DTC-visitor-project-report

Shaowu Bao

01/30/2022

Contents

1. Abstract

The Hurricane Analysis and Forecast System (HAFS) model is an effort under the Next Generation Global Prediction System and Unified Forecast System initiatives to create the next generation of hurricane prediction-and-analysis system based on the Finite Volume Cubic Sphere (FV3) Global Forecast System (GFS). It has been validated extensively using traditional verification indicators such as tracker error and biases, intensity error and biases, and the radii of gale, damaging and hurricane strength winds. Satellite images have been used to verify hurricane model forecasts, but not on HAFS. The community radiative transfer model (CRTM) is used to generate model synthetic satellite images from HAFS model forecast state variables. The multiple forecast snapshots in the whole, early, mature, and decay stages of hurricanes Dorian in 2019 and Teddy and Laura in 2020 are used to generate composite model synthetic Geostationary Operational Environmental Satellite R Series (GOES-R) infrared brightness images. These composite synthetic images are compared to the corresponding observed composite images to evaluate the model forecast Tropical Cyclone (TC) vortex intensity, size, and asymmetric structures.

Results show that the HAFS forecasts agree reasonably well with the observation, but the forecast intensity is weaker, its overall vortex size smaller, and the radii of its eye and maximum winds are larger than the observed. Also revealed by the evaluation is that when the Hurricane Weather Research Forecast (HWRF) physics suite was used to replace the GFS physics suite, the HAFS model simulated significantly larger vortices. Future research will be required to determine the reason for this distinct difference. The evaluation results are considered useful by model developers for further model improvement. While these results are consistent with those obtained by traditional verification methods, evaluations based on composite satellite images provide an additional benefit with richer information because they have near-real-time spatially and temporally continuous high-resolution data with global coverage. Composite satellite infrared images could be used routinely to supplement traditional verification methods in the HAFS and other hurricane model evaluations. Note since this study only evaluated three

hurricanes, caution should be exercised to extrapolate these conclusions to expect model biases in predicting other TCs. Nonetheless, the consistency of the evaluation using composite satellite images and the traditional metrics indicates that this method has the potential to be applied to other storms in future studies.

2. Introduction

Tropical cyclone (TC) forecasts are critical for mitigating damage to coastal communities. The Hurricane Weather Research Forecast (HWRF) model is currently the operational TC forecast model in the United States. It is based on the Non-Hydrostatic Mesoscale Model (NMM) core of the Weather Research and Forecast (WRF) model and has a physical suite suitable for the TC process. As part of the National Weather Service's (NWS) Next Generation Global Prediction System (NGGPS), the Global Forecast System (GFS) underwent a significant upgrade in 2019 to incorporate a new non-hydrostatic dynamic core known as the Finite Volume Cubic Sphere (FV3) developed by the National Oceanic and Atmospheric Administration's (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL). NOAA developed the Hurricane Analysis and Forecast System (HAFS) as the next-generation hurricane forecasting system based on the new FV3 dynamic core to advance TC forecasts within the Unified Forecasting System's (UFS) unified global and regional modeling framework. The global nested FV3 outperformed the Spectral GFS in terms of track and intensity prediction during the 2017 hurricane season (Hazelton et al., 2021). However, the intensity prediction performance is not as good as that of the operational HWRF. Recently, the HAFS-SAR model was implemented on a regional grid with external lateral boundary conditions, named Standalone Regional HAFS. HAFS-SAR is designed to cover a single large area in the North Atlantic Basin, with an improved planetary boundary layer (PBL), a surface flux parameterization scheme for TC, and a three-kilometer grid spacing to account for convection.

The HAFS model based on FV3 has been extensively evaluated and found to have superior track capabilities compared to several operational hurricane models, as well as comparable intensity and structure capabilities (Dong et al., 2020; Hazelton et al., 2021). Most of these assessments relied on traditional metrics, such as mean track and intensity errors and biases calculated using the maximum wind speed and the radii of the critical gale (34 kt), damaging (50 kt) and hurricane (64 kt) strength winds. Along with these traditional TC forecast metrics, Dong et al. (2020) and Hazelton et al. (2021) examined the structure of forecast TCs using several radar reflectivity images. Although these radar images can provide detailed information about cloud and hydrometeor fields, their limited spatial coverage and insufficient sampling prevent them from being used as a routine model evaluation tool.

The Advanced Baseline Imager (ABI) sensors onboard NOAA's Geostationary Operational Environmental Satellite (GOES) generate full-disk global images of the Earth every 15 minutes or every 5 minutes in continuous-disk mode, with a spatial resolution of 0.5–2 kilometers. These continuous and high-resolution images provide valuable information about the spatial and temporal distribution of clouds and water vapor and can be used to evaluate routine weather models. The regions within and surrounding TCs frequently lack in-situ measurements, making satellite observations particularly valuable in these data-poor regions because they can fill in

many gaps by covering large areas, including open oceans, with high resolution. In recent years, these GOES satellites' (GOES-13, GOES-16, and GOES-17) infrared brightness temperature (BT) images have been used to evaluate and understand the structural evolution of forecast TCs generated by WRF, HWRF, COAMPS, and GFS and other models (Bao, 2018; Bao et al., 2020; Cintineo et al., 2014; Jin et al., 2014; Novak & Bao, 2019; Otkin et al., 2017). These studies employ a "model-to-satellite" methodology (Otkin et al., 2017), using a forward radiative transfer model to convert model-simulated fields such as temperature, water vapor, and cloud mixing ratios to synthetic GOES-R infrared (IR) BT images.

In this report, we present the results of such an evaluation because GOES-R BT data have not previously been used to assess the TC structure and intensity of HAFS forecasts. Rather than comparing observed and model synthetic images obtained for individual forecast snapshots, we will compare composite images by superimposing all available observation and model images. In this way, we hope to eliminate the impact of random forecast errors on the evaluation so we can focus on identifying the HAFS model's systematic biases. We aim to answer the following three research questions: (1) Can the observed and synthetic images reveal the systematic biases of the HAFS model in its predictions of TC intensity, size, and structure? (2) Can the evaluation results corroborate those obtained using traditional evaluation methods? (3) Can the GOES-R BT images provide additional benefits that are not available from the traditional metrics? If the answers to all these questions are affirmative, we can propose that GOES-R BT images be added as a routine model evaluation tool in addition to the traditional ones, because satellite BT images provide rich and temporally and spatially continuous coverage. This work also compared the HAFS and the HWRF models to understand their systematic biases.

Note that since this study only evaluated three hurricanes, with some in-depth analyses (radii, intensity, size, and asymmetry etc.) conducted only on one of them, caution should be exercised when extrapolating the conclusions to anticipate model biases in predicting other TCs. Future studies using larger sized case samples may be needed to get more generalized conclusions.
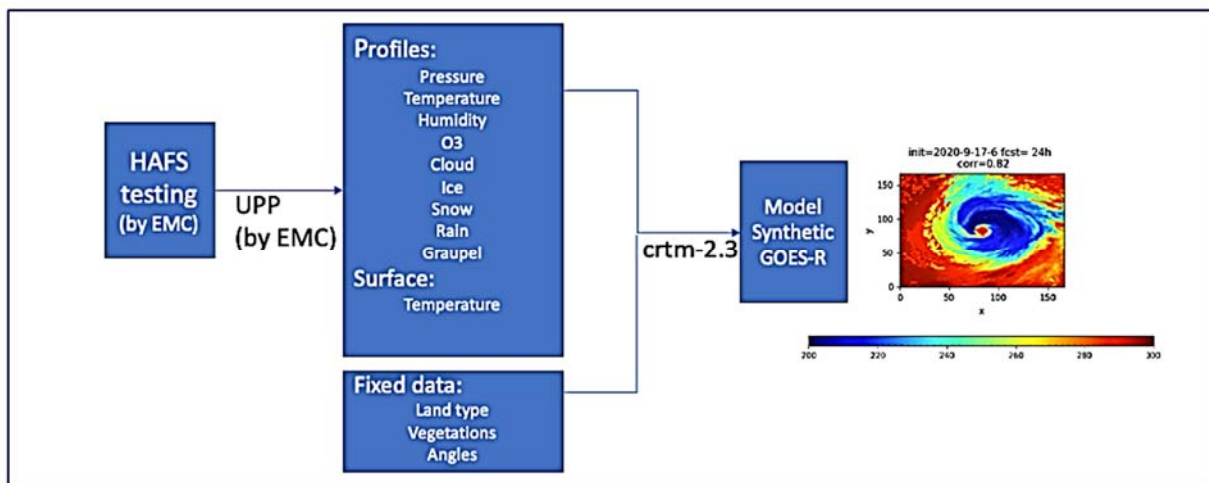

Figure 1 diagram of the method to create model synthetic satellite images

3. Data sources

At the Environmental Modeling Center (EMC), several high-impact hurricanes, including Hurricane Dorian in 2019, Teddy and Laura in 2020, were used to test the HAFS model version 0.2A. Figure 1 shows the procedures to generate synthetic satellite images. The native output files of these tests are first post-processed using Unified Post-Processing System (UPP, at *https://dtcenter.org/community-code/unified-post-processor-upp*) to interpolate state variables to pressure levels and to a horizontal latitude-longitude grid with a 0.03-degree grid spacing. Then the three-dimensional pressure, temperature, and humidity, ozone, cloud water, ice, snow, rain, and graupel, as well as surface characteristic variables, including surface temperature, land type, vegetation, and sensor angles calculated from grid latitude and longitude positions, are used as the input to the Python interface (named PyCRTM *https://github.com/JCSDA/pycrtm*) of the CRTM v 2.3; Han, 2006) which has been developed by the Joint Center for Satellite Data Assimilation (JCSDA), to convert the HAFS output processed by UPP into the mode synthetic images with a wavelength of µ=10.3 µm that matches GOES-R's ABI observation data in channel 13. The corresponding observed images with the same wavelength of 10.3 µm are obtained via the NOAA data Registry of Open Data on AWS at *https://registry.opendata.aws/noaa-goes*

4. Model

The HAFS Version 0.2A is a hurricane application of NOAA's UFS. HAFS V0.2A employs a stand-alone regional and coupled atmosphere-ocean configuration that incorporates air-sea interaction processes via a Community Mediator for Earth Prediction Systems (CMEPS)-based coupling of the FV3 atmospheric and Hybrid Coordinate Ocean Model (HYCOM) oceanic components. The FV3 atmospheric component makes use of a 90-second time step, a three-kilometer regional Extended Schmidt Gnomonic (ESG) grid, and an advanced physics suite optimized for TC forecasting, including GFDL microphysics, RRTMG radiation, scale-aware SAS convection, Noah LSM, GFS surface layer with HWRF exchange coefficients, and the GFSv16 scale-aware TKE-EDMF PBL scheme. The atmospheric initial and lateral boundary conditions for HAFS V0.2A are derived from the operational GFS v16 and the oceanic initial/lateral boundary conditions are derived from the operational Real-Time Ocean Forecast System (RTOFS). It forecasts TCs in the basins of the North Atlantic, the eastern North Pacific, and the western North Pacific. The model consists of 91 vertical levels, with the top level at a pressure of 10 hPa. A comprehensive model documentation can be found at *https://github.com/hafs-community/HAFS.*

5. Methods

We employ a vortex-following method when we interpolate the observed and model synthetic images. The forecast TC tracks are generated by the GFDL vortex tracker (Marchok, 2010) and include 6-hourly TC eye positions (latitude/longitude), maximum wind speed (MAX_wind), minimum sea level pressure (MSLP), and average radii of 34-, 50-, and 64-kt winds in the four quadrants of the northeast (NE), southeast (SE), northwest (NW) and southwest (SW) directions. The model synthetic images are interpolated onto 10° x 10° grids with a 0.06-degree

grid spacing, with the TC centers in the center of the domain grids. For the observed images, best tracks are used, which contain post-TC analyses provided by the National Hurricane Center and consist of official historical records of a storm's 6-hourly eye positions (latitude-longitude), MAX_wind, MSLP, and average radii of 34-, 50-, and 64-kt winds in the four quadrants. The eye positions in the best tracks are used to interpolate the observed images into 10° x 10° grids with a 0.06-degree grid spacing with the TC eyes in the center. Using this vortex-following method, we avoid the impact of track errors on the evaluation, allowing us to focus on evaluating the intensity, size, and structure of the model forecast TCs. Figure 2 shows an example of the observed and model synthetic images after these images are interpolated using the above-mentioned vortex-following method. In Figure 2, the model synthetic images are taken at the same valid time as the observed but with different forecast lead times (24h, 48h, and 72h).

While comparing observed images to their corresponding model synthetic images can reveal differences in cloud BT, vortex size, and structure, it is often difficult to determine whether the differences are due to model
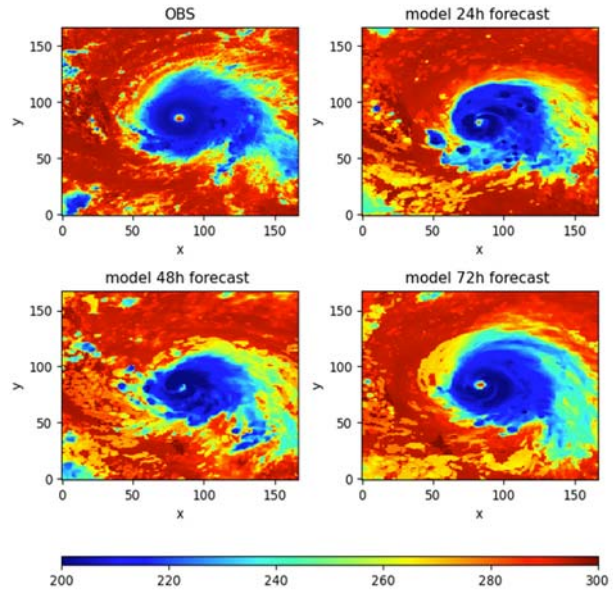

Figure 2 observed GOES-R IR BT image of hurricane Dorian at 2019-09-01 18Z and model synthetic images at valid time of 2019-09-01 18Z with a lead time of 24, 48 and 72 hours

systematic biases or random errors, given the high variability of the model forecast BT fields. Use Figure 2 as an illustration: When compared to the observed image, the 24h forecast vortex has a slightly smaller cloud covered area, the 48h forecast vortex is significantly smaller, but the 72h forecast vortex is larger (they are all at the same valid time with the observed image). As a result, evaluating individual images with varying forecast lead times may result in disparate evaluation conclusions. Therefore, we used a method of superimposing multiple images to create composite observed and model synthetic images with the goal of eliminating random errors and ensuring that any difference between the composite observed and model synthetic images is solely due to model biases, a strategy like the traditional TC verification methods, which uses average track and intensity errors from multiple cases rather than a single forecast snapshot.

In addition to visually comparing the composite images, we also examined the BT's probability density function (PDF) plots and scatter plots made using the observed and model synthetic image data. Note that the pdf gives a sense not only of the shape of the distribution but also the uncertainty in the (avg.) composites.

Traditional verifications are also conducted, including the mean track and intensity errors and biases calculated using the maximum wind speed, the minimum sea level pressure, and the radii of the gale (34-kt), damaging (50-kt), and hurricane (64-kt) winds. Conclusions drawn from traditional verification and the satellite images are compared to determine their consistency.

6. Results

6.1.   Individual images

Figure 3 depicts three samples of individual snapshots of observed and model synthetic satellite images. In general, the model accurately predicts hurricane vortices. However, there appears to be a pattern in which the observation contains more spread of clouds and hydrometeors than the model. Nonetheless, because there is a lot of variations between cases and different snapshots, visually inspecting these individual figures to see system biases is difficult. In the following section, we will look at composite observed and model synthetic images.
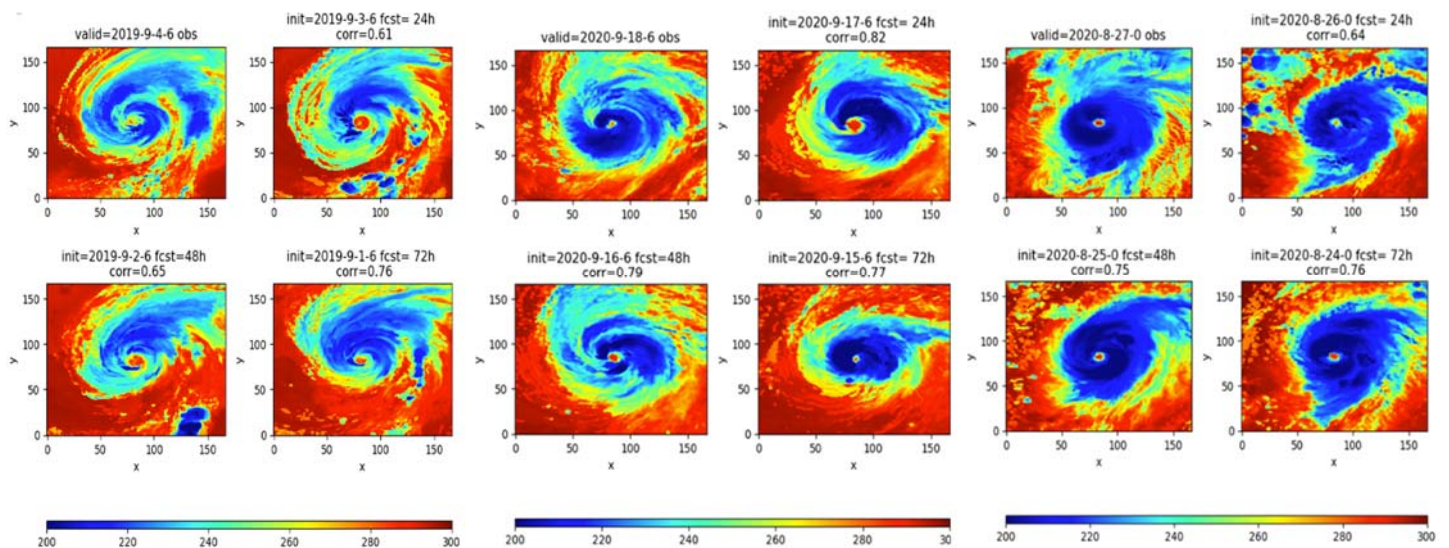
Figure 3 observed and model synthetic satellite images for hurricane Dorian in 2019 (left), Teddy (center) and Laura in 2020 (right).

6.2.   HAFS composite images comparison

6.2.1. Dorian (2019)

1. All-stages

Hurricane Dorian's forecast from 2019-08-24-18 to 2019-09-09-18, a composite of 44 forecast snapshots with a forecast lead = 48 h is shown in Figure 4. It is important to note that using multiple forecast lead times will cause the model to be over-smoothed than the observation because there is only one observation image for a given valid time, whereas the model has

multiple different images with different forecast lead times. Therefore, using all of these model's forecast lead times will result in unrealistic comparisons. In this study, we chose a lead time of 48 hours for the comparisons.

Because the variability of composite images is less than that of individual images, which appear to be small noises, the composite image makes it easier to identify systematic biases. It is clear from these 44 snapshots that the vortices predicted by the model are slightly smaller than the observed vortices. The observed BT has a contour of 220K that covers a larger area than HAFS (a HAFS warm BT bias). This pattern is also visible in the scatter plot and the PDF plot of BT.

Figure 4 all-stages hurricane Dorian observed and model synthetic satellite images. Lower panels show the scatter plot and the PDF plot of BT.

2. Early stage

The early stage of hurricane Dorian, from 2019-08-24-18 to 2019-08-30-12, includes 12 forecast snapshots with a forecast lead = 48 h. From the early-stage composite image (Figure 5), we can see the model forecast vortex tends to be smaller than the observed vortex. The scatter and PDF plots also show that the model has a warm bias.

3. Mature stage

The mature stage of hurricane Dorian is from 2019-08-30-18 to 2019-09-06-18, a composite of 23 forecast snapshots with a forecast lead = 48 h. The mature stage of hurricane Dorian is analyzed in more detail in terms of its (a) intensity, (b) size, and (c) asymmetric structure.

(a) Intensity

Figure 6 depicts the observed and model forecast composite images of hurricane Dorian from 2019 August 30 18Z to 2019 September 06 18Z, which is its mature stage with clear rain bands around its outer edges, the eye, and circular eyewall structure. These two composite images are created by superimposing the 24 observed

Figure 5 early-stage hurricane Dorian observed and model synthetic satellite images. Lower panels show the scatter plot and the PDF plot of BT.

snapshots during this period and the corresponding forecast snapshots. In general, the model's forecast vortex agrees fairly well with the observed. However, the observed composite image contains a nearly closed circular area of 1222 pixels with a BT < 220 K, more than the model synthetic image, in which only 543 pixels have a BT < 220 K. The observed image contains 212
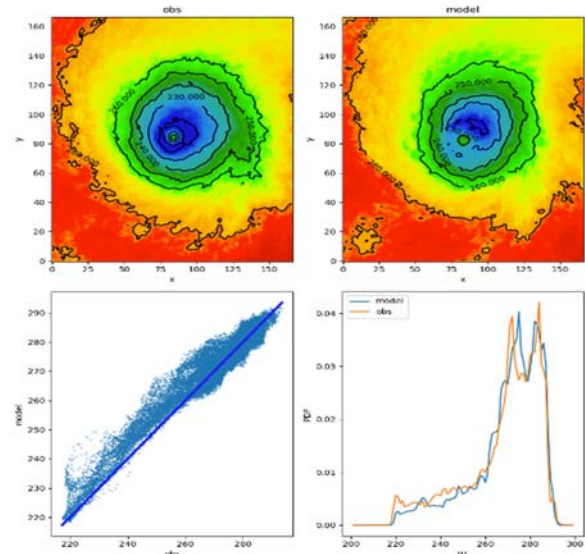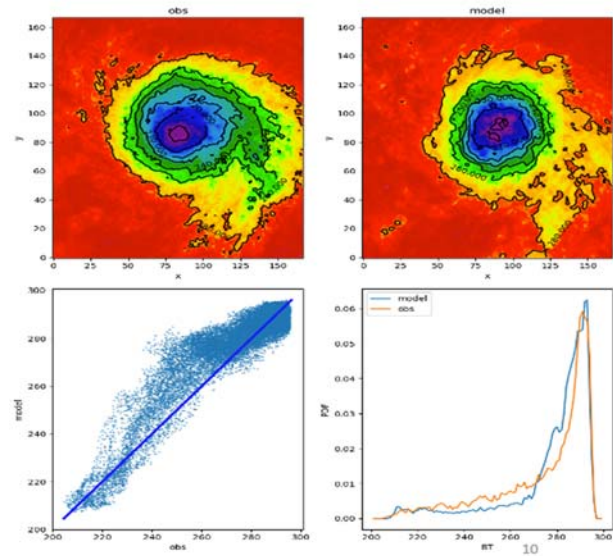
8

pixels with a BT < 210 K, whereas the model synthetic image contains none. As indicated by the red crosses in Figure 6, the minimum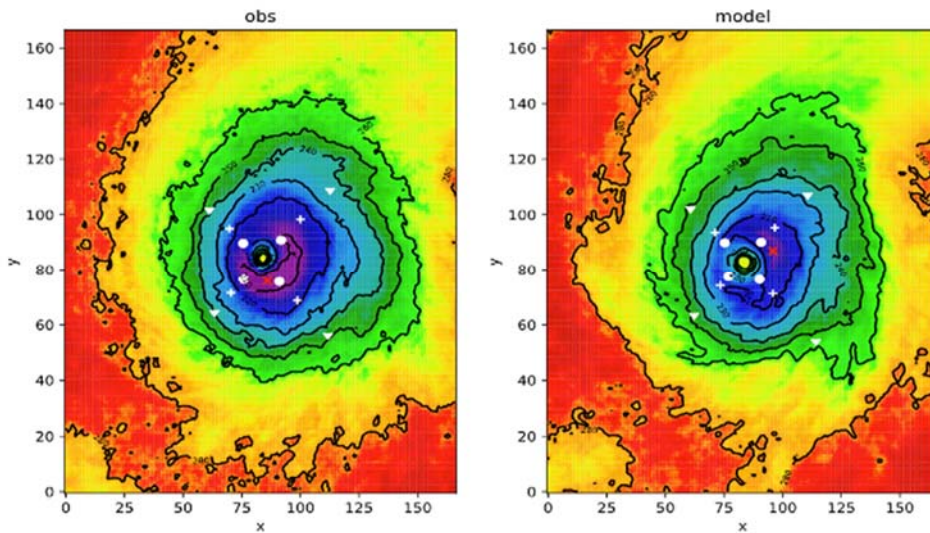 observed BT is 206K, while the model's synthetic minimum BT is 211K. The model's warm bias is also evident in the scatter plot in Figure 7, where the warm bias shifted pixels away from the blue 45° line toward the model side, particularly at the cold end of the spectrum.

Figure 6 observed (left) and model synthetic (right) GOES-R IR BT (in K) fields of the mature stage of hurricane Dorian from 08/30 18Z to 09/06 18Z of 2019. The white markers are the radii of the 34kt (triangles), 50kt (+ signs) and 64kt (circles) winds from best track (left) and the model forecast track (right). The red crosses indicate the location of minimum observed BT of 206K and minimum model synthetic BT of 211K

The cloud top BT represents radiation emitted by cloud tops into space; a lower BT indicates a higher cloud top, which indirectly indicates the strength of the convective activity and thus the intensity of TC. Therefore, the HAFS model's prediction of less cold GOES-R BT – a model warm bias – suggests the model underpredicts the cloud top, and thus the TC intensity. This evaluation result is confirmed by the traditional TC intensity metrics in Figure 8, which shows that the model forecasted a higher TC minimum sea level pressure and a lower maximum wind speed than the best tracks. As a result, the evaluation from the GOES BT images is consistent with the traditional TC intensity metrics.

Satellite data including BT have been used previously to estimate TC strengths, starting from the 1960s and 1970s when the first polar-orbiting weather satellites and geostationary weather satellites were launched. Today, the majority of TC intensity estimates are also based on satellite observations. The Dvorak technique (Dvorak, 1975, 1975, 1984) is widely used
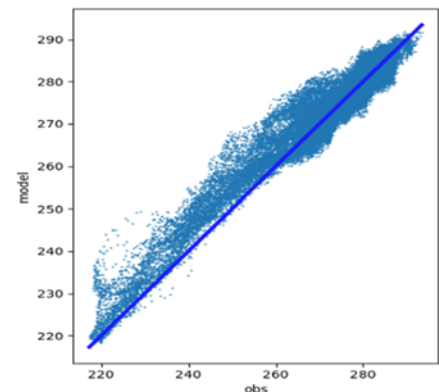

Figure 7 scatter plot of the observed and model forecast IR BT. The blue line indicates the positions for obs=model.

9

to estimate the intensity of TCs at specific times by analyzing TC cloud structures derived from
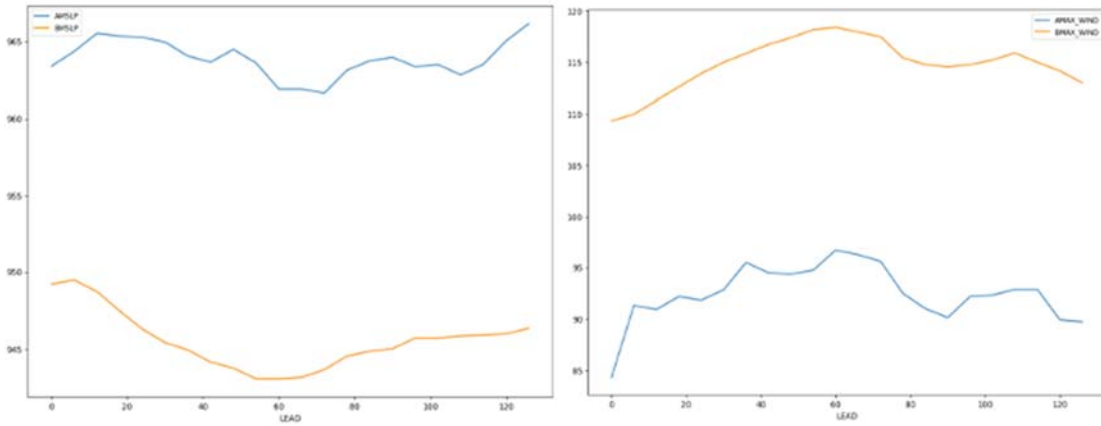

Figure 8 observed (brown) and model predicted (blue) mean sea level pressure (left) and maximum wind speed (right) of hurricane Dorian (2019) in its mature stage, as a function of the forecast lead time

visible and infrared satellite images. DeMaria and Kaplan (DeMaria et al., 2002; DeMaria & Kaplan, 1994) developed the Statistical Hurricane Intensity Prediction Program (SHIPS), which predicts TC intensity over the Atlantic Ocean using satellite BT-related predictors. This article demonstrates that comparing the composite observed and model synthetic images, during which the random differences are removed, can reveal systematic BT biases that are highly correlated with the TC intensity biases.

(b) Size

Along with intensity, the GOES-R BT images also reveal the model's forecast TC size bias. In Figure 6, the observed contour lines, 280k, 260k, 240k, 230k, and 220k cover a larger area than the corresponding lines in the model synthetic images, indicating that the model forecasts less cloud and hydrometeor coverage and thus a smaller TC vortex than observed. Traditional metrics corroborate this conclusion as well. Figure 9 demonstrates that the radii of 50-kt and 64-kt winds in the best tracks are longer than, or at least equal to, those in model forecast tracks in all 4 quadrants. The best track direction-averaged 50 kt (64 kt) wind radii are 21% (13%) larger than predicted by the model. The observed radii of 34 kt winds are larger than the model forecast in the NE quadrant and shorter in the other three
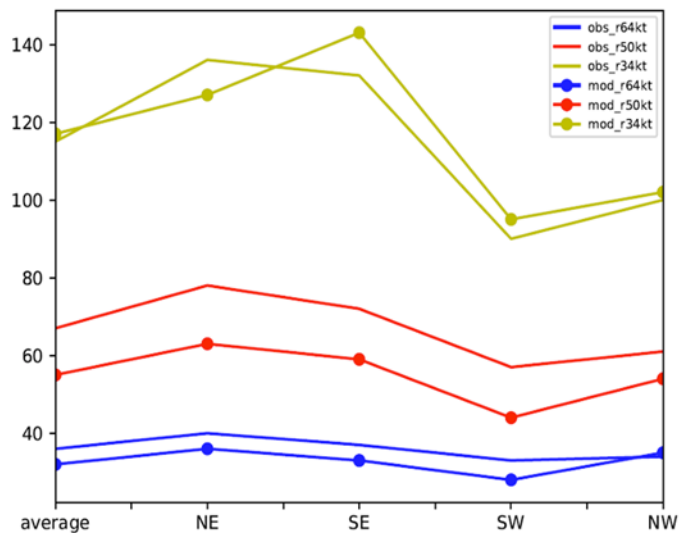

Figure 9 the radii (nautical miles) of winds at 64kt, 50kt and 34kt in the 4 quadrants of NE, SE, SW and NW and the direction-average for hurricane Dorian 2019.

10

directions, but the direction-averaged radii of 34 kt winds in best track are nearly identical to the model forecast, with a difference of only 1.7%. From the 3 critical wind speed radii, it is clear that the observed vortex is larger than the one forecast by the model. In other words, the model seems to under-predict the vortex size, which is consistent with the conclusion drawn from the verification using the composite satellite images.

While the model underpredicted the vortex's overall size, it overpredicted the eye size, as illustrated in Figure 6 by the yellow high BT area in the model TC center that is larger than the observed. Using the distance from the TC center to the sharpest BT gradient to estimate the radius of the TC eye, the observed image shows an eye radius of 20 km, whereas the model synthetic image shows an eye radius of 33 km. Similarly, in the observed image, the distance from the TC center to the minimum BT location is 50 km, while the one in the model synthetic image is 90 km, as indicated by the red crosses in Figure 6. Following the technique in Lajoie & Walsh (2008), we estimate the radius of maximum wind (RMW) in the observed image to be 31 km and the RMW in the model synthetic image is 55 km.

Figure 10 Hurricane Dorian (2019) 34-kt wind radii in best track (brown) and HAFS model verification (blue) by valid time.

In Dong et al. (2020) and Hazelton et al. (2021) it was found that the HAFS model overpredicted the 34-kt wind radii for all forecast lead times, a finding that is not observed in this paper. Dong and Hazelton's studies are based on an earlier version of HAFS than the version used in this study. The HAFSv0.2A performance has been greatly improved on top of the configuration used by Dong et al. and

Figure 11 Hurricane Dorian (2019) 34kt wind radii in best track (brown) and model forecast (blue) by forecast lead time

Hazelton et al. on almost all aspects, including the 34-kt wind radius size errors and biases. However, we do still see larger model-predicted 34 kt radii than those in the best tracks. So, there may be another reason why there is more systematic overestimation of the 34-kt wind radius in those studies, since the HAFS physics schemes have changed between their studies and this one. Nevertheless, there is a possibility that it is because in this report we focused on the mature stage of the hurricanes.
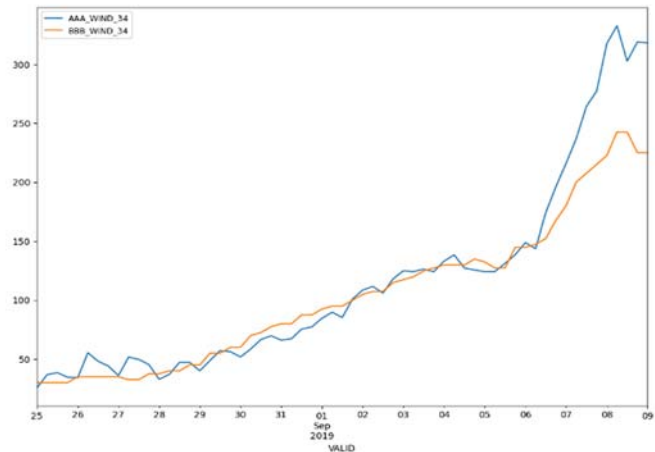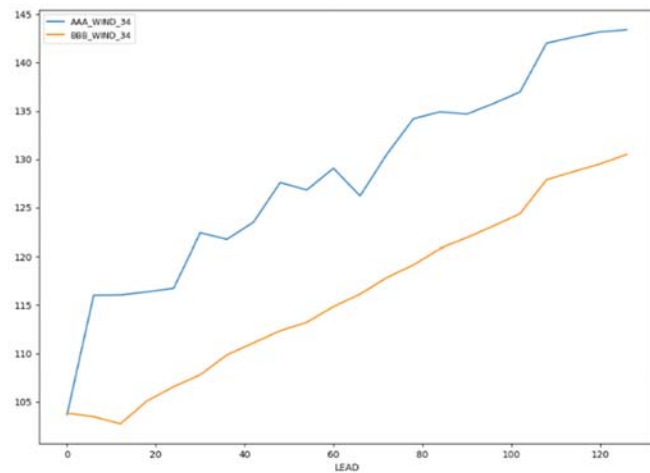
11

Figure 10 shows that the 34-kt wind radii in the best track and model forecast agreed well from August 25 to September 06, 2019, but once the hurricane entered its decay stage, the model substantially overpredicted the 34-kt wind radii by up to 100 NM. When the radii time series are plotted with the x-axis being the forecast lead times, the overpredicted 34-kt wind radii in the decay stage are evenly shared by all forecast lead times (Figure 11), giving a misleading impression that the HAFS model has a large positive bias of 34-kt wind radii in all its forecast times, when in fact the large positive bias occurred only in the TC decay stage. The model's distinct behavior in forecasting mature versus decay stage is illustrated in Figure 14, which demonstrates that during the decay stage of hurricane Dorian, the HAFS model had a cold BT bias (strong intensity), rather than a

Figure 12 The number of pixels in the observed and model synthetic GOES-R IR BT images that are less than the certain thresholds (220K,230K, 240K, 250K, 260K and 270K) in the 4 quadrants of NE SE SW and NW and the direction-averaged number.

warm BT bias (weaker intensity) during its mature stage as shown in Figure 6. This change may have caused the overprediction of 34-kt wind radii during the decay stage, as exaggerated by plots that use forecast lead times as x-axis. This did not occur for the 50-kt and 64-kt wind radii, most likely because these stronger winds did not or rarely exist during the decay stage.
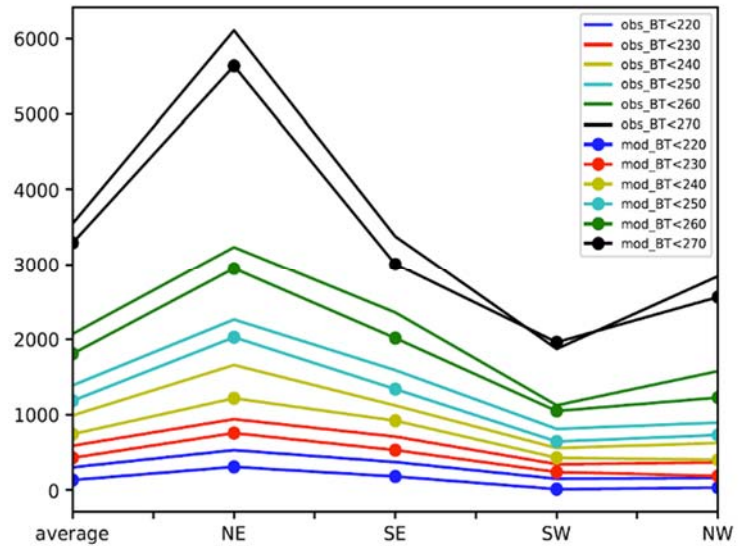
It should be noted that the wind radii from the best track and the output from the GFDL vortex tracker may not be entirely comparable. If there are some small areas of 34-kt winds in a rain band, for instance, the tracker will note that as the 34-kt wind radius. However, in the real world, these winds might remain undetected or be interpreted as unrepresentative, especially if there are no reconnaissance measurements. This uncertainty may have a greater impact in the decay stage when the TCs may have lost their coherent circular structures. It would be of interest in a future study to examine the contour maps of the 10-m wind speed in the hurricane during the decay stage to determine if the 34-kt radius diagnosed by the tracker is indeed associated with the TC core itself, rather than an extratropical feature or a small area in a rain band. Nevertheless, this analysis provides useful information on how we might further improve the HAFSv0.2A performance for storm size prediction by closer examination of the TC vortex size in TC's decay stages.

(c) Asymmetric structure

Figure 6 also depicts similar asymmetric structures between the observed and model forecast TC vortices. The areas covered by clouds and hydrometeors in both composite observed and

model images, as indicated by the BT contour lines (BT < 280K), extended more toward and even beyond the domain boundaries in the NE and SE directions, whereas large cloud-free areas (BT > 280K) are exposed in the NW and SW directions. The same asymmetric structure can be quantified in Figure 12, which demonstrates that for all thresholds, the NE quadrant always contains the most pixels while the SW quadrant contains the fewest, with the SE and NW quadrants in between and the SE quadrant containing more than the NW, highlighting the asymmetric structure of the TC vortices that extended toward the NE direction with the eyes located in the SW part of the areas covered by clouds and hydrometers. The asymmetric TC vortex structure revealed by the composite observed and model synthetic images can also be confirmed by the traditional TC metrics. The radii are plotted in Figure 6 as triangle markers (34-kt radii), "+" markers (50-kt radii) and "x" markers (64-kt radii). These markers clearly show a spatial pattern with the largest distance from the eye in the NE direction and the smallest in the SW; more specifically, $r\_NE > r\_SE > r\_NW > r\_SW$, which is consistent with the asymmetric structure depicted in Figure 12. The distribution of radii is illustrated more quantitatively in Figure 9, which depicts the largest wind radii in the NE direction and the smallest in the SW direction, as well as the wind radii in the SE and NW directions in between. The only exception is that the model-predicted 34-kt wind radius is larger in the SE than the NE.

Although both the observed and the model images had similar asymmetric TC vortex structures, the coldest BT in the observed image is in the direction to the SE, whereas the coldest BT in the model image points to the NE direction.

(d) Comparison with HWRF

For the mature stage of hurricane Dorian, which consists of 23 forecast snapshots with a forecast lead = 48h from 2019-08-30-18 to 2019-09-06-18, we also compared the composite satellite images from observation and the HAFS model with the HWRF 2019 operational version results. From Figure 13, the HAFS forecast vortex matched observations better than HWRF. HWRF has a much larger and colder inner vortex but a smaller outer vortex. The PDF lines in Figure 13 also indicate a higher correlation between observation and HAFS than between observation and HWRF.
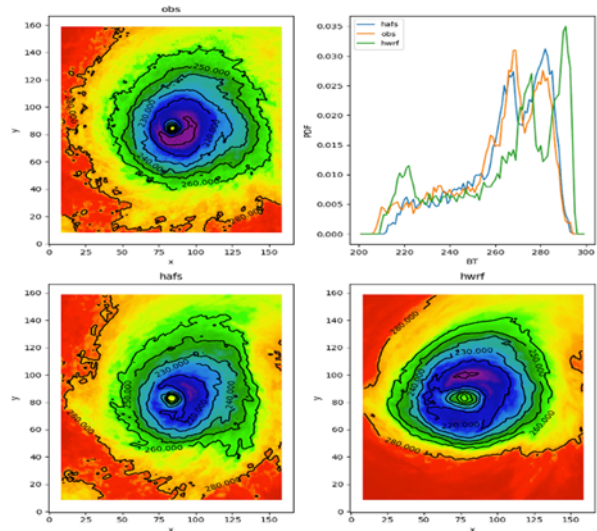

Figure 13 observed (top left) and HAFS (bottom left) and HWRF (bottom right) model synthetic satellite images for the mature stage of hurricane Dorian. PDF lines plot is on the top right.

13

4. Decay stage

The decay stage of hurricane Dorian is from 2019-09-07-00Z to 2019-09-09-18Z, consisting of 9 forecast snapshots. There is no clear annular vortex structure in Figure 14, which shows that the HAFS forecast BT is cooler than the observed BT during the decay stage. The scatter and PDF plots also show that the model has a cold bias in the decay stage. This is a markedly different behavior than that observed in the mature stage and throughout the HAFS model forecast of Hurricane Dorian.

6.2.2. Teddy (2020)

1. All-stages

The composite image of Hurricane Teddy (2020) consists of 27 forecast snapshots from
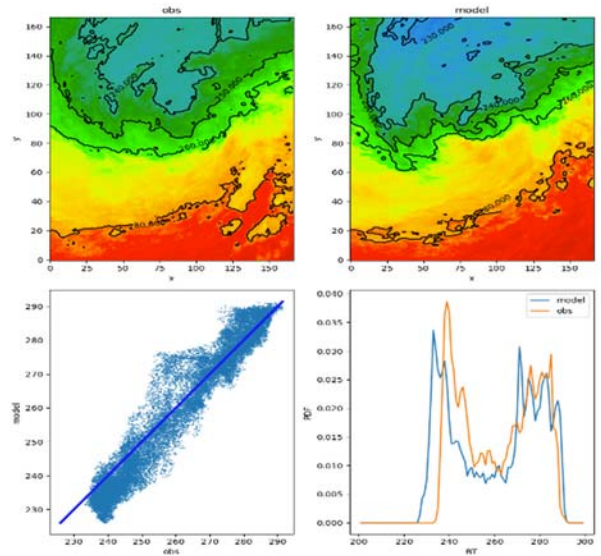
Figure 14 composite observed and model synthetic satellite images in the decay stage of hurricane Dorian (2019)

2020-09-13-06 to 2020-09-20-12 with a forecast lead =48 h. From Figure 15 one can see that the HAFS forecast vortex appears to have a similar size to the observed. The observed BT has a 220

K contour that is missing from the HAFS synthetic image, suggesting the HAFS BT is less cold near the eye-wall region. The scatter and PDF plots also show that the model has a warm bias in the range of 210-220K.
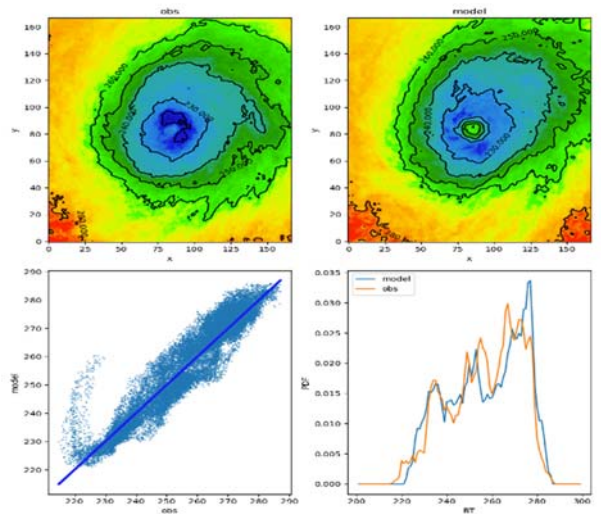
2. Early stage

The early stage composite images are generated from 3 snapshots from 2020-09-13-06Z to 2020-09-16-06Z. It is hard to see systematic patterns from such a small number of samples (Figure 16).

Figure 15 composite observed and model synthetic satellite images in the all-stages of hurricane Teddy (2020)

14

3. Mature stage

The mature stage of hurricane Teddy consists of 13 forecast snapshots and their corresponding
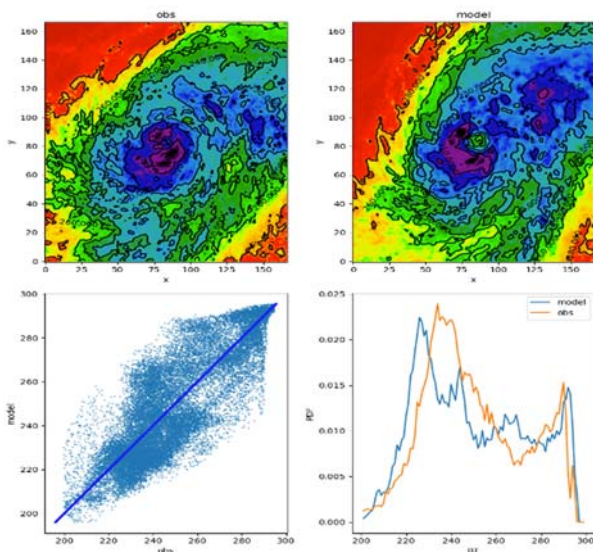

Figure 16 composite observed and model synthetic satellite images in the early-stage of hurricane Teddy (2020)
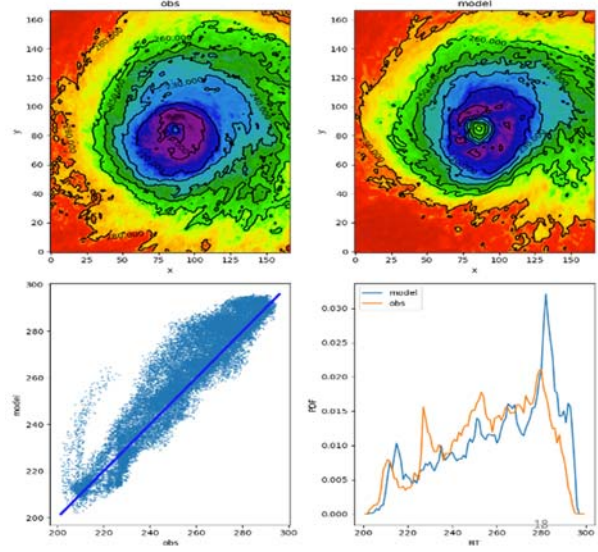

Figure 17 composite observed and model synthetic satellite images in the mature-stage of hurricane Teddy (2020)

observations from 2020-09-16 to 2020-09-20. The composite image Figure 17 shows the HAFS forecast vortex is smaller than observed. The observed BT has a 210 K contour covering a larger area than those in the HAFS vortex, suggesting the HAFS is less cold. The scatter and PDF plots also show that the model has a warm bias in its predicted BT field.

4. Decay stage

The decay stage of hurricane Teddy lasted from 2020-09-20-18 to 2020-09-23-18, consisting of
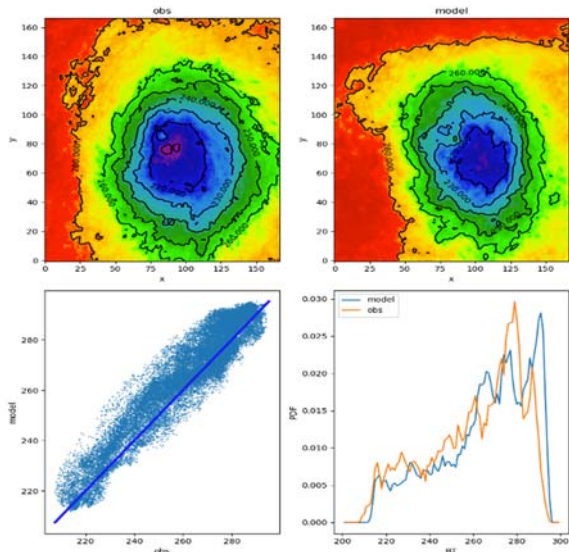

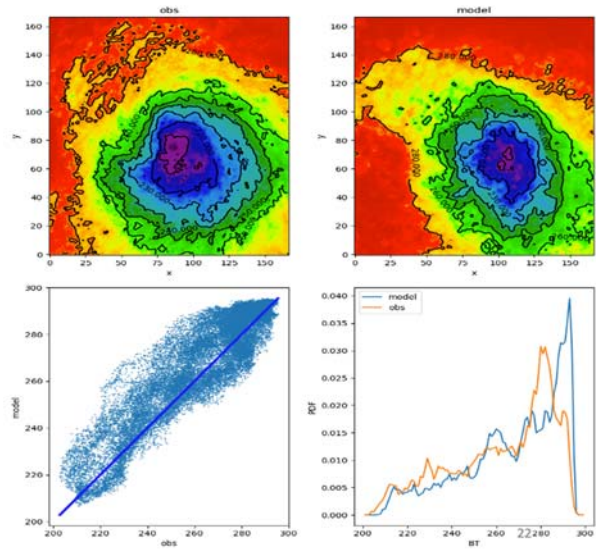Figure 18 composite observed and model synthetic satellite images in the all-stages of hurricane Laura (2020)

Figure 19 composite observed and model synthetic satellite images in the early-stage of hurricane Laura (2020)

1

11 forecast snapshots. The HAFS BT is colder than the observed BT during the decay period. The scatter and PDF plots also show that the model has a cold bias in its decay stage (not shown).

6.2.3. Laura (2020)

1. All-stages

The forecast for hurricane Laura consists of 21 forecast snapshots with a forecast lead=48 h from 2020-08-20-18 to 2020-08-28-00. From Figure 18, one can see The HAFS forecast vortex appears to be slightly smaller than observed. The observed BT has a 210 K contour that is missing in the HAFS BT, suggesting that the HAFS BT is less cold near the eyewall region. The scatter and PDF plots also show that the model has a warm bias in the forecast BT field.

2. Early stage

The early state of hurricane Laura lasted from 2020-08-20-18 to 2020-08-25-18, consisting of 12 forecast snapshots. Figure 19 showed the similar pattern seen in Figure 18.

3. Mature stage

Hurricane Laura's mature stage spanned from 2020-08-26-00 to 2020-08-27-12, consisting of 7 forecast snapshots, a relatively small number of samples. Figure 20 shows that the HAFS vortex is smaller than observed; however, its inner core appears to be slightly larger than observed. The observed BT has a 210-K contour that covers a smaller area than that in the HAFS BT, suggesting that the HAFS BT is colder near the eyewall region. Also, the scatter plot and PDF show that the model has a warm bias in the intermediate range of 220-280K but a cold bias in the cold range of BT<220 K. The mature stage of Hurricane Laura in 2020 is an exception to the HAFS behavior found in the other stages of Laura and other storms.

Figure 20 composite observed and model synthetic satellite images in the mature-stage of hurricane Laura (2020)

4. Decay stage

Like hurricanes Dorian (2019) and Teddy (2020), hurricane Laura's decay stage, which consists of only 2 forecast snapshots, showed that the HAFS BT is colder than the observed BT during the decay period (not shown).
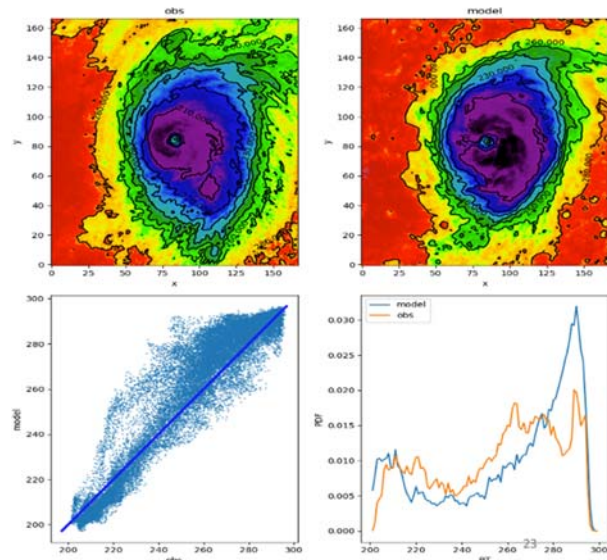
16

6.3. Two-physics-suite comparison

This past year, the DTC team evaluated and tested HAFS with two physics suites: the NCEP GFS suite and the HWRF suite. The original goal of this DTC Visitor Project was to compare the testing results to observations to determine which physics suite, GFS or HWRF, produces more realistic hurricane structure forecasts. The model synthetic imagery was created by the UFS UPP. Thousands of images were compared, and the results revealed that when the HWRF physics suite was used in HAFS, the model generated exceptionally large hurricane vortices. See Figure 21 for an example of hurricane Dorian in 2019, located along the US East Coast.
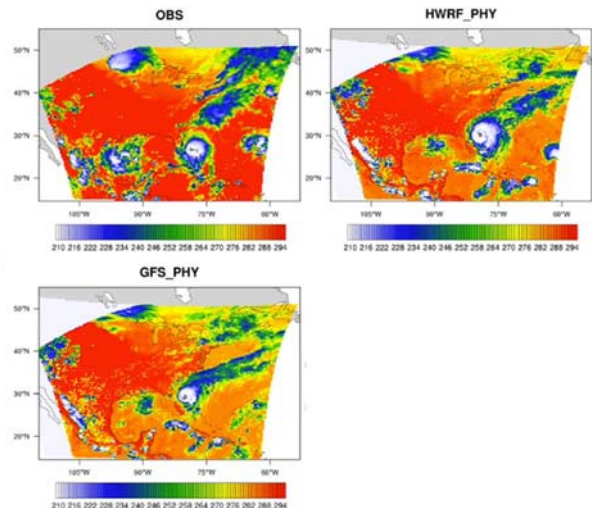

Figure 21 Hurricane Dorian from the observation, HAFS model forecast using HWRF physics suite and HAFS model forecast using GFS physics suite.

The HAFS model is being actively developed at the EMC. A seasonal test of HAFS (using a highly modified version of the GFS physics suite plus other advancements such as coupling with an ocean model) was performed at EMC in the summer of 2021. EMC suggested that it would be beneficial to the HAFS model development if we could broaden the scope of the DTC Visitor project to include the evaluation mentioned above. That is why we switched the focus of this work from evaluating the two-physics-suite tests to those more recent cutting-edge modeling test results.

6.4. Initial conditions

EMC model developers expressed interest in understanding if there are systematic biases of the vortex intensity and structure in the HAFS model's initial conditions. The initial conditions of HAFS use the FV3-GFS forecast cloud and hydrometeor fields, which allows for comparing the synthetic satellite images to the observed ones at initial times (forecast lead time = 0).
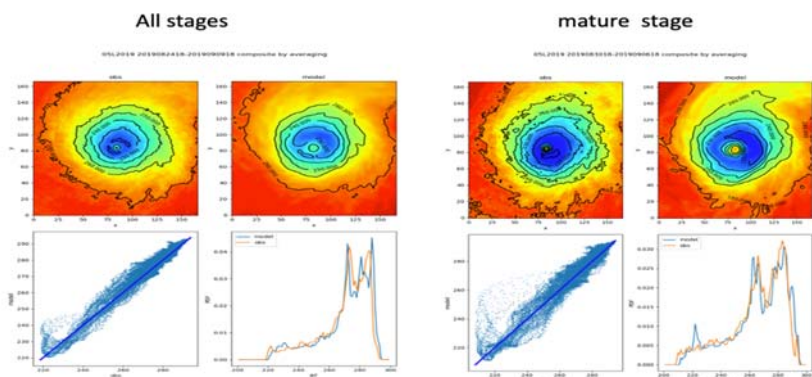

Figure 22 HAFS model initial time synthetic satellite images compared with the observed ones for the case of hurricane Dorian (2019).

Figures 22-24 show the comparisons between the model synthetic versus the observed BT images at initial time (forecast lead time = 0). The comparisons included the all-stages and mature-stage of hurricanes Dorian (2019), Teddy (2020) and Laura (2020).

From these figures one can see that that at the initial time: (1) as in the forecast at other lead times, the HAFS initial conditions have a BT that is less cold than the observed BT. Use hurricane Dorian (2019)'s all-stages composite images as an example. The observed BT has a contour of 220 K, which is missing in the HAFS initial condition. And in its mature stage, the observed BT has a contour of 210 K, which is also missing in the HAFS initial condition. This discrepancy is understandable considering that the lower-resolution FV3GFS cloud and hydrometeor files are used to fill the HAFS initial conditions. (2) the vortex size in the HAFS initial conditions matches the observed vortex size. Similar patterns are also found for hurricanes Teddy and Laura in 2020.
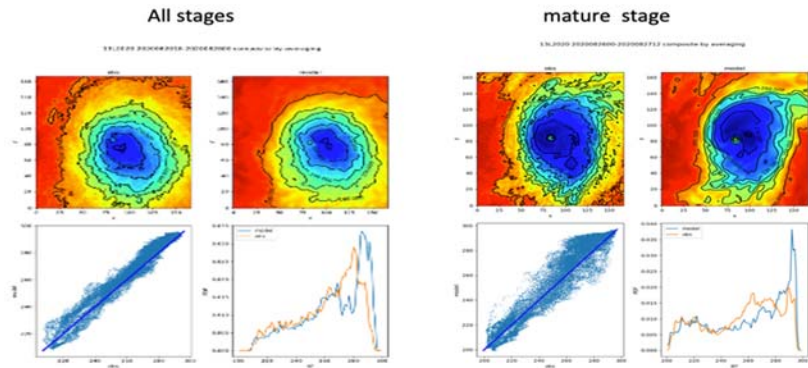

Figure 23 HAFS model initial time synthetic satellite images compared with the observed ones for the case of hurricane Teddy (2020).
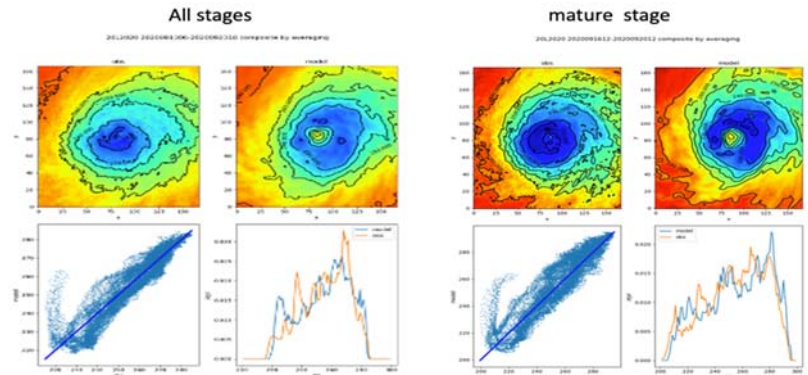

Figure 24 HAFS model initial time synthetic satellite images compared with the observed ones for the case of hurricane Laura (2020).
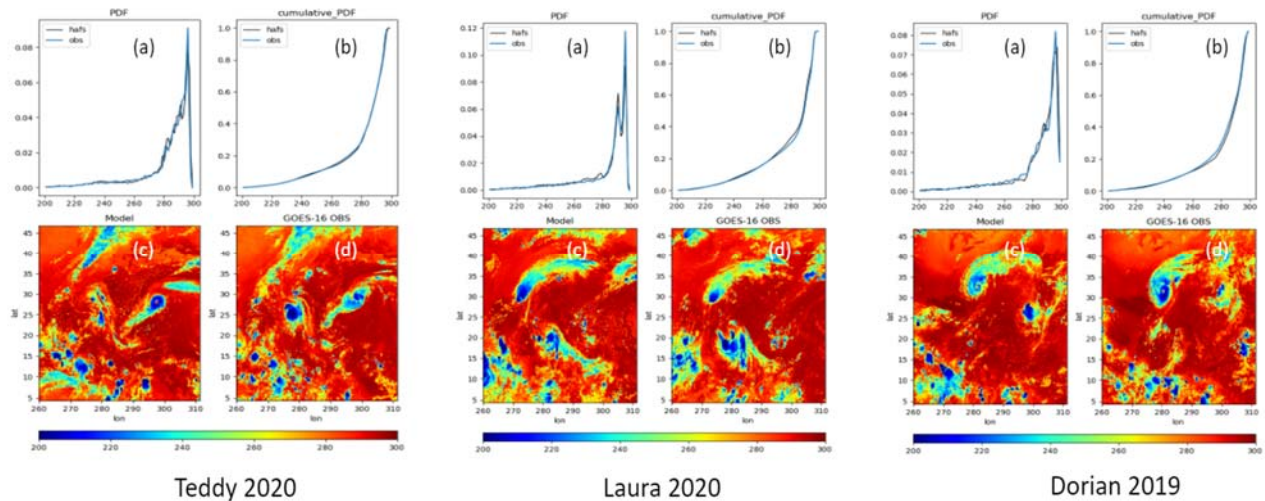

Figure 25 The PDF (a), cumulative PDF (b), model synthetic satellite image (c), and observed satellite image (d) for hurricanes Teddy (2020), Laura (2020) and Dorian (2019) on a synoptic scale.

6.5.   *Synoptic* satellite images comparison

The above analysis focused on the hurricane vortexes. The GFS and HAFS model developers also expressed interest in comparing the observed and model synthetic satellite images on a synoptic scale. Figure 25 shows the synoptic scale evaluation results for the three hurricanes. Because it is difficult to compare the composite images on a synoptic scale, we compared one forecast snapshot for each hurricane. The observed and model synthetic satellite images are interpolated onto a common domain grid whose latitudes span from 5N to 45N and longitudes from 260E to 310E, covering a large area over the North Atlantic basin. The PDF and cumulative PDF from HAFS BT and the observed BT matched very well on a synoptic scale, despite the differences shown earlier on a vortex scale. This may indicate that although there exist biases in HAFS forecast TC intensity and size, the overall physics schemes are realistic and robust, probably indicating that hurricane- and meso-(or even smaller-)scales processes with respect to BT are more sensitivity to the model physics evaluated in this study while the synoptic and large-scale processes relatively insensitive to the model physics.

Summary

HAFS is the next-generation hurricane prediction and analysis system based on the FV3 dynamic core with TC-specific physics schemes. A tool has been developed to create model synthetic

satellite images from HAFS and GFS grib2 data (the post-processed forecasts), utilizing the user-friendly python interface of the CRTM radiative transfer model to convert the HAFS forecasts to synthetic GOES-R BT images, which are then compared to the corresponding observed images. Uncertainties in TC forecasts include both random errors and systematic biases. Rather than comparing observed and model synthetic images at individual discrete times, which will be affected by random errors, we compared the composite (or averaged) observed and model synthetic images, assuming that any differences caused by random errors will be eliminated during the averaging process, leaving any remaining difference mainly due to the model's systematic biases. The composite images of hurricanes Dorian (2019), Teddy and Laura (2020) in their various stages are used for evaluation.

Evaluation results show that overall, the HAFS synthetic satellite images and the observed ones agree reasonably well. Results also revealed that the model exhibits a warm BT bias, probably indicating lower cloud altitude and thus weaker convection and TC intensity. The warm bias can also be seen in the scatter plot and the probability density function (PDF). The result is consistent with traditional metrics of mean sea level pressure (MSLP) and maximum wind speed (WAX_WIND). Furthermore, the model underpredicted the vortex's overall size, as the predicted cloud-covered areas were smaller than the observed. This conclusion agrees with the traditional indicators of the radii of 34-kt, 50-kt, and 64-kt winds. Thirdly, the images revealed the asymmetric structure of the observed and model forecast vortices, which is also confirmed by the differences in the 34-kt, 50-kt, and 64-kt wind radii in the four quadrants. The GOES-R images provided additional value to the traditional verification metrics. For instance, the satellite images can be used to compare the observed and model forecast TC eye sizes and radii of WAX_WIND,

which is not always available in traditional verification products. Also, the images reveal that the HAFS model overpredicted intensity during its decay stage, possibly resulting in significantly larger radii of 34-kt wind than in best track, which may propagate into the traditional vortex size verification when the 34-kt wind radii are presented as a time series varying with forecast lead times.

Note that since this study only evaluated three hurricanes, with the in-depth analyses (radii, intensity, size, and asymmetry etc.) conducted only on one of them, the above conclusions could be applicable only to the model behavior of the three hurricane cases used in this study, and caution should be exercised when extrapolating these conclusions to anticipate model biases in predicting other TCs. Nonetheless, the consistency between the evaluation using the composite satellite images and the traditional metrics, of hurricane Dorian, shows that this method has the potential to be applied to other storms in future studies. The composite satellite images provide a wealth of information thanks to their near-real-time spatially and temporally continuous high-resolution global coverage. Composite satellite infrared images could be routinely used as a physics-based verification technique to supplement traditional verification methods in the HAFS and other hurricane model evaluations.

Deliverables.

1. Published a manuscript titled "The Use of Composite GOES-R Satellite Imagery to Evaluate a TC Intensity and Vortex Structure Forecast by an FV3GFS-Based Hurricane Forecast Model" to the journal ***atmosphere***, in which the DTC visitor's program is acknowledged.

2. Presentations: a DTC-EMC joint presentation was made on October 14, 2021, with the title "GOES-R brightness temperature to evaluate the vortex structure forecast by the Hurricane Analysis and Forecast System (HAFS)"

The initially planned visits to NCAR DTC and NOAA EMC were not realized due to the COVID-19 pandemic.

Acknowledgement

Reference

Bao, S., Zhang, Z., Kalina, E., & Liu, B. (2022). The Use of Composite GOES-R Satellite Imagery to Evaluate a TC Intensity and Vortex Structure Forecast by an FV3GFS-Based Hurricane Forecast Model. *Atmosphere*, *13*(1), 126.

Bao, S. (2018). Use Satellite Infrared Brightness Temperature Data to Evaluate HWRF Ferrier-Aligo microphysics scheme. DTC visitor project report available at https://dtcenter.org/sites/default/files/visitor-projects/dtc-visitor-report-2018-Bao-final.pdf

Bao, S., Bernardet, L., Thompson, G., Kalina, E., Newman, K., & Biswas, M. (2020). Impact of the Hydrometeor Vertical Advection Method on HWRF's Simulated Hurricane Structure. Weather and Forecasting, 35(2), 723–737.

Cintineo, R., Otkin, J. A., Xue, M., & Kong, F. (2014). Evaluating the Performance of Planetary Boundary Layer and Cloud Microphysical Parameterization Schemes in Convection-Permitting Ensemble Forecasts Using Synthetic GOES-13 Satellite Observations. Monthly Weather Review, 142(1), 163–182. https://doi.org/10.1175/MWR-D-13-00143.1

DeMaria, M., & Kaplan, J. (1994). A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. Weather and Forecasting, 9(2), 209–220.

DeMaria, M., Zehr, R. M., Kossin, J. P., & Knaff, J. A. (2002). The use of GOES imagery in statistical hurricane intensity prediction. Preprints, 25th Conf. on Hurricanes and Tropical Meteorology, San Diego, CA, Amer. Meteor. Soc, 120, 121.

Dong, J., Liu, B., Zhang, Z., Wang, W., Mehra, A., Hazelton, A. T., Winterbottom, H. R., Zhu, L., Wu, K., & Zhang, C. (2020). The evaluation of real-time Hurricane Analysis and Forecast System (HAFS) Stand-Alone Regional (SAR) model performance for the 2019 Atlantic hurricane season. Atmosphere, 11(6), 617.

Dvorak, V. F. (1975). Tropical cyclone intensity analysis and forecasting from satellite imagery. Monthly Weather Review, 103(5), 420–430.

Dvorak, V. F. (1984). Tropical cyclone intensity analysis using satellite data (Vol. 11). US Department of Commerce, National Oceanic and Atmospheric Administration ….

Hazelton, A., Zhang, Z., Liu, B., Dong, J., Alaka, G., Wang, W., Marchok, T., Mehra, A., Gopalakrishnan, S., & Zhang, X. (2021). 2019 Atlantic Hurricane Forecasts from the Global-Nested Hurricane Analysis and Forecast System: Composite Statistics and Key Events. Weather and Forecasting, 36(2), 519–538.

Jin, Y., Wang, S., Nachamkin, J., Doyle, J. D., Thompson, G., Grasso, L., Holt, T., Moskaitis, J., Jin, H., Hodur, R. M., Zhao, Q., Liu, M., & DeMaria, M. (2014). The Impact of Ice Phase Cloud Parameterizations on Tropical Cyclone Prediction. Monthly Weather Review, 142(2), 606–625. https://doi.org/10.1175/MWR-D-13-00058.1

Lajoie, F., & Walsh, K. (2008). A Technique to Determine the Radius of Maximum Wind of a Tropical Cyclone. Weather and Forecasting, 23(5), 1007–1015. https://doi.org/10.1175/2008WAF2007077.1

Marchok, T. (2010). Use of the GFDL vortex tracker. WRF Tutorial for Hurricanes, NOAA/GFDL, 40.

Novak, K., & Bao, S. (2019). Evaluation of FV3 Model Using Satellite Brightness Temperature Data. AGU Fall Meeting Abstracts, 2019, A21O-2771.

Otkin, J. A., Lewis, W. E., Lenzen, A. J., McNoldy, B. D., & Majumdar, S. J. (2017). Assessing the Accuracy of the Cloud and Water Vapor Fields in the Hurricane WRF (HWRF) Model Using Satellite Infrared Brightness Temperatures. Monthly Weather Review, 145(5), 2027–2046. https://doi.org/10.1175/MWR-D-16-0354.1