Evaluating the Performance of Planetary Boundary Layer and Cloud Microphysical Parameterization Schemes in Convection-Permitting Ensemble Forecasts Using Synthetic *GOES-13* Satellite Observations

REBECCA CINTINEO AND JASON A. OTKIN

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

MING XUE AND FANYOU KONG

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

(Manuscript received 30 April 2013, in final form 29 August 2013)

ABSTRACT

In this study, the ability of several cloud microphysical and planetary boundary layer parameterization schemes to accurately simulate cloud characteristics within 4-km grid-spacing ensemble forecasts over the contiguous United States was evaluated through comparison of synthetic Geostationary Operational Environmental Satellite (GOES) infrared brightness temperatures with observations. Four double-moment microphysics schemes and five planetary boundary layer (PBL) schemes were evaluated. Large differences were found in the simulated cloud cover, especially in the upper troposphere, when using different microphysics schemes. Overall, the results revealed that the Milbrandt–Yau and Morrison microphysics schemes tended to produce too much upper-level cloud cover, whereas the Thompson and the Weather Research and Forecasting Model (WRF) double-moment 6-class (WDM6) microphysics schemes did not contain enough high clouds. Smaller differences occurred in the cloud fields when using different PBL schemes, with the greatest spread in the ensemble statistics occurring during and after daily peak heating hours. Results varied somewhat depending upon the verification method employed, which indicates the importance of using a suite of verification tools when evaluating high-resolution model performance. Finally, large differences between the various microphysics and PBL schemes indicate that large uncertainties remain in how these schemes represent subgrid-scale processes.

1. Introduction

Clouds strongly modulate the amount of solar radiation reaching the earth's surface and profoundly affect surface heating and atmospheric stability through changes in their horizontal extent and vertical structure (e.g., Xie et al. 2012). Cloud formation processes can also lead to explosive cyclogenesis (Sanders and Gyakum 1980) through latent heat release during condensation. Therefore, to accurately forecast the evolution of the atmosphere, including thunderstorm initiation and development, it is important for numerical weather prediction (NWP) models to realistically simulate cloud morphology.

Cloud microphysics and planetary boundary layer (PBL) parameterization schemes employed by NWP models strongly impact the structure and evolution of the simulated clouds and precipitation. Microphysics schemes typically account for changes in several liquid and frozen hydrometeor species and the complex interactions that occur between them, and vary greatly in sophistication and computational expense. Many existing schemes are single-moment schemes (i.e., only predict cloud and hydrometeor mass mixing ratios); however, increasing computational resources and the move toward higher spatial resolution have led to the development of more complex double-moment schemes that predict both the mass mixing ratio and total number concentration for each species. Double-moment schemes have been shown to improve simulated cloud characteristics (e.g., Milbrandt and Yau 2005a), though there is still much uncertainty in how to include various processes (e.g., drop breakup and ice-phase categories), and

considerable variability is seen among such schemes (e.g., Morrison and Milbrandt 2011). PBL schemes are used to parameterize the subgrid-scale vertical transfer of heat, moisture, and momentum between the surface and atmosphere due to turbulence that is too small to be explicitly resolved by the model, and thus impact the depth of the PBL and stability of the model atmosphere. This then influences cloud development, particularly during the daytime when radiative forcing is stronger, and the extent and evolution of the cloud cover in turn affects heat and moisture fluxes generated by the PBL scheme through changes in net radiation.

Prior work has evaluated the accuracy of simulated cloud fields in research and operational NWP models through comparisons of real and model-derived synthetic satellite observations (Morcrette 1991; Karlsson 1996; Rikus 1997; Tselioudis and Jakob 2002; Lopez et al. 2003; Sun and Rikus 2004; Otkin et al. 2009). The model-to-satellite approach has been used to validate and improve the accuracy of cloud microphysics schemes (Grasso and Greenwald 2004; Chaboureau and Pinty 2006; Otkin and Greenwald 2008; Grasso et al. 2010; Jankov et al. 2011). Synthetic satellite radiances derived from high-resolution NWP models have also been used as a proxy for future satellite sensors (Otkin et al. 2007; Grasso et al. 2008; Feltz et al. 2009) and have been shown to be a valuable forecast tool at convective scales (Bikos et al. 2012).

This study builds on previous work that has used satellite observations to compare the accuracy of different cloud microphysics and PBL schemes in the Weather Research and Forecasting Model (WRF). In high-resolution (4 km) simulations of a maritime extratropical cyclone over the North Atlantic, Otkin and Greenwald (2008) found that differing assumptions made by microphysics and PBL schemes have a substantial impact on the simulation of cloud properties. Overall, the double-moment microphysics schemes examined during their study produced a broader cirrus shield than the single-moment schemes and more closely matched the observations. It was also found that the Mellor–Yamada–Janjic (MYJ; Mellor and Yamada, 1982) PBL scheme gave more realistic results than the Yonsei University (YSU; Hong et al. 2006) scheme. A more recent study by Jankov et al. (2011) compared the performance of five microphysics schemes in WRF, including both single- and double-moment schemes, by evaluating the accuracy of synthetic 10.7-μm *Geostationary Operational Environmental Satellite-10* (*GOES-10*) imagery of an atmospheric river event over the western United States. They found that the relatively simple Purdue–Lin microphysics scheme (Chen and Sun 2002) was the least accurate scheme with similar results for the other schemes.

TABLE 1. List of ensemble members.

Member	Microphysics scheme	PBL scheme
M–Y	Millbrant–Yau	MYJ
MORR	Morrison	MYJ
WDM6	WDM6	MYJ
THOM or MYJ	Thompson	MYJ
MYNN	Thompson	MYNN
ACM2	Thompson	ACM2
QNSE	Thompson	QNSE
YSU	Thompson	YSU

In this study, synthetic infrared brightness temperatures are used to evaluate the performance of several microphysics and PBL schemes employed by the convection-permitting (4 km) WRF ensemble run by the Center for the Analysis and Prediction of Storms (CAPS) during the 2012 National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) Spring Experiment (Kong et al. 2012a). Unlike previous studies, all of the microphysics schemes evaluated here are double-moment for at least one of the cloud species, more than one event is investigated, and a larger domain encompassing the contiguous United States is used. The datasets used in this study are described in section 2, the verification results are presented in section 3, and conclusions are discussed in section 4.

2. Data and methodology

a. CAPS real-time ensemble forecasts

As part of the NOAA HWT Spring Experiment (Clark et al. 2012), CAPS at the University of Oklahoma has been using national supercomputing resources to produce high-resolution ensemble forecasts in near–real time over the contiguous United States (CONUS) since 2007 (Xue et al. 2007; Kong et al. 2007). The ensemble configuration and domain size have varied as computational resources have expanded and postexperiment analysis has investigated the performance of the ensemble. For the 2012 experiment, the CAPS ensemble contained 28 members employing different dynamical cores, parameterization schemes, and initialization techniques (Kong et al. 2012b). Within this ensemble, eight of the members contained identical grid and initial/boundary condition configurations, but employed different microphysics and PBL parameterization schemes (Table 1). In this study, we will use synthetic satellite observations to examine the accuracy of the simulated cloud fields for these members to investigate WRF sensitivity to different physics options.

Version 3.3.1 of WRF (Skamarock et al. 2008) was used during the 2012 Spring Experiment. The eight
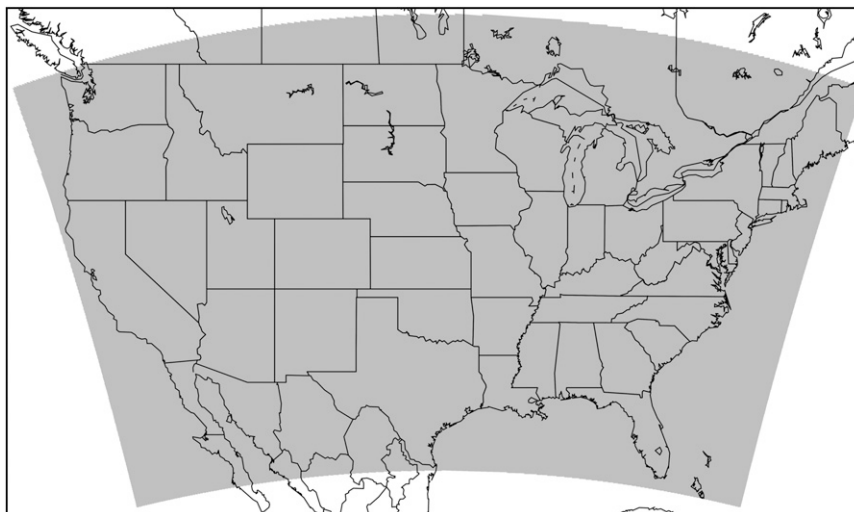
FIG. 1. The 2012 HWT Spring Experiment CAPS ARW-WRF domain is shaded in gray.

CAPS ensemble members examined here cover the entire CONUS with 4-km horizontal grid spacing (Fig. 1) and 51 vertical levels, and were initialized at 0000 UTC each day using enhanced analyses based on the 12-km North American Mesoscale Model (NAM) analysis background. Radar data and surface observations were assimilated into the initial conditions using the Advanced Regional Prediction System (ARPS) 3D variational data assimilation and cloud analysis system (Xue et al. 2003; Gao et al. 2004; Hu et al. 2006). All ensemble members used the Rapid Radiative Transfer Model (RRTM) longwave radiation (Mlawer et al. 1997), Goddard shortwave radiation (Chou and Suarez 1994), and the Noah land surface model to parameterize land surface processes. No cumulus parameterization scheme was used. The forecasts were run for 36 h starting at 0000 UTC, with data output once per hour. To allow sufficient time for model spinup, the first 6 h of each forecast were excluded from the analysis. Data from 20 forecasts during a 4-week period in May–June 2012 were used for this study (Table 2). Forecasts were not available for all days during the 4-week period since the Spring Experiment forecasts are run only on weekdays or whenever severe weather is expected to occur during the weekend. This 4-week period contained several episodes of severe weather; however, thunderstorm activity was below normal across much of the central United States.

b. Microphysics ensemble members

To investigate the impact of the microphysics schemes on the cloud field, four ensemble members employing different microphysics schemes were examined. These members were otherwise identical, including use of the MYJ PBL scheme. All of the evaluated schemes are double-moment for at least one cloud species. The WRF double-moment 6-class (WDM6) microphysics scheme, which is based on the WRF single-moment 6-class scheme (Hong et al. 2004; Hong and Lim 2006), is double-moment only for the warm rain processes, with the cloud condensation nuclei and cloud and rainwater number concentrations predicted (Lim and Hong 2010). The Thompson microphysics scheme includes prognostic equations for cloud water, cloud ice, snow, rain, and graupel mass mixing ratios. To maintain computational efficiency while increasing the accuracy of the scheme, only the cloud ice and rainwater species are double moment (Thompson et al. 2008). All other species are single moment. The Morrison microphysics scheme (Morrison et al. 2009) also predicts mass mixing ratios for cloud droplets, cloud ice, snow, rain, and graupel, with number concentrations predicted for all species except for cloud water. Last, the Milbrandt–Yau scheme is double moment for all species, with separate classes for graupel and hail (Milbrandt and Yau 2005a,b). In summary, the microphysics schemes evaluated herein differ in how many and which species are treated using two moments, and differ in their treatments of various cloud processes. These members will hereafter be referred to as the WDM6, THOM, MORR, and M–Y members.

TABLE 2. Dates from 2012 included in the study.

15 May	22 May	28 May	4 Jun
16 May	23 May	29 May	5 Jun
17 May	24 May	30 May	6 Jun
18 May	25 May	31 May	7 Jun
21 May	27 May	1 Jun	8 Jun

c. *PBL ensemble members*

For the PBL comparisons, five ensemble members are used, including the THOM member described in the previous section. Each of these members employed different PBL schemes, but were otherwise identical, including the use of the THOM microphysics scheme. The PBL schemes can be divided into local schemes, where turbulent fluxes are calculated at each grid point in the vertical column using only data from adjacent levels, and nonlocal closure schemes, which calculate fluxes based on variables throughout the depth of the model.

First, we describe the local schemes. The MYJ scheme is based on the Mellor–Yamada turbulence model that includes prognostic equations for turbulent kinetic energy (TKE; Mellor and Yamada 1982). The quasi-normal scale elimination (QNSE) scheme (Sukoriansky et al. 2005, 2006) is similar to the MYJ scheme except that it determines the mixing length differently for a stable PBL to better handle stable stratification. The Mellor–Yamada–Nakanishi–Niino (MYNN) scheme (Nakanishi 2001; Nakanishi and Niino 2004, 2006, 2009) is also a local closure scheme based on the Mellor–Yamada turbulence model. To overcome the underestimation of boundary layer depth and TKE and other problems associated with Mellor–Yamada models, the MYNN includes improvements to the closure constants and mixing length scale based on large-eddy simulation data.

Two nonlocal PBL schemes are also evaluated. The YSU scheme allows nonlocal mixing with explicit entrainment processes at the top of the PBL (Hong et al. 2006; Hong 2010). The asymmetric convective model, version2 (ACM2), adds a local mixing component to a nonlocal transport in an attempt to realistically capture both the subgrid and supergrid-scale fluxes, yet still be computationally efficient (Pleim 2007a,b). If conditions are stable or neutral, such as typically occurs at night, only the local closure is used in the ACM2. These members will hereafter be referred to as the MYJ, QNSE, MYNN, YSU, and ACM2 members. Note that the MYJ and THOM acronyms refer to the same ensemble member and will hereafter be used interchangeably. Different descriptors are used for this member because, as configured, it is included in both the microphysics and PBL groups.

d. *Synthetic satellite data and forward model description*

The accuracy of each ensemble member is evaluated through comparison of real and synthetic *GOES-13* infrared brightness temperatures. The Successive Order of Interaction (SOI) forward radiative transfer model (Heidinger et al. 2006; O'Dell et al. 2006) was used to compute the synthetic satellite observations. Gas optical depths are computed for each model layer using CompactOPTRAN code included in the Community Radiative Transfer Model (Han et al. 2006). Scattering and absorption properties, including single scatter albedo, extinction efficiency, and full scattering phase function, for each frozen species (i.e., ice, snow, and graupel) are obtained from Baum et al. (2005). A lookup table based on Lorenz–Mie theory is used for the cloud water and rainwater species. Infrared cloud optical depths are computed for each species by scaling visible optical depths by the ratio of the extinction efficiencies. No aerosol particles were assumed in the calculation of the synthetic brightness temperatures. The surface emissivity over land is obtained from the Seemann et al. (2008) dataset. WRF fields used by the forward model include water vapor mass mixing ratio, atmospheric and surface skin temperatures, 10-m wind speed, and the mass mixing ratios and number concentrations for each cloud species predicted by a given microphysics scheme. When only the mass mixing ratio is predicted for a specific species, the corresponding number concentration is diagnosed from the assumed fixed intercept parameter value (e.g., Otkin et al. 2007). The SOI code accounts for the assumptions made by the microphysics schemes, including each species' particle density, slope intercept parameter, and particle size distribution (e.g., Marshall–Palmer, generalized gamma). Previous studies have shown that the SOI model produces accurate brightness temperatures in both clear and cloudy conditions (e.g., Otkin and Greenwald 2008; Otkin et al. 2009; Bikos et al. 2012) and the accuracy is expected to be within 1 K for infrared brightness temperatures (Otkin 2010).

Synthetic *GOES-13* brightness temperatures are computed on the WRF grid each forecast hour for two infrared bands, and subsequently remapped to the *GOES-13* projection to allow for direct comparison of the real and synthetic data. The spectral bands that are analyzed include the 6.7-μm band sensitive to water vapor in the mid- and upper troposphere and the 10.7-μm infrared window band, which is sensitive to either cloud top properties or to the surface depending upon whether or not clouds are present.

3. Results

a. *Example from 7 June 2012*

Representative examples of observed and simulated 6.7- and 10.7-μm brightness temperatures for 0400 UTC 7 June 2012 (forecast hour 34) are shown in Figs. 2 and 3. At this time, thunderstorms had developed across portions of the central United States in response to a trough
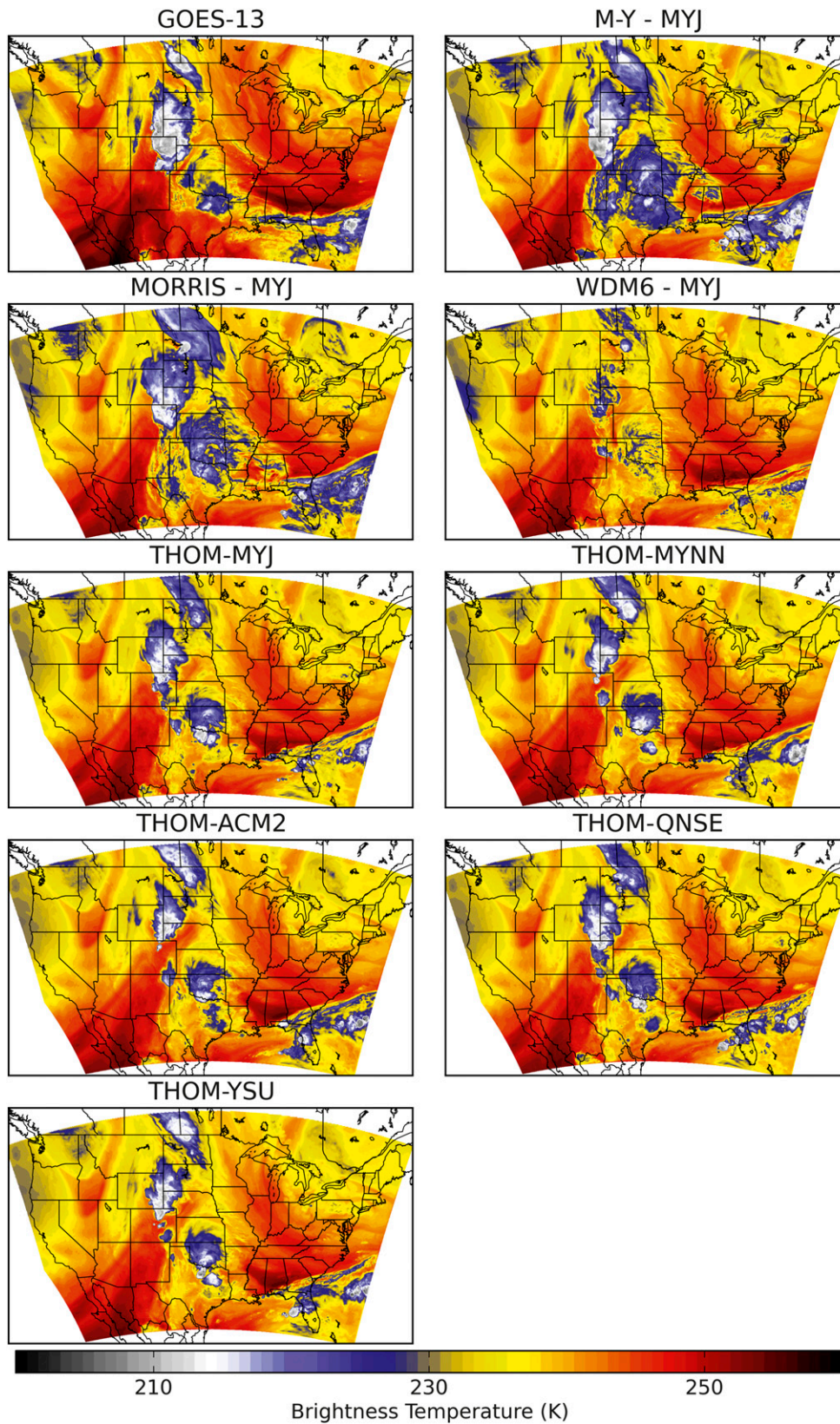
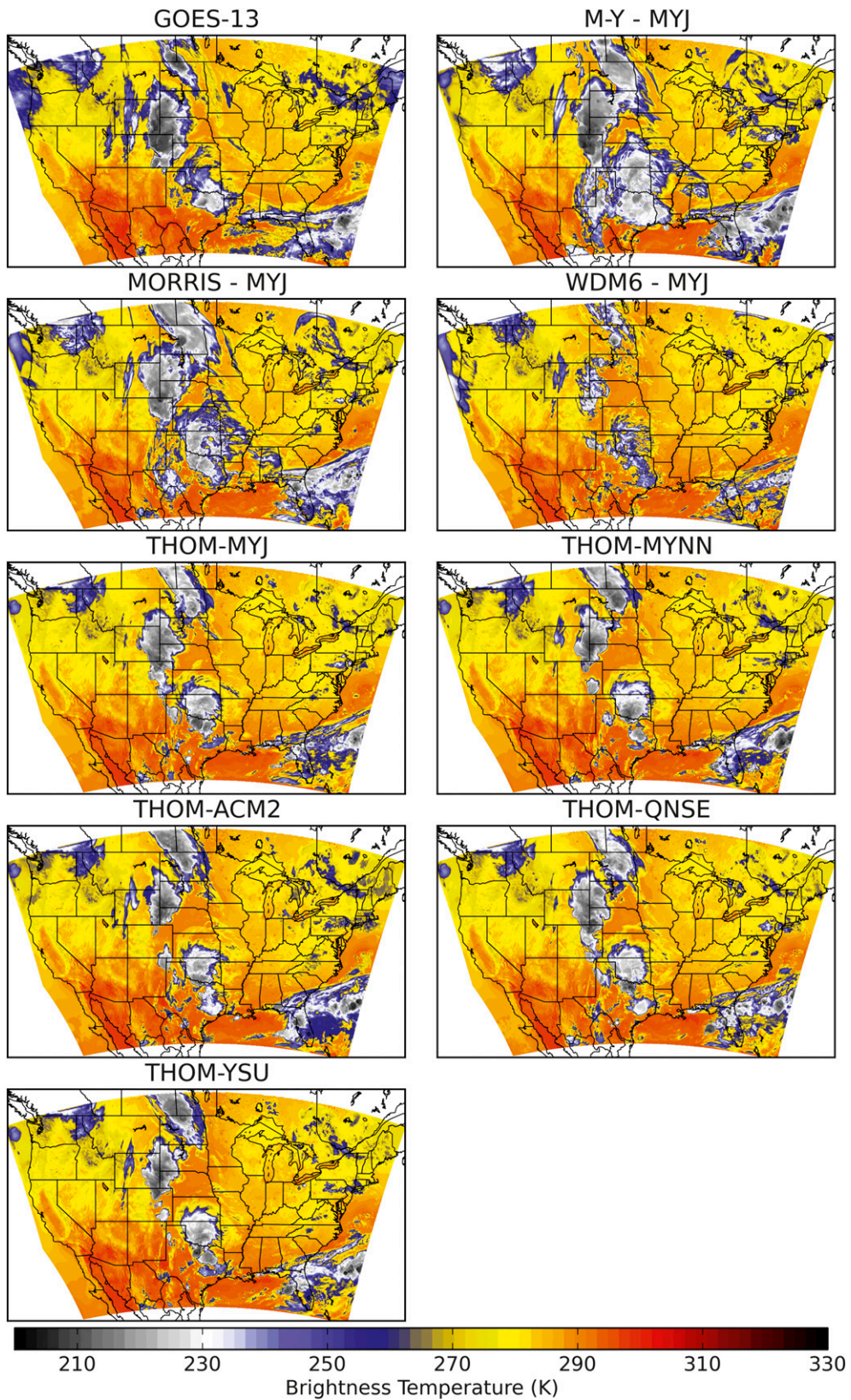FIG. 2. Observed and simulated 6.7-μm brightness temperatures for 0400 UTC 7 Jun 2012.

FIG. 3. Observed and simulated 10.7-μm brightness temperatures for 0400 UTC 7 Jun 2012. Note that the color scale is different from Fig. 2.

moving eastward from the Rocky Mountains. According to the Storm Prediction Center, several tornadoes were reported in eastern Colorado and Wyoming, with severe wind and hail reports scattered across the central United States from the Texas Gulf Coast to Minnesota. Overall, the model forecasts contain clouds in the correct general location and with a similar structure to the observations, though all of the members produced erroneous cloud cover over west Texas. However, there are variations between the ensemble members as to the extent and temperature of the clouds. For instance, the M–Y and MORR members both contain more extensive cloud cover than observed, while the WDM6 underforecasts the spatial extent of the cloud field, especially for colder clouds. The variations between the PBL schemes are not as readily apparent as those between the microphysics schemes. These results were similar to what was generally seen when examining the brightness temperatures during other times of active deep convection.

b. Frequency distributions

To examine the temporal evolution of each GOES channel during the forecast period, averaged over all 20 days during the HWT, normalized brightness temperature histograms were computed for each ensemble member as a function of time using all grid points in the model domain. The observed frequency distributions were subtracted from the forecast distributions to better highlight differences in the ensemble members.

Figure 4 shows the probability distributions for the 6.7-μm water vapor band. Overall, large systematic errors are present in all of the ensemble members. For instance, there are too many grid points with brightness temperatures between 230 and 240 K, and too few with warmer brightness temperatures between 250 and 260 K. The combined effect of these biases suggests that there is a moist bias in the mid- to upper troposphere that is unrelated to the microphysics or PBL schemes. Indeed, Coniglio et al. (2013) also found that there is a moist bias in the lower and midtroposphere in the WRF initial conditions, which indicates that this initial moist bias persists during the forecast period. Comparison of the microphysics schemes shows that the M–Y and MORR members contain much smaller errors between 230 and 240 K; however, the broader distribution of positive biases between 210 and 240 K indicates that these schemes are producing too many clouds in the upper troposphere. The WDM6 and THOM schemes, however, underpredict the number of pixels colder than 230 K, which indicates that these schemes did not produce enough upper-level cloud cover. Relatively small differences are apparent in the water vapor band distributions when comparing the

PBL schemes. The most notable difference is that the MYJ and QNSE schemes produce too many cold pixels during the late evening between 2000 and 0300 UTC. The ACM2 and YSU schemes also produce too many cold pixels during that time, but to a much lesser degree. This bias of the MYJ and QNSE schemes may indicate that they are increasing convective initiation, which would lead to too much convection and the development of excessive upper-level cloudiness during and after peak daytime heating. This is supported by the presence of a warm and moist bias below 0.75 AGL for each of these schemes in the vicinity of deep convection (Coniglio et al. 2013).

For the 10.7-μm window band (Fig. 5), the largest differences in the PBL schemes occur for the warmest pixels (>290 K), which indicates that there are large differences in surface temperatures and low-level cloudiness. Most of the ensemble members do not contain enough grid points with brightness temperatures >310 K between forecast hours 15 and 26, meaning that the model surface is either too cool or contains too many clouds in certain regions, such as the southwestern United States (not shown). The exception is the MYNN PBL scheme, which actually produces slightly too many of the warmest brightness temperatures, on average. The MYJ and QNSE schemes have the largest daytime cold bias, which is consistent with results from prior studies (e.g., Hu et al. 2010; Xie et al. 2012). Overall, the MYNN scheme appears to best handle surface heating during the day. Coniglio et al. (2013) also found that the MYNN scheme produced the smallest biases for PBL depth, moisture, and potential temperature.

Very large differences are also evident in the microphysics members, particularly for brightness temperatures <300 K. The M–Y and MORR microphysics schemes again contain too many cold pixels throughout the forecast cycle, but especially during the peak convective hours after 2000 UTC, whereas the WDM6 scheme does not contain enough. Because the differences for colder brightness temperatures are solely due to changes in cloud cover, these results indicate that the M–Y and MORR schemes produce too many upper-level clouds, while the WDM6 does not produce enough. Errors in the 280–300-K range are notably smaller when the M–Y and MORR schemes are used, but the improved accuracy is misleading because increases in upper-level cloud cover for these members shades some of the grid points where these errors occur, thus artificially improving their performance. Of the microphysics schemes, the THOM scheme generally produces the best distribution of brightness temperatures <240 K, though some small errors remain.
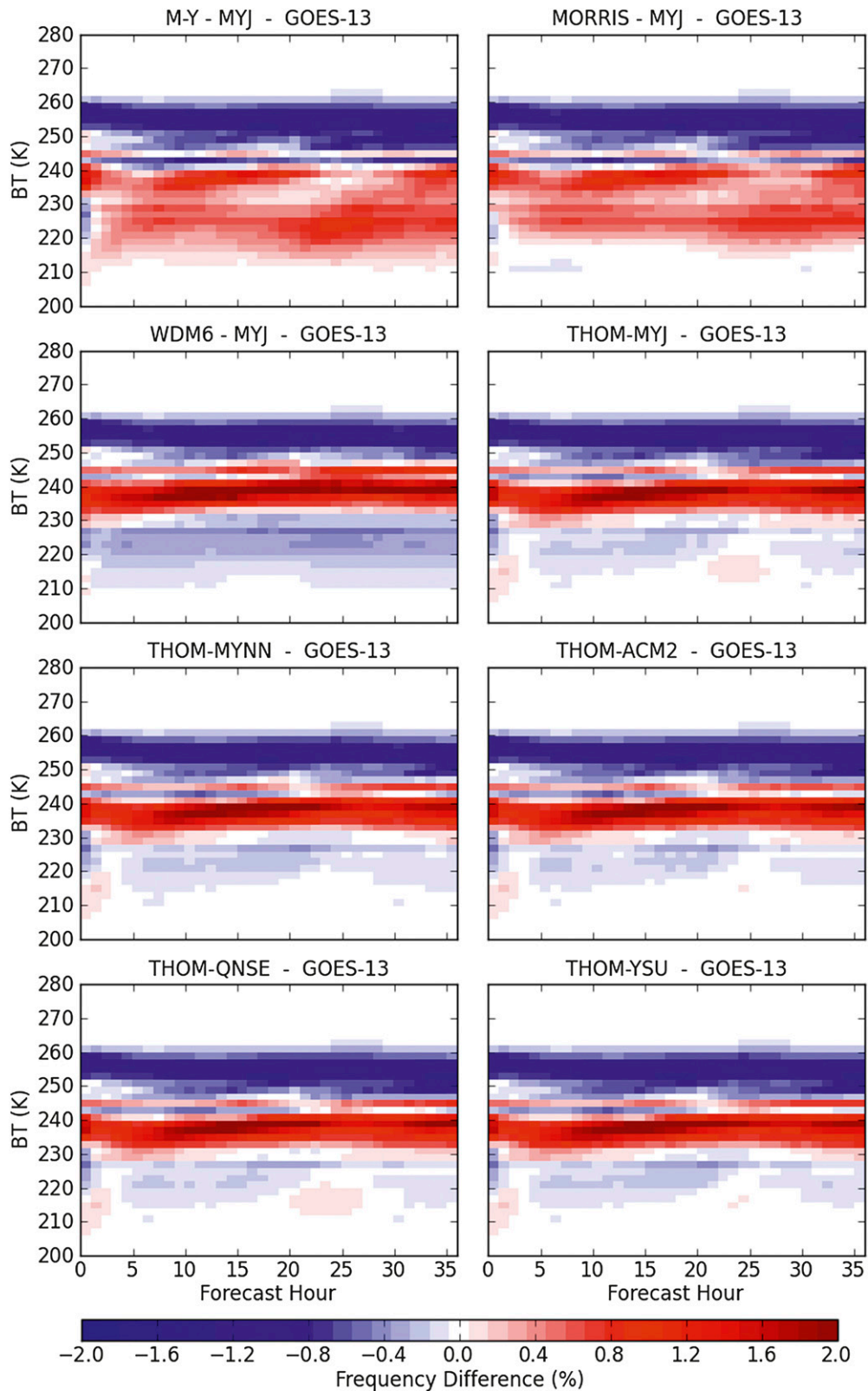
FIG. 4. 6.7-μm band brightness temperature (BT) frequency distribution differences between forecasts and observations during the forecast period averaged over all 20 days.
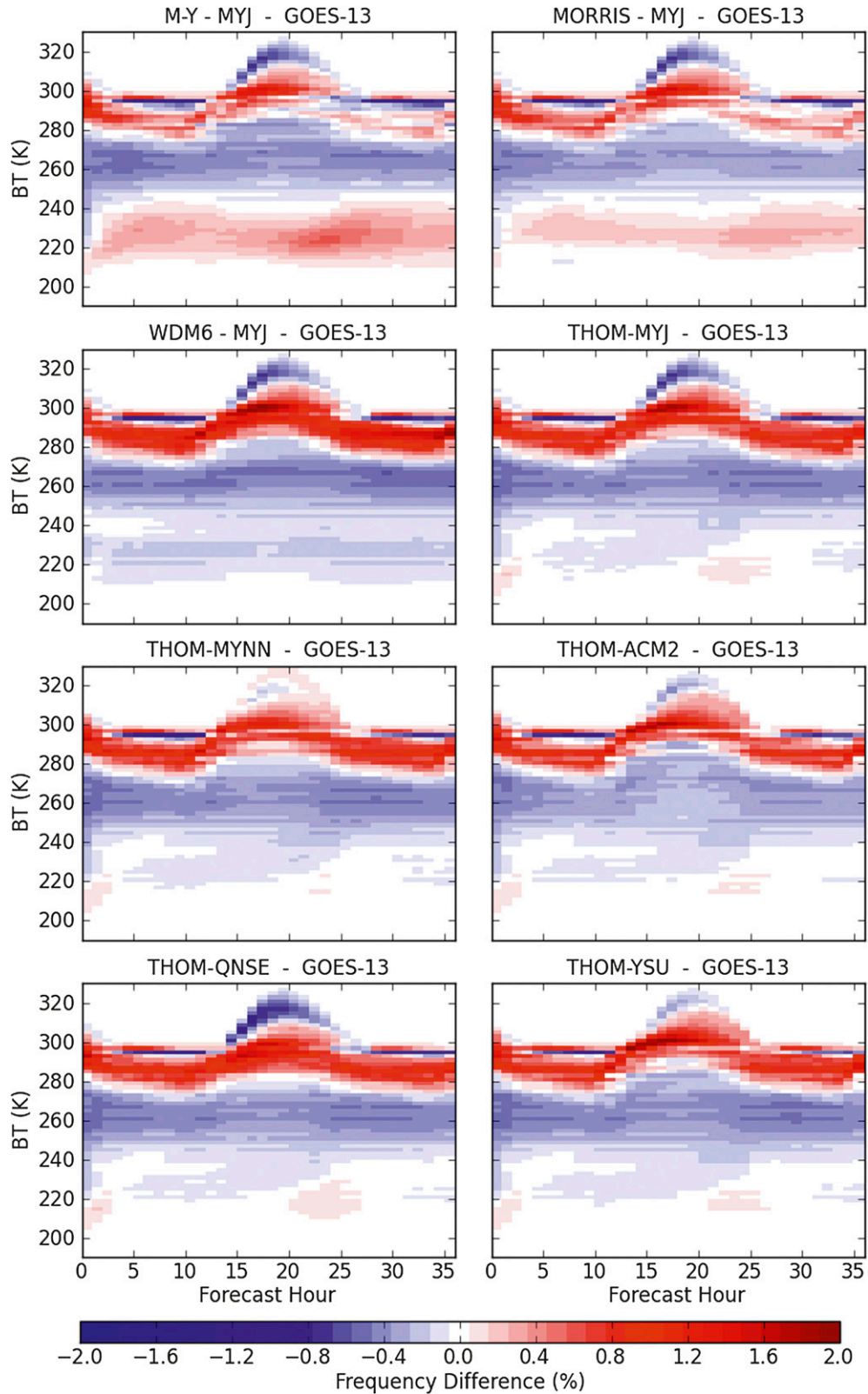
FIG. 5. 10.7-μm band brightness temperature (BT) frequency distribution differences between forecasts and observations during the forecast period averaged over all 20 days.
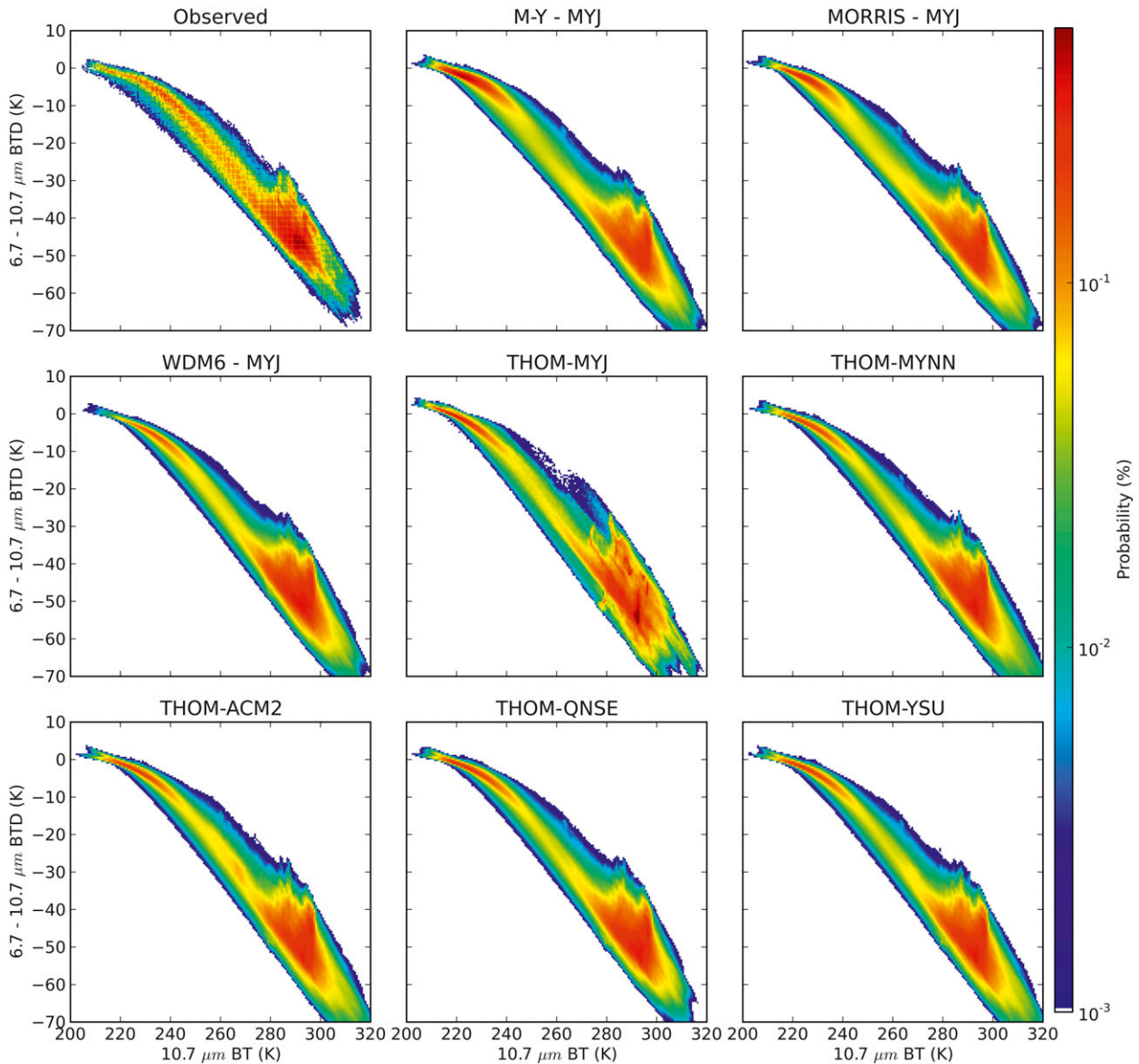
FIG. 6. Two-dimensional histogram of 6.7–10.7-μm brightness temperature differences (BTD) vs 10.7-μm brightness temperatures (BT) between 2100 and 0300 UTC over all 20 days.

c. 6.7–10.7-μm brightness temperature differences

Brightness temperature differences (BTD) between 6.7 and 10.7 μm can be used to examine how well each microphysics and PBL scheme forecast cloud height during the forecast period. Because of strong water vapor absorption at 6.7 μm and a general decrease in temperature with height in the troposphere, 6.7–10.7-μm BTDs are usually negative, with the largest differences occurring in clear sky regions with high surface temperatures (Mecikalski and Bedka 2006). Figure 6 shows two-dimensional histograms of 6.7–10.7-μm BTD versus

10.7-μm brightness temperature. Overall, the histogram shapes for each ensemble member match up fairly well with observations, though there are large differences in magnitude in some parts of the distributions. The microphysics schemes have greater variation between histograms than the PBL schemes, with the M–Y scheme generating far too many cold clouds with BTDs > -10 K and the WDM6 producing too few, as was seen in Figs. 4 and 5. BTD values near or above 0 K match up with 10.7-μm brightness temperatures less than 220 K, which are the highest clouds and can indicate overshooting tops in thunderstorms. This result agrees with other

studies that have also found that a 6.7–10.7-μm BTD exceeding 0 K indicates clouds at or above the tropopause (Ackerman 1996; Schmetz et al. 1997; Mecikalski and Bedka 2006). Last, the local BTD maximum between −55 and −35 K is too low in all of the ensemble members due to excessive upper-level cloud cover. The relative lack of upper-level clouds in the WDM6 scheme allows it to perform well in this part of the distribution.

Next, various 6.7–10.7-μm BTD thresholds are used to examine the areal extent of clouds in the lower, mid-, and upper troposphere. Similar to Mecikalski and Bedka (2006), the thresholds used here are −30 to −10 K for low- to midlevel cloud tops (about 850–500 hPa), −10 to 0 K for upper-level cloud tops, and >−2 K for overshooting tops, which is relaxed from 0 K to account for the coarser effective spatial resolution of the simulated data compared to observations (Skamarock 2004). In addition, the relatively coarse vertical resolution (600 m) near the cloud top makes it more difficult for WRF to properly simulate overshooting cloud tops. The spatial area encompassed by each threshold is determined and the ratio of the forecast-to-observed area for each ensemble member is plotted as a function of forecast time in Fig. 7. The area ratios are computed using data from the whole domain averaged over all forecasts; therefore, spatial displacement errors are not considered. Ratios greater (less) than one indicate that a given ensemble member contains more (fewer) clouds than observed.

The aerial extent of low- to midlevel clouds (Fig. 7a) is consistently too low during the entire forecast period for all of the ensemble members. Comparison of the microphysics schemes reveals that the MORR area is closest to the observed, followed by the M–Y scheme. For the PBL schemes, the ACM2 is much more accurate than the other schemes, particularly during the afternoon and evening. A similar lack of lower-level clouds was also found by Jankov et al. (2011) when NAM analyses were used for initial and lateral boundary conditions. Despite their deficiency, the simulated low- to midlevel cloud areas display a similar temporal evolution to what was observed (Fig. 8a), with the smallest spatial area occurring in the morning before cloud cover increases during the day and reaches a maximum during the late evening hours.

For upper-level clouds (Fig. 7b), their aerial extent is best captured by the THOM microphysics scheme, which remains fairly close to the 1:1 ratio throughout the forecast period. The M–Y scheme, however, greatly overestimates their extent by nearly a factor of 2, while the MORR scheme also overforecasts their extent by nearly 50%. This behavior stands in sharp contrast to the WDM6 scheme, which underpredicts their extent by

about 40% on average. When both low- to mid- and upper-level cloud cover is considered, it is apparent that the M–Y (WDM6) scheme generates the most (fewest) clouds throughout the depth of the troposphere. It is noteworthy that even though the M–Y and MORR schemes overestimate the upper-level cloud cover extent, they still contained the most low level clouds (Fig. 7a). Without the presumed shading effect of the excessive upper-level clouds that occurs with this top-down observing method, more of the low-level clouds could be seen. This suggests that their low-level cloudiness would have been closer to what is observed had there not been as much upper-level cloud cover.

Similar ratios occurred for all of the PBL schemes, with the largest differences occurring during the evening hours. The relatively small sensitivity compared to the microphysics schemes indicates that the PBL schemes have a much smaller impact on upper-level cloud cover. Though the ratios are close to 1 for all of the PBL schemes, this is primarily because they are paired to the THOM scheme. Their performance would have likely been different if a different microphysics scheme had been used. All of the schemes underforecast the upper-level cloud cover during the morning; however, by late evening, most schemes produced too many clouds, before settling near or slightly below the 1:1 ratio during the second overnight period (forecast hours 30–36). This suggests that the erroneously low cloud cover extent during the first 18 h of the forecast may be partially due to spinup issues. The forecast upper-level cloud area follows the temporal evolution of the observed upper-level cloud area fairly well, but with a slight delay between the peak of the observations and the peak of most of the forecasts (Fig. 8b). Only the WDM6 member peaks prior to what was observed.

Last, we examine clouds with a BTD exceeding −2 K (Fig. 7c) that are near or above the tropopause, and are often associated with deep convection. Among the microphysics schemes, the best forecast occurred when the THOM scheme was used, though the MORR scheme also performed well. The M–Y scheme, however, substantially overpredicts the cloud amount at all times. At the peak convective time of the day, all of the ensemble members overforecast the spatial area of the highest clouds except for the WDM6 microphysics. The vastly different area ratios indicate that the vertical transport of cloud condensate is much more vigorous in the M–Y scheme than it is in the WDM6 scheme. These differences could also be due to faster fall speeds and larger particles in the WDM6 scheme, causing upper-level clouds to dissipate more rapidly. The increase in area ratios to a maximum around 2000 UTC is due in part to the model developing the highest clouds too quickly. The forecast
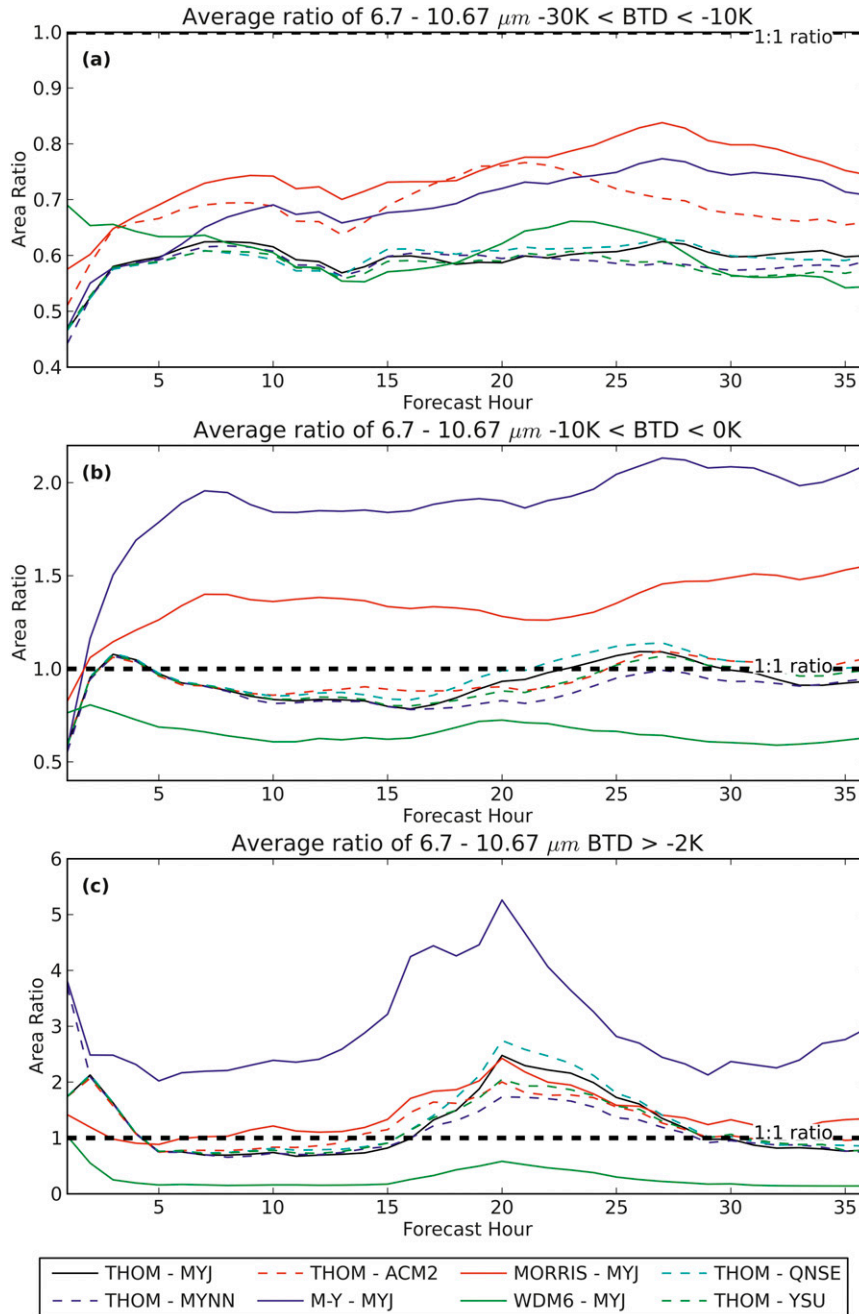
FIG. 7. Ratio of forecast to observed 6.7–10.7-μm brightness temperature difference area during the forecast period within a threshold of (a) −30 to −10 K, (b) −10 to 0 K, and (c) above −2 K.

areas from all members increase too soon compared with observations and reach their greatest extents about 3–5 h prior to the peak of the observed area (Fig. 8c).

d. Traditional objective verification scores

Traditional point-to-point verification measures, including bias, root-mean-square error (RMSE), and mean absolute error (MAE), were also computed for each ensemble member for the 6.7- and 10.7-μm bands. The domain-average bias for each band (Figs. 9a and 10a) is consistent with the earlier analysis and with a visual inspection of the brightness temperature datasets. In the 6.7-μm water vapor band, all of the ensemble members have a cold bias, consistent with the tendency for too
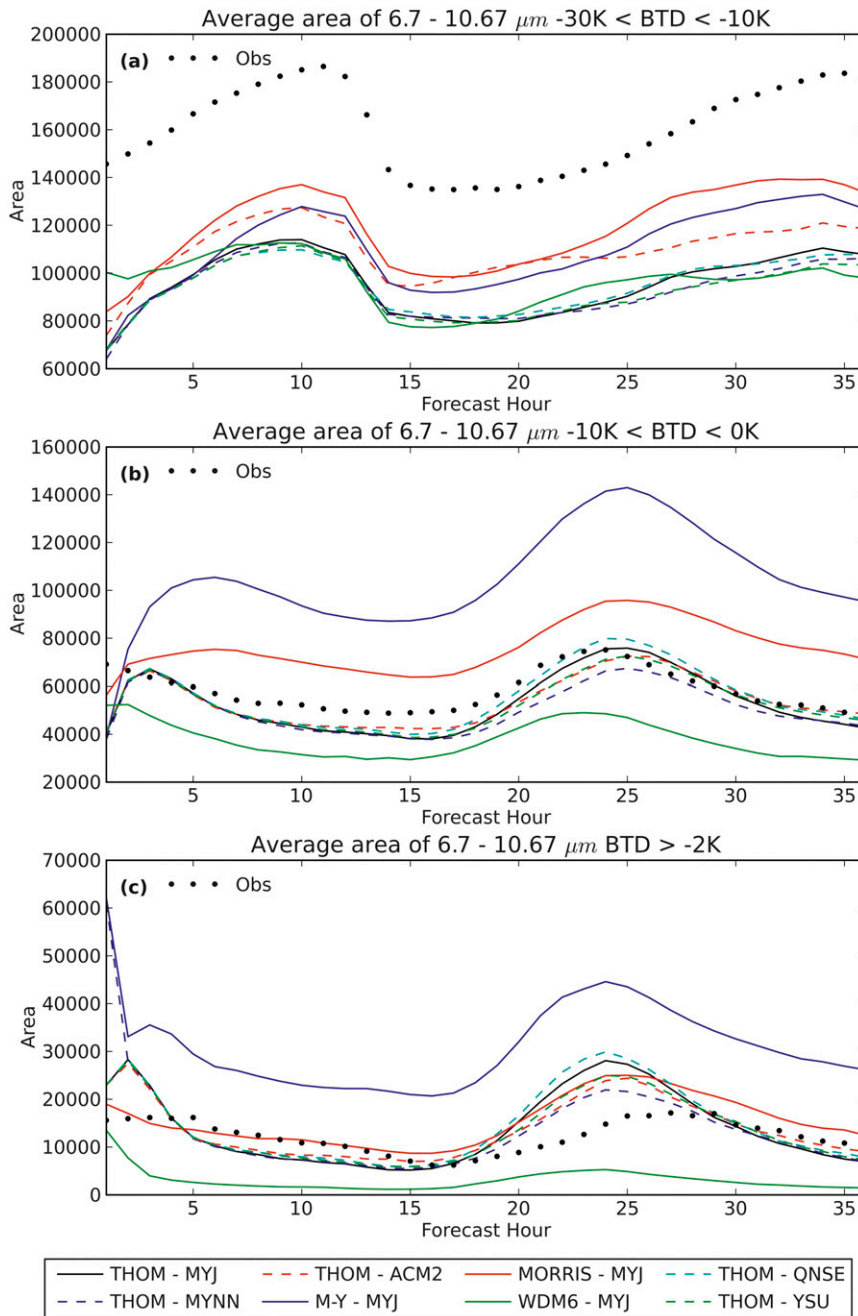
FIG. 8. Forecast and observed 6.7–10.7-μm brightness temperature difference areas during the forecast period within a threshold of (a) −30 to −10 K, (b) −10 to 0 K, and (c) above −2 K.

much water vapor in the mid- and upper levels shown in section 3a. The M–Y scheme has the largest cold bias, followed by the MORR scheme. The WDM6 scheme is characterized by the smallest cold bias due to its smaller areal extent of upper-level clouds. Similar results were found for the 10.7-μm band (Fig. 10a). The differences in bias between the PBL schemes are not as great, though the MYNN has the smallest cold bias in the 6.7-μm band

during the convective hours, while the QNSE has the smallest bias for the 10.7-μm band.

The RMSE and MAE (Figs. 9b,c and 10b,c) demonstrate that the M–Y and MORR microphysics schemes have the largest errors at all times for both bands, while the WDM6 and THOM schemes have the smallest errors. These results indicate that the excessive cloud cover in the M–Y and MORR schemes leads to greater
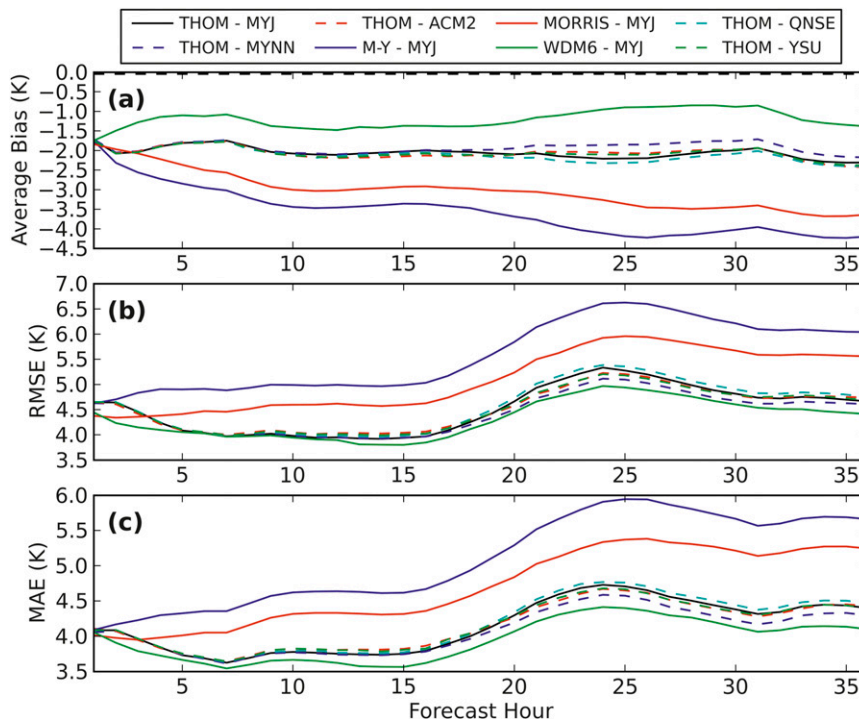
FIG. 9. 6.7-μm band (a) domain average bias, (b) RMSE, and (c) MAE during the forecast period, averaged over all 20 days. The black dotted line in (a) indicates no bias.

errors than the underforecast in cloud cover by the WDM6 scheme. The differences between the PBL schemes are not as dramatic as for the microphysics comparison. The ACM2 has the smallest RMSE and MAE for the 10.7-μm band, while the MYNN has the smallest errors for the 6.7-μm band. Out of the PBL schemes, the QNSE generally has the largest errors for the 10.7-μm band during the peak convective hours in the late afternoon and evening.

Overall, these results give a different picture than the other verification methods. Traditional point-to-point verification statistics, such as the RMSE and MAE, however, may not be optimal when examining predictions of mesoscale features by higher-resolution models (e.g., Mass et al. 2002; Davis et al. 2006; Roberts and Lean 2008). For instance, the statistics are penalized by double-error counting when small features are temporally or spatially displaced by a small amount, even though the structure of the forecast feature itself is more realistic. The better verification statistics for the WDM6 are likely because the underforecast of cloud cover leads to less double-error counting than the schemes that produce more clouds. Therefore, using point-to-point verification statistics can give misleading results and should be used along with other objective methods when examining the accuracy of high-resolution models, particularly when clouds are involved.

e. Fraction skill score

The fraction skill score (FSS; Roberts and Lean 2008; Roberts 2008) is a neighborhood-based verification method that gives the skill of a forecast at different spatial scales, taking into account the area around each point and thereby reducing the large errors from small spatial displacements associated with traditional point-to-point statistics. The fractional areas above a threshold within a neighborhood about each point in a forecast are compared with the fractional areas above the same threshold within the neighborhoods surrounding the corresponding points in the observation. The mean square error (MSE) of the fractional areas is computed and then used to calculate the FSS:

$$FSS = 1 - \left(\frac{MSE}{MSE_{ref}} \right), \qquad (1)$$

where MSE_{ref} represents the largest MSE that could be calculated from the given fractions. The FSS values range from 0 to 1, with a value of 1 corresponding to a perfect forecast.

The FSS is computed for each ensemble member using progressively larger neighborhood sizes from 1 to 20 grid points, centered on each grid point. The 10.7-μm band is examined in the remainder of this section, with
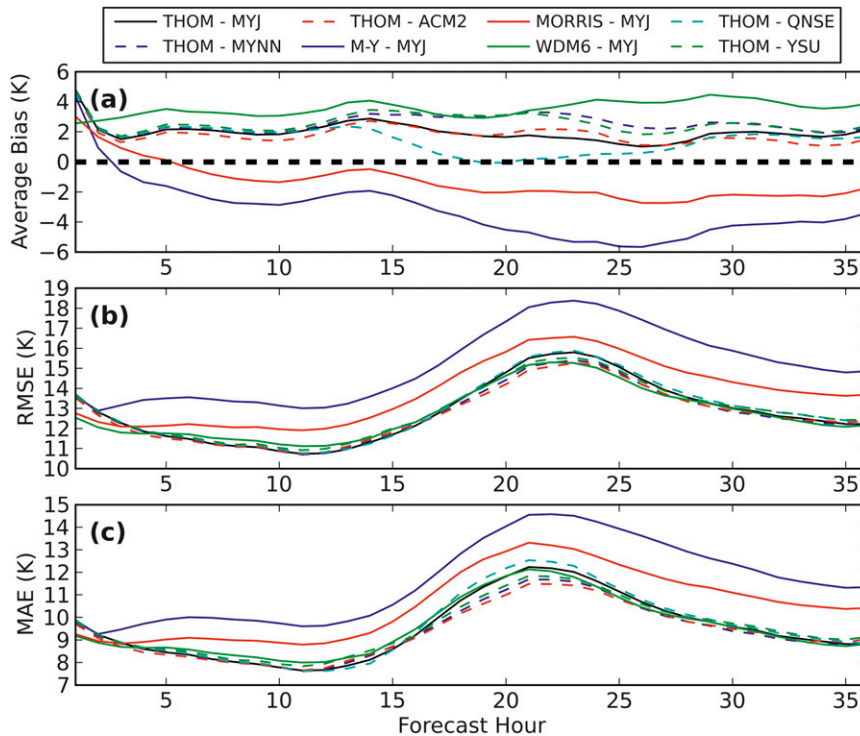
FIG. 10. 10.7-μm band (a) domain average bias, (b) RMSE, and (c) MAE during the forecast
period, averaged over all 20 days. The black dotted line in (a) indicates no bias.

a threshold of 270 K used to delineate all clouds, in-
cluding low-level clouds, and a threshold of 240 K used
to delineate only upper-level clouds. The FSS is com-
puted using data from the full day (0700–1200 UTC), as
well as for preconvective (0700–2000 UTC) and con-
vective (2100–0300 UTC) time periods.

Inspection of the results for the 270-K threshold for
the preconvective (Fig. 11b), and convective hours
(Fig. 11c) shows that the forecast skill is higher during
the peak convective hours. For the PBL schemes, the
ACM2 has the greatest skill for all clouds during both
forecast periods. Comparison of the microphysics
schemes shows that the THOM scheme performs best
during preconvective hours, but that the M–Y scheme is
most accurate during the convective hours. The 270-K
threshold results indicate that the WDM6 microphysics
produces the worst scores at all spatial scales, which is
due to its large underprediction of cloud cover.

For the 240-K threshold (Fig. 12), none of the en-
semble members display very good skill during either
period for all of the neighborhood sizes. The WDM6 has
the worst skill of the microphysics schemes for both
times. When all of the forecasts are considered, the
THOM scheme has the best skill among the micro-
physics schemes at both times; however, this is not
necessarily true for each individual day, with the MORR

scheme often being more accurate (not shown). When
only the peak convective hours are included (Fig. 12c),
the THOM scheme is the most accurate microphysics
scheme, while the WDM6 is the worst. For the PBL
schemes, the ACM2 has the highest FSS for the entire
forecast period (Fig. 12a), followed by the YSU, MYJ,
QNSE, and MYNN schemes. During the preconvective
hours (Fig. 12b) the ACM2 and MYNN schemes are
most accurate; however, the QNSE, MYJ, and YSU
perform better than these schemes during convective
hours (Fig. 12c). The higher accuracy of the ACM2 and
QNSE schemes during the preconvective hours likely
occurs because they have the best area forecast of the
high clouds during the preconvective hours (not shown).

4. Discussion and conclusions

In this study, the ability of several PBL and cloud
microphysics parameterization schemes employed by
the CAPS high-resolution ensemble forecasts during the
2012 HWT Spring Experiment to accurately simulate
cloud characteristics over the contiguous United States
was evaluated through comparison of real and syn-
thetic *GOES-13* infrared brightness temperatures. Four
double-moment microphysics schemes predicting both
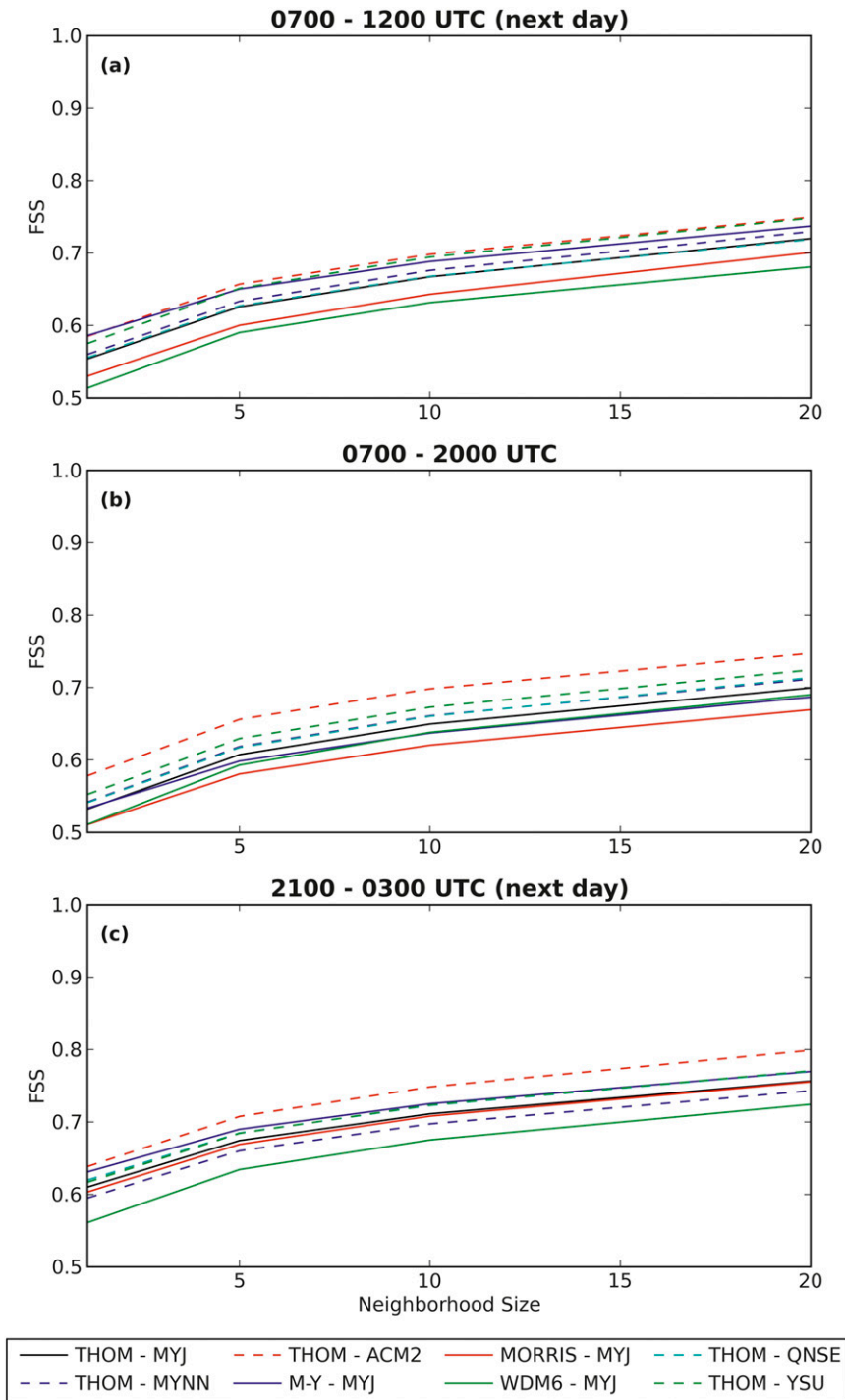the mass mixing ratio and number concentration for at

FIG. 11. Fraction skill score (FSS) for a 270-K threshold at (a) all times, (b) during pre-convective hours, and (c) during peak convective hours.

least one cloud species and five PBL schemes using either local or nonlocal mixing were evaluated. A sophisticated forward radiative transfer model was used to convert the model-simulated temperature, water vapor,

and cloud fields into synthetic infrared brightness temperatures. The results discussed here regarding the performance of the microphysics and PBL schemes are valid for cloud cover produced by the model, and may
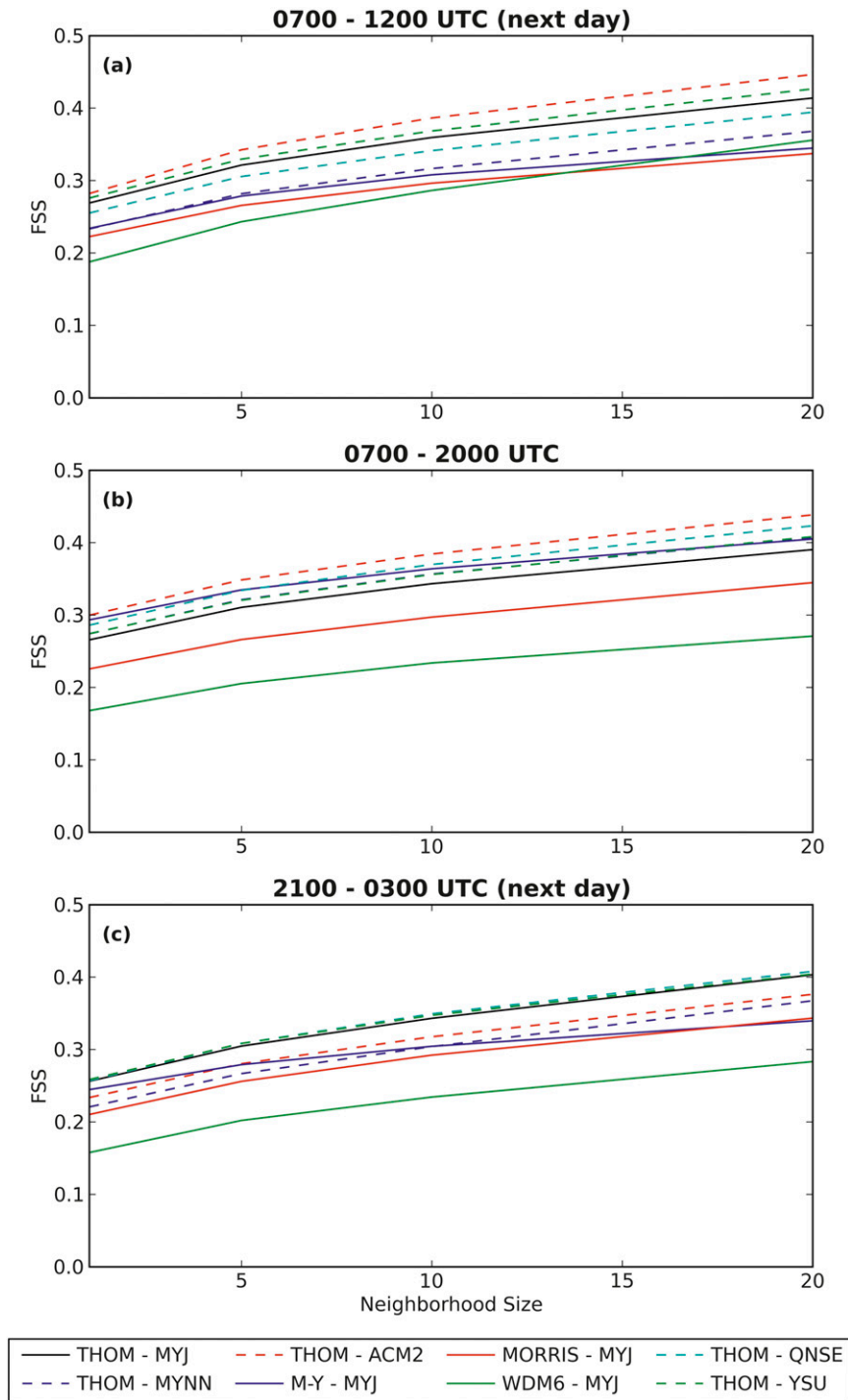
FIG. 12. Fraction skill score (FSS) for a 240-K threshold at (a) all times, (b) during pre-convective hours, and (c) during peak convective hours.

not necessarily be applicable to other forecast quantities, such as reflectivity and precipitation.

The M–Y and MORR schemes provided the best results among the microphysics schemes investigated

when only the low-level clouds were considered. Though all of the schemes underforecasted the spatial extent of the low-level clouds, the schemes tended to produce a cloud cover extent closer to observations, as seen in the

6.7–10.7-μm BTD area ratios. Overall, however, the M–Y and MORR microphysics schemes generated too much upper-level cloud cover, which resulted in the highest RMSE and MAE. Grasso and Greenwald (2004) found that pristine ice contributed the most to their synthetic 10.7-μm brightness temperatures, so an over-production of ice in the M–Y and MORR schemes could lead to too much cloud cover in the simulated satellite data. The maximum ice number concentration allowed in those two schemes is $10\,cm^{-3}$, which is much larger than the $0.25\,cm^{-3}$ permitted by the Thompson scheme. That could lead to more upper-level ice particles being created by the M–Y and MORR schemes than the Thompson scheme, which may be the reason for some of the disparity between the upper-level cloud cover they produced. The WDM6 scheme, on the other hand, consistently underforecasted the spatial extent of the lower- and upper-level cloud cover, leading to a worse FSS, but better RMSE and MAE, than the other microphysics schemes. Van Weverberg et al. (2013) found that the WSM6, which treats ice identically to the WDM6, produced a smaller ice number concentration and larger ice cloud particles in the upper troposphere that fell out faster than in the Thompson or Morrison schemes. This behavior helps explain the lack of cloud cover in the WDM6 forecasts. The smaller RMSE and MAE for this scheme represents a weakness in using traditional verification metrics since the smaller errors are primarily due to reduced susceptibility to the "double penalty" problem discussed earlier rather than to a more accurate cloud forecast. Overall, the THOM scheme was characterized by the most accurate upper-level cloud distribution based on the metrics investigated, including the 240-K threshold FSS, RMSE, MAE, and 6.7–10.7-μm BTD area ratios. It also performed well for the low-level clouds, though there were slightly fewer clouds than with the M–Y and MORR schemes. Even though THOM is not as complex as the M–Y and MORR schemes, these results indicate that the THOM scheme produced the most accurate cloud forecasts during the 2012 HWT Spring Experiment.

Comparison of the PBL results revealed that though there was a general lack of low-level clouds in all ensemble members, these clouds were most accurately forecast by the ACM2 scheme, as indicated by the 6.7–10.7-μm BTD distribution and the higher FSS for the 10.7-μm 270-K threshold. The FSS for the upper-level clouds was also highest for the ACM2 scheme in the preconvective hours; however, the QNSE and MYJ schemes, which are both local mixing schemes that are similar in how they handle unstable boundary layers, had the highest FSS during peak convective hours, when they produced more upper-level clouds than the rest.

The MYNN had the worst FSS and produced a high cloud area smaller than observed, but provided the most accurate distribution of clouds at or above the tropopause, as shown by the area ratio for 6.7–10.7-μm BTD $< -2\,K$. The other PBL schemes, particularly the MYJ and QNSE, had too many overshooting tops. These results suggest the convection produced by the ensemble members that use the MYJ and QNSE may have been too vigorous. Of the PBL schemes, the MYNN also had the most accurate surface temperatures as inferred from 10.7-μm brightness temperatures, showing improvement in the forecast of surface heating during the day over the other two local schemes—the MYJ and QNSE—which were too cool and greatly underforecast the warmest surface temperatures. In general, however, differences due to the PBL schemes are not as dramatic as those due to the microphysics schemes and the results are less consistent among the evaluation methods used, making it difficult to determine which PBL scheme performed best. In most cases, the ACM2, YSU, and MYNN PBL schemes provide the best forecast of satellite imagery, while the MYJ and QNSE generally have lower skill.

The differences in the results from the various verification methods employed in this study suggest the need for utilizing a suite of verification tools when evaluating high-resolution model performance, especially when relying on objective verification scores. Finally, the large differences between the various cloud and PBL schemes indicate that there is still large uncertainty in how these schemes represent subgrid-scale processes affecting cloud morphology, with attendant uncertainty in how these changes impact precipitation. Given the high spatial and temporal resolution of geosynchronous satellite observations, it is incumbent to continue using these valuable datasets to help improve the accuracy of the simulated cloud field generated by high-resolution numerical models. A more accurate depiction of cloud structure could improve forecasts of high-impact weather events and could also aid the renewable energy sector through better forecasts and integration of solar energy production within the electrical grid.

National Institute for Computational Science (http://www.nics.tennessee.edu/). The University of Oklahoma Supercomputing Center for Research and Education (OSCER) supercomputer was also used for ensemble postprocessing.

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.

Baum, B. A., P. Yang, A. J. Heymsfield, S. Platnick, M. D. King, and S. T. Bedka, 2005: Bulk scattering models for the remote sensing of ice clouds. Part II: Narrowband models. *J. Appl. Meteor.,* **44,** 1896–1911.

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

Chaboureau, J.-P., and J.-P. Pinty, 2006: Validation of a cirrus parameterization with Meteosat Second Generation observations. *Geophys. Res. Lett.,* **33,** L03815, doi:10.1029/2005GL024725.

Chen, S.-H., and W.-Y. Sun, 2002: A one-dimensional time dependent cloud model. *J. Meteor. Soc. Japan,* **80,** 99–118.

Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo. 104606, Vol. 3, 85 pp.

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.

Coniglio, M. C., J. Corria Jr., P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting,* **28,** 842–862.

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

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.

Gao, J., M. Zue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.,* **21,** 457–469.

Grasso, L. D., and T. Greenwald, 2004: Analysis of 10.7-mm brightness temperatures of a simulated thunderstorm with two-moment microphysics. *Mon. Wea. Rev.,* **132,** 815–825.

——, 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.

——, ——, and M. DeMaria, 2010: Comparison between observed and synthetic 6.5 and 10.7 mm *GOES-12* imagery of thunderstorms that occurred on 8 May 2003. *Int. J. Remote Sens.,* **31,** 647–663.

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

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

Hong, S.-Y., 2010: A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon. *Quart. J. Roy. Meteor. Soc.,* **136,** 1481–1496.

——, and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.,* **42,** 129–151.

——, J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.,* **132,** 103–120.

——, Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.,* **134,** 2318–2341.

Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.,* **134,** 675–698.

Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.,* **49,** 1831–1844.

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.

Karlsson, K.-G., 1996: Validation of modeled cloudiness using satellite-estimated cloud climatologies. *Tellus,* **48A,** 767–785.

Kong, F., and Coauthors, 2007: Preliminary analysis on the real-time storm-scale ensemble forecasts produced as a part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction,* Salt Lake City, UT, Amer. Meteor. Soc., 3B.2. [Available online at https://ams.confex.com/ams/22WAF18NWP/techprogram/paper_124667.htm.]

——, M. Xue, K. W. Thomas, Y. Wang, K. Brewster, and X. Wang, 2012a: CAPS storm-scale ensemble forecast in NOAA Hazardous Weather Testbed Spring Experiments. *Extended Abstracts, Third WMO/WWRP Int. Symp. on Nowcasting and Very Short Range Forecasting,* Rio de Janeiro, Brazil, WMO/WWRP. [Available online at http://www.labhidro.iag.usp.br/wsn12/papers/num4.pdf.]

——, and Coauthors, 2012b: CAPS storm-scale ensemble forecasting system for the NOAA HWT 2012 Spring Experiment: Impact of IC/LBC perturbations. *Extended Abstracts, 26th Conf. on Severe Local Storms,* Nashville, TN, Amer. Meteor. Soc., P138. [Available online at https://ams.confex.com/ams/26SLS/webprogram/Paper211966.html.]

Lim, K.-S. S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.,* **138,** 1587–1612.

Lopez, P., K. Finkele, P. Clark, and P. Mascart, 2003: Validation and intercomparison of three FASTEX cloud systems: Comparison with coarse-resolution simulations. *Quart. J. Roy. Meteor. Soc.,* **129,** 1841–1871.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.,* **83,** 407–430.

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.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.,* **20,** 851–875.

Milbrandt, J. A., and M. K. Yau, 2005a: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.,* **62,** 3051–3064.

——, and ——, 2005b: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.,* **62,** 3065–3081.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102** (D14), 16 663–16 682.

Morcrette, J. J., 1991: Evaluation of model-generated cloudiness: Satellite-observed and model-generated diurnal variability of brightness temperatures. *Mon. Wea. Rev.,* **119,** 1205–1224.

Morrison, H., and J. Milbrandt, 2011: Comparison of two-moment bulk microphysics schemes in idealized supercell thunderstorm simulations. *Mon. Wea. Rev.,* **139,** 1103–1130.

——, G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.,* **137,** 991–1007.

Nakanishi, M., 2001: Improvement of Mellor-Yamada turbulence closure model based on large-eddy simulation data. *Bound.-Layer Meteor.,* **99,** 349–378.

——, and H. Niino, 2004: An improved Mellor-Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.,* **112,** 1–31.

——, and ——, 2006: An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.,* **119,** 397–407.

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

O'Dell, C. W., A. K. Heidinger, T. Greenwald, P. Bauer, and R. Bennartz, 2006: The successive-order-of-interaction radiative transfer model. Part II: Model performance and applications. *J. Appl. Meteor. Climatol.,* **45,** 1403–1413.

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

——, and T. J. Greenwald, 2008: Comparison of WRF model-simulated and MODIS-derived cloud data. *Mon. Wea. Rev.,* **136,** 1957–1970.

——, 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 measurement simulation and product algorithm development. *J. Atmos. Oceanic Technol.,* **24,** 585–601.

——, 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.

Pleim, J. E., 2007a: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteor. Climatol.,* **46,** 1383–1395.

——, 2007b: A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *J. Appl. Meteor. Climatol.,* **46,** 1396–1409.

Rikus, L., 1997: Application of a scheme for validating clouds in an operational global NWP model. *Mon. Wea. Rev.,* **125,** 1615–1637.

Roberts, N. M., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.,* **15,** 163–169.

——, and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high- resolution forecasts of convective events. *Mon. Wea. Rev.,* **136,** 78–97.

Sanders, F., and J. R. Gyakum, 1980: Synoptic-dynamic climatology of the ''bomb.'' *Mon. Wea. Rev.,* **108,** 1589–1606.

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.

Seemann, S. W., E. E. Borbas, R. O. Knuteson, G. R. Stephenson, and H.-L. Huang, 2008: Development of a global infrared land surface emissivity database for application to clear-sky sounding retrievals from multispectral satellite radiance measurements. *J. Appl. Meteor. Climatol.,* **47,** 108–123.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.,* **132,** 3019–3032.

——, and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, Mesoscale and Microscale Meteorology Division, NCAR, Boulder, CO, 125 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]

Sukoriansky, S., B. Galperin, and I. Staroselsky, 2005: A quasi-normal scale elimination model of turbulent flows with stable stratification. *Phys. Fluids,* **17,** 85–107.

——, ——, and V. Perov, 2006: A quasi-normal scale elimination model of turbulence and its application to stably stratified flows. *Nonlinear Processes Geophys.,* **13,** 9–22.

Sun, Z., and L. Rikus, 2004: Validating model clouds and their optical properties using geostationary satellite imagery. *Mon. Wea. Rev.,* **132,** 2006–2020.

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

Tselioudis, G., and C. Jakob, 2002: Evaluation of midlatitude cloud properties in a weather and a climate model: Dependence on dynamic regime and spatial resolution. *J. Geophys. Res.,* **107,** 4781, doi:10.1029/2002JD002259.

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.

Xie, B., J. C. H. Fung, A. Chan, and A. Lau, 2012: Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model. *J. Geophys. Res.,* **117,** D12103, doi:10.1029/2011JD017080.

Xue, M., D. H. Wang, J. D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and assimilation. *Meteor. Atmos. Phys.,* **82,** 139–170.

——, and Coauthors, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction,* Salt Lake City, UT, Amer. Meteor. Soc., 3B.1. [Available online at https://ams.confex.com/ams/22WAF18NWP/techprogram/paper_124587.htm.]