

Article

The Evaluation of Real-Time Hurricane Analysis and Forecast System (HAFS) Stand-Alone Regional (SAR) Model Performance for the 2019 Atlantic Hurricane Season

Jili Dong ^{1,2,*}, Bin Liu ^{1,2}, Zhan Zhang ^{1,2}, Weiguo Wang ^{1,2}, Avichal Mehra ², Andrew T. Hazelton ^{3,4} , Henry R. Winterbottom ^{1,2}, Lin Zhu ^{1,2}, Keqin Wu ^{1,2}, Chunxi Zhang ^{1,2}, Vijay Tallapragada ², Xuejin Zhang ⁴, Sundararaman Gopalakrishnan ⁴ and Frank Marks ⁴ 

- ¹ I. M. System Group, Inc. (IMSG), Rockville, MD 20852, USA; Bin.Liu@noaa.gov (B.L.); Zhan.Zhang@noaa.gov (Z.Z.); Weiguo.Wang@noaa.gov (W.W.); Henry.Winterbottom@noaa.gov (H.R.W.); Lin.L.Zhu@noaa.gov (L.Z.); Keqin.Wu@noaa.gov (K.W.); Chunxi.Zhang@noaa.gov (C.Z.)
- ² National Oceanic and Atmospheric Administration (NOAA)/Environmental Modelling Center (EMC), College Park, MD 20740, USA; Avichal.Mehra@noaa.gov (A.M.); Vijay.Tallapragada@noaa.gov (V.T.)
- ³ University of Miami Cooperative Institute for Marine and Atmospheric Studies (CIMAS), Miami, FL 33149, USA; Andrew.Hazelton@noaa.gov
- ⁴ NOAA/Atlantic Oceanographic and Meteorological Laboratory (AOML)/Hurricane Research Division (HRD), Miami, FL 33149, USA; Xuejin.Zhang@noaa.gov (X.Z.); Sundararaman.G.Gopalakrishnan@noaa.gov (S.G.); Frank.Marks@noaa.gov (F.M.)
- * Correspondence: Jili.Dong@noaa.gov

Received: 15 May 2020; Accepted: 7 June 2020; Published: 11 June 2020



Abstract: The next generation Hurricane Analysis and Forecast System (HAFS) has been developed recently in the National Oceanic and Atmospheric Administration (NOAA) to accelerate the improvement of tropical cyclone (TC) forecasts within the Unified Forecast System (UFS) framework. The finite-volume cubed sphere (FV3) based convection-allowing HAFS Stand-Alone Regional model (HAFS-SAR) was successfully implemented during Hurricane Forecast Improvement Project (HFIP) real-time experiments for the 2019 Atlantic TC season. HAFS-SAR has a single large 3-km horizontal resolution regional domain covering the North Atlantic basin. A total of 273 cases during the 2019 TC season are systematically evaluated against the best track and compared with three operational forecasting systems: Global Forecast System (GFS), Hurricane Weather Research and Forecasting model (HWRF), and Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic model (HMON). HAFS-SAR has the best performance in track forecasts among the models presented in this study. The intensity forecasts are improved over GFS, but show less skill compared to HWRF and HMON. The radius of gale force wind is over-predicted in HAFS-SAR, while the hurricane force wind radius has lower error than other models.

Keywords: tropical cyclones; numerical weather prediction; high resolution tropical cyclone forecasts; finite-volume cubed sphere (FV3); Hurricane Analysis and Forecast System (HAFS); Stand-Alone Regional (SAR) models

1. Introduction

A tropical cyclone (TC) is one of the major devastating natural disasters often resulting in loss of lives and property damage. Numerical weather prediction (NWP) developments, specific to TCs, have progressed significantly over the last decade. In the United States, the National Hurricane Center (NHC) makes use of both global and high-resolution regional dynamical models for TC forecast

guidance during the hurricane season, including the Navy Operational Global Atmospheric Prediction System (NOGAPS) model, the United Kingdom METeorological (UKMET) model, the Canadian Meteorological Centre (CMC) model, the Japan Meteorological Agency's (JMA) Global Spectral Model (GSM) model, and the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) model as global models and the Navy's Coupled Ocean/Atmosphere Mesoscale Prediction System for Tropical Cyclones (COAMPS-TC) model as regional models, among others.

Currently, three NWP models are operated by the National Centers for Environmental Prediction (NCEP) at National Weather Service (NWS), in order to provide the essential guidance for TC prediction in the United States. These models include a global model known as the Global Forecast System (GFS), and two regional high-resolution NWP models, both with moving nests, known as the Hurricane Weather Research and Forecasting model (HWRF), and the Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic model (HMON).

The operational GFS underwent a significant upgrade during 2019 to implement a new dynamical core, known as the Finite-Volume cubed sphere (FV3) developed at the National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL). FV3 applies the finite volume method on the gnomonic grid on a cubed sphere with computational efficiency. The FV3-based global model developed in GFDL (fvGFS), using a 13 km resolution, demonstrated better performance of FV3 for both TC track and intensity forecasts, compared to the previous GFS spectral model [1]. However, the horizontal resolution of the above-mentioned global models is too coarse for simulating convective-scale TC structures [2–4]. The high-resolution convection-allowing NWP models are able to resolve important inner-core processes, such as eyewall replacement cycles and the structures within the eye, eyewall, and spiral rainbands of the TC [5–11].

The operational high-resolution regional HWRF and HMON models have been recently upgraded, resulting in substantial improvements for TC track and intensity forecasts [12]. The very high resolutions of the innermost nested domains of HWRF and HMON (1.5 km and 2 km, respectively) allow the forecast to better resolve deep convection and the TC inner core structures. At the same time, the moving nest introduces complications and uncertainties which include, but are not limited to, the re-initialization (zeroing) of the nested domain variables when storm moves, the discrepancy from the interpolation of parent domain to the moving nest, the treatment of topography and land-sea mask with the moving nest, and the sensitivity of TC forecasts skill to the dimension of the moving nest. The current operational HWRF and HMON are also not designed to process multiple movable nests within the same parent domain, which limits their ability to forecast multiple TCs simultaneously, while also possibly missing important TC to TC interactions during the instances of multiple TCs. The experimental basin-scale HWRF developed in Atlantic Oceanographic and Meteorological Laboratory (AOML), in collaboration with Environmental Modeling Center (EMC) and Developmental Testbed Center (DTC), is able to address this deficiency of HWRF and HMON TCs interaction, using multiple moving nests in a larger (e.g., basin-scale) parent domain. This configuration, as demonstrated, improves track forecasts when compared to the operational HWRF during instances of multiple TCs [13,14].

To accelerate improvements in TC intensity and track forecasts within a unified global and regional modelling framework, known as the Unified Forecast System (UFS), the next generation hurricane forecasting system, the Hurricane Analysis and Forecast System (HAFS), has been developed amongst collaborators within NOAA. HAFS utilizes the FV3-based global-regional modelling system for TC prediction. The system can be applied in either a high-resolution stand-alone regional configuration (HAFS-SAR), or using a uniform global model with a high-resolution static nest configuration (HAFS-global-nest). A high-resolution FV3 with a setup similar to the HAFS-global-nest has been successfully tested for both seasonal deterministic TC forecasts [15] and ensemble forecast case study [16]. The global-nest FV3 has demonstrated skillful track forecasts and improved intensity forecasts, when compared to the spectral GFS, during the 2017 hurricane season [15]. The intensity performance, however, is less skillful than HWRF. While the global-nest FV3 suggests promising results, running both a global model and a nest domain concurrently is computationally expensive. HAFS-SAR

applies the FV3 dynamic core on a regional grid with external lateral boundary conditions, and can reduce the computational cost significantly, by removing the necessity of the parent global domain used for the HAFS-global-nest configuration. The limited area HAFS-SAR model has been configured for the North Atlantic (NATL) basin for TC forecasts (Figure 1), using a single large domain with improved planetary boundary layer (PBL) and surface flux parameterization schemes designed/calibrated specifically for TC simulations. A resolution of 3 km is chosen for convection-allowing TC forecasts.

The HAFS-SAR was successfully implemented during 2019 Hurricane Forecast Improvement (HFIP) real-time experiments for the NATL TC season. This was the first time the FV3-based SAR was systematically tested specifically for TC forecasts. In this study, the performance and forecast skill of HAFS-SAR will be evaluated and compared to the three NOAA operational NWP models. In Section 2, the real-time experiment is described, while the intensity, track and storm size (wind-radii) forecasts are evaluated in Section 3. The HAFS-SAR performance is further examined for two high-impact TC events during the 2019 NATL TC season in Section 3, while Section 4 summarizes the results of this study.

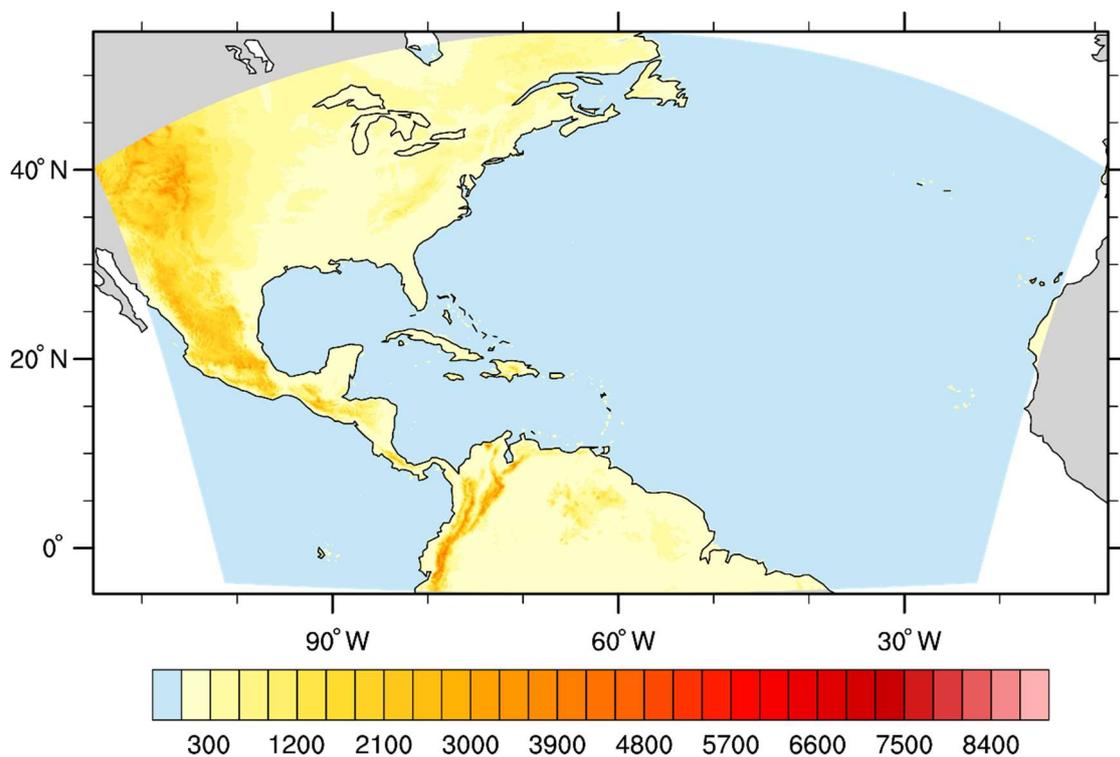


Figure 1. Topography (m) of the high-resolution stand-alone regional configuration (HAFS-SAR) domain.

2. Experiments

The FV3 dynamical core [17,18] for HAFS-SAR includes a non-hydrostatic finite volume solver using a Lagrangian vertical coordinate [19]. The HAFS-SAR has a single large domain of 2880×1920 grid-cells. The domain is centered at 62° W, 22° N, spanning 60° from south to north and 106.5° from east to west. The north Atlantic basin covers the main development region (MDR) for TCs, and extends northward as far as to Newfoundland, in order to capture TC genesis and extratropical transition. The HAFS-SAR is based on one of the 6 faces of the global cubed sphere C768 (~ 13 km grid spacing globally) and is further refined by a factor of 4. The computational grid is gnomonic, and has an average cell size (defined as square root of cell area) of 3.2 km. The cell size gradually increases from 2.6 km around the domain edges to 3.6 km in the domain center. The HAFS-SAR uses 64 vertical levels on a sigma-pressure hybrid coordinate with the lowest model level at about 25 m

above the surface and the top isobaric level of 0.2-hPa. The terrain data are interpolated from the 30-s (~1 km) USGS GMTED2010 dataset.

Initial and boundary conditions were interpolated from the 2019 operational global FV3-based GFS (~13-km) onto the 3-km HAFS SAR domain. Lateral boundary conditions (LBCs) were provided every 3 h from the same global GFS forecasts. In this study, no data assimilation is performed for the HAFS-SAR.

The HAFS-SAR physics parameterizations include the Hybrid Eddy-Diffusivity Mass-Flux (HEDMF) PBL scheme [20]. The eddy diffusivity over water in the HEDMF PBL scheme is modified based on observations, to better resolve the TC boundary layer processes in TC, similar to that used in HWRF [21]. The exchange coefficient of momentum C_d and heat C_k for the surface flux parameterization follows the same formula of the operational HWRF over water under strong wind conditions, in order to be more consistent with the observations [22]. The microphysics scheme is the GFDL 6-category hydrometeors scheme [23], while the land surface parameterization is the Noah land surface scheme [24]. The radiation schemes are RRTMG longwave and shortwave parameterizations [25,26], and the cumulus convection is turned off at the 3-km resolution. Finally, the SST is provided by the GFS Near SST (NSST) scheme, which predicts the vertical profile of sea temperature between the surface and a reference level.

The 2019 real-time HAFS-SAR experiments started 0000 UTC 12 July and ended 0000 UTC 01 November, covering eighteen storms consisting of TC Barry through TC Rebekah. The real-time experiments were initialized at 0000, 0600, 1200, and 1800 UTC each day and ran for 126 h. The GFDL vortex tracker [22,27] was used to generate the TC track information, including TC center locations, the maximum 10-m wind and minimum pressure intensities, and TC wind radii of gale force wind (17.5 m s^{-1}), damaging force wind (25.7 m s^{-1}) and hurricane force wind (32.9 m s^{-1}). The model products in GRIB2 format and the Automated Tropical Cyclone Forecasting System [28] (ATCF: the application to automate and optimize TC forecasts for operational centers) track information were delivered at 0930, 1530, 2130, and 0330 (next day) UTC every day during the TC season. A total of 273 TC cases are included. In order to evaluate the performance of HAFS-SAR real-time experiments, the forecasted TC track, intensity, and size are verified against the best track data from the NHC and compared with the operational GFS, HWRF, and HMON systems. The models for comparison are listed in Table 1. Both the HWRF and HMON models have vortex initializations implemented, while additional inner core data assimilation is used in HWRF (Table 1). The HWRF model is coupled with the Princeton Ocean Model (POM) and the HMON model is coupled with the Hybrid Coordinate Ocean Model (HYCOM). As demonstrated in the HWRF model, both the inner core data assimilation and the ocean coupling can potentially further improve the TC track and intensity forecasts [29,30]. The data assimilation/vortex initialization and ocean coupling capability are being developed for HAFS-SAR. These new capabilities are expected to present the more realistic vortex structure and ocean response for the HAFS-SAR in the future.

Table 1. Four models for verification in the 2019 North Atlantic tropical cyclone (TC) season.

	HAFS-SAR	GFS	HWRF	HMON
Global/regional	Regional	Global	Regional	Regional
Dynamic core	FV3	FV3	WRF-NMM ¹	NMMB ²
Resolution	~3 km	~13 km	13.5-4.5-1.5 km	18-6-2 km
VI³/inner core DA	N/A	N/A	VI+inner core DA	VI
Ocean coupling	N/A	N/A	POM ⁴	HYCOM ⁵

¹ Nonhydrostatic Mesoscale Model-(NMM). ² Nonhydrostatic Mesoscale Model on B grid-(NMMB). ³ Vortex initialization (VI) ⁴ Princeton Ocean Model (POM) ⁵ Hybrid Coordinate Ocean Model (HYCOM).

3. Results

The seasonal statistics of HAFS-SAR track, intensity and size forecasts are presented and compared with the GFS, HWRF, and HMON in this section.

3.1. Track Forecast

The track forecasts are verified by computing the track error defined as the great circle distance between the predicted TC center and the best track location. Track errors are plotted every forecast 6 h until 48 h, and then every 24 h until 120 h for each of the four models. The initial location errors for HWRF and HMON are smaller than GFS and the HAFS-SAR (Figure 2a), because of the vortex relocation methods of HWRF and HMON [22]. Both HAFS-SAR and GFS have relatively small initial track errors of 22 km, even without relocating TC, and likely benefit from the cycled satellite data assimilation in GFS. Beginning from the 6 h forecast, the track error growth of HAFS-SAR is slower than the other three models. The resulting track error of HAFS-SAR is consistently the smallest from 12–120 forecast hours. The 95% confidence intervals (CI) are also plotted for the track error as the vertical bars (Figure 2a). The error of HAFS-SAR is significantly smaller than the other three models during 24–72 h forecast periods, when the upper limit of the 95% CI of HAFS-SAR is always lower than the lower limit of the other three models (Figure 2a). Among the three operational models, HWRF appears to have the largest track error, while GFS was the second best performer for track forecasts.

The track forecast skill is also calculated by comparing the track error of each model to HWRF as follows:

$$\text{Skill} = (\text{error}_{\text{HWRF}} - \text{error}_{\text{model}}) / \text{error}_{\text{HWRF}} \quad (1)$$

A positive (negative) value represents a smaller (larger) forecast error and better (worse) performance compared to HWRF. HAFS-SAR, GFS and HMON all show positive track forecast skill at most forecast hours (Figure 2b), indicating improvement over HWRF. The track forecast improvement of HAFS-SAR, when compared to HWRF, is roughly 20% during the 12–120 h forecast period.

The total track forecast error can be further decomposed into along-track and cross-track directional error, to examine the contribution from the two respective components. The along-track error of HAFS-SAR is very close to HMON and smaller than GFS and HWRF prior to 96 h (Figure 2c). HAFS-SAR demonstrates more advantage in reducing cross-track error (Figure 2d). After 24 h, the HAFS-SAR cross-track error is always the smallest of the four models out to 120 h, with the error at 72 h being only half those of HWRF and HMON, thus suggesting that cross-track component contributes more to the track forecast improvement than along-track component.

The along- and cross-track biases are illustrated in Figure 2e,f, to further examine track forecast skills. Positive (negative) along-track bias means the predicted storm moved faster (slower) in the along-track direction than the observed TC (as denoted by the best track). Positive (negative) cross-track bias represents how far to the right (left) the predicted storm is relative to the observed track. The cross-track bias is very low for HAFS-SAR along with GFS prior to 72 h (Figure 2f), suggesting HAFS-SAR stays closely in the observed TC moving direction. After 72 h, a positive cross-track bias is observed for both HAFS-SAR and GFS, implying that both models move to the right of the observed track. The along-track bias is negative for most of the models during most of the forecast lead times (Figure 2e), suggesting slower moving speed along the observed TC direction.

In summary, HAFS-SAR performs noticeably better for the track forecast metric, when compared to the operational GFS, HWRF, and HMON. The improvement is mostly from the reduction of the cross-track errors. The comparison of HAFS-SAR to GFS is of particular interest, given that both models use the same dynamical core and initial conditions. The boundary conditions for HAFS-SAR are also from the GFS forecasts. The 3-km high resolution likely contributes to the track forecast improvements (Figure 2 of Xue et al. [4]). Further, HAFS-SAR also applies a different horizontal advection scheme than GFS, and also does not apply a cumulus convective parameterization scheme. Both factors may affect the large scale environment forecast and, thus, the TC track forecasts.

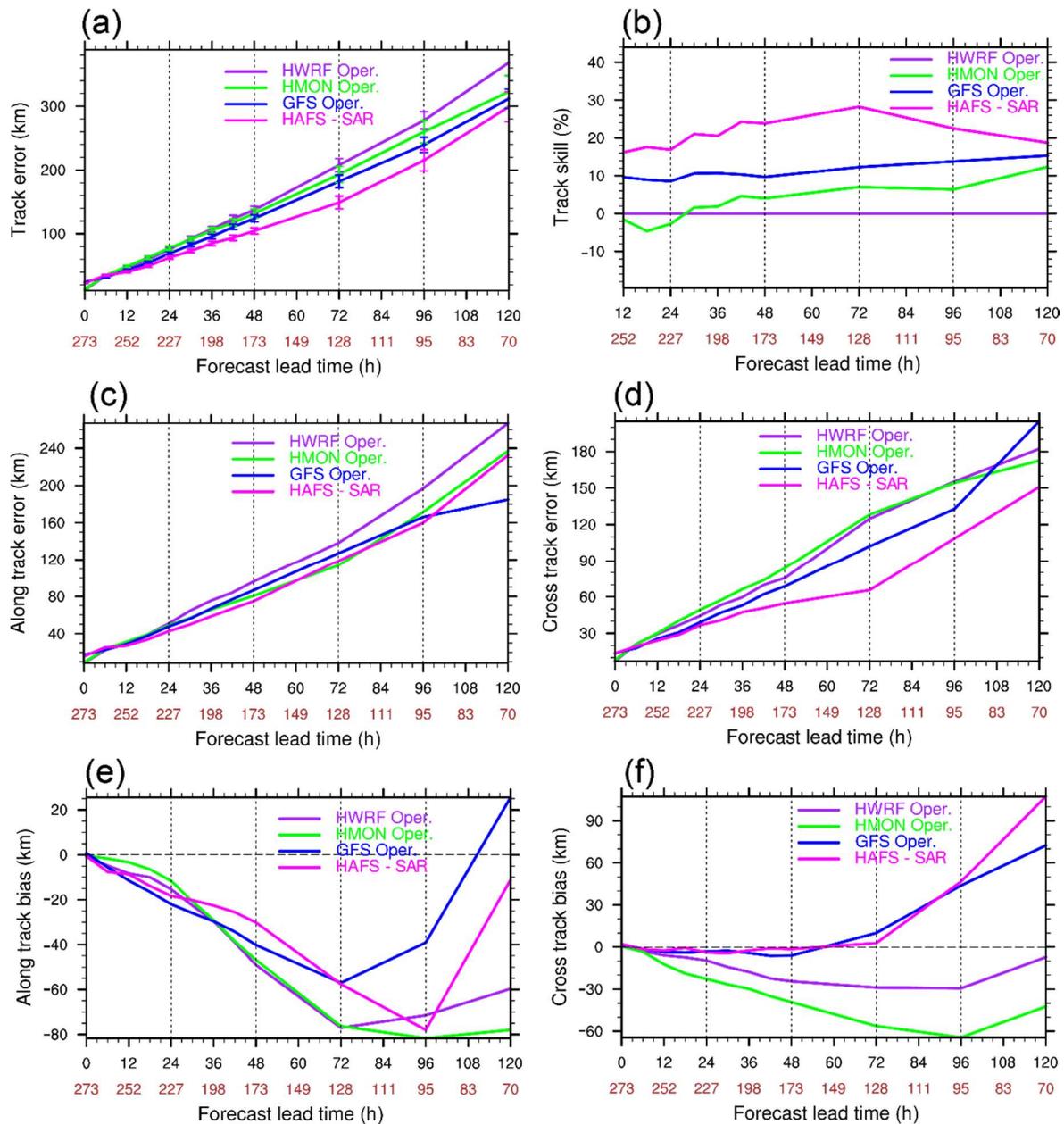


Figure 2. The 2019 seasonal track statistics for HAFS-SAR, Global Forecast System (GFS), Hurricane Weather Research and Forecasting model (HWRF), and Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic model (HMON): (a) track forecast error; (b) track forecast skill relative to HWRF; (c) along-track error; (d) cross-track error; (e) along-track bias; (f) cross-track bias. The error bar in (a) represents a 95% confidence interval. Case number is denoted in brown.

3.2. Intensity Forecast

Intensity forecasts, described by the maximum 10-m wind speed, for HAFS-SAR are examined and compared to GFS, HWRF, and HMON (Figure 3). Due to the lack of inner-core data assimilation and vortex initialization contained in HWRF and HMON [22], both HAFS-SAR and GFS have a relatively large initial intensity error of 6.2 m s^{-1} , compared to errors around 1.5 m s^{-1} for HWRF and HMON (Figure 3a). The simulated TC of HAFS-SAR spins up quickly with its intensity error decreasing to 4.6 m s^{-1} in 6 h, and reduced to 4.1 m s^{-1} at 12 h. The error of HAFS-SAR grows during the 12–72 h forecast period, appears to saturate by 72–96 h, and decreases during 96–120 h period. The use of high-resolution models is expected to improve the intensity forecasts by simulating more realistic

convective-scale inner-core structures [3]. Beginning at 6 h, the intensity error of HAFS-SAR is always lower than GFS, and is only 56% of GFS at 120 h. The performance of HAFS-SAR is significantly better than GFS from 6–24 h and at the lead time of 42 h (error bars in Figure 3a). When compared with HWRF and HMON, the intensity forecasts of the 3-km HAFS-SAR falls behind the finer resolution models before day 5. At 120 h, HAFS-SAR has the lowest intensity error among all four models. It is also likely a result of its vastly improved track forecast at this forecast lead time (Figure 2). The intensity forecast skill (Figure 3b) is consistent with the trend of intensity errors. HAFS-SAR has a negative forecast skill compared to HWRF except at 120 h and also has a lower skill compared to HMON before 72 h. However, the skill is always higher than GFS at all forecast lead times.

The wind bias indicates whether the respective NWP model predicts a stronger (positive) or weaker (negative) storm. Most of the models under-predict the intensity with negative wind bias in most of forecast hours (Figure 3c). The bias of HAFS-SAR is similar to HWRF during the 12–72 h period and turns into a positive 1.5 m s⁻¹ at 120 h, suggesting a slight over-prediction of TC intensity. The number of verifying forecasts is lower at 120 h as well, which may cause a sampling bias (e.g., stronger storms last longer).

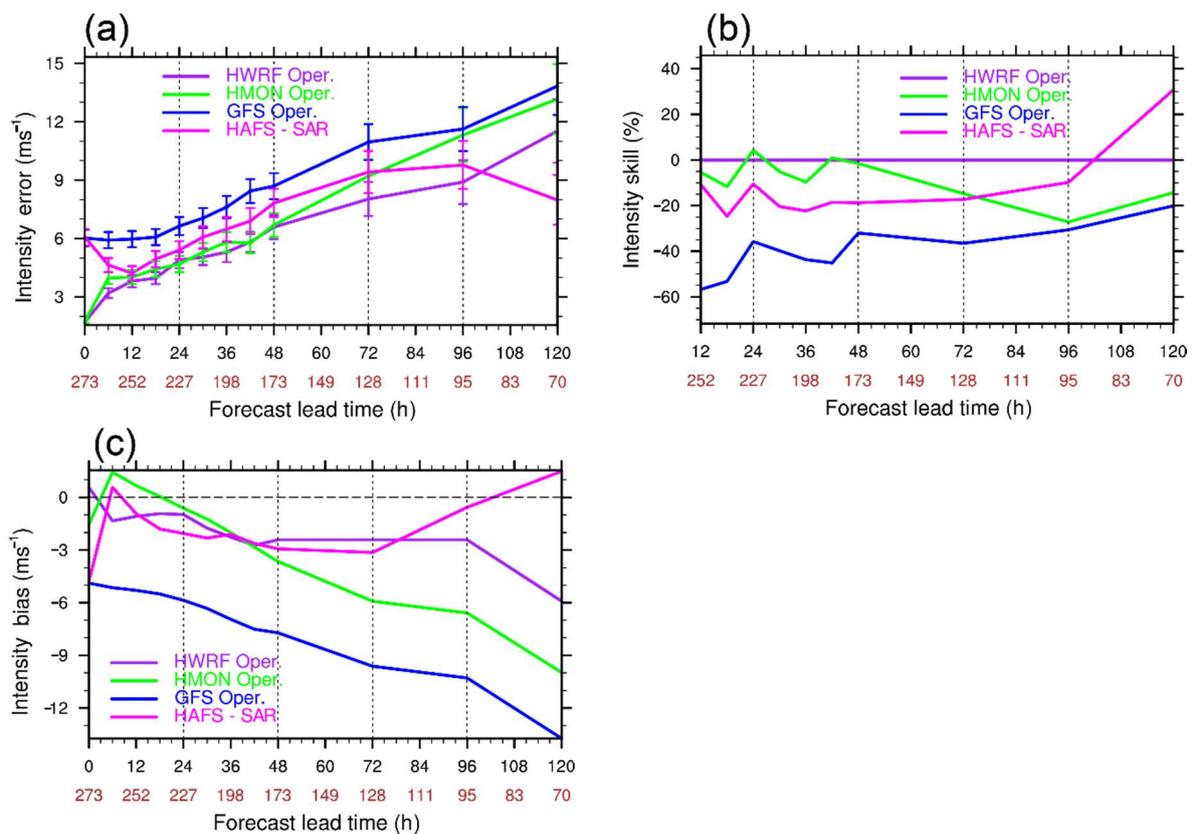


Figure 3. The 2019 seasonal intensity statistics for HAFS-SAR, GFS, HWRF and HMON: (a) intensity forecast error; (b) intensity forecast skill relative to HWRF; (c) intensity bias. The error bar in (a) represents a 95% confidence interval. Case number is denoted in brown.

The comparison of intensity error between HAFS-SAR, and other models, demonstrates advantages and limitations of the 3-km resolution grid for TC intensity forecasts. When the resolution is reduced from the ~13-km of GFS to the ~3-km of HAFS-SAR, the improvement is clear, since the GFS resolution is too coarse to resolve the inner core storm structure. However, a grid-spacing resolution of 3 km is considered marginally cloud resolving, and appears to lag behind the higher resolution of 1.5-km and 2-km in HWRF and HMON intensity forecasts, respectively. Increasing the resolution of HAFS-SAR to 1.5–2 km has the potential to further improve the intensity forecasts for strong TC events (not shown).

The inner core data assimilation, when implemented in the future, could also potentially improve the intensity forecasts.

3.3. Size Forecast

Accurate TC size forecasts can help storm surge prediction and estimation of damaging wind areas. The maximum radial extents of gale force wind (17.5 m s^{-1}), damaging force wind (25.7 m s^{-1}), and hurricane force wind (32.9 m s^{-1}) are verified against the observed radii (e.g., the best track; Figure 4) to evaluate TC size prediction. The initial gale and damaging force wind radii errors of HAFS-SAR are relatively small (Figure 4a,c), but increase in 6 h, leading to larger forecast errors than GFS, HWRF, and HMON in most lead times. The wind radii bias is also illustrated in Figure 4. A positive (negative) bias indicates that the model predicts a larger (smaller) storm at a specific wind radius. The gale force wind radius bias of HAFS-SAR is always positive (Figure 4b), indicating an over-prediction of storm size or the outer radius when other models underpredict the gale force wind radii. The difference between HAFS-SAR and other models is statistically significant from 6–96 h (error bars in Figure 4a). The damaging force wind radius bias of HAFS-SAR is comparable to other models (Figure 4d). For the hurricane force wind radius, and closer to the inner core region, HAFS-SAR performs better than the other models during the 12–120 h period, as illustrated in Figure 4e. The bias is also low at the hurricane force wind radius in HAFS-SAR forecasts, compared to HWRF and GFS (Figure 4f).

The relatively large size error for gale force wind radius from HAFS-SAR is possibly related to the horizontal advection scheme that is more diffusive than that of the GFS. The advection scheme in GFS is a fast unlimited fifth-order scheme with the built-in $2\Delta x$ filter, while HAFS-SAR chooses a Piecewise Parabolic Method (PPM) scheme with an intermediate-strength monotonicity constraint, due to the instability issue [31]. The scale-aware cumulus convection parameterization, which is turned off in the HAFS-SAR experiments, can also help to reduce the gale force wind radius error, as demonstrated by other experiments (not shown).

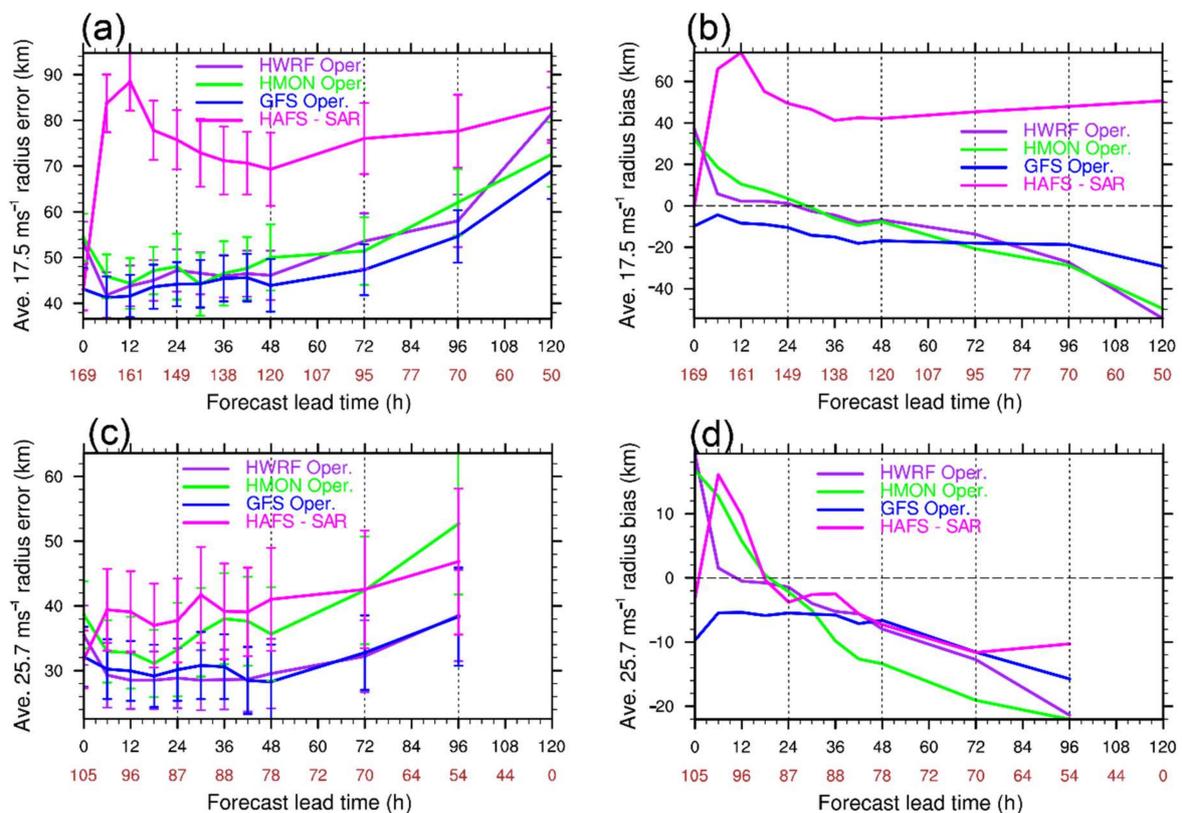


Figure 4. Cont.

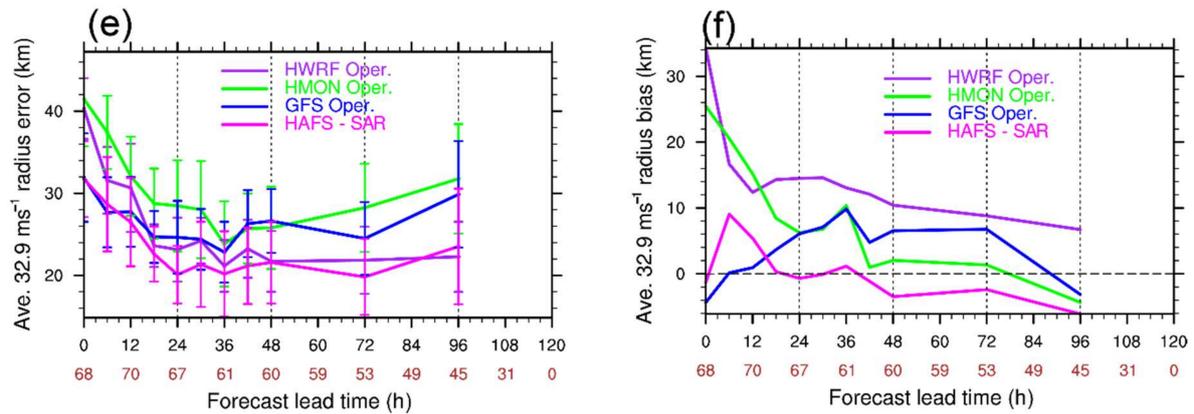


Figure 4. The 2019 seasonal storm size statistics for HAFS-SAR, GFS, HWRF and HMON: (a,c,e) gale force (17.5 m s⁻¹), damaging force (25.7 m s⁻¹) and hurricane force (32.9 m s⁻¹) wind radii error; (b,d,f) gale force (17.5 m s⁻¹), damaging force (25.7 m s⁻¹), and hurricane force (32.9 m s⁻¹) wind radii bias. The error bar in (a,c,e) represents a 95% confidence interval. Case number is denoted in brown.

3.4. Weak vs. Strong Storms

In this section, TCs are categorized as strong or weak storms, based on the initial maximum 10-m wind speed intensity, above or below 25.7-m s⁻¹ (damaging force wind). Track and intensity forecast errors are examined for these two groups in order to understand the performance of HAFS-SAR for different storm intensities (Figure 5). The track error for the strong storms is similar to or slightly smaller than that of all storms (Figure 5a). For the weak storms, the track error of HAFS-SAR is larger. HAFS-SAR has an average track error of 482 km (Figure 5c) on 120 h, which is much higher than the average track error of 259 km for the strong storms (Figure 5a). Further, the track error of HAFS-SAR is higher than GFS for weak storms, but still lower than HWRF and HMON. In general, GFS has the best performance of track forecast for the weak storms. It also should be noted that the case number of weak storms is much fewer than strong storms after 60 h.

For the intensity forecasts of strong storms, HAFS-SAR has the intensity error close to that of all storms (Figure 5b). The HAFS-SAR intensity error for weak storms is, however, much lower than that of strong storms (Figure 5d). The error at 72 h forecasts is 4.6 m s⁻¹, only approximately half of the 10.3 m s⁻¹ error for strong storms. The intensity error for HAFS-SAR is comparable to HWRF and HMON during the 12–48 h forecast lead times. From 72–96 h, the intensity error of HAFS-SAR is even lower than HWRF and HMON (Figure 5d).

As shown above, strong storms dominate both seasonal track and intensity forecast error for HAFS-SAR, with more cases at later forecast lead times. HAFS-SAR shows better forecast skill for weak storms intensity forecasts and strong storms track forecasts. The results are encouraging given that the high-resolution TC models (e.g., HWRF) tend to over-predict the intensity for weak storms [32].

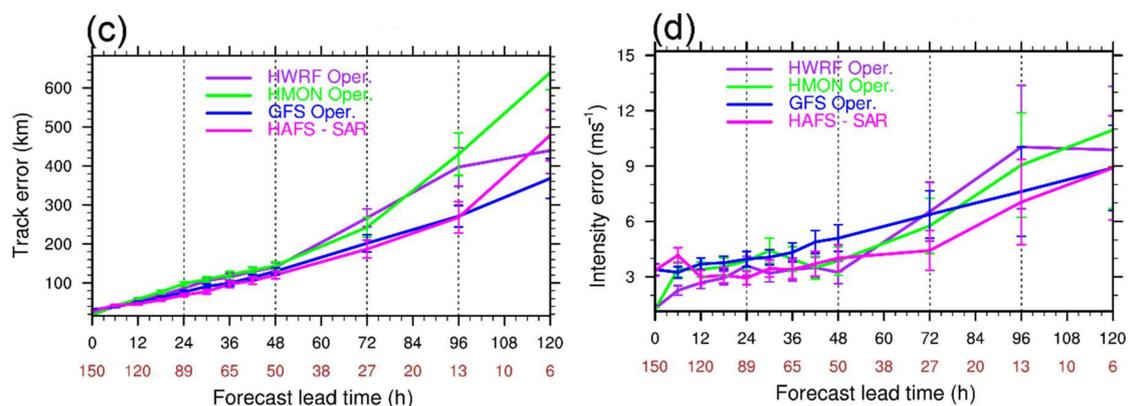


Figure 5. Cont.

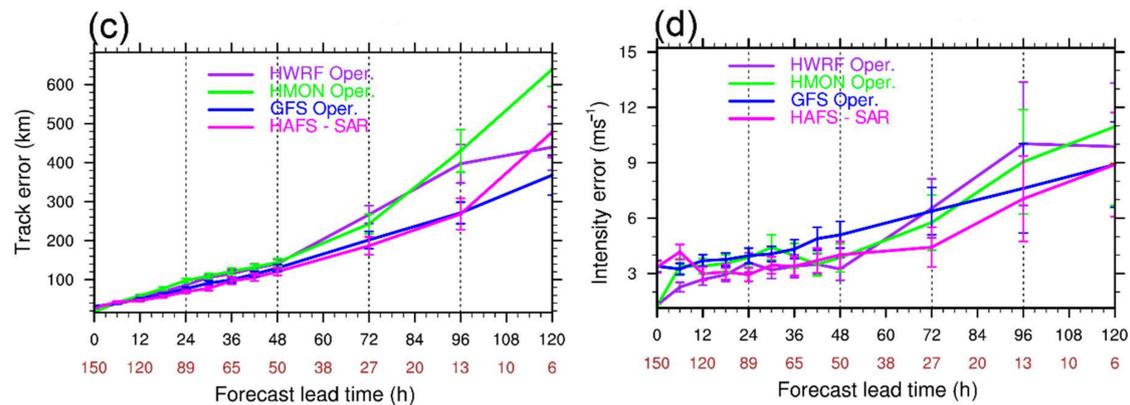


Figure 5. Track error for the 2019 hurricane season for strong storms (a) and weak storms (c); intensity error for the 2019 hurricane season for strong storms (b) and weak storms (d). See Section 3.4 for the definitions of strong and weak storms. The error bar in (b,d) represents a 95% confidence interval. Case number is denoted in brown.

3.5. TC Barry and TC Dorian

During the 2019 NATL TC season, two hurricanes made landfall in the United States and caused considerable damage and economic loss. The track and intensity forecasts of HAFS-SAR for TC Barry and Dorian are verified against the best track, along with the comparison to GFS, HWRF, and HMON, in order to evaluate the performance of these two high-impact events.

3.5.1. TC Barry

TC Barry began as a mesoscale convective vortex over the central US, and became a tropical depression in the northern Gulf of Mexico [33]. After intensifying into a tropical storm, Barry reached category 1 hurricane status on 13 July, and made landfall at Marsh Island, Louisiana, with a very asymmetric structure. The track and intensity forecast errors for TC Barry are illustrated in Figure 6. The HAFS-SAR track forecast shows clear improvements over GFS, HWRF, and HMON within the 24–96 h forecast lead times (Figure 6a). The HAFS-SAR track forecast error is always below 74 km, while the GFS error is above 222 km after 72 h. Both HWRF and HMON have track errors over 370 km at 96 h. The composite Barry track forecasts of all cases for the four models are shown in Figure 7. GFS, HWRF, and HMON all have a rightward bias in the track forecasts (Figure 7a–c). HAFS-SAR follows the observed track closely, and predicts the Louisiana landfall location in close proximity to the observed location in multiple forecast cases (Figure 7d). HAFS-SAR also performs well for Barry in intensity forecasts, as shown in Figure 6b. The maximum wind error of HAFS-SAR is below those of other models presented here at most forecast lead times. We notice that the intensity errors of HWRF and HMON are much higher than both GFS and HAFS-SAR for Barry, which is likely due to their erroneous track forecasts particularly at longer lead times (Figure 7b,c).

Affected by both the northerly shear and dry air intrusion at mid-levels, TC Barry maintained a very asymmetric structure at landfall (Figure 6c). The rainband structure at 1200 UTC 13 July is mostly along the east and the south side of the TC center. HAFS-SAR is able to capture the convective asymmetry at the 60 h forecast lead time for the forecast initialized at 0000 UTC 11 July, with the rainband extending from the borders of Mississippi and Alabama into the Gulf of Mexico. However, the convection to the south of TC center is over-predicted in HAFS-SAR (Figure 6d). The forecasted TC center (29.1° N, 92.3° W) was at the southwest of the observation (29.3° N, 91.9° W), and the distance between the model and observed locations is 44 km at the valid time.

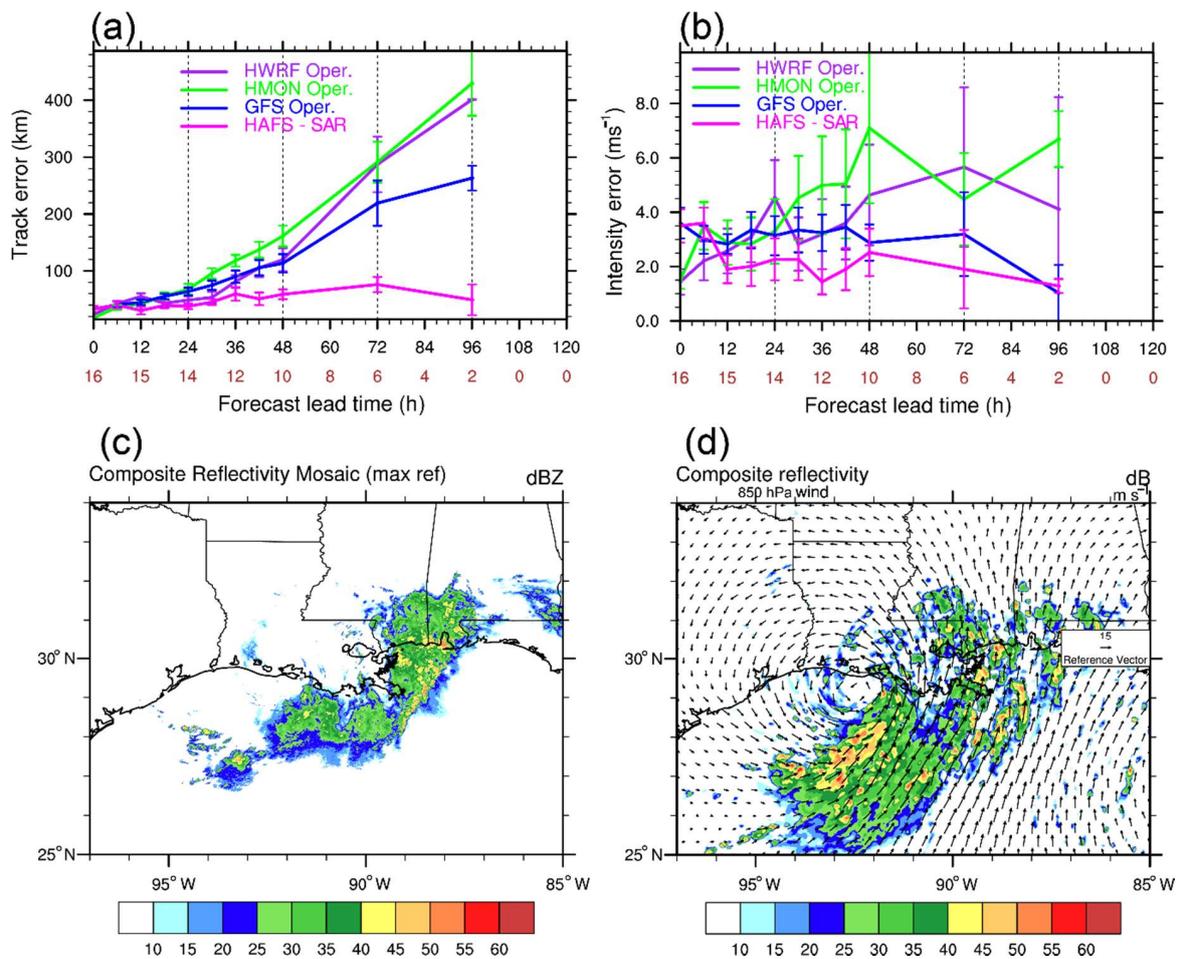


Figure 6. Track (a) and intensity (b) errors for TC Barry; (c) observed composite reflectivity of TC Barry at 1200 UTC 13 July; (d) composite reflectivity of HAFS-SAR forecasts initialized at 11 July 0000 UTC and valid at 13 July 1200 UTC, overlapped with wind vectors at 850 hPa. The error bar in (a,b) represents a 95% confidence interval. Case number is denoted in brown.

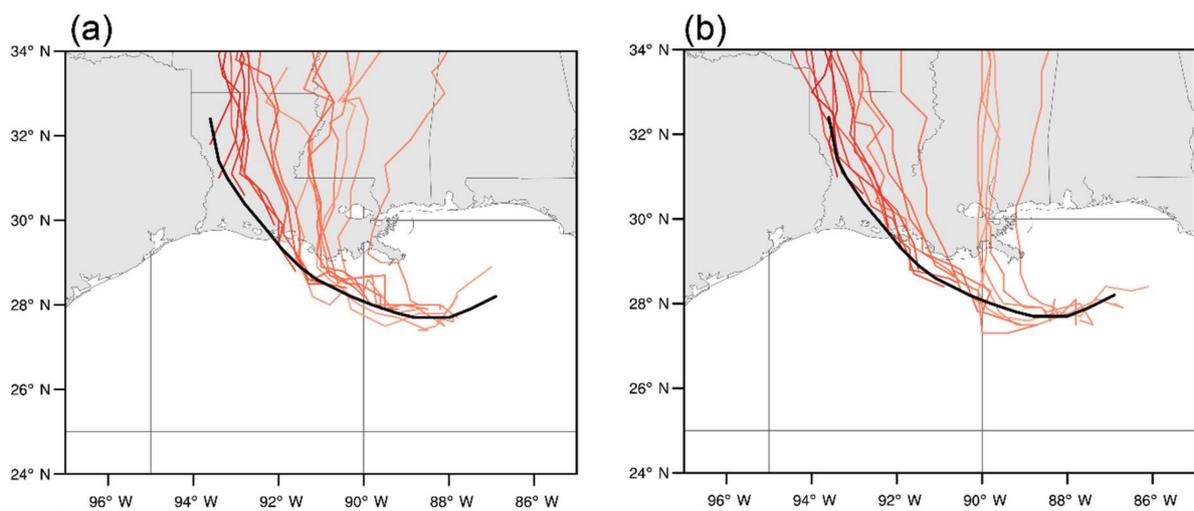


Figure 7. Cont.

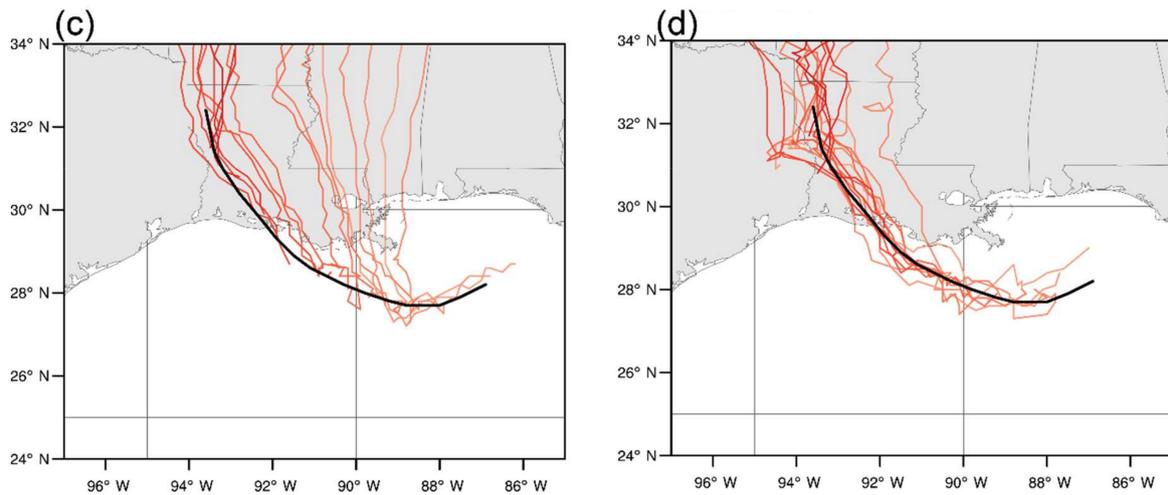


Figure 7. The composite track forecasts of (a) GFS, (b) HWRF, (c) HMON and (d) HAFS-SAR for TC Barry from 1200 UTC 10 July to 1200 UTC 14 July 2019. The observed best track is in black. The darker red colors represent later initialized times.

3.5.2. TC Dorian

TC Dorian originated from a tropical wave off of the West Coast of Africa on 19 August, and intensified rapidly as a powerful hurricane as it moved near to the Bahamas [34]. Dorian passed east of Puerto Rico and made Bahamian landfall as a very strong category 5 hurricane. After stalling over the Bahamas for approximately 24 h, Dorian recurved to the northwest and moved along the Florida coast, before it made a United States landfall at Cape Hatteras, North Carolina. Dorian finally underwent an extratropical transition during 6–7 September.

The HAFS-SAR track forecast outperforms GFS, HWRF, and HMON, as illustrated in Figure 8a. The HAFS-SAR forecast error is the lowest among all models during the 24–120 h period. The superiority of HAFS-SAR is mostly achieved from the cross-track component at shorter lead times (before 96 h), and from along-track component at longer lead times (not shown). GFS, HWRF, and HMON incorrectly predicted the Florida landfall of Dorian in multiple cases (Figure 9a–c). HAFS-SAR appears to predict the recurvature along the Florida coast more accurately, with none of its cases making landfall in Florida (Figure 9d). The intensity forecast error of HAFS-SAR for Dorian is consistent with the total seasonal statistics for strong storms statistics (Figure 8b), when compared with other models.

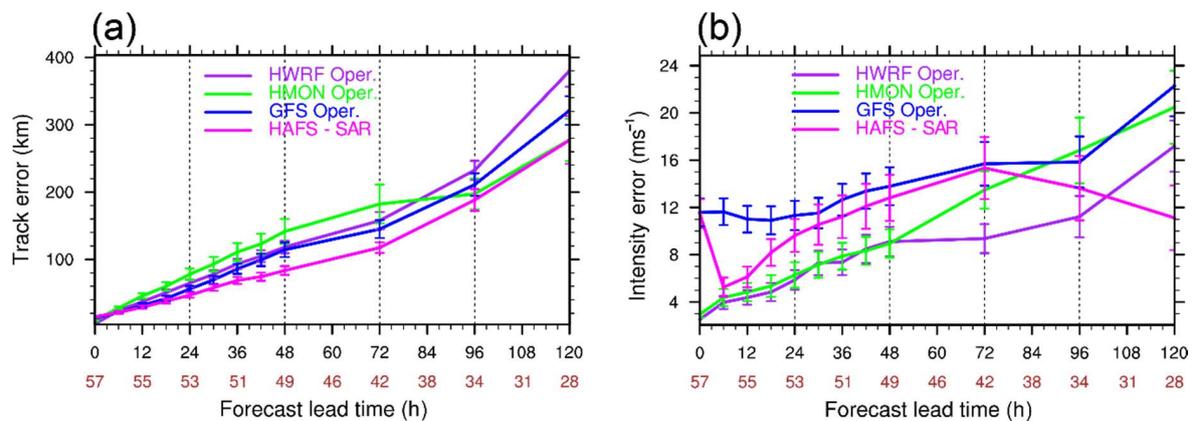


Figure 8. Track (a) and intensity (b) errors for TC Dorian. The error bar represents a 95% confidence interval. Case number is denoted in brown.

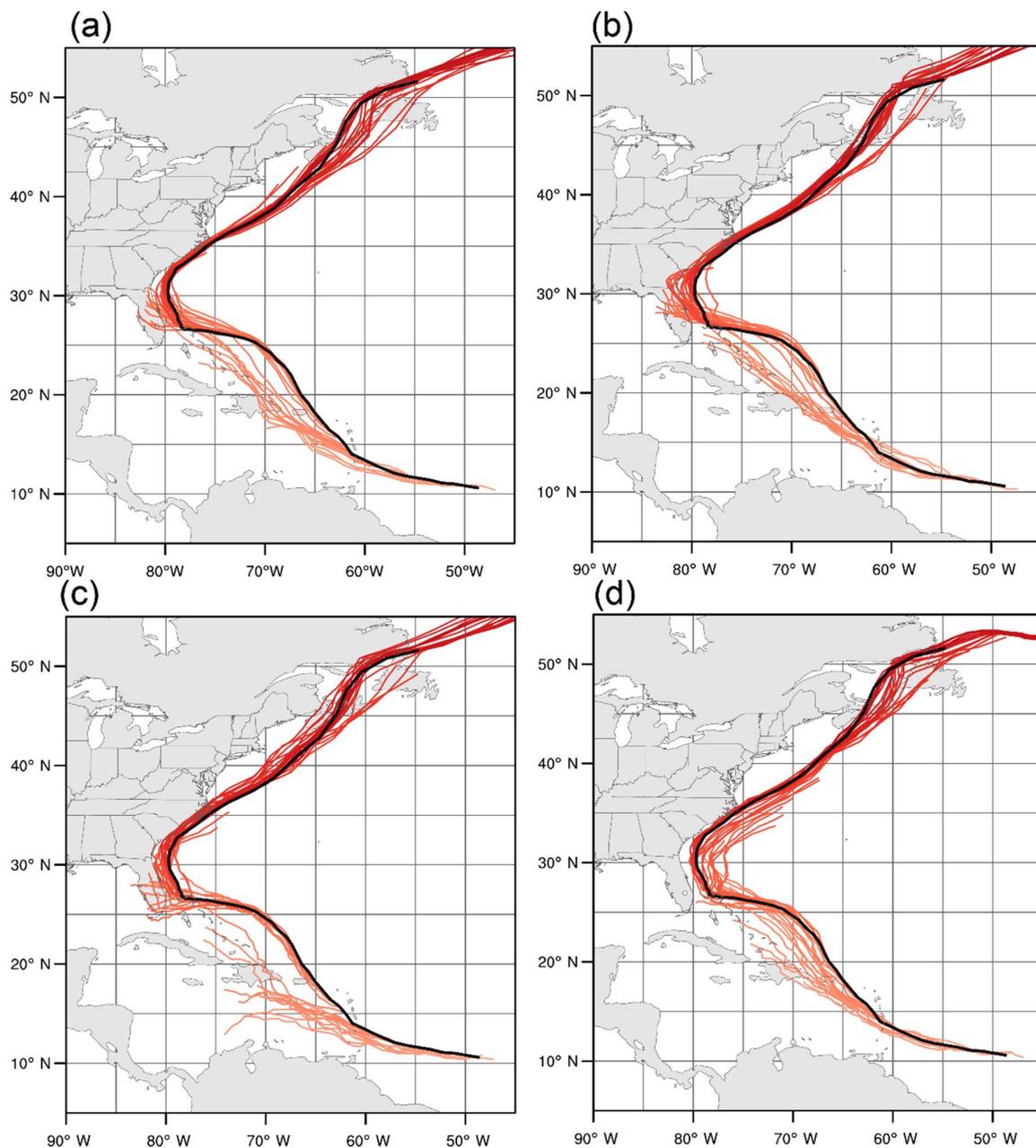


Figure 9. The composite track forecasts of (a) GFS, (b) HWRF, (c) HMON and (d) HAFS-SAR for TC Dorian from 1200 UTC 24 August to 1800 UTC 7 September 2019. The observed best track is in black. The darker red colors represent later initialized times.

4. Conclusions

With the ongoing advancements for the HAFS, the convection-allowing FV3-based HAFS-SAR was successfully implemented during the 2019 real-time NATL TC season. It is of great interest to examine the performance of the new regional hurricane forecast system with the same FV3 dynamic core in NOAA's unified modelling framework. The intensity, track, and storm size forecasts were systematically evaluated for 273 cases, and compared to both the observed (e.g., best) track data and the GFS, HWRF, and HMON.

HAFS-SAR demonstrate noticeable improvements of track forecasts compared to GFS, HWRF, and HMON at almost all forecast lead times. The track forecast skill of HAFS-SAR improves roughly 20% over HWRF. HAFS-SAR is more effective at reducing the cross-track error than the along track

component. The cross-track bias of HAFS-SAR is close to zero prior to 72 h, and has a rightward bias for longer forecast lead times.

The initial intensity error (6.2 m s^{-1}) of HAFS-SAR is similar to GFS, due to the lack of inner core data assimilation and/or vortex adjustment as done in HWRF and HMON. The spin-up during the first 6 h reduces the HAFS-SAR intensity error to 4.1 m s^{-1} , and the improvement over GFS is consistent at longer forecast lead times. When compared to both HWRF and HMON, the HAFS-SAR intensity forecasts demonstrate less skill, likely related to the superior horizontal resolution of the operational HWRF and HMON. The intensity forecasts of HAFS-SAR are improved in additional experiments when the horizontal resolutions are increased from 3 km to 2 km and 1.5 km for individual storms. The sensitivity of intensity forecasts to vortex initialization, model physics, dynamics, and grid specification are also being further investigated.

The verification of gale force, damaging force, and hurricane force wind radii reveals that HAFS-SAR performs better than the other three models for the hurricane force wind (32.9-m s^{-1}) radius. HAFS-SAR over-predicts the size of the TC (e.g., the 17.5 m s^{-1} gale force wind radius), compared to GFS, HWRF, and HMON. The storm size of HAFS-SAR is sensitive to the horizontal advection schemes of FV3, and the use of a less diffusive scheme is being investigated to improve the HAFS storm size forecasts.

The seasonal statistics for HAFS-SAR track and intensity forecasts were also examined for strong and weak storms, based on the initial intensity. The track forecast error is lower for strong storms, while weak storms have better intensity forecast performance. HAFS-SAR also demonstrates promising results for the track forecasts during two high-impact 2019 TC events: TC Barry and Dorian.

The successful implementation of HAFS-SAR during 2019 HFIP real-time demonstration provides an important step toward the efforts for building up the next generation high-resolution TC forecasting system. The systematic evaluation of HAFS-SAR demonstrates great potential for the FV3-based convection-allowing regional models toward improving TC forecasts. A new data assimilation system is under development for HAFS, to further improve the initial vortex structure by assimilating TC inner core observations, which are expected to reduce the intensity forecasts errors. The future development of ocean coupling for HAFS will also enhance the TC forecasting capability, by introducing realistic ocean response to TCs.

Author Contributions: A.M., V.T., S.G. and F.M. conceived HAFS experiments for the HFIP real-time demonstration. J.D., Z.Z. and B.L. designed the real-time HAFS-SAR experiments. J.D. conducted the real-time experiments, analyzed the results, plotted the figures and wrote the original manuscript. B.L. provided workflow software support. W.W. and C.Z. provided physics upgrade for HAFS-SAR. K.W. provide online visualization for HAFS-SAR in real-time. J.D., Z.Z. and L.Z. provided support for HAFS-SAR initialization. F.M., H.R.W., Z.Z., W.W., C.Z., A.T.H., S.G. and X.Z. reviewed and edited the manuscript. All authors interpreted and discussed the results. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by HFIP/Next Generation Global Prediction System (NGGPS) programs at NOAA/NWS and the Disaster Relief Act (2018).

Acknowledgments: The authors thank Zaizhong Ma for his helpful comments to this manuscript. We also want to thank four anonymous reviewers for their insightful comments and suggestions. The authors are grateful to help from NOAA/EMC and NOAA/GFDL on setting up FV3 and the stand-alone regional model. Thanks to Dan Iredell on data transfer for the real-time demonstration. The authors acknowledge the NOAA Research and Development High Performance Computing Program for providing computing and storage resources that have contributed to the research results reported within this paper.

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

References

1. Chen, J.-H.; Lin, S.-J.; Zhou, L.; Chen, X.; Rees, S.; Bender, M.; Morin, M. Evaluation of Tropical Cyclone Forecasts in the Next Generation Global Prediction System. *Mon. Weather Rev.* **2019**, *147*, 3409–3428. [[CrossRef](#)]
2. Davis, C.; Wang, W.; Chen, S.S.; Chen, Y.; Corbosiero, K.; DeMaria, M.; Dudhia, J.; Holland, G.; Klemp, J.; Michalakes, J.; et al. Prediction of Landfalling Hurricanes with the Advanced Hurricane WRF Model. *Mon. Weather Rev.* **2008**, *136*, 1990–2005. [[CrossRef](#)]
3. Davis, C.; Wang, W.; Dudhia, J.; Torn, R. Does Increased Horizontal Resolution Improve Hurricane Wind Forecasts? *Weather Forecast.* **2010**, *25*, 1826–1841. [[CrossRef](#)]
4. Xue, M.; Schleif, J.; Kong, F.; Thomas, K.W.; Wang, Y.; Zhu, K. Track and Intensity Forecasting of Hurricanes: Impact of Convection-Permitting Resolution and Global Ensemble Kalman Filter Analysis on 2010 Atlantic Season Forecasts. *Weather Forecast.* **2013**, *28*, 1366–1384. [[CrossRef](#)]
5. Yau, M.K.; Liu, Y.; Zhang, D.-L.; Chen, Y. A Multiscale Numerical Study of Hurricane Andrew (1992). Part VI: Small-Scale Inner-Core Structures and Wind Streaks. *Mon. Weather Rev.* **2004**, *132*, 1410–1433. [[CrossRef](#)]
6. Judt, F.; Chen, S.S. Convectively Generated Potential Vorticity in Rainbands and Formation of the Secondary Eyewall in Hurricane Rita of 2005. *J. Atmos. Sci.* **2010**, *67*, 3581–3599. [[CrossRef](#)]
7. Qiu, X.; Tan, Z.-M.; Xiao, Q. The Roles of Vortex Rossby Waves in Hurricane Secondary Eyewall Formation. *Mon. Weather Rev.* **2010**, *138*, 2092–2109. [[CrossRef](#)]
8. Chen, H.; Zhang, D.-L.; Carton, J.; Atlas, R. On the Rapid Intensification of Hurricane Wilma (2005). Part I: Model Prediction and Structural Changes. *Weather Forecast.* **2011**, *26*, 885–901. [[CrossRef](#)]
9. Menelaou, K.; Yau, M.K.; Martinez, Y. On the Origin and Impact of a Polygonal Eyewall in the Rapid Intensification of Hurricane Wilma (2005). *J. Atmos. Sci.* **2013**, *70*, 3839–3858. [[CrossRef](#)]
10. Qin, N.; Zhang, D.-L. On the Extraordinary Intensification of Hurricane Patricia (2015). Part I: Numerical Experiments. *Weather Forecast.* **2018**, *33*, 1205–1224. [[CrossRef](#)]
11. Xiao, J.; Tan, Z.-M.; Chow, K.-C. Structure and formation of convection of secondary rainbands in a simulated typhoon Jangmi (2008). *Meteorol. Atmos. Phys.* **2019**, *131*, 713–737. [[CrossRef](#)]
12. Mehra, A.; Tallapragada, V.; Zhang, Z.; Liu, B.; Zhu, L.; Wang, W.; Kim, H.-S. Advancing the State of the Art in Operational Tropical Cyclone Forecasting at Ncep. *Trop. Cyclone Res. Rev.* **2018**, *7*, 51–56. [[CrossRef](#)]
13. Alaka, G.J.; Zhang, X.; Gopalakrishnan, S.G.; Goldenberg, S.B.; Marks, F.D. Performance of Basin-Scale HWRF Tropical Cyclone Track Forecasts. *Weather Forecast.* **2017**, *32*, 1253–1271. [[CrossRef](#)]
14. Alaka, G.J.; Zhang, X.; Gopalakrishnan, S.G.; Zhang, Z.; Marks, F.D.; Atlas, R. Track Uncertainty in High-Resolution HWRF Ensemble Forecasts of Hurricane Joaquin. *Weather Forecast.* **2019**, *34*, 1889–1908. [[CrossRef](#)]
15. Hazelton, A.T.; Bender, M.; Morin, M.; Harris, L.; Lin, S.-J. 2017 Atlantic Hurricane Forecasts from a High-Resolution Version of the GFDL fvGFS Model: Evaluation of Track, Intensity, and Structure. *Weather Forecast.* **2018**, *33*, 1317–1337. [[CrossRef](#)]
16. Hazelton, A.T.; Zhang, X.; Gopalakrishnan, S.; Ramstrom, W.; Marks, F.; Zhang, J.A. High-Resolution Ensemble HFV3 Forecasts of Hurricane Michael (2018): Rapid Intensification in Shear. *Mon. Weather Rev.* **2020**, *148*, 2009–2032. [[CrossRef](#)]
17. Lin, S.-J. A “Vertically Lagrangian” Finite-Volume Dynamical Core for Global Models. *Mon. Weather Rev.* **2004**, *132*, 2293–2307. [[CrossRef](#)]
18. Putman, W.M.; Lin, S.-J. Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.* **2007**, *227*, 55–78. [[CrossRef](#)]
19. Chen, X.; Andronova, N.; Leer, B.V.; Penner, J.E.; Boyd, J.P.; Jablonowski, C.; Lin, S.-J. A Control-Volume Model of the Compressible Euler Equations with a Vertical Lagrangian Coordinate. *Mon. Weather Rev.* **2013**, *141*, 2526–2544. [[CrossRef](#)]
20. Han, J.; Witek, M.L.; Teixeira, J.; Sun, R.; Pan, H.-L.; 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](#)]

21. Wang, W.; Sippel, J.A.; Abarca, S.; Zhu, L.; Liu, B.; Zhang, Z.; Mehra, A.; Tallapragada, V. Improving NCEP HWRF Simulations of Surface Wind and Inflow Angle in the Eyewall Area. *Weather Forecast.* **2018**, *33*, 887–898. [[CrossRef](#)]
22. Biswas, M.K.; Abarca, S.; Bernardet, L.; Ginis, I.; Grell, E.; Iacono, M.; Kalina, E.; Liu, B.; Liu, Q.; Marchok, T.; et al. Hurricane Weather Research and Forecasting (HWRF) Model: 2018 Scientific Documentation. Available online: <https://dtcenter.org/community-code/hurricane-wrf-hwrf/documentation> (accessed on 9 June 2020).
23. Chen, J.-H.; Lin, S.-J. Seasonal Predictions of Tropical Cyclones Using a 25-km-Resolution General Circulation Model. *J. Clim.* **2013**, *26*, 380–398. [[CrossRef](#)]
24. Ek, M.B.; Mitchell, K.E.; Lin, Y.; Rogers, E.; Grunmann, P.; Koren, V.; Gayno, G.; Tarpley, J.D. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res. Atmos.* **2003**, *108*. [[CrossRef](#)]
25. Iacono, M.J.; Mlawer, E.J.; Clough, S.A.; Morcrette, J.-J. Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. *J. Geophys. Res. Atmos.* **2000**, *105*, 14873–14890. [[CrossRef](#)]
26. Clough, S.A.; Shephard, M.W.; Mlawer, E.J.; Delamere, J.S.; Iacono, M.J.; 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](#)]
27. Marchok, T.P. How the NCEP tropical cyclone tracker works. In Proceedings of the 25th Conference on Hurricanes and Tropical Meteorology, San Diego, CA, USA, 28 April–3 May 2002; pp. 21–22.
28. Sampson, C.R.; Schrader, A.J. The Automated Tropical Cyclone Forecasting System (Version 3.2). *Bull. Am. Meteorol. Soc.* **2000**, *81*, 1231–1240. [[CrossRef](#)]
29. Tong, M.; Sippel, J.A.; Tallapragada, V.; Liu, E.; Kieu, C.; Kwon, I.-H.; Wang, W.; Liu, Q.; Ling, Y.; Zhang, B. Impact of Assimilating Aircraft Reconnaissance Observations on Tropical Cyclone Initialization and Prediction Using Operational HWRF and GSI Ensemble–Variational Hybrid Data Assimilation. *Mon. Weather Rev.* **2018**, *146*, 4155–4177. [[CrossRef](#)]
30. Kim, H.-S.; Lozano, C.; Tallapragada, V.; Iredell, D.; Sheinin, D.; Tolman, H.L.; Gerald, V.M.; Sims, J. Performance of Ocean Simulations in the Coupled HWRF–HYCOM Model. *J. Atmos. Ocean. Technol.* **2014**, *31*, 545–559. [[CrossRef](#)]
31. Lin, S.-J.; Putman, W.; Harris, L. The GFDL Finite-Volume Cubed-Sphere Dynamical Core Technical Description. Available online: <https://www.gfdl.noaa.gov/fv3/fv3-documentation-and-references/> (accessed on 9 June 2020).
32. Gopalakrishnan, S.; Toepfer, F.; Forsythe-Newell, S.; Gall, R.; Marks, F.; Rappaport, E.N.; Tallapragada, V.; Franklin, J.; Alaka, G.; Abarca, S.; et al. 2016 HFIP R&D Activities Summary: Recent Results and Operational Implementation. Available online: <http://www.hfip.org/documents/> (accessed on 9 June 2020).
33. Cangialosi, J.P.; Hagen, A.B.; Berg, R. National Hurricane Center Tropical Cyclone Report: Hurricane Barry. Available online: <https://www.nhc.noaa.gov/data/tcr/> (accessed on 9 June 2020).
34. Avila, L.A.; Stewart, S.R.; Berg, R.; Hagen, A.B. National Hurricane Center Tropical Cyclone Report: Hurricane Dorian. Available online: <https://www.nhc.noaa.gov/data/tcr/> (accessed on 9 June 2020).

