Verification of Operational Forecast Models in Cases of Extratropical Transition of North Atlantic Hurricanes

ERICA BOWER,[a] KEVIN A. REED,[a] GHASSAN J. ALAKA JR.,[b] AND ANDREW T. HAZELTON[c,b]

[a] *School of Marine and Atmospheric Sciences, Stony Brook University, State University of New York, Stony Brook, New York*
[b] *NOAA/Atlantic Oceanographic and Meteorological Laboratory, Hurricane Research Division, Miami, Florida*
[c] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida*

(Manuscript received 18 January 2024, in final form 17 July 2024, accepted 15 August 2024)

ABSTRACT: Operational forecast models are necessary for the prediction of weather events in real time. Verification of these models must be performed to assess model skills and areas in need of improvement, particularly with different types of weather events that may occur. Despite the devastating impacts that can be caused by tropical cyclones (TCs) that undergo extratropical transition (ET) and become post-tropical cyclones (PTCs), these storms have not been extensively studied in the context of short-term weather prediction. This study completes the first analysis of the Global Forecast System (GFS) and a preoperational version of the newly operational Hurricane Analysis and Forecast System (HAFS) models in forecasting the occurrence of ET and the rainfall associated with ET storms in the North Atlantic basin. GFS's skill exceeds that of HAFS in forecasting the occurrence of ET, but HAFS tends to have lower track and rain-rate errors in the fully tropical phase of ET storms' life cycles. Both models simulate rain rates that are often too high near the storm center and fail to capture the larger area of moderate rain rates that greatly contributes to total rainfall accumulation. The discrepancies in rain rates between the models and Integrated Multi-satellitE Retrievals for GPM (IMERG) could be attributed to the models' tendency to keep storms too intense and too compact with an overly strong warm core, even throughout the ET process.

KEYWORDS: Extratropical transition; Precipitation; Tropical cyclones; Forecast verification/skill

1. Introduction

The prediction of tropical cyclones (TCs) is vital to protecting lives and property in areas affected by these storms. TC hazards include powerful winds, heavy rainfall, and storm surge that can lead to both coastal and inland flooding. To produce more accurate forecasts and effectively communicate risks to the general public, many broadcast meteorologists and emergency managers rely on National Weather Service products and numerical weather prediction models (Morss et al. 2022). These model forecasts guide decision-making and the response to the rapidly evolving conditions in their respective regions (Morss et al. 2022). As TCs move into the midlatitudes, some storms undergo an extratropical transition (ET) into post-tropical cyclones (PTCs) and lose their tropical characteristics (i.e., warm core and symmetric nature; Hart 2003). Throughout the ET process, the wind field and precipitation associated with the TC undergo dramatic shifts and expansion (Jones et al. 2003) as the TC interacts with midlatitude features. This phenomenon is fairly common, with nearly 50% of all North Atlantic (NATL) hurricanes undergoing at least one ET during their life cycles (Hart and Evans 2001; Zarzycki et al. 2017; Bieli et al. 2019). Furthermore, ET complicates the forecasting process due to scale interactions (Keller et al. 2011; Leonardo and Colle 2017; Balaguru et al. 2020) and precipitation shifts (Jones et al. 2003; Evans et al. 2017), especially if ET occurs during or close to landfall. An example of this situation is Hurricane Sandy (2012), which was responsible for 72 direct fatalities and $50 billion in damage in the Northeast and mid-Atlantic regions (Blake et al. 2013).

Many operational forecast models are used to forecast TCs and their impacts. Global operational models such as the National Oceanic and Atmospheric Administration's (NOAA) Global Forecast System (GFS) are used by forecasters to aid in the prediction of TC- and PTC-related events (Morss et al. 2022). Global models tend to have coarse resolution compared to regional models, which can limit their ability to forecast the smaller-scale processes that drive phenomena like TCs. Regional hurricane models excel in the prediction of storm track, intensity, and precipitation while storms are in their fully tropical phase (Dong et al. 2020; Hazelton et al. 2021; Ko et al. 2020; Alaka et al. 2022). These models have shown continued improvement in track and intensity forecasts in the last 20 years (Gopalakrishnan et al. 2021). Since 2007, the Hurricane Forecast Improvement Project (HFIP) has been devoted to improving regional hurricane models to reduce track and intensity errors as well as to improve the capability of predicting rapid intensification events (Gall et al. 2013). Over the course of the project, regional hurricane models have been developed, implemented operationally, and improved in their forecasts of track and intensity by as much as 20% at all lead times in the last decade (Gopalakrishnan et al. 2021).

Since 2019, NOAA has been developing and testing the Hurricane Analysis and Forecast System (HAFS) regional

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-24-0011.s1.

Corresponding author: Erica Bower, erica.bower@noaa.gov

DOI: 10.1175/WAF-D-24-0011.1

model as a next-generation replacement for the current operational regional hurricane models—the Hurricane Weather Research and Forecasting Model (HWRF) and the Hurricanes in a Multiscale Ocean-coupled Nonhydrostatic model (HMON) (Gopalakrishnan et al. 2021). HAFS not only shares the same finite-volume cubed-sphere (FV3) dynamical core as the most recent GFS (versions 15 and 16) but also includes smaller moving nests centered on active storms to simulate hurricanes at higher resolution than their surroundings (Alaka et al. 2022). Extensive development over the last several years resulted in the model undergoing real-time testing during the 2021 and 2022 hurricane seasons (Hazelton et al. 2022). In the 2019 season, HAFS had slightly less skillful intensity forecasts compared to HWRF and HMON at early forecast hours, falling farther behind by day 5 (Dong et al. 2020). In the near-real-time 2020 model runs, HAFS outperformed several other operational regional and global models until day 5, at which point the GFS forecasts had the lowest track errors (Hazelton et al. 2022).

Development of the model continued, leading to the current version of HAFS—which became operational in the 2023 hurricane season—improving track and intensity errors by as much as 10% over HWRF for NATL TCs (Gall et al. 2013; Mehra and Zhang 2023). Though these studies examined the performance of regional hurricane models like HAFS in the forecasting of fully tropical systems, no previous work details the performance of these types of models during and after the ET process. The spatial constraints of these limited domain regional models could increase the challenges of forecasting ET due to the inability to simulate upstream midlatitude features (Gall et al. 2013), allowing some global models to outperform the regional models at long lead times (Hazelton et al. 2021).

Regional and global operational forecasts of PTC-related precipitation have not been extensively studied, yet impacts on the United States alone from PTC precipitation can be devastating. In recent years, Hurricanes Irene (2011), Sandy (2012), and Ida (2021) have caused a combined $140 billion in damage in their PTC phases alone (Avila and Cangialosi 2013; Blake et al. 2013; Beven et al. 2022). While not all transitioning storms produce this level of devastation, 1–2 of these events occur each year on average along the East Coast of the United States (Bower et al. 2022). Despite these widespread flooding impacts, the majority of relevant prior work focused on the track errors that emerge during ET. Kehoe et al. (2007) examined west North Pacific ET events, discovering that most of the large track errors at 96 and 120 h were due to the direct interaction between an extratropical cyclone and a TC. Evans et al. (2006) similarly discovered that cluster forecast success significantly decreased after a TC began the ET process. Leonardo and Colle (2020) also found that along-track errors in ensemble forecasts were greater in ET cases than in non-ET cases at similar latitudes. Translational speed errors tend to worsen as well when ET occurs or is possible (Leonardo and Colle 2020). This could partly be due to timing errors in forecasting ET, which Bieli et al. (2020) note is an issue with their statistical and operational models for forecasting the ET process.

The combination of track and translational speed errors creates a challenging situation for extreme precipitation and flash flood forecasting, yet the effects of these model biases on precipitation forecasts have not been widely studied. Aside from Leonardo and Colle (2020), who assessed the standardized differences among models' representation of precipitation in cases where cross-track errors were excessive, most prior work on precipitation verification has focused instead on the tropical phase of the storms' life cycles. TC-related precipitation forecasts have been verified extensively, particularly the predictions from regional hurricane models such as HWRF (Ko et al. 2020; Bachmann and Torn 2021; Wang and Pu 2021); however, most of these analyses end when a storm begins to lose its tropical characteristics or only study landfalling storms (Ko et al. 2020; Bachmann and Torn 2021; Wang and Pu 2021; Stackhouse et al. 2023).

The present study focuses on GFS and HAFS forecasts of the precipitation resulting from transitioning TCs at all phases of the storm life cycle. Both of these operational forecast models have shown some skill in forecasting TC tracks, and GFS has demonstrated the ability to forecast ET events (Manikin et al. 2019; Liu et al. 2020, 2021). This study also aims to quantify the skill of GFS and HAFS in forecasting the occurrence and timing of ET. By combining the analysis of biases in both track and translational speed with an assessment of biases in rainfall rate forecasts, causes for specific precipitation forecast errors become evident. The remainder of the paper is structured as follows: section 2 describes the data and methods used in this study. The results concerning ET occurrence and track forecasts are presented in section 3. Rainfall errors are discussed in section 4. Section 5 details the large-scale environments surrounding the storms, and a discussion of the results and conclusions are included in section 6.

2. Data and methods

a. *Datasets*

1) GFS

The global operational model examined in this study is the GFS. Data are obtained from the University Corporation for Atmospheric Research (UCAR) Research Data Archive GFS Global Forecast Grid Historical Archive, which contains the deterministic forecast fields on a 0.25° horizontal resolution grid (NCEP 2007). The first 168 h (7 days) of each GFS forecast, available in 3-hourly temporal resolution, are evaluated to limit the impact of model error growth. All four daily model runs are utilized in the analysis, initialized at 0000, 0600, 1200, and 1800 UTC daily for the selected days in the 2019–21 seasons.

In June 2019, GFSv15 was implemented using an FV3 dynamical core (Harris et al. 2021) as developed by the Geophysical Fluid Dynamics Laboratory (GFDL) (Manikin et al. 2019, 2020). This update intended to address specific issues with the prior operational version of the GFS, including TC track forecasts at longer lead times, false alarms for TC genesis, and other issues not related to TC activity, such as improving the representation of the diurnal cycle of warm season

precipitation in the United States (Manikin et al. 2019). When specifically looking at TC simulation, the NOAA Environmental Modeling Center (EMC) Model Evaluation Group (MEG) reports a low bias in precipitation for certain extreme NATL TC cases (e.g., Hurricane Florence, 2018), a reduction in false alarms for TC genesis, and improvements in the short- and medium-range track errors for TCs (Manikin et al. 2019, 2020). Intensity forecasts were improved at all lead times in this version of the GFS, but the long-range track errors for days 6 and 7 remained problematic. The MEG notes that GFSv15 has a tendency to simulate TCs that have a greater vertical extent than the observed TCs, making them more susceptible to steering by the upper-level winds (Dawson et al. 2018).

An additional update to the GFS was implemented in March 2021, making GFSv16 operational for the 2021 hurricane season. GFSv16 increased the number of vertical levels from 64 to 127 while maintaining the same horizontal resolution and continuing to use the FV3 core. Some parameterizations were updated in this version but should have minimal impacts on the precipitation forecasts, including radiative flux calculations, the planetary boundary layer scheme, and gravity wave parameterizations. Most importantly, GFSv16 is one-way coupled to the WAVEWATCH-III wave model (Manikin et al. 2021). The update to v16 had some impact on TC forecasting as assessed by the MEG (Manikin et al. 2021); for example, potential TC precursors were identified at longer lead times at the expense of a higher false alarm rate for TC genesis. NATL TC track forecasts were improved particularly in the medium range, and the weak bias in intensity was mitigated as well. Finally, MEG noted a strong right-of-track bias at longer lead times in NATL TCs, which is especially of interest to this study of ET cases.

While it is not specifically designed to simulate hurricanes, GFS has many strengths, such as the ability to simulate over a week into the future and capture the atmospheric state of the entire globe. GFS can be skillful in situations with large-scale dynamical forcing due to its global domain and dynamical core. This may prove advantageous in simulating interactions between TCs and the midlatitude flow. However, other operational forecast models are specifically designed to simulate hurricanes.

2) HAFS

The regional model assessed in this study is the HAFS (Liu et al. 2020, 2021). HAFS was run in real time during the 2019–22 hurricane seasons leading up to its operational implementation in 2023. The model offers high spatial and temporal resolution, with 3-km horizontal grid spacing and 91 vertical levels up to a 10-hPa model top every 3 h until forecast hour 120. The model is two-way coupled to the Hybrid Coordinate Ocean Model (HYCOM; see HYCOM Documentation, https://www.hycom.org/hycom/documentation). The ocean model receives 10-m wind, 2-m wind and humidity, air–sea momentum flux, shortwave and longwave radiative fluxes, and precipitation data from the atmosphere model, while the atmosphere model ingests sea surface temperature (SST) data from the ocean model. Note that for the 2021 season, the orographic gravity wave drag (GWD) parameterization was turned on, while it had been turned off for the 2020 season. Both seasons have the convective GWD parameterization turned off (Liu et al. 2020, 2021).

HAFS 0.1A (2020 version) performed similarly to the HWRF regional hurricane forecast model in the 2020 season for North Atlantic hurricanes' track errors but improved on eastern North Pacific hurricane tracks. Some experiments using other versions of HAFS for the 2020 season showed that the inclusion of a scale-aware convection scheme further improved track errors over HAFS 0.1A. Furthermore, the convection scheme enabled HAFS to simulate stronger and deeper vortices than the HAFS 0.1A model used in this study (Liu et al. 2020). The scale-aware convection scheme was included in the 2021 season's HAFS 0.2A experiments, which played a role in the improvements over track forecasts from HAFS 0.1A. The HAFS 0.2A experiment track errors were nearly identical to the operational GFS in 2021, but HAFS intensity errors consistently outperformed those of GFS. Finally, comparing HAFS 0.2A to the operational HWRF in 2021 showed that HAFS 0.2A improved central pressure forecasts at all lead times as well as intensity errors beyond 48 h into the forecast (Liu et al. 2021). These two experimental runs of HAFS became the building blocks for the moving-nest precursor (HAFS 0.3) to the 2023 fully operational version of HAFS (Mehra and Zhang 2023; Hazelton et al. 2023).

3) IMERG

The Integrated Multi-satellitE Retrievals for GPM (IMERG) Final Run v06 product offers 0.1° spatial resolution and 30-min temporal resolution precipitation data (Huffman et al. 2019, 2020). Because of its high resolution, IMERG has been used in recent years for the analysis of TCs (Rios Gaona et al. 2018; Bower et al. 2022; Stansfield and Reed 2023). This dataset is constructed from several satellites' passive microwave radiometer precipitation estimates, which are then postprocessed for quality control and bias correction. Calibration is initially completed using the GPM dual-frequency precipitation radar, followed by intercalibration with the Combined Radar–Radiometer Algorithm (CORRA) product. Additional calibration to monthly Global Precipitation Climatology Project (GPCP) satellite-gauge estimates is done as well, followed by processing through two morphing algorithms to produce the high spatial and temporal resolution final product (Huffman et al. 2019, 2020). The IMERG product is downsampled to 6-hourly accumulation estimates to match the temporal resolution of the observed trajectories for this study. The spatial resolution is also downsampled to match that of ERA5 (see below) for collocation of TCs and PTCs with their associated precipitation.

4) IBTrACS

Observational TC data are obtained from the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2018). This study uses IBTrACS to record storm position, central pressure, and maximum sustained winds for NATL TCs

from 2019 through 2021. North Atlantic data are annually updated from the Atlantic Hurricane Database (HURDAT2) created by the National Hurricane Center (Landsea and Franklin 2013).

5) ERA5

IBTrACS TC trajectories include some but not all PTC trajectory points, depending on how the storm was classified and how long it was tracked after completing ET. For this reason, we also employ the use of the Extratropical Transition Tracker (ExTraTrack; Zarzycki et al. 2017) to standardize and objectively identify the ET process and all points afterward until the storm dissipates according to several mathematical criteria described in section 2b(2). This program requires geopotential height data for the full vertical column, low-level winds, and sea level pressure fields. To enable objective ET tracking, we match up IBTrACS trajectories with these fields in the ERA5 reanalysis dataset (Hersbach et al. 2020). This approach follows that of Bower et al. (2022). ERA5 offers global high-resolution data at roughly 31-km spatial and hourly temporal resolution at 37 vertical levels. This is then downsampled to a 6-hourly temporal resolution to match IBTrACS trajectory points.

Despite its strengths, ERA5 has some limitations as well. Bell et al. (2021) note that IBTrACS data are not assimilated into ERA5 for the period from 1979 to the present day. This contributes to TC intensities (using central pressure) being underestimated in many cases (Bell et al. 2021). However, Dulac et al. (2024) find that the errors in the central pressure of TCs in ERA5 are much lower than those in the low-level wind speeds. This point is further reinforced by Bourdin et al. (2022). In this paper, the wind–pressure relationship in IBTrACS was compared to that in ERA5, showing that at the same central pressures, the wind speeds recorded in ERA5 are lower than those in IBTrACS. Therefore, we use central pressure as a measure of intensity for the remainder of this analysis. However, it is worth noting that ERA5 central pressures tend to be slightly higher than those of IBTrACS (Bourdin et al. 2022).

b. *Analysis tools*

1) TempestExtremes

TempestExtremes is utilized throughout the analysis methods for various tracking functions, similar to Bower et al. (2022). The specific commands and settings used for the present analysis can be found in the online supplemental material.

2) ExTraTrack

ExTraTrack supplements the tracking capabilities of TempestExtremes, using a cyclone phase space (CPS) as in Hart (2003) to objectively identify the occurrence of ET in model and reanalysis datasets. ExTraTrack also extends the pre-existing TC trajectories to include post-ET points in the life cycle that are not included in the original IBTrACS archive. A storm is considered to have dissipated when the storm's central pressure rises above 1020 hPa or when 14 days have passed since ET occurred. Additionally, ExTraTrack imposes directional change limitations on the storm based on its translational speed, with larger changes being allowed when storms are slow moving. Finally, some postprocessing is included in this software package to construct climatologies of the ET events, types of ET, and other calculations. Additional details are provided in Zarzycki et al. (2017) and Bower et al. (2022).

3) MET

The Model Evaluation Tools (MET; Jensen et al. 2020) is a software package developed by the Developmental Testbed Center (DTC) that enables the verification of forecasts using various methods, including point-based, object-based, or grid-based approaches. Different functions can be configured to the user's needs, making the program highly versatile. In this study, the TC-Pairs function is used for verification of the forecast TC and PTC trajectories. The function compares TC locations and intensities point by point, quantifying forecast errors throughout the storm life cycle. Object-based precipitation tracking is completed in TempestExtremes rather than MET in order to implement the Z500 mask extraction method.

c. *Methods*

Storms to be considered in the analysis of the GFS model are those in the 2019–21 North Atlantic hurricane seasons that completed at least one ET during the storms' lifetimes. Due to data availability, only the 2020–21 North Atlantic hurricane seasons are considered for the analysis of the HAFS model. Tropical transitions are excluded due to the various pathways of transition that could be taken (Davis and Bosart 2004), as well as the lack of consistency in proximity to the dominant midlatitude flow. Storms that are never fully tropical (e.g., subtropical storms as recorded in IBTrACS) and those that only retain tropical storm characteristics for less than 24 h prior to ET as determined by ExTraTrack (short-lived, sometimes hybrid storms) are also excluded. Hurricane Sam was excluded due to the lack of data availability for IMERG v06, which was no longer produced after 30 September 2021 (Huffman et al. 2020). A detailed list of the storms used in this analysis is included in Table 1. GFS model runs are selected to include 5 days prior to and 2 days after the storm's observed ET onset as determined by ExTraTrack, with four model runs from each day, totaling 29 model runs per storm. The same time frame is selected for the HAFS model, but only the 0000 and 1200 UTC runs are available for analysis for some of these selected days. The number of HAFS simulations studied for each storm is in parentheses in Table 1, as the number varies by storm. The inclusion of 5 days prior to ET allows for the examination of the ability of GFS to forecast ET at longer lead times along with any track errors at the medium range lead times. The inclusion of 2 days after the observed ET onset retains model runs initialized while the storm may still begin with tropical characteristics and allows for the assessment of ETs with long durations.

First, the IBTrACS trajectories for the storms of interest are obtained (see Fig. 1 left) based on the criteria listed above. These trajectories are then extended to include all

TABLE 1. Storms that completed at least one ET in the 2019–21 NATL hurricane seasons. Storms that were excluded either made a tropical transition, remained subtropical throughout their entire life cycles, or only maintained tropical characteristics for less than 24 h before ET. Storms that were also included in the analysis of HAFS have the number of model runs in the HAFS analysis included in parentheses.

2019		2020		2021	
Included	Excluded	Included	Excluded	Included	Excluded
Dorian	Melissa	Arthur	Bertha	Elsa (11)	Bill
Erin	Nestor	Cristobal	Edouard	Ida (12)	Claudette
Gabrielle	Olga	Isaias (12)	Kyle	Larry (15)	Julian
Humberto		Paulette (15)			Odette
Lorenzo		Sally (15)			Sam
Sebastien		Teddy (12)			
		Beta (9)			
		Delta (3)			
		Epsilon (15)			
		Zeta (11)			
		Eta (15)			

points during and after ET has occurred using ExTraTrack [section 2b(2)]. ExTraTrack also uses the IBTrACS locations of storms and calculates the CPS parameters from ERA5 variables at each IBTrACS point. GFS and HAFS data are then obtained for each of the storms listed in Table 1, and all active TCs within each model run are tracked using TempestExtremes. These TC trajectories are then extended using ExTraTrack, with the same criteria in place as for the extension of the observed trajectories. The GFS (HAFS) trajectories for simulations that successfully predicted ET occurrence within 24 h of the observed ET onset are shown in Fig. 1 in the center (right) in gray with the corresponding observed trajectories shown in blue.

Next, all precipitation objects are tracked in both the IMERG dataset and each model run. Sensitivity analysis was conducted using thresholds of 0.5, 1, and 5 mm h^{-1} rain rates (see Figs. S1 and S2). Previous studies have used various thresholds for extracting specific types of precipitation; for example, 5 mm h^{-1} can be used to study deep convection (Zawislak 2020), or 2.33 mm h^{-1} (Skok et al. 2009) can be used to study generic tropical convection. Wu et al. (2015) and Lonfat et al. (2004) determined that the mean rain rate in the outer portions of TCs is 1 mm h^{-1}. Therefore, the 1 mm h^{-1} rain rate threshold was selected based on the ability of this metric to capture the lighter extratropical precipitation during the PTC phase while not capturing extraneous precipitation on the outer edges of the TC or PTC.

The storm track of interest is then isolated by comparing TC tracks from the model run to the observational trajectory of the storm. Then, a mask for the determination of TC- or PTC-related precipitation is created using the 500-hPa geopotential height (Z500 mask) as in Bower et al. (2022). The Z500 mask shape can be irregular, taking on the shape of the 500-hPa geopotential height field where the value increases by 10 m from the value at the storm center. If this value occurs outside of the 5° great circle distance (GCD) search radius, then the mask defaults to the search radius size. A 1° GCD minimum mask is also imposed to ensure that intense TCs are included in the analysis even when the 10-m geopotential height change occurs at distances smaller than the grid resolution of the dataset. This Z500 mask is overlaid with the tracked precipitation objects, and all objects that overlap the Z500 mask at any point are included in the analysis. This method uses the storm center as a starting point and then searches for precipitation objects that overlap with the area defined by the geopotential height field. This accommodates for slight differences in the IBTrACS and ERA5 storm locations as well as the effects of shear and interaction with the midlatitude flow.

The timing and location of ET are determined by ExTraTrack as an objective, standardized method of calculating these metrics. The program, described in section 2b(2), uses the CPS (Hart 2003) to track the ET process. The tropical phase of the storm life cycle is defined as times at which the storm has both a warm core nature and is thermally symmetric (bottom-right quadrant of the CPS). ET begins when one of those two criteria is no longer satisfied (top-right or bottom-left quadrant of the CPS). Finally, ET is considered complete (or the storm is classified as extratropical) when it is thermally asymmetric and has a cold core (top-left quadrant of the CPS). The analysis is broken down by storm phase: before ET onset (pre-ET), during the ET process (during-ET), and after ET completion (post-ET), as defined using the CPS quadrants (see Fig. 11 for an example). This breakdown enables a more targeted assessment of the model's performance in precipitation, track, translational speed, and timing errors relative to the ET process. Composites of storm rainfall are then created at 3-h intervals for GFS and HAFS and 6-h intervals for IMERG and ERA5. Command settings are found in the supplemental material. The compositing is based on the average of a particular field on a storm-centered grid. The HAFS fields are regridded to the ERA5 grid for verification before composites are made. This method is used for analyzing storm precipitation rate and accumulation errors along the entire TC and PTC trajectory. Though the acronym "PTC" in the operational community may sometimes refer to a

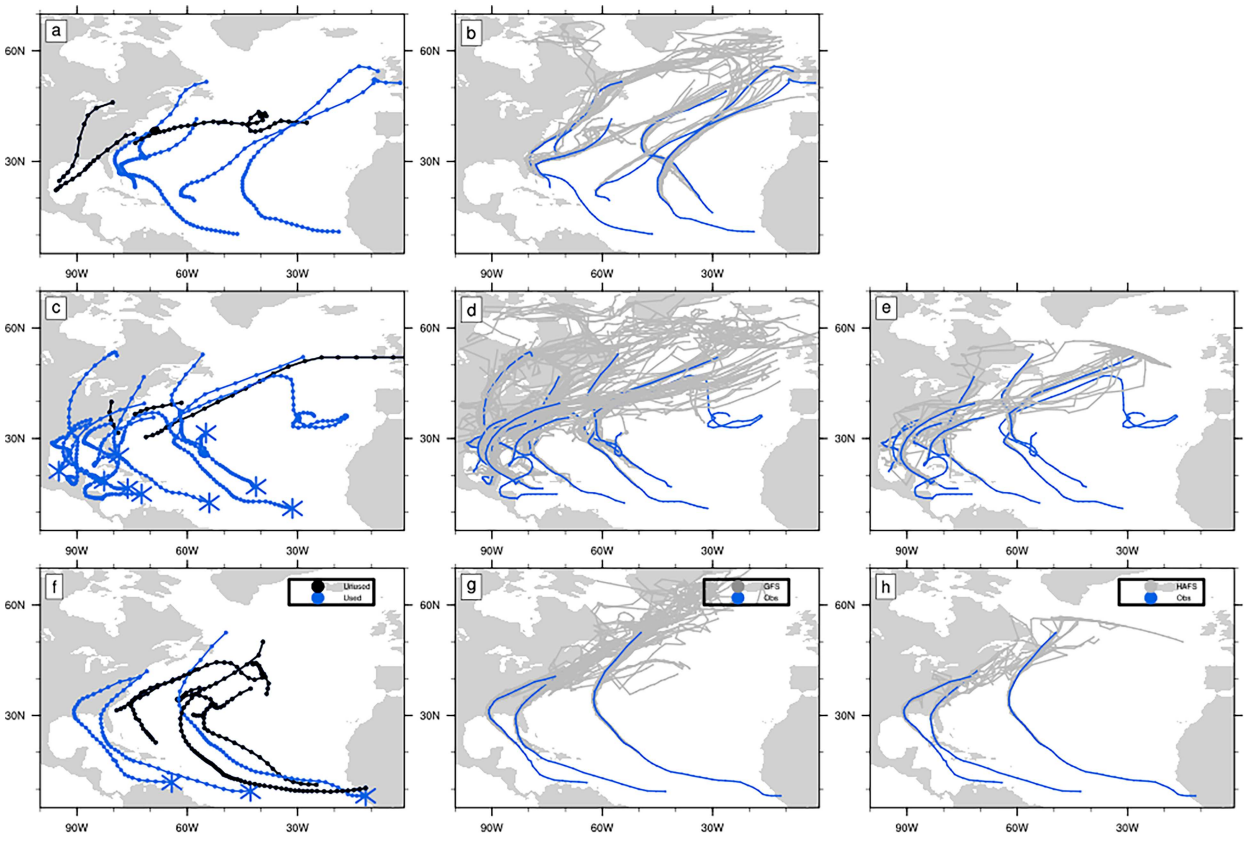

FIG. 1. Trajectories of the storms assessed in this study for the (a),(b) 2019, (c)–(e) 2020, and (f)–(h) 2021 seasons. (left) Observational trajectories. Black trajectories indicate storms that were excluded from the analysis. Trajectories that begin with stars belong to storms that are included in both the GFS and HAFS analyses. (center) Observational trajectories are in blue along with the GFS trajectories from simulations that correctly forecast ET within 24 h of the observed ET (gray). (right) As in the center panels, but for HAFS.

potential tropical cyclone, in this manuscript, it will be used to signify post-tropical cyclones and refers to the during- and post-ET phases combined. Throughout the results, the term "lead time" will refer to the initialization time as the number of days relative to the observed ET onset and will be represented by the symbol T_L. The term "forecast hour" will refer to the number of hours into a given simulation, represented by the symbol T_F.

3. ET occurrence and track errors

Before examining the rainfall predictions from GFS and HAFS in ET scenarios, we assess the ability of each model to predict ET occurrence. Figure 2 shows the percentage of the model runs analyzed that successfully predicted ET as a function of initialization time, or T_L. Each color line represents a different level of accuracy in the representation of ET timing in the model simulations. For example, 5 days before the observed ET (far left of Fig. 2a; $T_L = -5$), 20% of the evaluated GFS simulations were able to correctly forecast ET onset within 12 h of the observed ET onset (red line). The sample size for each initialization time for GFS is 20 (one simulation for each storm), but the sample size at each initialization time for HAFS varies due to the availability of the simulations for each storm.

Only 35% of the GFS simulations initialized 5 days before observed ET were able to correctly simulate ET for their respective storms with the timing of ET being accurate to within 24 h of the observed ET (Fig. 2a). As T_L decreases (initializing closer to the observed ET onset), the GFS simulations tend to become more successful at not only predicting ET but pinpointing the timing of the process as well (Fig. 2a). By $T_L = -1$, 70% of the simulations were able to correctly forecast ET with errors as low as 12 h. The HAFS simulations were not as successful in predicting ET. At $T_L = -5$, only 8% of the simulations studied were able to correctly forecast ET occurrence (Fig. 2b), but those that did were accurate to within 24 h of the observed ET onset. None of these simulations were able to predict ET occurrence within 12 h of the observed timing at long lead times. This percentage rises as lead times shorten, reaching a maximum of 42% of simulations correctly forecasting ET within 24 h of the observed ET around $T_L = -3$ and -2. Similar to GFS (see Fig. 2a), there is a sharp drop-off in this percentage around 12 h prior to ET onset in observations. The deterioration in skill is likely due to the development of hybrid characteristics in storms as the onset of ET approaches. In this case, the storm initialized in the model never has fully tropical characteristics and therefore is not identified as a candidate for ET in that specific model run.

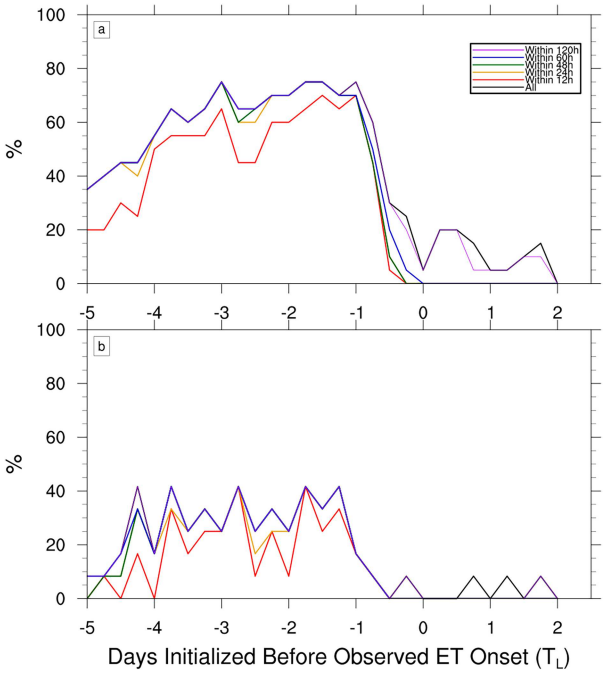

FIG. 2. Percentage of simulations at each lead time (T_L) that forecast ET. (top) For GFS ET cases in Table 1, the percentage of simulations that correctly forecast ET occurrence for the correct storm within various time frames of observed ET. (bottom) As in the top panel, but for HAFS cases in Table 1. The sample size for each lead time for GFS (top panel) is 20 simulations, and for HAFS (bottom panel), the sample size varies for each lead time.

In both models, most of the simulations with successful ET predictions forecast its onset within 24 h of the observed ET onset. For this reason, some of the following analyses include only the model runs that correctly predict ET within 24 h of the observed ET onset. This restriction retains a larger sample while excluding the few model runs with high ET timing errors (those greater than 24 h).

The errors in the simulated tracks of both models compared to the IBTrACS observed trajectories (extended using ExTraTrack) are shown in Fig. 3. In this figure, track errors from the GFS (HAFS) runs that correctly simulated ET within 24 h of the observed ET onset are included in the black (blue) box-and-whisker diagrams, while the remaining GFS (HAFS) runs' track errors (poorly simulated ET timing as well as simulations that did not forecast ET at all) are included in the red (green) boxes. Error values are shown as a function of forecast hour (T_F), grouping all data points from within each 24-h forecast period of every simulation into a box-and-whisker chart. For example, simulations initialized at −120, −114, −108, and −102 h are binned into the day −5 lead time ($T_L = -5$) category, while forecast hours 0, 6, 12, and 18 are binned into the first box-and-whisker chart at $T_F = 24$.

In general, track errors increase with increasing forecast hours at all lead times (Fig. 3), as shown in previous work examining all forecast points between forecast hours 72 and 120 (Leonardo and Colle 2020). In addition to being a function of forecast hour, ET timing plays a role in the magnitude of track errors. It is possible both that the larger track errors cause the ET timing errors due to different placement of the TC, and thus different interactions with the midlatitude flow, or vice versa, errors in ET timing lead to larger track errors after the interaction occurs due to different translational speeds of the observed and simulated TCs. Regardless of the order of events, the median total track errors (TTEs; also sometimes referred to as absolute track errors) for both models' simulations with poor ET timing forecasts or no ET forecast to occur tend to be higher than the median TTEs for simulations that forecast ET timing well. This difference becomes more pronounced at the time that observed ET occurs—for example, at $T_F = 120$ for $T_L = -5$ and so on. At short lead times and long forecast hours (e.g., Fig. 3e), TTEs cease to increase with increasing forecast hours as sample size rapidly diminishes. It should also be noted that 69% of the data points in the red (GFS) boxes come from simulations that do not forecast ET at all, with the other 31% representing simulations that have ET timing forecast errors of greater than 24 h. For all cases, both models tended to simulate ET later than the observed (68% of GFS and 70% of HAFS simulations) and thus typically farther poleward (not shown).

To gain insight into the causes of these track errors, we also break the TTEs down into their components: along- and cross-track errors (ATEs and CTEs), shown in Fig. 4. The tendency of "good timing" ET simulations to outperform the "poor timing" simulations persists for the ATEs at all lead times. CTEs lack a consistent and definitive pattern based on the lead time, forecast hour, or ET timing forecast. ATEs tend to be negative overall, indicating a slow bias in the storm forward motion. CTEs from both models are centered, suggesting that at most lead times and forecast hours, GFS and HAFS are as likely to exhibit a right-of-track error as a left-of-track error. HAFS does tend to have more positive CTEs than GFS, with the largest differences at long forecast hours. Similar to TTEs, the ATEs and CTEs at longer forecast days in the shorter lead-time simulations tend to be noisy and lack a consistent pattern, particularly for $T_L = -2$ (Fig. 4i) for the CTEs and $T_L = -1$ (Fig. 4e) for the ATEs. Much larger ATEs occur in simulations that poorly forecast ET compared to the well-forecast ET simulations, especially at long lead times. Translational speed errors are likely the primary driver for the track errors that GFS and HAFS experience in these cases, consistent with prior studies that find increased forecast uncertainty tied to the changing forward speed of storms during ET (Evans et al. 2017). The results for GFS and HAFS are closely linked, partially because the initial and lateral boundary conditions for HAFS are provided by the concurrent GFS simulation.

4. Rainfall forecast evaluation

Next, we investigate the skill of GFS and HAFS in simulating rainfall associated with TCs that undergo ET at all phases of their life cycles. For reference, the total accumulated rainfall from the storms studied in the PTC (during and post ET) phase of their life cycles from IMERG, GFS, and HAFS is

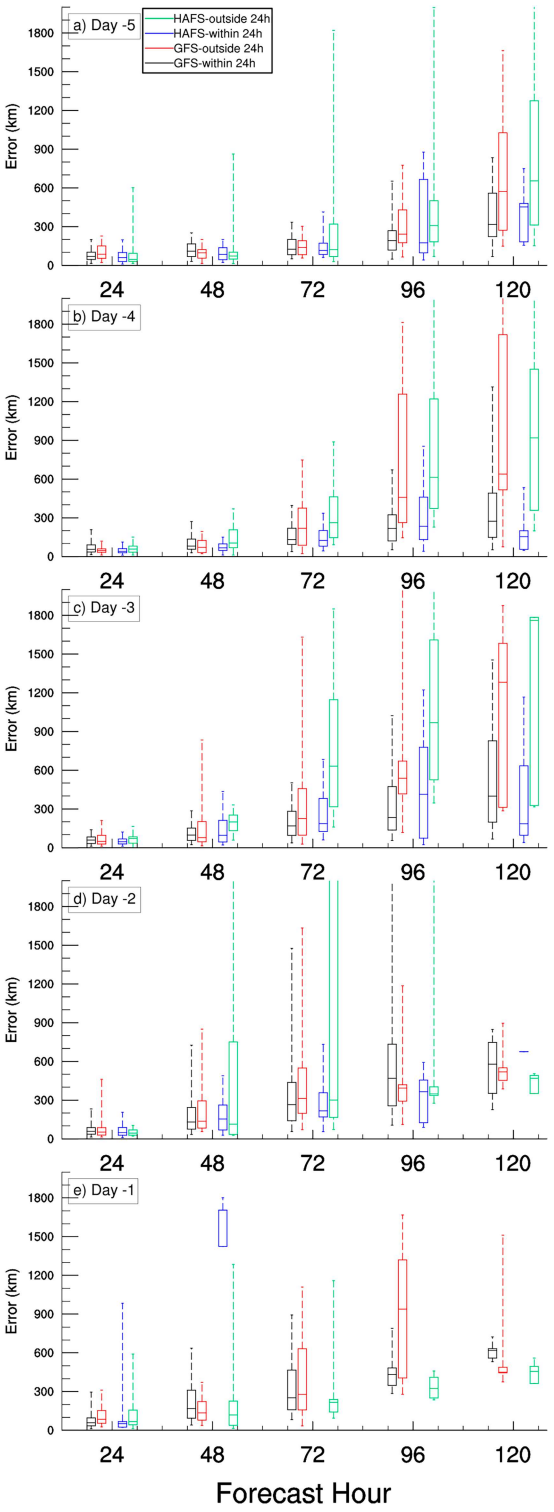

FIG. 3. TTEs for the GFS and HAFS simulations used in this study. Black boxes include track errors for all GFS simulations that forecasted ET occurrence within 24 h of the observed ET time. Blue boxes include the same for HAFS simulations. Red boxes include track errors from GFS simulations that had ET timing errors of greater than 24 h or did not predict ET at all. Green boxes include the same quantity for the HAFS simulations. Each panel includes simulations initialized at that lead time grouped into days. For example, $T_L = -5$ includes simulations initialized at −120, −114, −108, and −102 h. Forecast time T_F increases on the x axis.

shown in Figs. S3 and S4. The total rainfall accumulation as predicted by both models is considerably lower than the amounts recorded by IMERG due to the challenges in simulating these storms. However, many factors contribute to differences in the accumulated rainfall from each storm, such as storm translational speed, rain rates at individual time steps, and the distribution of rainfall around the storm center. To learn more about the reasons for these patterns, we next turn to the rain rates that are simulated by each model.

Examining composites of the rain rates at individual trajectory points along the storms' paths provides further insight into the rainfall accumulation errors in the GFS and HAFS simulations. Figures 5 and 6 show the composite rain rates for storms in the pre-ET (left), during-ET (center), and post-ET (right) phases for the IMERG values (top), GFS or HAFS simulated storms (middle), and the statistically significant differences between the model and IMERG average composite rain rates (bottom). Statistical significance is determined using a two-sample t test for the difference of means with a 95% significance level, and only precipitating points (rain rates of at least 1 mm h^{-1}) are included in the analysis.

In the pre-ET phase (left, Fig. 5), GFS-simulated rain rates tend to be higher than IMERG in the inner core of the storm. In the during-ET phase (center), this overestimation shifts to the left side of the storm and becomes asymmetric. GFS tends to forecast a more compact area with the highest rain rates than IMERG shows during ET, with the IMERG dataset spreading higher rain rates farther from the storm center. Both GFS and IMERG match in their depiction of the shift of precipitation to the north of the storm center during ET. Finally, once ET is complete, GFS predicts similar rain rates to IMERG, with higher values not extending far enough away from the storm center coupled with a more uniform area of lower rain rates. Furthermore, GFS does not show the dry southwest quadrant of the storm that is evident in IMERG. Similar to GFS, HAFS (Fig. 6) overestimates the inner core rain rates in the pre- and during-ET phases when compared to IMERG. These overestimations are less severe than those of GFS in the pre-ET phase, but both models share the tendency to simulate rainfall that is too compact, symmetric, and close to the storm center, especially in the during-ET phase. The overestimation of rain rates near the center of storms in the post-ET phase is more pronounced in the HAFS simulations compared to GFS.

Across all phases of the storm life cycle, GFS and HAFS tend to simulate higher rain rates than seen in IMERG but over a smaller core area than covered by the maximum rain rates shown in IMERG. This pattern is especially pronounced once the ET process has begun. When examining the total rainfall accumulation resulting from the observed and simulated storms, the small area of high rain rates leads to a

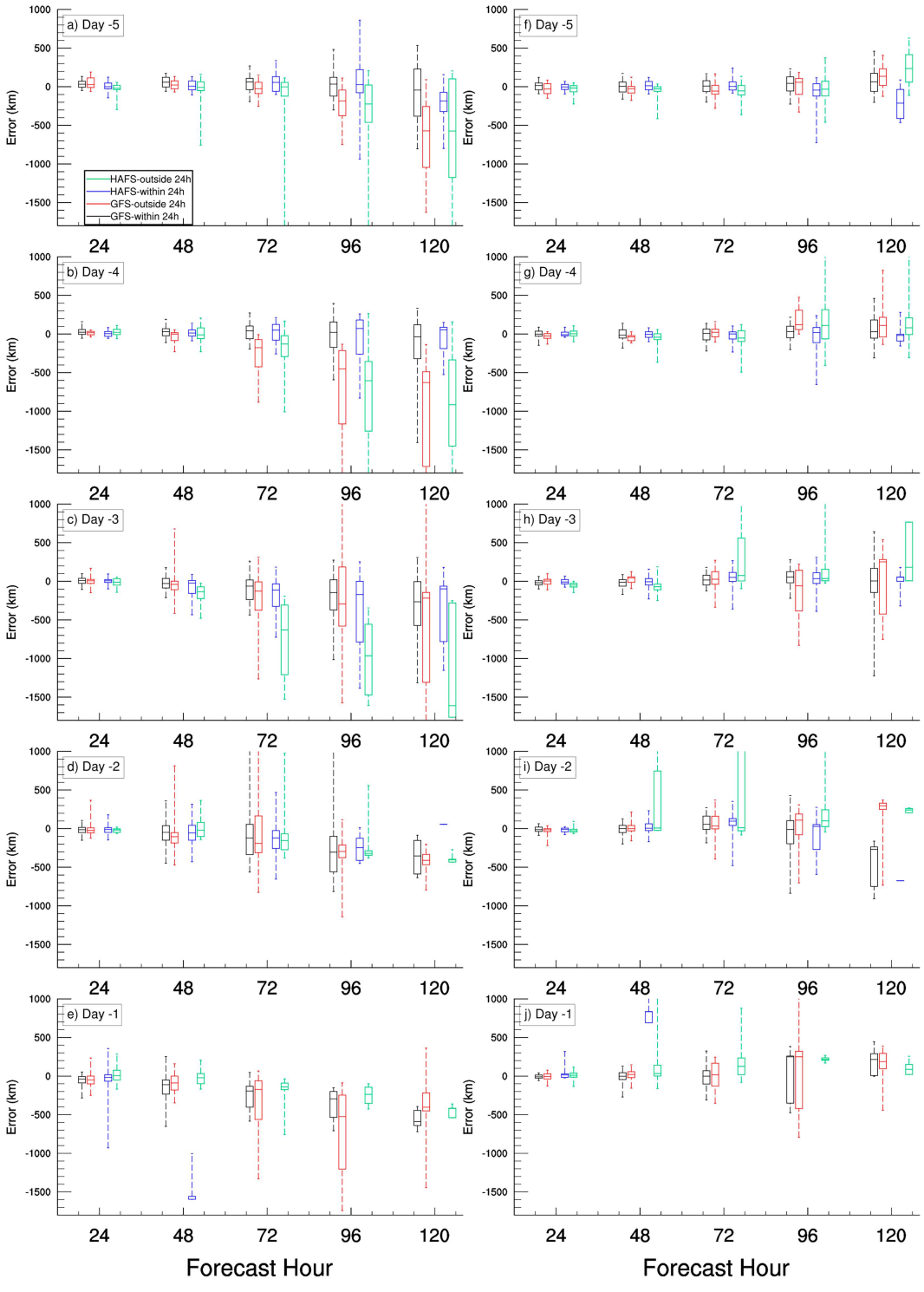

FIG. 4. As in Fig. 3, but for (left) ATEs and (right) CTEs.

smaller total accumulation in the models than the larger area of more moderate rain rates in IMERG. This underestimation of the total accumulated rainfall by both models is exacerbated in the during- and post-ET phases, suggesting that the structural evolution of the storm from being purely fueled by warm core processes to being fueled by baroclinic forcing differs from the real-world evolution. This concept will be explored further in the following section.

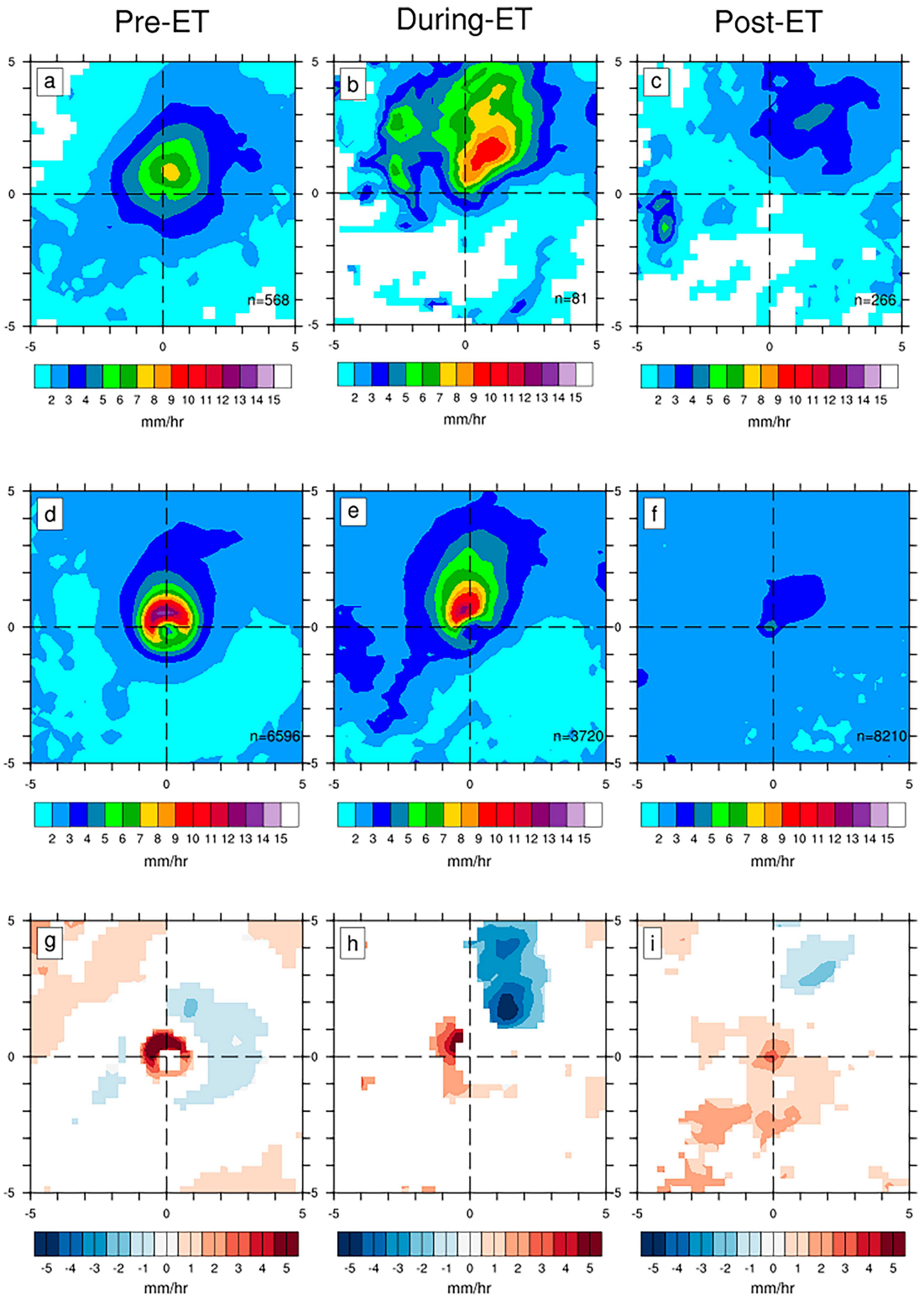

FIG. 5. Rain rate composites from the storms included in Table 1. (left) Pre-ET phase. (center) During-ET phase. (right) Post-ET phase. (top) Average IMERG 6-hourly rain rates across all storms. (middle) As in the top panels, but for GFS simulations. (bottom) Statistically significant difference between the GFS simulation average and the IMERG average rain rates at each phase of the TC life cycle. Statistical significance is determined using a two-sample *t* test for the difference of means.

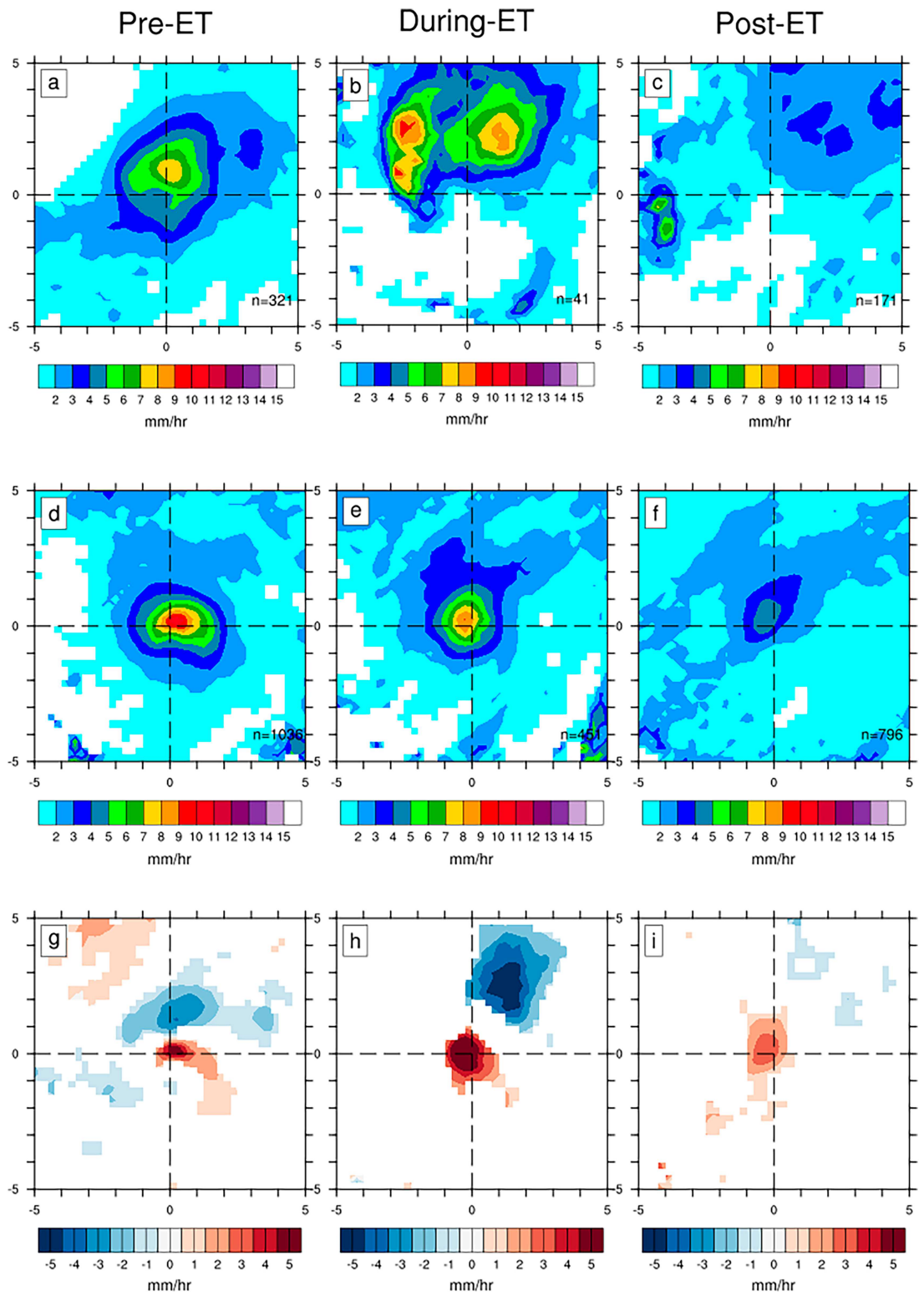

FIG. 6. Rain rate composites from the storms included in the HAFS analysis. As in Fig. 5, but for the HAFS simulations.

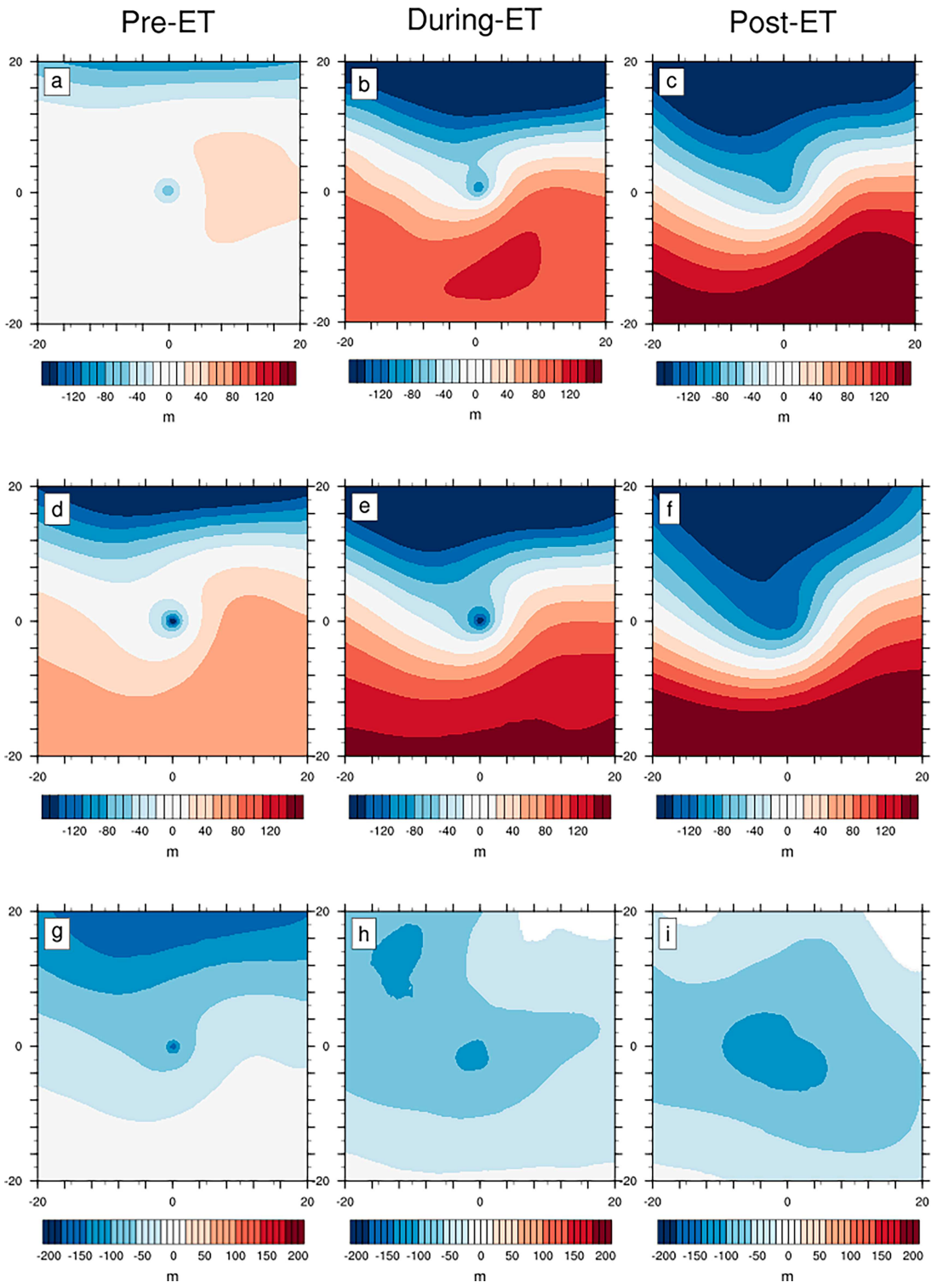

FIG. 7. Composites of 500-hPa geopotential height (Z500; m) for (top) the ERA5 reanalysis, (middle) GFS model runs, and (bottom) the difference between the original GFS and ERA5 average fields. ERA5 and GFS are shown as anomalies from the average value of the 40° × 40° composite field for the (left) pre-ET, (center) during-ET, and (right) post-ET phases.

5. Mean state and storm structure/evolution

To shed light on the potential cause of the patterns in the GFS and HAFS simulations of the storm rain field during and after ET, we examine the environment surrounding the simulated TC compared to the environment depicted in ERA5 reanalyses. To begin this investigation, we examine the geopotential height at 500 hPa (Z500). Z500 encompasses both midlatitude features, which are frequently identifiable at the midtropospheric levels, as well as TCs, which have a warm core extending from the mid- to upper levels.

Figure 7 shows the anomaly of the Z500 average composite from the average value of that field for ERA5 (top) and all GFS simulations (middle) along with the difference between the original GFS and ERA5 Z500 average composite fields in the bottom row. The pre-ET phase is shown on the left, the during-ET phase is shown in the center, and the post-ET is shown on the right. The equator-to-pole gradient of Z500 anomalies and the actual Z500 values are stronger in GFS simulations than in reanalysis in the pre- and during-ET phases of the storms' life cycles (Fig. 7). Once the ET process is complete, GFS shows a more meridional orientation of the midlatitude feature (Fig. 7f) as opposed to the more zonal pattern on the poleward side of the ERA5 composite (Fig. 7c). This indicates improper phasing of the TC and the upstream trough in GFS simulations compared to ERA5 reanalysis. Additionally, the Z500 anomaly within the storm has a higher magnitude in the GFS simulations as opposed to the ERA5 reanalysis. This could indicate a more intense storm or a more resilient warm core, which may resist the deterioration of deep convection and support the higher rain rates forecast by GFS, as discussed in section 4. A case study of the differences in the Z500 field between the models and the reanalysis is explored in the supplemental material (see Fig. S5).

To further investigate, we next examine composites of mean sea level pressure (SLP) for both the GFS simulations and ERA5. Figure 8 is as in Fig. 7 except for SLP for ERA5, GFS, and the difference between the two. SLP was chosen as a measure of intensity for the observed storms due to ERA5's unreliability in representing accurate low-level winds, particularly in TCs (Schenkel and Hart 2012; Stansfield et al. 2020; Bourdin et al. 2022). GFS consistently shows more intense storms than ERA5 with stronger gradients of SLP and higher magnitude anomalies than in the reanalyses. This overestimation of storm intensity could be contributing to the high rain rates seen in GFS simulations, as previous literature has demonstrated a link between TC intensity and rain rates (Jiang and Zipser 2010; Alvey et al. 2015).

When completing the same analysis for the HAFS simulations (Figs. 9 and 10), we find some marked similarities between the two models' behaviors. Like GFS, HAFS tends to simulate a sharper equator-to-pole gradient of Z500 than does ERA5, especially in the pre-ET phase (Fig. 9d). The storm center also has lower Z500 values in HAFS than in the reanalysis. The main difference between the HAFS and GFS Z500 fields is in the post-ET phase (Fig. 9f). While GFS simulates a more meridional feature once ET is finished, HAFS shows a zonal flow with lower amplitude than ERA5. When examining SLP for the HAFS simulations, once again the simulated storms have lower SLP anomalies than in ERA5 (Fig. 10), keeping the storms too intense especially before ET begins.

The strong gradient of Z500 in both models' simulations still requires investigation into the strength of the simulated warm cores of the storms of interest. A strong warm core would remain more resistant to the effects of the baroclinic environment (i.e., the transition to a cold core). A CPS density chart is shown in Fig. 11 to determine the strength of the warm core (as determined by the CPS values of $-V_T^L$, which represent the thermal anomaly in the atmosphere) and the degree of asymmetry seen in GFS and HAFS simulations compared to the ERA5 reanalysis. The left panel shows the percentage of all points in the GFS simulations used in this study within the CPS that occur in each grid square. The middle shows the same for all ERA5 points, and the right shows the same for all HAFS simulations. While the storms in ERA5 tend to remain warm core and thermally symmetric for longer, GFS storms can retain a very strong warm core (higher $-V_T^L$ values) while simultaneously becoming thermally asymmetric (developing fronts or remaining warm core in the comparatively cold baroclinic zone, thus achieving higher B values). The HAFS simulations also tend to simulate very strong, symmetric warm cores (TCs; bottom-right quadrant) most often. While HAFS does simulate some instances of asymmetric warm cores (top-right quadrant), only a small percentage of the storms achieve a cold core, regardless of thermal symmetry. The retention of the strong warm core would support deep convection in the simulated storms at times when the observed storm's warm core is deteriorating and causing the demise of deep convection at the storm center. These discrepancies between the model simulations and ERA5 reanalyses as a whole contribute to the rain rate errors seen in the simulations.

6. Discussion and conclusions

This study completed verification of GFS and HAFS in their abilities to forecast ET occurrence and the rainfall associated with storms that completed ET. While GFS is able to correctly forecast ET occurrence within 24 h of the observed ET 35% of the time at 5 days lead time, HAFS has almost no skill in forecasting ET at medium-range lead times. At 1 day prior to observed ET onset, the percentage of GFS simulations correctly forecasting ET with accurate timing rises to 70%, but a maximum of only 42% of HAFS simulations correctly forecast ET 2 days before observed ET onset. Noteworthy is that both models' ET timing, when forecast, is usually accurate to within 24 h of the observed ET onset at any lead time. ET occurrence and timing influence the track errors for each model's simulations as well. TTEs tend to be higher in simulations that do not forecast ET for the storm of interest or forecast ET with poor timing relative to the observed ET onset. Track errors also rise with increasing forecast hour, as model errors are amplified further into the simulation. TTEs tend to be driven by ATEs, or forward speed biases, in both models, which follow the same patterns as the TTEs. These ATEs are driven by the model's ability to predict ET (Fig. 4) as well as by errors in the phasing between the TC and the

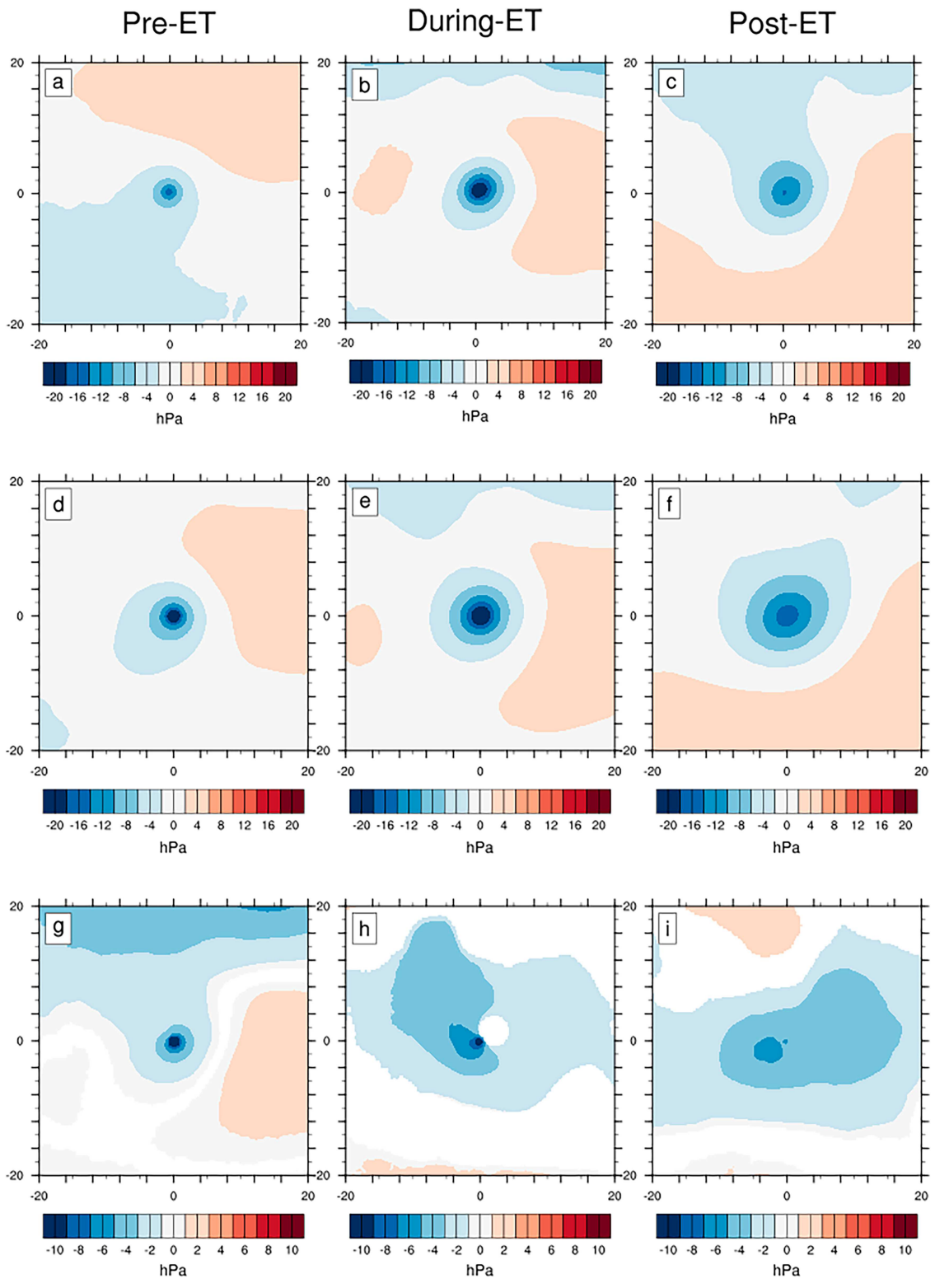

FIG. 8. As in Fig. 7, but for mean SLP (hPa).

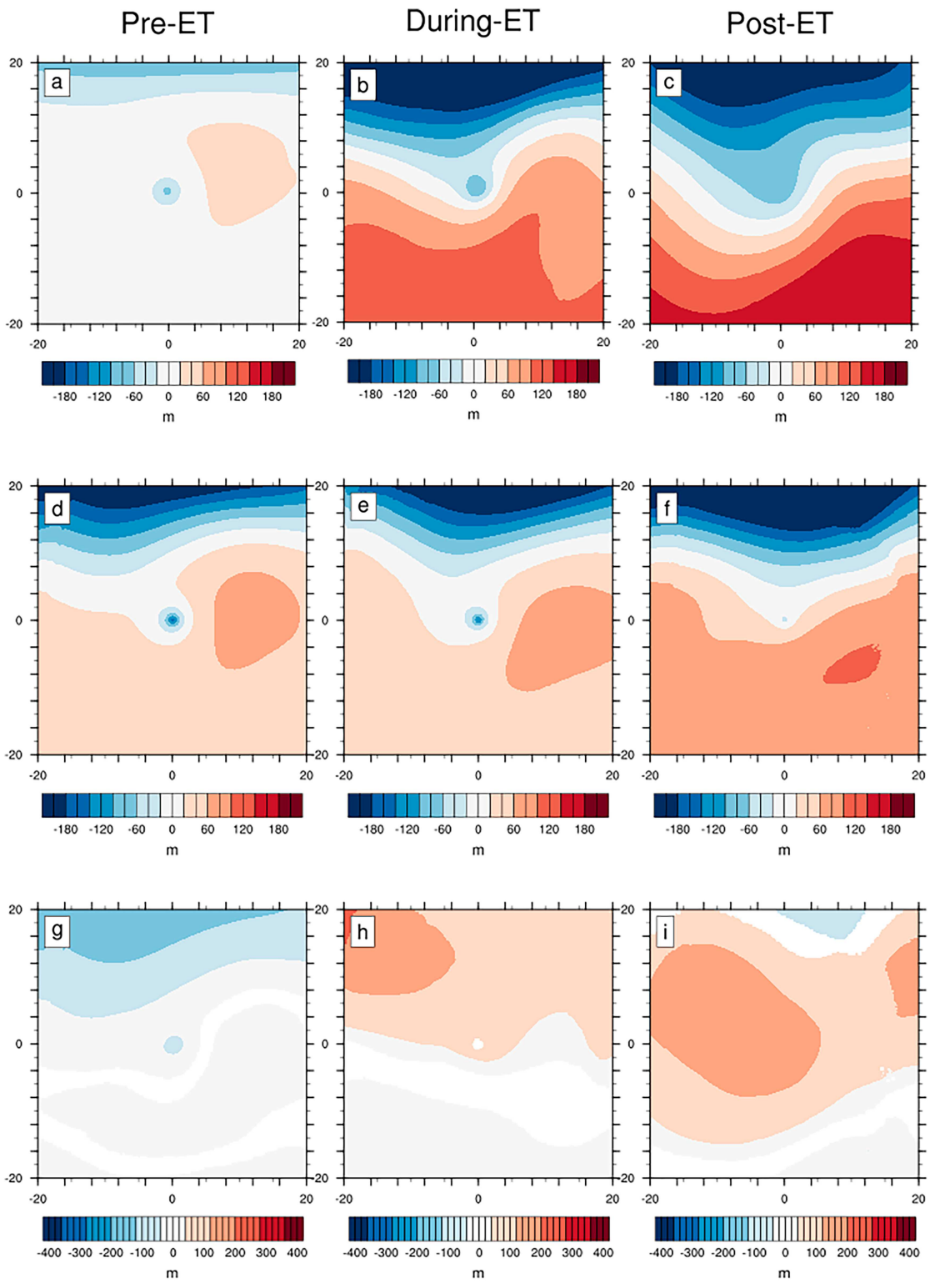

FIG. 9. As in Fig. 7, but for HAFS simulations. Note that snapshots in which the storm center was located within 20° of the domain edge were excluded.

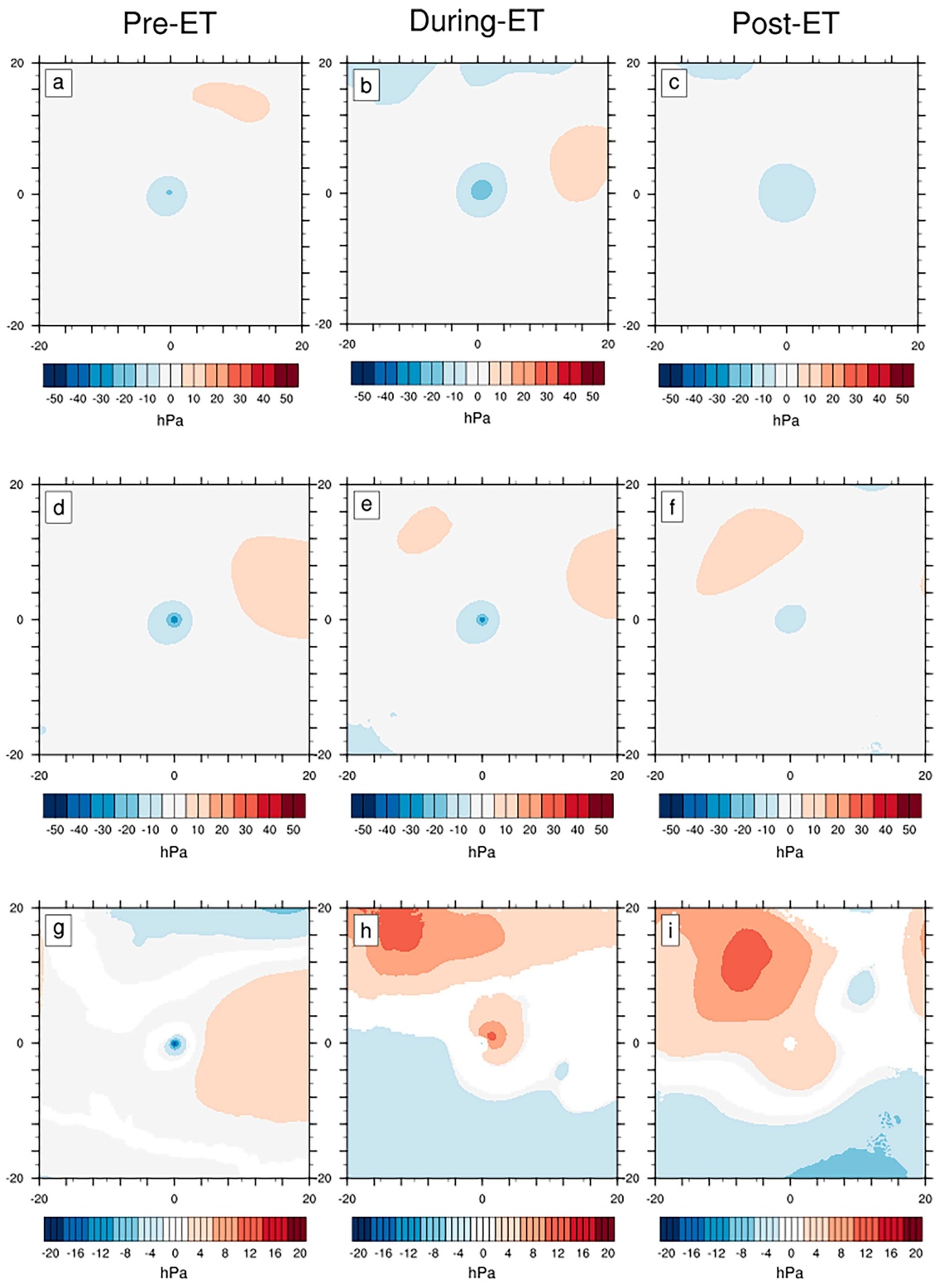

FIG. 10. As in Fig. 9, but for mean SLP (hPa).

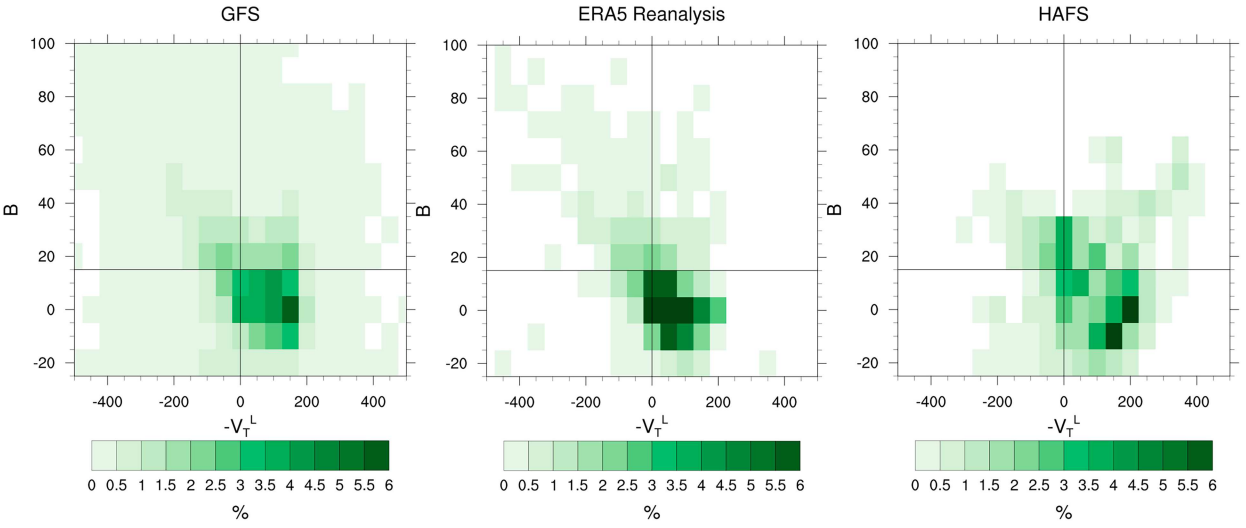

FIG. 11. Diagram showing the distribution of samples across a CPS as a percentage. (left) GFS simulations used in this study. (center) ERA5 values from the storms listed in Table 1. (right) HAFS simulations used in this study. The B values are sorted into bins of 10, while $-V_T^L$ values are grouped into bins of 50.

upstream midlatitude features (see Z500 field as shown in Figs. 7 and 9). While the models do not share the same errors in the Z500 field, they both produce errors that would alter the overall steering flow, which again could impact either the forecast occurrence of ET (or lack thereof) or phasing between the TC and surrounding features such as an upstream trough. Some of the differences between the models could be attributed to the different spatial resolutions, as grid spacing affects the representation of convective processes and TC outflow and in turn the interaction between the TC and its environment. While the poor HAFS ET forecasts indicate that the high spatial resolution is not necessarily important for the prediction of ET, with improved model physics, it may aid the simulation of rain rates due to better representation of convective processes.

Both GFS and HAFS underestimate the total accumulated rainfall from the storms studied when examining individual initialization lead times. Further investigation reveals that both models tend to overestimate rain rates in the inner core of the storms at individual time steps when compared to IMERG, which can sometimes underestimate rain rates in the inner cores of TCs as shown in Ko et al. (2020). The simulation of high rain rates over a small area still yields less accumulated rainfall than a larger area of more moderate rain rates, as seen in IMERG. For both models, the highest magnitude errors in rain rates occur in the during-ET phase, which appears to be attributed to errors in the placement and shape of a rapidly evolving precipitation shield.

The overestimation of rain rates by GFS and HAFS, particularly in the pre-ET phase, can be partially attributed to the overintensification of these storms by the model. Both the Z500 and SLP fields simulated by both models show a stronger storm in the simulations compared to the ERA5 reanalysis. The high rain rates that remain close to the storm center during and even after the ET process similarly can be partially attributed to the higher intensity of simulated storms (Jiang and Zipser 2010; Alvey et al. 2015). However, the tendency of the models to maintain a strong warm core, even well into the ET process, is likely causing the simulated storms to retain deep convection and high rain rates near the TC's inner core at times when the observed storm's rain rates are decreasing. These common discrepancies in both models could be related to the FV3 core model in general, as the dynamic core of models has been shown to influence TC intensity (Reed and Jablonowski 2011; Reed et al. 2015). Additionally, both GFS and HAFS use the GFDL microphysics scheme and utilize a scale-aware convection scheme. Further evaluation of the cause of the retention of a strong warm core is needed to determine any improvements that can be made to the operational models, whether involving the FV3 dynamical core or the model physics.

This analysis of the abilities of GFS and HAFS to simulate TCs that complete ET can be used for further model development in the future. For HAFS, the size of the regional domain is a primary concern; the outermost nest of the current operational HAFS could be expanded to include the entire globe. We hypothesize that this could improve the detection of ET potential at medium-range lead times. The high resolution of the model must be maintained around the storm to preserve model performance with storm intensity (as measured by winds or central pressure) as well as its lower errors compared to GFS in the pre-ET phase of the storm life cycle. This high resolution also aids in the representation of the smaller-scale processes that occur within a TC and can influence its interactions with large-scale features, such as the representation of deep convection and TC outflow (Alaka et al. 2022). In fact, multistorm applications with multiple moving nests have shown improvements to TC–environment interactions and, consequently, TC intensity predictions; innovations like these could be leveraged to study and improve ET forecasts as well.

Despite the strengths and weaknesses of each model, additional updates to the model parameterizations are likely needed to improve the representation of rain rates in both models, particularly in the deep convective inner core of the storm. The combined skills of both global and regional models would likely be ideal to simulate the proper trajectory, intensity, and rainfall of a TC throughout its entire lifetime.

Acknowledgments. Bower and Reed acknowledge the funding support of NASA under Grant 80NSSC19K0717 and NSF under AGS2244917, as well as the high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. The authors also would like to acknowledge Bin Liu, Zhan Zhang, and Matthew Green (NOAA Environmental Modeling Center and Weather Prediction Center) for providing the HAFS data for analysis, courtesy of the Hurricane Forecast Improvement Project (HFIP).

Data availability statement. The data that support the findings of this study are available at the following URLs: IMERG Final Run: https://gpm1.gesdisc.eosdis.nasa.gov/data/GPM_L3/GPM_3IMERGHH.06/, IBTrACS v4: https://www.ncdc.noaa.gov/ibtracs/index.php?name=ib-v4-access, ERA5 reanalysis: https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset, GFS forecast archive available from NCAR Research Data Archive: https://doi.org/10.5065/D65Q4TSG, HAFS data provided to Bower by Bin Liu, Zhan Zhang, and Matthew Green (NOAA EMC and WPC), TempestExtremes: https://github.com/ClimateGlobalChange/tempestextremes, and ExTraTrack: https://github.com/zarzycki/ExTraTrack.

REFERENCES

Alaka, G. J., Jr., X. Zhang, and S. G. Gopalakrishnan, 2022: High-definition hurricanes: Improving forecasts with storm-following nests. *Bull. Amer. Meteor. Soc.*, **103**, E680–E703, https://doi.org/10.1175/BAMS-D-20-0134.1.

Alvey, G. R., III, J. Zawislak, and E. Zipser, 2015: Precipitation properties observed during tropical cyclone intensity change. *Mon. Wea. Rev.*, **143**, 4476–4492, https://doi.org/10.1175/MWR-D-15-0065.1.

Avila, L. A., and J. P. Cangialosi, 2013: Tropical cyclone report: Hurricane Irene (21–28 August 2011). NHC Tech. Rep. AL092011, 45 pp., http://www.nhc.noaa.gov/data/tcr/AL092011_Irene.pdf.

Bachmann, K., and R. D. Torn, 2021: Validation of HWRF-based probabilistic TC wind and precipitation forecasts. *Wea. Forecasting*, **36**, 2057–2070, https://doi.org/10.1175/WAF-D-21-0070.1.

Balaguru, K., and Coauthors, 2020: Characterizing tropical cyclones in the Energy Exascale Earth System Model version 1. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002024, https://doi.org/10.1029/2019MS002024.

Bell, B., and Coauthors, 2021: The ERA5 global reanalysis: Preliminary extension to 1950. *Quart. J. Roy. Meteor. Soc.*, **147**, 4186–4227, https://doi.org/10.1002/qj.4174.

Beven, J. L., II, A. Hagen, and R. Berg, 2022: Tropical cyclone report: Hurricane Ida (26 August–1 September 2021). NOAA/NHC Tech. Rep. AL092021, 163 pp., https://www.nhc.noaa.gov/data/tcr/AL092021_Ida.pdf.

Bieli, M., S. J. Camargo, A. H. Sobel, J. L. Evans, and T. Hall, 2019: A global climatology of extratropical transition. Part I: Characteristics across basins. *J. Climate*, **32**, 3557–3582, https://doi.org/10.1175/JCLI-D-17-0518.1.

——, A. H. Sobel, S. J. Camargo, and M. K. Tippett, 2020: A statistical model to predict the extratropical transition of tropical cyclones. *Wea. Forecasting*, **35**, 451–466, https://doi.org/10.1175/WAF-D-19-0045.1.

Blake, E. S., T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Beven II, 2013: Tropical cyclone report: Hurricane Sandy (22–29 October 2012). NHC Tech. Rep. AL182012, 157 pp., https://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf.

Bourdin, S., S. Fromang, W. Dulac, J. Cattiaux, and F. Chauvin, 2022: Intercomparison of four algorithms for detecting tropical cyclones using ERA5. *Geosci. Model Dev.*, **15**, 6759–6786, https://doi.org/10.5194/gmd-15-6759-2022.

Bower, E., K. A. Reed, P. A. Ullrich, C. M. Zarzycki, and A. G. Pendergrass, 2022: Quantifying heavy precipitation throughout the entire tropical cyclone life cycle. *J. Hydrometeor.*, **23**, 1645–1662, https://doi.org/10.1175/JHM-D-21-0153.1.

Davis, C. A., and L. F. Bosart, 2004: The TT problem: Forecasting the tropical transition of cyclones. *Bull. Amer. Meteor. Soc.*, **85**, 1657–1662, https://doi.org/10.1175/BAMS-85-11-1657.

Dawson, L., A. Bentley, T. Dorian, and G. Manikin, 2018: Conclusion of the FV3GFS evaluation. NCEP/EMC Model Evaluation Group, https://vlab.noaa.gov/web/emc-model-evaluation-group/home/-/message_boards/category/3251785.

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*, **6**, 617, https://doi.org/10.3390/atmos11060617.

Dulac, W., J. Cattiaux, F. Chauvin, S. Bourdin, and S. Fromang, 2024: Assessing the representation of tropical cyclones in ERA5 with the CNRM tracker. *Climate Dyn.*, **62**, 223–238, https://doi.org/10.1007/s00382-023-06902-8.

Evans, C., and Coauthors, 2017: The extratropical transition of tropical cyclones. Part I: Cyclone evolution and direct impacts. *Mon. Wea. Rev.*, **145**, 4317–4344, https://doi.org/10.1175/MWR-D-17-0027.1.

Evans, J. L., J. M. Arnott, and F. Chairomonte, 2006: Evaluation of operational model cyclone structure forecasts during extratropical transition. *Mon. Wea. Rev.*, **134**, 3054–3072, https://doi.org/10.1175/MWR3236.1.

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.

Gopalakrishnan, S., and Coauthors, 2021: 2020 HFIP research and development activities summary: Recent results and operational implementation. HFIP Tech. Rep. HFIP2021-1, 49 pp., https://hfip.org/sites/default/files/documents/hfip-annual-report-2020-final_0.pdf.

Harris, L., X. Chen, W. Putman, L. Zhou, and J.-H. Chen, 2021: A scientific description of the GFDL finite-volume cubed-sphere dynamical core. NOAA Tech. Memo. OAR GFDL 2021-001, 109 pp., https://doi.org/10.25923/6nhs-5897.

Hart, R. E., 2003: A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon. Wea. Rev.*, **131**, 585–616, https://doi.org/10.1175/1520-0493(2003)131<0585:ACPSDF>2.0.CO;2.

——, and J. L. Evans, 2001: A climatology of the extratropical transition of Atlantic tropical cyclones. *J. Climate*, **14**, 546–564, https://doi.org/10.1175/1520-0442(2001)014<0546:ACOTET>2.0.CO;2.

Hazelton, A., 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, 2022: 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.

——, and Coauthors, 2023: 2022 real-time Hurricane forecasts from an experimental version of the Hurricane Analysis and Forecast System (HAFSV0.3S). *Front. Earth Sci.*, **11**, 1264969, https://doi.org/10.3389/feart.2023.1264969.

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

Huffman, G. J., E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and J. Tan, 2019: GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree × 0.1 degree V06. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 30 October 2022, https://doi.org/10.5067/GPM/IMERG/3B-HH/06.

——, D. T. Bolvin, E. J. Nelkin, E. F. Stocker, and J. Tan, 2020: V06 IMERG release notes. NASA GPM Note, 15 pp., https://gpm.nasa.gov/sites/default/files/2020-10/IMERG_V06_release_notes_201006_0.pdf.

Jensen, T., and Coauthors, 2020: Model evaluation tools version 9.0.1 user's guide. Developmental Testbed Center Doc., 482 pp., https://dtcenter.org/sites/default/files/community-code/met/docs/user-guide/MET_Users_Guide_v9.0.1.pdf.

Jiang, H., and E. J. Zipser, 2010: Contribution of tropical cyclones to the global precipitation from eight seasons of TRMM data: Regional, seasonal, and interannual variations. *J. Climate*, **23**, 1526–1543, https://doi.org/10.1175/2009JCLI3303.1.

Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: Forecast challenges, current understanding, and future directions. *Wea. Forecasting*, **18**, 1052–1092, https://doi.org/10.1175/1520-0434(2003)018<1052:TETOTC>2.0.CO;2.

Kehoe, R. M., M. A. Boothe, and R. L. Elsberry, 2007: Dynamical tropical cyclone 96- and 120-h track forecast errors in the western North Pacific. *Wea. Forecasting*, **22**, 520–538, https://doi.org/10.1175/WAF1002.1.

Keller, J. H., S. C. Jones, J. L. Evans, and P. A. Harr, 2011: Characteristics of the TIGGE multimodel ensemble prediction system in representing forecast variability associated with extratropical transition. *Geophys. Res. Lett.*, **38**, L12802, https://doi.org/10.1029/2011GL047275.

Knapp, K. R., H. Diamond, J. Kossin, M. C. Kruk, and C. J. Schreck III, 2018: International Best Track Archive for Climate Stewardship (IBTrACS) Project, version 4. NOAA National Centers for Environmental Information, accessed 20 June 2022, https://doi.org/10.25921/82ty-9e16.

Ko, M.-C., F. D. Marks, G. A. Alaka Jr., and S. G. Gopalakrishnan, 2020: Evaluation of Hurricane Harvey (2017) rainfall in deterministic and probabilistic HWRF forecasts. *Atmosphere*, **11**, 519–538, https://doi.org/10.3390/atmos11060666.

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.

Leonardo, N. M., and B. A. Colle, 2017: Verification of multimodel ensemble forecasts of North Atlantic tropical cyclones. *Wea. Forecasting*, **32**, 2083–2101, https://doi.org/10.1175/WAF-D-17-0058.1.

——, and ——, 2020: An investigation of large along-track errors in extratropical transitioning North Atlantic tropical cyclones in the ECMWF ensemble. *Mon. Wea. Rev.*, **148**, 457–476, https://doi.org/10.1175/MWR-D-19-0044.1.

Liu, B., and Coauthors, 2020: The HAFSv0.1A HFIP real time parallel experiment: A regional and ocean-coupled configuration. HFIP Doc., 23 pp., https://hfip.org/sites/default/files/events/65/100-liu-hafsv01a.pdf.

——, and Coauthors, 2021: The HAFSv0.2A regional coupled HFIP real time parallel experiment. HFIP Doc., 19 pp., https://hfip.org/sites/default/files/events/269/1200-liu-hafsv02apptx.pdf.

Lonfat, M., F. D. Marks Jr., and S. S. Chen, 2004: Precipitation distribution in tropical cyclones using the Tropical Rainfall Measuring Mission (TRMM) microwave imager: A global perspective. *Mon. Wea. Rev.*, **123**, 1645–1660, https://doi.org/10.1175/1520-0493(2004)132<1645:PDITCU>2.0.CO;2.

Manikin, G., A. Bentley, L. Dawson, and T. Dorian, 2019: Implementation of GFSv15. NCEP/EMC Model Evaluation Group, https://vlab.noaa.gov/web/emc-model-evaluation-group.

——, ——, S. Shields, P. Papin, L. Dawson, and C. Macintosh, 2020: GFSv16 evaluation kickoff. NCEP/EMC Model Evaluation Group, https://vlab.noaa.gov/web/emc-model-evaluation-group.

——, ——, ——, C. MacIntosh, L. Dawson, M. Caron, and P. Papin, 2021: Review of the GFSv16 upgrade. NCEP/EMC Model Evaluation Group, https://vlab.noaa.gov/web/emc-model-evaluation-group.

Mehra, A., and Z. Zhang, 2023: 2023 HAFS v1 implementation decisional briefing.

Morss, R. E., J. Vickery, H. Lazrus, J. Demuth, and A. Bostrom, 2022: Improving tropical cyclone forecast communication by understanding NWS partners' decision timelines and forecast information needs. *Wea. Climate Soc.*, **14**, 783–800, https://doi.org/10.1175/WCAS-D-21-0170.1.

NCEP, 2007: NCEP Global Forecast System (GFS) analyses and forecasts. National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 30 April 2023, https://doi.org/10.5065/D65Q4TSG.

Reed, K. A., and C. Jablonowski, 2011: Assessing the uncertainty in tropical cyclone simulations in NCAR's Community Atmosphere Model. *J. Adv. Model. Earth Syst.*, **3**, M08002, https://doi.org/10.1029/2011MS000076.

——, J. T. Bacmeister, N. A. Rosenbloom, M. F. Wehner, S. C. Bates, P. H. Lauritzen, J. E. Truesdale, and C. Hannay, 2015: Impact of the dynamical core on the direct simulation of tropical cyclones in a high-resolution global model. *Geophys. Res. Lett.*, **42**, 3603–3608, https://doi.org/10.1002/2015GL063974.

Rios Gaona, M. F., G. Villarini, W. Zhang, and G. A. Vecchi, 2018: The added value of IMERG in characterizing rainfall in tropical cyclones. *Atmos. Res.*, **209**, 95–102, https://doi.org/10.1016/j.atmosres.2018.03.008.

Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J. Climate*, **25**, 3453–3475, https://doi.org/10.1175/2011JCLI4208.1.

Skok, G., J. Tribbia, J. Rakovec, and B. Brown, 2009: Object-based analysis of satellite-derived precipitation systems over the low- and midlatitude Pacific Ocean. *Mon. Wea. Rev.*, **137**, 3196–3218, https://doi.org/10.1175/2009MWR2900.1.

Stackhouse, S. D., S. E. Zick, C. J. Matyas, K. M. Wood, A. T. Hazelton, and G. J. Alaka Jr., 2023: Evaluation of experimental high-resolution model forecasts of tropical cyclone precipitation using object-based metrics. *Wea. Forecasting*, **38**, 2111–2134, https://doi.org/10.1175/WAF-D-22-0223.1.

Stansfield, A. M., and K. A. Reed, 2023: Global tropical cyclone precipitation scaling with sea surface temperature. *npj Climate Atmos. Sci.*, **6**, 60, https://doi.org/10.1038/s41612-023-00391-6.

——, ——, C. M. Zarzycki, P. A. Ullrich, and D. R. Chavas, 2020: Assessing tropical cyclones' contribution to precipitation over the eastern United States and sensitivity to the variable-resolution domain extent. *J. Hydrol.*, **21**, 1425–1445, https://doi.org/10.1175/JHM-D-19-0240.1.

Wang, Y., and Z. Pu, 2021: Assimilation of radial velocity from coastal NEXRAD into HWRF for improved forecasts of landfalling hurricanes. *Wea. Forecasting*, **36**, 587–599, https://doi.org/10.1175/WAF-D-20-0163.1.

Wu, Q., Z. Ruan, D. Chen, and T. Lian, 2015: Diurnal variations of tropical cyclone precipitation in the inner and outer rainbands. *J. Geophys. Res. Atmos.*, **120** (1), 1–11, https://doi.org/10.1002/2014JD022190.

Zarzycki, C. M., D. R. Thatcher, and C. Jablonowski, 2017: Objective tropical cyclone extratropical transition detection in high-resolution reanalysis and climate model data. *J. Adv. Model. Earth Syst.*, **9**, 130–148, https://doi.org/10.1002/2016MS000775.

Zawislak, J., 2020: Global survey of precipitation properties observed during tropical cyclogenesis and their differences compared to nondeveloping disturbances. *Mon. Wea. Rev.*, **148**, 1585–1606, https://doi.org/10.1175/MWR-D-18-0407.1.