Seasonal Analysis of Cloud Objects in the High-Resolution Rapid Refresh (HRRR) Model Using Object-Based Verification

SARAH M. GRIFFIN AND JASON A. OTKIN

Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin

CHRISTOPHER M. ROZOFF

Research Applications Laboratory, National Center for Atmospheric Research,[a] Boulder, Colorado

JUSTIN M. SIEGLAFF AND LEE M. CRONCE

Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin

CURTIS R. ALEXANDER

NOAA/Earth System Research Laboratory, Boulder, Colorado

TARA L. JENSEN AND JAMIE K. WOLFF

Research Applications Laboratory, National Center for Atmospheric Research, and Developmental Testbed Center, Boulder, Colorado

ABSTRACT

In this study, object-based verification using the method for object-based diagnostic evaluation (MODE) is used to assess the accuracy of cloud-cover forecasts from the experimental High-Resolution Rapid Refresh (HRRRx) model during the warm and cool seasons. This is accomplished by comparing cloud objects identified by MODE in observed and simulated Geostationary Operational Environmental Satellite 10.7-μm brightness temperatures for August 2015 and January 2016. The analysis revealed that more cloud objects and a more pronounced diurnal cycle occurred during August, with larger object sizes observed in January because of the prevalence of synoptic-scale cloud features. With the exception of the 0-h analyses, the forecasts contained fewer cloud objects than were observed. HRRRx forecast accuracy is assessed using two methods: traditional verification, which compares the locations of grid points identified as observation and forecast objects, and the MODE composite score, an area-weighted calculation using the object-pair interest values computed by MODE. The 1-h forecasts for both August and January were the most accurate for their respective months. Inspection of the individual MODE attribute interest scores showed that, even though displacement errors between the forecast and observation objects increased between the 0-h analyses and 1-h forecasts, the forecasts were more accurate than the analyses because the sizes of the largest cloud objects more closely matched the observations. The 1-h forecasts from August were found to be more accurate than those during January because the spatial displacement between the cloud objects was smaller and the forecast objects better represented the size of the observation objects.

1. Introduction

Because detailed information about the horizontal distribution of clouds can be obtained from satellite infrared brightness temperatures (BT), satellites' BTs have been used in prior studies to evaluate the accuracy of the cloud field in high-resolution forecasts from numerical weather prediction models (e.g., Morcrette 1991; Otkin and Greenwald 2008; Otkin et al. 2009; Cintineo et al. 2014; Thompson et al. 2016; Griffin et al. 2017; Grasso and Greenwald 2004; Grasso et al. 2008, 2014; Feltz et al. 2009; Bikos et al. 2012; Jankov et al. 2011; Van Weverberg et al. 2013; Jin et al. 2014). Differences between the observed and forecast cloud fields

DOI: 10.1175/JAMC-D-17-0004.1

can be quantified using traditional verification metrics such as root-mean-square error (Wilks 2006) or neighborhood-based statistics such as the fractions skill score (Roberts and Lean 2008; Söhne et al. 2006). Even though these verification methods provide useful information concerning the model accuracy, additional information about errors in the spatial distribution of the cloud field can be obtained through use of more-sophisticated object-based verification tools such as the method for object-based diagnostic evaluation (MODE; Davis et al. 2006a,b).

Several recent studies have employed MODE to assess forecast skill. MODE has been used to evaluate Rossby waveguides from the Integrated Forecast System and ERA-Interim dataset from the European Centre for Medium-Range Weather Forecasts (Giannakaki and Martius 2016). Davis et al. (2009) used MODE to compare the 1-h rainfall accumulation from separate dynamic cores of the Weather Research and Forecasting Model. MODE has also been used to assess precipitation forecasts from the High-Resolution Rapid Refresh (HRRR; Bytheway and Kummerow 2015; Cai and Dumais 2015) as well as the Global Forecast System and North American Mesoscale models (Wolff et al. 2014).

Although prior studies have demonstrated the utility of satellite infrared BTs for cloud verification, object-based verification tools until recently have not typically been used to assess the forecast accuracy of clouds. Mittermaier and Bullock (2013) used MODE to assess the spatial and temporal characteristics of cloud-cover forecasts in high-resolution models over the United Kingdom. Their study, however, focused on evaluating the characteristics of clear objects within the satellite imagery. They showed that object-based verification provides a useful way to evaluate different attributes of the forecast. Griffin et al. (2017) most recently assessed the accuracy of the experimental HRRR (HRRRx) model using simulated BTs and traditional, neighborhood-based, and object-based statistics. They found that object-based verification provides information about the simulated cloud structure, such as errors in the location, shape, orientation, and spatial extent of the cloud field, that is not readily obtained using traditional or neighborhood-based verification metrics. Therefore, object-based statistics provide a more unique assessment of the forecast accuracy.

The purpose of the paper is to assess the accuracy of the HRRRx cloud-cover forecasts during two 1-month periods covering August 2015 and January 2016. These periods were chosen for this analysis to investigate potential differences in cloud characteristics during the warm and cool seasons. Given the fine spatial resolution

(3 km) of the HRRRx model, qualitatively small differences between the observed and forecast cloud fields may lead to large quantitative errors that differ greatly from a subjective forecaster evaluation of the model forecast accuracy. Therefore, these cloud-field differences are quantified using MODE instead of traditional or neighborhood-based statistics, which mainly measure spatial displacement between the observed and forecast cloud fields (Griffin et al. 2017). Using the method of assessing model cloud cover by comparing simulated and observed infrared BT, model cloud validation will be accomplished by comparing objects created from HRRRx simulated infrared BTs with objects created from observed infrared BT.

The paper is organized as follows: The datasets used in this study are described in section 2, and MODE will be described in more detail in section 3. The method used to evaluate the HRRRx forecast accuracy will be presented in section 4. Results and conclusions will be presented in sections 5 and 6, respectively.

2. Data

a. Simulated brightness temperatures from the High-Resolution Rapid Refresh model

This study uses simulated *Geostationary Operational Environmental Satellite 13* (*GOES-13*) 10.7-μm BTs computed using output from the HRRRx model. The 10.7-μm BTs are sensitive to cloud ice particles and thus are an ideal dataset to assess the accuracy of the cloud field in the upper troposphere. The HRRRx is an hourly updating model that covers the contiguous United States (CONUS) with 3-km horizontal grid spacing and 51 vertical levels. The version of the HRRRx model used for this study was implemented at the Earth System Research Laboratory on 4 May 2015 (http://rapidrefresh.noaa.gov/). It uses initial conditions from the Rapid Refresh (RAP) model and then applies data assimilation at 3-km resolution, including the assimilation of radar reflectivity. The HRRRx is a convection-allowing model that does not include deep convective parameterization (Benjamin et al. 2016). The HRRRx uses the Thompson aerosol-aware, version 3.6.1, microphysics scheme (Thompson and Eidhammer 2014); 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 (RRTMG) shortwave and longwave radiation schemes (Iacono et al. 2008; http://rapidrefresh.noaa.gov/).

Simulated *GOES-13* BTs are computed for forecast hours 0–24 and are valid at the beginning of the hour.

These BTs are computed using HRRRx output and version 2.0.7 of the Community Radiative Transfer Model (CRTM; Han et al. 2006) in the Unified Post Processor (UPP). The UPP incorporates the correct *GOES-13* viewing-angle geometry. For clear grid points, simulated BTs are computed using several model-predicted fields, including surface skin temperature, 10-m wind speed, pressure, and vertical profiles of temperature and water vapor. For cloudy grid points, additional information about cloud radiative properties is required to simulate BTs. Vertical profiles of mixing ratio and number concentration are used to compute the effective particle diameters for each hydrometeor species (cloud water, cloud ice, rainwater, snow, and graupel) predicted by the Thompson aerosol-aware microphysics scheme. In the CRTM, standard lookup tables for cloud absorption and scattering properties, such as extinction, single-scatter albedo, and the full scattering phase function, are used to assign values to each hydrometeor species as a function of the cloud effective diameter computed using the particle size distribution assumptions for that scheme (e.g., Otkin et al. 2007). Cloud optical properties are computed for each species and model layer and then are combined into an effective set of properties for each layer before computing the simulated infrared BTs.

b. Observed brightness temperatures

The satellite validation data used during this study are from the *GOES-13* imager. The 10.7-μm *GOES-13* BTs have a 4-km spatial resolution at nadir and are remapped to the 3-km HRRRx grid using an area-weighted average of all of the observed pixels that overlap a given HRRRx grid box. The *GOES-13* imager typically completes a scan over the CONUS every 15 min, and simulated HRRR BTs will be compared with the scan beginning at the top of the hour. An exception occurs every 3 h (0000, 0300 UTC, etc.), when the scan at the top of the hour is skipped so that the full disk scan that started 15 min earlier can be completed. In these instances, simulated HRRR BTs will be compared with the *GOES-13* scan starting 15 min prior to the HRRRx forecast time. Although this may introduce some uncertainty into the analysis, it is expected to be minor.

c. Seasonal comparison

To assess the HRRRx forecast accuracy, simulated and observed *GOES-13* 10.7-μm BTs from two 1-month-long periods, 1–31 August 2015 and 1–31 January 2016, were evaluated. These time periods were chosen so that the forecast accuracy of the HRRRx model could be assessed during both warm and cool seasons given

potential differences in meteorological regimes and associated cloud characteristics. Representative snapshots of the *GOES-13* 10.7-μm BT are shown in Fig. 1. The image on the left is from 2 August 2015, and the image on the right is from 22 January 2016. Cloud features in the upper troposphere associated with the coldest BTs were generally smaller in August than in January, although both large and small objects occurred during both months.

3. Method for object-based diagnostic evaluation

This study uses MODE (Davis et al. 2006a,b) to analyze the accuracy of the HRRRx forecast cloud field. MODE is a technique for identifying and matching objects in two different fields (e.g., observation and model forecast fields). These objects represent "regions of interest," which for this study are upper-level cloud features that contain the coldest infrared BTs. The MODE process is fully described in Davis et al. (2006a); a short outline as applied to cloud systems is provided here for context:

1) Smooth and threshold the forecast and observed BT fields using a process called convolution thresholding to identify objects.
2) Calculate various object attributes for each observed and forecast cloud object.
3) Match forecast and observed cloud objects using a fuzzy-logic algorithm and calculate attributes of paired objects, such as intersection area and distance.
4) Output attributes for individual objects and matched object pairs for assessment.

The convolution radius used for both the observed and forecast fields in step 1 was chosen to be five grid points (15 km) on the basis of Griffin et al. (2017). This radius allows for the analysis of small-scale storms, since Cai and Dumais (2015) state a range of two–eight grid points as identifying convective storm objects in ~4-km-resolution radar imagery. Object merging, in which larger objects, identified with a lower threshold, fully contain multiple individual objects, was not performed in the individual observation and forecast fields.

Step 1 of the above process requires choosing a BT threshold to define the edges of the cloud objects. As seen in Fig. 1, however, cloud objects contain different BTs depending upon the season, weather regime, and location. In addition, Griffin et al. (2017) identified a cold bias in the HRRRx forecast BTs that varies as a function of forecast hour. To illustrate these differences, probability distributions for the observed and simulated 10.7-μm BT from August and January at forecast hours 0 and 3 are shown in Fig. 2. Although the coldest BTs
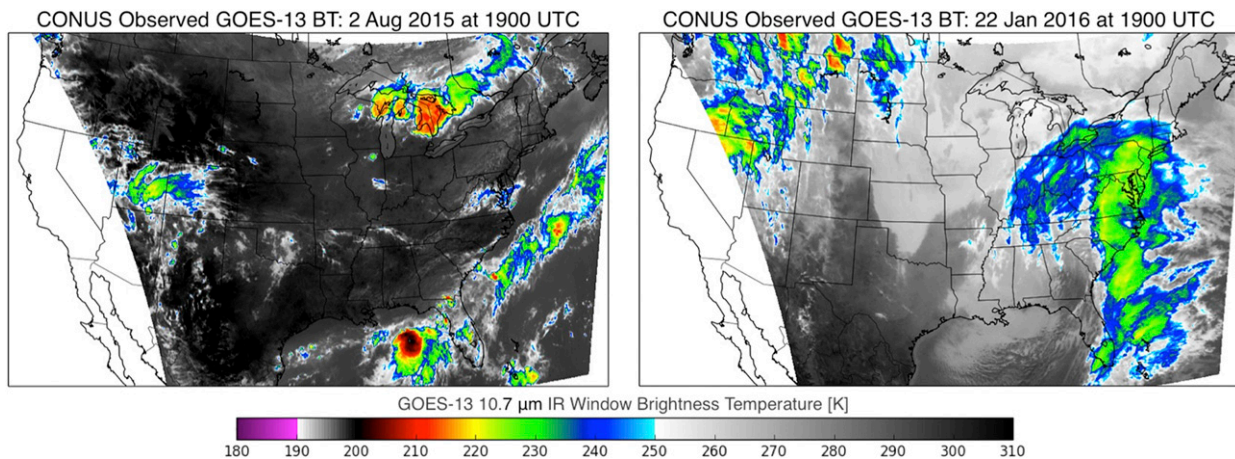
FIG. 1. Representative example IR-window BT image from the two months of interest in the study: (left) 1900 UTC 2 Aug 2015 and (right) 1900 UTC 22 Jan 2016.

occurred during August, the entire BT distribution is overall colder in January than it was during August. In addition, the coldest HRRRx simulated BTs are colder in the 3-h forecast than in the 0-h analysis, with the mean bias error from the 0-h analysis in August decreasing from 5.74 to 0.19 K by the 3-h forecast and the mean bias error from the 0-h analysis in January decreasing from 4.30 to 0.17 K by the 3-h forecast. Thus, to account for these seasonal and forecast-hour differences, the BT threshold used for each observed (forecast) BT field is based on the 10th percentile of the observed (simulated) BTs accumulated over the entire model domain covered by *GOES-13* observations during the 10-day period prior to and including the valid (forecast) time. In Fig. 1, the 10th percentile of *GOES-13* observed BTs is 252.0 K for 2 August 2015 and 241.0 K for 22 January 2016. The 10th percentile is used so that the analysis focuses on the coldest cloud objects occurring in the upper troposphere. Ten days are used to determine the BT threshold for upper-level clouds because it is possible that no upper-level clouds are present if a shorter time period is used. The threshold applied to the observed *GOES-13* BT is computed for each hour of the day to account for different cloud characteristics associated with the diurnal cycle. The threshold applied to the HRRRx simulated BT for a given HRRRx forecast is calculated from HRRRx forecasts with the same forecast hour and initialized at the same time of day as the given HRRRx forecast to account for any potential variations between different HRRRx initialization times. Since current versions of MODE are unable to compute the BT thresholds, they are calculated prior to running MODE.

When matching forecast and observation objects, MODE measures the correspondence between two objects by computing the interest value between matched objects (Developmental Testbed Center 2014). Interest values are a single number ranging from 0 to 1, with a perfect match having an interest value of 1. The interest value is a weighted combination of the object-pair attributes, where various object-pair attribute weights are defined by the user. The attributes and user-defined weights applied in this study are shown in Table 1. Attribute weights were chosen after testing multiple combinations using 5-h HRRRx forecasts from 1400 and 2000 UTC 23 July. The 5-h HRRRx forecasts are used to avoid potential errors associated with model spinup. Overall, the matching of observation and forecast objects in this analysis prioritizes the distance and size comparison between the objects. Two distance attributes are highly weighted relative to other attributes, with the minimum distance (boundary_dist) between the edges of objects having a slightly lower weight than the centroid distance (centroid_dist) that measures the distance between the objects' "centers of mass" so that more emphasis is placed on the displacement between the objects' centers of mass rather than their edges. Note, however, that, since the MODE centroid distance weight is the user-defined weight multiplied by the ratio of the objects' areas, the boundary distance between objects has greater weight when the ratio between the observation and forecast object area is less than 0.75. The area attributes receive the same user-defined weight as the distance attributes. The ratio of the objects' areas (area_ratio) is weighted higher than the ratio of the intersection area of objects to the observation or forecast object's area (int_area_ratio) because the int_area_ratio value can be artificially high when a small object is fully enclosed within a larger object.

Sample MODE objects defined using the observed and simulated BT thresholds from the 0-h HRRRx analyses are shown in Fig. 3. A solid black line surrounding
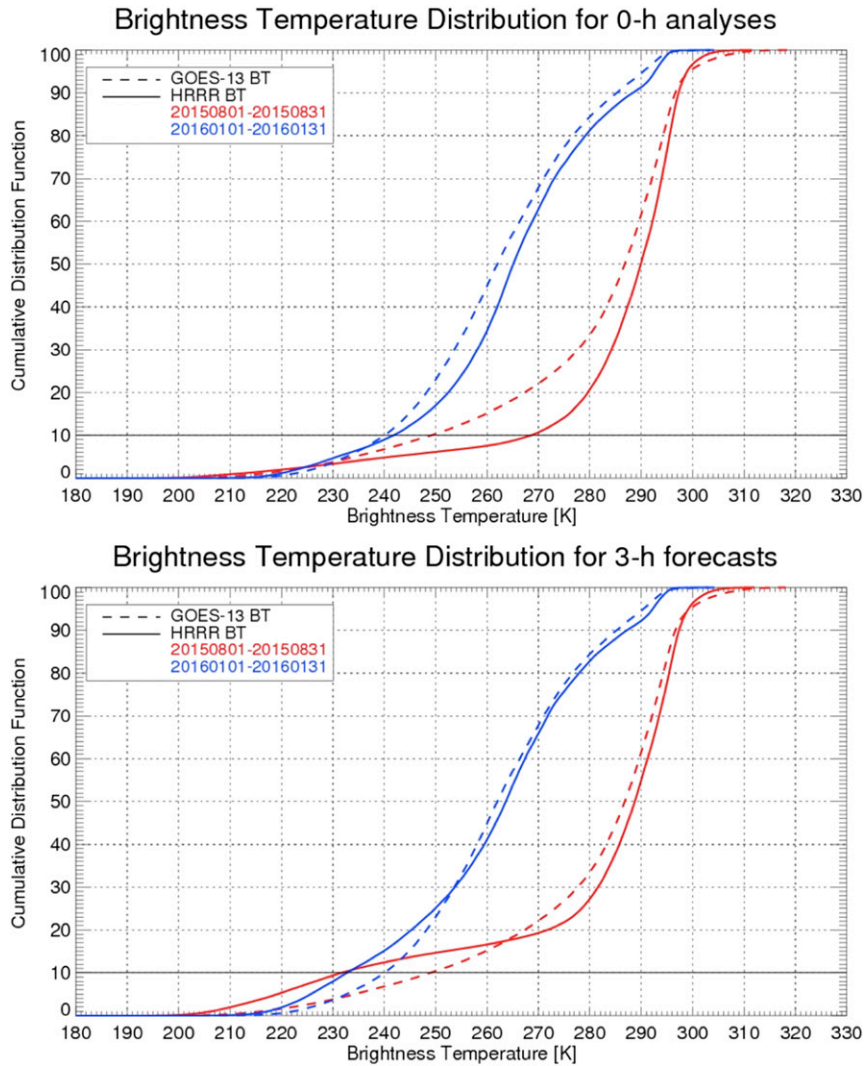
FIG. 2. Cumulative distribution of BT for August (red) and January (blue) for *GOES-13* observed BT (dashed curves, identical in both panels), along with the cumulative distribution of simulated HRRRx BT (solid curves) for (top) the 0-h analysis and (bottom) the 3-h forecast.

one or more objects, along with forecast and observation objects with the same color for a particular time, identifies an object cluster. A cluster is defined as a set of one or more observation (forecast) objects matching

the same one or more forecast (observation) objects (Developmental Testbed Center 2014). In this analysis, clusters are defined as object matches exceeding an interest value threshold of 0.65 [from Griffin et al. (2017)].

TABLE 1. User-defined weights and brief description of the object-pair attributes used in this analysis.

| Object pair attribute	User-defined weight (percent)	Description
Centroid_dist	4 (25.0)	Distance between objects' centers of mass
Boundary_dist	3 (18.75)	Min distance between the objects
Convex_hull_dist	1 (6.25)	Min distance between the polygons surrounding the objects
Angle_diff	1 (6.25)	Orientation-angle difference
Area_ratio	4 (25.0)	Ratio of the forecast and observation objects' areas (or its reciprocal, if it yields a lower value)
Int_area_ratio	3 (18.75)	Ratio of the objects' intersection area to the lesser of the observation or forecast area (whichever yields a lower value)
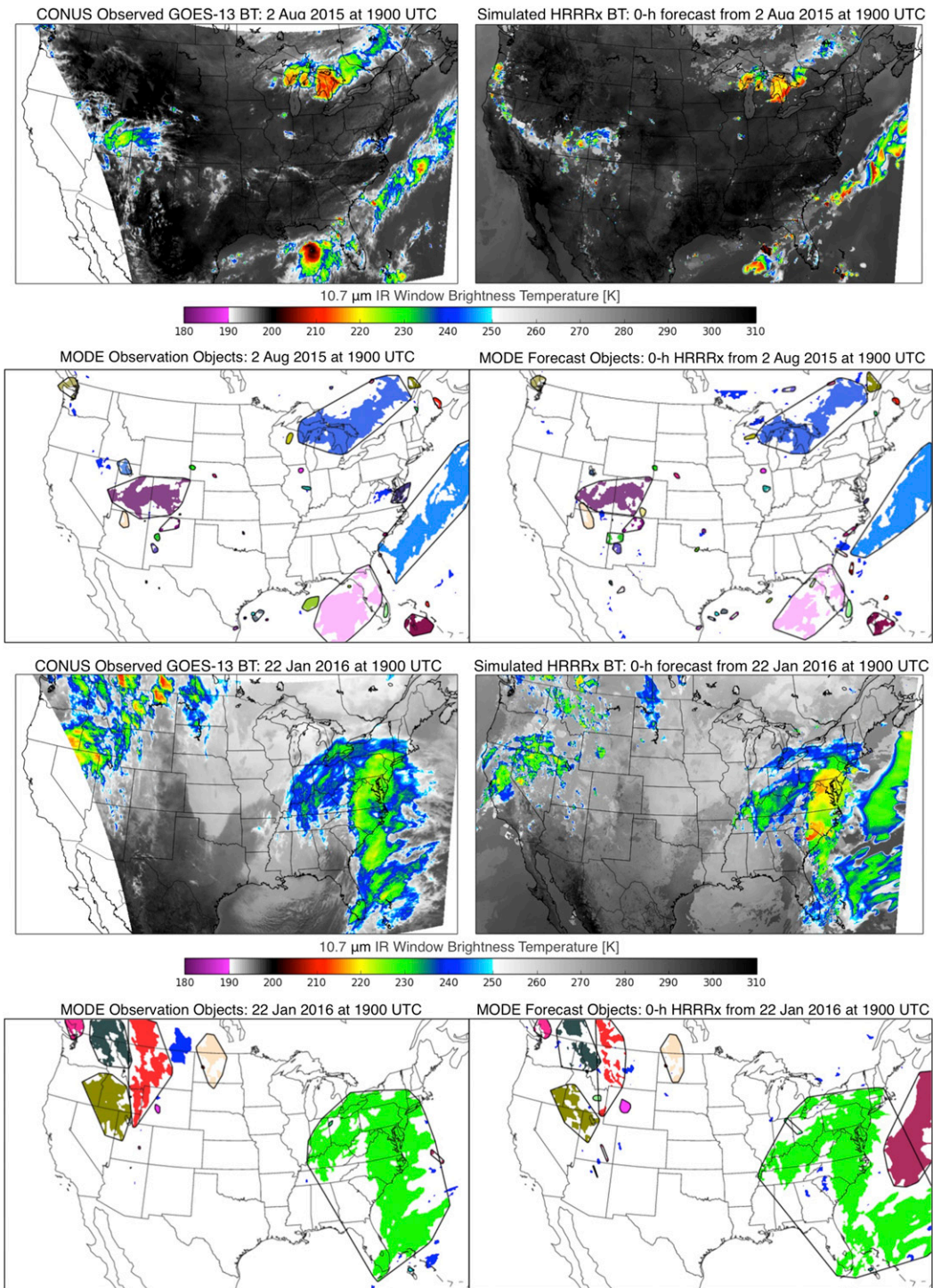
FIG. 3. (left) *GOES-13* observed and (right) HRRRx simulated IR-window BT images from (top) 1900 UTC 2 Aug 2015 and (bottom middle) 1900 UTC 22 Jan 2016, and the (top middle),(bottom) respective corresponding MODE objects.

TABLE 2. Relationship between HRRRx forecast object grid points and *GOES-13* observation object grid points in a 2 × 2 contingency table.

		GOES-13 observation object	
		Yes	No
HRRRx forecast object	Yes	*h*	*f*
	No	*m*	*z*

4. Method

a. Dichotomous (yes/no) verification

The relationship between the locations of MODE-identified observation and forecast cloud objects can be assessed using a standard 2 × 2 contingency table (Table 2) to depict HRRRx forecast accuracy. Thus, this analysis assesses overlapping observation and forecast object grid points, not matching object pairs. A grid point that is identified as both an observed and forecast cloud object is considered to be a hit, whereas a grid point that is not identified as either a forecast or observed object is considered to be a "correct no." A grid point that is identified as an observation cloud object but not a forecast object is considered to be a miss. A grid point that is identified as a forecast cloud object but not an observation object is considered to be a false alarm.

Output generated from the 2 × 2 contingency table will be displayed using a performance diagram that displays four different metrics on one image: probability of detection (POD), success ratio (SR), bias, and critical success index (CSI; Roebber 2009). The equations for these metrics are provided in Table 3 and are based on those found in Wilks (2006). The POD is the ratio of the number of observation object grid points coincident with an HRRRx object to the total number of observation object grid points. The SR is 1 minus the false-alarm ratio, the number of incorrect HRRRx object grid points divided by the total number of HRRRx object grid points. Both the POD and SR have a range from 0 to 1, with higher numbers indicating a better forecast. Bias is the ratio of the number of HRRRx object grid points to observation object grid points. Unbiased forecasts have a bias equal to 1, with biases of greater or less than 1 indicating that the HRRRx model is over-forecasting or underforecasting, respectively, the number of observation object grid points. Overall forecast accuracy is assessed using the CSI, the number of corresponding observation and HRRRx object grid points relative to the number of observation object grid points and false HRRRx object grid points. It has a range from 0 to 1, with 1 representing a perfect forecast. Since the CSI assumes correct-no forecasts are of no consequence

TABLE 3. Forecast metrics displayed on performance diagrams using the 2 × 2 contingency table in Table 2.

Forecast metric	Equation
POD	$\dfrac{h}{h+m}$
SR	$1 - \dfrac{f}{h+f}$
Bias	$\dfrac{h+f}{h+m}$
CSI	$\dfrac{h}{h+m+f}$

(Schaefer 1990), it will not be artificially high because of the larger percentage of grid points that are not associated with observation or forecast objects. An additional metric, the equitable threat score (ETS), will also be used to assess forecast accuracy. The equation for the ETS can be seen in Table 4. Unlike the CSI, the ETS does account for correct-no forecasts when using the sample forecast size (Wilks 2006).

b. MODE composite score

The extensive output generated from MODE to describe objects and object matches can be condensed using the MODE composite score (MCS; Griffin et al. 2017). The MCS is an area-weighted calculation that combines the interest values computed by MODE into a single value while still preserving the unique information provided via object-based verification methods. The MCS includes two components, one for cloud clusters and the other for cloud objects, and is computed as follows:

$$\text{MCS} = \sum_{i=1}^{N_C} \frac{\text{Area}_{\text{Observation\_Cluster}}(i)}{\text{Total\_Area}} \times \text{Interest\_Value}(i)$$
$$+ \sum_{j=1}^{N_O} \frac{\text{Area}_{\text{Observation\_Object}}(j)}{\text{Total\_Area}} \times \text{Interest\_Value}(j). \quad (1)$$

Here, the total area in the denominator is defined as the area of the observed cloud objects plus the area of the forecast cloud objects that are unmatched to the observed cloud objects. The N_C and N_O in the summation

TABLE 4. ETS using the 2 × 2 contingency table in Table 2.

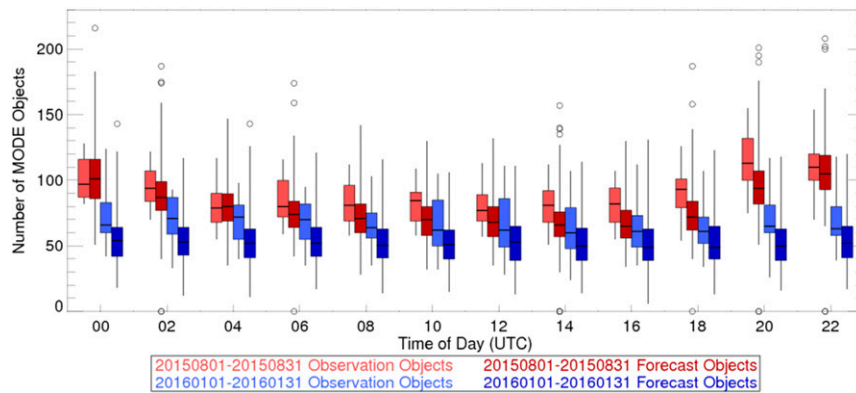| Forecast metric	Equation
ETS	$\dfrac{h - h_{\text{ref}}}{h - h_{\text{ref}} + f + m}$
h_{ref}	$\dfrac{(h+f)(h+m)}{n}$
n	$h + m + f + z$

FIG. 4. Box-and-whisker plot of number of MODE objects for August (red) and January (blue) as based on time of day. Lighter reds and blues represent the number of *GOES-13* observation objects, and darker reds and blues represent the number of HRRRx forecast objects from all HRRRx forecast hours.

represent the number of observation clusters and objects, respectively. Given that the maximum centroid distance in this analysis is chosen to be 200 km, all objects in the observed field will be matched and therefore have a nonzero interest value, with a forecast object that is within 200 km. The MCS has a range from 0 to 1. An MCS equal to 1 indicates that all HRRRx objects perfectly match observation objects. An MCS equal to 0 indicates that no HRRRx and observation objects are within a centroid distance of 200 km. This is unlikely to occur, however.

Since multiple objects in the observation (forecast) field may correspond to a single object in the forecast (observation) field, the MCS calculation first analyzes clusters. Once the MCS accounts for the clusters, only object pairs for which both the individual observation and forecast objects do not already belong to a matched cluster will be used in the MCS calculation. Because each observation and forecast object is only used once, remaining object matches are analyzed from highest to lowest interest value to calculate the highest possible MCS. Object matches with an area ratio of less than 5% are not included in the MCS calculation.

5. Results

a. Object comparison

The first step in assessing the accuracy of the model forecasts is to determine whether they accurately represent the number of upper-level cloud objects that are present in the *GOES-13* observations. Figure 4 shows the number of observed and forecast objects plotted as a function of time of day aggregated over all forecast hours and cycles during August (red) and January (blue). Overall, there are more observation (and forecast)

cloud objects in August, with a more prominent diurnal cycle. This behavior is consistent with the more predominant small-scale convective cloud features found during the summer as compared with the larger synoptic-scale cloud systems more frequently observed during the winter.

Although the different characteristics of the diurnal cycle are accurately captured in both the August and January forecasts, there are typically not enough forecast objects in both months. For example, the median number of forecast objects, identified by the middle line in the box plot, is smaller than the number of observed objects for each time of day and season, with the only exception occurring near 0000 UTC during August (Fig. 4). Even though the median number of forecast objects is smaller, however, the whiskers and outliers indicate that the largest number of forecast objects is typically larger than the largest number of observed objects. Figure 5 shows the number of observation and forecast objects plotted as a function of forecast hour. The typical situation is that too many forecast cloud objects are associated with the 0-h model analyses. This difference is attributed to multiple small convective cores in the model analyses versus a larger cloud cluster in the observations (not shown). In later forecast hours, there are generally fewer forecast objects identified relative to observation objects in the *GOES-13* imagery. The median number of forecast objects reaches a minimum around forecast hour 4 during August before slowly increasing with time. During January, however, there is a steady drift toward fewer forecast objects with increasing forecast hours as the largest forecast cloud objects have the tendency to increase in size with time.

The total size of the MODE objects as a function of time of day can be observed in Fig. 6. With a few
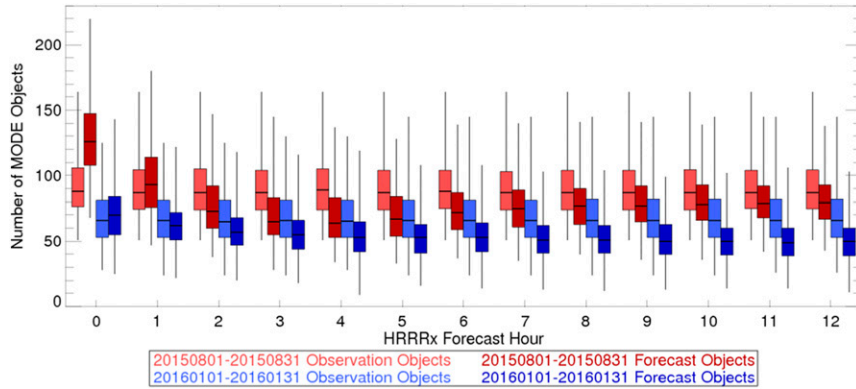
FIG. 5. As in Fig. 4, but as based on HRRRx forecast hour.

exceptions, the January objects encompass more of the HRRRx domain area than do the August objects. Because August has a higher number of objects than January (e.g., Fig. 4), this result indicates that the median cloud-object size for August is smaller than that for January, reflecting the cellular nature of convection in the summer. The number of grid points that the MODE objects encompass differs by season, with a greater number for January than for August. Because the BT threshold to determine MODE objects is calculated using data from the previous 10 days, the area of the MODE objects will not necessarily account for 10% of the total domain at any one time. In addition, the 10th-percentile BT threshold is applied to the raw BT data. Objects are then created from the raw data using a fuzzy-logic algorithm (Developmental Testbed Center 2014). During this process, the percentage of grid points identified as objects is reduced relative to the percentage of raw grid points identified using the BT threshold. This reduction is highly correlated with the number of objects, with correlations of 0.913 and 0.994 for observation forecast objects, respectively. Therefore, the average reduction in observed object grid points is

greater in August (0.61%) than in January (0.46%) because more objects are observed. The reduction in forecast object grid points is even greater in August (0.95%) than in January (0.58%), which could inherently cause a low bias in the number of forecast object grid points relative to observation object grid points.

b. Dichotomous (yes/no) verification

Results from comparing the collocation of the observation and forecast object grid points can be seen in the performance diagram for August and January in Fig. 7. Each of the boxes corresponds to the 99% confidence interval (CI) for a given forecast hour for every available HRRR initialization in a given month. The CI for the given variable X is calculated using

$$CI(X) = \text{mean}(X) \pm 2.58 \frac{\text{Standard\_Deviation}(X)}{\text{Sample\_Size}(X)},$$

where 2.58 is based on the 99% cutoff probability of a two-tailed Gaussian distribution (Wilks 2006). Colored squares in the left and right panels represent the data from August and January, respectively, and the open
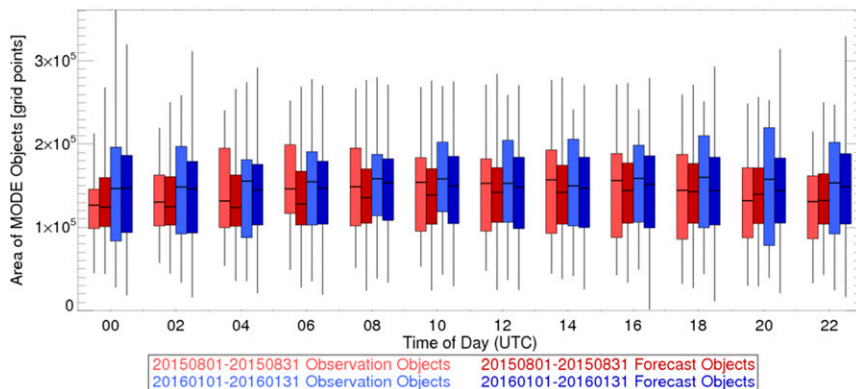

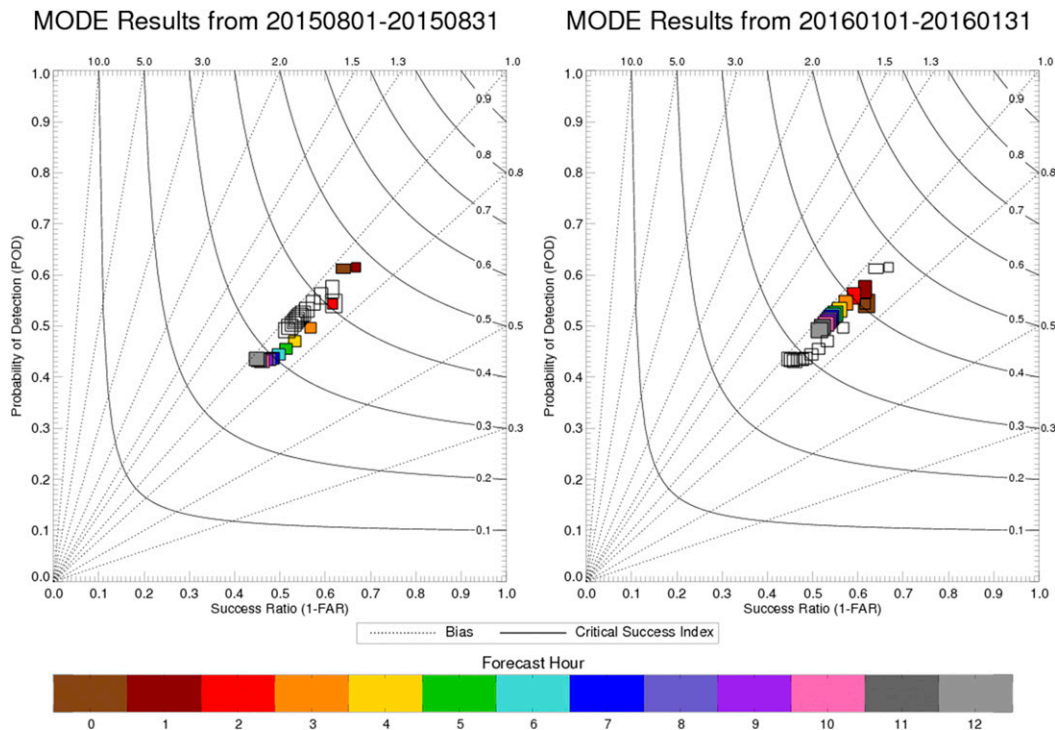FIG. 6. As in Fig. 4, but for the overall size of MODE objects.

FIG. 7. Performance diagram for (left) August and (right) January, as indicated by the colored squares. The open squares represent the data from the other month for comparison. Square size represents the CI.

squares in each panel represent the data from the other month (panel) to ease comparisons between the months. Forecast skill increases to the upper right in each image. For both months, the forecast skill typically decreases with forecast hour. One exception to this behavior is that the 1-h forecast is more accurate than the 0-h analysis. August forecasts are less skillful than the January forecasts at later forecast times, as based on the performance diagrams and the ETS in Table 5. This is consistent with the difficulty in accurately predicting the evolution of the smaller convective cells that are more common during the summer (Wolff et al. 2014). Of interest, though, is that the most skillful forecasts were actually obtained during August. The 0-h analyses and 1-h forecasts have the highest POD and SR when compared with all other forecast hours and thus have the highest CSI. This result suggests that the strong reliance of the HRRRx and RAP data assimilation systems on radar observations is beneficial during the summer when there is convection covering a larger part of the grid. The current radar data assimilation emphasizes higher-reflectivity observations, and therefore the lower-reflectivity winter precipitation (often stratiform in stable conditions) does not have as much impact during assimilation. These improvements then drop off within a couple of hours because of more limited predictability of thunderstorms as compared with the large-scale cloud

and weather systems that prevail during the winter. The POD for August begins to remain steady around the 6-h forecasts because the number of HRRRx object grid points increases with increasing forecast hour (as seen in Fig. 5). These additional HRRRx object grid points are not associated with observation object grid points, however, likely because the displacement between matching objects increases. January has a steady decline in forecast accuracy because the number of HRRRx object grid points remains fairly

TABLE 5. Average CSI and ETS for August and January by forecast hour. The highest value is in boldface type.

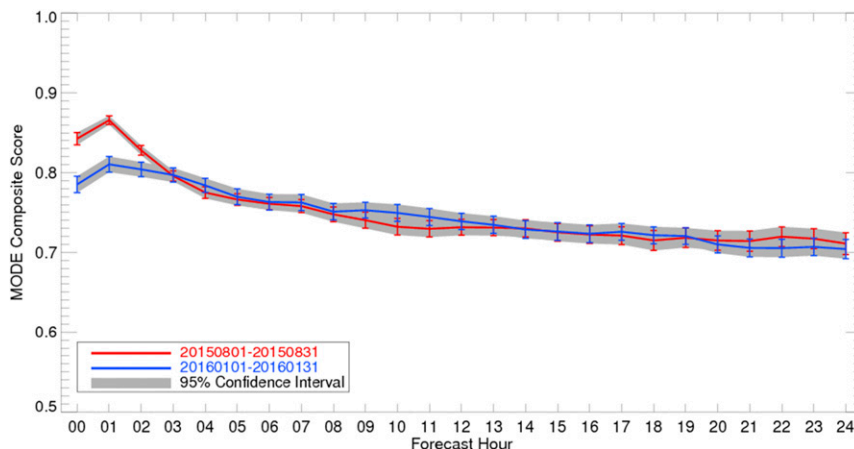| Forecast hour	Aug CSI	Aug ETS	Jan CSI	Jan ETS
0	0.452	0.410	0.406	0.363
1	**0.470**	**0.431**	**0.425**	**0.382**
2	0.405	0.364	0.406	0.362
3	0.359	0.317	0.391	0.346
4	0.332	0.288	0.376	0.330
5	0.317	0.273	0.367	0.321
6	0.307	0.262	0.362	0.315
7	0.299	0.253	0.357	0.311
8	0.295	0.249	0.353	0.306
9	0.292	0.246	0.350	0.303
10	0.288	0.241	0.346	0.299
11	0.286	0.239	0.340	0.293
12	0.285	0.237	0.334	0.288

FIG. 8. Mean MCS values plotted as a function of forecast hour for August (red line) and January (blue line). Gray represents the 95% confidence interval.

constant with forecast lead time, but the displacement between object pairs increases.

According to the performance diagram, both August and January have a bias of less than 1. This result indicates that there are fewer grid points associated with forecast objects than with observation objects. This outcome is expected on the basis of the results shown in Fig. 6 and is due to the fact that the reduction in grid points from raw data to objects is larger for forecasts than for observations. The larger reduction in the forecasts is related to the larger BT range exhibited in the HRRRx, as the logarithmic regression between the reduction in percentage of area and BT range for *GOES-13* and HRRRx BTs has a correlation of 0.821.

c. MODE composite score

In this section, the accuracy of forecast cloud-object characteristics relative to the observation cloud objects is assessed using the MCS. Figure 8 shows the mean MCS plotted as a function of forecast hour for August and January, with the gray shading indicating the 95% confidence interval around the mean MCS. Overall, the 1-h forecast is the most accurate for both months, with discernable skill out to 2 h and a steady decrease thereafter as predictability decreases. This figure indicates that the 1-h forecast objects better represent the observation objects than do other forecast hours. Also, the 0-h analyses and 1-h forecasts are more accurate for August than for January. Forecast hours 2 and beyond have similar accuracy for each season, with the MCS decreasing with increasing forecast hour. This result is consistent with the results from the performance diagram, which also indicated that the 0-h analyses and 1-h forecasts from August are the most accurate when compared with HRRRx forecasts from August and January.

d. Why is August more accurate than January?

As shown in the dichotomous verification and MCS, the 1-h forecast for August is more accurate than the 1-h forecast for January. This result seems counterintuitive because the small-scale features observed in August should be more difficult to forecast (Wolff et al. 2014; Griffin et al. 2017). This behavior is not simply related to the larger January objects being unmatched in the MCS calculation. For the 1-h forecast, the MCS for August and January are 0.871 and 0.830, respectively. After removing the impact of unmatched objects, and therefore confining the total area to the area of the matched observation objects, the MCS for August (0.930) is still higher than that for January (0.903).

To identify why the August forecasts tend to be more accurate than the January ones, attribute-based MCS values are calculated for the four main attributes related to MODE object pairs in this analysis: centroid distance, boundary distance, area ratio, and intersection area ratio. The attribute-based MCS is calculated using Eq. (1), with the attribute-specific interest value replacing the total interest value. Two assumptions are made when calculating the attribute-based MCS. First, the MODE centroid distance weight is the user-defined weight in Table 1 for ease of comparison. Second, the total area is only the area of the matched observation objects since unmatched objects do not affect which forecast hour is more accurate. Attribute interest values are used instead of the actual attribute magnitude (for example, a centroid distance of 170 km) because attribute interest values are calculated using interest functions that are based on the actual attribute magnitude. For example, the centroid distance interest function is

Centroid Distance Interest Value

$$= \begin{cases} 1 & \text{if} \quad X \le 20\,\text{km} \\ \dfrac{150 - X}{150 - 20} & \text{if} \quad 20 < X < 150\,\text{km}, \\ 0 & \text{if} \quad X \ge 150\,\text{km} \end{cases}$$

where X is the centroid distance between two matched objects. Therefore, a centroid distance of 170 km has the same attribute interest value as a centroid distance of 199 km even though it is smaller.

The comparison between the August and January attribute-based MCSs can be observed in Table 6. For all four attributes, the August attribute-based MCS is higher than in January, with the greatest differences observed for the centroid distance and area ratio attributes. The interest values for these attributes, as based on observation object size, are shown in Fig. 9. For each season, attribute interest values from matched object pairs are divided into 30 bins on the basis of the observation object size. Each bin has approximately the same number of matched object pairs, 1175 for August and 812 for January. The interest values for each bin are then averaged. For almost all observation object bins, the centroid distance interest values are greater in August than in January. This indicates that the displacement between centers of mass for matched object pairs is smaller in August than in January, potentially as a result of the stronger flow associated with dynamic systems yielding larger displacement errors in January than in August. This smaller displacement is one reason for the larger POD and CSI in August (Fig. 7), because more forecast object grid points overlap with more observation object grid points. For the area ratio attribute, the interest values for August and January are similar for more than one-half of all object matches. For larger objects, however, the area ratio interest values are greater for August; for smaller objects, the area ratio interest values are greater for January. Therefore, the forecast objects are not accurately depicting the size of the larger January observation cloud objects or the smaller August observation cloud objects relative to the August objects. The larger observation objects have a greater impact on the MCS because the MCS is an area-weighted calculation, leading to a more accurate forecast for August than for January.

e. Why is the 1-h forecast the most accurate?

Both the dichotomous verification and MCS indicate that the 1-h HRRRx forecast is the most accurate forecast for both August and January. Again, this is also not purely the result of more matched forecast and

TABLE 6. Attribute-based MCS for four object-pair attributes for 1-h forecasts from August and January. It is assumed that the MODE centroid distance weight is the user-defined weight in Table 1.

Object-pair attribute	Aug	Jan	Max value
Centroid distance	0.165	0.145	0.25
Boundary distance	0.178	0.176	0.1875
Area ratio	0.206	0.198	0.25
Intersection area ratio	0.176	0.172	0.1875

observation objects. Removing unmatched objects from the MCS calculation increases the MCS for the 0-h analyses from 0.898 to 0.917 and the MCS for the 1-h forecasts from 0.902 to 0.930. To identify the reasons that the 1-h forecast is more accurate than the 0-h analysis, the MODE interest values for the distance and area attributes from the 1-h forecasts were compared with those from the 0-h analyses. The comparison between the 0-h analysis and 1-h forecast attribute-based MCSs can be observed in Table 7. The centroid distance attribute and area ratio attribute have higher attribute-based MCSs for the 1-h forecast, whereas the boundary distance and intersection area ratio actually have higher attribute-based MCSs for the 0-h analysis.

To identify the cause of the differences between attribute-based MCSs, corresponding observation object sizes and attribute interest values are compared. Because the observation object sizes are the same, the attribute interest values for matched object pairs are arranged on the basis of the size of the matching observation object. Because the number of matches for a given observation cloud-object size can differ as a result of an observation object only having a matched forecast object in either the 0-h analysis or 1-h forecast, however, the percent of occurrences is calculated. The percent of occurrences is the ratio of the number of instances in which observation objects of a particular size and corresponding attribute interest value are observed relative to the total number of instances in which an observation object size is observed in matched pairs. The difference in percent of occurrences is then defined as

Difference in Percent of Occurrences

$$= \frac{\text{Number(score, obs\_size)}_{0\text{-h\_analyses}}}{\text{Total(obs\_size)}_{0\text{-h\_analyses}}}$$
$$- \frac{\text{Number(score, obs\_size)}_{1\text{-h\_forecasts}}}{\text{Total(obs\_size)}_{1\text{-h\_forecasts}}}, \qquad (2)$$

where ''score'' is the attribute-specific interest value and ''obs_size'' is the size of the observation object in the matched object pair.
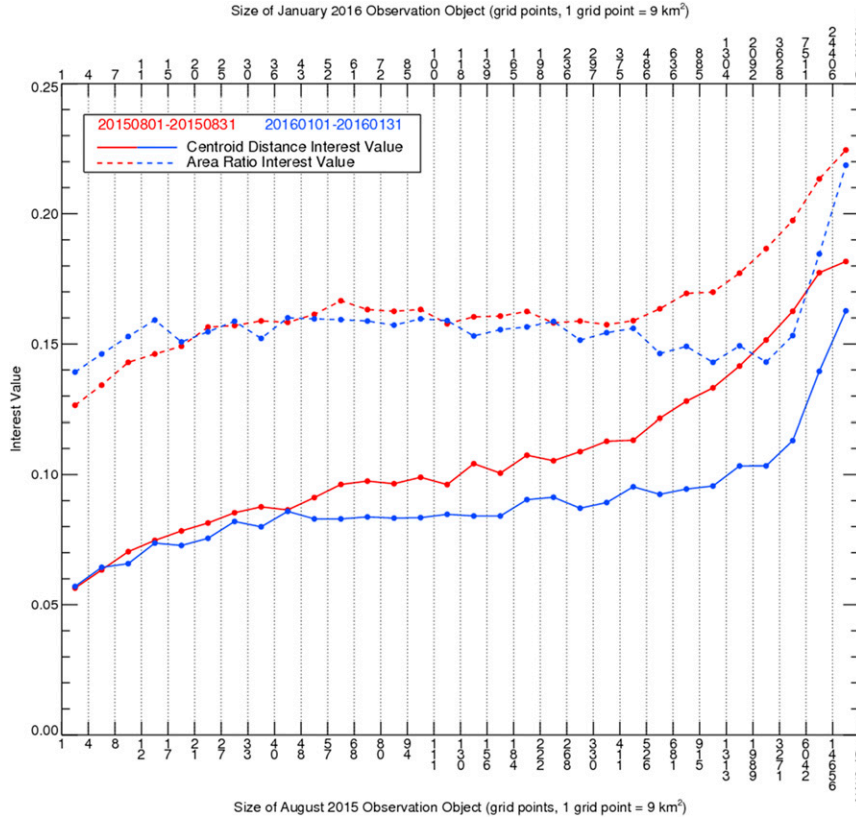
FIG. 9. The centroid distance (solid) and area ratio (dashed) attribute interest values, as based on observation object size, for August (red) and January (blue). The matched observation objects are divided into 30 bins, on the basis of observation object size, so that each bin has approximately the same number of objects. The interest values for each bin are averaged.

The difference in percent occurrences for the centroid distance attribute is shown in Fig. 10. Blue or red colors indicate that the centroid distance interest value and observation object size combination occurs more frequently in the 1-h-forecasts or in the 0-h-analyses object pairs, respectively. For small objects, a higher percentage of observation objects have lower interest values in the 1-h forecasts than in the 0-h analyses. This result is not unexpected because displacement errors tend to increase with increasing forecast hours (Griffin et al. 2017) and lower interest values are associated with greater displacement errors between the objects' centers of mass. Higher interest values are also observed more frequently in the 1-h forecast for the largest cloud objects, however, which indicates that these objects have centers of mass that are closer to their matched forecast objects in the 1-h forecast. Because this improvement in the centroid distance interest value occurs for the largest objects, it has a greater impact on the MCS because the MCS is an area-weighted calculation. The smaller displacement for large objects at the 1-h forecasts could correspond to a higher percentage of

overlapping grid points for these objects and, ultimately, higher SR relative to the 0-h analyses. This is potentially the effect of convective-scale growth toward the end of the "spinup" period.

Differences in the percent of occurrences in the boundary distance interest values are shown in Fig. 11. Overall, the highest interest values are associated with the 0-h analyses, which is a result that indicates that the model initialization is more accurate than the 1-h forecasts for this particular attribute. As was seen in the centroid distance scores in Fig. 10, lower interest values are more common for the small cloud objects in the 1-h forecasts. This result is due to greater displacement

TABLE 7. As in Table 6, but for the 0-h analysis and 1-h forecast from August.

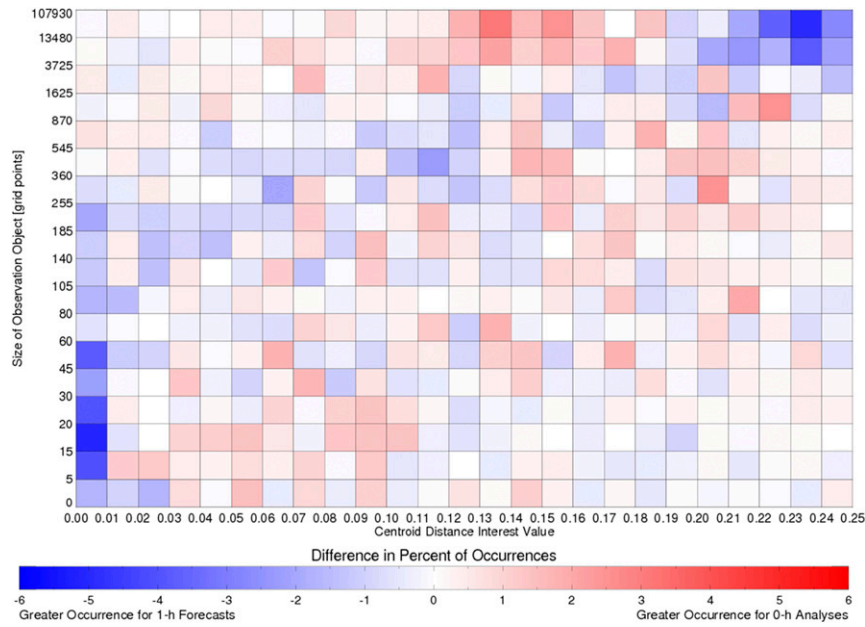| Object pair attribute	0-h analysis	1-h forecast	Max value
Centroid distance	0.154	0.165	0.25
Boundary distance	0.181	0.180	0.1875
Area ratio	0.201	0.206	0.25
Intersection area ratio	0.178	0.176	0.1875

FIG. 10. Percent difference in the occurrence of MODE centroid interest scores between the 0-h analysis and 1-h forecasts during August plotted as a function of MODE interest score along the *x* axis and observation object size along the *y* axis. Blue colors indicate that the given centroid distance interest score and observation object size combination occurs more frequently in the 1-h forecast than in the 0-h analysis, and red colors indicate that they occur less frequently.

between the small cloud objects. Unlike the centroid distance interest values, however, there are only very small differences for the larger cloud objects. Because the boundary distance is the minimum distance between

the edges of the observed and forecast cloud objects, larger objects need a larger displacement between cloud objects for no overlapping to occur. The difference in percent of occurrences for the intersection area ratio
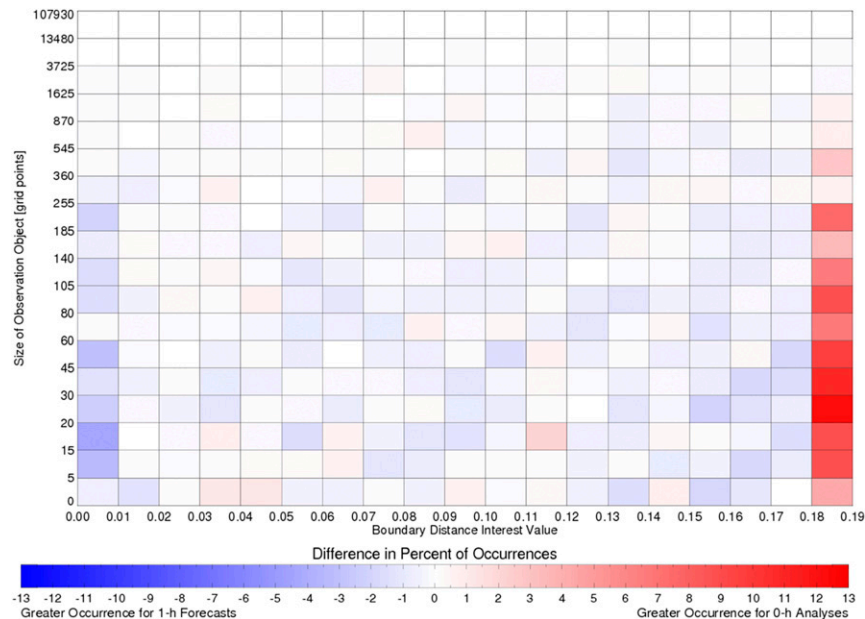

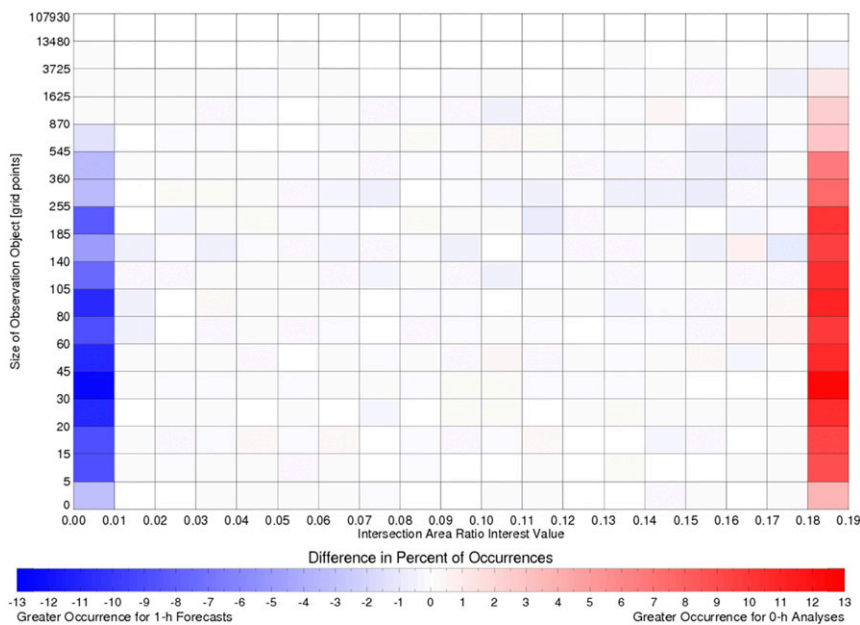FIG. 11. As in Fig. 10, but for the MODE boundary distance attribute.

FIG. 12. As in Fig. 10, but for the MODE intersection area ratio attribute.

interest value (Fig. 12) is very similar to the boundary distance interest values for this same reason. As the displacement between objects increases, the size of the overlapping area, and thus the intersection area ratio interest value, decrease. Matched pairs with larger observed objects are still overlapping, however, and therefore no changes are evident between the 0-h analysis and 1-h forecasts for the larger objects.

The difference in the percent of occurrence for the area ratio attribute is shown in Fig. 13. Overall, the differences between the 0-h analysis and 1-h forecasts are relatively small for this attribute. An important exception, however, is the largest observation objects, for which higher area ratio interest values are more likely to be associated with the 1-h forecast than with the 0-h analysis. This increase in the area ratio interest value for
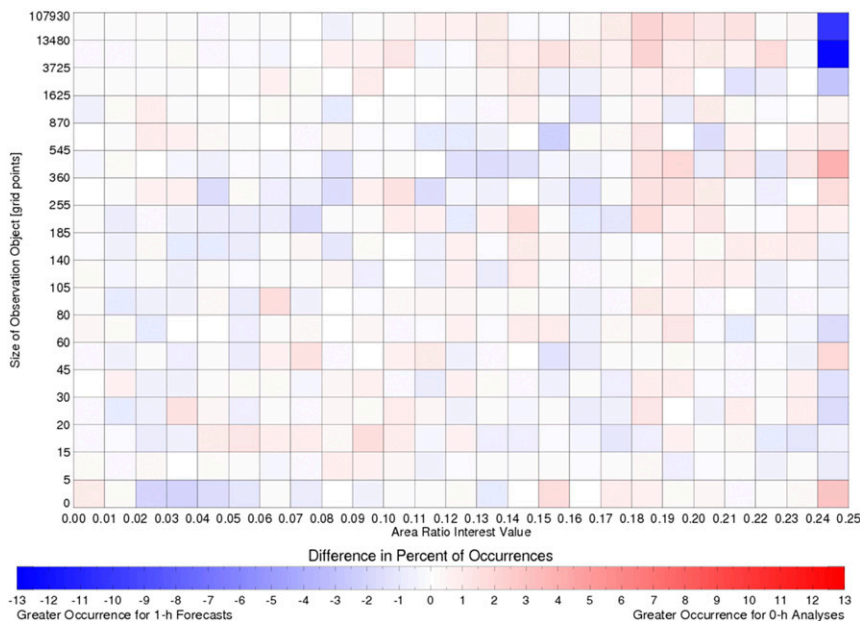

FIG. 13. As in Fig. 10, but for the MODE area ratio attribute.

the 1-h forecasts indicates that the size of the cloud objects in the 1-h forecast better represents the observed objects than does that in the 0-h analysis. Because the observed object sizes remain unchanged, it is assumed that this better representation of large observed objects is a result of the model spinup and forecast objects increasing in size. This would explain how the intersection area ratio remains unchanged between the 0-h analysis and 1-h forecasts for larger cloud objects. In addition, the center-of-mass distance between the objects could decrease even though displacement occurs because forecast objects are larger. Therefore, the 1-h forecasts are more accurate than the 0-h analyses because the areas of the largest cloud objects are more accurate and thus the distance between these object pairs is smaller.

6. Conclusions

In this study, the accuracy of HRRRx forecast cloud objects in the upper troposphere was assessed during warm- and cool-season months using MODE, a sophisticated object-based verification tool. This was accomplished by comparing MODE-identified cloud objects in simulated and observed *GOES-13* 10.7-μm BT imagery for two 1-month periods, 1–31 August 2015 and 1–31 January 2016. To account for biases in the HRRRx simulated BTs and differences in cloud characteristics between the warm and cool seasons, objects were defined using the 10th percentile of the BT distribution accumulated using data over a 10-day period. Overall, more objects were identified in August. The total area encompassed by the August objects was also smaller than that for the January objects, which is consistent with the more predominant small-scale convective cloud features that are found during the summer as compared with the larger synoptic-scale cloud systems that are more frequently observed during the winter.

The HRRRx forecast accuracy was assessed using two methods. The first method is dichotomous verification, in which the locations of the observation and forecast cloud objects are compared. On the basis of the CSI, the 1-h forecasts for both August and January are the most accurate for their respective months. Therefore, these forecasts have the highest proportion of correctly identified observation object grid points. Again on the basis of the CSI, the 1-h forecasts from August are the most accurate overall. The second method for assessing the HRRRx forecast accuracy is the MCS, which uses the MODE interest value from matched pairs of observation and forecast objects. Again, the 1-h forecasts for both August and January are the most accurate for their respective months, with 1-h forecasts from August being the most accurate overall. Thus, the 1-h forecast objects

better match the observation objects in location and size when compared with other forecast hours, with the 1-h forecast objects from August having the best object portrayal overall.

By using MODE to calculate interest scores for individual attributes, the reasons for why the 1-h forecasts from August are more accurate than those from January were identified. For matched observation and forecast objects, the displacement between the objects' centers of mass is smaller in August than in January. The smaller displacement results in more grid points corresponding to both observation and forecast objects, increasing the POD and ultimately the CSI. The forecast objects from the August 1-h forecasts also better represent the size of larger observation cloud objects. Combined with the smaller displacements, the improved object depictions result in a more accurate forecast by increasing the MODE interest values and therefore the MCS.

Individual MODE attribute interest scores were also used to identify the reasons for why the 1-h forecasts are more accurate than the 0-h analyses. As expected, displacement errors between observation and forecast objects increase between the 0-h analyses and 1-h forecasts, except for larger observation objects for which the centroid distance errors are smaller for the 1-h forecasts. The sizes of the 1-h forecast objects better represent the large observed *GOES-13* objects, however, which is consistent with model convective-scale spinup. This improvement for the largest cloud objects has a more positive impact on the MCS and forecast accuracy than the degradations that occurred for the smaller cloud objects.

This study highlights many interesting results related to seasonal characteristics of the HRRRx model. The overall decrease in HRRRx accuracy is potentially of interest to forecasters. As of the time of writing, forecasters can access real-time comparisons of simulated and observed GOES BTs and statistical analyses from both the HRRR and HRRRx models online (http://cimss.ssec.wisc.edu/hrrrval/). Because the simulated BTs for a given HRRRx forecast cycle typically are not available in real time until about 100 min after the initialization time, the verification results between the 0-h analysis and 1-h forecasts presented in this study are more relevant to researchers who are working on data assimilation and parameterization schemes than to forecasters. It is important to note, however, that the data latency is substantially lower for the operational HRRR model ($<$1 h).

Future work will use object-based verification to understand better why the seasonal characteristics of the HRRRx model exist. For example, differences in the number of observation and HRRRx cloud objects

observed during model spinup could also be due to the microphysics scheme producing too many or too few clouds or clouds that have an incorrect cloud optical depth. In addition, the accuracy of simulated HRRR BTs will be assessed for individual weather regimes to identify whether the HRRRx model does better in certain situations and worse in others. The overall smaller displacement errors between object pairs in August may be related to more-intense vertical motions identifying the correct cloud locations, whereas the smaller vertical motions that are more associated with frontal systems in January take longer to appropriately spin up frontal-system cloud features. This could also be related to the radar-reflectivity data-assimilation approach used by the HRRRx model that focuses on higher reflectivity features that are more common in the warm season when model response to convective forcing (latent heating) is more significant. It is also possible that all-sky satellite infrared radiance assimilation could improve the analysis and forecast accuracy through a more accurate depiction of the spatial distribution of clouds, especially in regions that lack higher radar reflectivities.

REFERENCES

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

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

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, doi:10.1002/2015MS000497.

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, doi: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, doi:10.1175/MWR-D-13-00143.1.

Davis, C. A., B. G. Brown, and R. G. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, doi: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, doi: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, doi:10.1175/2009WAF2222241.1.

Developmental Testbed Center, 2014: Model Evaluation Tools version 5.0 (METv5.0): User's guide 5.0. Developmental Testbed Center Rep., 241 pp. [Available online at http://www.dtcenter.org//met/users/docs/users_guide/MET_Users_Guide_v5.0.pdf.]

Feltz, W. F., K. M. Bedka, J. A. Otkin, T. Greenwald, and S. A. Ackerman, 2009: Understanding satellite-observed mountain wave signatures using high-resolution numerical model data. *Wea. Forecasting*, **24**, 76–86, doi:10.1175/2008WAF2222127.1.

Giannakaki, P., and O. Martius, 2016: An object-based forecast verification tool for synoptic-scale Rossby waveguides. *Wea. Forecasting*, **31**, 937–946, doi:10.1175/WAF-D-15-0147.1.

Grasso, L. D., and T. Greenwald, 2004: Analysis of 10.7-μm brightness temperatures of a simulated thunderstorm with two-moment microphysics. *Mon. Wea. Rev.*, **132**, 815–825, doi:10.1175/1520-0493(2004)132<0815:AOMBTO>2.0.CO;2.

——, M. Sengupta, J. F. Dostalek, R. Brummer, and M. DeMaria, 2008: Synthetic satellite imagery for current and future environmental satellites. *Int. J. Remote Sens.*, **29**, 4373–4384, doi:10.1080/01431160801891820.

——, D. T. Lindsey, K.-S. S. Lim, A. J. Clark, D. Bikos, and S. R. Dembek, 2014: Evaluation of and suggested improvements to the WSM6 microphysics in WRF-ARW using synthetic and observed *GOES-13* imagery. *Mon. Wea. Rev.*, **142**, 3635–3650, doi:10.1175/MWR-D-14-00005.1.

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

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. [Available online at https://www.star.nesdis.noaa.gov/sod/sst/micros/pdf/CRTM_v1_NOAAtechReport-1.pdf.]

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, doi:10.1029/2008JD009944.

Jankov, I., and Coauthors, 2011: An evaluation of five ARW-WRF microphysics schemes using synthetic GOES imagery for an atmospheric river event affecting the California coast. *J. Hydrometeor.*, **12**, 618–633, doi:10.1175/2010JHM1282.1.

Jin, H., M. S. Peng, Y. Jin, and J. D. Doyle, 2014: An evaluation of the impact of horizontal resolution on tropical cyclone predictions using COAMPS-TC. *Wea. Forecasting*, **29**, 252–270, doi:10.1175/WAF-D-13-00054.1.

Mittermaier, M. P., and R. Bullock, 2013: Using MODE to explore the spatial and temporal characteristics of cloud cover forecasts from high-resolution NWP models. *Meteor. Appl.*, **20**, 187–196, doi:10.1002/met.1393.

Morcrette, J. J., 1991: Evaluation of model-generated cloudiness: Satellite-observed and model-generated diurnal variability of brightness temperature. *Mon. Wea. Rev.*, **119**, 1205–1224, doi:10.1175/1520-0493(1991)119<1205:EOMGCS>2.0.CO;2.

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, doi: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, doi:10.2151/jmsj.87.895.

Otkin, J. A., and T. J. Greenwald, 2008: Comparison of WRF Model-simulated and MODIS-derived cloud data. *Mon. Wea. Rev.*, **136**, 1957–1970, doi:10.1175/2007MWR2293.1.

——, D. J. Posselt, E. R. Olson, H.-L. Huang, J. E. Davies, J. Li, and C. S. Velden, 2007: Mesoscale numerical weather prediction models used in support of infrared hyperspectral measurements simulation and product algorithm development. *J. Atmos. Oceanic Technol.*, **24**, 585–601, doi:10.1175/JTECH1994.1.

——, 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, doi:10.1175/2009JAMC2142.1.

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

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

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, doi:10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.

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, doi:10.1175/MWR-D-15-0198.1.

Söhne, N., J.-P. Chaboureau, S. Argence, D. Lambert, and E. Richard, 2006: Objective evaluation of mesoscale simulations of the Algiers 2001 flash flood by the model-to-satellite approach. *Adv. Geosci.*, **7**, 247–250, doi:10.5194/adgeo-7-247-2006.

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, doi:10.1175/JAS-D-13-0305.1.

——, M. Tewari, K. Ikeda, S. Tessendorf, C. Weeks, J. A. 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, doi: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, doi:10.1175/JAS-D-12-0104.1.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.

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, doi:10.1175/WAF-D-13-00135.1.