. LEWIS[a]

[a] *Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin*

(Manuscript received 23 June 2023, in final form 6 November 2023, accepted 9 November 2023)

ABSTRACT: In this study, we evaluate the ability of the High-Resolution Rapid Refresh (HRRR) model to forecast cloud characteristics through comparison of observed and simulated satellite brightness temperatures (BTs) and radar reflectivity during different weather phenomena in December 2021: the Mayfield, Kentucky, tornado on 11 December, a heavy snow event in Minnesota from 10 to 11 December, and the Midwest derecho on 15 December. This is done to illustrate the importance of examining model accuracy across a range of weather phenomena. Observation and forecast objects were created using the Method for Object-Based Diagnostic Evaluation (MODE). HRRR accurately depicted the spatial displacements between observation cloud (defined using BTs) and radar reflectivity objects, namely, the centers of cloud objects are to the east of the radar objects for the tornado and derecho events, and generally west of the radar objects for the snow event. However, HRRR had higher (less intense) simulated BTs and higher (more intense) radar reflectivity than the observations for the tornado event. Simulated radar reflectivity is higher and BTs are lower than the observations during the middle of the snow event. Also, simulated radar reflectivity is higher and BTs are lower than the observations during the derecho event. Of the three weather events, the HRRR forecasts are most accurate for the snow event, based on the object-based threat score, followed by the derecho and tornado events. The tornado event has lower accuracy because matches between paired simulated and observation objects are worse than for the snow event, with less similarity in size forecast objects and greater distance between paired object centers.

SIGNIFICANCE STATEMENT: The purpose of this study is to assess the accuracy of forecast cloud and radar objects, defined using simulated satellite brightness temperatures and radar reflectivity, from the High-Resolution Rapid Refresh (HRRR) model. This assessment was conducted for a tornado, snow, and derecho event from December 2021. Results from these three events indicate that the HRRR model accurately represents the observed displacement between the center of cloud and radar objects for the tornado and derecho events, and is the most accurate overall for the snow event.

KEYWORDS: Forecast verification/skill; Radars/Radar observations; Satellite observations; Clouds; Model evaluation/performance; Regional models

1. Introduction

Operational weather forecasters typically rely on numerical weather prediction (NWP) models when predicting future weather events. One of the most widely used regional NWP models is the Weather Research and Forecasting (WRF) Model (Powers et al. 2017). A version of the WRF Model, the High-Resolution Rapid Refresh (HRRR), was the first operational hourly updating convection-allowing model to provide guidance covering the contiguous United States (James et al. 2022). Therefore, many studies have evaluated the accuracy of HRRR forecasts (e.g., Ikeda et al. 2013, 2017; Dougherty et al. 2021; English et al. 2021; Griffin et al. 2017a,b; Pinto et al. 2015; Bytheway et al. 2017). Some studies, such as Griffin et al. (2017a,b), have used simulated satellite infrared (IR) brightness temperatures (BTs) to assess forecasts of clouds, whereas other studies (Duda and Turner 2021; James et al. 2022) have specifically evaluated simulated HRRR radar reflectivity.

One method for evaluating NWP models is through object-based verification. Object-based methods provide a powerful way to assess forecast accuracy because traditional metrics typically penalize forecasts for displacement errors between the forecast and observation objects. Therefore, many studies have used object-based methods to assess forecast accuracy (Griffin et al. 2017a,b, 2020; Jones et al. 2018, 2020; Senf et al. 2018; Skinner et al. 2016, 2018; Henderson et al. 2021). In this study, objects are identified using the Method for Object-Based Diagnostic Evaluation (MODE; Davis et al. 2006a,b, 2009; Bullock et al. 2016). MODE can be used to analyze attributes of objects, such as their size and location, as well as spatial displacement errors between paired forecast and observation objects. Many studies have used MODE to assess forecast skill, such as analyzing convective precipitation forecasts (Bytheway and Kummerow 2015; Cai and Dumais 2015; Ji et al. 2020; Wolff et al. 2014), intense snowfall (Radford et al. 2019; Ganetis et al. 2018), atmospheric rivers (DeHaan et al. 2021), and cold pools (Squitieri and Gallus 2020).

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-23-0109.s1.

Corresponding author: Sarah M. Griffin, sarah.griffin@ssec.wisc.edu

DOI: 10.1175/WAF-D-23-0109.1

The purpose of this study is to develop methods for validating the HRRR Model's depiction of cloud characteristics. This is done by comparing observed and simulated satellite BTs and composite radar reflectivity over three different weather phenomena events that occurred over a one-week period in December 2021: the Mayfield, Kentucky, tornado and associated severe weather outbreak on 11 December 2021 (https://www.weather.gov/pah/December-10th-11th-2021-Tornado), heavy snow in Minnesota from 10 to 11 December 2021 (https://www.weather.gov/mpx/Dec10WinterStorm), and the Midwest derecho on 15 December 2021 (https://www.weather.gov/dmx/StormyandWindyWednesdayDecember152021). Simulated BTs can be a proxy for cloud cover when assessing forecast accuracy (Otkin et al. 2009; Cintineo et al. 2014; Thompson et al. 2016; Griffin et al. 2017a,b, 2021; Bikos et al. 2012; Van Weverberg et al. 2013), while simulated composite radar reflectivity is calculated based on modeled cloud hydrometeor populations (James et al. 2022). Therefore, this study, though limited in time, differs from prior studies because it assesses and compares the HRRR forecast accuracy for both convective weather features (severe weather and derecho) and more synoptically driven features (heavy snowfall).

While satellite BTs and radar reflectivity measure different meteorological processes, another purpose of this study is to examine connections between the HRRR simulated radar reflectivity and infrared BTs to identify if satellite imagery can be a proxy for radar reflectivity. This purpose also differs from other studies, which only used one model field for verification. Our reasoning for comparing satellite BTs and radar reflectivity is that observed radar reflectivity is not readily available everywhere, most notably over the oceans and in radar coverage gaps. Satellite BTs (both observed and simulated) and simulated radar reflectivity, though, can be globally available. By analyzing and comparing the connections between observed and simulated satellite BTs and radar reflectivity, one can hopefully infer the accuracy of simulated radar reflectivity when no radar observations are available.

This paper is organized as follows. The data used in this analysis are described in section 2 while the methodology is presented in section 3. Section 4 describes the results of this study and section 5 presents a discussion and conclusions.

2. Data

The starting and ending analysis times for each event in this study can be seen in Table 1. A display of HRRR forecast initializations and forecast hours can be found in Fig. 1.

TABLE 1. Start and end analysis times for each event.

Event	Start time	End time
Tornado	0000 UTC 11 Dec 2021	0900 UTC 11 Dec 2021
Snow	1200 UTC 10 Dec 2021	1200 UTC 11 Dec 2021
Derecho	1800 UTC 15 Dec 2021	0600 UTC 16 Dec 2021

a. *Infrared brightness temperatures*

1) OBSERVED BRIGHTNESS TEMPERATURES

This study uses observed 10.3-μm BTs from the *Geostationary Operational Environmental Satellite 16* (*GOES-16*) Advanced Baseline Imager (ABI) sensor. As these BTs are sensitive to cloud ice particles, they are an ideal dataset to assess the accuracy of the cloud field in the upper troposphere. These BTs, which have a 2-km pixel spacing at nadir, are calculated by parallax correcting the observed *GOES-16* ABI radiances using the *GOES-16* Level 2 Cloud Top Height Product (ACHAC). This cloud height product is available at 10-km spatial resolution and is remapped to the 2-km GOES grid for parallax correction. The observed radiances are then remapped to the HRRR grid using an area-weighted average of all the observed pixels overlapping a given grid box. Once that is completed, the radiances are converted to BTs.

2) SIMULATED BRIGHTNESS TEMPERATURES

The simulated BTs used in this study are the simulated *GOES-13* 10.7-μm BTs from the real-time HRRR version 4 model run at the NOAA Global Systems Laboratory. HRRR uses the Thompson aerosol-aware version 3.9.1 microphysics scheme, which includes mixing ratios of cloud water, rainwater, cloud ice, snow, and graupel, as well as the number concentration for cloud ice (Thompson and Eidhammer 2014). HRRR also uses the Mellor–Yamada–Nakanishi–Niino (MYNN) version 3.6+ planetary boundary layer scheme (Nakanishi and Niino 2004, 2009); RUC, version 3.6+ land surface model (Smirnova et al. 2016); and the Rapid Radiative Transfer Model for general circulation models shortwave and longwave radiation schemes (Iacono et al. 2008). Simulated BTs are computed using version 2.0.7 of the Community Radiative Transfer Model (Han et al. 2006) in the Unified Post Processor, which incorporates the *GOES-13* viewing-angle geometry. Output from the HRRR Model used to calculate simulated BTs includes vertical profiles of temperature, specific humidity, and cloud hydrometeor mixing ratios and particle effective diameters (Ding et al. 2011; Thompson et al. 2016).

To account for slight differences in the spectral characteristics of the observed *GOES-16* 10.3-μm and simulated *GOES-13* 10.7-μm BTs used during this study, observed *GOES-16* 10.3-μm BTs and observed *GOES-13* 10.7 μm BTs from 1 November 2017 to 31 December 2017 were compared, as both satellites were providing imagery over the United States. For this study, we are interested in upper-level clouds and have therefore chosen BT thresholds of 210, 225, and 235 K for the *GOES-13* BTs when identifying the edges of cloud objects (determining these thresholds will be discussed in section 4a). Percentile matching of the observed *GOES-16* and *GOES-13* BT distributions over the 1 November–31 December 2017 time period was then used to determine the corresponding *GOES-13* 10.7-μm BT for each *GOES-16* 10.3-μm BT threshold (Table 2). It can be seen that only small differences (<1 K) are present between the 10.3- and 10.7-μm BTs, which provides confidence that we can obtain robust results when comparing simulated *GOES-13* 10.7-μm BTs from the HRRR to observed *GOES-16* 10.3-μm BTs. The *GOES-16* and *GOES-13* BT thresholds in

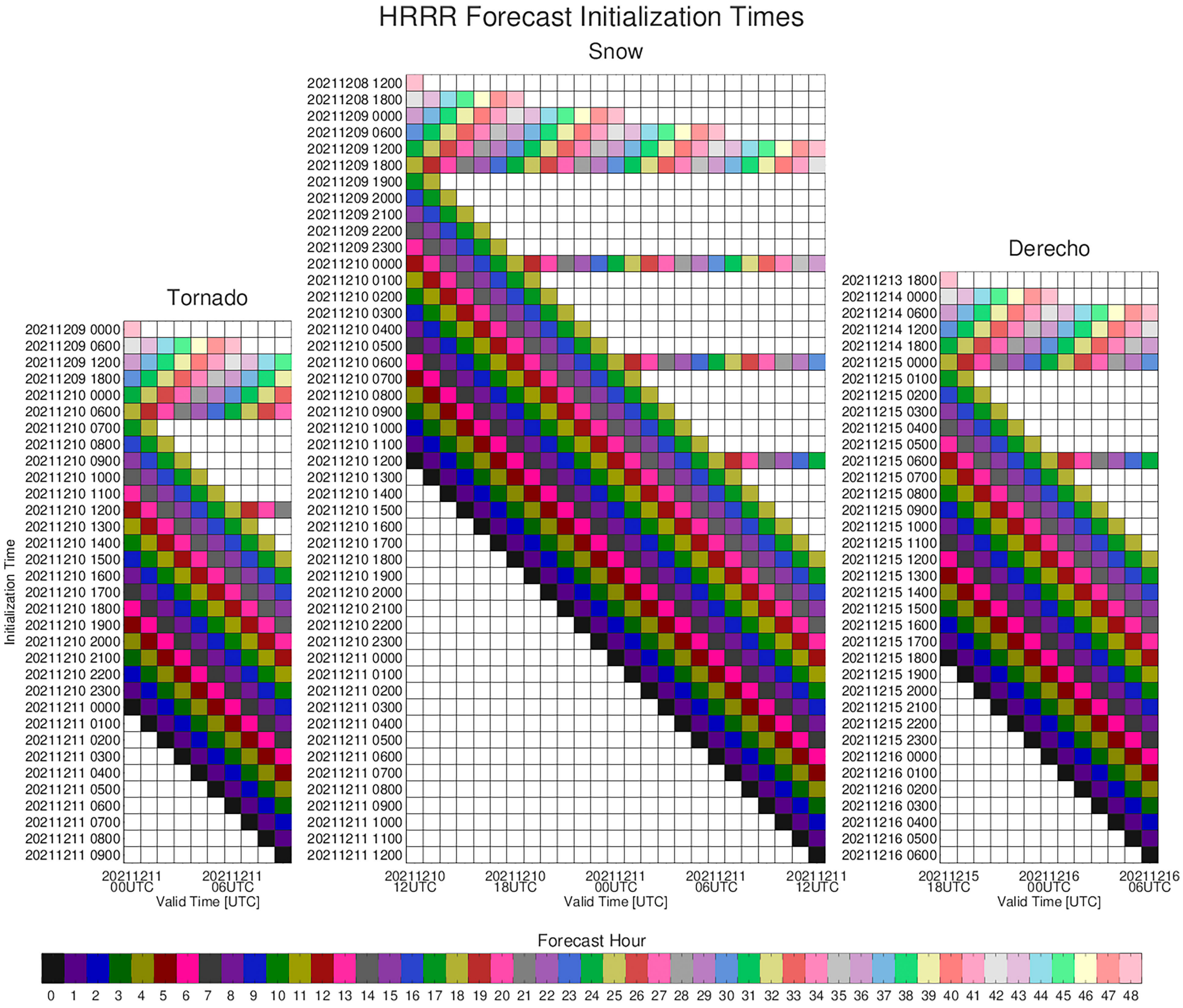


FIG. 1. HRRR forecast initializations and forecast hours used for each event in this study.

Table 2 are used to identify the cloud objects in the observations and HRRR forecasts, respectively.

b. Radar reflectivity

1) OBSERVED RADAR REFLECTIVITY

The observed radar reflectivity used in this study is the composite reflectivity product from the Multi-Radar Multi-Sensor (MRMS) project (Smith et al. 2016; Zhang et al. 2016), which is constructed from a regional mosaic of reflectivity values from the WSR-88D radars around the United States and Canada. MRMS data from the time closest to the top of each hour are used. MRMS data from the 0.01° latitude–longitude grid is remapped to the 3-km resolution HRRR grid using an area-weighted average of all the observed pixels overlapping a given grid box.

TABLE 2. Corresponding 10.7-μm BT from *GOES-13* to *GOES-16* 10.3-μm BT thresholds used in this analysis. Simulated HRRR BTs have a wavelength of 10.7 μm.

GOES-16	10.3 μm	210 K	225 K	235 K
GOES-13	10.7 μm	210.52 K	224.26 K	234.09 K

2) SIMULATED RADAR REFLECTIVITY

This study uses the composite radar reflectivity field from the HRRR Model, which is the maximum reflectivity in the column. Simulated composite radar reflectivity is calculated based on model microphysical hydrometeor populations (James et al. 2022) using the Rayleigh approximation, and assumes a 10-cm wavelength. The Thompson microphysics includes a bright-band enhancement, where melted snow/graupel is treated as water-coated ice spheres using routines from Blahak (2012). Hereafter, composite radar reflectivity will simply be called radar reflectivity.

3. Methodology

Satellite objects, which will be determined based on BTs and are a proxy for cloud objects, and radar reflectivity

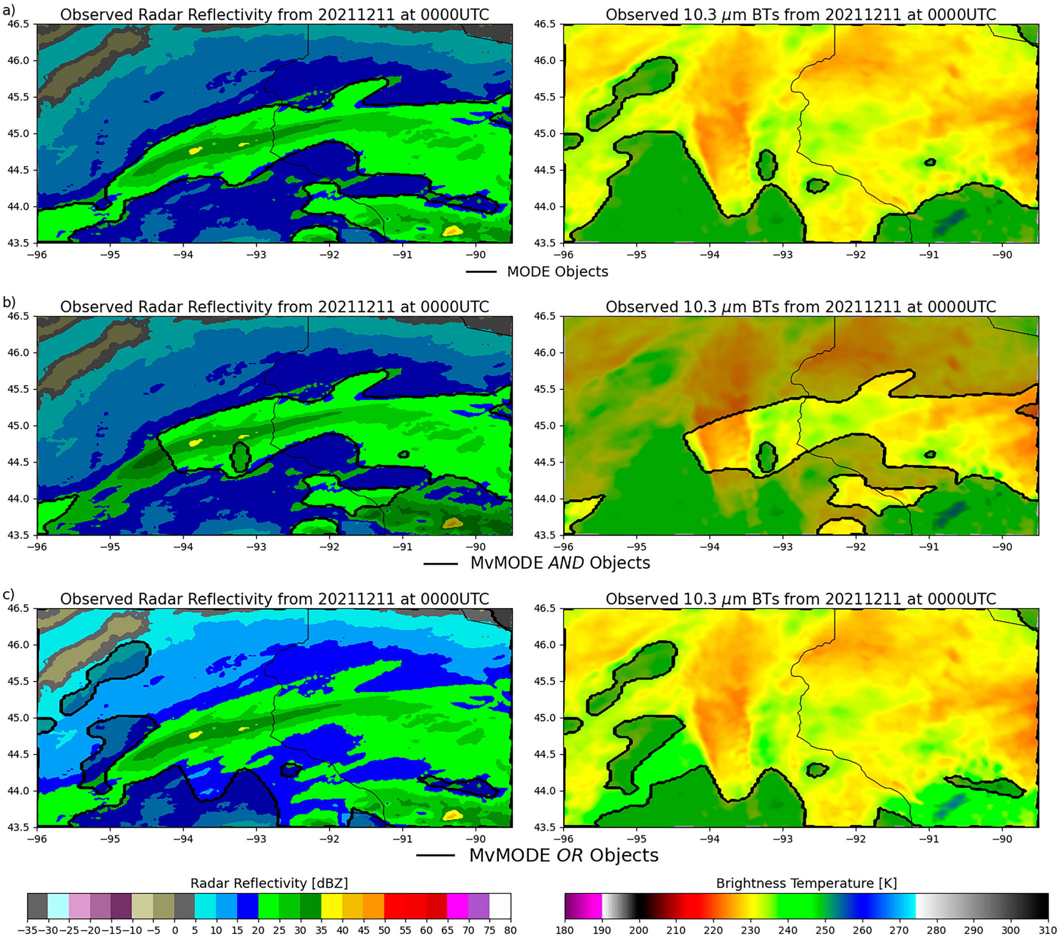


FIG. 2. Example of (a) MODE radar reflectivity and BT objects, (b) MvMODE objects where both radar reflectivity AND BTs exceed their given thresholds, and (c) MvMODE objects where both radar reflectivity OR BTs exceed their given thresholds. Object areas on each panel are unshaded.

objects will be identified using MODE (Davis et al. 2006a,b, 2009) and Multivariate MODE (MvMODE; Blank et al. 2021). While MODE uses a single field when identifying objects, MvMODE allows super objects to be defined using multiple fields. For example, MvMODE can identify areas where both radar reflectivity "AND" BTs exceed their given thresholds. These AND object areas are unshaded in Fig. 2b, and are much smaller than the MODE objects (unshaded areas in Fig. 2a). Unlike the MODE objects (Fig. 2a), the MvMODE objects are also the same for radar reflectivity and BTs. MvMODE objects can also be defined as areas where either radar reflectivity "OR" BTs exceed their given thresholds, and are unshaded in Fig. 2c. These OR objects are much larger than the MODE objects and MvMODE AND objects. Objects are identified in MODE and MvMODE using a process called convolutional thresholding. During this process, BT and radar reflectivity fields are smoothed using a convolution radius of 5 grid points (15 km). Therefore, this radius allows for the analysis of both large and small-scale objects, since a range of 2–8 grid points is necessary to identify resolved convective storm objects in the model forecasts.

Once objects are identified, various object attributes, such as location and size, are computed for each object. Then, an interest score is calculated to ascertain the similarity of the paired observation and forecast objects (Bullock et al. 2016; Jensen et al. 2020). Interest scores ranges in value from 0 to 1 with 1 being a perfect match. The interest score for an object pair is the weighted sum of the values of object pair attributes. Each object pair attribute is given a value between 0 and 1 based on the defined interest function. The attributes, user-defined weights, and interest functions used in this study (Table 3) are the same as those employed by Griffin et al. (2017a,b, 2020, 2021). It should be noted that when calculating the interest score, the centroid distance weight is actually the user-defined weight multiplied by the paired objects' area ratio. Overall, this analysis prioritizes size comparisons and distance between the forecast and observation objects. More emphasis is placed on the displacement between the centroids of the objects and the ratio of the objects' areas because the impact of both the boundary distance and ratio of the intersection area on the interest score can be high when large objects only slightly overlap each other or a larger object fully encloses a smaller object, respectively.

TABLE 3. User-defined weights and brief description of the object pair attributes used in this analysis.

Object pair attribute	Weight (%)	Description	Interest function
centroid_dist	4 (25.0)	Distance between objects' "center of mass"	1 if $x \le 20$ km $1 - \frac{1}{(150 - 20)} \times (x - 20)$ for $20 \le x \le 150$ km 0 if $x \ge 150$ km
boundary_dist	3 (18.75)	Minimum distance between the boundaries of two objects	$1 - \frac{1}{(133.33)} \times x$ for $x \le 133.33$ km 0 if $x > 133.33$ km
convex_hull_dist	1 (6.25)	Minimum distance between the polygons surrounding the objects	$1 - \frac{1}{(133.33)} \times x$ for $x \le 133.33$ km 0 if $x > 133.33$ km
angle_diff	1 (6.25)	Orientation angle difference	1 if $x \le 30$ $1 - \frac{1}{(90 - 30)} \times (x - 30)$ for $x > 30$
area_ratio	4 (25.0)	Ratio of the forecast and observation objects' areas (whichever yields a lower value)	1 if $x \ge 0.8$ $x \times \frac{1}{0.8}$ if $x < 0.8$
int_area_ratio	3 (18.75)	Ratio of intersection area to the lesser of the forecast and observation object areas (unitless)	$\frac{0.5}{0.1} \times x$ if $x \le 0.1$ $(x - 0.1) \times \frac{0.5}{0.15} + 0.5$ if $0.1\ x \le 0.25$ 1 if $x > 0.25$

Interest scores are also used to identify clusters of objects. Clusters occur when one or more observation objects are paired with one or more forecast objects (Bullock et al. 2016; Jensen et al. 2020). For example, a large observation object can be paired with multiple smaller forecast objects, therefore allowing these smaller forecast objects to be paired with a larger observation object. When clusters are identified, the attributes of the cluster are used instead of the individual objects' attributes. If clusters are not used, smaller objects might not be part of an object pair. In this analysis, clusters have a minimum interest score of 0.65, consistent with Griffin et al. (2017a,b, 2020, 2021).

Object-based threat score

The accuracy of the forecast cloud and radar reflectivity objects is assessed using the object-based threat score (OTS; Johnson and Wang 2013). The OTS is calculated using object sizes and interest scores between paired objects based on the equation:

$$\mathrm{OTS} = \frac{1}{A_f + A_o}\left[\sum_{p=1}^{P} I^p(a_f^p + a_o^p)\right]. \qquad (1)$$

In Eq. (1), A_f and A_o represent the area of all forecast and observation objects, both paired and unpaired, respectively. The term P represents the number of paired forecast and observation object pairs. The term I^p represents the interest score between the paired forecast and observation object/cluster calculated using the attributes in Table 3, and a_f^p and a_o^p represent the areas of the forecast and observation objects/clusters in the pair, respectively. As the OTS is based on a one-to-one correspondence, objects can only be used once in the calculation. Since an object can only be used once, either as an individual object or as part of a cluster, cluster and object pairs are sorted from the highest to the lowest interest score to calculate the highest possible OTS. Interest scores are used until no more paired objects and clusters remain. The OTS has a scale from 0 to 1, with one representing a perfect forecast.

4. Results and discussion

a. Comparison of observed radar and satellite objects

The first step in comparing object-based radar reflectivity to satellite BTs is to establish observation reflectivity and BT object thresholds. These thresholds will allow us to identify the relationship between observation BTs and radar reflectivity and assess the HRRR's ability to reproduce this relationship. An example of the observed BTs and composite radar reflectivity from these events is shown in Fig. 3. The domains used for each event are indicated in Table 4. Based on observed radar reflectivity for each event, it was decided to use 40 dBZ for the tornado event. This threshold was chosen because it captured the discrete locations of the most intense radar reflectivity. A threshold of 20 dBZ is used for the snow event, as it captures the full extent of heavy snow. The derecho event is defined using a threshold of 30 dBZ to encompass the convective areal extent of the derecho. MODE objects identified using these thresholds are outlined on the radar reflectivity images in Fig. 3. All analysis will be conducted over the full domain for each event in Fig. 3.

Once the radar thresholds are established, different BT thresholds are analyzed to determine which one is best collocated with the radar reflectivity threshold. To determine the corresponding observed satellite BT for each event, the areal extent of MvMODE AND objects and MvMODE OR objects over the full event domain is compared using the equation:

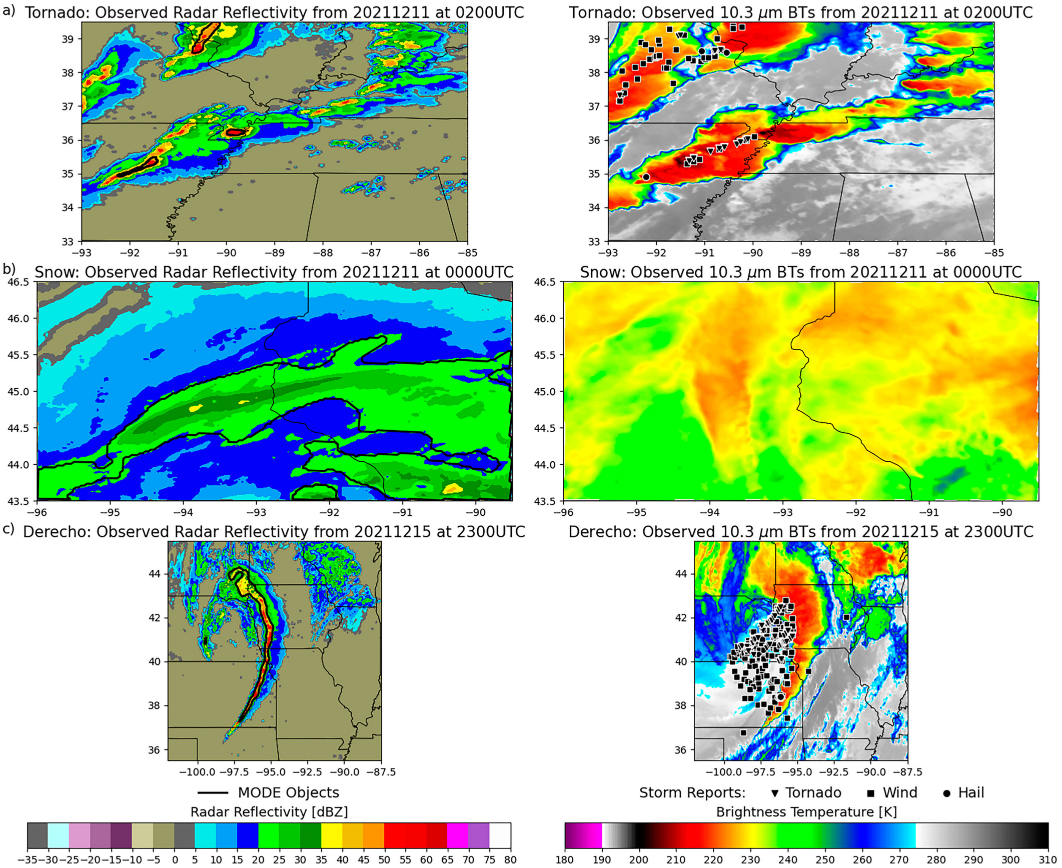


FIG. 3. Representative examples of the composite radar reflectivity and observed BTs for the (a) tornado, (b) snowfall, and (c) derecho events used in this analysis. Outlines on the radar reflectivity images represent MODE objects, and markers on the BT images represent storm reports from the Storm Prediction Center. These images also show the regions over which the model verification statistics are computed in this study.

$$\mathrm{SAC} = \begin{cases} \dfrac{\mathrm{area}_{\mathrm{OR}}}{\mathrm{area}_{\mathrm{radar}}} - \dfrac{\mathrm{area}_{\mathrm{AND}}}{\mathrm{area}_{\mathrm{radar}}}, & \mathrm{area}_{\mathrm{BT}} \geq \mathrm{area}_{\mathrm{radar}} \\[2ex] \dfrac{\mathrm{area}_{\mathrm{AND}}}{\mathrm{area}_{\mathrm{radar}}} - \dfrac{\mathrm{area}_{\mathrm{OR}}}{\mathrm{area}_{\mathrm{radar}}}, & \mathrm{area}_{\mathrm{BT}} < \mathrm{area}_{\mathrm{radar}} \end{cases} . \qquad (2)$$

The satellite area comparison (SAC) for the observation objects can be seen in Fig. 4b. A positive (negative) SAC indicates a larger (smaller) BT areal extent compared to radar. Perfectly overlapping radar and BT objects have a SAC value of 0.

Based on Fig. 4b, we chose to use a BT threshold of 210 K for the tornado event, even though this threshold does not have observation satellite objects for the first two hours when a radar object is present. However, the radar object is quite small at 0000 UTC (as seen in Fig. 4a). This threshold of 210 K is slightly lower than the 215 K used to identify overshooting tops in Bedka et al. (2010). Though a threshold of 215 K does have BT objects at 0100 UTC, it has a much larger areal extent than the radar object at later valid times. In this analysis, we would like the SAC to be as close to zero as possible. For the snowfall event, a BT threshold of 235 K will be used, which is also within the range of BTs associated with heavy snow from Hanna et al. (2008). For the derecho, a BT threshold of 225 K will be used. This is also the threshold used by Bedka et al. (2010) to determine anvils surrounding overshooting tops. The number of observed radar and satellite objects identified using these thresholds can be seen in Fig. 4c. While BT thresholds of 210, 235, and 225 K are used for the full event duration in the following analysis, we also

TABLE 4. Latitude and longitude ranges for each domain in Fig. 3.

Event	Minimum latitude (°)	Maximum latitude (°)	Minimum longitude (°)	Maximum longitude (°)
Tornado	33.0	39.5	−93.0	−85.0
Snow	43.5	46.5	−96.0	−89.5
Derecho	35.5	45.5	−102.0	−87.5

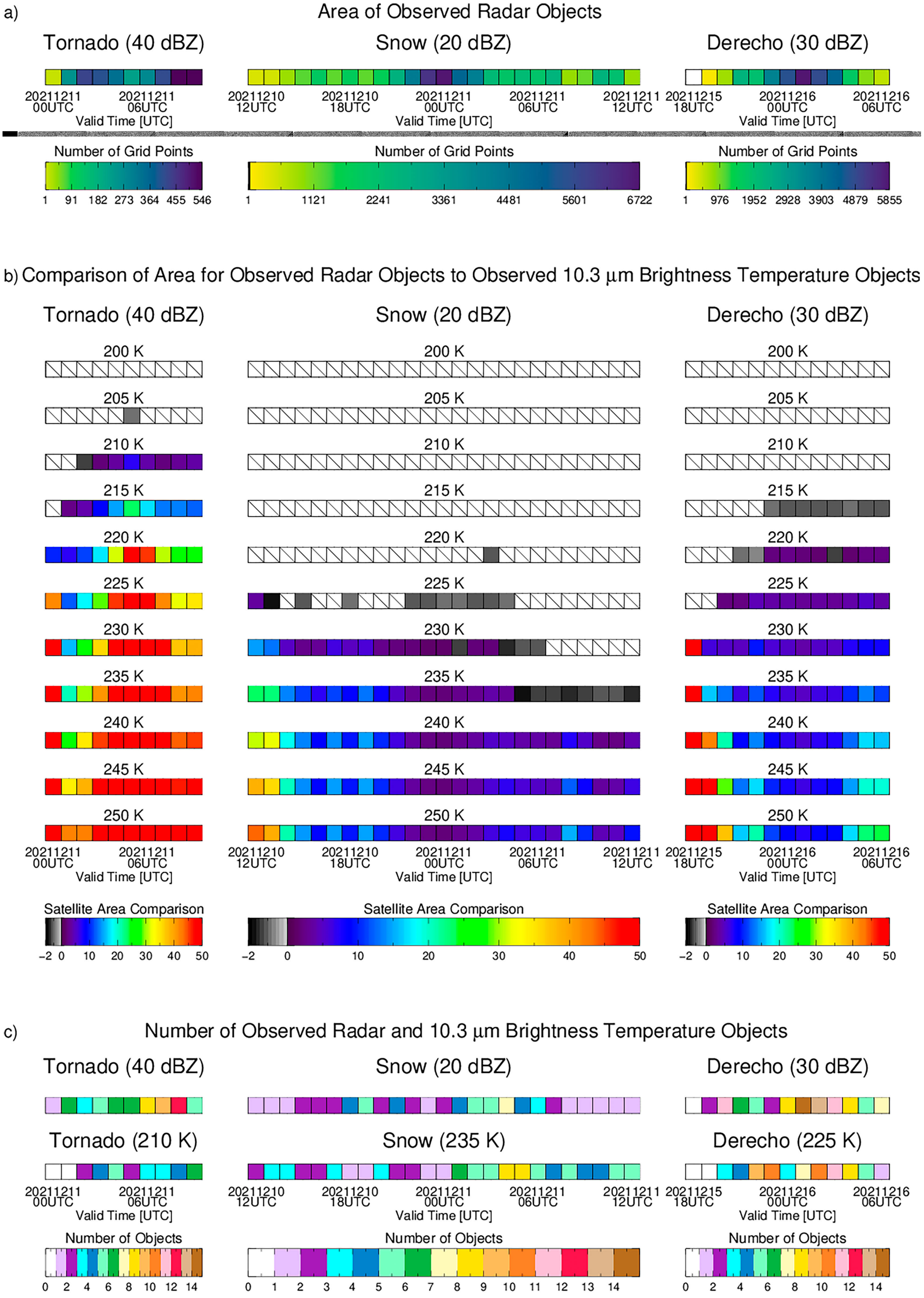


FIG. 4. (a) Areal extent of observation radar objects. (b) Comparison of areal extent for observation radar objects and satellite objects for different BT thresholds based on the satellite area comparison [Eq. (2)]. Instances with no satellite objects are identified by slashes, and white squares indicate no observed BT or radar object for that threshold are present. (c) Number of observed radar and satellite objects.

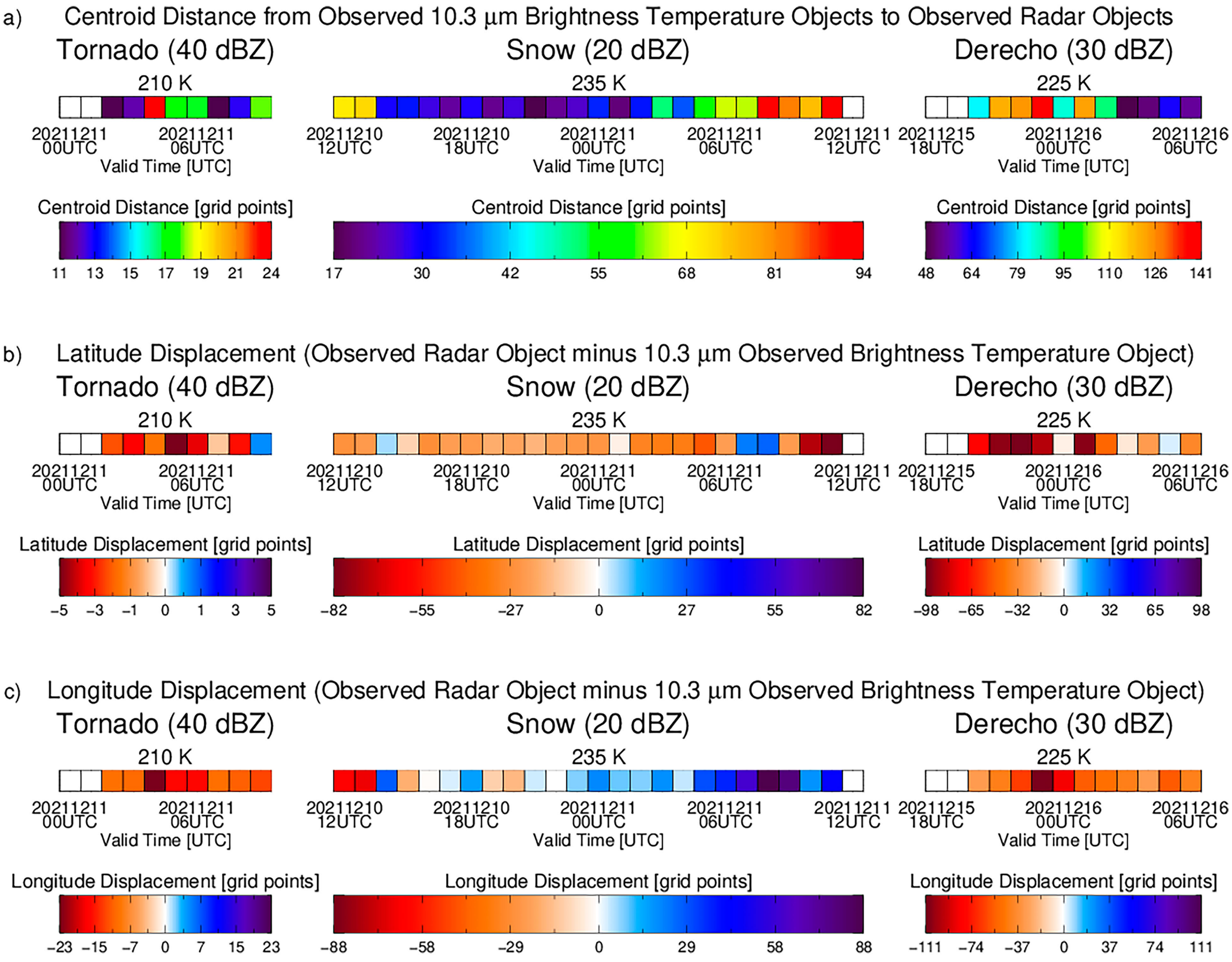


FIG. 5. (a) Centroid distance, (b) latitude displacement, and (c) longitude displacement between the center of paired radar and satellite objects.

did the same analysis as below but selected BT thresholds based on which BT had the lowest positive SAC value for each valid time. However, using those valid time BTs thresholds, (which can be seen in Fig. 8) did not impact the overall conclusions of this research.

The displacement between observation radar and satellite objects can be seen in Fig. 5. For the tornado event, the centroid distance (Fig. 5a) varies as the observed event progresses. A larger average centroid distance is associated with a higher number of BT objects (seen in Fig. 4c). More BT objects allow for more possible object pairs, which could have lower interest scores and be farther from each other. For the snow event, the centroid distance is lowest during the middle of the event, when the magnitude of the SAC is closest to zero. A lower SAC indicates fewer additional satellite object grid points compared to radar grid points, resulting in a smaller centroid distance. When the SAC is less than zero, the radar and satellite objects are not overlapping in this event (not shown), and the centroid distance is larger. For the derecho event, the highest centroid distances are present at the beginning of the event, when the radar objects are much smaller but more numerous than the satellite objects, and the threshold of 225 K struggles to identify objects at the beginning of the event. While a lower BT threshold of 220 K more accurately identifies the location of the 30-dB*Z* objects at 2100 UTC 15 December 2021, a 225-K threshold better encompasses its areal extent (not shown).

The latitude and longitude displacement between the center of paired observation objects can be seen in Figs. 5b and 5c. This displacement, in grid points, is calculated as

$$\text{displacement} = \text{center}_{\text{Radar}} - \text{center}_{\text{Satellite}}. \qquad (3)$$

So positive (negative) values of latitude displacement indicate the satellite object center is to the south (north) of the radar object, and positive (negative) values of longitude displacement indicate the satellite object center is west (east) of the radar object center (western hemisphere longitude is considered negative in this analysis). For the three events in this study, the center of the satellite objects is to the north of the radar objects. Satellite objects are to the east of the radar objects for the tornado and derecho events. This eastward displacement is consistent with Rasmussen and Straka (1998), which suggests upper-level cloud cover in convection tends to

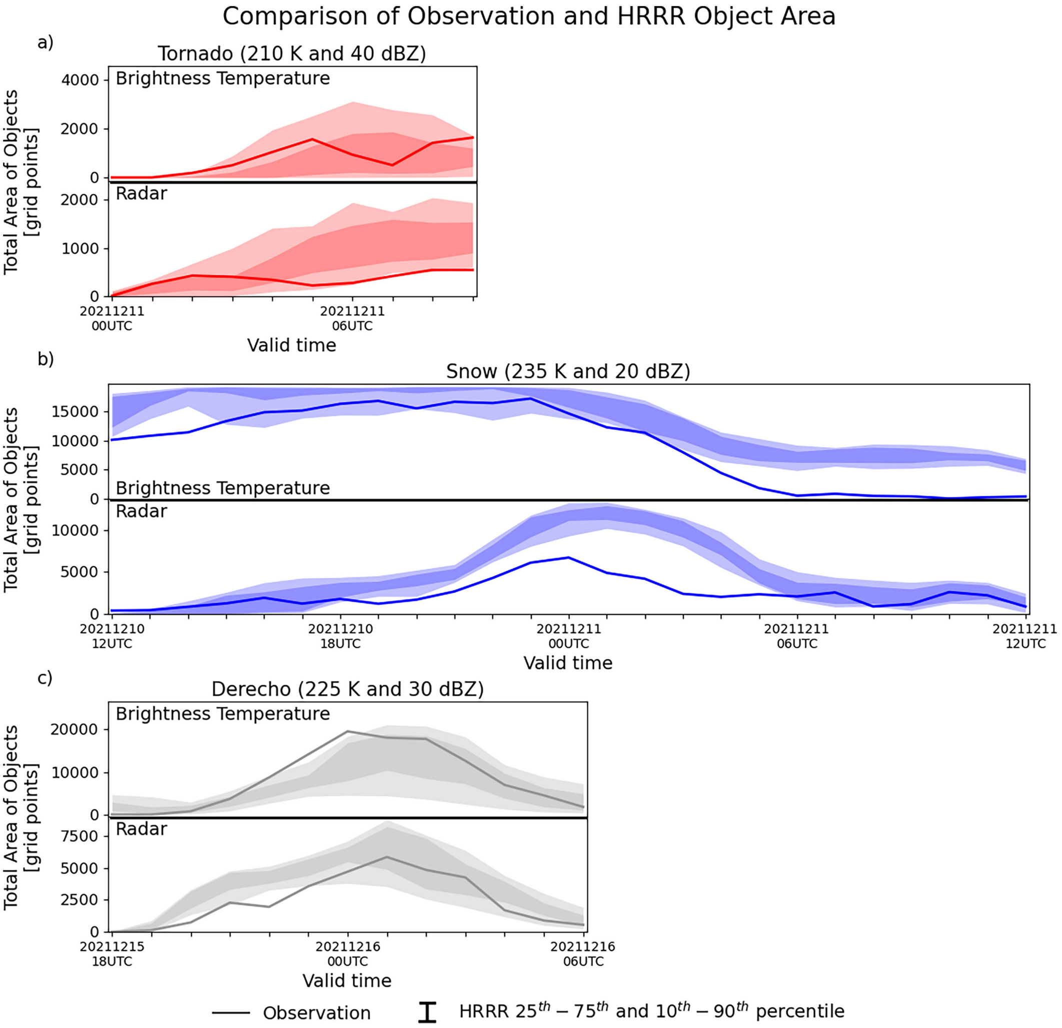


FIG. 6. Area of observation (line) and forecast (envelope) satellite BT and radar objects in the Fig. 2 domains for the (a) tornado, (b) snow, and (c) derecho events. The interquartile range for forecast objects area over the forecasts hours for each valid time is displayed as the darker shaded envelope, while the lighter shading represents the 10th–90th percentile range.

be located eastward of radar reflectivity. This displacement is possibly due to vertical wind shear associated with supercell storms (Brooks et al. 1994). For the snow event, the center of satellite objects is generally west of the radar object. These differences suggest the displacement between satellite and radar object can be due to the different meteorological regimes, though a larger sample would provide an additional confirmation.

b. *Validation of HRRR satellite and radar objects*

1) OBJECT AREA

To assess how accurately HRRR can simulate the areal extent of satellite and radar objects, a comparison between the area of observation and forecast satellite and radar objects defined using the thresholds from the previous section can be seen in Fig. 6. The HRRR Model generally underpredicts the area of BT and radar objects at the beginning of the tornado event (Fig. 6a), which could be due to delayed convective initiation, as the number of HRRR BT and radar objects is lower than the observations (not shown). Later in this event, though, the areal extent of all radar objects is too large compared to the observations, while the area of the BTs is still too small. One hypothesis for the smaller forecast satellite object area in the tornado event is that the HRRR atmospheric temperature is higher at a given level than the observations. This hypothesis assumes cloud-top temperature can be derived from cloud-top pressure, similar to the way in which the cloud-top pressure in the AWG cloud height algorithm (ACHA) is derived from the cloud-top temperature product and the atmospheric temperature profile provided by numerical weather prediction data (Heidinger and Li 2018). Figure 7 shows a comparison between observed BTs in the left column and the 4-h forecast HRRR BTs in the right column. While

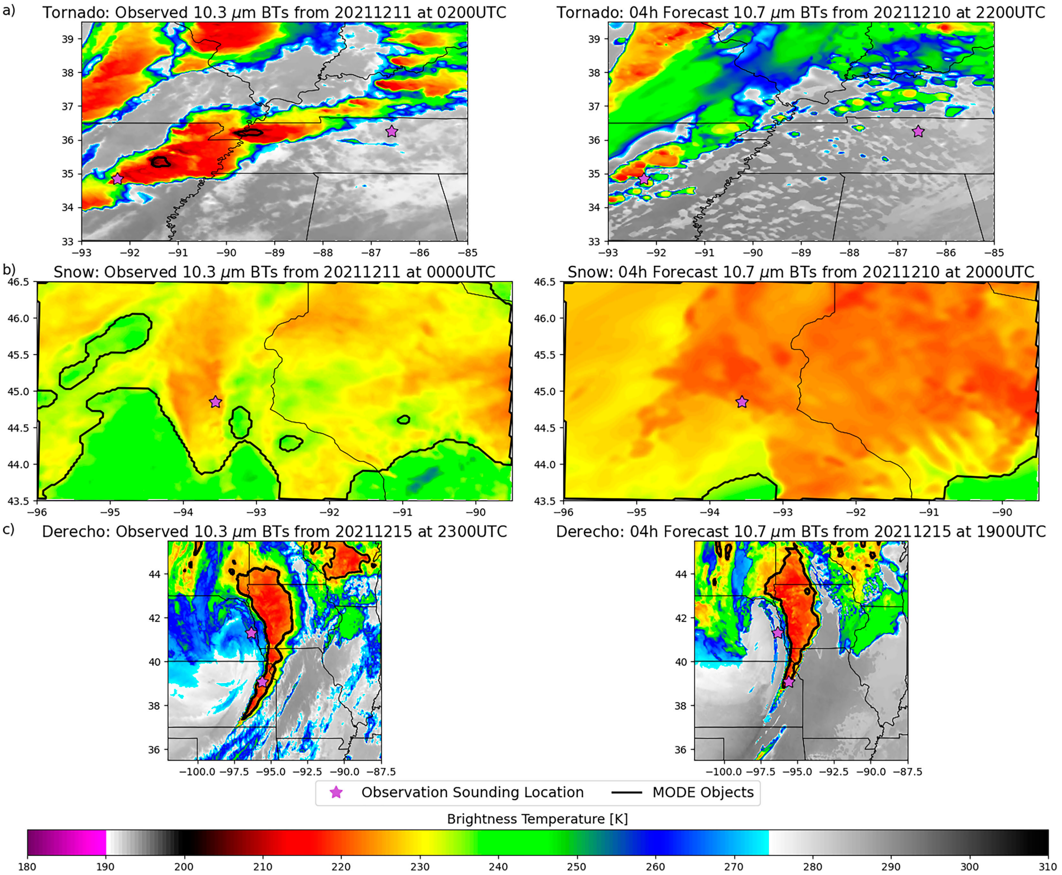


FIG. 7. (left) Observed and (right) 3-h forecast HRRR simulated BTs. The purple stars represent the location of observed temperature profiles.

the HRRR atmospheric temperature at 200 hPa, the approximate cloud-top pressure at Little Rock, Arkansas (left star in Fig. 7a), is 0.57 K lower than the observed temperature at 0000 UTC 11 December 2021, the HRRR atmospheric temperature at 150 hPa is 2.24 K higher. This higher 150-hPa HRRR temperature could be increasing the simulated HRRR BT compared to the observations, and therefore reducing the size of forecast satellite objects as seen in Fig. 7a. In comparison, the HRRR atmospheric temperature at 150 hPa for Nashville, Tennessee (right star in Fig. 7a), where few clouds are present, is 0.53 K lower than the observations. For the derecho event (Fig. 6c), again HRRR generally underpredicts the areal extent of satellite objects but overpredicts the areal extent of the radar objects. An exception occurs at the beginning of the event, when the areal extent of satellite objects is higher than the observations. The number and average areal extent of HRRR BT objects is also higher at the beginning of the derecho event (not shown). Therefore, the total underprediction of the areal extent of satellite objects throughout the duration of the event does not appear to be the result of delayed convective initiation. Instead, the HRRR atmospheric temperature could be increasing the simulated HRRR BT compared to the observations, as was hypothesized in the tornado event. At 0000 UTC 16 December 2021 the HRRR atmospheric temperature at 150 hPa is also 0.47 K higher than the temperature sounding for Topeka, Kansas (bottom star in Fig. 7c), while the HRRR atmospheric temperature at 500 hPa for Omaha, Nebraska (top star in Fig. 7c), is about 2.11 K lower. Thus, the higher HRRR atmospheric temperature could again be increasing the simulated BTs of upper-level clouds and reducing the areal extent of the objects (Fig. 7c).

This trend in smaller satellite object area for convective events differs from the snow event (Fig. 6b), where the areal extent of the simulated BT is much larger than the observation, though the areal extent of the radar objects is still lower than the observations at the beginning of the event. Compared to the temperature sounding from Minneapolis, Minnesota (star in Fig. 7b), at 1200 UTC 10 December 2021, the HRRR atmospheric temperature at the cloud-top pressure of 300 hPa is about 0.80 K lower, and the HRRR atmospheric temperature at the cloud-top pressure of 300 hPa is about 1.05 K lower than the temperature sounding from Minneapolis at 0000 UTC 11 December 2021 (cloud-top pressure data are not available for 1200 UTC 12 December 2021). These lower atmospheric temperatures could result in larger objects by lowering the simulated BTs compared to the observations (Fig. 7b). Thus, in this study, the bias in areal extent of

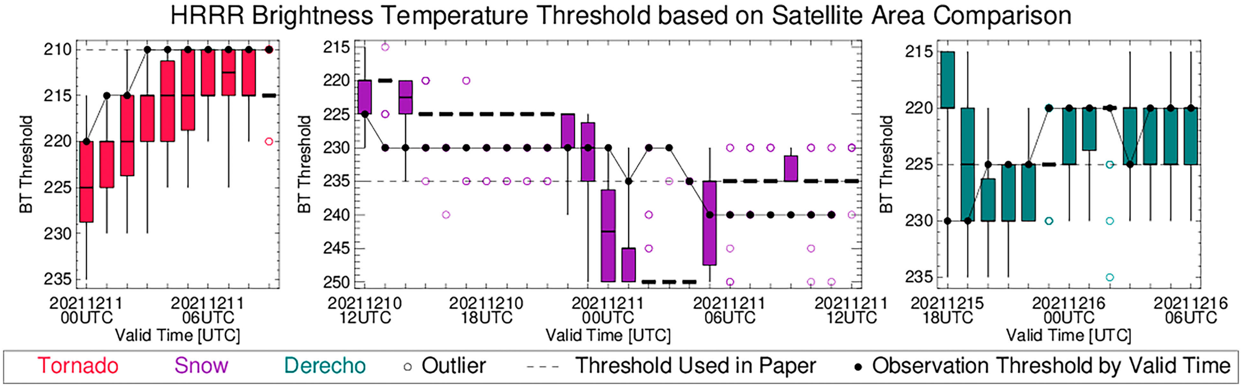


FIG. 8. Comparison of areal extent for HRRR radar objects and satellite objects based on the satellite area comparison [Eq. (2)]. The box-and-whisker plots represent the range of optimal HRRR BT thresholds over the forecast hours, with colored open circles representing outliers. When no box is present, it indicates the optimal HRRR BT threshold is the same within the interquartile range. The dashed line indicates the BT threshold used in this study, while the circles indicate the observation BT threshold if it were to change by valid time.

HRRR simulated BTs and radar objects differs greatly based on event type. Convective events have a smaller areal extent of BT objects, but the areal extent of both BT objects is larger for the synoptically driven snowfall event, which could be due to errors in the upper-tropospheric forecast temperatures. All events have a larger areal extent of radar objects.

As expected, the differing proportion of simulated satellite and radar reflectivity object area compared to the observations would impact the optimal BT threshold for each radar reflectivity threshold defined by the SAC. The HRRR BT threshold by valid time for each event can be seen in Fig. 8, where the box and whiskers represent the distribution of optimal BT threshold over the forecast valid at each time (hour of the day). For the tornado event, the HRRR BT threshold would be higher than that used in this study (dashed line) and the optimal observation threshold for each valid time (circle). This is consistent with the lower total area of forecast satellite objects but higher radar object area in Fig. 6a. The opposite is observed in the snow event, as the optimal HRRR BT threshold is lower than the observations at the beginning of the event, due to more satellite object grid points than the observations. The HRRR BT threshold for the derecho event is similar to the observation threshold, consistent with similar forecast and observation satellite and radar object total area in Fig. 6c. The use of these thresholds will be discussed in section 4b(5).

2) OBJECT INTENSITY

Another method for comparing forecast and observation objects is to calculate the average intensity of these objects, i.e., the average BT or radar reflectivity of the objects. The average BT for the observation objects is seen in Fig. 9a and for the forecast objects in Fig. 9b. A positive (negative) difference indicates the average forecast object BT is lower (higher) than the observation and thus more (less) intense than the observation object. For the tornado event, the forecast object BTs are similar to the observations when the forecast objects exist, though there are many instances with no forecast objects. For the snow event, the forecast object BTs are generally lower and more intense than the observations. Higher simulated BTs are only seen in the first two forecast hours during the snow event, similar to the spin up noted in Griffin et al. (2017b). The derecho event objects also have lower (more intense) simulated BTs. Forecast objects also occur before the observation object is identified, as evident by the colored squares to the left of the thick black line. Based on these events, simulated BT object intensity appears to vary based on object scale. The simulated BTs are similar to the observations for the smaller scale tornado event objects, but lower than the observations for the larger snow and derecho event objects. Again, a larger sample is necessary to confirm this trend.

The difference in intensity of forecast and observation radar objects can be seen in Fig. 10. Since instances of lower BTs and higher radar reflectivity are considered more intense, the difference calculation in Fig. 10b is switched compared to Fig. 9b so a positive (negative) difference indicates the average forecast object reflectivity is higher (lower) than the observations. Therefore, in both Figs. 9b and 10b, red colors indicate the forecast is more intense than the observations. Something noticeable when comparing Figs. 9b and 10b is the intensity difference changes orientation for the tornado event. For example, the forecast radar objects are more intense than the observations in instances where the simulated BTs are less intense for 39% of the tornado event forecast hours, compared to only 3% and 16% for the snow and derecho event, respectively. Only the 0-h forecast consistently has lower radar reflectivity than the observations, possibly due to radar spin-up as noted by Duda and Turner (2021). Figure 11a displays an example of more intense forecast radar objects, where the forecast radar object intensity is higher than the observations for the tornado event despite fewer forecast object grid points. This could be due to an error in the storm mode, as HRRR has predominately cellular objects while the observation objects are more elongated. Figure 11 is the same forecast hour and valid time as the BT displayed in Fig. 7. There are no identified simulated satellite objects for the tornado event (Fig. 7a, right).

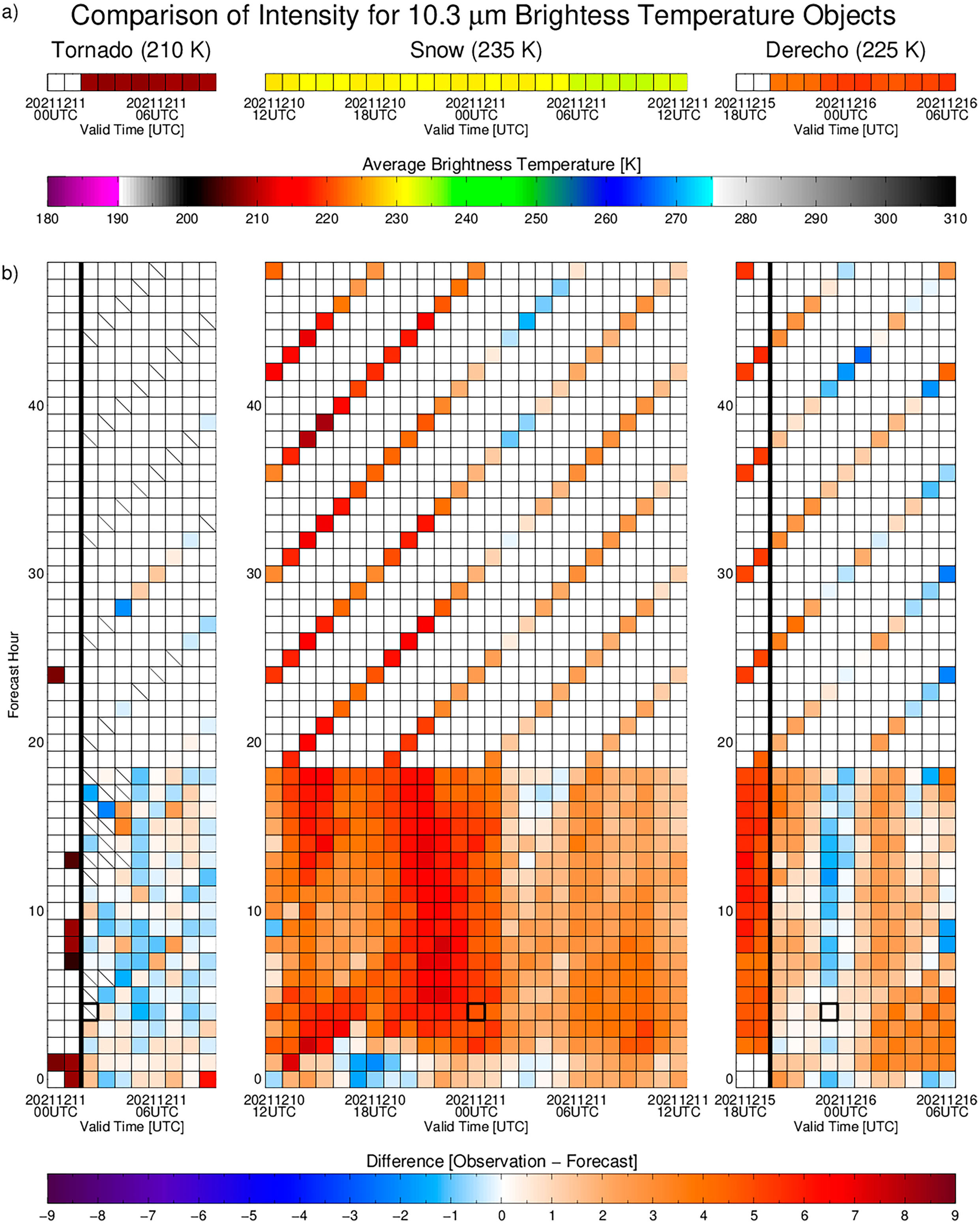


FIG. 9. (a) Average BT of observation objects. (b) Difference in average BT of observation and forecast objects by forecast hour. A positive (negative) difference indicates the average forecast object BT is lower (higher) and thus more (less) intense than the observation. Instances with no observation objects but existing forecast objects are left of the thick black line and plotted using the colors in (a). The cells corresponding to the right images in Fig. 7 are indicated by black squares.

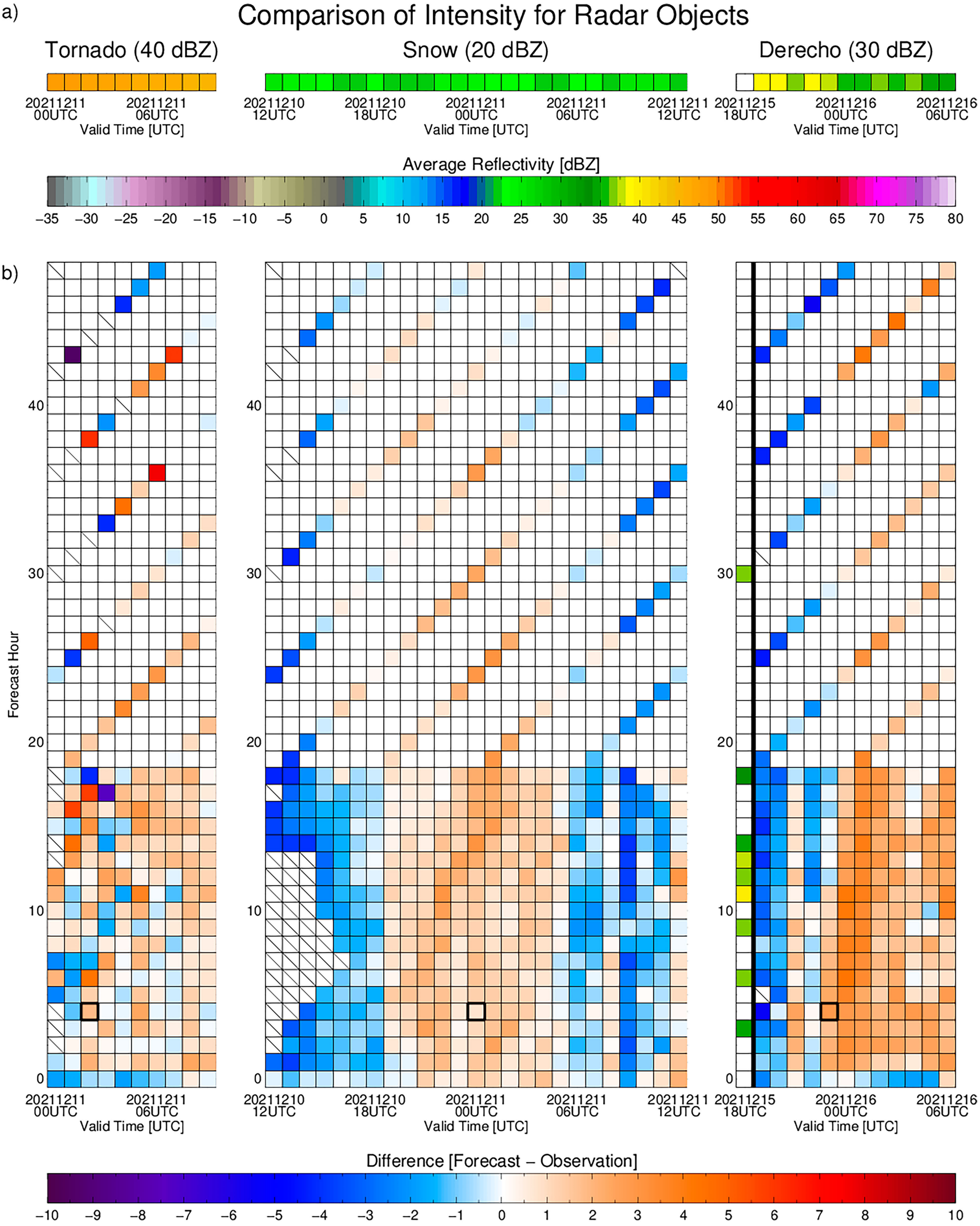
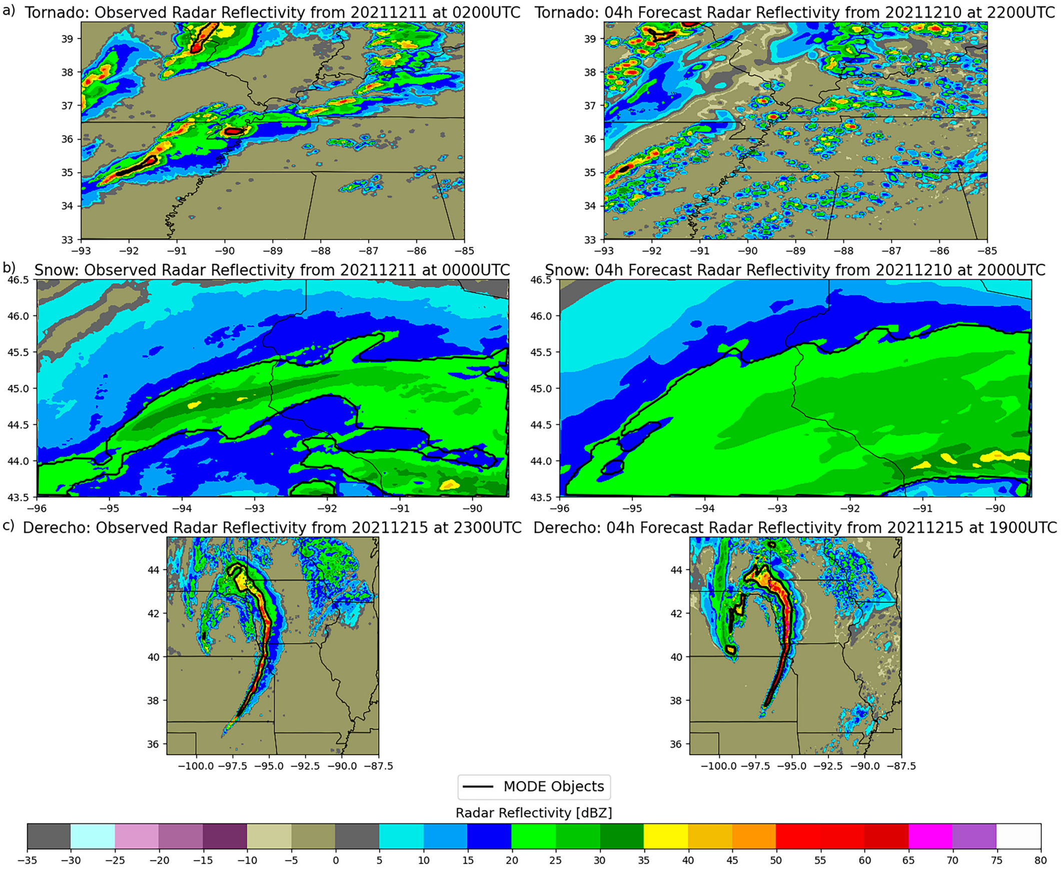


FIG. 10. (a) Average radar reflectivity of observation objects. (b) Difference in average radar reflectivity of forecast and observation objects by forecast hour. A positive (negative) difference indicates the average forecast object radar reflectivity is higher (lower) and thus more (less) intense than the observation object. Instances with no observation objects but existing forecast objects are left of the thick black line and plotted using the colors in (a). The cells corresponding to the right images in Fig. 11 are indicated by black squares.


a) Tornado: Observed Radar Reflectivity from 20211211 at 0200UTC

Tornado: 04h Forecast Radar Reflectivity from 20211210 at 2200UTC

b) Snow: Observed Radar Reflectivity from 20211211 at 0000UTC

Snow: 04h Forecast Radar Reflectivity from 20211210 at 2000UTC

c) Derecho: Observed Radar Reflectivity from 20211215 at 2300UTC

Derecho: 04h Forecast Radar Reflectivity from 20211215 at 1900UTC

FIG. 11. (left) Observed and (right) 4-h forecast HRRR simulated radar reflectivity.

Again, the higher HRRR atmospheric temperature could be increasing the simulated BTs as previously discussed. For the snow event, the simulated radar reflectivity is less intense than the observations during both the beginning of the event, with the empty boxes indicating no forecast radar objects were identified even though observation radar objects were present, and at the end of the event. However, in the middle of the event, both simulated radar reflectivity and satellite BTs are more intense. The lower HRRR atmospheric temperature could be lowering the simulated BTs at this time, as seen with the 0.80- and 1.05-K lower atmospheric temperature at 300 hPa at 1200 UTC 10 December 2021 and 0000 UTC 11 December 2021, respectively, in Minneapolis, Minnesota. The derecho event also has more intense object simulated BTs and radar reflectivity, except for within the first few hours of the event.

Another hypothesis for higher or lower simulated BTs in Fig. 9 is the increase or decrease of cloud hydrometeor mixing ratios compared to observations. The correlation between the total cloud mixing ratio (sum of ice, snow, graupel, cloud water and rainwater) above 500 hPa at every object grid point and the simulated BTs is generally negative for all events (supplemental Fig. 1 in the online supplemental material), indicating higher total cloud mixing ratio is associated with lower BTs. However, it is difficult to assess if the difference between the HRRR total cloud water mixing ratio and the observations is causing the difference in object intensity because there is no reliable measure of observed total cloud mixing ratio. Moreover, simulated BTs are also sensitive to the size and shape of the cloud hydrometers (Senf and Deneke 2017), which are also unavailable.

3) OBJECT LOCATION

A comparison of the displacement between satellite and radar objects can be seen in Fig. 12. As seen in Fig. 12a, the average centroid distance between the HRRR BTs objects to the HRRR radar objects for the tornado event trends upward as the event progresses. This is similar to the end of the event in the observations. The centroid distance is lower than the observations for the middle of the event. However, since the Pearson correlation coefficient between the centroid distance and forecast hour is generally negative for this event, the centroid distance is larger for smaller forecast hours. Again, this is possibly because there are more object pairs for earlier forecast hours. For the snow event, the average HRRR centroid distance follows the same pattern as the observations, which is higher at the beginning and end of the event and smaller in the middle. The centroid distance for the derecho also is consistent with the observations. These events do not have consistent correlations, so there generally is not a stronger relationship between forecast hour and centroid distance

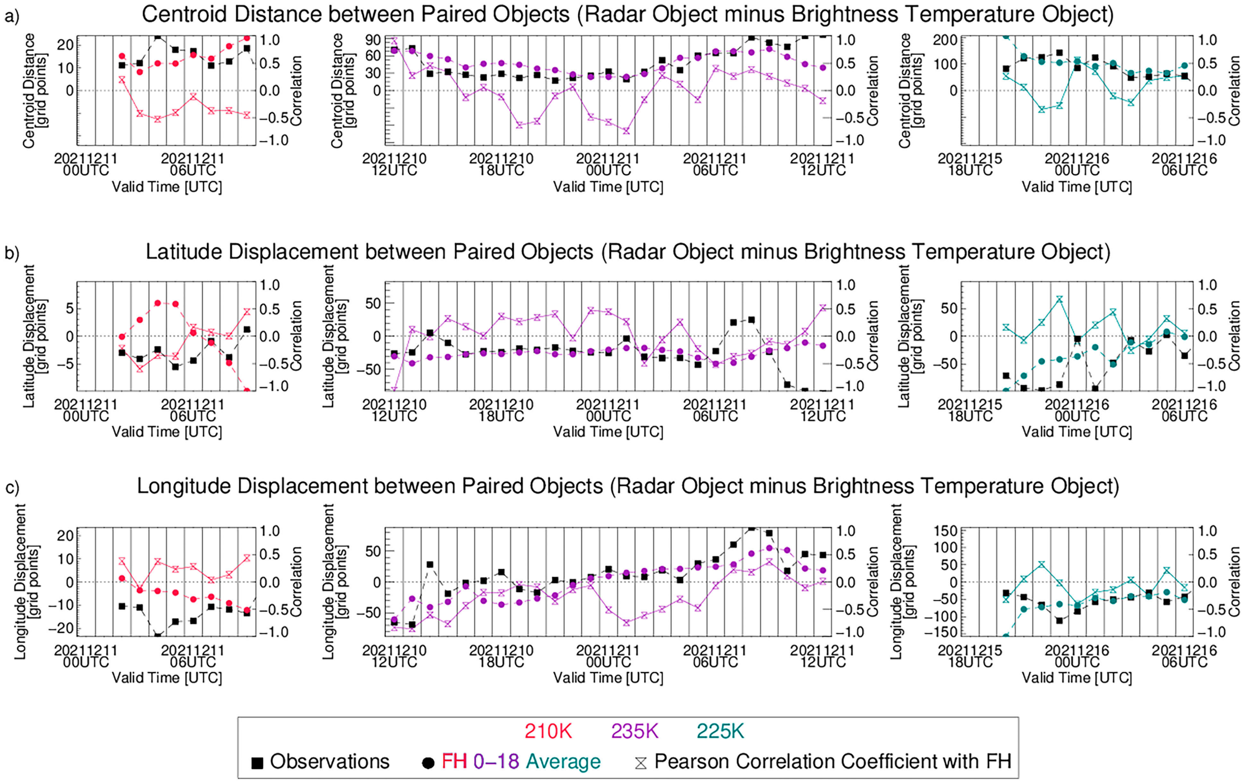


FIG. 12. (a) Centroid distance, (b) latitude displacement, and (c) longitude displacement between the center of paired observation (squares) and HRRR (circle) radar and satellite objects from the equation displacement = center$_{\text{Radar}}$ − center$_{\text{Satellite}}$. Positive values indicate the satellite object center is south of the radar object center in (b) and west of the radar object center in (c). The hourglass symbol denotes the Pearson correlation coefficient between HRRR forecast hour and the value plotted in (a)–(c). For example, the hourglass symbol in (a) indicates the Pearson correlation coefficient between the centroid distance and the forecast hour for each HRRR forecast valid at a given time.

for these larger areal extent events, unlike the tornado event. These correlation differences suggest the distance between the centers of forecast satellite and radar objects is more difficult to assume based on forecast hour alone for larger meteorological events, though a larger sample would provide an additional confirmation.

The latitude and longitude displacement between paired HRRR satellite and radar objects in Figs. 12b and 12c are also generally similar to the location displacement noted in the observations, with the exception of the tornado event latitude. The center of HRRR satellite objects is mainly to the east of the radar object for the tornado and derecho event, therefore HRRR is capturing the eastward displacement between satellite and radar objects of convective lines as seen in the observations (Fig. 5c) and Rasmussen and Straka (1998). For the snow and derecho events, HRRR satellite objects are north of the forecast radar objects, which is again consistent with the observations. The longitude displacement of the forecast satellite objects for the snow event is east at the beginning of the event and west at the end, also consistent with the observations. However, the center of HRRR satellite objects for some valid times is to the south of the radar object in the tornado event, while it is to the north for the observations. Many instances of southward displacement between the center of the HRRR satellite and radar objects occur in the early forecast hours, as evident by the negative Pearson correlation coefficient. These early forecast hours have a larger number of objects than later forecast hours. For example, the 4-h forecast valid at 0400 UTC 11 December 2021 has four objects, while the 15-h forecast only has one. This again could be resulting in fewer high interest score pairs, as one object pair from the 4-h forecast has a southward displacement of 101 grid points, and an interest score of only 0.23. However, the one object pair for the 15-h forecast has a southward displacement of 8 grid points and interest score of 1, so this southward displacement between the satellite and radar objects is not confined to lower interest scores and is instead a potential issue with the HRRR model at convective initiation.

4) OBJECT-BASED THREAT SCORE

The overall accuracy of the forecast cloud objects, based on the OTS, is seen in Fig. 13. For the forecast satellite objects (Fig. 13a), the snow event has the highest OTS, followed by the derecho and tornado events. This is not surprising, as smaller convective cells are harder to forecast (Wolff et al. 2014). As seen when comparing Fig. 13 to Fig. 6, the OTS is generally higher when the total object areas are higher. This

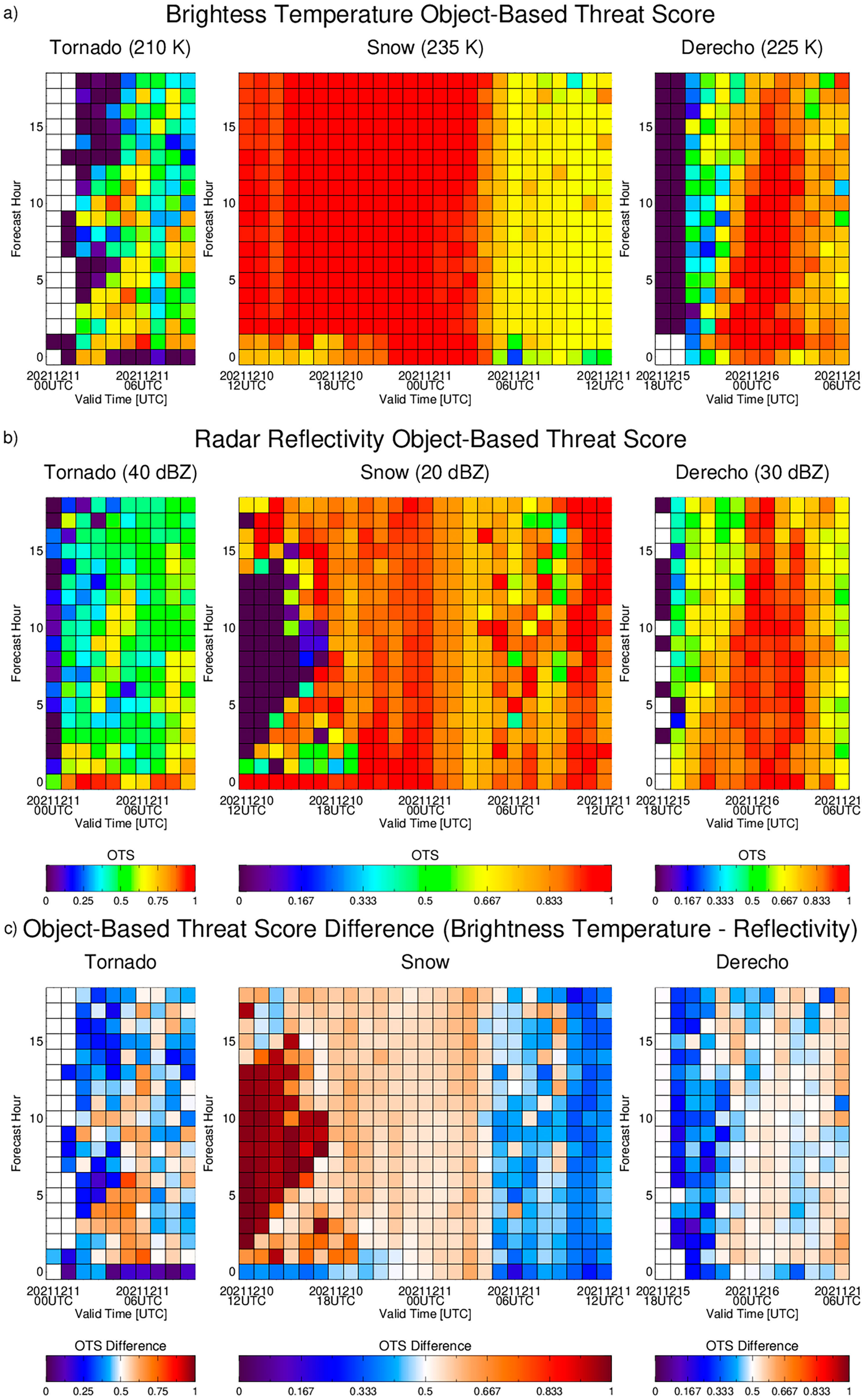


FIG. 13. (a) Object-based threat score (OTS) for HRRR forecast satellite objects. (b) OTS for HRRR forecast radar reflectivity objects. (c) Satellite object − radar reflectivity object OTS.

is especially evident in the derecho case, as the highest OTS values are seen in the middle of the event, when the object area is the largest. For forecast radar objects (Fig. 13b), the snow event also has the highest OTS, followed by derecho and tornado. Thus, the size and location of clouds and hydrometeors appears to be more accurate for events where the chosen thresholds produce larger scale objects. For the snow event, the average OTS over the entire time range in Fig. 13c is higher for the forecast cloud objects than the radar objects, whereas the opposite occurs for the tornado and derecho events. Over all events, OTS and area of observation satellite objects has a Pearson correlation coefficient of 0.66, indicating that larger objects are associated with a higher OTS.

The percent of observation area paired (a_o/A_o), percent of forecast area paired (a_f/A_f), and average interest score is seen in Fig. 14, as these three metrics are components of the OTS. In Fig. 14, blue represents the satellite objects and red represents the radar objects. For the tornado event, the percent of observation object area paired (Fig. 14a) is much higher for the radar objects than the satellite objects, possibly due to the larger area of forecast radar reflectivity objects than observation radar reflectivity objects (Fig. 6a). Combined with the higher average interest scores early in the event (Fig. 14c), this higher percentage of paired object areas probably increases the OTS for radar objects compared to the satellite objects. Compared to the tornado event, a higher percentage of observation and forecast area are paired for both satellite and radar objects in the snow event (Figs. 14d,e). Thus, even though the areal extent of the forecast objects is much higher than the observation objects (Fig. 6b), these additional forecast object areas are paired with observation objects. The increased area of paired forecast objects increases the OTS compared to the tornado event. In addition, the average interest score for the snow event (Fig. 14f) is also higher than the tornado event (Fig. 14c). For the derecho event, the percent of paired object area is larger than the tornado event. Thus, a higher OTS is observed for the derecho event (Figs. 12a,b), even though the average interest scores are similar. There is no noticeable difference between the percent of forecast and observation area paired and average interest scores for satellite and radar objects, which is why the difference in satellite and radar OTS is closest to zero for the derecho event (Fig. 13c).

A noticeable difference in the average interest score between the radar and satellite objects occurs in the snow event, as the average interest score is lower for the radar objects during the snow event for over half of the event, while there is little distinction between the average interest scores for the convective events a few hours after those events begin. This lower average interest score results in the lower OTS. To identify why this difference in average interest score occurs, the interest score is broken down into its four main contributors: centroid distance, boundary distance, area ratio, and intersection area ratio interest scores. This breakdown can be seen in Fig. 15, where the maximum possible attribute interest scores are the weights from Table 3 (0.25 for area ratio and centroid distance, and 0.1875 for boundary distance and intersection area ratio). For the snow event (Fig. 15a), though, the radar area ratio interest score is much lower than the satellite area ratio interest score until 0300 UTC 11 December 2021. This lower radar area ratio interest score for the snow event indicates that the overproduction of object grid points is greater for radar objects than for satellite objects, which can be seen by comparing Figs. 7c and 11c. While the total area of the HRRR forecast objects is much larger than the total area of the observation objects in both fields, the average observation-area-weighted area ratio is 0.62 for the radar objects and 0.85 for the satellite objects. This larger areal extent for radar objects reduces the OTS even when the forecast object area has a paired observation object. Less weight on the area ratio interest score could decrease the difference between the OTS for the radar object compared to the satellite object OTS. In contrast, the radar and satellite area ratio interest score for the tornado and derecho events are similar (Figs. 15b,c). Thus, the over or under production of object grid points in paired objects is consistent between satellite and radar objects for the convective events. Changing the weight of this attribute would impact the OTS, but it would not impact whether the HRRR satellite or radar objects are more accurate overall for these two events. In Figs. 15d–f, the satellite and radar object centroid distance, boundary distance, and intersection area ratio interest scores for all events are also similar, and thus are depicted together. Changing the weights of these attributes would also not overall impact whether the HRRR satellite or radar objects are more accurate. The lower radar area ratio interest score for the radar events indicates there is the overproduction of object grid points is greater for radar objects than satellite objects, and reduces the OTS for radar objects even when the forecast object area has a paired observation object.

5) ACCOUNTING FOR BIASES IN HRRR OBJECTS

Previous work has identified biases in the HRRR data, such as lower cloud BTs (Griffin et al. 2017a,b) or too many radar reflectivity objects (Duda and Turner 2021). To determine if HRRR biases exist during the three events used in this analysis, the percentile of the observation object threshold over the domains in Table 4 is calculated. Then, the HRRR simulated BT or radar reflectivity threshold corresponding to this percentile over the same domain is identified. These thresholds can be seen in Fig. 16, where the observation threshold is the solid line and the range of HRRR thresholds over all forecast hours is shaded. As seen in Fig. 16a, the corresponding HRRR simulated BTs are generally higher than the observations for the tornado event but lower for the snow event. The average HRRR simulated BT for the derecho event is similar to the observations (Table 5). For radar reflectivity (Fig. 16b and Table 6), the average HRRR simulated radar reflectivity is similar to the observations for the tornado and snow events but higher for the derecho event. Therefore, another method for validating HRRR satellite and radar objects would be to account for these biases by defining objects using the average HRRR thresholds in Tables 5 and 6. These objects will be known as "percentile objects," while the

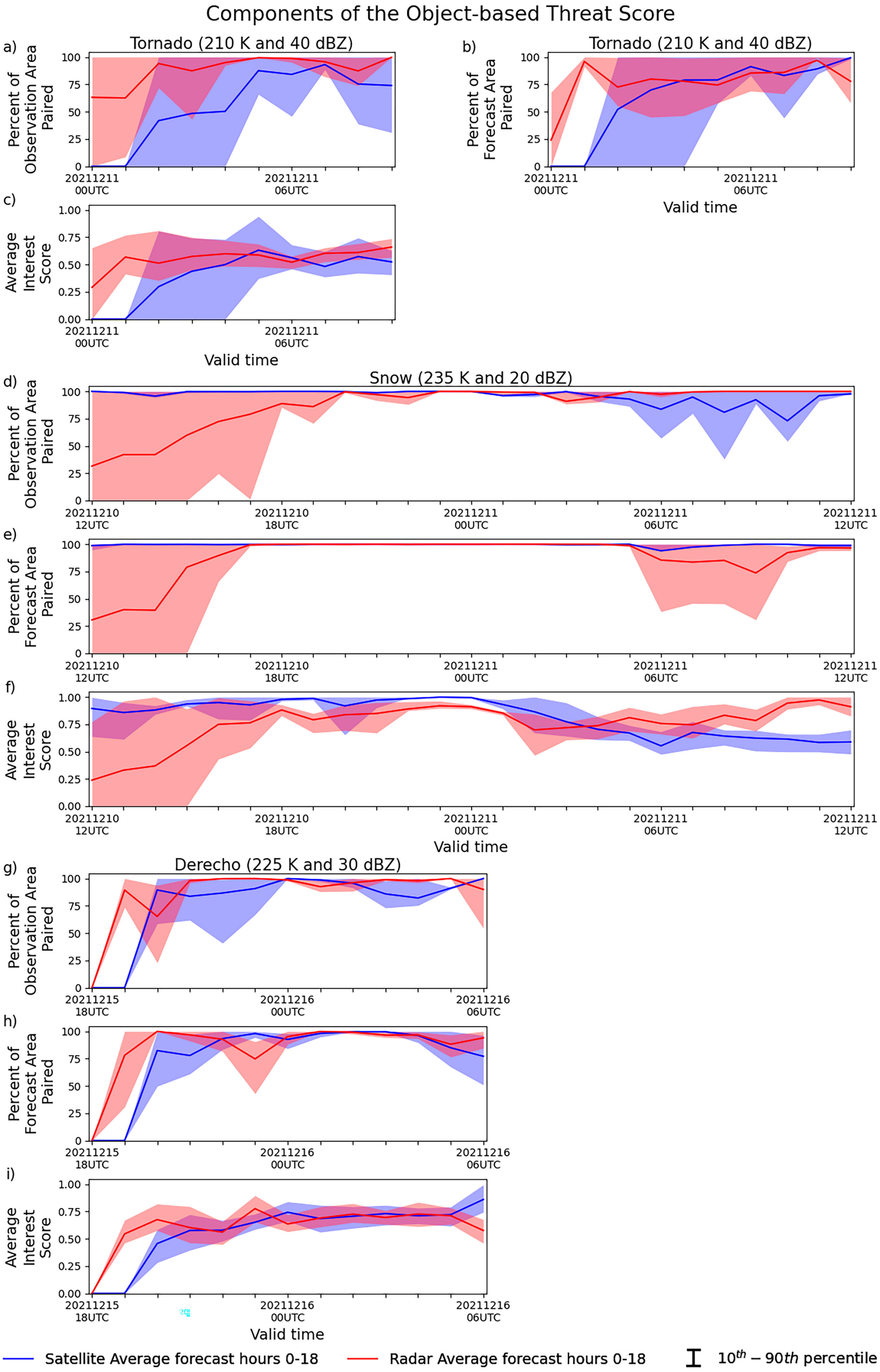


FIG. 14. (a),(d),(g) Percent of observation object area paired; (b),(e),(h) percent of forecast object area paired; and (c),(f),(i) average interest score for satellite (blue) and radar (red) objects over the tornado event in (a)–(c), snow event in (d)–(f), and derecho event in (g)–(i). The envelope represents the 10th–90th percentile for the forecasts valid at the given time.

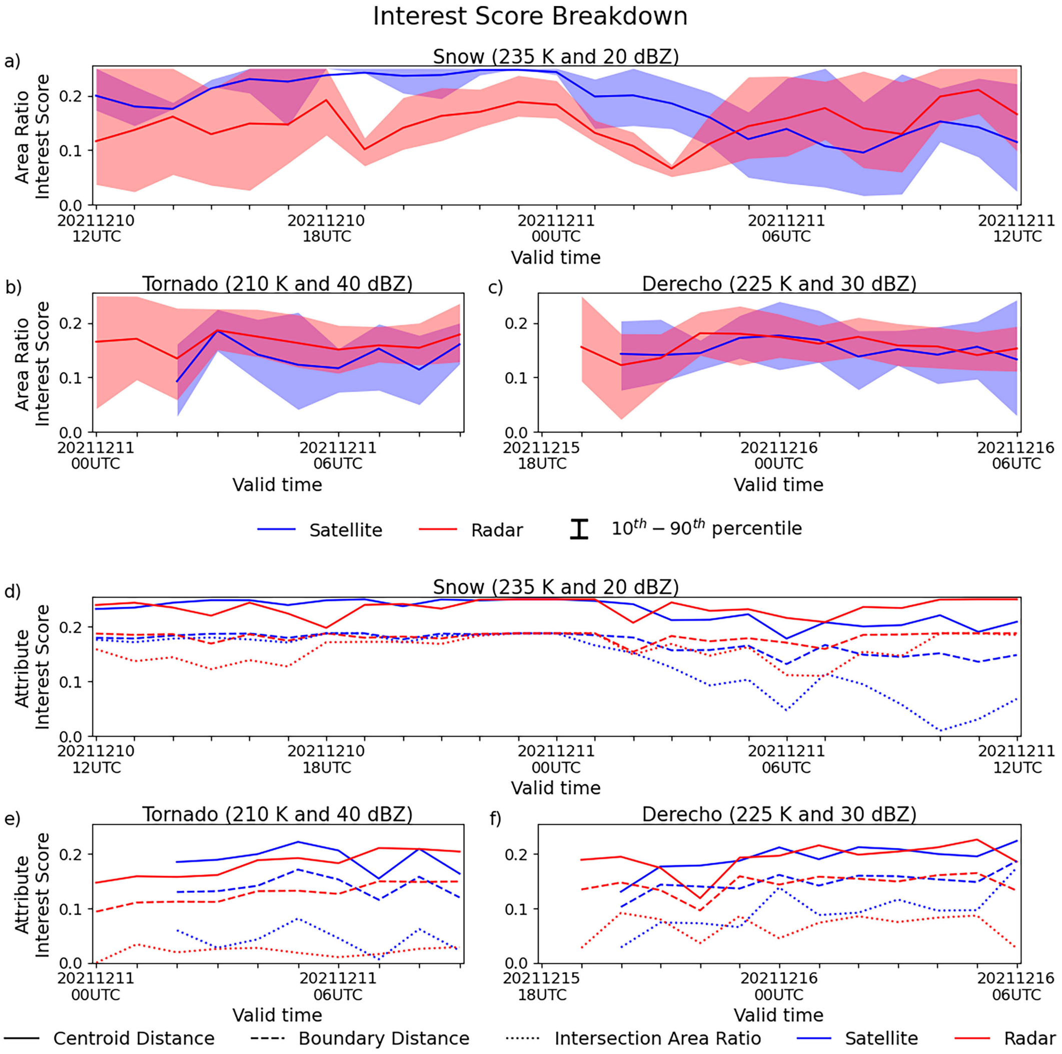


FIG. 15. Breakdown of the interest score into centroid distance interest score (solid line), boundary distance interest score (dashed line), and intersection area ratio interest score (dotted line) for the (a) tornado, (b) derecho, and (c) snow events. These scores are depicted together due to the similarity between the satellite and radar objects for each event. The breakdown of the area ratio interest score (solid line) for the (d) tornado, (e) derecho, and (f) snow events is plotted separately with envelopes to highlight the difference for the snow event.

objects from the previous sections will be known as "threshold objects."

A comparison of the OTS for percentile objects minus the threshold objects can be seen in Fig. 17, where negative values indicate the OTS is higher for the threshold objects. Aside from the simulated BTs for the tornado event, the percentile objects generally have a lower OTS on average than the threshold objects. Thus, in these three events, the percentile objects would be considered less skillful than the threshold objects over the event lifetimes. While this result appears counterintuitive, it is consistent with a previous study (Griffin et al. 2021). One reason for the increased OTS for the simulated BTs during the tornado event is the increase in the percent of paired observation area, probably because the higher simulated BT threshold for the percentile object resulted in a much larger area encompassed by forecast objects. This larger area, though, is still paired with observation objects. The average interest score is also slightly larger (difference of 0.037), mainly driven by the higher intersection area ratio interest score (difference of 0.021). In comparison, the areal extent of the radar reflectivity percentile objects is smaller than the threshold objects due to the higher thresholds for all three events. All events, therefore, saw a decrease in the percent area of paired observation objects, which negatively impacted the OTS. The average interest score is also lower or similar for percentile radar objects than threshold objects, partially due to a lower intersection area ratio interest score from the smaller forecast objects.

Further accounting for HRRR biases can increase the OTS compared to the threshold objects for some events. A few patterns in the difference in OTS in Fig. 17 can be observed. First, the OTS for radar objects is lower for the percentile objects than the threshold objects at the beginning of the snow

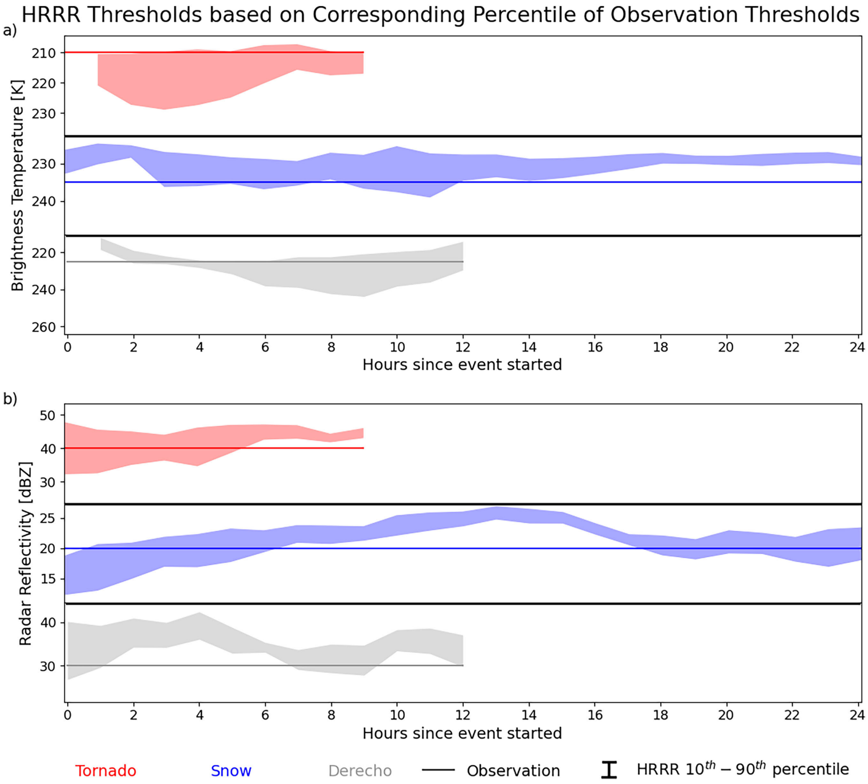


FIG. 16. (a) HRRR BT corresponding to the percentile of the observed BT (envelope representing 10th–90th percentile over the forecast hours valid at the given time). (b) As in (a), but for radar reflectivity.

event when HRRR threshold objects are present [HRRR did not identify objects at the beginning of the snow event (Fig. 10b)] and the end of the event for later forecast hours. Since the HRRR radar reflectivity threshold used to define the percentile objects is higher than the observations, but the HRRR radar reflectivity in Fig. 16 is similar or lower than the observation threshold at the beginning and end of the snow event, defining the HRRR threshold for objects based on valid time (hereafter referred to as varying percentile objects) does result in an increased OTS compared to threshold radar objects for the beginning and the end of the snow event (supplemental Fig. 2b). Varying percentile objects have a higher average OTS for both BTs and radar reflectivity compared to the percentile objects in the tornado event; however, the increase is quite small (<0.01). Finally, the OTS is lower for BT objects for early forecast hours of the snow event. This is probably due to BT spin-up as noted by Griffin et al. (2017a). Defining a HRRR threshold based on the percentile corresponding to the observed BT for each HRRR forecast hour (hereafter referred to as FH varying percentile objects) does increase these OTS values compared to the threshold objects (supplemental Fig. 3a). Thus, further accounting for the HRRR biases can increase OTS. However, even the FH varying percentile objects still have a lower OTS for BT objects at the end of the snow event. FH varying percentile objects have a higher average OTS for both the tornado and derecho events as well when compared to the percentile objects (0.05 and 0.01, respectively). However, accounting for these biases could be considered unsustainable in operational usage, and the OTS for these percentile and varying percentile objects is only more accurate on average compared to the threshold OTS for half of the fields and events.

TABLE 5. HRRR simulated BT corresponding to the percentile of the observation object BT threshold over the entire event.

	Tornado	Snow	Derecho
Observation BT (K)	210	235	225
Average HRRR BT (K)	214.3	230	225.9

5. Summary and conclusions

In this study, object-based methods for assessing cloud and precipitation characteristics were assessed, by comparing

TABLE 6. HRRR simulated radar reflectivity corresponding to the percentile of the observation object radar reflectivity threshold over the entire event.

	Tornado	Snow	Derecho
Observation radar reflectivity (dBZ)	40	20	30
Average HRRR radar reflectivity (dBZ)	42.6	21.4	34.8

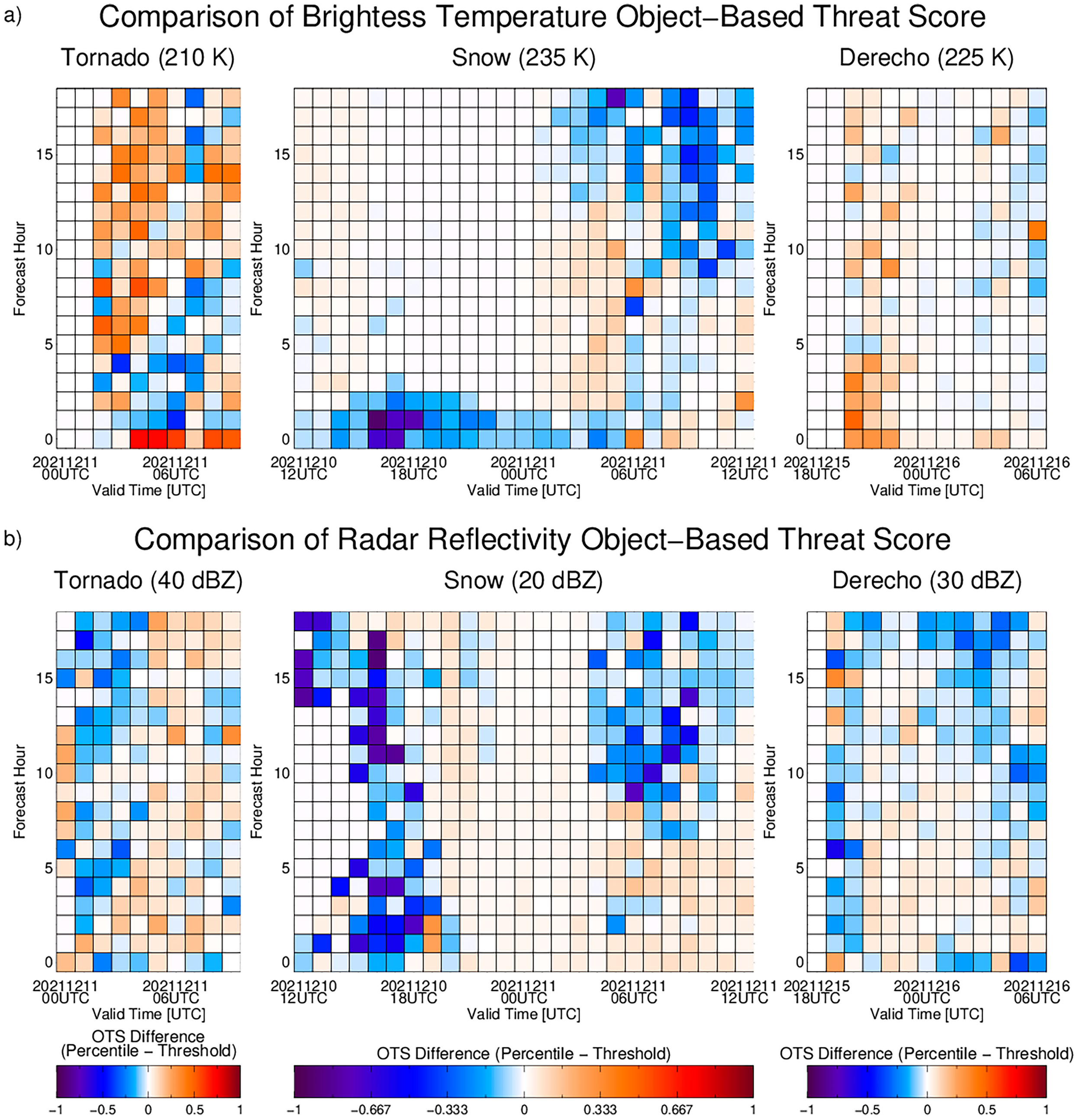


FIG. 17. (a) Difference in OTS for HRRR-simulated BT percentile objects minus the threshold objects. Negative values indicate the OTS is higher for the threshold objects. (b) Difference in OTS for HRRR-simulated radar reflectivity percentile objects minus the threshold objects.

simulated HRRR BTs and composite radar reflectivity to each other and their corresponding observations. This was done for during three weather phenomena which occurred during one week in December 2021: the Mayfield, Kentucky, tornado on 11 December 2021, heavy snow in Minnesota from 10 to 11 December 2021, and the Midwest derecho on 15 December 2021. The accuracy of simulated BTs and composite radar reflectivity was assessed using objects defined based on the observation thresholds. After subjectively determining a radar reflectivity threshold for each event, MvMODE was used to define objects based on that radar reflectivity threshold and 11 different BT thresholds. The optimal BT threshold was determined based on a comparison of the areal extent of objects where both observed radar AND satellite fields exceeding the given threshold and objects where either the radar OR satellite fields exceeding the given threshold. Finally, objects defined based on the HRRR value corresponding to the percentile of the observation threshold were also assessed.

Overall, by using the centroid of MODE objects, it was found that the HRRR model accurately depicted the spatial displacement between BTs and composite radar reflectivity. The centroid longitude of observation satellite objects is to the east of radar objects for the tornado and derecho event, and this orientation is accurately depicted by the HRRR objects. For the snow event, the centroid longitude of observation satellite objects is east of the radar objects at the beginning of the event and west at the end, which is also captured by the HRRR objects. However, the center of HRRR simulated satellite objects was south of the radar objects at the beginning of the tornado event, while the center of the satellite observations is to the north of the radar object. Thus, HRRR struggled a bit capturing this orientation between satellite and radar objects in this study at convective initiation. There are many instances of the southward displacement between simulated HRRR satellite and radar objects being associated with lower interest score pairs for early forecast lead times based on the correlation between forecast hour and displacement; however, higher interest score pairs also exhibited this southward displacement.

While HRRR did not fully capture the orientation between satellite and radar objects for the snow event, the OTS indicates HRRR had the highest accuracy for both satellite and radar objects during this event. The HRRR forecasts for the tornado event had the lowest OTS and are therefore the least accurate. By analyzing the components of the OTS: percent of observation and forecast area paired and the average interest score, it can be seen that the snow event is more accurate because it has a higher value for those components compared to the tornado and derecho events. Breaking down the interest score into its four main components reveals that the simulated satellite objects have a higher OTS for the snow event than the radar objects because the average interest score for radar objects is reduced compared to the satellite objects. This is consistent with the areal extent of the simulated radar reflectivity being much larger than the observations. While using objects defined based on the percentile of radar reflectivity reduces the areal extent of radar reflectivity objects, the OTS is still lower, on average, than when defining objects using the observation thresholds. Analyzing the components of the OTS indicates the lower value with the percentile objects, those defined using a threshold based on the HRRR biases over the full event, is due to a lower percentage of paired observation object area. This is true for satellite and radar objects for all events, except for the satellite objects during the tornado event.

Another outcome of this work is the hypothesis that while it is difficult to use simulated satellite BTs as an exact proxy for simulated radar reflectivity, displacements between observation BTs and radar reflectivity are accurately depicted by the HRRR model. For example, the location of observation and forecast satellite and radar objects are similar based on the displacement between the object centers. Thus, if the simulated BT object is accurately located, this narrow study suggests the location of the simulated radar reflectivity is also accurate. However, the intensity of the observed and simulated BT and radar reflectivity objects is not similar. HRRR generally had higher simulated BTs for the tornado event, which are less intense than the observations, but the simulated composite radar reflectivity was higher and more intense. One hypothesis for the lower, more intense BTs in the snow event is the lower HRRR atmospheric temperatures compared to the observations could be decreasing the simulated BTs compared to the observations, though lower simulated BTs are observed in the derecho event even with higher HRRR temperatures. Future work could analyze if lower/higher simulated BTs are coincident with lower/higher simulated atmospheric temperatures in a larger dataset.

As this study only examines weather phenomena from a single week, future work includes applying this methodology to expand the areas and length of analysis and examining if the tendencies noted in this study exist in a larger dataset. For example, if the simulated BT object is accurately located, is the location of the simulated radar reflectivity also accurate? A larger study can confirm if this relationship exists more broadly. While different radar reflectivity and BT thresholds may need to be used in a study like this, MvMODE could be used to establish the corresponding BT (radar reflectivity) threshold for a given radar reflectivity (BT) threshold. Also, does accounting for biases in the HRRR simulated radar reflectivity and BTs over the full event reduce skill? Additional future work includes expanding this methodology to the Rapid Refresh Forecast System (RRFS), which is targeted to replace HRRR in 2024. RRFS model output will include BTs for all of the *GOES-16* and *GOES-18* IR bands, as well as lightning data to compare to the *GOES-16* lightning mapper (GLM).

Acknowledgments. Funding for this project was provided by the NOAA Joint Technology Transfer Initiative (JTTI) program via Grants NA19OAR4590233 and NA21OAR4590168. Thank you to three anonymous reviewers for their comments and insights that have substantially improved the quality of the manuscript.

Data availability statement. *GOES-16* BTs and MRMS composite radar reflectivity are available at https://noaa-goes16.s3.amazonaws.com/index.html and https://noaa-mrms-pds.s3.amazonaws.com/index.html#CONUS/MergedReflectivityQCComposite_00.50/, respectively. HRRR simulated BTs (SBTXXXX) and composite radar reflectivity (REFC) are available in the "wrfsfc" files at https://noaa-hrrr-bdp-pds.s3.amazonaws.com/index.html. Temperature and pressure profiles are available at https://weather.uwyo.edu/upperair/sounding.html.

REFERENCES

Bedka, K., J. Brunner, R. Dworak, W. Feltz, J. Otkin, and T. Greenwald, 2010: Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients. *J. Appl. Meteor. Climatol.*, **49**, 181–202, https://doi.org/10.1175/2009JAMC2286.1.

Bikos, D., and Coauthors, 2012: Synthetic satellite imagery for real-time high-resolution model evaluation. *Wea. Forecasting*, **27**, 784–795, https://doi.org/10.1175/WAF-D-11-00130.1.

Blahak, U., 2012: RADAR_MIE_LM and RADAR_MIELIB—Calculation of radar reflectivity from model output. Tech. Doc., 167 pp., https://download.dwd.de/pub/DWD/Forschung_und_Entwicklung/RADVOP/mielib.pdf.

Blank, L. R., R. Bullock, D. Fillmore, J. Halley Gotway, T. L. Jensen, and J. K. Wolff, 2021: Automated identification of drylines in the United States using multivariate MODE. *11th Conf. on Transition of Research to Operations*, online, Amer. Meteor. Soc., 5A.9, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/383663.

Brooks, H. E., C. A. Doswell III, and R. B. Wilhelmson, 1994: The role of midtropospheric winds in the evolution and maintenance of low-level mesocyclones. *Mon. Wea. Rev.*, **122**, 126–136, https://doi.org/10.1175/1520-0493(1994)122<0126:TROMWI>2.0.CO;2.

Bullock, R. G., B. G. Brown, and T. L. Fowler, 2016: Method for object-based diagnostic evaluation. NCAR Tech. Note NCAR/TN-532+STR, 84 pp., https://doi.org/10.5065/D61V5CBS.

Bytheway, J. L., and C. D. Kummerow, 2015: Toward an object-based assessment of high-resolution forecasts of long-lived convective precipitation in the central U.S. *J. Adv. Model. Earth Syst.*, **7**, 1248–1264, https://doi.org/10.1002/2015MS000497.

——, ——, and C. Alexander, 2017: A features-based assessment of the evolution of warm season precipitation forecasts from the HRRR Model over three years of development. *Wea. Forecasting*, **32**, 1841–1856, https://doi.org/10.1175/WAF-D-17-0050.1.

Cai, H., and R. E. Dumais Jr., 2015: Object-based evaluation of a numerical weather prediction model's performance through forecast storm characteristic analysis. *Wea. Forecasting*, **30**, 1451–1468, https://doi.org/10.1175/WAF-D-15-0008.1.

Cintineo, R., J. A. Otkin, M. Xue, and F. Kong, 2014: Evaluating the performance of planetary boundary layer and cloud microphysical parameterization schemes in convection-permitting ensemble forecasts using synthetic *GOES-13* satellite observations. *Mon. Wea. Rev.*, **142**, 163–182, https://doi.org/10.1175/MWR-D-13-00143.1.

Davis, C., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

——, ——, and ——, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795, https://doi.org/10.1175/MWR3146.1.

——, ——, ——, and J. Halley-Gotway, 2009: The Method for Object-based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC spring program. *Wea. Forecasting*, **24**, 1252–1267, https://doi.org/10.1175/2009WAF2222241.1.

DeHaan, L. L., A. C. Martin, R. R. Weihs, L. Delle Monache, and F. M. Ralph, 2021: Object-based verification of atmospheric river predictions in the northeast Pacific. *Wea. Forecasting*, **36**, 1575–1587, https://doi.org/10.1175/WAF-D-20-0236.1.

Ding, S., P. Yang, F. Weng, Q. Liu, Y. Han, P. Van Delst, J. Li, and B. Baum, 2011: Validation of the community radiative transfer model. *J. Quant. Spectrosc. Radiat. Transfer*, **112**, 1050–1064, https://doi.org/10.1016/j.jqsrt.2010.11.009.

Dougherty, K. J., J. D. Horel, and J. E. Nachamkin, 2021: Forecast skill for California heavy precipitation periods from the High-Resolution Rapid Refresh Model and the coupled ocean–atmosphere mesoscale prediction system. *Wea. Forecasting*, **36**, 2275–2288, https://doi.org/10.1175/WAF-D-20-0182.1.

Duda, J. D., and D. D. Turner, 2021: Large-sample application of radar reflectivity object-based verification to evaluate HRRR warm-season forecasts. *Wea. Forecasting*, **36**, 805–821, https://doi.org/10.1175/WAF-D-20-0203.1.

English, J. M., D. D. Turner, T. I. Alcott, W. R. Moninger, J. L. Bytheway, R. Cifelli, and M. Marquis, 2021: Evaluating operational and experimental HRRR model forecasts of atmospheric river events in California. *Wea. Forecasting*, **36**, 1925–1944, https://doi.org/10.1175/WAF-D-21-0081.1.

Ganetis, S. A., B. A. Colle, S. E. Yuter, and N. P. Hoban, 2018: Environmental conditions associated with observed snowband structures within northeast U.S. winter storms. *Mon. Wea. Rev.*, **146**, 3675–3690, https://doi.org/10.1175/MWR-D-18-0054.1.

Griffin, S. M., J. A. Otkin, C. M. Rozoff, J. M. Sieglaff, L. M. Cronce, and C. R. Alexander, 2017a: Methods for comparing simulated and observed satellite infrared brightness temperatures and what do they tell us? *Wea. Forecasting*, **32**, 5–25, https://doi.org/10.1175/WAF-D-16-0098.1.

——, ——, ——, ——, ——, ——, T. L. Jensen, and J. K. Wolff, 2017b: Seasonal analysis of cloud objects in the High-Resolution Rapid Refresh (HRRR) Model using object-based verification. *J. Appl. Meteor. Climatol.*, **56**, 2317–2334, https://doi.org/10.1175/JAMC-D-17-0004.1.

——, ——, G. Thompson, M. Frediani, J. Berner, and F. Kong, 2020: Assessing the impact of stochastic perturbations in cloud microphysics using *GOES-16* infrared brightness temperatures. *Mon. Wea. Rev.*, **148**, 3111–3137, https://doi.org/10.1175/MWR-D-20-0078.1.

——, ——, S. E. Nebuda, T. L. Jensen, P. S. Skinner, E. Gilleland, T. A. Supinie, and M. Xue, 2021: Evaluating the impact of planetary boundary layer, land surface model, and microphysics parameterization schemes on cold cloud objects in simulated GOES-16 brightness temperatures. *J. Geophys. Res. Atmos.*, **126**, e2021JD034709, https://doi.org/10.1029/2021JD034709.

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, 33 pp., https://www.star.nesdis.noaa.gov/pub/smcd/jcsda/qliu/REL_2.1.3_June21_2013/CRTM_v1-NOAA_Tech_Report_NESDIS122.pdf.

Hanna, J. W., D. M. Schultz, and A. R. Irving, 2008: Cloud-top temperatures for precipitating winter clouds. *J. Appl. Meteor. Climatol.*, **47**, 351–359, https://doi.org/10.1175/2007JAMC1549.1.

Heidinger, A., and Y. Li, 2018: AWG Cloud Height Algorithm (ACHA). NOAA NESDIS Center for Satellite Applications and Research Algorithm Theoretical Basis Doc., version 3.2, 60 pp., https://www.ssec.wisc.edu/~daves/ACHA_ATBD.pdf.

Henderson, D. S., J. A. Otkin, and J. R. Mecikalski, 2021: Evaluating convective initiation in high-resolution numerical weather prediction models using *GOES-16* infrared brightness temperatures. *Mon. Wea. Rev.*, **149**, 1153–1172, https://doi.org/10.1175/MWR-D-20-0272.1.

Iacono, M. J., 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.

Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating High-Resolution Rapid Refresh Model. *Wea. Forecasting*, **28**, 921–939, https://doi.org/10.1175/WAF-D-12-00085.1.

——, ——, and G. Thompson, 2017: Examination of mixed-phase precipitation forecasts from the High-Resolution Rapid Refresh Model using surface observations and sounding data. *Wea. Forecasting*, **32**, 949–967, https://doi.org/10.1175/WAF-D-16-0171.1.

James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, https://doi.org/10.1175/WAF-D-21-0130.1.

Jensen, T., B. Brown, R. Bullock, T. Fowler, J. Halley Gotway, and K. Newman, 2020: Model evaluation tools version 9.0.1 user's guide. Developmental Testbed Center, 482 pp., https://dtcenter.org/sites/default/files/community-code/met/docs/user-guide/MET_Users_Guide_v9.0.1.pdf.

Ji, L., X. Zhi, C. Simmer, S. Zhu, and Y. Ji, 2020: Multimodel ensemble forecasts of precipitation based on an object-based diagnostic evaluation. *Mon. Wea. Rev.*, **148**, 2591–2606, https://doi.org/10.1175/MWR-D-19-0266.1.

Johnson, A., and X. Wang, 2013: Object-based evaluation of a storm-scale ensemble during the 2009 NOAA Hazardous Weather Testbed Spring Experiment. *Mon. Wea. Rev.*, **141**, 1079–1098, https://doi.org/10.1175/MWR-D-12-00140.1.

Jones, T. A., P. Skinner, K. Knopmeier, E. Mansell, P. Minnis, R. Palikonda, and W. Smith Jr., 2018: Comparison of cloud microphysics schemes in a Warn-on-Forecast System using synthetic satellite objects. *Wea. Forecasting*, **33**, 1681–1708, https://doi.org/10.1175/WAF-D-18-0112.1.

——, and Coauthors, 2020: Assimilation of *GOES-16* radiances and retrievals into the Warn-on-Forecast System. *Mon. Wea. Rev.*, **148**, 1829–1859, https://doi.org/10.1175/MWR-D-19-0379.1.

Nakanishi, M., and H. Niino, 2004: An improved Mellor–Yamada Level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1–31, https://doi.org/10.1023/B:BOUN.0000020164.04146.98.

——, and ——, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912.

Otkin, J. A., T. J. Greenwald, J. Sieglaff, and H.-L. Huang, 2009: Validation of a large-scale simulated brightness temperature dataset using SEVIRI satellite observations. *J. Appl. Meteor. Climatol.*, **48**, 1613–1626, https://doi.org/10.1175/2009JAMC2142.1.

Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the High-Resolution Rapid Refresh Model's ability to predict mesoscale convective systems using object-based evaluation. *Wea. Forecasting*, **30**, 892–913, https://doi.org/10.1175/WAF-D-14-00118.1.

Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Radford, J. T., G. M. Lackmann, and M. A. Baxter, 2019: An evaluation of snowband predictability in the High-Resolution Rapid Refresh. *Wea. Forecasting*, **34**, 1477–1494, https://doi.org/10.1175/WAF-D-19-0089.1.

Rasmussen, E. N., and J. M. Straka, 1998: Variations in supercell morphology. Part I: Observations of the role of upper-level storm-relative flow. *Mon. Wea. Rev.*, **126**, 2406–2421, https://doi.org/10.1175/1520-0493(1998)126<2406:VISMPI>2.0.CO;2.

Senf, F., and H. Deneke, 2017: Uncertainties in synthetic Meteosat SEVIRI infrared brightness temperatures in the presence of cirrus clouds and implications for evaluation of cloud microphysics. *Atmos. Res.*, **183**, 113–129, https://doi.org/10.1016/j.atmosres.2016.08.012.

——, D. Klocke, and M. Brueck, 2018: Size-resolved evaluation of simulated deep tropical convection. *Mon. Wea. Rev.*, **146**, 2161–2182, https://doi.org/10.1175/MWR-D-17-0378.1.

Skinner, P. S., L. J. Wicker, D. M. Wheatley, and K. H. Knopfmeier, 2016: Application of two spatial verification methods to ensemble forecasts of low-level rotation. *Wea. Forecasting*, **31**, 713–735, https://doi.org/10.1175/WAF-D-15-0129.1.

——, 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.

Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, https://doi.org/10.1175/MWR-D-15-0198.1.

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.

Squitieri, B. J., and W. A. Gallus Jr., 2020: On the forecast sensitivity of MCS cold pools and related features to horizontal grid spacing in convection-allowing WRF simulations. *Wea. Forecasting*, **35**, 325–346, https://doi.org/10.1175/WAF-D-19-0016.1.

Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1.

——, M. Tewari, K. Ikeda, S. Tessendorf, C. Weeks, J. Otkin, and F. Kong, 2016: Explicitly-coupled cloud physics and radiation parameterizations and subsequent evaluation in WRF high-resolution convective forecasts. *Atmos. Res.*, **168**, 92–104, https://doi.org/10.1016/j.atmosres.2015.09.005.

Van Weverberg, K., and Coauthors, 2013: The role of cloud microphysics parameterization in the simulation of mesoscale convective system clouds and precipitation in the tropical western Pacific. *J. Atmos. Sci.*, **70**, 1104–1128, https://doi.org/10.1175/JAS-D-12-0104.1.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, https://doi.org/10.1175/WAF-D-13-00135.1.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.