Performance of 2020 Real-Time Atlantic Hurricane Forecasts from High-Resolution Global-Nested Hurricane Models: HAFS-globalnest and GFDL T-SHiELD

ANDREW HAZELTON,[a,b] KUN GAO,[c,d] MORRIS BENDER,[c,d] LEVI COWAN,[e] GHASSAN J. ALAKA JR.,[b] ALEX KALTENBAUGH,[f] LEW GRAMER,[a,b] XUEJIN ZHANG,[b] LUCAS HARRIS,[c] TIMOTHY MARCHOK,[c] MATT MORIN,[f] AVICHAL MEHRA,[g] ZHAN ZHANG,[g] BIN LIU,[g,h] AND FRANK MARKS[b]

[a] *University of Miami, Cooperative Institute for Marine and Atmospheric Studies, Miami, Florida*
[b] *NOAA/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida*
[c] *NOAA/GFDL, Princeton, New Jersey*
[d] *Princeton University, Princeton, New Jersey*
[e] *Joint Typhoon Warning Center, Honolulu, Hawaii*
[f] *University Corporation for Atmospheric Research, Boulder, Colorado*
[g] *NOAA/EMC, College Park, Maryland*
[h] *I.M. Systems Group, College Park, Maryland*

(Manuscript received 16 June 2021, in final form 8 November 2021)

ABSTRACT: The global-nested Hurricane Analysis and Forecast System (HAFS-globalnest) is one piece of NOAA's Unified Forecast System (UFS) application for hurricanes. In this study, results are analyzed from 2020 real-time forecasts by HAFS-globalnest and a similar global-nested model, the Tropical Atlantic version of GFDL's System for High-resolution prediction on Earth-to-Local Domains (T-SHiELD). HAFS-globalnest produced the highest track forecast skill compared to several operational and experimental models, while T-SHiELD showed promising track skills as well. The intensity forecasts from HAFS-globalnest generally had a positive bias at longer lead times primarily due to the lack of ocean coupling, while T-SHiELD had a much smaller intensity bias particularly at longer forecast lead times. With the introduction of a modified planetary boundary layer scheme and an increased number of vertical levels, particularly in the boundary layer, HAFS forecasts of storm size had a smaller positive bias than occurred in the 2019 version of HAFS-globalnest. Despite track forecasts that were comparable to the operational GFS and HWRF, both HAFS-globalnest and T-SHiELD suffered from a persistent right-of-track bias in several cases at the 4–5-day forecast lead times. The reasons for this bias were related to the strength of the subtropical ridge over the western North Atlantic and are continuing to be investigated and diagnosed. A few key case studies from this very active hurricane season, including Hurricanes Laura and Delta, were examined.

KEYWORDS: Tropical cyclones; Hurricanes/typhoons; Forecast verification/skill; Numerical analysis/modeling

1. Introduction

Tropical cyclone (TC) forecasts have improved over the last decade, due to advancements in numerical modeling, observations, data assimilation, and forecasting techniques. Track forecast errors have steadily decreased for several decades now (Landsea and Cangialosi 2018), and improvements in forecasts of TC intensity have also been realized during the last decade (Cangialosi et al. 2020). However, biases in TC track forecasts persist, and intensity forecasts remain challenging, especially in instances of rapid intensification (e.g., Kaplan et al. 2015). Furthermore, track forecast errors make it difficult to interpret intensity errors, since it is hard to determine if those errors are caused by the TC being in the wrong location or by issues with the TC intensification processes simulated by the model (Tien et al. 2013). Improvement of predictions of TC rapid intensity change (both rapid intensification and rapid weakening) is one of the key goals of the Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013; Gopalakrishnan et al. 2020). In addition, improving prediction of TC structure, including radii for certain wind thresholds (e.g., Cangialosi and Landsea 2016), is important for prediction of TC evolution and impacts, including storm surge.

The Hurricane Analysis and Forecast System (HAFS) is a next-generation TC modeling system being developed as a collaborative effort within NOAA and partner institutions, with the goal of improving TC forecasts in all of the areas mentioned above. HAFS is based on the two-way-nested version of the finite-volume cubed sphere (FV3) dynamical core (e.g., Lin and Rood 1996; Lin 2004; Harris and Lin 2013). Recent work has shown the ability of two-way nested FV3-based models to produce skillful track, intensity, and structure forecasts for TCs. Hazelton et al. (2018b) analyzed the performance of a real-time global-nested FV3 modeling system called hfvGFS in the 2017 Atlantic hurricane season. The hfvGFS was configured with a static nest that had a grid spacing of 3 km and was positioned over the North Atlantic basin. The hfvGFS was also able to skillfully predict TC structure, based on comparisons with NOAA P-3 airborne radar data (Hazelton et al. 2018a). In addition, nested FV3-based modeling systems for short-term to subseasonal prediction (e.g., Gao et al. 2019) are the focus of ongoing developments at

Corresponding author: Andrew Hazelton, Andrew.Hazelton@noaa.gov

DOI: 10.1175/WAF-D-21-0102.1

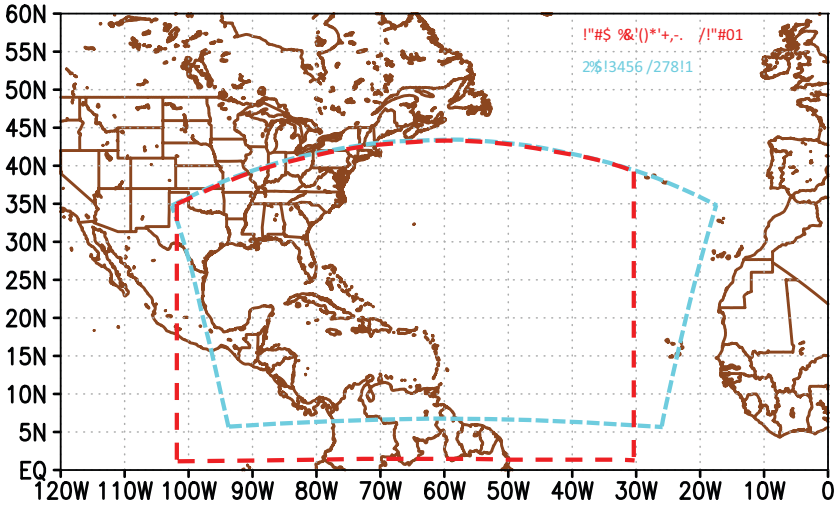

FIG. 1. Atlantic grid layouts for HAFS-globalnest (red) and T-SHiELD (light blue).

NOAA's Geophysical Fluid Dynamics Laboratory (GFDL), including the System for High-resolution prediction on Earth-to-Local Domains (SHiELD; Harris et al. 2020). In particular, GFDL developed the Tropical Atlantic version of GFDL's System for High-resolution prediction on Earth-to-Local Domains (T-SHiELD). T-SHiELD and hfvGFS have nearly identical configurations, although the physical parameterizations used in each model have important differences (as discussed below).

Building from these prototypes, forecasts from two versions of HAFS were produced during the 2019 North Atlantic hurricane season as a part of HFIP Real-time Experiments: a stand-alone regional version of HAFS (HAFS-SAR; Dong et al. 2020) and a global-nested version of HAFS (HAFS-globalnest; Hazelton et al. 2021). Both versions showed promising track forecast skill and, also, the ability to successfully predict rapid intensification in some cases. HAFS-globalnest showed better track forecasts than HAFS-SAR near the boundary of the nest, probably due to the two-way feedback with the concurrently running global model. These 2019 experiments formed the initial baseline for further development of HAFS.

The record-breaking, active 2020 North Atlantic hurricane season provided a unique opportunity to evaluate ongoing upgrades to HAFS-globalnest and the similarly configured GFDL T-SHiELD in a wide variety of forecasts produced on NOAA supercomputers. The model configurations for 2020, including changes from the 2019 version, are discussed in the next section, along with a description of the 2020 forecast cases. The large set of storms during the 2020 season provided a unique opportunity to evaluate and compare these two models, to understand how the different configurations affect the performance in forecasting TC track, structure, and intensity, as well as the large-scale environment. A few selected case studies then highlight details of the model's performance, including some of the track biases noted. Finally, conclusions and plans for future development are discussed.

TABLE 1. Configurations of HAFS-globalnest, T-SHiELD, and the other operational and experimental models they are compared with.

Model (plot abbreviation)	Domain	Boundary conditions	Finest grid spacing	PBL physics	Microphysics	Dynamic ocean (Y/N)
HAFS-Globalnest (HAFB)	Global with Atlantic static nest	Global HAFS	3 km	Modified EDMF-TKE	GFDL	N
GFDL T-SHiELD (T20H)	Global with Atlantic static nest	Global T-SHiELD	3 km	YSU	GFDL	Y (1D)
GFS (GFSO)	Global	N/A	13 km	EDMF	GFDL	N
HAFS-SAR (HAFA)	Atlantic static nest	GFS	3 km	Modified EDMF	GFDL	Y (3D)
HWRF (HWRF)	Storm-following nest	GFS	1.5 km	Modified EDMF	Ferrier–Aligo	Y (3D)
HMON (HMON)	Storm-following nest	GFS	2 km	Modified EDMF	Ferrier–Aligo	Y (3D)

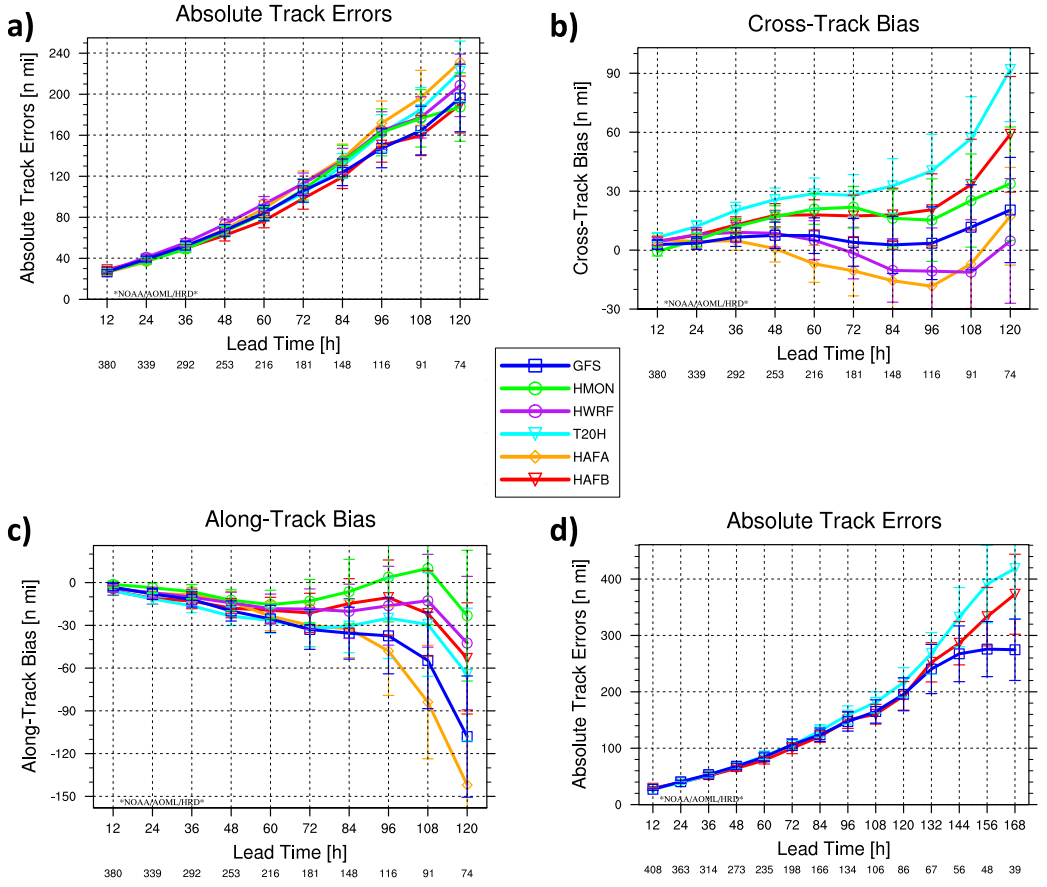

FIG. 2. (a) 5-day mean track errors (n mi; 1 n mi = 1.852 km) from HAFS-globalnest (red), T-SHiELD (light blue), HAFS-SAR (orange), GFS (dark blue), HWRF (purple), and HMON (green). (b) As in (a), but for the cross-track track bias (n mi). (c) As in (a), but for the along-track track bias (n mi). (d) As in (a), but out to 7 days for the models that forecast out to 7 days (HAFS-globalnest, T-SHiELD, and GFS). The number of homogenous cases at each forecast hour is shown at the bottom. The bars show the 95% confidence intervals.

2. Model configuration and 2020 changes—Data used

a. 2020 HAFS-globalnest and T-SHiELD domains

The 2020 HAFS-globalnest domain was slightly smaller than that used in the 2019 North Atlantic hurricane season (Fig. 1), with a static nest covering all of the western and central North Atlantic and extending to about 30°W (near the Cabo Verde Islands). The smaller nest this year was made necessary by an increase in the number of vertical levels from 64 to 75, with most of the new levels focused in the planetary boundary layer (PBL). This kind of increase in vertical resolution in lower levels has been shown to produce stronger TCs in other models, such as the Hurricane Weather Research and Forecasting (HWRF) Model (Zhang et al. 2015). The horizontal grid spacings on the global and nested domains were 13 and 3 km, respectively, as in the previous version of HAFS-globalnest. Thus, the smaller nest allowed for the computational efficiency needed for the quasi-real-time experiments run during the 2020 Atlantic hurricane season. TCs outside this nested domain were tracked on the global domain (but not the nested domain).

T-SHiELD used a similar nested domain that was shifted slightly east compared to that of HAFS-globalnest. Both models covered the vast majority of the tropical Atlantic basin, and captured almost all of the TCs during the 2020 Atlantic hurricane season, when most of the activity occurred in the Caribbean, Gulf of Mexico, and western and central Atlantic. The nests for both T-SHiELD and HAFS-globalnest exchanged data with their global domains in two-way feedback.

b. Model physics and 2020 upgrades

The 2020 HAFS-globalnest model physics were generally similar to the 2019 version of HAFS-globalnest (Hazelton et al. 2021), including use of the 6-class GFDL microphysics (Chen and Lin 2013; Zhou et al. 2019) and a scale-aware convective scheme for the global domain only (Han et al. 2017). One key difference was the introduction of the eddy diffusivity mass flux with prognostic turbulent kinetic energy (EDMF-TKE) PBL scheme (Han and Bretherton 2019). This scheme will be included in the operational Global Forecast System (GFS) version 16 in 2021. Some

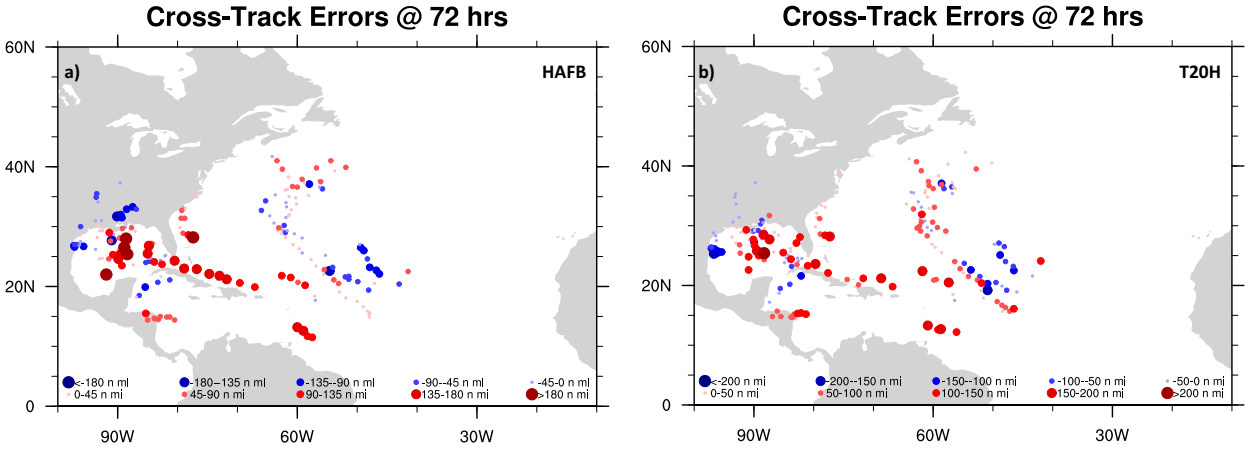

FIG. 3. (a) Spatial map of cross-track bias (positive is a right-of-observed bias, negative is a left-of-observed bias; n mi) for HAFS-globalnest for 72-h forecasts. The errors are shown at the forecast position. (b) As in (a), but for T-SHiELD. Each dot represents an individual forecast.

slight modifications to the scheme were made to improve the representation of eddy diffusivity, surface wind inflow angle, and general structure of the TC environment, based on evaluation of this and other schemes through comparison with observational data. These changes are summarized in Gopalakrishnan et al. (2021).

The physics for T-SHiELD were, for the most part, similar to those used in HAFS-globalnest, including the use of the 6-class

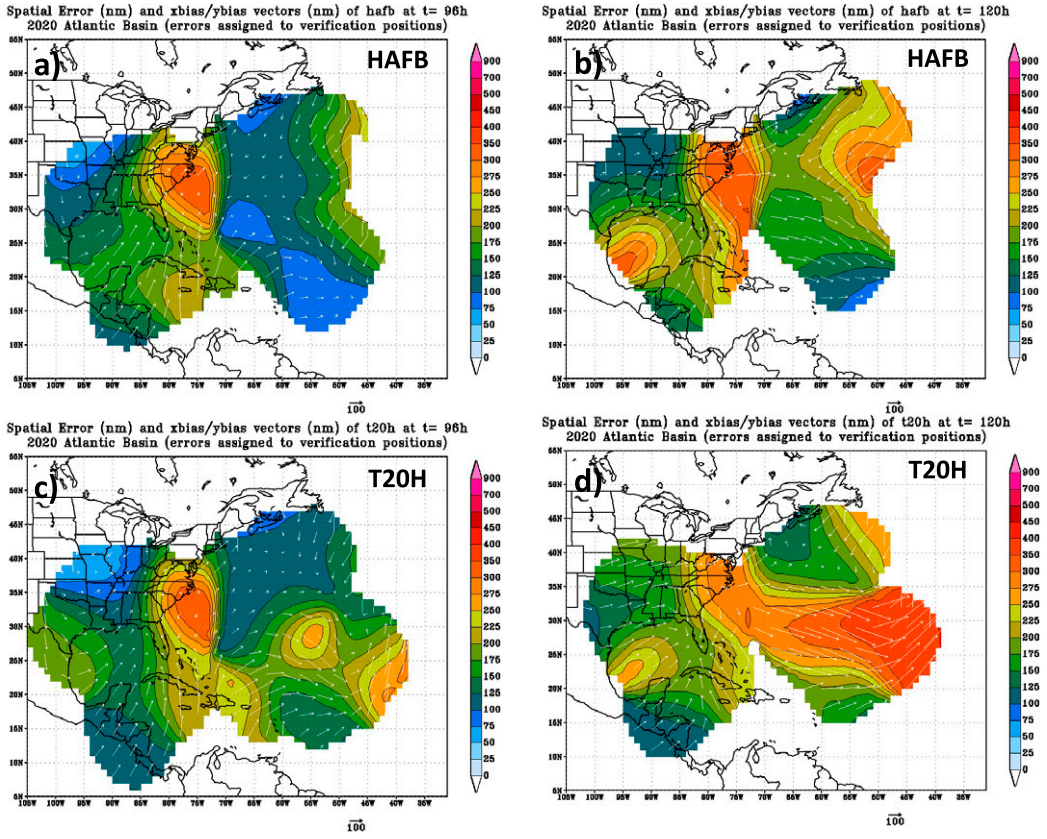

FIG. 4. (a) Spatial distribution of track errors and biases (both in n mi) for HAFS-globalnest at 96 h. The shading shows the magnitude of the track errors, and the vectors show the magnitude and direction of the biases. The unit arrow at the bottom right of each panel indicates the length of arrow for every 100 n mi of track bias. (b) As in (a), but at 120 h. (c) As in (a), but for T-SHiELD. (d) As in (c), but for 120 h.

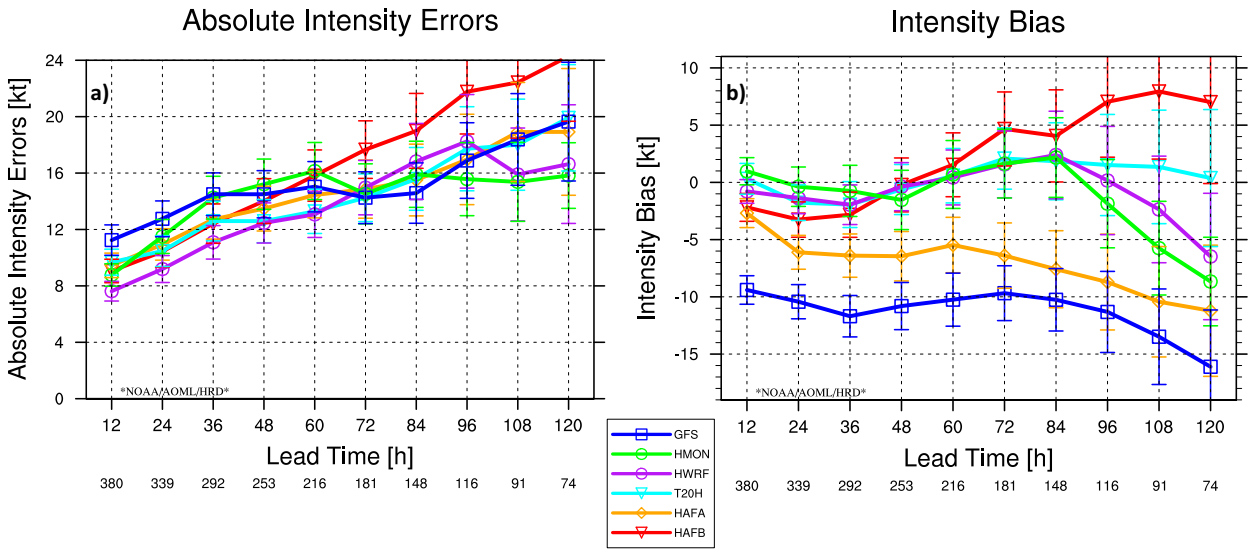

FIG. 5. (a) Mean intensity errors (kt) from HAFS-globalnest (red), T-SHiELD (light blue), HAFS-SAR (orange), GFS (dark blue), HWRF (purple), and HMON (green). The number of homogeneous cases at each forecast hour is shown at the bottom. (b) As in (a), but for mean intensity bias. The bars show the 95% confidence intervals.

GFDL microphysics. However, there were a few noteworthy differences that likely led to some of the forecast differences that will be discussed below. For example, T-SHiELD used the Yonsei State University (YSU; Hong et al. 2006) PBL scheme. A final key difference was the use of a one-dimensional mixed layer ocean model (e.g., Pollard et al. 1973), initialized with the climatological mixed layer depth, in T-SHiELD (but not HAFS) to account for some of the ocean cooling due to turbulent mixing induced by the hurricane's strong winds. A coupled ocean model is in development for HAFS-globalnest, but was not yet available in 2020.

c. *Forecast period and cases covered*

The runs began in early July 2020 and extended into mid-November, capturing the near record breaking late-season activity in the Caribbean. This period provided an extensive dataset of a variety of TCs, including rapid intensification cases (e.g., Hurricane Laura, Hurricane Delta, Hurricane Eta), late-developing cases that peaked right before landfall (e.g., Hurricane Sally and Hurricane Zeta), and weaker TCs that dissipated due to hostile environmental conditions (e.g., Tropical Storm Gonzalo and Tropical Storm Josephine). This variety of cases provided an excellent opportunity to evaluate the performance in a wide range of meteorological situations, which will be a helpful baseline for further model improvements. Modeled TCs were tracked using the GFDL tracker (Marchok 2002, 2021), and these data were compared with the observations from the best track data (Landsea and Franklin 2013) from the National Hurricane Center. HAFS-globalnest and T-SHiELD are compared with both global and

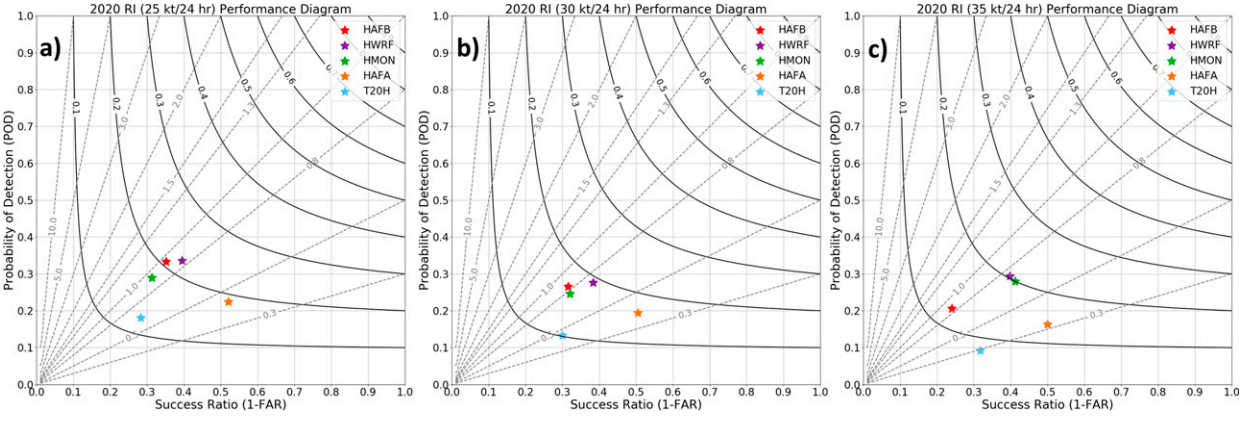

FIG. 6. Performance diagrams for rapid intensification forecasts for the following metrics: (a) 25, (b) 30, and (c) 35 kt (24 h)$^{-1}$. The x axis shows the success ratio (1 minus the false alarm ratio), and the y axis shows the probability of RI detection. The dashed lines are the bias scores, and the solid lines are the critical success index (CSI). See Roebber's (2009) Fig. 2 for details of the dashed lines plotted here (bias scores) and solid lines (critical success index).

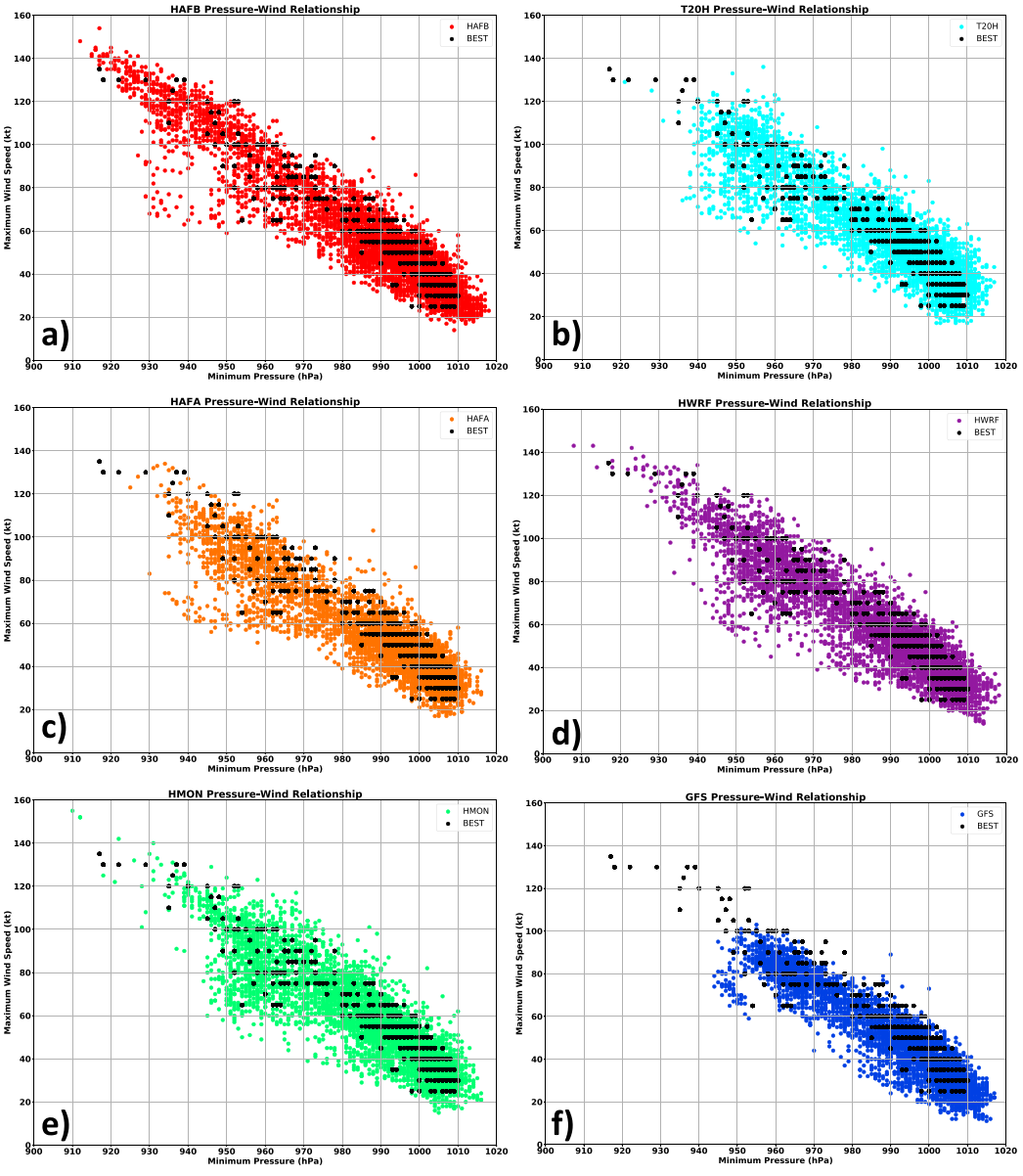

FIG. 7. (a) Pressure–wind relationship for HAFS-globalnest (red) and best track (black). (b) As in (a), but for T-SHiELD (cyan). (c) As in (a), but for HAFS-SAR (orange). (d) As in (a), but for HWRF (purple). (e) As in (a), but for HMON (green). (f) As in (a), but for GFS (blue).

regional operational models including the GFS (Han et al. 2016), HWRF (Tallapragada et al. 2016), and Hurricanes in a Multiscale Ocean-coupled Nonhydrostatic model (HMON; Mehra et al. 2018). HAFS-globalnest and T-SHiELD forecasts are also compared to the previously mentioned HAFS-SAR (Dong et al. 2020). These models are summarized in Table 1, where descriptions of model configurations are given as well as their four-letter Automated Tropical Cyclone Forecast (ATCF) abbreviations that are used in some of the plots (e.g., HAFS-globalnest is designated as HAFB and T-SHiELD is designated as T20H).

3. Results

a. Verification statistics

1) TRACK

Figure 2 shows the track errors for HAFS-globalnest, T-SHiELD, the experimental HAFS-SAR (HAFA), and the three operational models (GFS, HWRF, and HMON). Figure 2a shows the average 5-day track errors, Figs. 2b,c separate those errors into their along- and cross-track components, and Fig. 2d shows the 7-day mean errors for HAFS-globalnest, T-

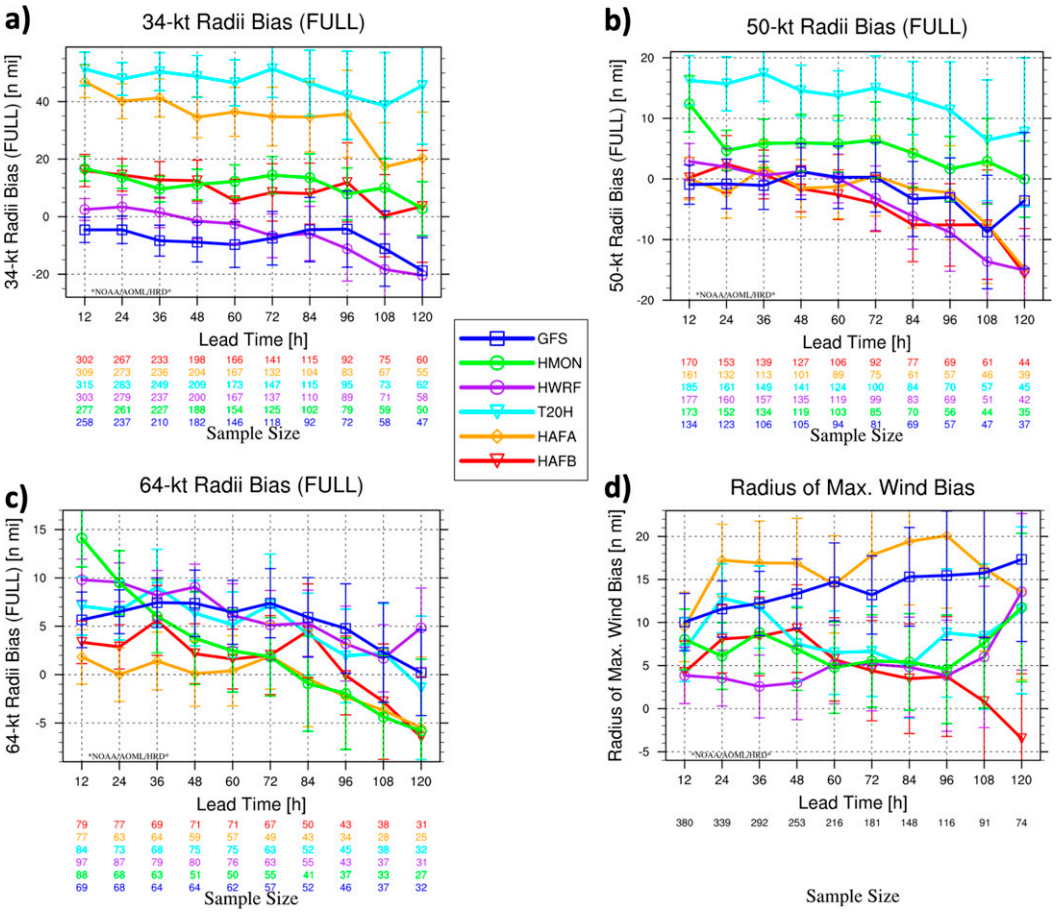

FIG. 8. (a) Mean 34-kt wind radii bias (n mi) out to 120 h from HAFS-globalnest (red), T-SHiELD (light blue), HAFS-SAR (orange), GFS (dark blue), HWRF (purple), and HMON (green). (b) As in (a), but for 50-kt wind radii. (c) As in (a), but for 64-kt wind radii. (d) As in (a), but for radius of maximum winds (RMW). The number of homogenous cases at each forecast hour is shown at the bottom. The bars show the 95% confidence intervals.

SHiELD, and the GFS (the other three regional models were only run to 126 h).

HAFS-globalnest had the lowest overall track errors at all forecast hours out to day 5, outperforming operational HWRF and GFS as well as T-SHiELD. At longer lead times (days 6 and 7), GFS tended to perform better than both HAFS-globalnest and T-SHiELD. However, it should be noted that the sample size at day 7 is around 10% of that at the initial time, due to the fact that there were not very many long-tracking TCs in 2020, so these longer-range results should be interpreted carefully.

Examining the along-track and cross-track bias (Figs. 2b,c) shows that all models had a slight slow bias, but HAFS-globalnest and T-SHiELD performed better in the along-track direction than HAFS-SAR and GFS, which both had large negative (slow) along-track biases by day 5. The main track bias for HAFS-globalnest and T-SHiELD was a significant right-of-track (positive) bias. This was noticeable at early lead times (as early as 12 h for T-SHiELD) and increased over time. It is worth noting that HMON had a similar bias at early leads but this bias did not grow as quickly after 72 h. Figure 3 shows maps of the cross-track bias at 72 h (about the time when the biases increased the most) for HAFS-globalnest and T-SHiELD. These maps show that the largest right-of-track (positive) biases mostly occurred over the western Atlantic and Southeastern Gulf of Mexico, near the edge of the climatological Atlantic subtropical ridge (Davis et al. 1997). Figure 4 further illustrates the track biases from HAFS-globalnest and T-SHiELD by showing gridded plots of the spatial distribution of track biases at 96 and 120 h. Both models had a large east bias over the Western Atlantic near the edge of the climatological subtropical ridge. T-SHiELD also had a fairly substantial right bias over the Eastern Atlantic due to low height biases there (shown later), which contributed to the overall east bias being larger than that in HAFS-globalnest. Later sections will examine the biases in the subtropical ridge and large-scale flow in this region, and also highlight case studies showing examples with a rightward bias in the region.

2) INTENSITY

As mentioned previously, one of the key goals of HFIP (Gopalakrishnan et al. 2020) has been to improve TC intensity

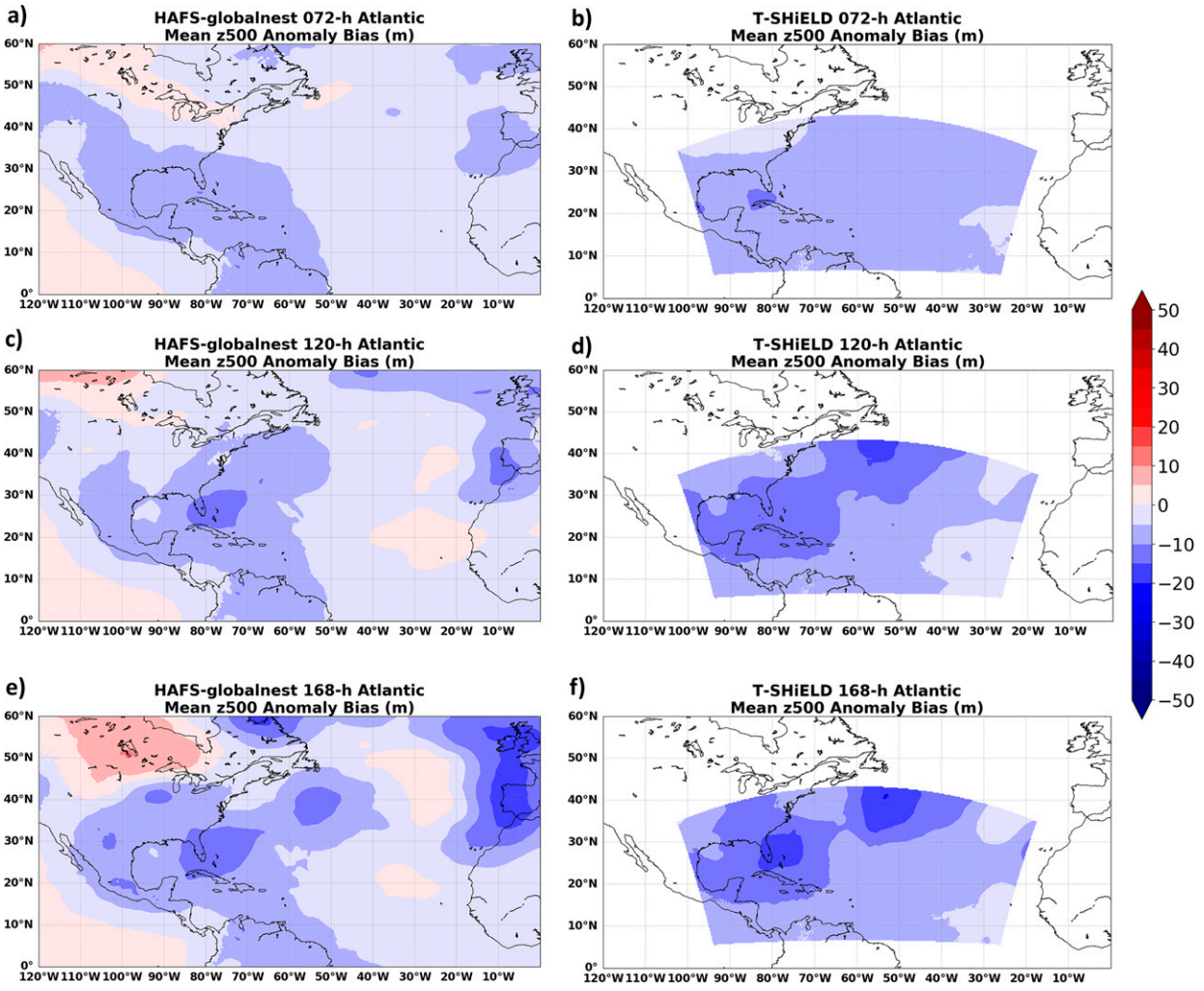

FIG. 9. (a) Mean 500-hPa height bias (m) for HAFS-globalnest over the North Atlantic for all 72-h forecasts. (b) As in (a), but for T-SHiELD. (c) As in (a), but for 120-h forecasts. (d) As in (b), but for 120-h forecasts. (e) As in (a), but for 168-h forecasts. (f) As in (b), but for 168-h forecasts.

forecast skill. Although Official intensity forecasts have improved during the past decade thanks to improved model forecasts (Cangialosi et al. 2020), there is still a lot of work to be done, particularly with regard to rapid intensification, and it is hoped that HAFS will make a major contribution to this effort.

Figure 5 shows the intensity errors and biases for HAFS-globalnest, T-SHiELD, and the other models evaluated in this study. For the first 2–3 days, the intensity errors are all generally similar among most of the different models, although HWRF has slightly lower errors. HAFS-globalnest and T-SHiELD errors (along with HAFS-SAR errors) are worse than HWRF and HMON errors at short leads (0–12 h) presumably due to the lack of vortex-scale data assimilation and initialization. However, T-SHiELD and HWRF performed slightly better at the 2- and 3-day lead times. At longer lead times, T-SHiELD has a mean intensity error slightly worse than HWRF/HMON (similar overall to HAFS-SAR), but better than HAFS-globalnest, which has the highest overall error at 120 h. Examination of intensity biases (Fig. 5b) provides further details about the intensity forecast performance. HAFS-globalnest and T-SHiELD (along with operational HWRF and HMON) have fairly small bias throughout the first 60 h of the forecast. This is in contrast to HAFS-SAR, which has a low bias at all lead times that is noticeable although smaller than that of the global (lower-resolution) operational GFS. At longer lead times, T-SHiELD intensity bias magnitude remains small while HAFS-globalnest has a high bias that gets worse with lead time. This is partly due to the lack of any dynamical ocean coupling (e.g., Yablonsky and Ginis 2009; Bender and Ginis 2000) in this version of HAFS-globalnest.[1] T-SHiELD, on the other hand, uses a one-dimensional mixed layer ocean model that accounts for some of the turbulent mixing due to strong TC winds acting on the ocean, which likely improved

[1] HAFS-SAR was coupled to a regional dynamical ocean model, but this capability is still in development for HAFS-globalnest due to complications associated with two-way feedback and likelihood of artifacts and instabilities along the edge of the regional domain.

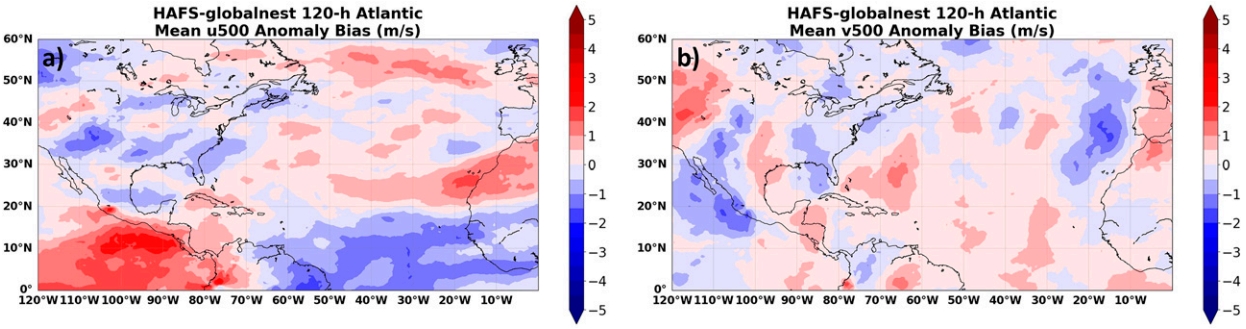

FIG. 10. (a) Mean 500-hPa zonal wind bias (m s^{-1}) for all 120-h forecasts. (b) As in (a), but for mean 500-hPa meridional wind bias (m s^{-1}).

the intensity bias at longer lead times. Options for ocean coupling in HAFS-globalnest are currently being developed to reduce this bias and improve future forecasts.

One of the key HFIP goals with regard to intensity is to improve the performance of forecasts of rapid intensification (RI). There were several rapidly intensifying TCs in the Atlantic during the active 2020 season, providing a useful opportunity to evaluate the skill of the HAFS-globalnest and T-SHiELD which could potentially offer clues to areas of future model development that needs to be addressed. Figure 6 shows the forecast skill of rapid intensification for several thresholds of 24-h intensity change [25, 30, and 35 kt per 24 h; 1 kt ≈ 0.51 m s^{-1}; Kaplan et al. (2010)] based on the methodology outlined in Fig. 2 of Roebber (2009) for these three thresholds. This plot shows the skill of rapid intensification based on probability of detection (i.e., correctly forecasting rapid intensification events that occurred in reality) and false alarm rate (i.e., incorrectly forecasting rapid intensification events that did not occur in reality) and combines these two into a single skill score, the critical success index (CSI). For this analysis, rapid intensification was considered for all 24-h periods in the dataset to give a larger sample size for analysis.

For the 25 kt $(24\ h)^{-1}$ metrics (i.e., how many forecasts or observed intensification periods showed at least 25 kt of intensification in 24 h), HAFS-globalnest has skill comparable to that of HWRF and HMON, while T-SHiELD has somewhat lower skill. The probability of detection from these forecasts is generally better than that shown in Kaplan et al. (2015) for models from the 2008 to 2013 period, indicating progress in RI prediction since that time thanks to model improvements. For the 30 kt $(24\ h)^{-1}$ threshold (the "standard" threshold of RI), HAFS-globalnest is slightly worse than HWRF and HMON, while T-SHiELD again has lower skill, due mostly to a much lower probability of detection. For the 35 kt $(24\ h)^{-1}$ threshold, the skill of HAFS-globalnest and T-SHiELD are both much lower than that of HWRF and HMON, with HAFS-globalnest having a large false alarm rate. It is also interesting to note that the large differences in RI skill between HAFS-globalnest, T-SHiELD, and HAFS-SAR are similar to the overall intensity skill. Besides the differences in ocean coupling, one of the main differences between these three FV3-based models was the PBL scheme. A separate analysis is currently being performed to examine the impact of using a different PBL scheme in both HAFS-globalnest and T-SHiELD. Overall, the results

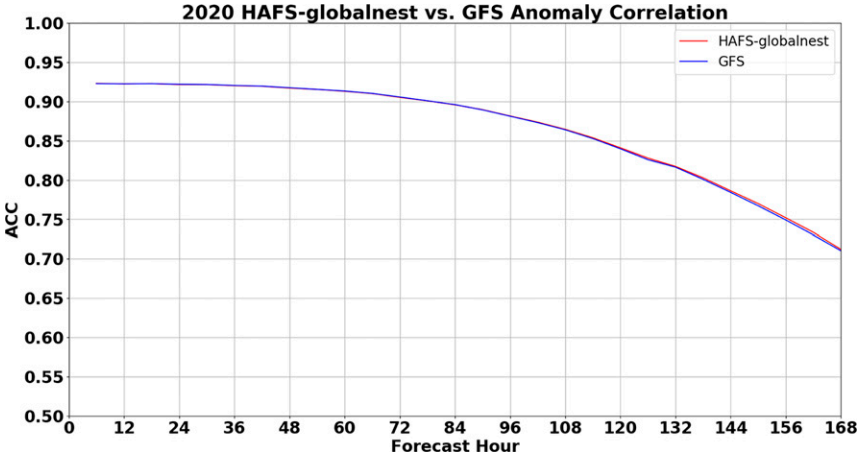

FIG. 11. 500-hPa anomaly correlation for all forecasts during the 2020 Atlantic season for HAFS-globalnest (red) and the corresponding GFS forecasts (blue).

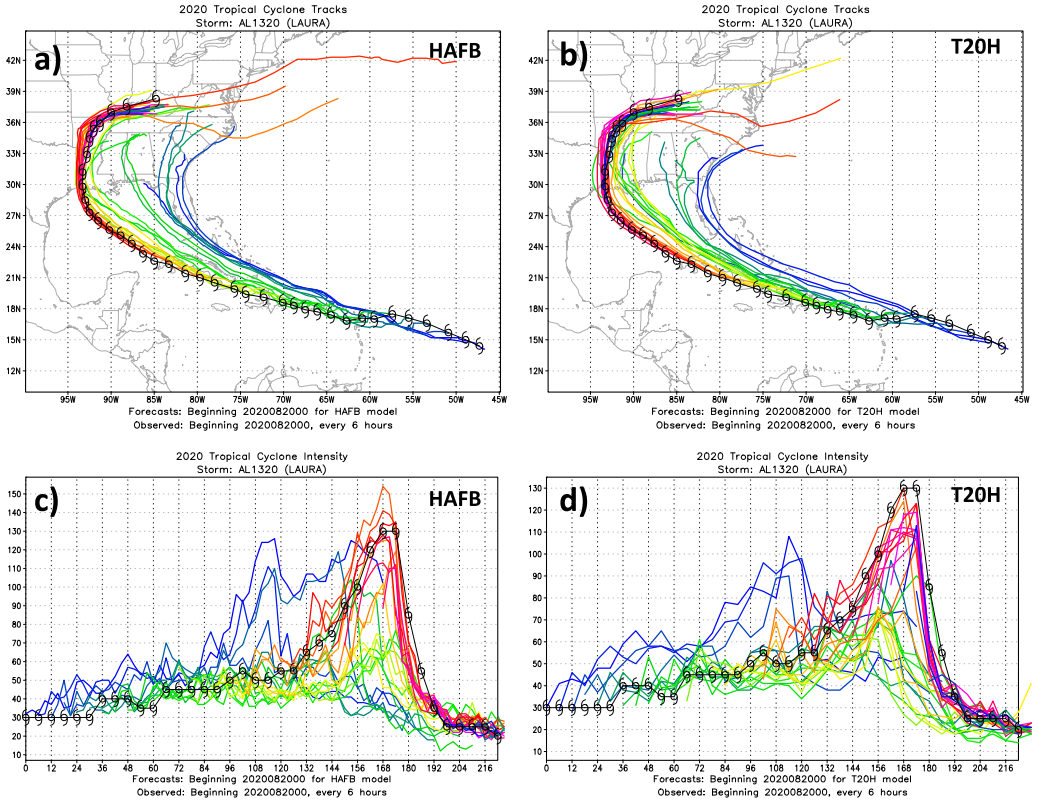

FIG. 12. (a) All HAFS-globalnest (HAFB) tracks for Hurricane Laura, starting at 0000 UTC 20 Aug 2020. The colors specify different initialization times. (b) As in (a), but for T-SHiELD (T20H). (c) All HAFS-globalnest (HAFB) intensity forecasts tracks for Hurricane Laura, starting at 0000 UTC 20 Aug 2020. The colors specify different initialization times. (d) As in (a), but for T-SHiELD (T20H).

show that HAFS-globalnest and T-SHiELD can compete with current operational models in some RI forecasts, but further improvements are needed. In particular, as already mentioned, HAFS-globalnest needs to reduce its high intensity bias, which will be addressed with the introduction of ocean coupling.

One other important aspect of intensity forecast evaluation is the pressure–wind relationship (e.g., Holland 2008). This can provide insights into possible biases and differences between models. Figure 7 shows the pressure–wind relationship from each of the models evaluated as well as the observed pressure–wind relationship from best track. Overall, most of the models are similar at higher pressures and lower wind speeds, and fairly well aligned with observations. The biggest differences come at high wind speeds and lower pressures. The operational GFS, not surprisingly, has trouble capturing the relationship at extreme values due to its coarser resolution (e.g., Moon et al. 2021). There are some differences in the extreme values for HAFS-globalnest versus T-SHiELD and the other hurricane models as well, however. In particular, HAFS-globalnest had many more cases above 120 kt/below 940 hPa than T-SHiELD, consistent with the stronger intensity and positive intensity bias seen previously.

3) WIND RADII

One of the most noticeable biases in the 2019 version of HAFS-globalnest, HAFS-SAR and T-SHiELD was a tendency for the 34-kt wind radii (R34, i.e., "gale force" winds) to quickly become too large in the forecasts (Hazelton et al. 2021). Figure 8 evaluates this metric, as well as 50-kt, 64-kt, and radius of maximum wind (RMW) biases for HAFS-globalnest, T-SHiELD, and the comparable operational and experimental models. The verification of radii is performed against the radii in the "best track" data. It should be noted that this data can have some large uncertainties, particularly when there is no aircraft reconnaissance data (Cangialosi and Landsea 2016). Also, although the 34-, 50-, and 64-kt radii estimates are quality-controlled in the postseason, the RMW values are not, and are mostly similar to the real-time values. However, many of the cases in 2020 were near land and had aircraft data, which increases the confidence in this verification. The radii are based on the maximum radius in each quadrant (southeast, southwest, northeast, and northwest), to be consistent with the operational definition. It should be noted that although the forecast samples were homogeneous, the radii statistics are not perfectly homogeneous because different forecasts produced different wind structures at times.

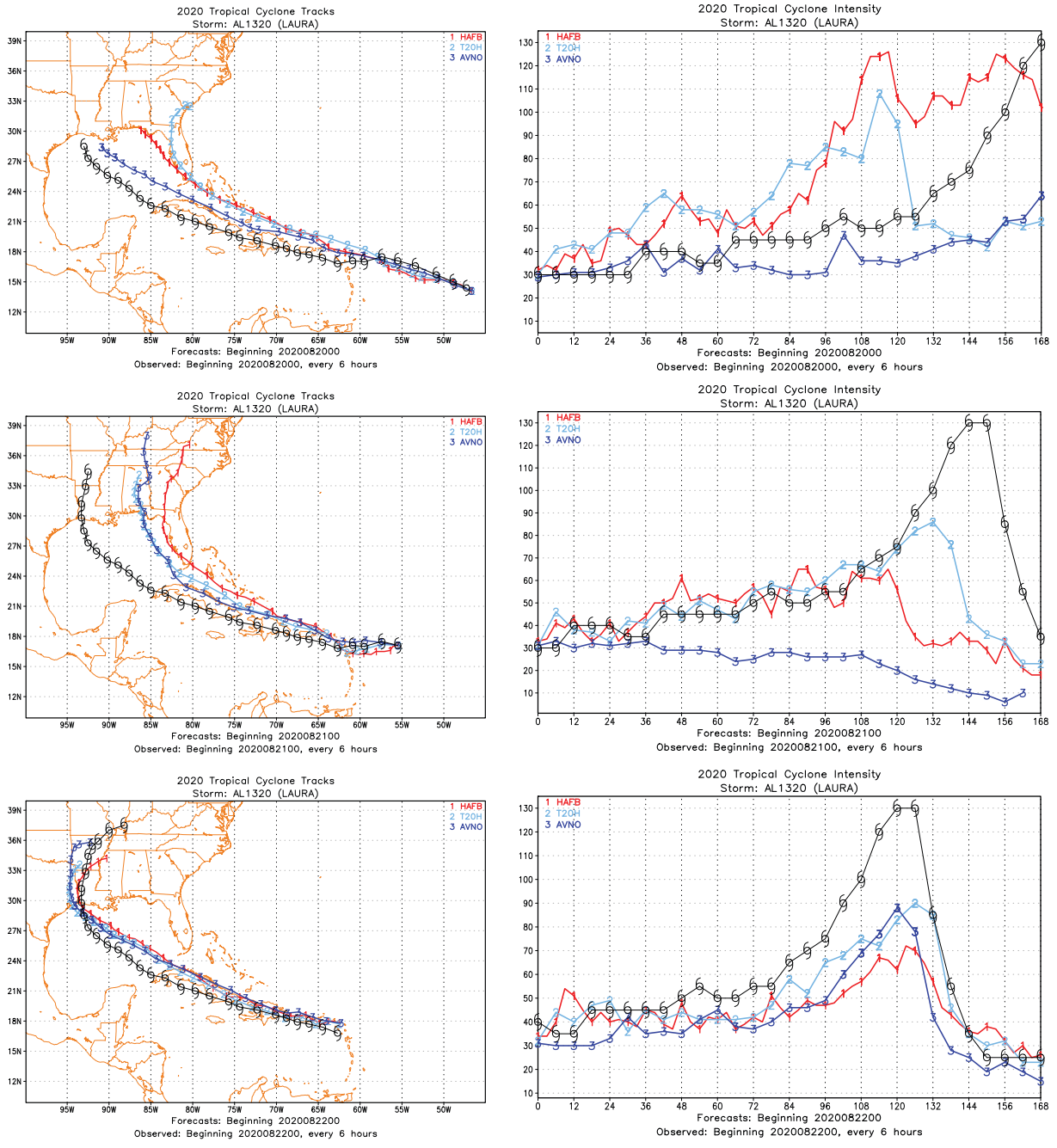

FIG. 13. (a) Tracks of Hurricane Laura from forecasts initialized at 0000 UTC 20 Aug 2020 and run out to 168 h from HAFS-globalnest (red), T-SHiELD (light blue), and the operational GFS (dark blue). The observed track is in black. (b) As in (a), but for intensity forecasts (kt). (c) As in (a), but initialized at 0000 UTC 21 Aug 2020. (d) As in (b), but initialized at 0000 UTC 21 Aug 2020. (e) As in (a), but initialized at 0000 UTC 22 Aug 2020. (f) As in (b), but initialized at 0000 UTC 22 Aug 2020.

For the 34-kt wind radii (Fig. 8a), HAFS-globalnest has a bias that is slightly larger than HWRF but comparable to HMON. However, both T-SHiELD and HAFS-SAR have a very large R34 bias similar to that observed in 2019, with a sharp increase during the first 12 h. Given that the main configuration difference between these three nested-FV3 models is the PBL scheme (with the new modified EDMF-TKE in HAFS-globalnest), this suggests that PBL dynamics may be contributing to these differences in R34. In particular, the lower eddy diffusivity (~25 m^2 s^{-1} for wind speeds of 60 m s^{-1}) in the modified EDMF-TKE scheme (Gopalakrishnan et al. 2021) leads to a sharper inflow angle (Zhang

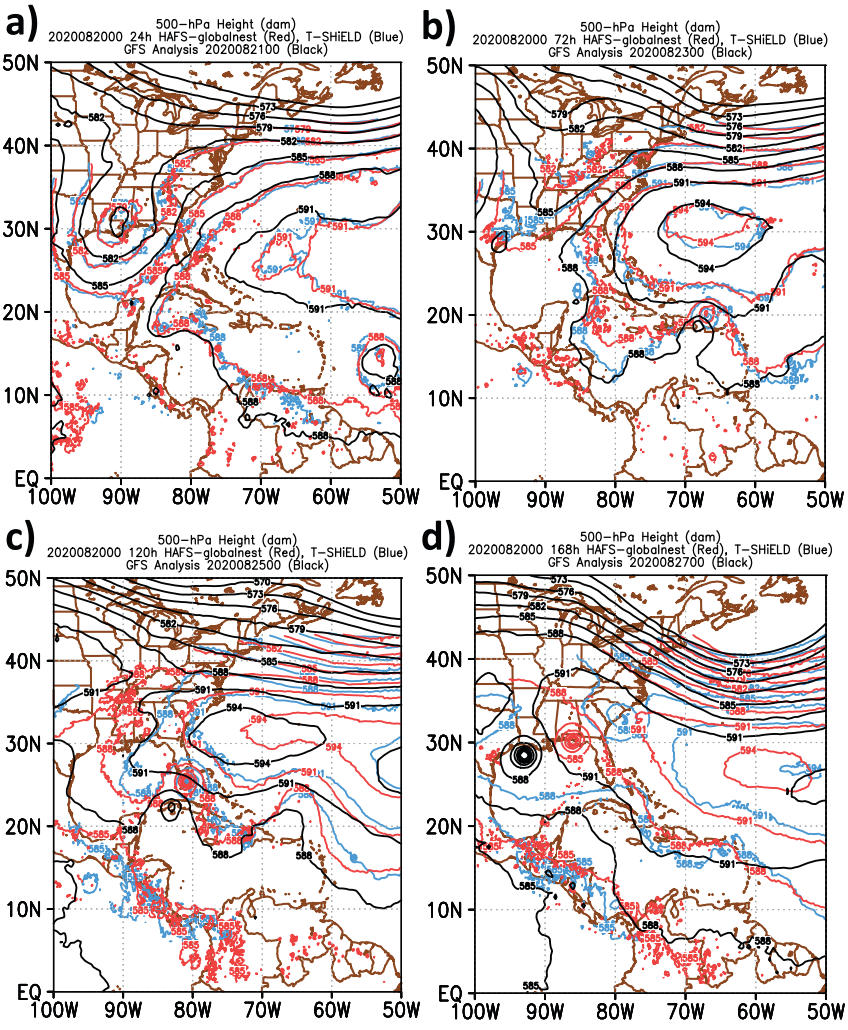

FIG. 14. (a) 500-hPa height (dam) from HAFS-globalnest (red) and T-SHiELD (blue) forecasts initialized at 0000 UTC 20 Aug 2020 and valid at 24 h. The GFS analysis of 500-hPa height at 0000 UTC 21 Aug 2020 is shown in black. (b) As in (a), but for 72-h forecasts valid at 0000 UTC 23 Aug 2020. (c) As in (a), but for 120-h forecasts valid at 0000 UTC 25 Aug 2020. (d) As in (a), but for 168-h forecasts valid at 0000 UTC 27 Aug 2020.

et al. 2017) and stronger/shallower PBL inflow that can impact both the outer structure and inner-core intensity by increasing the inward transport of angular momentum (e.g., Montgomery et al. 2014). A detailed analysis was made of the evolution of the gale-force radii in T-SHiELD during the first 8 h of integration for several of the storms (not shown). This analysis showed development of a strong ring of convection outside of the storm core as the gale radii expanded (not shown), suggesting a vortex adjustment was occurring and/indicating a need for development of a vortex DA to fully rectify this problem. T-SHiELD also had a positive bias in 50-kt wind radii, but the rest of the models were similar to each other. There was little difference between HAFS-globalnest, T-SHiELD, and other models in the mean R64 bias. HAFS-globalnest and T-SHiELD also had relatively small biases in RMW, especially HAFS-globalnest, which had an overall bias near 0 at day 5. The impact of different model PBL schemes on the radii biases, including analysis of differences in convection and inflow angle, is being examined in a separate paper. Further evaluation of storm size metrics, including considerations of observational uncertainty (Cangialosi and Landsea. 2016) as well as whether it is more appropriate to consider quadrant-mean or quadrant-maximum radii, is also a subject of ongoing investigation.

b. *Large-scale biases*

To quantify the large-scale performance of the global-nested modeling systems and also further explore some possible reasons for the track biases that were observed, large-scale 500-hPa geopotential height biases over the North Atlantic were examined in both HAFS-globalnest and T-SHiELD. The

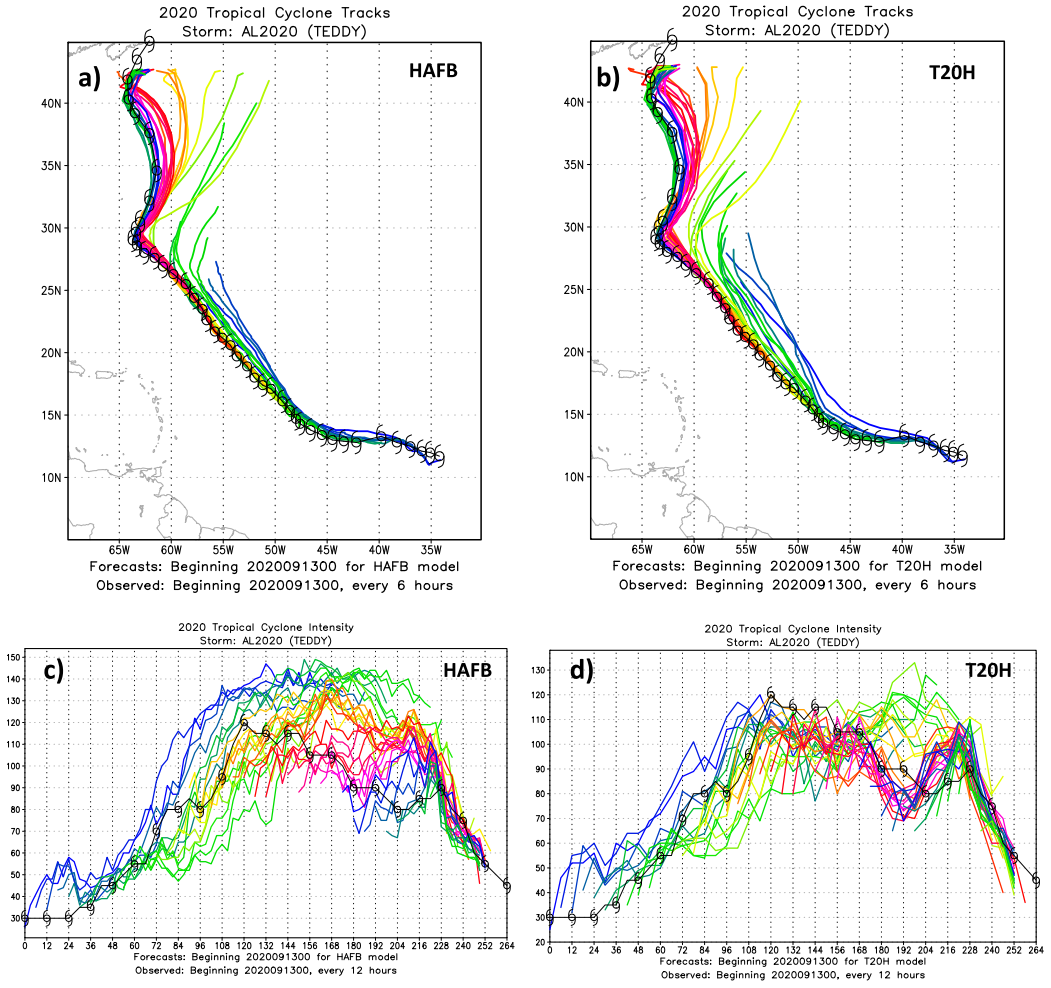

FIG. 15. All HAFS-globalnest (HAFB) tracks for Hurricane Teddy, starting at 0000 UTC 13 Sep 2020. The colors specify different initialization times. (b) As in (a), but for T-SHiELD (T20H). (c) All HAFS-globalnest (HAFB) intensity forecasts tracks for Hurricane Teddy, starting at 0000 UTC 13 Sep 2020. The colors specify different initialization times. (d) As in (a), but for T-SHiELD (T20H).

data for HAFS-globalnest comes from the global grid, while for T-SHiELD was only available for the nest, but the area of interest over the Atlantic is examined in both models. Figure 9 shows the 500-hPa height biases (calculated relative to the ERA5 reanalysis (Hersbach et al. 2020) for T-SHiELD and CFSR (Saha et al. 2010) reanalysis for HAFS-globalnest) over the Atlantic for 72, 120, and 168 h. Both HAFS-globalnest and T-SHiELD have a negative bias in 500-hPa height, particularly over the north-central and southwest Atlantic, both of which are near the edges of the subtropical high. In both models, the negative bias in heights grows over time. The bias for T-SHiELD is worse than that for HAFS, which probably helps explain why the rightward bias in tracks was also worse in T-SHiELD. This negative bias in heights is systematic across forecast lead times, and some possible culprits of this bias, including the model convective scheme, PBL scheme, and vertical level configurations, are being examined in further detail. This persistent bias also appears in individual cases, including one examined in more detail in a later section.

To further explore the changes in steering induced by the bias in the subtropical ridging, the mean 500-hPa wind biases are also calculated (Fig. 10). Consistent with the western edge of the subtropical high being too weak or far to the east, there was a positive (northward) bias in 500-hPa meridional wind over much of the southwest Atlantic. There was also a positive (eastward) bias in the 500-hPa zonal wind across the Antilles and stretching across the eastern Atlantic, indicating that the westward steering currents were consistently too weak basinwide. The reasons for the negative (westward) biases in the equatorial Atlantic combined with positive (eastward) bias in the equatorial Pacific are not apparent, but may be a bias induced by anomalous convection over Central America. This point needs further investigation in a future study.

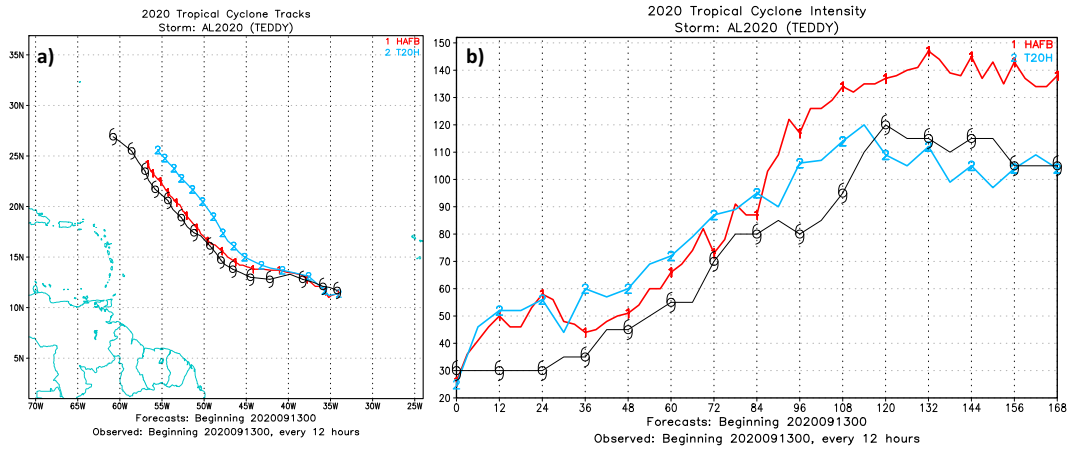

FIG. 16. (a) Tracks of Hurricane Teddy initialized at 0000 UTC 20 Aug 2020 and run out to 168 h from HAFS-globalnest (red) and T-SHiELD (light blue). The observed track is in black. (b) As in (a), but for intensity forecasts.

One of the questions raised by running a global-nested configuration with two-way feedback, as is done in HAFS-globalnest and T-SHiELD, is whether there is any significant large-scale impact from the inclusion of the large static nest? Figure 11 illustrates the large-scale global 500-hPa forecast skill, in the form of the global anomaly correlation (e.g., Tracton et al. 1989) for HAFS-globalnest and operational GFS, from 6 to 168 h. The skill is very similar at all lead times, indicating that there was no significant degradation or improvement from the global-nested configuration on the large-scale forecast, despite some of the regional biases that may have contributed to some of the track biases.

c. *Case studies*

Next, two key case studies from the 2020 hurricane season are examined in detail, to highlight details of the performance of HAFS-globalnest and T-SHiELD, and also to examine in more detail the biases revealed through the verification analysis, especially the right-of-track bias seen in some cases. This will help motivate further development and improvement of HAFS-globalnest.

1) HURRICANE LAURA

Hurricane Laura was the third-strongest hurricane of the 2020 Atlantic hurricane season. It made landfall south of Lake Charles, Louisiana, with 130-kt (150 mph) winds, causing significant storm surge and wind damage over southwest Louisiana. Laura became a tropical cyclone over the central Atlantic Ocean, but struggled to organize early. After multiple days of little intensification as it interacted with the Greater Antilles, Laura rapidly intensified until just prior to landfall in the Gulf of Mexico.

Figure 12 shows all HAFS-globalnest and T-SHiELD track and intensity forecasts for Hurricane Laura. For both models, several of the early forecasts showed a large right bias, keeping the TC north of the Antilles and incorrectly showing a strong hurricane making landfall in Florida. This led to a notable right bias in these cases, as well as a positive intensity bias, since the real TC stayed much weaker for longer due to interaction with the Greater Antilles. There may have also been a feedback between the track and intensity errors in this case, with an excessively strong TC tending to erode the ridge more and move farther to the right. Later forecasts, initialized as the TC was moving along the island chain, homed in on the correct track of Laura, with a landfall in Western Louisiana. Several of these later forecasts also correctly forecast the rapid intensification of Laura in the Gulf of Mexico, where environmental conditions were much more conducive to intensification.

Next, several of the early forecasts that had a significant right bias are examined in detail. Figure 13 shows 7-day track and intensity forecasts from HAFS-globalnest, T-SHiELD, and the operational GFS initialized at 0000 UTC 20, 21, and 22 August. For the forecasts initialized on the 20th, all three models had a right bias, but it was particularly notable in HAFS-globalnest and T-SHiELD, which both showed the landfall of a major hurricane in southeast Florida. This forecast resulted in a 114-h high intensity bias of ~80 kt (41 m s^{-1}) for HAFS-globalnest and ~65 kt (33 m s^{-1}) for T-SHiELD, with the TC incorrectly moving through the warm waters of the Bahamas rather than along the Antilles. This case highlights how the sources of longer-term intensity errors are often dominated by track errors (Emanuel and Zhang 2016). For the forecast initialized 24 h later, the right bias was still present. In this case, T-SHiELD and GFS were similar and HAFS-globalnest was even a little farther right. For this forecast, the intensity bias was not as large early, and actually switched to a negative bias at longer lead times. For the 0000 UTC 22 August forecasts, HAFS-globalnest and T-SHiELD have finally corrected to a track closer to what was observed, although there was still a right bias at early lead times. These forecasts all showed some intensification in the Gulf of Mexico, although not as much as was observed. Later forecasts did a better job of capturing Laura's RI in the Gulf.

To further examine the right bias seen in several of these forecasts for Hurricane Laura, Fig. 14 shows 500-hPa height forecasts at 24, 72, 120, and 168 h from the HAFS-globalnest and T-SHiELD forecasts initialized at 0000 UTC 20 August.

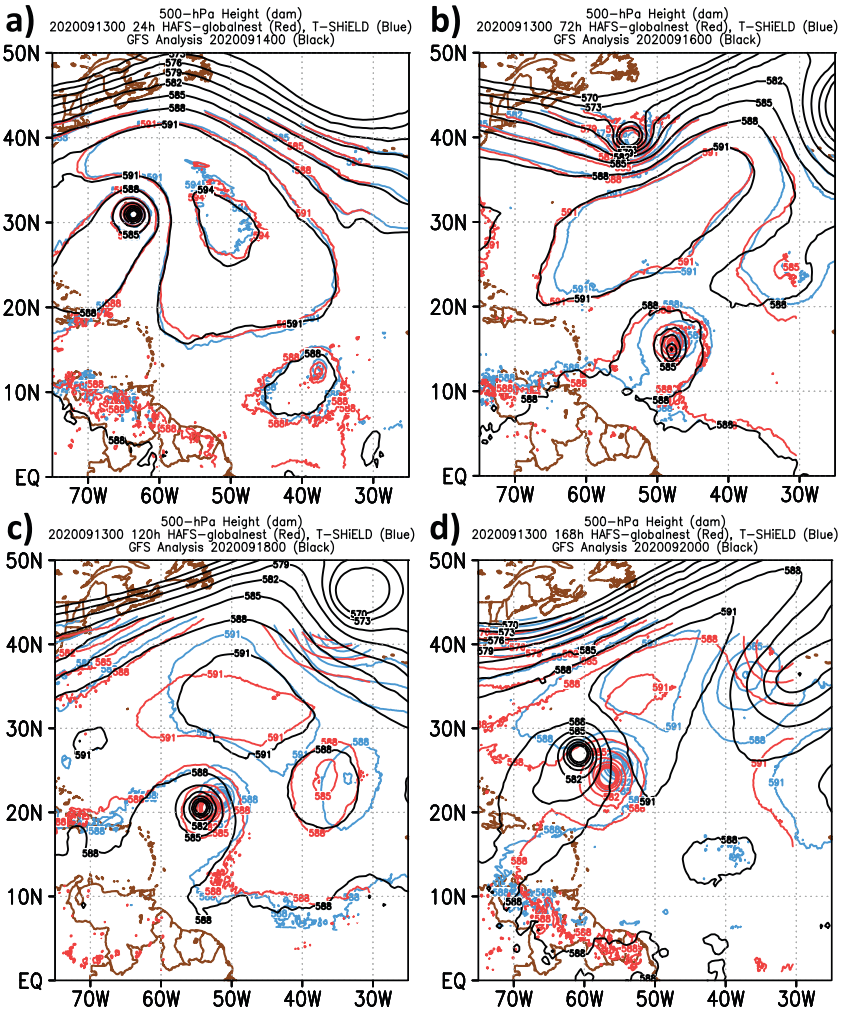

FIG. 17. (a) 500-hPa height (dam) from HAFS-globalnest (red) and T-SHiELD (blue) forecasts initialized at 0000 UTC 13 Sep 2020 and valid at 24 h and the GFS analysis of 500-hPa height valid at 0000 UTC 14 Sep 2020 (black). (b) As in (a), but for 72-h forecasts valid at 0000 UTC 16 Sep 2020. (c) As in (a), but for 120-h forecasts valid at 0000 UTC 18 Sep 2020. (d) As in (a), but for 168-h forecasts valid at 0000 UTC 20 Sep 2020.

The GFS analyses are also shown for validation. Large differences are seen between the model forecasts and observations in the structure of the subtropical ridge over the western Atlantic. Even at 24 h, the 591-dam line north of the Greater Antilles was much too far north and east in HAFS-globalnest and T-SHiELD, indicating excessive weakness in the subtropical ridge (this also indicates that there was likely a bias in the ridge even before the TC began to excessively intensify). By 72 h, the model forecasts of the storm position have diverged to the right of the observed TC, with the bias in the subtropical ridge growing. By day 7, the differences became huge, with the ridge too weak in the model forecasts, with T-SHiELD having an even weaker ridge than HAFS-globalnest. As a result, Laura recurved to the northeast much too early in T-SHiELD. These biases in the subtropical ridge are consistent with the composite biases, and are responsible for the right bias seen in this case and several others throughout the season. Analysis is being conducted to identify some of the potential sources of this bias (which shows up early in the forecasts) in order to correct this issue in future versions of these models.

2) HURRICANE TEDDY

Hurricane Teddy was one of only a few Cabo Verde hurricanes that formed and intensified over the east-central Atlantic during the 2020 season. It intensified into a category-4 hurricane while moving northwest over the central Atlantic, before weakening and transforming into a powerful extratropical cyclone as it impacted eastern Canada.

Figure 15 shows all the track and intensity forecasts for HAFS-globalnest and T-SHiELD for Hurricane Teddy. Looking at the tracks, the early forecasts for HAFS-

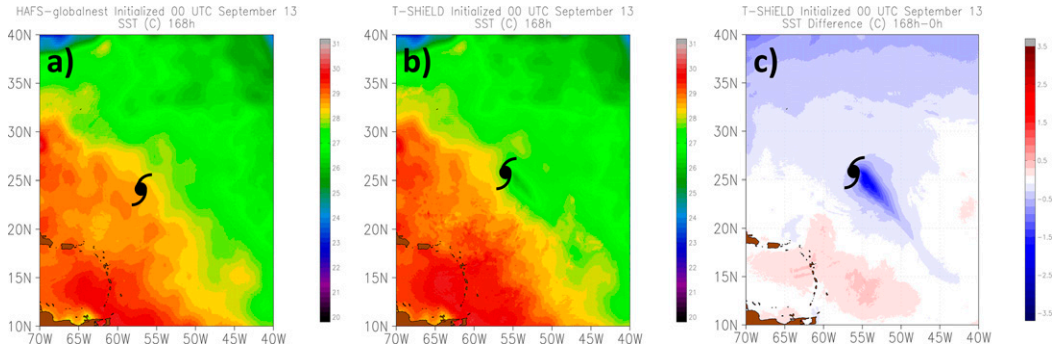

FIG. 18. (a) Sea surface temperature (°C) from HAFS-globalnest, initialized at 0000 UTC 13 Sep 2020, and valid at 0000 UTC 20 Sep 2020. (b) As in (a), but for T-SHiELD. (c) SST difference (°C) between the 168-h forecast and the initial T-SHiELD SST field from this same forecast. In all panels, the forecast position of the TC is shown in black.

globalnest verified well (with a slight east bias), while those for T-SHiELD were consistently too far to the east. This appears to be another case where the right bias in T-SHiELD was even more pronounced than that in HAFS-globalnest. This bias in an early forecast of Teddy is examined below. Later forecasts of both HAFS-globalnest and T-SHiELD were generally close to the observed track, although there was a slight right bias for both models as the TC interacted with the midlatitude flow over the north-central Atlantic. For the intensity forecasts, the difference between the forecasts was stark. Both models generally captured the rapid intensification during the first 2–5 days of Teddy's life cycle. While T-SHiELD generally kept the peak intensity correct close to the best track estimate of 120 kt (62 m s^{-1}), HAFS-globalnest excessively deepened the storm to 130–140 kt (67–72 m s^{-1}) in several forecasts, and was also too slow to show the TC decaying toward the end of its life cycle. The forecast cycle examined below will highlight one of the forecasts with this large divergence in intensity between the two models. It is also worth noting that Hurricane Teddy was one of a few long-lived long-track TCs during 2020, meaning that the high bias from this one case likely contributed to the overall high bias seen in the longer-range intensity forecasts for HAFS-globalnest.

Figure 16 shows the 7-day track and intensity forecasts for Hurricane Teddy from HAFS-globalnest and T-SHiELD initialized at 0000 UTC 13 September 2020. In both models, the TC is slightly right of the observed track and slower than observed. However, the track forecast from HAFS-globalnest is closer to observed and generally parallels the best track. In terms of intensity, the T-SHiELD and HAFS-globalnest forecasts are both very close to each other and to the observed for the first half (84 h) of the forecast. After that, both models correctly show the period of pronounced intensification into a major hurricane that was seen in the real storm, but HAFS-globalnest shows too much intensification, whereas T-SHiELD correctly shows the TC peaking as a category-4 hurricane before levelling off. This case is an excellent microcosm of the seasonal results for both models, with a right bias that was slightly worse in T-SHiELD, and a high intensity bias in HAFS-globalnest.

As was the case with Hurricane Laura, both HAFS-globalnest and T-SHiELD had similar biases in the prediction of the subtropical ridge that was steering Hurricane Teddy (Fig. 17). The bias appeared by 72 h, with too little ridging northeast of the TC in both HAFS-globalnest and especially T-SHiELD. At 120 h, both HAFS-globalnest and T-SHiELD showed too small/weak of a ridge, but HAFS-globalnest did a better job with the evolution of an upper-level low to the east of Teddy, and its track was also closer to the observed. By the end of the 7-day forecast, the ridge was too weak in both HAFS-globalnest and T-SHiELD, leading to solutions that were too slow and too far east in both forecasts. The edge of the 500-hPa ridge (as seen in the 588-dam contour) was farther southwest in HAFS-globalnest, leading to a track that was slightly closer to the observed.

Teddy was a case that illustrated how the inclusion of a simple one-dimensional ocean coupling in T-SHiELD (vs no dynamical ocean in HAFS-globalnest), impacted the intensity forecasts. Whereas HAFS-globalnest produced no cold wake for Hurricane Teddy, T-SHiELD showed cooling of 1°–2°C along the storm track (Fig. 18). At day 7, HAFS-globalnest placed Teddy over water with SSTs of 28°–29°C, while T-SHiELD showed the TC over water with SSTs of ~27°C. This difference due to the ocean response was likely exacerbated by both the sharp zonal gradient of SST in this region and also the recent passage of Paulette to the north. This case underscores the need for further development of HAFS to include a dynamic ocean and account for such processes.

Finally, due to a large expansion in wind radii as it moved north and the fact that it was one of the longest-lived TCs of the 2020 season, Teddy provided a useful case to examine the details of the radii forecasts from HAFS-globalnest and T-SHiELD. Figure 19 shows the wind radii verification statistics for Hurricane Teddy. HAFS-globalnest and T-SHiELD both kept the RMW of Teddy too small at longer forecast leads, although the bias characteristics for R34 were very different: HAFS-globalnest was like most operational guidance in showing the bias decreasing with lead time, while T-SHiELD was too large. Teddy was a case that was relatively well sampled by reconnaissance aircraft,

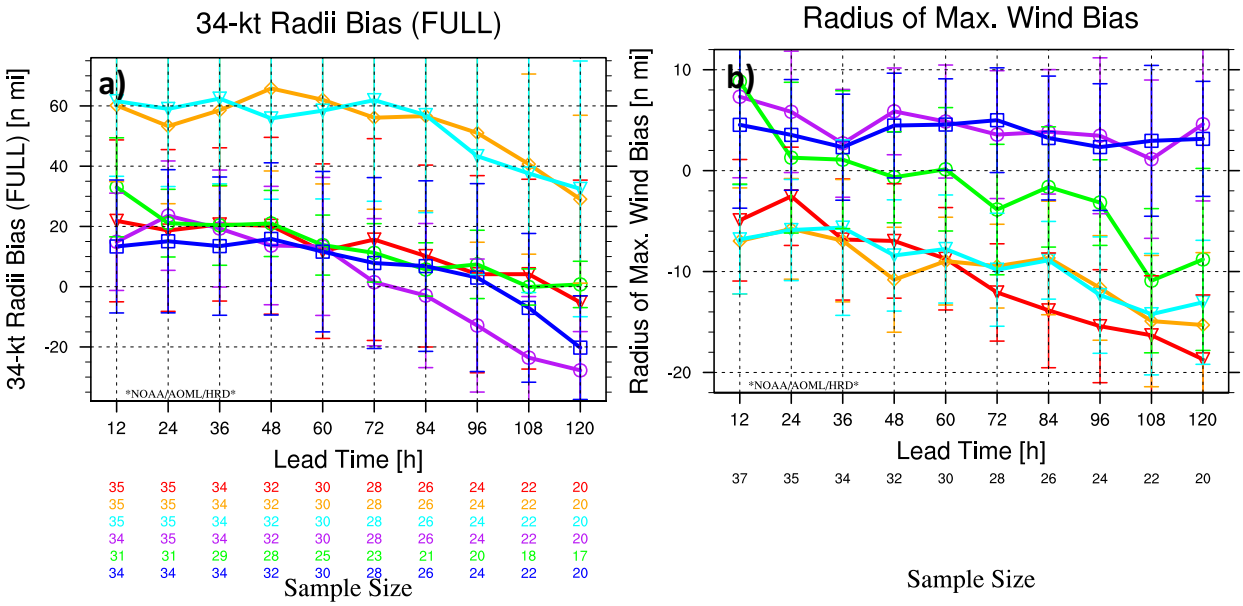

FIG. 19. (a) Mean 34-kt wind radii bias (n mi) out to 120 h from HAFS-globalnest (red), T-SHiELD (light blue), HAFS-SAR (orange), GFS (dark blue), HWRF (purple), and HMON (green) for Hurricane Teddy cases. (b) As in (a), but for radius of maximum winds (RMW). The number of homogenous cases at each forecast hour is shown at the bottom. The bars show the 95% confidence intervals.

which gives confidence to these wind radii biases that were generally similar to those in the full 2020 sample. Work is ongoing to understand how model characteristics lead to differences in storm structure, including storm size (R34) and eyewall size (RMW).

4. Discussion and conclusions

Evaluation of a large set of near-real-time model forecasts from the 2020 Atlantic Hurricane Season was made using both the HAFS-globalnest model developed at NOAA's Hurricane Research Division (HRD) and GFDL's global-nested T-SHiELD. The purpose of this study was to provide insightful analysis of these two new modeling systems in order to enhance future development of HAFS. Both models showed promising track forecast results relative to other operational and experimental forecast models. Despite this overall skillful track performance, a right bias seen in several storms appears to be due to an erroneously weak representation of the subtropical ridge over the western Atlantic as the forecasts evolved. This bias showed up in the case study of Hurricane Laura as well as in a composite mean for both models over the entire 2020 season. The bias was slightly worse in T-SHiELD, where even the eastern portion of the subtropical ridge became too weak in some cases, as shown in Hurricane Teddy. One avenue of future and ongoing work is to continue to identify the exact source of this bias (whether in the model dynamics, physics, or initialization) in order to mitigate it in future versions of these models. In particular, the convection scheme is being explored, to determine whether activating this scale-aware scheme on the 3-km nest improves the ridge bias and track forecasts. Preliminary testing indicates that the subtropical ridge position and intensity, and therefore TC track, is sensitive to activating and tuning both the deep and shallow convection schemes, and these results will be considered in a future analysis.

One key difference between the HAFS-globalnest and T-SHiELD intensity forecasts appeared in the intensity bias. HAFS-globalnest tended to have a high intensity bias at longer lead times, while T-SHiELD did not have a large bias in either direction. One of the contributing factors to this difference were the different PBL schemes used, although both the modified EDMF-TKE and YSU schemes should reasonably predict TC structure. A few subsequent tests (not shown) using EDMF-TKE in T-SHiELD have suggested that slight differences in the distribution of vertical levels in the boundary layer could play an important role in regulating the intensity forecasts. This is not surprising, as past studies have found that TC intensity forecasts are particularly sensitive to the distribution of vertical levels, especially in the PBL (e.g., Ma et al. 2012; Zhang et al. 2015). The differences in the PBL scheme and vertical resolution could have also contributed to the differences in the 34-kt wind radii, where T-SHiELD had radii that were consistently too large while HAFS-globalnest had a much smaller bias. The role of the PBL scheme in regulating TC structure and intensity in HAFS-B is the subject of ongoing investigation. While the intensity was too high in HAFS-globalnest, the physics and vertical resolution changes introduced in 2020, compared to the 2019 version led to improved structure forecasts.

Another key factor influencing the intensity differences between the two models is the one-dimensional ocean model employed in T-SHiELD, which accounted for some of the ocean turbulent mixing induced by the TC–ocean interaction

and reduced the intensity toward a more realistic value compared to HAFS-globalnest (Bender and Ginis 2000). A couple of tests of T-SHiELD with the 1D ocean model turned off (not shown) confirm that the simple one-dimension coupling significantly improves the intensity forecasts for strong TCs. While a detailed examination of each factor (e.g., PBL physics, ocean coupling) in a more controlled experiment is an important topic of ongoing and future research, the results do provide motivation for the ongoing adoption of ocean coupling in future upgrades to HAFS-globalnest, as it appears that this will lead to significant reduction of the high bias.

A significant amount of research and development is ongoing and planned for the HAFS-globalnest and T-SHiELD configurations. As mentioned above, ocean coupling is being developed and tested for HAFS-globalnest for future seasons, to hopefully mitigate much of the high intensity bias that was seen in some 2020 forecasts. Preliminary testing shows promise in reducing this high bias in cases like Teddy as well as Paulette from 2020. In addition, a data assimilation system is being developed that will be used in both the HAFS-SAR and HAFS-globalnest configurations. When complete, this will significantly improve the simulations over the current coarse GFS initialization, and should reduce much of the early low bias and spinup issues that were seen in some forecasts. In addition, research is ongoing to modify and improve the EDMF-TKE PBL scheme for use in HAFS and T-SHiELD, based on observational data. This will continue to improve the structure and intensity forecasts. Another project being pursued is modification of the GFDL microphysics based on microphysics observations collected in NOAA P-3 flights.

Acknowledgments. The lead author was supported by NOAA Grant NA19OAR0220187. The authors thank Sim Aberson, John Kaplan, and three anonymous reviewers for their helpful comments that improved an earlier version of the manuscript.

Data availability statement. Graphical data from these forecasts are available on the AOML Hurricane Model Viewer (https://storm.aoml.noaa.gov/basin/?projectName=BASIN). Track files and other raw data are available on the NOAA RDHPCS computer system, or by request.

REFERENCES

Bender, A. A., and I. Ginis, 2000: Real-case simulations of hurricane–ocean interaction using a high-resolution coupled model: Effects on hurricane intensity. *Mon. Wea. Rev.*, **128**, 917–946, https://doi.org/10.1175/1520-0493(2000)128<0917:RCSOHO>2.0.CO;2.

Cangialosi, J. P., and C. W. Landsea, 2016: An examination of model and official National Hurricane Center tropical cyclone size forecasts. *Wea. Forecasting*, **31**, 1293–1300, https://doi.org/10.1175/WAF-D-15-0158.1.

——, E. Blake, M. DeMaria, A. Penny, A. Latto, E. Rappaport, and V. Tallapragada, 2020: Recent progress in tropical cyclone intensity forecasting at the National Hurricane Center. *Wea. Forecasting*, **35**, 1913–1922, https://doi.org/10.1175/WAF-D-20-0059.1.

Chen, J.-H., and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, https://doi.org/10.1175/JCLI-D-12-00061.1.

Davis, R. E., B. P. Hayden, D. A. Gay, W. L. Phillips, and G. V. Jones, 1997: The North Atlantic subtropical anticyclone. *J. Climate*, **10**, 728–744, https://doi.org/10.1175/1520-0442(1997)010<0728:TNASA>2.0.CO;2.

Dong, J., and Coauthors, 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**, 617, https://doi.org/10.3390/atmos11060617.

Emanuel, K., and F. Zhang, 2016: On the predictability and error sources of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **73**, 3739–3747, https://doi.org/10.1175/JAS-D-16-0100.1.

Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, https://doi.org/10.1175/BAMS-D-12-00071.1.

Gao, K., L. Harris, J.-H. Chen, S.-J. Lin, and A. Hazelton, 2019: Improving AGCM hurricane structure with two-way nesting. *J. Adv. Model. Earth Syst.*, **11**, 278–292, https://doi.org/10.1029/2018MS001359.

Gopalakrishnan, S., and Coauthors, 2020: 2019 HFIP R&D activities summary: Recent results and operational implementation. HFIP Tech. Rep. HFIP2020-1, 42 pp., https://repository.library.noaa.gov/view/noaa/26468.

——, A. Hazelton, and J. A. Zhang, 2021: Improving hurricane boundary layer parameterization scheme based on observations. *Earth Space Sci.*, **8**, e2020EA001422, https://doi.org/10.1029/2020EA001422.

Han, J., and C. S. Bretherton, 2019: TKE-based moist eddy-diffusivity mass-flux (EDMF) parameterization for vertical turbulent mixing. *Wea. Forecasting*, **34**, 869–886, https://doi.org/10.1175/WAF-D-18-0146.1.

——, M. L. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. K. Fletcher, and C. S. Bretherton, 2016: 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. *Wea. Forecasting*, **31**, 341–352, https://doi.org/10.1175/WAF-D-15-0053.1.

——, W. Wang, Y. C. Kwon, S.-Y. Hong, V. Tallapragada, and F. Yang, 2017: Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Wea. Forecasting*, **32**, 2005–2017, https://doi.org/10.1175/WAF-D-17-0046.1.

Harris, L., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev.*, **141**, 283–306, https://doi.org/10.1175/MWR-D-11-00201.1.

——, and Coauthors, 2020: GFDL SHiELD: A unified system for weather-to-seasonal prediction. *J. Adv. Model. Earth Sys.*, **12**, e2020MS002223, https://doi.org/10.1029/2020MS002223.

Hazelton, A. T., L. Harris, and S.-J. Lin, 2018a: Evaluation of tropical cyclone structure forecasts in a high-resolution version of the multiscale GFDL fvGFS model. *Wea. Forecasting*, **33**, 419–442, https://doi.org/10.1175/WAF-D-17-0140.1.

——, M. Bender, M. Morin, L. Harris, and S.-J. Lin, 2018b: 2017 Atlantic hurricane forecasts from a high-resolution version of the GFDL fvGFS model: Evaluation of track, intensity, and

structure. *Wea. Forecasting*, **33**, 1317–1337, https://doi.org/10.1175/WAF-D-18-0056.1.

——, and Coauthors, 2021: 2019 Atlantic hurricane forecasts from the global-nested Hurricane Analysis and Forecast System (HAFS): Composite statistics and key events. *Wea. Forecasting*, **36**, 519–538, https://doi.org/10.1175/WAF-D-20-0044.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Holland, G., 2008: A revised hurricane pressure–wind model. *Mon. Wea. Rev.*, **136**, 3432–3445, https://doi.org/10.1175/2008MWR2395.1.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Kaplan, J., M. DeMaria, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **25**, 220–241, https://doi.org/10.1175/2009WAF2222280.1.

——, and Coauthors, 2015: Evaluating environmental impacts on tropical cyclone rapid intensification predictability utilizing statistical models. *Wea. Forecasting*, **30**, 1374–1396, https://doi.org/10.1175/WAF-D-15-0032.1.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, https://doi.org/10.1175/MWR-D-12-00254.1.

——, and J. P. Cangialosi, 2018: Have we reached the limits of predictability for tropical cyclone track forecasting? *Bull. Amer. Meteor. Soc.*, **99**, 2237–2243, https://doi.org/10.1175/BAMS-D-17-0136.1.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2.

——, and R. B. Rood, 1996: Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Wea. Rev.*, **124**, 2046–2070, https://doi.org/10.1175/1520-0493(1996)124<2046:MFFSLT>2.0.CO;2.

Ma, Z., J. Fei, X. Huang, and X. Cheng, 2012: Sensitivity of tropical cyclone intensity and structure to vertical resolution in WRF. *Asia-Pac. J. Atmos. Sci.*, **48**, 67–81, https://doi.org/10.1007/s13143-012-0007-5.

Marchok, T., 2002: How the NCEP tropical cyclone tracker works. *25th Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., P1.13, https://ams.confex.com/ams/25HURR/techprogram/paper_37628.htm.

——, 2021: Important factors in the tracking of tropical cyclones in operational models. *J. Appl. Meteor. Climatol.*, **60**, 1265–1284, https://doi.org/10.1175/JAMC-D-20-0175.1.

Mehra, A., V. Tallapragada, Z. Zhang, B. Liu, L. Zhu, W. Wang, and H.-S. Kim, 2018: Advancing the state of the art in operational tropical cyclone forecasting at NCEP. *Trop. Cyclone Res. Rev.*, **7**, 51–56, https://doi.org/10.6057/2018TCRR01.06.

Montgomery, M. T., J. A. Zhang, and R. K. Smith, 2014: An analysis of the observed low-level structure of rapidly intensifying and mature Hurricane Earl (2010). *Quart. J. Roy. Meteor. Soc.*, **140**, 2132–2146, https://doi.org/10.1002/qj.2283.

Moon, J., J. Park, and D.-H. Cha, 2021: Does increasing model resolution improve the real-time forecasts of western North Pacific tropical cyclones? *Atmosphere*, **12**, 776, https://doi.org/10.3390/atmos12060776.

Pollard, R. T., P. B. Rhines, and R. O. R. Y. Thompson, 1973: The deepening of the wind-mixed layer. *Geophys. Fluid Dyn.*, **4**, 381–404, https://doi.org/10.1080/03091927208236105.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Tallapragada, V., and Coauthors, 2016: Hurricane Weather Research and Forecasting (HWRF) Model: 2015 scientific documentation. NCAR Tech. Note NCAR/TN-522+STR, 122 pp., https://doi.org/10.5065/D6ZP44B5.

Tien, D. D., T. Ngo-Duc, H. T. Mai, and C. Kieu, 2013: A study of the connection between tropical cyclone track and intensity errors in the WRF model. *Meteor. Atmos. Phys.*, **122**, 55–64, https://doi.org/10.1007/s00703-013-0278-0.

Tracton, M. S., K. Mo, W. Chen, E. Kalnay, R. Kistler, and G. White, 1989: Dynamical Extended Range Forecasting (DERF) at the National Meteorological Center. *Mon. Wea. Rev.*, **117**, 1604–1635, https://doi.org/10.1175/1520-0493(1989)117<1604:DERFAT>2.0.CO;2.

Yablonsky, R. M., and I. Ginis, 2009: Limitation of one-dimensional ocean models for coupled hurricane–ocean model forecasts. *Mon. Wea. Rev.*, **137**, 4410–4419, https://doi.org/10.1175/2009MWR2863.1.

Zhang, D., L. Zhu, X. Zhang, and V. Tallapragada, 2015: Sensitivity of idealized hurricane intensity and structures under varying background flows and initial vortex intensities to different vertical resolutions in HWRF. *Mon. Wea. Rev.*, **143**, 914–932, https://doi.org/10.1175/MWR-D-14-00102.1.

Zhang, J. A., R. F. Rogers, and V. Tallapragada, 2017: Impact of parameterized boundary layer structure on tropical cyclone rapid intensification forecasts in HWRF. *Mon. Wea. Rev.*, **145**, 1413–1426, https://doi.org/10.1175/MWR-D-16-0129.1.

Zhou, L., S.-J. Lin, J.-H. Chen, L. M. Harris, X. Chen, and S. L. Rees, 2019: Toward convective-scale prediction within the next generation global prediction system. *Bull. Amer. Meteor. Soc.*, **100**, 1225–1243, https://doi.org/10.1175/BAMS-D-17-0246.1.