Sensitivity of HAFS-B Tropical Cyclone Forecasts to Planetary Boundary Layer and Microphysics Parameterizations

ANDREW HAZELTON,[a,b] XIAOMIN CHEN,[c] GHASSAN J. ALAKA JR.,[b] GEORGE R. ALVEY III,[a,b] SUNDARARAMAN GOPALAKRISHNAN,[b] AND FRANK MARKS[b]

[a] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida*
[b] *NOAA/AOML/HRD, Miami, Florida*
[c] *University of Alabama in Huntsville, Huntsville, Alabama*

(Manuscript received 26 July 2023, in final form 5 January 2024, accepted 8 February 2024)

ABSTRACT: Understanding how model physics impact tropical cyclone (TC) structure, motion, and evolution is critical for the development of TC forecast models. This study examines the impacts of microphysics and planetary boundary layer (PBL) physics on forecasts using the Hurricane Analysis and Forecast System (HAFS), which is newly operational in 2023. The "HAFS-B" version is specifically evaluated, and three sensitivity tests (for over 400 cases in 15 Atlantic TCs) are compared with retrospective HAFS-B runs. Sensitivity tests are generated by 1) changing the microphysics in HAFS-B from Thompson to GFDL, 2) turning off the TC-specific PBL modifications that have been implemented in operational HAFS-B, and 3) combining the PBL and microphysics modifications. The forecasts are compared through standard verification metrics, and also examination of composite structure. Verification results show that Thompson microphysics slightly degrades the days 3–4 forecast track in HAFS-B, but improves forecasts of long-term intensity. The TC-specific PBL changes lead to a reduction in a negative intensity bias and improvement in RI skill, but cause some degradation in prediction of 34-kt (1 kt $\approx$ 0.51 m s^{-1}) wind radii. Composites illustrate slightly deeper vortices in runs with the Thompson microphysics, and stronger PBL inflow with the TC-specific PBL modifications. These combined results demonstrate the critical role of model physics in regulating TC structure and intensity, and point to the need to continue to develop improvements to HAFS physics. The study also shows that the combination of both PBL and microphysics modifications (which are both included in one of the two versions of HAFS in the first operational implementation) leads to the best overall results.

SIGNIFICANCE STATEMENT: A new hurricane model, the Hurricane Analysis and Forecast System (HAFS), is being introduced for operational prediction during the 2023 hurricane season. One of the most important parts of any forecast model are the "physics parameterizations," or approximations of physical processes that govern things like turbulence, cloud formation, etc. In this study, we tested these approximations in one configuration of HAFS, HAFS-B. Specifically, we looked at two different versions of the microphysics (modeling the growth of water and ice in clouds) and boundary layer physics (the approximations for turbulence in the lowest level of the atmosphere). We found that both of these sets of model physics had important effects on the forecasts from HAFS. The microphysics had notable impacts on the track forecasts, and also changed the vertical depth of the model hurricanes. The boundary layer physics, including some of our changes based on observed hurricanes and turbulence-resolving models, helped the model better predict rapid intensification (periods where the wind speed increases quickly). Work is ongoing to improve the model physics for better forecasts of rapid intensification and overall storm structure, including storm size. The study also shows the combination of both PBL and microphysics modifications overall leads to the best results and thus was used as one of the two first operational implementations of HAFS.

KEYWORDS: Boundary layer; Hurricanes/typhoons; Tropical cyclones; Cloud microphysics; Forecast verification/skill; Operational forecasting

1. Introduction

The Hurricane Analysis and Forecast System (HAFS) is the tropical cyclone (TC) component of the National Oceanic and Atmospheric Administration (NOAA) Unified Forecast System (UFS). After several years of testing and evaluation, including multiple years of real-time experiments (e.g., Dong et al. 2020; Hazelton et al. 2021, 2022a) through the support of the Hurricane Forecast Improvement Project (HFIP; Gopalakrishnan et al. 2021), HAFS has been operationally implemented by NOAA during the summer of 2023. In the 2023 operational implementation, there are two different configurations (HAFS-A/HAFS-B), which replaced the operational Hurricane Weather Research and Forecasting (HWRF) Model and Hurricanes in a Multiscale Ocean-coupled Nonhydrostatic Model (HMON), respectively. One of them, HAFS-B, makes use of physics options that, prior to 2022, had not been previously tested in real-time HAFS experiments. In particular, HAFS-B uses different microphysics and PBL formulations than HAFS-A. The details will be discussed below.

It is expected that different cloud microphysics schemes would noticeably impact TC track, intensity, and structure forecasts within HAFS. Willoughby et al. (1984) was one of the first studies to examine the impacts of microphysics on TCs,

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

DOI: 10.1175/WAF-D-23-0124.1

comparing microphysics with ice and without (warm rain only) in an axisymmetric, nonhydrostatic model. This simple model produced notable intensity and structure differences, with more concentric rings in the simulations with ice processes included. Fovell and Su (2007) tested different microphysics schemes in ensemble simulations of Hurricane Rita (2005), and found a clear sensitivity of the track forecasts to choice of microphysics parameterizations, as well as an impact on the TC size forecasts. Li and Pu (2008) examined the rapid intensification of Hurricane Emily (2005) in the Advanced Research version of the Weather Research and Forecasting (WRF) Model (WRF-ARW) and found that there were large differences of up to 19 hPa in simulated TC intensity in forecasts with different microphysics. That study also examined the sensitivity to PBL physics, and found that eyewall convective heating and low-level fluxes and θ_e were among the variables that were most important for TC intensification in the simulations. Tyner et al. (2018) showed that forecasts of secondary eyewall formation (SEF) in the HWRF model were sensitive to the radial and vertical distribution and transport of solid-phase hydrometeors, based on sensitivity tests altering particle fall speed. A recent study by Park et al. (2020) compared track, intensity, and structure forecasts in the west Pacific using WRF with 3-class and 6-class microphysics schemes, and found that the more sophisticated scheme produced more realistic forecasts due to increased latent heat release.

There has also been some limited work specifically evaluating the performance of the Thompson microphysics in TCs. For example, Brown et al. (2016) examined several microphysics schemes through comparison with polarimetric radar in 2014 TCs Hurricane Arthur and Tropical Storm Ana, and found that the Thompson scheme produced the simulation that most closely matched observations. Wu et al. (2021) evaluated rainband structure in TC forecasts using Thompson and two other microphysics schemes, and found that all schemes produced too little ice above 12 km, and Thompson in particular had more development of snow in the 8–12-km layer. This body of prior work suggests that it is worthwhile to evaluate how the Thompson scheme impacts overall TC forecast skill in HAFS, by isolating its contribution to the forecasts of TC track, intensity, and structure in HAFS-B.

Prior studies have also clearly demonstrated that PBL physics parameterizations have significant effects on numerical forecasts of TC evolution. Braun and Tao (2000) was an early example of such a study (using the MM5 model), finding notable sensitivity of TC structure and intensity to the choice of PBL scheme. They noted that the precipitation structure seemed to be as sensitive to the choice of PBL scheme as the cloud microphysics. Nolan et al. (2009) examined two PBL schemes in simulations of Hurricane Isabel (2003), and found that modifications to the parameterizations of ocean roughness improved the simulations. Gopalakrishnan et al. (2013) examined idealized simulations of the HWRF model, and found that reducing the eddy diffusivity in the PBL scheme based on aircraft observations led to stronger inflow, a more robust warm core, and a smaller TC core with a stronger intensity. Several subsequent studies (Tallapragada et al. 2014; Zhang et al. 2015, 2017) found that this improved PBL scheme led to more accurate predictions of TC intensity and structure in HWRF forecasts, including in operational HWRF. Based on the observational framework established using HWRF in the studies mentioned above, an initial study in HAFS demonstrated how two different PBL schemes converged to similar improvements in TC structure with observations-based improvements to eddy diffusivity and mixing length (Gopalakrishnan et al. 2021). One of the schemes, the modified eddy-diffusivity mass flux with prognostic turbulent kinetic energy (EDMF-TKE) scheme, was subsequently used in experimental runs of HAFS and showed value in improving forecasts of TC structure and rapid intensification (Hazelton et al. 2022b). Recent evaluation of the EDMF-TKE in hurricane conditions based on large eddy simulations (LES) further revealed issues of excessive vertical mixing and mismatched PBL and surface-layer parameterizations; modifications to the original EDMF-TKE, termed as TC-specific modifications (or the "tc-pbl" option), were proposed to address these issues, followed by an implementation into HAFS (Chen et al. 2022, 2023). One unique aspect of the LES is that they were performed under realistic thermodynamic conditions of mature hurricanes.

In this study, we have the unique opportunity to evaluate the relative impacts of cloud microphysics and TC-specific PBL physics in a large retrospective study covering several hundred cases in multiple TCs over multiple years, expanding beyond prior case studies. This will lead to a better understanding of how the microphysics and PBL schemes impact TC track, intensity, and structure across a wider spectrum of cases that vary in size, location, and intensity. The results from this analysis will also be uniquely positioned to inform ongoing developments in HAFS, as future upgrades seek to improve the model representation of TC structure for better prediction of track, structure, and hazards such as rainfall (through physics upgrades) and storm surge (through better prediction of the surface wind field).

2. Model setup and experiment design

a. Model domain, configuration, and physics

The configuration of HAFS used in this study is one of the two "HAFS Version 1" preoperational baselines that was run for a full 3-yr (2020–22) retrospective set in collaboration between NOAA's Environmental Modeling Center (EMC) and the Hurricane Research Division (HRD) of NOAA's Atlantic Oceanographic and Meteorological Laboratory (AOML). The specific version of HAFS used is HAFS-B (the other pre operational version is HAFS-A). This version of HAFS-B features a large static nest (approximately 72° × 72°) with 6-km grid spacing and a smaller (approximately 12° × 12°) moving nest (Ramstrom et al. 2023, manuscript submitted to *Front. Earth Sci.*) that follows the TC, with 2-km grid spacing. The outer, static nest of HAFS is coupled to the Hybrid Coordinate Ocean Model (HYCOM; Bleck 2002), which runs concurrently with the atmospheric model and uses a fixed ocean domain at 9-km grid spacing that covers the Atlantic and east Pacific basins. Figure 1a shows the grid configuration for an example case of HAFS-B (note that there is no ocean coupling north of the northern boundary of HYCOM around 45°N). HAFS uses

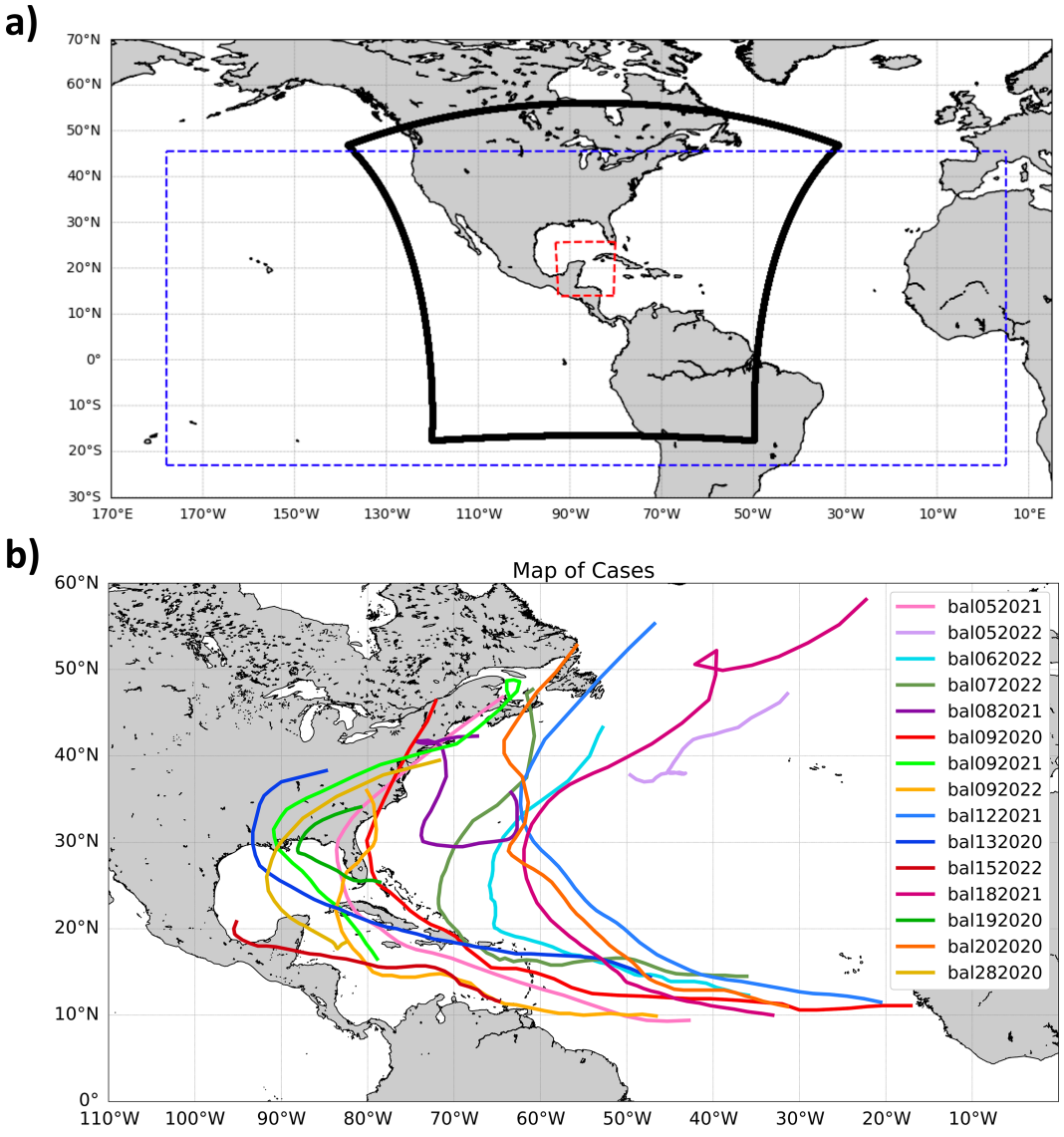

FIG. 1. (a) HAFS-B grid configuration. The red box shows the 2-km moving nest centered on the TC. The black box shows the 6-km static outer nest. The blue box shows the static HYCOM ocean model domain at 9-km grid spacing. (b) Map of all 15 TCs analyzed in this study.

81 hybrid (sigma-pressure) levels in the vertical, with levels more densely packed in the PBL and at mid- to upper levels for TC steering.

As mentioned in section 1, the key differences between HAFS-B and HAFS-A are aspects of the model physics, as HAFS-B uses the Thompson cloud microphysics (Thompson et al. 2004) and also uses TC-specific modifications to the EDMF-TKE PBL scheme (Han and Bretherton 2019) based on observations and large eddy simulations (Chen et al. 2021, 2022). These modifications are discussed in more detail below. HAFS-B and HAFS-A also use the scale-aware version (Han et al. 2017) of the simplified Arakawa–Schubert (SAS) convective scheme on both the 6- and 2-km domains, and the Rapid Radiative Transfer Model for General Circulation Models (RRTMG) scheme described in Iacono et al. (2008). The surface drag contains adjustments at high wind speeds (Bender et al. 2007), similar to those previously used in the HWRF model.

b. *Description of physics sensitivity tests*

The three physics sensitivity tests that were performed for a group of cases, and compared with the "default" HAFS-B retrospective runs, are described in Table 1. The sensitivity tests were performed by varying the microphysics or PBL physics in HAFS-B or by varying both. For two of the tests (HFSB_GFDLTC and HFSB_GFDLNoTC), the 6-class single moment Geophysical Fluid Dynamics Laboratory (GFDL) microphysics scheme (Zhou et al. 2022) was used instead of the Thompson scheme. Note that the GFDL microphysics scheme is used in

TABLE 1. Description of the differences between the four experiments analyzed, including the differences in microphysics and PBL physics.

Experiment description	Experiment name	Microphysics	PBL physics
HAFS-B (Default)	HFSB_ThompTC	Thompson	EDMF-TKE with tc-pbl, mixing length cap = 75 m
HAFS-B (GFDL MP)	HFSB_GFDLTC	GFDL	EDMF-TKE with tc-pbl, mixing length cap = 75 m
HAFS-B (No TC-PBL)	HFSB_ThompNoTC	Thompson	Default EDMF-TKE
HAFS-B (GFDL MP + No TC-PBL)	HFSB_GFDLNoTC	GFDL	Default EDMF-TKE

the operational HAFS-A, and was used extensively in previous HAFS real-time experiments. The GFDL microphysics is a single-moment scheme, while the Thompson scheme is double-moment. They produce the same classifications of hydrometeors, although the Thompson scheme has been shown to produce more snow (e.g., Wu et al. 2021) while the GFDL scheme tends to produce more graupel. As a result, the Thompson scheme tends to produce lower simulated radar reflectivities above the melting level.

For two of the tests (HFSB_ThompNoTC and HFSB_GFDLNoTC), the PBL scheme varied between the default EDMF-TKE scheme and the EDMF-TKE scheme with the TC-specific modifications described in Chen et al. (2021, 2022), using the "tc-pbl" option. These modifications include changes to the mixing length profile in the PBL and surface layer based on LES data, and also some modification to the vertical mass flux calculation. The mixing length changes also include a cap of 75 m on the asymptotic mixing length, which is similar to the value of 100 m (based on observations) used in more simple modification experiments (Gopalakrishnan et al. 2021; Hazelton et al. 2022a). For the default scheme, a value of 250 m was used as in HAFS-A. Note that the larger mixing length tends to make the PBL scheme more diffusive. Thus, HAFS-B tends to be less diffusive than HAFS-A, which can promote more inflow, stronger influx of angular momentum, and a stronger TC (e.g., Hazelton et al. 2022b; Chen et al. 2023). It should be noted that the mixing length modifications are only enabled on the inner nest, in order to minimize the impact on large-scale prediction. The tc-pbl changes are not turned off when the system moves over land, but were designed to focus on the high-wind environment in TCs over water.

c. *Cases analyzed*

For this analysis, a subset of notable, high-impact, and generally long-lived cases from the 2020–22 Atlantic hurricane seasons were selected to maximize the sample size from days 0 to 5. The forecasts were initialized every 6 h, and were run for the entire life cycle of the storms in order to get a sampling of forecasts for storms in both weak versus strong as well as early-stage versus mature parts of the TC life cycle. The list of cases that were analyzed, along with the number of initialization times for each, is listed in Table 2. Figure 1b shows the tracks (observed "best track") of all these cases as well. All reached hurricane intensity at some point, which may slightly bias the results toward stronger TCs. These cases were chosen to maximize the sample size, as they were longer-lived TCs. They occurred in a variety of environments in different parts of the basin, ranging from TCs in marginal environments with high shear (Elsa, Isaias) to TCs in very favorable environments (e.g., Sam, Teddy). This gives confidence in the representativeness of these results for different kinds of Atlantic TCs.

3. Results

a. *Verification results*

The first set of evaluations of these physics sensitivity tests was performed by performing verifications of standard metrics (track, intensity, and wind radii) using National Hurricane Center (NHC) verification methodologies (e.g., Cangialosi 2022). The TCs were tracked using the Geophysical Fluid Dynamics Laboratory (GFDL) vortex Tracker (Marchok 2021) and were verified against the best track data from NHC (Rappaport et al. 2009). The verification graphics include the "consistency metric" described in Ditchek et al. (2023), which provides a measure of whether the verification statistics at a given forecast hour are robust and consistent, or are influenced largely by outlier cases. The "baseline" for skill is HFSB_GFDLNoTC, which contains neither of the physics changes that were implemented in HFSB_ThompTC, and is closer to the HAFS-A version also being implemented in operations (with some slight differences). This is done in order to assess the impact of the individual and combined physics changes in the model. It should be noted that for the verification analysis in this study, we used the "late" guidance–the raw, noninterpolated model output–for comparison with best track. Finally, it is worth noting that the verification results presented here do include the periods where the TCs are over land

TABLE 2. Cases analyzed in the study, including the storm ID, peak intensity, and number of initialization times.

Storm	Storm ID	Peak intensity	No. of cases
Isaias (2020)	AL092020	75 kt (39 m s^{-1})	32
Laura (2020)	AL132020	130 kt (67 m s^{-1})	33
Sally (2020)	AL192020	95 kt (49 m s^{-1})	22
Teddy (2020)	AL202020	120 kt (62 m s^{-1})	44
Zeta (2020)	AL282020	100 kt (51 m s^{-1})	16
Elsa (2021)	AL052021	75 kt (39 m s^{-1})	36
Henri (2021)	AL082021	65 kt (33 m s^{-1})	29
Ida (2021)	AL092021	130 kt (67 m s^{-1})	17
Larry (2021)	AL122021	110 kt (57 m s^{-1})	43
Sam (2021)	AL182021	135 kt (69 m s^{-1})	49
Danielle (2022)	AL052022	80 kt (41 m s^{-1})	29
Earl (2022)	AL062022	90 kt (46 m s^{-1})	32
Fiona (2022)	AL072022	115 kt (59 m s^{-1})	42
Ian (2022)	AL092022	135 kt (69 m s^{-1})	33
Lisa (2022)	AL152022	70 kt (36 m s^{-1})	23
Total cases: 480			

(Fig. 1b), as is typical for standard verifications. It might be worthwhile in a future study to stratify the results into cases where the TCs are only over water.

1) TRACK RESULTS

Figure 2 shows the track verification results from each experiment. Although the track results are generally neutral overall (Figs. 2a,b), HFSB_ThompTC does not perform quite as well as some of the other configurations. It has negative overall mean skill relative to HFSB_GFDLNoTC, and at least marginally consistent degradation is found as well. The best overall performance comes from HFSB_GFDLTC, which implies that the microphysics are likely leading to the slight degradation of track skill in HFSB_ThompTC more than the tc-pbl option. Examining the across-track (Fig. 2d) and along-track (Fig. 2c) components of track error, HFSB_ThompNoTC and HFSB_GFDLNoTC both have notable left-of-track bias, indicating that the tc-pbl option is actually improving the across-track bias in this sample. HFSB_ThompTC and HFSB_ThompNoTC have larger negative along-track bias than HFSB_GFDLTC and HFSB_GFDLNoTC, which indicates that the Thompson microphysics is likely leading to a slow bias in TC tracks. It is not immediately clear whether this is due to differences in intensity and structure (outlined next) or differences in the large-scale flow. Further examination of this point would be a useful subject for future work.

Figure 3 shows the percentage point contribution (PPC) of each storm at each forecast hour to the overall track skill (where skill is taken relative to HFSB_GFDLNoTC). Thus, this graphic reveals which storms contributed the most positively or negatively to the overall forecast skill for each experiment, and also provides a measure of whether the skill is driven mostly by outliers or a wide range of cases. Some cases that stand out in Fig. 3 are as follows: 1) Hurricane Sam, where the two experiments that used tc-pbl (HFSB_ThompTC and HFSB_GFDLTC; Figs. 3a and 3b, respectively) were notably degraded relative to HFSB_GFDLNoTC. 2) Hurricane Ian, where the two experiments that used GFDL microphysics (HFSB_GFDLTC and HFSB_GFDLNoTC) outperformed HFSB_ThompTC (Fig. 3a) and HFSB_ThompNoTC (Fig. 3c). 3) Hurricane Laura, where the other three experiments showed notable improvement in skill relative to the HFSB_GFDLNoTC baseline at the longer lead times. The details from some of these cases will be examined in a follow up study. In all three cases, there was a left-of-track bias that was noted in these HAFS experiments.

2) INTENSITY RESULTS

Next, the intensity forecasts are examined. Understanding the physics configurations that lead to optimal intensity forecast skill will be important to ensuring that HAFS contributes to further improvement in intensity forecast skill, like the improvement that has been realized over the last decade (Cangialosi et al. 2020). The results (Fig. 4) demonstrate that, in contrast to the track forecast skill, the combination of physics selected in operational HAFS-B (HFSB_ThompTC) produces the best intensity forecast results. While both HFSB_ThompNoTC and HFSB_GFDLTC produce marginally to fully consistent improvement over HFSB_GFDLNoTC, HFSB_ThompTC (with both physics changes in place) produces the best and most consistent intensity skill improvement. In addition, it appears that the PBL and microphysics contribute differently to the overall intensity skill. For example, HFSB_ThompNoTC (with Thompson microphysics but no tc-pbl) produces similar overall skill and consistency results to HAFS-B at longer lead times, suggesting the microphysics are important for the medium range intensity skill (note the large drop off in skill for HFSB_GFDLTC at days 4–5). However, the skill around days 2–3 is better in HFSB_GFDLTC (with GFDL microphysics but keeping the tc-pbl option). In addition, HFSB_GFDLTC has notably better (smaller negative) intensity bias at all lead times than the other two sensitivity experiments (Fig. 4c). This indicates that the PBL changes are leading to stronger TCs overall. Reasons for this difference are explored in the structure section later in the paper. Interestingly, all of the experiments had a negative pressure bias (Fig. 4d) despite the negative Vmax bias, which indicates that the pressure–wind relationship in HAFS needs to be optimized (and may indicate some size biases, which is examined next). HFSB_ThompNoTC had the largest pressure bias despite also having a larger negative bias in wind speed, which does indicate that the PBL changes are improving the pressure–wind relationship.

A storm-by-storm examination of the intensity skill results for the sensitivity tests using the PPC graphics (Fig. 5) indicates that, just like for track, Hurricane Laura was a case where some of the sensitivity tests differed sharply from HFSB_ThompTC. In particular, the two experiments that used GFDL microphysics (HFSB_GFDLTC and HFSB_GFDLNoTC; Fig. 5b and the baseline) were notably worse than HFSB_ThompTC (Fig. 5a) for this case. There were also some notable differences in 2022 Hurricanes Ian, Fiona, and Earl.

Another key metric of intensity forecast performance is the skill of forecasts undergoing TC rapid intensification (RI). Improvement of RI forecasts is one of the key goals of HFIP, and so assessing the RI skill of various configurations of HAFS will be critical toward reaching this goal. The traditional wind-based definition of rapid intensification is a 30-kt (15.4 m s^{-1}) increase in maximum sustained wind speed within a 24-h period (Kaplan and DeMaria 2003). The HFIP definition of RI (DeMaria et al. 2021) is similar, but expands the sample size by evaluating RI when any model from the sample or the best track underwent RI during the 24-h verification period.

The RI skill of the physics sensitivity experiments are assessed in two ways. First, the intensity forecast verification is performed for the subset of cases fitting the HFIP-RI definition. Figures 6a–c shows the intensity error, skill (relative to HFSB_GFDLNoTC), and bias for the forecasts meeting this definition. Note that while the sample size is fairly small (5%–10% of the total) as RI does not occur frequently, it is still a comprehensive assessment over 3 years (2020–22) of notable, high-impact, and generally long-lived cases. As with the overall intensity performance, the intensity skill for HFSB_ThompTC is better than the other experiments, although all showed marginally consistent to fully consistent improvement at most forecast hours. At early lead times, the skill and bias of HFSB_GFDLTC are closest to that of HFSB_ThompTC,

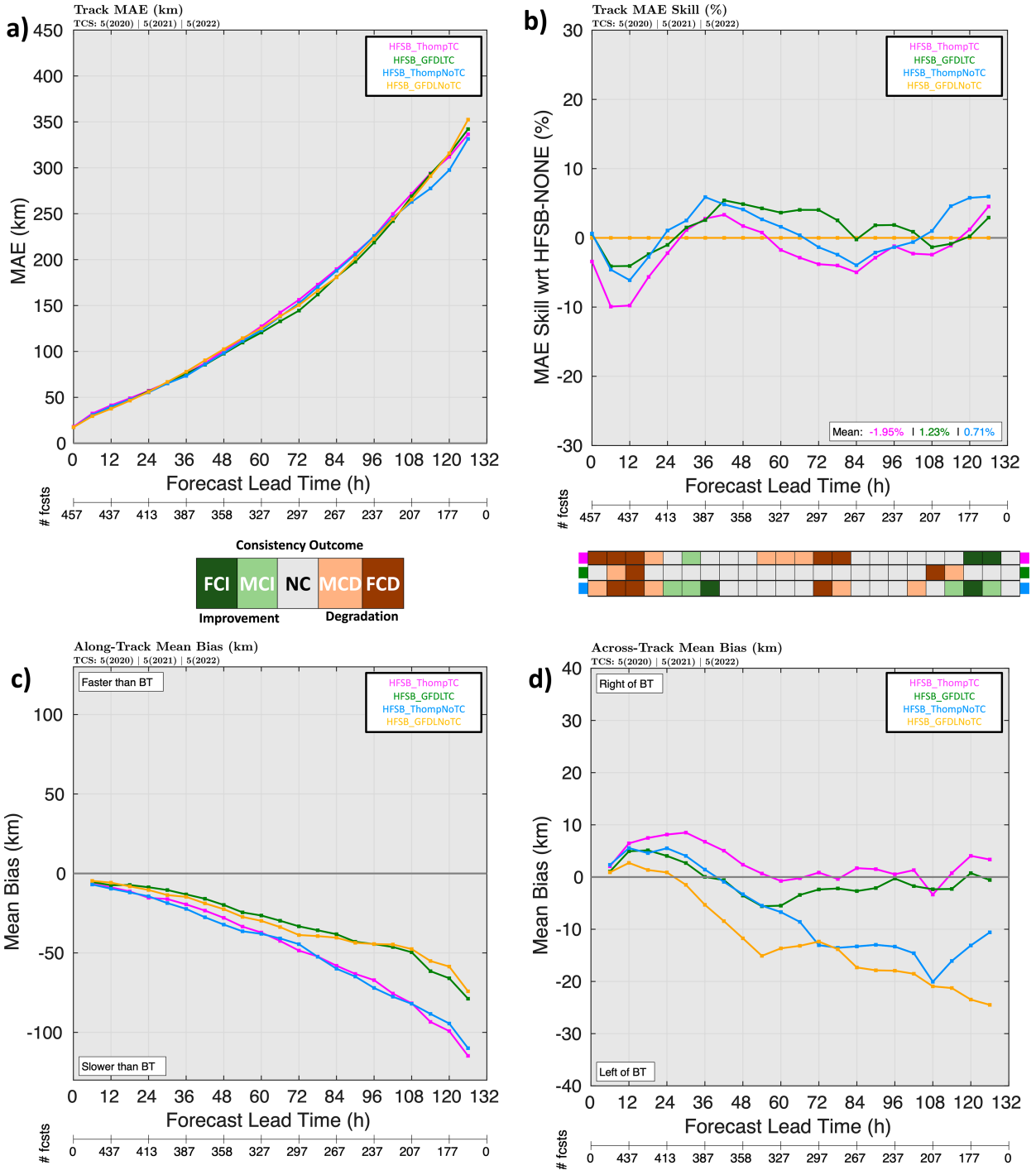

FIG. 2. (a) Mean absolute error (MAE; km) for TC track for HFSB_ThompTC (magenta), HFSB_GFDLTC (green), HFSB_ThompNoTC (blue), and HFSB_GFDLNoTC (yellow). (b) Track skill relative to HFSB_GFDLNoTC. This also shows the consistency metric for the track forecast, ranging from dark green (fully consistent improvement) to dark brown (fully consistent degradation). (c) As in (a), but for along-track bias. (d) As in (a), but for across-track bias.

indicating the importance of the TC-specific PBL changes to predictions of RI. Some possible reasons for this will be explored in a later section, and these results are consistent with the previous findings of Chen et al. (2022, 2023). At later lead times, the increased skill from HFSB_ThompTC and HFSB_ThompNoTC indicates the importance of the Thompson scheme to RI performance in the medium range.

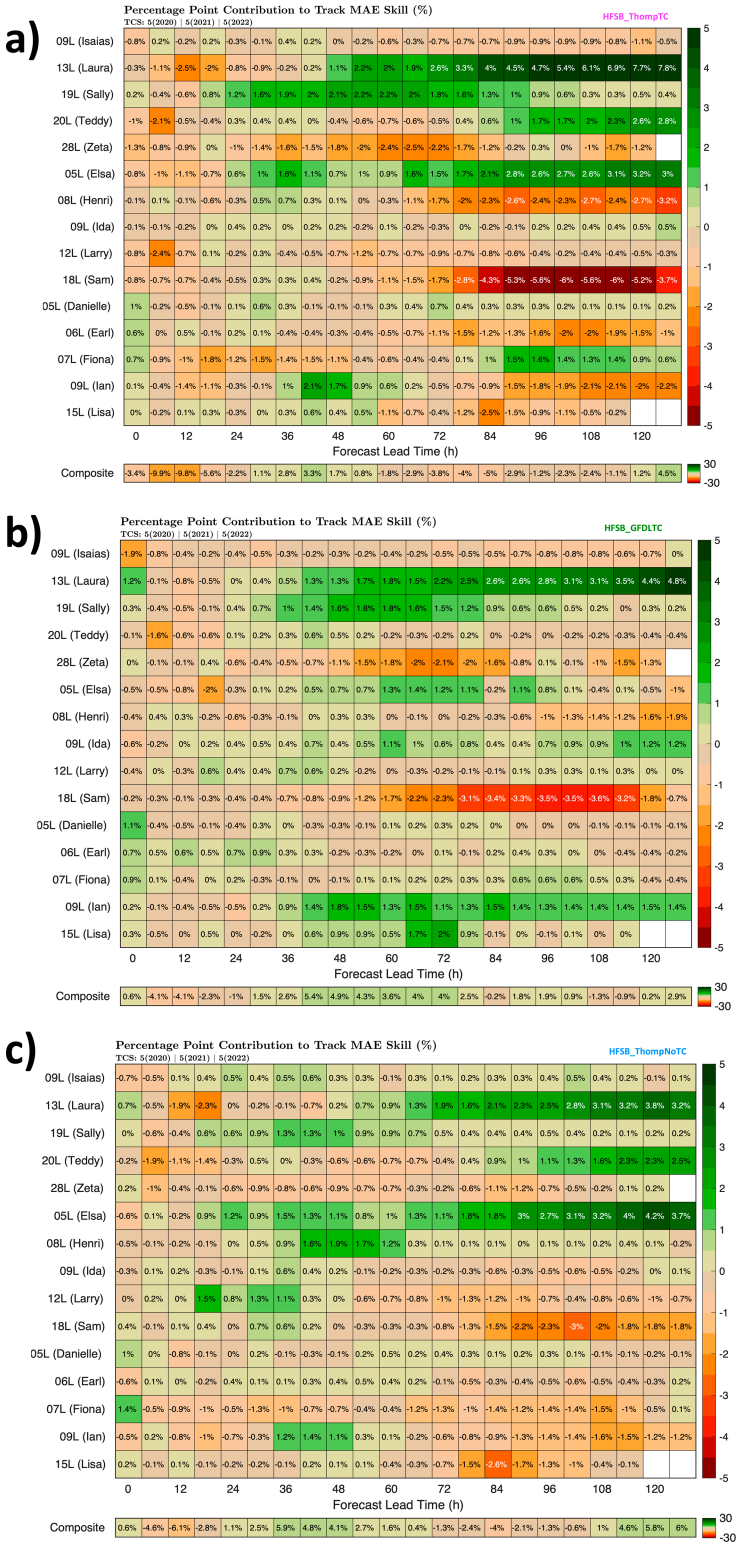

FIG. 3. (a) Percentage-point-contribution (PPC) of each TC at each forecast hour for track skill for HFSB_ThompTC relative to HFSB_GFDLNoTC. Green indicates improvement for HFSB_ThompTC relative to HFSB_GFDLNoTC, and brown indicates degradation. (b) As in (a), but for HFSB_GFDLTC. (c) As in (a), but for HFSB_ThompNoTC.

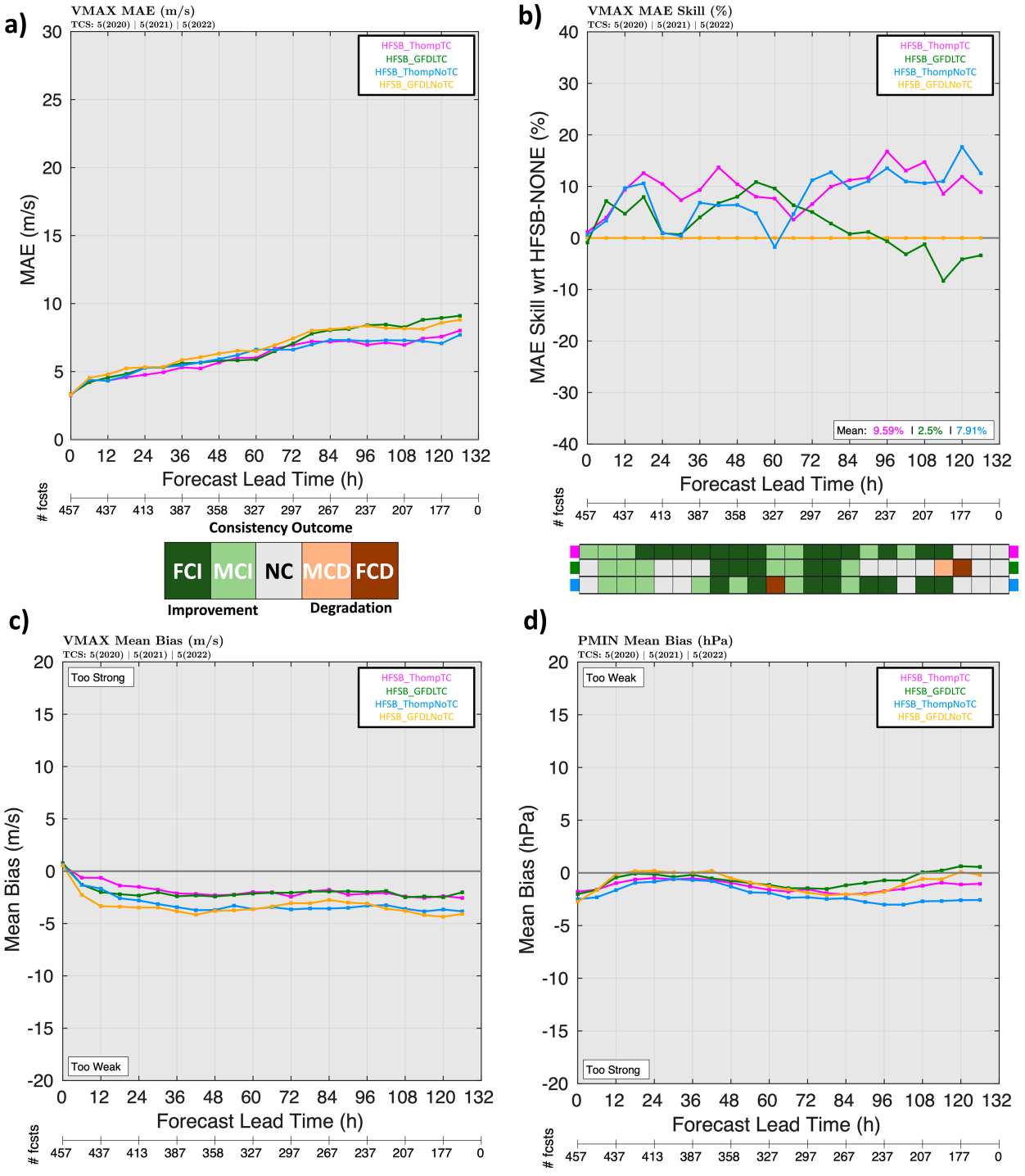

FIG. 4. As in Fig. 2, but for TC intensity (m s^{-1}).

Second, skill diagrams (Roebber 2009) were created for the 30 kt (24 h)$^{-1}$ RI threshold for each HAFS configuration (Fig. 6d). We also tested 25 kt (24 h)$^{-1}$ (Fig. 6e) and 35 kt (24 h)$^{-1}$ (Fig. 6f) as sensitivity experiments (1 kt ≈ 0.51 m s^{-1}). These skill diagrams incorporate RI forecasts from any 24-h period during the forecast. The skill diagrams highlight false alarm rate (FAR), probability of detection (POD), bias, and critical success index (CSI). The results from this analysis again confirms that HFSB_ThompTC had the best overall performance, and HFSB_GFDLTC was a close second (as in the standard verification). The comparison of HFSB_GFDLNoTC with HFSB_GFDLTC and HFSB_ThompNoTC with HFSB_ThompTC

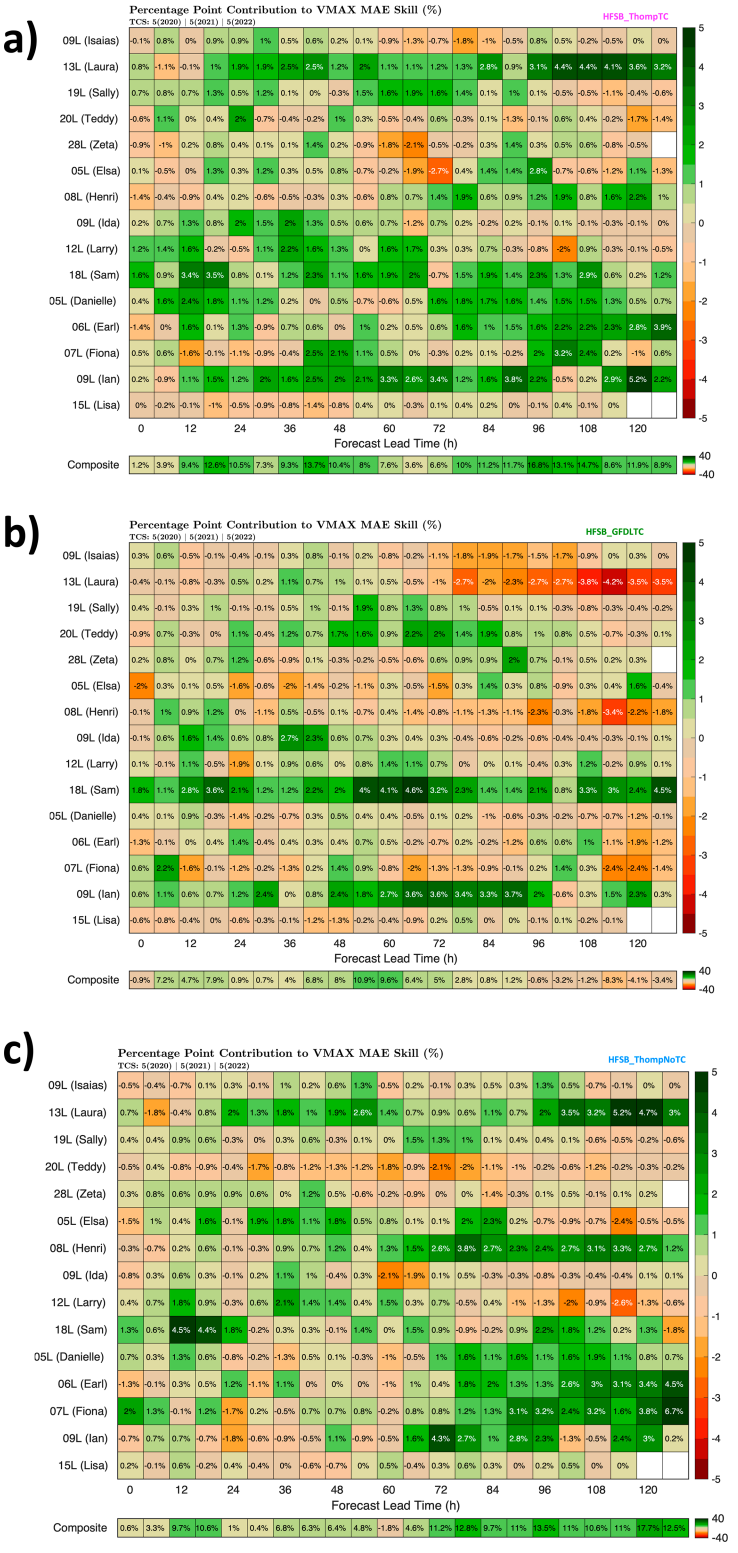

FIG. 5. As in Fig. 3, but for intensity skill. Green indicates improvement for HFSB_ThompTC relative to HFSB_GFDLNoTC, and brown indicates degradation. (b) As in (a), but for HFSB_GFDLTC. (c) As in (a), but for HFSB_ThompNoTC.

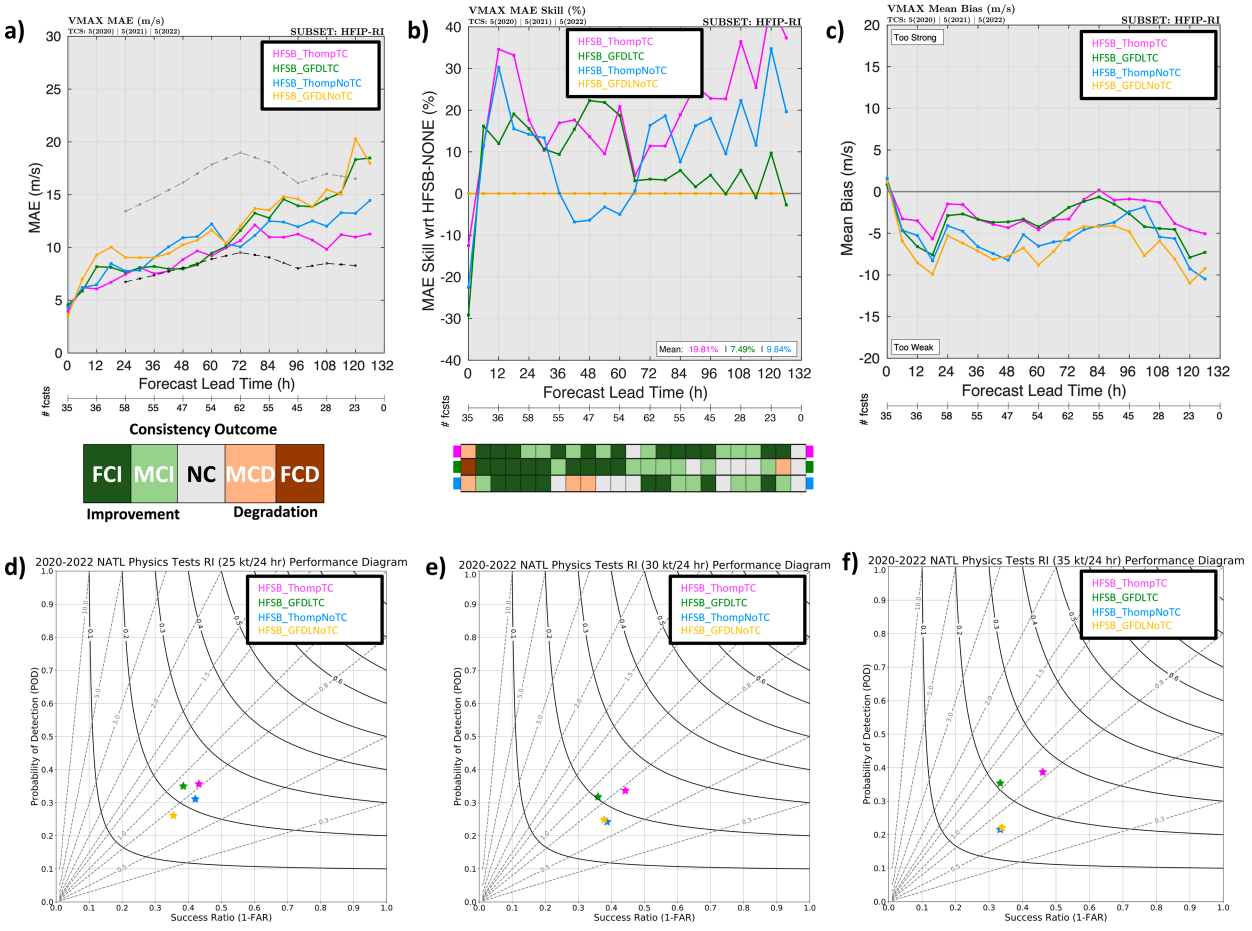

FIG. 6. (a)–(c) As in Fig. 2, but for TC intensity for the stratification of cases meeting the HFIP-RI threshold. The HFIP intensity error baseline is shown in the gray dashed line, and the HFIP goals are shown in the black dashed line. (d) Performance diagram for Atlantic basin RI skill [25 kt (24 h)$^{-1}$ threshold]. The diagram shows 1 − FAR on the *x* axis, the POD on the *y* axis, the bias in the diagonal lines, and the CSI in the curved lines. Better performance is up and to the right. (e) As in (d), but for the 30 kt (24 h)$^{-1}$ threshold. (f) As in (d), but for the 35 kt (24 h)$^{-1}$ threshold.

confirms the importance of the PBL-specific changes in RI detection. Comparing HFSB_GFDLTC with HFSB_ThompTC (lower FAR in the latter) indicates that the Thompson scheme may produce fewer RI false alarms than the GFDL microphysics, particularly at days 4–5 (based on the results in the verification). Overall, these findings confirm that the current configuration of HFSB_ThompTC is performing comparatively well for RI prediction in terms of currently available physics. Research to further improve model physics for RI prediction is ongoing.

3) RADII RESULTS

A verification variable that has gotten more attention in recent years is wind radius (e.g., Cangialosi and Landsea 2016). Size forecasts are important because TC size tends to be a factor in the size of a TC storm surge (e.g., Irish et al. 2008). In addition, the initial impact of tropical storm force winds (34 kt or 17.5 m s^{-1}) is often an important threshold for evacuations and other preparation measures. Thus, accurate forecasts of TC size are increasingly important for prediction of TC hazards. Figure 7 shows verifications of four standard TC size metrics, calculated using the updated GFDL tracker: radius of 34-kt (R34), 50-kt (R50), 64-kt (R64), and maximum (RMW) winds, all verified against best track. Best track tends to have more accurate wind radii for cases with aircraft reconnaissance data, which was the case for most of the TCs in this dataset (although prior to 2021, radii in the best track data were based on real-time analyses). The radii are averaged across four quadrants (southeast, southwest, northwest, northeast), and if a given radius does not exist in a certain quadrant, it is excluded from the average. R34 in the best track data are given to the nearest 10 n mi (1 n mi = 1.852 km), while R50, R64, and RMW are provided to the nearest 5 n mi. This degree of precision is important for interpreting small differences in verification between the different experiments.

The size bias is shown in each plot (the bias showed more notable and useful differences than the mean absolute errors). It should be noted that the sample sizes shown in Fig. 7 include the total number of cases from each quadrant

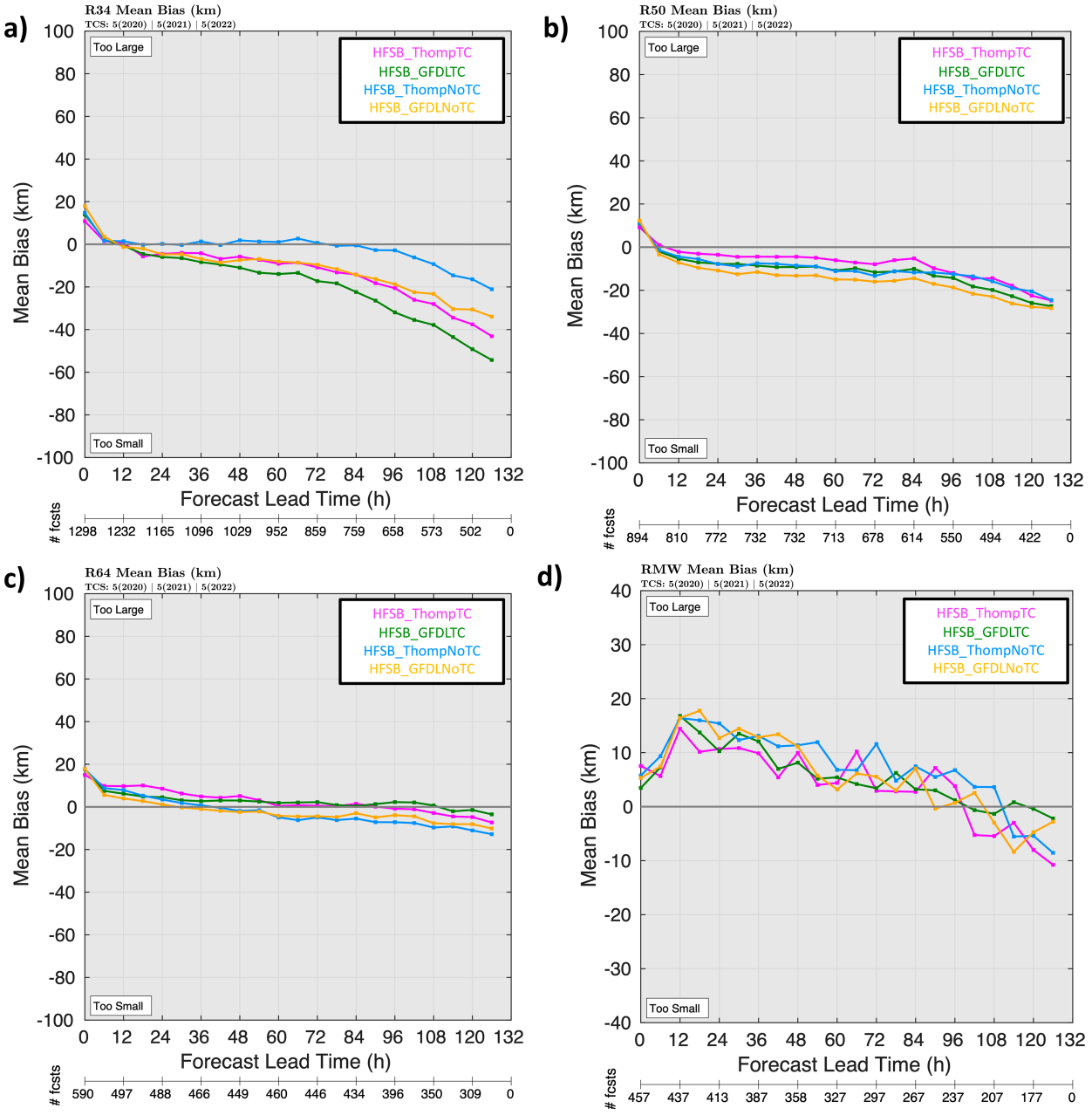

FIG. 7. (a) Mean 34-kt wind radii bias (R34; km) for HFSB_ThompTC (magenta), HFSB_GFDLTC (green), HFSB_ThompNoTC (blue), and HFSB_GFDLNoTC (yellow). (b) As in (a), but for 50-kt wind radii (R50). (c) As in (a), but for 64-kt wind radii (R64). (d) As in (a), but for the radius of maximum wind (RMW).

(i.e., treating each of the four TC quadrants separately). The largest differences are seen in R34 (Fig. 7a), where HFSB_ThompTC has a large negative (i.e., small) bias at longer lead times. HFSB_GFDLTC is even more pronounced, indicating that the TC-specific PBL changes, while leading to notable improvements in intensity bias and improvement of RI forecasts, led to some degradation in forecasts of R34. HFSB_ThompTC was better than HFSB_GFDLTC, and HFSB_ThompNoTC was better than HFSB_GFDLNoTC, which indicates that R34 forecasts are positively impacted by the Thompson Microphysics. The results are somewhat different for other radii. For R50 (Fig. 7b), HFSB_ThompTC has the smallest bias, while HFSB_ThompTC and HFSB_GFDLTC have the smallest biases for R64 (Fig. 7c). While the RMW bias is generally positive overall, RMW appears to be slightly better in HFSB_ThompTC and HFSB_GFDLTC than HFSB_ThompNoTC and HFSB_GFDLNoTC (Fig. 7d). This indicates that while the PBL changes negatively impact R34, they lead to increased prediction skill of

inner core radii. To better forecast TC size and structure, work is ongoing to evaluate and modify the model physics, including further improving the mass-flux parameterizations in high-wind conditions in tc-pbl and examining the connection with other physics schemes (e.g., the convective parameterization). It is also worth noting that, in general, the R34, R50, R64 biases are smaller than the observational uncertainty in operational estimates of wind radii even with aircraft data (30, 25, and 15 n mi for R34, R50, and R64, respectively, per Landsea and Franklin 2013), so small differences between model configurations may not be significant in light of this uncertainty.

b. *Composite results from each experiment*

To examine the physical reasons for the intensity and structure differences seen in the verification statistics, we examine composites of several key variables for each experiment, as well as the differences between the experiments. Before compositing, the data were first azimuthally averaged, and the radial coordinate was normalized by the 2-km radius of maximum wind (RMW), where $r^* = r/\mathrm{RMW}_{2\mathrm{km}}$. Similar techniques have been used extensively in model and observational composites of TC structure (e.g., Rogers et al. 2013; Zhang et al. 2015; Hazelton et al. 2022b), and prevent structures from being smoothed out solely due to TC size differences. The size differences are important, of course, and the structure distributions will also be examined in a later section. For these composites, only cases of hurricane intensity are included in order to ensure that the inner core is developed enough to make the RMW-relative compositing useful (the rest of the results, however, included the times when the storms were below hurricane intensity as well). We also removed cases where the distance to land was less than 5 times the 2-km RMW to filter out cases where land interactions were biasing the results. These filters trim the dataset to about 1/3 of the original size, but it still covers over 3000 separate 6-h forecast increments for each configuration. This analysis considers all forecast hours together. In the future, it might be worthwhile to examine how composites evolve with lead time in the forecast.

Figures 8a–d shows the composite tangential wind for each experiment. The overall structure pattern does not show many differences, but the difference plots (Figs. 8e–j) unveil some subtle ones. HFSB_ThompTC-HFSB_GFDLTC (Fig. 8e) and HFSB_ThompNoTC-HFSB_GFDLNoTC (Fig. 8j) differences indicate that the Thompson microphysics leads to a more vertically developed cyclonic circulation, but the tangential winds are weaker near and just inward of the RMW at lower levels. The two runs with the Thompson microphysics scheme also seem to have a slightly stronger anticyclone aloft at $r^* = 4$–5 (seen by comparing HFSB_ThompTC with HFSB_GFDLTC and HFSB_ThompNoTC with HFSB_GFDLNoTC). HFSB_ThompTC-HFSB_ThompNoTC (Fig. 8f) and HFSB_GFDLTC-HFSB_GFDLNoTC (Fig. 8i) show a subtly stronger PBL tangential wind just inward of $r^* = 1$, indicating that the TC-specific PBL changes are leading to spin up of the low-level vortex. The HFSB_ThompTC-HFSB_GFDLNoTC (Fig. 8g) differences, which combine the PBL and microphysics effects on the primary circulation, indicates that the combined changes lead to a more vertically developed and narrower cyclonic circulation, which is consistent with the intensity bias results and also the radii results seen in the verification results above. The HFSB_ThompNoTC-HFSB_GFDLNoTC results (Fig. 8j) show stronger tangential winds in the eye (and at $r^* = 3$–4) with the Thompson microphysics, indicating differences in the radial structure of the vortex. This will be further quantified later with some of TC structure metrics.

Figure 9 shows the radial wind composites for each experiment. The two experiments with the tc-pbl option (HFSB_ThompTC and HFSB_GFDLTC. Figs. 9a,b) again feature stronger PBL inflow in the lowest 1 km for $r^* = 1$–3, as well as stronger outflow at the top of the PBL, which has been shown to be associated with supergradient flow (e.g., Kepert and Wang 2001). Both of these structures seem to be indicative of larger inward transport of angular momentum inside the RMW, which also tends to promote spinup of the low-level vortex (e.g., Smith and Montgomery 2015). This also explains the smaller negative intensity bias in HFSB_ThompTC and HFSB_GFDLTC (Fig. 4c). The HFSB_ThompTC-HFSB_GFDLTC (Fig. 9e), HFSB_GFDLTC-HFSB_ThompNoTC (Fig. 9h) and HFSB_ThompNoTC-HFSB_GFDLNoTC (Fig. 9j) differences again show that the Thompson Microphysics seem to promote a deeper vortex with outflow at a higher altitude (the outflow differences also seem consistent with the slightly stronger anticyclone at $r^* = 4$–5 in the runs with Thompson microphysics). Yet, low-level inflow near the RMW is stronger with the GFDL microphysics (see the positive values near $r^* = 1$ in the difference plots in Figs. 9e,j).

Another composite shown is reflectivity (Fig. 10). The main difference in the composites is the shallower echo tops in the runs with Thompson Microphysics, compared to those with the GFDL microphysics. This is probably at least partly due to the excessive snow bias above the freezing level in the Thompson scheme (e.g., Wu et al. 2021), which was also verified in a representative case study of Hurricane Larry (2021, not shown). This pattern is more clearly seen and is very pronounced in the difference plots, with HFSB_ThompTC-HFSB_GFDLTC (Fig. 10e) and HFSB_ThompNoTC-HFSB_GFDLNoTC (Fig. 10j) showing very similar patterns with 5–10 dBZ lower reflectivity, on average, near $r^* = 1.25$ around $z = 10$ km. Interestingly, the Thompson runs (HFSB_ThompTC and HFSB_ThompNoTC) tended to have higher reflectivity in the outer regions of the TC ($r^* = 2$–5). Note that the effects of tc-pbl on reflectivity are not quite as pronounced in magnitude as those of the microphysics (not surprisingly), but still produce a notable effect on the distribution of convection and precipitation in the simulated TCs. The HFSB_ThompTC-HFSB_ThompNoTC (Fig. 10f) and HFSB_GFDLTC-HFSB_GFDLNoTC (Fig. 10i) differences show similar spatial patterns, with stronger eyewall convection (as noted by the higher dBZ values sloping outward for $r^* = 1$–2) and weaker outer core reflectivity. The eye has lower reflectivity in the runs with the PBL modifications, suggesting stronger compensating subsidence and a stronger TC. Additionally, The HFSB_GFDLTC-HFSB_GFDLNoTC (Fig. 10i) differences are more pronounced aloft than the HFSB_ThompTC-HFSB_ThompNoTC (Fig. 10f) differences. This

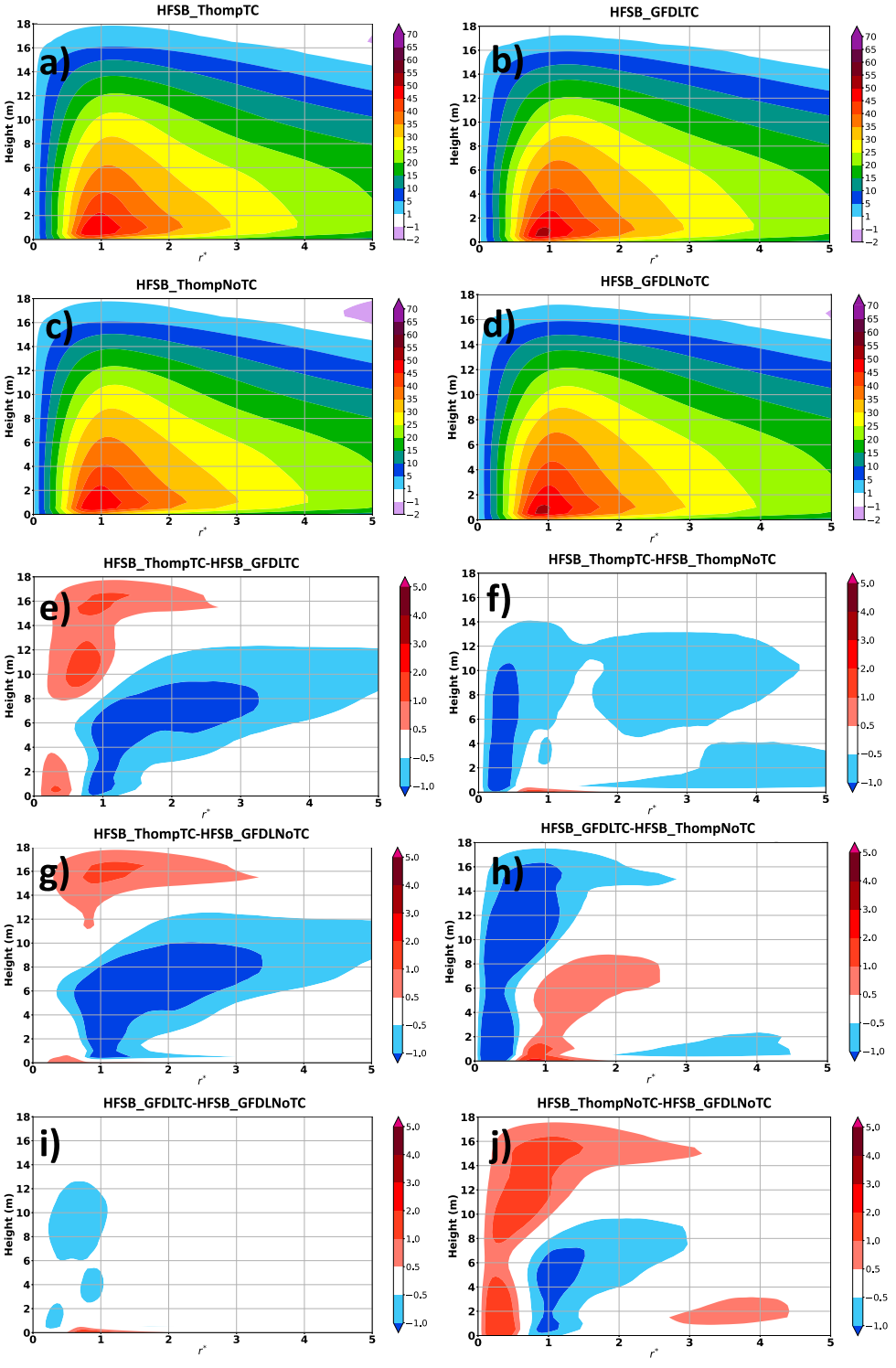

FIG. 8. (a) Composite tangential wind (m s^{-1}) for HFSB_ThompTC. The radial coordinate is normalized by the 2-km RMW. (b) As in (a), but for HFSB_GFDLTC. (c) As in (a), but for HFSB_ThompNoTC. (d) As in (a), but for HFSB_GFDLNoTC. (e) Difference in composite tangential wind between HFSB_ThompTC and HFSB_GFDLTC (HFSB_ThompTC − HFSB_GFDLTC). (f) As in (e), but for HFSB_ThompTC − HFSB_ThompNoTC. (g) As in (e), but for HFSB_ThompTC − HFSB_GFDLNoTC. (h) As in (e), but for HFSB_GFDLTC − HFSB_ThompNoTC. (i) As in (e), but for HFSB_GFDLTC − HFSB_GFDLNoTC. (j) As in (e), but for HFSB_ThompNoTC − HFSB_GFDLNoTC.

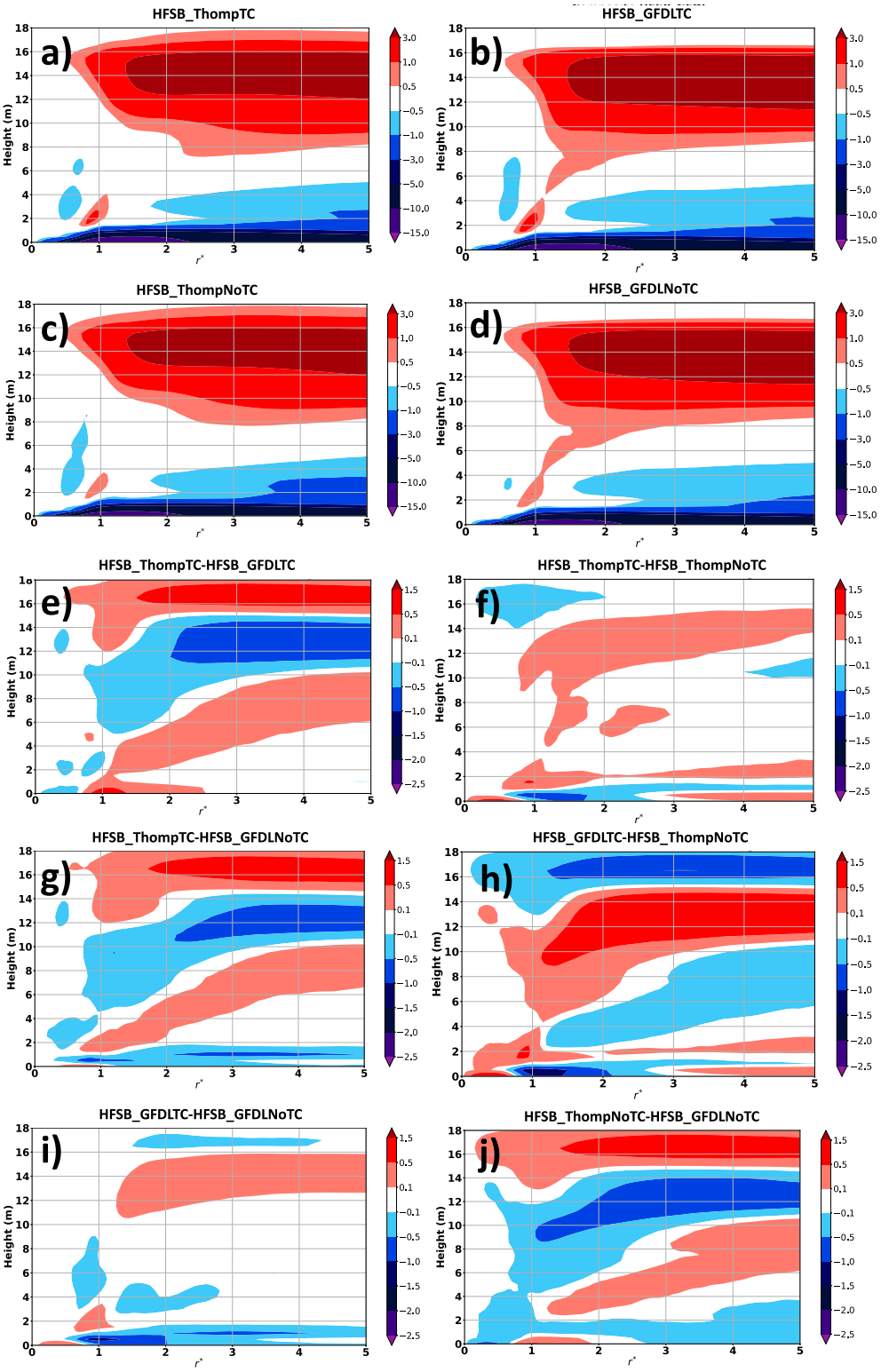

FIG. 9. As in Fig. 8, but for composite radial wind (m s^{-1}).

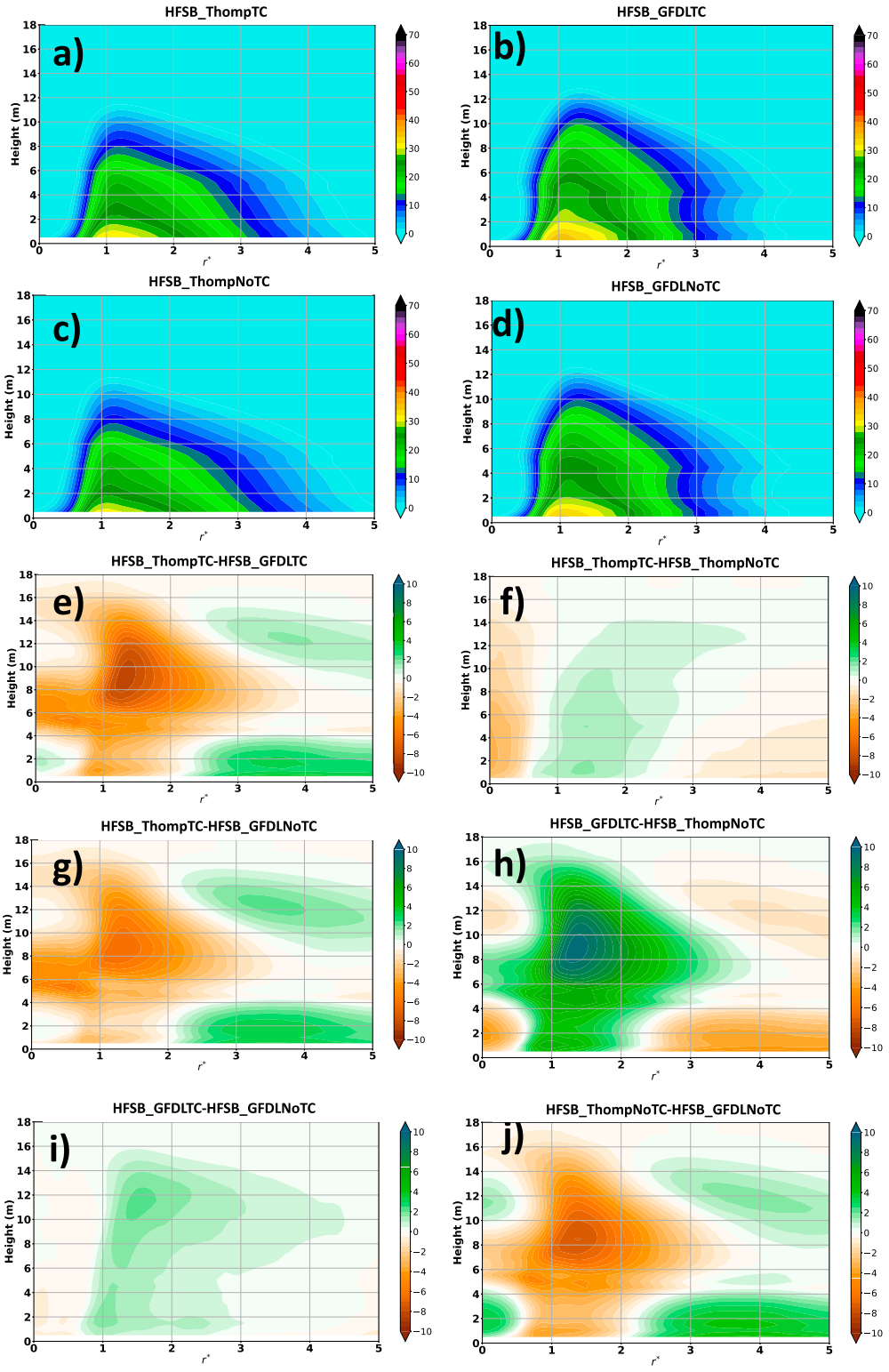

FIG. 10. As in Fig. 8, but for composite reflectivity (dBZ).

combination of differences in the radial and vertical distribution of precipitation illustrates the complex interplay between different physics parameterizations in TCs.

The final azimuthally averaged composite shown is the vertical velocity distribution (Fig. 11). The overall distribution of upward motion is similar in all of the composites (Figs. 11a–d). However, there are some notable differences that appear and are even clearer in the difference plots (Figs. 11e–j). Consistent with the reflectivity, the eyewall updraft appears to be more narrowly focused in HFSB_ThompTC compared to HFSB_ThompNoTC (Fig. 11f) and HFSB_GFDLTC compared to HFSB_GFDLNoTC (Fig. 11i), indicating stronger inner-core convection and a stronger secondary circulation with tc-pbl. There are also some clear differences between Thompson and GFDLMP (as seen from HFSB_ThompTC-HFSB_GFDLTC and HFSB_ThompNoTC-HFSB_GFDLNoTC; Figs. 11e,j). Updraft vertical velocities appear to be larger overall with the GFDL microphysics, and also spread out over a larger area (see the differences around $r^* = 2$–3). This may indicate that outer rainband activity is slightly more pronounced with the GFDL microphysics. Another possibility is that there are differences in the vertical velocity at outer radii due to differences in cloud–radiation feedback (e.g., Bu et al. 2014) with the two very distinct microphysics schemes. Specifically, the greater vertical velocity at outer radii in the Thompson scheme is consistent with the vertical circulation induced by in-cloud warming in that study. Further exploration of this point in a future study would be valuable.

One other comparison seeks to examine more details of the precipitation differences between the GFDL and Thompson microphysics, since some of those details may have been smoothed out in the azimuthally averaged composites. To accomplish this, we construct contoured frequency by altitude diagrams (CFADs; Yuter and Houze 1995) of reflectivity for HFSB_ThompTC and HFSB_GFDLTC (Fig. 12) to further illustrate the differences. The differences seen in the composites are even more stark in this visualization. GFDL microphysics has a wider range of reflectivities aloft, consistent with more graupel in the region above the melting level (not shown). The area where the Thompson microphysics has a larger distribution is around 5–6 km (approximately the melting level), indicating a stronger bright band (perhaps due to melting snow). Due to these larger reflectivities at the top of the melting level, the overall composite reflectivity values peak at a larger magnitude in the Thompson microphysics (as high as 50+ dBZ) than in GFDL. While a detailed exploration of the full differences between these schemes is beyond the scope of the paper, it is clear that they lead to differences in TC structure. A follow-up paper will examine some of these processes in more detail in several individual cases.

c. TC structure metrics

To further assess the structural differences between the different physics sensitivity experiments, we next calculate several structure metrics for each set of cases. These metrics are chosen to provide vertical and horizontal examinations of both kinematic and thermodynamic structure in the core, and have been shown in past studies (Zhang et al. 2015; Hazelton et al. 2022b) to be useful for examining differences in model behavior due to different physics. The specific metrics are as follows:

1) maximum wind speed;
2) 2-km radius of maximum winds (RMW);
3) dynamic vortex depth, defined as the height at which the tangential wind along the RMW decays to 40% of the value at $z = 2$ km (Hazelton et al. 2018; DesRosiers et al. 2023);
4) static vortex depth, defined as the height at which the tangential wind along the RMW decays to a value of 24 m s^{-1} (DesRosiers et al. 2023);
5) "alpha parameter" describing the rate of radial decay of the tangential wind between the RMW and 3RMW (Mallen et al. 2005);
6) RMW slope based on a linear best fit (Stern and Nolan 2009; Stern et al. 2014);
7) local Rossby number (Fig. 11f, Chen et al. 2018);
8) warm-core magnitude, defined as the maximum temperature difference between the inner 15-km and the outer 200–300-km ring. (Zhang et al. 2015); and
9) height of the maximum warm core.

Figure 13 shows the frequency distribution of each structure metric for each experiment. The intensity distributions (Fig. 13a) are consistent with the intensity biases: HFSB_ThompTC has the largest right (stronger) tail, and HFSB_GFDLTC has a notably higher peak in the distribution near 110–115 kt than the other experiments. The RMW distributions (Fig. 13b) are generally similar overall, but HFSB_ThompTC and HFSB_ThompNoTC are more bimodal than HFSB_GFDLTC and HFSB_GFDLNoTC. This may indicate more of a tendency for eyewall replacement with the Thompson microphysics, but this would need to be quantified further. Overall, dynamic (Fig. 13c) and static (Fig. 13d) vortex depth are generally similar between all of the experiments, although HFSB_ThompTC and HFSB_ThompNoTC are slightly taller, consistent with the composite radial structure noted previously. The RMW slope (Fig. 13e) is generally larger in HFSB_GFDLTC and HFSB_GFDLNoTC. Vortex Rossby Number (Fig. 13f) is slightly larger overall in HFSB_GFDLTC and HFSB_ThompNoTC, consistent with the intensity distributions and highlighting the importance of the modified PBL physics in leading to stronger/narrower systems with improved RI forecast skill. Some of the more notable differences in structure are the magnitude (Fig. 13g) and height (Fig. 13h) of the warm-core anomaly. This is driven by the microphysics, as HFSB_ThompTC and HFSB_ThompNoTC have a distribution with multiple maxima at both lower and very high upper levels while HFSB_GFDLTC and HFSB_GFDLNoTC have a single peak in the mid- to upper levels. The lower maximum is consistent with the findings of Stern and Nolan (2012), who also found some sensitivity of the warm core height and structure to using different microphysics parameterizations in their WRF simulations. The range of maximum warm core heights is much larger in the runs with Thompson microphysics than those with GFDL Microphysics. Whether this is due to differences in eyewall latent heating or the compensating

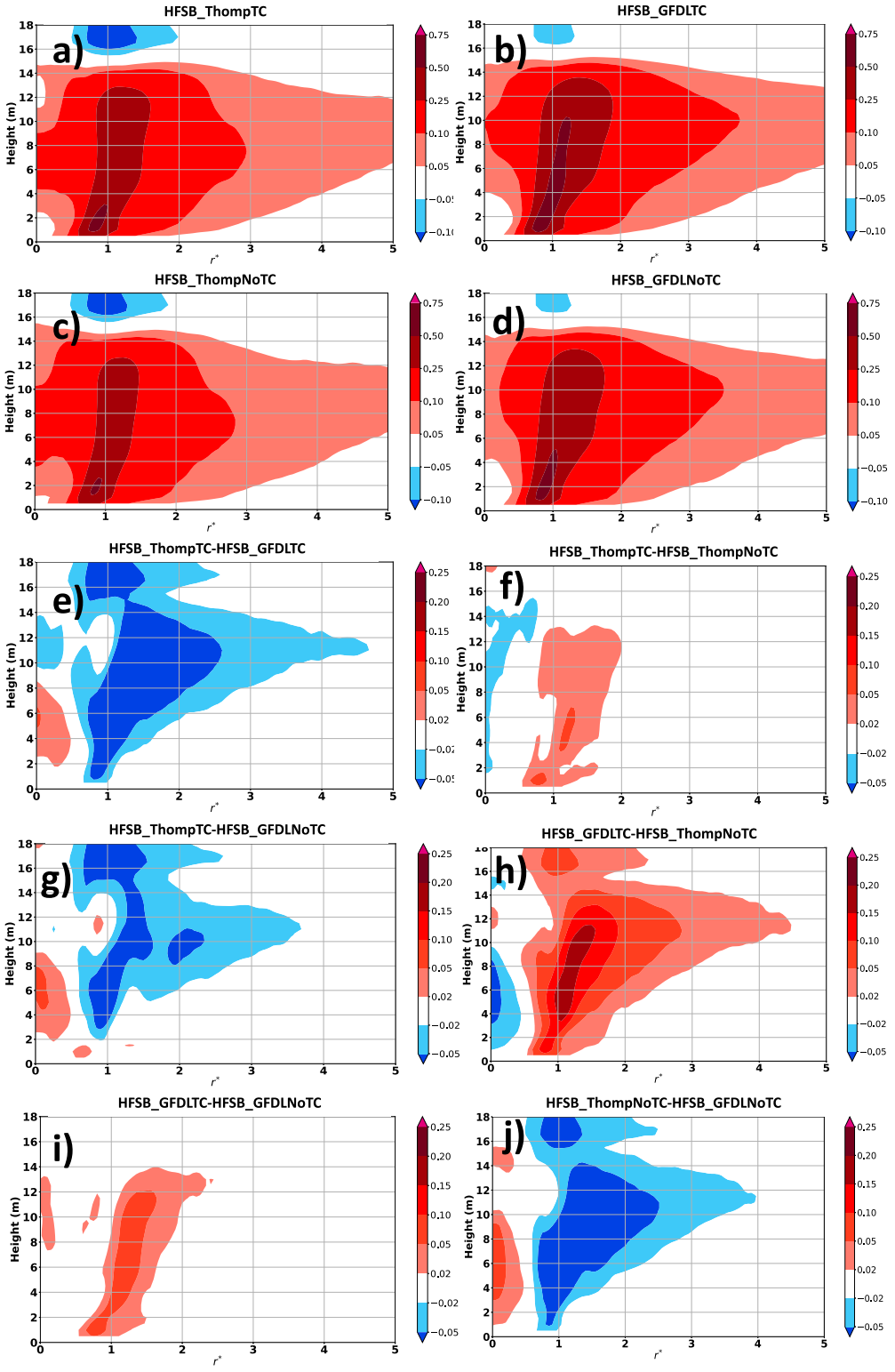

FIG. 11. As in Fig. 8, but for composite vertical velocity (m s^{-1}).

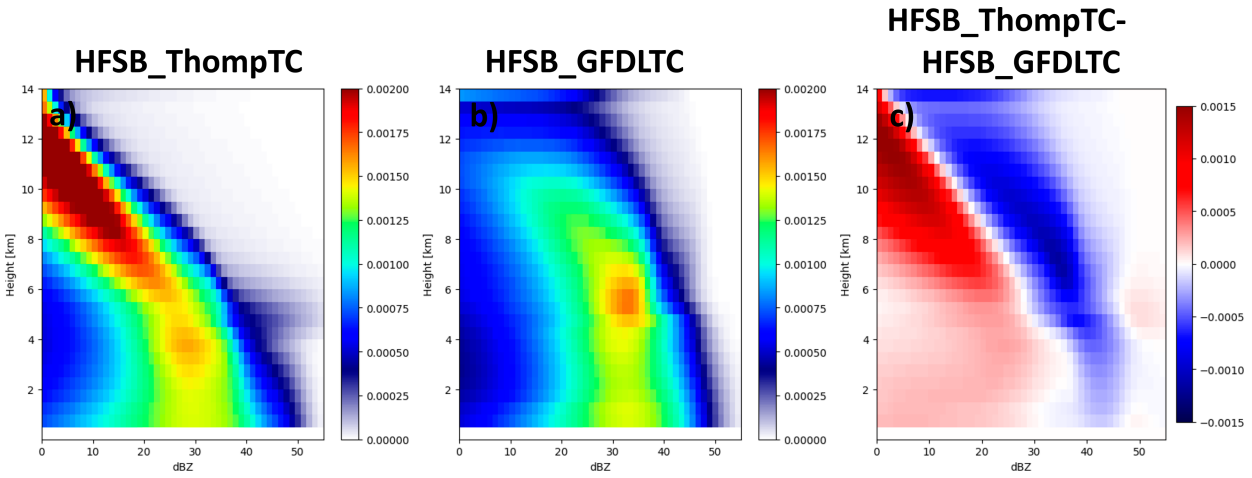

FIG. 12. (a) Contoured frequency by altitude (CFAD) diagram of reflectivity (dBZ) for the HFSB_ThompTC runs. The numbers indicate the frequency, wherein 0.01 = 1%. (b) As in (a), but for HFSB_GFDLTC. (c) Difference in the CFADS between HFSB_ThompTC and HFSB_GFDLTC.

subsidence in the eye is an interesting question for future work. The alpha differences (Fig. 13i) are generally small between the different groups, although the runs with GFDL microphysics (HFSB_GFDLNoTC and HFSB_GFDLTC) seem to have slightly narrower vortices (larger alpha), consistent with the composite tangential wind differences shown earlier.

To further quantify the differences in structure between the different experiments, statistical tests were performed to determine which metrics were statistically different. To prevent artificial significance due to the large sample size (over 3500 total when compositing all forecast hours), we use a bootstrap resampling method to construct confidence intervals on the mean for each of the metrics. The results are shown in Table 3 along with marks indicating which differences are statistically significant.

For Vmax, HFSB_ThompTC is higher than HFSB_ThompNoTC/HFSB_GFDLNoTC and lower than HFSB_GFDLTC. HFSB_GFDLTC is also higher than HFSB_ThompNoTC and HFSB_GFDLNoTC. This confirms the importance of the TC-specific PBL modifications in leading to more robust intensification. For 2-km RMW, there are fewer significant differences, although HFSB_GFDLTC is lower than HFSB_ThompTC and HFSB_ThompNoTC. This indicates that Thompson microphysics might be leading to a slightly larger RMW. Somewhat surprisingly, few of the vortex depth differences were significant, although HFSB_ThompNoTC is larger than HFSB_GFDLTC and HFSB_GFDLNoTC, which shows that Thompson microphysics produces a taller storm (consistent with the radial wind composite). One interesting relationship is the fact that the RMW slope is generally larger in the runs with GFDL MP than the runs with Thompson MP, and the warm core is also greater in magnitude. This is consistent with the connection between eyewall slope and the magnitude of the warm core in thermal wind balance in TCs described in the Sawyer–Eliassen axisymmetric model (e.g., Shapiro and Willoughby 1982), wherein a stronger warm core anomaly tends to be associated with a larger eyewall slope. The differences in eyewall slope and warm core structure, along with the differences in vortex depth, indicate that the different microphysics schemes are producing fundamental differences in balanced vortex structure that should be explored further. The warm core height is also notably lower in HFSB_ThompTC than in the sensitivity experiments, consistent with the distributions shown in Fig. 10. Somewhat surprisingly, few of the vortex Rossby number differences are significant, although HFSB_GFDLTC is larger than HFSB_GFDLNoTC, consistent with stronger/narrow vortices with the PBL modifications. Some of the alpha differences are also statistically significant, with GFDL microphysics producing a generally sharper vortex.

d. *Comparison with observations*

To further assess how realistic the predicted TC structure was for the various physics schemes, we next compared the composite forecast structure from each of the experiments with a composite of airborne radar observations. For this study, we used the TC-RADAR dataset (Fischer et al. 2022), which compiles the observations from the NOAA P3 Tail Doppler Radar (TDR) flights between 1997 and 2022. For the purposes of this analysis, we only use the years 2020–22, and specifically select the flights where there was a forecast from all four experiments, in order to have a homogeneous dataset that can be directly compared with the model data. In addition, we filter the HAFS data by removing points from individual cases where the model reflectivity is less than 0 dBZ, in order to have a reasonable approximation (in the model data) of the gaps that the TDR has where there are no scatterers. To be included in the composite (for the radar or model data), a given point (in normalized radius/height coordinates) has to be non-NaN (not a number) for at least 50% of cases, in order to minimize spurious data. For this composite, we selected 24-h forecasts where there was radar data available, to minimize differences due to track errors.

Figure 14a shows the composite tangential wind structure from the TDR data, and Figs. 14b–e show the difference field between the composite for this experiment. The overall

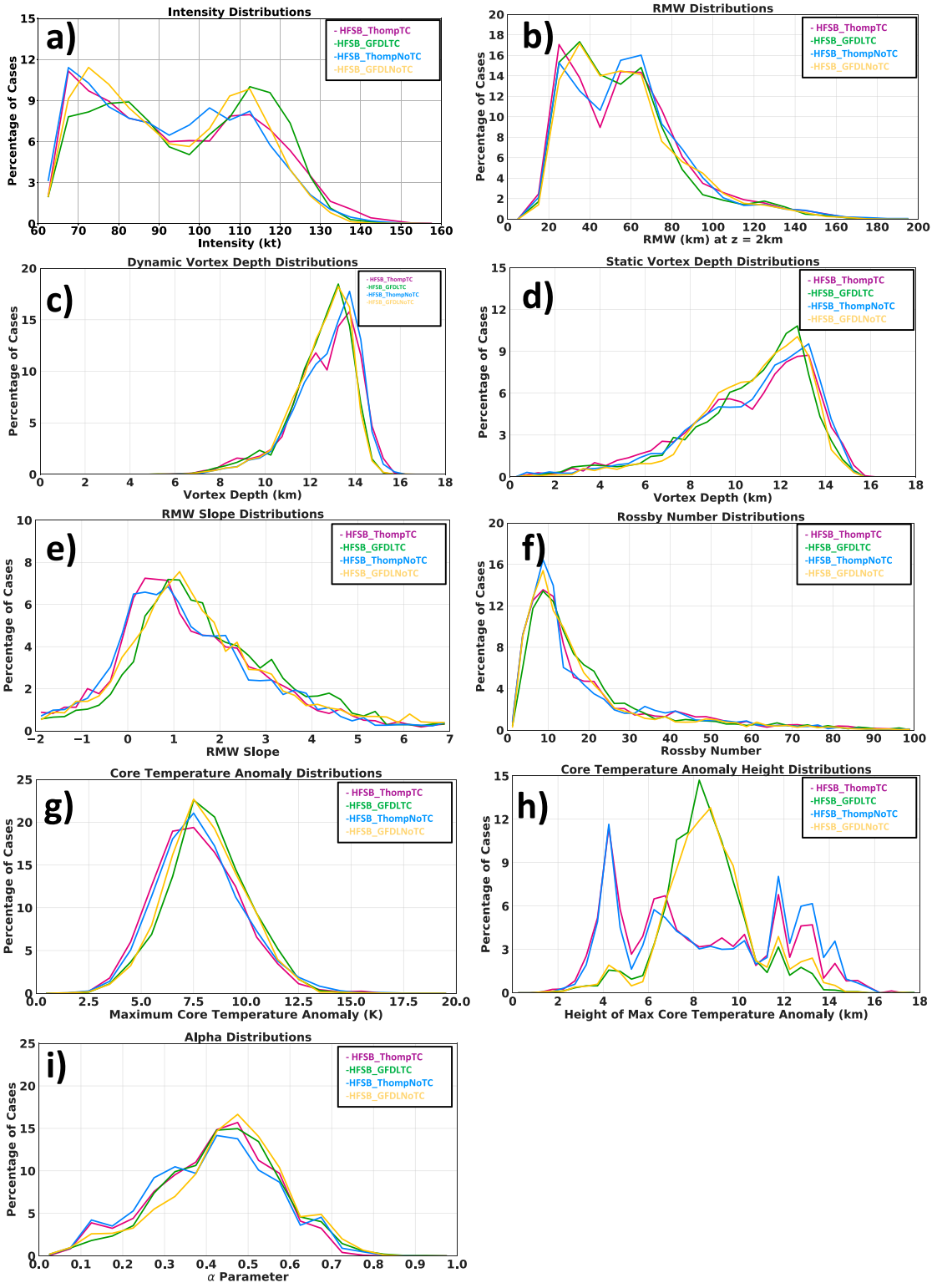

FIG. 13. Frequency distribution of structure metrics for each experiment: (a) maximum wind, (b) 2-km RMW, (c) dynamic vortex depth, (d) static vortex depth, (e) RMW slope, (f) vortex Rossby number, (g) core temperature anomaly, (h) height of maximum core temperature anomaly, and (i) alpha parameter.

TABLE 3. Means and confidence interval (based on a 1000-iteration bootstrap method) for each structure variable. Significant differences (at the 95% level) are shown with the "X."

	Vmax	2-km RMW	Dynamic vortex depth	Static vortex depth	RMW slope	Alpha	Vortex Rossby number	Core temp anomaly	Core temp anomaly height
Mean HFSB_ThompTC	93.9	57.5	12.4	10.3	1.08	0.42	20.4	7.7	7.8
Mean HFSB_GFDLTC	95.8	55	12.2	10.4	1.34	0.44	20.1	8.2	8.3
Mean HFSB_ThompNoTC	91.8	58.1	12.5	10.5	1.01	0.41	19.5	7.8	8.4
Mean HFSB_GFDLNoTC	92.6	56.4	12.2	10.6	1.28	0.45	18.6	8	8.5
Lower CI HFSB_ThompTC	93.3	56.6	12.3	10.2	0.98	0.41	19.8	7.6	7.7
Higher CI HFSB_ThompTC	94.6	58.5	12.4	10.4	1.17	0.42	21	7.7	8
Lower CI HFSB_GFDLTC	95.1	54.1	12.2	10.3	1.23	0.43	19.5	8.1	8.2
Higher CI HFSB_GFDLTC	96.4	55.9	12.3	10.5	1.44	0.44	20.6	8.2	8.3
Lower CI HFSB_ThompNoTC	91.2	57.2	12.4	10.4	0.91	0.41	18.9	7.7	8.3
Higher CI HFSB_ThompNoTC	92.5	59.1	12.5	10.6	1.12	0.42	20.1	7.9	8.5
Lower CI HFSB_GFDLNoTC	92	55.4	12.2	10.5	1.19	0.44	18	8	8.4
Higher CI HFSB_GFDLNoTC	93.2	57.3	12.3	10.6	1.38	0.45	19.1	8.1	8.5
HFSB_ThompTC/HFSB_GFDLTC	X	X			X	X		X	X
HFSB_ThompTC/HFSB_ThompNoTC	X								X
HFSB_ThompTC/HFSB_GFDLNoTC	X			X	X	X		X	X
HFSB_GFDLTC/HFSB_ThompNoTC	X	X	X		X	X		X	
HFSB_GFDLTC/HFSB_GFDLNoTC	X						X		X
HFSB_ThompNoTC/HFSB_GFDLNoTC			X		X	X		X	

difference patterns are fairly similar. All four experiments seem to show a vortex that is too narrow, with wind stronger than the observed composite inside $r^* = 1$ and weaker than the observed composite from $r^* = 1$–3. This is consistent with the negative R34 bias discussed previously. This bias seems to be reduced somewhat in the GFDLTC (GFDL MP with tc-pbl) configuration. In general, the vortex seems to be too weak even near the eyewall for the GFDLNoTC configuration, which again highlights how the TC-specific PBL modifications serve to produce more realistic spinup of the tangential winds near the eyewall.

4. Discussion, conclusions, and future work

In this study, we quantitatively examine the relative impacts of microphysics and PBL physics differences on TC forecasts in a newly operational hurricane model HAFS. This analysis helps to quantify how model physics interact to produce changes in TC structure and evolution, and also motivates ongoing and future work to continue to make improvements to the model physics based on observational data. For the purposes of this evaluation, we used the HAFS-B 2020–22 retrospective forecasts and compared them with three sensitivity tests: one using GFDL microphysics instead of the Thompson microphysics, one turning off the TC-specific PBL modifications that were incorporated in the EDMF-TKE scheme, and one combining the PBL and microphysics modifications. This evaluation also helps us to see how the individual physics modifications are impacting the track, intensity, and structure forecasts in HAFS-B. This study focused on 15 TCs across the 2020–22 seasons that all reached at least hurricane intensity, which may slightly bias the results toward stronger TCs but maximizes the sample size for comparison between the different experiments.

Interestingly, all three sensitivity tests outperformed HFSB_ThompTC for track, especially the configurations with GFDL microphysics, indicating that Thompson microphysics might be negatively impacting some of the TC steering. This will be examined in future work. In contrast, none of the sensitivity experiments performed as well for intensity as HFSB_ThompTC. Further, removing tc-pbl led to a larger negative intensity bias and less ability to detect rapid intensification, confirming the importance of correctly modeling these PBL processes to be able to accurately forecast RI. At the same time, the Thompson microphysics led to fewer RI false alarms than the GFDL microphysics. Together, these results indicate that HFSB_ThompTC (most similar to the version scheduled for implementation) has the optimal physics configuration for RI prediction (among these four). Wind radii prediction, however, is an area where the physics needs to be further optimized. In particular, the tc-pbl changes lead to a negative bias in R34, and changes to the scheme to address this issue are ongoing.

Examining composites of structure and quantitative structure metrics provides more detail about how the model physics altered the TC evolution in HAFS. The Thompson microphysics have less precipitation aloft in the eyewall (as seen in the reflectivity composites). Thompson microphysics also leads to a TC warm core that peaks at lower levels compared to the GFDL scheme. The PBL changes led to stronger low-level inflow near and inside the RMW, and also tended to focus precipitation closer to or inside the radius of maximum winds. These differences are consistent with the enhanced radial inflow with the PBL modifications, leading to a stronger overall secondary circulation, which acts to 1) transport more angular momentum into the TC core and spin the storm up (e.g., Smith and Montgomery 2015; Chen and Bryan 2021), and 2) promote the development of

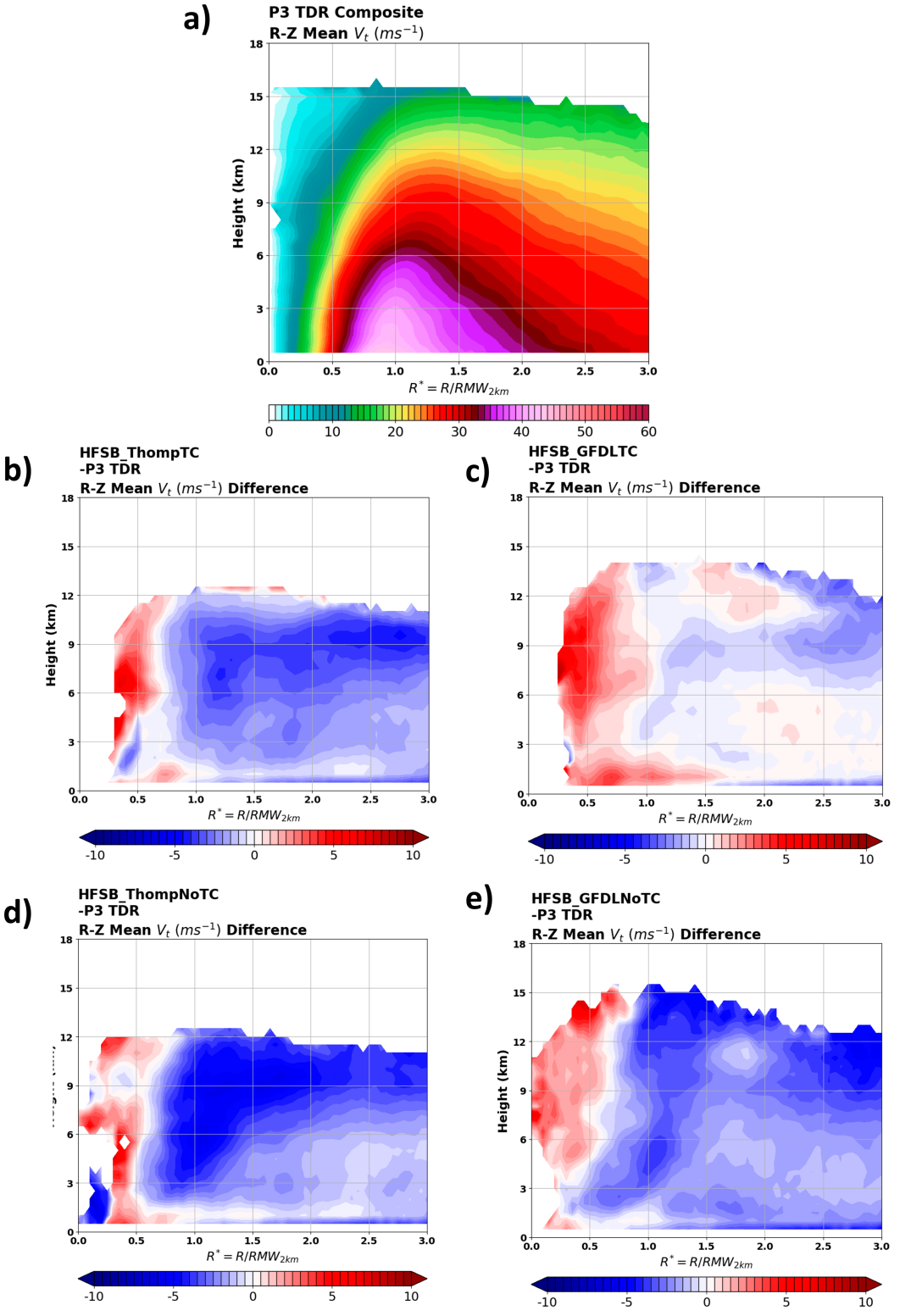

FIG. 14. (a) Composite tangential wind (m s^{-1}) from TC-RADAR from the TDR for cases where there was a 24-h forecast available from all four experiments. The radial coordinate is normalized by the 2-km RMW. (b) Difference between the HFSB_ThompTC experiment and TDR composites. (c) Difference between the HFSB_GFDLTC experiment and TDR composites. (d) Difference between the HFSB_ThompNoTC experiment and TDR composites. (e) Difference between the HFSB_GFDLNoTC experiment and TDR composites.

convection and convective bursts inside the RMW (e.g., Rogers et al. 2013), where diabatic heating is more efficient and can support continued pressure falls. Finally, comparison with observed TDR data from the TC-RADAR dataset provides three-dimensional structure validation of the schemes, highlighting, for example, how runs with tc-pbl produce more realistic tangential wind in the lower levels near the eyewall, but all configurations of HFSB seem to produce a vortex that is too narrow throughout the depth of the troposphere. Whether this is a consequence of missing structural features such as secondary eyewalls that tend to expand wind radii would be a useful subject of future investigation.

Several avenues of ongoing and future research seek to explore ways to evaluate and improve the physics parameterizations in HAFS. One is exploring how the modified PBL physics impact forecasts at different scales. This work should lead to better large-scale and track forecasts from HAFS, which will be particularly important if operational forecasts are extended to 7 days. Relatedly, we are exploring the unification of the PBL and convective scheme mass flux calculations to better predict mixing in the TC environment, which should improve wind radii forecasts. In general, we plan to continue to understand how the PBL physics impact TC structure (such as wind radii and various measures of vertical structure), beyond the "single number" that intensity represents. It would be particularly interesting to explore how the physics changes lead to differences in prediction of complex structural evolutions such as eyewall replacement cycles.

We are performing a comprehensive examination of the HAFS-B retrospective forecasts through comparison with airborne radar data, an analysis that will lead to further understanding of the structure biases in the model and inform physics adjustments to improve the model for TC prediction and research. In addition, we plan to examine how the Thompson microphysics impact the large-scale flow and steering near TCs. We also plan to conduct more detailed examination of the hydrometeors and latent heating profiles in a follow up case study using the different microphysics schemes. It will be useful to examine the details of how these differences lead to different outcomes in terms of both kinematic and thermodynamic inner-core structure. Finally, we are examining ways in which model physics can be varied (whether systematically as in this paper or stochastically) to generate an ensemble with realistic track, intensity, and structure spread for probabilistic prediction.

Acknowledgments. The authors thank the NOAA RDHPCS computer staff for maintaining the supercomputers used to perform these runs. We also thank the crews of the NOAA P3 aircraft for their tireless efforts to collect the observational data that was used to evaluate the model structure forecasts. Lead author Andrew Hazelton was supported by NOAA Grants NA19OAR0220187 and NA22OAR4050668D. The GRaphics for OS(s)Es and Other modeling applications on TCs (GROOT) verification package developed by Dr. Sarah Ditchek and funded by the Quantitative Observing System Assessment Program (QOSAP) and the FY18 Hurricane Supplemental (NOAA Award ID NA19OAR0220188) was used to generate graphics for this publication. This research was carried out in part under the auspices of the Cooperative Institute for Marine and Atmospheric Studies, a cooperative institute of the University of Miami and the National Oceanic and Atmospheric Administration (NOAA) Cooperative Agreement NA20OAR4320472.

Data availability statement. Raw model output from these HAFS runs (including the full retrospective data and the sensitivity experiments) is available on the NOAA RDHPCS archive system, or upon request. Sample HAFS data are available from EMC (https://www.emc.ncep.noaa.gov/gc_wmb/vxt/zack/HAFS_Sample_Files/HFSA_sample/hafs/v1.0/hfsa.20220920/00/). Graphics for retrospective forecasts are available on the AOML Hurricane Model Viewer (https://storm.aoml.noaa.gov/viewer/).

REFERENCES

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. *Mon. Wea. Rev.*, **135**, 3965–3989, https://doi.org/10.1175/2007MWR2032.1.

Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modell.*, **4**, 55–88, https://doi.org/10.1016/S1463-5003(01)00012-9.

Braun, S. A., and W.-K. Tao, 2000: Sensitivity of high-resolution simulations of Hurricane Bob (1991) to planetary boundary layer parameterizations. *Mon. Wea. Rev.*, **128**, 3941–3961, https://doi.org/10.1175/1520-0493(2000)129<3941:SOHRSO>2.0.CO;2.

Brown, B. R., M. M. Bell, and A. J. Frambach, 2016: Validation of simulated hurricane drop size distributions using polarimetric radar. *Geophys. Res. Lett.*, **43**, 910–917, https://doi.org/10.1002/2015GL067278.

Bu, Y. P., R. G. Fovell, and K. L. Corbosiero, 2014: Influence of cloud–radiative forcing on tropical cyclone structure. *J. Atmos. Sci.*, **71**, 1644–1662, https://doi.org/10.1175/JAS-D-13-0265.1.

Cangialosi, J. P., 2022: National Hurricane Center forecast verification report: 2021 Hurricane season. NHC Tech. Rep., 76 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2021.pdf.

——, 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, X., and G. H. Bryan, 2021: Role of advection of parameterized turbulence kinetic energy in idealized tropical cyclone simulations. *J. Atmos. Sci.*, **78**, 3559–3574, https://doi.org/10.1175/JAS-D-21-0088.1.

——, M. Xue, and J. Fang, 2018: Rapid intensification of Typhoon Mujigae (2015) under different sea surface temperatures: Structural changes leading to rapid intensification. *J. Atmos. Sci.*, **75**, 4313–4335, https://doi.org/10.1175/JAS-D-18-0017.1.

——, G. H. Bryan, J. A. Zhang, J. J. Cione, and F. D. Marks, 2021: A framework for simulating the tropical cyclone boundary layer using large-eddy simulation and its use in evaluating PBL parameterizations. *J. Atmos. Sci.*, **78**, 3559–3574, https://doi.org/10.1175/JAS-D-20-0227.1.

——, ——, A. Hazelton, F. D. Marks, and P. Fitzpatrick, 2022: Evaluation and improvement of a TKE-based eddy-diffusivity mass-flux (EDMF) planetary boundary layer scheme in hurricane conditions. *Wea. Forecasting*, **37**, 935–951, https://doi.org/10.1175/WAF-D-21-0168.1.

——, A. Hazelton, F. D. Marks, G. J. Alaka, and C. Zhang, 2023: Performance of an improved TKE-based eddy-diffusivity mass-flux (EDMF) PBL scheme in 2021 hurricane forecasts from the Hurricane Analysis and Forecast System. *Wea. Forecasting*, **38**, 321–336, https://doi.org/10.1175/WAF-D-22-0140.1.

DeMaria, M., J. L. Franklin, M. J. Onderlinde, and J. Kaplan, 2021: Operational forecasting of tropical cyclone rapid intensification at the National Hurricane Center. *Atmosphere*, **12**, 683, https://doi.org/10.3390/atmos12060683.

DesRosiers, A. J., M. M. Bell, P. J. Klotzbach, M. S. Fischer, and P. D. Reasor, 2023: Observed relationships between tropical cyclone vortex height, intensity, and intensification rate. *Geophys. Res. Lett.*, **50**, e2022GL101877, https://doi.org/10.1029/2022GL101877.

Ditchek, S. D., J. A. Sippel, P. J. Marinescu, and G. J. Alaka Jr., 2023: Improving best-track verification of tropical cyclones: A new metric to identify forecast consistency. *Wea. Forecasting*, **38**, 817–831, https://doi.org/10.1175/WAF-D-22-0168.1.

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.

Fischer, M. S., P. D. Reasor, R. F. Rogers, and J. F. Gamache, 2022: An analysis of tropical cyclone vortex and convective characteristics in relation to storm intensity using a novel airborne Doppler radar database. *Mon. Wea. Rev.*, **150**, 2255–2278, https://doi.org/10.1175/MWR-D-21-0223.1.

Fovell, R. G., and H. Su, 2007: Impact of cloud microphysics on hurricane track forecasts. *Geophys. Res. Lett.*, **34**, L24810, https://doi.org/10.1029/2007GL031723.

Gopalakrishnan, S. G., F. Marks, J. A. Zhang, X. Zhang, J.-W. Bao, and V. Tallapragada, 2013: A study of the impacts of vertical diffusion on the structure and intensity of the tropical cyclones using the high-resolution HWRF system. *J. Atmos. Sci.*, **70**, 524–541, https://doi.org/10.1175/JAS-D-11-0340.1.

——, 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.

——, 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.

Hazelton, A. T., L. Harris, and S.-J. Lin, 2018: 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.

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

——, and Coauthors, 2022a: Performance of 2020 real-time Atlantic hurricane forecasts from high-resolution global-nested hurricane models: HAFS-globalnest and GFDL T-SHiELD. *Wea. Forecasting*, **37**, 143–161, https://doi.org/10.1175/WAF-D-21-0102.1.

——, J. A. Zhang, and S. Gopalakrishnan, 2022b: Comparison of the performance of the observation-based hybrid EDMF and EDMF-TKE PBL schemes in 2020 tropical cyclone forecasts from the global-nested Hurricane Analysis and Forecast System. *Wea. Forecasting*, **37**, 457–476, https://doi.org/10.1175/WAF-D-21-0124.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Irish, J. L., D. T. Resio, and J. J. Ratcliff, 2008: The influence of storm size on hurricane surge. *J. Phys. Oceanogr.*, **38**, 2003–2013, https://doi.org/10.1175/2008JPO3727.1.

Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108, https://doi.org/10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2.

Kepert, J., and Y. Wang, 2001: The dynamics of boundary layer jets within the tropical cyclone core. Part II: Nonlinear enhancement. *J. Atmos. Sci.*, **58**, 2485–2501, https://doi.org/10.1175/1520-0469(2001)058<2485:TDOBLJ>2.0.CO;2.

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.

Li, X., and Z. Pu, 2008: Sensitivity of numerical simulation of early rapid intensification of Hurricane Emily (2005) to cloud microphysical and planetary boundary layer parameterizations. *Mon. Wea. Rev.*, **136**, 4819–4838, https://doi.org/10.1175/2008MWR2366.1.

Mallen, K. J., M. T. Montgomery, and B. Wang, 2005: Reexamining the near-core radial structure of the tropical cyclone primary circulation: Implications for vortex resiliency. *J. Atmos. Sci.*, **62**, 408–425, https://doi.org/10.1175/JAS-3377.1.

Marchok, T., 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.

Nolan, D. S., D. P. Stern, and J. A. Zhang, 2009: Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in situ observations and high-resolution simulations of Hurricane Isabel (2003). Part II: Inner-core boundary layer and eyewall structure. *Mon. Wea. Rev.*, **137**, 3675–3698, https://doi.org/10.1175/2009MWR2786.1.

Park, J., D.-H. Cha, M. K. Lee, J. Moon, S.-J. Hahm, K. Noh, J. C. L. Chan, and M. Bell, 2020: Impact of cloud microphysics schemes on tropical cyclone forecast over the western North Pacific. *J. Geophys. Res. Atmos.*, **125**, e2019JD032288, https://doi.org/10.1029/2019JD032288.

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419, https://doi.org/10.1175/2008WAF2222128.1.

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

Rogers, R., P. Reasor, and S. Lorsolo, 2013: Airborne Doppler observations of the inner-core structural differences between intensifying and steady-state tropical cyclones. *Mon. Wea. Rev.*, **141**, 2970–2991, https://doi.org/10.1175/MWR-D-12-00357.1.

Shapiro, L. J., and H. E. Willoughby, 1982: The response of balanced hurricanes to local sources of heat and momentum.

J. Atmos. Sci., **39**, 378–394, https://doi.org/10.1175/1520-0469(1982)039<0378:TROBHT>2.0.CO;2.

Smith, R. K., and M. T. Montgomery, 2015: Toward clarity on understanding tropical cyclone intensification. *J. Atmos. Sci.*, **72**, 3020–3031, https://doi.org/10.1175/JAS-D-15-0017.1.

Stern, D. P., and D. S. Nolan, 2009: Reexamining the vertical structure of tangential winds in tropical cyclones: Observations and theory. *J. Atmos. Sci.*, **66**, 3579–3600, https://doi.org/10.1175/2009JAS2916.1.

——, and ——, 2012: On the height of the warm core in tropical cyclones. *J. Atmos. Sci.*, **69**, 1657–1680, https://doi.org/10.1175/JAS-D-11-010.1.

——, J. R. Brisbois, and D. S. Nolan, 2014: An expanded dataset of hurricane eyewall sizes and slopes. *J. Atmos. Sci.*, **71**, 2747–2762, https://doi.org/10.1175/JAS-D-13-0302.1.

Tallapragada, V., C. Kieu, Y. Kwon, S. Trahan, Q. Liu, Z. Zhang, and I.-H. Kwon, 2014: Evaluation of storm structure from the operational HWRF during 2012 implementation. *Mon. Wea. Rev.*, **142**, 4308–4325, https://doi.org/10.1175/MWR-D-13-00010.1.

Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542, https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2.

Tyner, B., P. Zhu, J. A. Zhang, S. Gopalakrishnan, F. D. Marks Jr., and V. Tallapragada, 2018: A top-down pathway to secondary eyewall formation in simulated tropical cyclones. *J. Geophys. Res. Atmos.*, **123**, 174–197, https://doi.org/10.1002/2017JD027410.

Willoughby, H. E., H.-L. Jin, S. J. Lord, and J. M. Piotrowicz, 1984: Hurricane structure and evolution as simulated by an axisymmetric, nonhydrostatic numerical model. *J. Atmos. Sci.*, **41**, 1169–1186, https://doi.org/10.1175/1520-0469(1984)041<1169:HSAEAS>2.0.CO;2.

Wu, D., F. Zhang, X. Chen, A. Ryzhkov, K. Zhao, M. R. Kumjian, X. Chen, and P.-W. Chan, 2021: Evaluation of microphysics schemes in tropical cyclones using polarimetric radar observations: Convective precipitation in an outer rainband. *Mon. Wea. Rev.*, **149**, 1055–1068, https://doi.org/10.1175/MWR-D-19-0378.1.

Yuter, S. E., and R. A. Houze Jr., 1995: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part II: Frequency distribution of vertical velocity, reflectivity, and differential reflectivity. *Mon. Wea. Rev.*, **123**, 1941–1963, https://doi.org/10.1175/1520-0493(1995)123<1941:TDKAME>2.0.CO;2.

Zhang, J. A., D. S. Nolan, R. F. Rogers, and V. Tallapragada, 2015: Evaluating the impact of improvements in the boundary layer parameterization on hurricane intensity and structure forecasts in HWRF. *Mon. Wea. Rev.*, **143**, 3136–3155, https://doi.org/10.1175/MWR-D-14-00339.1.

——, 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., and Coauthors, 2022: Improving global weather prediction in GFDL SHiELD through an upgraded GFDL cloud microphysics scheme. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002971, https://doi.org/10.1029/2021MS002971.