

Article

Evaluating the Impact of Planetary Boundary Layer, Land Surface Model, and Microphysics Parameterization Schemes on Simulated GOES-16 Water Vapor Brightness Temperatures

Sarah M. Griffin *  and Jason A. Otkin 

Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, Madison, WI 53706, USA; jasono@ssec.wisc.edu

* Correspondence: sarah.griffin@ssec.wisc.edu

Abstract: The impact of several land surface models (LSMs) and microphysics (MP), planetary boundary layer (PBL), and surface layer schemes on the accuracy of simulated brightness temperatures (BTs) from water vapor (WV) sensitive bands was examined via comparison with observations from the GOES-16 Advanced Baseline Imager. Nine parameterization configurations were evaluated. Analysis revealed that, compared to the Thompson MP scheme, the National Severe Storms Laboratory MP scheme produced lower simulated WV BTs in the upper troposphere but higher WV BTs in the middle and lower troposphere. The configuration with the Geophysical Fluid Dynamics Laboratory MP and hybrid eddy-diffusivity mass-flux (EDMF) PBL instead of Mellor–Yamada–Nakanishi–Niino (MYNN) PBL produced higher BTs. Yet, changing the PBL from MYNN to Shin–Hong or EDMF reduced the simulated WV BTs. Changing the LSM from Noah to RUC also resulted in lower simulated WV BTs, which were further enhanced with the MYNN surface layer instead of the GFS. The location and orientation of upper-level jet streams and troughs was assessed using the location of WV gradient objects. Every configuration had an increased translation speed compared to the observations, as forecast WV gradient objects were west of the observation objects early in the forecast and then east later in the forecast.

Keywords: model verification; water vapor brightness temperatures; satellite observations



Citation: Griffin, S.M.; Otkin, J.A. Evaluating the Impact of Planetary Boundary Layer, Land Surface Model, and Microphysics Parameterization Schemes on Simulated GOES-16 Water Vapor Brightness Temperatures. *Atmosphere* **2022**, *13*, 366. <https://doi.org/10.3390/atmos13030366>

Academic Editor: Yoshihiro Tomikawa

Received: 19 January 2022

Accepted: 18 February 2022

Published: 22 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There are many operational uses for satellite infrared brightness temperatures (BTs) from bands sensitive to water vapor (WV) and clouds (hereafter referred to as WV bands). WV BTs can be used to identify areas of possible aircraft icing [1,2], track disturbances that produce severe weather [3], and derive atmospheric motion vectors [4,5] that can be used to compute tropical cyclone outflow and estimate future intensification [6]. In addition, many upper-level synoptic features such as jet streams [7,8] and troughs [9] can be identified in WV BTs. The location of upper-level jet streams and troughs can help delineate areas of potential aircraft turbulence, as wind shear near upper-level jet streams is a known cause of aircraft turbulence [10]. Severe weather can be associated with the exit and entrance regions of jet streams [11] or negatively tilted 500 hPa troughs [12,13]. These examples illustrate the fact that efforts to verify and improve the accuracy of WV BTs in high-resolution numerical weather prediction (NWP) models are very beneficial for a range of applications.

Many studies using satellite BTs to validate upper-level cloud forecasts (e.g., [14–28]) have noted that the cloud accuracy is dependent on the choice of model parameterization scheme [23,29]. However, despite the operational uses of WV BTs, little previous research has been conducted comparing simulated and observed WV BTs. Otkin et al. [27] compared simulated WV BTs from a large-scale Weather Research and Forecasting (WRF) model simulation to observations from the Spinning Enhanced Visible and Infrared Imager. Soden and Bretherton [30] compared Geostationary Operational Environmental Satellite (GOES)

6.7 μm BTs to simulated BTs from the European Centre for Medium-Range Weather Forecasts model and the National Center for Atmospheric Research Community Climate Model. Since forecasts of WV are sensitive to the microphysics scheme [31], Cintineo et al. [29] and Chung et al. [32] investigated the impact of the microphysics scheme on WV BTs. However, Cintineo et al. [29] only analyzed the 6.7 μm BTs from the GOES-13 imager, whereas observations from three WV bands are now available from the Advanced Baseline Imager (ABI) on the new GOES satellites [33]. Chung et al. [32] examined how different microphysics schemes affect all three WV bands; however, they did not assess the impact of other parameterization schemes such as the planetary boundary layer (PBL) on the forecasts.

Model accuracy, especially in global models, is often assessed using metrics such as 500 hPa height anomaly correlations [34,35] that capture errors in the large-scale flow patterns. Such statistics can be computed using conventional observation methods such as radiosondes or through comparisons with gridded model analyses; however, both approaches have limitations. Conventional observations suffer from poor temporal and spatial resolution with large data voids over the oceans. Model analyses provide continuous coverage; however, they contain biases and are only as accurate as the data assimilation system and model used to generate them. Moreover, these limitations are even more acute in regions where observations are scarce or nonexistent. Satellite data, in contrast, provide nearly continuous global observations with high spatial resolution that are useful for verifying the accuracy of both regional-scale and global NWP models. WV BTs in particular could be used to assess the location of upper-level features such as jet streams and troughs since they are often co-located with gradients in WV BTs [36–38]. Mohan et al. [39] have recently shown that the location of these upper-level features in model forecasts is impacted by the microphysics scheme.

The purpose of this study is to assess the impact of different parameterization schemes on the accuracy of WV BTs, and to use gradients in WV BTs to assess the accuracy with regard to upper-level jet streams and troughs. These schemes include different microphysics (MP) and planetary boundary layer (PBL) schemes, as well as different land surface models (LSM) and surface layers, that are being considered for inclusion in future operational versions of the Finite-Volume Cubed-Sphere [40,41] limited area model (FV3-LAM). This assessment uses output from a suite of experimental FV3-LAM simulations run by the Center for Analysis and Prediction of Storms during the 2019 Hazardous Weather Testbed (HWT) Spring Forecasting Experiment [42,43]. Similar assessments are conducted using output from the Environmental Modeling Center (EMC) and National Severe Storms Laboratory (NSSL) for the 2020 HWT Spring Forecasting Experiment.

2. Data

2.1. Model Configurations

This analysis used output from a suite of FV3-LAM configurations employing different parameterization schemes during the 2019 and 2020 NOAA HWT Spring Forecasting Experiments. All forecasts were initialized at 00 UTC and then run for 60 h. Data from the 2019 HWT included seven configurations initialized each weekday from 14–31 May 2019, while data from the 2020 HWT included three configurations initialized each weekday from 27 April to 29 May 2020. Forecast hours 0 to 5 were excluded from the analysis to reduce the impact of model spin-up processes on the forecast cloud field, because the model starts from a cloud-free state.

For the 2019 HWT dataset, all forecasts were run on a domain covering CONUS at 3 km grid spacing. A description of each model configuration is included in Table 1. The Control configuration is the baseline from which the other model configurations vary in the 2019 HWT dataset. As described in Griffin et al. [23], it employs the Thompson MP scheme with the aerosol-aware capability turned off [44,45], the Mellor–Yamada–Nakanishi–Niino (MYNN) v3.6 PBL [46,47], the Global Forecasting System (GFS) surface layer, the Noah LSM [48,49], and the North American Model (NAM) initial and lateral boundary conditions. A second configuration explores the sensitivity to the cloud MP

schemes by replacing the Thompson scheme with the National Severe Storms Laboratory (NSSL) scheme [50], hereafter referred to as MP-NSSL. A third set of experiments replaced the MYNN PBL scheme with either the Shin–Hong (SH) [51] or hybrid eddy-diffusivity mass-flux (EDMF) [52] schemes, hereafter referred to as the PBL-SH and PBL-EDMF configurations, respectively. A final set of configurations employed the RUC v3.6+ [53] LSM with either the GFS or MYNN surface layers, denoted as LSM-RUC_SFC-GFS and LSM-RUC_SFC-MYNN, respectively. The rapid radiative transfer model for general circulation models (RRTMG) [54] radiation scheme was used in all of the simulations when computing radiative fluxes at the surface and in the atmosphere, and in-cloud fluxes are directly coupled to the cloud properties via the prognostic mixing ratios and diagnosed effective radii for each cloud species [28].

Table 1. Description of model configurations for the 2019 HWT dataset used in this analysis.

Name	Microphysics Scheme	Planetary Boundary Layer Scheme	Surface Layer	Land Surface Model
Control	Thompson	MYNN	GFS	Noah
MP-NSSL	National Severe Storms Laboratory	MYNN	GFS	Noah
PBL-SH	Thompson	Shin–Hong	GFS	Noah
PBL-EDMF	Thompson	EDMF	GFS	Noah
LSM-RUC_SFC-GFS	Thompson	MYNN	GFS	RUC
LSM-RUC_SFC-MYNN	Thompson	MYNN	MYNN	RUC

The configurations for the 2020 HWT simulations are shown in Table 2. The EMC FV3-LAM configuration uses the Geophysical Fluid Dynamics Laboratory (GFDL) MP scheme [55], the hybrid EDMF PBL [51], the GFS surface layer, and the Noah LSM. The hybrid EDMF PBL scheme uses the EDMF scheme for strongly unstable PBLs, while the eddy-diffusivity counter-gradient (EDCG) scheme is used for weakly unstable PBLs mostly over the tropics [51]. EMC FV3-LAMx has the same configuration as Control from the 2019 HWT dataset: the Thompson MP scheme, the MYNN v3.6 PBL, the GFS surface layer, and the Noah LSM. Lastly, the NSSL FV3-LAM configuration is identical to the EMC FV3-LAMx configuration except that it uses the MYNN surface layer. Unlike the 2019 experiments, these configurations employ the GFS initial and lateral boundary conditions. These forecasts were also run on a domain covering CONUS at 3 km grid spacing.

Table 2. Description of configurations for the 2020 HWT dataset used in this analysis.

Name	Microphysics Scheme	Planetary Boundary Layer Scheme	Surface Layer	Land Surface Model
EMC FV3-LAM	Geophysical Fluid Dynamics Laboratory	Hybrid EDMF	GFS	Noah
EMC FV3-LAMx	Thompson	MYNN	GFS	Noah
NSSL FV3-LAM	Thompson	MYNN	MYNN	Noah

2.2. Water Vapor Brightness Temperatures

The verification dataset used in this study consisted of the observed 6.2, 6.9, and 7.3 μm BTs from the GOES-16 ABI sensor. These three bands are generally considered upper-level, mid-level, and low-level WV bands with the weighting function peaking at approximately 325 hPa, 475 hPa, and 675 hPa, respectively, in clear-sky scenes [33]. These BTs had a 2 km pixel spacing at the nadir, or approximately 3–4 km over the study domain.

For the observations, the GOES-16 radiances were remapped to the model grid using an area-weighted average of all the observed pixels overlapping a given model grid box, and then converted to BTs.

Simulated ABI BTs were generated for all configurations using the Community Radiative Transfer Model (CRTM V2+) and vertical profiles of temperature, specific humidity, and cloud hydrometeors [56,57]. A diagnosed effective radius for each microphysical species (cloud water, rainwater, ice, snow, and graupel) that was consistent with the assumptions made by the parameterization scheme [28,58] was input to the CRTM for cloudy grid points. The National Polar-orbiting Operational Environmental Satellite System (NPOESS) emissivity database was used to model land surface emissivity [56].

2.3. Pressure-Level Data

In addition to simulated BTs, pressure-level data are available for the EMC FV3-LAM. These data are available for 45 pressure levels from 2 hPa to 1000 hPa. Specifically, the zonal and meridional wind components and geopotential height were used to evaluate the utility of the gradient-based verification method for synoptic and mesoscale features. Specific humidity was also used to calculate the water vapor mixing ratio using the equation:

$$\text{Water Vapor Mixing Ratio} = \frac{\text{Specific Humidity}}{(1 - \text{Specific Humidity})}. \quad (1)$$

The water vapor mixing ratio was used to compare the simulated BTs to the amount of water vapor in the atmosphere.

3. Methodology

3.1. Grid Point Metrics

Overall model accuracy was assessed using the mean absolute error (MAE). The MAE was used instead of the root mean square error because errors in the simulated BTs did not follow a normal distribution based on a Shapiro–Wilk test [59,60]. The MAE was calculated using the equation

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |F_i - O_i|, \quad (2)$$

where F (O) represents the forecast (observed) BTs. An MAE of zero represents a perfect forecast.

The model bias was assessed using two metrics. The first metric was the mean bias error (MBE), which is calculated using the equation

$$\text{MBE} = \frac{1}{N} \sum_{i=1}^n (F_i - O_i). \quad (3)$$

Since the MBE calculates the difference between the forecast and observed BTs before calculating the average, it requires the same number of grid points for both the forecast and observed BTs. The second metric was the mean difference (MD), which is calculated using the equation:

$$\text{MD} = \frac{1}{N_F} \sum_{i=1}^{N_F} F_i - \frac{1}{N_O} \sum_{i=1}^{N_O} O_i. \quad (4)$$

The MD differs from the MBE because the average is computed separately for the forecast and observed BTs before calculating the difference between these two quantities. This approach is useful for situations when the forecast and the observations do not have the same number of grid points, which may occur if the forecasts have more or fewer clouds than the observations. It is used when evaluating error characteristics for clear grid points. A positive (negative) MBE and MD indicates that the simulated BTs are too high (low) compared to the observed BTs.

3.2. Identifying Upper-Level Jet Streams and Troughs

As the upper-level features such as jet streams between 150 and 350 hPa and 500 hPa short-wave troughs are often co-located with gradients in WV BTs [36–38], this analysis focused on the location of spatial changes in 6.2 μm BTs. These spatial changes, hereafter referred to as WV gradients, were calculated using the equation: $\nabla\text{BT} = \sqrt{(\Delta\text{BT}(x) + \Delta\text{BT}(y))^2}$, where ΔBT is the difference between BTs at ± 5 grid points surrounding a given point. An example of the simulated 6.2 μm BTs and corresponding BT gradients valid at 08 UTC on 02 May 2020 are shown in Figure 1a,b, respectively. As can be seen in Figure 1c,d, this was a complex scene characterized by areas of enhanced upper-level winds associated with a large ridge over the central U.S., troughs over the West and East Coasts, and a short-wave trough riding the top of the ridge over the north-central U.S. and southern Canada.

Comparison of Simulated 6.2 μm BT, Maximum 150-350-hPa Wind Speed and 500-hPa Heights valid 20200502 at 08UTC from 20200502 at 00 UTC

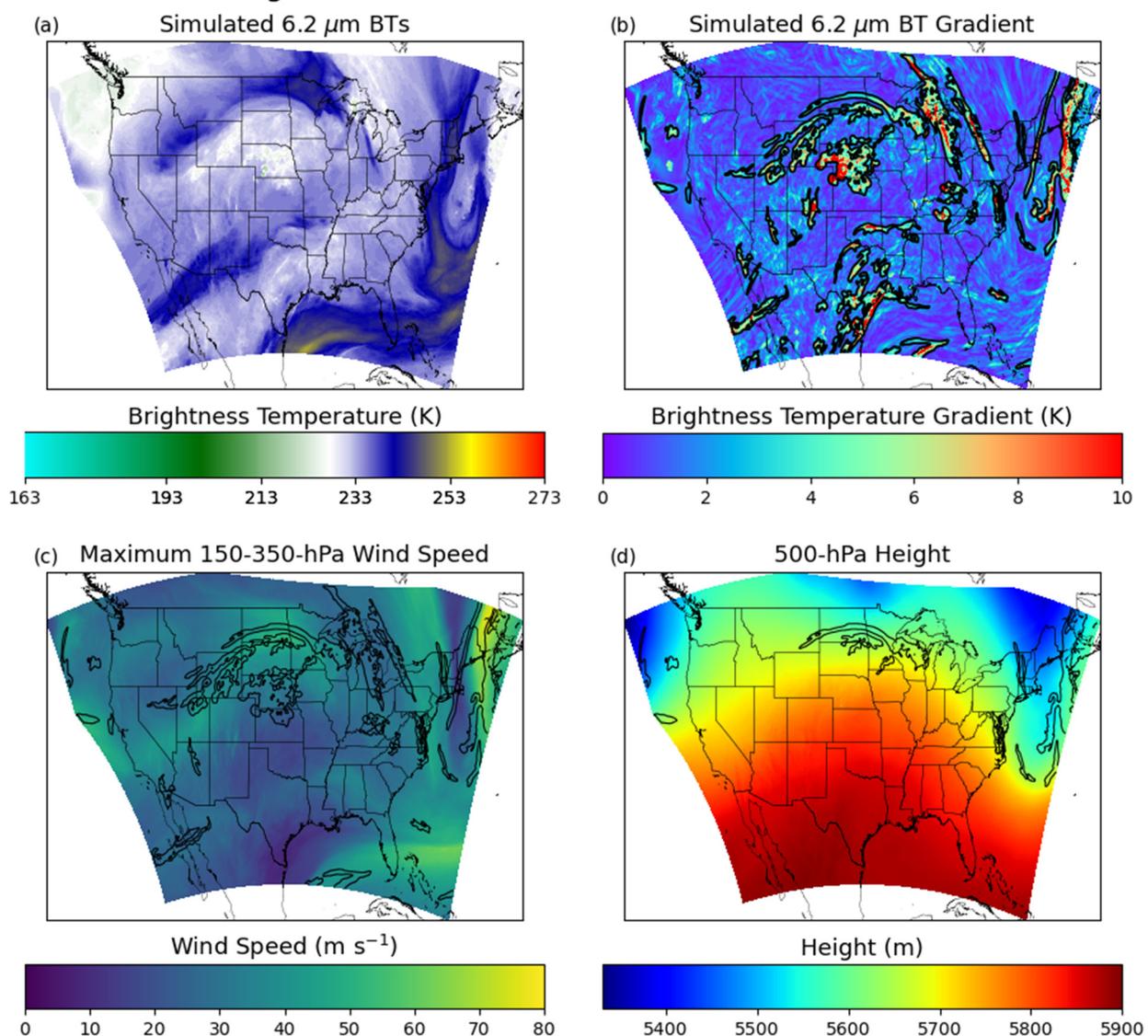


Figure 1. (a) Simulated 6.2 μm BTs, (b) 6.2 μm BT gradients, (c) maximum EMC FV3-LAM wind speed between 150 hPa and 350 hPa, and (d) EMC FV3-LAM 500 hPa heights valid 20200502 at 08 UTC. The black outlined objects in (b) indicate MODE objects defined using a BT gradient threshold of 2 K. The black outlined objects in (c,d) indicate objects that are co-located with either a jet stream or a trough, respectively.

The Method for Object-Based Diagnostic Evaluation (MODE) [61–64] was used to evaluate the attributes of WV gradient objects. Using a process called convolution thresholding, MODE identifies objects in the smoothed observed and forecast WV gradient fields and then calculates object attributes such as location or axis angle. In this analysis, BTs were smoothed using a convolution radius of 5 grid points (15 km). The black outlines in Figure 1b indicate MODE objects identified using a minimum gradient threshold of 2 K, which was chosen to lessen the impact of small-scale noise. In addition to object attributes, MODE matches forecast and observation objects, with the similarity between matched objects represented by an interest score that ranges from 0 to 1, with 1 representing a perfect match. MODE also identifies attributes for matched object pairs, such as the centroid distance (distance between objects' centers of mass) and area ratio (ratio of matched object areas).

In this study, we used WV gradient objects from observed and simulated 6.2 μm BTs to assess the accuracy of forecast upper-level jet streams and troughs (ULJSTs). Because the spatial resolution of radiosonde observations is much coarser than that of the BTs, we used high-resolution wind and geopotential height analyses from the EMC FV3-LAM configuration to identify the location of observed ULJST objects. Therefore, this object-based analysis was limited to 2020, because similar high-resolution analyses are not available for 2019. Jet streams were identified as maximum wind speeds greater than 26 m s^{-1} and the edges of troughs were identified as areas where the geostrophic wind speed was greater than 25 m s^{-1} . At least 50% of a given MODE forecast object had to overlap an observed jet stream between 150 and 350 hPa or a 500 hPa trough to be considered co-located with an ULJST. EMC FV3-LAM forecast WV gradient objects that were co-located with jet streams or troughs are shown in the black outlines in Figure 1c,d, respectively. To allow comparison with all configurations, we assumed that the observed 6.2 μm WV gradient objects with the highest interest scores above 0.5 with respect to WV gradient objects in the EMC FV3-LAM forecasts were co-located with an ULJST. This is a reasonable approximation that allows these objects to be used to evaluate all of the configurations.

3.3. Mean Error Distance

The distance between matching WV gradient objects was calculated using the mean error distance (MED). The MED was used in addition to the centroid distance provided by MODE because the centroid distance could be more strongly influenced by other features such as cloud distribution. The MED calculates the shortest distance between every grid point identified as an observation object to the closest grid point identified as a forecast object [65]. This could be more beneficial for forecasts of aircraft turbulence and severe weather on the edges of jet streams or troughs. The MEDs from the observation objects (O) to the forecast objects (F) were calculated using:

$$\text{MED}(F, O) = \frac{1}{N_O} \sum_{x \in O} d(x, F) \quad (5)$$

where N_O is the number of grid points in the observation objects and $d(x, F)$ is the shortest distance from a grid point in an observation object to a forecast object. The MED has a perfect value of zero, which occurs when all grid points in the observation and forecast objects overlap.

4. Results

4.1. Grid Point Metrics

The MAE and MBE for the 6.2 μm BTs for the 2019 HWT dataset are shown in Figure 2a,b, respectively, with the rightmost image displaying the statistical significance from Control over all forecast hours based on Student's t-test. Smaller values indicate that the difference between the average MAE/MBE for the configuration and Control is more statistically significant, as seen for the MBEs. Control had the lowest average MAE, and therefore had the most accurately simulated 6.2 μm BTs overall. Control also

had the least-negative average MBE, indicating it has less of a cold bias than the other configurations. PBL-SH and PBL-EDMF had the next-highest MAEs and more negative MBEs than Control. The fact that PBL-SH and PBL-EDMF had more-negative MBEs than Control is consistent with results from Cintineo et al. [29] indicating that the MYNN PBL scheme had the smallest cold bias in the 6.7 μm WV BTs from GOES-13. MP-NSSL had the next-highest average MAE, though the MBE was smaller than that of PBL-SH. While MP-NSSL had the highest MAE for early forecast hours, the largest average MAEs were for LSM-RUC_SFC-GFS and LSM-RUC_SFC-MYNN, and the difference in MAE between Control and LSM-RUC_SFC-MYNN was statistically significant. Therefore, changing the land surface model from Noah to RUC decreased the accuracy of the 6.2 μm BTs, with further reductions in accuracy occurring when using the MYNN surface layer instead of GFS. LSM-RUC_SFC-MYNN also had the most-negative MBE of all configurations. These results when using the RUC land surface model were similar to those observed in Griffin et al. [23] for the 10.3 μm infrared window channel, which suggests that the negative MBE for LSM-RUC_SFC-MYNN was due to excessive cloud cover.

Since WV BTs are sensitive not only to WV but also to clouds, Figure 2c displays the MD for clear grid points only, defined as grid points where the 10.3 μm BT is greater than 270 K. The MD was used since the number of clear grid points varies among the model configurations and observations. The negative MD indicates that the model forecasts were too moist in the upper troposphere, which was also shown in Otkin et al. [66]. All configurations also had more clear grid points than the observations, as seen in Figure 2d which displays the number of grid points as a function of BT computed using all forecast cycles and times. The delineation between cloudy and clear grid points is the lowest observed clear-sky BT averaged over all forecast initializations and hours. MP-NSSL had the largest MD, with an average MD that was statistically significant compared to Control. MP-NSSL also had the most clear grid points, which is consistent with the reduction in clouds for the NSSL MP scheme compared to Thompson in Griffin et al. [23]. Only PBL-SH had an average MD for clear grid points that was statistically significant compared to Control, even though the percentage of clear pixels was similar for both PBL-SH and Control. LSM-RUC_SFC-MYNN had the lowest number of clear grid points and the third-smallest average MD, confirming that some of the higher MBE was due to the excessive cloud cover. However, the average MD was not statistically significant compared to Control. All configurations had more grid points than the observations at the lowest BTs (<200 K) and fewer for the highest BTs (>245 K), resulting in the negative MBEs seen in Figure 2b.

The MAEs and MBEs for the 6.2 μm BTs for the 2020 HWT dataset are shown in Figure 3. Overall, the EMC FV3-LAM configuration had the smallest MAE but the highest MBE. EMC FV3-LAMx was the least-accurate configuration. While EMC FV3-LAMx employs the same parameterization schemes as the Control configuration from the 2019 HWT, the initial and lateral boundary conditions are different. This could be why EMC FV3-LAMx was the least-accurate configuration even though Control was the most-accurate configuration from the 2019 HWT (Figure 2a). EMC FV3-LAMx also had the least-positive MBE (Figure 3b), with the difference between the average MBE being statistically significant compared to EMC FV3-LAM and NSSL FV3-LAM. This was possibly due to an overproduction of the coldest clouds, as suggested by the increased number of grid points with BT < 210 K compared to the observations (Figure 3d). EMC FV3-LAMx had a lower MBE than NSSL FV3-LAM, which indicates that using the GFS surface layer instead of MYNN reduces the production of clouds, based on Figure 3d.

For clear grid points (Figure 3c), all configurations had a positive average MD, which is the opposite of what occurred in the 2019 HWT dataset. The positive MD for the 2020 HWT could be due to the use of GFS analyses rather than NAM analyses for the initial and lateral boundary conditions. EMC FV3-LAM had the lowest MD, and the difference from EMC FV3-LAMx was statistically significant. Therefore, either the GFDL MP or hybrid EDMF PBL schemes reduced the clear-sky 6.2 μm BTs compared to the Thompson MP scheme and MYNN PBL. Assuming the EDMF PBL scheme is used by EMC FV3-LAM for

these grid points, the possibility of the EDMF PBL reducing the 6.2 μm BTs compared to MYNN is consistent with Figure 2. However, since the difference between the average MD for clear grid points was five times greater between EMC FV3-LAM and EMC FV3-LAMx in 2020 compared to Control and PBL-EDMF in 2019 (Figure 2), it is probable that the GFDL MP scheme also reduced the 6.2 μm BTs. EMC FV3-LAMx had a slightly lower average MD for clear grid points than NSSL FV3-LAM. Therefore, using the MYNN surface layer instead of the GFS does not necessarily result in less-accurate clear-sky 6.2 μm BTs, as is suggested by the 2019 HWT dataset, and this may instead depend on the LSM or initial and lateral boundary conditions used. Running LSM-RUC_SFC-MYNN and NSSL FV3-LAM for the same dates would provide a more accurate comparison, but that is beyond the scope of this study.

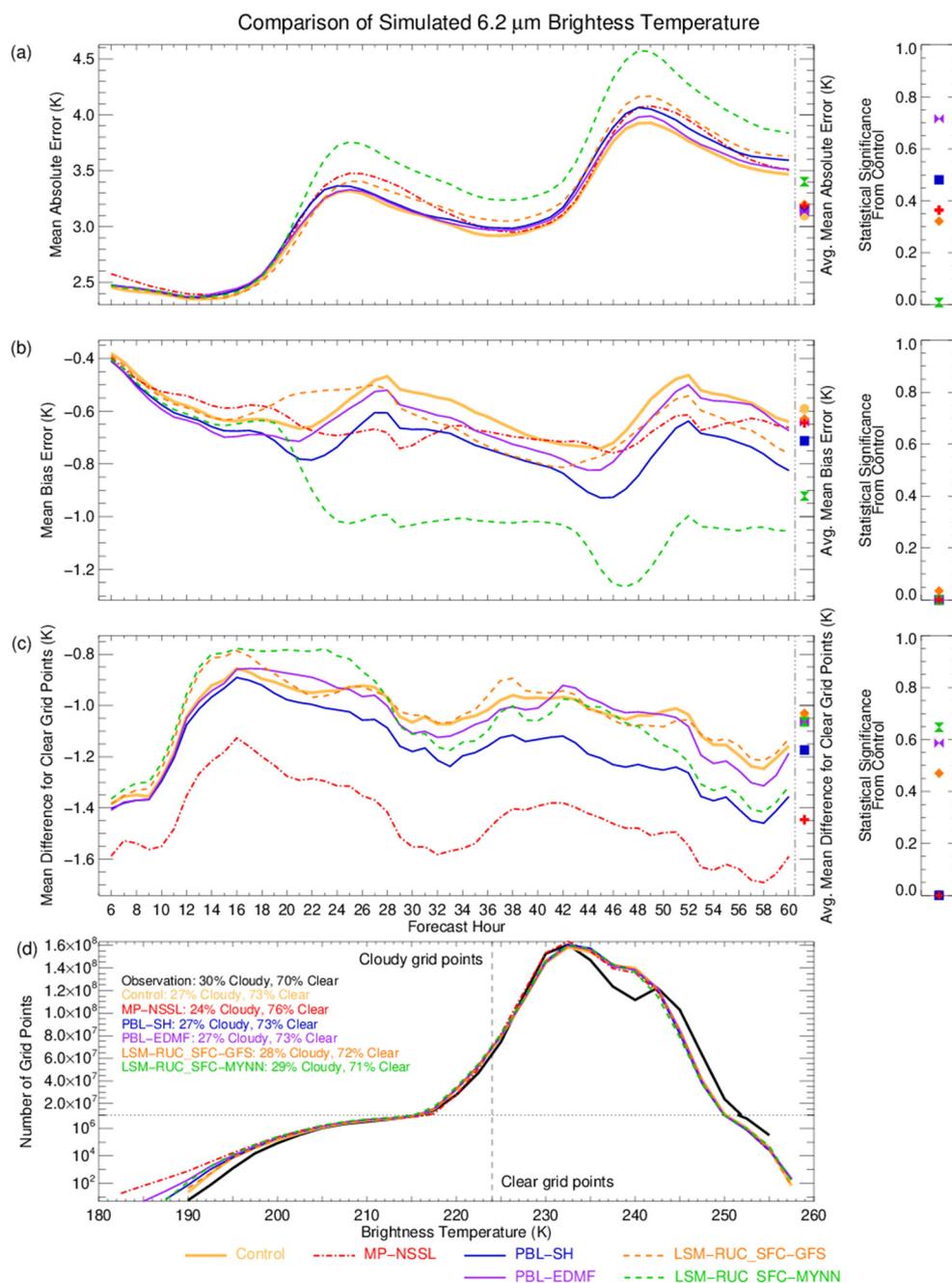


Figure 2. Line plot of GOES-16 6.2 μm brightness temperature: (a) mean absolute error, (b) mean bias error, (c) mean difference for clear grid points, and (d) distribution of number of grid points by

brightness temperature for configurations employed during the 2019 Hazardous Weather Testbed. The colored symbols on the right side of (a–c) show the average errors for each configuration over the entire time series and the statistical significance from Control. The delineation between cloudy and clear grid points in (d) is the lowest observed clear-sky BT, averaged over all forecast initializations and hours.

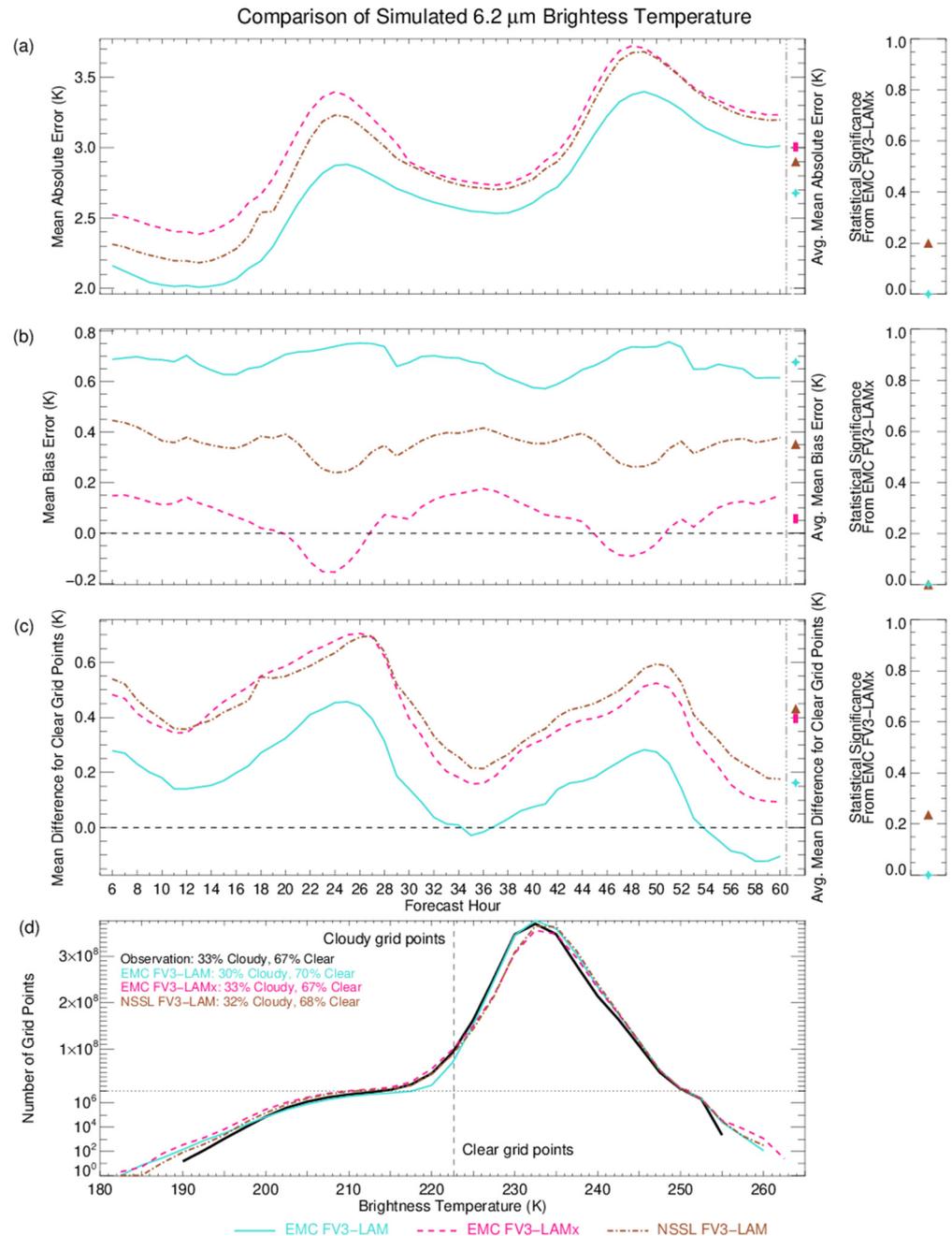


Figure 3. As in Figure 2, but using configurations employed during the 2020 Hazardous Weather Testbed. The statistical significance in (a–d) is calculated compared to EMC FV3-LAMx.

To evaluate WV biases in the middle and lower troposphere, Figure 4 shows the average MBE and MD for clear grid points for the 6.2, 6.9, and 7.3 μm WV BTs during the 2019 and 2020 HWT forecasts. Only biases are shown in Figure 4, because the relationship between the average MAE and the different configurations is the same for each WV band. Compared to the 6.2 μm BTs (Figure 4a), one noticeable difference is that the MBEs for the 6.9 μm BTs (Figure 4c) were more negative, with PBL-SH and PBL-EDMF again having

more-negative MBEs than Control. The MD for clear grid points was also more negative for the 6.9 μm BTs compared to the 6.2 μm BTs (Figure 4b,d). This is especially evident in the 2020 HWT datasets, as the MDs for clear grid points were negative in this case instead of positive. For the 7.3 μm BTs sensitive to WV in the middle and lower troposphere, the MBEs were the highest of any WV band. For the 2019 HWT forecasts, LSM-RUC_SFC_MYNN was the only configuration with a negative MBE, and the difference between Control and LSM-RUC_SFC_MYNN increased with increasing wavelength. MP-NSSL had the most-positive MBE (Figure 4e) for the 7.3 μm BTs, whereas Control had the highest MBE for 6.2 and 6.9 μm BTs. The MD for clear grid points was once again positive for the 2020 HWT forecasts (Figure 4f), indicating a dry bias in the lower troposphere. The MDs for the 2019 HWT forecasts were negative, but they were the highest in any of the three WV bands.

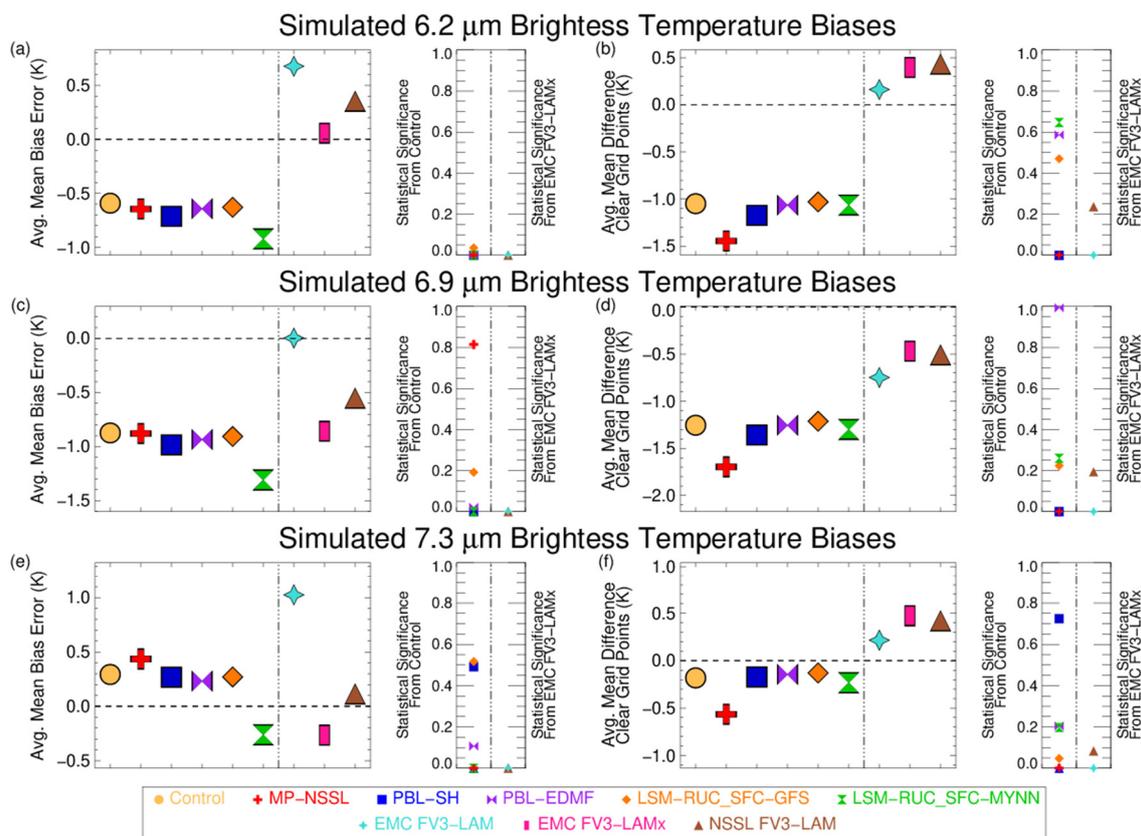


Figure 4. Average mean bias error (left: a,c,e) and mean difference for clear grid points (right: b,d,f) and statistical significance for 6.2 μm (top: a,b), 6.9 μm (middle: c,d), and 7.3 μm (bottom: e,f) brightness temperatures.

To investigate what can be inferred from the MBEs from EMC FV3-LAM, Figure 5 compares the simulated WV BTs to the average total WV mixing ratio above a certain pressure level for all grid points with a given BT, averaged over every 12 and 24 h forecast for all three WV bands. As expected, lower BTs are associated with more WV above a given pressure level, regardless of band or time of day. The positive MBE in the 6.2 μm BTs therefore indicates a dry bias in the upper troposphere whereas the 6.9 μm BTs indicate a moist bias in the middle troposphere. These results allow us to infer that a moist bias was also present in the upper tropopause in the 2019 dataset, based on the negative 6.2 μm BTs MBE. Because this moist bias occurred in all configurations in the 2019 dataset, it was possibly unrelated to which MP or PBL schemes were used, though modifications with respect to the Control configuration did influence the magnitude of the moist bias, as was similarly noted by Cintineo et al. [29]. In comparison, the 2020 HWT dataset did not have a

moist bias in the upper troposphere, which was probably related to the fact that the GFS initial conditions were used instead of the NAM.

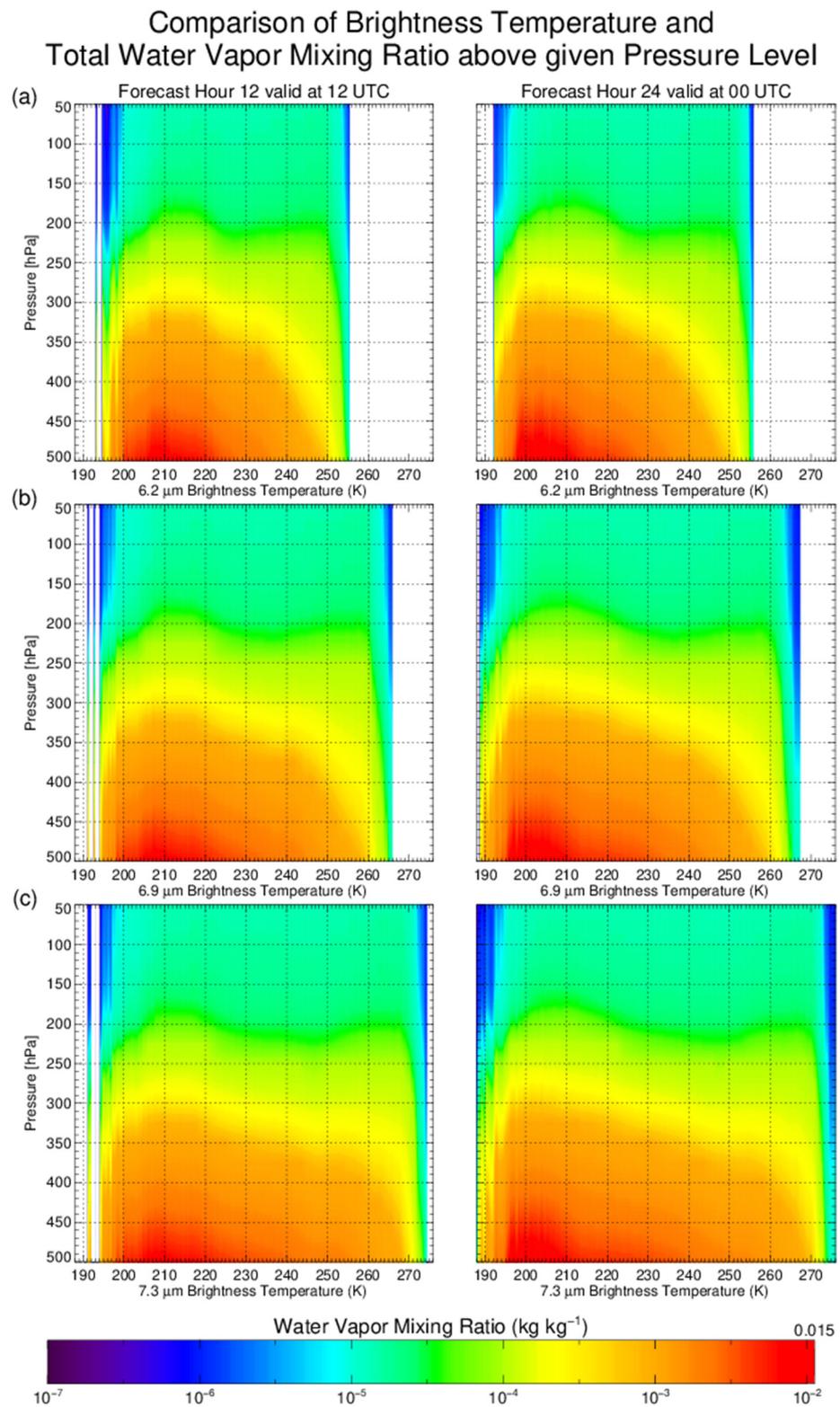


Figure 5. Comparison of the averaged total water vapor mixing ratio above a given pressure level for simulated (a) 6.2 μm , (b) 6.9 μm , and (c) 7.3 μm brightness temperatures for all EMC FV3-LAM grid points, averaged over all 12 h (left) and 24 h (right) forecasts.

The positive bias in the 7.3 μm BTs for both years indicates a dry bias in the lower troposphere. Dry biases in the lower troposphere are thought to inhibit convection [67]. For the 2019 HWT dataset, only LSM-RUC_SFC-MYNN had a moist bias and also had the highest area of upper-level clouds, as identified in Griffin et al. [23]. Both MP-NSSL and EMC FV3-LAM had the largest dry bias and also the smallest area of upper-level clouds. They also had the highest number of clear grid points, which could contribute to the dry bias. MP-NSSL and EMC FV3-LAM had the lowest MD for clear grid points, i.e., the largest negative and least-positive values, respectively. Since more intense storms can occur with greater amounts of low-level moisture [67], this moist bias in the clear grid points could be contributing to the higher number of grid points observed with the lowest BTs, especially for MP-NSSL, as shown in Figure 4e.

As mentioned previously, the simulated 6.9 μm BTs for the 2020 HWT dataset had a low bias while the 6.2 μm BTs had a high bias for clear grid points. To examine this more closely, Figure 6 presents a cumulative distribution function (CDF) of the observed and simulated 6.2 μm and 6.9 μm BTs for clear grid points for the 2020 HWT dataset. The pressure, shown on the upper x -axis, is the level where the atmospheric temperature reaches the given WV BT [5] based on pressure profiles from the EMC FV3-LAM configuration. For the 6.2 μm BTs, the CDF for the observed BTs was shifted to the left of all the configurations, consistent with the positive MBE. The opposite was observed for the 6.9 μm BTs, as the configurations had slightly lower BTs at all percentiles. Based on the corresponding atmospheric pressures, these CDFs indicate a lack of WV above approximately 350 hPa, as this is where the CDFs for the 6.2 μm BTs become more similar to the observations, and an excess of WV below 350 hPa, as this is where the CDFs for the 6.9 μm BTs begin to differ even more. This excess of WV below 350 hPa is possibly associated with the increase in cloudy grid points seen in Figure 3d, as Lu and Takle [68] found an increase in WV saturation below 350 hPa to be correlated with increased precipitation.

4.2. Brightness Temperature Differences

Brightness temperature differences (BTDs) between two satellite bands can be used to examine how well model configurations represent observed cloud properties. The first BTD used was that at 6.9–11.2 μm , which examines cloud top height [29]. The largest 6.9–11.2 μm BTDs generally occur in clear-sky regions [69], and BTDs become progressively smaller as the cloud top height increases. Due to the strong WV absorption at 6.9 μm combined with 11.2 μm BTs that generally decrease with height, these BTDs are usually negative. The second BTD used was at 8.4–11.2 μm , which discriminates between liquid and ice clouds. Absorption for water is lower at 8.4 μm compared to 11.2 μm , while the absorption for ice is higher at 8.4 μm than 11.2 μm [70–72]. Therefore, ice clouds are characterized by a positive 8.4–11.2 μm BTD, while the reverse is true for water clouds [27].

Figures 7 and 8 present two-dimensional histograms of 6.9–11.2 μm BTDs and 8.4–11.2 μm BTDs, respectively, compared to the corresponding 11.2 μm BTs. Red (blue) colors indicate a higher (lower) number of grid points for a given configuration compared to the observations. These are the only colors used in Figures 7a and 8a, as well as in Figures 7g–i and 8g–i since these figures only compare a single configuration to the observations. Figures 7b–f and 8b–f show histograms comparing the model configurations to both the Control and observations. This novel visualization approach was used to identify the impacts of the different parameterization schemes on the cloud compositions more easily, with respect both to the observations and the Control configuration, without the need for multiple figures. In these panels, the blue shading indicates that there are fewer configuration grid points and no Control grid points for a given BTD and 11.2 μm BT combination compared to the observations. Purple shading indicates that the configuration has fewer grid points at the given BTD and 11.2 μm BT combination than the observations and Control, while pink indicates that the configuration has more grid points than the observations and Control. Green indicates the configuration also has more grid points than the observations at the given BTD and 11.2 μm BT but fewer than Control, and

orange indicates the configuration has more grid points than Control but fewer than the observations. The results are shown for each model configuration during 2019 and 2020.

Brightness Temperature Distribution For Clear Grid Points

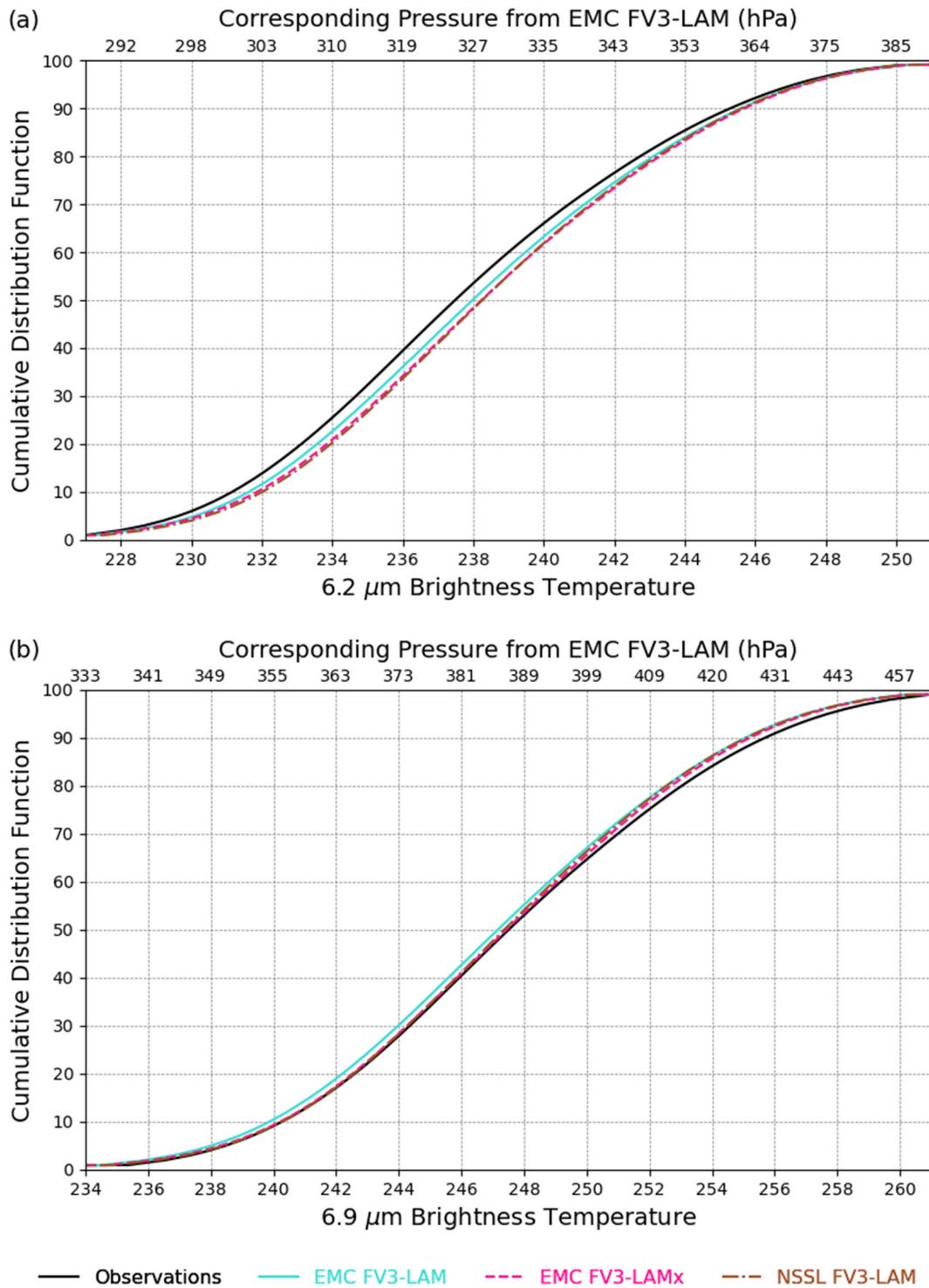


Figure 6. Cumulative distribution function of (a) 6.2 μm brightness temperatures and (b) 6.9 μm brightness temperatures for the 2020 HWT dataset. The corresponding pressure on the upper x-axis is the level where the atmospheric temperature reaches the given WV BT in the EMC FV3-LAM.

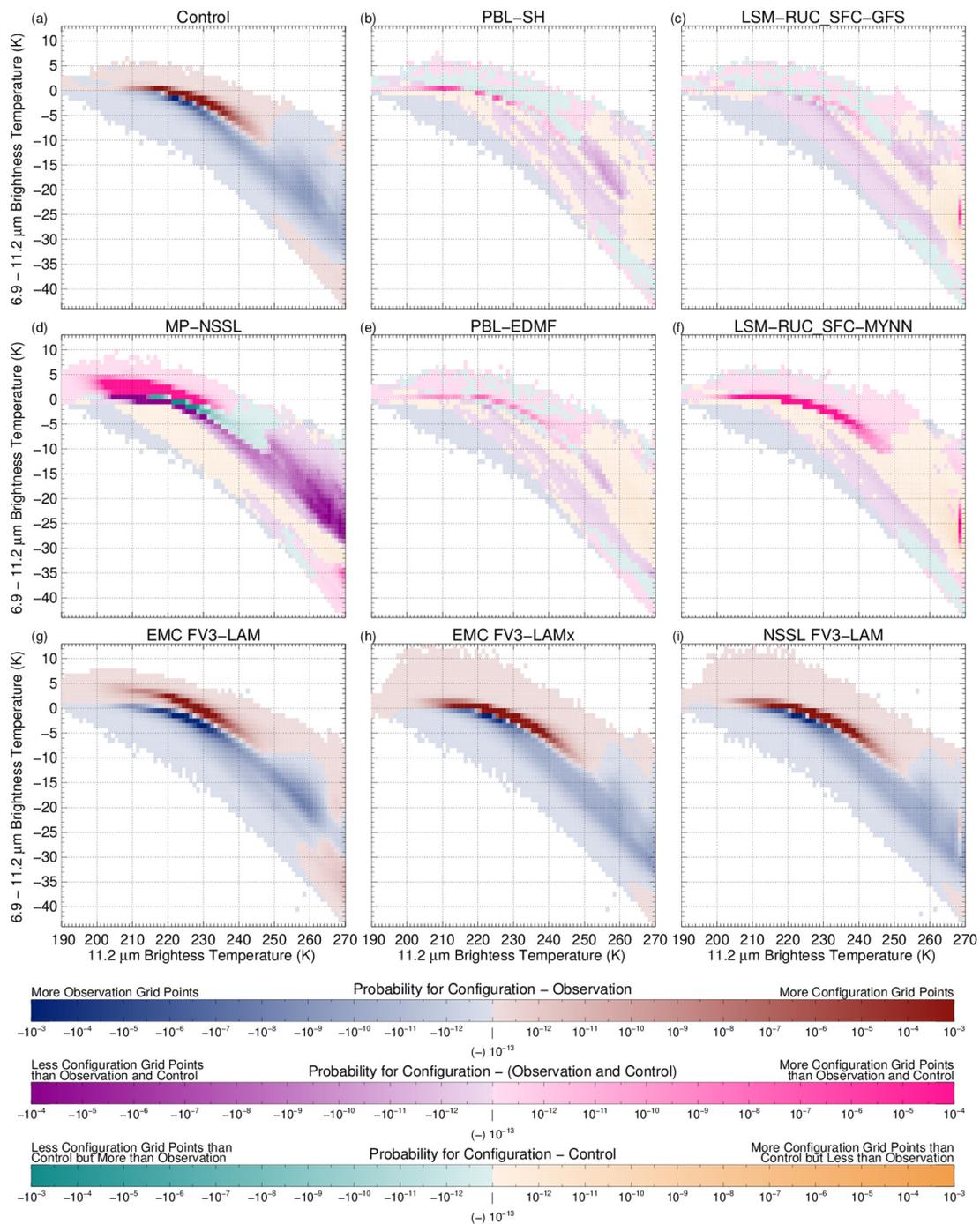


Figure 7. Histograms of GOES-16 ABI 6.9–11.2 μm brightness temperature differences plotted as a function of the ABI 11.2 μm brightness temperature (a–i). Red (blue) shading in the first color bar indicates a higher (lower) number of grid points with a given BTD and 11.2 μm BT combination for the configuration compared to the observations. Purple (pink) shading in the second color bar indicates a given configuration has fewer (more) grid points at the given BTD and 11.2 μm BT combination than both the observations and Control. In the third color bar, green indicates the configuration also has more grid points than the observations for a given combination but fewer than Control, while orange indicates that the configuration has more grid points than Control but fewer than the observations.

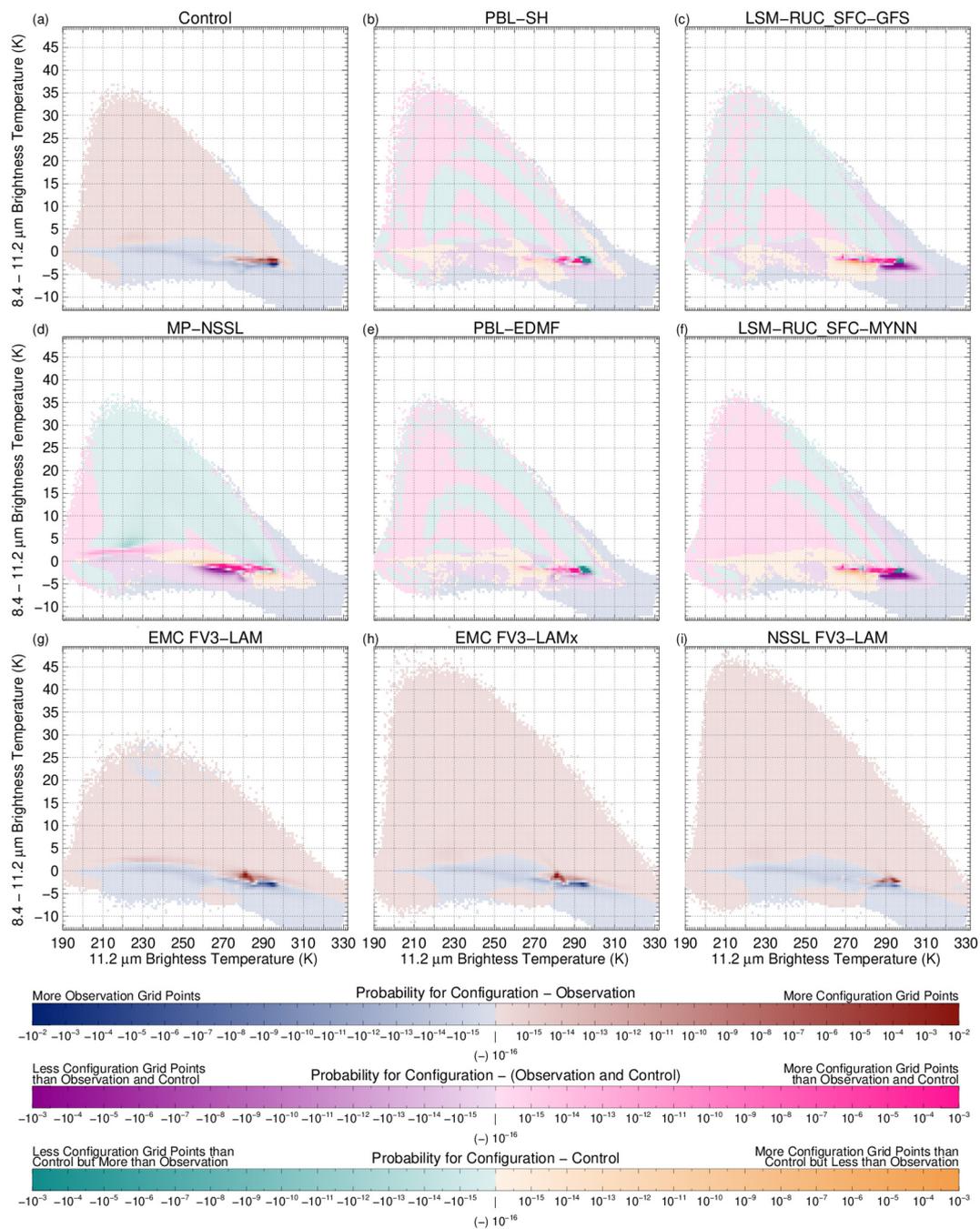


Figure 8. Histograms of GOES-16 ABI 8.4–11.2 μm brightness temperature differences plotted as a function of the ABI 11.2 μm brightness temperature (a–i). Red (blue) shading in the first color bar indicates a higher (lower) number of grid points with a given BTD and 11.2 μm BT combination for the configuration compared to the observations. Purple (pink) shading in the second color bar indicates a given configuration has fewer (more) grid points at the given BTD and 11.2 μm BT combination than both the observations and Control. In the third color bar, green indicates the configuration also has more grid points than the observations for a given combination but fewer than Control, while orange indicates that the configuration has more grid points than Control but fewer than the observations.

All configurations from both years had an increased number of grid points with a positive 6.9–11.2 μm BTD (Figure 7). This increase can be identified by the red shading for positive BTD in Figure 7a,g–i and pink shading in Figure 7b–f. A positive 6.9–11.2 μm BTD is indicative of overshooting clouds exceeding the tropopause height [73]. Using the NSSL

MP scheme instead of the Thompson scheme increased the number of grid points with a positive BT, as indicated by the pink color in Figure 7d. LSM-RUC_SFC-MYNN also had more grid points with a positive BT than Control, as shown in Figure 7f, consistent with the excess in clouds noted by Griffin et al. [23]. The number of grid points with cloud tops exceeding the tropopause height was reduced when using the Shin–Hong PBL instead of the MYNN PBL, as shown in Figure 7b by the green colors. To a lesser extent, the EDMF PBL also reduced the number of grid points with a positive BT compared to the MYNN PBL, as shown in Figure 7e. This reduction can also be seen by comparing EMC FV3-LAM in Figure 7g to EMC FV3-LAMx and NSSL FV3-LAM in Figure 7h,i. However, regardless of which configurations were used, there were still more grid points with a positive BT in the forecasts than occurred in the observations.

As seen in the two-dimensional histogram of 8.4–11.2 μm BT compared to the 11.2 μm BT in Figure 8, all configurations produced too many ice clouds compared to the observations. These additional ice clouds can be seen in the red shading for positive 8.4–11.2 μm BT for Control in Figure 8a. Changing from the Thompson MP scheme to NSSL reduced the number of ice clouds (green shading in Figure 8d), as did changing the PBL from MYNN to Shin–Hong (Figure 8b) or EDMF (Figure 8e), though to a lesser extent. The reduction in the number of ice clouds when using the hybrid EDMF instead of MYNN PBL, as well as GFDL MP scheme instead of the Thompson scheme, can also be seen in the lower number of grid points with a positive BT (red shading) for EMC FV3-LAM (Figure 8g) compared to EMC FV3-LAMx (Figure 8h) and NSSL FV3-LAM (Figure 8i). As seen in Figure 8f, LSM-RUC_SFC-MYNN had an increased number of ice clouds (pink shading) compared to Control while LSM-RUC_SFC-GFS had fewer ice clouds (green shading, Figure 8c). While it appears that using the MYNN surface layer instead of GFS increases the number of ice clouds, NSSL FV3-LAM (Figure 8i), had a lower probability of ice clouds compared to the EMC FV3-LAMx (Figure 8h). NSSL FV3-LAM uses the MYNN surface layer like LSM-RUC_SFC-MYNN but it uses the Noah LSM like Control, and EMC FV3-LAMx has the same configuration as Control. Therefore, it is possible the conjunction of RUC LSM and MYNN surface layer could result in more ice clouds.

In addition to the increased production of ice clouds, all configurations underproduced the number of water clouds compared to the observations. This lack of water clouds can be seen in the blue, purple, and orange shading for negative BTs in Figure 8. Focusing on 11.2 μm BTs between 253 K and 273 K, where aircraft icing conditions are possible [2], the Thompson MP scheme produced more water clouds compared to the NSSL (purple shading in Figure 8d). This increase was expected, as one target of the Thompson scheme is improving forecasts of aircraft icing [74]. The GFDL MP scheme with the hybrid EDMF PBL, however, produced more water clouds than the Thompson MP scheme with the MYNN PBL and the observations. This indicates that changing from the MYNN PBL increases the number of water clouds within the icing range, as does using the RUC LSM instead of Noah.

4.3. Upper-Level Jet Streams and Troughs

To determine the location and orientation of upper-level jet streams and troughs (ULJSTs), objects from the gradients in the simulated and observed 6.2 μm BTs were identified using a threshold of 2 K in MODE. ULJSTs are often co-located with gradients in WV BTs; however, current validation of these features usually relies on sparse radiosonde observations or model re-analyses. The distances between matching WV gradient objects associated with ULJSTs can be seen in Figure 9 for the 2020 HWT dataset. To be included in this analysis, observation objects had to have a matching forecast object with an interest score and area ratio greater than 0.5 in all three configurations. Figure 9a depicts the MED from every grid point considered an observation object to the closest forecast object grid point. The MED increases with forecast lead time, probably due to larger displacement errors between the centroids of the matching objects (Figure 9b). EMC FV3-LAM had a larger average MED than EMC FV3-LAMx but a similar average centroid distance. The

increased MED for EMC FV3-LAM was possibly due to smaller gradient objects (not shown), and therefore a lower percentage of observation object grid points having an overlapping EMC FV3-LAM object. For NSSL FV3-LAM, the MED and overlapping percentage was smaller than EMC FV3-LAMx but larger than EMC FV3-LAM.

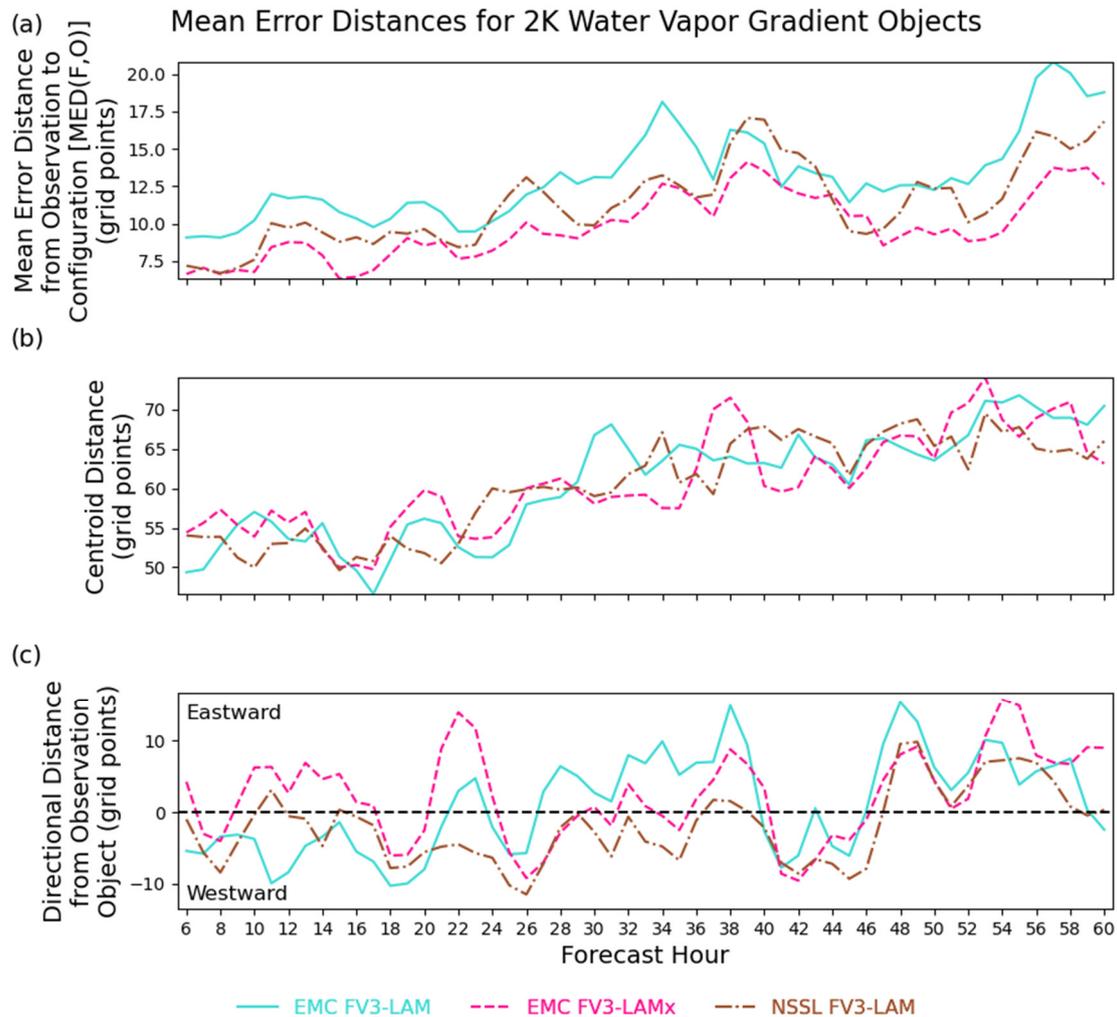


Figure 9. Time series of: (a) mean error distance (MED), (b) differences in centroid distance, and (c) zonal displacement errors between matching observation and forecast water vapor gradient objects associated with upper-level jet streams and troughs.

Figure 9c displays the west/east displacement between the configuration and observation WV gradient objects associated with ULJSTs. Especially noticeable for EMC FV3-LAM, the centers of mass for the WV gradient objects associated with ULJSTs were westward of the observation objects early in the forecast but shifted to being too far to the east later in the forecast. Therefore, the translation speed of ULJSTs appears to be faster in the configurations compared to the observations. This displacement was consistent among all configurations, which suggests that these translation errors are related to the dynamic core in the FV3-LAM rather than the various parameterization schemes being used.

To compare the progression of 500 hPa troughs, Figure 10a displays the difference in axis angle between matching forecast and observation WV gradient objects when both objects have a negative tilt, for the 2020 HWT dataset. The difference between the matching forecast and observation axis angles are mostly negative, and the overall average is also negative except for EMC FV3-LAMx. Therefore, the forecast object had a more negative tilt than the observation object. This negative tilt is possibly correlated with more convection, as evidenced by the larger number of very low 6.2 μm BTs for all configurations in Figure 3d,

and also the fact that the average area encompassed by MODE objects defined using a $10.3 \mu\text{m}$ BT of 235 K was larger than the observations for all configurations even though the average area of objects defined using 265 K was smaller (not shown). There were subtle differences in the axis angles, as EMC FV3-LAMx generally had a less-negative tilt in the matched objects compared to EMC FV3-LAM. EMC FV3-LAMx also had more area encompassed by cold cloud objects; therefore, the greater extent of these objects was not solely related to an increased negative tilt in 500 hPa troughs. Extreme differences in the axis angles around forecast hours 25 and 49 were associated with local minima in the number of matching objects, as seen in Figure 10b. These local minima were probably caused by increased cloud cover, similar to that observed in Griffin et al. [23], resulting in larger gradient objects that reduce the percentage of gradient object grid points overlapping with troughs. This diurnal cycle in total area of gradient objects can be seen in the light-blue boxes in Figure 11a. A diurnal cycle was not observed in the total area encompassed by or the number of EMC-FV3-LAM objects overlapping with troughs, as shown in in the grey boxes in Figure 11a,b, respectively. The number of EMC-FV3-LAM objects overlapping with troughs also did not decrease with forecast hour, as shown in Figure 11b. Therefore, the decrease in the number of matched objects in Figure 10b was probably due to displacement and other forecast errors increasing with time, resulting in fewer matching objects having an interest score and area ratio higher than 0.5.

Comparison of Axis Angles for Matched Objects when Objects are Negatively Tilted

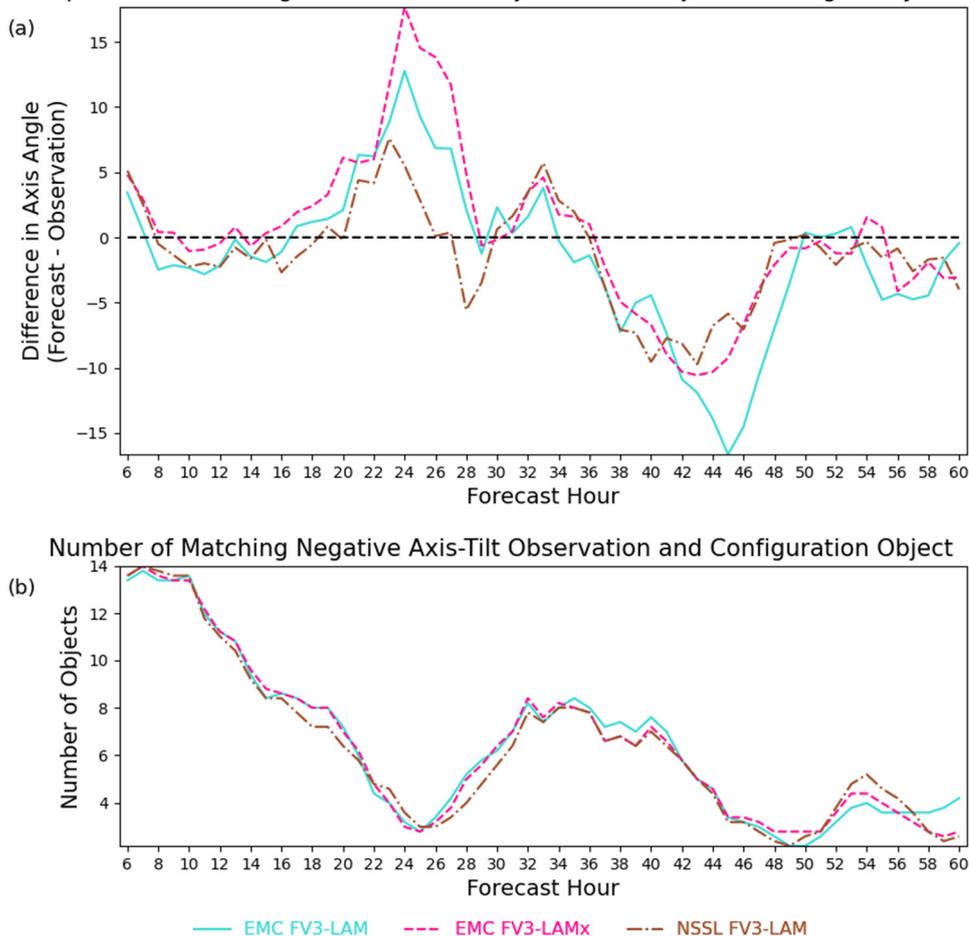


Figure 10. (a) Difference in axis angles for matching objects when both the observation and forecast objects have a negative axis angle. (b) Number of matched object pairs where both the observation and forecast objects have a negative axis angle.

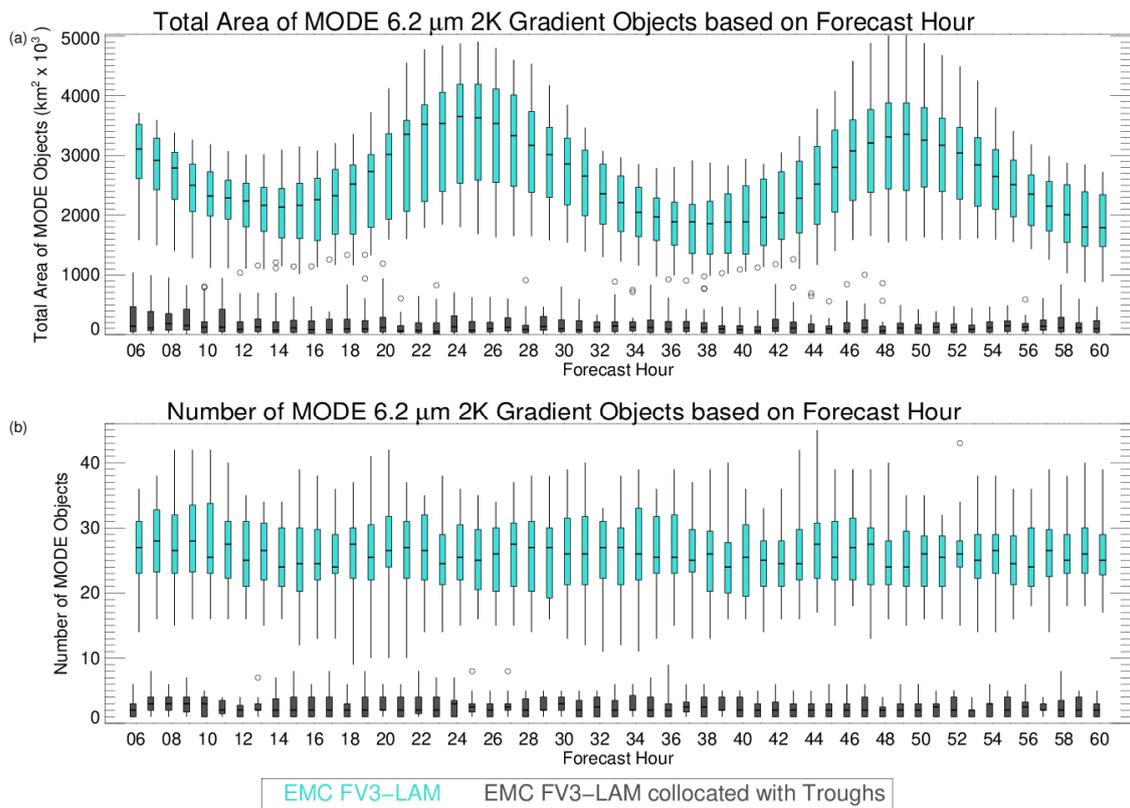


Figure 11. (a) Total area and (b) number of MODE gradient objects by forecast hour from EMC FV3-LAM (light blue) co-located with a 500 hPa trough (grey).

5. Discussion and Conclusions

In this study, simulated WV BTs from the GOES-16 ABI sensor were used to assess the impact of different parameterization schemes in the FV3-LAM on the WV and cloud fields, and on the location and orientation of ULJSTs. Nine model configurations from high-resolution simulations performed during the 2019 and 2020 HWT Spring Forecasting Experiment were evaluated. The accuracy and bias in the simulated WV BTs were assessed using MAE, MBE, and MD. Cloud composition was evaluated using BT differences. The locations of ULJSTs were evaluated by calculating the gradient in the 6.2 μm BTs and then creating MODE objects.

The impact of different MP schemes on the average water vapor brightness temperatures can be seen in Table 3. Changing the MP scheme from Thompson to NSSL in the 2019 HWT dataset increased WV in the upper atmosphere, which led to less-accurate 6.2 μm BTs over the full domain. In contrast, higher MBEs for the 6.9 and 7.3 μm BTs indicated less WV in the middle and lower troposphere, and these simulated BTs were also less accurate. The dry bias and lack of clouds in the lower troposphere when using the NSSL MP scheme is consistent with the higher MBEs in the 10.3 μm BTs found in Griffin et al. [23]. Although the NSSL MP scheme produced more overshooting clouds exceeding the tropopause height than the Thompson MP scheme, it also produced fewer ice clouds and water clouds potentially related to aircraft icing. Conversely, the GFDL MP scheme combined with the hybrid EDMF PBL produced more supercooled water clouds than the Thompson scheme with MYNN PBL in the 2020 HWT dataset, while still producing fewer ice clouds. Unfortunately, there is no direct comparison between the Thompson and GFDL MP schemes in either dataset, but since the Thompson MP and EDMF PBL also produced more of these water clouds compared to Thompson MP and MYNN in the 2019 HWT dataset, some of the additional clouds could be related to the change in PBL schemes. However, using the Thompson scheme with EDMF PBL instead of MYNN only led to a 5% increase in the number of supercooled water clouds, while the GFDL MP and

hybrid EDMF PBL showed an increase of 18%. Therefore, some of the increase in these water clouds could be due to the GFDL MP. Similarly, the decrease in ice clouds with the GFDL MP scheme and hybrid EDMF PBL was probably due to the GFDL MP scheme, since the Thompson MP and EDMF PBL configuration had a negligible reduction in ice clouds compared to the Thompson MP and MYNN PBL.

Table 3. Comparison of average mean absolute error (MAE), mean bias error (MBE), and mean difference (MD) of clear grid points between the Thompson MP scheme and the NSSL and GFDL MP schemes. Bold arrows indicate that the averages are statistically different at the 95th percentile based on Student's *t*-test.

Average MAE	NSSL	GFDL	Average MBE	NSSL	GFDL	Average MD for Clear Grid Points	NSSL	GFDL
6.2 μm	↑	↓	6.2 μm	↓	↑	6.2 μm	↓	↓
6.9 μm	↑	↓	6.9 μm	↑	↑	6.9 μm	↓	↓
7.3 μm	↑	↓	7.3 μm	↑	↑	7.3 μm	↓	↓

The reduction in ice clouds for both the NSSL and GFDL MP schemes could be due to their treatment of hail compared to the Thompson scheme. The NSSL MP scheme includes separate hail and graupel categories [50], while the GFDL MP combines graupel and hail into a single category like the Thompson scheme, but with a particle density almost twice as large [45,75]. This additional category or increase in particle density could impact the spatial coverage and composition of ice clouds due to cloud ice being considered as graupel or hail and the resultant impact on particle fall speed. The lack of ice clouds for the GFDL MP scheme with the hybrid EDMF PBL appears to result in more accurately simulated BTs for all three WV bands, even though the MBE over the full domain is the highest. The MBE is the highest because the MD for cloudy grid points is positive when using the GFDL MP and the hybrid EDMF PBL configuration but negative for the Thompson MP and MYNN PBL configuration, due to a lack of clouds with 6.2 μm BTs between 215 K and 223 K. For clear grid points, the MD is lowest for the GFDL MP and the hybrid EDMF PBL configurations, indicating an increase in WV at all levels of the troposphere when compared to the Thompson MP and MYNN PBL configuration. However, it is unclear whether the higher WV content is due to the GFDL MP scheme or the hybrid EDMF PBL, as the Thompson MP scheme with the EDMF PBL also has similarly lower BTs. Further work is needed to confirm whether the GFDL MP or the hybrid EDMF PBL scheme, or both, led to a higher WV content.

The impact of different PBL schemes on the average water vapor brightness temperatures can be seen in Table 4. Changing the PBL scheme from the MYNN to the Shin–Hong or EDMF schemes reduced the accuracy of all WV BTs in the 2019 HWT dataset. The higher MAE for the Shin–Hong PBL scheme compared to MYNN is consistent with the findings in Cintineo et al. [29], where the Thompson MP scheme and Yonsei University (YSU) PBL scheme, on which the Shin–Hong scheme is based [76], had a slightly higher MAE for 6.7 μm BTs than the Thompson MP scheme and MYNN PBL. Since the Shin–Hong and EDMF PBL schemes had more-negative MBEs compared to the MYNN scheme, this lower accuracy is possibly due to an increased amount of WV at all levels of the troposphere. This increase in WV does not appear to be related to the additional cloud cover produced by the Shin–Hong and EDMF PBL schemes [23], as the average MD for the Shin–Hong and EDMF PBL schemes for clear grid points is also lower than MYNN, except for the 7.3 μm BTs. While the MBEs increased between EMC FV3-LAM and EMC FV3-LAMx in the 2020 HWT dataset, this increase could possibly be due to the use of the GFDL MP scheme instead of the hybrid EDMF PBL scheme. Compared to the MYNN PBL, the Shin–Hong and EDMF PBL schemes also reduced the number of ice clouds and increased the number of water clouds potentially related to aircraft icing to numbers that better represented the observations. These results are consistent with those of Schwitalla et al. [77], who found

that replacing the MYNN PBL with YSU when using the Thompson MP scheme resulted in cloud liquid water that was more consistent with the observations.

Table 4. As in Table 3 but for a comparison between the MYNN PBL scheme and the Shin–Hong and EDMF PBL schemes.

Average MAE	Shin-Hong	EDMF	Average MBE	Shin-Hong	EDMF	Average MD for Clear Grid Points	Shin-Hong	EDMF
6.2 μm	↑	↑	6.2 μm	↓	↓	6.2 μm	↓	↓
6.9 μm	↑	↑	6.9 μm	↓	↓	6.9 μm	↓	↓
7.3 μm	↑	↑	7.3 μm	↓	↓	7.3 μm	↑	↑

The impact of different surface features on the average water vapor brightness temperatures can be seen in Table 5. For the configurations using different LSMs and surface layers, changing to the RUC LSM from Noah generally resulted in lower and less-accurate WV BTs over the full domain in the 2019 HWT dataset. These lower WV BTs were consistent with there being too much WV and too many clouds, especially when using the RUC LSM instead of Noah and the MYNN surface layer instead of the GFS. These additional clouds can be seen in the more frequent overshooting clouds produced by the LSM-RUC_SFC-MYNN and the fact that the MD for clear pixels was less negative compared to the full domain. However, the MD for clear grid points was still more negative when using the RUC LSM. Since Jin et al. [78] found that simulations using the RUC LSM had a higher surface air temperature than those using the Noah LSM, it is possible that these higher surface air temperatures resulted in higher atmospheric air temperatures, which would support the additional WV, based on the Clausius–Clapeyron equation. Interestingly, changing the LSM from Noah to RUC with the GFS surface layer did reduce the number of overshooting clouds and ice clouds to something better matching the observations. Furthermore, using the MYNN surface layer instead of the GFS with the Noah LSM resulted in more-positive MBEs in the 2020 HWT dataset. Therefore, the MYNN surface layer did not solely result in the additional clouds in LSM-RUC_SFC-MYNN compared to LSM-RUC_SFC-GFS. Instead, the additional clouds could be due to the combination of MYNN surface layer and the RUC LSM or the NAM initial and boundary conditions instead of the GFS.

Table 5. As in Table 3 but for a comparison between the Noah land surface model with the GFS surface layer and the RUC land surface model with the GFS or MYNN surface layer.

Average MAE	RUC GFS	RUC MYNN	Average MBE	RUC GFS	RUC MYNN	Average MD for Clear Grid Points	RUC GFS	RUC MYNN
6.2 μm	↑	↑	6.2 μm	↓	↓	6.2 μm	↑	↓
6.9 μm	↑	↑	6.9 μm	↓	↓	6.9 μm	↑	↓
7.3 μm	↑	↑	7.3 μm	↑	↓	7.3 μm	↑	↓

When analyzing the location of ULJSTs, it was found that every configuration in the 2020 HWT dataset had an increased translation speed compared to the observations, as the center of mass for the WV gradient objects associated with these features were west of the observation object early in the forecast but too far to the east later in the forecast. The axes of matching WV gradient objects associated with 500 hPa troughs were also more negative, when both observation and forecast objects had a negative tilt. However, while the location and orientation of these synoptic-scale features differed among the model configurations, there was no consistently optimal configuration.

Since the ULJST analysis required upper-level forecast information from at least one configuration to be accomplished, future work on this project will include expanding the identification of WV gradient objects associated with ULJSTs, so that the upper-level

forecast information will not be necessary. Identifying objects associated with ULJSTs may be accomplished by increasing the threshold used for defining gradient objects, looking at other object features such as length compared to width, or restricting objects to those north of a given latitude. Additional work on this project will also look at analyzing ULJSTs for different seasons. An important aspect of the object-based verification using satellite WV band BTs is that it could be applied globally and serve as an important verification method for synoptic-scale features in regions lacking conventional observations. Additional future work will also assess the impact of the GFDL MP scheme on simulated BTs compared to the Thompson MP scheme, without the confounding influence of changes in the PBL scheme.

Author Contributions: Conceptualization, S.M.G. and J.A.O.; methodology, S.M.G.; formal analysis, S.M.G.; investigation, S.M.G.; writing—original draft preparation, S.M.G.; writing—review and editing, J.A.O.; funding acquisition, J.A.O. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this project was provided by the NOAA Joint Technology Transfer Initiative (JTTI) program via grants NA19OAR4590233 and NA21OAR4590168.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Available from the first author on request.

Acknowledgments: The authors would like to thank John Halley-Gotway and Eric Gilleland for their help on understanding the MED in MET. Thanks also to Timothy Supinie, Ming Xue, and Sharon Nebuda for the data used in this analysis.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Smith, W.L., Jr.; Minnis, P.; Young, D.F. An icing product derived from operational satellite data. In Proceedings of the Ninth Conference on Aviation, Range and Aerospace Meteorology, Orlando, FL, USA, 13 September 2000; pp. 256–259.
2. Ellrod, G.; Bailey, A.P. Assessment of aircraft icing potential and maximum icing altitude from geostationary meteorological satellite data. *Weather Forecast.* **2007**, *22*, 160–174. [[CrossRef](#)]
3. Scofield, R.; Vicente, G.; Hodges, M. *The Use of Water Vapor for Detecting Environments That Lead to Convectively Produced Heavy Precipitation and Flash Floods*; NOAA Technical Reports NESDIS 99; US Department of Commerce: Washington, DC, USA, 2000; 64p.
4. Velden, C.S.; Hayden, C.M.; Nieman, S.; Menzel, W.P.; Wanzong, S.; Goerss, J. Upper-tropospheric winds derived from geostationary satellite water vapor observations. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 173–195. [[CrossRef](#)]
5. Laurent, H. Wind extraction from Meteosat water vapor channel image data. *J. Appl. Meteorol. Climatol.* **1993**, *32*, 1124–1133. [[CrossRef](#)]
6. Oyama, R. Relationship between Tropical Cyclone Intensification and Cloud-Top Outflow Revealed by Upper-Tropospheric Atmospheric Motion Vectors. *J. Appl. Meteorol. Climatol.* **2017**, *56*, 2801–2819. [[CrossRef](#)]
7. Ramond, D.; Corbin, H.; Desbois, M.; Szejwach, G.; Waldteufel, P. The Dynamics of Polar Jet Streams as Depicted by the METEOSAT WV Channel Radiance Field. *Mon. Weather Rev.* **1981**, *109*, 2164–2176. [[CrossRef](#)]
8. Velden, C.S. Satellite observations of Hurricane Elena (1985) using the VAS 6.7 μm “water vapor” channel. *Bull. Am. Meteorol. Soc.* **1987**, *68*, 210–215. [[CrossRef](#)]
9. Muller, B.M.; Fuelberg, H.E. A Simulation and Diagnostic Study of Water Vapor Image Dry Bands. *Mon. Weather Rev.* **1990**, *118*, 705–722. [[CrossRef](#)]
10. Sharman, R.D.; Trier, S.B.; Lane, T.P.; Doyle, J.D. Sources and dynamics of turbulence in the upper troposphere and lower stratosphere: A review. *Geophys. Res. Lett.* **2012**, *39*. [[CrossRef](#)]
11. Rose, S.F.; Hobbs, P.V.; Locatelli, J.D.; Stoelinga, M.T. A 10-Yr Climatology Relating the Locations of Reported Tornadoes to the Quadrants of Upper-Level Jet Streaks. *Weather Forecast.* **2004**, *19*, 301–309. [[CrossRef](#)]
12. Hurlbut, M.M.; Cohen, A.E. Environments of Northeast U.S. Severe Thunderstorm Events from 1999 to 2009. *Mon. Weather Rev.* **2014**, *29*, 3–22. [[CrossRef](#)]
13. Macdonald, N.J. On the Apparent Relationship between Convective Activity and the Shape of 500 mb Troughs. *Mon. Weather Rev.* **1976**, *104*, 1618–1622. [[CrossRef](#)]
14. Cutraro, F.; Galligani, V.S.; García Skabar, Y. Evaluation of synthetic satellite images computed from radiative transfer models over a region of South America using WRF and GOES-13/16 observations. *Q. J. R. Meteorol. Soc.* **2021**, *147*, 2988–3003. [[CrossRef](#)]

15. Feltz, W.F.; Bedka, K.M.; Otkin, J.A.; Greenwald, T.; Ackerman, S.A. Understanding satellite-observed mountain wave signatures using high-resolution numerical model data. *Weather Forecast.* **2009**, *24*, 76–86. [[CrossRef](#)]
16. Geiss, S.; Scheck, L.; de Lozar, A.; Weissmann, M. Understanding the model representation of clouds based on visible and infrared satellite observations. *Atmos. Chem. Phys.* **2021**, *21*, 12273–12290. [[CrossRef](#)]
17. Grasso, L.D.; Greenwald, T. Analysis of 10.7- μm brightness temperatures of a simulated thunderstorm with two-moment microphysics. *Mon. Weather Rev.* **2004**, *132*, 815–825. [[CrossRef](#)]
18. Grasso, L.D.; Sengupta, M.; Dostalek, J.F.; Brummer, R.; DeMaria, M. Synthetic satellite imagery for current and future environmental satellites. *Int. J. Remote Sens.* **2008**, *29*, 4373–4384. [[CrossRef](#)]
19. Grasso, L.D.; Lindsey, D.T.; Sunny Lim, K.-S.; Clark, A.J.; Bikos, D.; Dembek, S.R. Evaluation of and suggested improvements to the WSM6 microphysics in WRF-ARW using synthetic and observed GOES-13 imagery. *Mon. Weather Rev.* **2004**, *142*, 3635–3650. [[CrossRef](#)]
20. Griffin, S.M.; Otkin, J.A.; Rozoff, C.M.; Sieglaff, J.M.; Cronics, L.M.; Alexander, C.R. Methods for comparing simulated and observed satellite infrared brightness temperatures and what do they tell us? *Weather Forecast.* **2017**, *32*, 5–25. [[CrossRef](#)]
21. Griffin, S.M.; Otkin, J.A.; Rozoff, C.M.; Sieglaff, J.M.; Cronics, L.M.; Alexander, C.R.; Jensen, T.R.; Wolff, J.K. Seasonal analysis of cloud objects in the High-Resolution Rapid Refresh (HRRR) model using object-based verification. *J. Appl. Meteorol. Climatol.* **2017**, *56*, 2317–2334. [[CrossRef](#)]
22. Griffin, S.M.; Otkin, J.A.; Thompson, G.; Frediani, M.; Berner, J.; Kong, F. Assessing the Impact of Stochastic Perturbations in Cloud Microphysics using GOES-16 Infrared Brightness Temperatures. *Mon. Weather Rev.* **2020**, *148*, 3111–3137. [[CrossRef](#)]
23. Griffin, S.M.; Otkin, J.A.; Nebude, S.E.; Jensen, T.L.; Skinner, P.S.; Gilleland, E.; Supinie, T.A.; Xu, M. Evaluating the impact of planetary boundary layer, land surface model, and microphysics parameterization schemes on cold cloud objects in simulated GOES-16 brightness temperatures. *J. Geophys. Res. Atmos.* **2021**, *126*, e2021JD034709. [[CrossRef](#)]
24. Henderson, D.S.; Otkin, J.A.; Mecikalski, J.R. Evaluating Convective Initiation in High-Resolution Numerical Weather Prediction Models Using GOES-16 Infrared Brightness Temperatures. *Mon. Weather Rev.* **2021**, *149*, 1153–1172. [[CrossRef](#)]
25. Jones, T.A.; Skinner, P.; Knopfmeier, K.; Mansell, E.; Minnis, P.; Palikonda, R.; Smith, W., Jr. Comparison of Cloud Microphysics Schemes in a Warn-on-Forecast System Using Synthetic Satellite Objects. *Weather Forecast.* **2018**, *33*, 1681–1708. [[CrossRef](#)]
26. Otkin, J.A.; Greenwald, T.J. Comparison of WRF model-simulated and MODIS-derived cloud data. *Mon. Weather Rev.* **2008**, *136*, 1957–1970. [[CrossRef](#)]
27. Otkin, J.A.; Greenwald, T.J.; Sieglaff, J.; Huang, H.-L. Validation of a large-scale simulated brightness temperature dataset using SEVIRI satellite observations. *J. Appl. Meteorol. Climatol.* **2009**, *48*, 1613–1626. [[CrossRef](#)]
28. Thompson, G.; Tewari, M.; Ikeda, K.; Tessendorf, S.; Weeks, C.; Otkin, J.A.; Kong, F. Explicitly-coupled cloud physics and radiation parameterizations and subsequent evaluation in WRF high-resolution convective forecasts. *Atmos. Res.* **2016**, *168*, 92–104. [[CrossRef](#)]
29. Cintineo, R.; Otkin, J.A.; Kong, F.; Xue, M. Evaluating the accuracy of planetary boundary layer and cloud microphysical parameterization schemes in a convection-permitting ensemble using synthetic GOES-13 satellite observations. *Mon. Weather Rev.* **2014**, *142*, 107–124. [[CrossRef](#)]
30. Soden, B.J.; Bretherton, F.P. Evaluation of water vapor distribution in general circulation models using satellite observations. *J. Geophys. Res. Atmos.* **1994**, *99*, 1187–1210. [[CrossRef](#)]
31. Zeng, X.; Tao, W.K.; Lang, S.; Hou, A.Y.; Zhang, M.; Simpson, J. On the sensitivity of atmospheric ensemble states to cloud microphysics in long-term cloud-resolving model simulations. *J. Meteorol. Soc. Jpn.* **2008**, *86A*, 45–65. [[CrossRef](#)]
32. Chung, K.-S.; Chiu, H.-J.; Liu, C.-Y.; Lin, M.-Y. Satellite Observation for Evaluating Cloud Properties of the Microphysical Schemes in Weather Research and Forecasting Simulation: A Case Study of the Mei-Yu Front Precipitation System. *Remote Sens.* **2020**, *12*, 3060. [[CrossRef](#)]
33. Schmit, T.J.; Griffith, P.; Gunshor, M.M.; Daniels, J.M.; Goodman, S.J.; Lebar, W.J. A closer look at the ABI on the GOES-R series. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 681–698. [[CrossRef](#)]
34. Krishnamurti, T.N.; Rajendran, K.; Kumar, T.S.V.; Lord, S.; Toth, Z.; Zou, X.; Cocke, S.; Ahlquist, J.E.; Navon, I.M. Improved Skill for the Anomaly Correlation of Geopotential Heights at 500 hPa. *Mon. Weather Rev.* **2003**, *131*, 1082–1102. [[CrossRef](#)]
35. Elmore, K.L.; Baldwin, M.E.; Schultz, D.M. Field Significance Revisited: Spatial Bias Errors in Forecasts as Applied to the Eta Model. *Mon. Weather Rev.* **2006**, *134*, 519–531. [[CrossRef](#)]
36. Bikos, D.; Lindsey, D.T.; Otkin, J.; Sieglaff, J.; Grasso, L.; Siewert, C.; Correia, J., Jr.; Coniglio, M.; Rabin, R.; Kain, J.S.; et al. Synthetic satellite imagery for real-time high-resolution model evaluation. *Weather Forecast.* **2012**, *27*, 784–795. [[CrossRef](#)]
37. Ellrod, G.P. A water vapor image feature related to severe thunderstorms. *Natl. Weather Dig.* **1990**, *15*, 21–29.
38. Santurette, P.; Georgiev, C.G. Water vapour imagery analysis in 7. In $3\ \mu/6.2\ \mu$ for diagnosing thermo-dynamic context of intense convection. In Proceedings of the Joint 2007 EUMETSAT Meteorological Satellite Conference and the 15th AMS Satellite Meteorology & Oceanography Conference, Amsterdam, The Netherlands, 24–28 September 2007.
39. Mohan, P.M.; Srinivas, C.V.; Yesubabu, V.; Baskaran, R.; Venkatraman, B. Simulation of a heavy rainfall event over Chennai in Southeast India using WRF: Sensitivity to microphysics parameterization. *Atmos. Res.* **2018**, *210*, 83–99. [[CrossRef](#)]
40. Lin, S. A “Vertically Lagrangian” Finite-Volume Dynamical Core for Global Models. *Mon. Weather Rev.* **2004**, *132*, 2293–2307. [[CrossRef](#)]
41. Putman, W.M.; Lin, S.-J. Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.* **2007**, *227*, 55–78. [[CrossRef](#)]

42. Gallo, B.T.; Clark, A.J.; Jirak, I.; Kain, J.S.; Weiss, S.J.; Coniglio, M.; Knopfmeier, K.; Correia, J.; Melick, C.J., Jr.; Karstens, C.D.; et al. Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Weather Forecast.* **2017**, *32*, 1541–1568. [[CrossRef](#)]
43. Clark, A.J.; Jirak, I.L.; Gallo, B.T.; Roberts, B.; Knopfmeier, K.H.; Clark, R.A.; Vancil, J.; Dean, A.R.; Hoogewind, K.A.; Heinselman, P.L.; et al. A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bull. Am. Meteorol. Soc.* **2020**, *101*, E2022–E2024.
44. Thompson, G.; Rasmussen, R.M.; Manning, K. Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part I: Description and Sensitivity Analysis. *Mon. Weather Rev.* **2004**, *132*, 519–542. [[CrossRef](#)]
45. Thompson, G.; Field, P.R.; Rasmussen, R.M.; Hall, W.D. Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Mon. Weather Rev.* **2008**, *136*, 5095–5115. [[CrossRef](#)]
46. Nakanishi, M.; Niino, H. An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteorol.* **2004**, *112*, 1–31. [[CrossRef](#)]
47. Nakanishi, M.; Niino, H. Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteorol. Soc. Jpn.* **2009**, *87*, 895–912. [[CrossRef](#)]
48. Mitchell, K. *The Community Noah Land-Surface Model. User's Guide Public Release Version 2.7.1*; NOAA/NCEP; Environmental Modeling Center: College Park, MD, USA, 2005; 26p.
49. Niu, G.-Y.; Yang, Z.-L.; Mitchell, K.E.; Chen, F.; Ek, M.B.; Barlage, M.; Kumar, A.; Manning, K.; Niyogi, D.; Rosero, E.; et al. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.* **2011**, *116*. [[CrossRef](#)]
50. Mansell, E.R.; Ziegler, C.L.; Bruning, E.C. Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.* **2010**, *67*, 171–194. [[CrossRef](#)]
51. Shin, H.H.; Hong, S.-Y. Analysis of resolved and parameterized vertical transports in convective boundary layers at gray-zone resolutions. *J. Atmos. Sci.* **2013**, *70*, 3248–3261. [[CrossRef](#)]
52. Han, J.; Witek, M.L.; Teixeira, J.; Sun, R.; Pan, H.; Fletcher, J.K.; Bretherton, C.S. Implementation in the NCEP GFS of a Hybrid Eddy-Diffusivity Mass-Flux (EDMF) Boundary Layer Parameterization with Dissipative Heating and Modified Stable Boundary Layer Mixing. *Weather Forecast.* **2016**, *31*, 341–352. [[CrossRef](#)]
53. Smirnova, T.G.; Brown, J.M.; Benjamin, S.G.; Kenyon, J.S. Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) Available in the Weather Research and Forecasting (WRF) Model. *Mon. Weather Rev.* **2016**, *144*, 1851–1865. [[CrossRef](#)]
54. Clough, S.; Shephard, M.; Mlawer, E.; Delamere, J.; Iacono, M.; Cady-Pereira, K.; Boukabara, S.; Brown, P.D. Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Radiat. Transf.* **2005**, *91*, 233–244. [[CrossRef](#)]
55. Zhou, L.; Lin, S.; Chen, J.; Harris, L.M.; Chen, X.; Rees, S.L. Toward Convective-Scale Prediction within the Next Generation Global Prediction System. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 1225–1243. [[CrossRef](#)]
56. Han, Y.; van Delst, P.; Liu, Q.; Weng, F.; Yan, B.; Treadon, R.; Derber, J. *JCSDA Community Radiative Transfer Model (CRTM) Version 1*; NOAA Technical Report 122; The National Environmental Satellite, Data, and Information Service (NESDIS): Silver Spring, MD, USA, 2005.
57. Ding, S.; Yang, P.; Weng, F.; Liu, Q.; Han, Y.; Van Delst, P.; Li, J.; Baum, B. Validation of the community radiative transfer model. *J. Quant. Spectrosc. Radiat. Transf.* **2011**, *112*, 1050–1064. [[CrossRef](#)]
58. Otkin, J.A.; Posselt, D.J.; Olson, E.R.; Huang, H.-L.; Davies, J.E.; Li, J.; Velden, C.S. Mesoscale numerical weather prediction models used in support of infrared hyperspectral measurement simulation and product algorithm development. *J. Atmos. Ocean. Technol.* **2007**, *24*, 585–601. [[CrossRef](#)]
59. Willmott, C.J.; Johnson, M.L. Resolution errors associated with gridded precipitation fields. *Int. J. Climatol.* **2005**, *25*, 1957–1963. [[CrossRef](#)]
60. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]
61. Davis, C.A.; Brown, B.G.; Bullock, R.G. Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Weather Rev.* **2006**, *134*, 1772–1784. [[CrossRef](#)]
62. Davis, C.A.; Brown, B.G.; Bullock, R.G. Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Weather Rev.* **2006**, *134*, 1785–1795. [[CrossRef](#)]
63. Davis, C.A.; Brown, B.G.; Bullock, R.G.; Gotway, J.H. The Method for Object-based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program. *Weather Forecast.* **2009**, *24*, 1252–1267. [[CrossRef](#)]
64. Bullock, R.G.; Brown, B.G.; Fowler, T.L. *Method for Object-Based Diagnostic Evaluation*; No. NCAR/TN-532+STR; National Center for Atmospheric Research: Boulder, CO, USA, 2016.
65. Gilleland, E. A new characterization in the spatial verification framework for false alarms, misses, and overall patterns. *Weather Forecast.* **2017**, *32*, 187–198. [[CrossRef](#)]
66. Otkin, J.A.; Lewis, W.E.; Lenzen, A.; McNoldy, B.; Majumdar, S. Assessing the accuracy of the cloud and water vapor fields in the Hurricane WRF (HWRF) model using satellite infrared brightness temperatures. *Mon. Weather Rev.* **2017**, *145*, 2027–2046. [[CrossRef](#)]

67. Sherwood, S.C.; Roca, R.; Weckwerth, T.M.; Andronova, N.G. Tropospheric water vapor, convection, and climate. *Rev. Geophys.* **2010**, *48*, RG2001. [[CrossRef](#)]
68. Lu, E.; Takle, E. Contributions of Water Vapor and Temperature to the Interannual Variability of Precipitation: An Evaluation from North American Regional Reanalysis. In Proceedings of the 20th Conference on Climate Variability and Change, New Orleans, LA, USA, 23 January 2008.
69. Mecikalski, J.R.; Bedka, K.M. Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery. *Mon. Weather Rev.* **2006**, *134*, 49–78. [[CrossRef](#)]
70. Ackerman, S.A.; Smith, W.L.; Spinhirne, J.D.; Revercomb, H.E. The 27–28 October 1986 FIRE IFO Cirrus Case Study: Spectral properties of cirrus clouds in the 8–12 μm window. *Mon. Weather Rev.* **1990**, *118*, 2377–2388. [[CrossRef](#)]
71. Strabala, K.I.; Ackerman, S.A.; Menzel, W.P. Cloud properties inferred from 8–12-mm data. *J. Appl. Meteorol. Climatol.* **1994**, *33*, 212–229. [[CrossRef](#)]
72. Baum, B.A.; Soulen, P.F.; Strabala, K.I.; King, M.D.; Ackerman, S.A.; Menzel, W.P.; Yang, P. Remote sensing of cloud properties using MODIS Airborne Simulator imagery during SUCCESS. II. Cloud thermodynamic phase. *J. Geophys. Res. Atmos.* **2000**, *105*, 11781–11792. [[CrossRef](#)]
73. Schmetz, J.; Tjemkes, S.A.; Gube, M.; van de Berg, L. Monitoring deep convection and convective overshooting with METEOSAT. *Adv. Space Res.* **1997**, *19*, 433–441. [[CrossRef](#)]
74. Thompson, G.; Politovich, M.K.; Rasmussen, R.M. A numerical weather model's ability to predict characteristics of aircraft icing environments. *Weather Forecast.* **2017**, *32*, 207–221. [[CrossRef](#)]
75. Lin, Y.-L.; Farley, R.D.; Orville, H.D. Bulk parameterization of the snow field in a cloud model. *J. Appl. Meteorol. Climatol.* **1983**, *22*, 1065–1092. [[CrossRef](#)]
76. Huang, M.; Gao, Z.; Miao, S.; Chen, F. Sensitivity of urban boundary layer simulation to urban canopy models and PBL schemes in Beijing. *Meteorol. Atmos. Phys.* **2019**, *131*, 1235–1248. [[CrossRef](#)]
77. Schwitalla, T.; Branch, O.; Wulfmeyer, V. Sensitivity study of the planetary boundary layer and microphysical schemes to the initialization of convection over the Arabian Peninsula. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 846–869. [[CrossRef](#)]
78. Jin, J.; Miller, N.L.; Schegel, N. Sensitivity study of four land surface schemes in the WRF model. *Adv. Meteorol.* **2010**, 167436. [[CrossRef](#)]