2021-2022 HFIP R&D Activities Summary: Recent Results and Operational Implementation

November 2023

HFIP Technical Report: HFIP2023-1

https://doi.org/10.25923/exgj-1n68

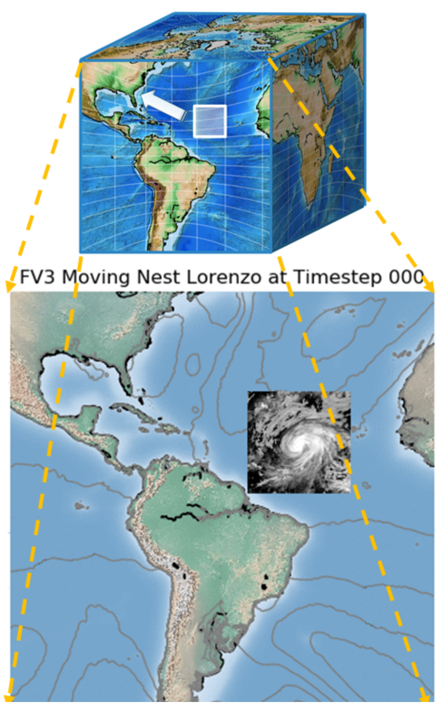

Image on the cover page shows FV3 moving nest on tile 6 in global domain.

2021-2022 HFIP R&D Activities Summary: Recent Results and Operational Implementation

A. Poyer[7], W. Komaromi[3], S. Gopalakrishnan[11], L. Wolf[7,18], F. Marks[11], G. Alaka Jr.[11], J. Anderson[3], V. Tallapragada[8], M. Brennan[9], A. Mehra[8], X. Zhang[11], Z. Zhang[8], A. Hazelton[11,23], D. A. Zelinsky[9], J. L. Franklin[9,19], A. Aksoy[11,23], C. Alexander[13], M. Bender[17], L. Bernardet[2,13], M. Biswas[2,6], J. Cangialosi[9], M. DeMaria[1,9], R. Dunlap[5], M. Ek[2,6], G. Eosco[16], L.Gramer[11,23], L. Harris[15], J. S. Hilderbrand[10], E. Kalina[2,13,22], H.-S. Kim[12], P. Kucera[6], B. Liu[3,8], P. McCaslin[13], T. Marchok[15], J. Moskaitis[4], K. Musgrave[1], L. Nance[2,6], K. Newman[2,6], M. Onderlinde[9], W. Ramstrom[11,23], D. Rosen[5], J. Sims[7], J. Sippel[11], D. Stern[21], R. Torn[20], X. Wang[24], W. Wang[3,8], Y. Weng[3,8], B. C. Zachry[9], C. Zhang[3,8], M. Zhang[2], and L. Zhu[3,8].

September 2023

[1] Colorado State University, CIRA, Fort Collins, CO
[2] Developmental Testbed Center (DTC), Boulder, CO
[3] I. M. Systems Group (IMSG) Inc, Rockville, MD
[4] Naval Research Laboratory, Washington D.C.
[5] NCAR/CGD/ESMF, Boulder, CO
[6] NCAR/RAL/JNT, Boulder, CO
[7] NOAA/NWS/OSTI, Silver Spring, MD
[8] NOAA/NWS/NCEP/EMC, College Park, MD
[9] NOAA/NWS/NCEP/NHC, Miami, FL
[10] NOAA/NWS/OPPSD, Silver Spring, MD
[11] NOAA/OAR/AOML/HRD, Miami, FL
[12] NOAA/OAR/AOML/PhOD, Miami, FL
[13] NOAA/OAR/ESRL/GSL, Boulder, CO
[14] NOAA/OAR/ESRL/PSL, Boulder, CO
[15] NOAA/OAR/GFDL, Princeton, NJ
[16] NOAA/OAR/WPO, Silver Spring, MD
[17] Princeton University, CIMES, Princeton, NJ
[18] Science and Technology Corporation (STC), Columbia, MD
[19] Systems Research Group, Miami FL
[20] University at Albany, State University of New York, Albany, NY
[21] University Corporation for Atmospheric Research, Monterey, CA
[22] University of Colorado, CIRES, Boulder, CO
[23] University of Miami, CIMAS, Miami, FL
[24] University of Oklahoma, Norman, OK

This page is intentionally left blank.

Table of Contents

This page is intentionally left blank.

List of Figures

Figure 12: Track (a) and intensity (b) forecast skills from the 2021 season for HWRF (purple), HMON (green), HAFS-A (cyan), HAFS-B (orange), HAFS-D (red), and HAFS-E (yellow).

Figure 13: Track (a) and intensity (b) forecast errors from the 2021 season for HWRF (purple), HAFS-A (cyan), and GFS(blue).

Figure 14: Wind-pressure relationship from HAFS-A (cyan), HWRF (purple), and best track (black).

Figure 15: Atlantic (a) mean track forecast errors and (b) mean absolute intensity errors from the 2021 season for HAFS-A (dark green), HAFS-B (red), GFDL T-SHiELD (light blue), operational GFS (dark blue), operational HWRF (purple), and operational HMON (light green).

Figure 16: Eastern North Pacific (a) track forecast errors from HAFS-B (the global domain, red), operational GFS (blue), and GFDL T-SHiELD (cyan); and (b) intensity forecast bias from HAFS-B (the global domain, red), operational GFS (blue), and operational HWRF (purple).

Figure 17: Track and intensity forecast errors from the experiments with (blue) and without (green) enhanced GOES-R AMV data assimilated.

Figure 18: Track and intensity forecast errors from the 2021 season for HWRF (purple), HAFS-A (green), and HAFS-D (blue).

Figure 19: Track and intensity forecast skills from the 2021 season for unperturbed ensemble control (HF00, blue), HAFS-A (green), all member ensemble mean (HFMN, red), subset ensemble mean (HS12, purple), and HAFS-E host model GFSF (AEMN, orange).

Figure 20: Examples of HFIP post processing and verification accomplishments in 2021: a) NHC director Ken Graham uses 3D graphics of aircraft Tail Doppler Radar during a public Facebook Live briefing ahead of the landfall of Hurricane Elsa; b) Experimental forecast using the "WTCM"-based wind speed probability model, which realistically highlights large differences between land and water points; c) output from a machine learning technique which aims to better quantify forecast uncertainty; d) Observations of eddy diffusivity vs wind speed at 500 m vs model output from the HAFS with a variety of parameterization schemes.

Figure 21: The purposeful, complementary design behind the projects.

Figure 22: Ongoing triangulation efforts to find similarities across projects.

Figure 23: Evolution of inner-core data assimilation techniques under HFIP.

Figure 24: Observed track forecast error (nmi; bar graph) at 48-h lead time, pre-HFIP in 2007, when HFIP goals reached the year 10 mark in 2017, and the Weather Act goals reached year 5 in 2022, compared to the original 10-year goal, the Weather Act 5-year goal, and the Weather Act 10-year goal (black, red, and green stars, respectively).

Figure 25: As in Figure 24, but for intensity error (kt).

Figure 26: Observed intensity forecast error (kt; bar graph) at 48-h lead time, conditional upon rapid intensity being observed. Bars correspond to pre-HFIP in 2007, when HFIP goals reached the year 10 mark in 2017, and the Weather Act goals reached year 5 in 2022, compared to the Weather Act 5-year goal, and the Weather Act 10-year goal (red and green stars, respectively).

List of Tables

Executive Summary

This technical report describes the activities and results of the Hurricane Forecast Improvement Program (HFIP) that occurred in the 2021 and 2022 hurricane seasons. The major focus of this report is the development of the Hurricane Analysis and Forecast System (HAFS) as a hurricane application of the Unified Forecast System (UFS) and its first operational implementation. As 2022 marks five years since the Weather Act of 2017 established new 5-year goals for HFIP, we will pay particular attention to progress HFIP has made in meeting these goals. We also report some significant improvements in forecasting rapid intensification (RI) of tropical cyclones (TCs), one of the primary goals of HFIP that was set in the beginning of the program.

The 2021 North Atlantic hurricane season was above average. There were 21 named storms, of which 7 developed into hurricanes, with 4 of those becoming major hurricanes. There were 8 landfalls in the U.S. from 6 tropical storms and 2 hurricanes. In the eastern North Pacific, there were 19 named storms, of which 8 developed into hurricanes with 2 major hurricanes. There were 6 observed RI events, defined as an intensification of 30 kt or more in 24 hours, from 5 tropical cyclones (Elsa, Grace, Ida, Larry and Sam) in the Atlantic basin and 3 reported events of RI in the eastern North Pacific (Felicia, Linda and Olaf).

The 2022 North Atlantic hurricane season was an average, but destructive, season, with 14 named storms, 8 hurricanes, and 2 major hurricanes. Hurricane Fiona was a category 4 hurricane that caused significant damage to Puerto Rico, the Dominican Republic, and Nova Scotia. Hurricane Ian was a category 5 hurricane at its peak, that slammed into southwest Florida at category 4 intensity, causing widespread damage. The 2022 eastern North Pacific hurricane season was active, with 19 named storms, 10 hurricanes, and 4 major hurricanes. Seven eastern North Pacific tropical cyclones made landfall, including two which crossed over from the Atlantic basin. There were 4 RI events from 3 tropical cyclones in the North Atlantic (Danielle, Martin, and Ian), and 8 RI events in the eastern North Pacific (Agatha, Blas, Darby, Estelle, Howard, Kay, Orlene, Roslyn).

The major highlights of 2021 were:

1. Significant progress was made toward meeting the HFIP RI performance targets. Comparison of the HFIP RI performance metric for 2019-21 against the 2015-17 baselines is encouraging. At 24 h the baseline error was reduced by 34%, at 48 h the baseline error was reduced by 27% , and at 72 h the baseline error was reduced by 27%.

2. A major accomplishment in 2021 was the accelerated development of NOAA’s next-generation HAFS through the Bipartisan Budget Act of 2018 - also referred to as the Hurricane Supplemental Appropriations funding. In particular, significant progress was made with the development of the moving nest in the global and regional versions of HAFS, and the regional development with one moving nest, capable of automatically tracking one hurricane at a time.

3. For the intensity guidance, the Hurricane Weather Research and Forecasting (HWRF) Model did better at early lead times until 48 h but lagged behind Hurricanes in a Multi-scale Ocean-coupled Non- hydrostatic (HMON) model in the North Atlantic basin. In the eastern North Pacific basin, HWRF was comparable to HMON beyond 72 h but lagged behind for early lead times.

4. HWRF was the second-best track model behind the GFS and had reasonable track skill in the Atlantic basin. HMON lagged behind HWRF in track skill. In the east Pacific basin, HWRF lagged behind GFS at most lead times. HMON intensity skill was better than HWRF at most lead times.

5. Four configurations of the HAFS model were run as part of the 2021 HFIP Real-time Experiment (HREx). They were (i) the ocean-coupled, high-resolution regional Limited Area Model; (ii) global model with a high-resolution nest; (iii) regional HAFS with data assimilation (DA); and (iv) HAFS ensemble with 21 members. In general, all the fours version showed improved performance over HWRF at various lead times The regional version showed some significant improvements over HWRF both in terms of track and intensity at several lead times, hence offering promise for further developments. The moving nest was implemented in this regional version and was tested during the 2022 hurricane season in advance of the operational implementation of HAFS in 2023.

6. Disaster Supplemental Appropriations provided a unique and important opportunity to integrate social, behavioral and economic sciences (SBES) into NOAA's tropical cyclone products and services, as well as incorporate risk communication research into the design of its products. To accomplish these goals, the Office of Oceanic and Atmospheric Research (OAR) collaborated with the National Weather Service (NWS) to identify relevant operational challenges, develop project descriptions, and fund four SBES projects.

Major highlights of 2022 were:

1. Amongst real-time intensity forecast guidance, HWRF and HMON performed particularly well in 2022 in the North Atlantic. In fact, HWRF outperformed most blended intensity guidance from 60-120 h, and it outperformed all model blends at 72 h, an extremely impressive feat for a deterministic mesoscale model. HMON also performed very well in 2022, and was the best individual model for short-range intensity forecasts, namely from 24-48 h.

2. Overall, a steady reduction in RI forecast errors continues to be quite promising, and in-line with preestablished HFIP goals. In the 2021-2022 combined sample, RI forecast errors are well below the baseline from 2007 of 27-37 kt error (varying by forecast lead time), and match up quite well with the target, or the 5-year goals from the Weather Act of 2017, with errors ranging from 15-23 kt.

3. Building upon results from 2021, the two most skillful configurations of HAFS in terms of track and intensity skill scores relative to HWRF, HAFS-A and HAFS-S (subsequently renamed HAFS-B), continued development, while previous less skillful configurations have been dropped in order to refocus all resources to developing the most promising model configurations. Running a three-year retrospective sample for both the North Atlantic and eastern North Pacific, HAFS-A and HAFS-S show 5-15% improvement with respect to HWRF for track, and comparable skill with the already quite skillful HWRF for intensity in the North Atlantic, and improvements of 10-20% at days 4-5 in the eastern North Pacific.

4. In 2022, an experimental 12-member HAFS ensemble made its debut as part of HREx. Relative to the unperturbed control member, the HAFS ensemble mean forecast produced neutral skill for track, but significant improvement for intensity. The ensemble mean intensity forecast is on the

order of 20% more skillful than the ensemble control from 48-96 h, and is 5-15% more skillful than both HAFS-v0.3A and HAFS-v0.3S for intensity from 42-72 h.

5. Under Sect. 104 of the Weather Research and Forecasting Innovation Act, HFIP will continue to address the new goals of further reducing track and intensity forecast errors by 20% within 5 years and 50% within 10 years and to extend forecasts out to 7 days, with a particular focus on RI guidance. In addition, the updated plan extends HFIP's purview to improving guidance on predicting storm structure and all hurricane hazards (e.g., storm surge, rain, associated severe weather like tornadic activity, and wind gusts, as well as sustained winds) at actionable lead times for emergency managers (e.g., 72 hours). While significant progress was made in 2022, especially for track and intensity predictions, further improvements are necessary for the HAFS system to fully address the HFIP goals.

1. Chapter I: HFIP Overview and Background

1.1. Introduction

This report describes the Hurricane Forecast Improvement Program (HFIP), its goals, proposed methods for achieving those goals, and the most recent results from the program, with an emphasis on advances in the skill of operational hurricane forecast guidance. Chapter I of this report describes the background, goals, and baselines for measuring success within the HFIP program. Chapter II focuses upon capturing state-of-the-art HFIP modeling accomplishments during the 2021 hurricane season, and continued development of the Hurricane Analysis and Forecasting System (HAFS) within the Unified Forecast System (UFS). Chapter III highlights high-resolution hurricane modeling successes from the 2022 hurricane season, continued development of the HAFS system, retrospective testing and evaluation of HAFS and preparation for transition in 2023, and future plans. For more background information, readers are referred to earlier reports available on the HFIP website.

1.2. The Hurricane Forecast Improvement Program (HFIP)

The Hurricane Forecast Improvement Program (HFIP) was established within NOAA in June 2007, in response to particularly damaging landfalling hurricanes (e.g., Charley, 2004; Wilma, Katrina, Rita, 2005) in the first half of that decade. HFIP's original 5-year (for 2014) and 10-year goals (for 2019) are:[1]

- Reduce average track errors by 20% in 5 years, and by 50% in 10 years for days 1-5.
- Reduce average intensity errors by 20% in 5 years, and 50% in 10 years for days 1-5.
- Increase the probability of detection (POD)[2] for RI to 90% at Day 1, decreasing linearly to 60% at day 5, and decreasing the false alarm ratio (FAR) for rapid intensity change to 10% for day 1, increasing linearly to 30% at day 5. [The focus on RI change is the highest-priority forecast challenge identified by the National Hurricane Center (NHC)].
- Extend the lead-time for hurricane forecasts out to Day 7 (with accuracy equivalent to that of the Day 5 forecasts when those were introduced in 2003).

For more than a decade, HFIP has been providing the unified organizational infrastructure and funding for NOAA and other agencies to coordinate the hurricane research needed to achieve the above goals, improve storm surge forecasts, and accelerate the transition of model codes, techniques, and products from research to operations. HFIP focuses on multi-organizational activities to research, develop, demonstrate, and implement enhanced operational modeling capabilities, dramatically improving the numerical forecast guidance made available to the NHC, as well as enhancing the interpretation of that guidance. Through HFIP, NOAA continues to improve the accuracy of hurricane forecasts, with applied research using advanced computer models.

---

[1] The current operational model HWRF and HMON is evaluated based on the 2014 HFIP strategic plan, while the next-gen hurricane model is being developed and evaluated based on the 2019 HFIP strategic plan.

[2] POD is equal to the total number of correct RI forecasts divided by the total number of forecasts that should have indicated RI: number of correctly forecasted ÷ (correctly forecasted RI + did not forecast RI, but should have). False Alarm Ratio (FAR) is equal to the total number of incorrect forecasts of RI divided by the total number of RI forecasts: forecasted RI that did not occur ÷ (forecasted RI that did occur + forecasted RI that did not occur).

In 2017, Congress passed the Weather Research and Forecasting Innovation Act including Section 104. The Hurricane Forecast Improvement Program instructed NOAA to maintain a project to improve hurricane forecasting with the goal of developing and extending accurate hurricane forecasts and warnings in order to reduce loss of life, injury, and damage to the economy. HFIP has a particular focus on improving the prediction of rapid intensification and track of hurricanes, improving the forecast and communication of surges from hurricanes, and incorporating risk communication research to create more effective watch and warning products. In response to this charge, the HFIP strategic plan was updated outlining the research and development needed to continue improving hurricane forecast guidance, enhance probabilistic hazard products, and design a more effective tropical cyclone (TC) product suite to better communicate risk to the public and emergency management community. Under the updated plan, HFIP will continue to address the original goals of reducing track and intensity forecast errors by 20% within 5 years and 50% within 10 years, and to extend forecasts out to 7 days, particularly with focus on rapid intensification guidance. In addition, the updated plan extends HFIP's purview to improving guidance on predicting storm structure and all hurricane hazards (surge, rain, associated severe weather, gusts as well as sustained winds) at actionable lead times for emergency managers (e.g., 72 hours). Improved hazard guidance will derive from dynamical model ensembles enabling probabilistic hazard products and improved track, intensity change and structure (radii to maximum and 35-knot winds) predictions before formation and throughout the storm's life cycle. Using social science research, HFIP will design a more effective tropical cyclone product suite to better communicate risk and transition all current tropical hazards products.

One of the key strategies defined in the revised hurricane forecast improvement strategic plan in response to the proposed framework for addressing the Weather Act of 2017, is to advance an operational HAFS. HAFS is a multi-scale model and data assimilation package capable of providing high-resolution analyses and forecasts of the inner core structure of the TC out to a lead time of 7 days, which is key to improving size and intensity predictions, as well as the large-scale environment that is known to steer TCs and provides favorable/unfavorable dynamic (e.g., vertical wind shear) and thermodynamic (e.g., mid-tropospheric moisture) conditions. HAFS will provide an operational analysis and forecast system out to 7 days for hurricane forecasters with reliable, robust and skillful guidance on TC track and intensity (including RI), storm size, genesis, storm surge, rainfall and tornadoes associated with TCs. It will provide an advanced analysis and forecast system for cutting-edge research on modeling, physics, data assimilation, and coupling to earth system components for high-resolution TC predictions within the UFS. HAFS is supported under several Hurricane Supplemental projects, (i) 1A-4a: Accelerate Development of Moving Nest for HAFS; (ii) 3A-1: Accelerate implementation of the updated HFIP Plan; (iii) 3A-2: Accelerate Re-engineering of HAFS; (iv) 2019 Disaster Supplemental Improving Forecasting of Hurricanes, Floods and Wildfires HU-2 project (v) 2022 Disaster Relief Supplemental Act HURR1 project.

HFIP is organized along two lines of activities: Stream-1 and Stream-2. While Stream-1 works within presumed operational computing resource limitations, Stream-2, also called as HFIP Real-time Forecasting Experiments (HREx; https://hfip.org/products), activities assume that resources will be provided to increase the available computer capability in operational settings, above the one that is already planned for the next five years. The purpose of Stream-2 is to demonstrate that the application of advanced and innovative science, technology, and increased computing will lead to the desired increase in accuracy, and other improvements in forecast performance. Because the level of computing necessary to perform such a demonstration is larger than can be accommodated by current operational computing resources, HFIP developed its own computing system at NOAA's Earth System Research Laboratories (ESRL) in Boulder, Colorado. For instance, in the 2021 season, four versions of HAFS were tested in near real-time within the stream-2 HREx. *(see section 9 for results)*

1.3. HFIP Baseline for measuring progress

To measure progress towards the above-defined HFIP goals, a baseline level of accuracy was established. The HFIP goals were to reduce track and intensity errors by 20% in 5 years and 50% within 10 years. A set of baseline track and intensity errors were developed by NHC, where the baseline is the consensus (average) from an ensemble of top-performing operational models evaluated over the period of 2006-2008 for the North Atlantic basin. For track, the ensemble members were the operational aids GFSI, GFDI, UKMI, NGPI, GFNI, and EMXI, while for intensity the members were GFDI, DSHP, and LGEM[3] (Cangialosi, June 2020). Results from HFIP model guidance are then compared with the baseline to assess progress. Figure 1 shows the mean absolute errors of the consensus over the period 2006-2008 for the North Atlantic basin. A separate set of baseline errors (not shown) was computed for the eastern North Pacific basin (Franklin, 2009, 2010).

To provide a more representative, longer-term perspective, the progress of HFIP models are also evaluated in terms of forecast skill. Because a sample of cases from a season might have a different inherent level of difficulty from the baseline sample of 2006-2008 (for example, because it had an unusually high or low number of rapidly intensifying storms), it is helpful to evaluate the progress of the HFIP models in terms of forecast skill as well as error. Here, that evaluation is determined with the percent improvement, relative to a statistical model for the same cases. A statistical model is one where a number of predictors are combined, using weights that are determined by correlation with past data and, consequently, performs better in relatively 'easy-to-predict' seasons, and worse in relatively 'difficult-to-predict' seasons. Figure 1 shows the skills of the baseline, baseline errors, and the 5- and 10-year goals - represented in blue and labeled on the right side of the graph. The goals are presented as the percentage improvement over the Decay-(Statistical Hurricane Intensity Forecast) SHIFOR5 and (Climatology and Persistence) CLIPER5 forecasts, for the same cases that were used to determine the mean absolute baseline error.

[3] See appendix A for details on operational aids (GFSI, GFDI, UKMI, NGPI, GFNI, EMXI,GFDI, DSHP, LGEM)

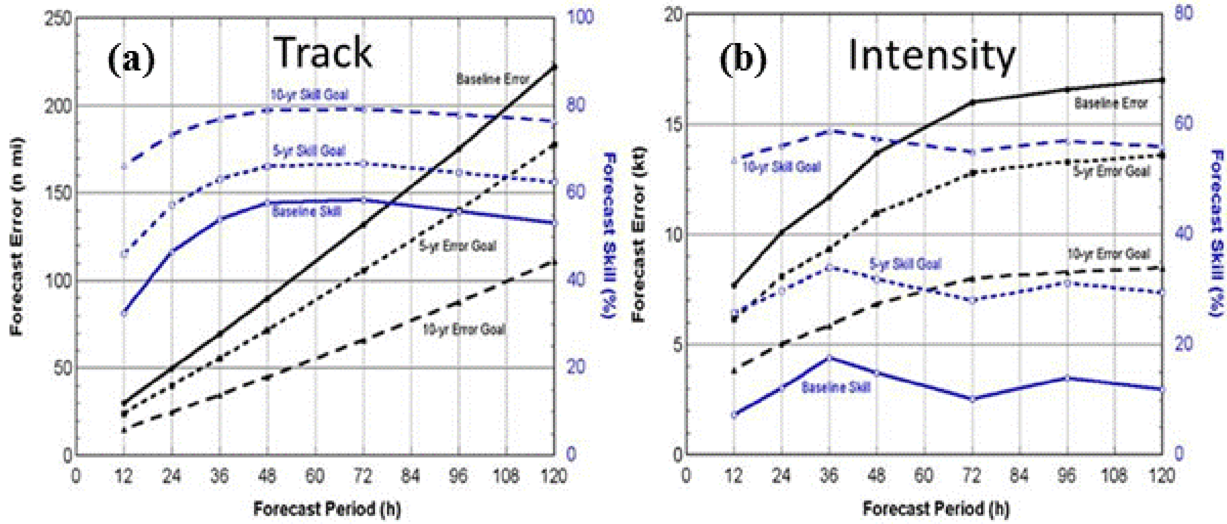

Figure 1: (a) Track and (b) Intensity Error Baseline and Goals, where the forecast errors are represented by black lines labeled on the left side of the graph, and the forecast skill is represented by blue lines labeled on the right side of the graph. Solid black lines represent baseline forecast errors, while solid blue lines represent baseline forecast skill. The 5 and 10 years goals are represented by dashed black lines for errors, and dashed blue lines for skill.

The skill baseline and goals for intensity at all lead times are roughly constant, with the baseline representing a ~10% improvement over Decay-SHIFOR5, and the 5- and 10-year goals representing ~30% and ~55% improvements, respectively. It's important to remember, however, that normalization by CLIPER or (especially) Decay-SHIFOR5 can fail to adequately account for forecast difficulty in some circumstances. A hurricane season that features extremely hostile environmental conditions will lead to very high Decay-SHIFOR intensity forecast errors (as climatology will be a poor forecast in such years), but relatively low errors in dynamical models and NHC official forecasts (as few storms will intensify rapidly, making it less challenging for both models and forecasters). This combination of baseline and model errors yields an unrealistic skill estimate. Hence, both skill and absolute errors are used to measure HFIP model improvements.

It is also important to note that HFIP performance baselines were determined from a class of operational aids known as "early" models. Early models are those that are available to forecasters quickly enough to meet forecast deadlines for the synoptic cycle. Nearly all the dynamical models currently used at tropical cyclone forecast centers, such as the Global Forecast System (GFS) and HWRF models, are considered "late" models because their results arrive too late to be used in the forecast for the current synoptic cycle. For example, the HWRF run for 12:00 Coordinated Universal Time or Zulu Time Zone (Z) does not become available to forecasters until around 16:00Z, whereas the NHC official forecast based on the 12:00Z initialization must be issued by 15:00Z, one hour before the HWRF forecast can be viewed. It's actually the older, 06:00Z run of the HWRF model that would be used as input for the 15:00Z official NHC forecast, through a procedure developed to adjust the 06:00Z model run, to match the actual storm location and intensity at 12:00Z. This procedure also adjusts the forecast position and intensity at some of the forecast times as well, and then applies smoothing to the adjusted forecast. This adjustment, called an "interpolation" procedure, creates the 12:00Z "early" aid HWRF with 6-hour interpolation (HWFI) that can be used for the 15:00Z NHC forecast. Model results so adjusted are denoted with an "I" (e.g., HWFI). The distinction between early and late models is important in assessments of model performance provided in subsequent sections, since late models have an advantage of more recent observations/analysis than their early counterparts.

2. Chapter II: HFIP in 2021

2.1. HFIP Model Systems

Accurate TC forecasts beyond a few days require a global domain, because influences on a forecast at a particular location can come from weather systems elsewhere, far from the particular location. Figure 2a shows the steep-step improvements to track predictions for 24, 48, 72, 96 and 120 hours since the 90's. Those advancements have come through developing improved dynamical global models (e.g., GFS), further improving resolution and physics in those models, and through advancing DA techniques. Most of the GFS developments have been at the National Center for Environmental Prediction (NCEP). Nevertheless, one of the first efforts in HFIP was to improve the existing operational global models. Early in the program, it was shown that forecasts were improved, particularly in the tropics, by using a more advanced DA scheme than the one employed operationally at that time. A version of this advanced DA went operational in the GFS model in May, 2012.

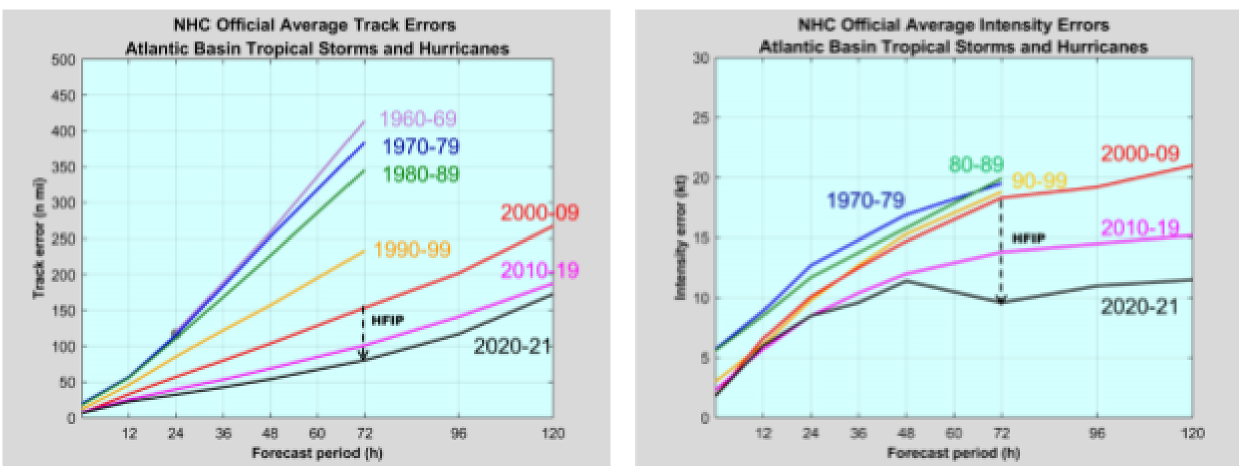

Figure 2: Official NHC (a) Track errors (1960-2021) and (b) Intensity errors (1970-2021) in the Atlantic basin. The downward arrow denotes the period HFIP is active.

While significant track improvements have been achieved since the 1960s, progress in reducing intensity errors had been slow until the onset of HFIP in 2009 (Figure 2). Part of the problem was inadequate model-grid resolution. It is generally assumed that the hurricane inner core (i.e., the eye-wall region) must be resolved to see consistently accurate hurricane intensity forecasts (NOAA SAB, 2006). It is believed that the best approach to improve hurricane track and intensity forecasts involves the use of high-resolution global models, with at least some being run as ensembles. However, global models and their ensembles are likely to be limited by computing capability, for at least the next five years, to a horizontal resolution no finer than about 8-10 km, which is inadequate to resolve the inner core of a hurricane. Maximizing improvements in hurricane intensity forecasts will, therefore, require high-resolution regional models, or global models with moveable high-resolution nests, perhaps also run as an ensemble. During the last 12 years, the focus has been on improving the intensity forecast, which for decades has significantly lagged behind the track forecast. For that purpose, regional models with (two-way interactive) moving nests capable of resolving the inner core structure of hurricanes are usually used for intensity predictions. The domains of the hurricane regional models are usually larger than their CONUS counterparts. The HWRF and HMON that were developed during HFIP are prime examples. Track predictions from these regional models, especially HWRF, have been shown to improve with larger domains (Zhang et. al., 2016; and Alaka et. al., 2017; 2022). The Basin-Scale HWRF (HWRF-B) has demonstrated the usefulness of expanding the regional domain for TC predictions and paving the way towards the advancements of Global-to-local scale HAFS.

The error and skill performance statistics below will highlight the last ten years, through 2021, with a particular emphasis on 2021. The 2021 North Atlantic hurricane season was above average. There were 21 named storms, of which 7 developed into hurricanes, with 4 of those becoming major hurricanes. There were 5 TCs that underwent RI in the North Atlantic basin: Elsa, Grace, Ida, Larry, and Sam. There were 8 landfalls in the U.S. from 6 tropical storms and 2 hurricanes. A total of 399 forecasts were issued in the Atlantic. In the eastern North Pacific, there were 19 named storms, of which 8 developed into hurricanes with 2 major hurricanes.

2.2. Operational HWRF and HMON systems (Stream 1)

a. HWRF System

One of the major accomplishments of HFIP has been the development of the storm-following, double-nested, high-resolution HWRF model, and its transition to operations. HWRF is a joint development between NOAA research and operations, with significant support from the Developmental Testbed Center (DTC), UCAR, and the TC community. It is one of the top-performing track prediction models, and is paving the way to improve operational intensity forecasts all over the globe. The HWRF model is based on the Non-Hydrostatic Mesoscale Model on an E-grid (NMM) dynamical core and can be coupled to Princeton Ocean Model (POM) or HYbrid Coordinate Ocean Model (HYCOM). It is a part of the general WRF infrastructure, but using the NMM dynamic core, which is more focused on supporting operations (Biswas et al., 2018; Tallapragada et. al., 2014). HFIP has coordinated the following HWRF improvements: (i) storm-following nesting, (ii) horizontal grid spacing (3 km in 2012, 2 km in 2015, and 1.5 km in 2018), (iii) physical parameterizations, and (iv) initial conditions enhanced by aircraft observations. These improvements have led to improved numerical guidance that TC forecasters use in real time. HWRF is also the main driving dynamical model of the Real-Time HFIP Corrected Consensus Approach (HCCA) for TC Intensity Guidance at NHC (Simon et. al., 2018) and has become the flagship intensity prediction tool for hurricane forecasting at NWS.

In the last ten years (2012-2021), the HWRF system was upgraded considerably under HFIP, including the following annual upgrades. The model code for each year is provided for reference.

- In 2012 (H212), for the first time, the double-nested, cloud-resolving version of HWRF was run at 3 km horizontal resolution (27/9/3 km version) with improved physics based on observations (Gopalakrishnan et. al., 2011; 2012; 2013; Goldenberg et. al., 2015).

- In 2013 (H213), upgraded physics and vortex initialization were adopted.
- In 2014 (H214), HWRF was run in real-time in all global basins beyond the North Atlantic.
- In 2015 (H215), HWRF implementation consisted of increased horizontal resolution from 27/9/3 km to 18/6/2 km across all domains, continued improvement of the Nest-Tracking-Algorithm, advanced vortex initialization, and improved products.
- In 2016 (H216), new SAS and GFS-EDMF physics suites were implemented. This was the watermark year for 5-year HFIP improvements.
- In 2017 (H217), a dramatically improved DA system was implemented.
- In 2018 (H218), the HWRF implementation incorporated a further increment of the horizontal resolution, from 18/6/2 km, to 13.5/4.5/1.5 km, as well as continued improvement of the Nest-Tracking-Algorithm, and advanced vortex initialization. With the 2018 upgrade in model resolution, the HWRF model is now the highest resolution hurricane model ever implemented for operations in the NWS.
- In 2019 (H219), HWRF was not operationally upgraded due to the NCEP Central Operations (NCO) moratorium.
- In 2020 (H220), HWRF was upgraded to two-way ocean coupling, one-way wave-model coupling and the high-resolution land-sea masks for the moving nests.
- In 2021 (H221), HWRF was synced with the latest UFS upgrades, but otherwise there were no HWRF-specific upgrades.

Figure 3 presents a summary of improvements since the start of HFIP. These improvements are measured in terms of mean forecast skill scores using climatology and persistence (OCD5) as a reference model for each respective season (e.g., 2011 forecasts of OCD5 are used as a baseline for H211 forecasts). The last three operational versions of HWRF (H219, H220, and H221) were chosen especially to illustrate improvements over the last decade (i.e., compare against H211) and to account for the variability of model performance from one year to the next.

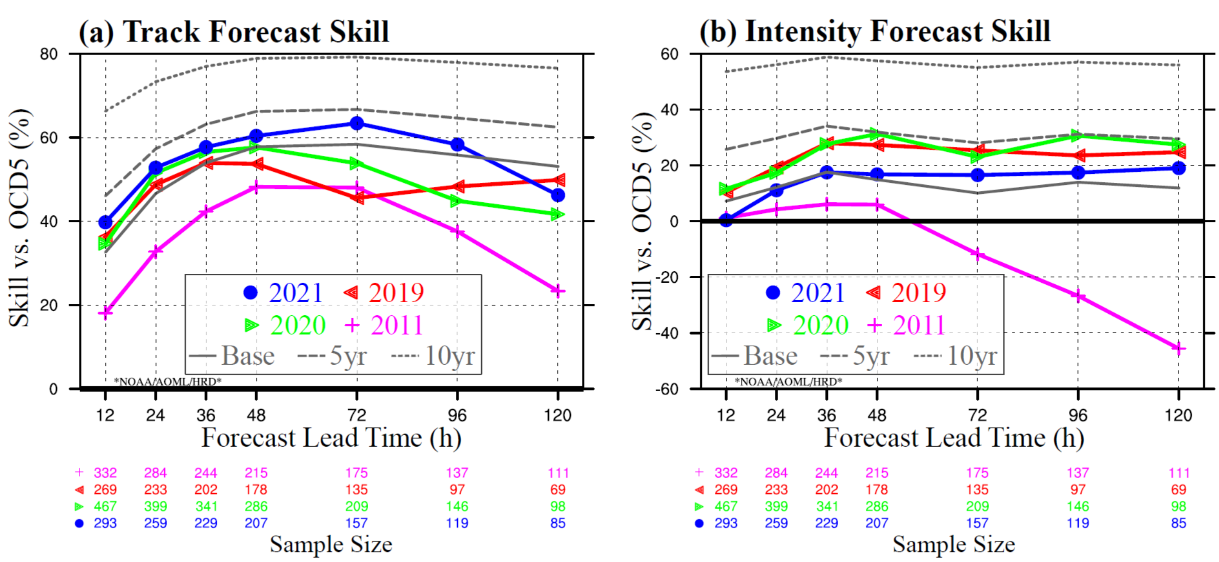

Figure 3: For H211 (pink cross), H219 (red left-pointing triangle), H220 (green right-pointing triangle) and H221 (blue circle), the following is shown: (a) track forecast skill relative to NHC's climatology-persistence

skill baseline (OCD5), and (b) intensity forecast skill relative to OCD5. The HFIP baselines (solid), HFIP 5-year goals (dashed), and HFIP 10-year goals (dot-dashed) are shown in gray for track, OCD5 is also known as CLIPER5, and, for intensity, OCD5 is also known as Decay-SHIFOR. The verification excludes actual and forecast positions that are inland.

Figure 3a illustrates the track forecast skill of the four HWRF versions relative to OCD5. Overall, track forecasts have steadily improved over the last decade, with average track predictions performing ~20% better at all forecast lead times in H221 compared to H211. H221 had better performance than the previous two versions of HWRF at most lead times, with H219 showing particularly poor performance due to difficult TC track predictions (e.g., Dorian stalling over the Bahamas). For the last three years, track skills have been improved between 30-60% at all lead times, with the H221 track skill maximizing above 60% at 72 h. Although HWRF track forecasts have clearly improved over the last decade, even H221's performance is barely above the HFIP baseline and below the HFIP 5-year goal. We believe further improvements may be possible with HAFS (section 9).

Figure 3b portrays the progress of HWRF in forecasting maximum wind speed (i.e., intensity), measured in terms of skill relative to OCD5. Through 2011, HWRF operated with a single 9 km-resolution moving nest that could automatically track hurricanes[4] (Gopalakrishnan et. al., 2006). Although this was a huge advancement for TC predictions, the resolution was too coarse to capture processes critical for intensification, and, consequently, H211's intensity forecast performance was quite poor, especially at longer lead times. For the last three years, the intensity skill has been positive, hovering between 15-30% at lead times of 36 h and longer. At those lead times, the intensity forecast performance exceeded the HFIP baseline.

[4] It should be noted that the plots between 2016 and 2020 showed no statistically significant differences. The differences could be due to year-to-year variability.

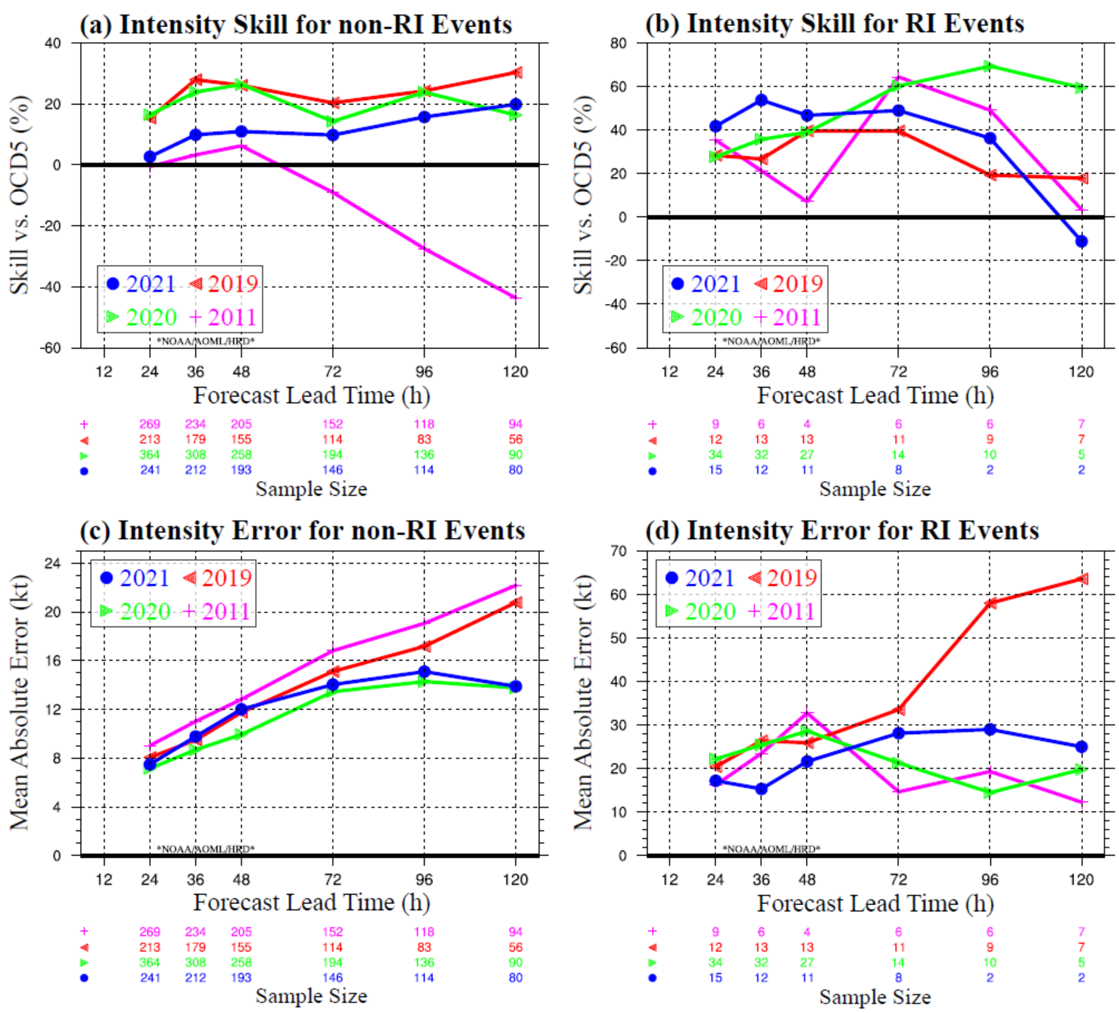

Figure 4: For H211 (pink cross), H219 (red left-pointing triangle), H220 (green right-pointing triangle) and H221 (blue circle), the following is shown: (a) intensity forecast skill for non-RI cases, (b) intensity forecast skill for RI cases (i.e., intensification of ≥30 kt in 24 h), (c) intensity errors (in kt) for non-RI cases, and (d) intensity errors for RI cases. Intensity changes are calculated over the preceding 24-h period for each forecast time ≥24 h. Forecast skill is computed relative to OCD5. The verification excludes actual and forecast positions that are inland.

Rapid Intensification of TCs are of major concern to HFIP since the start of the program. RI forecasts are particularly challenging because the timing, duration, and intensity change associated with each RI event are not well-predicted by numerical weather prediction models, in general. Even a high-resolution model like HWRF has struggled to strike an optimal balance between increasing RI detection while limiting false alarms.

In Figure 3b although HWRF performance has shown improvements in an overall sense, clearly there is noticeable degradation in performance of the 2021 HWRF in terms of skill -vs-OCD. We stratified the samples further in terms non-RI and RI events to understand this degradation in HWRF skill. Figure 4a and b (Top panel) shows the skill for non-RI and RI cases and Figure 4c and d (Bottom panel) shows the mean absolute errors for the stratified samples

The conclusions can be summarized as follows:

- In general, H221 skill was worse than the previous two seasons because climatology and persistence performed better in 2021, i.e., 2021 was an “easier” year for TC predictions (Figure 4).
- H221 intensity forecast errors were consistent with those from H220 for non-RI events, and both H221 and H220 had noticeably lower mean absolute errors at 96 h and 120 h compared with H219 (Figure 4c).
- RI predictions improved significantly in 2021 at short lead times (≤48 h). This is reflected in both mean absolute errors and forecast skill (Fig. 4b,d). However, RI predictions were the worst in the last 3 years at longer lead times, very likely contributing to loss of skill in HWRF intensity predictions in 2021.
- Only two RI events occurred that corresponded to 96 h and 120 h forecasts from H221, both of which were for Hurricane Grace in the Gulf of Mexico.
- Although mean absolute errors are larger for RI events than for non-RI events, skill is higher at most lead times for RI events. This indicates that HWRF is generally performing much better than climatology and persistence for RI events than for non-RI events.
- It appears that the use of mean absolute error is less prone to variability from year to year especially when the sample sizes are small.
- Sustained HFIP research and development is necessary for further improvements in intensity and intensity change predictions (RI and RW).

b. HMON System

Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic model (HMON) was developed to provide higher-resolution intensity and track forecast guidance to NHC, along with HWRF. HMON replaced the legacy (hydrostatic) Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model, being 2-way coupled to HYbrid Coordinate Ocean Model (HYCOM), which was used as the second dynamical model along with HWRF for intensity guidance until 2016. The HMON model is based on the Non-Hydrostatic Mesoscale Model on a B grid (NMMB) dynamic core, which is currently being used in NCEP operational systems - the North American Mesoscale (NAM) Model and the Short Range Ensemble Forecast (SREF) model. The HMON was built using shared infrastructure with unified model development within the NOAA Environmental Modeling System (NEMS) and could also be coupled with other (ocean, wave, land, surge, inundation, etc.) models within the NEMS infrastructure. Use of NEMS also paves the way for future use of physics packages like CCPP (Common Community Physics Package). HMON has been in operations since 2017 and has demonstrated forecast consensus improvement. In 2020, several upgrades were made to the model infrastructure and physics including an increase of the vertical level from 51 to 71.

2.3. Operational Hurricane Guidance Improvements

NHC uses several deterministic guidance models for their official intensity forecasts, including NCEP’s HWRF and HMON regional dynamical models, several global models, and the D-SHIPS (Decay-Statistical Hurricane Intensity Prediction Scheme) and LGEM (Logistics Growth Equation

Model) statistical models. As noted earlier, the dynamical models are not available in time to be used by the NHC forecasters, so a method to interpolate the predictions from the previous forecast cycle has been developed. The interpolated versions are called “early” models. In all of the discussion below, only early models are considered. Several consensus intensity models are also used as input to the NHC forecast. The simplest is IVCN (Intensity consensus of at least two forecasts), which is a linear average of the D-SHIPS and LGEM statistical models, the early versions of the HWRF and HMON regional models, and the U.S. Navy’s COAMPS-TC regional hurricane model that uses GFS (Global Forecast System) initial and boundary conditions also called CTCX. IVCN is computed whenever two or more of the above models (HWRF, HMON, CTCX, D-SHIPS and LGEM) are available. IVCN is used as the basis for performance measures for RI predictions instead of individual model guidance from HWRF and HMON (section 6c).

a. Track Guidance

In 2021, official Atlantic track forecasts (Figure 5a) were very skillful and close to or even better than the best-performing consensus aids - FSSE (Florida State University Super-Ensemble Corrected Consensus), HCCA (HFIP Corrected Consensus Approach) and TVCA (Track Variable Consensus of at least two forecasts) (Cangiolosi, 2021). GFSI (GFS with 6 hour interpolation) was the best dynamical model at all lead times. AEMI (GEFS with 6 hour interpolation), EMXI (ECMWF with 6 hour interpolation), HMNI (HMON with 6 hour interpolation) and HWFI (HWRF with 6 hour interpolation) came in second place, very close to one another. CTCI (COAMPS-TC 6 hour interpolation) and CMCI (Canadian Global Model 6 hour interpolation) were less good at longer lead times. NVGI (Navy Global Environmental Model 6 hour interpolation) lagged behind other models. Similar to the previous year, GFS outperformed EMXI track skills at all lead times.

In the eastern Pacific (Figure 5b), the official forecasts were very skillful, very close to TVCE, HCCA, FSSE consensus models. GFSI and AEMI were the best individual models through 72 h, EMXI was best at 96 and 120 h. HMNI, HWFI and CMCI were not very good. NVGI and EGRI lagged behind other models.

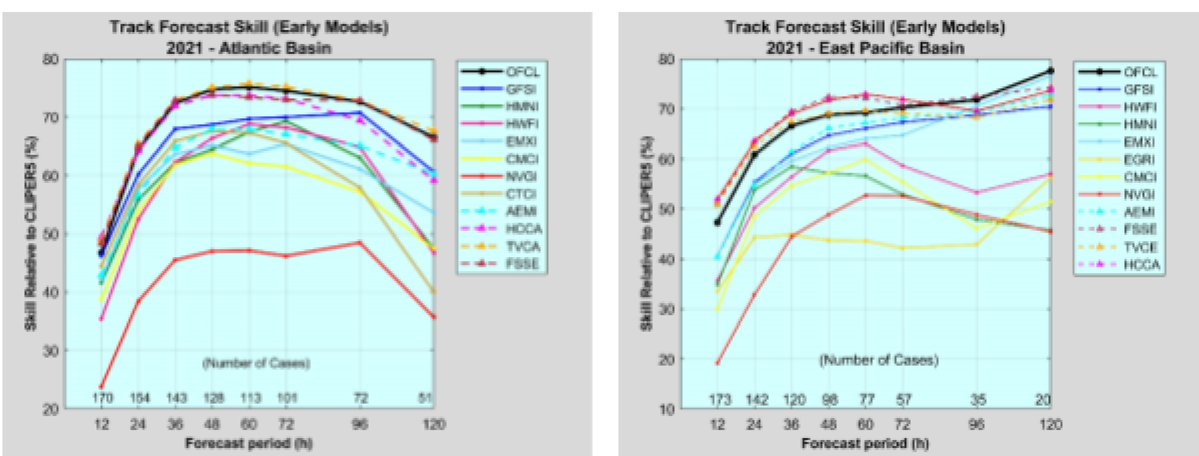

Figure 5: Official track forecast skill in 2021 for the (a) Atlantic (left) and (b) eastern Pacific (right) basins. Numbers immediately above the X-axis show the total number of cases covered by each data point.

b. Intensity Guidance

Intensity forecast skill for the 2021 season is shown in Figure 6. In the Atlantic basin (Figure 6a), official forecasts were very skillful, as good as or better than the consensus aids. Among the consensus models, FSSE was the best till 96 h. HMNI was a strong performer and the best individual model at most lead times. HWFI and CTCI did not do as well as HMNI. DSHP and LGEM were fair performers, but not as good as HMNII and consensus models. GFSI was somewhat competitive and EMXI was skillful only at longer lead time.

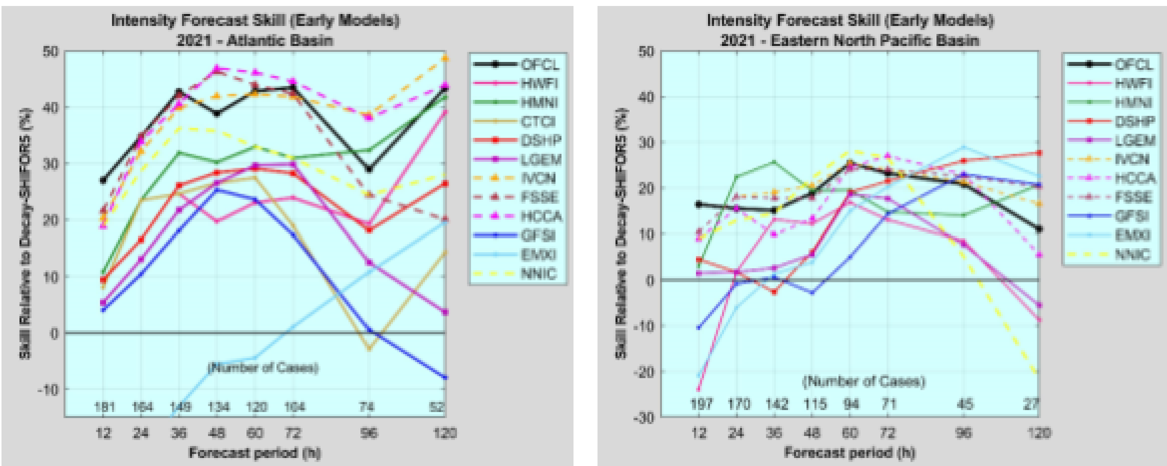

Figure 6: Official intensity forecast skill in 2021 for the (a) Atlantic Basin (left) and (b) East Pacific Basin (right). Numbers immediately above the X-axis show the total number of cases covered by each data point.

In the eastern Pacific (Figure 6b), official intensity forecast performance was good as the best consensus aids (IVCN, HCCA, FSSE). Consensus aids were generally best, except at 96 and 120 h where EMXI and DSHP had more skill. HMNI was a strong performer and better than HWFI. DSHP and LGEM were fair performers. GFSI and EMXI were competitive in this basin.

c. State-of-art in RI guidance

One of the HFIP goals is to "reduce intensity forecast guidance errors by 50% for RI events". After consideration of several metrics to measure RI progress, HFIP chose to use the mean absolute error for a subset of cases where RI was forecast or observed. The new metric is less prone to large year to year variability due to small sample sizes than other metrics such as probability of detection or false alarm rate. The HFIP RI performance metric, baseline, and initial progress toward the RI forecast goal are discussed below.

The RI metric is the mean absolute error (MAE) of the IVCN consensus, for the Atlantic and eastern Pacific basins combined, evaluated for only those verification times when RI was either ongoing or was forecast. Specifically, this means the verifying time must satisfy at least one of the following criteria:

- A 30-kt or larger intensity increase in the best-track intensity, relative to the best-track intensity 24-h prior to the verification time.

- A 30-kt or larger forecast intensity increase in any of the IVCN member models, relative to the forecast intensity 24-h prior to the verification time.

With this as the metric, HFIP then defined the baseline sample as those 24-, 36-, 48-, 72-, 96-, and 120-hr forecasts satisfying the above criteria for the combined Atlantic and eastern Pacific basins over the period 2015-17. When non-consensus forecasts (e.g., an individual model such as HWFI, or the NHC official forecast, OFCL) are evaluated relative to the RI baseline or target, criteria (2) above should be applied to each of the models forming the homogeneous sample.

By considering both RI cases occurring in the best track and the RI cases being forecast, the metric ensures that overly aggressive models are penalized for false alarms. A full assessment of our ability to forecast RI requires consideration of false alarms as well as misses, and from an operational standpoint, a metric that considers both types of errors will be of greater value to forecasters who must gauge the credibility of a forecast of RI when one is presented to them.

The values of the RI baseline are presented in Table 1 and Figure 7. One complication in determining the baseline values was that the membership of IVCN at any particular forecast time is not recorded operationally nor readily determined after the fact, and the sample definition depends on checking each member's forecast for occurrences of RI. Furthermore, the composition of IVCN changed over the baseline period 2015-17. For these reasons, the HFIP baseline errors were determined from a single recomputed version of IVCN comprising models used in the operational IVCN at any time from 2015-17; these models were DSHP, LGEM, GHMI, HWFI, and CTCI. It is seen that our ability to predict RI during the baseline period was only weakly dependent on forecast lead time; the errors were high even at 24 h (26 kt) and saturated quickly. In terms of skill relative to climatology/persistence, a peak is seen from 72-96 h but skill was minimal throughout the 5-day forecast period. It's worth noting that the target MAEs in Table 1 are all large enough to be observationally detectible, in contrast to the overall (non-RI) intensity targets, which are small enough that it may be difficult to distinguish them from the best-track uncertainty.

Table 1: HFIP RI performance measures baseline and target errors. Baseline errors are the mean absolute errors over the period 2015-17 for the Atlantic and eastern North Pacific for the variable consensus comprising at least two of the models DSHP, LGEM, GHMI, HWFI, and CTCI. Target errors represent 50% of the baseline errors.

Verification Time (h)	Baseline (kt)	Target (kt)
24	26.1	13.1
36	28.6	14.3
48	31.4	15.7
72	36.9	18.5
96	31.3	15.6
120	32.1	16.1

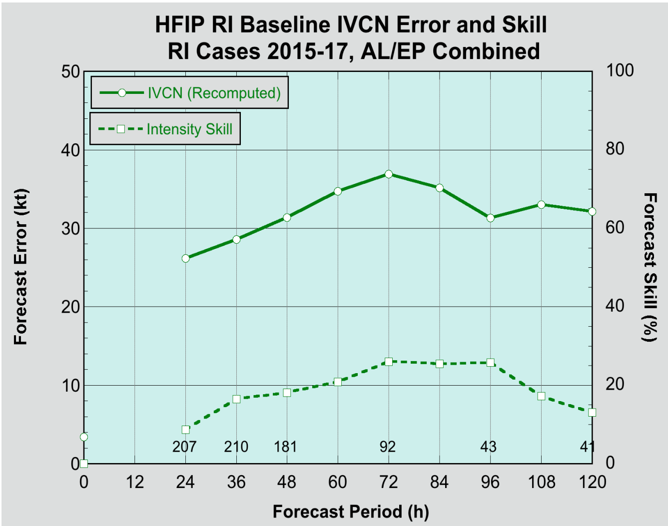

Figure 7: HFIP RI performance measures baseline errors and skill. Baseline errors are the mean absolute errors over the period 2015-17 for the Atlantic and eastern North Pacific for the variable consensus comprising at least two of the models DSHP, LGEM, GHMI, HWFI, and CTCI. Skill values are computed relative to OCD5.

Since the baseline period ended in 2017, NHC's operational intensity consensus has not changed, comprising DSHP, LGEM, HWFI, CTCI, and HMNI during each season from 2018-2021. Figure 8 shows a verification of the HFIP RI intensity metric for the 2021 season. Note that the NHC best tracks were not final at the time of these verifications, so the results are preliminary. The errors of the RI metric are seen to be well below the baseline errors at all forecast lead times: 42% below the baseline at 24 h, 44% below at 48 h, and 36% below at 72 h.

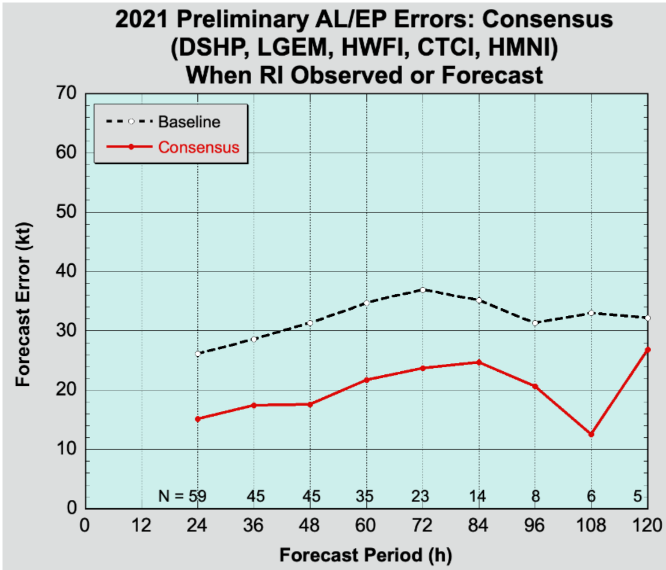

Figure 8: HFIP RI performance measure for 2021. Errors for the consensus from 24-120 h are shown by the red line, while HFIP baseline errors are shown by the dashed black line. Results are preliminary since the 2021 best tracks were not final at the time these verifications were performed. Number of cases for each forecast lead are given along the bottom of the diagram.

Figure 9 shows how the RI intensity metric has performed over the past few seasons. The consensus forecast shown here for each season corresponds to NHC's operational composition of IVCN for that season. MAEs for each season are shown at 24, 48, and 72 h, with the HFIP baseline values given by the three asterisks plotted at 2016, the midpoint of the baseline period. Comparison of the 2015-17 baselines to mean errors over the most recent three years 2019-21 is very encouraging: at 24 h the baseline error (26.1 kt) has been reduced by 34% (to 17.3 kt), at 48 h the baseline error (31.4 kt) has been reduced by 27% (to 22.9 kt), and at 72 h the baseline error (36.9 kt) has been reduced by 27% (to 26.8 kt).

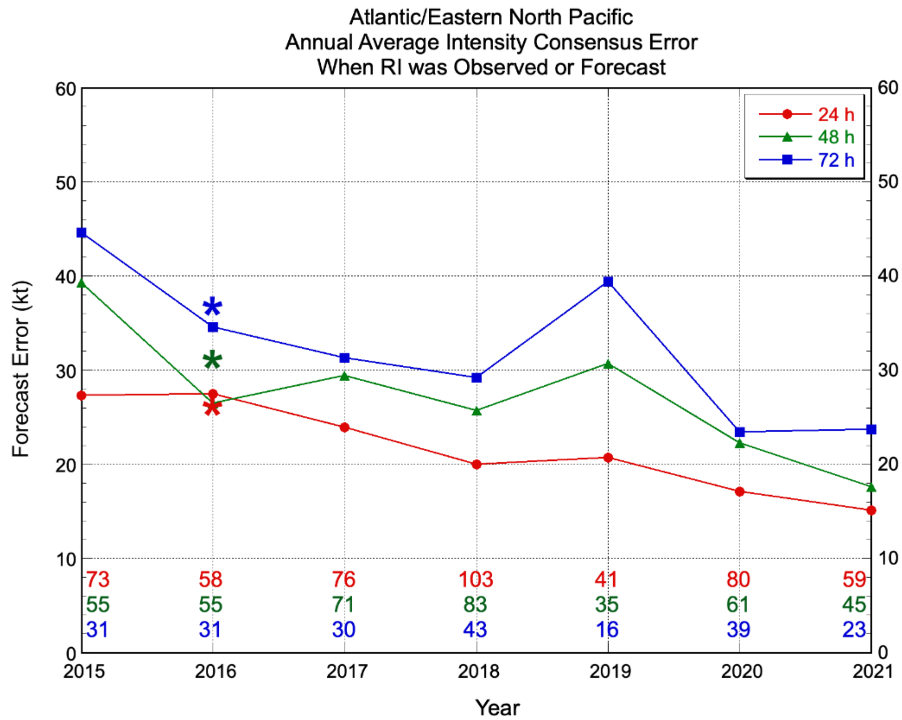

Figure 9: HFIP RI performance measure at 24, 48, and 72 h for 2015-21. The consensus evaluated for each season corresponds to NHC's operational composition of IVCN for that season. Results for 2021 are preliminary. HFIP baseline errors are given by the asterisks plotted for the year 2016. Number of cases for each forecast lead are given along the bottom of the diagram.

Examination of IVCN error distributions illustrates where the forecast improvements have been coming from. Figure 10 shows the 24-h error distributions for the baseline period (top panel) and for the past two seasons 2020-21 (bottom panel). During the baseline period, nearly all the errors for this metric were negative, with a mode at -25 kt, and a few errors as large as -70 to -80 kt. Over the past two seasons, however, the mode hasn't changed much, suggesting that missed RI cases are still an issue; however, the distribution has sharply shifted to the right and broadened, with a large cluster of errors now near zero. Clearly, the models are now capturing RI much more frequently than they were during the baseline period. Furthermore, during the past two seasons there have been fewer very large negative errors. Examination of the error distributions at 48 h (not shown) indicate similar but less dramatic changes.

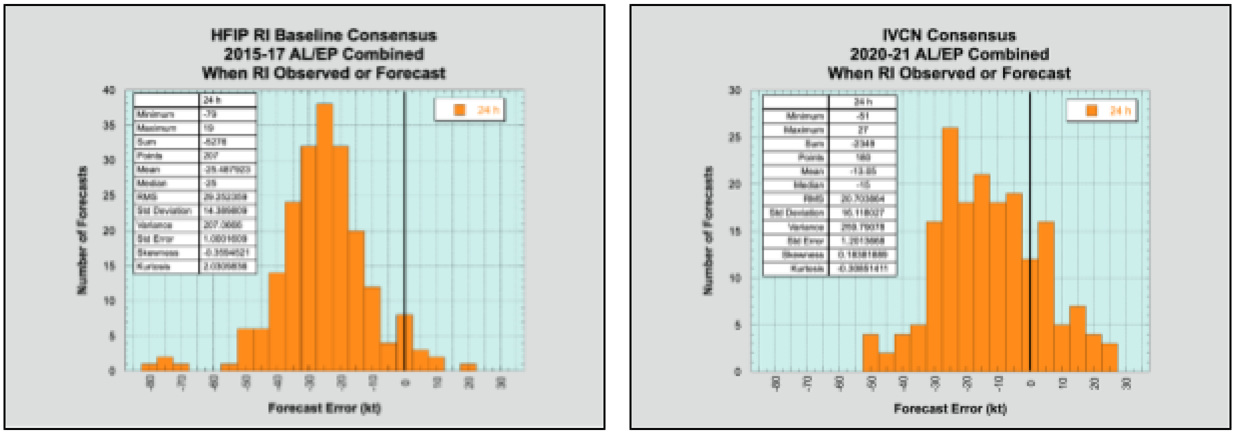

Figure 10: Error distributions of the HFIP RI performance measure at 24 h for the baseline period (2015-17), left, and for 2020-21, right. Results for 2021 are preliminary.

Collectively, these results indicate that very strong progress is being made toward reaching the HFIP RI goal of a 50% error reduction. If the 2021 season's results are representative, the HFIP RI target appears to be well within reach. It's also encouraging to see that the improvements in RI guidance are being reflected in NHC official forecast errors; Figure 11 shows downward OFCL error trends for RI cases that are very similar to the trends shown in Fig. 3 for the consensus.

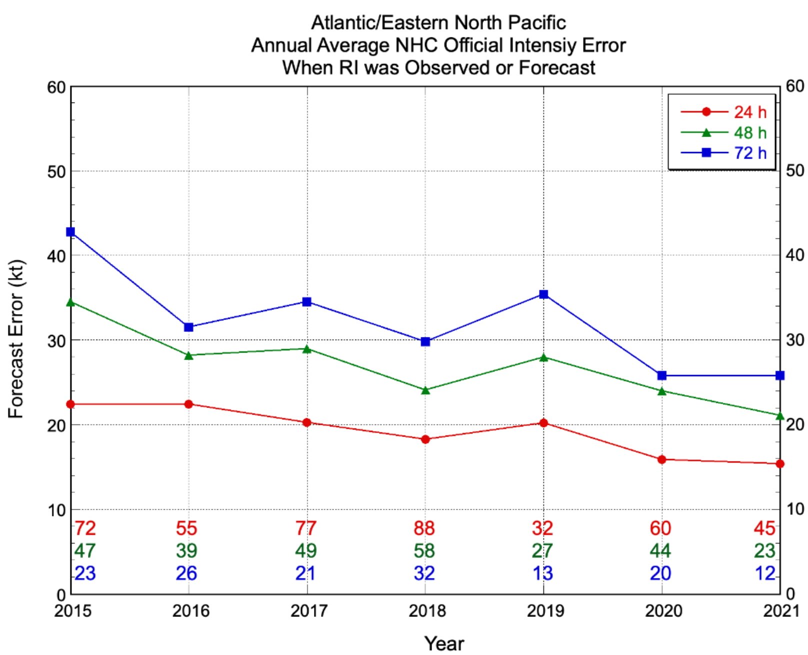

Figure 11: NHC official forecast error for RI cases at 24, 48, and 72 h for 2015-21. Results for 2021 are preliminary. Number of cases for each forecast lead are given along the bottom of the diagram.

During the past year a description of the HFIP RI performance metric was published, along with a history of operational forecasting of RI at the National Hurricane Center (DeMaria et al., 2021).

2.4. Next Generation HFIP Goals and Plans

The Weather Research and Forecasting Innovation Act of 2017 (also known as The Weather Act) required NOAA to prioritize research that improves forecasts and warnings for the protection of life, property, and the enhancement of the national economy. In response to Section 104 of the Weather Act, the new HFIP Strategic Plan detailing the specific research, development, and technology transfer activities necessary to sustain HFIP's next generation of science and R2O challenges has been approved.

To improve TC forecasting with the goal of developing and extending accurate TC forecasts and warnings in order to reduce loss of life, injury, and damage to the economy, the next generation of HFIP will focus on:

i. Improving the prediction of rapid intensification and track of TCs;
ii. Improving the forecast and communication of surges from TCs; and
iii. Incorporating risk communication research to create more effective watch and warning products.

In order to address the three primary focus areas outlined above, HFIP has developed a set of specific goals and metrics to improve the accuracy and reliability of TC forecasts and warnings and increase the confidence in those forecasts to enhance mitigation and preparedness decisions by emergency management officials at all levels of government and by individuals.

Improved model guidance for TC formation, track, intensity and size will be essential to address all three areas. Basic TC forecast parameters will be improved, including the formation time and location, position, maximum wind (i.e., intensity), and storm size. Estimates of the uncertainty of those parameters will also be enhanced, enabling better risk communication to end users through accurate probabilistic information (i.e., information that considers the likelihood, or probability, that an event will occur). Rapid intensification remains an especially important and challenging forecast problem. Specific goals and metrics are defined for the prediction of the basic TC forecast parameters, new extended range forecasts, rapid intensification, and TC formation.

HFIP will build upon the original goals of the project through the following specific goals and metrics:

- Reduce forecast guidance errors, including during rapid intensification, by 50 percent from 2017;
- Produce 7-day forecast guidance as good as the 2017 5-day forecast guidance;
- Improve guidance on pre-formation disturbances, including genesis timing, and track and intensity forecasts, by 20 percent from 2017; and
- Improve hazard guidance and risk communication, based on social and behavioral science, to modernize the TC product suite (products, information, and services) for actionable lead-times for storm surge and all other threats.

Six key strategies were developed to address these new goals, of which the main strategy is the ongoing development of a multi-scale modeling system, referred to as HAFS.

2.5. Development of Hurricane Analysis and Forecast System (HAFS)

The HAFS is NOAA’s next-generation multi-scale numerical model, with data assimilation package and ocean/wave coupling, which will provide an operational analysis and forecast out to seven days, with reliable and skillful guidance on Tropical Cyclone (TC) track and intensity (including RI), storm size, genesis, storm surge, rainfall and tornadoes associated with Tropical Cyclones. The UFS is a community-based, coupled comprehensive Earth system modeling system based on the FV3 dynamical core, whose numerical applications span local to global domains and predictive time scales from sub-hourly analyses to seasonal predictions. It is designed to support the Weather Enterprise and to be the source system for NOAA's operational numerical weather prediction applications. The HAFS will be a part of UFS geared for hurricane model applications.

The current version of HAFS includes following major components: (a) Cloud permitting high-resolution, storm-following moving nest; (b) Advanced vortex initialization; (c) Inner-core data assimilation; (d) scale-aware physics uniquely calibrated for TC application (e) Ocean coupling, and (f) High-resolution observations to support the DA.

a. Cloud permitting high-resolution moving nest

Central to the development of HAFS is the FV3 dynamical core with an embedded storm-following moving nest capable of tracking the inner core region of the hurricane at 1-2 km resolution (cover picture). Although the FV3 model dynamic core itself is fully tested with convection-allowing grid spacing and could be run both as global and regional models, the current nesting capabilities are very limited, at best to severe weather applications over CONUS. However, hurricane forecast applications require storm following, telescopic nests at about 1-2 km resolution that can be located anywhere in the globe or in a regional domain and should be capable of following tropical storms for several days. In addition, unlike for severe weather applications (eg. CAM), two-way interactive nests are essential for improving the accuracy of TC forecasts. AOML, in partnership with EMC and GFDL, is working on these developments to transition advances in HWRF to FV3-HAFS under hurricane supplemental (1A4 of the supplemental project).

b. Vortex initialization

VI is one of the key components in the hurricane model system. It consists of vortex relocation, and size and intensity corrections. VI procedure is necessary to provide accurate background fields for the data assimilation. Besides, VI improves the initial intensity of TCs where observation is spare. VI procedure is based on the model start option (cold start or warm start) and initial storm intensity (i.e., maximum wind speed). The basic strategy of this scheme is to extract the hurricane vortex from the previous 6-h hurricane model or GDAS forecast field and relocate and merge it to the model initial field after removing a weak storm vortex in the GDAS field. Especially, before the extracted hurricane vortex is blended with the model initial field, it undergoes size and intensity correction so that the adjusted hurricane vortex is well matched with the observation: the Tropical Cyclone Vitals Database (TCVitals).

c. Inner-core Data Assimilation

Hurricane data assimilation schemes do not have a counterpart. While global models focus on synoptic scale observations, and CAM applications rely on local and storm scale data, both inner core as well as synoptic scale observations are essential for further improving both track and intensity predictions. Central to producing a good analysis is the need for developments of a scale-spanning data assimilation scheme. Though great strides have recently been made in HWRF DA, more work remains to be done. In particular, there are a number of known problems in the current hurricane DA system that will require varying degrees of effort to resolve. These include: (i) Vortex initialization procedures need to work more seamlessly with the data assimilation system. The current procedure, while helpful in some ways, destructively interferes with the data assimilation system when inner-core observations are available. A possible alternative that needs to be explored is to assimilate synthetic observations to supplement inner-core observations. (ii) All state variables need to be carried from one cycle to the next, which is not currently the case in HWRF. Most crucially, HWRF currently does not cycle condensate or vertical motion, which is known to impact the analysis. (iii) The current self-cycled three-dimensional hybrid ensemble-variational (3DEnVAR) HWRF DA system improves upon the old DA system, but more development is needed to improve dynamic balance, particularly for intense hurricanes where inner core gradients are extremely large. Among necessary improvements are an upgrade to four-dimensional hybrid ensemble-variational data assimilation (4DEnVAR) from 3DEnVAR and also to cycle DA more frequently (e.g., every hour instead of every 6 hours). (iv) The current HWRF DA makes suboptimal use of observations. For example, though all reconnaissance data are now assimilated into HWRF, much of this data has had no assumed observation error tuning. Though the HWRF system assimilates satellite radiances, it currently uses bias correction from the global model, which is problematic since HWRF and the global model do not have the same biases. (v) The inner-core data assimilation capability for HAFS will be aligned with Joint Effort for Data Assimilation (JEDI) developments. AOML in joint partnership with EMC is working on these HAFS developments under hurricane supplemental effort.

d. Scale-aware Model Physics

Some of the HWRF, observation-based physics such as the surface and boundary layer, and microphysical parameterization schemes have been found to improve tropical cyclone structure and intensity predictions, which is critical for meeting the HFIP goals. For instance, the boundary layer and surface layer parameterization schemes have been proven to improve hurricane size predictions almost by 50% (Gopalakrishnan, et al., 2013 and Tallapragada et al., 2014). The HWRF physics is currently being transitioned to the HAFS system under 2018 Hurricane Supplemental funding. In addition, HFIP is seeking opportunities for unification of physics between various UFS applications in consultation with the UFS Physics Working Group (3A1 and 3A2 of the supplemental project).

e. Two-way Ocean coupling

The ocean model component of HAFS will use HYbrid Coordinate Ocean Model (HYCOM) that is based on 3D free-surface, primitive governing equations. Solutions are sought on Arakawa C-grids at resolutions of 1/12-degree and 41 hybrid z-sigma in horizontal and vertical, respectively. Initial and boundary conditions (ICs/BCs) are provided in real-time via subsetting NCODA-based nowcasts and forecasts from global Real-Time Forecast Ocean System (RTOFS), respectively. Subgrid turbulence mixing is simulated by KPP mixing. For better simulations of the upper ocean structure, particularly of freshwater barrier and freshwater lenses, use of model precipitation and river freshwater discharge will be included in the future. A plan for ocean DA is to employ RTOFS-DA based on the 3DVAR approach, which replaces the subset of global RTOFS nowcasts.

f. Observations

Apart from synoptic-scale observations used for NWP and in global model data assimilation schemes, airborne observations are critical for improving TC predictions. In the Atlantic basin, Air Force Reserve C-130 and NOAA WP-3D aircraft are used to sample TCs whenever possible to provide critical observations of the location, strength, and structure of the storm circulation. Sampling of the environment is typically accomplished by the NOAA G-IV aircraft. These manned aircraft are equipped with a variety of instruments that sample the wind, temperature, moisture, pressure, precipitation, and ocean surface and subsurface temperature and salinity, current, and wave fields within and around TCs (e.g., with flight-level measurements, dropwindsonde, airborne Doppler radar, Stepped Frequency Microwave Radiometer, lower fuselage radar, and airborne expendable bathythermographs/current profilers). Experimental airborne observing technologies, such as Light Detection and Ranging (LIDAR), have the ability to sample the wind field in the absence of precipitation scatterers. Unmanned aerial systems, such as the Coyote and Global Hawk can sample temperature, moisture, and pressure fields in the planetary boundary layer of hurricanes, and over vast areas at very high altitudes for extended periods of time, areas that can't be reached by manned aircraft because of safety and/or aircraft performance limitations. These experimental observing technologies could potentially fill gaps in the current observing system, providing critical measurements needed to more fully capture the structures important to TC structure and intensity change. Many of the inner-core observations provided by AOML have been used for not only improving DA but also for improving model parameterization schemes. HAFS will take advantage of advancements in these observing technologies to optimize sampling of the TC inner-core and environment and provide the needed support for forecast, analysis, model initialization and evaluation, current and future data impact studies (OSEs and OSSEs), and process studies.

Remote-sensing sea surface temperature (SST), sea surface salinity (SSS) and absolute dynamic height, temperature and salinity profiles from various observing platforms are routinely used for Ocean DA at this time. However, there are a couple of invaluable ocean observing programs, such as the US Integrated Ocean Observing System (IOOS) Program and Global Drifter Program (GDP), which at least provides synoptic oceanic conditions. Systematic ocean target observations collecting surface and subsurface temperature and salinity before, during and after a TC are ideal to provide more realistic enthalpy flux exchange and accurate assessments of TC ocean response at a TC scale. In particular, concurrent and co-located samples covering both the air and sea (including the air-sea boundary layer) near the TC field are absolutely crucial. Future sUAS observations (and SST sondes) could be helpful with several existing (and new/proposed) requirements.

While active developments of the HAFS system enlisted above are ongoing, four HAFS configurations were run under Stream-2. Some of the preliminary results where the operational models struggled, showed promise in the next generation hurricane forecast system i.e. HAFS.

2.6. Important HREx Results: HAFS Experimental systems

There has been steep-step progress in HAFS testing in the last three years. In 2019, HREx demonstrated the skill of two versions of HAFS in predictions of TC track and intensity (HAFS-SAR or HAFS-A; Dong et al. 2020 and HAFS-globalnest or HAFS-B; Hazelton et al. 2021). The 2020 experiments built off of this success with further improvements to HAFS. In 2020, both versions of HAFS have evolved into a unique testbed for different sets of activities. In 2021, the real-time experiments featured tests of several different configurations of HAFS, allowing for tests of different grid layouts, physics options, and initialization methods in advance of operational implementation in 2023. Four configurations of HAFS were conducted in the 2021 real-time hurricane season, with detailed information listed in Table 2.

Table 2: Model configurations for the 2021 real-time, HAFS-SAR (HAFA), HAFS-globalnest (HAFB), HAFS-SAR With DA (HAFD), and HAFS-SAR ensemble experiments.

	HAFS-A	HAFS-B	HAFS-D	HAFS-E
Resolution/ Model top	~3km (ESG), L91/10hPa	~13-3km global-nest, L75/2hPa	~3km (ESG)/L91, 10hPa	~6km, L64, 10hPa
Domain	~94º✕65º, 3121✕2161	Global ATM:(C768), Nest ATM:~79º✕43º OCN: ~330º✕89º	~94º✕65º, 3121✕2161	~86º✕58º, 1441✕1081
IC/BC	GFSv16/3hrly	GFSv16/3hrly	GFSv16/3hrly	GEFS/6hrly
Coupling Ocean IC	CMEPS-HYCOM RTOFSv2	CMEPS-HYCOM RTOFSv2	CMEPS-HYCOM RTOFSv2	No ocean model NSST
Data Assimilation	No	No	Yes (addl:TDR, METAR, meso GOES-R AMVs)	No
Radiation	RRTMG (30min)	RRTMG(30min)	RRTMG(30min)	RRTMG(60min)
PBL/Surf GWD	M-TKE-EDMF/M-GFS orographic GWD	M-TKE-EDMF/M-GFS saGWD	M-TKE-EDMF/M-GFS orographic GWD	M-TKE-EDMF/M-GFS orographic GWD
CP/MP	saSAS/GFDL	saSAS/GFDL	saSAS/GFDL	saSAS/GFDL
LSM	NOAH	NOAH	NOAH	NOAH

The hurricane track and intensity forecast skills of these four experiments are compared along with two NOAA's current operational tropical cyclone prediction systems, HWRF and HMON (figure 12). The results demonstrated that HAFS configurations have skillful track forecasts than HWRF, except for HAFS-E track forecasts after day-3, likely due to coarser horizontal and vertical resolutions than other configurations and due to lack of ocean coupling (Figure 15a). The intensity forecast skills are mostly improved in all HAFS experiments after day-2, but are still behind HWRF before day-2 (Figure 15b). It should be noted that the vortex initialization (VI) procedure was not included in these HAFS experiments. The results indicate the importance of VI procedure and inner-core DA for the intensity forecasts at earlier forecast hours.

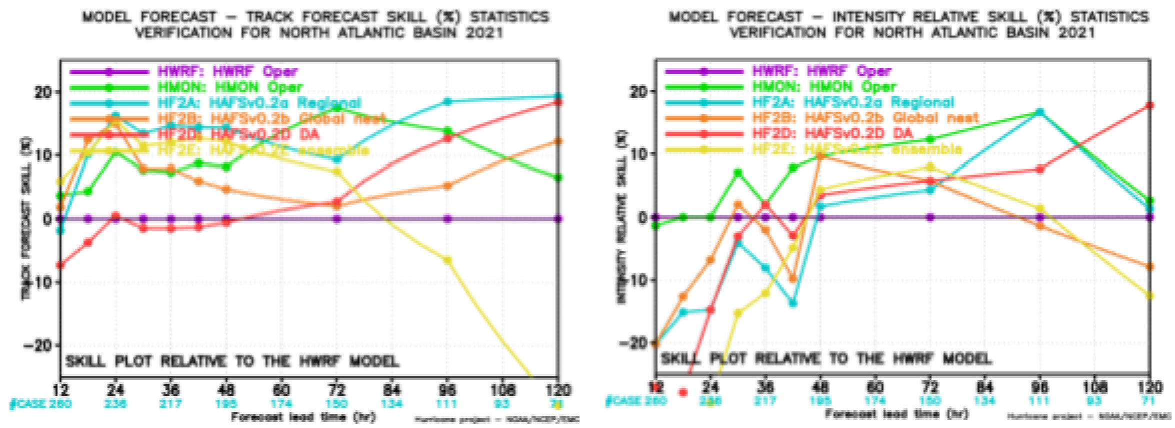

Figure 12: Track (a) and intensity (b) forecast skills from the 2021 season for HWRF (purple), HMON (green), HAFS-A (cyan), HAFS-B (orange), HAFS-D (red), and HAFS-E (yellow).

a. HAFS-A experiment (HAFS v0.2A)

HAFS-A was the stand alone regional (SAR) version of HAFS, featuring a stand-alone static nest domain covering the North Atlantic basin. HAFS-A is an atmosphere/ocean coupled system, and serves as a baseline for other configurations.

The atmospheric component (FV3) of HAFS-A configuration uses a C3091 (3-km) regional Extended Gnomonic Grid (ESG) with 91 vertical levels. The experiment runs four times a day at 00, 06, 12, and 18Z cycles, when there are active storms. Each cycle will produce a 126-hour forecast with 3-hourly outputs, including the ATCF format track file with the storm positions and intensities. The initial condition and 3-hourly lateral boundary conditions for the atmospheric component model come from the operational GFS netcdf and grib2 format input files.

The ocean component uses the HYCOM ocean model, which is at 1/12-degree horizontal resolution with 41 hybrid z-isopycnal layers. The ocean model takes initial conditions subsetting from the nowcast (for 00Z cycle) and forecast products (for 06, 12 and 18Z cycles) of the global RTOFS, and uses the persistent lateral boundary conditions. The ocean model products include 3-hourly HYCOM native binary data and 6-hourly z-level netCDF files that include water temperature, salinity, horizontal and vertical velocities, mixed layer depth, ocean heat content, the depth 20 and 26 degree C isotherm over the upper 350 m depth.

Figure 13 compares the track and intensity forecast errors between HAFS-A and two operational TC prediction systems, HWRF and GFS. The results show that the track forecast errors are lower than HWRF and comparable with GFS, while intensity forecast errors are lower than HWRF after 48 h forecast lead time. The HAFS-A configuration was also run in quasi real time for Northeastern Pacific and Northwestern Pacific basins, the verification results are similar as that in the North Atlantic basin (not shown). The wind-pressure relationship produced by HAFS-A is also compared with that from HWRF in figure 14, which clearly shows an improved wind-pressure relationship than HWRF and matches better with the best track data.

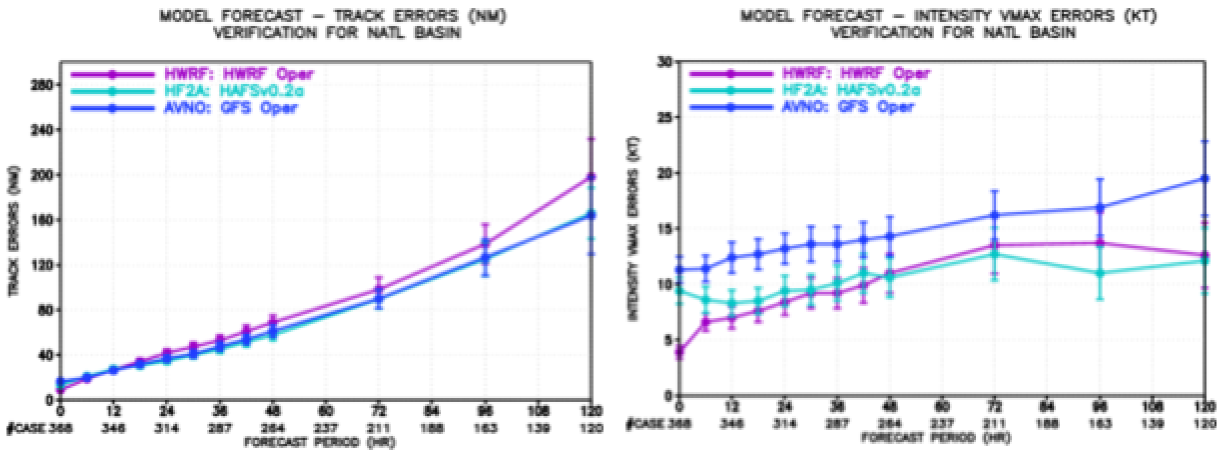

Figure 13: Track (a) and intensity (b) forecast errors from the 2021 season for HWRF (purple), HAFS-A (cyan), and GFS(blue).

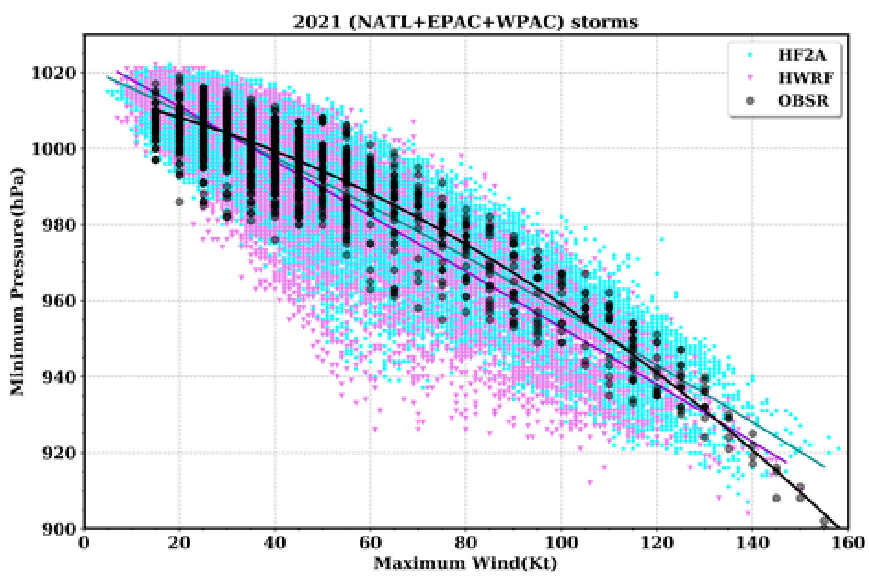

Figure 14: Wind-pressure relationship from HAFS-A (cyan), HWRF (purple), and best track (black).

b. HAFS-B experiment (HAFS v0.2B)

HAFS-B was the global-nested version of HAFS, featuring a 3-km static nest covering the North Atlantic basin, with 2-way feedback with a simultaneously-running 13-km global domain. For the first time, the 2021 version of HAFS-globalnest (HAFS-B) was coupled to an ocean model, with the nested domain coupled to the HYCOM ocean model, similar to the configuration used in HAFS-A. Other configuration options that were unique to HAFS-B included the use of a modified version of the EDMF-TKE scheme to better match observational estimates of eddy diffusivity and mixing length (Gopalakrishnan et al. 2021, Hazelton et al. 2022) as well as the use of a less diffusive tracer advection scheme (Gao et al. 2021). Figure 15a shows that the track results from HAFS-B were generally similar to HAFS-A over the North Atlantic. The intensity errors were also similar to HAFS-A (Figure 15b), although HAFS-B had slightly larger errors at longer leads, mostly due to a high bias in Hurricane Larry (not shown).

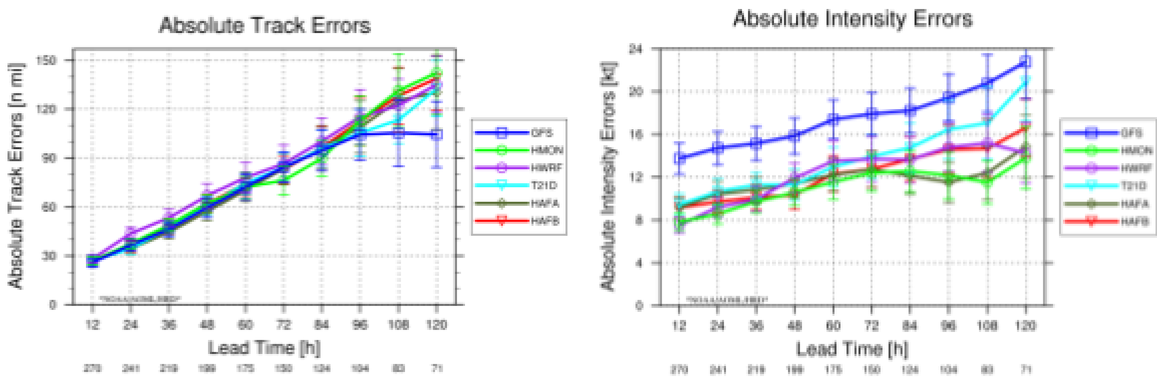

Figure 15: North Atlantic (a) mean track forecast errors and (b) mean absolute intensity errors from the 2021 season for HAFS-A (dark green), HAFS-B (red), GFDL T-SHiELD (light blue), operational GFS (dark blue), operational HWRF (purple), and operational HMON (light green).

Unlike HAFS-A, HAFS-B was run out to 7 days, and showed promising track performance in long-range track forecasts in 2021 (Figure 16a), with results comparable to or better than both the operational GFS and the GFDL T-SHiELD models. Another interesting aspect of HAFS-B was that the intensity biases from the East Pacific, on the global domain, were significantly better than the operational GFS and were similar to HWRF (Figure 16b). This indicates the importance of the high resolution forecasts of upstream disturbances (from the high resolution nest over the Atlantic) and motivates ongoing development towards an eventual multiple-moving-nest configuration in the global-nested version of HAFS.

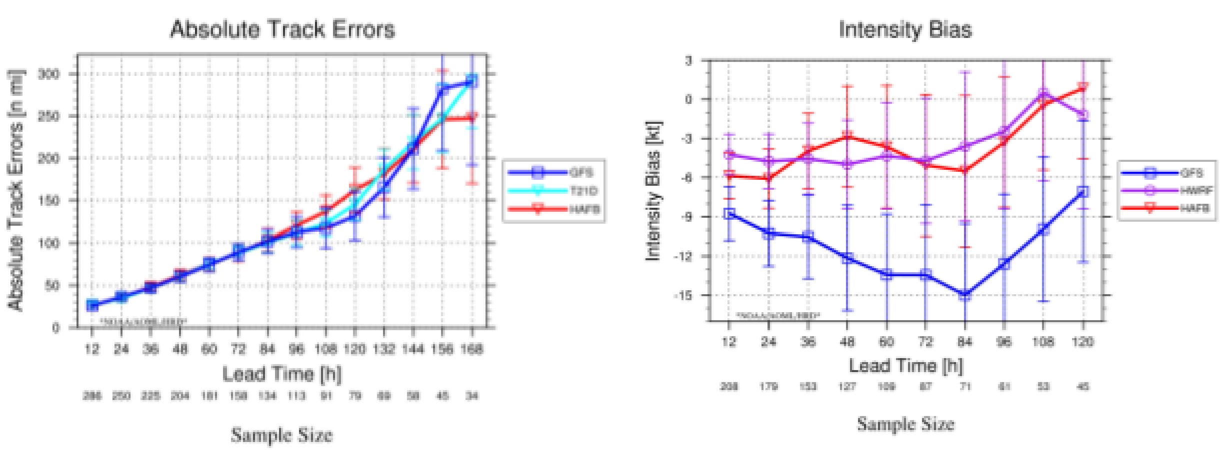

Figure 16: Eastern North Pacific (a) track forecast errors from HAFS-B (the global domain, red), operational GFS (blue), and GFDL T-SHiELD (cyan); and (b) intensity forecast bias from HAFS-B (the global domain, red), operational GFS (blue), and operational HWRF (purple).

c. HAFS-D experiment (HAFS-v0.2D)

HAFS-D was an experiment which was designed to see the impact of data assimilation. This experiment used the configuration as HAFS-A, but included the following DA capabilities: 3-hourly FGAT, 3DEnVar with GDAS ensembles, assimilate all observations ingested by the operational HWRF/GFS/GDAS systems. Because of impressive intensity improvement from retrospective run with enhanced GOES-16 AMVs assimilation (Figure 17), HAFS-D experiment included additional enhanced GOES-R AMVs. In real time, HAFS-D had comparable track skill to other HAFS and HWRF and good intensity skill after 36 hrs (Figure 18).

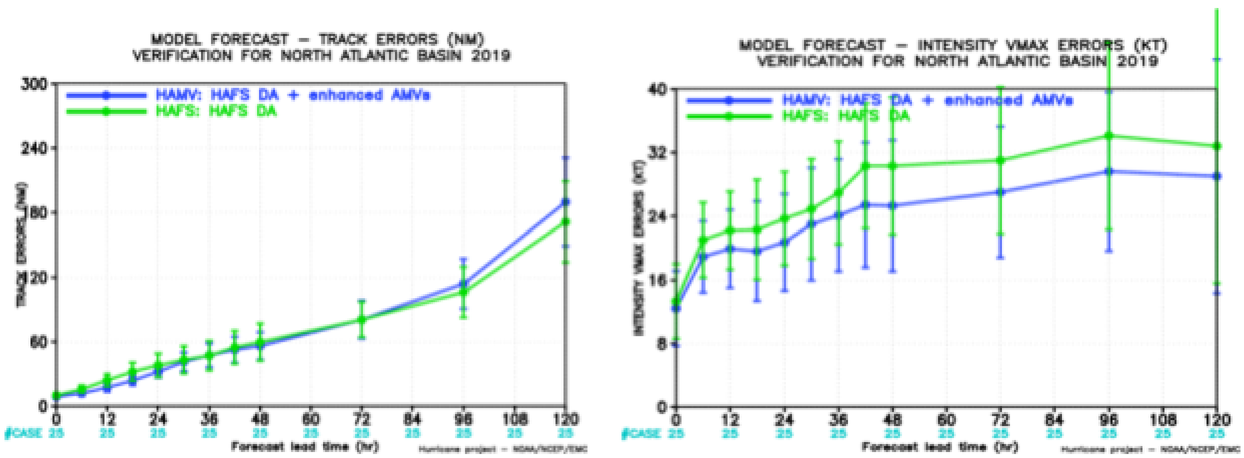

Figure 17: Track and intensity forecast errors from the experiments with (blue) and without (green) enhanced GOES-R AMV data assimilated.

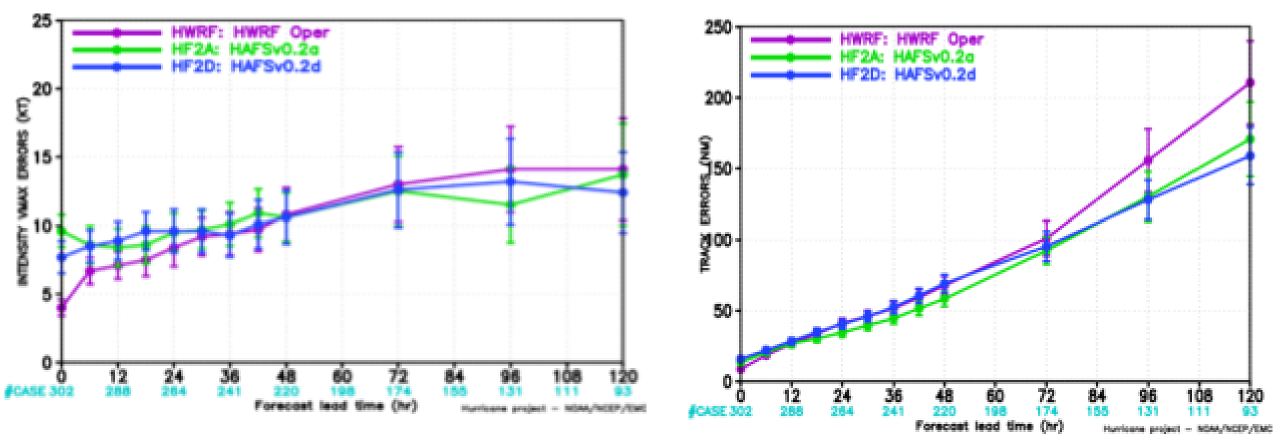

Figure 18: Track and intensity forecast errors from the 2021 season for HWRF (purple), HAFS-A (green), and HAFS-D (blue).

d. HAFS-E experiment (HAFS v0.2E)

HAFS-E was an ensemble experiment, which includes one unperturbed member (member 0) and 20 perturbed ensemble members. The HAFS-E configuration was based on HAFS-A configuration, except for the following modification to save computer resources: i) lower resolution for the static nest (~6 km and L64), ii) no ocean coupling, and iii) Slightly larger physics calling steps. Details can be found in Table 2. Global Ensemble Forecast System (GEFS) output files are used as initial and lateral boundary conditions for HAFS-E to account for large-scale flow uncertainties. Three types model physics perturbations, Stochastically perturbed physics tendencies (SPPT), Stochastic kinetic energy backscatter (SKEB), Stochastically perturbed PBL humidity (SHUM), are included to account for model physics uncertainties. Two ensemble mean methods, all ensemble member average (HFMN) and sub-setting ensemble average (HS12), are used to represent ensemble track and intensity forecasts. The ensemble results are compared with three experiments, unperturbed lower resolution deterministic member 0, high resolution HAFS v0.2A, and its host model GEFS. Figure 19 shows the track and intensity forecast skill comparison. The following points can be clearly seen -(a) HAFS-A is more skillful than unperturbed ensemble members in terms of both track and intensity at all lead times, and has better intensity bias. (b) Equally-weighted HAFS-E ensemble-mean improved the track forecast by ~5% at all lead times, the intensity forecasts by > 10% after day-2 over its deterministic model (HF00). The subset of ensemble-mean further improved track/intensity forecasts, especially before day-2. HAFS ensemble mean track forecasts outperformed its host model GEFS in the short lead hours (< 60h).

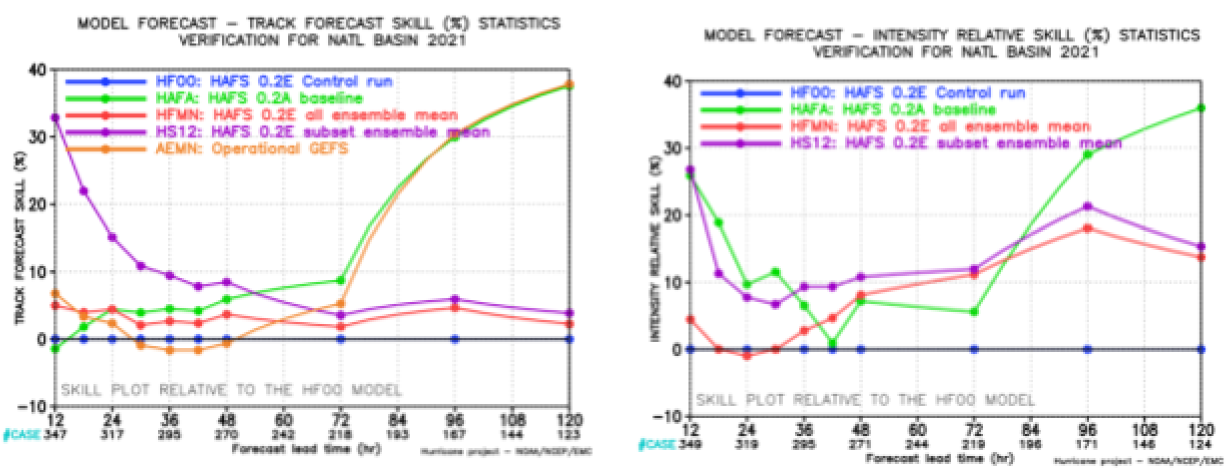

Figure 19: Track and intensity forecast skills from the 2021 season for unperturbed ensemble control (HF00, blue), HAFS-A (green), all member ensemble mean (HFMN, red), subset ensemble mean (HS12, purple), and HAFS-E host model GFSF (AEMN, orange).

2.7. New Products, Tools, and Services at NHC

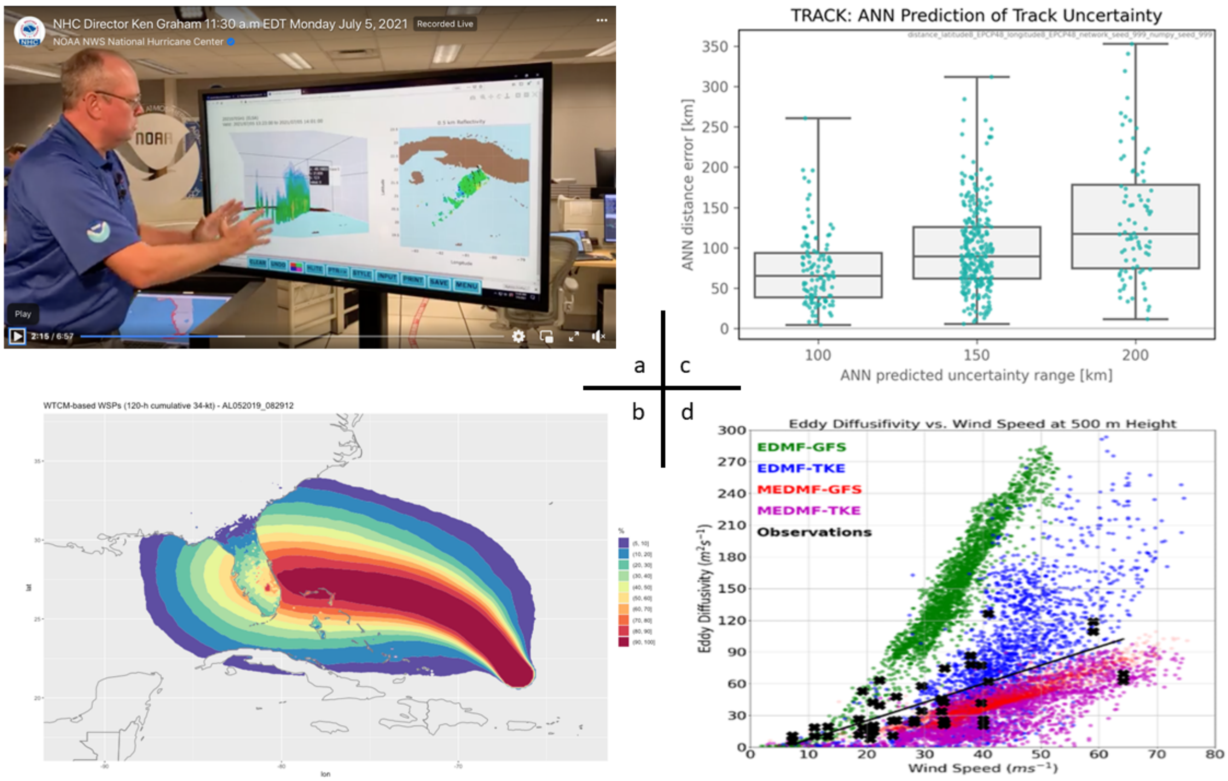

Figure 20: Examples of HFIP post processing and verification accomplishments in 2021: a) NHC director Ken Graham uses 3D graphics of aircraft Tail Doppler Radar during a public Facebook Live briefing ahead of the landfall of Hurricane Elsa; b) Experimental forecast using the "WTCM"-based wind speed probability model, which realistically highlights large differences between land and water points; c) output from a machine learning technique which aims to better quantify forecast uncertainty; d) Observations of eddy diffusivity vs wind speed at 500 m vs model output from the HAFS with a variety of parameterization schemes (from Gopalakrishnan et al. 2021).

a. Operational and Real-Time Applications

HFIP supported several efforts to improve operational and real-time products at the National Hurricane Center (NHC) in 2021. Updates were applied to the HFIP Corrected Consensus Approach (HCCA) model, the Statistical Hurricane Intensity Prediction Scheme (SHIPS), and the Logistic Growth Equation Model (LGEM), and an effort to migrate HCCA to permanent operations is underway. A major upgrade of the NWS operational probabilistic surge model was implemented in Spring 2021 to improve representation of the radius of maximum winds. The NHC and CIRA conducted evaluations of rapid intensification forecasts with large errors to improve the SHIPS-Rapid Intensification Index (SHIPS-RII) and also conducted an evaluation of the updated COAMPS-TC model. The NHC and CIRA are also testing new machine learning techniques to improve prediction of intensity, including rapid intensification. In addition, machine learning techniques are being developed to better quantify the uncertainty of forecasts (Figure 20c).

Progress was also made toward improving public forecast products and warnings from the NHC and the National Weather Service (NWS). This includes updates to the wind speed probability model and the "WTCM" – a gridded representation of the NHC forecast used by the NWS to keep gridded forecasts consistent between offices. One specific effort to improve the WTCM focused on the use of the NHC forecast wind radii, which are the maximum in a quadrant, but are converted to the average in a quadrant for the WTCM. HWRF surface wind forecasts from the past two years are being used to develop a more accurate conversion from maximum to average radii. An effort to merge the methodologies of the wind speed probability and WTCM models is also underway. This should improve the wind speed probabilities over land (Figure 20b). The improved wind speed probabilities will contribute to an effort to improve coastal and inland tropical storm and hurricane warnings using a new innovative collaboration process between NWS offices using the AWIPS-2 platform known as the Wind Hazard Recommender. A test of this new software was recently conducted in March 2022. HFIP also supported activities that will lead to improved public products. Development began on one of those in 2021, a probabilistic landfall intensity product that will meet a need in the emergency response community. Finally, NHC was able to use the result from a hurricane supplemental project focused on 3D visualization of model and aircraft for public briefings during the 2021 hurricane season. NHC director Ken Graham used the images during live briefings ahead of the landfall of Hurricane Elsa to highlight information about the storm and to demonstrate the work of the hurricane hunter aircraft (Figure 20a).

b. Display and Diagnostic Activities

As in past years, the HFIP community worked to improve model diagnostic and visualization techniques in 2021. Many of the tools will be used during the upcoming operational transition of HAFS by allowing model developers to evaluate the model beyond traditional track and intensity forecasts. NOAA's Hurricane Research Division (HRD) developed visualization tools to evaluate parameterization schemes in HWRF and HAFS and compare the model output to observations (Figure 20d; Gopalrkrishnan et al. 2021). The HRD also continued to directly compare model output with tail doppler data, investigated the impact of the Coyote unmanned aircraft observation platform on model initializations, and maintained a web viewer that hosted over 50 million real time graphical products from HFIP in 2021. Other visualization tools from ESRL and NCAR were supported and improved, including the brand new hfip.org, which debuted in 2021. Updates were applied to TC-specific web tools like the NCAR "NHC display" (products.hfip.org/nhc-display), which can assist both the NHC and the wider community with model evaluation and real-time forecasting. The NHC hopes to use the NCAR display tool to assist with post-storm analysis in the future as well. A new feature was added to the NCAR display tool in 2021 to allow multiple TCs to be displayed on the same plot to assist with post-storm analyses.

2.8. Community Involvement

Research to Operations (R2O) was one of the initial goals of the WRF program and is supported by HFIP in developing a repository for a community-based hurricane modeling system, which ensures the same code base can be used for research and in operations. During 2009-2016, both the EMC and the DTC worked to update the operational version of HWRF from version 2.0 to the community version of HWRF, version 3.9a. The 3.9a version made the operational model completely compatible with codes in community repositories, allowing researchers to access the operational codes. Hence, the improvements in HWRF, developed by the research community, were easily transferable into operations. DTC has played a significant role to help the HWRF community by conducting HWRF training sessions twice per year from

2010-2018, two of which were international. In addition, twelve Community Workshops on topics ranging from physics, observations, ensemble product development, satellite DA, to social science were conducted. In July 2018, the code version of the HWRF system v4.0a was available for the HWRF community. Since then DTC has continued to provide user support. Apart from US, there are about one thousand HWRF model users in about 200 countries[5]. User support was expanded with the Stream-2 efforts, the significant one being the Basin-Scale HWRF. This research system can support any number of high-resolution movable nests centered on TCs in either the Atlantic or eastern North Pacific basin. Working with HRD, the DTC also supported the transition of this research version to the latest community repository, enabling users to access all advancements in the HWRF system including the end-to-end Basin-Scale configuration (excluding ocean coupling and data assimilation). A similar testbed activity is recommended for transitioning the proposed HAFS.

2.9. NOAA Federally Funded Opportunity (FFO)

The following tables (Table 3 and Table 4) provide the list of projects supported by HFIP during 2018-2020 and 2020-2022.

Table 3: HFIP Supported Projects from Awards Round V 2018-2020.

PI Name	PI Institution	Project Title	Status
Agnes Lim	University of Wisconsin (UWI)	Advanced DA Techniques for Satellite-Derived Atmospheric Motion Vectors from GOES 16/17 in the HWRF	New assimilation techniques developed for GOES-16/17 AMVs will be offered for transition to operations.
Andrea Schumacher	Colorado State University (CSU)	Using Dynamically-Based Probabilistic Forecast Systems to Improve the NHC Wind Speed Products	A new version of the Monte Carlo wind speed probability model (MC model) that directly uses data from NCEP global and/or regional ensemble prediction systems was developed, validated, and is running in a semi-operational environment at CIRA.
Kerry Emanuel	Massachusetts Institute of Technology (MIT)	New Frameworks for Predicting Extreme Rapid Intensification	The development of Forecasts of hurricanes using large-ensemble output (FHLO) is completed.
Ping Zhu	Florida International University (FIU)	Rapid Intensification Changes: Improving Sub-Grid Scale Model	Static stability correction in the eyewall and rainbands, TKE turbulent mixing scheme combined

[5] https://www.emc.ncep.noaa.gov/gc_wmb/vxt/HWRF

PI Name	PI Institution	Project Title	Status
		Parameterization and Microphysical-Dynamical Interaction	with stability correction have been implemented in both HWRF and HAFS.
Ryan Torn	SUNY Albany	Evaluating Initial Condition Perturbation Methods in the HWRF Ensemble Prediction System	The milestone of validating probabilistic wind and precipitation forecasts from ensemble prediction systems is at RL-6; validating quantitative and probabilistic ePHRaM forecast is RL-5; and the ensemble-based precipitation sensitivity work is at RL-5, with the hope of advancing it toward RL-6.
Ting-Chi Wu	Colorado State University (CSU)	Enabling Cloud Condensate Cycling for All-Sky Radiance Assimilation in HWRF	A new development branch of GSI named icda_dev_cira was created and implemented code modifications to enable cloud condensate cycling via all-sky radiance assimilation in HWRF. Then a routine branch merge was conducted to ensure that the icda_dev_cira branch syncs with the latest change contained in the HWRF branch of GSI, which is used by the operational HWRF.

Table 4: HFIP Supported Projects from Round VI 2020-2022.

PI Name	PI Institution	Project Title	Status
Alan Brammer	CSU-CIRA	Extending the Tropical Cyclone Genesis Index to Global Ensemble Forecasts	The project has implemented the ensemble based genesis guidance to run on invests in all basins in real time.
Enrique Curchitser	Rutgers	Developing Regional Ocean Modeling Capabilities with MOM6 for use in the UFS	The project has made progress on targeted regional MOM6 domains for HAFS application with improved surface and boundary conditions.
Ryan Torn	SUNY Albany	Application of Innovation Statistics to Diagnose Biases in the HAFS System	The project has made substantial progress in the development of storm-centric innovation biases and identifying the relationship between

PI Name	PI Institution	Project Title	Status
			innovation biases between different vertical and horizontal locations using advanced statistical approaches.

2.10. Socio-economic Aspects of HFIP

The section 104 of the Weather Act 2017, as well as the hurricane supplemental funding provided NOAA with a unique and important opportunity to integrate the social, behavioral and economic sciences into NOAA's tropical products and services, as well as incorporate risk communication research into the design and communication of its products. To accomplish these goals, the Office of Oceanic and Atmospheric Research (OAR)'s Weather Program Office (WPO) worked side by side with the National Weather Service (NWS) to identify relevant operational challenges, develop project descriptions, and fund four social and behavioral science projects:

i. **There's a Chance of What? Assessing Numeracy Skills of Forecasters, Partners, and Publics to Improve Tropical Cyclone Product Uncertainty, IDSS, and Training.** The goal of this project was to examine how end-users, such as forecasters, emergency managers, and the public interpret and comprehend probabilistic tropical cyclone information. This was explored through the use of a concept known as numeracy, or one's ability to use and understand numerical information.

ii. **Minding the Gap: Modernizing the Tropical Cyclone Product Suite by Evaluating NWS Partner Information Needs.** By interviewing and surveying NWS partners, specifically emergency managers and broadcast meteorologists, this project was designed to help NWS prioritize their efforts to modernize their tropical cyclone product suite and identify gaps needed to enhance NWS partner decision-making.

iii. **Wait, that Forecast Changed? Assessing How Publics Consume and Process Changing Tropical Cyclone Forecasts Over Time.** This project explored how various publics consume and process changing tropical cyclone forecasts over time. To do this, this project developed a social science methodology to deploy surveys before, during, and after tropical cyclone events to measure the public's information-seeking behavior, risk reception, and protective action responses in real-time.

iv. **Optimizing Tropical Cyclone Information: An National Hurricane Center Web User Experience Study from a Public Perspective.** Using a combination of user-centered design and usability study approaches, the goal of this project was to evaluate the usability of National Hurricane Center's (NHC) webpage and help NOAA identify various design opportunities to modernize the NHC's web presence.

These four social and behavioral science research projects were developed with a purposeful, complementary design (Figure 21). Instead of creating individual projects that would offer discrete findings and recommendations, this complementary approach created an opportunity to build a collective body of research whereby the cross-cutting findings from each project could build on one another to provide more generalizable findings about the suite of tropical cyclone products and services. Similarly, the differences among each of the four projects was also strongly considered. The OAR-NWS social science team wanted to intentionally create projects that differed in the audience examined (i.e., general public, emergency managers, broadcast meteorologists, and/or forecasters), their theoretical focus, and their application to also provide unique findings and research-guided recommendations that addressed specific operational gaps or needs. Not only did this result in four projects that incorporated risk communication research in a meaningful way to empirically examine the NWS tropical cyclone product suite, it also created an opportunity for the four project teams to collaborate with one another as a research cohort.

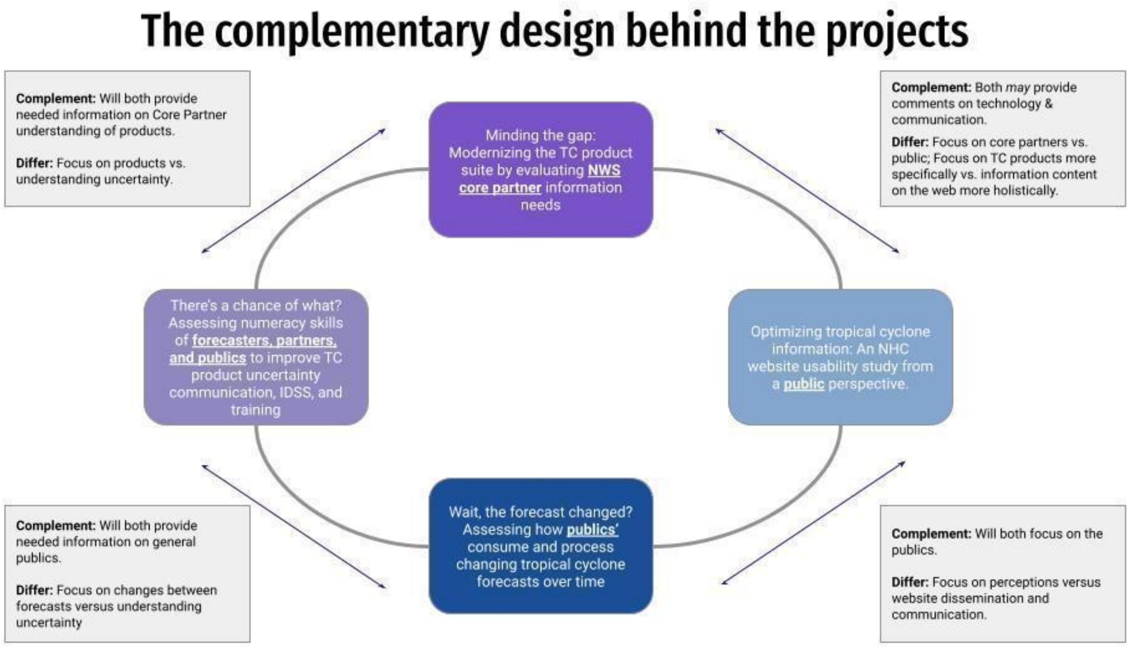

Figure 21: The purposeful, complementary design behind the projects.

After each project team was established, the OAR-NWS social science team brought the project teams together as a cohort early in the research process to nurture cross-learning, collaboration, and rapport development. These collaborations were first developed through a virtual Tropical Socio-Econ Virtual Workshop in June 2020. At this workshop, each project team provided an overview of their project and, after hearing from all four projects, offered suggestions on how they envisioned productively collaborating with other project teams within the cohort. These collaborative conversations with both the research teams and the OAR-NWS social science team continued throughout the project period. As the projects progressed and began collating early research findings, the cohort met more frequently to socialize their research findings, determine whether results from other research teams resonated with them, and if so, identify cross-cutting takeaways and findings across two or more projects. Although we are still waiting for all of the final reports, these engagements with the project teams *throughout* the award period provided the OAR-NWS social science team the unique opportunity to begin triangulating research findings across all four social and behavioral science projects.

Using conference presentations, draft reports, and some final reports, the OAR-NWS social science team began triangulating the preliminary findings from each project to identify high-level themes or concepts that were common among the four projects (Figure 22). Similar to the concept of triangulation in a navigational sense, triangulation can also be used in the context of social science research by using more than one method or approach to investigate a topic or research question. Because the OAR-NWS social science team developed the four projects with a purposeful, complementary design, it was possible to triangulate the research findings across all four projects. According to Heale and Forbes (2013)[6], "the combination of findings from two or more rigorous [social science] approaches provides a more comprehensive picture of the results than either approach could [provide] alone." The next few paragraphs will provide some big themes and preliminary takeaways from the four social and behavioral science hurricane supplemental projects. However, please keep in mind that these findings are still preliminary and that our triangulation efforts are still ongoing. These efforts will continue until we receive all four final reports near the end of FY22.

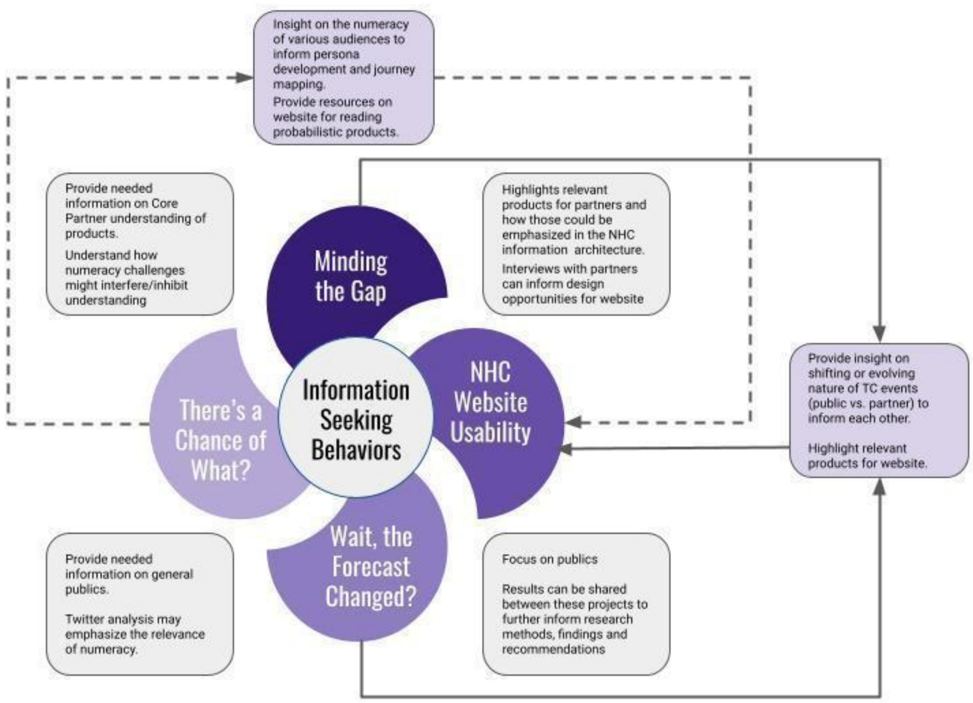

Figure 22: Ongoing triangulation efforts to find similarities across projects.

Broadly speaking, the biggest takeaway from the four projects is that broadcast meteorologists, emergency managers, and members of the public find NWS' tropical cyclone products and services useful and important. However, thanks to the purposeful and complementary nature of the four projects, each project also provides unique insight on how NWS products and services could be further enhanced to improve end-user usability, understanding, and decision-making. As a reminder, the big themes and takeaways are still preliminary and our triangulation efforts are still ongoing. Across all four projects,

[6] Heale R, and D. Forbes, 2013: Understanding triangulation in research. Evidence-Based Nursing, ***16***(4), pg. 98; doi:https://doi.org/10.1136/eb-2013-101494

there are eight high level themes that emerge. These themes are ordered based on how often they appeared across the four projects. As such, the first theme represents the most consistent findings compared to the final theme, which does not emerge as often in the presentations and reports.

- **Identify ways to localize and personalize information for end-users.** Broadcast meteorologists, emergency managers, and members of the public have a strong desire for NWS tropical cyclone products and services to be more specific and local to their area. End users, for example, want to be able to find themselves on various tropical cyclone graphical products. They want to be able to type in their zip code or zoom into their location to find additional information about their local area.
- **End users search for different types of tropical cyclone information during different phases in the lifecycle of a tropical cyclone threat.** All four projects provide specific information on the information-seeking tendencies of broadcast meteorologists, emergency managers, and members of the public. This may have implications on product development, refinement, and/or operational changes to the issuance of products/services to more clearly align with end user needs.
- **Timing is important for critical decision making, as as a result, the timing of when forecasts are issued is important too.** Timing plays an interesting and multifaceted role in these project findings. Broadcast meteorologists, emergency managers, and members of the public are all interested in the timing of the arrival of various tropical cyclone impacts. Timing was also prevalent in terms of partner's decision-making timelines. Broadcast meteorologists, for example, explained that they do not have much time to decipher the latest forecast information and make changes to their in-studio graphics before going live on-air.
- **Forecast uncertainty is important to communicate, but is not always communicated well.** Broadcast meteorologists and emergency managers believe forecast uncertainty is one of the most important pieces of information to communicate early in a tropical cyclone event. However, not all tropical cyclone products or services are effective at reaching low-numerate populations. Therefore, best practices and research-guided recommendations from previous research should be used to improve the communication of probabilistic and/or uncertainty information.
- **Graphical products are important for risk communication, but sometimes need to improve their depiction of risk and/or uncertainty.** Several projects highlighted the value of NWS graphical products for tropical cyclone risk communication. However, not all graphical products do this effectively. Findings from these projects suggested that graphical products are more valuable when meteorologists co-produce or co-develop products and services alongside partners and end-users. Co-development ensures that *all* individuals are able to access, understand, and use these products when making decisions.
- **There is a misperception among forecasters and partners that members of the public do not understand uncertainty information.** Instead of providing numerical information when communicating uncertainty information, these projects revealed that forecasters and partners often use vague words and phrases. This likely has a chain reaction, such that this watered-down uncertainty information does not offer beneficial information to members of the public. Because members of the public do not find this information helpful when making decisions, this likely fuels the perception that members of the public do not understand uncertainty information.

- **There is a misperception that emergency managers are as highly numerate as weather forecasters.** Although emergency managers are specialized users of weather information, it does not mean that they are as highly numerate as many weather forecasters. Findings from these projects revealed that emergency managers are generally more numerate compared to members of the public, but not to the level of weather forecasters. In fact, emergency managers' average numeracy ratings are closer to the ratings for members of the public.
- **NWS needs to increase the accessibility of tropical cyclone products and services.** The OAR-NWS social science team sees value in exploring the findings through an accessibility lens, especially given the current administration's focus on Diversity, Equity, Inclusion, and Accessibility. In particular, the NHC Website project offers various website and graphical design best practices to improve accessibility (e.g., Screen Reader Capability). However, the team also sees value in thinking about accessibility in terms of low-numerate individuals and ensuring those individuals have more access to tropical cyclone products and services.

Although our triangulation efforts are still ongoing, the OAR-NWS social science team, in collaboration with the NWS Tropical Roadmap Team, started thinking about the process for translating social science findings from the four projects for possible transition into operations. While translating research outputs into operations is a priority, it is important to first evaluate the social science findings' readiness for transition from both a research (i.e., generalizability) and operational perspective (i.e., operational viability or feasibility).These evaluations will determine whether a research output is ready for transition to operations, or whether additional physical and/or social science research and development (R&D) may be needed prior to implementation. In fact, some of the research findings that require more R&D are especially interesting to the OAR-NWS social science team. These social science projects, for example, point to physical science capabilities that end-users want, but are not yet operationally feasible. Therefore, in addition to providing research-guided recommendations on how to improve the tropical cyclone product suite, these projects also exemplify the interconnectedness of social and physical science R&D and how one can inform the other—and vice versa. In the interim, the OAR-NWS social science team plans to continue translating findings from the four social and behavioral sciences into relevant R&D needs and applications. We look forward to continuing our ongoing collaboration with the NWS Tropical Roadmap Team, as we explore these potential applications together.

2.11. HFIP State-of-the-art and HAFS developments

In 2009, NOAA established the 10-year HFIP to accelerate the improvement of forecasts and warnings of tropical cyclones and to enhance mitigation and preparedness by increasing confidence in those forecasts. Regional models with moving nests were created especially to address the problem of intensity changes in TCs. Global models cannot currently address the intensity forecast problem because horizontal resolutions are too coarse, limited by operational high performance computing (HPC) resources, to capture the hurricane eyewall and the inner-core structure of the hurricanes critical for predicting intensity changes (section 4).

Sustained HFIP investments in research and development (R&D) and HPC led to the creation and transitions of the high-resolution HWRF system from research to operations (R2O). This system is now paving the way around the globe, and removing the initial roadblocks associated with predicting intensity changes with the dynamical prediction, which was nearly non-existent until 2009 (Figure 2b). HWRF has improved by at least 15-30% since 2011 over the Atlantic basin (Figure 3b). Since 2014, HWRF has run operationally in all global basins and is used by forecasters for reliable intensity guidance worldwide. Significant improvements to the HWRF system are attributed to a number of major changes since 2012, including a new, higher- resolution moving nest capable of better resolving eyewall convection and scale interactions, improved planetary boundary layer and turbulence physics, an improved nest motion algorithm, and, above all, yearly upgrades, systematic testing and evaluation (T&E) that are based not only on single simulations and idealized case studies but on several seasons of testing.

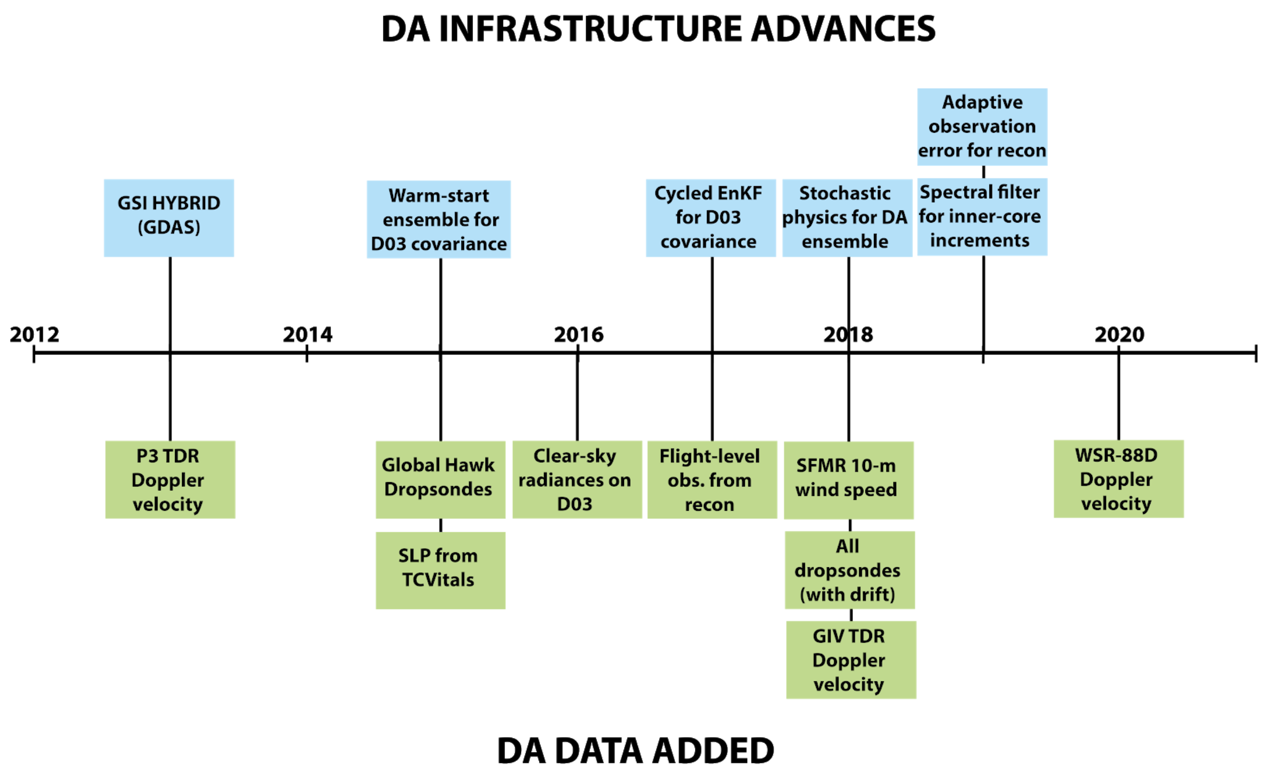

Figure 23: Evolution of inner-core data assimilation techniques under HFIP.

It should be noted that because high-resolution, storm-following nests are central to hurricane NWP, data assimilation (DA) requirements for hurricanes are uniquely different from other weather model applications. Apart from NWP model developments, some significant progress has also been made with inner core DA techniques, which not only demonstrated positive improvements to forecasts (Figure 3) but also will be foundational for next-generation hurricane models, both in terms of developments as well as in building a capacity. Figure 23 shows the progress associated with the developments of multiscale data assimilation techniques under HFIP.

A more advanced version of HWRF, called the Basin-Scale HWRF, an unparalleled capacity for addressing NOAA’s next generation forecasting needs within the unified forecasting system was created under HFIP. The Ocean-Coupled Basin-Scale HWRF, which was run in Stream 2 from 2013-2020, demonstrated how a basin wide domain with multiple-moving nests tracking several storms simultaneously in the North Atlantic and eastern North Pacific basins could improve storm-storm and land-storm interactions without using uniform high-resolution domain, hence providing an operational solution to further advance TC forecasting. Transitions of this multiple-moving-nested HWRF to next-generation global and regional modeling systems within the Unified Forecast System is underway and is expected to further expand the hurricane prediction capacity in NOAA.

These developments and T&E would not be possible without the support of HFIP JET-HPC in Boulder, which was dedicated for Hurricane R2O early in the program. HFIP has also built a capacity of model users, developers and hurricane scientists both within NOAA and academia to tackle the next-generation hurricane forecast improvements. It should be emphasized that nearly all major HWRF developments and R2O efforts, including the first high-resolution version of HWRF, were supported and tested in a real-time demonstration mode (i.e., HFIP Stream 2 or HREx) during the hurricane season and then transitioned to operations. In addition, there have been five Federally Funded Opportunities over the last 10 years for HFIP, awarding 40 grants to University PIs, totaling $10.5M. All these HFIP efforts have led to hundreds of publications related to HWRF within that period[7]. However, it should be noted that as of 2021, we are only about half way to the HFIP goals set over a decade ago.

HFIP's approach is designed to accelerate the implementation of promising technologies and techniques from the research community into operations. That approach has resulted in ~20% improvement of track forecast skill (Figure 3a), and more than 15-30% improvement of intensity forecast skill (Figure 3b) for tropical cyclone forecasts in the North Atlantic basin between 2011 and 2020. Importantly, 2020 HWRF intensity skill scores were 10-30% better than climatology and persistence at all forecast lead times (Figure 3b). Yet, as shown in Figure 3b, these improvements in intensity predictions only resulted in reaching closer to the 5-year-goals in 10 years of time. Part of the reason may be associated with the lack of progress with dynamical guidance until 2012. In fact, until 2011 intensity predictions lagged even the baseline (Figure 3b) primarily set on statistical-dynamical models (SHIPS and LGEM). In addition, predicting RI continues to be a challenge. In terms of track predictions, we have only reached closer to the original HFIP baseline (Figure 3a). It appears that global models with two-way interactive high-resolution nests may be the ultimate solution for both track and intensity predictions (Figures 15 and 16). Moreover, our needs for additional forecast improvements and products have grown since 2009.

Key to HFIP's success are six strategies outlined below:

1. Development of HAFS (DA, Obs, Development)
2. Probabilistic Guidance (goal: is to increase forecast lead time)
3. Improved Risk Communication
4. High-Performance Computing (10-15 million hours per month)
5. Transitions to testbeds, NOAA's transition plan
6. Support to the science community

Supported by the NOAA Hurricane Supplemental projects under the Bipartisan Budget Act of 2018 (P.L.115-123), accelerated developments of HAFS are ongoing. Those developments include high-resolution, telescoping two-way interactive moving nests, model physics to support high-resolution prediction, hurricane inner core data assimilation techniques, regional ensembles and products to support probabilistic forecasts. All developments are being seamlessly merged with the UFS developments (Section 8).

---

[7] http://hfip.org/documents

HFIP Real-time Experiment (HREx; formerly known as Stream 2) is a project undertaken during the hurricane season to demonstrate that the application of advanced science, technology, and increased computing will lead to the desired increase in accuracy, and other improvements in forecast model performance since 2012 as laid out in the HFIP strategic plan. New and innovative Numerical Weather Prediction and data assimilation techniques, model configurations and products must be at least at RL4 or higher to be selected for obtaining HFIP computational resources on the NOAA R&D machines, JET and Orion, following a call for proposal in early April. The HFIP real-time experiments start officially on August 1 and end on October 31. Progress of these real-time runs are evaluated after each season to identify techniques that appear particularly promising to operational forecasters and/or modelers. These potential advances are then blended into operational implementation plans through subsequent model upgrades, or further developed outside of operations with subsequent testing. Starting in the 2019 hurricane season, experimental versions of the UFS-based HAFS were introduced to the suite.

Four configurations of the HAFS model were run as part of the 2021 HFIP Real-time Experiment (HREx). They were (i) the ocean coupled, high resolution regional Limited Area Model (LAM) (HAFS v0.2A); (ii) global model with a high resolution nest (HAFS v0.2B); (iii) regional HAFS with data assimilation (DA) (HAFS v0.2D); and (iv) HAFS ensembles with 21 members (HAFS v0.2E). The results demonstrated that HAFS configurations have more skillful track forecasts than HWRF in the North Atlantic Basin, except for HAFS-E track forecasts which lagged behind HWRF after day-3. The intensity skills are mostly improved in all HAFS configurations yet lagged behind HWRF at the early lead times, but showed some skills after day-2. We believe further improvements may be possible with HAFS (section 9).

2.12. Future direction of HFIP

NOAA recognizes the broad scope of the scientific challenges associated with understanding and predicting hurricanes. Addressing these challenges and improving the forecasts of TC track and intensity will involve significant community interaction and access to the necessary expertise. The success of the next phase of HFIP in reaching the goals requires sufficient funding to support the activities outlined here. NOAA made significant progress toward achieving HFIP goals in the first 5-6 years of the program. Starting in FY 2015, however, NOAA dedicated fewer resources to HFIP due to competing budget priorities across the agency. This slowed the rate of progress towards HFIP goals (e.g. Tropical Cyclone Intensity and RI research) by restricting the capacity to test and evaluate new research and delaying transition of potential new analysis and forecast applications into operations. The lower funding levels also hindered engagement with the academic community that dramatically slowed model improvements.

With the passage of the Weather Act by Congress in 2017, NOAA is now dedicated to reinvigorating HFIP to move towards meeting the requirements of the Act. Resource requirements are still being considered within the agency and will be reflected in NOAA's future year budget requests. The FY18 Appropriations remained constant with the 2015 funding levels and does not address how to support the changes in HFIP priorities directed by the Section 104 of the Weather Act, which requires addressing new strategies, such as risk communication and improving probabilistic guidance. The original HFIP focused on model developments, in particular HWRF and building a capacity to accelerate the model development (HPC upgrades, DTC support for the model developers, EMC & NHC support, and accelerated R2O). The Bipartisan Budget Act of 2018 (P.L.115-123) appropriated funding to improve weather forecasting, hurricane intensity forecasting and flood forecasting and mitigation capabilities to support HAFS developments under HFIP from 2019-2022 and 2022 Disaster Relief Supplemental Act HURR1 project for further advancements in HAFS until 2024. This provided a firm resource for the development of HAFS and the next phase of HFIP, but the challenge remains to ensure sufficient funding is dedicated to reach HFIP goals beyond 2022.

3. Chapter III: HFIP in 2022

3.1. Introduction

This chapter summarizes the activities and results of the Hurricane Forecast Improvement Program (HFIP) that occurred in 2022. The major focus of this report is the development of the Hurricane Analysis and Forecast System (HAFS) within the Unified Forecast System (UFS) and its first operational implementation. As 2022 marks five years since the Weather Act of 2017 established new 5-year goals for HFIP, we will pay particular attention to progress HFIP has made in meeting these goals.

Much recent progress in tropical cyclone forecasting can be attributed to the success of HFIP over the last 15 years. In Section 3.2, we will provide more detailed background on the HFIP program and summarize the success of HFIP since its inception, highlighting the establishment of new goals as previous goals have been met. Section 3.3, we will summarize the performance of the National Hurricane Center (NHC) and available real-time forecast guidance, with particular emphasis on the HWRF and HMON mesoscale models that were developed primarily under the HFIP program. In Section 3.4, we will discuss the development of the state-of-the-science next generation of mesoscale models supported by HFIP: HAFS-A and HAFS-B. Lastly, in Section 3.5, we will discuss future objectives and provide concluding remarks.

3.2. Background and Successes of the HFIP Program

Following the wake of the unfathomable damage from hurricanes in 2004-2005, including Hurricanes Charley (2004), Katrina (2005), Rita (2005), and Wilma (2005), the Hurricane Forecast Improvement Program (HFIP) was established by the NOAA Executive Council on May 10, 2007, outlining a blueprint for the NWS and OAR to collaborate on hurricane forecast improvements. The vision for HFIP in 2007 was to organize the hurricane community to work together to drastically improve numerical forecast models and guidance for the NWS/National Hurricane Center in 5-10 years. HFIP established quantifiable goals, including: (1) reducing track and intensity forecast guidance errors by 20% within 5 years, and 50% within 10 years; (2) extending forecast guidance to 7 days, with skill comparable to that of 5-day forecasts in 2007; (3) increasing the probability of detection for rapid intensification to 90% at day 1, and 60% at day 5; and (4) improving storm surge prediction.

HFIP has been a quantifiable success. Since the inception of HFIP, model hurricane track errors have been reduced by 50%, intensity forecast errors have been reduced by 56%, and intensity errors during rapid intensification (RI) have been reduced by 47%. With the support of HFIP, the Hurricane Weather Research and Forecasting (HWRF) model became the top deterministic intensity guidance used worldwide in tropical cyclone prediction. In response to the Weather Act of 2017, a new set of HFIP goals were established in order to maintain ongoing research to improve hurricane forecasting. This new set of goals included: (1) reducing track and intensity forecast guidance errors further than the 2007 goal, by an additional 50%, including for rapid intensification; (2) improve forecasts and guidance for storm surge and other storm-induced hazards; and (3) incorporate risk communication research to create more effective watch and warning products.

HFIP has been a cross-cutting effort across NOAA. NWS/OSTI leads a collaborative effort to carry out the goals of HFIP, including, but not limited to, the invaluable collaboration between NWS/EMC for transitioning model innovations into operations, NWS/NHC for operational forecasts and products, and OAR/HRD for research and development. More recently, hurricane modeling has begun to look to the future, with a forthcoming transition to the Unified Forecast System (UFS) through development of the Hurricane Analysis and Forecast System (HAFS); expected to become operational in 2023.

Model track forecast errors are closing the gap to meet the original 2007 HFIP 10-year and 2017 Weather Act 5-yr error reduction goals (Figure 24). Further development of the HAFS model is needed to close the gaps between observed track error and the original goals, as well as meet the Weather Act 10-year goal by 2027. The results have been even more impressive for intensity. Model track error has met the original 10-year goal, and even exceeded the Weather Act 5-year goal (Figure 25). The Weather Act 10-year goal for intensity is ambitious, and further development of HAFS will be needed to meet this goal by 2027.

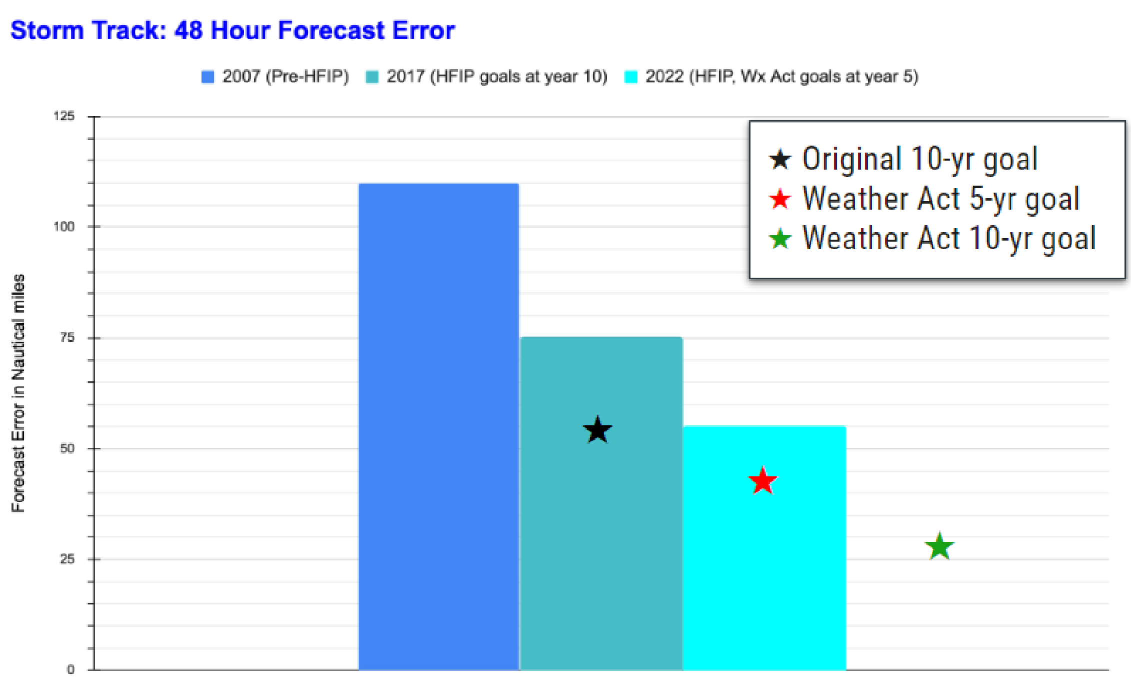

Figure 24: Observed track forecast error (nmi; bar graph) at 48-h lead time, pre-HFIP in 2007, when HFIP goals reached the year 10 mark in 2017, and the Weather Act goals reached year 5 in 2022, compared to the original 10-year goal, the Weather Act 5-year goal, and the Weather Act 10-year goal (black, red, and green stars, respectively).

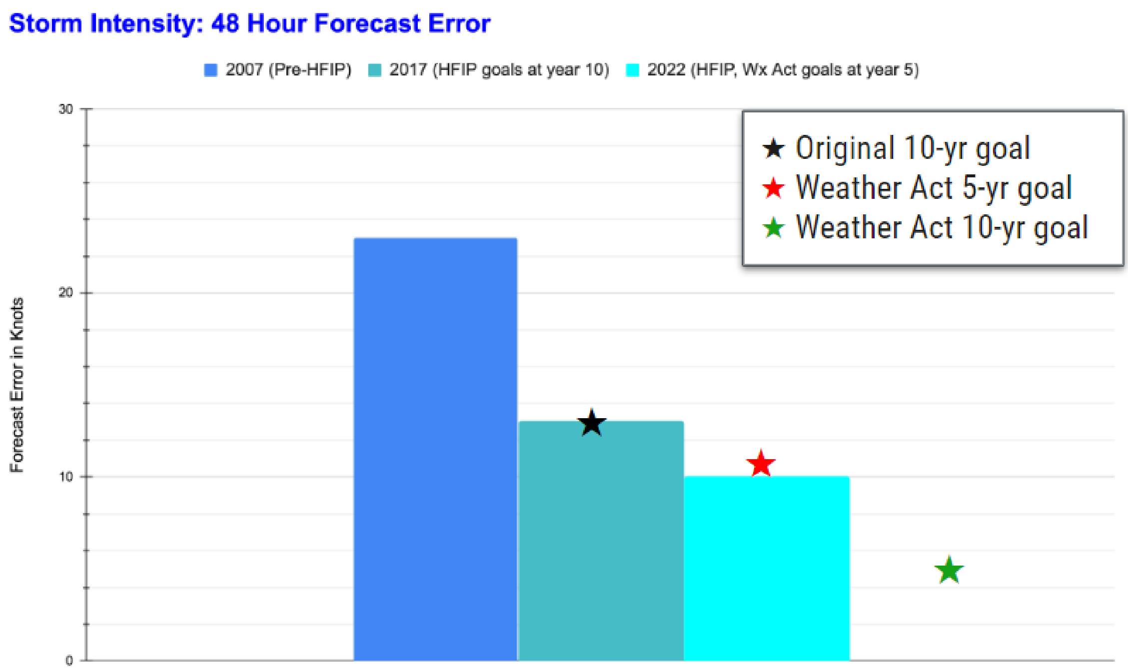

Figure 25: As in Figure 24, but for intensity error (kt).

The original HFIP 2007 goals pertained to the probability of detection of rapid intensification, as opposed to a specific improvement in error when RI occurs. Quantifiable 5-year and 10-year goals in terms of the reduction in intensity forecast error, conditional on RI being observed, were established in the Weather Act or 2017. Model-predicted intensity errors during periods of rapid intensification are currently approaching the Weather Act 5-year goal (Figure 26). As was the case for the HFIP track error objectives, further development of HAFS is needed to close the Weather Act 5-year goal gap and meet the Weather Act 10-year goal by 2027.

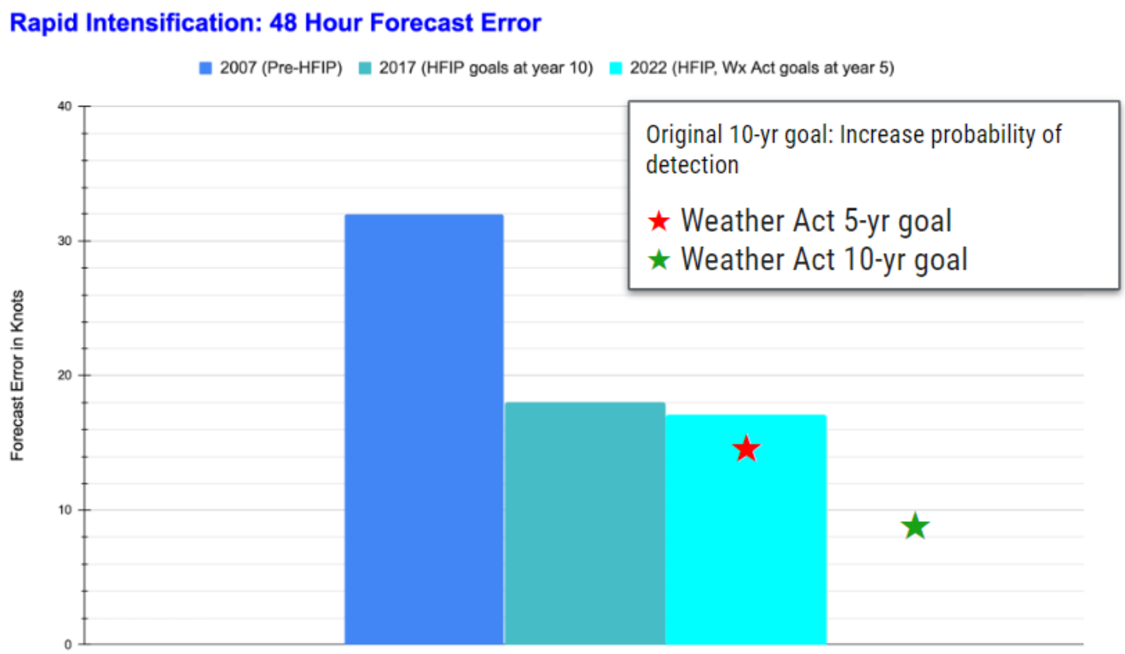

Figure 26: Observed intensity forecast error (kt; bar graph) at 48-h lead time, conditional upon rapid intensity being observed. Bars correspond to pre-HFIP in 2007, when HFIP goals reached the year 10 mark in 2017, and the Weather Act goals reached year 5 in 2022, compared to the Weather Act 5-year goal, and the Weather Act 10-year goal (red and green stars, respectively).

Aligned with the new 2017 Weather Act goals, HFIP is supporting a series of critical intermediary steps, including working towards having real-time (but not yet operational) predictive guidance from a HAFS ensemble by 2023, improved pre-formation disturbance guidance by 2026, and multiple moving nest capability in HAFS for all tropical ocean basins by 2027. Ongoing challenges, such as the recent Hurricane Ian disaster, highlight the need for continuing HFIP. In addition to focusing on the development of the next-generation HAFS probabilistic and ensemble systems, the future of HFIP also seeks to advance the social sciences component of risk communication in hurricane science. Critical advancements towards HFIP strategic goals related to risk communication are being made, including the operational implementation of Tropical Storm Force Winds - Time of Arrival product. HFIP will achieve Social Behavioral and Economic Science (SBES) goals to further improve risk communication through the tropical product suite by integrating research outcomes into new and existing internal and public facing tropical products and services. Recent work on the development of HAFS ensemble looks to address continued challenges in communicating probabilistic information to forecasters, emergency managers, and the public.

3.3. Operational Highlights from the 2022 Hurricane Season

The 2022 Atlantic hurricane season was an average, but destructive, season, with 14 named storms, 8 hurricanes, and 2 major hurricanes. Hurricane Fiona was a category 4 hurricane that caused significant damage to Puerto Rico, the Dominican Republic, and Nova Scotia. Hurricane Ian was a category 5 hurricane at its peak, that slammed into southwest Florida at category 4 intensity, causing widespread damage. The 2022 Eastern Pacific hurricane season was active, with 19 named storms, 10 hurricanes, and 4 major hurricanes. Seven Eastern Pacific tropical cyclones made landfall, including two which crossed over from the Atlantic basin. There were 4 RI events from 3 tropical cyclones in the Atlantic (Danielle, Martin, and Ian), and 8 RI events in the Eastern Pacific (Agatha, Blas, Darby, Estelle, Howard, Kay, Orlene, Roslyn).

Following the theme from 2021, the 2022 season broke records for NHC forecast accuracy. The 2022 season was NHC's best ever for track forecast accuracy from 36-120 h, and for intensity forecast accuracy from 12-60 h in the Atlantic. Higher intensity forecast errors from 72-120 h were primarily due to difficulty in predicting Hurricane Fiona. The Atlantic hurricane season was below average in terms of the number of forecasts issued: 234 forecasts versus the mean of 320. In the Eastern Pacific, no records were set for track forecast accuracy, but records were set for short-term intensity forecast accuracy at 12 and 24 h. The number of forecasts issued in the Eastern Pacific, 325, was very close to the long-term mean.

NHC's long-term (since 1990) statistics show that track forecasts in the Atlantic continue to improve at all forecast lead times, from 24-120 h. There was some concern around 2017-18 that forecast track errors were no longer decreasing meaningfully, but meaningful further decreases in track error since then have demonstrated that there has been opportunity for additional improvement. In 2022, as is often the case, blended consensus aids were the top performing track guidance (Figure 27a) in the Atlantic in 2022. Amongst individual models, the most accurate forecast varied by forecast lead time. GFSI was best short term model through 24 h, the HFIP-supported HMNI was the most accurate model from 48-72 h, and the Navy's CTCI was the top model from 96-120 h. HWFI lagged HMNI again this year for track, but again bested CMCI and NVGI.

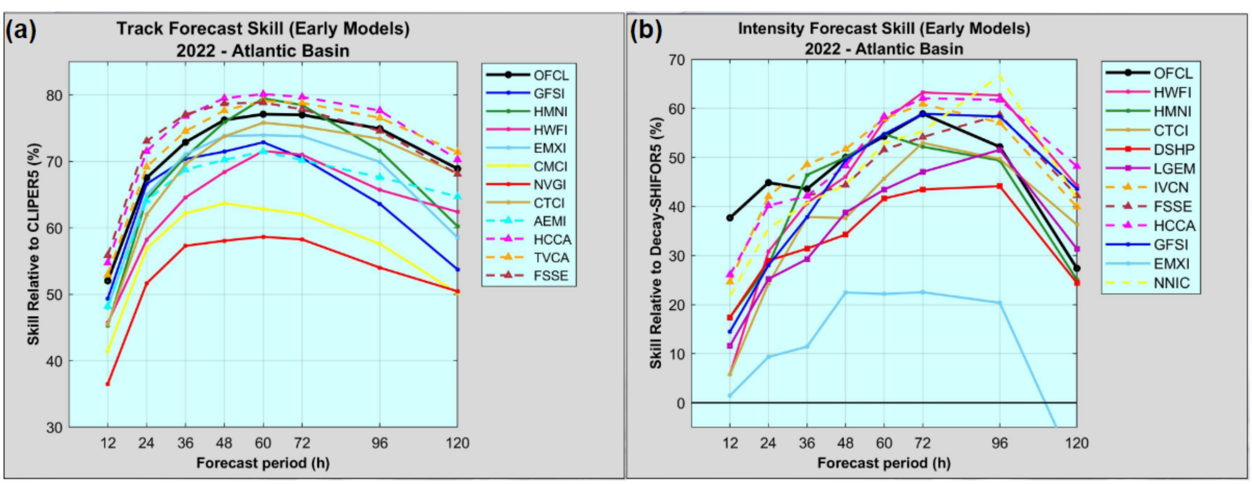

Figure 27: (a) Track and (b) intensity forecast skill (% improvement) of real-time operational forecast guidance relative to a climatological and statistical model baseline (CLIPER5 for track, SHIFOR5 for intensity) for the 2022 Atlantic hurricane season, as a function of forecast lead-time. Numbers immediately above the x-axis indicate the number of cases included at each forecast lead time.

Intensity error is typically associated with a greater season-to-season variance than track error. As such, the long-term trends in NHC official intensity forecast error is associated with much larger upward and downward swings than for track. In 2022, Atlantic intensity forecast errors dropped from 24-72 h with a strong downward long-term trend, but errors spiked up at 96 and 120 h. It is clear that TC intensity at days 4 and 5 remains an operational forecasting challenge. As such, improving the intensity forecast, and also understanding the forecast uncertainty at these timeframes, remains one of the top goals of HFIP. NHC official Atlantic intensity forecast errors in 2022 were below 5-yr means through 72 h, but above means at 96 and 120 h. OCD5 errors show that 72-120 h forecasts were more challenging than average. Amongst real-time intensity guidance, blended consensus aids performed best overall, as per usual, but HWFI and HMNI performed particularly strong in 2022 for intensity (Figure 27b). In fact, HWFI outperformed most blended intensity guidance from 60-120 h, and outperformed all model blends at 72 h, an extremely impressive feat for a deterministic mesoscale model. HMNI also performed very well in 2022, and was the best individual model for short-range intensity forecasts, namely from 24-48 h.

Examination of the long-term gains in track forecast accuracy in the Eastern Pacific are very impressive, with a 67% reduction in 48-h track error since 1990, and a 50% reduction in 120-h track error since 2001. However, unlike in the Atlantic, track error in the Eastern Pacific has not reduced meaningfully at any lead time since 2017. The blended consensus aids were the top guidance in terms of track forecast performance in the Eastern Pacific in 2022, followed by EMXI, the interpolated ECMWF deterministic model (Figure 28a). GFSI, AEMI, and CTCI were the next best models for track, while HMNI and HWFI performed closer to the middle of the pack, similar to their performance in 2021. One hypothesis for the difference in performance of HWRF between the Atlantic and Eastern Pacific basins is the fact that much of the forecast model improvements since 2017 have been made to the data assimilation system, particularly in terms of the assimilation of P-3 tail Doppler radar, flight-level winds, and dropsondes. These in-situ supplemental observations are frequently available in the Atlantic, but rarely available in the Eastern Pacific, due to comparative concerns about landfall, societal impacts, etc. As such, it is possible that improvements to the DA that have allowed for the assimilation of supplemental observations in the Atlantic have not meaningfully benefited the Eastern Pacific.

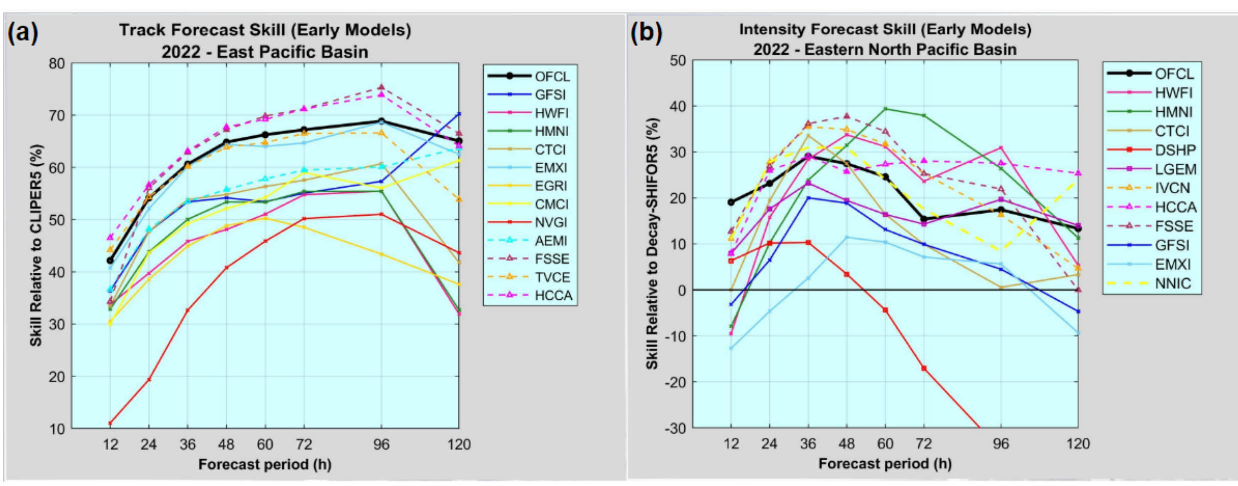

Figure 28: As in Figure 27, but for the Eastern Pacific.

Long-term trends in Eastern Pacific intensity forecast error show slow, but not steady, improvement. Errors decreased at all lead times in 2022 relative to 2021, which was a particularly challenging year amongst recent seasons for intensity forecasts in this basin. The blended consensus aids provided the best intensity guidance overall in the Eastern Pacific in 2022, followed closely by HWFI and HMNI (Figure 28b). In fact, HMNI bested all other models at 60 and 72 h, while HWFI was the top model at 96 h, even outperforming the blended consensus products. As is often the case, global models such as GFSI and EMXI underperformed for intensity.

Since rapid intensification is inherently a low-probability event by definition (a 30-kt intensification in 24-h corresponds to the 95% percentile of intensity change), statistical verification of RI can be somewhat of a challenge in a single basin over a single season due to sample size. As such, RI is often verified over multiple basins and/or multiple seasons. Here we examine the combined Atlantic and East Pacific intensity errors from 2022 as a function of lead time, conditional upon RI either being observed or in the forecast. The resulting multi-basin sample size is 52 cases at 24 hours and 15 cases at 120 hours. As demonstrated in Figure 29a, forecast errors are well below the 2007 "baseline". By further combining 2021 and 2022, the sample increases to 111 cases at 24 hours, and 20 cases at 120 hours (Figure 29b). In the two-year sample, RI forecast errors are again well below the "baseline" from 2007, and match up quite well with the "target", or the 5-year goals from the Weather Act of 2017. Overall, a steady reduction in RI forecast errors continues to be quite promising, and in-line with preestablished HFIP goals.

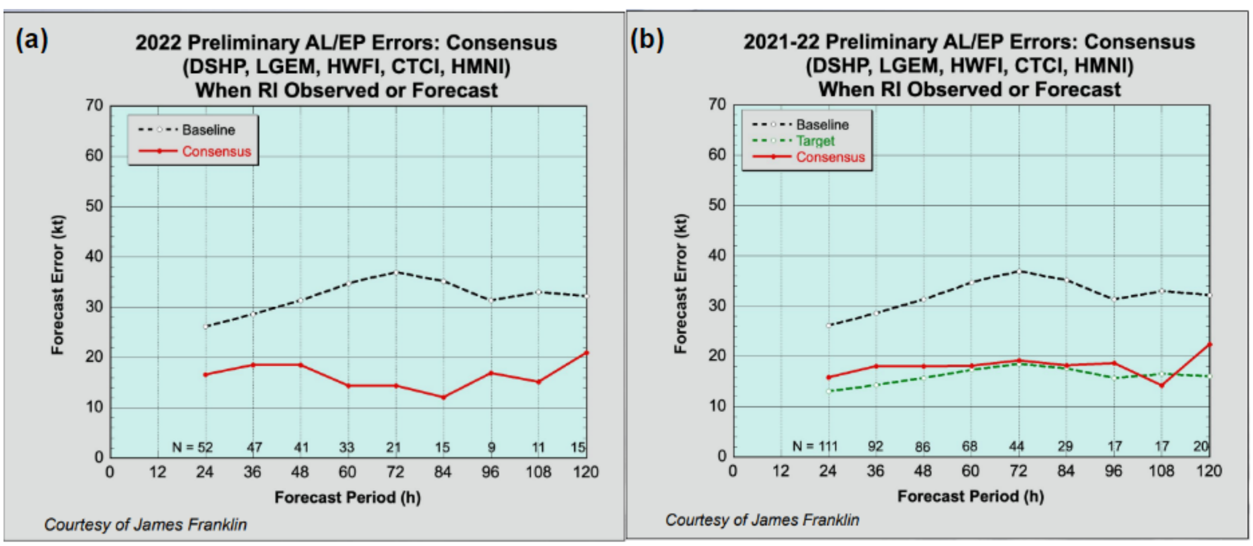

Figure 29: Intensity forecast error (kt) as a function of lead time (h), conditional upon RI (30-kt or more intensification in 24 hours) either being observed or forecast, for the combined Atlantic and East Pacific basins, from (a) 2022, and (b) 2021-2022. Included are the 2007 HFIP baseline (black dashed), the consensus forecast error (red), and the 2017 Weather Act 5-year goal (green dashed).

3.4. Development of the Next Generation of Mesoscale Models: HAFS-A and HAFS-B

a. HAFS Overview

One of the greatest challenges, and recent successes, of the HFIP program has been the development of the next-generation HAFS modeling system, which utilizes the FV3 core streamlining HAFS with the GFS as part of the Unified Modeling System (UFS). The focus of 2022 was to demonstrate HAFS as being ready for real-time operational implementation in the 2023 hurricane season, while also demonstrating skill on par with or exceeding the skill of the operational HWRF model in multi-year retrospective test samples. As demonstrated in 2021, two different variants of HAFS have shown complementary skill in a variety of differing forecast scenarios. These two variants were referred to as HAFS-v0.3A and HAFS-v0.3S during real-time testing and evaluation, and were subsequently renamed HAFS-A and HAFS-B, respectively. A comparison between the HAFS-v0.3A and HAFS-v0.3AS configurations run experimentally in 2022 appears in Table 5. Note that differences in physics options and vortex initialization (VI) provide model diversity, which has been found to be beneficial for helping understand forecast uncertainty. In this section, we will define the various HAFS subvariants, and summarize the performance of HAFS in comparison with the current operational HWRF and HMON models.

Table 5: Comparison between the experimental HAFS-v0.3A and HAFS-v0.3AS configurations run in 2022. Differences between the two configurations are highlighted in red.

HAFS-v0.3A Configuration	HAFS-v0.3S Configuration
Regional storm-centric 6-km parent with a 2-km storm-following moving nest	Regional storm-centric 6-km parent with a 2-km storm-following moving nest
L81 vertical levels with a 2-hPa model top	L81 vertical levels with a 2-hPa model top
Model physics time step of 90s and radiation time step of 900s	Model physics time step of 90s and radiation time step of 1800s
Positive-definition tracer advection scheme	Positive-definition tracer advection scheme
Turn on topography smoothing	Turn on topography smoothing
Use the HAFS CCPP physics suite with GFDL MP, modified PBL	Use the HAFS CCPP physics suite with Thompson MP, modified PBL
VI (VM for all storms) and inner-core DA for model initialization and warm-cycling	VI (VM for hurricane strength only) and inner-core DA for model initialization and warm-cycling
CMEPS-based ocean coupling with an extended HYCOM ocean domain	CMEPS-based ocean coupling with an extended HYCOM ocean domain
Upgraded GFDL vortex tracker	Upgraded GFDL vortex tracker

The HREx experiment in 2022 was a resounding success. HAFS was successful in modeling the dual eyewall structure of Hurricane Ian. Atlantic track forecast error showed a marked improvement in skill with the HAFS system versus HWRF in 2022. That said, there were also challenges to pursue into 2023. Intensity performance demonstrated periods of both improvement and degradation versus HWRF. We continue to investigate periods of intensity forecast degradation to better understand why this occurred, and how to remedy it. From a computing standpoint, issues with the NOAA HPC system "Jet" required running experimental HREx runs on WCOSS, which is typically reserved for operational computing. This occasionally resulted in longer wait times for open cores for the HREx runs. The HFIP team is pursuing a cloud computing option in 2023 in order to evaluate reliability and resource availability versus Jet.

The configurations of HAFS-v0.3A and HAFS-v0.3S that would be run in real-time and tested during the 2022 hurricane season were synced as of 26 May 2022. Both HAFS variants feature: a regional storm-centric 6-km parent with a 2-km storm-following moving nest, L81 vertical levels with a 2-hPa model top, a model physics time step of 90s, a positive-definition tracer advection scheme, topography smoothing, the HAFS CCPP physics suite, inner-core DA for model initialization and warm-cycling, CMEPS-based ocean coupling with an extended HYCOM ocean domain, and an upgraded GFDL vortex tracker. The two variants differ in a number of ways as well. HAFS-v0.3A features a radiation timestep of 900 s, GFDL microphysics with a modified PBL, and vortex initialization for all storm intensities. Alternately, HAFS-v0.3S is configured with a radiation time step of 1800 s, Thompson microphysics with a modified PBL, and vortex initialization for hurricane strength systems only.

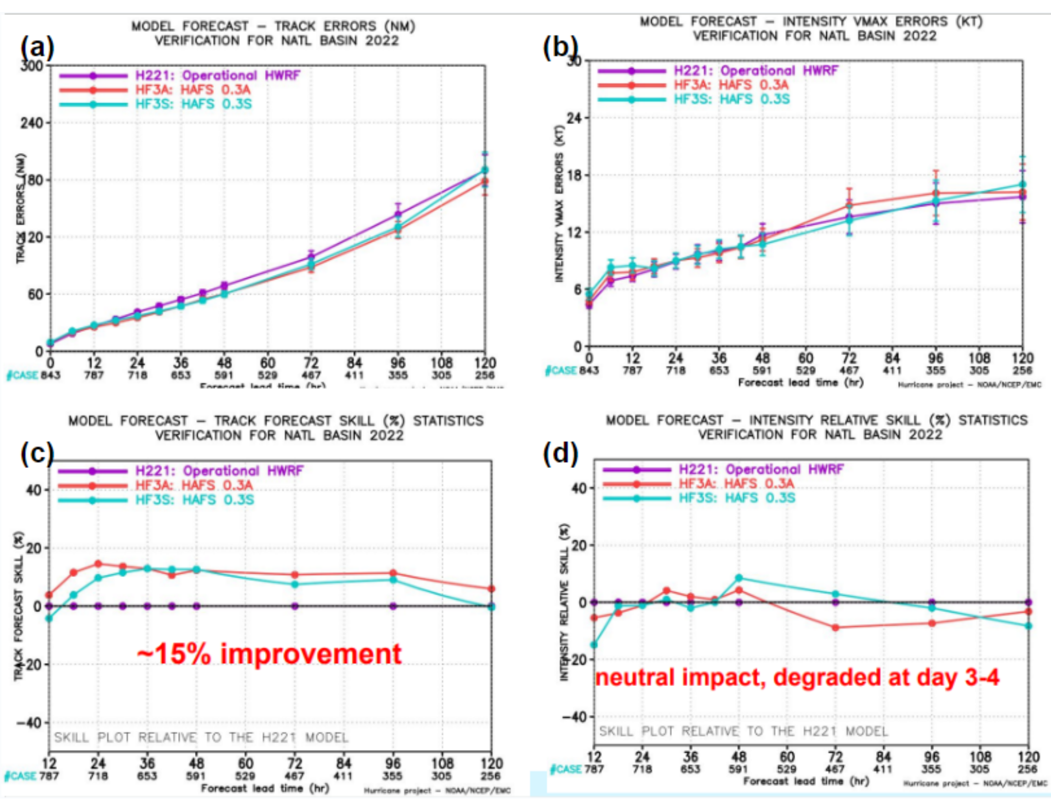

Figure 30: Track (a, c), and intensity (b, d) forecast errors (a, b) and relative skill (c, d) with respect to the operational HWRF (magenta) for HAFSv0.3A (red) and HAFSv0.3S (cyan) for the 2020-2022 retro sample in the Atlantic.

The two configurations of HAFS were evaluated using a three-year retrospective case sample (the 2020-2022 seasons) for the Atlantic and East Pacific basins. In the Atlantic, HAFS-v0.3A and HAFS-v0.3S provided improved track forecasts at all lead times with respect to HWRF, on the order of 5-15% (Figure 30a,c). In the East Pacific, both HAFS variants provided improved track forecasts at all lead times, except for 12 h (Figure 31a,c). For intensity, both HAFS configurations have comparable intensity forecast skills with HWRF from days 0-2, with some degradation of skill from days 3-5 in the Atlantic (Figure 30b,d). For the East Pacific, both HAFS configurations produced similar intensity forecast skills compared to HWRF at most lead times, but with improved skill on the order of 20% at day 5 (Figure 31b,d). EMC and AOML/HRD are currently working on improving DA and model physics schemes specifically to improve days 3-5 intensity forecast skill in the Atlantic.

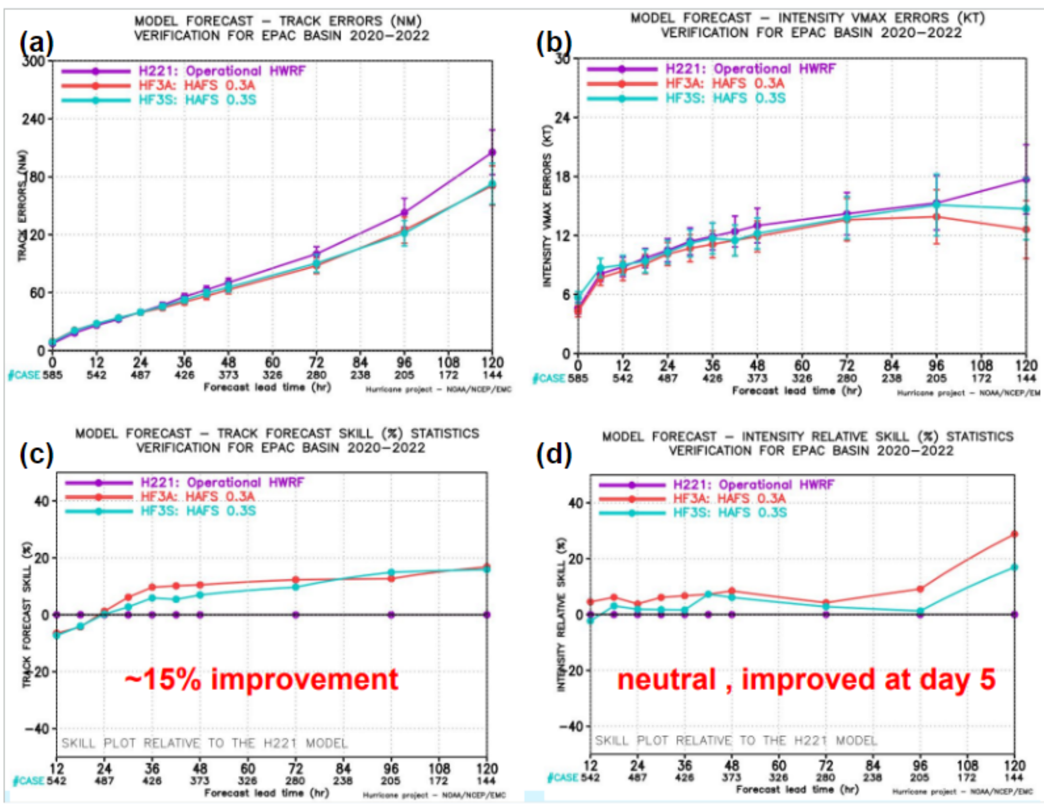

Figure 31: As in Figure 30, but for the East Pacific basin from 2020-2022.

It should be noted that following the real-time experiments in 2022, a software bug was discovered in the momentum flux exchange between the atmosphere and the ocean model. This bug resulted in a slightly degraded intensity forecast in both HAFS configurations. Subsequent re-runs with the bugfix resulted in improved intensity forecast skill compared to the real time experiments.

b. HAFS Experimental Ensemble

Additionally, an experimental HAFS ensemble was run in the Atlantic in 2022 in real time as a part of the HREX experiment. The ensemble configuration is based on HAFSv0.3A, but due to the increased computational cost associated with running multiple members in an ensemble, the model is run with lower horizontal resolution: ~6-km horizontal resolution without nesting, L76 vertical levels, and a smaller domain. GFS Near-Surface Sea Temperature Scheme (NSST) is used to initialize the ocean, but the ensemble runs without ocean coupling. The ensemble is configured with one unperturbed control member plus 11 perturbed ensemble members, running four cycles per day (00Z, 06Z, 12Z and 18Z). Initial condition and boundary condition perturbations are provided from the GEFS at 0.5x0.5 degree resolution, and physics perturbations consist of stochastically perturbed physics tendencies (SPPT), stochastic kinetic energy backscatter (SKEB), and stochastically perturbed PBL humidity (SHUM).

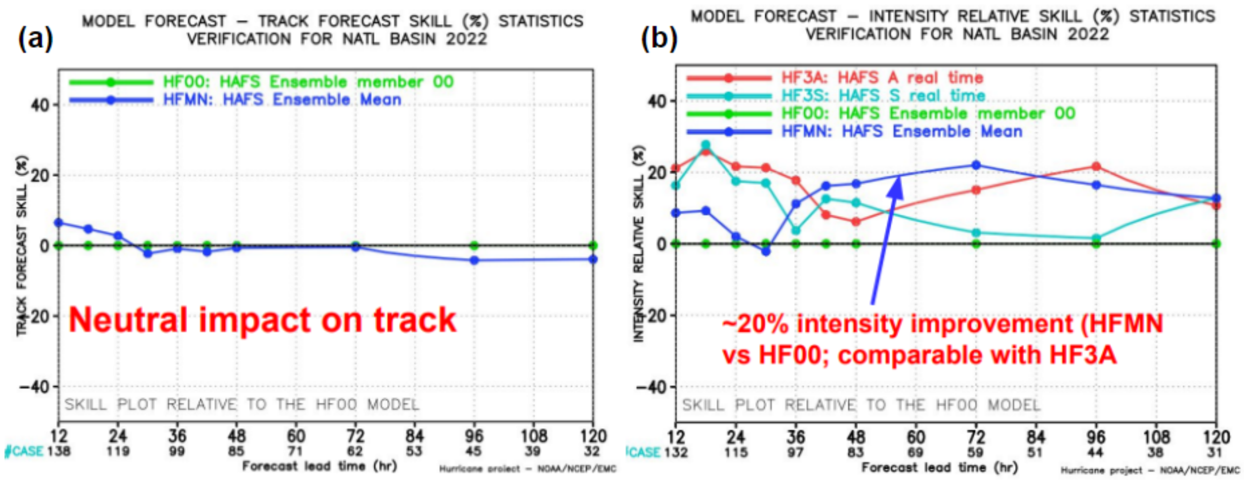

Figure 32: HAFS Experimental Ensemble 2022: (a) track forecast skill of ensemble mean (blue) with respect to unperturbed control member (green); and (b) intensity forecast skill of ensemble mean (blue) compared to HAFSv0.3A (red), HAFSv0.3S (cyan), and the unperturbed control member (green).

Relative to the unperturbed control member, the HAFS ensemble mean forecast produces neutral skill for track (Figure 32a), but significant improvement for intensity (Figure 32b). The ensemble mean intensity forecast is on the order of 20% more skillful than the ensemble control from 48-96 h, and exhibits 5-15% greater skill than HAFS-v0.3A and HAFS-v0.3S for intensity from 42-72 h.

c. Future Work and Preparing for Operational Transition

Additional work is currently underway to prepare EMC for the HAFSv1.0 release, including: merging the HAFSv0.3 code with the latest UFS weather model, implement of ESG grid with dynamic core diffusion tuning, model physics tuning to improve intensity forecasts, vortex initialization threshold optimization, 4DEnVar using GDAS ensemble, enhanced GOES-R AMVs and GOES-18/NOAA-21, NOAH MP Land Surface Model (LSM) with VIIRS Veg Type, Unified Gravity Wave Drag, uGWP, ocean coupling bug fix, ensuring that the code complies with NCO code standard, evaluating the model for stability on multiple platforms, testing and evaluation of JTWC basins, and evaluation of the impact of the new GFSv16.3

input for initial and boundary conditions. All of these steps are underway, and will be addressed prior to the HAFSv1.0 release scheduled for summer 2023.

3.5. Summary and Concluding Remarks

HFIP in 2021 and 2022 has demonstrated the utility of the new HAFS model in real-time hurricane forecasting scenarios. HAFS has demonstrated that it is at least as skillful in terms of predicting track, intensity, rapid intensity change, and structure as the existing top-tier HWRF system, and is occasionally 5-10% better for some storms. The focus for 2023 will be preparation and implementation of HAFS v1 into operations, following the timeline outlined in Figure 33. We will evaluate the potential for retirement of operational HWRF and HMON, while tailoring this decision to NHC needs and feedback. It is likely that there will be an initial period of overlap, in which HAFS will run operationally while HWRF and HMON continue to run, in order to ease the transition for NHC forecasters.

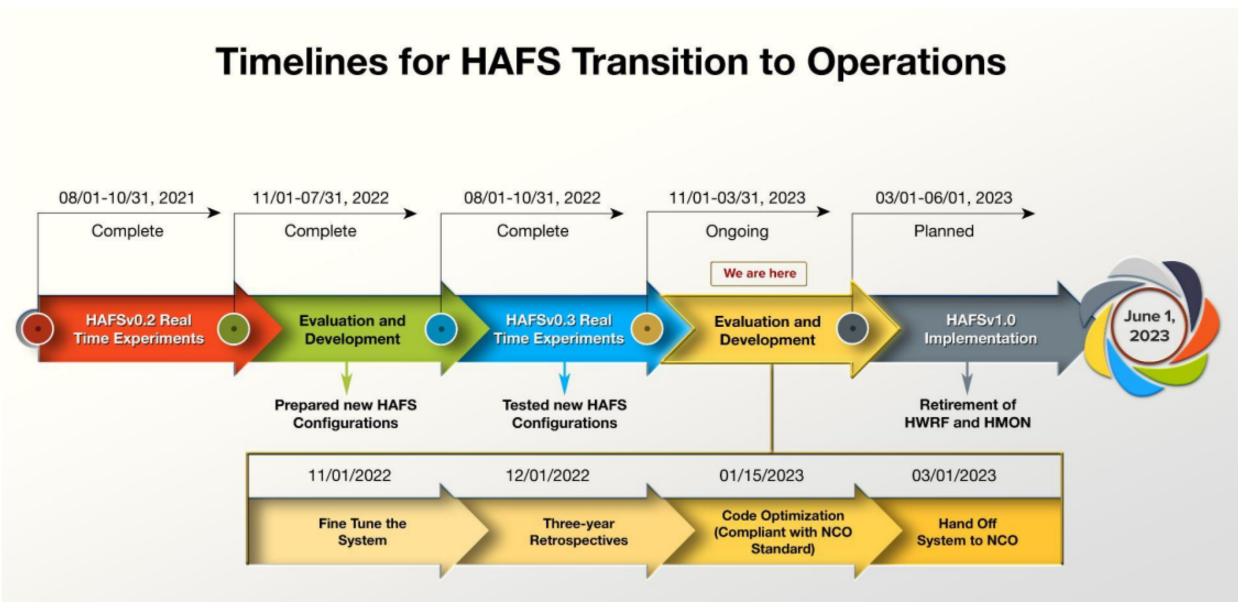

Figure 33: Timeline for testing, evaluation, and operational transition for HAFSv1.0.

In 2023, we will configure and test multiple HAFS pre-V2 configurations for the HFIP real-time experiment for the 2023 hurricane season. HFIP will also work closely with our EMC and HRD partners to develop a 21-member HAFS ensemble system, with 6-km horizontal resolution, encompassing the Atlantic and East Pacific domains. HFIP will continue to support and expand upon our existing Social, Behavioral, and Economic Sciences (SBES) into probabilistic guidance and hazard communications.

4. List of HFIP Publications

Ahern, K., R. E. Hart, and M. A. Bourassa, 2022: Asymmetric Hurricane Boundary Layer Structure during Storm Decay. Part II: Secondary Eyewall Formation. Mon. Wea. Rev., 150, 1915-1936, https://doi.org/10.1175/mwr-d-21-0247.1.

Alaka, G. J., Jr., X. Zhang, and S. G. Gopalakrishnan, 2022: High-Definition Hurricanes: Improving Forecasts with Storm-Following Nests. Bull. Amer. Meteor. Soc., 103, E680-E703, https://doi.org/10.1175/bams-d-20-0134.1.

Alvey, G. R., and A. Hazelton, 2022: How Do Weak, Misaligned Tropical Cyclones Evolve Toward Alignment? A Multi‐Case Study Using the Hurricane Analysis and Forecast System. J. Geophys. Res.: Atmos., 127, https://doi.org/10.1029/2022jd037268.

Chan, M., Chen, X., & Anderson, J. L., 2023: The potential benefits of handling mixture statistics via a bi-Gaussian EnKF: tests with all-sky satellite infrared radiances, Journal of Advances in Modeling Earth Systems, https://doi.org/10.1029/2022MS003357.

Chen, X., 2022: How Do Planetary Boundary Layer Schemes Perform in Hurricane Conditions: A Comparison With Large‐Eddy Simulations. J. Adv. Model. Earth Syst., 14, https://doi.org/10.1029/2022ms003088.

Chen, X., G. H. Bryan, A. Hazelton, F. D. Marks, and P. Fitzpatrick, 2022: Evaluation and Improvement of a TKE-Based Eddy-Diffusivity Mass-Flux (EDMF) Planetary Boundary Layer Scheme in Hurricane Conditions. Wea. Forecasting, 37, 935-951, https://doi.org/10.1175/waf-d-21-0168.1.

Chen, X., Leung, L. R., Feng, Z., & Yang, Q., 2023: Diurnal MCSs precede the genesis of tropical cyclone Mora (2017): the role of convectively forced gravity waves, Journal of the Atmospheric Sciences, 80, 1463–1479, https://doi.org/10.1175/JAS-D-22-0203.1.

Fan, D., Greybush, S. J., Chen, X., Lu, Y., Zhang, F., & Young, G. S., 2022: Exploring the Role of Deep Moist Convection in the Wavenumber Spectra of Atmospheric Kinetic Energy and Brightness Temperature, Journal of the Atmospheric Sciences, 79, 2721-2737, https://doi.org/10.1175/JAS-D-21-0285.1.

Fischer, M. S., P. D. Reasor, B. H. Tang, K. L. Corbosiero, R. D. Torn, and X. Chen, 2022: A Tale of Two Vortex Evolutions: Using a High-Resolution Ensemble to Assess the Impacts of Ventilation on a Tropical Cyclone Rapid Intensification Event. Mon. Wea. Rev., https://doi.org/10.1175/mwr-d-22-0037.1.

Gramer, L. J., J. A. Zhang, G. Alaka, A. Hazelton, and S. Gopalakrishnan, 2022: Coastal Downwelling Intensifies Landfalling Hurricanes. Geophys. Res. Lett., 49, https://doi.org/10.1029/2021GL096630.

Hartman, C. M., Chen, X., & Chan, M., 2023: Improving Tropical Cyclogenesis Forecasts of Hurricane Irma (2017) through the Assimilation of All-Sky Infrared Brightness Temperatures, Monthly Weather Review., https://doi.org/10.1175/MWR-D-22-0196.1.

Hazelton, A., 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.

Hazelton, A., G. J. Alaka, Jr., M. S. Fischer, R. Torn, and S. Gopalakrishnan, 2022: Factors Influencing the Track of Hurricane Dorian (2019) in the West Atlantic: Analysis of a HAFS Ensemble. Mon. Wea. Rev., Early Online Release, https://doi.org/10.1175/MWR-D-22-0112.1.

Hazelton, A., J. A. Zhang, and S. Gopalakrishnan, 2022: Comparison of the Performance of the Observation-Based Hybrid EDMF and EDMF-TKE PBL Schemes in 2020 Tropical Cyclone Forecasts from the Global-nested Hurricane Analysis and Forecast System. Wea. Forecasting, Early Online Release, https://doi.org/10.1175/WAF-D-21-0124.1.

Hazelton, A., K. Gao, M. Bender, L. Cowan, G. J. Alaka, A. Kaltenbaugh, L. Gramer, X. Zhang, L. Harris, T. Marchok, M. Morin, A. Mehra, Z. Zhang, B. Liu, and F. Marks, 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.

Kim, H.-S., J. Meixner, B. Thomas, B. 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.

Leighton, H., R. Black, X. Zhang, and F. D. Marks, 2022: The Relationship Between Reflectivity and Rainfall Rate From Rain Size Distributions Observed in Hurricanes. Geophys. Res. Lett., 49, https://doi.org/10.1029/2022gl099332.

Lim A., S. Nebuda, J. A. Jung, J. Daniels, W. Bresky, L. Bi and A. Mehra, 2022: Optimizing the assimilation of GOES-16 and 17 Atmospheric Motion Vectors in the Hurricane Weather Forecasting (HWRF) model. Remote Sens., 14, 3068, https://doi.org/10.3390/rs14133068.

Wang W., B Liu, Z. Zhang, A. Mehra, V. Tallapragada, 2022, Improving low-level wind simulations of tropical cyclones by a regional HAFS. WMO WGNE blue book 2022, Section 4, p9. Research activities in Earth system modeling. Working Group on Numerical Experimentation. Report No. 52. WCRP Report No.4/2022. Ed. E. Astakhova, July 2022, WMO, Geneva.

Wang, W., L. Zhu, B. Liu, Z. Zhang, A. Mehra, V. Tallapragada, 2023: A Forecast Cycle–Based Evaluation for Tropical Cyclone Rapid Intensification Forecasts by the Operational HWRF Model. Wea. Forecasting, 38, 125-138, DOI: https://doi.org/10.1175/WAF-D-22-0007.1.

Wang, W., and Coauthors, 2023: A review of recent advances (2018–2021) on tropical cyclone intensity change from operational perspectives, part 2: Forecasts by operational centers. Tropical Cyclone Research and Review, 12, 50-63, https://doi.org/10.1016/j.tcrr.2023.05.003.

Wu, S.-N., B. J. Soden, and G. J. Alaka, Jr., 2023: The Influence of Radiation on the Prediction of Tropical Cyclone Intensification in a Forecast Model. Geophys. Res. Lett., e2022GL099442, https://doi.org/10.1029/2022GL099442.

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

Zhao, X., 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, DOI: https://doi.org/10.1175/MWR-D-21-0303.1

5. References

Alaka, G.J. Jr., X. Zhang, and S.G. Gopalakrishnan (2022): High-definition hurricanes: Improving forecasts with storm-following nests. Bulletin of the American Meteorological Society, 103(3):E680-E703, https://doi.org/10.1175/BAMS-D-20-0134.1.

Alaka, G. J., X. Zhang, S. G. Gopalakrishnan, S. B. Goldenberg, and F. D. Marks, (2017): Performance of basin-scale HWRF tropical cyclone track forecasts. *Wea. Forecast*., **32**(3):1253-1271. https://doi.org/10.1175/WAF-D-16-0150.1.

Biswas M. K., S. Abarca, L. Bernardet, I. Ginis, E. Grell, M. Iacono, E. Kalina, B. Liu, Q. Liu, T. Marchok, A. Mehra, K. Newman, J. Sippel, V. Tallapragada, B. Thomas, W. Wang, H. Winterbottom, and Z. Zhang, (2018): Hurricane Weather Research and Forecasting (HWRF) Model: 2018 Scientific Documentation, Available at https://dtcenter.org/HurrWRF/users/docs/index.php.

Bleck, R., (2002): An oceanic general circulation model framed in hybrid isopycnic–Cartesian coordinates. Ocean Modell., 4 , 55–58. https://doi.org/10.1016/S1463-5003(01)00012-9.

Cangialosi, J. P., (2020): National Hurricane Center Forecast Verification Report, 2019 Hurricane Season. NOAA/NWS/National Hurricane Center, Annual Report, April 2020. [Available online https://www.nhc.noaa.gov/verification/pdfs/Verification_2019.pdf].

Cangialosi, J. P., (2021): National hurricane center 2021 Preliminary forecast verification. The 13th HFIP Annual Rev. Conf. presentation, Nov 15, 2021, Virtual. [Available online at https://hfip.org/sites/default/files/events/269/140-cangialosi-nhcverificationpptx.pdf.

Chen, H., and S.G. Gopalakrishnan, (2015): A study on the asymmetric rapid intensification of Hurricane Earl (2010) using the HWRF system. Journal of the Atmospheric Sciences, 72(2):531-550. https://doi.org/10.1175/JAS-D-14-0097.1.

DeMaria, M., J. Kaplan, (1994): A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Weather Forecasting,* **9**(2):209-220, doi:10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2.

DeMaria, M., J. L. Franklin, M. J. Onderlinde, and J. Kaplan, 2021: Operational forecasting of tropical cyclone rapid intensification at the National Hurricane Center. *Atmosphere*, 12(6), 683; https://doi.org/10.3390/atmos12060683.

Dong, J., B. Liu, Z. Zhang, W. Wang, A. Mehra, A. T. Hazelton, H. R. Winterbottom, L. Zhu, K. Wu, C. Zhang, V. Tallapragada, X. Zhang, S. Gopalakrishnan, F. Marks, 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* 2020, *11*, 617. https://doi.org/10.3390/atmos11060617.

Franklin, J. L., 2009, May 5; 2010, April 27: A proposal for HFIP performance baselines. Modified Apr. 27, 2010 to include the Eastern North Pacific baseline. Unpublished manuscript, National Hurricane Center, Miami, FL.

Gao, K., L. Harris, L. Zhou, M. A. Bender, and M. J. Morin, 2021: On the sensitivity of hurricane intensity and structure to horizontal tracer advection schemes in FV3. Journal of the Atmospheric Sciences, 78(9), 3007–3021. https://doi.org/10.1175/JAS-D-20-0331.1.

Goldenberg, S. B., S. Gopalakrishnan, V. Tallapragada, T. Quirino, F. Marks, S. Trahan, X. Zhang, and R. Atlas, (2015): The 2012 triply-nested, high-resolution operational version of the hurricane weather research and forecasting system (HWRF): Track and intensity forecast verifications. Wea. Forecast. https://doi.org/10.1175/WAF-D-14-00098.1

Gopalakrishnan S. G., N. Surgi, R. Tuleya, and Z. Janjic (2006): NCEP's Two-way-Interactive-Moving-Nest NMM-WRF modeling system for Hurricane Forecasting, 27th Conference on Hurricanes and Tropical Meteorology, Session 7A, Tropical Cyclone Prediction I - Model Development, Wednesday, 26 April 2006. (Available online at https://ams.confex.com/ams/pdfpapers/107899.pdf)

Gopalakrishnan, S.G., F. Marks, 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. Monthly Weather Review, 139(6):1762-1784. https://doi.org/10.1175/2010MWR3535.1

Gopalakrishnan, S.G., S. Goldenberg, T. Quirino, F. Marks, X. Zhang, K.-S. Yeh, R. Atlas, and V. Tallapragada (2012): Towards improving high-resolution numerical hurricane forecasting: Influence of model horizontal grid resolution, initialization, and physics. Weather and Forecasting, 27(3):647-666. https://doi.org/10.1175/WAF-D-11-00055.1.

Gopalakrishnan, S.G., F. Marks, J.A. Zhang, X. Zhang, J.-W. Bao, and V. Tallapragada, (2013): A study of the impacts of vertical diffusion on the structure and intensity of tropical cyclones using the high resolution HWRF system. Journal of the Atmospheric Sciences, 70(2):524-541. https://doi.org/10.1175/JAS-D-11-0340.1.

Gopalakrishnan et. al., (2019): 2018 HFIP R&D Activities Summary: Recent Results and Operational Implementation, HFIP Technical Report: HFIP 2019-1. [Available online at http://hfip.org/sites/default/files/documents/hfip-annualreport-fy2018.pdf]

Gopalakrishnan, S., A. Hazelton, and J. A. Zhang, (2021): Improving Hurricane Boundary Layer Parameterization Scheme Based on Observations. *Earth Space Sci.*, **8**, e2020EA001422, https://doi.org/10.1029/2020EA001422.

Harris, L., L. Zhou, S.-J. Lin, J.-H. Chen, X. Chen, K. Gao, M. Morin, S. Rees, Y. Sun, M. Tong, B. Xiang, M. Bender, R. Benson, K.-Y. Cheng, S. Clark, O. Elbert, A. Hazelton, J. J. Huff, Z. Liang, T. Marchok, H. H. Shin, and W. Stern, 2020: GFDL SHiELD: A unified system for weather‐to‐seasonal prediction. Journal of Advances in Modeling Earth Systems, 12 (10) https://doi.org/10.1029/2020MS002223.

Hazelton A. T. and M. Bender, (2018): 2017 Atlantic Hurricane Forecasts from a High-Resolution Version of the GFDL fvGFS Model: Evaluation of Track, Intensity, and Structure. Weather and Forecasting, 33, 1317-1337. https://doi.org/10.1175/WAF-D-18-0056.1.

Hazelton, A., Z. Zhang, B. Liu, J. Dong, G. Alaka, W. Wang, T. Marchok, A. Mehra, S. Gopalakrishnan, X. Zhang, M. Bender, V. Tallapragada, and F. Marks, 2021: 2019 Atlantic Hurricane Forecasts from The Global-Nested Hurricane Analysis and Forecast System: Composite Statistics and Key Events, Weather and Forecasting,36(2), 519-538. https://doi.org/10.1175/WAF-D-20-0044.1.

Hazelton, A. T., Gopalakrishnan, S., and J. A. Zhang, 2022: Comparison of The Hybrid EDMF and Modified EDMF-TKE PBL Schemes in 2020 Tropical Cyclone Forecasts from the Global-nested Hurricane Analysis and Forecast System, Wea. Forecasting, in press.

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). Journal of the Atmospheric Sciences, 75(1):275-295. https://doi.org/10.1175/JAS-D-17-0171.1..

NOAA SAB, 2006: Majority report: National oceanic and atmospheric administration science advisory board hurricane intensity research working group. NOAA SAB HIRWG, 05 July, 2006. [Available at http://www.sab.noaa.gov/Reports/HIRWG_final73.pdf]

Poterjoy, J., L. Wicker, M. Buehner, (2019): Progress toward the Application of a Localized Particle Filter for Numerical Weather Prediction, Monthly Weather Review, 147, 1107-1125. https://doi.org/10.1175/MWR-D-17-0344.1.

Simon, A., A. B. Penny, M. DeMaria, J. L. Franklin, R. J. Pasch, E. N. Rappaport, and D. A. Zelinsky, (2018): A Description of the Real-Time HFIP Corrected Consensus Approach (HCCA) for Tropical Cyclone Track and Intensity Guidance, Weather and Forecasting, 33, 37-57. https://doi.org/10.1175/WAF-D-17-0068.1.

Tallapragada, V., L. Bernardet, M. K. Biswas, S. Gopalkishnan, Y. Kwon, Q. Liu, T. Marchok, D. Sheinin, M. Tong, S. Trahan, R. Tuleya, R. Yablonsky, X. Zhang, (2014): Hurricane Weather Research and Forecasting (HWRF) Model: 2014 Scientific Documentation, HWRF v3.6a, Developmental Testbed Center (DTC). [Avialable online at https://dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.6a_ScientificDoc.pdf]

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.

Zhang, X., 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 verification. Weather and Forecasting, 31(6):2019-2034. https://doi.org/10.1175/WAF-D-16-0087.1.

Appendix A: List of Acronyms

AEMI	GEFS with 6 hour interpolation
AOML	Atlantic Oceanographic and Meteorology Laboratory
AVNI	GFS with 6 hour interpolation
AWIPS	Advanced Weather Interactive Processing System
CCPP	Common Community Physics Package
CLIPER	Climate and Persistence model
CMC	Canadian Meteorological Center model
CMCI	CMC with 6 hour interpolation.
COAMPS	Coupled Ocean/Atmosphere Mesoscale Prediction System-Tropical Cyclone
CONUS	Contiguous United States
CPHC	Central Pacific Hurricane Center
CTCI	COAMPS-TC 6 hour interpolation
CTCX	NRL's Coupled Ocean/Atmosphere Mesoscale Prediction System for Tropical Cyclones (COAMPS-TC) model
DA	Data Assimilation
DTC	Developmental Testbed Center
D-SHIPS	Decay-Statistical Hurricane Intensity Prediction Scheme
DTOPS	Deterministic to Probabilistic Statistical RI Index
ECMWF	European Centre for Medium-range Weather Forecasts model
EDMF	Eddy Diffusivity Mass Flux
EMC	Environmental Modeling Center
EGRI	UKMET with 6 hour interpolation
EM	Equally-weighted Ensemble Mean for models used in MMSE
EMXI	ECMWF with 6 hour interpolation
EnKF	Ensemble Kalman Filter
EFS	Experimental Forecast System
ESRL	Earth System Research Laboratory
FAR	False Alarm Rate
FSSE	Florida State University Super-Ensemble Corrected Consensus
FV3	Finite Volume Cubed-Sphere
GDP	Program and Global Drifter Program
GDAS	Global Data Assimilation System
GEFS	Global Ensemble Forecast System

GFDL	Geophysical Fluid Dynamics Laboratory
GFDI	GFDL with 6 hour interpolation
GFS	Global Forecast System
GFSI	Early GFS with 6 hour interpolation
GHMI	GFDL adjusted using a variable intensity offset with 6 hour interpolation
GIV	NOAA Gulf IV
GSI	Grid-point Statistical Interpolation
HAFS	Hurricane Analysis Forecast System
HCCA	HFIP Corrected Consensus Approach
HDOBS	High Density Observations
HFIP	Hurricane Forecast Improvement Program
HMON	Hurricanes in a Multi-scale Ocean coupled Non-hydrostatic model
HMNI	HMON with 6 hour interpolation
HNMMB	Hurricane Non-hydrostatic Multi-scale Model on B-grid
HPC	High Performance Computing
HRD	Hurricane Research Division
HWHI	Basin-scale HWRF with 6 hour interpolation
HWMI	HWRF Ensemble Mean Forecast Interpolated Ahead 6 hour
HWRF	Hurricane Weather and Research Forecasting
HWFI	HWRF with 6 hour interpolation
HYCOM	HYbrid Coordinate Ocean Model
IOOS	Integrated Ocean Observing System
IVCN	Intensity consensus of at least two of DSHP, LGEM, HWFI, CTCI, HMNI forecasts
JEDI	Joint Effort for Data Assimilation
JTWC	Joint Typhoon Warning Center
LGEM	Logistics Growth Equation Model
MAE	Mean Absolute Error
MMSE	FSU Multi-Model Ensemble
NAM	North American Mesoscale Model
NAVGEM	Center Navy Global Environmental Model
NWS	National Weather Service
NCEP	National Centers for Environmental Prediction
NCO	NCEP Central Operations
NCAR	National Center for Atmospheric Research

NEMS	NOAA Environmental Modeling System
NGGPS	Next Generation Global Prediction System
NGPI	NOGAPS with 6 hour interpolation
NGXI	NOGAPS with 6 hour interpolation
NHC	National Hurricane Center
NMM	Non-hydrostatic Mesoscale Model
NMMB	NMM on the B-grid
NMME	Non-Hydrostatic Mesoscale Model on an E-grid
NOGAPS	Navy Operational Global Atmospheric Prediction System
NNIC	Neural Network Intensity Combination
NVGI	Navy Global Environmental Model 6 hour interpolation
OAR	Oceanic and Atmospheric Research
OFCL	Official National Hurricane Center Forecast
OSEs	Observing system experiments
OSSE	Observing system simulation experiments
POD	Probability of Detection
POM	Princeton Ocean Model
RI	Rapid Intensification
RW	Rapid weakening
SAR	Stand Alone Regional
SFMR	Stepped-Frequency Microwave Radiometer
SIP	Strategic Implementation Plan
SHIFOR	Statistical Hurricane Intensity Forecast
SHIPS	Statistical Hurricane Intensity Prediction System
SPICE	Statistical Prediction of Intensity from a Consensus Ensemble
SPIN-UP	Slang terminology for vortex acceleration and/or initialization
SPIN-DOWN	Slang terminology for vortex deceleration and/or termination
SREF	Short Range Ensemble Forecast
SST	Sea surface temperature
SSS	Sea surface salinity
TAB	Trajectory And Beta (TAB) model for trajectory track using GFS input
TC	Tropical Cyclone
TVCA	Track Variable Consensus of at least two of AVNI, EGRI, EMXI, NGPI, GHMI, HWFI forecasts

TVCE	Variable Consensus of AVNI, EGRI, EMXI, NGPI, GHMI, GFNI, HWFI Model Track Forecasts
TVCI	Variable Consensus of AVNI, EGRI, EMXI, NGPI, GHMI, GFNI, HWFI Model Track Forecasts (6-hour interpolation)
TVCN	Track Variable Consensus
UFS	Unified Forecast System
UKMI	United Kingdom Meteorological Office model with 6 hour interpolation
UW4I	University of Wisconsin's Non-hydrostatic Modeling System (4 km)
UWNI	UW-NMS with 6 hour interpolation (UWNI)
UW-NMS	University of Wisconsin Non-hydrostatic Modeling System
WMO	World Meteorological Organization
WRF	Weather Research & Forecasting
WFO	Weather Forecast Office