atmosphere

MDPI

Article

The Use of Composite GOES-R Satellite Imagery to Evaluate a TC Intensity and Vortex Structure Forecast by an FV3GFS-Based Hurricane Forecast Model

Shaowu Bao [1,*], Zhan Zhang [2], Evan Kalina [3,4,5] and Bin Liu [2]

1 Department of Marine Science, Coastal Carolina University, 100 Chanticleer Dr E, Conway, SC 29528, USA
2 I.M. Systems Group, Inc., NOAA/NWS/NCEP/EMC, 5830 University Research CT, College Park, MD 20740, USA; zhan.zhang@noaa.gov (Z.Z.); bin.liu@noaa.gov (B.L.)
3 Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309, USA; evan.kalina@noaa.gov
4 Global Systems Laboratory (GSL), NOAA, Boulder, CO 80305, USA
5 Developmental Testbed Center (DTC), Boulder, CO 80305, USA
* Correspondence: sbao@coastal.edu

Abstract: The HAFS model is an effort under the NGGPS and UFS initiatives to create the next generation of hurricane prediction and analysis system based on FV3-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. While satellite images have been used to verify hurricane model forecasts, they have not been used on HAFS. The community radiative transfer model CRTM is used to generate model synthetic satellite images from HAFS model forecast state variables. The 24 forecast snapshots in the mature stage of hurricane Dorian in 2019 are used to generate a composite model synthetic GOES-R infrared brightness image. The composite synthetic image is compared to the corresponding composite image generated from the observed GOES-R data, to evaluate the model forecast TC vortex intensity, size, and asymmetric structure. Results show that the HAFS forecast TC Dorian agrees 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 larger than the observed. The evaluation results can be used to further improve the model. 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-times 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 one hurricane, the above conclusions are only applicable to the model behavior of the mature stage of hurricane Dorian in 2019, and caution is needed to extend these conclusions to expect model biases in predicting other TCs. Nevertheless, the consistency between the evaluation using 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.

Keywords: tropical cyclone; hurricane; remote sensing; model verification

check for updates

Citation: Bao, S.; Zhang, Z.; Kalina, E.; Liu, B. 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* **2022**, *13*, 126. https://doi.org/10.3390/atmos13010126

Academic Editor: Eric A. Hendricks

Received: 10 December 2021
Accepted: 7 January 2022
Published: 13 January 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

Tropical cyclone (TC) forecasts are critical for mitigating damage to coastal communities. The Hurricane WRF (HWRF) model is currently the operational TC forecast model in the United States. It is based on the 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 GFS dynamic core in order to advance tropical cyclone 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 [1]. 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 [1]. 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 [1,2]. 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) strengths winds. Along with these traditional TC forecast metrics, Dong et al. and Hazelton [1,2] 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 min or every 5 min in continuous-disk mode, with a spatial resolution of 0.5–2 km. These continuous 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 tropical cyclones (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 recents, 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 [3–8]. These studies employ a "model-to-satellite" methodology [8], 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) brightness temperatures (BT) images (hereinafter as "images").

Since GOES-R IR BT data have not previously been used to assess the TC structure and intensity of HAFS forecasts, we present the findings of such an evaluation in this paper. 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 model's systematic biases. We aim to answer the following three research questions. (1) Can the observed and synthetic images reveal HAFS systematic biases in its predictions of TC intensity, size, and structure? (2) Can the evaluation results corroborate those obtained through the use of traditional evaluation methods? (3) Can the GOES-R IR BT images provide additional benefits that are not available from the traditional metrics? If the answers to all of these questions are affirmative, we can propose that GOES-R IR BT images be added as a routine model evaluation tool in addition to the traditional ones, because satellite IR BT images provide rich temporally and spatially continuous coverage.

2. The Model

The Hurricane Analysis and Forecast System (HAFS) Version 0.2A is a hurricane application of the National Oceanic and Atmospheric Administration's (NOAA) Unified Forecast System (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 Finite-Volume Cubed-Sphere (FV3) atmospheric and Hybrid Coordinate Ocean Model (HYCOM) oceanic components. The FV3 atmospheric component makes use of a 90-s time step, a 3-km 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 GFSv16 scale-aware TKE-EDMF PBL scheme. The atmospheric initial/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 tropical cyclones 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 (accessed 7 January 2022).

3. Case and Data

At the Environmental Modeling Center (EMC), several high-impact hurricanes, including Hurricane Dorian in 2019, were used to test the HAFS model version 0.2. The native output files of these tests are first post-processed using the unified post-processing (UPP (https://dtcenter.org/community-code/unified-post-processor-upp) (accessed 7 January 2022)) 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 (PyCRTM (https://github.com/JCSDA/pycrtm) (accessed 7 January 2022)) of the CRTM v 2.3 [9] 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 the Geostationary Operational Environmental Satellite (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 (accessed 7 January 2022).

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 [10] 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 northeast (NE), southeast (SE), northwest(NW) and southwest (SW). The model synthetic images are interpolated onto $10^\circ \times 10^\circ$ grids with a 0.06-degree grid spacing, with the TC centers in the center of 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 to $10^\circ \times 10^\circ$ 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 1 shows an example of the observed and model synthetic images after these images are interpolated using the above-mentioned vortex-following method. In Figure 1,

the model synthetic images are taken at the same valid time as the observed but with different forecast lead times (24 h, 48 h and 72 h).

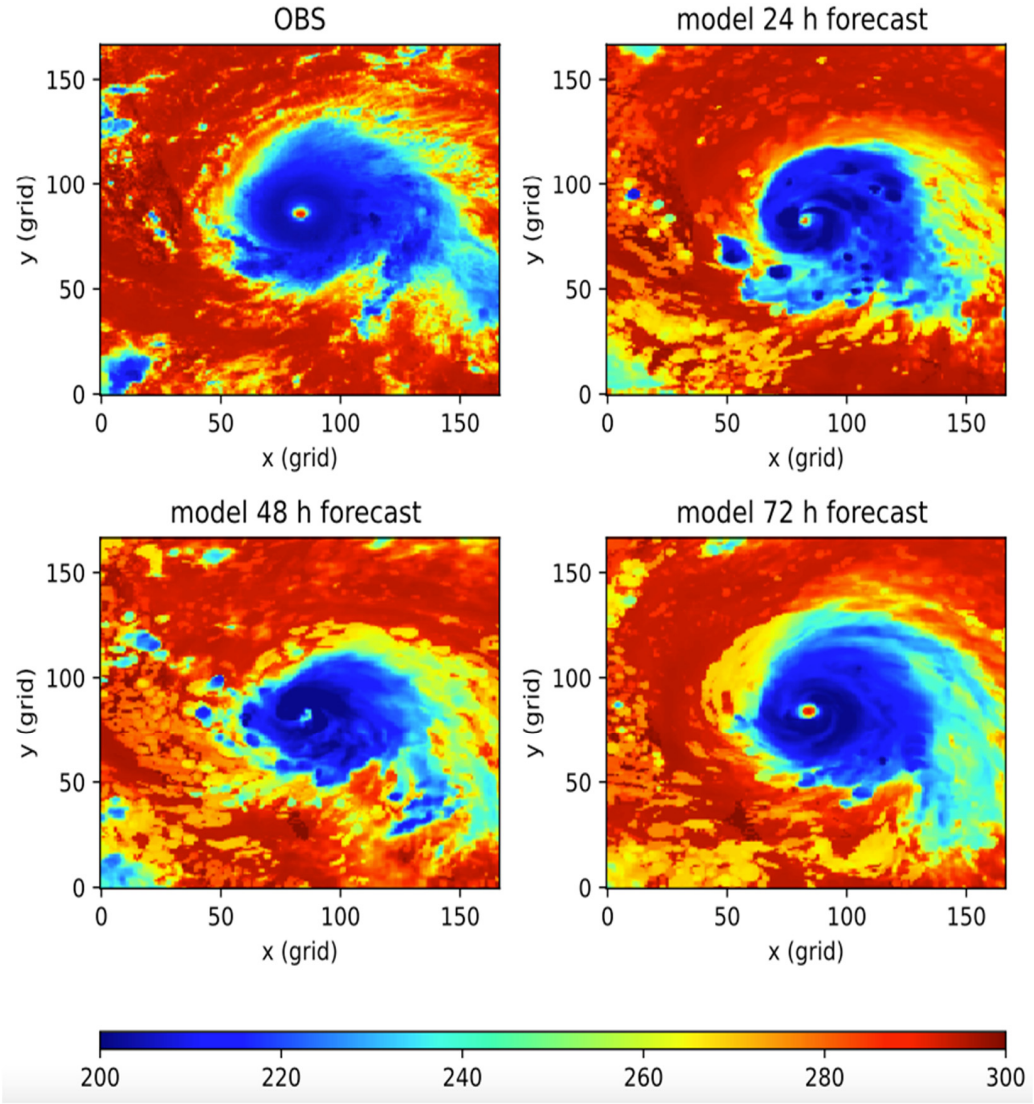


Figure 1. Observed GOES-R IR BT image of hurricane Dorian at 1 September 2019 18Z and model synthetic images at valid time of 1 September 2019 18Z with a lead time of 24, 48 and 72 h.

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 systematic biases or random errors, given the high variability of the model forecast IR BT fields. Use Figure 1 as an illustration: when compared to the observed image, the 24 h forecast vortex has a slightly smaller cloud covered area, the 48 h forecast vortex is significantly smaller, but the 72 h 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 similar to the traditional TC verification methods, which uses average track and intensity error from multiple cases rather than a single forecast snapshot.

4. Results and Discussion

4.1. TC Intensity

Figure 2 depicts the observed and model forecast composite images of hurricane Dorian from 30 August 2019 18Z to 6 September 2019 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 snapshots during this period and the corresponding forecast snapshots obtained from multiple model forecasts initialized at different times. 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 (Table 1) with BT < 220 K, more than the model synthetic image, in which only 543 pixels, most in the eastern half of the grid, have a BT < 220 K. The observed image contains 212 pixels whose BT < 210 K, whereas the model synthetic image contains none. As indicated by the red crosses in Figure 2, the minimum observed BT is 206 K, while the model synthetic minimum BT is 211 K. The model's warm bias is also evident in the scatter plot Figure 3 where the warm bias shifted pixels away from the blue 45° line toward the model side, particularly at the cold end of the spectrum. Similarly, one can see in Figure 4 that, while the two PDF lines match well overall, the model synthetic image accounts for a smaller fraction than the observed BT at the cold end of the BT spectrum.

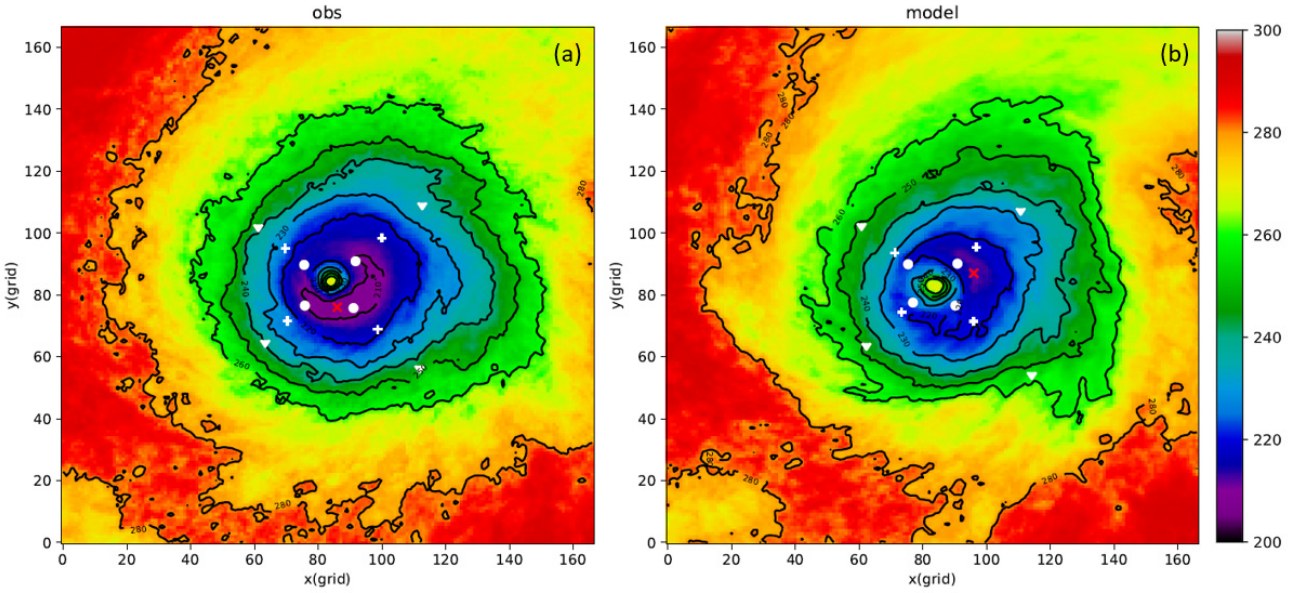


Figure 2. (**a**) observed and (**b**) model synthetic GOES-R IR BT (in K) fields of the mature stage of hurricane Dorian from 30 August 18Z to 6 September 18Z of 2019. The white markers are the radii of the 34 kt (triangles), 50 kt (+signs) and 64 kt (circles) winds from best track in (**a**) and the model forecast track in (**b**). The red crosses indicate the location of minimum observed BT of 206 K and minimum model synthetic BT of 211 K.

Table 1. The number of pixels whose BT are below the thresholds.

BT.	<210 K	<220 K	<230 K	<240 K	<250 K	<260 K	<270 K
Model Area	0	543	1722	2976	4745	7237	13,162
Obs Area	212	1222	2367	3968	5562	8291	14,196

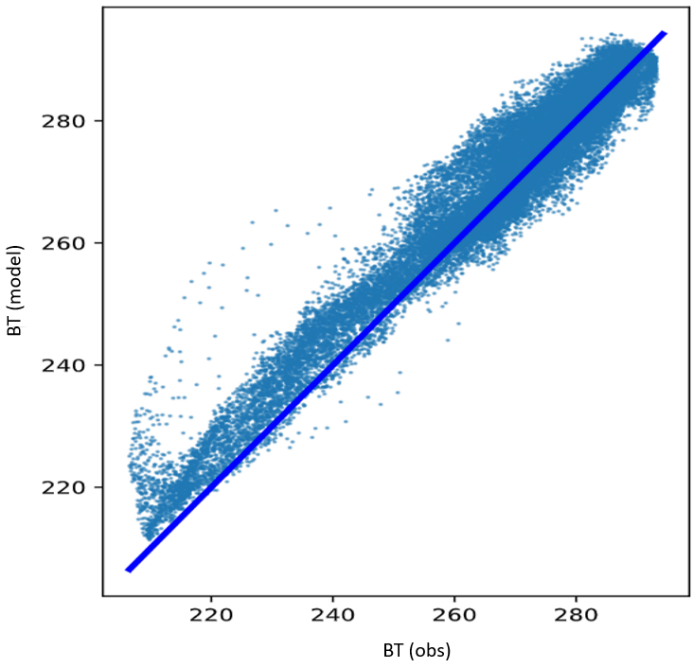


Figure 3. Scatter plot of the observed and model forecast IR BT. The blue line indicates the positions for obs = model.

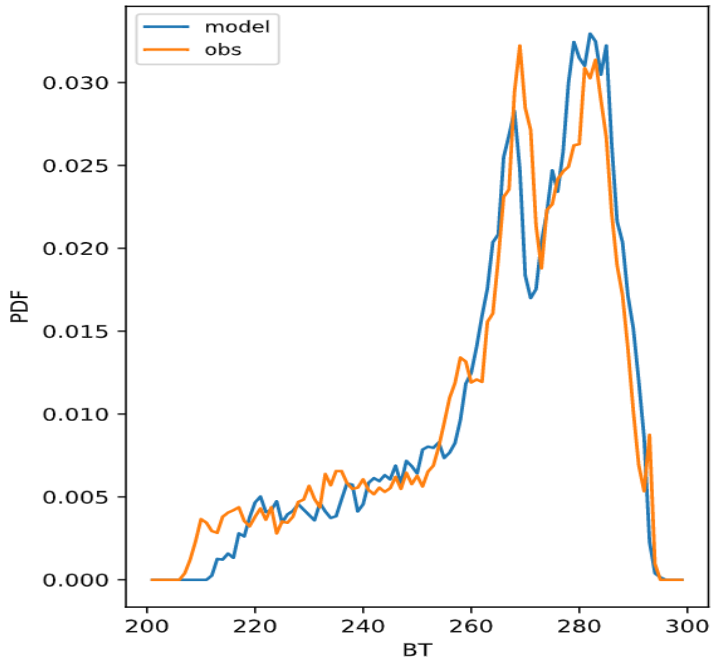


Figure 4. Probability density function (PDF) of the observed and model forecast IR BT.

The cloud top IR 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 IR 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 5, which shows that the model forecasted higher TC minimum sea level pressure and lower maximum wind speed than the best tracks. As a result, the evaluation from the GOES IR BT images are consistent with the traditional TC intensity metrics.

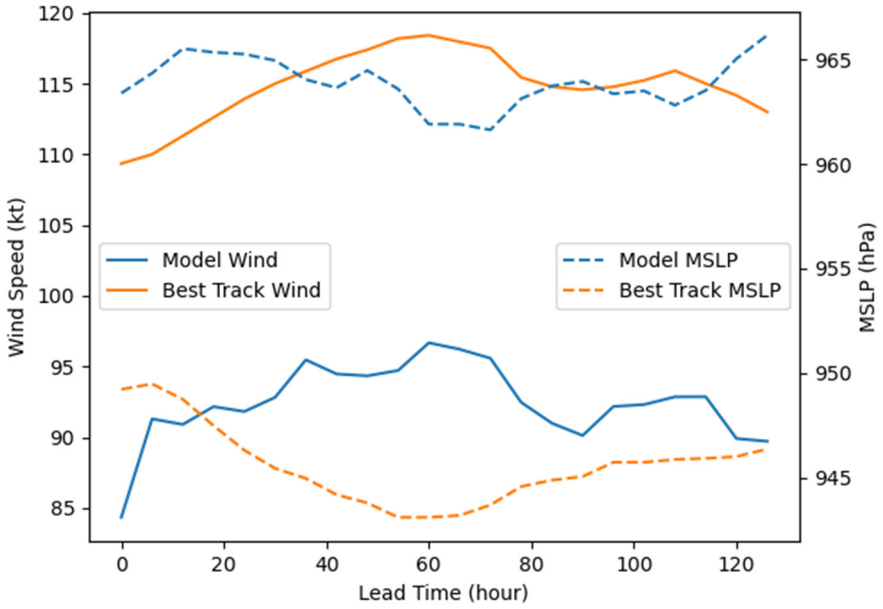


Figure 5. Observed (brown) and model predicted (blue) minimum sea level pressure (dashed line) and maximum wind speed (solid line) of hurricane Dorian (2019) in its mature stage, as a function of the forecast lead time.

Satellite data including IR BT have been used previously to estimate TC strengths, starting from 1960s and 1970s when the first polar-orbiting weather satellites and geostationary weather satellites were launched. Today, the majority of tropical cyclone intensity estimates are also based on satellite observations. The Dvorak technique [11,12] is widely used to estimate the intensity of tropical cyclones at specific times by analyzing tropical cyclone cloud structures derived from visible and infrared satellite images. DeMaria et al. [13,14] developed the Statistical Hurricane Intensity Prediction Program (SHIPS) which predicts TC intensity over the Atlantic Ocean using satellite BT-related predictors. The present study demonstrates that comparing the composite observed and model synthetic images, during which the random differences are removed, can reveal systematic IR BT biases that are highly correlated with the TC intensity biases.

4.2. Size

Along with intensity, the GOES-R IR BT images also reveal the model's forecast TC size bias. In Figure 2a, the contour lines [280 K, 260 K, 240 K, 230 K and 220 K] cover a larger area than the corresponding lines in Figure 2b (also see Table 1), indicating that the model predicted less cloud and hydrometeors coverage than the observed, thus a smaller TC vortex. Traditional metrics corroborate this conclusion as well. Figure 6 demonstrates that the radii of 50-kt and 64-kt winds in 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 model forecast in the NE quadrant and shorter in the other three 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%. It is clear from the 3 critical wind speed radii 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.

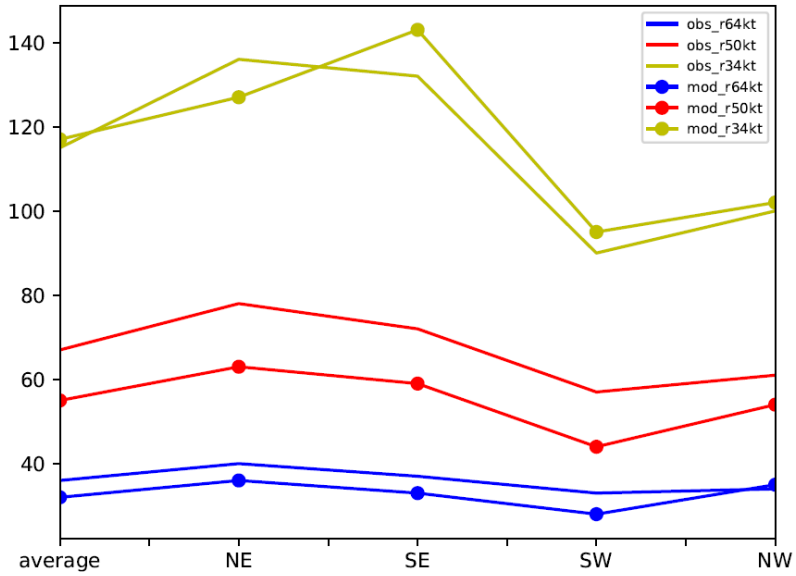


Figure 6. The radii (nautical miles) of winds at 64 kt, 50 kt and 34 kt in the 4 quadrants of NE, SE, SW and NW and the direction-average for hurricane Dorian 2019.

While the model under-predicted the vortex's overall size, it over-predicted the eye size, as illustrated in Figure 2 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 model synthetic image is 90 km, as indicated by the red crosses in Figure 2. Following the technique in Lajoie & Walsh [15], 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.

In [1,2] 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 results has been greatly improved compared to the configuration used by Dong et al. and Hazelton et al. on almost all aspects, including the 34-kt wind radius size errors/biases. However, we do still see larger model predicted R34 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 have changed between their studies and this one. Nevertheless, there is a possibility that it is because in this paper we focused on the mature stage of the Hurricane. Figure 7 shows that the 34-kt wind radii in best track and model forecast agreed well from August 25 to 6 September 2019, but once the hurricane entered its decay stage, the model significantly 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 over-predicted 34-kt wind radii in the decay stage were evenly shared by all forecast lead times (Figure 8), 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 9, which demonstrates that during the decay stage of hurricane Dorian, the HAFS model had a cold BT bias (strong intensity), rather than a warm BT (weak intensity) bias during its mature stage as shown in Figure 2. 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.

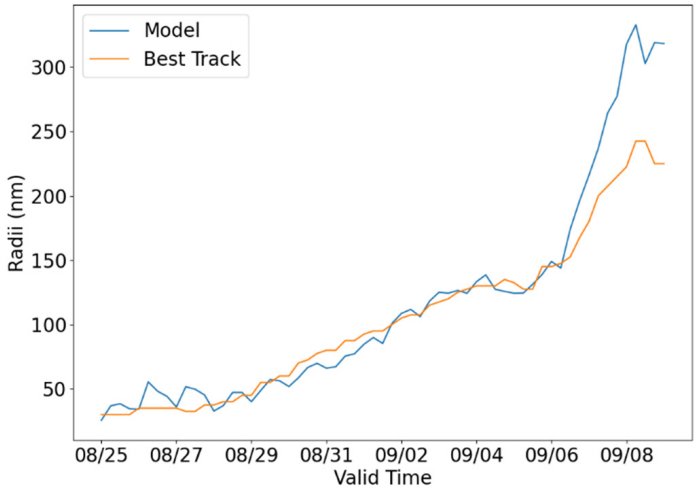


Figure 7. Hurricane Dorian (2019) 34-kt wind radii in best track (brown) and HAFS model verification (blue) by valid time.

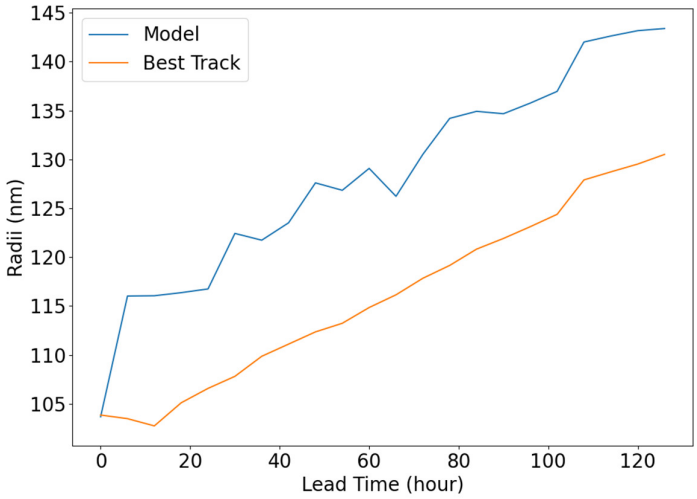


Figure 8. Hurricane Dorian (2019) 34 kt wind radii in best track (brown) and model forecast (blue) by forecast lead time.

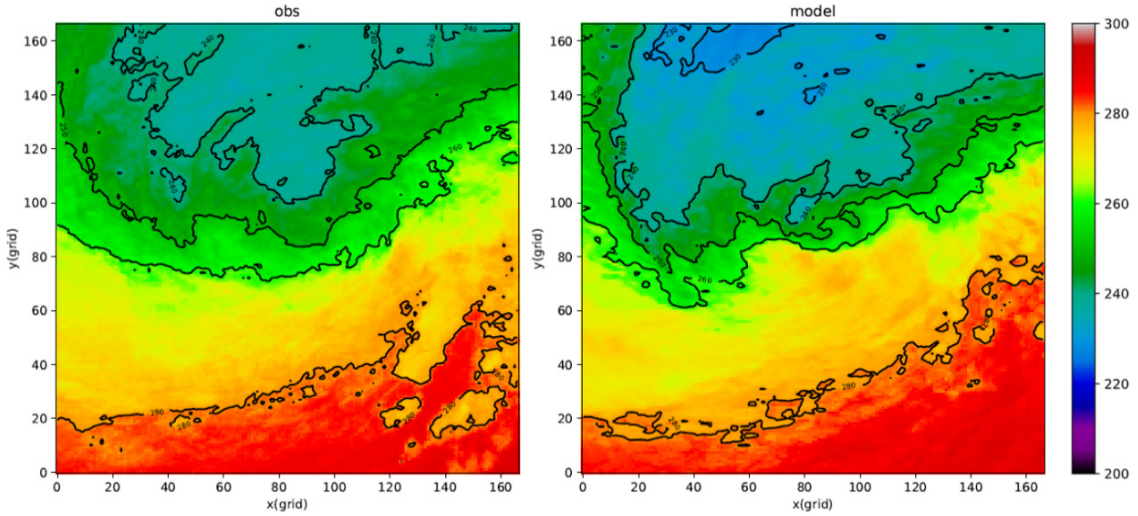


Figure 9. Hurricane Dorian observed and model synthetic GOES-R IR BT in the storm's decay stage from 7 September 2019 to 18 September 2019.

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 is 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. Nonetheless, 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.

4.3. Asymmetry

Figure 2 also depicts the similar asymmetric structures of the observed and model forecast TC vortices. The areas covered by cloud and hydrometeors in both composite observed and model images, as indicted by the IR BT contour lines (BT < 280 K), extended more toward and even beyond the domain boundaries in the NE and SE directions, whereas large cloud-free areas (BT > 280 K) are exposed in the NW and SW directions. The same asymmetric structure can be quantified in Figure 10, 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 NW, highlighting the asymmetric structure of the TC vortices that extended toward NE direction with the eyes located in the SW part of the areas covered by cloud and hydrometers. The asymmetric TC vortices 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 2 as triangle markers (34-kt radii), "+" markers (50-kt radii) and round markers (64-kt radii). These markers clearly show a spatial distribution with the largest distance from the eye in the NE direction and smallest in SW; more specifically, r_NE > r_SE > r_NW > r_SW, which is consistent with the asymmetric structure depicted in Figure 10. The distribution of radii is illustrated more quantitatively in Figure 6, which depicts the largest wind radii in the NE direction and the smallest in the SW direction, as well as the wind radii in SE and NW directions in between. The only exception is that the model predicted 34-kt wind radius is larger in SE than NE.

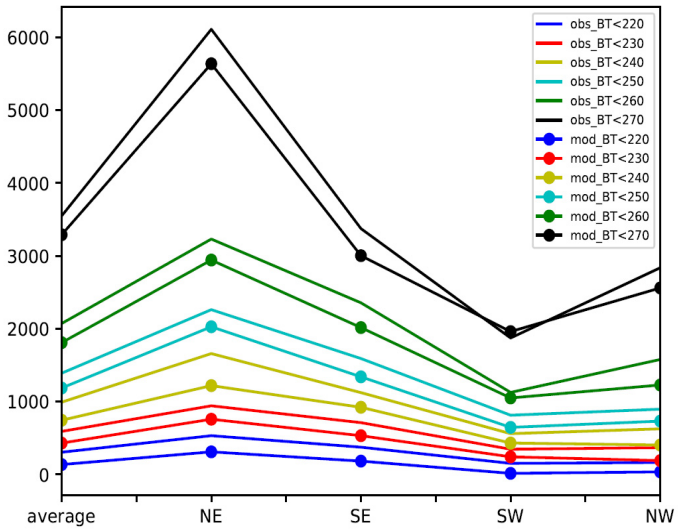


Figure 10. The number of pixels in the observed and model synthetic GOES-R IR BT images that are less than the certain thresholds (220 K, 230 K, 240 K, 250 K, 260 K and 270 K) in the 4 quadrants of NE SE SW and NW and the direction-averaged number.

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

Comparing observations and synthetic satellite images of the model can reveal the model's systematic deviations in predicting TC intensity, size, and asymmetry. These results are consistent with those obtained by traditional metrics. At the same time, satellite images can provide richer information, because they can provide continuous coverage in space and time, so they can provide near real-time and more detailed 2D structures within global coverage.

5. Conclusions

HAFS is the next generation hurricane prediction and analysis system based on the FV3-GFS dynamic core with tropical cyclone-specific physics schemes. The user-friendly python interface of the CRTM radiative transfer model is used to convert the HAFS forecast to synthetic GOES-R IR BT images, which are then compared to the corresponding observed images. Uncertainties in TC forecasts includes both random errors and systematic biases. Rather than comparing observed and model synthetic images at individual discrete times, which will be affected by the 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 model's systematic biases. The composite images from 24 forecasts of hurricane Dorian (2019) in its mature stage is used for the evaluation. Results revealed that the model exhibits a warm IR BT bias, indicating lower cloud altitude and thus weaker convections and TC intensity. The warm bias can also be seen in scatter plot and 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 are 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 the model forecast vortices, both showed similar patterns of vortices extending furthest to the NE quadrant, followed by SE, then NW, and shortest to SW quadrant, which is also confirmed by the differences in the 34-kt, 50-kt and 64-kt winds 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 maximum wind speed, which is not always available in traditional verification products. Furthermore, the images reveal that the hurricane Dorian overpredicted intensity during its decay stage, possibly resulting in significantly larger radii of 34-kt wind than in best track, which contaminated the traditional vortex size verification when the 34kt wind radii is plotted with x-axis being the forecast lead time (instead of valid time).

Note since this study only evaluated one hurricane, the above conclusions could be only applicable to the model behavior of the mature stage of hurricane Dorian in 2019, and caution should be exercised when extrapolating these conclusions to anticipate model biases in predicting other TCs or other stages. 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-times 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.

Author Contributions: Conceptualization, S.B.; methodology, S.B.; software, S.B.; validation, S.B., Z.Z., E.K. and B.L.; formal analysis, S.B., Z.Z., E.K. and B.L.; investigation, S.B., Z.Z., E.K. and B.L.; resources, S.B.; data curation, Z.Z. and B.L.; writing—original draft preparation, S.B.; writing—review and editing, S.B., Z.Z., E.K. and B.L.; visualization, S.B.; supervision, S.B.; project administration, S.B.; funding acquisition, S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NOAA grant number NA18NWS4680052 and the Developmental Tested Center's visitor program. This work received funding support from (1) NOAA NGGPS grant NA18NWS4680052 to use remote sensing data to evaluation FV3-GFS model and (2) the Developmental Testbed Center (DTC) Visitor Program. The DTC is funded by the National Oceanic and Atmospheric Administration, the Air Force, the National Center for Atmospheric Research, and the National Science Foundation. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Data Availability Statement: Data is available upon request.

Acknowledgments: The computation has been conducted using resources from the NOAA Research and Development High Performance Computing Program (URL: http://rdhpcs.noaa.gov (accessed 7 January 2022). The authors would like to thank the three anonymous reviewers for their thoughtful comments and suggestions towards improving the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hazelton, A.; Zhang, Z.; Liu, B.; Dong, J.; Alaka, G.; Wang, W.; Marchok, T.; Mehra, A.; Gopalakrishnan, S.; Zhang, X. 2019 Atlantic Hurricane Forecasts from the Global-Nested Hurricane Analysis and Forecast System: Composite Statistics and Key Events. *Weather Forecast.* **2021**, *36*, 519–538. [CrossRef]
2. Dong, J.; Liu, B.; Zhang, Z.; Wang, W.; Mehra, A.; Hazelton, A.T.; Winterbottom, H.R.; Zhu, L.; Wu, K.; Zhang, C. The Evaluation of Real-Time Hurricane Analysis and Forecast System (HAFS) Stand-Alone Regional (SAR) Model Performance for the 2019 Atlantic Hurricane Season. *Atmosphere* **2020**, *11*, 617. [CrossRef]
3. Bao, S. Use Satellite Infrared Brightness Temperature Data to Evaluate HWRF Ferrier-Aligo Microphysics Scheme. 2018. Available online: https://dtcenter.org/sites/default/files/visitor-projects/dtc-visitor-report-2018-Bao-final.pdf (accessed on 6 January 2022).
4. Bao, S.; Bernardet, L.; Thompson, G.; Kalina, E.; Newman, K.; Biswas, M. Impact of the Hydrometeor Vertical Advection Method on HWRF's Simulated Hurricane Structure. *Weather Forecast.* **2020**, *35*, 723–737. [CrossRef]
5. Cintineo, R.; Otkin, J.A.; Xue, M.; Kong, F. Evaluating the Performance of Planetary Boundary Layer and Cloud Microphysical Parameterization Schemes in Convection-Permitting Ensemble Forecasts Using Synthetic GOES-13 Satellite Observations. *Mon. Weather Rev.* **2014**, *142*, 163–182. [CrossRef]
6. Jin, Y.; Wang, S.; Nachamkin, J.; Doyle, J.D.; Thompson, G.; Grasso, L.; Holt, T.; Moskaitis, J.; Jin, H.; Hodur, R.M.; et al. The Impact of Ice Phase Cloud Parameterizations on Tropical Cyclone Prediction. *Mon. Weather Rev.* **2014**, *142*, 606–625. [CrossRef]
7. Novak, K.; Bao, S. Evaluation of FV3 Model Using Satellite Brightness Temperature Data. *AGU Fall Meeting Abstracts*. 2019, Volume 2019, p. A21O-2771. Available online: https://ui.adsabs.harvard.edu/abs/2019AGUFM.A21O2771N/abstract (accessed on 6 January 2022).
8. Otkin, J.A.; Lewis, W.E.; Lenzen, A.J.; McNoldy, B.D.; Majumdar, S.J. 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]
9. Han, Y. JCSDA Community Radiative Transfer Model (CRTM): Version 1. 2006. NOAA Technical Report NESDIS (122). Available online: https://repository.library.noaa.gov/view/noaa/1157 (accessed on 6 January 2022).
10. Marchok, T. Use of the GFDL Vortex Tracker. *WRF Tutor. Hurric. NOAAGFDL* **2010**, *40*. Available online: https://dtcenter.org/sites/default/files/community-code/hwrf/docs/presentations/tutorial2018/16-HWRFtutJan2018_VortexTracker_Marchok.pdf (accessed on 6 January 2022).
11. Dvorak, V.F. Tropical Cyclone Intensity Analysis and Forecasting from Satellite Imagery. *Mon. Weather Rev.* **1975**, *103*, 420–430. [CrossRef]
12. Dvorak, V.F. *Tropical Cyclone Intensity Analysis Using Satellite Data*; US Department of Commerce, National Oceanic and Atmospheric Administration, 1984; Volume 11. Available online: https://repository.library.noaa.gov/view/noaa/19322 (accessed on 6 January 2022).
13. DeMaria, M.; Kaplan, J. A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic Basin. *Weather Forecast.* **1994**, *9*, 209–220. [CrossRef]

14. DeMaria, M.; Zehr, R.M.; Kossin, J.P.; Knaff, J.A. The Use of GOES Imagery in Statistical Hurricane Intensity Prediction. In Proceedings of the 25th Conference on Hurricanes and Tropical Meteorology, San Diego, CA, USA, 29 April 2002; Volume 120, p. 121.
15. Lajoie, F.; Walsh, K. A Technique to Determine the Radius of Maximum Wind of a Tropical Cyclone. *Weather Forecast.* **2008**, *23*, 1007–1015. [CrossRef]