BAMS Article

Lifetime Performance of the Operational Hurricane Weather Research and Forecasting Model (HWRF) for North Atlantic Tropical Cyclones

Ghassan J. Alaka Jr.,[a] Jason A. Sippel,[a] Zhan Zhang,[b] Hyun-Sook Kim,[c] Frank D. Marks,[a] Vijay Tallapragada,[b] Avichal Mehra,[b] Xuejin Zhang,[a] Aaron Poyer,[d] and Sundararaman G. Gopalakrishnan[a]

KEYWORDS: Extreme events; Hurricanes/typhoons; Tropical cyclones; Short-range prediction; Model evaluation/performance; Numerical weather prediction/forecasting

ABSTRACT: The Hurricane Weather Research and Forecasting Model (HWRF) was the flagship hurricane model at NOAA's National Centers for Environmental Prediction for 16 years and a state-of-the-art tool for tropical cyclone (TC) intensity prediction at the National Weather Service and across the globe. HWRF was a joint development between NOAA research and operations, specifically the Environmental Modeling Center and the Atlantic Oceanographic and Meteorological Laboratory. Significant support also came from the National Hurricane Center, Developmental Testbed Center, University Corporation for Atmospheric Research, universities, cooperative institutes, and the TC community. In the North Atlantic basin, where most improvement efforts focused, HWRF intensity forecast errors decreased by 45%–50% at many lead times between 2007 and 2022. These large improvements resulted from increases in horizontal and vertical resolution, as well as advances in model physics and data assimilation. HWRF intensity forecasts performed particularly well over the Gulf of Mexico in recent years, providing useful guidance for a large number of impactful landfalling hurricanes. Such advances were made possible not only by significant gains in computing but also through substantial investment from the Hurricane Forecast Improvement Program.

SIGNIFICANCE STATEMENT: HWRF is considered one of the most important and successful ocean-coupled, regional numerical weather prediction models. This work documents HWRF forecast performance for TCs in the North Atlantic since its operational inception (2007–22). Track, intensity, and storm size (wind radii) forecasts all showed significant error reductions and skill improvements, including 45%–50% lower intensity errors. HWRF intensity forecasts showed marked improvement for rapid intensification events, especially in the Gulf of Mexico, where numerous impactful TCs made landfall throughout HWRF's lifetime. HWRF's impressive performance led to its expanded use as guidance for official TC forecasts and as a research tool across the TC community. These successes paved the way for NOAA's next-generation hurricane model, the Hurricane Analysis and Forecast System.

DOI: 10.1175/BAMS-D-23-0139.1

Corresponding author: Ghassan J. Alaka, ghassan.alaka@noaa.gov

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/BAMS-D-23-0139.s1.

Manuscript received 6 June 2023, in final form 7 March 2024, accepted 8 March 2024

AFFILIATIONS: [a] NOAA/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida; [b] NOAA/National Weather Service/National Centers for Environmental Prediction/Environmental Modeling Center, College Park, Maryland; [c] NOAA/Atlantic Oceanographic and Meteorological Laboratory/Physical Oceanography Division, Miami, Florida; [d] NOAA/National Weather Service/Office of Science and Technology Integration/Modeling Program Division, Silver Spring, Maryland

1. Introduction (the rise of dynamical hurricane models at NOAA)

Tropical cyclone (TC) forecast errors have markedly decreased over the past two decades, largely due to improved numerical weather prediction capabilities. For example, since the National Hurricane Center (NHC) of the National Oceanic and Atmospheric Administration (NOAA) began issuing 5-day TC forecasts in 2001, their track forecast errors have decreased by 55%–65% (Cangialosi 2023, Fig. 3). This improvement continues a well-known trend that has persisted for over half a century.[1] Landsea and Cangialosi (2018) showed that track forecast improvements have been linked with advances in guidance and consensus techniques. NHC intensity forecast errors have also decreased by 25%–40% since 2001 (Cangialosi 2023, Fig. 10), though much of that improvement came after NOAA established the Hurricane Forecast Improvement Project[2] (HFIP; Gall et al. 2013; Gopalakrishnan et al. 2021) in 2008. As discussed by DeMaria et al. (2014), the HFIP era marked the first time that intensity guidance became sufficiently skillful to strongly influence official forecasts issued by the NHC. They attributed the improvements to a movement away from simple climatology and persistence models to statistical–dynamical and dynamical models as well as consensus techniques.

[1] Also see the NHC verification page (https://www.nhc.noaa.gov/verification/), updated annually with the latest statistics.

[2] The Hurricane Forecast Improvement Project was transitioned to the Hurricane Forecast Improvement Program to address section 104 of the Weather Research Forecasting Innovation Act of 2017.

Cangialosi et al. (2020) showed that dynamical models have become increasingly useful to forecasters at the NHC as guidance for TC intensity prediction in the North Atlantic (NATL) basin. One of the models highlighted in their study was the NOAA/National Weather Service (NWS)/National Centers for Environmental Prediction (NCEP) HWRF (Gopalakrishnan et al. 2011, 2012, 2013; Bao et al. 2012; Tallapragada et al. 2014; Atlas et al. 2015; Mehra et al. 2018). In particular, Fig. 4 in Cangialosi et al. (2020) compared 48-h NATL intensity forecast skill among operational models available annually from 1996 to 2019. During HWRF's early years, statistical models soundly beat its generally unskillful intensity forecasts. With one exception in 2006, the logistic growth equation model (LGEM) and the Decay-Statistical Hurricane Intensity Prediction Scheme (DSHP) statistical models were the top annual performers for 48-h NATL intensity forecasts through 2011. In 2012, the Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model performed best, and then HWRF performed best for the first time in 2013. DSHP again outperformed other models in 2014. From 2015 to 2022, HWRF at last had the highest NATL intensity skill of any individual model in all except for 1 year (2020; see Cangialosi 2021).

HWRF produced its first operational forecast at 1200 UTC 19 June 2007 (Fig. 1) and was continually improved through 2022.[3] The development team included NOAA's Environmental Modeling Center (EMC), the Atlantic Oceanographic and Meteorological Laboratory (AOML), GFDL, the Developmental Testbed Center (DTC), and

[3] HWRF continued running in a limited capacity for select storms in 2023 and 2024 to facilitate the addition of new operational hurricane models.

NOUS41 KWBC 141245
PNSWSH

TECHNICAL IMPLEMENTATION NOTICE 07-40
NATIONAL WEATHER SERVICE HEADQUARTERS WASHINGTON DC
845 AM EDT THU JUN 14 2007

TO: FAMILY OF SERVICES /FOS/ SUBSCRIBERS
NOAA WEATHER WIRE SERVICE /NWWS/ SUBSCRIBERS
EMERGENCY MANAGERS WEATHER INFORMATION NETWORK /EMWIN/
SUBSCRIBERS
NOAAPORT SUBSCRIBERS
OTHER NWS PARTNERS...USERS AND EMPLOYEES

FROM: PAUL HIRSCHBERG
CHIEF...SCIENCE PLANS BRANCH
OFFICE OF SCIENCE AND TECHNOLOGY

SUBJECT: NATIONAL CENTERS FOR ENVIRONMENTAL PREDICTION /NCEP/ TO IMPLEMENT HURRICANE WEATHER RESEARCH AND FORECAST /HWRF/ MODEL: EFFECTIVE JUNE 19 2007

EFFECTIVE TUESDAY JUNE 19 2007...BEGINNING WITH THE 1200 COORDINATED UNIVERSAL TIME /UTC/ RUN...NCEP WILL BEGIN RUNNING A NEW HURRICANE MODEL...THE HWRF SYSTEM...IN ADDITION TO THE GEOPHYSICAL FLUID DYNAMICS LABORATORY /GFDL/ HURRICANE MODEL.

THE HWRF IS THE NOAA NEXT GENERATION OPERATIONAL HURRICANE FORECAST SYSTEM. IT HAS BEEN DEVELOPED TO PROVIDE ADVANCED HURRICANE GUIDANCE TO THE NATIONAL HURRICANE CENTER /NHC/ AT THE TROPICAL PREDICTION CENTER /TPC/. IT WILL PROVIDE FOR CONTINUED IMPROVEMENT OF TRACK FORECASTS AND IMPROVED FORECASTS OF INTENSITY AND STRUCTURE FORECASTS.

THE CODING INFRASTRUCTURE OF THE HWRF FOLLOWS THE 2000 NWS STRATEGIC PLAN WHICH CALLED FOR THE USE OF A COMMUNITY MODEL - I.E...THE WEATHER RESEARCH AND FORECASTING /WRF/ MODELING INFRASTRUCTURE. THE HWRF SHARES THE SAME DYNAMIC CORE AS THE NCEP MIDLATITUDE MESOSCALE WRF BASED SYSTEM...THE NORTH AMERICAN MODEL /NAM/...BUT IS SPECIFICALLY DESIGNED FOR THE HURRICANE PROBLEM TO INCLUDE AN ADVANCED ANALYSES OF THE HURRICANE CORE CIRCULATION...A MOVABLE NESTED GRID CONFIGURATION SIMILAR TO THE GFDL MODEL AND COUPLING TO THE OCEAN.

ON THE ABOVE IMPLEMENTATION DATE...NCEP CENTRAL OPERATIONS WILL BEGIN TO PROVIDE OUTPUT FROM THE HWRF MODEL ON ITS FILE TRANSFER PROTOCOL /FTP/ SERVER. USERS MAY ACCESS THESE DATA FROM THE LINK BELOW /USE LOWER CASE/:

FTP://FTPPRD.NCEP.NOAA.GOV/PUB/DATA/COM/HUR/PROD/HWRF.YYYYMMDD

Fig. 1. The NOAA TIN was released at 0845 eastern daylight time (EDT) 14 Jun 2007, highlighting the first operational implementation of HWRF at 1200 UTC 19 Jun 2007.

several academic institutions. The roadmap for community support and transitions of research to operations is discussed in more detail in Bernardet et al. (2015). As a result of this investment, HWRF became the premier operational and research tool for TC prediction.

This study examines HWRF's evolution and operational forecast errors in the NATL basin since its implementation in 2007. A major motivation is to illustrate the increasing importance of dynamical guidance, especially high-resolution, regional numerical weather prediction models, to TC prediction as NOAA enters the next generation of hurricane forecasting with the Hurricane Analysis and Forecast System (HAFS; Dong et al. 2020; Hazelton et al. 2021, 2022). We concentrate on the NATL basin for two main reasons: 1) the majority of operational and developmental HWRF resources were focused there and 2) the HWRF

configuration was degraded and/or ran for a shorter period in other basins (see sidebar). These differences made it difficult to fairly compare HWRF performance in the NATL against other basins. This work proceeds as follows: section 2 introduces a brief history of HWRF improvements; section 3 describes verification methods and considerations; section 4 reports the results, including verification trends; and section 5 contains a discussion and concluding remarks.

2. HWRF development history

HWRF became operational in 2007 after 5 years of development. Many of the initial components of HWRF were modeled after the operational GFDL hurricane model (Kurihara et al. 1998; Bender et al. 2007) with the addition of several capabilities, including forecast cycling (see below) and data assimilation (DA; Gopalakrishnan et al. 2010; Tallapragada 2016). The first NOAA Technical Information Notice (TIN; see Fig. 1) released for HWRF described the basic framework and initial developmental history of the model.

The initial operational configuration of HWRF contained general infrastructure that persisted through its use as an operational model. This included the following:

1) Locating the HWRF domains (both the parent and moving nests) based on the TC center position
2) Interpolating the analyzed NCEP Global Forecast System (GFS) model fields onto the HWRF parent domain and applying GFS as lateral boundary conditions
3) Removing the global model vortex and inserting a mesoscale vortex obtained from the HWRF 6-h forecast from the previous cycle (if available; referred to as "cycling") or from a synthetic vortex
4) Updating the location and structure of the vortex to match the latest observations through vortex initialization, including relocation and modification/adjustment
5) Accounting for new observations through DA
6) Simulating TCs with a high-resolution nest
7) Interacting dynamically with a three-dimensional ocean model

The above infrastructure was initially configured with a parent domain that spanned ~80° × ~80° (~27-km horizontal grid spacing) and a single, two-way interactive, storm-following nest that spanned ~6° × ~6° (~9-km grid spacing). Further, the initial DA system used a pure three-dimensional variational data assimilation (3DVAR) implementation that only assimilated large-scale data in the moving nest.

Many evidence-based upgrades of high utility transformed HWRF into a state-of-the-art TC prediction system. In the following summary, we focus on the more significant upgrades[4] made to the NATL configuration. More detailed information about HWRF upgrades through 2018 can be found in Biswas et al. (2018). In general, the major upgrades can be categorized as pertaining to domain configuration and dynamics (Fig. 2, left), physics (Fig. 2, right), and DA (Fig. 3).

[4] Significant upgrades referenced mostly came from NCEP Technical Information Notices and service change notices that accompanied upgrades. These notices can be found at https://www.weather.gov/notification/archive.

The HWRF grid configuration and dynamics dramatically evolved after 2010 (Fig. 2). A high-resolution, convection-allowing ~3-km moving nest was added in 2012 that resided telescopically within the preexisting ~9-km moving nest[5] (Tallapragada et al. 2014; Goldenberg et al. 2015; X. Zhang et al. 2016). Subsequent horizontal resolution increases brought the HWRF grid spacing for the parent domain and two moving nests to ~18/6/2 km in 2015 and ~13.5/4.5/1.5 km in 2018. The vertical resolution likewise increased from 43 to 61 levels in 2014 and to 75 levels in 2017. The storm-following domains

[5] For more information about HWRF's telescopic outer- and inner-moving nests, refer to Fig. 2 in X. Zhang et al. (2016).

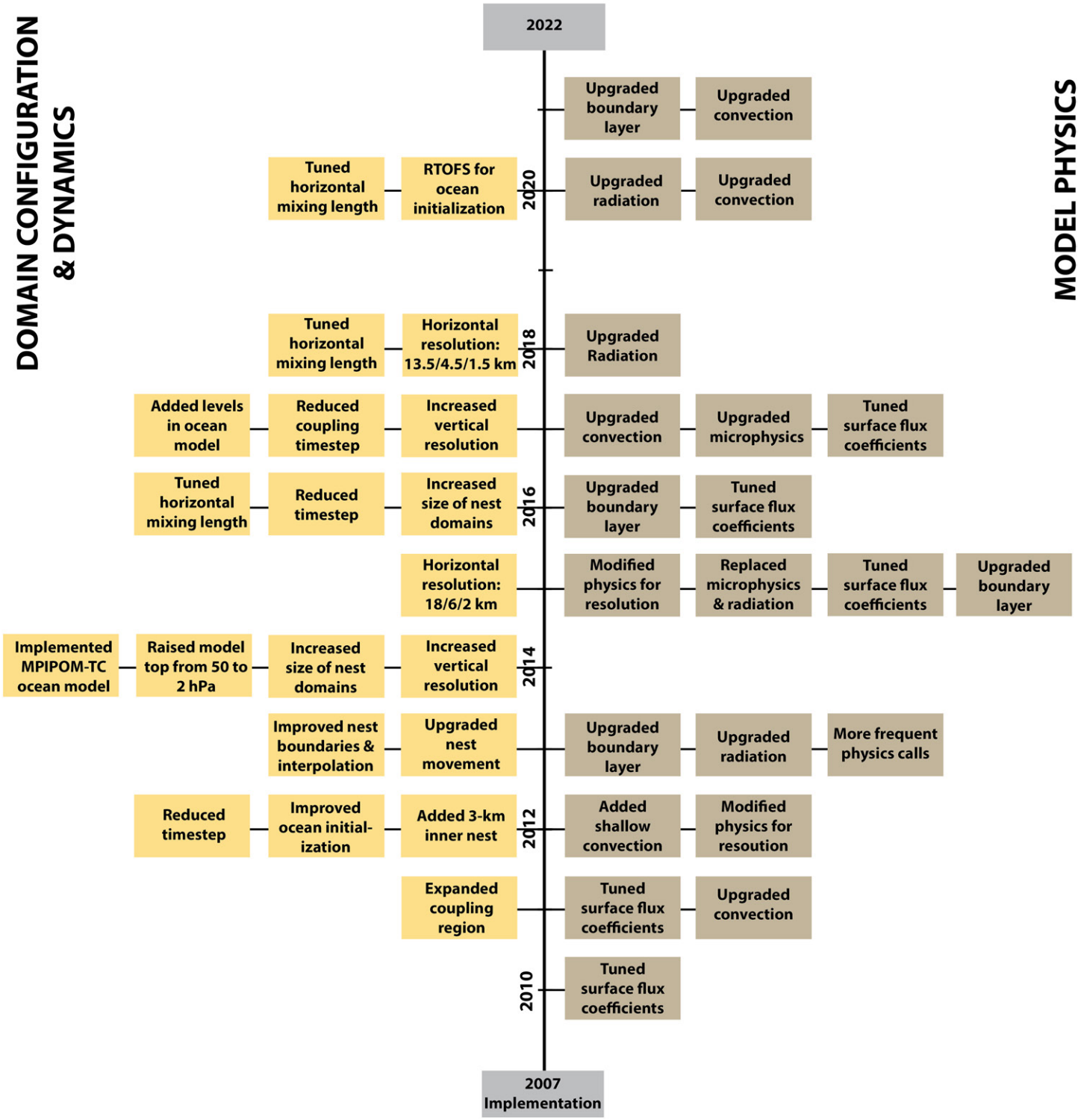

Fig. 2. Upgrades to (left) the HWRF grid configuration, dynamics, and ocean coupling and (right) HWRF physics.

also grew in size in 2013, 2014, and 2016. These upgrades allowed better prediction of sharp horizontal and vertical gradients and, consequently, the convective evolution of a TC. Previous work showed that major upgrades to the moving nests coincided with notable increases in TC intensity forecast skill (Alaka et al. 2020, 2022). Commensurate with resolution increases, HWRF dynamics also advanced (Fig. 2). The dynamic time step significantly decreased in both 2012 (Goldenberg et al. 2015) and 2016. Following the work of

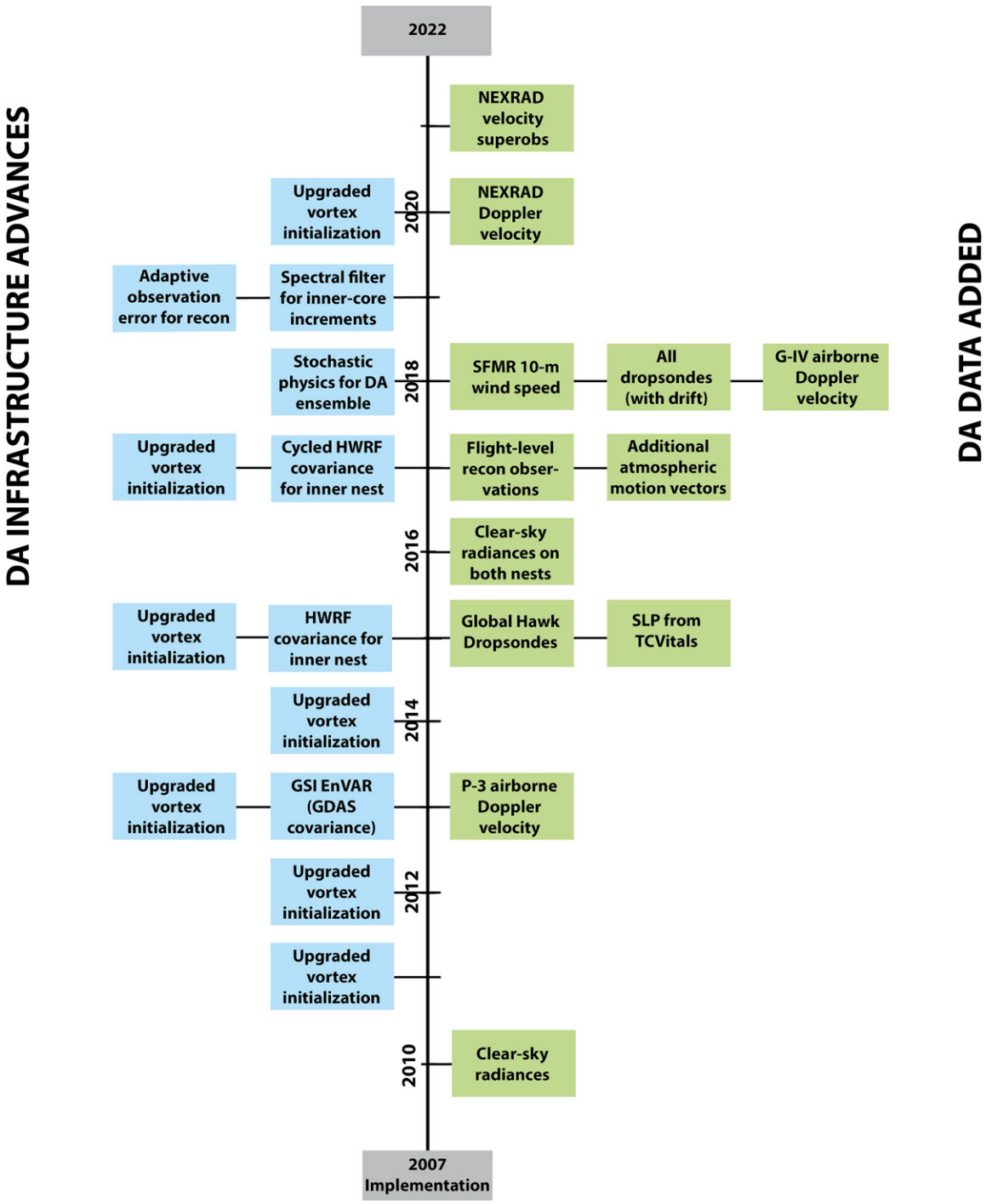

Fig. 3. HWRF DA upgrades, including upgrades to the (left) system infrastructure and (right) data assimilated.

Zhang and Marks (2015), horizontal diffusion was first modified in 2016 (Zhang et al. 2018) and then again in 2018 and 2020 to be consistent with model resolution. The formulation for horizontal mixing length was also updated in 2017 to account for local winds and horizontal wind gradients (Wang et al. 2021).

As shown in Fig. 2, physics parameterization schemes were also upgraded significantly since 2010, especially to improve performance at increasingly finer horizontal, vertical, and temporal resolutions.[6] The frequency of all model physics calls

[6] Please refer to Biswas et al. (2018) for comprehensive technical details and references for many of these upgrades.

increased significantly in 2013 (Goldenberg et al. 2015) to better capture tendencies associated with the rapid evolution of TC convection in the 3-km moving nest (Zhang et al. 2011). HWRF benefited immensely from upgrades to GFS physics, which motivated many updates to convection, radiation, microphysics, surface exchanges, and land surface in the HWRF physics suite. One such upgrade was the introduction of "scale awareness" in the simplified Arakawa–Schubert convection scheme by reducing cloud mass flux with increasing resolution (Han et al. 2017), leading to the improved prediction of convection across multiple scales in HWRF's three domains.

Many critical physics upgrades in HWRF focused on the planetary boundary layer (PBL). In general, the PBL scheme was modified to better represent smaller scales while remaining consistent with large-scale fields from GFS (e.g., Gopalakrishnan et al. 2013; Wang et al. 2018; Zhang et al. 2015; J. A. Zhang et al. 2017). The GFS PBL scheme used in HWRF's early years relied on the *K*-profile parameterization (KPP) in a well-mixed boundary layer, including the hurricane boundary layer. This scheme, also adopted by community research models such as MM5 and WRF, faced limitations in TC applications (Gopalakrishnan et al. 2013). By introducing an alpha (α) parameter to the nonlocal *K* approach, the match between modeled and observed eddy diffusivities improved. A reduction in eddy diffusivity aligned HWRF more closely with observations and reproduced the TC spinup mechanism in idealized simulations. As such, HWRF forecasts of intensity and storm size improved significantly when the modified scheme ($\alpha = 0.50$) was implemented into operations (e.g., Tallapragada et al. 2014). This work set the stage for several additional PBL improvements, including an upgrade to an eddy-diffusivity mass-flux scheme in 2016 (Han et al. 2016).

Alongside atmospheric model improvements, the ocean model and atmosphere–ocean coupling in HWRF also advanced significantly (Fig. 2). From 2007 to 2013, the HWRF atmospheric model was coupled to the feature-based Princeton Ocean Model (POM) for TCs with the Naval Oceanographic Office's 1/4° generalized digital environmental model (GDEM) temperature and salinity for the initial and boundary conditions (Yablonsky and Ginis 2008; Yablonsky et al. 2015). This ocean component was similar to the GFDL hurricane model, with initialization procedures that employed feature-based DA to add more realistic mesoscale variability to the climatology. The initialization focused on correcting the locations and the cross-frontal shape of features, especially in the Gulf of Mexico. Note that the shapes of the ocean features were determined subjectively by operational forecasters and then manually digitized. This was followed by assimilating 1/12° NCEP daily real-time global sea surface temperature high-resolution analysis (RTG_SST_HR; Gemmill et al. 2007) to modify the upper oceanic temperature field. The final initial condition was the product of a 2-day integration of POM to achieve dynamic adjustment, followed by a 3-day integration with the observed surface wind point observations provided by the NHC to simulate the TC cold wake and surface currents.

A message passing interface (MPI) version of POM for TCs (MPIPOM-TC) with 1/12° horizontal resolution and 24 vertical levels was implemented in 2014 to advance predictions and efficiency. MPIPOM-TC was coupled to the atmospheric parent domain and outer moving nest, and dynamically updated sea surface temperatures were downscaled to the inner-moving nest. In 2017, MPIPOM-TC adopted 41 vertical levels to be consistent with the Real-Time Ocean Forecast System (RTOFS; Mehra and Rivin 2010). Finally, a subset of the 41-layer RTOFS nowcasts replaced the NATL GDEM initialization in 2020.

There were also major changes to the atmospheric DA system since the earliest days of HWRF (Fig. 3). The vortex initialization procedure that provided the first guess for atmospheric DA was upgraded most years (Liu et al. 2020). This procedure updated the TC by first relocating its vortex to the observed position and then modifying its structure to match the

observed intensity and size (i.e., vortex adjustment). Strictly speaking, vortex adjustment and relocation are separate procedures, but they were always both employed in HWRF. Lu et al. (2017a) demonstrated that correctly positioning the first-guess vortex is critical when assimilating inner-core data, and not doing so can lead to unphysical analysis increments. Despite some shortcomings, vortex adjustment also improved both track and intensity forecasts (Tong et al. 2018). Vortex initialization is also useful to produce reasonable initial conditions in the absence of inner-core and/or aircraft reconnaissance data.

NCEP made many other upgrades to HWRF DA as well. The first substantial upgrade to the DA system itself was a change from 3DVAR to three-dimensional ensemble variational data assimilation (3DEnVar) in 2013, which introduced flow-dependent covariance from Global Data Assimilation System (GDAS) ensemble members. Using flow-dependent covariance instead of static covariance allows the DA system to more accurately analyze the structure of the TC vortex, especially when assimilating within the inner core (e.g., Lu et al. 2017b). Despite this advancement, the GDAS covariance struggled to represent background errors in hurricanes since the global model resolution is too coarse to faithfully simulate the inner core of stronger TCs. This issue in part led to unphysical analysis updates and subsequent spindown[7] when inner-core data were assimilated into hurricanes (Tong et al. 2018). In response to this problem, NCEP began completely removing low-level analysis increments within 150 km of the TC center in 2013. Initially, the increments were removed when the maximum 10-m winds exceeded 50 kt (1 kt ≈ 0.51 m s^{-1}), but that threshold increased to 60 kt in 2017.

[7] Subjectively, spindown is a rapid decrease in the forecast vortex strength during the first 12 h. Objectively, spindown is a decrease in maximum 10-m winds of 20 kt or more during that period.

Further DA system upgrades ultimately allowed inner-core data to have a greater positive impact on the analysis. The 2015 upgrade allowed HWRF inner-nest DA to use mesoscale covariance from a 6-h HWRF ensemble initialized off GDAS perturbations and was a first step toward a more advanced DA system. In 2017, NCEP finally implemented an ensemble Kalman filter (EnKF) that provided cycled storm-scale covariance for 3DEnVar in the innermost nest. The configuration of this system was very similar to the experimental system described in Lu et al. (2017a,b). This upgrade allowed for more physically realistic analysis updates when assimilating inner-core data (e.g., Lu et al. 2017b), though low-level, inner-core increments were still removed for the time being. Stochastic physics was added to this configuration in 2018, increasing the EnKF spread that was previously deficient (Z. Zhang et al. 2020). A further upgrade in 2019 changed how HWRF handled analysis increments in the inner core of TCs. In particular, a spectral filter was implemented to remove inner-core increments based on their wavenumber instead of removing all increments.[8] Commensurate with that upgrade, HWRF began treating aircraft reconnaissance data with adaptive observation errors that increased in the inner core. Together, these upgrades allowed for more data impact while reducing the potential for spindown.

[8] The spectral filter was applied when the storm intensity exceeded 50 kt. Analysis increments were converted into a polar coordinate system, increments exceeding a threshold wavenumber were truncated, and then the remaining increments were added to the first guess. For TCs between 50 and 60 kt, increments with a wavenumber > 1 were rejected, and for TCs ≥ 65 kt, only wavenumber 0 was retained. The truncated increments were applied within 150 km of the TC, and they relaxed to full increments at a radius of 300 km.

The advanced configuration with a cycled EnKF was computationally expensive to run, and sufficient operational resources only ever existed to run it for one storm at a time. If multiple storms were simultaneously active, the EnKF configuration was reserved for the storm with aircraft reconnaissance data or the largest threat to land or shipping, and the remaining storms used GDAS covariance for DA. Practically, activation or deactivation of this advanced configuration was sent by a data file through a web interface after consultation between Hurricane Research Division (HRD), EMC, and NHC.

Concomitant with DA system improvements, HWRF continuously increased the use of both satellite and reconnaissance data (Fig. 3, right). Satellite data usage was systematically

expanded to maintain consistency with GFS. In 2017, HWRF also began assimilating several new types of atmospheric motion vectors (AMVs) that were not assimilated in GFS (Lim et al. 2019). Most recently, the 2019 HWRF upgrade added GOES-R AMVs. Meanwhile, the infrastructure to assimilate reconnaissance data from the inner core was built into HWRF in 2013 (Tong et al. 2018), and tail Doppler radar data from the NOAA P-3 were assimilated starting that same year. HWRF began assimilating flight-level high-density observations (HDOBs) in the 2017 upgrade and stepped frequency microwave radiometer (SFMR) wind speed observations in 2018. Additional observations added in 2018 included tail Doppler radar data from the NOAA G-IV as well as all dropsonde data with locations that accounted for horizontal motion as described in Aberson et al. (2017). Prior to that time, dropsonde data were assimilated in a single vertical profile with an inaccurate location obtained from TEMP DROP[9] messages, and dropsonde winds within several degrees of the TC center were not used. With these upgrades, HWRF became the only operational model in the world to make use of all NOAA reconnaissance data transmitted in real time. Figure 4 provides an example of the reconnaissance data assimilated in a single well-sampled cycle of Hurricane Ian (2022). In 2020, HWRF also began assimilating high-resolution radial velocity data from the U.S. NEXRAD network (Wang and Pu 2021).

[9] For more information, please refer to https://www.aoml.noaa.gov/hrd/format/tempdrop_format.html.

Several external factors have impacted HWRF over the years. First, the GFS, which provides initial and boundary conditions in addition to error covariance for the HWRF DA system, has also significantly evolved since HWRF's initial implementation. As with HWRF, GFS TC track and intensity errors dramatically reduced due to improvements in the model and DA, including improved assimilation of TC reconnaissance data (e.g., Sippel et al. 2022). Further, the number of NATL TC reconnaissance missions from NOAA aircraft substantially increased after 2015. Given the known benefits of assimilating reconnaissance data (Zawislak et al. 2022 and references therein), these additional flights likely also improved forecasts.

Note that while the greatest efforts in HWRF development concentrated on the NATL, HWRF did run in other basins. See the sidebar "HWRF Forecasts Anywhere in the World" for information about the progress of HWRF elsewhere across the globe.

3. Verification methodology

Verification against the NHC best track (Rappaport et al. 2009; Cangialosi et al. 2020) was performed according to NHC standard procedures (Landsea and Franklin 2013). A forecast was verified only if the system was designated as a tropical or subtropical cyclone at the forecast initialization time and the verification time. In other words, if a system did not have a tropical or subtropical designation at a particular time, verification

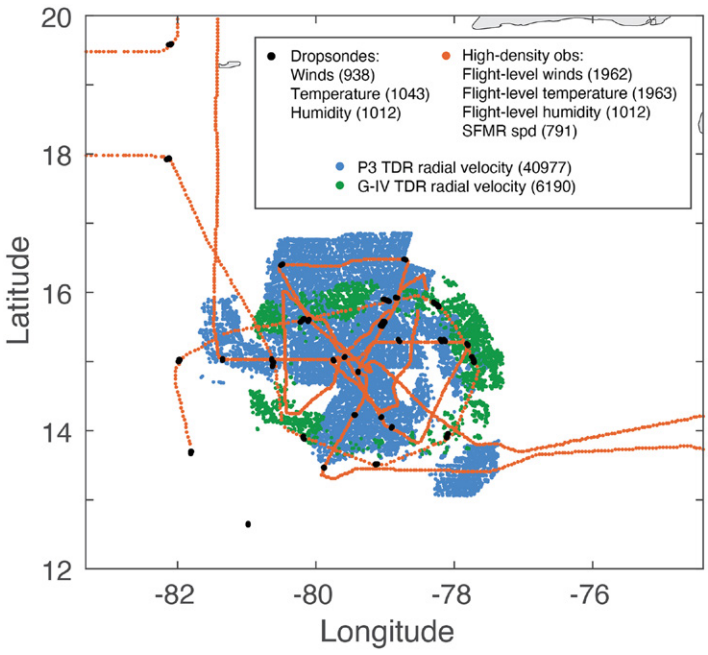

Fig. 4. Example of aircraft reconnaissance data assimilated into HWRF for a single cycle of Hurricane Ian (2022). Data are shown for the cycle initialized at 1200 UTC 25 Sep 2022 and include data from 0900 to 1500 UTC. The legend provides observation sources and the number of observations assimilated for each type.

HWRF Forecasts Anywhere in the World

Important HWRF developments occurred in worldwide basins beyond the NATL. From 2007 to 2013, NCEP ran HWRF for up to five storms between the NATL and eastern North Pacific (EPAC) basins. As part of HFIP's HREx, HWRF forecasts for western North Pacific (WPAC) TCs ran in parallel on NOAA's Jet supercomputer from 2012 to 2014 (Gopalakrishnan et al. 2020). The results from this demonstration (e.g., Tallapragada et al. 2016) yielded critical information for the 2015 upgrade. At that point, NCEP allotted up to seven instances of HWRF, allowing it to run operationally in any worldwide basin covered by the NHC, the Central Pacific Hurricane Center (CPHC), or the Joint Typhoon Warning Center (JTWC).

Important differences always remained in non-NATL configurations. For example, HWRF did not use DA in the EPAC until the 2016 upgrade, and DA was never used outside NHC's area of responsibility. Across all basins, the vertical resolution was not unified until 2018, and the ocean coupling was never unified. For example, the EPAC ocean component evolved from a simple, coarse-resolution, one-dimensional ocean model (POM) in 2012 to MPIPOM-TC in 2014. In the EPAC, the MPIPOM-TC configuration began using RTOFS in 2016 (4 years earlier than in the NATL basin) and adopted 41 vertical levels in 2017. Meanwhile, basins outside the NATL and EPAC were not coupled to an ocean model until 2017, at which point the Hybrid Coordinate Ocean Model (HYCOM; (Kim et al. 2014) was coupled to HWRF forecasts in the WPAC and north Indian Ocean. HWRF finally started using ocean coupling with HYCOM for Southern Hemisphere TCs in 2018. On the other hand, HYCOM was never introduced in the NATL or EPAC.

HWRF forecasts generally improved in the EPAC, despite being of secondary focus to the NATL. HWRF track errors in the EPAC decreased by a third to half from 2007 to 2022, and track skill improved by 1%–2% per year, resulting in a net improvement of ~25% at most lead times. Both of these numbers were similar to the improvement seen in the NATL. Meanwhile, VMAX errors fell by less than in the NATL (~30%–45%). Nonetheless, EPAC VMAX skill improved ~40%–50% at most lead times, a qualitatively similar result to VMAX skill in the NATL. As highlighted above, significant differences between the HWRF configuration in NATL and HWRF in non-NATL basins made it difficult to compare performance.

procedures ignored the entire forecast cycle initialized at that time as well as prior forecasts valid at that time. Additionally, verification was not performed if the GFDL TC tracker (Marchok 2002, 2021) was unable to find a vortex associated with a given system or, likewise, if there was no associated best track entry. The latter may occur when the model produces false alarms or continues to forecast a TC beyond its actual dissipation. To avoid inflated intensity and structure errors related to landfall timing issues that are not the focus of this study, forecasts were only verified if the model and the best track were both over water. We are more concerned with HWRF's performance over water and *leading up to* landfall because that was the primary focus of developments. Note that verification did include forecasts if a TC reemerged over water in both the model and the best track.

Among the best track metrics, this paper focuses on verification of track, maximum sustained 10-m winds (VMAX; a measure of TC intensity), and radii of 34- and 50-kt 10-m sustained winds as reported by the NHC (R34 and R50, respectively). As in Rappaport et al. (2009), track error is defined as the great-circle distance between the forecast and best track positions, and intensity error is defined as the difference between the tracker-calculated and best track VMAX. Verification metrics shown here include the mean absolute error (MAE; the standard NHC verification metric), mean error (bias), and the percent improvement (i.e., skill) of the VMAX and track MAE with respect to the climatology and persistence model (OCD5), as described in DeMaria et al. (2022). The VMAX bias is positive if a storm is too strong and negative if it is too weak. For significant wind radii, the bias is positive if a storm is too large and negative if it is too small. Along-track bias is positive if a forward motion is too fast and negative if it is too slow. Cross-track bias is positive if a storm is to the right of the actual track and negative if it is to the left. Note that the HWRF VMAX is an instantaneous output, whereas the NHC best track is inherently smoother. Smoothed versions of the high-frequency (every ~3.3 s) 10-m wind data from HWRF verified better against the NHC best track (i.e., errors were

lower) than the instantaneous output (Z. Zhang et al. 2021), though this smoothing technique was never used in operations.

While using OCD5 as a baseline to assess HWRF performance is useful to examine trends, it does carry caveats on smaller time scales. In particular, unusually low or high OCD5 errors can strongly impact skill calculations even when the error in the model being evaluated does not change. For example, OCD5 VMAX MAE was unexpectedly low in the NATL basin in 2011 and 2012 (Fig. 5b). In fact, OCD5 VMAX errors after 36 h in 2011 were the lowest on record up until that date, and the errors were even lower in 2012. The annual variability and downward trend in OCD5 errors indicate that every year features a unique set of TCs (location, origin, and environment), complicating interyear comparisons.

It is also useful to compare HWRF performance with NHC Official (OFCL) forecast errors, which are shown in Fig. 6 and also available in Cangialosi et al. (2023). For 2007–22, OFCL errors exhibited clear downward trends at all forecast lead times for track and intensity MAE. Since 2007, NHC's track forecast errors decreased by 30%–50% and intensity errors shrank by up to 55% (i.e., at 48 and 72 h). Annual variability is observed, although it is less pronounced than in OCD5 errors. For more information about OFCL verification, please refer to Cangialosi et al. (2023).

This study evaluates the "early-model" version of HWRF (i.e., HWFI), which is the primary aid for operational forecasts. The raw output from the vortex tracker (i.e., the "late model") is usually not available until after an official forecast is issued, so operational centers typically base their forecasts off postprocessed output from the most recent available cycle. The postprocessing procedure, known as "the interpolator," corrects the bias in the previous 6-h forecast for position and VMAX. The interpolator applies a constant position correction through the duration of the track forecast, whereas it applies VMAX correction

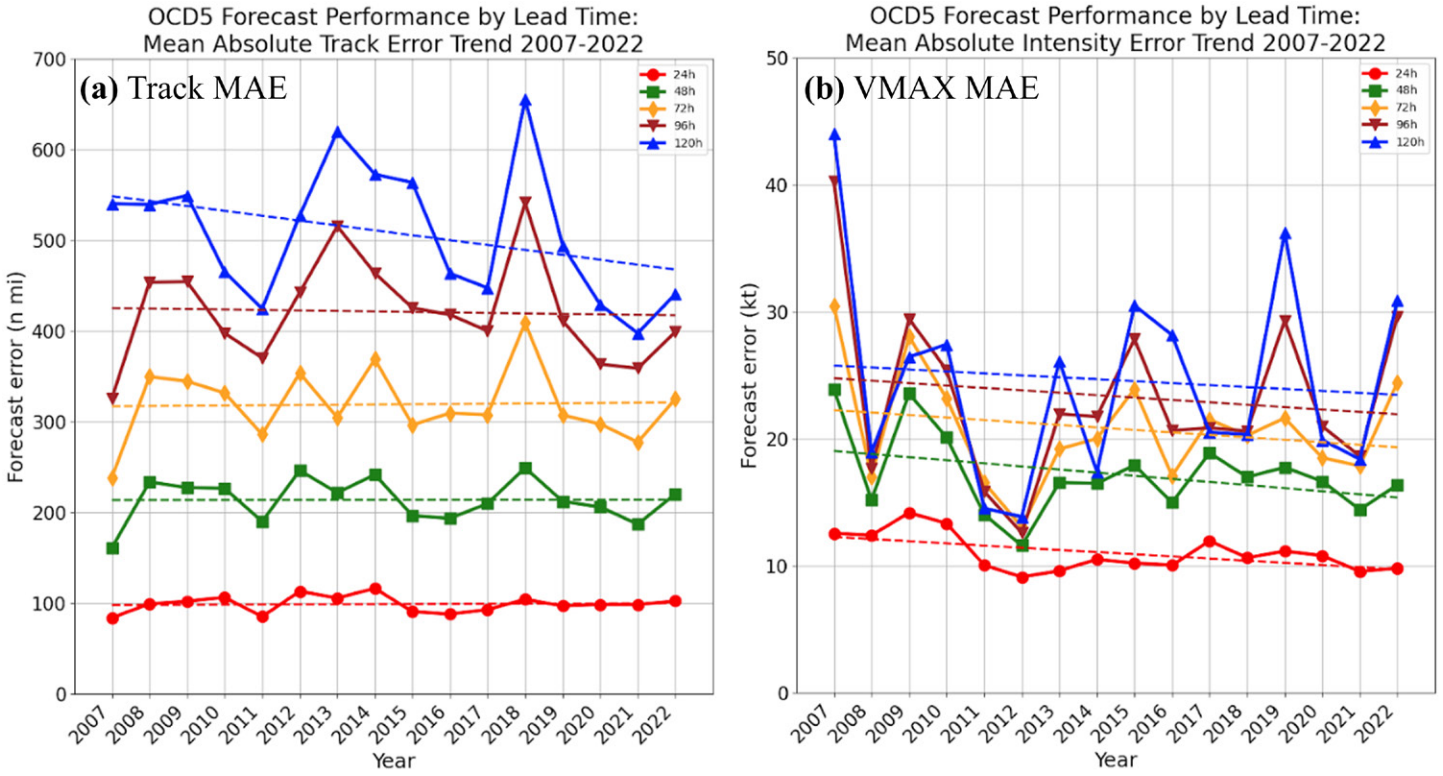

Fig. 5. The evolution of the 5-day climatological and persistence (OCD5) model forecast MAE during the evaluation period (2007–22) for NATL (a) track and (b) VMAX. The errors are evaluated at lead times of 24 h (red circle), 48 h (green square), 72 h (gold diamond), 96 h (brown upside-down triangle), and 120 h (blue triangle). The linear trend for each forecast lead time is shown by an accompanying dashed line of the same color.

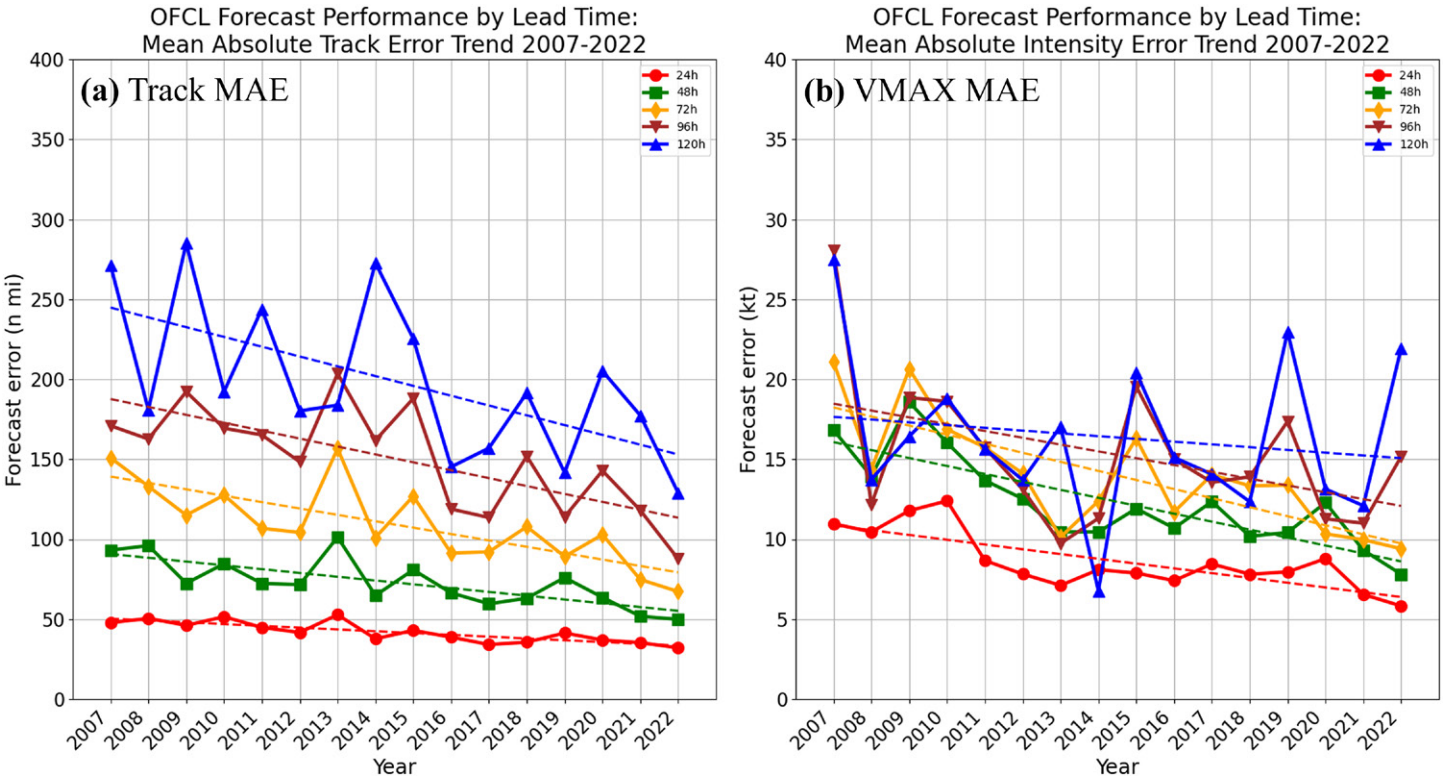

Fig. 6. As in Fig. 5, but for NHC official (OFCL) forecasts.

in three steps: 1) A full correction is applied from 0 h to t_A; 2) the correction weight linearly decreases to zero from t_A to t_B; and 3) no correction is applied from t_B to the end of the forecast. The correction offsets t_A and t_B have varied across years and basins to maximize the performance of HWFI. In 2022, HWFI used t_A = 0 h and t_B = 96 h in the NATL and eastern North Pacific (J. Martinez, NHC, 2023, personal communication), and it used t_A = 0 h and t_B = 6 h for all non-NHC basins (L. Cowan, JTWC, 2023, personal communication). In addition to the bias correction, the interpolator applies 10 passes of a 1:2:1 smoother to produce the early-model VMAX forecast.

Since a great deal of HWRF improvements resulted from HFIP, this study also compares HWRF errors and skill to HFIP goals that were established in 2009. The HFIP goals were set using a consensus of track and VMAX MAE from top-performing operational models during the 2006–08 hurricane seasons. The 5-yr goal (for 2014) was to reduce those errors by 20%, and the 10-yr goal (for 2019) was to reduce them by 50% (see Gall et al. 2013). The corresponding HFIP skill baseline and goals were calculated by comparing the consensus errors to OCD5 MAE from 2006 to 2008 as a reference.

Finally, this study also examines error trends in forecasts of rapid intensification (RI). RI events represent only 5% of the VMAX change distribution (i.e., the tail of the distribution), though they have historically been associated with fast error growth rates and large forecast errors (Halperin and Torn 2018; Kieu et al. 2018; Trabing and Bell 2020). The method for verifying RI forecasts follows from DeMaria et al. (2021), who counted a 24-h period as an RI event if the model and/or best track exhibit RI. Specifically, the standard RI definition (an increase of ≥30 kt in 24 h) is applied to the 24-h period leading up to the verification time for both the model and the best track.

4. Historical forecast verification and trends

Before presenting results, a critical point to keep in mind is that the sample size varied tremendously from year to year, with a noticeable upward trend at all lead times due to increased

TC activity in the NATL over the evaluation period (Fig. 7). The sample size exhibited strong interannual variability, with the number of 24-h cases in some years (e.g., 2009, 2013, and 2014) on par with the number at 120 h in other years (e.g., 2008 and 2012). Any interpretation of errors and skill must take this variability into consideration since MAE in years with small samples could be more strongly influenced by outliers. As such, three different 120-h sample size bins are shown in Figs. 8–10 to denote above-average (≥100), average (50–99), and below-average (<50) activity.

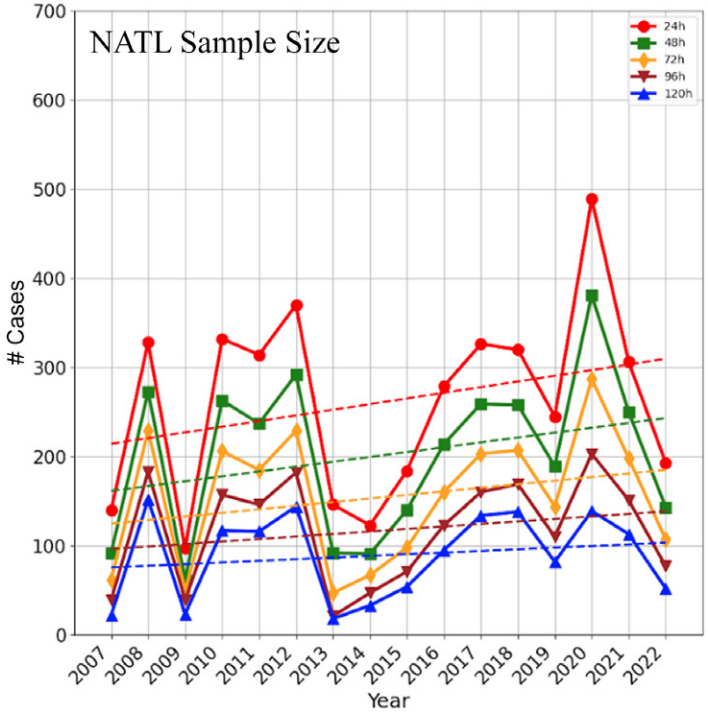

Fig. 7. The total sample size of "early" HWRF (HWFI) forecasts in the NATL basin for the evaluation period (2007–22) at lead times of every 24 h through 120 h. The linear trend for each forecast lead time is shown by an accompanying dashed line of the same color.

a. Overall verification trends. Perhaps the most important message in this review is that HWRF track and VMAX errors decreased remarkably in the NATL basin during the period of its operational use (Figs. 8 and 9). Track errors at 24 h fell ~25%, and at longer lead times, they fell 30%–40% (Fig. 8a). Track skill improved generally between 1% and 2% a year on average, yielding a net gain of ~30% at 24 h and roughly 20% at longer lead times (Fig. 8b). VMAX errors improved more quickly than track errors, especially after 24 h, where total error reduction ranged between 45% and 50% (Fig. 9a). The somewhat diminished improvement at 24 h is likely related to using the interpolator to produce early-model forecasts (i.e., it improves the skill of bad short-term forecasts the most). In terms of skill, the VMAX forecasts at lead times beyond 24 h improved about 2%–3% a year on average, resulting in a net improvement of around 40% through 96 h and 60% at 120 h (Fig. 9b). One can also look at HWRF error improvements in terms of lead time gained. For example, track errors at 120 h in 2022 were less than 72-h track errors in 2007, yielding a 2-day gain in lead time. Likewise, 120-h VMAX errors in 2022 were commensurate with 24-h VMAX errors in 2007, which represents a 4-day improvement in lead time. Upon closer inspection, VMAX MAE was cut in half at all lead times from 2007 to 2016, with a smaller decreasing trend thereafter (Fig. 9a). This precipitous error drop reflected how poor HWRF forecasts were prior to 2011, after which major upgrades to resolution, physics, and vortex initialization were implemented in operations.

HWRF forecasts of outer significant wind radii (R34 and R50) also improved considerably, though the errors demonstrated more interannual variability than VMAX or track errors did (Fig. 10). Note that the 2007 and 2008 data were removed from Fig. 10 due to drastically inflated errors that appear to reflect issues with the tracker and/or interpolation software [R34 errors in those years were 60–80 n mi (1 n mi = 1.852 km), while R50 errors were 34–45 n mi]. No apparent connection was found between sample size and the variability of wind radii errors. The error trends in Fig. 10a show R34 MAE decreased from about 30–40 n mi in 2009 to 20–30 n mi in 2022 or an improvement of about 30%–35%. R50 trends suggest MAE improved from about 20 to 15 n mi or ~25% (Fig. 10b).

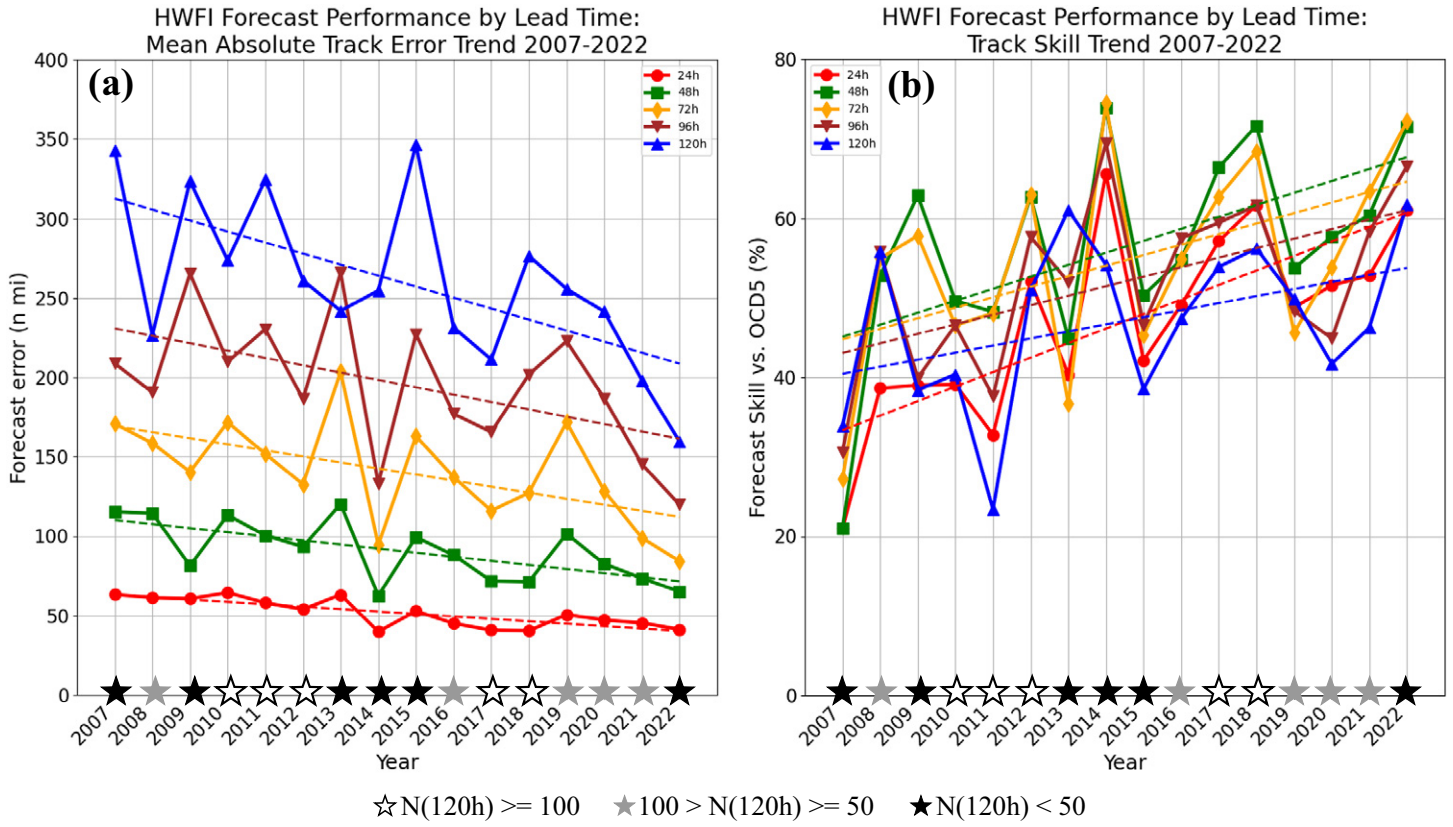

FIG. 8. As in Fig. 5, but for the evolution of HWFI forecast error and skill over the duration of HWRF's operational lifetime (2007–22): (a) track MAE and (b) track skill relative to OCD5. The 120-h sample size is denoted along the *x* axis, with a white star representing greater than or equal to 100 cases, a gray star representing 50–99, and a black star representing less than 50.

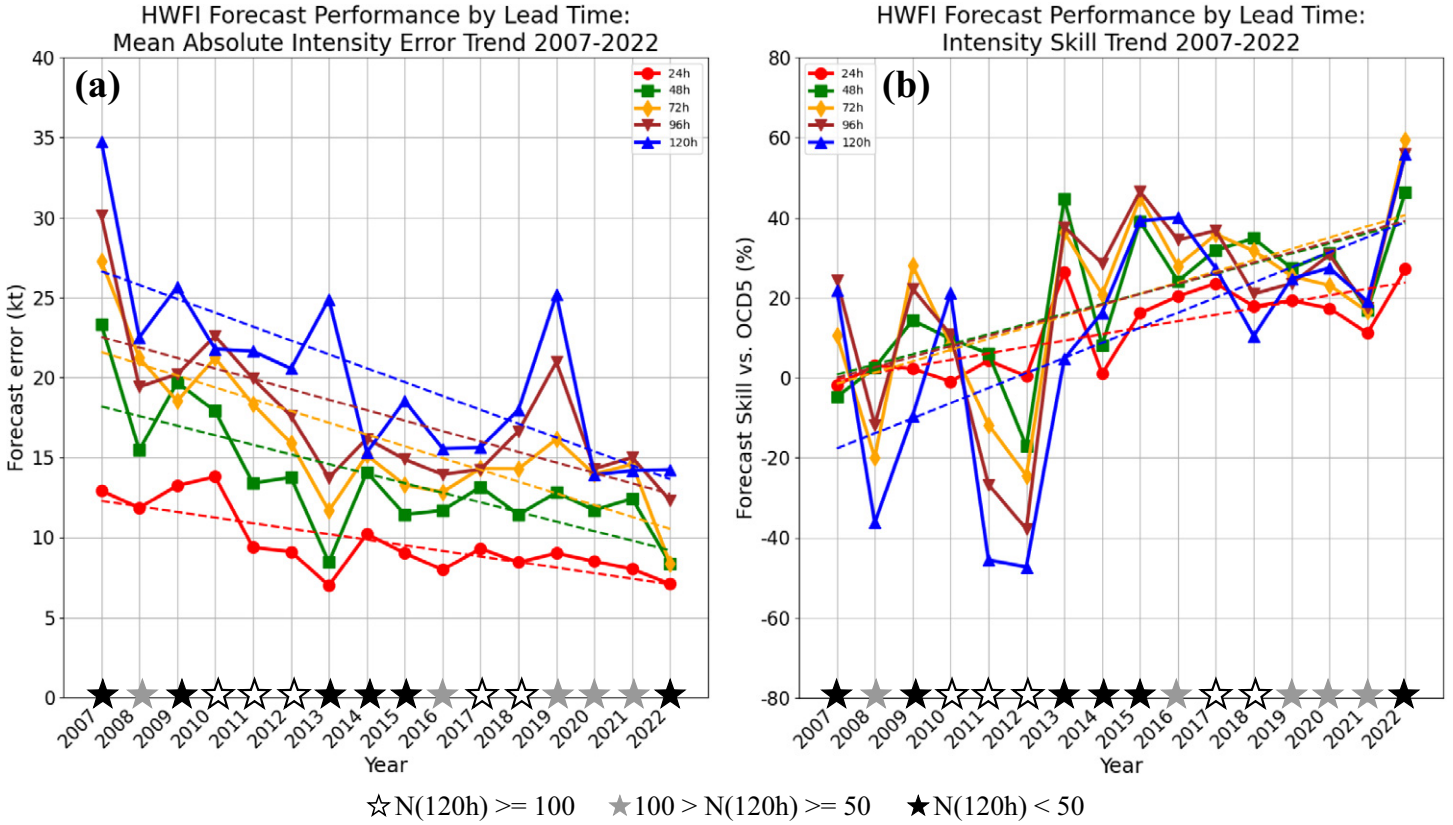

FIG. 9. As in Fig. 8, but for (a) VMAX MAE and (b) VMAX skill.

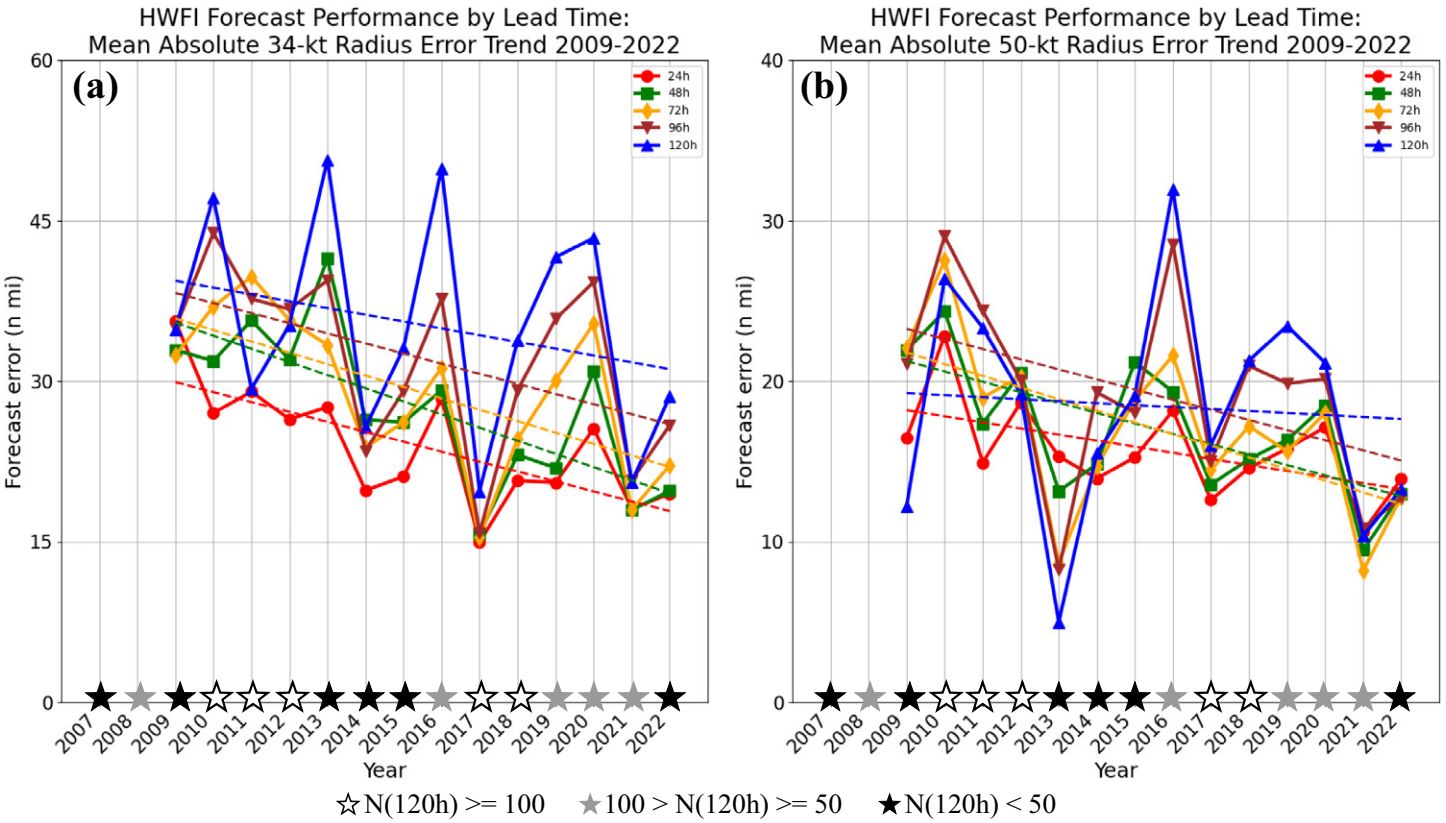

Fig. 10. As in Fig. 8, but for (a) R34 MAE and (b) R50 MAE.

Figures 11 and 12 highlight HWRF performance in three periods: 1) a preliminary period (2007–11), when HWRF featured only a single moving nest; 2) a maturing period (2012–17), when the addition of a second moving nest enabled significant physics and DA upgrades; and 3) a modern period (2018–22) that featured an acceleration of improvements based on past successes. Figure 11 focuses on the MAE and skill for track and intensity as well as the MAE for R34 and R50. Meanwhile, Fig. 12 shows the biases for these variables, including a breakdown of along- and cross-track bias. Note that these figures also include an evaluation for 2022 to highlight HWRF's final year of performance as NOAA's flagship hurricane model. HWRF's performance in each year individually can be found in Fig. S1 in the online supplemental material.

The most obvious signal is the very large VMAX MAE at all lead times in the preliminary period that was reduced in later years (Fig. 11d). This error was associated with TCs that were generally forecast to be too weak at shorter lead times and too strong at longer lead times (Fig. 12c), a consequence of insufficient grid spacing (9 km) to resolve the evolution of the TC inner core. The 3-km inner-moving nest was introduced in 2012 (Fig. 2), and the VMAX MAE in the maturing period was significantly smaller than that in the preliminary period (e.g., Tallapragada et al. 2014; Zhang et al. 2023). The VMAX bias in the maturing period continued to be positive (i.e., the forecast was too strong on average) at longer lead times. Overall, VMAX improved notably in the maturing and modern periods due to improvements in DA, physics, ocean coupling, and grid resolution (Figs. 2 and 3). By the modern period, the VMAX skill was >20% at 36 h and longer lead times (Fig. 11e), representing a significant improvement compared to earlier years.

HWRF intensity performance in 2022 was particularly noteworthy. The average VMAX bias was only slightly negative at most lead times, and the VMAX MAE was below 10 kt through 72 h. Remarkably, MAE < 10 kt is commensurate with uncertainty in the NHC best track (Landsea and Franklin 2013; Torn and Snyder 2012). The forecast VMAX skill in 2022

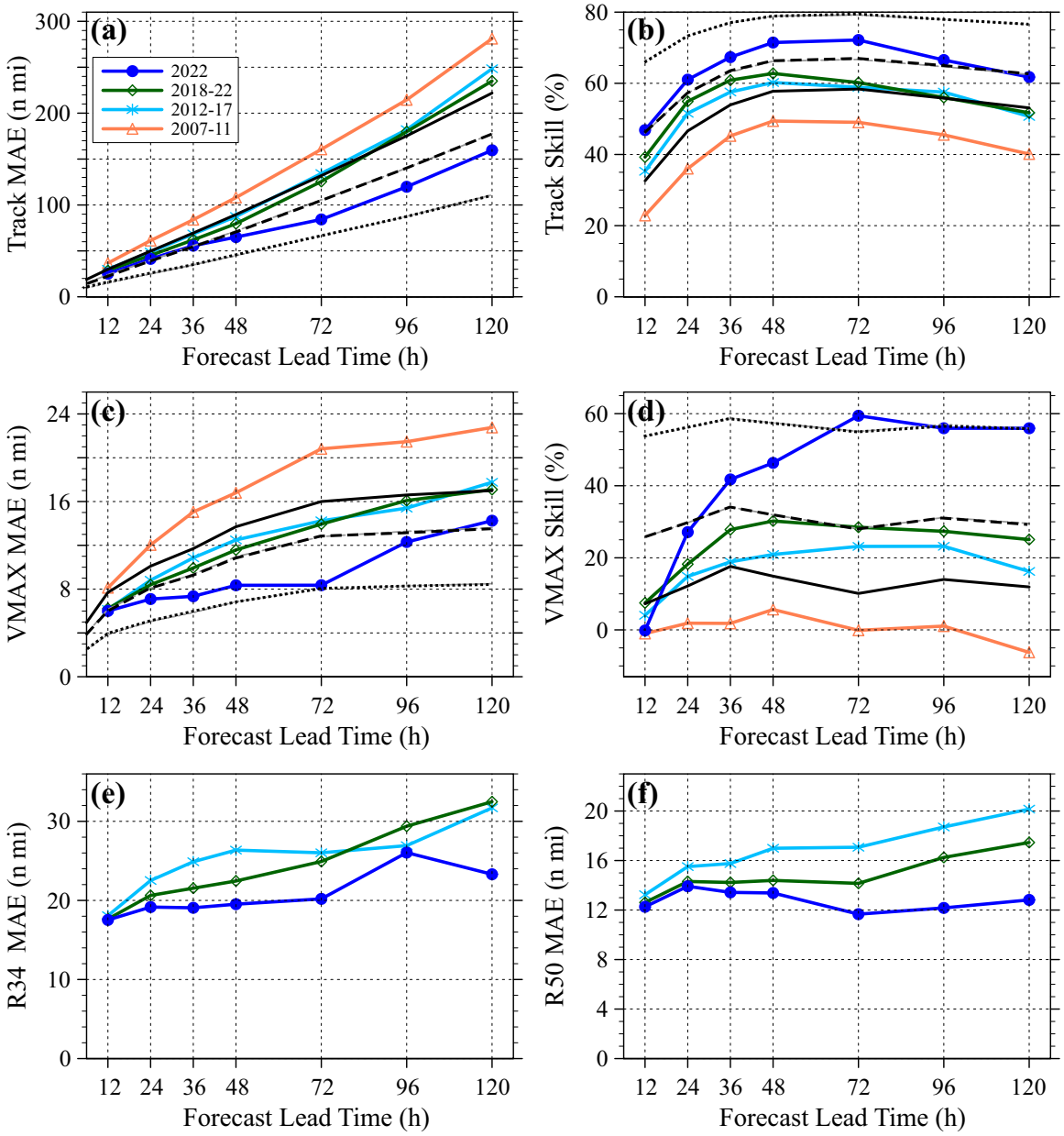

Fig. 11. Verification statistics for NATL TC forecasts from four HWFI subsets: a preliminary period (2007–11; orange triangle), a maturing period (2012–17; cyan asterisk), a modern period (2018–22; green diamond), and 2022 only (blue circle). Shown are (a) track MAE (n mi), (b) track skill relative to OCD5 (%), (c) VMAX MAE (kt), (d) VMAX skill relative to OCD5 (%), (e) R34 MAE (n mi), and (f) R50 MAE (n mi). Note that the preliminary period is omitted in the wind radii panels (e) and (f). In (a)–(d), the original HFIP baseline, 5-yr goal, and 10-yr goal are shown in black solid, dashed, and dotted lines, respectively. HFIP baseline/goals were not established for wind radii.

also superseded that of the prior years by a substantial margin and held steady at around 60% on days 3–5.

A comparison of biases for VMAX and wind radii suggests HWRF forecasts had structural errors in the preliminary and maturing periods that were resolved in later years (Fig. 12). For example, the VMAX bias in 2012–17 was strongly positive (Fig. 12c), whereas R50 and R34 were both negatively biased (Figs. 12d,e). Stronger storms are typically larger (Chavas and Emanuel 2010; Chan and Chan 2012; Ruan and Wu 2022), so this result defied that expectation. The situation changed in the modern period, when the biases in VMAX, R34, and R50 were more physically consistent. It should be noted that smaller TCs rapidly intensify at a

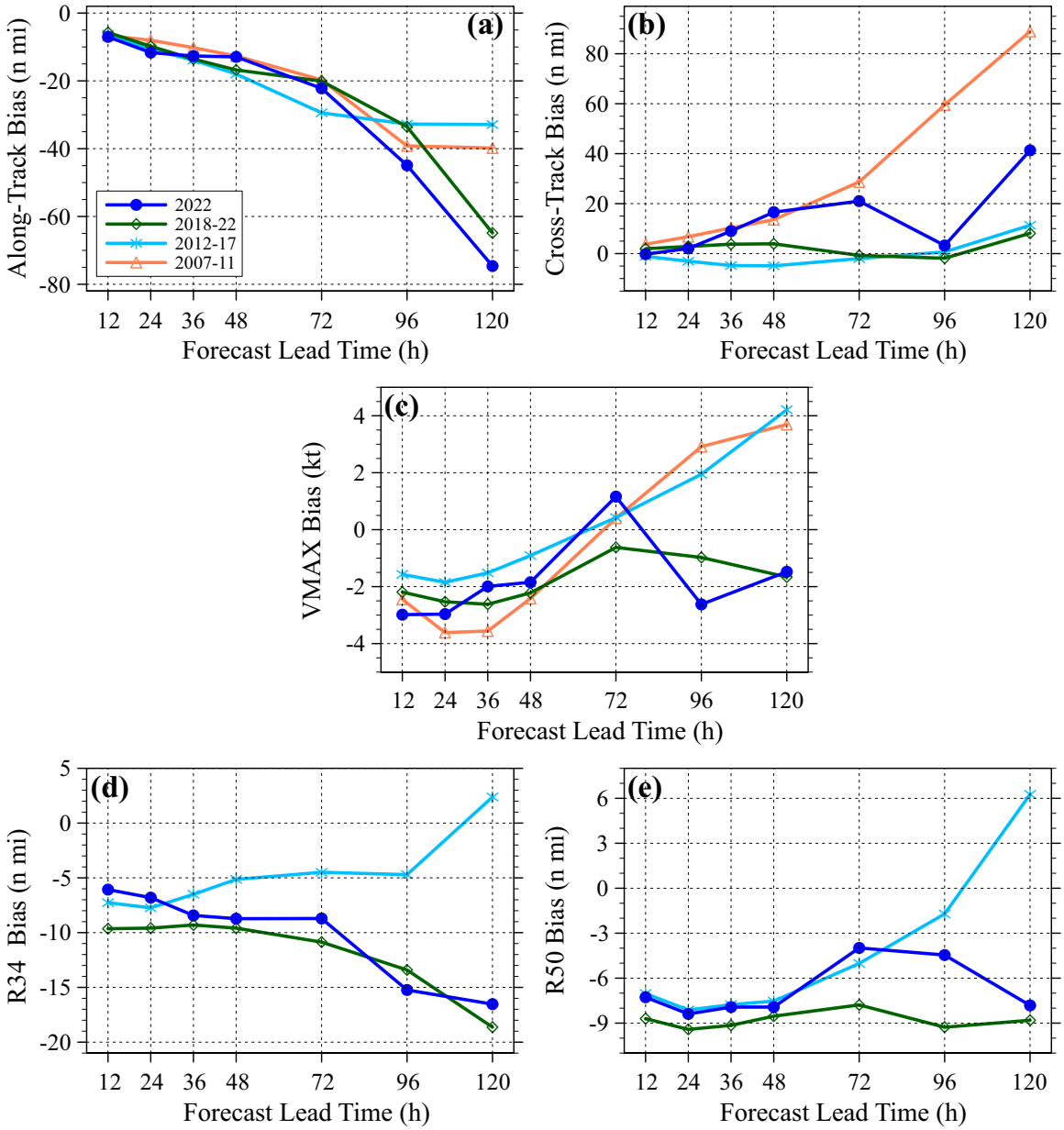

Fig. 12. As in Fig. 11, but for forecast biases: (a) along-track bias, (b) cross-track bias, (c) VMAX bias, (d) R34 bias, and (e) R50 bias. Note that the preliminary period (2007–11) is omitted in the wind radii panels in (d) and (e).

higher rate than larger TCs (e.g., Carrasco et al. 2014), so the VMAX bias could reflect a poor prediction of RI timing that improved over time.

HWRF track forecasts improved substantially from the preliminary period to the maturing period, but then more slowly thereafter. Large track MAE in the preliminary period was associated with much larger cross-track biases than in later years. Track skill improved by ~15% from the preliminary period to the maturing period, but subsequent improvements were rather small (≤5%). By 2022, the track skill had improved to 60%–70% at most lead times, which was only slightly better than the intensity skill. In this respect, HWRF was a bit unusual because it had developed exceptional intensity skill but only average track skill compared with other dynamical guidance (see Cangialosi 2023).

The greater improvement in HWRF VMAX compared with track has a few underlying causes. First, heavier weight was generally assigned to improving intensity than improving track. Concomitantly, early HWRF VMAX forecasts had much more room to improve than did its track forecasts. For instance, HWRF VMAX forecasts during its first few years had no skill,

whereas track skill at the time was ~40%. In comparison, official NHC forecasts around 2007 had about 20% skill for VMAX and 45% skill for track. Since then, NHC VMAX and track skill have, respectively, improved by about 20% and 25% (i.e., about the same amount).

HWRF also better met the HFIP goals established in 2009 for VMAX than for track. To demonstrate this more clearly, Fig. 11 overlays the original HFIP baseline established in 2009 along with the 5- and 10-yr HFIP goals intended for the 2014 and 2019 seasons, respectively. First, it is important to recognize that HWRF errors in 2007 far exceeded the baseline, which reflects HWRF's initial lack of skill compared to other models. Though the track improved over each major period, it was not until the last few years (~2021–22) that HWRF track MAE and skill improved beyond the original 5-yr goal at most lead times (Figs. 11a,b). Further, track forecasts never reached the 10-yr goal. On the other hand, VMAX forecasts began robustly exceeding the baseline in the maturing period, flirted with the 5-yr skill goals in the modern period, and surpassed the 10-yr VMAX skill goals in 2022 (Fig. 11d). In terms of VMAX MAE, HWRF did not exceed the 5-yr goals at most lead times until 2022 (Fig. 11c).

VMAX skill met HFIP goals faster than VMAX MAE because the 2006–08 OCD5 VMAX errors used to calculate these goals[10] were rather low compared to OCD5 errors in the 2007–22 period (Gall et al. 2010). In fact, only 5 years in the study period had OCD5 errors that were smaller than the 2006–08 OCD5 reference errors. As such, HFIP VMAX skill goals were easier to meet in years when OCD5 errors were larger than 2006–08 OCD5 errors (e.g., 2022). This is a cautionary example of evaluating model performance against a moving baseline.

[10] OCD5 reference errors were 11.5, 16.1, 17.8, 19.3, and 19.3 kt at 24, 48, 72, 96, and 120 h, respectively.

b. Rapid intensification. HWRF made tremendous progress in NATL RI forecasts starting around 2017 (Fig. 13), a result that might be tied to some key recent upgrades. Before 2017, VMAX errors for RI typically ranged from 20 to 40 kt and had no meaningful trend (see

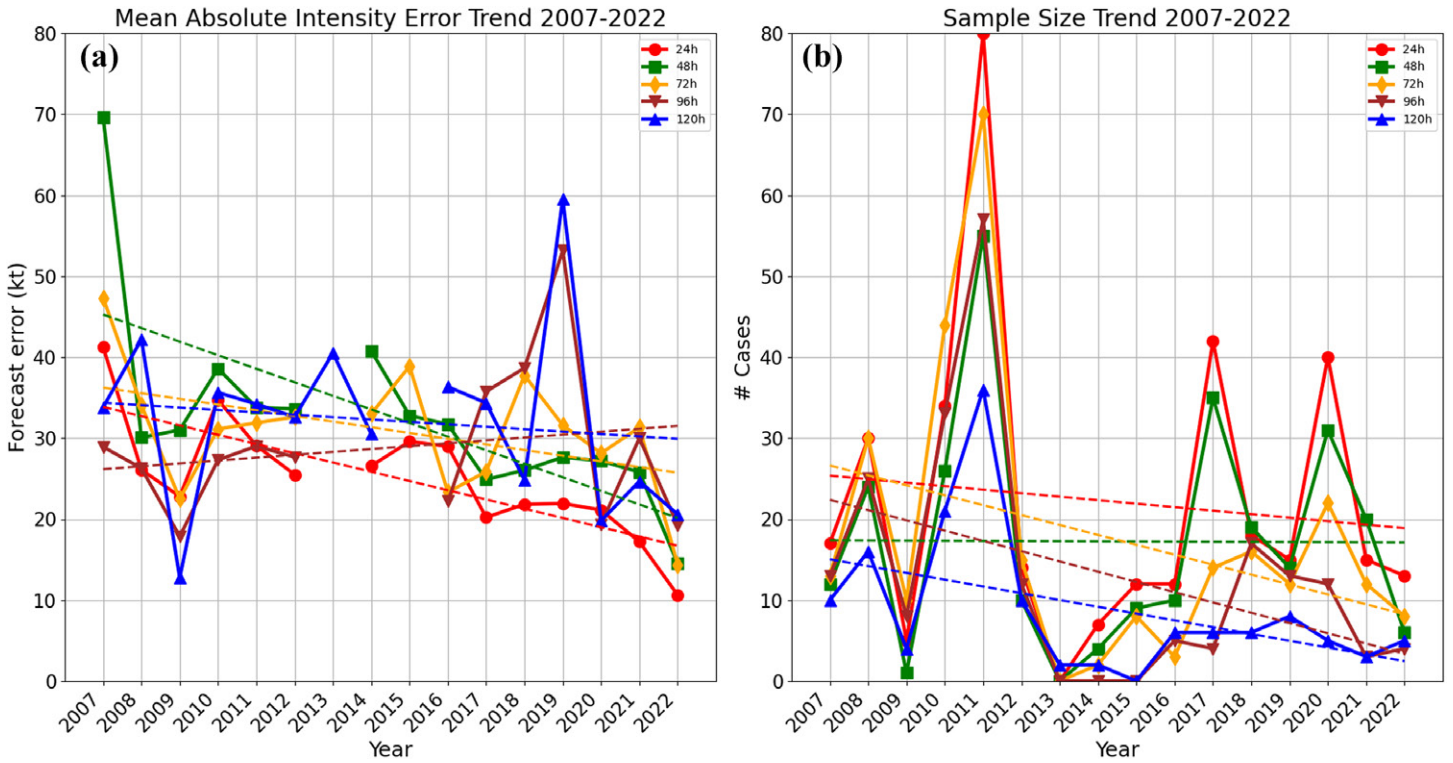

Fig. 13. (a) As in Fig. 9a, but for VMAX MAE stratified by over-water RI events for HWFI forecasts. (b) The corresponding sample size is shown.

section 2 for RI definition). These errors were roughly twice the errors for the entire sample (10–20 kt; Fig. 9a). However, they steadily decreased after 2017 and ranged from about 10 to 20 kt in 2022 (i.e., a roughly 50% reduction in 5 years). In other words, errors at the RI tail of the error distribution were reduced, contributing to the overall improvement of VMAX forecasts (see Fig. 9b). Note that this improvement coincides with two important changes in the timelines shown in Figs. 2 and 3 and described in section 2. First, the grid spacing of the innermost nest changed from ~2 to ~1.5 km in 2018 (Fig. 2). Second, HWRF had a major DA system upgrade in 2017 to start using mesoscale error covariance more appropriate for the TC core as well as other DA upgrades in 2018–19 that allowed for much more effective use of inner-core reconnaissance observations (Fig. 3). Commensurate with these upgrades, the model began assimilating more inner-core data: flight-level HDOBs in 2017, as well as SFMR and G-IV Doppler velocity in 2018 (see section 2). Further, more accurate estimations of dropsonde location in 2018 allowed important inner-core dropsonde data to be assimilated.

c. Regional performance variability and U.S. landfalls. Recently, the NATL basin has experienced very active hurricane seasons, and the availability of accurate and timely guidance near land has proven critical. From 2020 to 2022, the contiguous United States (CONUS) experienced 22 TC landfalls, including 10 hurricanes and 4 major hurricanes.

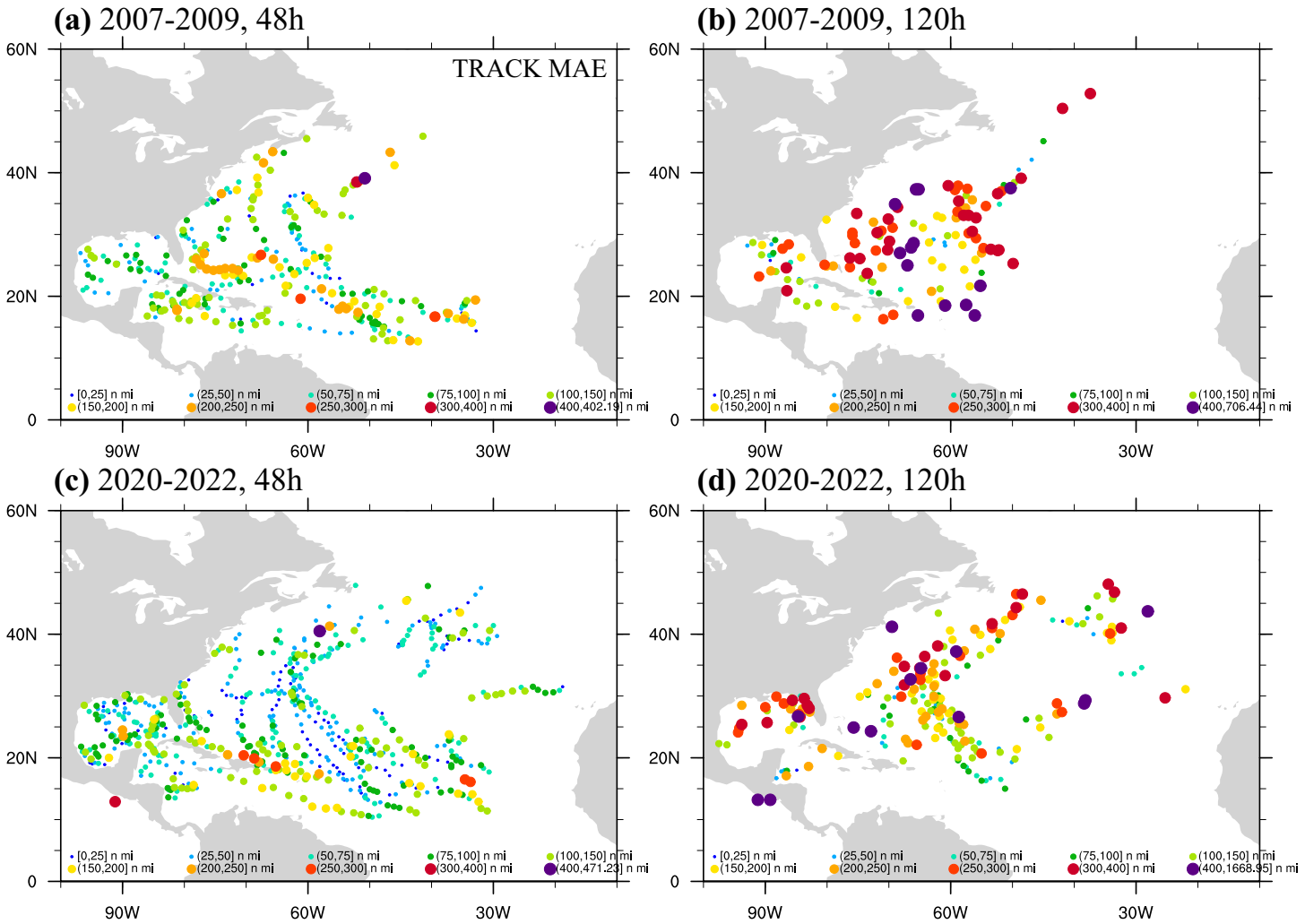

FIG. 14. **The spatial distribution of HWFI forecast track MAE (n mi) for (a),(b) 2007–09 and (c),(d) 2020–22. Specifically, the following lead times are shown: (a),(c) 48 and (b),(d) 120 h. Markers are placed at verifying locations, and marker size scales with error magnitude. Bluer/smaller markers denote smaller errors, while redder/larger ones denote larger errors. Error bins are identical in all panels: [0, 25], (25, 50], (50, 75], (75, 100], (100, 150], (150, 200], (200, 250], (250, 300], (300, 400], (400, E_{max}]. The term E_{max} represents the maximum MAE in each panel.**

Sixteen of those 22 landfalls were along the U.S. Gulf Coast (eight tropical storms, four hurricanes, and four major hurricanes). Though improving TC forecasts anywhere and anytime is useful, having skillful forecasts during the watch/warning time period (i.e., 48–72 h) as TCs approach land is critical to guide the most accurate official forecasts when it matters most.

HWRF forecast performance evolution varied as a function of location (Figs. 14 and 15). From 2007–09 to 2020–22, track forecasts slightly improved across the NATL at the 48-h lead time, and 120-h track forecasts improved notably in the open Atlantic (Fig. 14). However, track forecast performance did not show any obvious improvement near CONUS. On the other hand, Fig. 15 shows that HWRF VMAX forecasts improved in many locations across the NATL over time, especially near CONUS. In 2007–09, some of the largest VMAX errors (negative biases) were located in the northwest Caribbean Sea at both 48 and 120 h (Figs. 15a,b). This result changed substantially in recent years, and Figs. 15c and 15d show that by 2020–22 the largest errors had shifted far from CONUS to the main development region (MDR[11]). Though large VMAX busts in the NATL MDR were both weak and strong, positive biases were much more frequent (Fig. 15c). Meanwhile, errors

[11] Goldenberg et al. (2001) defined the MDR as the region between 10° and 20°N between Africa and the Windward Islands.

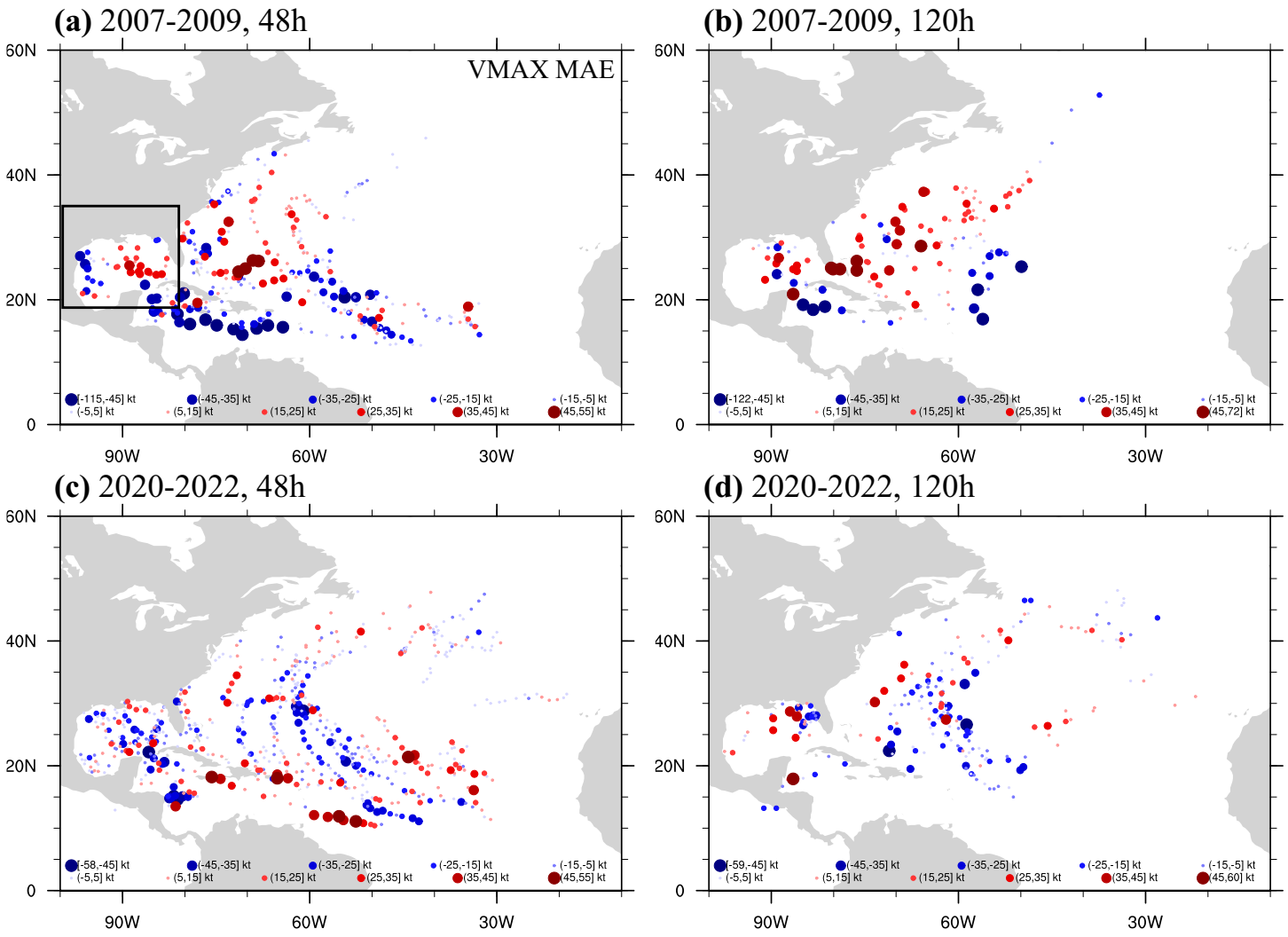

FIG. 15. As in Fig. 14, but for HWFI forecast VMAX bias (kt). Blue markers denote negative errors, while red markers denote positive errors. Error bins are identical in all panels: [E_{min}, –45], (–45, –35], (–35, –25], (–25, –15], (–15, –5], (–5, 5], (5, 15], (15, 25], (25, 35], (35, 45], (45, E_{max}]. The term E_{max} is the maximum bias in each panel, while E_{min} is the minimum bias in each panel. The black "Gulf box" in (a) includes the Gulf of Mexico and northwest Caribbean Sea and is bounded by 18°–35°N and 81°–100°W.

near CONUS had become fairly small at shorter lead times compared to errors in the rest of the basin.

Though the previous analysis has some caveats (e.g., the 2007–09 TCs were fundamentally different from those during 2020–22), Fig. 16 underscores HWRF's improvement near CONUS by comparing MAE of NHC forecasts with HWRF forecasts initialized over water and in the Gulf of Mexico and northwest Caribbean, i.e., in a "Gulf box" (18°–35°N, 81°–100°W; see Fig. 15a). Coincident with the large forecast busts in the northwest Caribbean, HWRF compared quite poorly to the NHC forecast in the Gulf box from 2007 to 2009 (Fig. 16a). However, this changed drastically by 2020–22, by which time HWRF VMAX forecasts were competitive with the NHC forecast (Fig. 16b). This improvement was particularly important given the plethora of landfalls the Gulf Coast experienced from 2020 to 2022. Note that track MAE in this region was generally consistent with the NHC official forecast (not shown). The success of HWRF in the Gulf region in particular might be because of the extremely frequent reconnaissance sampling used for landfalling U.S. TCs (e.g., Zawislak et al. 2022). As discussed earlier, inner-core data were not used at all in the earliest implementations of the model, but in the later years it had a substantial impact on model performance. In addition, HWRF began assimilating NEXRAD data in 2020, significantly expanding the amount of data available in the Gulf region.

Hurricane Laura (2020) is an excellent example of a Gulf storm where accurate intensity forecasts were quite valuable. Laura made landfall in western Louisiana as a strong category 4 major hurricane on 26 August 2020 after moving over the Greater Antilles as a tropical storm, taking 47 lives and causing ~$19 billion in damage (Pasch et al. 2020). Laura was also responsible for a devastating storm surge of up to 18 ft above ground, inundating coastal communities. When Laura emerged into the Gulf of Mexico on 24 August 2020, it was tropical storm strength. However, Laura proceeded to rapidly intensify from 50 to 130 kt in 2 days, quickly creating a dire scenario along the U.S. Gulf Coast that required accurate forecasts to optimize preparations.

Figures 17 and 18 show examples of operational HWRF forecasts from Hurricane Laura from 24 to 25 August that were included in the sample of Fig. 16. By this time, HWRF had been assimilating reconnaissance data for about 5 days with its more advanced assimilation

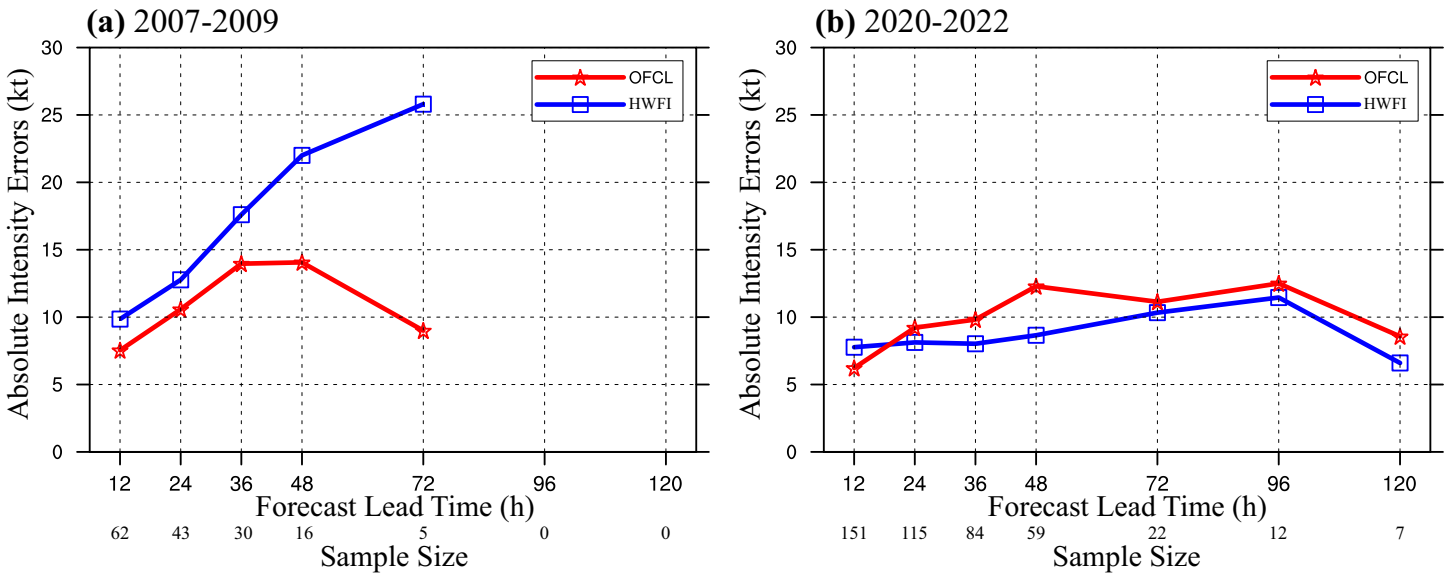

Fig. 16. Over-water VMAX MAE for NHC official (red star) and HWFI (blue square) TC forecasts within the "Gulf box" (i.e., the northwest Caribbean and Gulf of Mexico) for the following periods: (a) 2007–09 and (b) 2020–22. A TC forecast was included in the verification if its initial time and valid time positions were over water and in a box shown in Fig. 15a. Forecast times were ignored after landfall or dissipation. The sample size is shown for each lead time below each panel.

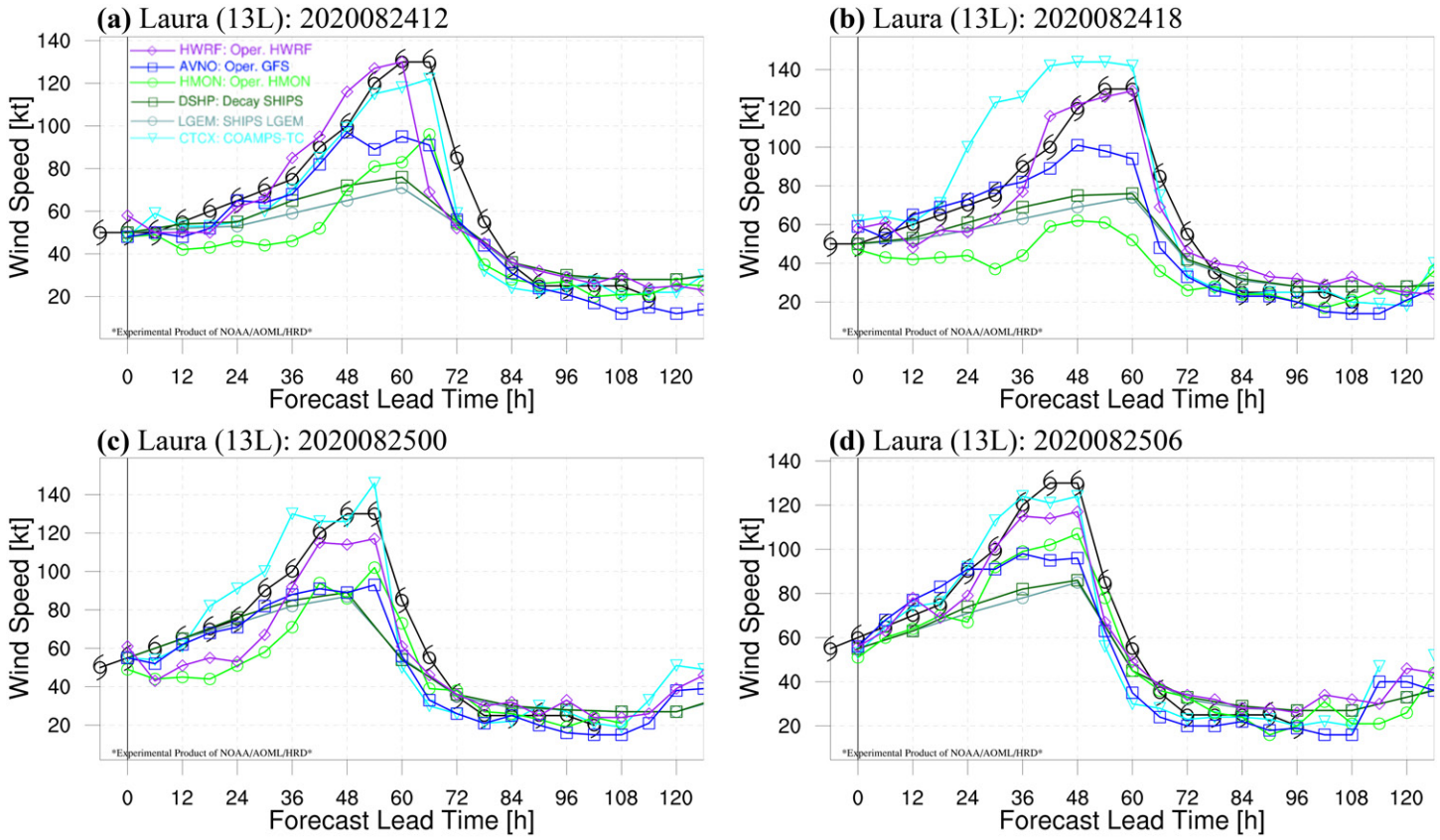

Fig. 17. Hurricane Laura (13L) VMAX forecasts from NOAA TC models [HWRF, operational GFS (AVNO), Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic (HMON) model, DSHP, LGEM, and CTCX] and the NHC best track (BEST; black line with TC symbol) from four consecutive forecast cycles: (a) 1200 UTC 24 Aug, (b) 1800 UTC 24 Aug, (c) 0000 UTC 25 Aug, and (d) 0600 UTC 25 Aug 2020.

mode, thereby improving the accuracy of inner-core analyses. In this critical period, HWRF consistently captured the timing and magnitude of the upcoming RI (Fig. 17). The only exception was the forecast initialized at 0000 UTC 25 August, which only slightly underpredicted Laura's maximum VMAX through landfall. The only other regional TC model to come close to HWRF's performance during this period was the Navy's COAMPS-TC model (CTCX), though HWRF began consistently showing RI potential sooner than CTCX did (not shown). Figure 18,

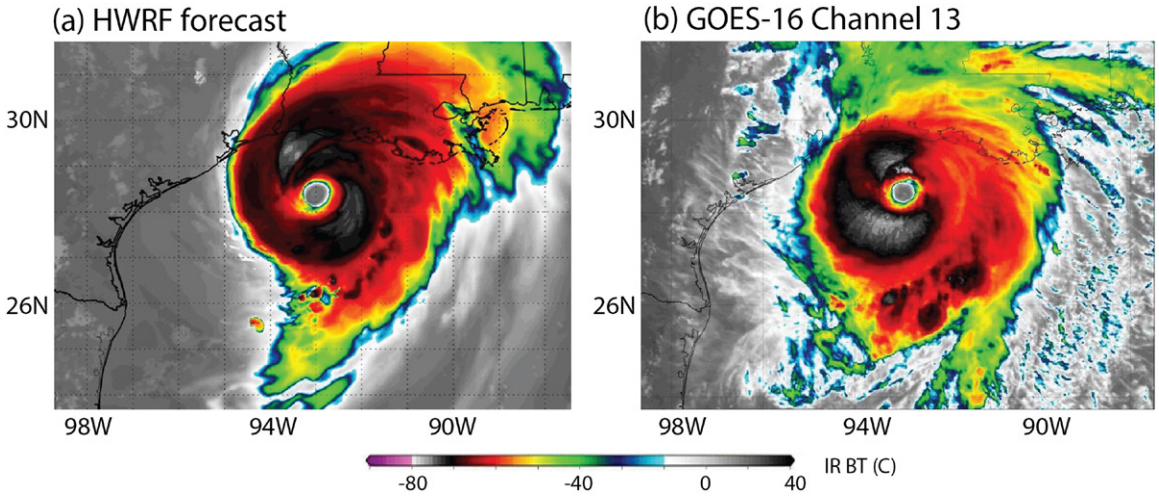

Fig. 18. For a valid time of 0000 UTC 27 Aug 2020, brightness temperature (°C) is compared for (a) 54-h HWRF forecast initialized at 1800 UTC 24 Aug 2020 and (b) infrared imagery from *GOES-16* channel 13. Images were modified from http://www.tropicaltidbits.com.

which compares observed and simulated brightness temperatures from the forecast in Fig. 17b, underscores HWRF's exceptional performance in the Gulf. Not only was the predicted maximum VMAX for this particular forecast nearly identical to the best track at 54-h lead time (Fig. 17b), but also the overall structure of the simulated storm closely resembled that actually observed in satellite data.

Outside of the MDR and areas near CONUS, HWRF VMAX forecasts performed relatively well at higher latitudes compared to the rest of the basin (Fig. 15). Even during the earliest years, HWRF VMAX errors north of about 30°N tended to be smaller than errors elsewhere in the basin. This trend continued through the late years, when errors became even smaller. The disparity in performance between the deep tropics and midlatitudes may reflect differences in RI frequency since RI is much more common south of about 30°N (Kaplan et al. 2010). Furthermore, TCs north of 30°N are typically weak or weakening in generally unfavorable environments and consequently are easier to predict.

5. Discussion and conclusions

The performance of the HWRF modeling system advanced considerably over a 16-yr period starting from its initial operational implementation (2007–22). Consequently, HWRF grew into a state-of-the-art, high-resolution, ocean-coupled numerical weather prediction tool for TCs that provided useful and increasingly accurate guidance to forecast centers/offices across the globe. Building upon earlier successes in the NATL, HWRF ran operationally for all other global TC basins from 2015 to 2022 (see sidebar). Over the study period, HWRF intensity forecast errors reduced dramatically in the North Atlantic (40%–50%), and intensity skill scores doubled (Fig. 9). Years of effort led to exceptional intensity forecast improvements near CONUS, especially in the Gulf of Mexico, boosted by increased and improved use of in situ reconnaissance data collected in the TC inner core (Figs. 14–18). In recent hurricane seasons, HWRF intensity forecast errors in the Gulf of Mexico were competitive with NHC official forecasts in that same region (Fig. 16b). In addition, the HWRF system made significant headway in predicting RI events in TCs, resulting in error reductions of nearly 50% during such events (Fig. 13). Forecast improvements also extended to track (Fig. 8) and wind radii (Fig. 10), demonstrating comprehensive improvements of TC forecasts for over a decade.

Significant improvements to the HWRF system can be attributed to a number of major changes, especially since 2012. These improvements and their impacts are summarized as follows:

1) Resolution (Fig. 2, left): Horizontal-resolution improvements were directly tied to moving nest developments and upgrades. The single most important development of HWRF was the addition of a second, inner, moving nest in 2012 that reduced grid spacing from ~9 to ~3 km. This development ultimately supported ~1.5-km grid spacing in 2018. The vertical resolution also became finer, with the number of vertical layers nearly doubling over HWRF's lifetime, and the model top was raised to 2 hPa. Finer resolutions enabled physics and DA upgrades (see below).
2) Physics parameterization schemes (Fig. 2, right): Physics upgrades focused on the PBL, cumulus, and radiation parameterizations, including the introduction of scale-aware schemes. The frequency of physics calls increased, as well. Invaluable turbulence measurements from NOAA P-3 flights were used for the first time to improve model physics. Nevertheless, there are still gaps in the model, especially in our understanding of the boundary layer where the ocean and atmosphere meet and where TCs gain and lose strength through an exchange of heat and energy between the ocean and atmosphere. New observations from uncrewed ocean and aircraft systems will help to improve our understanding of this interface and ultimately result in model improvements.

3) Data assimilation (Fig. 3): The HWRF DA system improved significantly, especially in terms of inner-core DA techniques. The most recent DA system was 3DEnVar with a fully cycled EnKF that provided storm-scale and flow-dependent covariances for the innermost moving nest. The improvements to the model and DA from 2016 to 2019 improved the impacts of inner-core data, thus allowing for increased use of inner-core reconnaissance data. These improvements along with the expanded use of satellite and NEXRAD data improved forecasts considerably.
4) Atmosphere–ocean coupling (Fig. 2, left): HWRF's ocean coupling evolved dramatically as the ocean model was upgraded from a one-dimensional POM to a more complex 3D MPIPOM-TC. Two major upgrades to MPIPOM-TC occurred when (i) the number of vertical levels increased and (ii) it was initialized with the RTOFS analysis instead of the GDEM climatology.
5) External factors. Improvements to the GFS, including resolution and physics parameterizations, led to more accurate initial and lateral boundary conditions for HWRF.

As discussed above, assimilating aircraft reconnaissance data substantially contributed to some of the recent HWRF forecast successes. Therefore, an obvious recommendation to further improve hurricane forecasts is to collect more of these data by conducting more airborne missions and adding new instruments, echoing the sentiment in recent publications (e.g., Ditchek et al. 2023; Ditchek and Sippel 2023). While airborne data have historically been mostly collected by crewed aircraft, future reconnaissance missions could be supplemented by uncrewed systems. In fact, a number of recent studies have shown the benefits of assimilating data from uncrewed missions (Christophersen et al. 2017, 2018; Wick et al. 2020; Aksoy et al. 2022; Sellwood et al. 2023).

One of the major challenges in the next generation of hurricane prediction will be to fill in the gaps where other data sources are not being used well. For example, groundbreaking research has shown that improving the use of oceanic data reduces forecast errors during RI (Dong et al. 2017; Domingues et al. 2021; Le Hénaff et al. 2021). Other work has shown that TC forecasts benefit from better use of satellite-based data, such as atmospheric motion vectors (e.g., Velden et al. 2017; S. Zhang et al. 2017; Sawada et al. 2019; Lim et al. 2019; Lewis et al. 2020; Li et al. 2020), satellite radiances (Zou et al. 2013; F. Zhang et al. 2016; Minamide and Zhang 2017, 2018; Wu et al. 2019; Y. Zhang et al. 2021), and the next generation of global navigation satellite system (GNSS) radio occultation bending angles (Miller et al. 2023). TC forecasts can also benefit from assimilating data from newer, experimental satellite sources such as the European Space Agency Aeolus mission (Marinescu et al. 2022), the NASA Time-Resolved Observations of Precipitation structure and storm Intensity with a Constellation of Smallsats (TROPICS) mission, and the Cyclone GNSS (CYGNSS) mission (Pu et al. 2023). These kinds of advancements can particularly benefit forecasts without in situ reconnaissance data, which include the vast majority of worldwide TCs.

HWRF progress and improvements were closely linked with the success of HFIP, established by NOAA in 2009 (Gall et al. 2013; Gopalakrishnan et al. 2021). HFIP accelerated the improvement of TC forecasts and warnings and enhanced mitigation and preparedness by increasing confidence in those forecasts. HFIP started as a funding mechanism that was successful and innovative, especially with respect to HWRF developments, through its use of a strong and diverse community of experts both within and external to NOAA. HWRF collaborators included the Developmental Testbed Center (DTC), the University Corporation for Atmospheric Research, the Naval Research Laboratory, several universities and cooperative institutes (Colorado State University and CIRA, The Pennsylvania State University, Princeton University and Cooperative Institute for Modeling the Earth System (CIMES), University at Albany, University of Colorado and CIRES, University of

Miami and Cooperative Institute for Marine and Atmospheric Studies (CIMAS), University of Oklahoma, University of Rhode Island, University of Wisconsin, and CIMSS), and others across the TC community. HFIP's timely and invaluable contributions to HWRF cannot be overstated; HFIP enabled the creation, sharing, and growth of robust, critical components (summarized below) that were then successfully transitioned from research to operations, specifically into HWRF. These efforts have led to hundreds of HWRF-related publications since HFIP began.

HFIP supported a number of critical projects and experiments to advance HWRF for research, enriching the TC community and enabling a clear pathway for the transition of research to operations (R2O). The HWRF experimental modeling system (HWRFx) tested the impact of higher model resolution on TC forecasts, leading to the development and implementation of the second inner-moving nest (Yeh et al. 2011; Zhang et al. 2011). An advanced version of HWRF, called the basin-scale HWRF (HWRF-B), was created to study the impacts of simulating multistorm interactions at high resolution while also supporting advances to DA (X. Zhang et al. 2016; Alaka et al. 2017, 2020, 2022; Poterjoy et al. 2021; Ditchek et al. 2023). HWRF-B demonstrated intensity forecast improvement when storm-following nests were configured for as many TCs as possible. Using HWRF-B, Ditchek et al. (2023) showed the value of using the advanced cycled EnKF for multiple storms simultaneously. DA efforts were also advanced considerably by the HWRF Ensemble Data Assimilation System (HEDAS; Aksoy et al. 2012; Aksoy 2013; Aberson et al. 2015; Aksoy et al. 2022; Christophersen et al. 2017, 2018), as well as experiments that used hybrid global/regional DA (Zhang et al. 2013; Lu et al. 2017a,b; Tong et al. 2018; Feng and Wang 2021; Green et al. 2022) and parent-domain cycling (Poterjoy et al. 2021). Using HWRF ensemble prediction systems, TC uncertainty was explored by perturbing initial and lateral boundary conditions and physics parameterizations (Zhang et al. 2014; Alaka et al. 2019; Zhao and Torn 2022). HFIP supported additional research using HWRF, with topics including rapid intensification (Kieu et al. 2014; Chen and Gopalakrishnan 2015; Smith et al. 2017; Leighton et al. 2018; Green et al. 2021; Ko et al. 2023; Wang et al. 2023), TC predictability (Kieu and Moon 2016; Kieu et al. 2018; Keshavamurthy and Kieu 2021), precipitation (Ko et al. 2020; Stackhouse et al. 2023), the boundary layer (J. A. Zhang et al. 2020), air–sea enthalpy fluxes (Kim et al. 2014, 2022; Ma et al. 2020; Gramer et al. 2022), eyewall replacement cycles (Zhu et al. 2015; Green et al. 2022), convective-scale heating (Biswas et al. 2014), atmospheric radiation (Wu et al. 2023), and ice water content (Wu et al. 2020).

This research and development, as well as subsequent testing and evaluation, would not have been possible without the HFIP real-time experiment (HREx). As discussed in Gopalakrishnan et al. (2021), HREx leverages HFIP investments in research and development and high-performance computing on the NOAA Jet supercomputer to conduct real-time experiments parallel to operations. Nearly all major HWRF developments and R2O efforts, including the first high-resolution version of HWRF, were initially supported as HREx demonstrations on Jet during NATL hurricane seasons.

Though NOAA plans to completely retire HWRF in the near future, its legacy will live on in the next-generation Hurricane Analysis and Forecast System (HAFS; Dong et al. 2020; Hazelton et al. 2021, 2022; Poyer et al. 2023). HAFS, NOAA's new state-of-the-art, multiscale numerical weather prediction model and DA system, was transitioned to operations at NOAA's NCEP on 27 June 2023. Indeed, HWRF imparted many lessons that have guided HAFS development. First and foremost, HWRF's improvement highlighted the vital role of HFIP to set goals and foster collaboration across the TC community. HAFS has already benefited from this strategy.

Long after HWRF retires, one of its enduring impacts will be on how reconnaissance is used to improve TC prediction. As the positive impacts of assimilating reconnaissance data

into HWRF became more apparent over recent years, the model was increasingly used for observing-system experiments that have evaluated the relative impacts of different datasets and collection strategies (Sippel 2018; Wick et al. 2020; Sellwood et al. 2023; Ditchek et al. 2023; Ditchek and Sippel 2023; Sippel et al. 2024). The results from these experiments have been crucial for the development of NOAA's aircraft recapitalization plan (NOAA Office of Marine and Aviation Operations 2022), which outlines NOAA aircraft procurement requirements between now and 2030. Within this plan, NOAA has established a desire to sample TCs earlier and more frequently with the next generation of reconnaissance aircraft.

HWRF also had some shortcomings that have impacted the way HAFS is being developed. For example, one major issue in HWRF was that the model code infrastructure was too rigid, limiting the operational implementation of new components that showed promise to further improve TC forecasts. This problem arose in part because HWRF was developed at a time when high-resolution TC modeling and inner-core DA were considered merely distant operational possibilities. Once the basic HWRF infrastructure was developed, some aspects of it became quite difficult to modify. In response, HAFS infrastructure is more flexible and efficient. Another shortcoming was that HWRF developments focused heavily on reducing VMAX errors over water without enough emphasis on hazards at the coastline or over land, including rainfall, wind gusts, and storm surge. Looking forward, TC hazard predictions are being stressed more holistically in HAFS than in HWRF, and as a result, the guidance available to forecasters to mitigate those hazards is expected to improve and expand.

HAFS is another HFIP product, and many scientists on the HAFS development team worked directly on HWRF. This continuity in institutional knowledge is allowing HAFS to build upon HWRF's accomplishments and address its shortcomings. Indeed, HWRF has left a lasting legacy to hurricane forecasting that will continue propelling HAFS and future models to new heights.

Acknowledgments. This work was supported by NOAA/OAR/AOML. We would especially like to recognize the HWRF Development Team, including the EMC Hurricane Team, the AOML/Hurricane Research Division Modeling Team, and DTC, who spearheaded efforts to develop, transition, implement, and improve the HWRF modeling system for the benefit of NOAA operations and the research community. A great deal of gratitude is extended to HFIP, its numerous collaborators, and its program managers at NWS's Office of Science and Technology Integration for providing the support that aided the acceleration of HWRF research and improvements. We are thankful for the invaluable input of reviewers, including Dr. Mark DeMaria (CSU/CIRA), Dr. Ryan Torn (SUNY/Albany), Dr. Sarah Ditchek (AOML/CIMAS), Dr. Sim Aberson (AOML), and John Cangialosi (NHC), for their efforts to improve this manuscript. HWRF is a community model that can be downloaded from the DTC website (https://dtcenter.org/community-code/hurricane-wrf-hwrf).

Data availability statement. HWRF is a community model that can be downloaded from the DTC website (https://dtcenter.org/community-code/hurricane-wrf-hwrf). All data used in this study can be found on NHC's ATCF website (https://ftp.nhc.noaa.gov/atcf).

References

Aberson, S. D., A. Aksoy, T. Vukicevic, K. J. Sellwood, and X. Zhang, 2015: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using HEDAS: Evaluation of 2008–11 HWRF forecasts. *Mon. Wea. Rev.*, **143**, 511–523, https://doi.org/10.1175/MWR-D-14-00138.1.

——, K. J. Sellwood, and P. A. Leighton, 2017: Calculating dropwindsonde location and time from TEMP-DROP messages for accurate assimilation and analysis. *J. Atmos. Oceanic Technol.*, **34**, 1673–1678, https://doi.org/10.1175/JTECH-D-17-0023.1.

Aksoy, A., 2013: Storm-relative observations in tropical cyclone data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **141**, 506–522, https://doi.org/10.1175/MWR-D-12-00094.1.

——, S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. Lorsolo, S. D. Aberson, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for highresolution data: The impact of airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843–1862, https://doi.org/10.1175/MWR-D-11-00212.1.

——, J. J. Cione, B. A. Dahl, and P. D. Reasor, 2022: Tropical cyclone data assimilation with coyote uncrewed aircraft system, observations, very frequent cycling and a new online quality control technique. *Mon. Wea. Rev.*, **150**, 797–820, https://doi.org/10.1175/MWR-D-21-0124.1.

Alaka, G. J., Jr., X. Zhang, S. G. Gopalakrishnan, S. B. Goldenberg, and F. D. Marks, 2017: Performance of basin-scale HWRF tropical cyclone track forecasts. *Wea. Forecasting*, **32**, 1253–1271, https://doi.org/10.1175/WAF-D-16-0150.1.

——, ——, ——, Z. Zhang, F. D. Marks, and R. Atlas, 2019: Track uncertainty in high-resolution HWRF ensemble forecasts of Hurricane Joaquin. *Wea. Forecasting*, **34**, 1889–1908, https://doi.org/10.1175/WAF-D-19-0028.1.

——, D. Sheinin, B. Thomas, L. Gramer, Z. Zhang, B. Liu, H.-S. Kim, and A. Mehra, 2020: A hydrodynamical atmosphere/ocean coupled modeling system for multiple tropical cyclones. *Atmosphere*, **11**, 869, https://doi.org/10.3390/atmos11080869.

——, 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.

Atlas, R., and Coauthors, 2015: Observing System Simulation Experiments (OSSEs) to evaluate the potential impact of an Optical Autocovariance Wind Lidar (OAWL) on numerical weather prediction. *J. Atmos. Oceanic Technol.*, **32**, 1593–1613, https://doi.org/10.1175/JTECH-D-15-0038.1.

Bao, J.-W., S. G. Gopalakrishnan, S. A. Michelson, F. D. Marks, and M. T. Montgomery, 2012: Impact of physics representations in the HWRFX on simulated hurricane structure and pressure–wind relationships. *Mon. Wea. Rev.*, **140**, 3278–3299, https://doi.org/10.1175/MWR-D-11-00332.1.

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.

Bernardet, L., and Coauthors, 2015: Community support and transition of research to operations for the Hurricane Weather Research and Forecasting Model. *Bull. Amer. Meteor. Soc.*, **96**, 953–960, https://doi.org/10.1175/BAMS-D-13-00093.1.

Biswas, M. K., L. Bernardet, and J. Dudhia, 2014: Sensitivity of hurricane forecasts to cumulus parameterizations in the HWRF model. *Geophys. Res. Lett.*, **41**, 9113–9119, https://doi.org/10.1002/2014GL062071.

——, L. Carson, K. Newman, D. Stark, E. A. Kalina, E. Grell, and J. Frimel, 2018: Community HWRF users' guide V4.0a. HWRF Doc., 163 pp., https://dtcenter.org/sites/default/files/community-code/hwrf/docs/users_guide/HWRF-UG-2018.pdf.

Cangialosi, J. P., 2021: National Hurricane Center forecast verification report: 2020 hurricane season. 77 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2020.pdf.

——, 2023: National Hurricane Center forecast verification report: 2022 hurricane season. 75 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2022.pdf.

——, 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.

Carrasco, C. A., C. W. Landsea, and Y. Lin, 2014: The influence of tropical cyclone size on its intensification. *Wea. Forecasting*, **29**, 582–590, https://doi.org/10.1175/WAF-D-13-00092.1.

Chan, K. T. F., and J. C. L. Chan, 2012: Size and strength of tropical cyclones as inferred from QuikSCAT data. *Mon. Wea. Rev.*, **140**, 811–824, https://doi.org/10.1175/MWR-D-10-05062.1.

Chavas, D. R., and K. A. Emanuel, 2010: A QuikSCAT climatology of tropical cyclone size. *Geophys. Res. Lett.*, **37**, L18816, https://doi.org/10.1029/2010GL044558.

Chen, H., and S. G. Gopalakrishnan, 2015: A study on the asymmetric rapid intensification of Hurricane Earl (2010) using the HWRF system. *J. Atmos. Sci.*, **72**, 531–550, https://doi.org/10.1175/JAS-D-14-0097.1.

Christophersen, H., A. Aksoy, J. P. Dunion, and K. J. Sellwood, 2017: The impact of NASA Global Hawk unmanned aircraft dropwindsonde observations on tropical cyclone track, intensity, and structure: Case studies. *Mon. Wea. Rev.*, **145**, 1817–1830, https://doi.org/10.1175/MWR-D-16-0332.1.

——, ——, ——, and S. D. Aberson, 2018: Composite impact of Global Hawk unmanned aircraft dropwindsondes on tropical cyclone analyses and forecasts. *Mon. Wea. Rev.*, **146**, 2297–2314, https://doi.org/10.1175/MWR-D-17-0304.1.

DeMaria, M., C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, https://doi.org/10.1175/BAMS-D-12-00240.1.

——, 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.

——, and Coauthors, 2022: The National Hurricane Center tropical cyclone model guidance suite. *Wea. Forecasting*, **37**, 2141–2159, https://doi.org/10.1175/WAF-D-22-0039.1.

Ditchek, S. D., and J. A. Sippel, 2023: A comparison of the impacts of inner-core, in-vortex, and environmental dropsondes on tropical cyclone forecasts during the 2017–2020 hurricane seasons. *Wea. Forecasting*, **38**, 2169–2187, https://doi.org/10.1175/WAF-D-23-0055.1.

——, ——, G. J. Alaka Jr., S. B. Goldenberg, and L. Cucurull, 2023: A systematic assessment of the overall dropsonde impact during the 2017–20 hurricane seasons using the basin-scale HWRF. *Wea. Forecasting*, **38**, 789–816, https://doi.org/10.1175/WAF-D-22-0102.1.

Domingues, R., and Coauthors, 2021: Ocean conditions and the intensification of three major Atlantic hurricanes in 2017. *Mon. Wea. Rev.*, **149**, 1265–1286, https://doi.org/10.1175/MWR-D-20-0100.1.

Dong, J., and Coauthors, 2017: Impact of assimilating underwater glider data on Hurricane Gonzalo (2014) forecasts. *Wea. Forecasting*, **32**, 1143–1159, https://doi.org/10.1175/WAF-D-16-0182.1.

——, 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.

Feng, J., and X. Wang, 2021: Impact of increasing horizontal and vertical resolution during the HWRF hybrid EnVar data assimilation on the analysis and prediction of Hurricane Patricia (2015). *Mon. Wea. Rev.*, **149**, 419–441, https://doi.org/10.1175/MWR-D-20-0144.1.

Gall, R., and Coauthors, 2010: 2009 HFIP R&D activities summary: Accomplishments, lessons learned, and challenges. NOAA HFIP Tech. Rep. HFIP2010-1,

60 pp., https://hfip.org/sites/default/files/documents/2009-hfip-report-combined-teams.pdf.

——, 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.

Gemmill, W., B. Katz, and X. Li, 2007: Daily real-time, global sea surface temperature—High-resolution analysis: RTG_SST_HR. NCEP/EMC Office Note, 39 pp., https://polar.ncep.noaa.gov/mmab/papers/tn260/MMAB260.pdf.

Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nunez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, **293**, 474–479, https://doi.org/10.1126/science.1060040.

——, S. G. Gopalakrishnan, V. Tallapragada, T. Quirino, F. Marks Jr., S. Trahan, X. Zhang, and R. Atlas, 2015: The 2012 triply nested, high-resolution operational version of the Hurricane Weather Research and Forecasting Model (HWRF): Track and intensity forecast verifications. *Wea. Forecasting*, **30**, 710–729, https://doi.org/10.1175/WAF-D-14-00098.1.

Gopalakrishnan, S., and Coauthors, 2010: Hurricane Weather Research and Forecasting (HWRF) model scientific documentation. HWRF Doc., 96 pp., https://www.aoml.noaa.gov/ftp/hrd/annane/osse/upgrade_to_operational/HWRFScientificDocumentation_v3.4a.pdf.

——, and Coauthors, 2020: 2019 HFIP R&D activities summary: Recent results and operational implementation. NOAA HFIP Tech. Rep. HFIP2020-1, 42 pp., https://www.hfip.org/documents/HFIP_AnnualReport_FY2019.pdf.

——, and Coauthors, 2021: 2020 HFIP R&D activities summary: Recent results and operational implementation. NOAA HFIP Tech. Rep. HFIP2021-1, 49 pp., https://hfip.org/sites/default/files/documents/hfip-annual-report-2020-final_0.pdf.

Gopalakrishnan, S. G., F. Marks Jr., X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2011: The experimental HWRF system: A study on the influence of horizontal resolution on the structure and intensity changes in tropical cyclones using an idealized framework. *Mon. Wea. Rev.*, **139**, 1762–1784, https://doi.org/10.1175/2010MWR3535.1.

——, S. Goldenberg, T. Quirino, X. Zhang, F. Marks Jr., K.-S. Yeh, R. Atlas, and V. Tallapragada, 2012: Toward improving high-resolution numerical hurricane forecasting: Influence of model horizontal grid resolution, initialization, and physics. *Wea. Forecasting*, **27**, 647–666, https://doi.org/10.1175/WAF-D-11-00055.1.

——, F. Marks Jr., 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.

Gramer, L. J., J. A. Zhang, G. Alaka, A. Hazelton, and S. Gopalakrishnan, 2022: Coastal downwelling intensifies landfalling hurricanes. *Geophys. Res. Lett.*, **49**, e2021GL096630, https://doi.org/10.1029/2021GL096630.

Green, A., S. G. Gopalakrishnan, G. J. Alaka, and S. Chiao, 2021: Understanding the role of mean and eddy momentum transport in the rapid intensification of Hurricane Irma (2017) and Hurricane Michael (2018). *Atmosphere*, **12**, 492, https://doi.org/10.3390/atmos12040492.

Green, T., X. Wang, and X. Lu, 2022: Impact of assimilating ground-based and airborne radar observations for the analysis and prediction of the eyewall replacement cycle of Hurricane Matthew (2016) using HWRF hybrid 3DEnVar system. *Mon. Wea. Rev.*, **150**, 1157–1175, https://doi.org/10.1175/MWR-D-21-0234.1.

Halperin, D. J., and R. D. Torn, 2018: Diagnosing conditions associated with large intensity forecast errors in the Hurricane Weather Research and Forecasting (HWRF) model. *Wea. Forecasting*, **33**, 239–266, https://doi.org/10.1175/WAF-D-17-0077.1.

Han, J., M. L. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. K. Fletcher, and C. S. Bretherton, 2016: Implementation in the NCEP GFS of a hybrid eddy-diffusivity mass-flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Wea. Forecasting*, **31**, 341–352, https://doi.org/10.1175/WAF-D-15-0053.1.

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

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.

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

Keshavamurthy, K., and C. Kieu, 2021: Dependence of tropical cyclone intrinsic intensity variability on the large-scale environment. *Quart. J. Roy. Meteor. Soc.*, **147**, 1606–1625, https://doi.org/10.1002/qj.3984.

Kieu, C., V. Tallapragada, and W. Hogsett, 2014: Vertical structure of tropical cyclones at onset of the rapid intensification in the HWRF model. *Geophys. Res. Lett.*, **41**, 3298–3306, https://doi.org/10.1002/2014GL059584.

——, K. Keshavamurthy, V. Tallapragada, S. Gopalakrishnan, and S. Trahan, 2018: On the growth of intensity forecast errors in the operational Hurricane Weather Research and Forecasting (HWRF) model. *Quart. J. Roy. Meteor. Soc.*, **144**, 1803–1819, https://doi.org/10.1002/qj.3344.

Kieu, C. Q., and Z. Moon, 2016: Hurricane intensity predictability. *Bull. Amer. Meteor. Soc.*, **97**, 1847–1857, https://doi.org/10.1175/BAMS-D-15-00168.1.

Kim, H.-S., C. Lozano, V. Tallapragada, D. Iredell, D. Sheinin, H. L. Tolman, V. M. Gerald, and J. Sims, 2014: Performance of ocean simulations in the coupled HWRF–HYCOM model. *J. Atmos. Oceanic Technol.*, **31**, 545–559, https://doi.org/10.1175/JTECH-D-13-00013.1.

——, J. Meixner, B. Thomas, B. G. Reichl, B. Liu, A. Mehra, and A. Wallcraft, 2022: Skill assessment of NCEP three-way coupled HWRF–HYCOM–WW3 Modeling System: Hurricane Laura case study. *Wea. Forecasting*, **37**, 1309–1331, https://doi.org/10.1175/WAF-D-21-0191.1.

Ko, M.-C., F. D. Marks, G. J. Alaka Jr., and S. G. Gopalakrishnan, 2020: Evaluation of Hurricane Harvey (2017) rainfall in deterministic and probabilistic HWRF forecasts. *Atmosphere*, **11**, 666, https://doi.org/10.3390/atmos11060666.

——, X. Chen, M. Kubat, and S. Gopalakrishnan, 2023: The development of a consensus machine learning model for hurricane rapid intensification forecasts with Hurricane Weather Research and Forecasting (HWRF) data. *Wea. Forecasting*, **38**, 1253–1270, https://doi.org/10.1175/WAF-D-22-0217.1.

Kurihara, Y., R. E. Tuleya, and M. A. Bender, 1998: The GFDL Hurricane Prediction System and its performance in the 1995 hurricane season. *Mon. Wea. Rev.*, **126**, 1306–1322, https://doi.org/10.1175/1520-0493(1998)126<1306:TGHPSA>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.

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

Le Hénaff, M., and Coauthors, 2021: The role of the Gulf of Mexico ocean conditions in the intensification of Hurricane Michael (2018). *J. Geophys. Res. Oceans*, **126**, e2020JC016969, https://doi.org/10.1029/2020JC016969.

Leighton, H., S. Gopalakrishnan, J. A. Zhang, R. F. Rogers, Z. Zhang, and V. Tallapragada, 2018: Azimuthal distribution of deep convection, environmental factors, and tropical cyclone rapid intensification: A perspective from HWRF ensemble forecasts of Hurricane Edouard (2014). *J. Atmos. Sci.*, **75**, 275–295, https://doi.org/10.1175/JAS-D-17-0171.1.

Lewis, W., C. Velden, and D. Stettner, 2020: Strategies for assimilating high-density atmospheric motion vectors into a regional tropical cyclone forecast model (HWRF). *Atmosphere*, **11**, 673, https://doi.org/10.3390/atmos11060673.

Li, J., J. Li, C. Velden, P. Wang, T. J. Schmit, and J. Sippel, 2020: Impact of rapid-scan-based dynamical information from GOES-16 on HWRF hurricane forecasts. *J. Geophys. Res. Atmos.*, **125**, e2019JD031647, https://doi.org/10.1029/2019JD031647.

Lim, A. H. N., J. A. Jung, S. E. Nebuda, J. M. Daniels, W. Bresky, M. Tong, and V. Tallapragada, 2019: Tropical cyclone forecasts impact assessment from the assimilation of hourly visible, shortwave, and clear-air water vapor atmospheric motion vectors in HWRF. *Wea. Forecasting*, **34**, 177–198, https://doi.org/10.1175/WAF-D-18-0072.1.

Liu, Q., and Coauthors, 2020: Vortex initialization in the NCEP operational hurricane models. *Atmosphere*, **11**, 968, https://doi.org/10.3390/atmos11090968

Lu, X., X. Wang, M. Tong, and V. Tallapragada, 2017a: GSI-based, continuously cycled, dual-resolution hybrid ensemble–variational data assimilation system for HWRF: System description and experiments with Edouard (2014). *Mon. Wea. Rev.*, **145**, 4877–4898, https://doi.org/10.1175/MWR-D-17-0068.1.

——, ——, Y. Li, M. Tong, and X. Ma, 2017b: GSI-based ensemble-variational hybrid data assimilation for HWRF for hurricane initialization and prediction: Impact of various error covariances for airborne radar observation assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 223–239, https://doi.org/10.1002/qj.2914.

Ma, Z., and Coauthors, 2020: Investigating the impact of high-resolution land–sea masks on hurricane forecasts in HWRF. *Atmosphere*, **11**, 888, https://doi.org/10.3390/atmos11090888.

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.

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

Marinescu, P. J., L. Cucurull, K. Apodaca, L. Bucci, and I. Genkova, 2022: The characterization and impact of Aeolus wind profile observations in NOAA's regional tropical cyclone model (HWRF). *Quart. J. Roy. Meteor. Soc.*, **148**, 3491–3508, https://doi.org/10.1002/qj.4370.

Mehra, A., and I. Rivin, 2010: A real time ocean forecast system for the North Atlantic Ocean. *Terr. Atmos. Ocean. Sci.*, **21**, 211–228, https://doi.org/10.3319/TAO.2009.04.16.01(IWNOP).

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

Miller, W. J., Y. Chen, S. Ho, and X. Shao, 2023: Evaluating the impacts of COSMIC-2 GNSS RO bending angle assimilation on Atlantic hurricane forecasts using the HWRF Model. *Mon. Wea. Rev.*, **151**, 1821–1847, https://doi.org/10.1175/MWR-D-22-0198.1.

Minamide, M., and F. Zhang, 2017: Adaptive observation error inflation for assimilating all-sky satellite radiance. *Mon. Wea. Rev.*, **145**, 1063–1081, https://doi.org/10.1175/MWR-D-16-0257.1.

——, and ——, 2018: Assimilation of all-sky infrared radiances from Himawari-8 and impacts of moisture and hydrometer initialization on convection-permitting tropical cyclone prediction. *Mon. Wea. Rev.*, **146**, 3241–3258, https://doi.org/10.1175/MWR-D-17-0367.1.

NOAA Office of Marine and Aviation Operations, 2022: Update, status, and implementation of the NOAA aircraft plan: Building and sustaining the 21st century fleet. 44 pp., https://www.omao.noaa.gov/sites/default/files/2022-12/2022_Aircraft_Recapitalization_Plan_Final.pdf.

Pasch, R. J., R. Berg, D. P. Roberts, and P. P. Papin, 2020: Tropical cyclone report: Hurricane Laura (AL132020), 20–29 August 2020. NHC Tech. Rep., 75 pp., https://www.nhc.noaa.gov/data/tcr/AL132020_Laura.pdf.

Poterjoy, J., G. J. Alaka Jr., and H. R. Winterbottom, 2021: The irreplaceable utility of sequential data assimilation for numerical weather prediction system development: Lessons learned from an experimental HWRF system. *Wea. Forecasting*, **36**, 661–677, https://doi.org/10.1175/WAF-D-20-0204.1.

Poyer, A., and Coauthors, 2023: 2021-2022 HFIP R&D activities summary: Recent results and operational implementation. NOAA HFIP Tech. Rep. HFIP2023-1, 73 pp., https://doi.org/10.25923/exgj-1n68.

Pu, Z., W. J. Blackwell, and C. S. Ruf, 2023: Assimilation of TROPICS and CYGNSS satellite data for improved numerical prediction of tropical cyclones. *2023 Annual Meeting*, San Francisco, CA, Amer. Geophys. Union, A52F-03, https://agu.confex.com/agu/fm23/meetingapp.cgi/Paper/1364432.

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.

Ruan, Z., and Q. Wu, 2022: Relationship between size and intensity in North Atlantic tropical cyclones with steady radii of maximum wind. *Geophys. Res. Lett.*, **49**, e2021GL095632, https://doi.org/10.1029/2021GL095632.

Sawada, M., Z. Ma, A. Mehra, V. Tallapragada, R. Oyama, and K. Shimoji, 2019: Impacts of assimilating high-resolution atmospheric motion vectors derived from Himawari-8 on tropical cyclone forecast in HWRF. *Mon. Wea. Rev.*, **147**, 3721–3740, https://doi.org/10.1175/MWR-D-18-0261.1.

Sellwood, K. J., J. A. Sippel, and A. Aksoy, 2023: Assimilation of coyote small uncrewed aircraft system observations in Hurricane Maria (2017) using operational HWRF. *Wea. Forecasting*, **38**, 901–919, https://doi.org/10.1175/WAF-D-22-0214.1.

Sippel, J., 2018: The value of extra dropsondes deployed by AFRES. *72nd Interdepartmental Hurricane Conf.*, Miami, FL, Interagency Council on Advancing Meteorological Services, https://www.icams-portal.gov/meetings/TCORF/ihc18/session_8/8-8-sippel.pdf.

Sippel, J. A., X. Wu, S. D. Ditchek, V. Tallapragada, and D. T. Kleist, 2022: Impacts of assimilating additional reconnaissance data on operational GFS tropical cyclone forecasts. *Wea. Forecasting*, **37**, 1615–1639, https://doi.org/10.1175/WAF-D-22-0058.1.

——, S. D. Ditchek, K. Ryan, and C. W. Landsea, 2024: The G-IV inner circumnavigation: A story of successful organic interactions between research and operations at NOAA. *Bull. Amer. Meteor. Soc.*, **105**, E218–E232, https://doi.org/10.1175/BAMS-D-23-0084.1.

Smith, R. K., J. A. Zhang, and M. T. Montgomery, 2017: The dynamics of intensification in a Hurricane Weather Research and Forecasting simulation of Hurricane Earl (2010). *Quart. J. Roy. Meteor. Soc.*, **143**, 293–308, https://doi.org/10.1002/qj.2922.

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.

Tallapragada, V., 2016: Overview of the NOAA/NCEP operational Hurricane Weather Research and Forecast (HWRF) modelling system. *Advanced Numerical Modeling and Data Assimilation Techniques for Tropical Cyclone Prediction*, U. C. Mohanty and S. G. Gopalakrishnan, Eds., Springer, 51–106.

——, 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.

——, and Coauthors, 2016: Forecasting tropical cyclones in the western North Pacific basin using the NCEP operational HWRF model: Model upgrades and evaluation of real-time performance in 2013. *Wea. Forecasting*, **31**, 877–894, https://doi.org/10.1175/WAF-D-14-00139.1.

Tong, M., and Coauthors, 2018: Impact of assimilating aircraft reconnaissance observations on tropical cyclone initialization and prediction using operational HWRF and GSI ensemble–variational hybrid data assimilation. *Mon. Wea. Rev.*, **146**, 4155–4177, https://doi.org/10.1175/MWR-D-17-0380.1.

Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Wea. Forecasting*, **27**, 715–729, https://doi.org/10.1175/WAF-D-11-00085.1.

Trabing, B. C., and M. M. Bell, 2020: Understanding error distributions of hurricane intensity forecasts during rapid intensity changes. *Wea. Forecasting*, **35**, 2219–2234, https://doi.org/10.1175/WAF-D-19-0253.1.

Velden, C., W. E. Lewis, W. Bresky, D. Stettner, J. Daniels, and S. Wanzong, 2017: Assimilation of high-resolution satellite-derived atmospheric motion vectors:

Impact on HWRF forecasts of tropical cyclone track and intensity. *Mon. Wea. Rev.*, **145**, 1107–1125, https://doi.org/10.1175/MWR-D-16-0229.1.

Wang, W., J. A. Sippel, S. Abarca, L. Zhu, B. Liu, Z. Zhang, A. Mehra, and V. Tallapragada, 2018: Improving NCEP HWRF simulations of surface wind and inflow angle in the eyewall area. *Wea. Forecasting*, **33**, 887–898, https://doi.org/10.1175/WAF-D-17-0115.1.

——, B. Liu, L. Zhu, Z. Zhang, A. Mehra, and V. Tallapragada, 2021: A new horizontal mixing-length formulation for numerical simulations of tropical cyclones. *Wea. Forecasting*, **36**, 679–695, https://doi.org/10.1175/WAF-D-20-0134.1.

——, L. Zhu, B. Liu, Z. Zhang, A. Mehra, and V. Tallapragada, 2023: A forecast cycle–based evaluation for tropical cyclone rapid intensification forecasts by the operational HWRF Model. *Wea. Forecasting*, **38**, 125–138, https://doi.org/10.1175/WAF-D-22-0007.1.

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.

Wick, G. A., and Coauthors, 2020: NOAA's Sensing Hazards with Operational Unmanned Technology (SHOUT) experiment observations and forecast impacts. *Bull. Amer. Meteor. Soc.*, **101**, E968–E987, https://doi.org/10.1175/BAMS-D-18-0257.1.

Wu, S.-N., B. J. Soden, and G. J. Alaka Jr., 2020: Ice water content as a precursor to tropical cyclone rapid intensification. *Geophys. Res. Lett.*, **47**, e2020GL089669, https://doi.org/10.1029/2020GL089669.

——, ——, and ——, 2023: The influence of radiation on the prediction of tropical cyclone intensification in a forecast model. *Geophys. Res. Lett.*, **50**, e2022GL099442, https://doi.org/10.1029/2022GL099442.

Wu, T.-C., M. Zupanski, L. D. Grasso, C. D. Kummerow, and S.-A. Boukabara, 2019: All-sky radiance assimilation of ATMS in HWRF: A demonstration study. *Mon. Wea. Rev.*, **147**, 85–106, https://doi.org/10.1175/MWR-D-17-0337.1.

Yablonsky, R. M., and I. Ginis, 2008: Improving the ocean initialization of coupled hurricane–ocean models using feature-based data assimilation. *Mon. Wea. Rev.*, **136**, 2592–2607, https://doi.org/10.1175/2007MWR2166.1.

——, ——, B. Thomas, V. Tallapragada, D. Sheinin, and L. Bernardet, 2015: Description and analysis of the ocean component of NOAA's operational Hurricane Weather Research and Forecasting model (HWRF). *J. Atmos. Oceanic Technol.*, **32**, 144–163, https://doi.org/10.1175/JTECH-D-14-00063.1.

Yeh, K.-S., X. Zhang, S. Gopalakrishnan, S. Aberson, R. Rogers, F. Marks, and R. Atlas, 2011: Performance of the experimental HWRF in the 2008 hurricane season. *Nat. Hazards*, **63**, 1439–1449, https://doi.org/10.1007/s11069-011-9787-7.

Zawislak, J., and Coauthors, 2022: Accomplishments of NOAA's airborne hurricane field program and a broader future approach to forecast improvement. *Bull. Amer. Meteor. Soc.*, **103**, E311–E338, https://doi.org/10.1175/BAMS-D-20-0174.1.

Zhang, F., M. Minamide, and E. E. Clothiaux, 2016: Potential impacts of assimilating all-sky infrared satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophys. Res. Lett.*, **43**, 2954–2963, https://doi.org/10.1002/2016GL068468.

Zhang, J. A., and F. D. Marks, 2015: Effects of horizontal diffusion on tropical cyclone intensity change and structure in idealized three-dimensional numerical simulations. *Mon. Wea. Rev.*, **143**, 3981–3995, https://doi.org/10.1175/MWR-D-14-00341.1.

——, 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.

——, F. D. Marks, J. A. Sippel, R. F. Rogers, X. Zhang, S. G. Gopalakrishnan, Z. Zhang, and V. Tallapragada, 2018: Evaluating the impact of improvement in the horizontal diffusion parameterization on hurricane prediction in the operational Hurricane Weather Research and Forecast (HWRF) model. *Wea. Forecasting*, **33**, 317–329, https://doi.org/10.1175/WAF-D-17-0097.1.

——, E. A. Kalina, M. K. Biswas, R. F. Rogers, P. Zhu, and F. D. Marks, 2020: A review and evaluation of planetary boundary layer parameterizations in hurricane weather research and forecasting model using idealized simulations and observations. *Atmosphere*, **11**, 1091, https://doi.org/10.3390/atmos11101091.

Zhang, M., M. Zupanski, M.-J. Kim, and J. A. Knaff, 2013: Assimilating AMSU-A radiances in the TC core area with NOAA operational HWRF (2011) and a hybrid data assimilation system: Danielle (2010). *Mon. Wea. Rev.*, **141**, 3889–3907, https://doi.org/10.1175/MWR-D-12-00340.1.

Zhang, S., Z. Pu, D. J. Posselt, and R. Atlas, 2017: Impact of CYGNSS ocean surface wind speeds on numerical simulations of a hurricane in observing system simulation experiments. *J. Atmos. Oceanic Technol.*, **34**, 375–383, https://doi.org/10.1175/JTECH-D-16-0144.1.

Zhang, X., T. Quirino, K.-S. Yeh, S. Gopalakrishnan, F. Marks, S. Goldenberg, and S. Aberson, 2011: HWRFx: Improving hurricane forecasts with high-resolution modeling. *Comput. Sci. Eng.*, **13**, 13–21, https://doi.org/10.1109/MCSE.2010.121.

——, S. G. Gopalakrishnan, S. Trahan, T. S. Quirino, Q. Liu, Z. Zhang, G. Alaka, and V. Tallapragada, 2016: Representing multiple scales in the Hurricane Weather Research and Forecasting modeling system: Design of multiple sets of movable multilevel nesting and the basin-scale HWRF forecast application. *Wea. Forecasting*, **31**, 2019–2034, https://doi.org/10.1175/WAF-D-16-0087.1.

Zhang, Y., and Coauthors, 2021: Ensemble-based assimilation of satellite all-sky microwave radiances improves intensity and rainfall predictions for Hurricane Harvey (2017). *Geophys. Res. Lett.*, **48**, e2021GL096410, https://doi.org/10.1029/2021GL096410.

Zhang, Z., V. Tallapragada, C. Kieu, S. Trahan, and W. Wang, 2014: HWRF based ensemble prediction system using perturbations from GEFS and stochastic convective trigger function. *Trop. Cyclone Res. Rev.*, **3**, 145–161, https://doi.org/10.6057/2014TCRR03.02.

——, and Coauthors, 2020: The impact of stochastic physics-based hybrid GSI/EnKF data assimilation on hurricane forecasts using EMC operational hurricane modeling system. *Atmosphere*, **11**, 801, https://doi.org/10.3390/atmos11080801.

——, J. A. Zhang, G. J. Alaka Jr., K. Wu, A. Mehra, and V. Tallapragada, 2021: A statistical analysis of high-frequency track and intensity forecasts from NOAA's operational Hurricane Weather Research and Forecasting (HWRF) modeling system. *Mon. Wea. Rev.*, **149**, 3325–3339, https://doi.org/10.1175/MWR-D-21-0021.1.

——, and Coauthors, 2023: A review of recent advances (2018–2021) on tropical cyclone intensity change from operational perspectives, part 1: Dynamical model guidance. *Trop. Cyclone Res. Rev.*, **12**, 30–49, https://doi.org/10.1016/j.tcrr.2023.05.004.

Zhao, X., and R. D. Torn, 2022: Evaluation of Independent Stochastically Perturbed Parameterization Tendency (iSPPT) scheme on HWRF-based ensemble tropical cyclone intensity forecasts. *Mon. Wea. Rev.*, **150**, 2659–2674, https://doi.org/10.1175/MWR-D-21-0303.1.

Zhu, P., and Coauthors, 2015: Impact of subgrid-scale processes on eyewall replacement cycle of tropical cyclones in HWRF system. *Geophys. Res. Lett.*, **42**, 10 027–10 036, https://doi.org/10.1002/2015GL066436.

Zou, X., F. Weng, B. Zhang, L. Lin, Z. Qin, and V. Tallapragada, 2013: Impacts of assimilation of ATMS data in HWRF on track and intensity forecasts of 2012 four landfall hurricanes. *J. Geophys. Res. Atmos.*, **118**, 11 558–11 576, https://doi.org/10.1002/2013JD020405.