Assessing the Impact of Stochastic Perturbations in Cloud Microphysics using *GOES-16* Infrared Brightness Temperatures

SARAH M. GRIFFIN AND JASON A. OTKIN

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

GREGORY THOMPSON, MARIA FREDIANI, AND JUDITH BERNER

National Center for Atmospheric Research, Boulder, Colorado

FANYOU KONG

Center for Analysis and Prediction of Storms, Oklahoma University, Norman, Oklahoma

(Manuscript received 9 March 2020, in final form 6 May 2020)

ABSTRACT

In this study, infrared brightness temperatures (BTs) are used to examine how applying stochastic perturbed parameter (SPP) methodology to the widely used Thompson–Eidhammer cloud microphysics scheme impacts the cloud field in high-resolution forecasts. Modifications are made to add stochastic perturbations to three parameters controlling cloud generation and dissipation processes. Two five-member ensembles are generated, one using the microphysics parameter perturbations (SPP-MP) and another where white noise perturbations were added to potential temperature fields at initialization time (Control). The impact of the SPP method was assessed using simulated and observed *GOES-16* BTs. This analysis uses pixel-based and object-based methods to assess the impact on the cloud field. Pixel-based methods revealed that the SPP-MP BTs are slightly more accurate than the Control BTs. However, too few pixels with a BT lower than 270 K result in a positive bias compared to the observations. A negative bias compared to the observations is observed when only analyzing lower BTs. The spread of the ensemble BTs was analyzed using the continuous ranked probability score differences, with the SPP-MP ensemble BTs having less (more) spread during May (January) compared to the Control. Object-based analysis using the Method for Object-Based Diagnostic Evaluation revealed the upper-level cloud objects are smaller in the SPP-MP ensemble than the Control but a lower bias exists in the SPP-MP BTs compared to the Control BTs when overlapping matching objects. However, there is no clear distinction between the SPP-MP and Control ensemble members during the evolution of objects, Overall, the SPP-MP perturbations result in lower BTs compared to the Control ensemble and more cloudy pixels.

1. Introduction

An accurate depiction of the spatial and temporal characteristics of clouds is necessary for skillful weather and climate forecasts. Predictions of clouds are useful for forecasting when and where severe weather may occur (Mecikalski and Bedka 2006; Sieglaff et al. 2011; Purdom 1993; Cintineo et al. 2013). Changes in cloud cover affect daily temperatures, as they are negatively correlated with the diurnal temperature range (Karl et al. 1993; Dai et al. 1999). Accurate cloud predictions are also necessary for climate modeling, as clouds have a net cooling effect in the earth radiation budget (Ramanathan et al. 1989; Harrison et al. 1990).

Clouds are very complex and often difficult to represent accurately in numerical weather prediction (NWP) and climate models. Nonlinear interactions between different cloud hydrometeor species and the local thermodynamic environment occur at scales that are much smaller than those represented by most models. In recent years, cloud microphysics schemes have become more sophisticated due to improved understanding of cloud processes and increased computational resources. One such microphysics scheme is the bulk microphysics parameterization scheme developed by Thompson et al. (2004, 2008).

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

DOI: 10.1175/MWR-D-20-0078.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

Thompson and Eidhammer (2014) enhanced this microphysics scheme to an "aerosol-aware" scheme that includes the explicit treatment of cloud droplet activation and ice nucleation via aerosols. However, large uncertainties still exist in microphysics schemes pertaining to parameter and process uncertainty (White et al. 2017).

While it is of high socioeconomic value to provide reliable forecasts of extreme and hazardous weather events at the storm-scale, forecasting systems generally have problems, with the actual observed event often lying often outside the uncertainty predicted by the ensemble spread (e.g., Berner et al. 2009, 2011). One reason for the lack in ensemble spread lies in the need to truncate the underlying equations at a particular grid scale without representing subgrid-scale variability (Palmer 2001; Berner et al. 2017). Several methods that have been developed to represent model error due to truncation errors are now used routinely in operational weather forecasting (e.g., Sanchez et al. 2016; Leutbecher et al. 2017). These can be categorized as multimodel ensembles, multiparameter ensembles, as well as stochastic parameterization schemes (Berner et al. 2017).

In this study, we will combine the multiparameter and stochastic parameterization approach and perturb key parameters in the Thompson–Eidhammer microphysics scheme with a spatially and temporally varying perturbation pattern (Berner et al. 2015). There is a long history of static parameter perturbations in weather and climate modeling (e.g., Berner et al. 2015; Christensen et al. 2015; Palmer 2019), with stochastic parameter perturbations have been previously evaluated in other ensemble systems (Bowler et al. 2008; Ollinaho et al. 2017). Since only the microphysical parameters are perturbed in this study, the resulting scheme is called Stochastic Parameter Perturbation (SPP) for Micro-Physics, or short: SPP-MP.

Stochastic parameterizations have been used in multiple forecast studies. Jankov et al. (2019) used SPP to perturb parameters in the planetary boundary layer scheme in convection-resolving simulations using the High Resolution Rapid Refresh model and found an improvement in low-level wind forecasts. Watson et al. (2017) and Subramanian and Palmer (2017) found that stochastic parameterizations improved the accuracy of tropical precipitation and convection, but Subramanian and Palmer (2017) also indicated the forecast for zonal winds was degraded when perturbing the boundary layer temperature. Connelly (2018) used a stochastically perturbed physical tendency (SPPT) scheme to analyze the predictability of finescale snowbands in a 40-member Weather Research and Forecasting (WRF) Model. However, no studies have directly assessed the impacts of the SPP method on cloud forecasts.

Detailed information about the horizontal distribution of clouds can be obtained from satellite (IR) brightness temperatures (BT). Prior studies have used satellites BTs to evaluate the accuracy of the cloud field in high-resolution numerical weather prediction model forecasts (e.g., Bikos et al. 2012; Cintineo et al. 2014; Feltz et al. 2009; Grasso and Greenwald 2004; Grasso et al. 2008, 2014; Griffin et al. 2017a,b; Jankov et al. 2011; Jin et al. 2014; Morcrette 1991; Otkin and Greenwald 2008; Otkin et al. 2009; Thompson et al. 2016; Van Weverberg et al. 2013). Traditional pixel-based metrics, such as mean absolute error, can be used to assess the differences between the observed and simulated cloud fields. While these methods are easier to implement, object-based statistics such as the method for object-based diagnostic evaluation (MODE; Davis et al. 2006a,b) can provide a more detailed assessment of the forecast accuracy.

The purpose of the paper is to assess the impact of SPP-MP applied to the Thompson and Eidhammer (2014) microphysics scheme (hereafter TE14) on cloud cover using high-resolution WRF Model forecasts, as it is essential to develop perturbation methods that are able to provide sufficient ensemble spread. This analysis utilizes data from May 2017 and January 2018 to investigate potential differences in cloud characteristics during the warm and cool seasons and to take advantage of the new (*Geostationary Operational Environmental Satellite*) *GOES-16* Advanced Baseline Imager (ABI), this analysis utilizes data from May 2017 and January 2018. Given the fine spatial resolution (3-km) of the WRF Model used during this study, differences between observed and simulated cloud fields will be assessed using both pixel-based and object-based metrics. Pixel-based metrics will be used to evaluate overall model accuracy as well as the spread in the ensemble BTs, while object-based statistics are used to assess the accuracy of cloud features without penalizing for displacement errors, as well as track the evolution of clouds through time.

The results from paper will be broken down into two sections. The first section analyzes the impact of the SPP-MP over the entire domain via a pixel-based methodology. The second section looks at the impact of the SPP-MP for individual cloud features using an object-tracking method. Sections for data and methodology will be presented before the results, and conclusions will follow.

2. Data

a. WRF Model

For the experiments performed during this study, we used the Weather Research and Forecasting (WRF) Model with the Advanced Research WRF dynamic

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

solver (Skamarock et al. 2008), version 3.9.1.1. The WRF Model has been used rather extensively in NOAA's annual Hazardous Weather Testbed Spring Forecast Experiment (HWT-SFE; e.g., Clark et al. 2012, 2018). In nearly the identical manner to prototype real-time WRF forecasts that support the HWT-SFE, we configured the model with 50 vertical levels and 3-km grid spacing covering nearly all of the contiguous United States (CONUS) along with the same physical parameterizations used in the operational High Resolution Rapid Refresh (HRRR) model. This includes TE14 microphysics scheme, the Rapid Radiative Transfer Model global version (RRTMG; Iacono et al. 2000), the Rapid Update Cycle land surface model (Smirnova et al. 2016), and the MYNN planetary boundary layer scheme (Nakanishi and Niino 2004). The experiment design was based on the HRRR model for the intended goal of being possible to transition to operations. Running at 3-km grid spacing, HRRR omits the use of a convective parameterization, since this resolution is relatively capable of explicitly representing convective-scale storms.

To study the sensitivity of brightness temperatures to uncertainties in the microphysics parameters, the SPP-MP scheme was developed with the aim of perturbing key parameters in the TE14 microphysics scheme. Previous work has found that the model can only hold on to perturbations that have spatial and temporal correlations (see Berner et al. 2017 and references therein), which are designed to represent organized nonlocal processes. Modifications were made to the code so that the SPP-MP method could be used to create a stochastically sampled, two-dimensional, time-varying field of correlated parameter values following locally a Gaussian distribution with a prescribed mean and standard deviation. This field is then used to add spatially and temporally varying stochastic perturbations to three processes listed below that control cloud generation and dissipation processes in the TE14 microphysics scheme. These parameters were chosen because they are known to be highly variable based on observational evidence while traditionally being set to a constant in nearly all existing bulk microphysical parameterizations.

1) The graupel spectra y-intercept parameter in TE14 is diagnostically determined by the graupel mass mixing ratio and supercooled liquid water amount at each grid point during the forecast. Most one-moment bulk microphysics schemes (e.g., Rutledge and Hobbs 1984; Hong et al. 2006) use a constant in space and time single value for y-intercept even though observations have found it to vary by as much as 2–3 orders of magnitude (e.g., Knight et al. 1982; Field et al. 2019). To account for this type of variability, the stochastic perturbation field is scaled such that the y-intercept value varies within plus or minus 1.5 orders of magnitude, while also being bounded from 5×10^3 to $5 \times 10^7\,\mathrm{m}^{-4}$. A low value of this parameter could result in hail-like hydrometeors that fall very rapidly while a very high value would be more typical of rimed snow. The implications to subsequent microphysical processes such as collection and spread of convective anvil clouds are rather dramatic as discussed in Gilmore et al. (2004).
2) The cloud water category in TE14 follows a generalized gamma distribution (Verlinde et al. 1990) with a shape parameter that gets diagnosed from the predicted droplet number concentration (following Martin et al. 1994). As such, this gives a shape parameter that can vary in space and time; however, the parameter remains highly uncertain in observations (Miles et al. 2000). To capture this variability, the SPP-MP field is scaled between ± 3 before being added to the previously diagnosed value. The gamma distribution shape parameter effectively shifts the mean size of cloud droplets such that it can directly impact the warm-rain formation process (via autoconversion), which can subsequently affect cloud longevity (Albrecht 1989), cloud albedo (Twomey 1974) and precipitation amounts (TE14).
3) In the atmosphere, nearly all cloud droplets and ice crystals form on an aerosol particle. While the TE14 scheme explicitly predicts the potential aerosols that serve as cloud condensation nuclei (CCN) and ice nuclei (IN), there is obviously inherent uncertainty with these predicted variables. Furthermore, there is uncertainty in the prediction of the model vertical velocity forcing, especially from considerations of eddies that occur at scales much smaller than the grid spacing. For this reason, the CCN and IN activation code of TE14 was modified to include the SPP field as an addition to the gridscale vertical velocity when retrieving a lookup table value of fraction of activated aerosols. The perturbations on vertical velocity were bounded between 0 and $0.5\,\mathrm{m\,s}^{-1}$ and all perturbation values were offset by the SPP field minimum since a downward vertical velocity would not result in supersaturation, which is required in the lookup tables for aerosol activation. As such, this experiment means that CCN and IN activation can only be increased from the experiment that does not use SPP-MP. Since the predicted number concentration of cloud droplets directly impacts the calculation of the gamma size distribution shape parameter mentioned in 2 above, this could potentially lead to a larger combined effect than would occur if they were perturbed individually.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

Temporally and spatially correlated perturbations were added to each of the above parameters using the SPP-MP method. Here, we provide a brief overview of the method with the reader referred to Thompson et al. (2020, manuscript submitted to *Mon. Wea. Rev.*) for a more detailed description. The perturbation patterns generated by the SPP-MP are fully determined by three parameters, including temporal decorrelation time, spatial length scale, and the variance in gridpoint space. These three variables used to define the wavenumber dependent variance of a Gaussian white-noise process (Thompson et al. 2020, manuscript submitted to *Mon. Wea. Rev.*). As described in Thompson et al. (2020, manuscript submitted to *Mon. Wea. Rev.*), the stochastic perturbations for a given variable and grid point are drawn from a univariate Gaussian distribution centered on the value of the deterministic parameter. The spatial correlation constraint ensures that the perturbations at adjacent grid points on average will have the same sign. The temporal correlations allow the SPP-MP perturbations to vary with time in a way that guarantees an assigned degree of memory based on the length of the decorrelation time. For the experiments performed during this study, the spatial and temporal decorrelation lengths were identical for each of the three parameters, and were set to 200 km and 2 h, respectively. These values were chosen because they are representative of the scales associated with deep convection and are consistent with the high-spatial resolution used during this study.

b. Ensemble configuration

Two ensembles are run for each forecast initialization time to assess the impact of the SPP perturbations on the cloud field. The first is the SPP-MP ensemble consisting of 5 members where time- and spatially varying SPP perturbations were added to each of the three parameters described above during the forecast. While we made it possible to test independently each of the three aspects, we decided to report only on the simulations with all three SPP aspects enabled together. The application of SPP within the WRF Model experiments is nearly the same as Jankov et al. (2019) application into the HRRR model except for the settings of time and spatial correlation scales and perturbation magnitudes. This SPP-MP ensemble does not have a representation of initial condition or lateral boundary condition uncertainties. To compare the impact of SPP-MP, a Control ensemble was generated by introducing white noise perturbations at the initialization time to four ensemble members with the unperturbed control initialization included as the fifth member. The white noise was entirely uncorrelated in (x, y, z) space with a maximum magnitude of 0.05°C to the potential temperature variable at any model level within 800 m of the land/water surface. The Control ensemble was designed to compare the impact of adding realistic perturbations to select cloud processes against small initial condition perturbations. It is important to note that the white noise perturbations have the potential to trigger convection while the SPP-MP perturbations can only have an impact when clouds are present. So while the white noise perturbations are small compared to realistic initial condition uncertainties, they are able to capture an aspect of forecast uncertainty that cannot be captured by SPP-MP perturbations alone. Furthermore, while it is more operationally relevant to produce initial and lateral boundary condition uncertainty, we wish to reveal how small initial potential temperature perturbations grow into much larger perturbations to model fields similar to the problem of "seeding chaos" noted by Ancell et al. (2018). While we expect that ensemble forecast spread can be caused by SPP-MP, we postulate that some portion of the spread is completely unrelated to SPP-MP and instead results from numerical error growth for which the white noise experiments may provide a baseline amount of spread assured to occur no matter how any perturbations were introduced. Our ensemble design of white noise experiments to analyze sensitivity in convection-permitting models is similar to Flack et al. (2019).

c. Brightness temperatures

We generate simulated BTs from WRF using the Community Radiative Transfer Model (CRTM; Han et al. 2006), version 2.3.0. Three-dimensional profiles of pressure, temperature, and water vapor mixing ratio, as well as cloud liquid, ice, rain, snow, and graupel mixing ratios are passed to the CRTM from the WRF Model. Cloud effective diameters are calculated consistent with the particle size distribution assumptions inherent to the microphysics scheme (Thompson et al. 2016). The file containing the cloud optical properties data for scattering calculations is CloudCoeff_TAMU-11 September 2014 bin (Yang et al. 2013). This file fixes errors in the CRTM's ice optical properties by updated the delta-fit coefficients (Grasso et al. 2018). In addition, two-dimensional variables of latitude, longitude, surface temperature, height and pressure, and land use are passed to the CRTM. Surface emissivity for each IR band is created using the University of Wisconsin High Spectral Resolution Emissivity Algorithm (Borbas and Ruston 2010).

The satellite validation data used for this study is from the *GOES-16* Advanced Baseline Imager (ABI). This sensor has a 2-km pixel spacing at nadir[1] for IR channels,

[1] Nadir is the location on Earth directly below the satellite.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

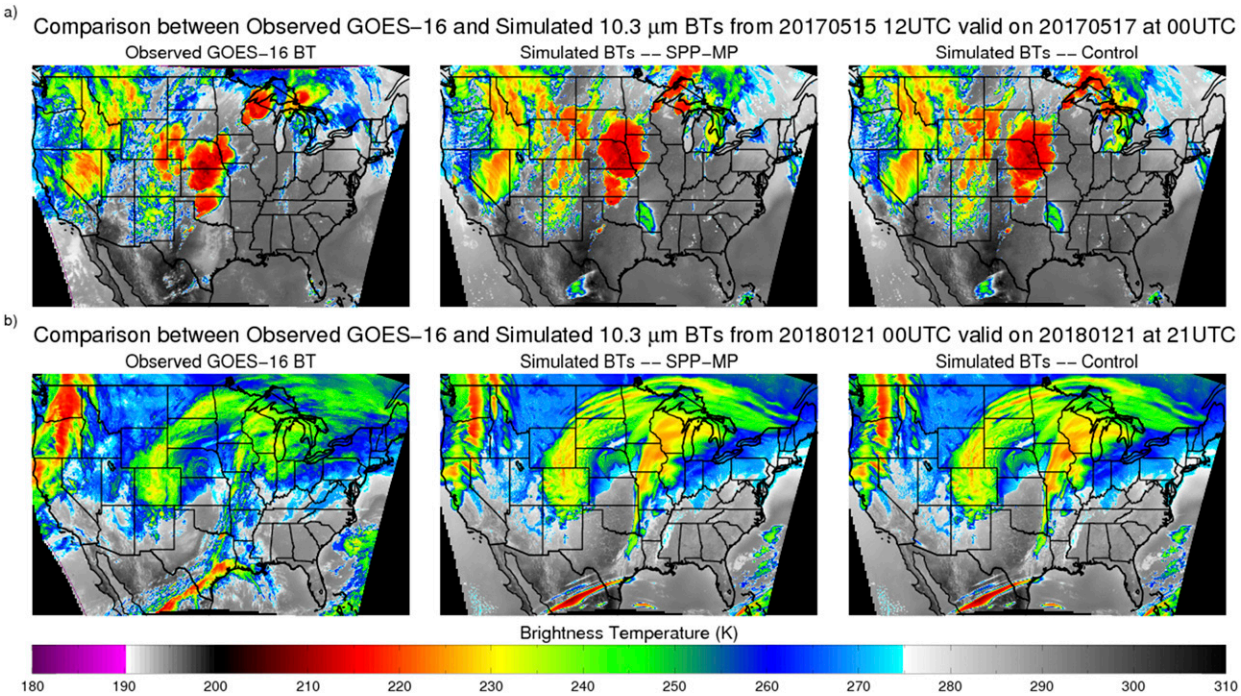

FIG. 1. Comparison of observed and simulated *GOES-16* ABI 10.3 μm brightness temperatures at (a) 0000 UTC 17 May 2017 and (b) 2100 UTC 21 Jan 2018. The simulated images in (a) are from 36-h forecasts initialized at 1200 UTC 15 May 2017 and a 21-h forecast initialized at 0000 UTC 21 Jan 2018 in (b). The simulated *GOES-16* ABI 10.3 μm brightness temperatures are from a randomly selected SPP-MP and Control ensemble member.

which is remapped to the 3-km WRF grid using an area-weighted average of all the observed pixels overlapping a given WRF Model grid box. Simulated ABI BTs are compared to the observed ABI BTs from the *GOES-16* satellite scan that starts just after the top of each hour.

d. Seasonal comparison

Specific days during two months were chosen to assess the impact of the SPP-MP microphysics changes on the simulated *GOES-16* BTs, May 2017 and January 2018. These months during warm and cool seasons were chosen given potential differences in meteorological regimes and associated cloud characteristics. A representative snapshot of the observed and simulated *GOES-16* 10.3 μm BTs can be seen in Fig. 1a for May 2017 and Fig. 1b for January 2018. Cloud features are generally colder and smaller for May 2017 compared to January 2018, though it is important to note that both large and small objects can occur during both months. The overall difference in cloud features during these two months, together with the two ensembles, will allow us to determine if the SPP perturbations have different forecast impacts depending upon if the clouds are primarily convectively or synoptically driven. A total of 10 different 48-h forecasts for each month are analyzed, with forecasts initialized at 1200 UTC on 1, 7, 9, 15, 17, 19, 21, 23, 25, and 27 May 2017 and at 0000 UTC on 7, 9, 11, 13, 19, 21, 23, 25, 27, and 29 January 2018. Forecast hours 0 to 5 are not included in this analysis to reduce the impact of model spinup on the forecast cloud fields, which start from a cloud-free analysis. The choice of starting with forecast hour 6 for the analysis is somewhat arbitrary as cloud spinup may not be complete until after this time.

3. Methodology

This analysis will utilize two types of metrics when assessing the impact of the SPP-MP perturbations on the simulated BTs: pixel-based metrics and object-based analysis. Pixel-based metrics are easy to implement; however, they are susceptible to the well-known "double penalty" problem if features such as clouds are spatially displaced. Object-based methods can potentially account for this displacement, as well as provide other interesting analysis options such as tracking objects through time.

a. Pixel-based metrics

1) DIMENSIONED METRICS

Two pixel-based dimensioned metrics are used to assess the impact of the SPP-MP on simulated BTs.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

These metrics are considered "dimensioned" as they have the same units as the variable of interest (Willmot and Matsuura 2005). Overall model accuracy will be assessed using the mean absolute error (MAE). The MAE is used instead of the root-mean-square error because errors in the ensemble BTs do not follow a normal distribution based on a Shapiro–Wilk test (Willmot and Matsuura 2005; Chai and Draxler 2014). MAE is calculated using the following equation:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{n}|F_i - O_i|, \tag{1}$$

where $F(O)$ represents the simulated (observed) BTs. A MAE of zero is perfect forecast.

Bias in the ensemble simulated BTs is calculated using the following equation:

$$\mathrm{Bias} = \frac{1}{N}\sum_{i=1}^{n}(F_i - O_i). \tag{2}$$

A positive (negative) bias indicates the simulated BTs are too high (low) compared to the observed BTs. For both the MAE and bias, a 95% confidence interval around the difference between the verification metrics indicates whether the forecasts are statistically different from each other. The confidence intervals are calculated using bootstrap sampling with replacement. Each bootstrap sample contains 10 000 data points, resampled 1000 times, and the resulting interval represent the 2.5th and 97.5th percentiles. If the confidence interval surrounding the differences of the metric in question, for example the SPP-MP ensemble mean BT bias minus the Control ensemble mean BT bias, does not encompass zero, a statistically significant difference exists between the ensemble mean BTs (Gilleland 2010). A 95% confidence interval, which is the most common, is used to identify statistical significance (Xu 2006).

2) Continuous ranked probability skill score

To identify how closely the BTs from the five members of each ensemble represent the observed BT, the continuous ranked probability score (CRPS) is utilized. The CRPS compares the cumulative distribution function (CDF) of the simulated ensemble BTs to the observed BT at a given pixel. The CRPS can be decomposed into $\mathrm{CRPS}_{\mathrm{reliability}}$ and $\mathrm{CRPS}_{\mathrm{potential}}$, and equations for these parameters can be found in Hersbach (2000). The continuous ranked probability skill score (CRPSS) is used to compare the CRPS for the SPP-MP and Control ensembles. The CRPSS is computed using the following equation:

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}_{\mathrm{SPP-MP}}}{\mathrm{CRPS}_{\mathrm{Control}}}. \tag{3}$$

The CRPSS ranges from −1 to 1, with a positive CRPSS indicating the SPP-MP ensemble BTs more closely represent the GOES BT than the Control ensemble BTs.

The CRPS can be decomposed into $\mathrm{CRPS}_{\mathrm{reliability}}$ and $\mathrm{CRPS}_{\mathrm{potential}}$. $\mathrm{CRPS}_{\mathrm{reliability}}$ is closely connected to the rank histogram of an ensemble. It tests the ensemble statistical consistency (the observation frequency within members is proportional to the ensemble size), in addition to weighting the bin width (sharper ensembles yield better reliability compared to equally accurate but less sharp ensembles). $\mathrm{CRPS}_{\mathrm{potential}}$ is sensitive to the range of ensemble BTs. The narrower the ensemble system is, the lower the $\mathrm{CRPS}_{\mathrm{potential}}$, assuming the ensemble system encloses the observation (Hersbach 2000). However, it is also sensitive to too many or too large of outliers. A higher $\mathrm{CRPS}_{\mathrm{potential}}$ will be seen for an ensemble system with outliers farther from the observation if the observation is not enclosed in the ensemble spread.

b. Object-based metrics

Upper-level cloud objects are identified using MODE (Davis et al. 2006a,b). MODE identifies and matches objects in two different fields (e.g., observed and simulated data). While the MODE process is fully described in Davis et al. (2006a), a short outline is provided here for context as was done in Griffin et al. (2017a,b):

1) Identify objects by smoothing the simulated and observed BT fields.
2) Calculate various object attributes like distance between object centers and ratio of object sizes for each observed and forecast cloud object.
3) Match forecast and observed cloud objects using a fuzzy logic algorithm.
4) Output attributes for individual objects and matched object pairs for assessment.

The convolution radius used for both the observed and simulated BTs is 5 grid points (15 km) based on Griffin et al. (2017a). This radius allows for the analysis of small-scale storms, since Cai and Dumais (2015) state a range of 2–8 grid points as identifying convective storm objects in ~4-km resolution radar imagery.

The similarity between matching simulated and observed in MODE objects is calculated by computing an interest value (Developmental Testbed Center 2014). Interest values are a weighted combination of the object pair attributes. They range from 0 to 1, with one being a perfect match. The attributes and user-defined weights applied in this study, similar to Griffin et al. (2017a,b), are

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

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

Object pair attribute	User-defined weight (%)	Description
centroid_dist	4 (25.0)	Distance between objects' "center of mass"
boundary_dist	3 (18.75)	Minimum distance between the objects
convex_hull_dist	1 (6.25)	Minimum distance between the polygons surrounding the objects
angle_diff	1 (6.25)	Orientation angle difference
area_ratio	4 (25.0)	Ratio of the simulated and observed objects' areas (or its reciprocal, whichever yields a lower value)
int_area_ratio	3 (18.75)	Ratio of the objects' intersection area to the lesser of the observed or simulated area (whichever yields a lower value)

shown in Table 1. Again, distance and size comparison between the objects is prioritized in this analysis. More emphasis is placed on the displacement between the centroids of the objects rather than their edges, and the ratio of the objects' areas instead of the ratio of the intersection area of objects. The ratio of the intersection area can be artificially high when a larger object fully encloses a smaller object.

Tracking the evolution of attributes associated with matching cloud object pairs is accomplished using an extension of MODE, known as MODE time domain (MODE-TD; Bullock 2011). Overall, this analysis includes 13 objects from May 2017 identified using varied subjectively determined BTs, which can be seen in Table 2. The sample is limited to 13 objects because the initiation of the object must be clearly identified in the GOES observations and therefore cloud patterns that progress throughout the forecast are not considered. Object tracking begins when either the observed or simulated object is identified by MODE-TD at the given threshold. An example of using MODE-TD to track objects can be seen in Fig. 2. In this figure, the simulated object is first tracked two hours before the observed object appears, and tracking stops when the forecast cycle ends. Object tracking stops when a given forecast cycle ends (seven cases), when either the simulated or observed objects merge with another object (three cases), or when the observed object dissipates (three cases). The thresholds in Table 2 are chosen to maximize the time objects remain discrete and therefore trackable. Objects from January 2018 are not included in this analysis because the initiation time of the larger-scale features could not be identified.

4. Results

a. Pixel-based metrics

1) DIMENSIONED METRICS

The MAE for the SPP-MP and Control ensemble mean 10.3 μm BTs for May and January can be seen in Fig. 3. In Fig. 3, the top of each plot presents the MAE while the lower plot indicates the confidence interval envelope. For May, The SPP-MP perturbations have little effect on the accuracy of the ensemble mean BTs. In January, the SPP-MP ensemble mean BTs have lower overall error. Both the SPP-MP and Control ensemble mean BTs are more accurate for January compared to May, due to the smaller-scale features observed in May being more difficult to forecast (Wolff et al. 2014; Griffin

TABLE 2. Start time, end time, and BT threshold of the 13 objects tracked using MODE-TD. Bold rows indicate objects with tracks terminated by the end of the forecast cycle and italicized rows indicate object tracks terminated due to merging either the simulated or observed object merging with another object.

Start date	Start time (UTC)	End date	End time (UTC)	BT threshold (K)
3 May 2017	**0100**	**3 May 2017**	**1200**	**220**
8 May 2017	1800	9 May 2017	0800	220
8 May 2017	**2100**	**9 May 2017**	**1200**	**220**
10 May 2017	1600	11 May 2017	0700	220
15 May 2017	*2200*	*16 May 2017*	*0600*	*220*
16 May 2017	*1800*	*17 May 2017*	*0400*	*220*
17 May 2017	0200	17 May 2017	0800	220
18 May 2017	*1800*	*19 May 2017*	*0800*	*220*
19 May 2017	**0800**	**19 May 2017**	**1200**	**215**
26 May 2017	**0900**	**27 May 2017**	**1200**	**225**
26 May 2017	**2100**	**27 May 2017**	**1200**	**220**
27 May 2017	**0300**	**27 May 2017**	**1200**	**220**
28 May 2017	**2000**	**29 May 2017**	**1200**	**220**

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

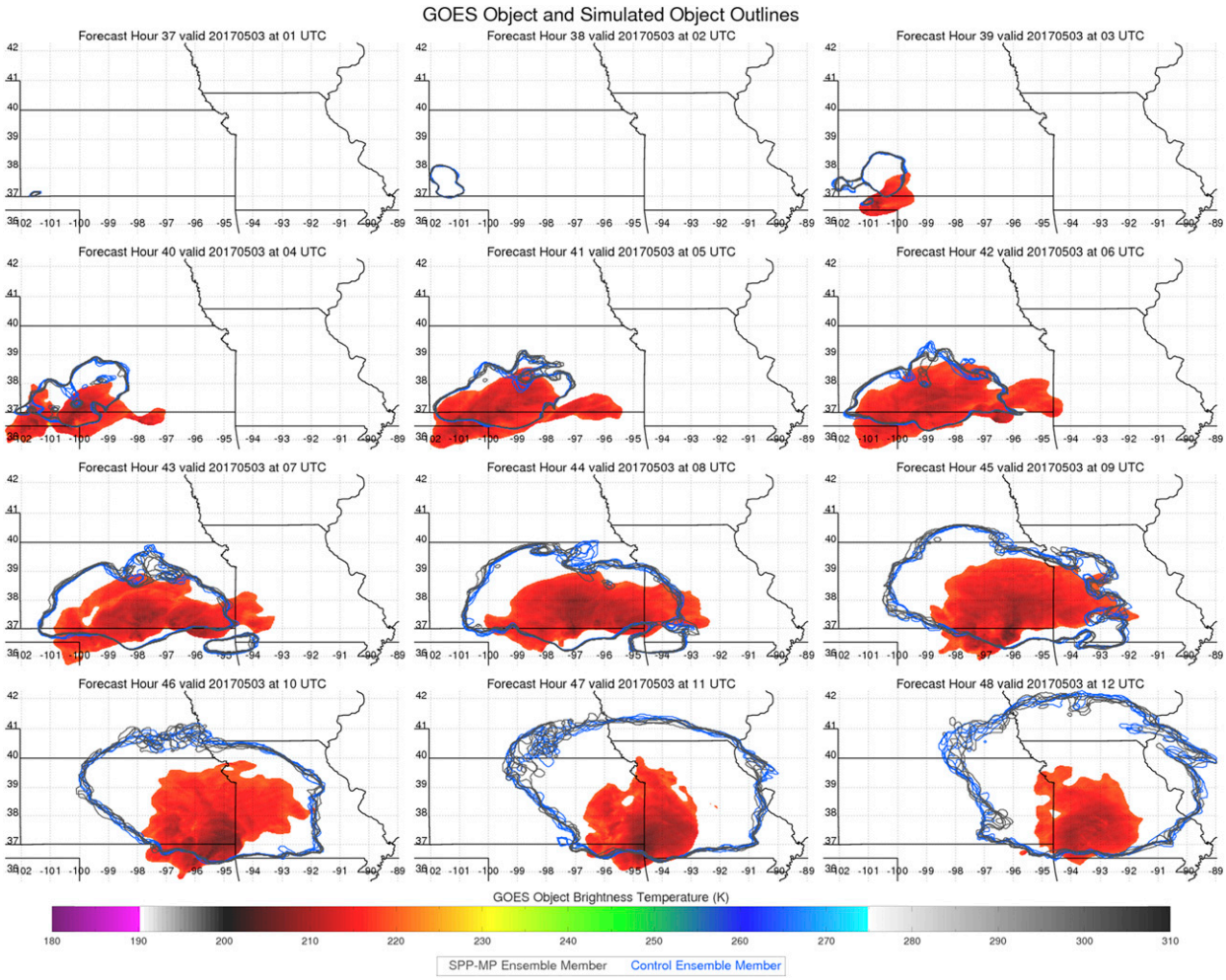

FIG. 2. Example of an object tracked using MODE-TD from convective initiation in the simulated imagery to the end of the forecast cycle. GOES object is plotted and SPP-MP (Control) ensemble member objects are outlined in gray (blue).

et al. 2017a), and a distinct diurnal cycle is observed with higher error between 2000 and 0000 UTC. For both months, the difference between the MAE is not statistically significant. However, this is not necessarily unexpected. The perturbations from SPP-MP are only active when clouds form, which in turn means that perturbations are limited spatially and temporally. Therefore, a conditional MAE is calculated for only pixels where either the ABI or any ensemble member 10.3 μm BT is lower than a given threshold. As this BT threshold decreases, the ensemble mean BT MAE increases and continues to be higher for May compared to January (not shown), further illustrate that smaller-scale and colder cloud features are harder to predict in both ensembles. The difference between the SPP-MP and Control ensemble mean BTs MAEs can be seen in Fig. 4. As the BT threshold decreases, the magnitude of the difference between the SPP-MP and Control ensemble MAE grows. The positive (negative) difference for May (January) indicates the SPP-MP is producing less (more) accurate BTs compared to the Control ensemble. However, these differences are also not statistically significant at the 95th percentile.

Bias in the SPP-MP and Control simulated 10.3 μm BTs can be seen in Fig. 5. The positive bias for both May and January indicates that the simulated BTs are too high overall compared to the observed GOES BTs. SPP-MP ensemble mean BTs are lower than the Control ensemble mean BTs, though, based on the smaller positive bias. For the January case, the ensemble mean SPP-MP BTs are lower than the Control ensemble mean BTs in a statistically significant way. At lower BT thresholds (not shown), the average bias for both May and January becomes rather negative, and the bias for the SPP-MP ensemble mean BTs is more negative compared to the bias for the Control ensemble mean BTs. To identify

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

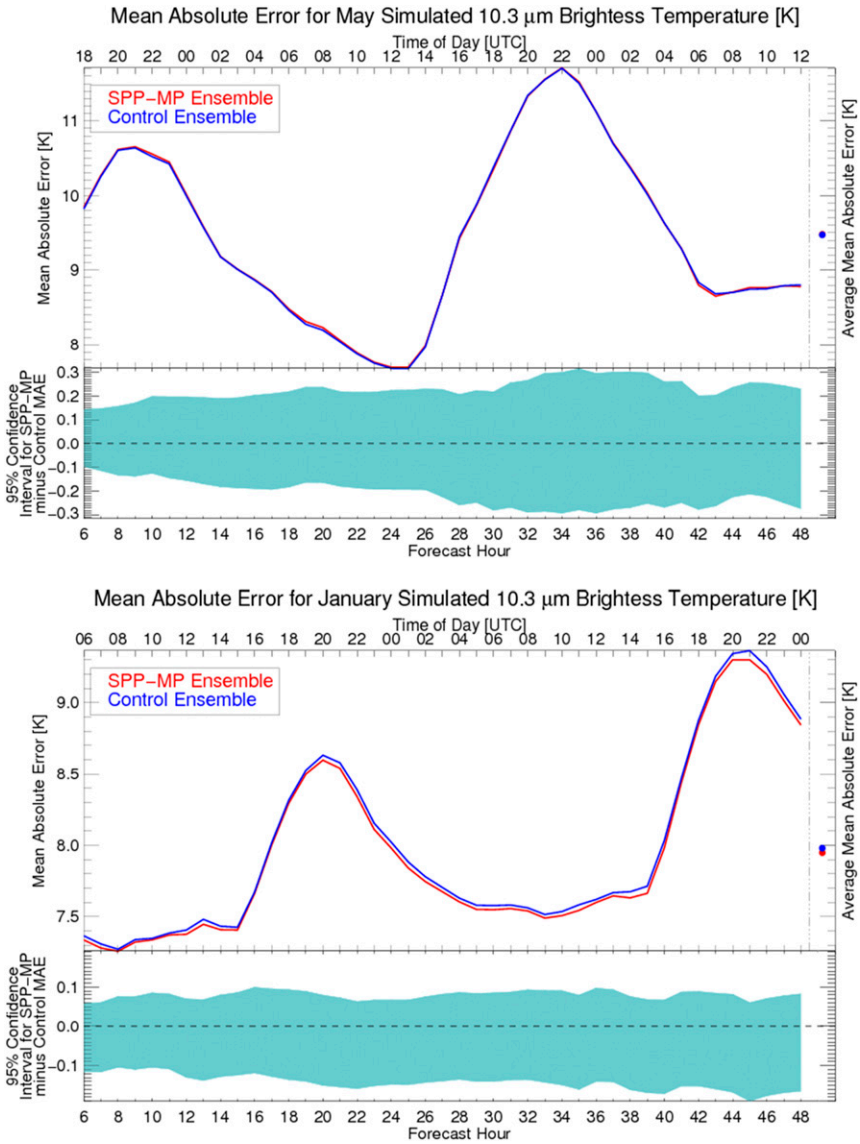

FIG. 3. Line plot of *GOES-16* ABI 10.3 μm brightness temperature mean absolute error (MAE) for (top) May 2017 and (bottom) January 2018 based on forecast hour. The cyan envelope represents the 95% confidence interval around the difference between the SPP-MP and Control ensemble MAE. If the envelope does not encompass zero, a statistically significant difference exists.

why the bias switches from positive to negative for lower BT thresholds, Fig. 6a displays the average difference between the percent of BTs lower than a given threshold for each ensemble and the GOES observation. Red (blue) squares indicate where a higher percentage of GOES (ensemble) pixels are lower than the BT threshold. Biases are mostly positive over the full domain because a higher percentage of GOES BTs are lower than 270 K, and therefore a larger percentage of ensemble BTs are higher than 270 K compared to GOES. For instances where a negative bias is evident over the full domain, more ensemble BTs are lower than 260 K compared to forecast hours with positive biases. Negative biases are observed at lower BTs because more ensemble BTs are lower than these thresholds. Therefore, both ensembles are not producing enough low-level clouds or optically thin cirrus clouds while producing too many thick upper-level clouds. However, the SPP-MP ensemble does produce fewer upper-level clouds for some forecast hours, as seen in the red squares for the lowest BTs. The SPP-MP ensemble produces more

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

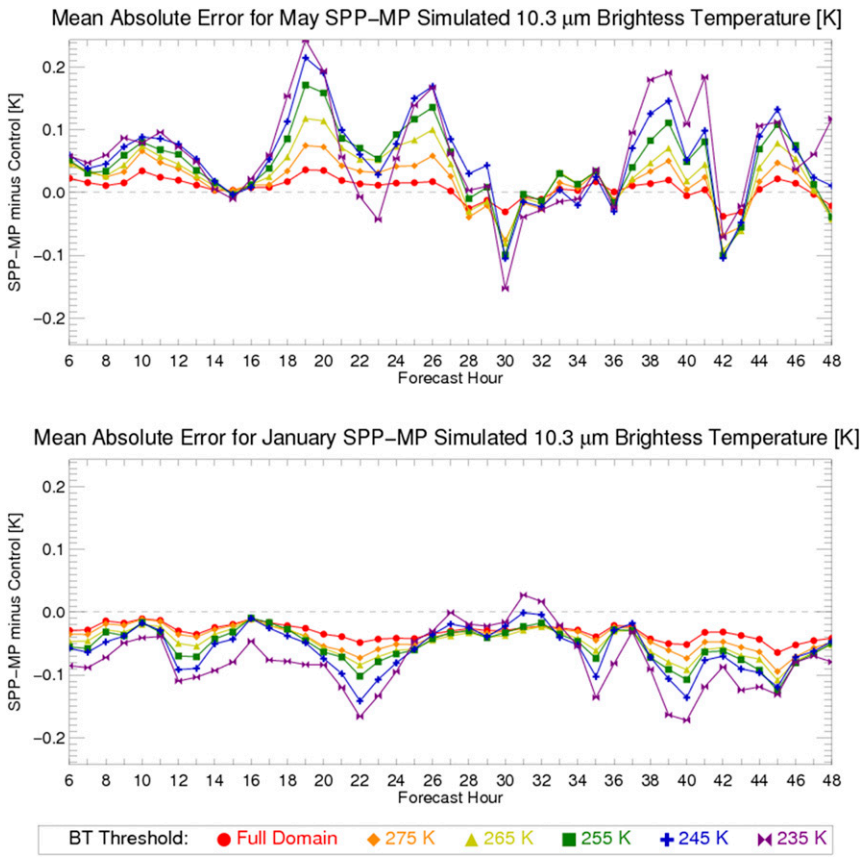

FIG. 4. SPP-MP minus Control Mean absolute error for ensemble mean 10.3 μm brightness temperature for 6 different brightness temperature thresholds.

BTs lower than 270 K compared to the Control ensemble, as seen in Fig. 6b.

There are two potential hypotheses for explaining the overall lower BTs with the SPP-MP compared to the Control BTs, as seen in the MBE in Fig. 5. The first hypothesis is that the SPP-MP ensemble contains more clouds, resulting in a lower domain-average BT. The second hypothesis is that the SPP-MP ensemble produces approximately the same number of clouds as the Control ensemble, but that these clouds are colder due to their microphysical composition or are located at a higher (e.g., colder) altitude. To determine if either or both of the processes contributes to the lower SPP-MP BTs, the distribution of cloud liquid, snow, and ice water content is analyzed for the January ensemble members. Figures 7a and 7b show composite profiles from the SPP-MP and Control ensembles for an arbitrarily selected forecast. The composite profile from each ensemble is calculated using the same 5000 data points from all 5 ensemble members, or 25 000 profiles. Collectively, the 5000 data points must have a BT lower than 250 K in all 10 ensemble members. In addition, the absolute difference between the ensemble mean BTs of these 5000 data points must be greater than 0.25 K.

As seen in Fig. 7c, larger snow content above 400 hPa is associated with lower BTs. To verify this is not a result of the arbitrarily selected forecast displayed in Fig. 7, the process used to create Fig. 7 was repeated to create composite profiles from 50 random forecasts where the forecast hour is 6 and higher. Of these 50 composite forecast profiles, 66% (33 profiles) have higher snow content associated with lower BTs. Therefore, it is assumed that higher snow content contributes to lower BTs. Based on these profiles, a snow content threshold of 10^{-6} kg kg^{-1} above 400 hPa is used to identify an upper-level cloud. This threshold identified 79.5% (63.0%) of SPP-MP (Control) profiles used to generate Fig. 7 as a cloud.

When analyzing the full dataset, the total snow content for cloudy pixels above 400 hPa is larger for the SPP-MP ensemble compared to the Control. However, the SPP-MP ensemble also has more pixels exceeding this snow content cloud threshold, similar to Figs. 7a and 7b. When dividing the total snow content by the number of pixels exceeding the snow content threshold, the SPP-MP ensemble has less

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

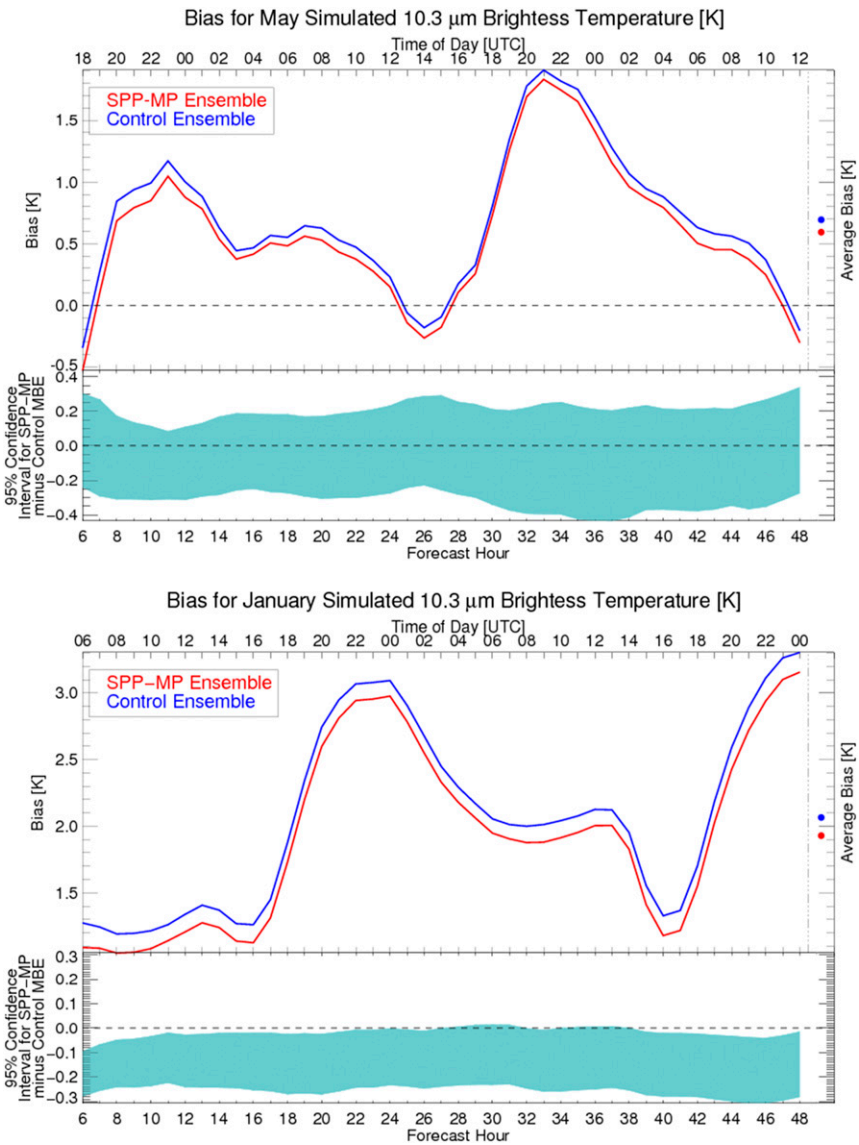

FIG. 5. As in Fig. 3, but for *GOES-16* ABI brightness temperature bias.

snow above 400 hPa per pixel than the Control. Therefore, the lower BTs in the SPP-MP are due to the SPP-MP forecasts producing more cloud pixels instead of lower BTs where clouds exist. This can be observed in Fig. 6b, as more Control ensemble pixels are seen for the lowest BTs.

2) CONTINUOUS RANKED PROBABILITY SKILL SCORE

The comparison between the CRPSS of the SPP-MP and Control ensembles for the 10.3 μm BTs can be seen in Fig. 8. During May, the CRPSS is negative for about 61% of all forecast hours, indicating that the SPP-MP ensemble has lower skill than the Control ensemble. For January, the SPP-MP ensemble is more skillful, shown by the positive CRPSS observed in over 75% of the forecast hours. A lower CRPSS can either be the result of a reduction in ensemble spread width or improvement in accuracy, or a combination of both. First, the SPP-MP ensemble spread could be wider than the spread in the Control ensemble. Another explanation for a negative CRPSS is that the observed BT is outside the range of the SPP-MP ensemble BT CDF more often compared to the Control.

Decomposing the CRPS into $\mathrm{CRPS}_{\mathrm{reliability}}$ and $\mathrm{CRPS}_{\mathrm{potential}}$, the SPP-MP ensemble has a lower $\mathrm{CRPS}_{\mathrm{reliability}}$ value than the Control ensemble for 35% (73%) of May (January) forecast hours (not

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

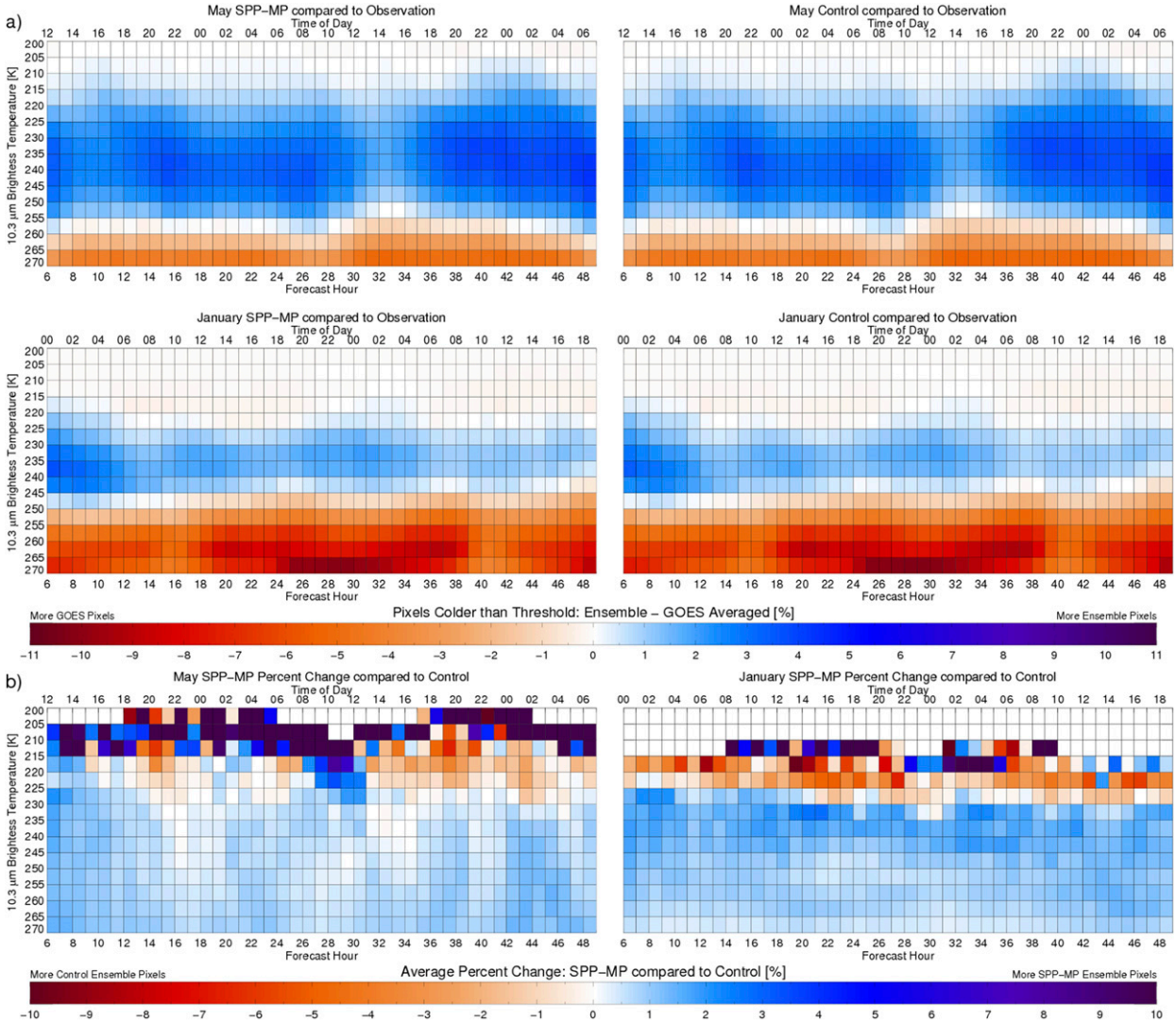

FIG. 6. (a) Average of percent of pixels with a 10.3 μm BT lower than a given threshold for each ensemble minus the percent of pixels with a BT lower than a given threshold for the GOES observations. Red (blue) squares indicate where a higher percentage of GOES (ensemble) pixel BTs are lower than the BT threshold. (b) Percent change between SPP-MP and Control ensemble for forecast hours averaging at least 1000 pixels smaller than the given threshold. Red (blue) squares indicate where a higher percentage of SPP-MP (ensemble) pixel BTs are lower than the BT threshold.

shown). These overall less reliable SPP-MP forecasts in May are likely due to the ensemble not enclosing the observed BTs, as the SPP-MP ensemble pixels do not enclose the observed BT in 70.4% (289 out of 410) of the forecast hours and initialization times. In January, this percentage is reduced to 49.8% (214 out of 430; more forecast hours are available in January as seen in Fig. 8). The $\mathrm{CRPS}_{\mathrm{potential}}$ is lower for the SPP-MP ensemble than the Control for 57% (48%) of May (January) forecast hours (not shown). This lower $\mathrm{CRPS}_{\mathrm{potential}}$ for the SPP-MP ensemble in May is due to a narrower spread in the ensemble BTs compared to the Control ensemble when both ensemble systems encompass the observation. Overall, more pixels exhibit a smaller spread in the SPP-MP ensemble spread compared to the Control for 75.6% (32.1%) of forecast hours and initializations in May (January). When the observed BT is outside of the ensemble system, these smaller outliers can also influence the $\mathrm{CRPS}_{\mathrm{potential}}$. Based on an example of the CRPS for the SPP-MP and Control ensembles in Fig. 9, valid at 0700 UTC 11 May 2017, the CRPS is the highest around the edges of the cloud objects. Therefore, it is possible the cloud edges in the SPP-MP ensemble are less accurately defined during May compared to the Control ensemble, which could result in a distribution of BTs

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

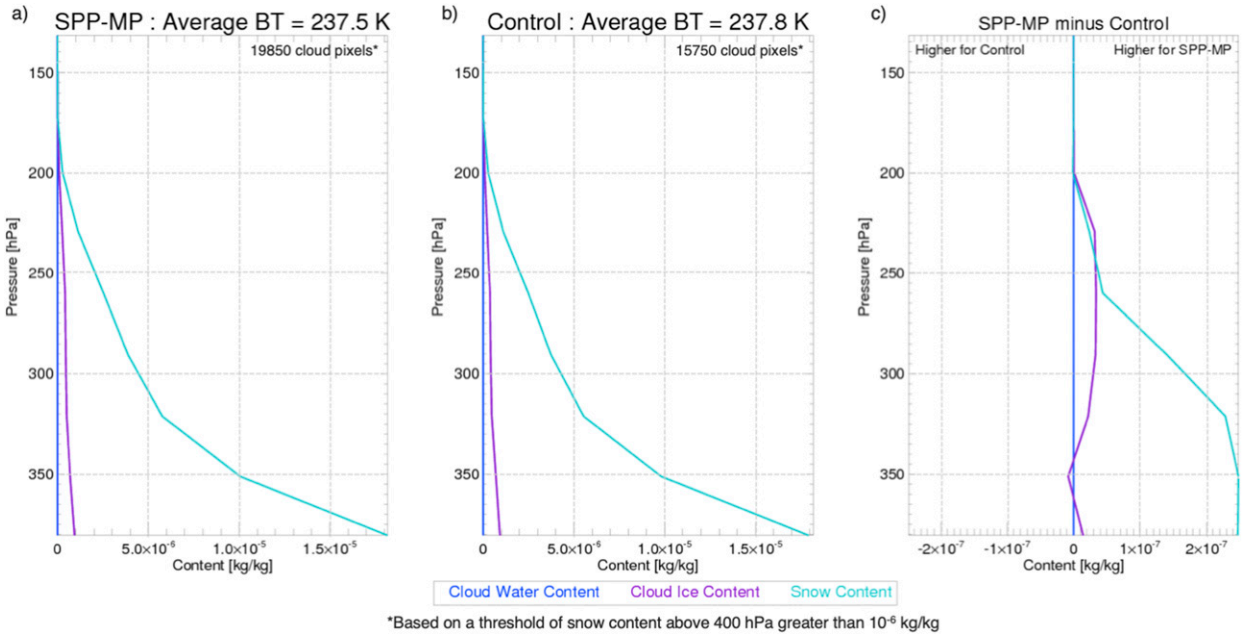

FIG. 7. Composite of 5000 random vertical profiles of cloud liquid water content, cloud ice water content, and cloud snow water content from a randomly chosen forecast. The *GOES-16* ABI 10.3 μm BT for each pixel must be lower than 250 K.

that does not encompass the observed BT and a poorer reliability.

3) BRIGHTNESS TEMPERATURE DIFFERENCES

Brightness temperature differences (BTD) between two satellite bands can be used to examine how well the SPP-MP and Control ensembles represent observed cloud properties, such as cloud top height and hydrometeor phase. Cloud top height is examined using a 6.9–11.2 μm BTD (Cintineo et al. 2014), where strong water vapor absorption at 6.9 μm combined with 11.2 μm BTs that generally decrease with height results in negative BTDs. The largest 6.9–11.2 μm BTDs generally occur in clear-sky regions (Mecikalski and Bedka 2006), with progressively smaller BTDs as the cloud top height increases. Discriminating between liquid and ice clouds can be done using an 8.4–11.2 μm BTD, as absorption for ice (water) is higher (lower) at 8.4 μm compared to 11.2 μm (Ackerman et al. 1990; Strabala et al. 1994; Baum et al. 2000). Ice clouds are characterized by a positive 8.4–11.2 μm BTD, while the reverse is true for water clouds (Otkin et al. 2009).

A two-dimensional histogram of 6.9–11.2 μm BTD compared to the 11.2 μm BT can be seen in Fig. 10. Only pixels with an 11.2 μm BT colder than 270 K are shown to focus on cloudy pixels. Overall, the shape of the SPP-MP and Control ensemble histograms matches the observation for both May and January. One notable exception is 6.9–11.2 μm BTDs that are greater than 0 K for the May BTs (Figs. 10g,h). A positive 6.9–11.2 μm BTD is indicative of overshooting clouds exceeding the tropopause height (Schmetz et al. 1997). Both the SPP-MP and Control ensembles have more pixels exceeding the 0 K BTD threshold than the observations. This indicates convection may be too vigorous in the model forecasts regardless of which perturbation method is used, which could be a consequence of the model's low vertical resolution (Roeckner et al. 2006). The SPP-MP has a lower probability of pixels exceeding the 0 K BTD threshold than the Control ensemble, but this difference is small and does not appear on the difference plot (Fig. 10i). For January, the peak in 6.9–11.2 μm BTDs greater than 0 K is centered between 230 and 250 K, with a positive 6.9–11.2 μm BTD existing for pixels as high as 270 K (Figs. 10d–f). Positive 6.9–11.2 μm BTD can also be a result of clear-sky inversions over cold surface 11.2 μm BTs (Ackerman 1996). The SPP-MP ensemble has more low-BT pixels with a less negative 6.9–11.2 μm BTD compared to the Control for both months (Figs. 10i,l). As the 11.2 μm is an IR window channel like 10.3 μm, this corresponds with the results indicating that the SPP-MP 10.3 μm BTs are overall lower than the Control BTs. The reduction in the lowest BTs for the SPP-MP ensemble can be seen in the Figs. 10i and 10l, as a higher probability of BTs lower than 220 K is observed for the Control

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

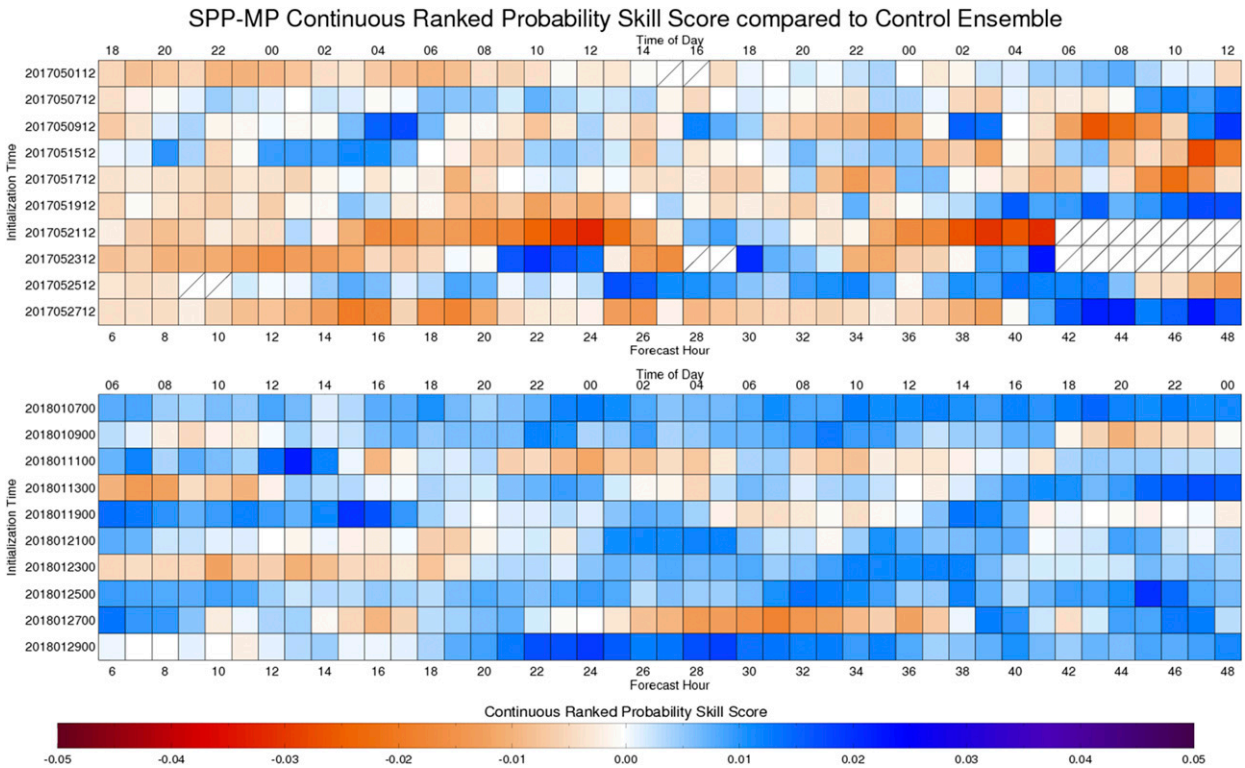

FIG. 8. Continuous ranked probability skill score (CRPSS) for (top) May 2017 and (bottom) January 2018. Blue indicates the CRPS for the SPP-MP ensembles is better than the CRPS for the Control ensembles. Squares with backslashes denote times with missing *GOES-16* ABI data, which were considered preliminary during this time period.

ensemble. The SPP-MP ensemble also extends the range of 6.9–11.2 μm BTD for a given 11.2 μm BT compared to the Control.

The discrimination between ice and water clouds can be seen in the two-dimensional histogram of 8.4–11.2 μm BTD compared to the 11.2 μm BT in Fig. 11. One immediate observation is the high probability of positive BTD for both May and January SPP-MP and Control ensembles compared to the observations (Figs. 11b,c,e,f). Therefore, both of the ensembles are producing too many ice clouds. The SPP-MP ensemble has slightly more of these ice clouds compared to the Control, especially for May (Fig. 11i), potentially due to a negative bias in the simulated 11.2 μm BTs. By averaging the difference between CDFs of the simulated and observed 8.4 μm BTs, and doing the same for the 11.2 μm BTs, it was found that the departure between the simulated and observed 11.2 μm BTs is lower than the 8.4 μm BTs. The 8.4 μm is centered on a weak water vapor absorption line (Ackerman et al. 1990), potentially making it less susceptible to negative biases from the clouds. These lower 11.2 μm BTs result in a positive 8.4–11.2 μm BTD and an overabundance of ice clouds. There is also a notable lack in water clouds in May for 11.2 μm BTs between 230 and 280 K. At higher BTs, both the SPP-MP and Control histograms better match the observations.

b. Object-based metrics

To determine whether the clouds from each ensemble forecast accurately represents cold, upper-level cloud objects in the *GOES-16* imagery, objects from simulated and observed 10.3 μm BTs are identified using a 235 K threshold in MODE. As seen in Fig. 12, on average slightly more cloud objects are simulated in the SPP-MP ensemble than in the Control during May; however, this difference in not statically significant. This would help explain the negative CRPSS, as more clouds could result in more cloud edges and therefore increase the CRPS. A diurnal cycle in the number of objects also exists in May. Both ensembles produce fewer cloud objects compared to the observations from approximately 1800–0000 UTC, or late afternoon local time, and too many objects during all other times. In January, SPP-MP has fewer objects compared to the Control and a mostly lower CRPS. Compared to the observations, the median number of simulated objects in both ensembles are similar during May but much lower in January. However, the area encompassed by the simulated objects is much larger than the area of observed objects for both ensembles in both

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

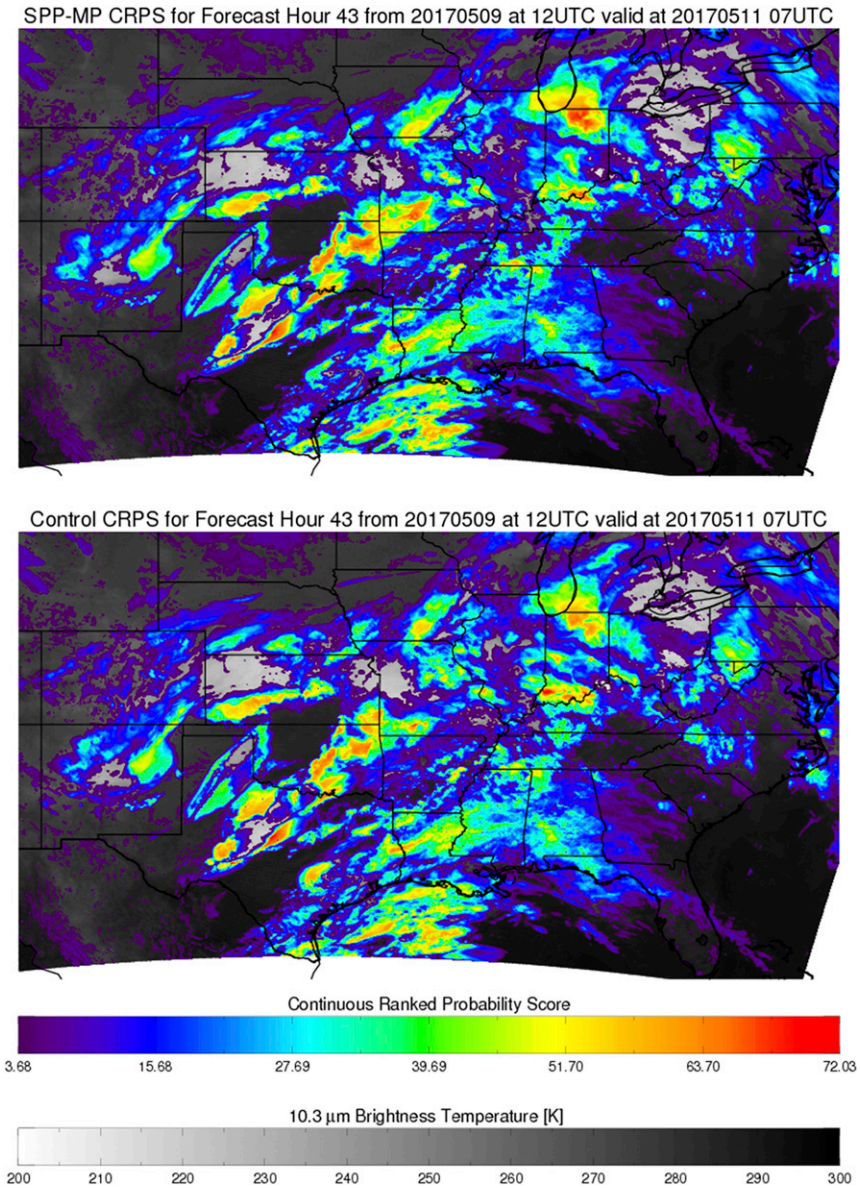

FIG. 9. Continuous ranked probability score valid at 1200 UTC 9 May 2017.

months (Fig. 13), consistent with the negative bias in the 10.3 μm BTs. During May, about 44% of the forecast hours have a smaller average object size in the SPP-MP ensemble compared to the Control ensemble. During January, this occurs in only 14% of forecast hours.

Spatial displacement errors between the observed and simulated cloud objects can be removed by centering objects using the object centroid latitude and longitude identified by MODE. An example can be seen in Fig. 14. In Fig. 14a, the observed object is about a half degree north of the simulated object. The differences between the observed and simulated BTs before and after the objects have been overlaid on each other can be seen in Figs. 15b and 15c, respectively. To keep this analysis homogeneous, neither the observed nor the 10 simulated BTs objects can touch the domain boundary or have an interest score of zero, and at least one match between the observed and simulated objects must have an interest score higher than 0.65. It is important to note that this object-based methodology requires an object to exist in both the observed and simulated BT. This analysis uses between 24 and 82 objects, depending upon forecast hour and month.

The 10.3 μm BT MAE for cloud objects colder than 235 K can be seen in Fig. 15. Unlike the assessment over

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

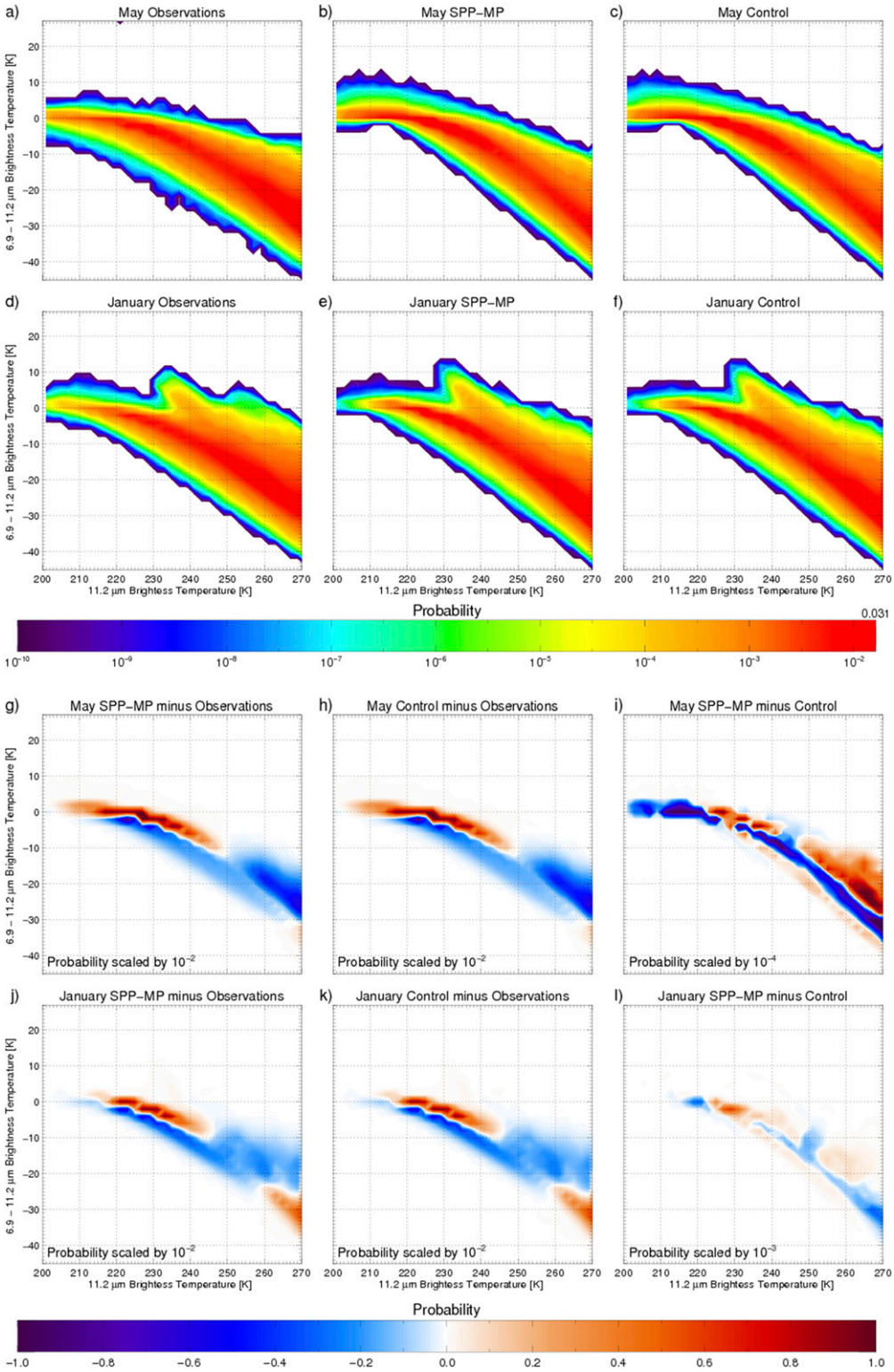

FIG. 10. Histogram of *GOES-16* ABI 6.9–11.2 μm brightness temperature differences plotted as a function of the ABI 11.2 μm brightness temperature for (top) May 2017 and (bottom) January 2018. The difference between the SPP-MP and Control histograms can be seen in the bottom row for (top) May 2017 and (bottom) January 2018.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

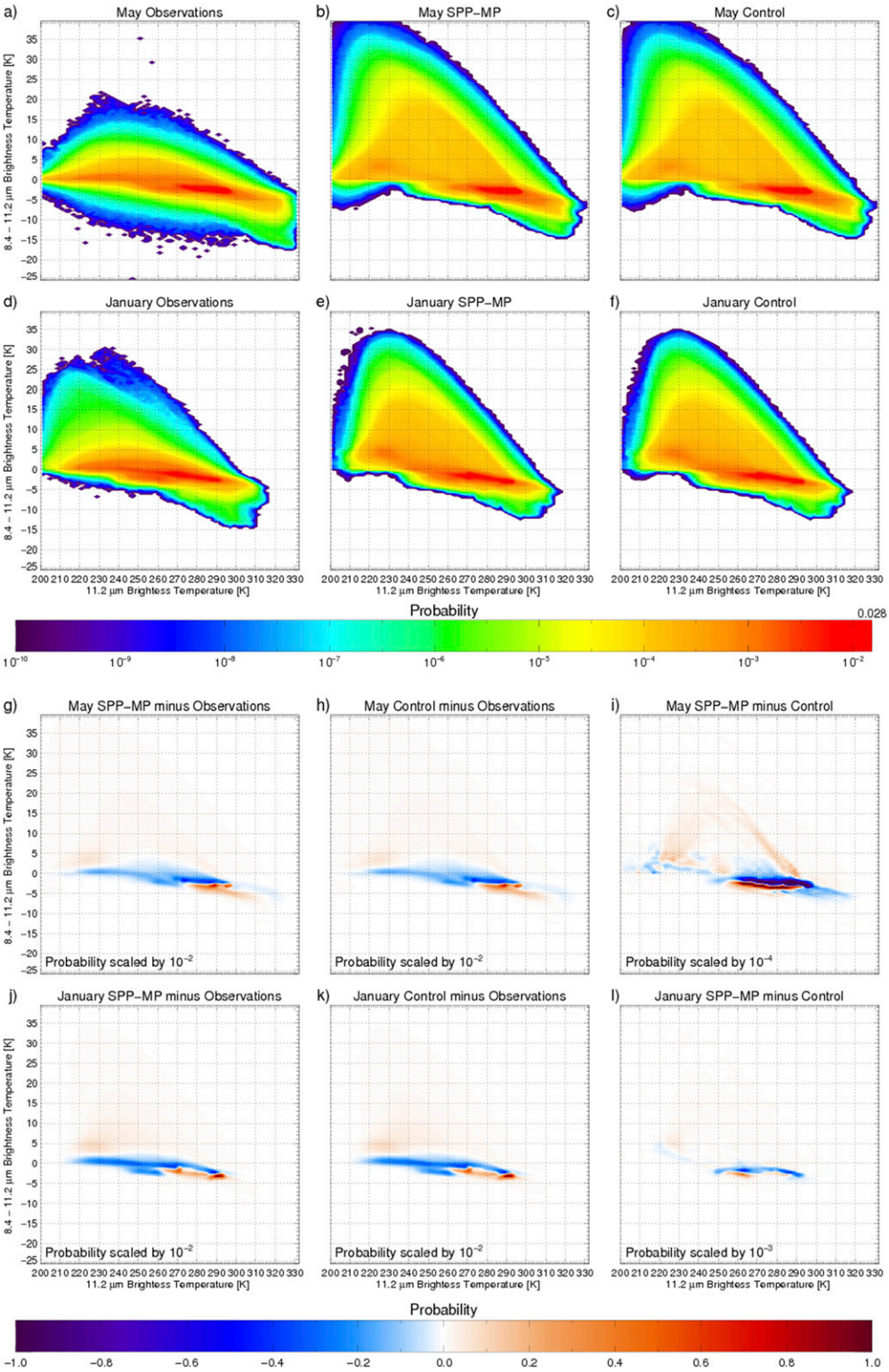

FIG. 11. As in Fig. 10, but for *GOES-16* ABI 8.4–11.2 μm brightness temperature differences.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

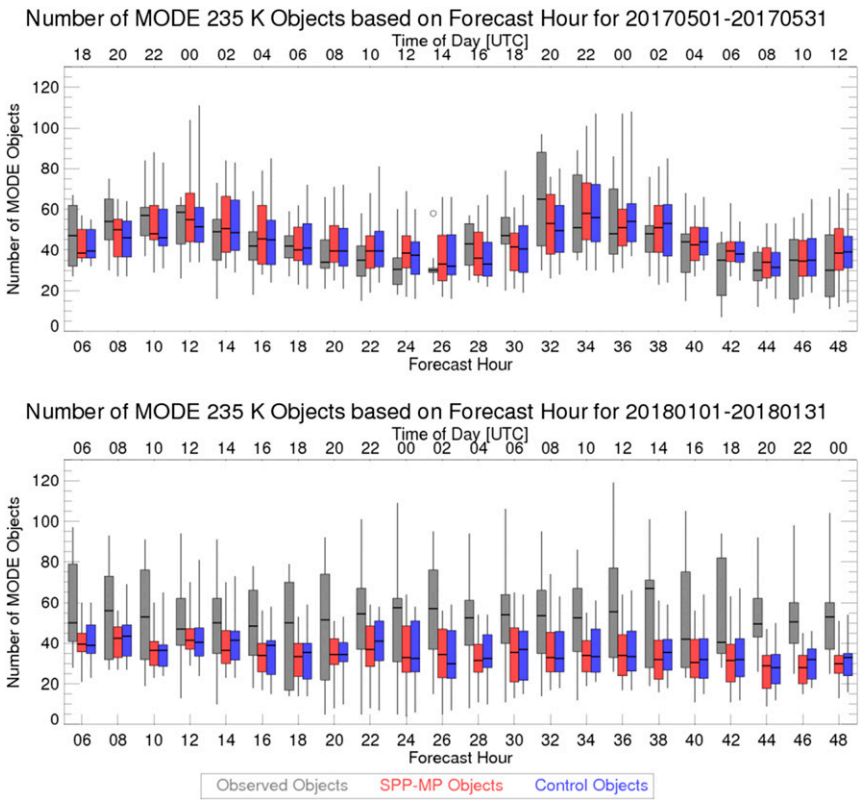

FIG. 12. Box-and-whisker plot of number of MODE objects for (top) May 2017 and (bottom) January 2018. Gray represents the number of *GOES-16* ABI objects while red (blue) indicate the number of SPP-MP (white noise) ensemble objects.

the full domain, the SPP-MP BTs are now on average more accurate than the Control ensemble BTs, indicating the SPP-MP BTs better represent the observed BTs for cold clouds once spatial displacement errors have been removed. Since the actual objects area is only the area within the black line, see Fig. 14a for an example, but analysis includes all the colored area, errors can be caused by differences in the actual observed and simulated object sizes. The ratio of the total observed object sizes to the simulated object sizes is called the simulated area ratio, and values less than one indicate the total simulated object size is larger. While most simulated area ratios are lower than 1, indicating larger simulated objects, there is no correlation between the MAE and simulated area ratio in May. Therefore, the error in the BTs identified by the MAE cannot be described by the difference in object sizes. However, in January, moderate correlation is observed, with higher MAEs associated with higher area ratios.

The 10.3 μm BT bias for the simulated cloud objects is shown in Fig. 16. For both May and January, the SPP-MP has a more negative/less positive bias than the Control, indicating that the BTs are lower for the SPP-MP. However, these biases are lower compared to those of pixels with BTs lower than 235 K over the full domain. The bias is highly correlated with the simulated area ratio. Smaller simulated area ratios usually result in more enhanced negative biases, as the simulated object is larger, and therefore potentially colder, than the observation. January cloud objects exhibit a positive bias, and the simulated area ratio is above one in some forecast hours.

The evolution of several attributes associated with cloud objects can be seen in Fig. 17. Figure 17a depicts the area ratio of matched observed and simulated objects. A ratio of 100%, represented by the dashed line in Fig. 17a, indicates that the observed and simulated objects are the same size. Simulated (observed) objects are larger below (above) the dashed line, with the area ratio becoming smaller until it reaches 0% at the top and bottom of the graph. Overall, once an observed object develops, it is on average larger than the simulated object. This continues for about the first 5 h of an observed object's lifetime before becoming smaller than the simulated

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

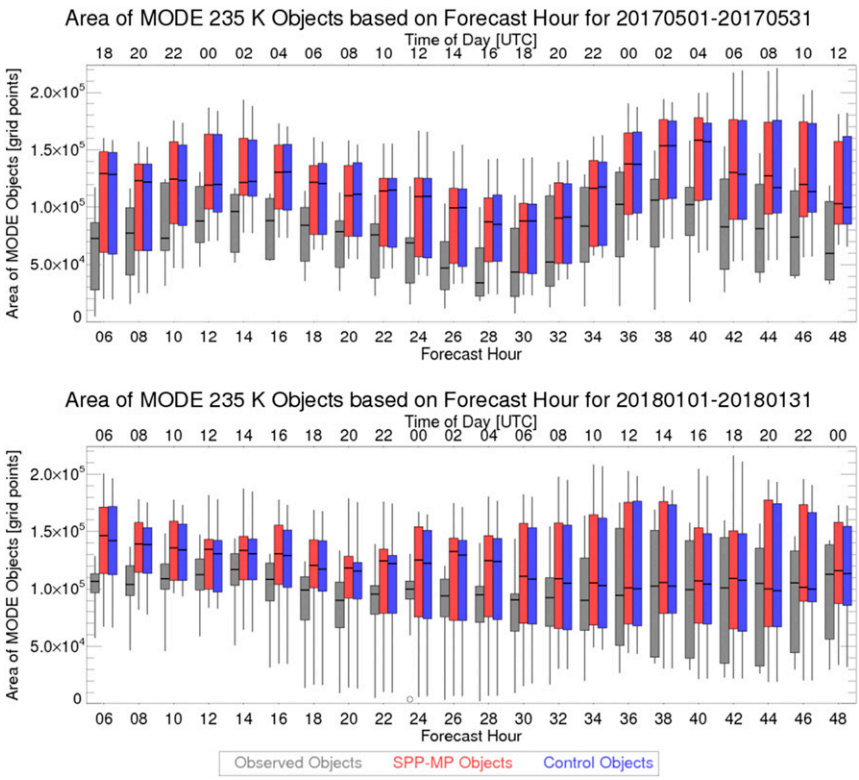

FIG. 13. As in Fig. 12, but for area of MODE objects.

object. The ratio of observed and simulated objects is smaller (closer to the top of the graph) in the SPP-MP ensemble members compared to the Control, which appears to corroborate the hypothesis that SPP-MP objects are smaller than the Control objects. Only later in an observed objects' life cycle is the paired Control object smaller than the SPP-MP object, on average. However, these differences between the SPP-MP and Control object sizes are not statistically significant based on a Welch's test (Welch 1947).

Figures 17b and 17c depict the distance between the centers of paired objects and overall interest scores of paired objects, respectively. Once the observed object has developed, the distance between the observed and simulated centers of objects is nearly constant for about the first nine hours before increasing with time. Interest scores are highest 2–9 h after the observed object develops, where the distance between the centers of objects is low and the area ratio is closest to 100%. The interest scores then decrease with increasing observed time. This result is not unexpected because lower interest values are associated with greater displacement errors between the simulated and observed objects, which tend to increase during the forecast (Griffin et al. 2017b). However, there is no clear distinction between the SPP-MP and Control ensembles in either interest score or distance between the center of objects.

5. Conclusions

In this study, the impact of using a stochastic perturbed parameterization (SPP) to add realistic perturbations to select cloud generation and dissipation processes in the TE14 microphysics scheme is analyzed. This is accomplished by comparing simulated and observed *GOES-16* ABI BTs from specific days during May 2017 and January 2018. Two different ensembles are created from the WRF output, and each ensemble is run for 48 forecast hours. The first is a five-member ensemble where the graupel *y*-intercept parameter, cloud water shape parameter, and the number of cloud condensation nuclei are allowed to vary, referred to as the SPP-MP ensemble. A second Control ensemble is generated by introducing spatially uncorrelated white noise perturbations in the lowest 800 m of the troposphere at the initialization time to four ensemble members and includes the control (unperturbed) initialization as the fifth member. The Control ensemble was designed to compare the impact of adding realistic perturbations to select cloud processes against small initial condition perturbations. It is

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

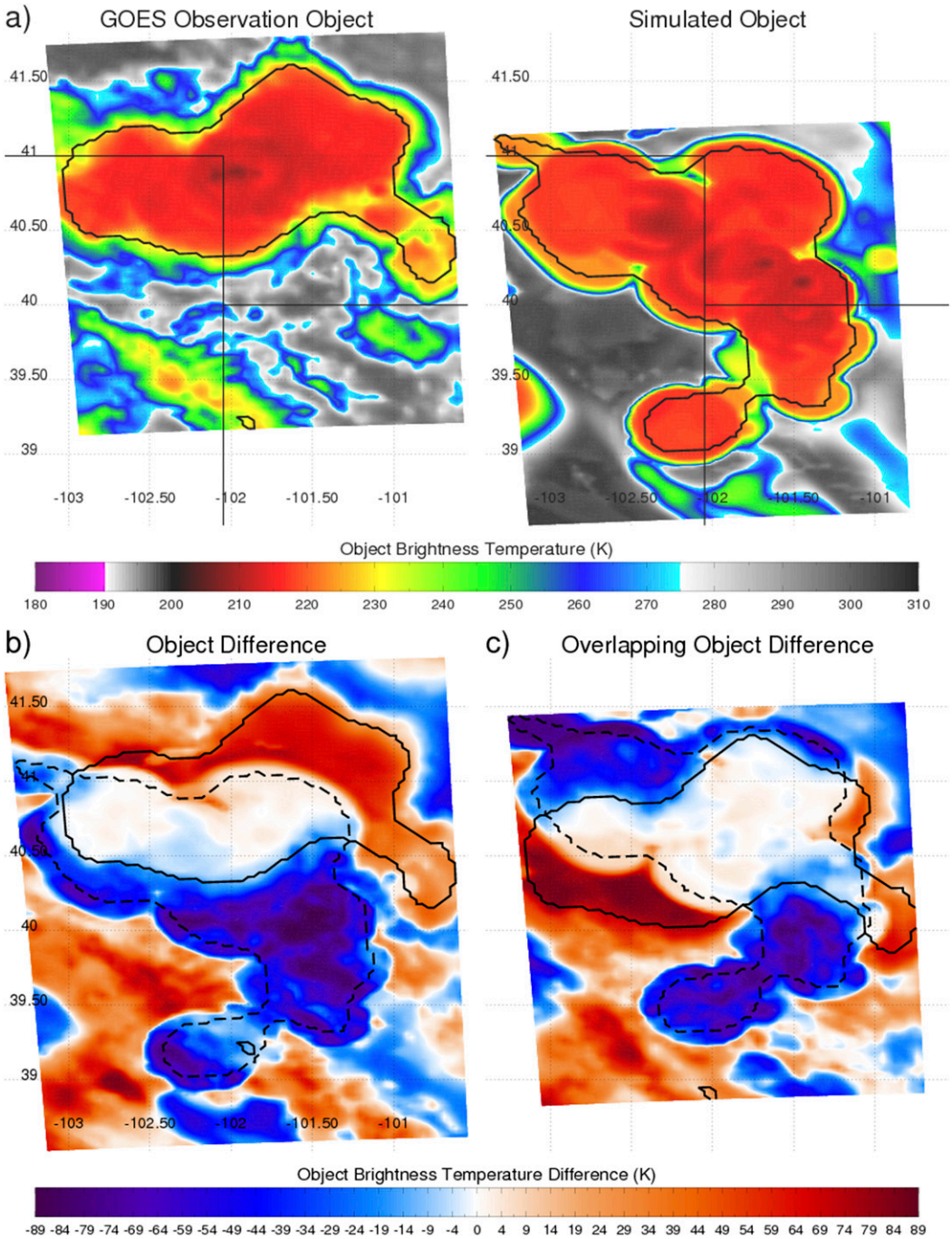

FIG. 14. Example of removing displacement using the center latitude and longitude of MODE images. (a) Two matching object that are displaced. Difference between BTs when the objects (b) are not and (c) are overlapped.

important to note that the SPP-MP perturbations can only impact the forecast when clouds are present, whereas the white noise perturbations have the potential to trigger convection in areas where no convection was present. Therefore, we did not attempt a full description of forecast uncertainty, which is additionally influenced by initial and lateral boundary condition uncertainty, as well as uncertainty in all other physical parameterization schemes. Instead, we focus on the quantification of uncertainty related to parameter uncertainties solely in the microphysics scheme.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

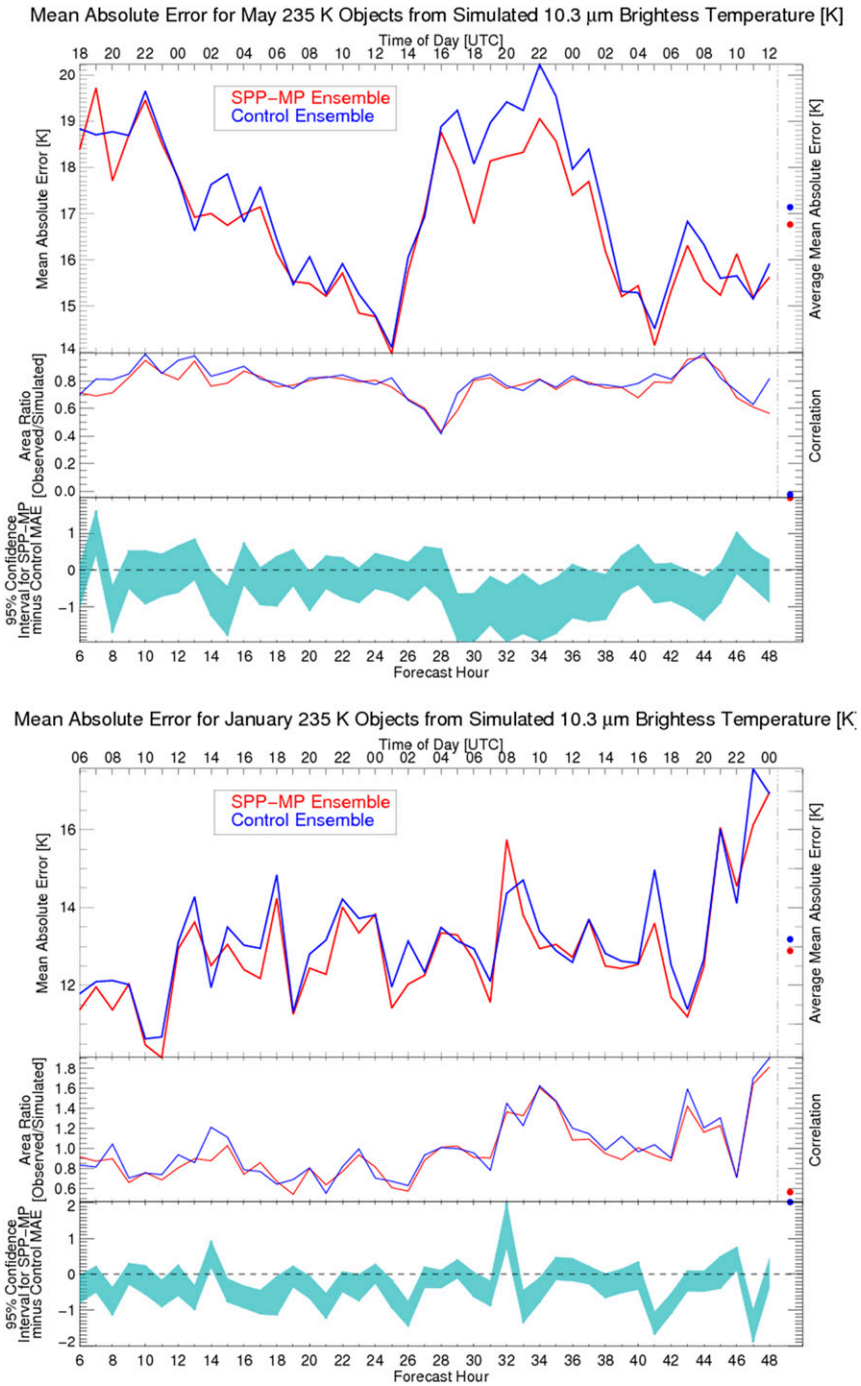

FIG. 15. Line plot of mean absolute error (MAE) and area ratio for the cloud objects containing ABI 10.3 μm brightness temperatures lower than 235 K as identified by MODE for (top) May 2017 and (bottom) January 2018 based on forecast hour. Objects are overlaid using the method from Fig. 14. The cyan envelope represents the 95% confidence interval around the difference between the SPP-MP and Control ensemble MAE. If the envelope does not encompass zero, a statistically significant difference exists.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

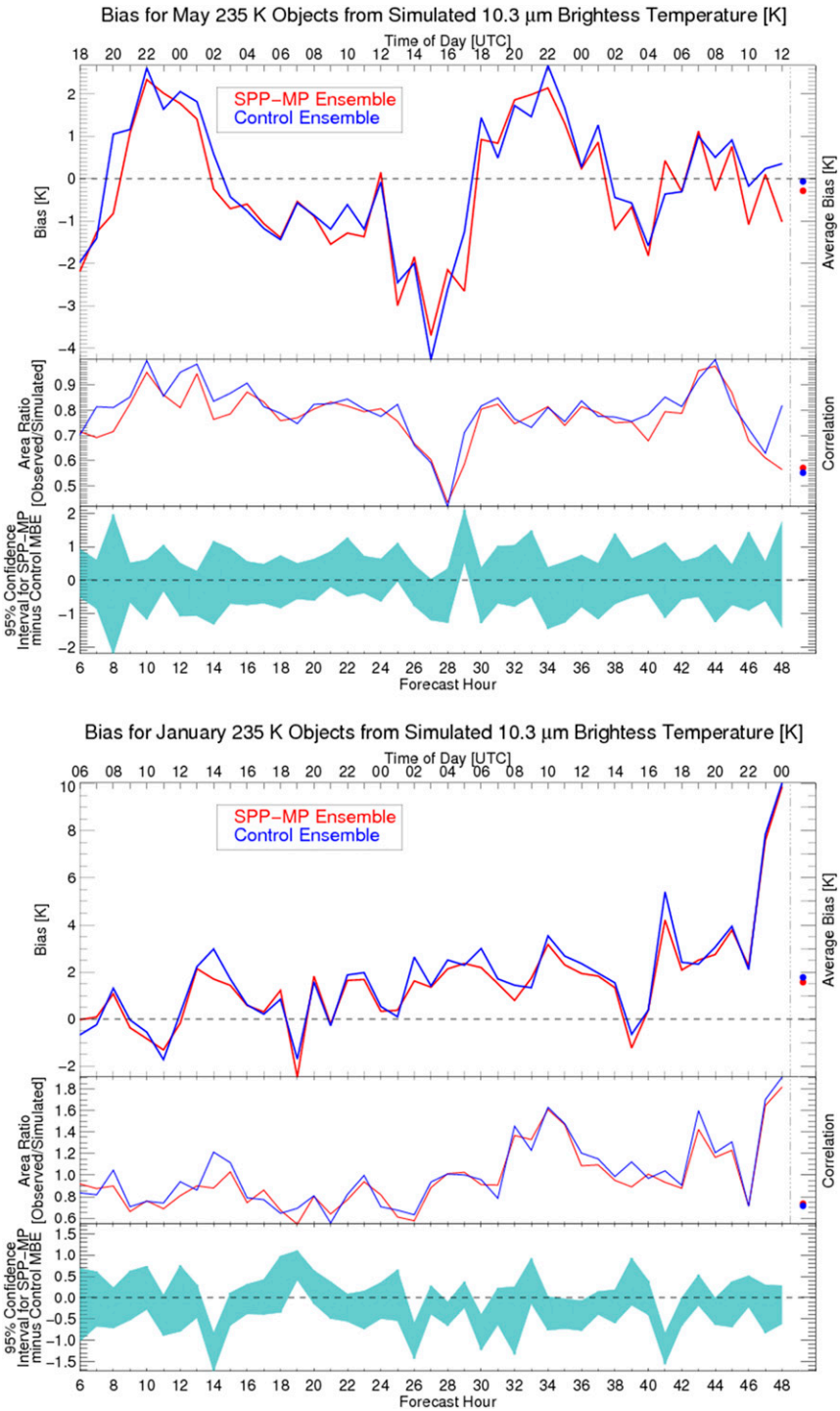

FIG. 16. As in Fig. 15, but for Bias.

Overall, it is found that the SPP-MP perturbations result in lower BTs compared to the Control ensemble and more cloudy pixels. Some specific findings related to these lower BTs are summarized as follows:

1) Over the full domain, the SPP-MP ensemble mean BTs have a lower mean absolute error (MAE) in January 2018 and similar MAEs in May 2017 when compared to the Control ensemble. The SPP-MP ensemble members have a lower overall positive bias

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

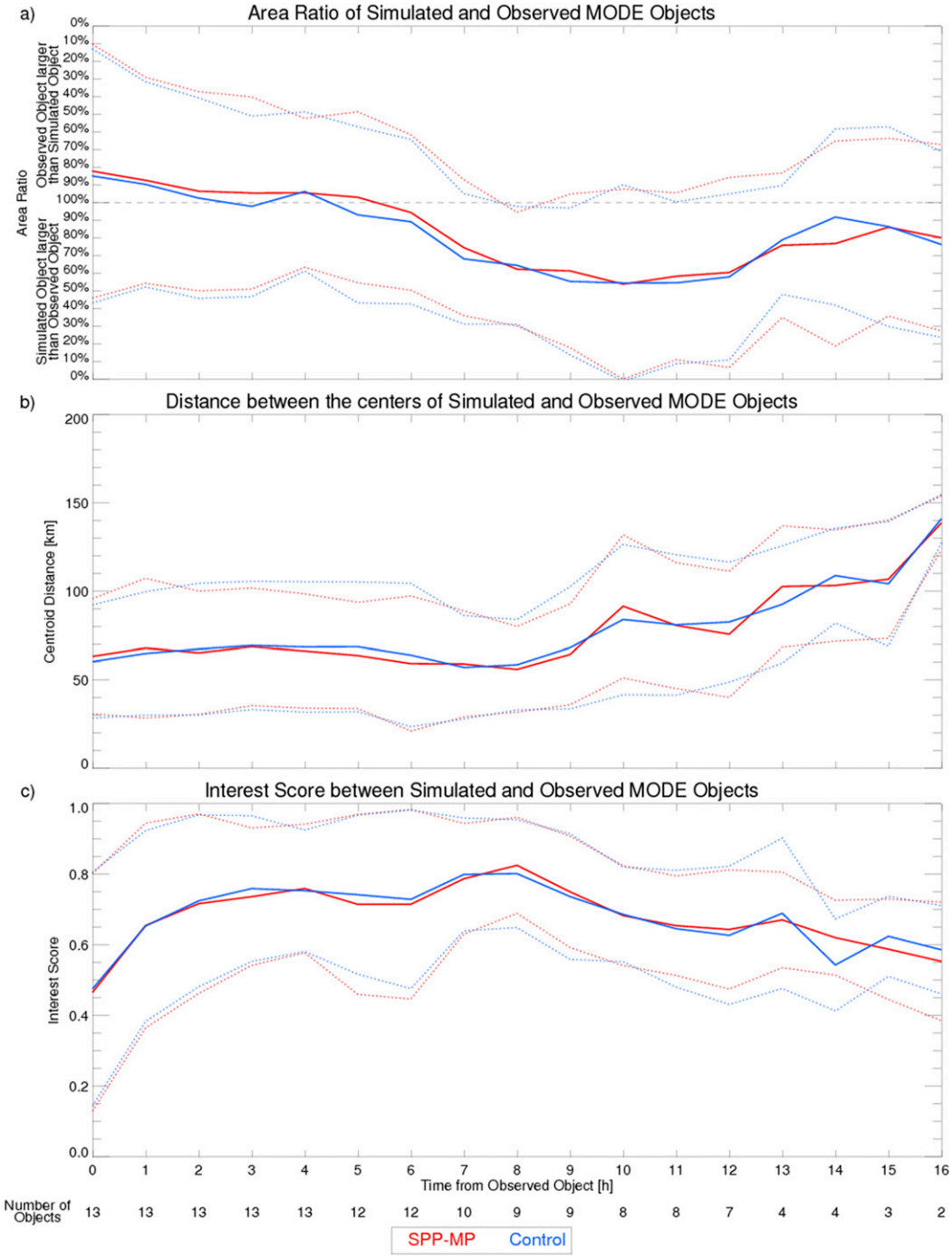

FIG. 17. (a) Area ratio between, (b) distance between the centers of, and (c) interest score between paired observed and simulated objects as a function of observed object life cycle. Red (blue) lines represent that SPP-MP (Control) ensemble members. The solid line is the average over all ensemble members, with the dashed line representing ±one standard deviation.

compared to the Control by producing more low-level clouds or optically thin cirrus. While both ensembles have an excess of thick upper-level cloud, the SPP-MP ensemble does produce fewer clouds with 10.3 μm BTs lower than 225 K compared to the Control.

2) The SPP-MP ensemble produces more ice clouds than the Control ensemble and observations, especially

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

during May. This is evidenced by the low bias in the 11.2 μm BTs compared to the 8.4 μm and a positive 8.4–11.2 μm brightness temperature difference (BTD). Using a 6.9–11.2 μm BTD, it is observed that SPP-MP ensembles produce less vigorous convection than the Control ensembles, but convection in both ensembles is too vigorous compared to the observations.

3) More cloud objects are produced by the SPP-MP in May 2017 compared to the Control, with the opposite observed in January 2018. These additional cloud objects potentially result in a higher continuous ranked probability score (CRPS) when compared to the Control. In May, the SSP-MP lower skill is likely due to the ensemble not enclosing the observed BTs, especially along the edges of observed clouds. In January, the SPP-MP ensemble BT distribution better represents the observed BT than the Control ensemble BT distribution.
4) When looking at matched simulated and observed cloud objects defined using a 235 K threshold (that do not touch the domain boundary), the SPP-MP ensemble has a lower MAE. Since the SPP-MP can only impact existing clouds, instead of triggering new convection like the Control ensembles, removing displacement errors between the observed and forecast clouds can help identify how the SPP-MP improves the cloud characteristics. In May, cloud objects are too cold compared to the observations, with the opposite occurring in January. The bias in the ensemble BTs can be described by the difference in sizes between the observed and simulated object, where observed objects smaller (larger) than the simulated objects are moderately correlated with negative (positive) biases.
5) Tracking the evolution of cloud objects, at the initiation of the observed object the SPP-MP and Control objects are smaller than the corresponding GOES objects. This persists for about the first 5 h of an observed object's lifetime before simulated objects become larger than the observed object. Only later in an observed object's life cycle is the paired Control object smaller than the SPP-MP object, on average. No clear distinction exists between the SPP-MP and Control ensemble interest score and distance between the centers of objects.

In the proverbial stratospheric view of the overall impact of SPP-MP as a method of increasing ensemble forecast spread in the context of NOAA's desire to move toward a unified model system with single physics plus various perturbation methods, SPP-MP should probably be combined with SPP applied to other parameterization schemes (i.e., PBL, LSM, radiation, etc.) similar to Jankov et al. (2019). As stated above, SPP-MP inherently cannot act on clear sky areas whatsoever so it shouldn't be expected to show improved weather forecast capabilities in the very short term as the internal perturbations take time to manifest as cloud property changes that affect incoming radiation, convective cold pool development, or anything else fundamental enough to change the forecast model results. As such, convective-scale "day 1" forecasts are unlikely to see much change when using SPP-MP in comparison to SPP-PBL for example. However, microphysical and dynamical feedbacks do occur using SPP-MP and its application might be better suited to forecast lead times beyond 24 h.

As this study mostly investigates the impact of the SPP method on the 10.3 μm BTs, future work will include extending this analysis to other *GOES-16* satellite BTs. For example, unlike the 10.3 μm BTs, the 6.2 or 6.9 μm BTs are sensitive to water vapor at different levels of the troposphere and therefore can indicate how the SPP impacts atmospheric water vapor. Additional studies will focus on examining relationships between the satellite brightness temperatures and other aspects of the model, such as the jet stream location and stability, as well as incorporate other observations such as radar reflectivity. In addition, the small ensemble size employed during this project is likely underdispersive with respect to the spread in the ensemble BTs, and therefore the inclusion of additional ensemble members should be explored.

Acknowledgments. This project was supported by the NOAA Joint Technology Transfer Initiative (JTTI) via Grant NA17OAR4590179. The authors thank the anonymous reviewers for their contributions to this manuscript.

REFERENCES

Ackerman, S. A., 1996: Global satellite observations of negative brightness temperature differences between 11 and 6.7 μm. *J. Atmos. Sci.*, **53**, 2803–2812, https://doi.org/10.1175/1520-0469(1996)053<2803:GSOONB>2.0.CO;2.

——, W. L. Smith, J. D. Spinhirne, and H. E. Revercomb, 1990: The 27–28 October 1986 FIRE IFO cirrus case study: Spectral properties of cirrus clouds in the 8–12 μm window. *Mon. Wea. Rev.*, **118**, 2377–2388, https://doi.org/10.1175/1520-0493(1990)118<2377:TOFICC>2.0.CO;2.

Albrecht, B. A., 1989: Aerosols, cloud microphysics, and fractional cloudiness. *Science*, **245**, 1227–1230, https://doi.org/10.1126/science.245.4923.1227.

Ancell, B. C., A. Bogusz, M. J. Lauridsen, and C. J. Nauert, 2018: Seeding chaos: The dire consequences of numerical noise in NWP perturbation experiments. *Bull. Amer. Meteor. Soc.*, **99**, 615–628, https://doi.org/10.1175/BAMS-D-17-0129.1.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

Baum, B. A., P. F. Soulen, K. I. Strabala, M. D. King, S. A. Ackerman, W. P. Menzel, and P. Yang, 2000: Remote sensing of cloud properties using MODIS airborne simulator imagery during SUCCESS: 2. Cloud thermodynamic phase. *J. Geophys. Res.*, **105**, 11 781–11 792, https://doi.org/10.1029/1999JD901090.

Berner, J., G. Shutts, M. Leutbecher, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, https://doi.org/10.1175/2008JAS2677.1.

——, S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, https://doi.org/10.1175/2010MWR3595.1.

——, K. R. Fossell, S.-Y. Ha, J. P. Hacker, and C. Snyder, 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295–1320, https://doi.org/10.1175/MWR-D-14-00091.1.

——, and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1.

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.

Borbas, E. E., and B. C. Ruston, 2010: The RTTOV UWiremis IR land surface emissivity module. Version 1, NWP SAF Mission Rep. NWPSAF-MO-VS-042, 24 pp., https://nwpsaf.eu/vs_reports/nwpsaf-mo-vs-042.pdf.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, https://doi.org/10.1002/qj.234.

Bullock, R., 2011: Development and implementation of MODE time domain object-based verification. *24th Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 96, https://ams.confex.com/ams/91Annual/webprogram/Paper182677.html.

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.

Chai, T., and R. R. Draxler, 2014: Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, **7**, 1247–1250, https://doi.org/10.5194/gmd-7-1247-2014.

Christensen, H. M., I. M. Moroz, and T. N. Palmer, 2015: Simulating weather regimes: Impact of stochastic and perturbed parameter schemes in a simple atmospheric model. *Climate Dyn.*, **44**, 2195–2214, https://doi.org/10.1007/s00382-014-2239-9.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, and A. K. Heidinger, 2013: Evolution of severe and nonsevere convection inferred from GOES-derived cloud properties. *J. Appl. Meteor. Climatol.*, **52**, 2009–2023, https://doi.org/10.1175/JAMC-D-12-0330.1.

Cintineo, R., J. A. Otkin, F. Kong, and M. Xue, 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**, 107–124, https://doi.org/10.1175/MWR-D-13-00143.1.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, https://doi.org/10.1175/BAMS-D-11-00040.1.

——, and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1433–1448, https://doi.org/10.1175/BAMS-D-16-0309.1.

Connelly, R., 2018: Predictability of snow multi-bands in the cyclone comma head using a 40-member WRF ensemble. M.S. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, 153 pp.

Dai, A., K. E. Trenberth, and T. R. Karl, 1999: Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range. *J. Climate*, **12**, 2451–2473, https://doi.org/10.1175/1520-0442(1999)012<2451:EOCSMP>2.0.CO;2.

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, 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.

Developmental Testbed Center, 2014: Model Evaluation Tools Version 5.0 (METv5.0) User's Guide 5.0. Developmental Testbed Center, 241 pp., 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, https://doi.org/10.1175/2008WAF2222127.1.

Field, P. R., A. J. Heymsfield, A. G. Detwiler, and J. M. Wilkinson, 2019: Normalized hail particle size distribution from the T-28 storm-penetrating aircraft. *J. Appl. Meteor. Climatol.*, **58**, 231–245, https://doi.org/10.1175/JAMC-D-18-0118.1.

Flack, D. L. A., S. L. Gray, and R. S. Plant, 2019: A simple ensemble approach for more robust process-based sensitivity analysis of case studies in convection-permitting models. *Quart. J. Roy. Meteor. Soc.*, **145**, 3089–3101, https://doi.org/10.1002/qj.3606.

Gilleland, E., 2010: Confidence intervals for forecast verification. NCAR Tech. Note NCAR/TN-4791STR, 78 pp.

Gilmore, M. S., J. M. Straka, and E. N. Rasmussen, 2004: Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. *Mon. Wea. Rev.*, **132**, 2610–2627, https://doi.org/10.1175/MWR2810.1.

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, https://doi.org/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, https://doi.org/10.1080/01431160801891820.

——, D. T. Lindsey, K.-S. Sunny 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, https://doi.org/10.1175/MWR-D-14-00005.1.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

——, ——, Y.-J. Noh, C. O'Dell, T.-C. Wu, and F. Kong, 2018: Improvements to cloud-top brightness temperatures computed from the CRTM at 3.9 μm. *Mon. Wea. Rev.*, **146**, 3927–3944, https://doi.org/10.1175/MWR-D-17-0342.1.

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.

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., http://www.star.nesdis.noaa.gov/sod/sst/micros/pdf/CRTM_v1_NOAAtechReport-1.pdf.

Harrison, E. F., P. Minnis, B. R. Barkstrom, V. Ramanathan, R. D. Cess, and G. G. Gibson, 1990: Seasonal variation of cloud radiative forcing derived from the Earth Radiation Budget Experiment. *J. Geophys. Res.*, **95**, 18 687–18 703, https://doi.org/10.1029/JD095iD11p18687.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate mode, CCM3. *J. Geophys. Res.*, **105**, 14 873–14 890, https://doi.org/10.1029/2000JD900091.

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, https://doi.org/10.1175/2010JHM1282.1.

——, J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterizations in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, https://doi.org/10.1175/MWR-D-18-0092.1.

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, https://doi.org/10.1175/WAF-D-13-00054.1.

Karl, T. R., and Coauthors, 1993: A new perspective on recent global warming: Asymmetric trends of daily maximum and minimum temperature. *Bull. Amer. Meteor. Soc.*, **74**, 1007–1023, https://doi.org/10.1175/1520-0477(1993)074<1007:ANPORG>2.0.CO;2.

Knight, C. A., W. A. Cooper, D. W. Breed I. R. Paluch, P. L. Smith, and G. Vali, 1982: Microphysics. *Hailstorms of the Central High Plains, Vol 1: The National Hail Research Experiment*, C. A. Knight and P. Squires, Eds., Colorado Associated University Press, 151–193.

Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, https://doi.org/10.1002/qj.3094.

Martin, G. M., D. W. Johnson, and A. Spice, 1994: The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds. *J. Atmos. Sci.*, **51**, 1823–1842, https://doi.org/10.1175/1520-0469(1994)051<1823:TMAPOE>2.0.CO;2.

Mecikalski, J. R., and K. M. Bedka, 2006: Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery. *Mon. Wea. Rev.*, **134**, 49–78, https://doi.org/10.1175/MWR3062.1.

Miles, N. L., J. Verlinde, and E. E. Clothiaux, 2000: Cloud droplet size distributions in low-level stratiform clouds. *J. Atmos. Sci.*, **57**, 295–311, https://doi.org/10.1175/1520-0469(2000)057<0295:CDSDIL>2.0.CO;2.

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, https://doi.org/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, https://doi.org/10.1023/B:BOUN.0000020164.04146.98.

Ollinaho, P., and Coauthors, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quart. J. Roy. Meteor. Soc.*, **143**, 408–422, https://doi.org/10.1002/qj.2931.

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

——, ——, 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.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, https://doi.org/10.1002/QJ.49712757202.

——, 2019: Stochastic weather and climate models. *Nat. Rev. Phys.*, **1**, 463–471, https://doi.org/10.1038/s42254-019-0062-2.

Purdom, J. F. W., 1993: Satellite observations of tornadic thunderstorms. *The Tornado: Its Structure, Dynamics, Prediction, and Hazards, Geophys. Monogr.*, Vol. 79, Amer. Geophys. Union, 265–274.

Ramanathan, V., R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and D. Hartmann, 1989: Cloud-radiative forcing and climate: Results from the earth radiation budget experiment. *Science*, **243**, 57–63, https://doi.org/10.1126/science.243.4887.57.

Roeckner, E., and Coauthors, 2006: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. *J. Climate*, **19**, 3771–3791, https://doi.org/10.1175/JCLI3824.1.

Rutledge, S. A., and P. V. Hobbs, 1984: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. XII: A diagnostic modeling study of precipitation development in narrow cold-frontal rainbands. *J. Atmos. Sci.*, **41**, 2949–2972, https://doi.org/10.1175/1520-0469(1984)041<2949:TMAMSA>2.0.CO;2.

Sanchez, C., K. D. Williams, and M. Collins, 2016: Improved stochastic physics schemes for global weather and climate models. *Quart. J. Roy. Meteor. Soc.*, **142**, 147–159, https://doi.org/10.1002/qj.2640.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020

Schmetz, J., S. A. Tjemkes, M. Gube, and L. van de Berg, 1997: Monitoring deep convection and convective overshooting with METEOSAT. *Adv. Space Res.*, **19**, 433–441, https://doi.org/10.1016/S0273-1177(97)00051-3.

Sieglaff, J. M., L. M. Cronce, W. F. Feltz, K. M. Bedka, M. J. Pavolonis, and A. K. Heidinger, 2011: Nowcasting convective storm initiation using satellite-based box-averaged cloud-top cooling and cloud-type trends. *J. Appl. Meteor. Climatol.*, **50**, 110–126, https://doi.org/10.1175/2010JAMC2496.1.

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

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.

Strabala, K. I., S. A. Ackerman, and W. P. Menzel, 1994: Cloud properties inferred from 8–12-μm data. *J. Appl. Meteor.*, **33**, 212–229, https://doi.org/10.1175/1520-0450(1994)033<0212:CPIFD>2.0.CO;2.

Subramanian, A. C., and T. N. Palmer, 2017: Ensemble superparameterization versus stochastic parameterization: A comparison of model uncertainty representation in tropical weather prediction. *J. Adv. Model. Earth Syst.*, **9**, 1231–1250, https://doi.org/10.1002/2016MS000857.

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.

——, R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542, https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2.

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

——, 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, https://doi.org/10.1016/j.atmosres.2015.09.005.

Twomey, S., 1974: Pollution and the planetary albedo. *Atmos. Environ.*, **8**, 1251–1256, https://doi.org/10.1016/0004-6981(74)90004-3.

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.

Verlinde, J., P. J. Flatau, and W. R. Cotton, 1990: Analytical solutions to the collection growth equation: Comparison with approximate methods and application to cloud microphysics parameterization schemes. *J. Atmos. Sci.*, **47**, 2871–2880, https://doi.org/10.1175/1520-0469(1990)047<2871:ASTTCG>2.0.CO;2.

Watson, P. A. G., J. Berner, S. Corti, P. Davini, J. von Hardenberg, C. Sanchez, A. Weisheimer, and T. N. Palmer, 2017: The impact of stochastic physics on tropical rainfall variability in global climate models on daily to weekly time scales. *J. Geophys. Res. Atmos.*, **122**, 5738–5762, https://doi.org/10.1002/2016JD026386.

Welch, B. L., 1947: The generalisation of 'Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28–35, https://doi.org/10.1093/BIOMET/34.1-2.28.

White, B., E. Gryspeerdt, P. Stier, H. Morrison, G. Thompson, and Z. Kipling, 2017: Uncertainty from the choice of microphysics scheme in convection-permitting models significantly exceeds aerosol effects. *Atmos. Chem. Phys.*, **17**, 12 145–12 175, https://doi.org/10.5194/acp-17-12145-2017.

Willmot, C. J., and K. Matsuura, 2005: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.*, **30**, 79–82, https://doi.org/10.3354/cr030079.

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.

Xu, K.-M., 2006: Using the bootstrap method for a statistical significance test of differences between summary histograms. *Mon. Wea. Rev.*, **134**, 1442–1453, https://doi.org/10.1175/MWR3133.1.

Yang, P., L. Bi, B. A. Baum, K. N. Liou, G. W. Kattawar, M. Mishchenko, and and B. Cole,2013: Spectrally consistent scattering, absorption, and polarization properties of atmospheric ice crystals at wavelengths from 0.2 to 100 μm. *J. Atmos. Sci.*, **70**, 330–347, https://doi.org/10.1175/JAS-D-12-039.1.

Downloaded from http://journals.ametsoc.org/mwr/article-pdf/148/8/3111/4965255/mwrd200078.pdf by guest on 08 July 2020